

STOCHASTIC MODELS AND THE TEMPORAL DYNAMICS OF ECOLOGICAL COMMUNITIES

Vicente Luis Jiménez Ontiveros

Per citar o enllaçar aquest document:
Para citar o enlazar este documento:
Use this url to cite or link to this publication:
<http://hdl.handle.net/10803/672831>



<http://creativecommons.org/licenses/by-nc-sa/4.0/deed.ca>

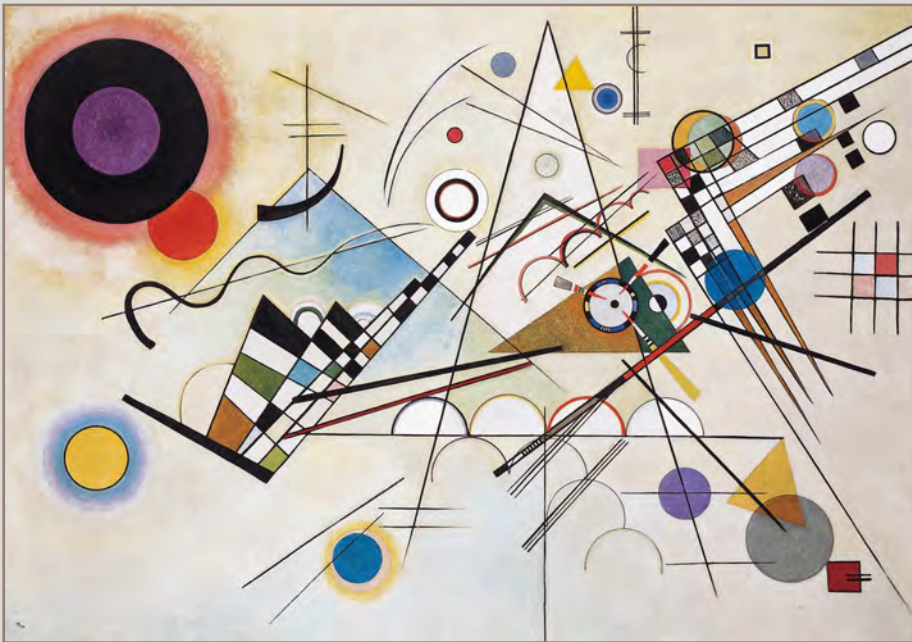
Aquesta obra està subjecta a una llicència Creative Commons Reconeixement-
NoComercial-CompartirIgual

Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial-
CompartirIgual

This work is licensed under a Creative Commons Attribution-NonCommercial-
ShareAlike licence

DOCTORAL THESIS

STOCHASTIC MODELS AND THE TEMPORAL DYNAMICS OF ECOLOGICAL COMMUNITIES



Vicente Luis Jiménez Ontiveros

2020



DOCTORAL THESIS

STOCHASTIC MODELS AND THE TEMPORAL
DYNAMICS OF ECOLOGICAL COMMUNITIES

Vicente Luis Jiménez Ontiveros

2020

DOCTORAL PROGRAMME IN WATER SCIENCE
AND TECHNOLOGY

Supervised by: David Alonso
& Emilio O. Casamayor

Tutor: Anna M. Romani

Presented to obtain the degree of PhD at the
University of Girona

Ontiveros, VJ (2020). Stochastic models and the temporal dynamics of ecological communities. PhD thesis. Universitat de Girona. 278 p.

Cover image:
Wassily Kandinsky. *Composition VIII*. 1923, The Solomon R. Guggenheim Museum, New York.

Dr David Alonso, of the Centro de Estudios Avanzados de Blanes – CSIC, and
Dr Emilio O. Casamayor, of the Centro de Estudios Avanzados de Blanes – CSIC,

WE DECLARE:

That the thesis titled “Stochastic models and the temporal dynamics of ecological communities”, presented by Vicente Luis Jiménez Ontiveros to obtain a doctoral degree, has been completed under our supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, we hereby sign this document.

Signature

David Alonso



Signature

Emilio O. Casamayor

Blanes (Girona), November 25th 2019

A mi padre, Vicente, y a mi madre, María Enriqueta,

Acknowledgements

La andadura de mi tesis comenzó en diciembre de 2013, con una llamada de un número extraño. Decidí cogerlo aunque estaba ocupado en otros menesteres. Se me ofrecía comenzar un doctorado en aproximaciones teóricas a la ecología microbiana, en Blanes. Sin dudarle mucho, acepté y empecé a preparar papeleo. Después de algún retraso y a pesar de lo improbable que era, en verano de 2014 llegué al CEAB. Han sido cinco años que han cambiado por completo mi realidad y por los que tengo que agradecer a muchísima gente.

Empiezo por mi codirector, Emili. Él fue quien hizo esa llamada desde Girona que dudé en responder. Sin él no estaría aquí. Ojalá se me pegue algo más de su pragmatismo y eficiencia. Gracias por la oportunidad. En segundo lugar, mi director de tesis, David. ¡Qué decir de él! Al principio me daba un poco de miedo... nada más alejado de la realidad. Cercano, creativo y más que caótico, estocástico, una constante que me ha llevado hasta este nuevo principio.

También quiero agradecer su presencia y su amistad a mis compañeros de grupo. Ellos fueron mi primer contacto real con lo que conllevaba hacer una tesis. Desde los primeros días resolviendo dudas de R, han estado ahí. Joan, hilando fino, Xevi, el incansable, Tomas, siempre avanzando, Rudi y su creatividad, la armonía de Gerard, y Mateu, su inquietud y su búsqueda de la perfección. Y nuestro compañero de BRIDGES en la sala blanca, Capi, casi un segundo codirector, con sus cuentas y su gracia.

Las reuniones del grupo de ecología teórica y computacional también

fueron fructíferas e inspiradoras. En ellas se ha visto más de una versión preliminar de los capítulos de esta tesis. Gracias Fede, Joan, Roger, Zuzana, Jose Manuel, Meritxell, Dani, Mikel, John, Julia, Pol, Katarina, Jose, Natali, y todos los miembros que poco a poco se incorporan a nuestro grupo.

También hubo ratos para desayunar, merendar, pasarlo bien y hablar de la vida. Gracias a Leire, Magda, Sara, Carla, Celia, Marc, Carlos, Miguel, Xavi, Miquel, Mireia, Aina, Rubén... Y también hubo ratos para jugar y muchas cosas más, gracias Luisfran, Nayeli, Héctor, Ibor, Dani, Victor, Eneko, Nerea, Nuria... Gracias en especial al piso Ventura y sus habitantes, que tantas veces me ha acogido. María también tuvo una especial importancia en que haya llegado a escribir esta tesis, gracias. Jana, por hablar sin reservas. Marta y sus dotes audiovisuales. Sin olvidarme de Aitziber – Almíbar – Cuchiviri, Danilo, Cris, Xavi, Marta, Arnau, Jorge, Mario, Adriá, Dani, Carmen, Federica, Jong Moh, Candela, Ana, Anna S., Anna T., Aurora, Francesco, Chiara, Simone, Gerard, Guillem, Joao, los Jordis, Maria Elena, Mirco, Paula o Sandra. Y al primo Jose, que me encontró en el CEAB, lo que fue una grata sorpresa.

Durante mi tesis tuve la oportunidad de disfrutar cuatro estancias en el extranjero. Con Carlos Melián, aprendí sobre la incertidumbre, escribir y los clásicos. Con Daniel Stouffer, sobre interacciones y algoritmos genéticos. Con Stefano Allesina, matrices y ecuaciones maestras. Y con Mercedes Pascual, epidemiología y procesos estocásticos. Y tuve la oportunidad de conocer y disfrutar de Miguel Leal, Carlos Serván, Matthias Dehling, Rogini, Malyon, Jacopo Grilli, Shai Pilosof, Sergio Alcalá, Victoria, Rahul, Mauricio, Ruby y Dan Maynard.

Quiero agradecer su apoyo y colaboración a Rohan, Teresa, Ferrán, Xavier, Fernando y Hanna. A Helena y Anna Romaní por guiarme en la UdG, y Miquel Sola por sus consejos. Y llevaré a gala que siempre me trataron bien en administración, gracias Carmela, Gemma, Marta, Susanna, Concha, Margarita y Sergio.

A los amigos de siempre, también estaban ahí en la lejanía o un poco más cerca (los que vinieron). Juan que siempre sabe aconsejarme y reconducirme. Jesús, Alba, y nuestras carreras. Manuel, queriendo ir de vacaciones. Guille que siempre me recordará que soy un campeón y que me quiere. Álex, Argent y Tere, Tito y Laura, Ana y Rafa, Juan, Escolano, Ali, Álvaro, Nuria, Fao, Mafo, los gemelos, Laura, Manu...

Debo reconocer a mis fuentes de financiación también. Esta tesis ha sido financiada por un contrato FPI (BES2013066798) del Ministerio de Economía y Competitividad, vinculada al proyecto SITES (CGL201239964) de David Alonso. También he recibido financiación del proyecto BRIDGES (CGL201569043P, D.A., E.O.C., J.A.C.).

Finalmente, gracias por aguantarme y hacerme sonreír, Jenny. A mi padre y a mi madre, que me hicieron quien soy. Y a mis hermanos, Paco y Eloy, con los que crecí, me peleé, jugué y todas esas cosas que hacen los hermanos.

No quiero acabar sin expresar mi grata sorpresa con la obra de Kandinsky, ni a propósito podría ilustrar mejor estas páginas. Acabaré recordando la frase de Newton, a hombros de gigantes. Ya sean MacArthur, Wilson o todas las personas que han hecho posible esta tesis, incluyendo dos revisores anónimos que mejoraron versiones previas. La ciencia es un proyecto colaborativo, y es un honor aportar mi granito de arena.

List of publications

PUBLISHED:

Chapter 3: Ontiveros, VJ, Capitán, JA, Arthur, R, Casamayor, EO, Alonso, D. (2019). Colonization and extinction rates estimated from temporal dynamics of ecological communities: The `island` R package.

Methods in Ecology and Evolution 10:1108–1117.

<https://doi.org/10.1111/2041-210X.13176>

UNDER REVIEW:

Chapter 5: Ontiveros, VJ, Capitán, JA, Casamayor, EO, Alonso, D.

Colonization – persistence trade-offs in natural microbial communities. *Ecology Letters*

SUBMITTED:

Chapter 4: Ontiveros, VJ, Capitán, JA, Casamayor, EO, Alonso, D.

The characteristic time of ecological communities.

IN PREPARATION:

Chapter 6: Ontiveros, VJ, Caliz, J, Triadó-Margarit, X., Alonso, D, Casamayor, EO. General decline in the diversity of the airborne microbiome under future climatic scenarios.

Chapter 7: Ontiveros, VJ, Matthews, B, Melián, CJ, Alonso, D. Individualistic and organismic phytoplankton communities in the reoligotrophication process of a lake.

List of Figures

| | | |
|-----|--|-----|
| 1.1 | The Theory of Island Biogeography | 9 |
| 1.2 | Conceptual Landmarks of Ecology | 11 |
| 2.1 | Example of simulated dynamics, Kadmat atoll | 35 |
| 2.2 | Example of stochastic IBD dynamics | 45 |
| 2.3 | Species Area Relationship | 49 |
| 3.1 | Effect of imperfect detectability on estimated colonization and extinction rates | 61 |
| 3.2 | Simulations of a sagebrush steppe plant community in Idaho, USA. | 63 |
| 3.3 | Network of significant co-occurrences of corallivorous fishes in the Lakshadweep Archipelago. | 65 |
| 4.1 | Temporal dynamics of the arthropod community in the Florida Keys | 81 |
| 4.2 | Richness and community composition patterns recovered by a simple stochastic model of island biogeography | 83 |
| 4.3 | Variation in community composition for two habitats. | 84 |
| 4.4 | Community dynamics across habitats and taxa. | 85 |
| 5.1 | Colonization and extinction rates precisely differentiate dy- namics | 101 |
| 5.2 | The core members of the community follow a log-normal distribution | 105 |

| | | |
|-----|---|-----|
| 5.3 | Microbial communities show a colonization–persistence trade-off | 106 |
| 5.4 | Colonization and persistence distinguish components of microbial communities | 108 |
| 6.1 | Effect of environmental variables associated with climate or origin on colonization and extinction dynamics | 124 |
| 6.2 | Trends in microbial richness and evenness for the period 2021 – 2100 | 125 |
| 6.3 | Seasonality changes for the period 2081 – 2100 | 126 |
| 6.4 | Richness and seasonality predicted for plant and human putative eukaryal pathogens | 128 |
| 7.1 | Environmental gradients in lake Zürich, Switzerland | 144 |
| 7.2 | Abrupt community changes in composition | 145 |
| 7.3 | Observed community dynamics and simulations with environmental dependence | 147 |
| 7.4 | Proportion of cooccurrence that deviate from expectation | 149 |
| 8.1 | A general spatially explicit model | 171 |
| B.1 | Temporal dynamics of two classic communities | 231 |
| B.2 | Evolution of the Jaccard index with time under simulations | 232 |
| B.3 | Relaxation time and characteristic time | 233 |
| B.4 | Properties of c , e , and T_C as estimators | 234 |
| B.5 | Influence of sampling frequency over estimates of characteristic time | 235 |
| B.6 | Autocorrelation is associated with characteristic time | 236 |
| C.1 | Fits for Bacteria and Eukarya | 241 |
| C.2 | Fit for bacterial groups | 242 |
| C.3 | Fit for eukaryal groups | 243 |
| C.4 | Predicted increase in temperatures by the ensemble model in the period 2021 – 2100 | 244 |

| | |
|---|-----|
| C.5 Prediction for bacterial groups | 245 |
| C.6 Prediction for eukaryal groups | 246 |
| C.7 Decadal prediction for seasonal bacterial OTU richness . . | 247 |
| C.8 Decadal prediction for seasonal eukaryal OTU richness . . | 248 |
| C.9 Bacterial groups prediction of seasonal OTU richness in the period 2081 – 2100 | 249 |
| C.10 Eukaryal groups prediction of seasonal OTU richness in the period 2081 – 2100 | 250 |

List of Tables

| | | |
|-----|--|-----|
| 2.1 | Example of colonization and extinction events | 25 |
| 2.2 | Data entry for regular sampling schemes | 27 |
| 2.3 | Data entry for irregular sampling schemes | 27 |
| 2.4 | Data entry for parameter estimation under imperfect de- tectability | 42 |
| 2.5 | Stochastic IBD models included in package 'island' | 44 |
| 3.1 | Vignettes and main functions of R package <code>island</code> | 57 |
| 3.2 | Comparison of seven models for coral reef fishes | 62 |
| 4.1 | Datasets studied in Chapter 4 | 77 |
| 5.1 | Model selection procedure for the dynamics of microbial communities in lakes of the Pyrenees | 102 |
| 5.2 | Colonization and persistence in the three communities stud- ied are related | 103 |
| 5.3 | Core and satellite sub-communities show differential rela- tionships for colonization and persistence | 107 |
| 7.1 | Indicator taxa of summer and winter phytoplankton com- munities | 146 |
| 7.2 | Types of taxa in lake Zürich | 148 |
| C.1 | Model selection for airborne microbial OTUs | 237 |
| C.2 | Models and OTU richness for bacterial groups | 238 |
| C.3 | Models and OTU richness for eukaryal groups | 239 |

| | |
|--|-----|
| C.4 Model selection for airborne microbial pathogens | 240 |
| C.5 Models and OTU richness for two pathogen groups | 240 |

| | |
|------------|--|
| IBD | Immigration, birth, death |
| OTU | Operational taxonomic unit |
| SAR | Species-Area relationship |
| STR | Species-Time relationship |
| TIB | Theory of island biogeography |
| CMT | Classic metapopulation theory |
| NT | Neutral theory |
| SS | Species sorting |
| PD | Patch dynamics |
| ME | Mass effects |
| SDM | Species distribution model |
| ARIMA | Autoregressive integrated moving average |
| ETIB | Equilibrium Theory of Island Biogeography |
| ODE | Ordinary differential equations |
| AIC | Akaike information criterion |
| NLL | Negative Log-likelihood |
| p.g.f. | Probability generating function |
| PDF | Partial differential equations |
| CRAN | Comprehensive R archive network |
| IBT | Island biogeography theory |
| NGS | Next-generation sequencing |
| ILTER | Long-term ecological research |
| RCP | Representative concentration pathway |
| m.a.s.l. | meters above sea level |
| ANC | Acid neutralizing capacity |
| Cond. | Conductivity |
| DIC | Dissolved inorganic carbon |
| DOC | Dissolved organic carbon |
| Hum. | Humidity |
| Irr. | Irradiance |
| Samp. Eff. | Sampling effort |
| Tmax | Maximum temperature |
| Tmed | Mean temperature |
| Tmin | Minimum temperature |
| TN | Total nitrogen |
| TP | Total phosphorus |
| t-SNE | t-distributed stochastic neighbour embedding |
| m.l.e. | Maximum likelihood estimator |

Table of Contents

| | Page |
|---|-----------|
| Summary | 1 |
| 1 Introduction | 7 |
| 1.1 Spatio-temporal distribution of biodiversity | 7 |
| 1.1.1 The niche concept | 12 |
| 1.2 Microbial communities in space and time | 14 |
| 1.2.1 The concept of niche in microbial communities . | 15 |
| 1.2.2 Selection and environmental filter in microbial community assembly | 16 |
| 1.3 The dynamic model of island biogeography | 17 |
| 1.4 Objectives of this thesis | 19 |
| 2 General methodology | 21 |
| 2.1 The Equilibrium Theory of Island Biogeography | 22 |
| 2.2 Package 'island' | 26 |
| 2.2.1 Data entry | 26 |
| 2.2.2 Regular sampling schemes | 28 |
| 2.2.3 Irregular sampling schemes | 30 |
| 2.2.4 Rates Vs. transition probabilities | 32 |
| 2.2.5 Estimating confidence intervals | 32 |
| 2.2.6 A few words on model selection | 33 |
| 2.2.7 Simulating colonization and extinction dynamics | 34 |
| 2.2.8 Environmental fit | 36 |
| 2.2.9 Imperfect detectability | 38 |
| 2.2.10 Model selection grouping | 43 |
| 2.2.11 Immigration, Birth, and Death stochastic models | 43 |
| 2.3 The master equation approach for n species | 45 |
| 3 Colonization and Extinction Rates estimated from Temporal Dynamics of Ecological Communities: The island R Package | 53 |

| | | |
|----------|---|------------|
| 3.1 | Detectability and model selection | 60 |
| 3.2 | Environmental variability and community dynamics . . | 62 |
| 3.3 | Species co-occurrence networks | 64 |
| 3.4 | Discussion | 66 |
| 4 | The characteristic time of ecological communities | 73 |
| 4.1 | Materials and Methods | 76 |
| 4.1.1 | Data samples | 76 |
| 4.1.2 | Community characteristic time | 76 |
| 4.1.3 | Temporal β diversity: characteristic Jaccard index | 79 |
| 4.2 | Results | 81 |
| 4.3 | Discussion | 85 |
| 5 | Colonization – persistence trade-offs in natural mi- | |
| | crobial communities | 93 |
| 5.1 | Materials and Methods | 96 |
| 5.1.1 | Data samples | 96 |
| 5.1.2 | Colonization and extinction rates | 97 |
| 5.1.3 | Core and satellite members of the community . . | 98 |
| 5.2 | Results | 99 |
| 5.2.1 | A colonization - persistence trade-off | 99 |
| 5.2.2 | The species equivalence assumption | 101 |
| 5.2.3 | Relaxing the equivalence assumption | 102 |
| 5.3 | Discussion | 107 |
| 6 | General decline in the diversity of the airborne mi- | |
| | crobiome under future climatic scenarios | 115 |
| 6.1 | Material&Methods | 117 |
| 6.1.1 | Data | 117 |
| 6.1.2 | Estimation of colonization-extinction dynamics . | 118 |
| 6.1.3 | Prediction of microbial responses to climate change | 120 |
| 6.1.4 | Statistical analyses | 122 |
| 6.2 | Results | 122 |

| | | |
|----------|--|------------|
| 6.3 | Discussion | 128 |
| 7 | Individualistic and organismic phytoplankton communities in the reoligotrophication process of a lake | 135 |
| 7.1 | Material & Methods | 139 |
| 7.1.1 | Data | 139 |
| 7.1.2 | Community composition | 140 |
| 7.1.3 | Colonization and extinction dynamics and realized niches | 140 |
| 7.1.4 | Positive feedbacks | 142 |
| 7.2 | Results | 143 |
| 7.3 | Discussion | 149 |
| 8 | General discussion | 153 |
| 8.1 | An effective dynamic model | 154 |
| 8.2 | Dispersal and selection in microbial communities | 155 |
| 8.3 | Estimating niches | 157 |
| 8.4 | Stochastic models of microbial communities | 159 |
| 8.5 | Limitations and implications of the stochastic dynamic model of island biogeography | 162 |
| 8.5.1 | Implications for conservation, management, and restoration | 163 |
| 8.6 | Extensions and perspectives | 165 |
| 8.6.1 | Phenology | 166 |
| 8.6.2 | Trait extensions | 166 |
| 8.6.3 | Dormancy | 167 |
| 8.6.4 | Open spatially realistic metacommunities | 170 |
| 9 | General conclusions | 177 |
| | Bibliography | 179 |
| | APPENDIXES | 212 |

| | | |
|----------|---|------------|
| A | Colonization and Extinction Rates estimated from Temporal Dynamics of Ecological Communities: The island R Package | 213 |
| A.1 | Revisiting Island Biogeography Theory | 213 |
| B | The characteristic time of ecological communities | 219 |
| B.1 | Estimation of colonization and extinction rates | 219 |
| B.2 | Community composition patterns | 224 |
| B.3 | Properties of c , e , and T_c as estimators | 227 |
| B.4 | Sources of error in T_c | 228 |
| B.5 | Supplementary figures | 231 |
| C | General decline in the diversity of the airborne microbiome under future climatic scenarios | 237 |
| C.1 | Supplementary tables | 237 |
| C.2 | Supplementary figures | 241 |

The publication of *The Theory of Island Biogeography* by MacArthur and Wilson (1967) marks a shift in Ecology from a descriptive science towards a more analytical one that tries to find underlying principles and laws. Its main idea is that species richness in an island shows a dynamical equilibrium between the colonization of species from the mainland and the extinction of species in the site. Colonization depends on the distance to the mainland and extinction on the area of the island. The Theory of Island Biogeography was highly successful as a research program, and rapidly extended its domain of application to several types of ecological communities, generating ideas and paradigms that inspire today's Ecology. However, the temporal aspects of the theory remained largely unexplored. Recognition of the intrinsic stochastic character of the main processes of the theory, controlled by colonization and extinction rates, allowed recent theoretical approaches to estimate them from presence-absence temporal studies under the assumptions of species *equivalence* and *independence*. Based on those advances, this thesis aims to develop a quantitative approach to understand and predict the spatio-temporal distribution of biodiversity. It presents the R package 'island' that: *i*) estimates colonization and extinction rates in a variety of temporal settings, *ii*) estimates the influence of the environment over these processes, and *iii*) simulates the dynamics of the associated models. The R package can be applied to any community, although it is applied preferentially to microbial communities in this thesis. Recent methodological advances have overcome some of the inherent difficulties in studying these types of communities, generating broader insights for community ecology. Particularly, this work examines the temporal dynamics of ecological communities, the niche concept in microbial communities, and the influence of selection and environmental filter in microbial community assembly. We have found that a *characteristic time* governed

community temporal dynamics, that indicates the temporal scale of change in biodiversity, helping to design longitudinal studies. Also, we characterized the evolution in time of compositional change through analytical expressions. Concerning the concept of niche, this work ascertains the usefulness of the stochastic approach to study microbial communities, as it obtains parameters coherent with environmental knowledge. In this sense, several natural microbial communities displayed a colonization-persistence trade-off, largely dominated by rare species. Furthermore, we confirm that the importance of environmental filtering for airborne bacterial communities is higher than for airborne eukaryal communities, more influenced by community origin. Thanks to the estimation of airborne microbial niches, we have predicted, by 2080 – 2100, general airborne bacterial declines in richness, idiosyncratic responses for the eukaryal component, changes in seasonality, and declines in putative eukaryotic pathogen richness. Finally, we have devised a four-step approach to distinguish between organismic communities, mostly structured by biotic interactions, and individualistic communities, where adaptation to local environmental conditions predominate. This approach was tested with a long time-series of phytoplankton in lake Zürich, Switzerland, finding that both kinds of communities can be characterized in the process of reoligotrophication of the lake. The value of the stochastic version of the theory of island biogeography reinforces its use as a effective model where the effects of dispersal, selection, and drift are averaged in model parameters. This thesis also emphasizes the importance of long-term research to obtain new insights into community ecology. I hope it will contribute to improving our understanding of species distributions and community dynamics in space and time.

RESUMEN

La publicación de *La Teoría de Biogeografía de Islas* por MacArthur y Wilson (1967) marca un cambio en ecología, de una ciencia descriptiva a una ciencia más analítica que trata de encontrar los principios y leyes subyacentes. Su idea principal es que la riqueza de especies en una isla muestra un equilibrio dinámico entre la colonización de las especies desde el continente y la extinción de las especies en la isla. La colonización depende de la distancia al continente y la extinción del área de la isla. La Teoría de Biogeografía de Islas fue extremadamente exitosa como programa de investigación y rápidamente extendió su dominio de aplicación a diversos tipos de comunidades ecológicas, generando ideas y paradigmas que inspiran la ecología actual. Sin embargo, los aspectos temporales de la teoría permanecieron en gran parte inexplorados. El reconocimiento del carácter intrínsecamente estocástico de los procesos de la teoría, la colonización y la extinción, ha permitido recientemente enfoques teóricos para estimar dichos procesos a partir de estudios temporales de presencia-ausencia, bajo los supuestos de *equivalencia e independencia* entre especies. Profundizando en dichos avances, esta tesis trata de desarrollar una aproximación cuantitativa para entender y predecir la distribución en el tiempo y el espacio de la biodiversidad. Se presenta el paquete de R 'island' que: *i*) estima tasas de colonización y extinción en una variedad de muestreos temporales, *ii*) estima la influencia del ambiente sobre estos procesos, y *iii*) simula la dinámica de los modelos asociados. El paquete de R se puede aplicar a cualquier tipo de comunidad, aunque en esta tesis se aplica preferentemente a las comunidades microbianas. Superando las dificultades inherentes de este tipo de comunidades, se han generado nuevas percepciones de mayor alcance en ecología de comunidades. En particular, este trabajo examina la dinámica temporal de las comunidades ecológicas, el concepto de nicho en las comunidades microbianas y la influencia de la selección y el filtro ambiental en el ensamblaje de las comunidades microbianas. Se ha encontrado que la dinámica temporal de una comu-

nidad se rige por un *tiempo característico*, que indica la escala temporal de cambio en biodiversidad, lo cual puede ayudar a diseñar adecuadamente estudios longitudinales. Además, se proporcionan expresiones analíticas para la evolución en el tiempo del cambio composicional. Con respecto al concepto de nicho, este trabajo determina la utilidad del enfoque estocástico para estudiar comunidades microbianas, ya que obtiene parámetros coherentes con el conocimiento ambiental. En este sentido, se muestra que varias comunidades microbianas naturales muestran un compromiso (*trade-off*) entre colonización y persistencia, en gran parte dominado por especies raras. Además, confirmamos que la importancia del filtro ambiental para las comunidades bacterianas aerotransportados es mayor que para sus análogos eucariotas, más influenciados por el origen de la comunidad. Gracias a la estimación de nichos microbianos en comunidades aerotrasportadas, hemos pronosticado, para 2080 – 2100, la disminución general de la riqueza bacteriana, respuestas idiosincráticas para el componente eucariótico, cambios en la estacionalidad y disminuciones en la riqueza de patógenos putativos eucarióticos. Finalmente, hemos ideado un procedimiento de cuatro pasos para distinguir entre comunidades orgánicas, en su mayoría estructuradas por interacciones bióticas, y comunidades individualistas, donde predomina la adaptación a las condiciones ambientales locales. Este enfoque se probó con una serie temporal de fitoplancton en el lago Zürich, Suiza, y se descubrió que ambos tipos de comunidades se pueden identificar durante el proceso de reoligotrofización del lago. El valor de la versión estocástica de la teoría de la biogeografía de islas refuerza su uso como un modelo efectivo que promedia los efectos de dispersión, selección y deriva, lo que facilita el manejo de dinámicas temporales complejas. Esta tesis también enfatiza la importancia de los estudios a largo plazo para obtener nuevos conocimientos sobre ecología de comunidades y nos deja un paso más cerca para comprender la distribución de las especies en el espacio y el tiempo.

RESUM

La publicació de *La Teoria de Biogeografia d'Illes* per MacArthur i Wilson (1967) marca un canvi en ecologia d'una ciència descriptiva a una ciència més analítica que tracta de trobar els principis i lleis subjacents. La seva idea principal és que la riquesa d'espècies en una illa està en un equilibri dinàmic entre la colonització de les espècies des del continent i l'extinció de les espècies a l'illa. La colonització depèn de la distància al continent i l'extinció de l'àrea de l'illa. La Teoria de Biogeografia d'Illes va ser extremadament reeixida com a programa d'investigació i ràpidament va estendre el seu domini d'aplicació a diversos tipus de comunitats ecològiques, generant idees i paradigmes que transpiren l'ecologia actual. No obstant això, els aspectes temporals de la teoria van romandre en gran part inexplorats. El reconeixement del caràcter intrínsecament estocàstic dels processos principals de la teoria, la colonització i l'extinció, ha permès enfocaments teòrics recents per estimar aquests processos a partir d'estudis temporals de presència-absència sota els supòsits de *equivalència* i *independència* entre espècies. Basada en aquests avenços, aquesta tesi tracta de desenvolupar una aproximació quantitativa per entendre i predir la distribució en el temps i l'espai de la biodiversitat. Es presenta el paquet de R 'island' que: *i*) estima taxes de colonització i extinció en una varietat de mostres temporals, *ii*) estima la influència de l'ambient sobre aquests processos, i *iii*) simula la dinàmica dels models associats. El paquet de R es pot aplicar a qualsevol tipus de comunitat, encara que s'aplica preferentment a les comunitats microbianes en aquesta tesi. Superant les dificultats inherents d'aquest tipus de comunitats, s'han generat noves percepcions de més abast en ecologia de comunitats. En particular, aquest treball examina la dinàmica temporal de les comunitats ecològiques, el concepte de nínxol en les comunitats microbianes i la influència de la selecció i el filtre ambiental en l'acoblament de les comunitats microbianes. Hem trobat que la dinàmica temporal d'una comunitat es regeix per un *temps característic*, que indica l'escala

temporal de canvi en biodiversitat, ajudant a dissenyar adequadament estudis longitudinals, i caracteritzem l'evolució en el temps del canvi composicional a través d'expressions analítiques. Pel que fa el concepte de nínxol, aquest treball determina la utilitat de l'enfocament estocàstic per estudiar comunitats microbianes, ja que obté paràmetres coherents amb el coneixement ambiental. En aquest sentit, es mostra que diverses comunitats microbianes naturals mostren un compromís (*trade-off*) entre colonització i persistència, en gran part dominat per espècies rares. A més, confirmem que la importància del filtre ambiental per a les comunitats bacterianes aerotransportats és més gran que pels seus anàlegs eucariotes, més influenciats per l'origen de la comunitat. Gràcies a l'estimació de nínxols microbians en comunitats aerotrasportadas, hem pronosticat, per 2080 - 2100, disminucions generals en la riquesa bacteriana, respostes idiosincràtiques per al component eucariòtic, canvis en l'estacionalitat i disminucions en la riquesa de patògens putatius eucariòtics. Finalment, hem ideat un procediment de quatre passos per distingir entre comunitats orgàniques, majoritàriament estructurades per interaccions biòtiques, i comunitats individualistes, on predomina l'adaptació a les condicions ambientals locals. Aquest enfocament es va provar amb una llarga sèrie de fitoplàncton al llac Zürich, Suïssa, i es va descobrir que ambdós tipus de comunitats es poden identificar durant el procés de reoligotrofizació de l'estany. El valor de la versió estocàstica de la teoria de la biogeografia d'illes reforça el seu ús com un model efectiu que mitjana els efectes de dispersió, selecció i deriva, el que facilita el maneig de dinàmiques temporals complexes. Aquesta tesi també emfatitza la importància dels estudis a llarg termini per obtenir nous coneixements sobre ecologia de comunitats i ens deixa un pas més a prop per comprendre la distribució de les espècies en l'espai i el temps.

Introduction

This thesis intended to develop a quantitative approach to understand and predict the spatio-temporal distribution of biodiversity. This approach is applicable across ecological communities of any kind. However, the application of these methods to microbial communities, in particular, represents a big challenge due to the inherent characteristics of the microbial world. This is the reason why most examples in this thesis deal with microbial communities. Throughout my work, I emphasize that overcoming these difficulties has also generated new insights of broader scope into community ecology and conservation biology.

In the following sections, first, I introduce basic concepts in community ecology, such as a working definition of an ecological community and well-known patterns in the spatio-temporal distributions of species, together with the concept of ecological niche. After that, I revisit these concepts, especially within the field of microbial ecology, giving particular emphasis to the role of selection (*sensu* Vellend, 2010; 2016) and environmental filtering in community assembly. Then, I present the dynamic model of island biogeography that has been used throughout this thesis and the rationale behind it. Finally, I enumerate the objectives of this thesis formally, and I summarize the structure of the chapters that follow.

1.1 Spatio-temporal distribution of biodiversity

Biodiversity on Earth is organized in ecological communities. But, what is an ecological community? A working definition could read: "An

ecological community is a group of sympatric species that potentially interact" (Stroud et al., 2015). This simple definition hides the whole complexity of spatio-temporal studies of biodiversity within ecological communities. First, if species share the same environment, it is because they have developed, independently or not, suitable adaptations to it. Second, if species coexist within a distinct, long-lasting community, it is because they interact in a way that no major extinctions occur on average. And third, potential species interactions emphasize the spatio-temporal nature of communities and the importance of observational scales. All these aspects cannot be overlooked: species come and go, while the community remains only as a highly dynamic entity. Therefore, adaptation, coexistence, the scale of observation, interaction, and dynamics are keywords both in community ecology and throughout this thesis.

The first studies concerning the spatial and temporal distribution of biodiversity can be traced back to Preston (1948; 1960). Preston recognized that the relation between species and area follows Arrhenius law (1921),

$$N = kA^z, \tag{1.1}$$

where N is the number of species on area A , and k and z are constants recovered from the data at hand. He also realized that the same expression might be applied to time, rendering both space and time exchangeable, in what has been termed 'ergodic conjecture' (Rosenzweig, 2001). We now call these relationships as Species-Area Relationship (SAR) and Species-Time Relationship (STR). The first one has been extensively studied, while STRs did not get much attention until recent times.

SARs inspired one of the most influential theories in Ecology, the Theory of Island Biogeography (TIB). The joint work of Robert H. MacArthur and Edward O. Wilson (1963; 1967) produced the famous depiction of the equilibrium model of faunas in islands of different area and distance from the continent (see Figure 1.1). The equilibrium

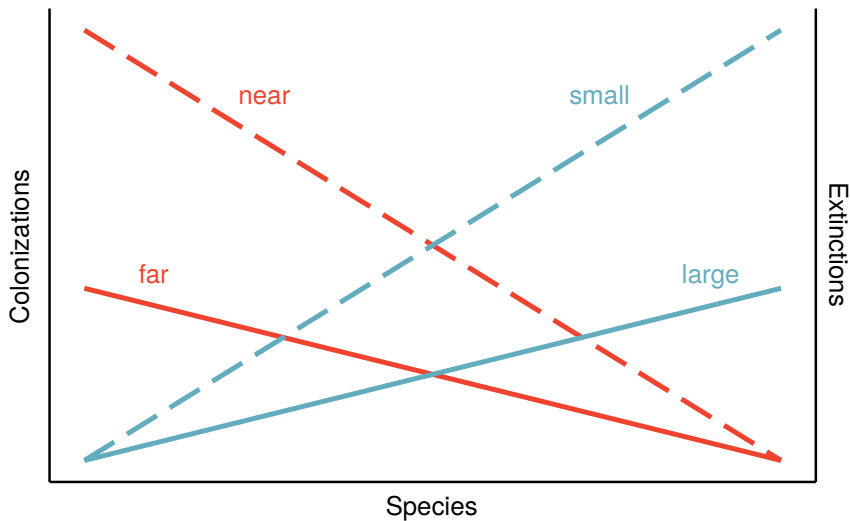


Figure 1.1: **The Theory of Island Biogeography.** The colonization and extinction curves cross in a point, determining a constant number of species in a dynamic equilibrium, as colonization and extinction compensate each other. MacArthur & Wilson hypothesized that colonization depends on the distance to the mainland and extinction on island area. Species colonizations and extinctions are considered independent random processes described by constant rates (therefore the straight lines).

is attained as total external colonization and local species extinction balance out. The equilibrium point can be calculated as two curves, the colonization and extinction curves, intersect, which determine the number of species at equilibrium. The first experimental confirmation did not have to wait (Simberloff, 1969; Simberloff and Wilson, 1969), and more examples of equilibrium in species number ensued (as shown in Losos and Ricklefs 2010). I will deal more extensively with the simple stochastic model of island biogeography in subsection 1.3. Many extensions spawned (e.g., Brown and Kodric-Brown 1977 or Diamond and May 1977), and island biogeography became predominant in Conservation Biology in the 1980s (Hanski and Simberloff, 1997).

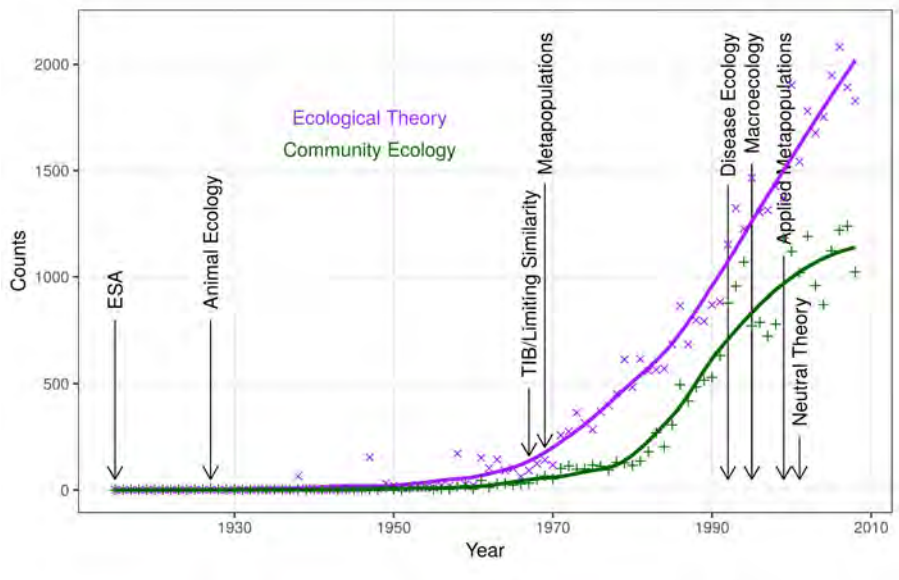
Still in the 1960s, Levins (1969) proposed the classic metapopulation theory (CMT). Although not properly in the domain of community ecology, its subsequent importance in the study of spatio-temporal

dynamics makes it deserve a quick overview. Metapopulations are local patches potentially hosting viable populations connected by dispersal, which allows species to undergo colonization and extinction dynamics. The basic Levins model was formulated as follows:

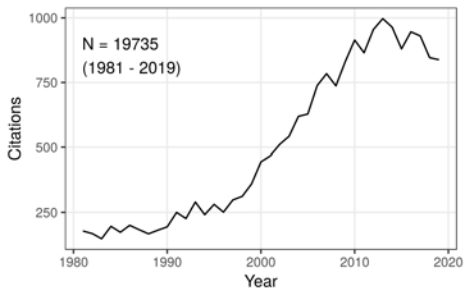
$$\frac{dN}{dt} = mN(1 - N/T) - EN \quad (1.2)$$

where N is the number of occupied patches, m is a migration (colonization) rate per patch, T the total number of patches, and E an extinction rate. This model is a species-specific model that assumes homogeneous patches and constant extinction and colonization rates (per patch). Metapopulation theory gained momentum in the 1980s and 1990s and shifted the paradigm in Conservation Biology (Hanski and Simberloff, 1997), even becoming spatially explicit (Hanski, 1998, 1999, 2001) (see Fig 1.2).

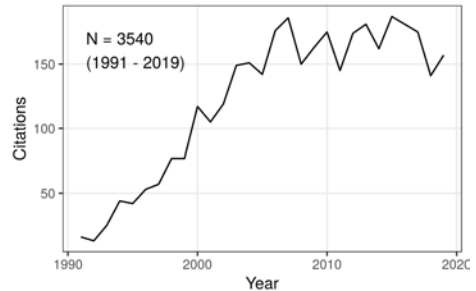
The 2000s saw the advent of two major research programs: neutral theory (NT) and metacommunity ecology. NT, as proposed by Hubbell (2001), was born as an attempt to draw a new general theory of biodiversity in a geographical context, and, as such, it is directly inspired by TIB. It is based on a radical assumption, that *per capita* birth and death rates are equivalent for all species. Despite its controversial assumption, NT has proven useful (Alonso et al., 2006; Chave, 2004; Leigh, 2007) for reasons such as i) its heuristic power as a generator of predictions for well-established patterns for which explanations had remained elusive to ecologists for years; ii) being a precursor of theories that, along with stochasticity, sampling, dispersal limitation or speciation, extend and incorporate more realistic aspects of all these processes; and iii) being capable of acting as a null model. However, the relevance of NT in conservation biology is still limited and unrecognized (Rosindell et al., 2011). Metacommunity ecology (Holyoak et al., 2005; Leibold and Chase, 2017), on his part, draws from metapopulation theory and can be regarded as an attempt to reconcile major paradigms from the past, namely species sorting (SS), patch dynamics (PD), mass



(a) ngrams: Google books



(b) Theory of Island Biogeography



(c) Metapopulation Model

Figure 1.2: During the 20th century several key concepts contributed to the theoretical development of ecology as a science, indicated with arrows. Here I emphasize the importance of two of these landmarks. The upper panel shows the growth of the use of the digrams "Ecological Theory" and "Community Ecology" in millions of digitiled books (Michel et al., 2011). The lower ones, (b) and (c), show the parallel increase in the number of citations per year of MacArthur & Wilson monography (1967), and Levins seminal work on metapopulations (1969), respectively.

effects (ME) and NT. However, while metapopulation theory has found valuable applications to conservation Hanski (1999), this is not the case for metacommunity theory, whose use, in practice, is also still limited (Bengtsson, 2010).

Rio Declaration on Environment and Development in 1992 internationally recognized that one of the major threats humanity currently faces is the loss of biological diversity (Ripple et al., 2017). Therefore, the development of new theory in ecology should provide robust methods and clear-cut guidance to conservation biology. Conservation biology has developed tools as population viability analysis (Morris and Doak, 2002) to assess species extinction risk. It also has benefited from technological advances in genomics, giving rise to fields such as evolutionary conservation genetics (Hoglund, 2009). In addition, with the rise of Geographical Information Systems since the 1990s and 2000s (Guisan and Zimmermann, 2000), conservation biology has also embraced the concept of 'niche', which, in practice, has become widely accepted with the use of Species Distribution Models (SDMs), which are somewhat disconnected from ecological theory (Elith and Leathwick, 2009).

1.1.1 The niche concept

The concept of niche has a long tradition in Ecology. Grinnell (1917) interpreted it as the habitat of the species, Elton (1927) added trophic relations, and Hutchinson expanded the definition greatly, considering it an N-dimensional hypervolume (Hutchinson, 1957). All points that allow indefinite persistence in this hypervolume are called the *fundamental niche*. In contrast, the *realized niche* of a species corresponds to the points in a habitat that coincide with the *fundamental niche* and where other species do not exclude our focal species. Niche theory has further developed to what today is termed as contemporary niche theory (Chase and Leibold, 2003), a term that can be considered parallel to modern coexistence theory (Chesson, 2000; Letten et al., 2017).

SDMs are defined as models that relate species distribution data with information on the environmental, spatial, or both characteristics of those locations (Elith and Leathwick, 2009). SDMs may be considered as depictions of the realized niche (Booth et al., 1988), which produces the increasing recognition of the importance of explicit biotic interactions (Araújo and Luoto, 2007). However, most techniques are species-centered, this is, they rely on species independence (but see Ferrier and Guisan 2006), and ignore the dynamic nature of species populations and ecological communities (as can be seen in Naimi and Araújo 2016 and references therein). One of the few approaches that take into account species dynamics is the one by MacKenzie et al. (2003, 2017). This approach estimates and models species occupancy, and it is inspired by Diamond seminal use of incidence functions (Cody and Diamond, 1975), and Hanski's studies of metapopulations, which also assume that the observed occupancies can be regarded as the stationary states of a simple Markov processes (Hanski, 1994, 1998, 1999). As MacKenzie's approach is fully probabilistic, time is considered discrete, rather than continuous, which imposes certain restrictions on data acquisition and usage.

Changes in species distribution and responses to climate change have been extensively demonstrated (Parmesan, 2006). Thus, ecology needs to become predictive to respond, manage, and palliate the effects of climate and global change (Clark et al., 2001). The prediction of species responses to climate change relies on statistical approaches devoid of a mechanistic basis (Urban et al., 2016). This is so, in part, because it seems very costly to include mechanisms and represent ecological processes in predictive models. At a community level, trait-based approaches seem the way forward to predict responses to and effects of global change (Lavorel and Garnier, 2002; Suding et al., 2008). However, we still lack a null community model able to provide a baseline, as this is of paramount interest for conservation and management (Hillebrand et al., 2018).

1.2 Microbial communities in space and time

Microbial Ecology is very young (Konopka et al., 2015). Strongly influenced by early mottos like Baas-Becking's (1934) "*everything is everywhere, but the environment selects*", progress in the understanding of microbial species distributions across habitats is still very limited. However, microbial species do show biogeographic patterns and dispersal limitation (Bell et al., 2005; Reche et al., 2005; Thompson et al., 2017). Microbial communities can also be considered metacommunities (Nemergut et al., 2013), that is, local communities linked by dispersal. However, the study of microbial metacommunities has relied on statistical approaches, such as correlational, hypothesis-testing, or variance-partitioning frameworks (Soininen, 2012). Dynamical approximations are generally lacking, as metapopulation or TIB dynamical approaches have been seldom used for microbes. Consequently, this has led to a lack of broad principles that potentially explain microbial community dynamics (Konopka et al., 2015). Nevertheless, recent work on microbial community assembly (Faust et al., 2015) has started to consider temporal dynamics.

A study on temporal changes in microbial communities shows that they do change with time, presenting STRs with higher exponents than macroscopic communities (Shade et al., 2013). STRs also show that the turnover of microbial taxa decreases as selective pressures increase (Van Der Gast et al., 2008). Other traditional approaches such as time-series analysis need long time series with short and regular sampling intervals, which are difficult to find in microbial ecology (Faust et al., 2015), although some have looked for interactions and alternative stable states (Gerber, 2014; Gonze et al., 2017; Lahti et al., 2014). Dynamic models have also been used to study microbial dynamics, as ARIMA models have been used to estimate microbial community dynamics, including the effect of environmental variables and interactions (Ridenhour et al., 2017). Additionally, taxa-specific measures of time to recovery after

an antibiotics treatment have been developed (Gerber et al., 2012). Still, the connection with ecological theory of these models is thin. However, generalized Lotka-Volterra models have been recently fitted to microbial community data, although they are challenging to scale up and do not include the influence of immigration (Gonze et al., 2018).

1.2.1 The concept of niche in microbial communities

Traditionally, microbial species have been classified in attention to their morphology, biochemistry, physiology and metabolism. However, the advent of new genotypic and sequencing tools has revolutionized the field discovering vast amounts of organisms that are not possible to cultivate, whose main characteristics remain hidden (Kämpfer and Glaeser, 2013). These unknown characteristics comprise aspects such as nutritional requirements, biochemical abilities, or tolerance to environmental conditions that allow them to perform particular functions in an ecosystem (Schlegel and Jannasch, 2013). However, there exist computational tools that estimate the functional profile of microbial communities based on the known ones (Aßhauer et al., 2015; Langille et al., 2013).

A possible limitation of these approaches in natural environments comes from the existence of core and flexible parts in the genome due to mechanisms as recombination (Fraser et al., 2009, 2007). Microbes form multiple subpopulations in the same habitat (Kashtan et al., 2014; Polz et al., 2013), which challenges the traditional view of canonical species-level niches and favors a view where different genotypes arise due to selection to microhabitat conditions and act as populations with exclusive niches and preferential gene flow (Arevalo et al., 2019; Cordero and Polz, 2014; Niehus et al., 2015). Tools that associate environmental and habitat metadata with specific OTUs (Sinclair et al., 2016) can serve as a broad approximation to their niche and their response or tolerance to environmental variables. A first approximation to model these responses would be to find linear relations with the environmental

factors, although in the case of microbial species other relations might reflect better the actual mechanisms underlying microbial dynamics, for example, Monod's equation in the case of microbial growth (Wade et al., 2016).

1.2.2 Selection and environmental filter in microbial community assembly

The importance of selection (*sensu* Vellend 2016) and environmental filtering in microbial community assembly has been studied traditionally in comparison with other mechanisms in at least three similar frameworks, represented by the dichotomies niche-neutral, SS-NT or deterministic-stochastic. As I find the metacommunity framework more amenable, I will stick to SS-NT. The two main views of this debate are that either NT predominates over SS for the whole communities studied (Lee et al., 2013; Ofiteru et al., 2010; Woodcock et al., 2007), particularly for generalists (Langenheder and Székely, 2011), or that SS predominates over NT (Dumbrell et al., 2010; Hanson et al., 2012; Lindström and Langenheder, 2012; Morrison-Whittle and Goddard, 2015; Nemergut et al., 2013; Stegen et al., 2012; Van der Gucht et al., 2007). However, some studies found acting both at different spatial scales (van der Gast et al., 2011) or components of the community (Caruso et al., 2011; Lindström and Langenheder, 2012). Moreover, a shift from NT to SS in succession has been reported (Dini-Andreote et al., 2015; Ferrenberg et al., 2013; Langenheder and Székely, 2011; Lee et al., 2013; Nemergut et al., 2013), although the role of NT may be more influential in temporal changes (Stegen et al., 2012). Besides, the effect of high dispersal (ME) has been shown to overcome the power of SS experimentally (Lindström and Östman, 2011; Souffreau et al., 2014), although observations do not report the same outcome (Van der Gucht et al., 2007) and a recent simulation of a community of decomposers needed high dispersal to show the effect of SS (Evans et al., 2017). Finally, several studies express concerns such as *i*) an

overrepresentation of SS being it easier to measure than other mechanisms (Lindström and Langenheder, 2012), *ii*) inability of partitioning the effects of dispersal, selection or drift (Evans et al., 2017), but see van der Plas et al. (2015), or *iii*) the inadequacy of naive interpretations of these dichotomies to produce experimental results, which should be backed on mathematical modeling (Pholchan et al., 2013).

One interesting aspect of microbial communities is that they are usually found to be phylogenetically clustered (Goberna et al., 2014a,b; Soininen, 2012). Under the terminology of modern coexistence theory (Chesson, 2000), clustering may be produced by niche differences (environmental filtering) or fitness differences (competitive exclusion of deeply branching clades) (Mayfield and Levine, 2010). Disentangling the effects of these environmental and competitive filters may be achieved using knowledge of the underlying traits, more related to either tolerance or competition (Goberna et al., 2014b; van der Plas et al., 2015), or through null model approaches in environmental gradients (Triadó-Margarit et al., 2019). This knowledge will help to comprehend the mechanisms of the responses of microbial communities to the environment and climate change (Amend et al., 2016).

1.3 The dynamic model of island biogeography

The dynamic model of island biogeography, sometimes called the Equilibrium Theory of Island Biogeography, can be formulated as follows:

$$\frac{dS}{dt} = c(S_P - S) - eS \quad (1.3)$$

where S is the number of species in a site, S_P the number of species in the regional pool and c and e colonization and extinction rates. Alonso *et al.* (2015) found that this equation can be easily solved for a single species (see also Diamond and May 1977), and that the solution allows for treating these dynamics as a Markov chain between two states for the species, present and absent, and to estimate the colonization

and extinction rates from the data, maximizing the likelihood of the observed dynamics regardless its sampling scheme, whether regular or not (Ontiveros et al., 2019). Thus, our approach tries to understand and highlight community dynamics and aims to provide a null model of richness and compositional change, which is of potential interest in the management and conservation of biodiversity (Hillebrand et al., 2018). Our model assumes *species equivalence*, since all species share the same rates, and *species independence*. Still, it has to be emphasized the average character of the model, as it encompasses the mean effects of drift, selection, and dispersal over the species of the community. These two assumptions are reasonable in the case of *horizontal communities*, that is, groups of species that share similar needs of resources and space. However, we can always relax the equivalence assumption and apply the model on a *per* species basis, given a sufficiently thorough sampling. Moreover, colonization and extinction rates can depend explicitly on environmental covariates, offering us a dynamic first approximation to the niche of the considered species. These approximations would make our model close to SDMs, but incorporating, explicitly, temporal dynamics. Using these models, from the extremely simplified average model to the more advanced species-specific, environmentally-driven models, I seek to address the following objectives.

1.4 Objectives of this thesis

- Developing theoretical models to explore, quantify, and evaluate the mechanisms that control spatio-temporal variability, especially in microbes.
- Evaluating the importance of dispersal in microbial community assembly.
- Evaluating the concept of ecological niche in microbial ecology.
- Evaluating the importance of selection and the environmental filter in microbial community assembly.

A description of the structure of my thesis follows. The next chapter, **Chapter 2**, corresponds to the general methodology of this thesis. It delves into the dynamic model of island biogeography and explains the general functioning of the tool that implements it, package 'island'.

Chapter 3 corresponds to the published application paper of package 'island', entitled 'Colonization and extinction rates estimated from temporal dynamics of ecological communities: The island R package' (Ontiveros et al., 2019), and presents the main uses of the package.

Chapter 4, 'The characteristic time of ecological communities', explores the use of the basic community model to describe the temporal dynamics of ecological communities and generate predictions about their STRs, turnover, and changes in community composition, also identifying the underlying temporal scale at which these communities change.

Chapter 5, 'Colonization – persistence trade-offs in natural microbial communities', reports and investigates a colonization – persistence trade-off in microbial communities largely driven by rare taxa, highlighting the importance of equalizing mechanisms for microbial coexistence.

Chapter 6, 'General decline in the diversity of the airborne microbiome under future climatic scenarios', examines the usefulness of the

model to discover the contrasting responses of airborne Bacteria and Eukarya to predicted climate change.

Chapter 7, 'Individualistic and organismic communities in the reoligotrophication process of a lake', expands the methods of the previous chapter to include non-linear responses to the environment, which leads to propose a general approach to determine the existence of both individualistic and organismic communities, which is then applied to the plankton of Lake Zürich.

Chapter 8 corresponds to the general discussion and future perspectives from this work, with an emphasis on further developments of the model, and **Chapter 9** corresponds to the general conclusions of this thesis. Finally, after the bibliography, I conclude with three appendixes that correspond to Chapters 3, 4, and 6.

General methodology

Island Biogeography is one of the most influential and successful theories in ecology, a rich source of inspiration for new research and ideas since it was first suggested. MacArthur and Wilson originally proposed that the number of species on an island is determined by the size of the island and its distance from the mainland. These factors result in a dynamic equilibrium between colonization from the mainland and extinction of species once they arrive on the island. The theory initially presented by MacArthur and Wilson was called Equilibrium Theory of Island Biogeography (ETIB, Simberloff 1974). Wilson and Simberloff's (Simberloff, 1969; Simberloff and Wilson, 1969; Wilson and Simberloff, 1969) study of the experimental defaunation of mangrove islands in the Florida Keys was the first validation of the theory, which explicitly emphasizes its dynamic aspects, and it showed how different mangrove islands reached an equilibrium with a species richness equivalent to the number of species present before the experimental defaunation.

The following sections are devoted to explaining the general methodology of this thesis, that is, the dynamic stochastic model of island biogeography. First, I revisit the basic ETIB and how the use of a stochastic approach for a single species enables us to estimate its basic parameters. Then, I introduce the R package 'island', which implements the previous approach and expands it in several directions. Finally, I explain a novel way to apply the stochastic approach to multiple species following ETIB dynamics.

2.1 The Equilibrium Theory of Island Biogeography

The seminal series of Simberloff & Wilson (1969; 1969; 1969) on the experimental defaunation of mangrove islands in the Florida Keys was the first experimental validation of the ETIB, which predicted that colonization and extinction on an island would balance out and reach a dynamic equilibrium. The basic model underlying the theory can be expressed as in Eq. 1.3:

$$\frac{dS}{dt} = c(S_P - S) - eS$$

where S is the number of species present at a site, S_P is the number of species in the regional pool, and c and e are colonization and extinction rates, respectively. This model states that the temporal variation in the number of species in a site over time is equal to the total rate of arrival of species from the pool that are not already at the site minus the total rate of species loss from the site. Since all species are equivalent in their rates, S/S_P can be interpreted as the probability of presence of any species on the island. Therefore, by dividing the last equation by S_P , we obtain an equation for the probability of any single species being present in the island, which can be easily solved (Alonso et al., 2015).

Alonso and colleagues used what has been called a *master equation* (van Kampen, 1992). The formalism known as master equation has been used in the last decades to build a strong theoretical foundation for population and community ecology. It has several advantages as it *i*) represents an extension of classical deterministic approaches, *ii*) emphasizes the natural continuous flow of time, *iii*) recognizes the essential discreteness of ecological events such as colonization or extinction, and *iv*) models based on this approach are usually amenable to mathematical analysis. A master equation is essentially a set of ODEs expressing the temporal evolution of the probabilities of having the system at each configurational state.

A general master equation, for instance counting the number of species, n , in a site, can be written as a system of ODEs (which is called a continuous-time birth-death process in a mathematical context):

$$\frac{dP(n, t)}{dt} = g_{n-1}P(n-1, t) + r_{n+1}P(n+1, t) - (g_n + r_n)P(n, t) \quad (2.1)$$

where g_n is the probability rate at which n increases in one species and r_n the probability rate at which n decreases in one species. Notice that there is an equation for any possible configuration in the system. Let us assume we only have one species in the pool, $S_P = 1$, then the configuration of the system is determined by only a Boolean variable (presence/absence). The system in Eqs (2.1) then collapses into a system of only two equations ($g_0 = c$, $g_1 = 0$, $r_1 = e$, and $r_0 = 0$):

$$\frac{dP(1, t)}{dt} = cP(0, t) - eP(1, t) \quad (2.2)$$

$$\frac{dP(0, t)}{dt} = eP(1, t) - cP(0, t) \quad (2.3)$$

Since $P(1, t) + P(0, t)$ should be 1 for any time, we can write:

$$\frac{dP(1, t)}{dt} = c(1 - P(1, t)) - eP(1, t) \quad (2.4)$$

The full solution of Eq. 2.4 can be easily calculated. It depends on an initial condition, C , as usual, that corresponds with p_0 , that is, the probability of being present at $t = 0$.

$$P(1, t) = \frac{c}{c+e}C \exp(-(c+e)t); \quad P(1, 0) = p_0$$

$$\text{if } p_0 = 0 \text{ then } P(1, t) = \frac{c}{c+e}(1 - \exp(-(c+e)t)) \quad (2.5)$$

Note that we have obtained an expression for the transition between the two states that can have a species in a community, absent and present, given that the species was absent at the initial time. Alonso denotes this as T_{10} . It is known that $T_{10} + T_{00} = 1$. So just repeating the process above for $P(0, t)$ and $p_0 = 1$, we find T_{01} , the transition probability of going from present to absent. Again, $T_{01} + T_{00} = 1$.

So, with these probabilities, we can construct a matrix of transition probabilities for the associated Markov model:

$$\begin{cases} T_{00} &= 1 - \frac{c}{e+c}(1 - \exp(-(e+c)\Delta t)) \\ T_{10} &= \frac{c}{e+c}(1 - \exp(-(e+c)\Delta t)) \\ T_{01} &= \frac{e}{e+c}(1 - \exp(-(e+c)\Delta t)) \\ T_{11} &= 1 - \frac{e}{e+c}(1 - \exp(-(e+c)\Delta t)) \end{cases} \quad (2.6)$$

where Δt is the time-step between consecutive samplings, T_{01} is the probability of extinction, T_{10} the probability of colonization, T_{11} of repeated presence, and T_{00} of repeated absence.

Two key assumptions, the assumption of *species equivalence* (the same parameters apply for a group of species), and the assumption of *species independence* (the presence of other species does not affect the dynamics of a focal species), enables us to generalize this Markov process to groups of species or even communities. In particular, we can apply the formalism to recover Eq. 1.3. Since all species are equivalent and independent, the average of species present in the island at any given time would be $S_P P(1, t)$, which is the continuous variable S in Eq 1.3. This equation is just recovered by multiplying Eq. (2.4) times the number of species in the pool, S_P .

Importantly, by using the expressions given in Eqs (2.6), we can accurately calculate the likelihood of a given data set under the colonization-extinction model with the following expression for a regular sampling scheme, that is, when samplings are repeated at equivalent time-steps, given specific colonization and extinction rates, c and e :

$$P(M|c, e) = (1 - T_{10})^{N_{00}} T_{10}^{N_{10}} T_{01}^{N_{01}} (1 - T_{01})^{N_{11}} \quad (2.7)$$

where N_{00} is the number of events of repeated absence, N_{10} events of colonization, N_{01} events of extinction, and N_{11} events of repeated presence. In the case of an irregular sampling scheme, we need to calculate the transition probabilities and the matching likelihood for each transition.

As an example, imagine that we have sampled arthropods on an island for three seasons and we have the following data:

Table 2.1: **Example of colonization and extinction events.** Ten species (A-J) were sampled in seasons 1-3, and found to be absent (0) or present (1).

| Sp. | 1 | 2 | 3 |
|-----|---|---|---|
| A | 0 | 0 | 1 |
| B | 0 | 1 | 1 |
| C | 1 | 1 | 1 |
| D | 0 | 0 | 0 |
| E | 0 | 1 | 0 |
| F | 0 | 0 | 0 |
| G | 1 | 0 | 0 |
| H | 0 | 0 | 0 |
| I | 1 | 1 | 0 |
| J | 0 | 0 | 0 |

Table 2.1 shows that between samples 1 and 2 we have $N_{01} = 1$ (G), $N_{11} = 2$ (C, I), $N_{00} = 5$ (A, D, F, H, J), and $N_{10} = 2$ (B, E); we can similarly calculate events between samples 2 and 3. Assuming we already know the true value of transition probabilities ($T_{01} = 0.4$, $T_{11} = 0.6$, $T_{00} = 0.7$, and $T_{10} = 0.3$), we can easily calculate the likelihood of the observed dataframe using Eq. 2.7:

$$0.4^3 \cdot 0.6^4 \cdot 0.7^{10} \cdot 0.3^3 = 6.326 \cdot 10^{-6}$$

This is a small likelihood, and such low numbers are susceptible to numerical errors. So, I will use log-likelihoods from now on.

$$\log(0.4) \cdot 3 + \log(0.6) \cdot 4 + \log(0.7) \cdot 10 + \log(0.3) \cdot 3 = -11.971$$

This is the log-likelihood associated with the given set of transition probabilities. Notice that we obtain the same value with both approaches.

2.2 Package 'island'

R package 'island' follows the stochastic implementation of Simberloff's model (Simberloff, 1969) developed by Alonso et al. (2015), which uses a likelihood approach to estimate colonization and extinction rates for communities that have been repeatedly sampled through time. The package allows:

1. Estimating colonization and extinction rates with regular and irregular sampling schemes, for whole communities or groups within those communities, taking into account or not imperfect detectability.
2. Converting those rates into transition probabilities.
3. Performing Akaike Information Criterion (AIC) -based model selection.
4. Estimating the influence of environmental variables on colonization and extinction dynamics.
5. Simulating the dynamics of the Equilibrium Theory of Island Biogeography, as well as three other population models.
6. Evaluating model error and R^2 .

In the following paragraphs, I will address the practicalities of the package. A list of the main functions of the package can be found in Table 3.1.

2.2.1 Data entry

In order to estimate colonization and extinction rates with package 'island', we need a dataframe of repeatedly sampled communities, organized in time and with at least one species (or the relevant taxonomic unit for our analysis). In this package, we adopt the convention of indicating sampling times in columns and species in rows of the

dataframe. Table 2.2 shows an example that uses a regular sampling scheme for 4 years (2000 to 2003).

Table 2.2: **Dataframe example of regular sampling schemes.** All the species shown pertain to genus *Acanthurus*, present in Kadmat atoll (dataset `alonso15`). T.L., trophic level.

| Species | T.L. | Kd.2000 | Kd.2001 | Kd.2002 | Kd.2003 | Kd.2010 | Kd.2011 | Guild |
|------------------------|------|---------|---------|---------|---------|---------|---------|--------------|
| <i>A. auranticavus</i> | NA | 1 | 1 | 1 | 1 | 1 | 1 | Algal Feeder |
| <i>A. leucosternon</i> | 2 | 1 | 1 | 1 | 1 | 1 | 1 | Algal Feeder |
| <i>A. lineatus</i> | 2 | 1 | 1 | 1 | 1 | 1 | 1 | Algal Feeder |
| <i>A. nigrofuscus</i> | 2 | 1 | 1 | 1 | 1 | 1 | 1 | Algal Feeder |
| <i>A. triostegus</i> | 2,78 | 1 | 1 | 1 | 1 | 1 | 1 | Algal Feeder |
| <i>A. xanthopterus</i> | 2,87 | 0 | 0 | 0 | 1 | 1 | 1 | Algal Feeder |

In the case of irregular sampling schemes, we adopt a convention of recording subsequent time intervals in the column header. For example, if our sampling started on day 17, and the next sample was taken 20 days after, that is, on day 37, the column heads for those samples should read 17 and 37. These conventions have been followed in `data(simberloff)`, for which an extract is shown in Table 2.3:

Table 2.3: **Dataframe example of irregular sampling schemes.** Data for island E2 in `simberloff` dataset. PRE indicates if the taxa were present or absent in the sampling previous to the defaunation of the islands.

| Taxa | PRE | 31 | 51 | 69 | 86 | 111 |
|-----------------------------|-----|----|----|----|----|-----|
| <i>Aglaopteryx</i> sp. | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Latiblattella</i> n. sp. | 1 | 0 | 0 | 0 | 0 | 0 |
| <i>Latiblattella rehni</i> | 1 | 0 | 0 | 0 | 0 | 1 |
| <i>Cycloptilum</i> sp. | 1 | 0 | 0 | 0 | 0 | 0 |
| <i>Cyrtoxipha</i> sp. | 0 | 0 | 0 | 0 | 1 | 1 |

In both examples, we can see that we have columns for the taxa (or species) studied, some columns with additional information (such as Guild or Taxonomic Unit considered), and consecutive columns with data on presence (1) or absence (0). Studying ecological communities

can be messy. Temporal studies that track changes over many sites may often result in sampling schemes that are irregular both in time and space. In cases where we have different sampling schemes in our data, we require the different sampling schemes to be treated as different dataframes inside a list. An example would be the dataset `simberloff`.

2.2.2 Regular sampling schemes

The ideal scenario for ecological studies is one in which we sample our subject of study (say the community of fishes on a coral reef) regularly and frequently enough to observe changes in the community. Even though this is rare in ecology, it facilitates the estimation of rates and is less computationally intense. The estimation of rates for these regular sampling schemes has an analytically exact expression, while for irregular sampling schemes, we need to rely on heuristic methods or numerical solvers (see subsection 2.2.3).

Briefly, we can come back to table 2.1 to finish the example there. Say that we did not know the transition probabilities that generate the example. Using function `regular_sampling_scheme`, we obtain the m.l.e. of the rates producing the data. These are $c = 0.3769$ and $e = 0.7000$. With these rates, we can now calculate the associated transition probabilities, using function `cetotrans`, and obtaining $T_{01} = 0.4286$ and $T_{10} = 0.2308$. The equations $T_{11} = 1 - T_{01}$ and $T_{00} = 1 - T_{10}$ will give us a complete set of transition probabilities estimated from the data.

I include here another example to discuss several functionalities of the package. I will use `data(alonso15)`, a dataset of three lists that have information on presence-absence of the community of coral reef fish communities in atolls in the Lakshadweep Archipelago (India). I will use data only from one of the atolls, Kadmat. We have already seen an extract of the data in Table 2.2. There are several columns of data. The first column of the `data.frame` lists the species studied, while the second, its associated trophic level. The third to the eighth

columns contain presence-absence data of temporal samples collected in consecutive years between 2000 to 2003 and once again in 2010-2011. Finally, the last column has data on the guild of the species studied. For example, the third entry in the `data.frame` has data for *Acanthurus lineatus*, an algal feeder present in all the years studied.

From here, we can start estimating colonization and extinction rates for the entire community using the function `regular_sampling_scheme`. First, we have to specify a single `data.frame` in the function using argument `x =` and the name of our `data.frame`. Next, we need to tell the function in which columns the presence-absence data located with the argument `vector =`, in this case, columns 3 to 6. We will not use the data from 2010 and 2011. So let's estimate the colonization and extinction rates.

| c | c_up | c_low | e | e_up | e_low | N | NLL |
|--------|------|-------|--------|------|-------|-----|--------|
| 0.6035 | NA | NA | 0.3506 | NA | NA | 156 | 276.58 |

As we can see the colonization rate is $c = 0.6035$ and the extinction rate $e = 0.3506$. In addition, the output of the function tells us that we have examined 156 species and that the Negative Log-Likelihood of this model is $NLL = 276.58$. What if we want to know the rates for each guild? We just have to add to use arguments `level`, as in `level = "Guild"`, indicating the name of the column that divides the data into groups, and `n`, as in `n = 5` indicating the minimum number of species that a group needs in order to estimate its rates. When we do so, we obtain the following results.

| Group | c | c_up | c_low | e | e_up | e_low | N | NLL |
|------------------|--------|------|-------|--------|------|-------|----|-------|
| Algal Feeder | 0.7685 | NA | NA | 0.1630 | NA | NA | 30 | 38.96 |
| Corallivore | 0.4427 | NA | NA | 0.3935 | NA | NA | 20 | 35.72 |
| Macroinvertevore | 1.0152 | NA | NA | 0.5647 | NA | NA | 41 | 78.09 |
| Microinvertevore | 0.4960 | NA | NA | 0.5120 | NA | NA | 21 | 39.37 |
| Omnivore | 0.5091 | NA | NA | 0.4072 | NA | NA | 9 | 16.34 |
| Piscivore | 0.5263 | NA | NA | 0.6000 | NA | NA | 21 | 40.03 |
| Zooplanktivore | 0.5125 | NA | NA | 0.0919 | NA | NA | 14 | 15.94 |

Note that now we have relaxed the *equivalence* assumption.

2.2.3 Irregular sampling schemes

Given the complexities and challenges inherent to collecting real-world ecological data, temporal monitoring is often patchy, making irregular sampling schemes common in ecology. Data typically includes one or even several data sets with different sampling schemes. For accommodating the messiness of real-world data, `island` has two functions, `irregular_single_dataset` and `irregular_multiple_datasets`. Irregular sampling schemes force to alter how we calculate colonization and extinction rates, precluding the use of the algebraic (exact) method for estimating rates. We can still calculate the likelihood of the dynamic model using two methods with different approaches: a *heuristic* method and a *semianalytical* method.

The *heuristic* or *semianalytical* approaches of calculating rates are implemented in functions `irregular_single_dataset` and `irregular_multiple_dataset`, and we can switch between approaches using the argument `jacobian`. The *heuristic* method uses R's built-in optimization routine, `optim`, to obtain estimates for the colonization and extinction rates. The function `optim` uses an implementation of the Nelder-Mead algorithm to find the optimum of the likelihood function. However, *heuristic* methods do not guarantee to find the global optimum of the objective function, and they do not have

a mechanism to evaluate how good the optimum is for the selected point. In contrast, the *semianalytical* method guarantees that the optimum calculated is a root of the gradient of the log-likelihood function when the algorithm converges. In this package, we call routine `multiroot` from package `rootSolve` in order to find the semianalytical optimum value for our likelihood function. Even though we encourage the use of the gradient-based method, it may not always converge. The function will notify possible problems, and we recommend using the heuristic method if difficulties are encountered. One way to find good initial values for the gradient-based method is to use the estimates of the heuristic method as a starting point. The gradient-based method was termed incorrectly Jacobian-based when the arguments were named.

The functions called `irregular_single_dataset` and `irregular_multiple_datasets` reproduce all the functionalities of `regular_sampling_scheme`. Note that `irregular_single_dataset` requires numbers for the names of the columns with presence-absence data. The only other difference with the function for the regular sampling schemes is that we need to enter priors for c and e , our colonization and extinction rates, using arguments `c =` and `e =`.

Sampling schemes for similar communities can be very different due to multiple reasons, usually associated with the challenges of ecological research. The function `irregular_multiple_datasets` allows us to use data from different sampling schemes and estimate joint parameters for these data sets. In order to use these data sets, they need to comply with the general structure for irregular sampling schemes, and they should be combined in a list. The argument `list` will collate the data sets we want to study jointly, and the argument `vectorlist` must contain the vectors (ordered in time) that indicate where the presence-absence data is located. The remaining arguments work, as discussed in earlier sections.

2.2.4 Rates Vs. transition probabilities

There are two main differences between transition probabilities and rates. First, transition probabilities are dimensionless numbers, while rates have dimensions of time^{-1} . Second, with rates, we can directly estimate the transition probabilities for any time interval. In contrast, transition probabilities are associated with a specific time interval, so if we have $T_{10} = 0.6$ for a specific time interval and we need the transition probability for twice that interval $T'_{10} \neq 2 \times 0.6$. However, if a colonization rate per week is $c = 0.6 \text{ week}^{-1}$ the colonization rate after 2 weeks, this is, biweekly, would be $c' = 2 \text{ week}/1 \text{ (bi-week)} \times 0.6 \text{ week}^{-1} = 1.2 \text{ (bi-week)}^{-1}$, this is, only the units in which we express the rate change and everything else holds. This property allows us to work with irregular sampling schemes—much more common in ecology than regular ones—naturally, simply using a pair of colonization and extinction rates.

Function `cetotrans` produces transition probabilities for equivalent or different time intervals, given colonization and extinction rates. Arguments `c`, `e`, and `dt` must be vectors of the same length, in the case that we want to obtain transition probabilities that correspond to different time-steps.

2.2.5 Estimating confidence intervals

The estimates of colonization and extinction rates can not be regarded as fixed values, but have an associated uncertainty due, at least in part, to the amount of data we use to calculate the estimates. In interpreting these results, we recommend calculating confidence intervals for the rates we estimate. Confidence intervals can be calculated using function `regular_sampling_scheme` with the argument `CI = TRUE`. This option provides confidence intervals whose limits have a difference in log-likelihood of 1.96 units around the estimator.

We employ two separate methods to estimate confidence intervals: a

sequential method, which is computationally expensive, and a binary search seeded with the Hessian of the likelihood function. The rationale of the sequential method is the following: we start from one of the parameters, say c , and we sequentially add a small amount `step` to the parameter, obtaining the log-likelihood of the new value of the parameter, c' , until the difference in log-likelihood between c and c' climbs above 1.96. This gives us the upper boundary of c . The lower boundary is calculated similarly, but with sequential subtraction. This is a straightforward method, but it has to balance computational time and accuracy with the size of the interval `step`.

The alternative method uses the Hessian of the likelihood function as a seed for a binary search to determine the intervals. The Hessian provides an asymptotic estimator of the confidence interval, thus being a good starting point to look for the actual interval. The method works by calculating the likelihood of this asymptotic estimator and expanding the search (if the difference with our estimate is less than 1.96) or narrowing it (if it is greater than 1.96). This method typically finds the confidence interval in about ten steps and is significantly less computationally intense than the sequential method. However, in some cases, the value of the Hessian reduces almost to zero or produces a non-invertible matrix. In these cases, we recommend using the argument `step`.

2.2.6 A few words on model selection

Model selection identifies the best model that fits a given data set. In this package, we propose the use of AIC and the associated measure of *weight of evidence* (Burnham and Anderson, 2002) to choose the model that best fits the data and compares it against several alternative models. The functions in package 'island' provide Negative Log-Likelihoods (NLLs) that can easily be transformed to AIC with functions `akaikeic` or `akaikeicc`. We can then calculate the `weight_of_evidence` for each competing model that potentially ex-

plains our observations. As a reminder, AICs are only comparable when we try to explain the same data with different models because it makes no sense comparing the AIC of two different data sets with the same model since the data is different in both cases.

As a little example, consider the likelihoods of the two competing models that explain the community in Kadmat atoll. The model with only one pair of rates had a $NLL = 276.58$, while for the model that divides into guilds, it is $NLL = 264.45$. The log-likelihood of the model with different dynamics for each guild is over 12 points better than the model with a single dynamic group, which means that considering the guilds separately better explains the data than lumping together all guilds in a single group. However, the "guilds" model has 14 parameters while the "community" model has only 2. It is unsurprising that the more parameters the model has, the better it explains the data. For this reason, we calculate the AIC of each model using function `akaikeic`, that, to preserve parsimony, accounts for the number of parameters with the argument `k`. AIC values suggest that the "guilds" model is only marginally better than the "community" model, with a difference of 0.25. This is a tiny difference, and we cannot conclusively say that one of the models has more support than the other. This analysis is backed by the function `weight_of_evidence`, which provides a measure of the support that each competing model has. We found that the "guilds" model has a weight of evidence of 53% while the "community" model has 47%, and, as already noted, we cannot conclude that any model is clearly better.

2.2.7 Simulating colonization and extinction dynamics

The package `island` enables us to simulate ecological data, driven by colonization and extinction dynamics. Functions `data_generation` and `PA_simulation` allow us to generate species richness or presence-absence data from an initial vector of presence-absence for specific transition probabilities that can vary between different simulated tran-

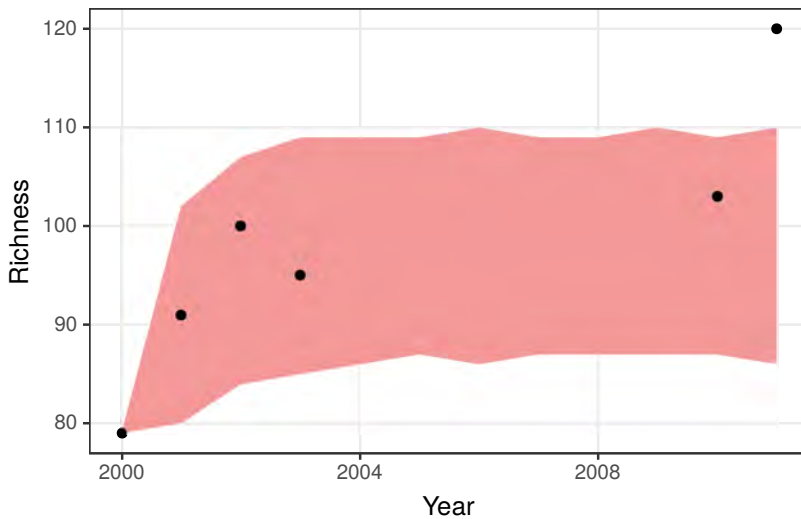


Figure 2.1: **Temporal evolution in Kadmat atoll.** Points indicate the observed species richness, while the red shade indicates the 95% distribution of simulations.

sitions.

As an example, we will simulate the temporal evolution in richness for Kadmat atoll. To do so, first, we have to transform the colonization and extinction rates to transition probabilities with function `cetotrans`. Next, I will use function `data_generation`. This function needs a starting point to simulate the dynamics. I will use the first column of presence-absence data, that has to be indicated with arguments `x =`, to indicate the `data.frame` of the initial data, and `column =`, giving the number of the column that contains the initial data. You also have to identify the transitions probabilities with argument `transitions =`. Finally, the number of realizations of the dynamics is specified with argument `iter =` and the number time-steps with argument `times =`. So, I simulated the dynamics in 999 realizations, up to the last sampling recorded. Figure 2.1 shows the result.

In the simulation, I used colonization and extinction rates estimated with the first four samples to simulate community dynamics from 2000 to 2011. We can use this to evaluate how reliably our parameter

estimates of colonization-extinction dynamics fit observed data. The plot shows the actual observations and the 95% distribution of the simulations of the dynamics. The actual data falls well within the boundaries delimited by the simulation except for the last sample.

We can estimate goodness-of-fit calculating model R^2 . Any estimation of R^2 requires a null model to compare with. R^2 measures how well our model performs compared to a null model of choice. As a null model, here we assume that, at any given time, we can have a number of species present from 0 to S_P , the number of species in the pool, which is drawn from a uniform probability distribution. In order to estimate goodness-of-fit, we use the function `r_squared`, which requires us to specify the arguments `observed` species richness, `simulated` species richness, and the number of species in the pool `sp`. The output, $R^2 = 0.9652$, indicates that we have a good fit compared to our null model. However, changes in the null model would lead to different estimates of R^2 . For this reason, we also include the function `simulated_model` that produces the quadratic error of a model given some data. We can estimate a new R^2 with a different null model and the previous function using the equation below:

$$R^2 = 1 - \frac{\epsilon^2}{\epsilon_0^2}, \quad (2.8)$$

where ϵ^2 corresponds to the quadratic error of a model given the data and ϵ_0^2 to the quadratic error of a null model.

2.2.8 Environmental fit

Apart from several biotic factors (e.g., the presence of top predators or the abundance of primary producers), the number of species in a community can also be influenced by several abiotic factors (e.g., temperature or rainfall) - referred to as environmental covariates. The package ‘island’ includes functions `all_environmental_fit`, `greedy_environmental_fit`, `custom_environmental_fit`, and `rates_calculator` to analyze the

influence of environmental covariates on the colonization and extinction dynamics of ecological communities. The functions assume a linear functional response for environmental covariates since it is the simplest way to build this dependency into colonization and extinction rates,

$$c_t = \alpha_0 + \sum_{i=1}^F \alpha_i Y_{it}, \quad e_t = \beta_0 + \sum_{i=1}^F \beta_i Y_{it},$$

where Y_{it} is the value of the i -th environmental covariate ($i = 1, \dots, F$) at time t .

We employ R expressions in these functions. An **expression** is an R object, frequently a call, symbol, or constant, that has to be evaluated before its use. We have minimized the use of expressions to make function calls easier to understand. However, we make use of them internally and provide function `custom_environmental_fit` that needs two expressions in order to hone the results from `all_environmental_fit` and `greedy_environmental_fit` as well as developing custom dependencies with environmental covariates that can be specified by advanced users.

All three functions for environmental fit need a single argument `dataset` with the same structure as the one for irregular schemes and an argument `vector` indicating the columns containing presence-absence data. Also, all three functions need another `data.frame` with related environmental covariates in columns. The names of these columns have to be specified to functions `all_environmental_fit` and `greedy_environmental_fit` with the argument `env`. These two functions also need arguments `c`, `e`, and `aic`, this is, tentative values for colonization and extinction rates and a tentative AIC value for the model. In practice, this AIC value should be chosen very large (of the order of 10^8 , but this depends on the size of the data set). The difference between `all_environmental_fit` and `greedy_environmental_fit` is that the latter employs a greedy algorithm that sequentially adds environmental covariates, one at a time, to the previously best set (using AIC), i.e., the algorithm finds first the best environmental

covariate and then the best combination with two covariates including the previously selected, and so forth until AIC does not justify the addition of new covariates. In contrast, `all_environmental_fit` tries all combinations of environmental variables, starting with the minimum number of environmental covariates to the maximum. Since the number of combinations rapidly becomes unmanageable, this method is not recommended except when we have a few environmental variables. An example of the use of these functions can be found in section 3.2.

The function `custom_environmental_fit` calls R optimizer function to find the m.l.e. for the parameters of the model with environmental variables, honing the estimation done by the previous functions. Besides arguments `dataset` and `vector`, it requires an argument `params` that corresponds to the parameters estimated for expressions `c_expression` and `e_expression`, for colonization and extinction. To simulate the dynamics under environmental variation, we need to calculate the value of the colonization and extinction rates for each sampling. To do so, we use `rates_calculator`, that uses arguments `params`, `c_expression`, and `e_expression` as in the previous function plus argument `t` that indicates the number of colonization and extinction rates needed. This function calculates the actual rates at each time in order to be able to simulate the dynamics of the colonization and extinction process resulting from these parameters. Note that these rates have dimensions of time^{-1} , so when converting rates to transition probabilities, we have to indicate the interval of time between samples for each rate.

2.2.9 Imperfect detectability

Most real-world ecological studies are characterized by imperfect detectability, i. e. the inability to detect a species or taxon despite its presence in a location. Imperfect detectability is a potential source of bias that must be avoided or at least estimated, particularly since it influences estimates of colonization and extinction. Unfortunately,

it is not always possible to avoid or estimate the effects of imperfect detectability. We should be cautious in interpreting estimates derived from the methods that assume perfect detectability. However, when we have a replicated sampling design, we can account for detectability while estimating colonization and extinction rates (MacKenzie et al., 2003).

MacKenzie et al. (2003) present a likelihood function to estimate site occupancy, colonization, and local extinction when a species is detected imperfectly. The method relies on replicate observations per sampling time. The implementation of this likelihood is not trivial because there might be several underlying colonization-extinction trajectories that are compatible with the same observed detection history. For example, a detection history such as $\{101\ 100\}$ means that we have three replicates in the first sampling time, 101, where we detected our hypothetical species twice, and a second sampling time, where we observed 100, this is, we detected the species only once. Since we detected it at least once at both time 0 and time 1, there is only one underlying colonization-extinction trajectory compatible with it, which we take the convention of collapsing it into $(1\ 1)$. However, imagine we fail to detect the species at time 1, being then our detection history $\{101\ 000\}$. In this case, two underlying trajectories are compatible with this observation, since the species could have or could not have gone extinct at time 1. These are $(1\ 1)$ and $(1\ 0)$. Therefore, the probability of the observed detection history $\{101\ 000\}$ should sum over the two ways in which that detection history could have been observed, either through the trajectory $(1\ 1)$ or $(1\ 0)$. For simplicity, let us analyze first what is the probability for the observed detection history $\{101\ 100\}$. The first sampling time always considers the probability of the species being present at the site, P_0 , as the fourth model parameter, and given that, the probability of making two out of three possible detections, $d^2 \cdot (1 - d)$. d is the detectability *per* replicate or probability of detecting a species when it is present per observation. The probability of being

also present at the time 1 given that the species was present at time 0 is given by T_{11} , and given that, the probability of making only one out three possible detections is $d \cdot (1 - d)^2$. Taking all together, this leads us to the following probability for the full detection history:

$$Pr(\{101\ 100\}) = P_0 \cdot d^2 \cdot (1 - d) \cdot T_{11} \cdot d \cdot (1 - d)^2$$

Now, let us examine the detection history $\{101\ 000\}$. As mentioned, we have two possibilities for the second sampling time: the species could be present and have not been detected or could have been truly absent. Notice then that the probability of the full detection history should sum over the two underlying colonization-extinction histories, $\{1\ 1\}$ and $\{1\ 0\}$. It would be:

$$Pr(\{101\ 000\}) = P_0 \cdot d^2 \cdot (1 - d) \cdot T_{11} \cdot (1 - d)^3 + P_0 \cdot d^2 \cdot (1 - d) \cdot T_{10}$$

where T_{10} is the probability of colonization.

As a final example, consider the detection history $\{001\ 000\ 101\ 000\ 111\}$. Four underlying colonization-extinction trajectories can produce this detection history. These are: $(1\ 1\ 1\ 1\ 1)$, $(1\ 0\ 1\ 1\ 1)$, $(1\ 1\ 1\ 0\ 1)$, and $(1\ 0\ 1\ 0\ 1)$. The probability of this detection history should sum over these four possible underlying colonization-extinction trajectories because all are compatible with it. Below we detailed the four conditional probabilities:

$$Pr(\{001\ 000\ 101\ 000\ 111\} | (1\ 1\ 1\ 1\ 1)) =$$

$$P_0 \cdot d \cdot (1-d)^2 \cdot T_{11} \cdot (1-d)^3 \cdot T_{11} \cdot d^2 \cdot (1-d) \cdot T_{11} \cdot (1-d)^3 \cdot T_{11} \cdot d^3$$

$$Pr(\{001\ 000\ 101\ 000\ 111\} | (1\ 0\ 1\ 1\ 1)) =$$

$$P_0 \cdot d \cdot (1-d)^2 \cdot T_{01} \cdot T_{10} \cdot d^2 \cdot (1-d) \cdot T_{11} \cdot (1-d)^3 \cdot T_{11} \cdot d^3$$

$$Pr(\{001\ 000\ 101\ 000\ 111\} | (1\ 1\ 1\ 0\ 1)) =$$

$$P_0 \cdot d \cdot (1-d)^2 \cdot T_{11} \cdot (1-d)^3 \cdot T_{11} \cdot d^2 \cdot (1-d) \cdot T_{01} \cdot T_{10} \cdot d^3$$

$$Pr(\{001\ 000\ 101\ 000\ 111\} | (1\ 0\ 1\ 0\ 1)) =$$

$$P_0 \cdot d \cdot (1-d)^2 \cdot T_{01} \cdot T_{10} \cdot d^2 \cdot (1-d) \cdot T_{01} \cdot T_{10} \cdot d^3$$

The algorithm implemented in the package 'island' would sum over these four conditional probabilities to calculate the total probability for the initial detection history, $Pr(\{001\ 000\ 101\ 000\ 111\})$. Note that, for real-life examples, when a species goes fully undetected for many sampling times, the full total sum becomes unfeasible because the number of compatible trajectories undergoes a combinatorial explosion rapidly. This may happen in practice if detectability per replicate is very low. In this case, only approximated likelihoods can be given. Alternatively, one could get around this problem by redesigning the full survey and taking more replicates per sampling time. Transition probabilities T_{00} , T_{10} , T_{01} , T_{11} are functions of the rates c and e for a given time interval Δt between observations. Therefore, we have all the elements required to estimate the likelihood of any detection history, even if time intervals between observations vary, which allows to find maximum likelihood estimates for the four model parameters, colonization and extinction rates, c and e , along with the detectability, d , and the probability of initial presence, P_0 .

Table 2.4: **Dataframe example for parameter estimation under imperfect detectability.** All the species shown pertain to the genus *Acanthurus*. 0.1 indicates missing values.

| Species | Atoll | Guild | X2000 | X2000.1 | X2001 | X2001.1 | X2001.2 | X2001.3 |
|------------------------|--------|----------------|-------|---------|-------|---------|---------|---------|
| <i>A. auranticavus</i> | Agathi | Algal Feeder | 1 | 1 | 1 | 0 | 1 | 0.1 |
| <i>A. leucosternon</i> | Agathi | Algal Feeder | 1 | 1 | 1 | 0 | 0 | 0.1 |
| <i>A. lineatus</i> | Agathi | Algal Feeder | 1 | 1 | 1 | 0 | 0 | 0.1 |
| <i>A. nigrofuscus</i> | Agathi | Algal Feeder | 0 | 0 | 0 | 0 | 0 | 0.1 |
| <i>A. thompsoni</i> | Agathi | Zooplanktivore | 0 | 0 | 0 | 0 | 0 | 0.1 |
| <i>A. triostegus</i> | Agathi | Algal Feeder | 1 | 1 | 0 | 1 | 1 | 0.1 |

Data entry

In order to estimate detectability, we need to provide presence-absence data with replicated samples for the same sampling time, as in the example below extracted from data set `lakshadweepPLUS`, where column `X2000` and `X2000.1` correspond to two replicate transects sampled in the same year. Besides, the data can have groups that can be treated as levels of a factor, as in the column "Guild".

Estimating colonization and extinction rates with imperfect detectability

Functions `sss_cedp`, `mss_cedp` allow the estimation of colonization and extinction rates with imperfect detectability with single and multiple sampling schemes, respectively. The function `sss_cedp` allows estimation for a single sampling scheme with repeated measures that has to be specified with arguments `Time`, that contains the unique sampling times, and argument `Transects`, that specifies the number of transects per sampling time. By contrast, `mss_cedp` allows the estimation of rates with perfect or imperfect detectability for multiple sampling schemes, via the use of flags for missing values specified by argument `MV_FLAG`, for the whole data set or groups of factors. A full sampling scheme should be specified with argument `Time`, which is internally used to calculate the particular sampling schemes associated

with each separate row with the help of the missing value flags on the columns that have not been sampled. When specifying the argument `Guild`, we relax the equivalence assumption entirely and estimate the four parameters for each group. An example can be found in section 3.1.

2.2.10 Model selection grouping

Model selection aims to select the best model for a given phenomenon with a reasonable number of parameters describing it and avoiding over-fitting. Our procedure is intended to distinguish, for example, guilds or islands with different colonization and extinction dynamics.

The function `upgma_model_selection` incorporates a UPGMA algorithm based model selection procedure intended to find an optimal partition that minimizes AIC values. The algorithm needs a vector of tags in order to estimate the partition. This function allows the estimation of colonization and extinction rates with or without imperfect detectability. It also generates two output files in latex format (.tex) with: a) the parameters of the best model found under the model selection procedure and b) the summary of the procedure. Table 3.2 shows an example.

2.2.11 Immigration, Birth, and Death stochastic models

Colonization and extinction rates can be considered *effective rates* in the sense that they are average approximations to the intrinsic individual-level processes driving species-specific dynamics. This function includes three simple stochastic population models simulated with Gillespie's exact algorithm (Gillespie, 1977). All three models assume individuals as discrete entities subject to a range of processes. These functions aim to show the correspondence between parameters driving individual deaths and births, and colonization-extinction rates at the population level. The population models included in function `ibd_models` are the seminal model by Kendall (1948), a mainland-island model by Alonso &

Table 2.5: **Stochastic IBID models included in package 'island'.**

| Model | Immigration | Birth | Death |
|--------------------------|---------------------------|----------------------------|---|
| | $\emptyset \rightarrow A$ | $A \rightarrow A + A$ | $A \rightarrow \emptyset$ |
| Kendall (1948) | μ | βn | δn |
| Alonso & McKane (2002) | $\mu(K - n)$ | $\beta n(1 - \frac{n}{K})$ | δn |
| Haegeman & Loreau (2010) | μ | βn | $n(\delta + (\beta - \delta)\frac{n}{K})$ |

McKane (2002), and a basic population model with density-dependent deaths by Haegeman & Loreau (2011).

The three models have three basic processes that differ in their parametrization: immigration (arrival of new individuals from outside the population), individual births, and individual deaths. In table 2.5, we specify the processes and the transitions (with probability rates in units of time^{-1}) associated with each model. Please note that Kendall's model may "explode" if the birth rate is higher than the death rate. Notice also that the death rate in the Haegeman & Loreau model may become ill-defined, this is, negative for death rate values higher than the value for the birth rate, given that probability rates can never be negative.

The function `ibd_models` enables us to simulate the stochastic population dynamics described in the previous section. The function requires that we specify arguments `n0` (initial population size), `beta` (birth rate), `delta` (death rate), `mu` (immigration rate), and `K` (carrying capacity) when required. We also need to specify a vector of sampling times `time_v` and the model we plan to use via argument `type`. To illustrate its stochastic nature, I show a simulation of the dynamics of Alonso and McKane's (2002) model, with equal birth and death rates and low immigration (Figure 2.2). An example of the use of IBID stochastic models is shown in Box 1 (Chapter 3).

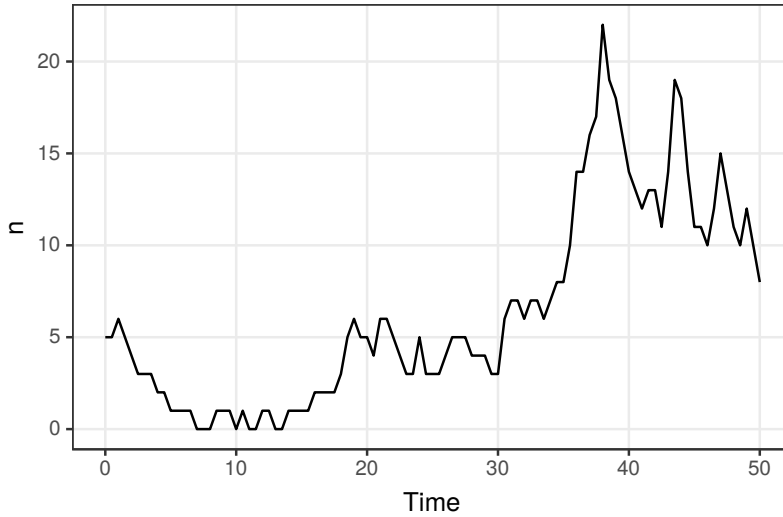


Figure 2.2: **Simulation of the dynamics of the model by Alonso & McKane.** In black, a stochastic realization with the following parameters $n_0 = 5$, $\mu = 0.005$, $\beta = 0.3$, $\delta = 0.3$ and $K = 100$.

2.3 The master equation approach for n species

I will finish this chapter with a result that, despite the extensive literature on the TIB, has not been reported yet, to the best of my knowledge. Consider the simplest stochastic model underlying TIB again. This model describes the temporal evolution for the number of species on an island, n . We could use n as our state variable in the master equation. Now, species colonize the island at rate λ from the mainland and undergo extinction at rate δ . The mainland harbors a total number of potentially colonizing species, S_P . With these elements, the master equation, that is, the evolution over time of the probability of finding each state n , can be written as follows:

$$\begin{aligned}
 \frac{dP(n, t)}{dt} = & \overbrace{\lambda(S_P - n - 1)P(n - 1, t)}^{\text{probability of a colonization with } n - 1 \text{ species}} \\
 & + \overbrace{\delta(n + 1)P(n + 1, t)}^{\text{probability of a extinction with } n + 1 \text{ species}} \\
 & - \overbrace{\lambda(S_P - n)P(n, t)}^{\text{probability of a colonization with } n \text{ species}} \\
 & - \overbrace{\delta n P(n, t)}^{\text{probability of a extinction with } n \text{ species}}
 \end{aligned} \tag{2.9}$$

Solving a master equation directly is usually difficult, if not impossible. However, this equation can be solved through the associated probability generating function. Probability generating functions (p.g.f.) are a useful way to describe a random variable, which allows easy calculation of the expected value and higher moments, as the variance. A p.g.f., $G(x)$, is formally defined as:

$$G(x) = \sum_{j=0}^{\infty} p_j x^j \tag{2.10}$$

Normalization requires that $G(1) = 1$ (meaning that the sum of probabilities of all events must be 1), and the mean or expected value is known to be $E(x) = G'(1)$, as well as the variance is $Var(x) = G''(1) + G'(1) - [G'(1)]^2$.

Now, I will derive the p.g.f. associated with the master equation of the simplest model of TIB. Note that the temporal evolution of the p.g.f. would arise multiplying the terms of the master equation 2.9 by x^n and summing terms over n . It would be as follows:

$$\begin{aligned}
 \frac{dP(n, t)x^n}{dt} = & \sum \lambda(S_P - (n - 1))P(n - 1, t)x^n \\
 & + \sum \delta(n + 1)P(n + 1, t)x^n \\
 & - \sum \lambda(S_P - n)P(n, t)x^n \\
 & - \sum \delta n P(n, t)x^n
 \end{aligned}$$

which reordering a bit corresponds to:

$$\begin{aligned} \frac{dP(n, t)x^n}{dt} &= \lambda S_P x \sum P(n-1, t)x^{n-1} - \lambda x^2 \sum (n-1)P(n-1, t)x^{n-2} \\ &\quad + \delta \sum (n+1)P(n+1, t)x^n \\ &\quad - \lambda S_P \sum P(n, t)x^n + \lambda x \sum nP(n, t)x^{n-1} \\ &\quad - \delta x \sum nP(n, t)x^{n-1} \end{aligned}$$

That in terms of $G(x, t)$ is expressed as:

$$\begin{aligned} \frac{dG(x, t)}{dt} &= \lambda S_P x G(x, t) - \lambda x^2 \frac{\partial G(x, t)}{\partial x} \\ &\quad + \delta \frac{\partial G(x, t)}{\partial x} \\ &\quad - \lambda S_p G(x, t) + \lambda x \frac{\partial G(x, t)}{\partial x} \\ &\quad - \delta x \frac{\partial G(x, t)}{\partial x} \end{aligned}$$

So we have that:

$$\begin{aligned} \frac{dG(x, t)}{dt} &= (\lambda x + \delta)(1-x) \frac{\partial G(x, t)}{\partial x} + \lambda S_p (x-1)G(x, t) \\ \frac{dG(x, t)}{dt} - (\lambda x + \delta)(1-x) \frac{\partial G(x, t)}{\partial x} &= \lambda S_p (x-1)G(x, t) \end{aligned}$$

The solution to the equation above is not readily available. However, we can use the method of characteristics to solve it. The method of characteristics is a general method to solve linear, first-order, partial differential equations (PDE). It finds curves (called characteristic curves) to define the solution of the initial PDE. The characteristic curves of this equation are:

$$\frac{dx}{(\lambda x + \delta)(x-1)} = \frac{dt}{1} = \frac{dG}{\lambda S_p (x-1)G(x, t)} \quad (2.11)$$

So, we will find now the first characteristic curve solving for the left and right characteristic curves:

$$\frac{dx}{(\lambda x + \delta)(x-1)} = \frac{dG}{\lambda S_p (x-1)G(x, t)}$$

$$\text{reordering } \frac{\lambda S_p(x-1)dx}{(\lambda x + \delta)(x-1)} = \frac{dG}{G(x,t)};$$

$$\text{simplifying and integrating } \lambda S_p \int \frac{dx}{(\lambda x + \delta)} = \int \frac{dG}{G(x,t)} \Rightarrow$$

$$S_p \ln(\lambda x + \delta) + k = \ln(G); \quad \ln[k(\lambda x + \delta)^{S_p}] = \ln(G);$$

$$\text{and exponentiating } K_1(\lambda x + \delta)^{S_p} = G(x,t)$$

The integration constant K_1 can be determined using the normalization requirement of the p.g.f.:

$$G(x,0) = 1 \Rightarrow K_1(\lambda x_0 + \delta)^{S_p} = 1; \quad K_1 = (\lambda x_0 + \delta)^{-S_p} \Rightarrow$$

$$G(x,t) = (\lambda x_0 + \delta)^{-S_p} (\lambda x + \delta)^{S_p}$$

We now must solve for x_0 to have our p.g.f. ready. For that, we are going to use the characteristic curves (equation 2.11) on the left and center:

$$\frac{dx}{(\lambda x + \delta)(x-1)} = \frac{dt}{1};$$

$$\text{integrating } \int \left(\frac{1/(\lambda + \delta)}{x-1} + \frac{-\lambda/(\lambda + \delta)}{\lambda x + \delta} \right) dx = \int \frac{dt}{1} \Rightarrow$$

$$\frac{1}{\lambda + \delta} \ln\left(\frac{x-1}{\lambda x + \delta}\right) = t + k; \quad \left(\frac{x-1}{\lambda x + \delta}\right)^{\frac{1}{\lambda + \delta}} = K_2 e^t; \quad e^{-t} \left(\frac{x-1}{\lambda x + \delta}\right)^{\frac{1}{\lambda + \delta}} = K_2$$

K_2 can be found easily with the initial condition, and it will be possible to find x_0 :

$$K_2 = \left(\frac{x_0 - 1}{\lambda x_0 + \delta}\right)^{\frac{1}{\lambda + \delta}}; \quad e^{-t} \left(\frac{x-1}{\lambda x + \delta}\right)^{\frac{1}{\lambda + \delta}} = \left(\frac{x_0 - 1}{\lambda x_0 + \delta}\right)^{\frac{1}{\lambda + \delta}};$$

$$\text{we solve for } x_0 = \frac{\lambda x + \delta + \delta(x-1)e^{-t(\lambda + \delta)}}{\lambda x + \delta - \lambda(x-1)e^{-t(\lambda + \delta)}}$$

Substituting in our expression for $G(x,t)$:

$$G(x,t) = (\lambda x + \delta)^{S_p} \left(\lambda \frac{\lambda x + \delta + \delta(x-1)e^{-t(\lambda + \delta)}}{\lambda x + \delta - \lambda(x-1)e^{-t(\lambda + \delta)}} + \delta \right)^{-S_p} \quad (2.12)$$

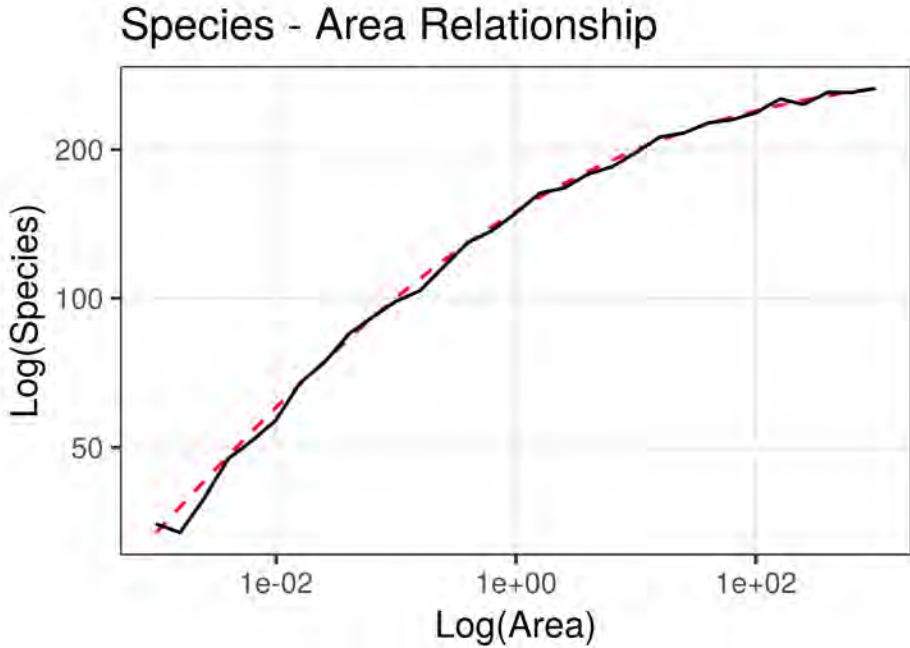


Figure 2.3: **Species Area Relationship.** In black, mean of 5 stochastic realizations of the stochastic model of island biogeography with the following parameters $S_P = 300$, $\lambda = 0.3$, $\delta = 0.3$, $\alpha = 0.2$, $\beta = 0.1$, over a range of areas. In red, theoretical expectation of the same dynamics. Axes in Log10.

Equation 2.12 is the p.g.f. that we seek. Now we can find the expected value for the temporal evolution of the dynamics and its variance easily.

$$E(x) = G'(1, t) = S_P \frac{\lambda}{\lambda + \delta} (1 - e^{-t(\lambda + \delta)}) \quad (2.13)$$

$$Var(t) = S_P \frac{\lambda(1 - e^{-t(\lambda + \delta)})(\delta + \lambda e^{-t(\lambda + \delta)})}{(\lambda + \delta)^2} \quad (2.14)$$

Note that equation 2.13 is akin to the known number of species at equilibrium when $\lambda \equiv c$, $\delta \equiv e$ and time goes to infinity.

To finish this part, I will do a little exercise to demonstrate the potential use of the above expressions to derive the Species Area Relationship. Usually bigger islands harbor more species. Assuming that extinction and colonization rates scale with area according to:

$$\begin{aligned}\lambda &= \lambda_0 A^\beta \\ \delta &= \delta_0 A^{-\alpha}\end{aligned}$$

where α and β are positive exponents, we could find the expression, via Eq. 2.13, for the SAR. If we just assume that $t = \infty$, so we have reached equilibrium, the expression for the number of species in the equilibrium reduces to $E[n] = S_P \frac{\lambda}{\lambda + \delta}$. The expression for the SAR would be:

$$E[n(A)] = S_P \frac{\lambda_0 A^\beta}{\lambda_0 A^\beta + \delta_0 A^{-\alpha}}$$

Simulations with R Package 'island' confirm this approximation, as can be seen in Figure 2.3. The simulations followed the expectation closely, fluctuating around the expected number of species.

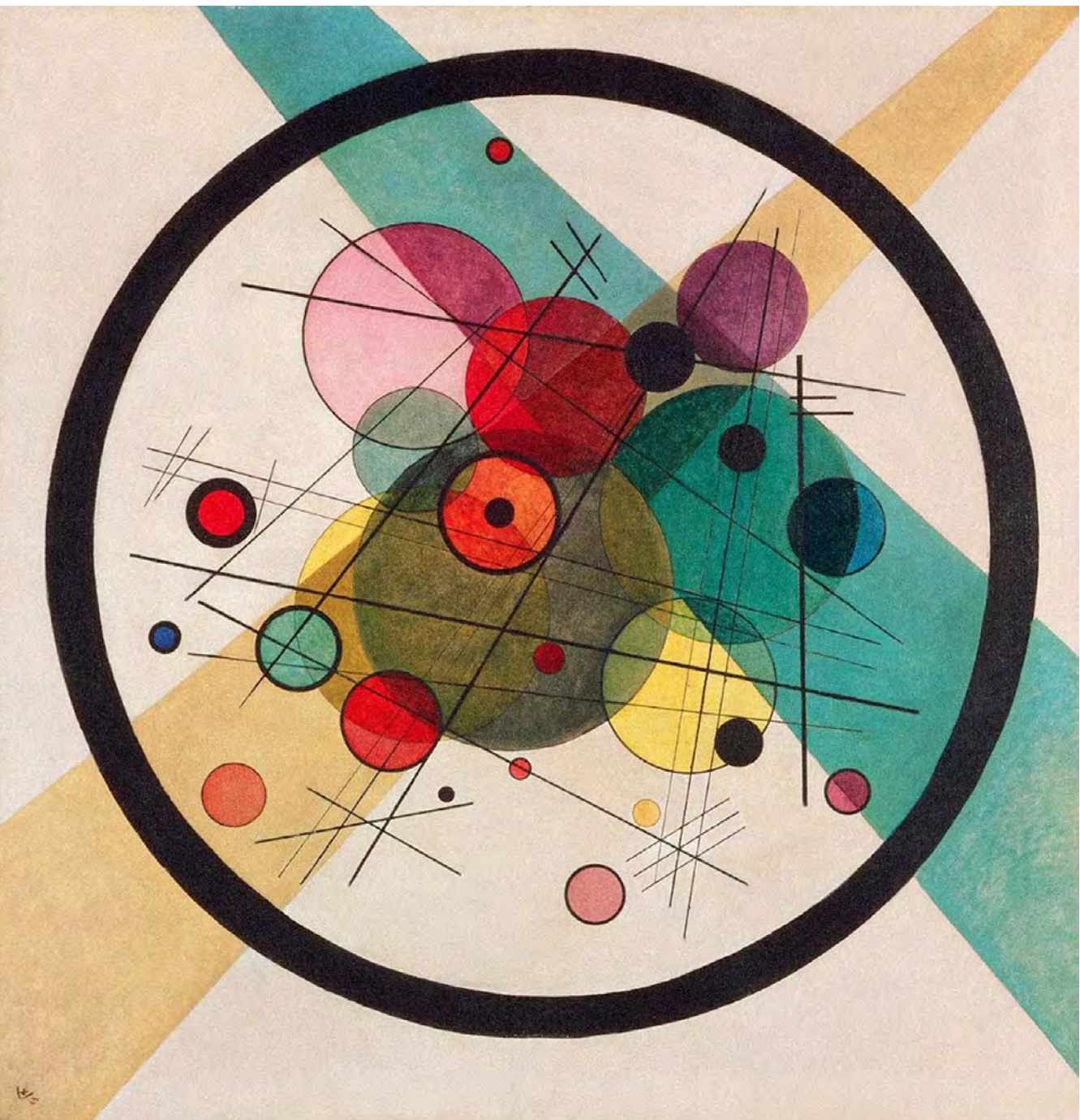


Image:
Wassily Kandinsky. *Circles in a Circle*. 1923, Philadelphia Museum of Art, Philadelphia.

This chapter, as published:
Ontiveros, VJ, Capitán, JA, Arthur, R, Casamayor, EO, Alonso, D. (2019). Colonization and extinction rates estimated from temporal dynamics of ecological communities: The `island` r package. *Methods in Ecology and Evolution* 10:1108–1117. <https://doi.org/10.1111/2041-210X.13176>

Colonization and Extinction Rates estimated from Temporal Dynamics of Ecological Communities: The `island` R Package

Abstract

1. The temporal dynamics of ecological communities are inherently complex, as dispersal processes, interspecific biotic interactions and environmental fluctuations interact to shape species assemblages. We developed an R package, `island` — a community ecology tool to analyse temporal patterns and explore potential drivers of community dynamics.
2. The R package `island` is inspired by the simplicity of the Theory of Island Biogeography. Our package includes a suite of functions to easily confront models of increasing complexity to community data from the simplest stochastic model underlying Island Biogeography Theory to more elaborate models that include species responses to environmental variables and imperfect detectability, obtained at regular or irregular sampling times, using maximum-likelihood and AIC-based model selection.
3. Models implemented in `island` should be regarded as simple approximations of true temporal community dynamics. In this sense, we call them *effective models* which we use to estimate *effective parameters*. These models are simple, but dynamic, as opposed to purely statistical. We illustrate the concept of dynamic *effective models* by using an immigration-birth-death

stochastic model of population dynamics. We generated simulated community data to demonstrate the correspondence between the underlying true model parameters values and our effective parameter estimates.

4. We present three examples showcasing the uses of our package: detectability estimation and model selection, influence of environmental covariates on community dynamics and estimation of co-occurrence networks. Three detailed vignettes with a full set of examples accompany the R Package `island` available on CRAN.

Keywords: temporal dynamics, community ecology, colonization, extinction, stochastic processes, island biogeography, R package, effective models.

The dynamics of community assembly play out over ecological time scales. Its central unresolved problems require datasets that span these temporal scales. Long-term data hold the key to unraveling the respective roles that ecological interactions and environmental factors play in community assembly over space and time. Our understanding of how species diversity and community structure change with time has been influenced by opposing theories emphasizing either dispersal and individual species adaptations to local environments (Gleason, 1926), or biotic interactions (Clements, 1916). Earlier approaches in studying the relative importance of community assembly drivers have largely been based on randomization procedures of species within the assembly (package `STPCAM`, van der Plas et al. (2015), Chase et al. (2011)). Related approaches use several community-level patterns, such as the species–abundance distribution and the aggregation of conspecific individuals, to constrain community assembly, and study the behaviour of different biodiversity metrics across scales (package `mobsim`, May et al. (2018)). Although these approaches are useful for some purposes, we need models incorporating assumptions at the individual and/or species level since community-level patterns

truly emerge from bottom up ecological processes affecting the fate of individuals and species (Allouche and Kadmon, 2009; Hortal et al., 2009; Hubbell, 2001; May et al., 2015; Vellend, 2016). The range of possible mechanistic simulation models for community dynamics is infinite (Cabral et al., 2017; Hortal et al., 2012), and such models are usually difficult to decipher (May et al., 2015), often clouding rather than enlightening our understanding of community assembly processes. It is our contention that there is still considerable power in the dynamic simplicity of Island Biogeography Theory (MacArthur and Wilson, 1963, 1967) that is yet unexplored and largely overlooked. Here, we work with the stochastic model underlying this theory and develop a dynamic framework to analyze the relative importance of competing processes in structuring real-world communities.

The ultimate test of ecological theory lies in confronting it with real-world data over ecologically relevant time scales to see if its basic assumptions still hold. Ecological monitoring programmes are only now beginning to generate rich long temporal datasets (Hortal et al., 2015), for a variety of taxa around the world (see, for instance, Condit (1998) and Hubbell et al. (2005)), and developing robust methods to deal with their complexities is often a challenge. For many systems, the speed of data accumulation far outstrips our ability to explore and generate new conceptual ideas emerging from these data (Barberán et al., 2014). This is perhaps unsurprising given the lack of common conceptual frameworks providing simple effective models to confront new theoretical ideas with community data.

We present an R package, `island`, that helps analyze temporal patterns in community assembly. Inspired by Island Biogeography Theory (IBT), we focus on estimating colonization and extinction rates from species presence data over time and space. Functions in `island` implement a number of methods to characterize communities and guilds within communities, and test conceptual ideas of the drivers of community assembly through AIC-based model selection. While

other R packages exist that are geared to estimate colonization and extinction dynamics, such as R package `unmarked` (function `colect()`), they typically focus on single populations in space and time rather than a community-level approach. In addition, using existing packages for sparse data or irregular sampling schemes is not straightforward. Although recent community-level approaches that expand IBT to include biological interactions are of great value (Cazelles et al., 2016b; Gravel et al., 2011), they do not provide robust maximum-likelihood methods for the analysis of community time series data. Instead, they are centered on the distribution of species presences at stationarity, under different interaction structures, across scales, and over environmental gradients (Cazelles et al., 2016b).

A detailed description of the theoretical underpinnings of the package is provided in the Supplementary Material S0. Our basic model is the simplest stochastic formulation of the theory of island biogeography (Alonso et al., 2015). Since temporal dynamics of real-world communities are influenced by an entangled web of factors, we emphasize that the colonization-extinction models of R package `island` should only be regarded as a simple way to effectively approximate community dynamics under certain assumptions. Colonization and extinction pairs (c, e) represent effective parameters characterizing communities as if population changes were driven by independent species dynamics. These parameters should be regarded as coarse-grained approximations of the intrinsic, species-specific parameters driving population processes and species interactions. In Box 1, we illustrate the correspondence between effective (c, e) rates and the underlying parameters driving community dynamics. Species population dynamics can be modeled as a stochastic process driven by elemental events such as immigration of new species to local communities, local births and deaths, and interactions between individuals of the same or different species (Capitán et al., 2015; Capitán et al., 2017; Haegeman and Loreau, 2011; Solé et al., 2002; Vellend, 2010, 2016). In Box 1, abundances were generated

Table 3.1: Vignettes and main functions of R package `island`.

| Main functions | Description |
|--|---|
| <code>regular_sampling_scheme()</code> , <code>irregular_single_dataset()</code> , <code>irregular_multiple_datasets()</code> <code>sss_cepdc()</code> , <code>mss_cepdc()</code> | Estimates colonization and extinction rates under regular (temporally equispaced) sampling schemes. Estimates colonization and extinction rates under irregular sampling schemes for single or multiple datasets. Estimates colonization and extinction rates, as well as detectability and initial occupancy probabilities, with data in single or multiple replicated datasets. |
| <code>upgma_model_selection()</code> | Conducts a model selection procedure that searches for an optimal partition of the data (based on a categorical variable) for its colonization and extinction parameters (or, alternatively, for detectability and initial occupancy). |
| <code>greedy_environmental_fit()</code> | Estimates the best model for colonization and extinction rates and their dependency on environmental variables using a greedy algorithm that sequentially adds the variables with most influence in the data. |
| <code>cetotrans()</code> | Transforms colonization and extinction rates to transition probabilities after a specified interval of time. |
| <code>PA_simulation()</code> , <code>data_generation()</code> | Simulates species richness dynamics or presence-absence matrices according to specified transition probabilities. |
| <code>r_squared()</code> , <code>simulated_model()</code> <code>ibd_models()</code> | Returns the value of R^2 or the average squared error of the predictions of model simulations. Simulates stochastic population dynamics under immigration, birth and death processes under three different models (Alonso and McKane, 2002; Haegeman and Loreau, 2011; Kendall, 1948). |
| Vignette | Description |
| <code>vignette("island")</code> | Introduces the main concepts applied in the package, details data preparation, and illustrates the use of the functions that estimate colonization and extinction rates, as well as the influence of environmental variables, and simulation of the associated dynamics. |
| <code>vignette("detectability")</code> | Explains how colonization and extinction rates are estimated under imperfect detectability, and details data entry and use of functions under this scenario. Also explains model selection to detect groups of species of equivalent colonization-extinction dynamics. |
| <code>vignette("IBDmodels")</code> | Shows the correspondence between parameters driving deaths, births and immigrations at the individual level and colonization-extinction rates at the population level. |

from a species immigration-birth-death (IBD) model (Haegeman and Loreau, 2011), and then transformed into presence-absence data in order to estimate colonization-extinction probabilities. Other stochastic community models can be also used (for detailed information see `help("ibd_models")` and `vignette("IBDmodels")`).

Box 1. The colonization-extinction model and community dynamics.

Here we generated synthetic, simulated communities using an IBD stochastic model (Haegeman and Loreau, 2011). Births occur with probability r^+a_i per unit time (a_i is the i -th species abundance), and immigrations at rate μ . Species abundances decrease at rate r^-a_i (intrinsic mortality) or ra_i^2/K (intraspecific competition), where K represents a carrying capacity, and $r = |r^+ - r^-|$. These assumptions result in a deterministic logistic population model with immigration ($da_i/dt = ra_i(1 - a_i/K) + \mu$ for $i = 1, 2, \dots, S$). In Fig B1, we show the correspondence between the population carrying capacity, K , and the colonization probability, T_{10} , and the extinction probability, T_{01} (see Eq. S0-6), in panel **a** and **b**, respectively. In panels **c** and **d**, we show how an empirically-derived functional dependence between colonization and extinction probabilities, estimated from community time series data, would constrain the scaled rates of the IBD that are compatible with it. Simulation averages over 3000 stochastic replicates of the IBD community model yield two probability surfaces, $T_{01}(\mu/r^+, r^-/r^+)$ and $T_{10}(\mu/r^+, r^-/r^+)$, as functions of scaled immigration and mortality rates (**c**). The intersection $T_{01} = f(T_{10})$ provided an estimation of the region in parameter space where the intrinsic IBD species rates are compatible with the observed, macroscopic rates (**d**).

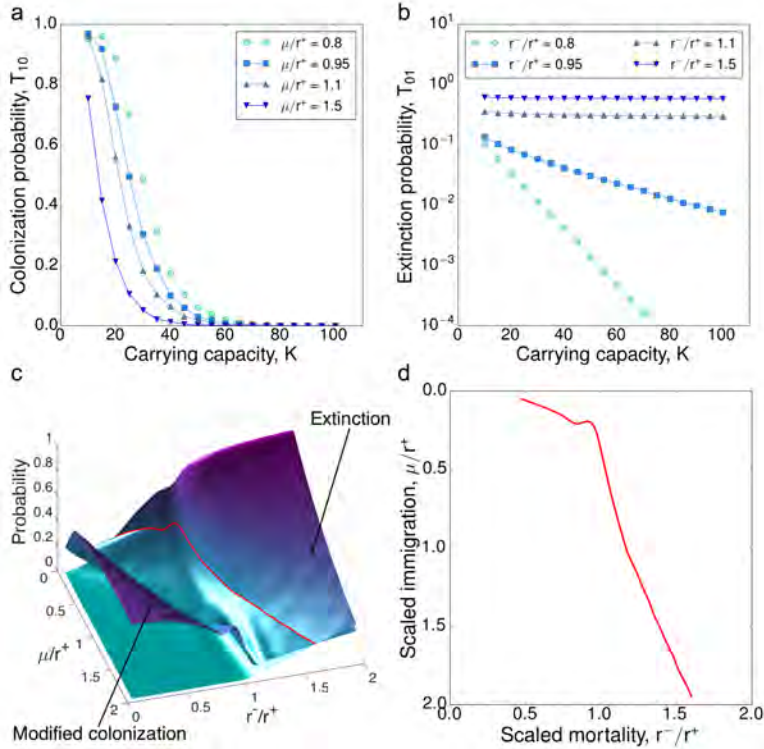


Figure B1. All curves in panel **a** are shown for the same scaled mortality ($r^-/r^+ = 0.8$), while all curves in panel **b** share the same value for the scaled immigration ($\mu/r^+ = 0.5$). In all cases simulation were done for $S = 20$ species. The assumed functional dependence between colonization and extinction probabilities is $T_{01} = f(T_{10}) = 0.6 - 0.5T_{10}$ (panels **c** and **d**).

Without exhausting the numerous possibilities of the package (Table 3.1 summarizes its functions and vignettes), we discuss three worked examples showcasing the capabilities of `island`. In the first, we explore how imperfect detectability influences model parameter estimates, and show how our package functions identify species groups or guilds and help investigate the extent to which species ecological equivalence underlies temporal community dynamics (Alonso et al., 2015). We then extend the basic model to analyze the influence of environmental variability on community dynamics. Finally, we show how functions in `island` can help unveil the potential interaction structure of species

assemblages using a novel way of assigning significance to empirical species co-occurrences.

3.1 Detectability and model selection

When working with presence datasets we often make the simplifying assumption of perfect detectability, assuming that sufficient replication at the site level accounts for genuine absences. In reality, it is difficult to discard interspecies differences in detectability, which could have profound consequences when interpreting assemblage patterns. MacKenzie developed a likelihood approach for parameter estimation when species detectability is not perfect. Mackenzie's seminal idea can be used to obtain better estimates of site occupancies (MacKenzie et al., 2002) as well as non-biased estimates of colonization-extinction model parameters (MacKenzie et al., 2003). Package `island` provides a new implementation of MacKenzie's likelihood for uneven time intervals and sparse sampling schemes.

Figure 3.1 shows parameter estimation, using function `regular_sampling_scheme`, compared to estimates obtained with imperfect detectability (function `sss_cedp`). We generated a set of data matrices based on stochastic iterations of a colonization-extinction model, whose parameters are known, drawn randomly from a uniform distribution between 0 and 1. From these data, we compared true parameter values with parameter estimates, both with perfect and imperfect detectability, as detectability increases. We show that detectability should be above 0.9 per site to yield parameter estimates close to true parameter values. By contrast, detectability per transect above 0.25 seems to be enough to obtain reliable parameter estimates for models including imperfect detectability (see function `mss_cepd` and `vignette("detectability")`).

In addition, we used `data("lakshadweepPLUS")` on coral reef fishes (Alonso et al., 2015) to investigate whether trophic guilds are *effectively*

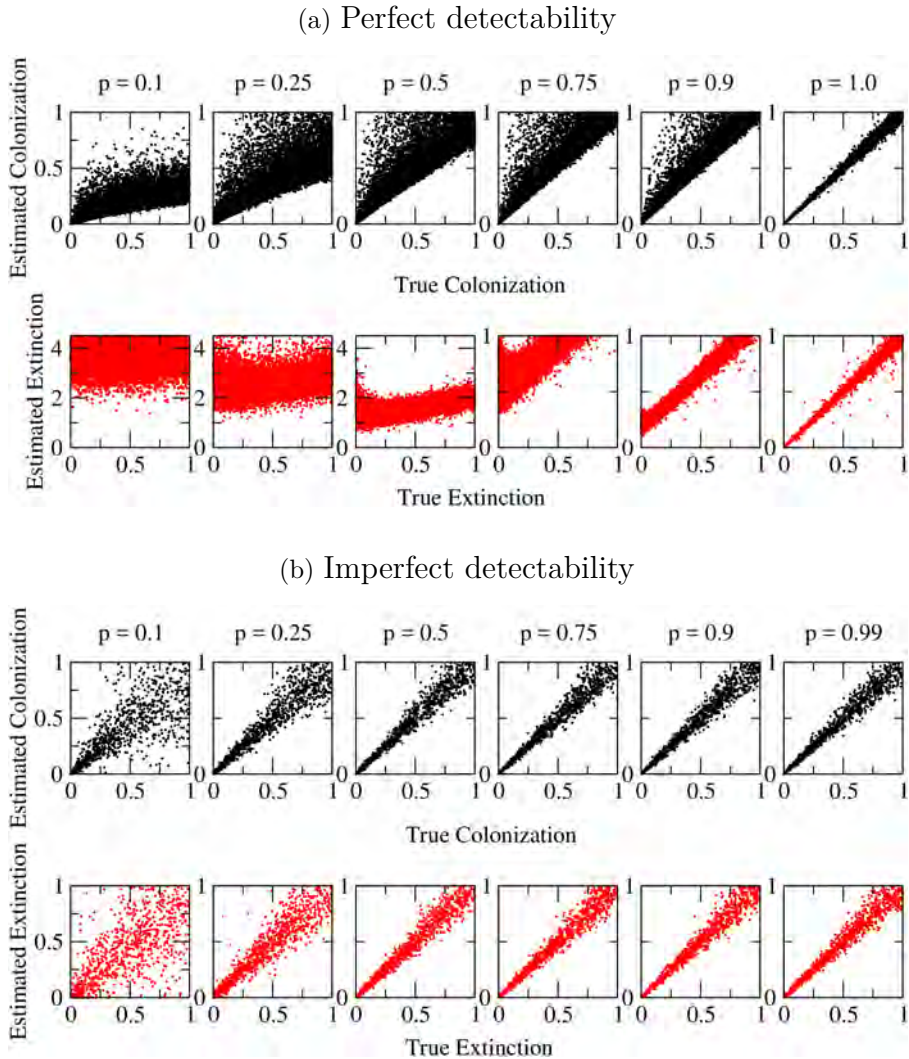


Figure 3.1: **Effect of imperfect detectability on estimated colonization and extinction rates.** We generated colonization-extinction dynamics for communities of 100 species ($S = 100$) for known colonization-extinction (true) values. In (a) we generated $S \times T$ data matrices, where the number of rows ($S = 100$) is interpreted as the number of species in a community undergoing independent colonization-extinction dynamics (or, alternatively, as the number of sites that have been monitored over $T = 100$ sampling times for just one species). True presences were then filtered according to detectability levels to generate the observed data matrices. 10000 different colonization-extinction pairs are represented. In (b) we generated $S \times N$ data matrices, with $S = 100$ and $T = 10$, but 4 different transects per sampling time, so that $N = 40$. Again true presences were filtered according to detectability levels to generate the observed data matrices. 1000 different colonization-extinction pairs are represented.

Table 3.2: **Comparison of seven models for coral reef fishes.** The column k represents the number of model parameters. The first model, with only two parameters (1st row, $k = 2$), considers all species described by the same colonization and extinction rates. In this sense, they are all equivalent, and, therefore, this is the most neutral one. The second model considers two groups (2nd row, $k = 4$), where algal feeders are characterized by a different colonization-extinction pair, and the remaining trophic guilds share the same colonization-extinction rates, and so on, until the last model (7th row, $k = 14$), where every guild is described by a different colonization-extinction pair.

| Model | k | AIC_c | AIC_{dif} | $AIC_{weights}$ |
|---|-----|---------|-------------|----------------------|
| {Alg Cor Mac Mic Omn Pis Zoo} | 2 | 12800 | 127.57 | $1.8 \cdot 10^{-28}$ |
| {Cor Zoo Mac Omn Pis Mic} {Alg} | 4 | 12713 | 40.641 | $1.4 \cdot 10^{-09}$ |
| {Pis Mic Zoo Mac Cor} {Alg} {Omn} | 6 | 12715 | 42.61 | $5.1 \cdot 10^{-10}$ |
| {Mic Zoo Mac Cor} {Pis} {Alg} {Omn} | 8 | 12680 | 7.9 | $1.7 \cdot 10^{-2}$ |
| {Mic Mac Cor} {Zoo} {Pis} {Alg} {Omn} | 10 | 12672 | 0 | 0.91 |
| {Mic Mac} {Cor} {Zoo} {Pis} {Alg} {Omn} | 12 | 12678 | 5.4 | 0.06 |
| {Mic} {Mac} {Cor} {Zoo} {Pis} {Alg} {Omn} | 14 | 12681 | 8.8 | 0.01 |

equivalent in terms of their *effective* colonization-extinction dynamics across the sampling period so they can be characterized by the same colonization-extinction-detectability- Φ_0 model parameters across all groups, or if they need to be grouped differently to optimize the trade off between model simplicity and prediction ability according to Akaike Information Criterion. The function `upgma_model_selection` performs this analysis, with the input argument `PerfectDetectability = FALSE` (see `vignette("detectability")` for more details), yielding, as an output, table 3.2.

3.2 Environmental variability and community dynamics

The influence of environmental variability on community dynamics has been explored using a variety of approaches (Cazelles et al., 2016b; Legendre et al., 1985; Parmesan, 2006). Colonization and extinction rates may depend, in general, on a number of relevant environmental covariates (understanding covariate as an independently-measured variable with a potential influence on the dynamics). The simplest

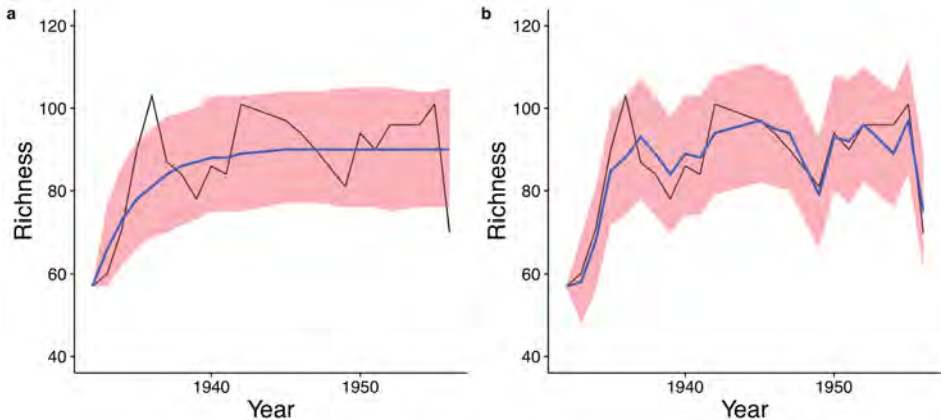


Figure 3.2: **Simulations of a sagebrush steppe plant community in Idaho, USA.** The black line represents the local richness aggregated to all quadrats every sampled year. The blue line stands for the median overall richness averaged over 1000 simulations, and the 95% confidence lines are depicted in red. **a**, These simulations have been calculated with the maximum-likelihood estimated pair of colonization and extinction rates for this dataset. **b**, Here we considered colonization and extinction rates as functions of the environmental covariates selected sequentially maximizing its AIC at each step. The use of the environmental covariates improved substantially the performance of our simulations.

form incorporating environmental factors reduces to assume a linear dependency,

$$c_t = \alpha_0 + \sum_{i=1}^F \alpha_i Y_{it}, \quad e_t = \beta_0 + \sum_{i=1}^F \beta_i Y_{it}, \quad (3.1)$$

Y_{it} being the value of the i -th environmental covariate ($i = 1, \dots, F$) at time t . Note that colonization and extinction rates are once again considered as *effective* parameters since they encompass the mean field response of the community to the environment, regardless of species identity.

We include in `island` a subset of a historical dataset, `idaho`, consisting of a series of permanent 1 m^2 quadrats located on the sagebrush steppe in eastern Idaho, USA, sampled annually from 1923 to 1973 (Zachmann et al., 2010). It also contains records of covariates such as monthly precipitation, mean temperature and snowfall, for which we estimated annual summaries.

We evaluated the influence of the environmental covariates on the colonization and extinction rate of the whole community, between 1932 and 1956, using function `greedy_environmental_fit`, that sequentially selects the environmental covariate that better improves the AIC of the model in each iteration. The model selected included the influence over colonization of snow in November and March and the influence over extinction of snow in May and June and temperature in May, December and April. We simulated the dynamics of the accumulated species richness in the quadrats using these rates. Figure 3.2 shows that the model with environmental variables better approximates observed species richness over time than a single pair of colonization and extinction rates.

We estimated the goodness-of-fit of the environmental-based model in relation to the null expectation yielded by the model driven solely by averaged colonization and extinction rates using the quantity $R^2 = 1 - \epsilon^2/\epsilon_0^2$, where ϵ^2 (ϵ_0^2 , respectively) is the (null) model quadratic error, estimated with `simulated_model`. We get $R^2 = 0.673$ for the model based on environmental covariates relative to the null model. Calculations and model are detailed in the R script `tvjo2of3.R` (see Supplementary Materials S1), and more information can be found in `vignette("island")`.

3.3 Species co-occurrence networks

The inference of interactions from patterns of presence/absence goes back at least to Diamond's 'checkerboard' pattern to infer competitive exclusion (Diamond, 1975). These research efforts continue today (Barberán et al., 2012; Cazelles et al., 2016a; Chase et al., 2011). Here we developed a procedure to test the species-independence assumption. Our method highlights species pairs that co-occur more or less than expected under independent colonization-extinction dynamics —the null hypothesis. These significant pair-wise associations can then be

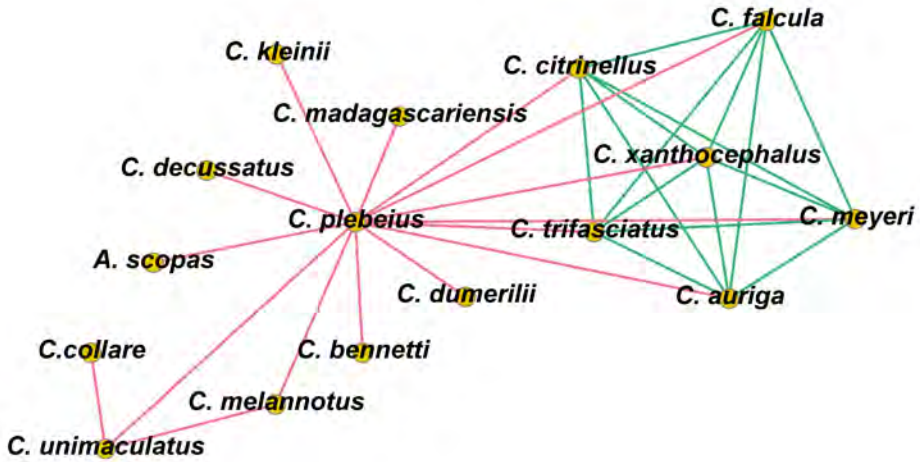


Figure 3.3: Network of significant co-occurrences of corallivorous fishes at the Lakshadweep Archipelago. Pairs of species linked with a red line co-occurred less than expected, while the green line indicates species that co-occurred more than expected according to the null expectation provided by the IBT stochastic model based on colonization and extinction rates characterizing each trophic guild. All species belong to genus *Chaetodon*, except *Cantherhines dumerilii* and *Amaneses scopas*.

used to build a network.

We used the dataset `alonso15`, which describes community re-assembly after a coral mass mortality event in the relatively unfished Lakshadweep Archipelago (Alonso et al., 2015). For simplicity, we only show results for the guild of corallivores. To obtain co-occurrence networks, we first calculated empirical colonization and extinction rates for each island and guild. Second, we simulated presence/absence matrices for the species in the guild based on previously calculated rates. Third, we obtained the distribution of expected co-occurrences given the stochastic colonization-extinction dynamics (computed after a number of model realizations). Finally, we compared actual co-occurrences with the distribution of occurrences across replicates under the null model to determine if observed co-occurrence values were significantly lower or higher (confidence level 2.5% or 97.5%) than null model predictions, respectively. The results show that some species

of corallivores are much more likely to be found together, while others (like *Chaetodon plebius*) tend to exclude others in the guild (Figure 3.3). The file `tvjo3of3.R` provides the code necessary to conduct this analysis (see Supplementary Materials S2).

3.4 Discussion

Long time series data on multi-species assemblages are rich in information but notoriously complex to analyse. Having tools that can robustly explore these rich datasets is critical to testing and advancing theories of community dynamics. In this contribution we present an R package that implements models and maximum-likelihood methods to analyze community dynamics based on the simplest stochastic model underlying IBT. The functions developed in `island` allow for implementation of colonization-extinction models with and without perfect detectability.

Most R packages to analyze community data provide metrics based on statistical approaches, such as diversity indices and ordination techniques (Oksanen et al., 2016), or indices of functional diversity (Laliberté et al., 2014) that can be applied over time (May et al., 2018). Many packages develop methods to specifically deal with and describe temporal community dynamics (Hallett et al., 2016), but few of these methods focus on the potential dynamic processes underlying community assembly over time. Promising recent approaches provide a clear link between models and basic dynamic ecological theory, which also use IBT as a point of departure (Cazelles et al., 2016a,b; Gravel et al., 2011; Massol et al., 2017) —as we have done in `island` package. However, they do not develop maximum-likelihood methods for community time series data analysis in practical settings. More work is needed in this direction.

Our package complements some of the capabilities of the otherwise comprehensive `unmarked` R package (Fiske and Chandler, 2011), which

implements MacKenzie's (2003) colonization-extinction dynamic model (see `colext` function). However, unlike the `unmarked` R package, our basic model parameter estimates are rates (in T^{-1} units) rather than dimensionless probabilities (Fiske and Chandler, 2011, 2015; MacKenzie et al., 2003). As a consequence, methods of `island` R package work well with irregular sampling times, helping address the problem of missing data, which so often constrains the analysis of long-term data series.

The three case studies showcase the use of our methods and functions to analyze whole community dynamics monitored for a period of time across a range of sampling locations (see package vignettes for other applications and alternative examples). Colonization-extinction models to analyze single species over a range of sites—a metapopulation—have been widely used (Hanski, 1994, 1999, 2001; Hanski and Ovaskainen, 2000). By contrast, comprehensive community temporal data sets have been and still are expensive to obtain and, consequently, methods to analyze community dynamics have received much less attention. Such methods are still scarce in spite of an entire body of theory devoted to the study of metacommunities (Holyoak et al., 2005). Our examples analyze data frames that include at least three factors ("species", "sampling location", and "species guild") that can be studied separately with standard model selection techniques, and "sampling time", which is our model dynamic variable. Our model estimates can be considered effective parameters that provide aggregated information reflecting true community dynamics. The idea of an *effective model* has important theoretical and ecological implications. While all models implemented in the packages rely on "the species independence assumption", merely finding good agreement between them and community data does not imply that between-species interactions are unimportant. The success of fit implies instead that all the biotic interactions influencing a focal species can be *effectively* represented by its average model parameters. Species dynamics are only effectively decoupled to achieve a mean-field

description of true community dynamics (Solé et al., 2002). Even if the overall performance of a given model assuming species independence is good, our approach can be used to unveil significant pair-wise species inter-dependencies or associations, as we show in our third case study.

By using a process-based stochastic immigration-birth-death model to generate synthetic data, we show how *effective* colonization and extinction rates can be compared with intrinsic immigration and death rates (Box 1). In this sense, the parameters calculated by our `island` R package provide effective coarse-grain measures of the intrinsic quantities driving processes affecting individuals and their interactions to end up determining community dynamics over time.

Our first worked example highlights the importance of accounting for imperfect detectability in community datasets. We clearly show that maximum likelihood estimation based on perfect detectability can yield highly biased estimates unless we work with detectability per sampling time higher than 0.9. By contrast, models that explicitly account for imperfect detectability produce reliable estimates when detectability per transect is just over 0.25. Ideally, we recommend models that account for imperfect detectability, while recognizing that we often do not have the luxury of sufficiently replicated data sets, limiting us to simpler approaches. In these situations, the potential biases of imperfect detectability should be clearly acknowledged while interpreting model results. As a rule of thumb, with at least 4 replicates per sampling time and a detectability over 0.5, it is possible to use the simpler, more straightforward methods from `island` to estimate almost unbiased colonization-extinction parameters.

Our second worked example shows how environmental variables can be taken into account to better explain community dynamics through effective colonization and extinction rates. Notice that rates were simply written as linear combinations of environmental covariates. Since our methods estimate rates rather than probabilities, we do not need to make use of the *logit* link to consider covariates (Fiske and Chandler,

2011, 2015). In addition, our methods can cope with highly flexible sampling schemes including non-evenly spaced sampling time intervals. Simple model selection techniques can be used to emphasize the relatively stronger influence of a small set of environmental covariates on the dynamics of community-level species richness.

Finally, our third case study uses the community colonization-extinction model, which is based on species independence, as a null working hypothesis to assign a level of significance to species pair-wise co-occurrence. Most robust methods to find when species co-occurrence is significantly higher or lower than random expectations are based on static probabilistic models (Barberán et al. 2012; Griffith et al. 2016; Veech 2013, `cooccur` R package, but see also Cazelles et al. 2016a and Cirtwill and Stouffer 2016). Instead we used a dynamic community assembly model as a null model, where independent species extinctions and colonizations occur at the rates previously estimated from community data. In addition, because we can generate stochastic replicates with the same sampling structure as the given dataset, we do not require the stationarity assumption. As a result, as shown in the resulting co-occurrence network (see Fig. 3.3), where links represent only significant co-occurrences, we identified a group of species that tend to occur together, and a species, *Chaetodon plebeius*, that occurred less than expected with most species in the guild under the null assumption of species-independent colonization-extinction dynamics. This clearly warrants further ecological investigation. This example highlights a novel procedure to investigate species interactions from community data over time, which can go easily unnoticed or biased if only static probabilistic methods are used.

Our understanding of what shapes ecological communities requires a constant conversation between empirical observations and theory. Some of the most basic questions of community assembly such as the stationarity or temporal dynamism of communities have yet to be resolved (Warren et al., 2015). Questions such as these require real

world data to test assumptions and predictions of ecological theory. As valuable multi-species temporal datasets are becoming increasingly available, having adequate tools to explore them is vital. We hope that the methods developed in our `island` R package allow for more detailed analyses of ecological community temporal datasets and an advancement of basic ecological theory.

Supplementary materials

Appendix A. An appendix revisiting Island Biogeography Theory, its fundamental assumptions, its connection to colonization-extinction models, and the simplest likelihood function of package `island`.

S1. R file `tvjo2of3.R` with the code to reproduce worked example "Environmental variability and community dynamics".

S2. R file `tvjo3of3.R` with the code to reproduce worked example "Species co-occurrence networks".

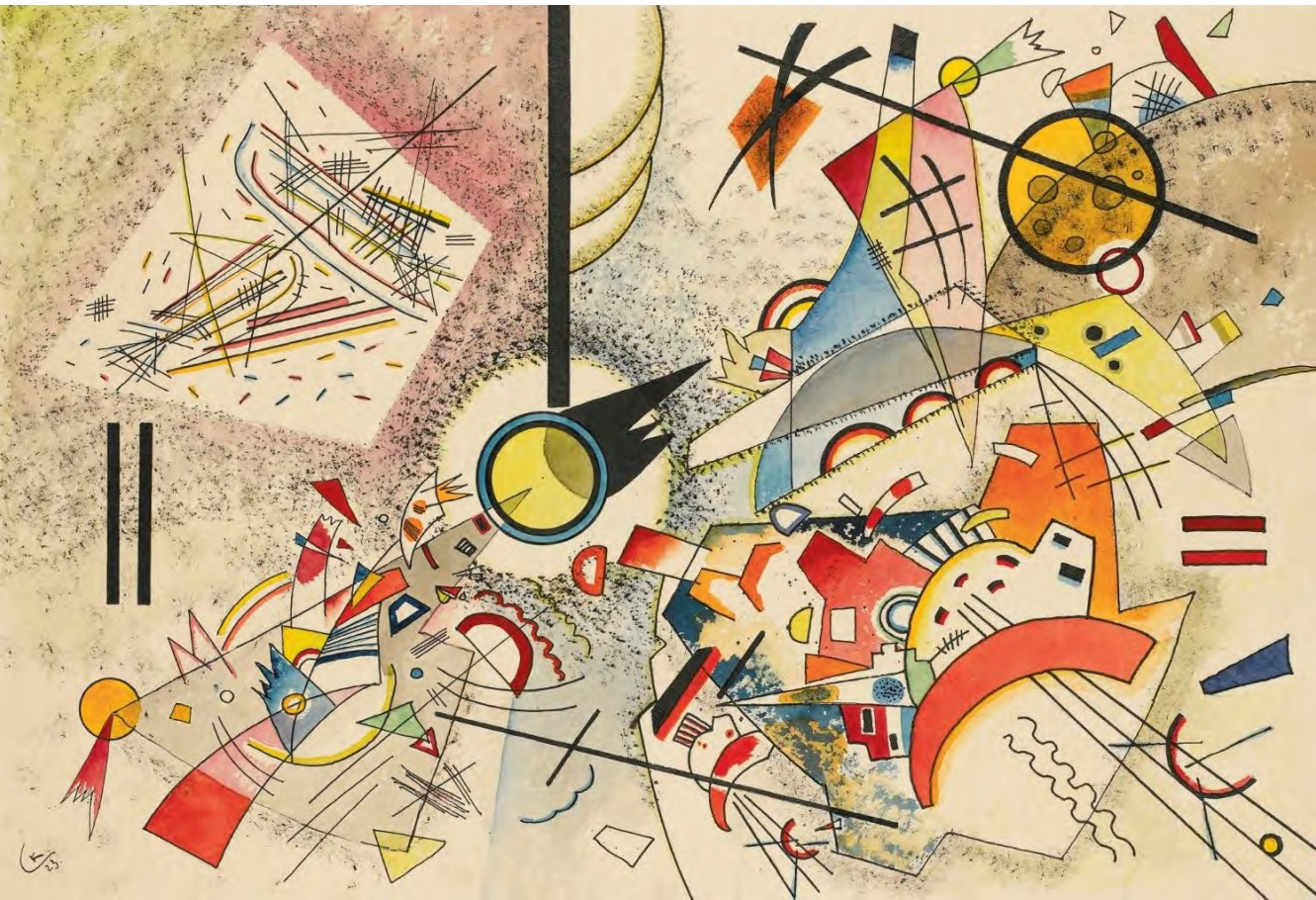


Image:

Wassily Kandinsky. *No title*. 1923, private collection.

This chapter:

Ontiveros, VJ, Capitán, JA, Casamayor, EO, Alonso, D. The characteristic time of ecological communities.

Submitted to *Ecology*.

The characteristic time of ecological communities

Abstract

The temporal dynamics of ecological communities are key to comprehend community assembly and responses to disturbances or improve conservation measures. Although temporal changes develop at multiple scales, summarising the dynamics of species richness and community composition would allow the comparison and understanding of these patterns across ecological communities. We used the simplest stochastic model of Island Biogeography to estimate a community characteristic time, that sets the temporal scale at which a null, non-interacting model acts. We showed its applicability to classic time-series in ecology and found its relation to changes in community composition, defining a characteristic Jaccard Index. More thoroughly sampled communities, relative to characteristic time, presented more similar composition. Additionally, we estimated the characteristic time of different microbial and macroscopic assemblages, with a mix of *a priori* fast and slow communities. These results concurred with current knowledge, but indicate that some habitats have been sampled less frequently than required. Our work makes evident that it is essential to account for time in ecology and highlights the usefulness of simple approximations to the complex dynamics of ecological communities.

A central goal of community ecology is to understand temporal dynamics in species diversity and composition. Temporal dynamics are essential to gain a mechanistic comprehension of the drivers of community assembly, improve conservation and management measures, monitor the effect of disturbances, forecast the consequences of climate change or even design and control microbial communities (Fisher et al., 2010; Gonze et al., 2018; Supp and

Ernest, 2014; Warren et al., 2015). This knowledge is still limited (Dornelas et al., 2014; Hortal et al., 2015), although its importance in ecology has been increasingly acknowledged (Dornelas et al., 2013; Levin, 1992; White et al., 2006). A central issue is to establish the temporal scale at which communities function. Previous research has established that temporal changes develop at multiple scales, as biotic and abiotic drivers of community change interact in different complex, scale-dependent ways (Levin, 1992). However, community ecologists still try to describe temporal dynamics with simple measures that summarize the inherent dynamics of ecological communities. These measures may help us address pressing questions in ecology, such as the search for generalities in the dynamics of ecological communities or the identification of baselines for detecting ecological changes (Sutherland et al., 2013).

Although the importance of temporal dynamics was hinted already by Darwin (Magurran, 2008), it can be considered that Preston was the first to study the temporal scale. Preston introduced the concept of Species-Time Relationships (Preston, 1960), the relationship between time and the number of species we find in a sample as time increases. Other measures of richness or composition change have been used since then, as species turnover (Diamond and May, 1977; Hallett et al., 2016; Hillebrand et al., 2018; Russell et al., 1995) or the slope of the Jaccard index (or other similarity indices) against time (Dornelas et al., 2014). However, some of these works (Diamond and May, 1977; Dornelas et al., 2014; Russell et al., 1995) also try to compare turnover or community similarity with models rooted in the Theory of Island Biogeography (TIB).

The TIB of MacArthur and Wilson (1967) can be considered as the first dynamic mechanistic attempt to understand community ecology (May, 2010). TIB focus on how a dynamic equilibrium between colonization and extinction determines both the average number of species at equilibrium and community temporal turnover even after

this equilibrium has been attained (Simberloff, 1969). Afterwards, Diamond and May (1977) built on TIB a species-specific model of turnover as a function of census interval to overcome the *uncertainty principle* associated with the choice of time interval between censuses, as censuses apart in time might underestimate repeated colonization and extinction events. Further exploration of a similar community turnover model indicates that a simple measure of mean turnover is not enough to allow comparisons among communities (Russell et al., 1995). Another attempt developed a null model based on random colonizations and extinctions to compare the observed slopes of the Jaccard index against time with the ones obtained for that null model (Dornelas et al., 2014). However, this random null model was not fitted to the data. Outside of TIB, other dynamic models of diversity try to understand temporal change in ecological communities, based on either niche (Chisholm et al., 2014) or neutral (Allen and Savage, 2007; Azaele et al., 2006; Kalyuzhny et al., 2015) community dynamics, and require variation in species abundances. However, long temporal data on species abundances across whole communities are expensive to obtain and maintain (but cf. Condit 1998; Dornelas et al. 2018). Besides, when we only have accounts of presence and absence, these methods do not apply.

The main aim of this study is to investigate the potential of a simple stochastic model (Alonso et al., 2015; Ontiveros et al., 2019) to characterize temporal scales of species richness change and community dynamics. We make use of the stochastic formulation of this theory to define two measures of temporal community change: (1) the community characteristic time, and (2) the characteristic Jaccard index. Both measures are based on simple colonization-extinction models that only need presence-absence data. The first one is a characteristic time that, in practice, can be used to inform about the optimal frequency at which communities should be sampled, solving the *uncertainty principle*. The second one, which is based on the Jaccard index, is a measure of

community composition change over time. Therefore, it can be regarded as an alternative measure of stability or temporal beta diversity. Our methods allow us to characterize and compare temporal community dynamics across taxa and habitats. In the following paragraphs, we define these metrics and examine the ability of the model to recover community dynamics, richness, and composition temporal patterns, applying it to different classic datasets. Then, we explore the relation of the observed patterns with sampling frequency and finish with estimates of characteristic time for an array of different communities, including both macroscopic and microbial taxa.

4.1 Materials and Methods

4.1.1 Data samples

For our study, we have selected several datasets spanning from micro- to macro-organisms, aiming to represent a wide range of communities and including *a priori* fast- and slow-changing communities. Among them, we have included three classic datasets that studied temporal aspects of ecological communities originally. Preston (1960) used Neotoma birds to propose the STR, Simberloff and Wilson (1969) validated the TIB dynamics with island arthropods, and Diamond and May (1977) studied turnover rates for the Farne Islands birds. For details on the communities studied, see Table 1.

4.1.2 Community characteristic time

We have applied the simplest stochastic model of TIB (Simberloff, 1969; Simberloff and Wilson, 1969) as a dynamic model that explains the variation in richness in a study site, estimating the colonization and extinction processes in that place. It has been shown that it is possible to calculate colonization and extinction rates for whole communities easily, assuming the independence and equivalence of the species (Alonso et al., 2015). These temporal rates give us an

Table 4.1: **Datasets used in this study.** Duration expressed in years. Different sites were considered separately. The two hydroperiods of the shallow saline lakes were considered separately, as the lagoons dried in Summer, restarting the dynamics. ¹ available in GenBank.

| Habitat | Taxa | Location | Duration | Reference |
|----------------------|------------|--------------------------------|-----------|-------------------------------|
| Deciduous forest | Birds | Neotoma, Ohio, USA | 17 | Preston (1960) |
| Island | Birds | Farne Islands, England | 28 | Diamond and May (1977) |
| Marine | Bacteria | English Channel | 6 | Gilbert et al. (2012) |
| Alpine lakes | Bacteria | Pyrenees, Spain | 1 | PRJNA566370 ¹ |
| Shallow saline lakes | Bacteria | Monegros desert, Spain | 2 | Triadó-Margarit et al. (2019) |
| Aeroplankton | Bacteria | Pyrenees, Spain | 7 | Cáliz et al. (2018) |
| Soils | Bacteria | Switzerland | 4 | Hartmann et al. (2014) |
| Islands | Arthropods | Florida Keys, USA | 1.5 | Simberloff and Wilson (1969) |
| Sagebrush steppe | Plants | Idaho, USA | 50 | Zachmann et al. (2010) |
| Coral reef | Fishes | Lakshadweep Archipelago, India | 11 | Alonso et al. (2015) |
| Human | Bacteria | – | 0.5 – 1.5 | Caporaso et al. (2011) |

approximation to the diversity dynamics in the site. Through these rates, we can, therefore, define the characteristic time of a community, a measure that gives us the temporal scale at which species richness and composition vary in that community (see Box 1 for details). These rates also allow us to simulate the dynamics of the communities. We have done so for three classic examples of community dynamics, Neotoma birds, Farne Islands birds, and Florida Keys arthropods. For Neotoma, we also studied the STR that our simulations produced. All estimates have been obtained using the function `irregular_single_dataset` of package 'island' (Ontiveros et al., 2019), which allows calculating colonization and extinction rates for irregular sampling schemes, and simulations were performed using function `PA_simulation`.

Box 1. Estimating temporal scales with the Theory of Island Biogeography

Theory of Island Biogeography. In the 60s, MacArthur and Wilson proposed the Theory of Island Biogeography with the ultimate goal of explaining the geographical distribution of diversity (MacArthur and Wilson, 1967). The basic equation underlying TIB can be expressed as:

$$\frac{dS}{dt} = c(S_P - S) - eS \quad (4.1)$$

where S stands for the number of species in a site, S_P for the number of species in the regional pool, and c/e for the colonization/extinction rates, which in the original theory were related to distance to the mainland and area of the island respectively. The rates represent the proportion of sites where species colonize or undergo extinction per unit of time, but as we are considering only one site, the rates indicate the pace at which species enter or leave the site, or in the specific case of microbial communities, how the species rise or fall over the detection limit. Estimation of the rates is detailed in Appendix B, following Alonso et al. (2015) and Ontiveros et al. (2019).

Characteristic time. The model formulation shows clearly that its temporal evolution is governed by $c + e$ (see Eq. B.6). Therefore, we can define a temporal scale at which diversity changes, that we denominate *characteristic time*, T_c , and its confidence interval length, ΔT_c , with the following equations:

$$T_c = \frac{1}{c + e} \quad (4.2)$$

$$\Delta T_c = \frac{|\Delta c| + |\Delta e|}{2(c + e)^2}, \quad (4.3)$$

where Δc (Δe) corresponds to the confidence interval of c (e). As shown in Alonso et al. (2015), in the case of regular sampling schemes, characteristic time depends directly on the time

interval among samples, Δt , as well as on the colonization and extinction transition probabilities (see Eq. B.24 in Appendix B). With all these elements, we define the number of samples in a characteristic time as ν_s , the *relative sampling frequency* (in units of characteristic time):

$$\nu_s = \frac{T_c}{\Delta t} \quad (4.4)$$

ν_s can be used as an indicator of how well we are sampling in order to estimate the dynamics of the community.

4.1.3 Temporal β diversity: characteristic Jaccard index

As a measure of β diversity or compositional change, we used the Jaccard index, which accounts for richness but is invariant to abundance, being the perfect match for our measures. We also explored turnover (Diamond and Marshall, 1976; Russell et al., 1995), as it is an almost complementary measure to Jaccard's index. We define both measures and find approximate expressions for their evolution over time in Box 2. We examined the evolution in time for these measures in the classic data of the Farne Islands (turnover) and Florida Keys (Jaccard index).

Additionally, we tested how well our measures of temporal scale related to compositional change. Compositional change not only depends on characteristic time but also on its relationship with the mean time among samples, i.e., how often we sample, so following Box 1, we calculated ν_s , the *relative sampling frequency*. We calculated the mean Jaccard index between consecutive samples for each site in two habitats with enough replicates: arthropod communities in mangrove islands and shallow saline lakes from the Monegros desert.

Box 2. Community composition dynamics: Characteristic Jaccard Index

One of the most used indices to study the change in community composition is the Jaccard index. This index let us compare sites using presence-absence data, and it is defined as follows:

$$\mathcal{J}_{ij} = \frac{C}{A + B + C} \quad (4.5)$$

where A is the number of species present at time i and not at time j , B the number of species present at time j and not at i , and C the number of species present at both times. Now, let p_i be the proportion of species present at time i and Δt the interval of time between samples i and j . As shown in Appendix B, we can approximate the variation of the Jaccard index, for a relatively high number of species, in terms of colonization and extinction rates as:

$$\mathcal{J}_i(\Delta t) \approx \frac{p_i [c + e \exp(-(e + c)\Delta t)]}{p_i [e + c \exp(-(e + c)\Delta t)] + c [1 - \exp(-(e + c)\Delta t)]} \quad (4.6)$$

Therefore, the characteristic Jaccard index would correspond to the value of the expression above when $\Delta t = T_c$.

Eq. 4.6 also hints that the Jaccard index reaches an asymptote after some time. Therefore, we have defined an asymptotic Jaccard index, $\tilde{\mathcal{J}}^*$, starting from the occupancy at equilibrium:

$$\tilde{\mathcal{J}}^* = \lim_{\Delta t \rightarrow \infty} \mathcal{J}^*(\Delta t) \approx \frac{c}{c + 2e}. \quad (4.7)$$

We calculated bounds of the ratio $\frac{\Delta t}{T_c} = \frac{1}{\nu_s}$ for the system to reach approximately the asymptotic Jaccard index (see Appendix B). Specifically, for $0.22 \leq \frac{e}{c} \leq 2.31$, the relative error $|\mathcal{J}^*(\Delta t) - \tilde{\mathcal{J}}^*|/\tilde{\mathcal{J}}^*$ equals 10^{-2} if Δt satisfies the bounds $3.6 \leq \frac{\Delta t}{T_c} \leq 5.6$. This means that, for a wide range of extinction to colonization ratios, the system takes from about 4 to 5 characteristic times to

reach the asymptotic compositional state starting from an initial proportion of species $p_i = \frac{c}{c+e}$.

The same reasoning can be extended to turnover, \mathcal{T} , as defined by Diamond and May (1977). Translating this definition to the same components A , B and C as above, we have that

$$\mathcal{T}_{ij} = \frac{A + B}{A + B + 2C} \quad (4.8)$$

from which we obtain a similar expression for the evolution of turnover over time:

$$\mathcal{T}_i(\Delta t) \approx \frac{[\exp(\Delta t(e + c)) - 1] [(1 - p_i)c + p_i e]}{c [\exp(\Delta t(c + e))(p_i + 1) + p_i - 1] + p_i e [\exp(\Delta t(c + e)) + 1]} \quad (4.9)$$

The derivation of these and additional metrics can be found in Appendix B.

4.2 Results

The first question we asked is whether the model can reproduce, at least qualitatively, the temporal dynamics of ecological communities. To answer that question, we revisited three classic studies on community dynamics, namely: arthropod recolonization in island E3 of the Florida Keys, Farne Island birds, and *Neotoma* birds. First, we estimated colonization and extinction rates in the three communities, obtaining colonization rates of 5.64×10^{-3} , 2.04×10^{-4} and $7.47 \times 10^{-4} \text{ day}^{-1}$, and extinction rates of 1.14×10^{-2} , 4.10×10^{-4} , $2.84 \times 10^{-4} \text{ day}^{-1}$, for arthropods, island birds and deciduous forest (*Neotoma*) birds, respectively. These rates allowed us to simulate the dynamics of the communities. In figures 4.1 and B.1, we show that the observed richness is inside the 95% confidence interval of the simulations in the arthropods

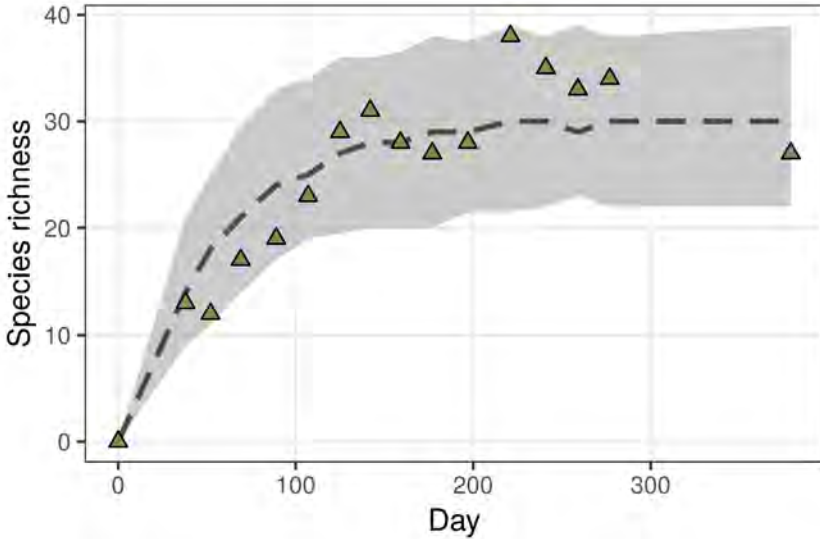


Figure 4.1: **Temporal dynamics of the arthropods community in the Florida Keys.** Shaded areas correspond to the 95% confidence interval in model simulations, whereas the dashed line corresponds to the median. Green triangles correspond to the observations. Here the characteristic time is estimated as $T_c \approx 59$ days.

community (see Appendix B for the performance of our model for bird communities).

Now, we turn to richness and community composition patterns. Figure 4.2 presents the different patterns studied for three classic datasets. We satisfactorily recovered the STR for Neotoma birds, turnover over time for the transient species in the Farne Islands, and the evolution of the Jaccard index over time for island E3 of the Florida Keys. Only small deviations were found. We also calculated the characteristic Jaccard index for island E3, that is, the expected Jaccard index after a characteristic time, yielding a value of 0.406. Simulations also validate our approach to estimating the variation of the Jaccard index over time, as can be seen in figure B.2.

In order to show that we can determine an average time scale at which communities function, we examined the relationship between characteristic time and community composition (Figure 4.3). We found

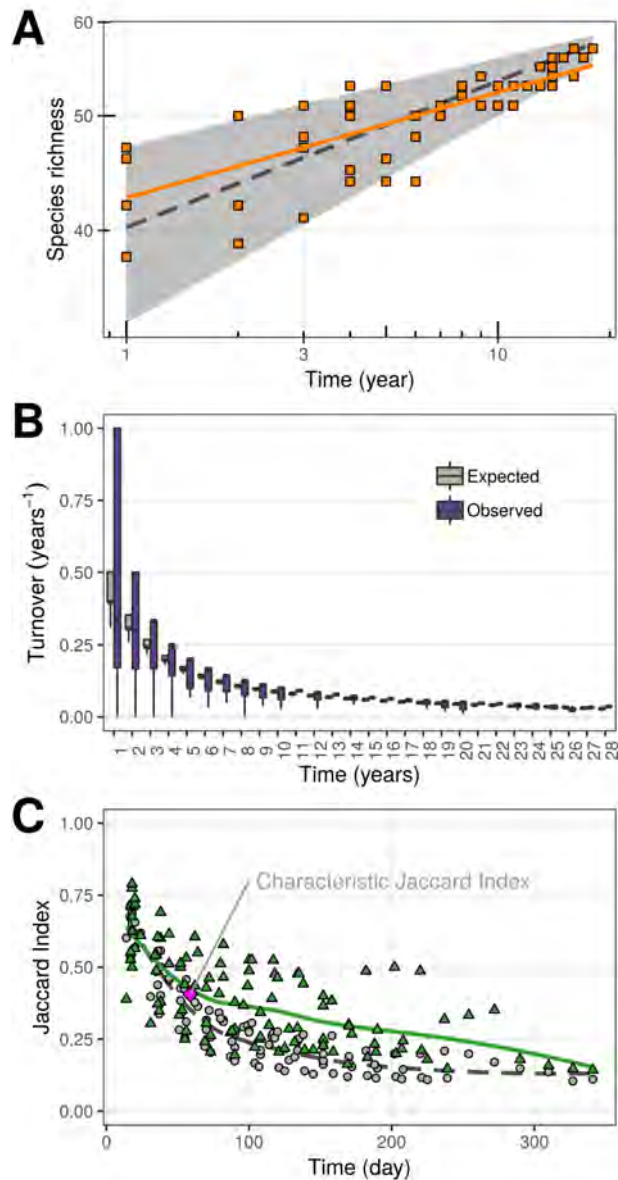


Figure 4.2: **Richness and community composition patterns recovered by a simple stochastic model of island biogeography.** A) Species Time Relationship of deciduous forest birds community at Neotoma. B) Turnover rate of island birds in the Farne Islands. C) Community similarity measured with the Jaccard index for arthropods in Florida Keys. Grey elements (shaded areas, boxplots, or circles) correspond to model predictions. Colored elements correspond to the observed measures.

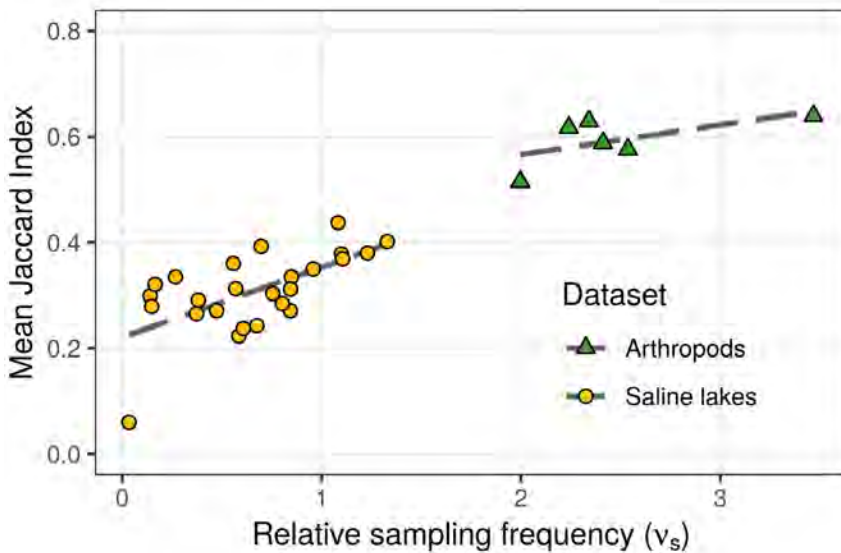


Figure 4.3: **Variation in community composition for multiple sites and two habitats.** Arthropod communities showed higher values of the mean of Jaccard indices between consecutive samples than shallow saline lakes, as the former were sampled more thoroughly.

higher values of the Jaccard index for arthropod communities (0.50 – 0.65) than for Monegros shallow saline lakes (0.05 – 0.45). Additionally, we found that Jaccard’s index increased with the relative sampling frequency ν_s (how often we sampled the community compared to the characteristic time of the system), as demonstrated by their Spearman’s ρ ($\rho = 0.976$, p-value < 0.001 for saline lakes; $\rho = 0.486$, p-value = 0.3556 for arthropods —statistically not significant because of small sample sizes).

Finally, we examined the temporal scales of different communities estimating their colonization/extinction rates and their associated characteristic times. Among microbial communities, soils had the slowest dynamics with a characteristic time of above seven months, while the communities found in humans, especially those on hands, had a faster characteristic time —about a few days. The macroscopic communities presented longer characteristic times of years, except the macroinvertebrates that yielded several weeks as the community

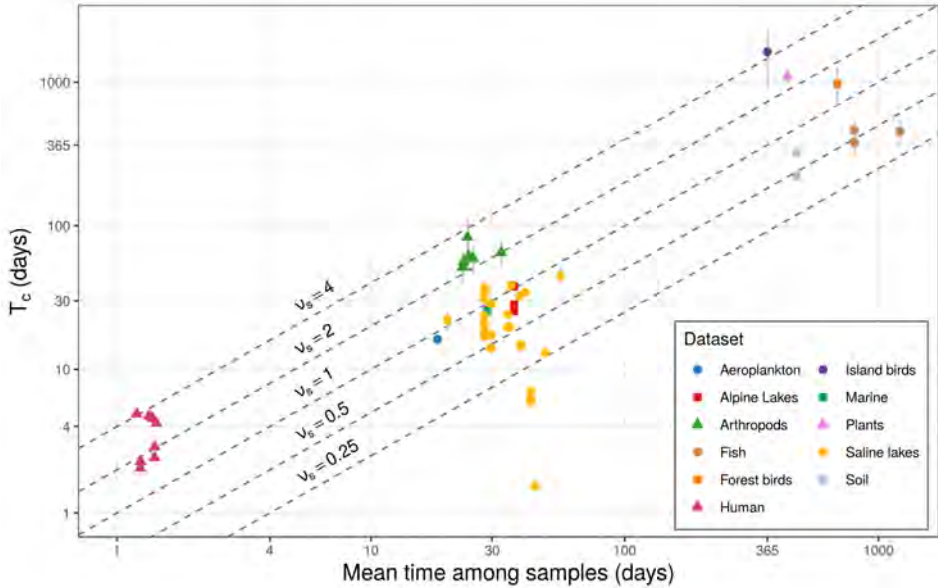


Figure 4.4: **The characteristic time of different communities.** We found that birds and plants had the slowest dynamics, while the fastest habitat is found in humans, especially in hands. Each point represents a site, and errorbars (gray) indicate the characteristic time error estimation. Along the dotted lines, the relative sampling frequency is constant.

characteristic time (Figure 4.4). Among the eleven studied datasets, five of them (humans, invertebrates, plants, birds) were sampled more frequently than the characteristic time indicated, while coral reef fishes, soil microbes and some saline lakes were slightly undersampled (Figure 4.4).

4.3 Discussion

The current study has found that the simplest stochastic model of Island Biogeography can recover richness and composition patterns such as the STR, turnover, or Jaccard index. Moreover, we show that the observed patterns depended on an inherent temporal scale, that we have called characteristic time. This measure may not only help us compare compositional change among different communities

through the characteristic Jaccard index, but it also aids in assessing whether we are sampling a community frequently enough not to miss its dynamics, solving the *uncertainty principle* effectively.

However, we are conscious of exceptions and limitations in our approach. A note of caution is due in the case of the estimation of turnover in the Farne Islands. When we considered all species in the community, our estimates of turnover were higher than the observed/predicted ones (Diamond and May, 1977). We attribute this discrepancy to four species that were always present in the community, whereas the rest of the species were occasional, effectively violating the assumption of *species equivalence*. Excluding these species solved this discrepancy, although it left a small number of species in each sampling, producing fluctuations that account for the observed deviations. As a general rule, we advise caution with the model when its two main assumptions, *species equivalence* and *independence*, are violated, and recommend its application preferentially to horizontal communities (Loreau, 2010; Vellend, 2010). Mild deviations from these assumptions, like the ones reported for the arthropod data previously (Cirtwill and Stouffer, 2016), are unlikely to affect much the estimates as can be observed here. Thus, the model acts as an *effective model* that integrates mild effects of interactions and niche differences in community dynamics. We have identified two additional issues when estimating characteristic times. The first one arises when we sample much less than characteristic times indicate. It might be happening for the less thoroughly sampled hydroperiods of shallow saline lakes. The second one evidences that characteristic time and temporal autocorrelation are associated, and consist in a linear relationship between characteristic time and mean inter-sample time, that we have observed for bacterial communities in humans (Caporaso et al., 2011). We have devised procedures to identify these cases, as detailed in Appendix B.

Our results coincide with previous studies in several ways. First, we recovered richness and composition patterns, such as Preston's original

STR (Preston, 1960), through simulation. Second, our approximation to turnover is analogous to that found in a previous study (Russell et al., 1995). Third, our estimates of temporal scales for different communities coincide with general intuition. For example, we found slower dynamics for bacterial communities in soil than in aquatic environments, or birds compared with arthropods. Also, the characteristic time for the marine microbial community accords with the corresponding estimates in a previous study (Benincà et al., 2008). However, we also identified a contrasting result. A previous meta-analysis found that the dynamics of lacustrine assemblages is faster than the dynamics of marine ones for macroscopic species (Korhonen et al., 2010), whereas we observed the contrary in microbial communities. Further work is needed to discern if characteristic times differentiate microbial communities from those of larger organisms. In any case, the lack of estimates of timescales in ecology is outstanding, which may reflect unintentional biases towards evident natural cycles (i.e., seasonal or annual) or merely the lack of appropriate information on temporal trends of species richness (Hortal et al., 2015) due to long-standing institutional disincentives (Wolfe et al., 1987). This circumstance, applied to organisms that are inconspicuous but increasingly accessible, such as microorganisms, may prevent adequate knowledge of the drivers of ecosystem functioning (Shade and Gilbert, 2015). In this sense, we think that the knowledge of the temporal dynamics of microbial communities might benefit from more frequent samplings.

Our study may help to address several aspects of interest. The most evident of them is the resolution of the *uncertainty principle*. The estimation of characteristic times across habitats and taxa might allow us to study changes in community composition and richness at appropriate timescales, for which theoretical frameworks are currently lacking (Hastings, 2010). Although characteristic times can not be estimated *a priori*, wide application of the method might help ecologists to establish rules of thumb derived from our measures. Besides, char-

acteristic times are intimately associated with patterns of β -diversity. We propose the characteristic Jaccard index as a measure to compare horizontal communities (cf. Vellend 2016), in contrast with previous studies that indicate that single measures are not enough to characterize composition dynamics (Russell et al., 1995). At the very least, this simple stochastic model of TIB can be used as a null model of community dynamics, estimated from presence–absence data, representing a baseline for detecting ecological change (Sutherland et al., 2013). Such a null model, improving previous attempts that do not fit the parameters to the data (as in Dornelas et al. 2014), is interesting for monitoring and conservation, as it portrays random drift in community richness and composition (Hillebrand et al., 2018). Deviations from it may indicate departures of equivalence, independence, and/or the influence of abiotic factors on the community. The influence of abiotic factors can also be modeled as previously demonstrated (Ontiveros et al., 2019), which potentially could help predict community dynamics under climate change, a much-needed venue of research (Dornelas et al., 2013; Fisher et al., 2010). Our model may also capture the effect of disturbances. Defaunation experiments (Simberloff, 1969) and simulations (not shown) indicate that, roughly after four characteristic times, a completely defaunated island reached equilibrium in richness (see Figure 4.1). Concerning community composition, a similar time is needed to approach the asymptotical baseline of the Jaccard index, as shown in Box 1 (see also Figures 4.2C and B.2).

Temporal scales in ecology are seldom studied nor characterized. In this chapter, we have shown the usefulness of a stochastic model of Island Biogeography and derived measures of timescale and compositional change to recover richness and community dynamics. This theoretical work sheds light on the inherent temporal scales of ecological communities. Further work is needed to understand the temporal aspects of different communities and habitats, and whether our estimates hold as additional datasets are analyzed. Moreover, we expect that the

approximations of community compositional change we have developed would allow us to compare it across habitats and groups, searching for generalities. We hope that putting the focus into dynamic patterns of community ecology will aid in improving our understanding of the processes and forces that drive diversity in any kind of community.

Supplementary material

Additional tables and figures can be found in Appendix B.

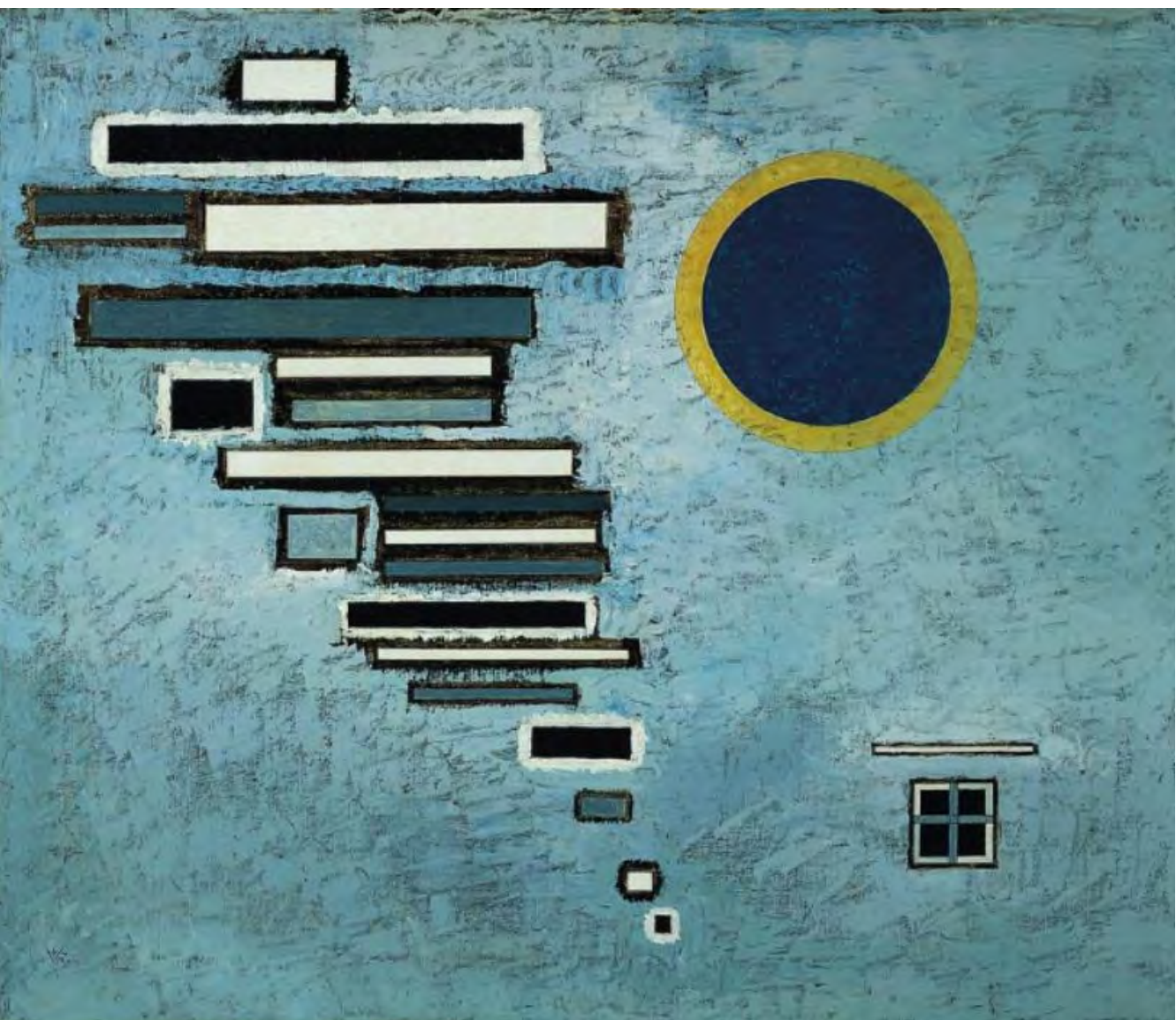


Image:

Wassily Kandinsky. *Unequal*. 1932, Norton Simon Museum, Pasadena, California.

This chapter:

Ontiveros, VJ, Capitán, JA, Casamayor, EO, Alonso, D. Colonization-persistence trade-offs in natural microbial communities.

Under review in *Ecology Letters*.

Colonization – persistence trade-offs in natural microbial communities

Abstract

Fitness equalizing mechanisms, such as trade-offs, are recognized as one of the main factors promoting species coexistence in community ecology. However, they have rarely been explored in microbial communities. Although microbial communities are highly diverse, the coexistence of their multiple taxa is largely attributed to niche differences and high dispersal rates, following the principle “everything is everywhere, but the environment selects”. We use a dynamical stochastic model based on the Theory of Island Biogeography to study highly diverse microbial communities over time across three different systems (soils, alpine lakes, and shallow saline lakes). Here we report for the first time a colonization-persistence trade-off in natural microbial communities. We conclude that this trade-off is mainly driven by rare taxa, which are occasional and more likely to follow independent colonization/extinction dynamics. Our work also emphasizes the fundamental value of dynamical models for understanding temporal patterns and processes in highly diverse communities.

Contemporary coexistence theory indicates that there are two major classes of mechanisms that promote coexistence: *stabilizing* mechanisms that increase negative frequency-dependent selection, and *equalizing* mechanisms that reduce fitness differences among species (Chesson, 2000). Stabilizing mechanisms comprise resource partitioning, disease, or storage effects (HilleRisLambers et al., 2012), while equalizing mechanisms such as trade-offs are generally

produced by life-history trait evolution in a context of historical contingency (Hubbell, 2001). Although many examples of trade-offs can be found in macroscopic communities (e.g., Connell 1961; Siepielski et al. 2010; Werner and McPeck 1994), few have been shown for microbes, usually in experimental metacommunities (e.g., Cadotte et al. 2006; Livingston et al. 2012). To the best of our knowledge, equalizing mechanisms in natural microbial communities have not been carefully evaluated yet. In this paper, our goal is to examine the role of a colonization – persistence trade-off in promoting coexistence across natural microbial communities in both terrestrial and aquatic ecosystems.

The coexistence of a high number of species is a recurring theme in ecology (Hutchinson, 1959). Despite the key insight of Chesson (2000), ecologists are unable to predict species coexistence in an open area (Sutherland et al., 2013). Metacommunity ecology (Holyoak et al., 2005; Leibold and Chase, 2017) tries to understand species coexistence and biodiversity, recognizing the importance of scale and spatio-temporal processes. Metacommunity ecology is characterized by four distinct archetypes: species sorting (SS), which focuses on how local environmental conditions enable some species to coexist; neutral theory (NT), which centers on dispersal limitation and demographic stochasticity; patch dynamics (PD), which concentrates on the balance of colonization and extinction processes in relatively homogeneous patches; and mass effects (ME), which emphasizes that dispersal may outweigh competitive forces in a set of heterogeneous patches. Adler et al. (2007) relate SS and NT with stabilizing and equalizing mechanisms, respectively. SS is related to niche differences, while in NT, dispersal limitation and stochasticity associated to demographic processes override fitness differences resulting in equalization. In the PD archetype, species diversity is maintained by equalizing mechanisms, such as trade-offs in colonization and competitive ability (Calcagno et al., 2006; Solé et al., 2004; Tilman, 1994), or survival/fecundity and competition (Chave et al., 2002; Muller-Landau, 2010).

Microbial communities are highly diverse, and their dynamics have been explained traditionally with the principle "everything is everywhere, but the environment selects" (Barberán et al., 2014; Becking, 1934). Consequently, microbial diversity is usually understood appealing to the formation of highly interacting microbial associations maintained by niche differences, thus emphasizing that stabilizing mechanisms underlie microbial coexistence. This interpretation of the principle neglects the effects of dispersal (Barberán et al., 2014; Nemergut et al., 2013), a potential equalizing mechanism. Classical microbial ecology has made almost no mention of coexistence-promoting mechanisms when analyzing microbial communities. However, microbial ecologists have recently started to talk in terms of general theoretical frameworks in ecology, such as community assembly and metacommunity ecology (Costello et al., 2012; Nemergut et al., 2013). In fact, there has been considerable debate on whether SS or NT dominates as an assembly mechanism in microbial communities, with a somewhat inconclusive result (Lee et al., 2013; Van der Gucht et al., 2007; Woodcock et al., 2007).

Interestingly, it has been conjectured that the relative importance of assembly mechanisms might differ for distinct components of the microbial communities (Langenheder and Székely, 2011; Lindström and Langenheder, 2012). Along similar lines, Hanski (1982) already proposed the *core-satellite hypothesis*, framed within the PD archetype, where stochastic variation in colonization and extinction rates leads to species falling into two distinct categories: core species, abundant and persistent, and satellite species, occasional and rare. Magurran and Henderson (2003) extended the relevance of the core-satellite hypothesis into the temporal domain, finding that core species display a species abundance distribution compatible with a log-normal distribution, while satellite species follow a log-series. These differences were associated with distinct functional roles for these two components. Microbial ecologists have also identified core and satellite species (van der

Gast et al., 2011). Thus, the maintenance of species coexistence in highly dynamic communities, such as the microbial ones, should not be constrained to a single dominant mechanism.

Here we suggest that equalizing mechanisms of coexistence in microbial communities are more important than currently acknowledged and that the relative importance of stabilizing *vs.* equalizing mechanisms is different in core and satellite taxa. In this article, we first validate the use of the simplest stochastic model underlying island biogeography to estimate colonization and extinction rates from temporal series of microbial metacommunities. Then we report a novel colonization – persistence trade-off characterizing these metacommunities coherently at different taxonomical levels. Moreover, we found that this trade-off is mainly driven by satellite species, the rare component of the community. The identification of core and satellite taxa allowed us to conclude that the relative influence of coexistence promoting mechanisms is different for these two components. Recognizing the importance of equalizing mechanisms may render a better understanding of the functioning of highly diverse microbial communities.

5.1 Materials and Methods

5.1.1 Data samples

We analyzed temporal samples from (*i*) the water column of four high altitude lakes in the Spanish Pyrenees, monthly followed during one year (Auguet et al., 2011, 2012), (*ii*) 12 shallow saline lakes in the Spanish Monegros desert plateau, monthly sampled along three years and covering different dry-wet periods (Triadó-Margarit et al., 2019), and (*iii*) two sites in Switzerland, after a soil compaction experiment lasting four years (Hartmann et al., 2014). Microbial communities were studied after NGS 16S rRNA amplicons analyses, clustered at 97%

OTU identity, and transformed to presence-absence data. Sequences processing and genetic data analyses were carried out as reported in the original studies where additional ecological and environmental information are also available. The complete genetic datasets are available in GenBank under BioProject record IDs PRJNA566370 (Pyrenean lakes), PRJNA429605 (Monegros), and as supplemental material for the Swiss soils (Hartmann et al., 2014).

5.1.2 Colonization and extinction rates

Throughout this work, we applied the simplest stochastic model underlying TIB (Alonso et al., 2015; Simberloff, 1969; Simberloff and Wilson, 1969). This dynamic model explains the average level of richness and its variation in a study site (or *island*) in terms of colonization and extinction processes, on the one hand, and the total number of potentially colonizing species in the regional pool, or metacommunity richness, on the other hand. As Hanski (2001) showed, this model can be derived from an ensemble of single-species models of presence-absence dynamics, under the assumptions of both species independence and equivalence (Alonso et al., 2015). So, we can estimate the model parameters for the dynamics of the whole community from presence-absence temporal data, and, therefore, we characterize the entire microbial community by a single colonization-extinction pair. Alternatively, we can subdivide the community in guilds, relaxing the equivalence assumption, and estimate a distinct and characteristic colonization-extinction pair for each of them (Alonso et al., 2015; Ontiveros et al., 2019). To calculate colonization (c) and extinction (e) rates from the observed presences and absences, as the microbial communities were sampled following an *irregular sampling scheme* (samples separated by unequal time intervals), we used the functions `irregular_single_dataset` and `irregular_multiple_datasets` from R package ‘*island*’ (Ontiveros et al., 2019).

As a way to assess the applicability of the method to microbial

communities, we started estimating colonization-extinction rates for several independent sites. For the Pyrenees dataset, we compared three lakes from the same basin (Lakes Llebre, Llong and Redó d’Aiguestortes; Auguet et al. 2011) and one in a different basin (Lake Redon; Auguet et al. 2012). We followed a model selection procedure, based on the Akaike Information Criterion and the weight of evidence, w_i (Burnham and Anderson, 2002), to develop a series of models with different sets of partitions of the four lakes, and estimate, for each of these partitions, a pair of colonization and extinction rates. Besides, we used data from the Swiss soils to test the precision of the method when confronted with replicates of the same community. Once we assessed the correct performance of the method, we subdivided whole communities into different taxonomic levels, which we considered as ecologically equivalent guilds, for the three habitats under study. Note that the estimation of colonization-extinction rates for very labile taxa might be biased. Therefore, we excluded, from subsequent analyses, those taxa with an estimated persistence value, defined as the inverse of the extinction rate ($p_i \equiv 1/e_i$), much shorter than the minimal inter-event sampling time (less than approximately a quarter of this time).

5.1.3 Core and satellite members of the community

Multiple methods have been applied to distinguish between core and satellite members of a community. While core species are abundant and persistent, satellite species usually show up at lower abundances and are occasional or even accidental. These two components of ecological communities feature distinct functional characteristics. The fact that persistent members of the community usually follow a log-normal abundance distribution, while accidental species follow the log-series (Magurran and Henderson, 2003), can be potentially used as a method to sort out the community core from the rest. However, when processing sequence data from microbial samples, it is common practice

to discard OTU sequences appearing only once to minimize potential errors. Therefore, the log-series is difficult or impossible to assess since it requires to record all real singleton species possibly observed in the sample. Instead of using abundance distributions directly, we have applied Chow tests to identify structural breaks in the relation between logarithmic maximum abundances and occupancy (defined as the probability that a species appears in the community over time). The Chow test (Chow, 1960) aims to identify unexpected changes in the parameters of linear regression models along the range of the independent variable. We first identified the intermediate breakpoint with the highest Chow test's statistic, this leading to two different slopes in the abundance-occupancy relation. Then we estimated the mean occupancy between consecutive ends of the two regression lines. We defined as core members of the community those OTUs with occupancy values higher than the aforementioned mean occupancy. OTUs with occupancy values below this threshold were identified as satellite members of the community. We performed Chow tests using the R package 'strucchange' (Zeileis et al., 2003) and log-normal fits for the core sub-community using the R package 'vegan' (Oksanen et al., 2019).

5.2 Results

5.2.1 A colonization - persistence trade-off

Trade-offs in ecology arise due to multiple mechanisms, such as competition, perturbations, or physiological constraints. Trade-offs tend to equalize fitness across species. In the context of colonization-extinction models, the colonization to extinction ratio can be regarded as a good measure of species fitness (Solé et al., 2004). In fact, it represents the number of new colonization events during the average time a species remains present in the system before extinction. If two species share this number, they should reach the same importance in the system,

either measured in terms of average abundance or average presence. This is true when species follow Levin’s metapopulation dynamics (Solé et al., 2004), or simple colonization-extinction independent dynamics, as we used in this paper. Under the assumption of species dynamics independence, species metacommunity dynamics can be formulated as

$$\frac{d\pi_i}{dt} = c_i(1 - \pi_i) - e_i\pi_i, \quad (5.1)$$

where (c_i, e_i) stands for the colonization-extinction rate pair for species i belonging to a pool of size S_P ($i = 1, 2, \dots, S_P$), and π_i is the probability that species i is found in a community (i.e., the occupancy of that species). Therefore, the probability of species i being present at equilibrium can be written as

$$\pi_i^* = \frac{k_i}{1 + k_i}, \quad (5.2)$$

where $k_i = c_i/e_i$ is the colonization to extinction ratio. Now we assume that equalizing mechanisms drive community dynamics, hence we expect that species fitness tends to equalize among species, $k_1 \approx k_2 \approx \dots \approx k_{S_P}$. Therefore, the probability π_i^* , which is also called expected occupancy at stationarity, tends to equalize for those species that share the same dimensionless colonization to extinction ratio, k . Conversely, if steady-state occupancies π_i^* are assumed to be roughly equal across species, then Eq. (5.2) trivially implies that all colonization to extinction ratios (k_i) will tend to be constant across species. Henceforth, we defined persistence as the inverse of the extinction rate ($p_i \equiv 1/e_i$). Because the hypothesis of equalizing mechanisms implies that all ratios c_i/e_i are approximately constant, $c_i/e_i = c_i p_i \approx k$, we find the following persistence-colonization fitness-equalizing trade-off:

$$p_i = k c_i^{-1}. \quad (5.3)$$

A generic colonization – persistence trade-off can be conceptualized as

$$p_i = k c_i^\alpha, \quad (5.4)$$

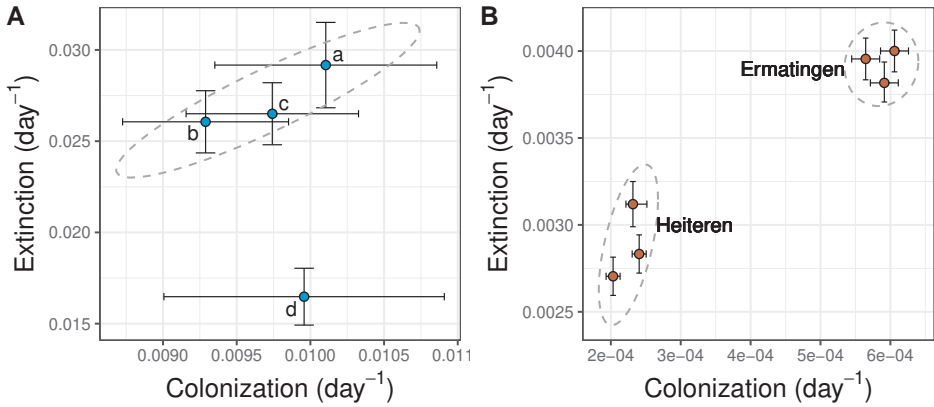


Figure 5.1: **Colonization and extinction rates precisely differentiate dynamics.** In panel A, we pooled the three consecutive Pyrenean lakes from the same basin together while a different colonization-extinction pair characterize the fourth lake from another basin. Lake a, Llebreta, b, Llong, c, Redó d’Aigüestortes, d, Redó. In panel B, two groups of soil samples cluster together around similar colonization-extinction values. These groups correspond to replicates from the same site. Colonization and extinction rates and their error bars were calculated with function `irregular_single_dataset` from the ‘`island`’ R package.

with exponent $\alpha < 0$. Therefore we conclude that if a community of equivalent species is close to performing colonization-extinction independent dynamics, the exponent α of the generic colonization – persistence trade-off above [Eq. (5.4)] should be -1. This is our theoretical prediction, the one we have checked across the three different microbial communities. Throughout this work, we have represented colonization and persistence axes on a logarithmic scale. This leads us to conclude that a trade-off between colonization and persistence compatible with independent colonization-extinction dynamics should display a slope equal to -1 on a log scale, as it is deduced from the equation: $\log p_i = K - \log c_i$.

5.2.2 The species equivalence assumption

Under this assumption, all species in the community are described by the same colonization-extinction pair. This approximation allowed us to explore whole community dynamics for microbes in lakes of the

Table 5.1: **Model selection procedure for the dynamics of microbial communities in lakes of the Pyrenees.** Letters correspond to different lakes, and the separation by points indicates different models with their specific pair of colonization and extinction rates.

| Model | NLL | Parameters | AIC | Δ AIC | w_i (%) |
|---------|----------|------------|----------|--------------|-----------|
| a.b.c.d | 19797.67 | 8 | 39611.35 | 2.65 | 11.1 |
| abcd | 19873.56 | 2 | 39751.20 | 142.50 | < 0.1 |
| abc.d | 19800.82 | 4 | 39609.64 | 0.94 | 26.2 |
| ab.c.d | 19800.09 | 6 | 39612.19 | 3.49 | 7.3 |
| ac.b.d | 19799.50 | 6 | 39610.99 | 2.29 | 13.4 |
| a.bc.d | 19798.35 | 6 | 39608.70 | 0 | 41.9 |

Pyrenees and soils in Switzerland. For the lakes in the Pyrenees (Figure 5.1A), we found that the dynamics of the three lakes in the same basin were so similar that they accumulated a weight of evidence of 89% (summing over all models considering at least two of the three lakes as having the same colonization and extinction), as opposed to the model with all lakes with different rates, which had only a weight of evidence of 11% (see Table 5.1). In the case of the soils in Switzerland (Figure 5.1B), the distance among replicas within the same soil type and site in colonization and extinction rates was smaller than between sites, showing that the replicas had similar dynamics on each site.

5.2.3 Relaxing the equivalence assumption

Next, we relaxed this assumption and considered the different taxonomic groups in these two bacterial communities, plus the meta-community in saline lagoons in the Monegros desert. We found that colonization-persistence patterns were coherent as we descended to lower taxonomic levels. So, the distances of genera and families within phyla, classes, or orders (intra-group) were lower than distances between different higher taxonomic levels (inter-group) in the three communities (Kruskal-Wallis test, all p-values < 0.1 in the Pyrenees and < 0.01 in saline lagoons and soils). Moreover, our estimates of colonization and

Table 5.2: **The logarithms of colonization and persistence in the three communities studied are related.** p-values refer to Spearman's ρ .

| Community | Level | Slope | Lower C.I. | Upper C.I. | Spearman's ρ | p-value | n |
|-----------|--------|--------|------------|------------|-------------------|---------|-----|
| Pyrenees | Phylum | -3.455 | -5.288 | -1.623 | -0.736 | 0.0058 | 13 |
| | Class | -2.156 | -3.778 | -0.533 | -0.643 | 0.0016 | 22 |
| | Order | -1.906 | -3.109 | -0.702 | -0.548 | 0.0010 | 34 |
| | Family | -1.636 | -2.678 | -0.593 | -0.507 | 0.0008 | 41 |
| | Genera | -1.619 | -2.687 | -0.552 | -0.516 | 0.0007 | 41 |
| Soils | Phylum | -1.402 | -2.192 | -0.611 | -0.468 | 0.0148 | 27 |
| | Class | -1.088 | -1.582 | -0.593 | -0.411 | 0.0009 | 63 |
| | Order | -1.023 | -1.377 | -0.669 | -0.427 | 1e-05 | 100 |
| | Family | -0.815 | -1.102 | -0.528 | -0.441 | 3e-07 | 127 |
| | Genera | -0.807 | -1.087 | -0.527 | -0.424 | 4e-07 | 136 |
| Monegros | Phylum | -1.072 | -1.445 | -0.700 | -0.709 | 5e-05 | 27 |
| | Class | -0.989 | -1.300 | -0.678 | -0.609 | 3e-06 | 52 |
| | Order | -0.929 | -1.199 | -0.660 | -0.557 | 8e-09 | 95 |
| | Family | -0.907 | -1.142 | -0.673 | -0.480 | 2e-10 | 160 |
| | Genera | -0.820 | -1.038 | -0.602 | -0.437 | 1e-10 | 202 |

persistence were negatively related conforming to a generic trade-off. An increase in colonization rates led to decreases in persistence, and this relation was maintained across taxonomic levels for the three communities (Table 5.2). Also, the slope of the linear models relating the logarithms of colonization and persistence was close to -1. We recall here that a slope of -1 would correspond to a colonization–persistence trade-off resulting from fitness equalization between different taxonomic groups, under the assumption of colonization-extinction independent dynamics.

However, the assumptions underlying a colonization – persistence trade-off with exponent -1 might be too severe to apply to whole communities. It is well-known that core members of a community may display different dynamics from the satellite components of it (Magurran and Henderson, 2003). Satellite species tend to be rare and accidental. Sometimes they are observed, sometimes they are not. These species may be good candidates to show a kind of behavior consistent with colonization-extinction independent dynamics, and,

therefore, the satellite subcommunity should tend to show, accordingly, a colonization – persistence trade-off with exponent -1. Instead, the core members of the community tend to be more abundant, and therefore the relative strength of niche processes, such as interactions and niche segregation, would be higher than in the case of satellite species. Then, core species would not necessarily show a trade-off with exponent -1 if they show any at all.

To test this hypothesis, we first identified the core and the satellite members of our communities. As abundance enhances occupancy, following a similar argument as in Magurran and Henderson (2003), we represented the linear relation among maximum abundance and occupancy at the Genus level. However, we used a Chow test analysis instead to separate the core from the satellite members of the community by identifying structural changes in the linear relation among maximum abundance and occupancy. Figure 5.2 shows the structural changes found in the three studied communities. The point with the biggest statistic allowed us to infer an occupancy threshold that separated the core from the satellite members of the community. As reported for macroscopic communities, the abundance distribution of the core of the communities followed a log-normal distribution closely.

In sum, the distinction between core and satellite members of the community allowed us to examine the relationship between colonization and persistence separately for these two components, as shown at the family taxonomic level (Figure 5.3). These two components presented significantly different slopes, as shown by testing the hypothesis that satellite and core species share the same slope of the linear model (in logarithmic axes) relating colonization and persistence. Moreover, the satellite component of the communities showed slopes very close to -1, while slopes were lower for the community core, except in the case of soils, where both core and satellite sub-communities showed exponents close to -1 (see Table 5.3). These results were maintained across taxonomic levels, up to the lowest one, the Genus. However,

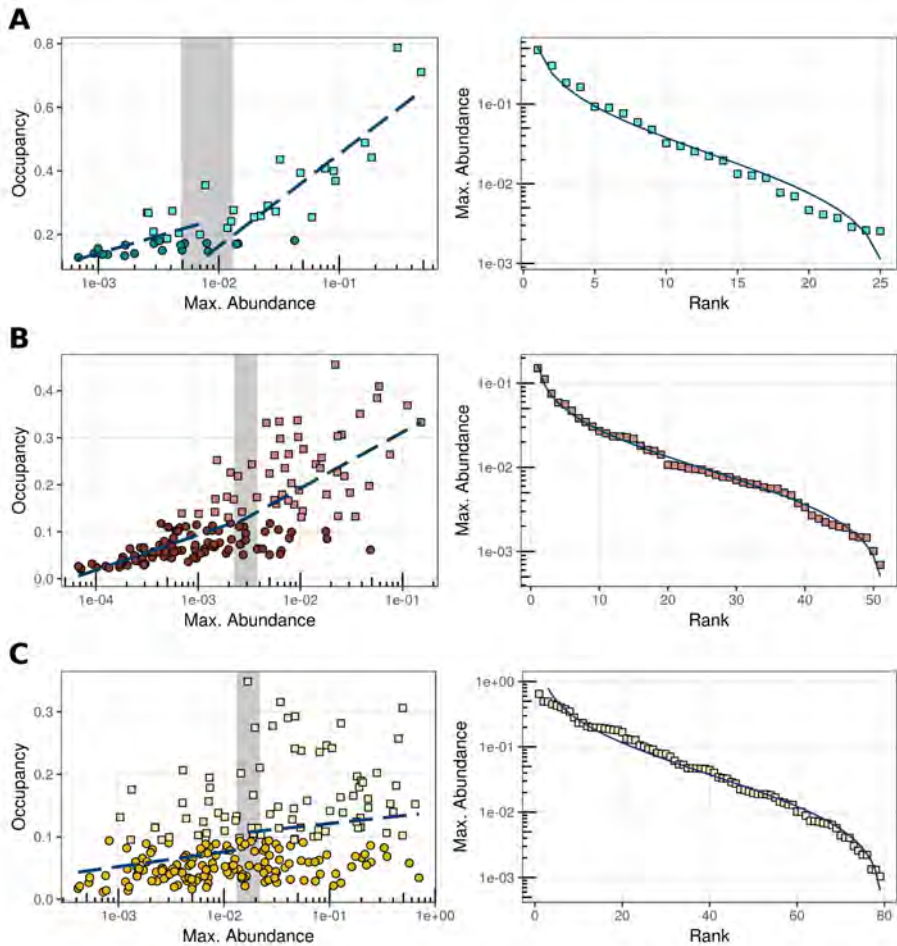


Figure 5.2: **The core members of the community follow a log-normal distribution.** A) Lakes in the Pyrenees, B) Soils, C) Shallow saline lakes in Monegros. *Left*, blue dashed lines represent the linear relationship between the highest abundance and occupancy at the genus level, which presents structural changes, determined by a Chow test with maximum values for the statistic in the grey shaded area. We have considered as core genus (squares) those that presented values of occupancy higher than the mean occupancy of the point with maximum structural change, while those with a lesser occupancy were considered satellite members (circles). *Right*, the core members of the communities present a log-normal distribution (solid blue line). Pyrenees deviance = 1.063; soils deviance = 0.666; Monegros deviance = 4.042. Log-normal distributions were fitted using function `rad.lognormal` of the R package ‘vegan’.

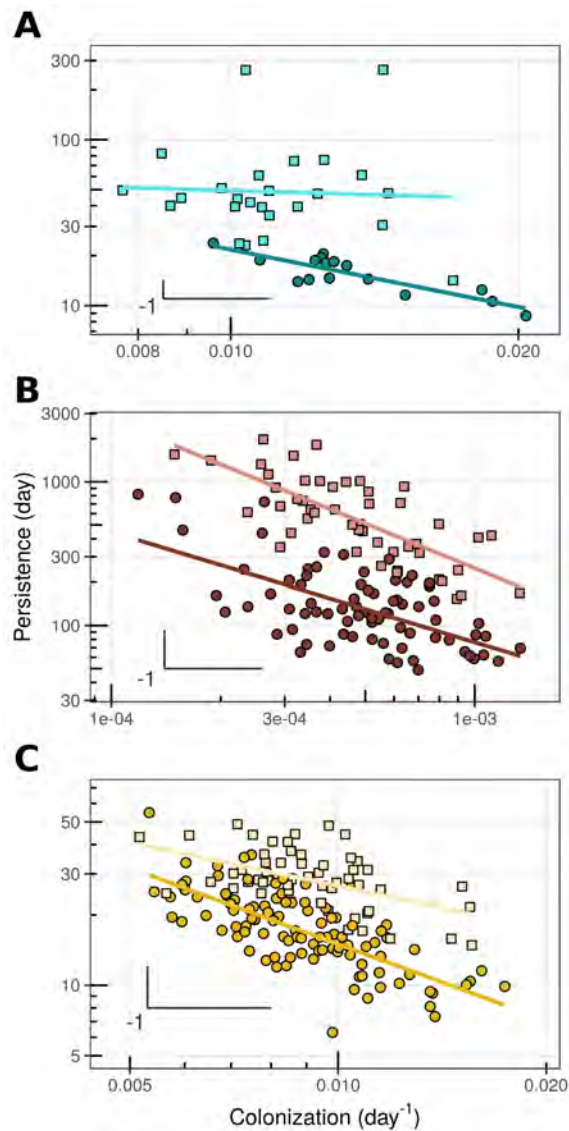


Figure 5.3: **Microbial communities show a colonization–persistence trade-off at the family level.** Three different habitats, alpine lakes (A), soils (B), and shallow saline lakes (C), display a linear relationship close to the theoretical expectation under a perfect colonization–persistence trade-off (not shown). The trade-off is maintained throughout the phylogeny, from *Phylum* to *Genus*. However, core (squares) and satellite (circles) members of the community show different relationships between persistence and colonization, being the satellite members closer to the theoretical expectation. The two legs indicate the -1 slope.

Table 5.3: **Core and satellite sub-communities show differential relationships for colonization and persistence.** A slope of -1 would correspond to a perfect trade-off between colonization and persistence. We have tested the hypothesis that the slope of the linear model for satellite taxa is equal to the slope obtained for core species (Student’s t-test), which was rejected in all cases. Associated p-values and t-scores are shown. Additionally, we report data for fitted slopes and their 95% confidence interval. *ns*, p-value higher than 0.1, *****, p-value lower than 0.001.

| Community | Tax. | p-value | t-score | Component | Slope | Lower C.I. | Upper C.I. | n |
|-----------|--------|---------|---------|-----------|-----------------------|------------|------------|-----|
| Pyrenees | Family | 8e-05 | -5.513 | Core | -0.168 ^{ns} | -1.651 | 1.315 | 25 |
| | | | | Satellite | -1.149 ^{***} | -1.531 | -0.767 | 16 |
| | Genera | 3e-05 | -5.913 | Core | -0.103 ^{ns} | -1.637 | 1.430 | 25 |
| | | | | Satellite | -1.148 ^{***} | -1.527 | -0.769 | 16 |
| Soils | Family | 0.0012 | 3.366 | Core | -1.038 ^{***} | -1.314 | -0.762 | 49 |
| | | | | Satellite | -0.688 ^{***} | -0.895 | -0.482 | 78 |
| | Genera | 0.0004 | 3.697 | Core | -1.056 ^{***} | -1.319 | -0.792 | 51 |
| | | | | Satellite | -0.683 ^{***} | -0.883 | -0.482 | 85 |
| Monegros | Family | 2e-05 | -4.432 | Core | -0.621 ^{***} | -0.905 | -0.336 | 60 |
| | | | | Satellite | -1.081 ^{***} | -1.288 | -0.875 | 100 |
| | Genera | 0.0001 | -4.032 | Core | -0.662 ^{***} | -0.938 | -0.385 | 79 |
| | | | | Satellite | -1.038 ^{***} | -1.223 | -0.853 | 123 |

as we go up in the taxonomy, losses in statistical power blur these relationships.

5.3 Discussion

In this study, we have shown that occasional and persistent taxa in microbial communities are characterized by colonization – persistence trade-offs. Across the three systems analyzed, we found that microbial taxa conform to an almost perfect colonization – persistence trade-off, especially for the occasional members of the community. The existence of this trade-off is consistent with the satellite component following, in a close approximation, colonization-extinction independent dynamics near to a steady-state. This trade-off also implies the existence of fitness equalization, which may be more important than previously recognized in microbial communities, particularly for the functioning

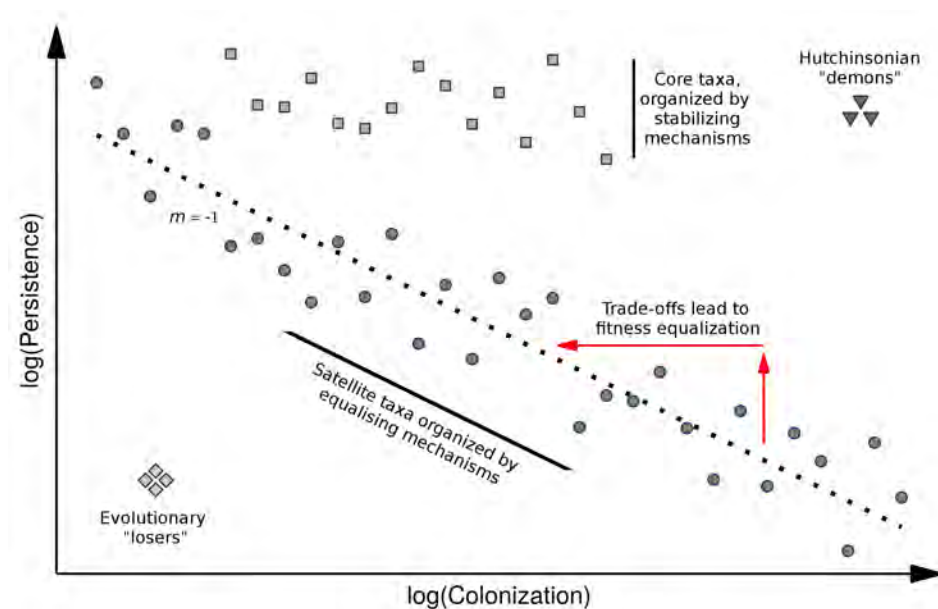


Figure 5.4: **Relationship among colonization and persistence.** The relative importance of coexistence promoting mechanisms allows us to distinguish several components in microbial communities. The dotted line indicates a perfect persistence – colonization trade-off, where equalizing mechanisms such as trade-offs lead to similar fitness among groups. Any attempt of the satellite taxa to increase their performance would likely result in a corresponding decrease due to life-history constraints. However, in core taxa stabilizing mechanisms dominate and niche differences are high, due to *e.g.* resource partitioning. Hutchinsonian "demons" would represent very persistent species with great colonization abilities, that would outcompete all other taxa, which is unlikely in microbial communities. Evolutionary "losers" stand for species with low fitness and low colonization abilities that would likely represent accidental dispersers not adapted to the environmental conditions of the community.

of the satellite sub-community. Satellite members would remain in the community by either evolving higher colonization rates but persisting shorter periods or developing the ability to stay longer in the community along with lower colonization rates. Examples of life-history trade-offs can be found easily among macroorganisms and experimental settings of microbes (Jessup and Bohannan, 2008), but, to the best of our knowledge, this is the first time that such a trade-off is reported in highly diverse natural microbial communities.

We have conceptualized the relationship between persistence and colonization in Figure 5.4. This pattern is reminiscent of the one reported by Cadotte et al. (2006). However, the main difference with figure 1 from Cadotte et al. (2006) is that here we identify core and satellite taxa. The identification of core and satellite species is not new in microbial ecology (van der Gast et al., 2011), although similar terms have arisen to refer to the less abundant component, such as the rare biosphere (Lynch and Neufeld, 2015), or conditionally rare taxa (Shade and Gilbert, 2015; Shade et al., 2014). While satellite taxa would follow the trade-off as a result of fitness equalizing mechanisms, core taxa would be driven by stabilizing mechanisms tending to maintain similar persistence across them, but higher than for satellite taxa. Conversely, within the satellite sub-community, in the presence of equalizing mechanisms, any increase in colonization (or persistence) ability would be followed by decreases in persistence (or colonization) ability. Moreover, core taxa are common, abundant species following a log-normal abundance distribution (Magurran and Henderson, 2003). As Cadotte et al. (2006) pointed out, in principle, other kinds of taxa could potentially exist: Hutchinsonian "demons", that would competitively exclude other taxa, and evolutionary "losers", that would not colonize nor persist in the community. The microbial communities we have analyzed appear to be compatible with this conceptual view. The three studied communities showed a log-normal abundance distribution for the core component, as expected, and our observation of ecological coherence in colonization and persistence within taxonomic levels might well indicate niche differences (Philippot et al., 2010) produced by stabilizing mechanisms. Besides, the satellite component of the aquatic communities under study showed a slope of approximately -1 compatible with a colonization – persistence trade-off under species colonization-extinction independent dynamics. Conversely, the soil community showed a similar trade-off (with a slope close to -1) for both core and satellite taxa. This might be due to the way a soil

compaction experiment affected this microbial community. While the microbial aquatic communities might be considered at a seasonally-driven steady state, the soil community was intentionally poised out from its natural steady state. For instance, soil compaction may have led to increased anaerobiosis driving the community out of and far away from a previous natural colonization-extinction equilibrium. Furthermore, the relaxation time to the new steady-state in response to this disturbance may have also differed for the different treatments (Hartmann et al., 2014).

In the context of metacommunities, stabilizing forces have been associated with SS, while equalizing forces to NT (Adler et al., 2007). SS and NT have been proposed alternatively as the major mechanisms controlling microbial community assembly. In fact, the importance of SS has been evaluated against other metacommunity archetypes as NT (Langenheder and Székely, 2011; Lee et al., 2013) or mass effects (ME) (Souffreau et al., 2014; Van der Gucht et al., 2007) with contrasting results. Also, NT has been tested and proposed as the dominant force structuring communities (Ofiteru et al., 2010; Woodcock et al., 2007). The dichotomies niche – neutral (Dumbrell et al., 2010; Ferrenberg et al., 2013) or stochastic – deterministic (Caruso et al., 2011; Wang et al., 2013; Zhou and Ning, 2017) are similar to the SS – NT divide, and are often used as synonyms. The most accepted view seems to be that initial steps in community assembly are dominated by neutral processes, while SS characterizes later stages, but this view is rarely put in the context of coexistence mechanisms. Our work adds to this discussion the fact that are precisely satellite species the ones governed by fitness equalization. In the light of our findings, this important component of natural communities would be integrated by ecologically equivalent, rare species, at the lower end of the abundance spectrum, undergoing a type of temporal dynamics consistent with simple colonization-extinction independent dynamics.

Equalizing mechanisms can evolve in species-rich communities with

strong dispersal and recruitment limitation (Hubbell, 2006), although microbial communities are unlikely affected by these limitations. However, experimental settings have repeatedly shown that microbial trade-offs evolve easily in controlled, species-poor microbial experiments (Huang et al., 2017; Yawata et al., 2014), and might be key in microbial communities (Litchman et al., 2015). A potential equalizing mechanism might be horizontal gene transfer, as it has been proposed that it produces highly flexible gene pools associated with specific habitats (Polz et al., 2013), that would equalize fitness and increase niche overlap. Also, nonlinear responses to fluctuating environments can act as equalizing or stabilizing mechanisms (Chesson, 2000). Stabilizing mechanisms are widespread in microbial communities, *e.g.*, resource partitioning, dormancy (Jones and Lennon, 2010), or cross-feeding (Goldford et al., 2018), although the processes underlying these mechanisms are rarely studied or understood at trait or biochemical levels. The strength of these stabilizing mechanisms may well allow the satellite members of the community to coexist in the presence of the core component.

The purpose of this study was to examine the importance of equalizing mechanisms for microbial coexistence. Our results rely on a dynamic stochastic model, rooted in classic ecological theory. Although its assumptions are drastically simplifying (*species equivalence* and *species independence*), it should be viewed as an approximation to the actual underlying dynamics of the community or its components (when relaxing the *equivalence* assumption). We used this model to estimate extinction and colonization rates from temporal datasets (Alonso et al., 2015; Ontiveros et al., 2019). However, the accuracy of these estimates should be assessed carefully. First, very rare species may be there, but under detectability levels (MacKenzie et al., 2003; Ontiveros et al., 2019; ?). Second, when persistence times are too short compared to inter-sampling times, these estimates may not be reliable. If taxa go in and out from the system too rapidly, their estimated

rates may be biased (see Appendix B). This possible bias is the reason why some labile taxa (less than 13 % in all cases) were removed from our analyses. The exclusion of these taxa did not change the overall patterns reported in this study.

The relevance of equalizing mechanisms for coexistence might have been overlooked in natural microbial communities. However, they may be relevant in highly diverse ecosystems, especially acting on occasional taxa, as previously suggested (Langenheder and Székely, 2011). Here we argue that, as a result of fitness equalization, occasional taxa should show a persistence–colonization trade-off with slope -1 in logarithmic axes. For satellite species, this pattern may well hold beyond the microbial world, which would be worth exploring in the future. Long-term temporal studies are needed to improve our knowledge of coexistence mechanisms. We hope that framing this discussion in the context of equalizing vs. stabilizing mechanisms would add clarity to current knowledge on the forces maintaining high microbial diversity on Earth ecosystems.



Image:

Wassily Kandinsky. *Sky Blue*. 1940, Centre Georges Pompidou, Paris.

This chapter:

Ontiveros, VJ, Caliz, J, Triadó-Margarit, X., Alonso, D, Casamayor, EO. General decline in the diversity of the airborne microbiome under future climatic scenarios.

In preparation.

General decline in the diversity of the airborne microbiome under future climatic scenarios

Abstract

Microorganisms have a profound impact on global functioning and huge dispersal potential. Many of them move attached to aerosols, where they can travel long distances, survive and even have metabolic activity. The fate of airborne microbes is influenced by environmental and climatic variables that are predicted to change, with unknown consequences. Using a 7 yr. temporal monitoring of airborne microbes collected in the Central Pyrenees (LTER node Aigüestortes), we have developed a new predictive method, firmly rooted in the Theory of Island Biogeography, that addresses the temporal evolution of biodiversity dynamically in response to associated environmental covariates linked to future climatic scenarios. In the worst-case scenario, general declines in bacterial richness, idiosyncratic responses for the eukaryal component, and changes in seasonality are expected, motivated by contrasting importances of the environmental filtering of the atmosphere and the origin of the aerosols. Also, we predicted lower richness of two major eukaryotic putative pathogen groups related to plant and human pathogens. Our results highlight the need for new theoretical and predictive tools, coupled with high-quality information collected by long-term monitoring, to predict and understand the complex consequences of future climatic scenarios.

Humanity is failing to solve ecological challenges as the emission of greenhouse gases, deforestation, and loss of species (Ripple et al., 2017), and approaches a mass extinction event (Barnosky et al., 2011). Loss of species is likely to accelerate further

(Johnson et al., 2017), while climate change already has effects on plants and animals in aquatic and terrestrial ecosystems (Parmesan, 2006). Hence, ecological forecasting is imperative in order to revert, avoid, or mitigate the effects of climate change (Clark et al., 2001). Throughout the history of Earth, microbes have changed the climate, and climate has changed the microbes (Cavalier-Smith et al., 2006; Hutchins et al., 2019). Microorganisms can modify ecosystem processes or biogeochemistry on a global scale, and we start to understand their role and potential involvement in changing the climate (Singh et al., 2010). However, the effects of climate change on microbial communities (i.e., diversity, dynamics, or distribution) are rarely addressed (see Cavicchioli et al. 2019). The development of microbial communities with climate change is central to identify and understand climate-ecosystem feedbacks.

During the last century, microorganisms are being moved at unprecedented scales changing species distributions due to waste disposal, tourism, and global transport (Zhu et al., 2017). Physical forces such as air mass circulations and oceanic water currents disperse vast amounts of microorganisms and interconnect remote environments. The origin of air masses from marine or terrestrial environments, as well as from anthropogenic-impacted systems, mainly shapes the atmospheric air microbiome (Cáliz et al., 2018). Airborne microorganisms can travel and survive between continents (Creamean et al., 2013), settle on remote environments, and create geographic patterns (Barberán et al., 2015). The circulation of atmospheric microorganisms contributes to global health and ecological concerns such as the dispersion of pathogens (Brown and Hovmøller, 2002; Kellogg and Griffin, 2006) and antibiotic resistances (Mazar et al., 2016), cloud formation and precipitation (Creamean et al., 2013), and colonization of pristine environments (Hervàs et al., 2009; Hervàs and Casamayor, 2009). Airborne microorganisms also play a role in the formation of the phyllosphere, which is one of the vastest habitats of the earth's surface (Griffin and

Carson, 2015; Vorholt, 2012) involved in nutrient cycling (Guerrieri et al., 2019; Nadkarni et al., 2011). Microbial mass movements can change the geographic distribution of microorganisms on a global scale, and therefore have ecological and socioeconomic consequences (Pecl et al., 2017).

Revisiting the aforementioned study (Cáliz et al., 2018), we have examined the effects of climatic variables and the source of aerosols (indicated by the chemical composition of depositions) on the diversity of atmospheric microbial communities. Airborne samples were fortnightly collected on a remote high-altitude mountain in the Central Pyrenees (Long Term Ecological Research site Aigüestortes, NE Spain) over seven years. We collected washed out aerosols in wet deposition that has recently been proven to be a proper method for monitoring the long-term intercontinental exchange of high-atmosphere airborne microorganisms (Triadó-Margarit et al., 2019). The dataset is unique to explore the temporal dynamics of the long-range dispersal of airborne bacteria, fungi, and protists. Based on the identification of the main drivers that determine the diversity of airborne microorganisms, we have simulated possible future responses for these communities according to three commonly used climatic scenarios. We anticipate here that simulations indicate general declines in bacterial richness and changes in seasonality for both eukaryal and bacterial components. Our results highlight the use of ecological theory to understand long-distance airborne dispersal of microbial taxa.

6.1 Material&Methods

6.1.1 Data

Briefly, aeroplankton samples were collected twice monthly at the LTER-AT node site, within the protected area of the Aigüestortes i Estany de Sant Maurici National Park (42°33'N 0°53'E; the Pyrenees, northeastern Spain). DNA was extracted using the Mobio PowerSoil

DNA Isolation Kit (Mobio Laboratories). PCR and high-speed multiplexed SSU rRNA gene Illumina MiSeq sequencing were carried out for 16S and 18S rRNA genes. Raw rRNA gene sequences were processed using the UPARSE pipeline, and OTUs (97% identity) were taxonomically assigned with SILVA_119. The entire genetic dataset is available in the National Center for Biotechnology Information Sequence Read Archive under accession no. PRJEB14358. Airflow trajectories, coupled with the chemical characterization of rain/snow samples, indicate that air mass origin drives the seasonality of aeroplankton and affects long-range dispersal of airborne microbes. For more details on the environmental metadata used, please see below and Cáliz et al. 2018.

The identification of potentially pathogenic microbes was conducted in the same way as previously done by air samples indoors (Triadó-Margarit et al., 2017) and biofilm and water samples (Auguet et al., 2017; Subirats et al., 2017). Briefly, BLAST analyses (Altschul et al., 1990) were used to examine representative OTU sequences against an in-house database of obligate and opportunistic pathogens. Only those OTUs having sequence identity values above 98% and highest BLAST alignment coverage values (threshold set at 90%) to inventoried pathogens were considered for downstream analysis. As stated earlier (Subirats et al., 2017; Triadó-Margarit et al., 2017), the results presented here should be regarded only as indicative of putative pathogenic microbes, as additional confirmation steps would be needed to corroborate the pathogenic capacity of the matched OTUs fully.

6.1.2 Estimation of colonization-extinction dynamics

We have followed Alonso et al. 2015 in order to estimate the air microbiome dynamics. Briefly, our methodology is based on the simplest stochastic model of Island Biogeography (Simberloff, 1969). This dynamic model indicates that the change in richness in a site follows equation 1.3:

$$\frac{dS}{dt} = c(S_P - S) - eS$$

where S corresponds to the OTU-richness in the site, S_P to the number of species in the regional pool (i.e., all OTUs identified) and c and e correspond to colonization and extinction rates, respectively. Alonso et al. (2015) show that it is possible to estimate these colonization and extinction rates from temporal community data. Additionally, these colonization and extinction rates might not be constant and could vary alongside several environmental variables. The easiest way to model this dependency is:

$$c_t = c_0 + \sum_{i=1}^F \alpha_i Y_{it}; \quad e_t = e_0 + \sum_{i=1}^F \beta_i Y_{it}$$

where Y_{it} represents the value of the environmental variable Y_i at time t . These coefficients α_i and β_i can be estimated with a greedy algorithm included in R Package `island` (Ontiveros et al., 2019), using function `greedy_environmental_fit`.

Model selection procedure

We started estimating the colonization and extinction rates of the simplest stochastic model of island biogeography (named as model S). This model can be considered as a null model in which all OTUs share the same dynamics, assuming equivalence and independence among them. We examined another model that relaxes the equivalence assumption (model P), based on the estimation of both rates for each group of OTUs ascribed to high taxonomic ranks (37 and 12 groups for bacteria and eukarya, respectively). A similar model (model A) to the previous one considered separately those groups of abundant taxonomic ranks as in Cáliz et al. 2018 (>1% relative abundance; 12 and 13 groups for bacteria and eukarya, respectively), while the rest of ranks (<1% relative abundance) were included in the same group named as "other" bacteria or eukarya. We also devised a model that examined the influence of environmental covariates in the dynamics of the whole community, assuming equivalence among the OTUs (model E). Finally, we combined the model A and E in the last model $A * E$

to consider the influence of the environmental variables within each group. For the subset of OTUs identified as pathogens, we separated them based on their putative host instead of the taxonomic ranks, in a model that relaxes the equivalence assumption (model G). We also combined models G and E in the last model, $G * E$, to consider the influence of the environmental variables within each host.

Sampling effort was considered in the models that included environmental effects, and was treated as another environmental covariate.

Validation of the model

We have used R^2 as a means for validation. R^2 is defined as $1 - \frac{\epsilon^2}{\epsilon_0^2}$ where ϵ^2 represents the mean quadratic error of a simulated model and ϵ_0^2 the mean quadratic error of a null model of choice. We have selected the mean observed richness as our null model. Given so, our R^2 would represent the relative performance of our dynamic simulations against the static estimator that is the mean value. We have also checked the ability of our models to predict unencountered data, estimating them with three-quarters of the temporal series, and simulating the remaining one.

6.1.3 Prediction of microbial responses to climate change

We downloaded regional climate models from EURO-CORDEX. We selected the simulations with monthly mean, maximum, and minimum temperatures that were bias-corrected and had predictions for three different Representative Concentration Pathways (RCP), RCP2.6, RCP4.5, and RCP8.5, for the European domain at a resolution of 0.11 degrees. The different RCP scenarios are named after their relative radiative forcing against the baseline of 1750. RCP2.6 represents a forcing of 2.6 W/m², a scenario with intense mitigation with future negative emissions, RCP4.5 corresponds to stabilization at 4.5 W/m², and RCP8.5 is the high emissions scenario, with radiative forcing 8.5 W/m² increasing even after 2100.

Our search gave, as a result, the models MPI-M-MPI-ESM-LR-MPI-CSC-REMO2009 (with two replicates that were averaged) and ICHEC-EC-EARTH-SMHI-RCA4. We selected only the prediction at the cell of the grid that corresponded to our study site, obtaining said temperatures. We obtained an ensemble model by averaging the two models, ranging from 2001 to 2100.

Future temperatures

In order to simulate future temperatures, we first obtained data of daily mean, maximum, and minimum temperatures (together with humidity and irradiance) for the years studied, 2007 – 2013, from the nearest meteorological station, corresponding to Boi (2535 m.a.s.l.). We obtained monthly averages, and we estimated the mean and standard deviation of the monthly averages. Then, we calibrated the ensemble temperatures for the same period (2007 – 2013), equalizing the mean and standard deviation of these temperatures with the observed ones, using the equation $y_i = \bar{x}_2 + x_i - \bar{x}_1 \cdot \frac{\sigma_2}{\sigma_1}$, where \bar{x}_1 and \bar{x}_2 are respectively the means of the ensemble and observed temperatures and σ_1 and σ_2 their standard deviations. Future temperatures were then transformed accordingly for the period 2021 – 2100.

Future environment

To obtain simulations of the future conditions of the Pyrenees, we first identified which variables correlated with any of the temperature variables, finding that pH, ANC, Cond., Cl, SO₄, NO₃, Na, K, Ca, Mg, DIC, DOC, TN, Hum., Irr. and Samp. Eff. did, while Rain and TP did not and were subsequently excluded from the following steps. Next, we found the generalized linear model of Tmax, Tmed, Tmin, and their interactions that best explained (lower AIC) each environmental covariate. Using these models and the previously predicted future temperatures, we obtained predicted values for the environmental covariates to which we added a random residual, chosen via a random

uniform number in the $[0, 1]$ that represented a quantile of the observed distribution of residuals.

In the case of the two excluded environmental covariates, the future values of the environmental covariates were chosen as the value given by the quantile associated with a uniform random number, taking into account seasonality for TP, as we did not detect seasonality for Total Rain. The results of this procedure were then the future values for the environmental variables. This procedure was repeated 100 times in order to obtain more reliable estimates and avoid unwanted biases in the prediction of future environmental conditions.

Prediction under future regimes

We predicted species richness for each group fortnightly during the period 2021 – 2100, with the previously estimated expressions for colonization and extinction and the future environmental conditions, using functions `rates_calculator`, `cetotrans`, and `data_generation` of R Package 'island'. We repeated this procedure 100 times for each of the replicates of the future environmental conditions.

6.1.4 Statistical analyses

We have calculated Pielou's evenness to measure how equal bacterial and eukaryal communities are. As we are dealing with species richness but not abundances, we have used a richness-based Shannon Index for each predicted sample, with S equal to the number of groups in the best model, and p_i being the proportion of OTUs of group i within the total number of OTUs in the sample.

6.2 Results

The model selection procedure indicated that the model $A * E$, which considered the most abundant high-rank taxonomic groups and the environmental variables, was the model with lowest AIC among all

examined for both bacteria and eukarya (Table C.1). Moreover, we used an estimation of R^2 to validate the fit of this model (C.1), which included each considered group in the best model (Figures C.2, C.3). For the whole temporal series, we found an R^2 of 0.323 (0.303) for the bacterial (eukaryal) model estimated with the whole dataset, while these values descended to 0.324 (0.233) for the model estimated with three-quarters of the temporal series. For the last quarter of the data, R^2 was 0.229 (0.304) for the bacterial (eukaryal) models estimated with the whole series, while the validation model performance was as good as the static null model that was the mean (R^2 0.003 for Bacteria and 0.058 for eukarya).

The dynamics of bacteria were strongly related to climatic factors such as temperature, irradiance, and humidity (Figure 6.1). Overall, bacterial dynamics reacted coherently to changes in minimum and mean temperatures. Moreover, the colonization of the richest bacterial groups (i.e., Alphaproteobacteria and Bacteroidetes) increased with DOC enriched depositions. Nitrogen compounds slightly influenced the dynamics of most bacteria. In contrast, the origin of aerosols, suggested by the chemical composition of depositions, determined the dynamics of the most abundant and richest eukarya (Figure 6.1); ANC, conductivity, Ca, and Cl strongly influenced the colonization of Basidiomycota and Ascomycota. The temperature fluctuations also determined the dynamics of other abundant eukarya such as Chytridiomycota and Rhizaria. The humidity and irradiance slightly influenced the dynamics of most eukarya. Model parameters and richness for each group are shown in tables C.2 and C.3.

We have found that temperature is the key driver of the dynamic of aeroplanktonic communities, either directly (Figure 6.1) or through its relation with the rest of the environmental variables considered in this study. The temperature in the study site is expected to steadily increase up to 5°C by the end of this century in the worst-case scenario and to have slight increases in the other two scenarios, as shown by the

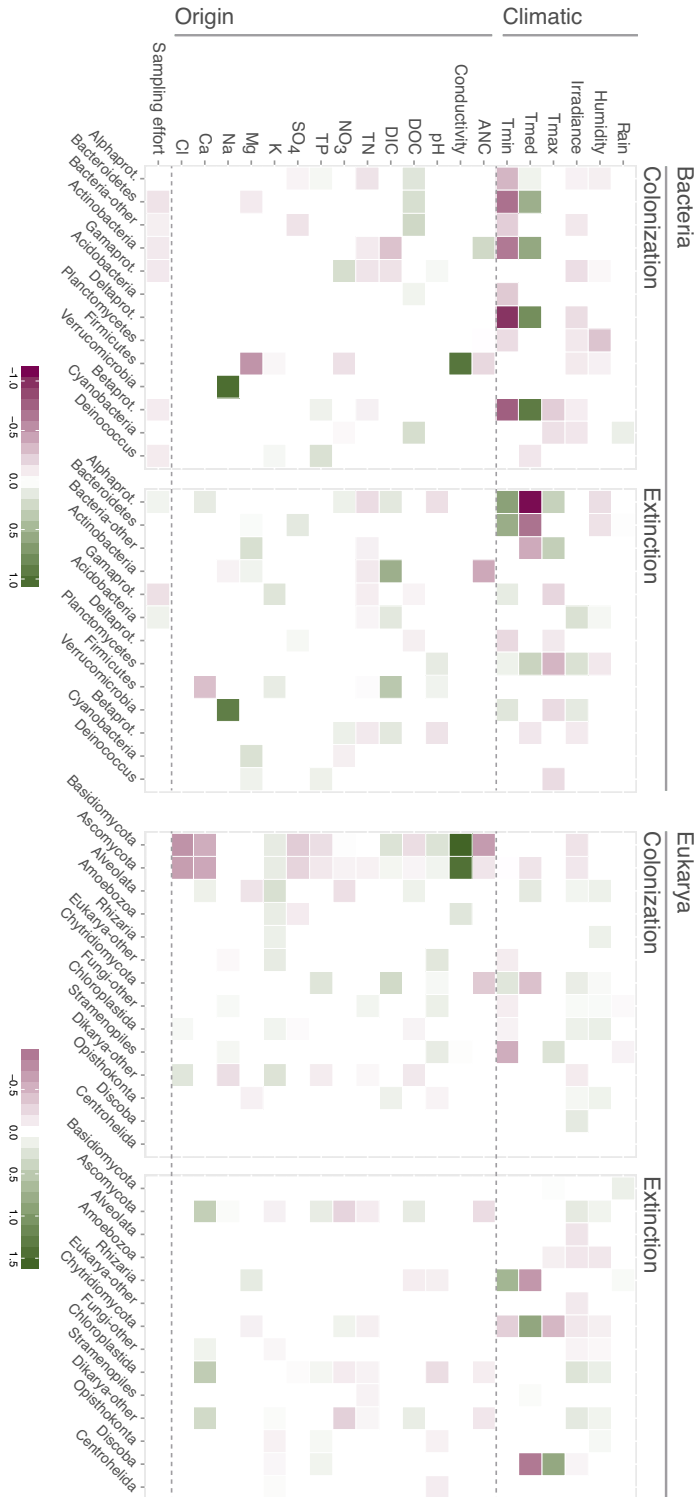


Figure 6.1: **Effect of environmental variables associated with climate or origin on colonization and extinction dynamics.** The effect is weighted by the logarithm of richness and relative to either colonization or extinction independent terms. Positive values indicate an increase in colonization (or extinction) when a given environmental variable increases.

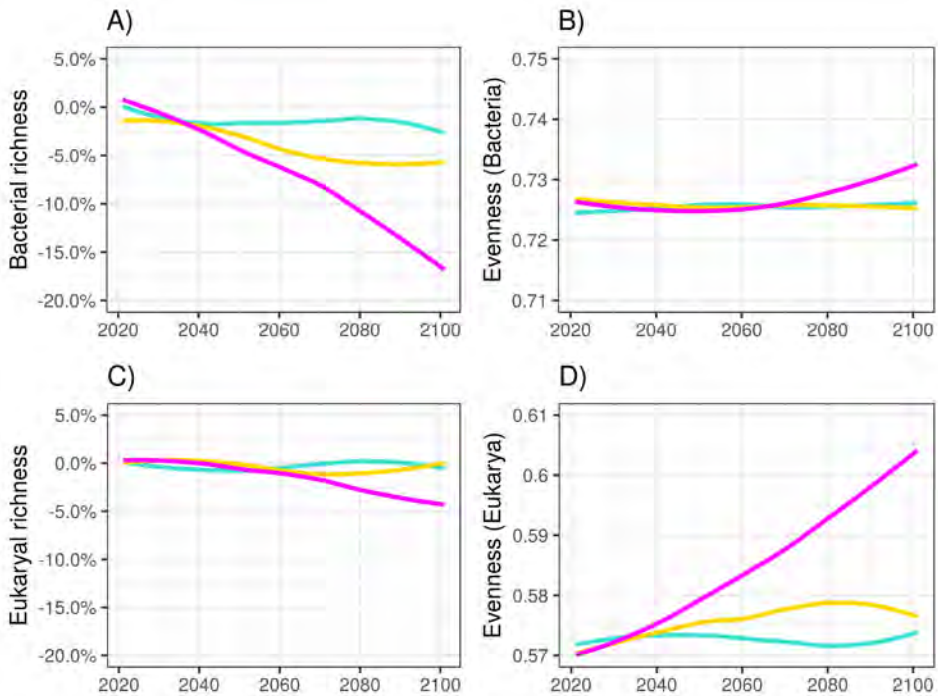


Figure 6.2: **Trends in richness and evenness for the period 2021 – 2100**, for projections of three Representative Concentration Pathways (RCPs), corresponding to emissions decline, stabilization or increase scenarios (2.6, 4.5, 8.5 respectively). A) Relative change from 2020 levels of bacterial OTU richness. B) Evenness for bacterial groups. C) Relative change from 2020 levels of bacterial OTU richness. D) Evenness for eukaryal groups.

ensemble of downscaled RCMs for the maximum, mean and minimum temperatures (Figure C.4). This increase in temperature is expected to influence the richness of microbial communities greatly, with a larger impact as the temperature anomaly gets higher (Figure 6.2A-C). In the worst-case scenario, RCP8.5, we found that bacterial richness would decrease more than 15%, while eukaryal richness would be stable until 2060 and decrease slightly then. These changes in richness would be accompanied by almost no change in bacterial evenness and an increase in eukaryal evenness (Figure 6.2B-D), meaning that the changes in bacteria would be uniformly distributed. In contrast, the changes in eukarya would affect mainly the richest groups. In the worst-case

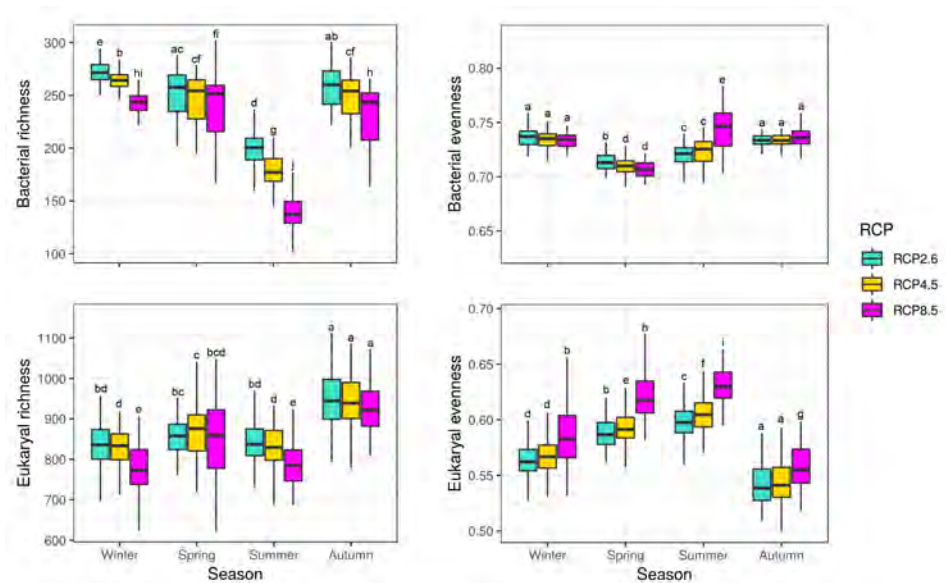


Figure 6.3: **Change in seasonality for the period 2081 – 2100 in the three different scenarios.** *Upper left*, bacterial OTU richness. *Upper right*, bacterial evenness. *Bottom left*, eukaryal OTU richness. *Bottom right*, eukaryal evenness. The boxes are delimited by quartiles Q1 and Q3, and whiskers correspond to 1.5 interquartile ranges from the corresponding quartile.

scenario, the majority of bacterial groups would decrease (Figure C.5), while eukarya would show opposite tendencies between the different groups (Figure C.6), producing stable levels of richness due to replacement. For example, Alphaproteobacteria and Actinobacteria richness would decrease over 15%, with a lesser decrease in Bacteroidetes and stability in Betaproteobacteria, while for eukaryal groups we predict more than 30% increase for Alveolata, stability for Chytridiomycota and decreases for Basidiomycota and Ascomycota. The intermediate scenario, RCP4.5, shows only small changes for Bacteria, and almost no change for Eukarya, both for the whole communities and the specific groups.

The changes in richness were not equally distributed seasonally and got amplified towards the end of our predictions, as Figures C.7 and C.8 suggest, so we show them for the period 2081 – 2100 (Fig-

ure 6.3). For Bacteria, the richness loss in winter and summer is more pronounced, even in the RCP4.5 scenario, also for the abundant groups Actinobacteria, Alphaproteobacteria, and Bacteroidetes, while for Betaproteobacteria we predict slight summer increases (Figure C.9). Evenness highlights again that these losses are almost equally distributed. In the case of Eukarya, winter and summer displayed more richness loss, while Spring and Autumn remained at the same levels in the worst-case scenario, a tendency in part determined by the richest groups, Ascomycota and Basidiomycota. At the same time, Alveolata tends to increase and Chytridiomycota to be somewhat stable (Figure C.10). The evenness for the eukaryal community shows that the richest groups are the most affected in this scenario, RCP8.5. In the intermediate scenario, the changes are negligible. We expect a more marked seasonality for both Bacteria and Eukarya if the worst-case scenario is confirmed.

We repeated the same procedure applied to the aeroplankton communities over bacterial and eukaryal putative pathogens. We found that the best models included the influence of the environment over pathogen groups (Table C.4). We focused on the specific case of plant and human eukaryal pathogens, as those are among the most studied and interesting groups for health concerns (Figure 6.4 and Table C.5). Putative eukaryal plant pathogens, contrary to the typical eukaryal OTU, are more affected by climate than by their origin. The main drivers of the dynamics of this group are minimum and maximum temperatures, SO_4 , and Cl. We expect that the richness of this group might decrease up to a 15% in the worst-case scenario and about a 5% in the intermediate scenario, with no decrease in the RCP2.6 scenario. These losses are higher in summer and winter. In the case of putative human pathogens, the drivers of their dynamics were TP, minimum temperature, K, TN, and Mg. We would expect a slight decrease in the richness of this group in the worst-case scenario, especially in summer, while in more benign scenarios, almost no differences are expected.

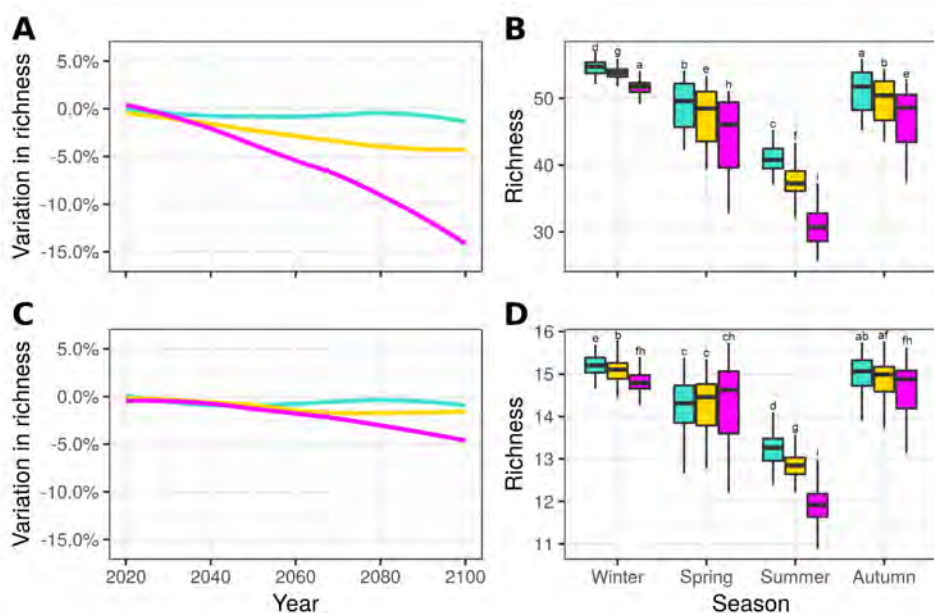


Figure 6.4: **Richness trends and seasonal predictions for plant and human putative eukaryal pathogens.** A) Variation in richness for plant pathogens. B) Seasonal prediction for plant pathogens in 2081 – 2100. C) Variation in richness for human pathogens. D) Seasonal prediction for human pathogens in 2081 – 2100.

6.3 Discussion

In this work, we have used a simple approach based on ecological theory to predict species richness dynamics of airborne microbial communities under three different climate change scenarios (RCP2.6, RCP4.5, RCP8.5), finding richness and composition changes for the most abundant groups. As far as we know, this is the first prediction on the fate of aeroplanktonic communities under climate change. In the following paragraphs, we discuss the main findings of this work, and next, we examine the possible consequences for related communities and ecosystem processes.

We have found that bacterial and eukaryotic communities are differently affected by climatic and chemical variables associated with the origin of the aeroplankton. Bacteria were more affected by climatic variables than by origin, while Eukarya depended more on the origin

than on climate. This result confirms previous research (Cáliz et al., 2018), and reflects the susceptibility of bacterial cells to atmospheric conditions (Chudobova et al., 2015; Smith et al., 2011) while eukaryotic cells may mainly correspond to resistance forms. Bacterial richness decreased whereas evenness remained stable. Conversely, eukaryal richness barely changed while eukaryal evenness increased, which indicates that the richest groups would be proportionally more affected than species-poor groups. Ascomycota and Basidiomycota are among the most affected groups, although some of them are classified as black-yeasts, which are known to be highly resistant against different environmental stresses. These two groups are known to discharge their ascospores and basidiospores under humid conditions (Elbert et al., 2007). Hence, the model may reflect reductions in spore discharge, or, changes in the origin of the aeroplankton due to altered air mass circulation patterns. The considerable decrease in eukaryal plant pathogens also points to this interpretation. Besides, it is interesting that SAR - Alveolata were expected to increase, as they include encysting groups that dominate arid and semi-arid regions (Bates et al., 2013), a circumstance that would support changes in the source.

One of the features of airborne microbial communities is their seasonality (Bowers et al., 2013; Cáliz et al., 2018; Franzetti et al., 2011). Our baseline (RCP2.6) predictions of richness show matches and mismatches with the previous analysis of this temporal series (Cáliz et al., 2018). The seasonal pattern of bacterial richness is confirmed in this study, and we report significantly higher richness in autumn for Eukarya, that were not found to be significant previously. For the analysis based on indicator values, which take into account abundance and occurrence, we found that groups as Acidobacteria, Planctomycetes, Alphaproteobacteria, or Gammaproteobacteria for the bacterial component of the community, and Amoebozoa, Chlorophyta and Alveolata matched their previously reported seasonal trends. In contrast, groups as Ascomycota, Bacteroidetes, Deinococcus-Thermus,

or Deltaproteobacteria do not match the patterns showcased by the indicator values. Other studies also indicate the seasonality of airborne microbes and they are confirmed in our baseline predictions, as the dominance of Actinobacteria in winter (Franzetti et al., 2011) or the dominance of Firmicutes in autumn or Gammaproteobacteria and Bacteroidetes in winter and spring (Bowers et al., 2012). However, the reported dominance of Alphaproteobacteria in summer for both studies was not met in our predictions. In the worst-case scenario (RCP8.5), many of the seasonal trends found would change, e.g., spring would be the richest season for Bacteria or winter and summer would present the lowest eukaryal richness. These changes would tend to affect equally all bacterial groups, which reflects again their increased susceptibility to the environmental filter that represents for them the atmosphere (Chudobova et al., 2015; Delort et al., 2017), whereas they would hit harder the richest eukaryal groups throughout all seasons, probably reflecting the change in origin and air circulation patterns that increased temperatures suggest. The consequences of these seasonal changes may risk diversity as complex underlying associations might be altered (Tonkin et al., 2017).

Many predictions of climate change for ecological communities imply loss or interchange of species. The effects on other communities and ecosystem processes of these richness losses are complex to predict, especially in the field of bioaerosols, where knowledge gaps are still wide (Fröhlich-Nowoisky et al., 2016). Now, we will outline some possible consequences of the richness losses predicted by our models. We will cover themes as pathogen dispersal, sentinel environments, or cloud formation. Remote alpine lakes are especially suited for ecological research and act as sentinel environments (Catalan et al., 2006). Airborne microbes have the potential of colonizing these pristine lakes. However, the importance of these invasions may change as increased temperatures are expected to decrease microbial richness (this study) and affect composition (Bowers et al., 2012) in airborne communi-

ties. Moreover, microbial abundance is expected to decrease due to increased temperatures in alpine lakes (Jiang et al., 2019). Additional elements may influence those communities, as expected increases in aridity and temperatures that reduce soil microbial abundance and diversity (Maestre et al., 2015), increasing dust loads in the atmosphere, changes in soil community composition (Matulich et al., 2015), or concentrated dust depositions in snow (Weil et al., 2017). We might end up with different microbial communities than the current ones, and these changes may percolate and alter the riverine ecosystems after the lakes, due to priority effects, as has been shown in the evolution of a freshwater network (Ortiz-Álvarez et al., 2019; Ruiz-Gonzalez et al., 2015). Bacteria have been shown to help form clouds, although it is unclear whether species identities and status influence their action (Delort et al., 2017). Increased dust loads may compensate partially or totally the possible effect on cloud formation of a less diverse aeroplankton, as we predict.

The phyllosphere is one of the vastest habitats in the Earth (Griffin and Carson, 2015; Vorholt, 2012) and may exert a great influence exchange processes of matter and energy with the atmosphere (Nadkarni et al., 2011), such as nitrification (Guerrieri et al., 2019). It is unclear the importance of airborne microbes on leaf community composition (Griffin and Carson, 2015), although the influence of airborne communities has been shown experimentally (Maignien et al., 2014), so the predicted changes may influence the phyllosphere at local scales. However, the role of airborne microbes as plant pathogens is widely recognized (Brown and Hovmøller, 2002). Our prediction for putative eukaryal plant pathogens indicates a sharp decline in richness for the RCP8.5 scenario, particularly in winter and summer, probably due to the influence of temperature. However, previous studies suggest that the importance of aeroplankton, including airborne bacterial food pathogens, may increase due to higher dust activity and faster winds associated with heavy storms or cyclones (DeLeon-Rodriguez et al., 2013;

Hellberg and Chu, 2015). For putative eukaryal human pathogens, expected decreases are much more limited. In this case, a previous study shows reduced concentrations of airborne pathogens (van Leuken et al., 2016). Nevertheless, further mechanistic approaches are needed to improve predictions of the fate of airborne pathogens under global change (Fröhlich-Nowoisky et al., 2016).

Overall, this pioneering study set out to investigate the dynamics of airborne microbial communities in relation with the environmental variables that affect them, and subsequently, the effect of climate warming on those. Our simulations have shown apparent modifications of the aeroplanktonic communities for three future scenarios that might serve as new hypotheses/warnings to be tested in future research. The most important limitation of this study lies in the fact of being centered in a single but comprehensive temporal dataset, being our findings related to general atmospheric circulation patterns. Notwithstanding the exploratory nature of our study, in the current state of climate 'emergency', it is our imperative as scientists to warn, advise and build predictive tools to help prevent, mitigate or revert the consequences of human actions. The development of mechanistic models based on ecological theory, together with high-quality, comprehensive datasets and transdisciplinary work, may provide us with the ability to cope with some of the ecological challenges that lie ahead at the dawn of the twenty-first century.

Supplementary material

Additional tables and figures can be found in Appendix C.



Image:
Wassily Kandinsky. *Striped*. 1934, The Solomon R. Guggenheim
Museum, New York.

This chapter:
Ontiveros, VJ, Matthews, B, Melián, CJ, Alonso, D. Individualistic
and organismic phytoplankton communities in the reoligotrophication
process of a lake.
In preparation.

Individualistic and organismic phytoplankton communities in the reoligotrophication process of a lake

Abstract

Freshwater ecosystems are threatened by biological invasions, pollution, climate change, and modifications in land use. In Western countries, after decades of increased nutrient loads, freshwater ecosystems have been undergoing the reverse process, reoligotrophication. This process is known to produce critical transitions and alternative stable states, which are difficult to predict and manage. In the context of community ecology, it has been theorized recently that an environmental gradient would reveal the nature of ecological communities, either organismic *sensu* Clements or individualistic *sensu* Gleason. Organismic communities emphasize the role of biotic interactions and positive feedback loops, while the individualistic ones stress the importance of species adaptations to local environments. We have devised a four-step approach to identify individualistic *versus* organismic communities in a long time-series of phytoplankton in lake Zürich, a deep perialpine lake undergoing reoligotrophication. The four steps are *i*) characterization of the environmental change as abrupt or gradual, *ii*) identification of abrupt compositional changes, *iii*) modelization of species responses, and *iv*) search for positive feedback loops. Our results show that the divide between winter and summer communities seems properly individualistic, while there are winter communities that respond abruptly to a gradual environmental trend showing evidence of positive feedback loops, characteristic of organismic communities. This raises the debate about the relevance of critical transitions and alternative stable states in phytoplankton communities and highlights the importance of long-term studies to advance our understanding of community dynamics.

General discussion

Throughout this thesis, I have developed theoretical aspects of TIB for the exploration and quantification of the temporal dynamics of ecological communities, with a particular emphasis on microbial communities. The companion R package 'island' has been presented thoroughly (Chapter 3), and it has been useful in a wide range of issues, such as temporal dynamics (Chapter 4), coexistence (Chapter 5), niche estimation (Chapters 6, 7), global change (Chapters 6, 7) or alternative stable states (Chapter 7). In general, this stochastic model of island biogeography helps to evaluate three of the high-level processes affecting ecological communities (Vellend, 2010, 2016), drift, dispersal, and selection, in an *effective* way. The remaining high-level process, speciation, lies outside of the scope of this thesis, although models linking colonization and extinction with speciation exist (Valente et al., 2015).

Here, first, I will revisit the concept of effective models, such as the ones I have developed throughout my work based on the stochastic dynamic model of island biogeography. Then I will cover three particular central topics in ecological research, with an especial emphasis on microbial communities and taxa: *i*) the relative importance of dispersal and selection, *ii*) the concept of niche, and *iii*) the role of stochasticity. After that, I will briefly examine the limitations and some implications for conservation of the model. Finally, as future work and perspectives, I will outline extensions of the basic stochastic model. Two of those can be equally used as a point of departure for likelihood-based data analysis. First, models that include a reservoir, which can account for inactive/dormant/resistant forms of taxa in, for instance, seed banks

in the soils or sediments, and second, model extensions that consider explicitly the spatial distribution of sampling sites linked by migration.

8.1 An effective dynamic model

The importance of TIB in ecology is paramount. Although recently revisited (Losos and Ricklefs, 2010), TIB is still an active field of research. The incorporation of trophic requirements (Cirtwill and Stouffer, 2016; Gravel et al., 2011; Massol et al., 2017), biotic interactions and environmental constraints (Cazelles et al., 2016b) has been accomplished more recently. Additionally, the derivation of the likelihood of the TIB colonization/extinction processes (Alonso et al., 2015) for generic sampling time schemes paved the way to the research applications and further theoretical developments in this thesis. However, specific questions in island biogeography remain elusive (Santos et al., 2016; Warren et al., 2015). Are communities close to a stationary state? How predictable is community assembly? I have explicitly addressed the first question in chapter 4, along with the temporal evolution of patterns of β -diversity, and the very topic of prediction has been thoroughly addressed throughout this thesis.

Despite the success of TIB, something has been overlooked. The stochastic theory of island biogeography can be used as an effective theory. In science, an *effective* theory is a scientific theory to describe a particular set of observations, but explicitly, without claiming or involving that the assumptions and mechanisms employed in the theory have a direct counterpart in the actual causes of the observed phenomena that the theory seeks to explain. The assumptions of our models are ideal or non-realistic, and, therefore, we are not claiming they are precisely true in nature. While it is straightforward to relax the species equivalence assumption (see Chapters 4, 5, 6, and 7), the assumption of species independence cannot be relaxed within this framework. Despite this limitation, our approach can be even used to look for potential

species interactions, as we have shown in Chapters 3 and 7. We characterized each focal species independently by its associated parameters, assuming that they represent the average influence of the whole range of other potentially interacting species. If we drew a parallel with physics, each species would be embedded in the mean field that the other species create. To some extent, our models are effective, as mean field theories are in other sciences. Future work might examine the influence of the violation of the equivalence assumption over parameter estimation.

8.2 Dispersal and selection in microbial communities

Selection refers to the set of ecological processes that drive differential species growth and death, determining differential overall species fitness and potentially controlling species assembly in distinct communities (Vellend, 2010, 2016). Selection, specifically environmental filtering, has been deemed the main process that structures microbial communities (Lozupone and Knight, 2007; Nemergut et al., 2013), that is, "the environment selects". However, Baas Becking (1934) consideration of "everything is everywhere" has been disputed (Martiny et al., 2006). Microbial taxa show biogeographical patterns (Bahl et al., 2011; Bates et al., 2013), endemic species (Barberán and Casamayor, 2011; Whitaker et al., 2003), and distance – decay relationships (Reche et al., 2005; Soininen et al., 2011). I interpret these pieces of evidence as microbial communities being poised close to a balance between selection and dispersal. Coming back to metacommunity archetypes, where dispersal is related to NT (PD and ME too) and selection to SS, I discuss this balance in the light of the dichotomy NT – SS, where the subtle balance between the intensity of these two processes (usually favoring SS) determines community assembly.

The results of Chapter 5 pointed to the existence of two distinct

components in the studied communities, core and satellite taxa. Core taxa are characterized by their abundance and persistence, and their coexistence is determined by *stabilizing mechanisms*, which tend to decrease niche overlap. On the contrary, satellite taxa are rare and subject to *equalizing mechanisms*, which tend to decrease fitness differences. In my opinion, our ability to detect the two components might be determined by the balance between selection and dispersal. In highly selective environments, an SS scenario, fitness differences among core and satellite taxa might decimate the latter, diminishing the relative importance of the satellite component in favor of the core component that may become stronger and richer in species. However, since rare taxa always outnumber core ones, in highly dispersal-driven environments, a ME scenario, dispersers would dominate, leaving only a few core species, if any, around, which would make very difficult to sort them out from satellite species. As a consequence, this would give rise to a common colonization – persistence signature for all species. A similar framework has been proposed in consumer-resource competition models (Leibold et al., 2017), that identifies *i*) competitively dominant species, with the lowest R^* (SS), *ii*) poorly competitive species, only maintained by high dispersal (ME), and *iii*) subdominant species, transient and probably subjected to trade-offs (PD) (Leibold and Chase, 2017).

Several examples suggest that the balance between selection and dispersal determines community assembly in aquatic microbial communities. Composition of microbial taxa in arctic freshwaters changes directionally from soil waters to lakes, through streams, being structured by initial inoculation and subsequent dispersal and species sorting downslope (Crump et al., 2012). A similar situation is found in boreal freshwater networks (Niño-García et al., 2016; Ruiz-Gonzalez et al., 2015) and lakes in the Pyrenees (Ortiz-Álvarez et al., 2019). The importance of nearby habitats to seed microbial communities seems clear, potentially determining taxa distribution. Sadly, these examples

do not take into account community dynamics, so direct translation to colonization and persistence plots is not possible. In any case, the phylogenetic clustering in colonization and persistence found in Chapter 5 is indicative of niche differences, which may correspond to micro-niches with specific requirements. Thus, both dispersal and selection play an important role and have to be acknowledged in microbial communities.

8.3 Estimating niches

Microbial habitats harbor multiple microhabitats enabling different metabolic niches (Louca et al., 2018), which are just part of the fundamental niches of the different taxa (Muller, 2019). These fundamental niches may change due to horizontal gene transfer between closely related taxa (Fraser et al., 2009, 2007) in ecological time. However, it seems clear that competition- and tolerance-related traits are phylogenetically clustered (Goberna et al., 2014a,b). We are far from fully understanding the genotype-phenotype map, even in microbes. Therefore, our understanding of microbial metabolism and how it is genetically coded and regulated would be a fruitful avenue in the future and will benefit a lot from computational tools. Still, simple approximations to taxa responses to the environment might help define their niches (Lozupone and Knight, 2007) or even make evident fundamental niches and species interactions (Lima-Mendez et al., 2015). But it can not be overlooked that species occupancy and distribution vary over time. This is one of the strengths of my work. Models presented throughout this thesis are dynamic as they consider temporal dependencies explicitly, which sheds new light on the causal explanations of our observations, at least, in an effective manner.

So, results in Chapter 5 suggested that the core sub-community had low niche overlap pointing to the validity of the niche concept in microbes. Then, Chapter 6 showed that models for the most abundant airborne microbes grouping their responses to the environment per-

formed significantly better than community models responding or not to the environment. These models inform about the relative influence of origin or climatic variables over airborne community dynamics and permit to simulate the effects of future climate change. Finally, in the case of phytoplankton (Chapter 7), we accurately modeled phytoplankton responses in a species-specific way. The majority of most likely responses were non-linear, which should be expected as non-linear phenomena control phytoplankton dynamics at different levels; for instance, Michaelis-Menten dynamics are usually considered to depict resource acquisition accurately by single cells (Tilman, 1977). Nevertheless, I would like to stress that the resulting models can be regarded as approximations to the actual species responses to the environment. As previously discussed (see also Chapter 2), our approach uses effective models. As such, we can only identify the effective species response to the environment, which will be closer to reality the closer species conform to the species-independent assumption.

One could ask: do microorganisms respond consistently to the environment? This question can be answered in a Bayesian framework (Hilborn and Mangel, 1997) based on our model selection procedure. Say that we are studying the environmental response of a microbial taxon with a temporal data set. For instance, we could fit models, which each of them can be assimilated to different hypotheses:

- A. Simple colonization – extinction dynamics, with c and e rates.
- B. Linear responses to an environmental covariate, as seen in Eq. 3.1.
- C. Non-linear responses to an environmental covariate, as in Eq. 7.1.

The data, given hypotheses A, B, and C, would have likelihoods $\mathcal{L}(\text{data} | H_A)$, $\mathcal{L}(\text{data} | H_B)$ and $\mathcal{L}(\text{data} | H_C)$. Also, each hypothesis may have a prior probability, $\{H_i\}$, that might be the same for all hypotheses. However, in science, we use data to gather evidence supporting a

particular hypothesis in front of alternative ones. Mathematically, we are interested in the probability or plausibility of a given hypothesis given the data at hand. Therefore, we would have to use Bayes' theorem:

$$Pr\{H_i | data\} = \frac{\mathcal{L}\{data | H_i\}Prior(H_i)}{\sum_j \mathcal{L}\{data | H_j\}Prior(H_j)} \quad (8.1)$$

We would update this probability, using $Pr\{H_i | data\}$ as our new prior probability, when we fit the corresponding models for other observations, to obtain thus a new posterior probability. After several iterations, we might find the hypothesis with a higher probability, or maybe find that our specific organism does not respond consistently to the environment.

As an alternative to the Bayesian approach, one would substitute the likelihoods by the weights of evidence, more in line with the model selection procedures proposed in package 'island', obtaining the following equation:

$$Evidence\{H_i | data\} = \frac{w_i Prior(H_i)}{\sum_j w_j Prior(H_j)} \quad (8.2)$$

This expression would take into account the different number of parameters of the models. Obtaining high values in these evaluations would provide a strong foundation for the prediction of future outcomes of global change, as knowledge about microbial taxa responses to the environment improves (Cavicchioli et al., 2019; Hutchins et al., 2019). In the light of current biodiversity redistribution (Pecl et al., 2017), knowledge about the consistency of responses to the environment is necessary. It might inform about realized and fundamental niches of the species, or even about their intraspecific variability, irrespective of their taxonomic filiation.

8.4 Stochastic models of microbial communities

In this section, I will discuss the suitability of simple stochastic models to understand microbial communities. Already in the late 70s stochastic

models of bacterial growth in chemostats existed (Crump and O'Young, 1979). More recently, the advent of the Unified Neutral Theory of Hubbell (2001) gave rise to models that could be applied to microbial communities. Sloan and colleagues (Ofiteru et al., 2010; Sloan et al., 2006, 2007) developed a model converting Hubbell's neutral theory to a continuous diffusion equation with immigration, reproduction, and death processes, with selective advantages or disadvantages that may depend on environmental variables. Completely neutral approaches have also been used (Lee et al., 2013). Zhou & Ning (2017) reviewed the role of stochasticity in microbial community assembly. They call attention to the stochastic nature of drift, dispersal, and diversification, while selection is deterministic. This appreciation alone should convince us to examine, at least, the relative importance of stochastic and deterministic drivers of the communities. These authors also point us to three ways to study stochasticity in microbial communities: multivariate analyses, neutral stochastic models (also recognized by Widder et al. 2016), and null model analyses. The simplest stochastic model studied in this thesis would correspond to the neutral model fashion. As explained, it should be considered as a "mean-field" model in the sense that colonization and extinction rates include the average effects of biotic and abiotic drivers over the whole community. Therefore, being deeply rooted in ecological theory, and capable of acting as a null model for richness and compositional change, the simple stochastic model is suited to describe the dynamics of microbial communities. Specifically, it has been useful for the understanding of the differential nature of core and satellite taxa (chapter 5), detecting alternative stable states in phytoplankton communities (chapter 7), or predicting the effects of environmental change in microbial communities (chapter 6).

Relaxing the *equivalence* assumption usually improves the likelihood of the estimators producing more accurate descriptions of the dynamics of the different taxa. It is so in the case of airborne microbes and,

particularly, for the core and satellite communities studied in Chapter 5. The distinction of transient, rare, and persistent taxa in microbial communities has been revisited elsewhere (Shade and Gilbert, 2015; Shade et al., 2014). The core – satellite subcommunity distinction aims to provide an ecological interpretation to a continuous pattern, establishing a link between coexistence theory (Chesson, 2000) and early metapopulation theory (Hanski, 1982). By using spatially realistic metacommunity models (see section 8.6.4 and Hanski 2001), further work on the spatio-temporal dynamics of microbial communities may help elucidate to what extent the relative balance of dispersal and selection underlies these rarity patterns.

Alternative stable states have been detected in the gut (Lahti et al., 2014) and vaginal microbiota (Gajer et al., 2012). For free-living microbes, there is evidence of alternative stable states in nectar microbes (Tucker and Fukami, 2014). Nevertheless, it is expected to find more examples of alternative stable states, as time-series studies are extended (Shade et al., 2012). The procedure indicated in Chapter 7 has helped check recent theoretical predictions relating the presence of alternative stable states to the structure of the competition matrix and the organismic nature of microbial communities.

Microbial ecology needs predictive modeling (Prosser et al., 2007). However, the required level of understanding to predict the dynamics and function of microbial communities has not been achieved (Widder et al., 2016). Several approaches have been proposed to overcome this limitation, e.g., the search for principles instead of laws in microbial ecology (Konopka et al., 2015), the use of synthetic microbial communities as model systems to understand taxa interactions (Großkopf and Soyer, 2014), or the development of synthetic ecosystems to control and reduce factors that influence community composition (De Roy et al., 2013). My thesis adds a complementary approach to this research program, which is based on effective dynamic models. Chapter 6 proposes to predict the responses of microbial taxa using linear relationships

with the observed environment, obtaining good fits. Chapter 7 shows non-linear responses in phytoplankton. I believe that, given sufficiently long time-series, predictions would be accurate enough to produce reasonable approximations. The degree of departure from species independence and the suitability of the estimated species responses to the environment would determine how accurate model predictions are and to what extent the outcome of this simulation exercises holds.

8.5 Limitations and implications of the stochastic dynamic model of island biogeography

In several meetings, I have been asked whether the approach presented here can be used for spatial samplings of communities. Sadly, the answer was no. But, as Hanski (2001) realized (see also Gotelli and Kelley 1993), TIB and Classic Metapopulation Theory can be unified within the same framework. Hanski even took a step further, introducing spatially realistic metapopulations (see section 8.6.4 below).

Perhaps, the question that I have heard most, whenever I have presented scientific results from my work here, was:

Does the model use abundances?

Again, the answer was no. Colonization and extinction rates should be regarded as coarse-grained approximations of individual events such as immigration, birth, and death (IBD) affecting populations. Chapter 3 shows how an empirical dependence among colonization and extinction constraints compatible values of intrinsic IBD population rates, and thus informs about compatible population dynamics under the assumption of species independence (see Box 1). In a similar line of inquiry, Fung et al. (2020) estimate species abundance distributions at equilibrium for Hubbell's spatially implicit neutral model, and they also approximate colonization and extinction rates from the elemental

parameters of the model. Although there is not a full correspondence between colonization/extinction rates and abundance-based IBD population dynamics, the relation between the two approaches can be accurately studied. In any case, it is important to recall that the assumptions of *equivalence* and *independence* are not met in the majority of ecological communities. The effective models presented throughout this work are only ideal representations of nature that help think about the causal relationships determining community assembly and structure.

Species richness in local communities is usually found to be in a dynamic equilibrium (see Brown et al. 2001; Dornelas et al. 2014; Velend et al. 2013). Although many dynamic models can be considered to explain this pattern (Chave et al., 2002; Hubbell, 2001; O'Dwyer and Chisholm, 2014), the conceptual simplicity of the stochastic model of island biogeography is still appealing as a first approximation to the true dynamics of ecological communities. In this sense, many of its assumptions may be wrong. Even its emphasis on locally bounded communities and regional species pools might render difficult its application on continuous habitats. However, most of the challenges that global change imposes have to be studied and managed at a community level (Simberloff, 2004). Thus, a null model of local dynamics is needed. Next, I outline some implications for conservation derived from my work in this thesis.

8.5.1 Implications for conservation, management, and restoration

Chapter 4 deals with the temporal dynamics of ecological communities and provides most of the intuitions that follow. First, this work can fill the need for a null model on compositional change, establishing a baseline to compare with (Hillebrand et al., 2018; Supp and Ernest, 2014). This baseline would allow for the early detection of community changes in monitoring schemes. We could identify peaks of community

dissimilarity exceeding expectations. Disturbance studies may also benefit from the null model approach, which provides an expectation of change in community composition (Murphy and Romanuk, 2014). The null model quantifies the temporal scale at which recovery should be evident, about 4-5 times the characteristic time of the community. It is usually recognized that community responses to disturbance seem to be compensated by colonization-extinction dynamics (Elmendorf and Harrison, 2011; Supp and Ernest, 2014), although species traits might be important in some settings affecting reassembly dynamics (Holt et al., 2017). Species with low colonization rates might need actions to reach a community again in the event of disturbances.

Close examination of colonization and extinction rates might produce further insights. On the one hand, very dynamic species, those with high colonization and extinction rates, might be affected to a greater extent by reductions on the connectivity of the community. On the other hand, very stable species, those with low colonization and extinction rates, might need active conservation measures, and being more prone to extinction debt.

Additionally, characteristic times would help solve the *uncertainty principle* associated with the choice of the time interval between censuses (Diamond and May, 1977). A few quick suggestions are mandatory. Ecologists should make sure that they sample at appropriate temporal scales, and the estimation of characteristic times might help assess it. Another suggestion comes from Chapter 3; try to take replicates to assess the detectability of the different species.

Conservation strategies very often require the use of specific models adapted to a particular focal species or community. In this sense, our methods can be used as species distribution models. Further enhancements are possible, and the addition of spatial structure should be a crucial ingredient, as I will discuss in the next section, together with other extensions.

8.6 Extensions and perspectives

Many extensions of TIB have been developed over the years. Since there is a comprehensive review of those extensions and associated predictions (Losos and Ricklefs, 2010), now I briefly review those extensions that appeared more recently, including speciation, trophic structure, interactions, environmental constraints, a balance between neutral and niche processes, and functional trait composition. Speciation has been included explicitly in island biogeography, whether through an individual-based model (Rosindell and Phillimore, 2011) or through likelihood methods that account for branching times obtained from dated phylogenies (Valente et al., 2015). The trophic theory of island biogeography (Gravel et al., 2011; Massol et al., 2017) tries to retain the simplicity of classic TIB assuming that consumer species can be present in a community only if one of their preys is present. Interactions and environmental constraints can be included in TIB through a probabilistic framework (Cazelles et al., 2016b) in discrete time, to the contrary to the continuous-time approach of this thesis. The balance between neutral and niche processes can be studied with a presence-absence model with a general immigration rate and extinction rate depending on the local site carrying capacity, noise and competition, and a simple scaling relation can be drawn between the niche and neutral phases (Dickens et al., 2016; Fisher and Mehta, 2014). The niche phase is favored at high carrying capacity and interaction strength, whereas the neutral phase is favored at low carrying capacity, low interaction diversity, and high stochasticity. Finally, a model has been developed relating the distribution of a trait in a regional community to that of local assemblages, linking trait value, and occupancy (Jacquet et al., 2017). A special mention has to be made; topics related to the relaxation of the assumption of species independence are of great interest. Future directions should aim to understand the effects on colonization and extinction dynamics of species interactions, multiple

trophic levels, and even disease.

In addition to all these extensions that already represent work of impressive quality, I still believe that there is room for further development of models based in TIB. Island biogeography theory would be further advanced by including phenology, traits and the estimation of their effect on colonization-extinction dynamics, dormancy, or realistic depictions of space in metacommunities. With these additions, I aim to make connections with other conceptual domains and advance in the mechanistic depiction of colonization and extinction processes, while retaining the conceptual clarity of TIB. I outline these perspectives in the following subsections.

8.6.1 Phenology

Besides the uses presented throughout my work, our model can also be applied to phenology. It will not be difficult to predict the probability of the daily arrival of bird species, for instance. Our modeling approach might be used for any dichotomic state of a species in a site, e.g., present – absent, masting – non-masting, flowering – non-flowering (see Satake et al. 2013 for a similar model).

8.6.2 Trait extensions

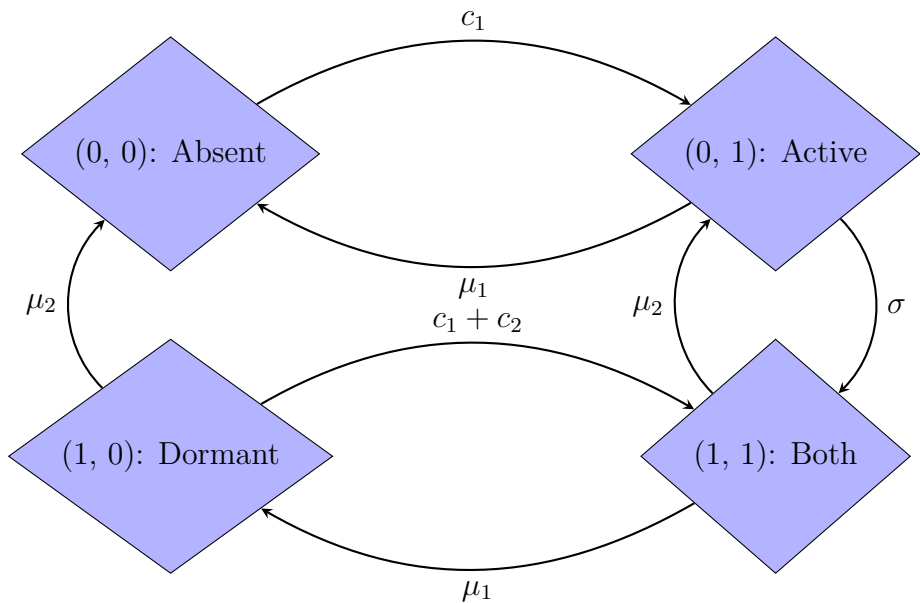
As we have seen earlier, continuous traits can be included in TIB (Jacquet et al., 2017). However, I will present here the most simple situation. Many microbial traits are binary, for instance, spore formation or tolerance to desiccation. The easiest way to take into account one of these traits would be relaxing the *equivalence assumption* and estimate their rates separately. This approach would obviate differences in the rates due to other factors. Thus, we could model the occupancy of a species explicitly depending on a trait as:

$$\frac{dp_i}{dt} = c(1 - p_i) - e'p_i \quad \text{where} \quad e' = e_0 + \sum e_j\psi_{ij} \quad (8.3)$$

where e_j is the effect of the trait over extinction, where ψ_{ij} is a boolean indicating the absence/presence of trait j for species i . It can be easily shown that the occupancy at equilibrium would correspond to $p_i^* = \frac{c}{c+e_0+\sum e_j\psi_{ij}}$. This approach can be extended to multiple independent species and multiple independent traits. It should not be difficult to estimate the influence of each trait. For continuous traits, we will need a function $e_j(\Psi_j)$ that translates the effect of the trait to extinction. This function can even depend on environmental variables, making more mechanistic assumptions and improving the results found in Chapters 6 and 7. More complicated situations have yet to be explored.

8.6.3 Dormancy

Dormancy may play an essential role in microbial communities and has been modeled previously in an environmental gradient without dispersal (Jones and Lennon, 2010). I was inspired by Locey (2010), that develops a graphic model addressing the influence of dormancy in an extension of TIB. To the best of my knowledge, no mathematical model has been developed that explicitly considers the dynamics of two states, dormant–not and active–not, in an island biogeography setting, like the following scheme:



This system is easily translated to the following master equation:

$$\frac{d\vec{P}}{dt} = \begin{pmatrix} -c_1 & \mu_1 & \mu_2 & 0 \\ c_1 & -(\sigma + \mu_1) & 0 & \mu_2 \\ 0 & 0 & -(c_1 + c_2 + \mu_2) & \mu_1 \\ 0 & \sigma & c_1 + c_2 & -(\mu_1 + \mu_2) \end{pmatrix} \vec{P} \quad (8.4)$$

Gladly, the stationary state of the system can be found:

$$\begin{aligned}
P(0, 0)^* &= \frac{\mu_1 \mu_2 (c_1 + c_2 + \mu_1 + \mu_2 + \sigma)}{\mu_2 (c_1 + \mu_1) (c_1 + c_2 + \mu_1 + \mu_2) + c_1 \sigma (c_1 + c_2 + \mu_1) + \mu_2 \sigma (c_1 + \mu_1)}; \\
P(0, 1)^* &= \frac{c_1 \mu_2 (c_1 + c_2 + \mu_1 + \mu_2)}{\mu_2 (c_1 + \mu_1) (c_1 + c_2 + \mu_1 + \mu_2) + c_1 \sigma (c_1 + c_2 + \mu_1) + \mu_2 \sigma (c_1 + \mu_1)}; \\
P(1, 0)^* &= \frac{c_1 \sigma \mu_1}{\mu_2 (c_1 + \mu_1) (c_1 + c_2 + \mu_1 + \mu_2) + c_1 \sigma (c_1 + c_2 + \mu_1) + \mu_2 \sigma (c_1 + \mu_1)}; \\
P(1, 1)^* &= \frac{c_1 \sigma (c_1 + c_2 + \mu_2)}{\mu_2 (c_1 + \mu_1) (c_1 + c_2 + \mu_1 + \mu_2) + c_1 \sigma (c_1 + c_2 + \mu_1) + \mu_2 \sigma (c_1 + \mu_1)}
\end{aligned}$$

The temporal solution of the system is known to exist and corresponds to $\mathbf{P}(t) = e^{t\mathbf{M}} \cdot \mathbf{P}_0$, where \mathbf{M} is the matrix in Eq. 8.4. However, its exact expression is cumbersome, but numerical methods help write a close solution for the temporal evolution of this probability, which allows for the development of a likelihood approach. We know that this likelihood should be well behaved, so there is no major complication in the full development of this pursuit. The explicit calculation of the likelihood given some data will always depend on whether or not observations allow us to define the configuration of the system fully at any time. For instance, a direct application of this model would allow a paleological extension of TIB with the study of a species community in a lake where a sediment compartment records the presence of resistant egg cells. This can be studied by merely dating a sediment core. In this case, for every species, four configurations are possible: (1) absent from the water column and sediment, (2) present in water, (3) present in the sediment, or (4) present in both.

Further work is needed to grasp the implications that dormancy may have on biogeographical patterns fully.

8.6.4 Open spatially realistic metacommunities

Cornerstones of ecology, TIB and CMT flourished independently (Hanski and Simberloff, 1997). However, these two separate foundations would be united when Hanski presented its spatially realistic theory of metapopulation ecology (Hanski 2001, see also Gotelli and Kelley 1993). Hanski developed a model that under certain assumptions can be reduced to the dynamics of either TIB or CMT, and, in addition, it takes into account the spatial location of the patches and its quality, which has been traditionally expressed as area. Hanski put the focus on the structure of the landscape and patch area, but the model was still centered in a single species and, unlike TIB, rarely in communities (but see Holyoak et al. 2005), neglecting migration from outside the system or even an explicit mainland.

Hanski's framework with its versatility can allow us to examine these patterns once we change the focus from a species to the community, as TIB already did by assuming independent and equivalent species, that is, species that do not interact with each other and respond in the same fashion to their environment. Therefore, shifting the focus from metapopulations to metacommunities, we can deduct expected patterns from a range of models that goes from MacArthur and Wilson to Levins' using the common framework of Hanski. The model can deal with different kinds of complexities, namely landscape structure, dispersal kernels, or patch quality (see Grilli et al. 2015). I summarized in Fig. 8.1 the structural complexity of the model, where the horizontal axis corresponds to the intensity of colonization from the mainland, and the vertical axis refers to the proportion of other patches that a species can reach, on average, from a patch, that is, the mean number of patches connected in relation to the total number of patches.

Below I review Levins, MacArthur&Wilson and Hanski models in the same notation. Moreover, I derive a procedure to estimate the stationary state in the joint model with an implicit mainland and an approximation to the mean patch occupancy. A basic model of TIB

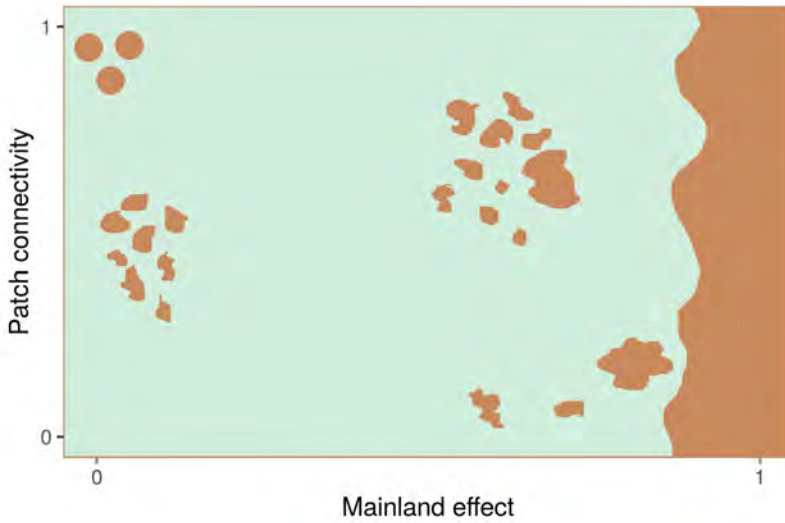


Figure 8.1: **A general spatially explicit model.** Relation among dynamical models emerging from Eq. 8.7, based on patch connectivity, that is, the number of effectively connected patches, N_e , over the total number of patches, N , and the intensity of colonization from the mainland over the total colonization in the system. CMT would correspond to the left-top corner, TIB to the right-bottom corner, the traditional Hanski model would be on the left side, and an open spatially realistic metapopulation model would be in the center-right.

can be formulated as follows:

$$\frac{dS}{dt} = c(S_P - S) - eS \quad (8.5)$$

where S_P accounts for the number of species in the pool, S the number of species present in a site, and c and e correspond respectively to colonization and extinction rates. The classical metapopulation model can be described as follows:

$$\frac{dP}{dt} = cP(N - P) - eP \quad (8.6)$$

where P is the fraction of occupied patches, and N the total number of patches. And the Hanski model can be expressed as follows:

$$\frac{dp_i}{dt} = c_i(1 - p_i) - e_i p_i \quad (8.7)$$

where p_i is the probability of patch i being occupied and c_i and e_i are patch-specific colonization and extinction rates. Note that Eq. 8.5 can

be obtained from Eq. 8.7 assuming a mainland pool of identical species S_P that colonize a patch with independent dynamics, and Eq. 8.6 can be obtained from Eq. 8.7 assuming a set of identical and equally connected patches (Hanski, 2001).

Now, I will follow Alonso and McKane (2002) in their rendering of Hanski's spatially realistic model. They explicitly treat space, assuming a set of patches with known areas. This circumstance allows for the parametrization of the model in a way that maintains its ability to undergo mainland-island and metapopulation dynamics at the same time. Thus, we can construct a model with an equation for each patch giving its occupancy probability:

$$\frac{dp_i}{dt} = (c_M + c_0 \sum_{i \neq j} \exp(-\alpha d_{ij}) A_j p_j)(1 - p_i) - p_i e / A_i \quad (8.8)$$

where c_M corresponds to the colonization rate from the mainland, c_0 is the colonization rate per fully connected patch of unitary area, $1/\alpha$ is the average migration distance, d_{ij} is the distance between patches, p_j the occupancy of the other patches and A corresponds to the area of each patch, that can be considered as a measure of its quality. Alonso and McKane (2002) only studied extinction dynamics for this model. In the following paragraphs, I devised new approaches to study the stationary state and mean patch occupancy of the model, simulate its dynamics, and estimate its parameters.

Stationary state

Although it is not possible to derive a fully analytic solution for the stationary state, one can devise an iterative process that converges to the vector \vec{p} of occupancy probabilities of each patch. Before sketching the procedure, I am going to rewrite Eq. 8.8 in a more convenient form:

$$\frac{d\vec{p}}{dt} = (c_M + c_0 M \vec{p}) \odot (1 - \vec{p}) - \vec{p} \odot (e \oslash \vec{A}) \quad (8.9)$$

where \vec{p} is the vector of occupancy probabilities, \vec{A} is the vector of areas and M is a matrix with diagonal 0, $m_{ij} = \exp(-\alpha d_{ij})A_j$, and \odot (\oslash) corresponds to the Hadamard product (division), the element-wise product (division) of two vectors. As we know that the occupancy probabilities will not change in the stationary state by definition, we can find the following relation after a bit of algebra:

$$\vec{p}^* = (c_0 M \vec{p}^* + c_M) \oslash (c_0 M \vec{p}^* + c_M + e \odot \vec{A}) \quad (8.10)$$

being \vec{p}^* the vector of occupancy probabilities in the equilibrium. Notice that assuming $\vec{A} = 1$, if we set $c_0 = 0$, we automatically have the stationary state for a single species in the TIB and setting $c_M = 0$, we can recover after some algebra an expression equivalent to Eq. 8.6. Also, this expression allows us to recover the approximations to the stationary state in the spatially realistic model of Hanski showed by Grilli *et al.* (2015).

Now we can easily devise an iterative procedure to estimate the vector of patch occupancies in the stationary state, \vec{p}^* :

1. Obtain an eigenvector λ_p of the leading eigenvalue of M_{ij} . Use it as an initial guess for patch occupancies.
2. In Eq. 8.10, substitute \vec{p} by λ_p and obtain a new vector of patch occupancies, \vec{p}' .
3. Using Eq. 8.10 iteratively, find a \vec{p}' that changes less than a threshold δ when compared to the previous \vec{p} .

Approximation to the mean patch occupancy \bar{p} in the equilibrium

Knowing that the dynamics of the model is well described with a one-dimensional approach Hanski (2001), we can assume that the mean patch occupancy \bar{p} will behave as the vector of occupancy probabilities, just receiving the effect of the mean of the sums of the rows in M , N_e ,

so we will have the following expression:

$$\bar{p} \approx \frac{c_0 N_e \bar{p} + c_M}{c_0 N_e \bar{p} + c_M + e/\bar{A}} \quad (8.11)$$

Again, this equation complies with the homologous expression in TIB. However, Hanski followed a slightly different approach, from the equation equivalent to 8.9 he multiplied both sides by \bar{A} . Assuming that the dynamics is given by the leading eigenvalue λ_D of matrix \hat{M} (M made symmetric when we multiplied by \bar{A}), that the mean patch occupancy \bar{p} is the mean of \vec{p}^* , and \bar{A} is the mean of the areas, we have the expression:

$$0 = \bar{A} \frac{d\bar{p}}{dt} = (c_M \bar{A} + c_0 \lambda_D \bar{p})(1 - \bar{p}) - e\bar{p} \quad (8.12)$$

Solving this quadratic equation, we have that

$$\bar{p} = \frac{c_0 \lambda_D - e - \bar{A} c_M + \sqrt{(c_0 \lambda_D - e - \bar{A} c_M)^2 + 4c_0 \lambda_D c_M \bar{A}}}{2c_0 \lambda_D} \quad (8.13)$$

which reduces to the expression derived by Ovaskainen and Hanski (2001), $\bar{p} = 1 - \frac{e}{c_0 \lambda_D}$, when $c_M = 0$.

Simulation and parameter estimation

The dynamics of the model can be simulated using a direct Gillespie algorithm, which relies on the list of possible events and their individual rates, and on its total rate, R , which would follow the next equation:

$$R = \sum_i \left((c_M + c_0 \sum_{i \neq j} \exp(-\alpha d_{ij}) A_j p_j) (1 - p_i) + p_i e/A_i \right) \quad (8.14)$$

Parameter estimation for this general model is another issue that deserves further research. It has been thoroughly studied in the case of populations without a mainland (as reviewed in Etienne et al. 2004), but despite that interest, no exact likelihood methods have been devised to estimate these parameters. However, I believe that an exact likelihood is possible, which would make model parameter estimation

perfectly feasible also for spatio-temporal data by using a likelihood-based procedure, as the general one in package 'island'. As species are considered equivalent and independent, one would just have to follow a heuristic optimization method of the likelihood to obtain the maximum likelihood estimates of model parameters.

Some general considerations

The procedures delineated above, the iterative and the metapopulation capacity approaches, allows us to estimate the mean patch occupancy accurately for the whole metacommunity at the dynamic equilibrium, without simulating the actual dynamics. However, metapopulation capacity is just an approximation that may fail if the first eigenvalue of the landscape matrix is close to the second eigenvalue, which means that we may have at least two disconnected metacommunities. This problem has been addressed in the field of epidemics, as susceptible-infected-susceptible models are analogous to metapopulation theory. Miegheem and Cator (2012) devised a method to correct the first-order approximation given by the leading eigenvalue.

However, simulations will be needed to study β -diversity patterns, an approach that has been proven useful with other models (Muneepeerakul et al., 2008). Interestingly, there are approximations to the spatial correlation among patches (i.e., Ovaskainen and Cornell 2006), which could help derive β -diversity and distance-decay patterns in metacommunities.

The importance of describing general patterns in ecology is paramount, and models like the one outlined here will help us to know the range of variation that specific processes can generate under different sets of parameters. Deductive reasoning through analytical, numerical, and *in silico* approaches, such as the ones presented throughout this thesis, allows us to gain new insights into how ecological communities work.

General conclusions

- We developed an R package, 'island', to analyze temporal patterns in community ecology. The package aims to make easy the application of the Theory of Island Biogeography, from simple models to more complex ones. These models are *effective*, as they are simple approximations to complex temporal dynamics. Now, we can estimate detectability, the influence of environmental variables, and co-occurrence networks in a dynamic temporal setting (Chapter 3).
- The simplest stochastic model of island biogeography studied in this thesis has a temporal scale, called here characteristic time, inversely related to its colonization and extinction rates, that can be estimated from the temporal dynamics of ecological communities and solve the so-called *uncertainty principle* (Chapter 4).
- The characteristic time is also related to compositional change, either measured via STRs, turnover, or the Jaccard Index. Our model can be used as a baseline for richness and compositional change, under the assumptions of *equivalent* and *independent* species, although its *effective* character includes the average effects of the possible deviations from the assumptions (Chapter 4).
- Colonization/extinction simple models can be applied to microbial communities. These models show parameters coherent with environmental knowledge and present adequate accuracy when confronted with replicated samplings (Chapters 4, 5, 6).

- There is a colonization-persistence trade-off in natural microbial communities, first reported here. This trade-off is largely driven by rare taxa that may be subjected preferentially to fitness – equalizing mechanisms (Chapter 5).
- The importance of the environmental filter for airborne bacterial communities is higher than for airborne eukaryal communities. In contrast, aerosol origin has higher importance for airborne eukaryal communities than for airborne bacterial communities (Chapter 6).
- Regional climate predictions would result in general airborne bacterial richness declines, idiosyncratic responses for the eukaryal component, changes in seasonality, and declines in the richness of two major eukaryotic pathogen groups in the worst-case scenario. The intensity of these changes would be related to the extent of climate change (Chapter 6).
- Individualistic and organismic communities can be teased apart in a four-step approach, involving the identification of gradual environmental change, abrupt changes in community composition, modelization of species responses, and characterization of positive feedback loops. Phytoplankton in lake Zürich presents both types of communities, with evidence of alternative stable states (Chapter 7).

- Adler, P. B., HilleRisLambers, J., and Levine, J. M. (2007). A niche for neutrality. *Ecology Letters*, 10(2):95–104.
- Adrian, R., Wilhelm, S., and Gerten, D. (2006). Life-history traits of lake plankton species may govern their phenological response to climate warming. *Global Change Biology*, 12(4):652–661.
- Agawin, N. S. R., Rabouille, S., Veldhuis, M. J. W., Servatius, L., Hol, S., van Overzee, H. M. J., and Huisman, J. (2007). Competition and facilitation between unicellular nitrogen-fixing cyanobacteria and non-nitrogen-fixing phytoplankton species. *Limnology and Oceanography*, 52(5):2233–2248.
- Allen, A. P. and Savage, V. M. (2007). Setting the absolute tempo of biodiversity dynamics. *Ecology Letters*, 10(7):637–646.
- Allouche, O. and Kadmon, R. (2009). A general framework for neutral models of community dynamics. *Ecology Letters*, 12(12):1287–1297.
- Alonso, D., Etienne, R. S., and McKane, A. J. (2006). The merits of neutral theory. *Trends in Ecology & Evolution*, 21(8):451–457.
- Alonso, D. and McKane, A. (2002). Extinction dynamics in mainland-island metapopulations: an N-patch stochastic model. *Bulletin of mathematical biology*, 64(5):913–958.
- Alonso, D., Pinyol-Gallemí, A., Alcoverro, T., and Arthur, R. (2015). Fish community reassembly after a coral mass mortality: higher trophic groups are subject to increased rates of extinction. *Ecology Letters*, 18(5):451–461.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.
- Amend, A. S., Martiny, A. C., Allison, S. D., Berlemont, R., Goulden, M. L., Lu, Y., Treseder, K. K., et al. (2016). Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. *The ISME Journal*, 10(1):109–118.

- Anneville, O., Gammeter, S., and Straile, D. (2005). Phosphorus decrease and climate variability: mediators of synchrony in phytoplankton changes among European peri-alpine lakes. *Freshwater Biology*, 50(10):1731–1746.
- Anneville, O., Ginot, V., Druart, J.-C., and Angeli, N. (2002). Long-term study (1974-1998) of seasonal changes in the phytoplankton in Lake Geneva: a multi-table approach. *Journal of Plankton Research*, 24(10):993–1008.
- Anneville, O., Souissi, S., Gammeter, S., and Straile, D. (2004). Seasonal and inter-annual scales of variability in phytoplankton assemblages: comparison of phytoplankton dynamics in three peri-alpine lakes over a period of 28 years. *Freshwater Biology*, 49(1):98–115.
- Araújo, M. B. and Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16(6):743–753.
- Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J., and Polz, M. F. (2019). A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations. *Cell*, 178(4):820–834.e14.
- Arrhenius, O. (1921). Species and Area. *The Journal of Ecology*, 9(1):95.
- Aßhauer, K. P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data: Fig. 1. *Bioinformatics*, 31(17):2882–2884.
- Auguet, J.-C., Nomokonova, N., Camarero, L., and Casamayor, E. O. (2011). Seasonal changes of freshwater ammonia-oxidizing archaeal assemblages and nitrogen species in oligotrophic alpine lakes. *Applied and environmental microbiology*, 77(6):1937–45.
- Auguet, J.-C., Triadó-Margarit, X., Nomokonova, N., Camarero, L., and Casamayor, E. O. (2012). Vertical segregation and phylogenetic characterization of ammonia-oxidizing Archaea in a deep oligotrophic lake. *The ISME Journal*, 6(9):1786–1797.
- Auguet, O., Pijuan, M., Borrego, C. M., Rodriguez-Mozaz, S., Triadó-

- Margarit, X., Giustina, S. V. D., and Gutierrez, O. (2017). Sewers as potential reservoirs of antibiotic resistance. *Science of the Total Environment*, 605-606:1047–1054.
- Azaele, S., Pigolotti, S., Banavar, J. R., and Maritan, A. (2006). Dynamical evolution of ecosystems. *Nature*, 444(7121):926–928.
- Bahl, J., Lau, M. C. Y., Smith, G. J. D., Vijaykrishna, D., Cary, S. C., Lacap, D. C., Lee, C. K., et al. (2011). Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications*, 2(1):163.
- Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal*, 6(2):343–351.
- Barberán, A. and Casamayor, E. O. (2011). Euxinic Freshwater Hypolimnia Promote Bacterial Endemicity in Continental Areas. *Microbial Ecology*, 61(2):465–472.
- Barberán, A., Henley, J., Fierer, N., and Casamayor, E. O. (2014). Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. *Science of The Total Environment*, 487:187–195.
- Barberán, A., Ladau, J., Leff, J. W., Pollard, K. S., Menninger, H. L., Dunn, R. R., and Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 112(18):5756–61.
- Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., et al. (2011). Has the Earth’s sixth mass extinction already arrived? *Nature*, 471(7336):51–57.
- Bates, S. T., Clemente, J. C., Flores, G. E., Walters, W. A., Parfrey, L. W., Knight, R., and Fierer, N. (2013). Global biogeography of highly diverse protistan communities in soil. *The ISME Journal*, 7(3):652–659.
- Becking, L. B. (1934). *Geobiologie of inleiding tot de milieukunde*.

- Number 18-19. WP Van Stockum & Zoon.
- Bell, T., Ager, D., Song, J.-I., Newman, J. A., Thompson, I. P., Lilley, A. K., and van der Gast, C. J. (2005). Larger Islands House More Bacterial Taxa. *Science*, 308(5730).
- Bengtsson, J. (2010). Applied (meta)community ecology: Diversity and ecosystem services at the intersection of local and regional processes. In *Community Ecology: Processes, Models, and Applications*, volume 9780199228973, pages 115–130. Oxford University Press.
- Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K. D., Branco, P., Van Nes, E. H., Scheffer, M., et al. (2008). Chaos in a long-term experiment with a plankton community. *Nature*, 451(7180):822–825.
- Blenckner, T., Elliott, A., Markensten, H., Pers, C., and Thackeray, S. (2010). Modeling the Effects of Climate Change on the Seasonal Dynamics of Phytoplankton. In *The Impact of Climate Change on European Lakes*, pages 275–292. Springer Netherlands, Dordrecht.
- Booth, T. H., Nix, H. A., Hutchinson, M. F., and Jovanic, T. (1988). Niche analysis and tree species introduction. *Forest Ecology and Management*, 23(1):47–59.
- Bowers, R. M., Clements, N., Emerson, J. B., Wiedinmyer, C., Hannigan, M. P., and Fierer, N. (2013). Seasonal Variability in Bacterial and Fungal Diversity of the Near-Surface Atmosphere. *Environmental Science & Technology*, 47(21):12097–12106.
- Bowers, R. M., McCubbin, I. B., Hallar, A. G., and Fierer, N. (2012). Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment*, 50:41–49.
- Brown, J. H., Ernest, S. M., Parody, J. M., and Haskell, J. P. (2001). Regulation of diversity: maintenance of species richness in changing environments. *Oecologia*, 126(3):321–332.
- Brown, J. H. and Kodric-Brown, A. (1977). Turnover Rates in Insular Biogeography: Effect of Immigration on Extinction. *Ecology*, 58(2):445–449.
- Brown, J. K. M. and Hovmøller, M. S. (2002). Aerial dispersal of

- pathogens on the global and continental scales and its impact on plant disease. *Science (New York, N.Y.)*, 297(5581):537–41.
- Bunin, G. (2017). Ecological communities with Lotka-Volterra dynamics. *Physical Review E*, 95(4):1–8.
- Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference : a practical information-theoretic approach*. Springer.
- Cabral, J. S., Valente, L., and Hartig, F. (2017). Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. *Ecography*, 40(2):267–280.
- Cadotte, M. W., Mai, D. V., Jantz, S., Collins, M. D., Keele, M., and Drake, J. A. (2006). On testing the competition-colonization trade-off in a multispecies assemblage. *The American naturalist*, 168(5):704–9.
- Calcagno, V., Mouquet, N., Jarne, P., and David, P. (2006). Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecology Letters*, 9(8):897–907.
- Cáliz, J., Triadó-Margarit, X., Camarero, L., and Casamayor, E. O. (2018). A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proceedings of the National Academy of Sciences of the United States of America*, 115(48):12229–12234.
- Capitán, J. A., Cuenda, S., and Alonso, D. (2015). How similar can co-occurring species be in the presence of competition and ecological drift? *J. R. Soc. Interface*, 12:20150604.
- Capitan, J. A., Cuenda, S., and Alonso, D. (2017). Stochastic competitive exclusion leads to a cascade of species extinctions. *Journal of Theoretical Biology*, 419:137–151.
- Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., et al. (2011). Moving pictures of the human microbiome. *Genome Biology*, 12(5):R50.
- Carey, C. C. and Rengefors, K. (2010). The cyanobacterium

- Gloeotrichia echinulata stimulates the growth of other phytoplankton. *Journal of Plankton Research*, 32(9):1349–1354.
- Caruso, T., Chan, Y., Lacap, D. C., Lau, M. C. Y., McKay, C. P., and Pointing, S. B. (2011). Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *The ISME journal*, 5(9):1406–13.
- Cassey, P., Blackburn, T. M., Lockwood, J. L., and Sax, D. F. (2006). A stochastic model for integrating changes in species richness and community similarity across spatial scales. *Oikos*, 115(2):207–218.
- Catalan, J., Camarero, L., Felip, M., Pla, S., Ventura, M., Buchaca, T., Bartumeus, F., et al. (2006). High mountain lakes: Extreme habitats and witnesses of environmental changes. *Limnetica*, 25:551–584.
- Cavalier-Smith, T., Brasier, M., and Embley, T. M. (2006). Introduction: how and when did microbes change the world? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1470):845–850.
- Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., Behrenfeld, M. J., et al. (2019). Scientists’ warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, pages 1–18.
- Cazelles, K., Araújo, M. B., Mouquet, N., and Gravel, D. (2016a). A theory for species co-occurrence in interaction networks. *Theor Ecol*, 9:39–48.
- Cazelles, K., Mouquet, N., Mouillot, D., and Gravel, D. (2016b). On the integration of biotic interaction and environmental constraints at the biogeographical scale. *Ecography*, 39(10):921–931.
- Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., and Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere*, 2(2):art24.
- Chase, J. M. and Leibold, M. A. (2003). *Ecological niches : linking classical and contemporary approaches*. University of Chicago Press.
- Chave, J. (2004). Neutral theory and community ecology. *Ecology*

- Letters*, 7(3):241–253.
- Chave, J., Muller-Landau, H. C., and Levin, S. a. (2002). Comparing classical community models: theoretical consequences for patterns of diversity. *The American naturalist*, 159(1):1–23.
- Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, 31(1):343–366.
- Chisholm, R. A., Condit, R., Rahman, K. A., Baker, P. J., Bunyavejchewin, S., Chen, Y.-Y., Chuyong, G., et al. (2014). Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. *Ecology Letters*, 17(7):855–865.
- Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3):591.
- Chudobova, D., Cihalova, K., Jelinkova, P., Zitka, J., Nejdl, L., Guran, R., Klimanek, M., et al. (2015). Effects of Stratospheric Conditions on the Viability, Metabolism and Proteome of Prokaryotic Cells. *Atmosphere*, 6(9):1290–1306.
- Cirtwill, A. R. and Stouffer, D. B. (2016). Knowledge of predator-prey interactions improves predictions of immigration and extinction in island biogeography. *Global Ecology and Biogeography*, 25(7):900–911.
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., et al. (2001). Ecological forecasts: an emerging imperative. *Science (New York, N.Y.)*, 293(5530):657–60.
- Clements, F. E. (1916). *Plant succession; an analysis of the development of vegetation*,. Carnegie Institution of Washington,, Washington,.
- Cody, M. L. and Diamond, J. M. (1975). *Ecology and evolution of communities*. Belknap Press of Harvard University Press.
- Condit, R. (1998). *Tropical Forest Census Plots*. Springer, Berlin.
- Connell, J. H. (1961). The Influence of Interspecific Competition and Other Factors on the Distribution of the Barnacle *Chthamalus Stellatus*. *Ecology*, 42(4):710–723.
- Cordero, O. X. and Polz, M. F. (2014). Explaining microbial genomic di-

- versity in light of evolutionary ecology. *Nature Reviews Microbiology*, 12(4):263–273.
- Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M., and Relman, D. A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science (New York, N.Y.)*, 336(6086):1255–1262.
- Creamean, J. M., Suski, K. J., Rosenfeld, D., Cazorla, A., DeMott, P. J., Sullivan, R. C., White, A. B., et al. (2013). Dust and biological aerosols from the Sahara and Asia influence precipitation in the western U.S. *Science (New York, N.Y.)*, 339(6127):1572–8.
- Crump, B. C., Amaral-Zettler, L. A., and Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *The ISME Journal*, 6(9):1629–1639.
- Crump, K. S. and O’Young, W.-S. C. (1979). Some stochastic features of bacterial constant growth apparatus. *Bulletin of Mathematical Biology*, 41(1):53–66.
- De Roy, K., Marzorati, M., den Abbeele, P., de Wiele, T., and Boon, N. (2013). Synthetic microbial ecosystems: An exciting tool to understand and apply microbial communities. *Environmental Microbiology*, 16:1472–1481.
- DeLeon-Rodriguez, N., Lathem, T. L., Rodriguez-R, L. M., Barazesh, J. M., Anderson, B. E., Beyersdorf, A. J., Ziemba, L. D., et al. (2013). Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7):2575–80.
- Delort, A. M., Văitilingom, M., Joly, M., Amato, P., Wirgot, N., Lallement, A., Sancelme, M., et al. (2017). Clouds: A Transient and Stressing Habitat for Microorganisms. In *Microbial Ecology of Extreme Environments*, pages 215–245. Springer International Publishing, Cham.
- Diamond, J. and Marshall, A. (1976). Origin of the New Hebridean

- avifauna. *Emu*, 76(4):187.
- Diamond, J. M. (1975). Assembly of species communities. In Cody, M. L. and Diamond, J. M., editors, *Ecology and Evolution of Communities*, pages 342–444. Harvard University Press, Cambridge, MA.
- Diamond, J. M. and May, R. M. (1977). Species Turnover Rates on Islands: Dependence on Cencus Interval. *Science*, 197(4300):266–270.
- Dickens, B., Fisher, C. K., and Mehta, P. (2016). Analytically tractable model for community ecology with many species. *Physical Review E*, 94(2):022423.
- Dini-Andreote, F., Stegen, J. C., van Elsas, J. D., and Salles, J. F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences*, page 201414261.
- Dornelas, M., Antão, L. H., Moyes, F., Bates, A. E., Magurran, A. E., Adam, D., Akhmetzhanova, A. A., et al. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, 27(7):760–786.
- Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., and Magurran, A. E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science (New York, N.Y.)*, 344(6181):296–9.
- Dornelas, M., Magurran, A. E., Buckland, S. T., Chao, A., Chazdon, R. L., Colwell, R. K., Curtis, T., et al. (2013). Quantifying temporal change in biodiversity: challenges and opportunities. *Proceedings. Biological sciences*, 280(1750):20121931.
- Dufrêne, M. and Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible assymetrical approach. *Ecological Monographs*, 67(3):345–366.
- Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C., and Fitter, A. H. (2010). Relative roles of niche and neutral processes in structuring a soil microbial community. *The ISME Journal*, 4(3):337–345.
- Elbert, W., Taylor, P. E., Andreae, M. O., and Pöschl, U. (2007).

- Contribution of fungi to primary biogenic aerosols in the atmosphere: wet and dry discharged spores, carbohydrates, and inorganic ions. *Atmospheric Chemistry and Physics*, 7(17):4569–4588.
- Elith, J. and Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):677–697.
- Elmendorf, S. C. and Harrison, S. P. (2011). Is plant community richness regulated over time? Contrasting results from experiments and long-term observations. *Ecology*, 92(3):602–609.
- Elton, C. S. (1927). *Animal ecology*. Macmillan Co., New York,.
- Etienne, R. S., ter Braak, C. J., and Vos, C. C. (2004). Application of Stochastic Patch Occupancy Models to Real Metapopulations. In *Ecology, Genetics and Evolution of Metapopulations*, pages 105–132. Elsevier.
- Evans, S., Martiny, J. B. H., and Allison, S. D. (2017). Effects of dispersal and selection on stochastic assembly in microbial communities. *The ISME Journal*, 11(1):176–185.
- Faust, K., Lahti, L., Gonze, D., de Vos, W. M., and Raes, J. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current Opinion in Microbiology*, 25:56–66.
- Ferrenberg, S., O’Neill, S. P., Knelman, J. E., Todd, B., Duggan, S., Bradley, D., Robinson, T., et al. (2013). Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *The ISME Journal*, 7(6):1102–1111.
- Ferrier, S. and Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43(3):393–404.
- Fisher, C. K. and Mehta, P. (2014). The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13111–6.
- Fisher, J. A. D., Frank, K. T., and Leggett, W. C. (2010). Dynamic macroecology on ecological time-scales. *Global Ecology and Biogeography*, 19(1):1–15.

- Fiske, I. and Chandler, R. (2011). unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10):1–23.
- Fiske, I. and Chandler, R. (2015). Overview of unmarked: An R Package for the Analysis of Data from Unmarked Animals. *R*, pages 1–5.
- Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R., and Bestetti, G. (2011). Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Applied Microbiology and Biotechnology*, 90(2):745–753.
- Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., and Hanage, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science (New York, N. Y.)*, 323(5915):741–6.
- Fraser, C., Hanage, W. P., and Spratt, B. G. (2007). Recombination and the Nature of Bacterial Speciation. *Science*, 315(5811):476–480.
- Fröhlich-Nowoisky, J., Kampf, C. J., Weber, B., Huffman, J. A., Pöhlker, C., Andreae, M. O., Lang-Yona, N., et al. (2016). Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. *Atmospheric Research*, 182:346–376.
- Fung, T., Verma, S., and Chisholm, R. A. (2020). Probability distributions of extinction times, species richness, and immigration and extinction rates in neutral ecological models. *Journal of Theoretical Biology*, 485:110051.
- Gajer, P., Brotman, R. M., Bai, G., Sakamoto, J., Schütte, U. M. E., Zhong, X., Koenig, S. S. K., et al. (2012). Temporal dynamics of the human vaginal microbiota. *Science translational medicine*, 4(132):132ra52.
- Garriga, J. and Bartumeus, F. (2019). *bigMap: Big Data Mapping*. R package version 2.1.0.
- Gerber, G. K. (2014). The dynamic microbiome. *FEBS Letters*, 588(22):4131–4139.
- Gerber, G. K., Onderdonk, A. B., Bry, L., Willsky, A., McOrist, A.,

- Raedt, L. D., and Wrobel, S. (2012). Inferring Dynamic Signatures of Microbes in Complex Host Ecosystems. *PLoS Computational Biology*, 8(8):e1002624.
- Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., et al. (2012). Defining seasonal marine microbial community dynamics. *The ISME Journal*, 6(2):298–308.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.
- Gleason, H. A. (1926). The Individualistic Concept of the Plant Association. *Bulletin of the Torrey Botanical Club*, 53(1):7.
- Goberna, M., García, C., and Verdú, M. (2014a). A role for biotic filtering in driving phylogenetic clustering in soil bacterial communities. *Global Ecology and Biogeography*, 23(12):1346–1355.
- Goberna, M., Navarro-Cano, J. A., Valiente-Banuet, A., Garcia, C., and Verdú, M. (2014b). Abiotic stress tolerance and competition-related traits underlie phylogenetic clustering in soil bacterial communities. *Ecology Letters*, 17(10):1191–1201.
- Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segrè, D., et al. (2018). Emergent simplicity in microbial community assembly. *Science*, 361(6401):469–474.
- Gonze, D., Coyte, K. Z., Lahti, L., and Faust, K. (2018). Microbial communities as dynamical systems. *Current Opinion in Microbiology*, 44:41–49.
- Gonze, D., Lahti, L., Raes, J., and Faust, K. (2017). Multi-stability and the origin of microbial community types. *The ISME Journal*, 11(10):2159–2166.
- Gotelli, N. J. and Kelley, W. G. (1993). A General Model of Metapopulation Dynamics. *Oikos*, 68(1):36.
- Gravel, D., Massol, F., Canard, E., Mouillot, D., and Mouquet, N. (2011). Trophic theory of island biogeography. *Ecology Letters*, 14(10):1010–1016.
- Griffin, E. A. and Carson, W. P. (2015). The Ecology and Natural

- History of Foliar Bacteria with a Focus on Tropical Forests and Agroecosystems. *The Botanical Review*, 81(2):105–149.
- Griffith, D. M., Veech, J. A., and Marsh, C. J. (2016). cooccur : Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software*, 69(Code Snippet 2):1–17.
- Grilli, J., Barabás, G., and Allesina, S. (2015). Metapopulation Persistence in Random Fragmented Landscapes. *PLOS Computational Biology*, 11(5):e1004251.
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*, 34(4):427–433.
- Großkopf, T. and Soyer, O. S. (2014). Synthetic microbial communities. *Current Opinion in Microbiology*, 18:72–77.
- Guerrieri, R., Lecha, L., Mattana, S., Cáliz, J., Casamayor, E. O., Barceló, A., Michalski, G., et al. (2019). Partitioning between atmospheric deposition and canopy microbial nitrification into throughfall nitrate fluxes in a Mediterranean forest. *Journal of Ecology*, pages 1365–2745.13288.
- Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3):147–186.
- Haegeman, B. and Loreau, M. (2011). A mathematical synthesis of niche and neutral theories in community ecology. *Journal of Theoretical Biology*, 269(1):150–165.
- Hallett, L. M., Jones, S. K., MacDonald, A. A. M., Jones, M. B., Flynn, D. F., Ripplinger, J., Slaughter, P., et al. (2016). codyn: An R package of community dynamics metrics. *Methods in Ecology and Evolution*, 7(10):1146–1151.
- Hanski, I. (1982). Dynamics of Regional Distribution: The Core and Satellite Species Hypothesis. *Oikos*, 38(2):210.
- Hanski, I. (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology*, 63(1):151–162.
- Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706):41–49.
- Hanski, I. (1999). *Metapopulation Ecology*. Oxford University Press,

- Oxford.
- Hanski, I. (2001). Spatially realistic theory of metapopulation ecology. *Naturwissenschaften*, 88(9):372–381.
- Hanski, I. and Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404:755–758.
- Hanski, I. and Simberloff, D. (1997). The Metapopulation Approach, Its History, Conceptual Domain, and Application to Conservation. *Metapopulation Biology*, pages 5–26.
- Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10(7):497.
- Hartmann, M., Niklaus, P. A., Zimmermann, S., Schmutz, S., Kremer, J., Abarenkov, K., Lüscher, P., et al. (2014). Resistance and resilience of the forest soil microbiome to logging-associated compaction. *ISME Journal*, 8(1):226–244.
- Hastings, A. (2010). Timescales, dynamics, and ecological understanding. *Ecology*, 91(12):3471–3480.
- Hastings, A. (2016). Timescales and the management of ecological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 113(51):14568–14573.
- Hellberg, R. S. and Chu, E. (2015). Effects of climate change on the persistence and dispersal of foodborne bacterial pathogens in the outdoor environment: A review. *Critical Reviews in Microbiology*, 42(4):1–25.
- Hervàs, A., Camarero, L., Reche, I., and Casamayor, E. O. (2009). Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environmental Microbiology*, 11(6):1612–1623.
- Hervas, A. and Casamayor, E. O. (2009). High similarity between bacterioneuston and airborne bacterial community compositions in a high mountain lake area. *FEMS Microbiology Ecology*, 67(2):219–228.
- Hilborn, R. and Mangel, M. (1997). *The ecological detective : con-*

- fronting models with data*. Princeton University Press.
- Hillebrand, H., Blasius, B., Borer, E. T., Chase, J. M., Downing, J. A., Eriksson, B. K., Filstrup, C. T., et al. (2018). Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology*, 55(1):169–184.
- HilleRisLambers, J., Adler, P., Harpole, W., Levine, J., and Mayfield, M. (2012). Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):227–248.
- Holt, R. E., Brown, C. J., Schlacher, T. A., Sheldon, F., Balcombe, S. R., and Connolly, R. M. (2017). Species traits and connectivity constrain stochastic community re-assembly. *Scientific Reports*, 7(1):14424.
- Holyoak, M., Leibold, M. A., and Holt, R. D., editors (2005). *Metacommunities: Spatial Dynamics and Ecological Communities*. University Of Chicago Press, Chicago.
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., and Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1):523–549.
- Hortal, J., De Marco Jr, P., Santos, A. M. C., and Diniz-Filho, J. A. F. (2012). Integrating biogeographical processes and local community assembly. *Journal of Biogeography*, 39(4):627–628.
- Hortal, J., Triantis, K. A., Meiri, S., Thébault, E., and Sfenthourakis, S. (2009). Island species richness increases with habitat diversity. *The American naturalist*, 174(6):E205–17.
- Hoglund, J. (2009). *Evolutionary conservation genetics*. Oxford University Press.
- Huang, W., Traulsen, A., Werner, B., Hiltunen, T., and Becks, L. (2017). Dynamical trade-offs arise from antagonistic coevolution and decrease intraspecific diversity. *Nature Communications*, 8(1):2059.

- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Hubbell, S. P. (2006). Neutral theory and the evolution of ecological equivalence. *Ecology*, 87(6):1387–1398.
- Hubbell, S. P., Condit, R., and Foster, R. B. (2005). Barro Colorado forest census plot data. URL: <http://ctfs.si.edu/webatlas/datasets/bci>.
- Hutchins, D. A., Jansson, J. K., Remais, J. V., Rich, V. I., Singh, B. K., and Trivedi, P. (2019). Climate change microbiology — problems and perspectives. *Nature Reviews Microbiology*, 17(6):391–396.
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0):415–427.
- Hutchinson, G. E. (1959). Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *The American Naturalist*, 93(870):145–159.
- Ibáñez, C., Alcaraz, C., Caiola, N., Rovira, A., Trobajo, R., Alonso, M., Duran, C., et al. (2012). Regime shift from phytoplankton to macrophyte dominance in a large river: Top-down versus bottom-up effects. *Science of The Total Environment*, 416:314–322.
- Jacquet, C., Mouillot, D., Kulbicki, M., and Gravel, D. (2017). Extensions of Island Biogeography Theory predict the scaling of functional trait composition with habitat area and isolation. *Ecology Letters*, 20(2):135–146.
- Jessup, C. M. and Bohannan, B. J. M. (2008). The shape of an ecological trade-off varies with environment. *Ecology Letters*, 11(9):947–959.
- Jiang, Y., Huang, H., Ma, T., Ru, J., Blank, S., Kurmayer, R., and Deng, L. (2019). Temperature Response of Planktonic Microbiota in Remote Alpine Lakes. *Frontiers in Microbiology*, 10:1714.
- Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., and Wilmschurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science (New York, N. Y.)*, 356(6335):270–275.

- Jones, S. E. and Lennon, J. T. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):5881–5886.
- Kalyuzhny, M., Kadmon, R., and Shnerb, N. M. (2015). A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities. *Ecology Letters*, 18(6).
- Kämpfer, P. and Glaeser, S. P. (2013). Prokaryote Characterization and Identification. In *The Prokaryotes*, pages 123–147. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., Ding, H., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science (New York, N.Y.)*, 344(6182):416–420.
- Kazamia, E., Helliwell, K. E., Purton, S., and Smith, A. G. (2016). How mutualisms arise in phytoplankton communities: building eco-evolutionary principles for aquatic microbes. *Ecology Letters*, 19(7):810–822.
- Kéfi, S., Holmgren, M., and Scheffer, M. (2016). When can positive interactions cause alternative stable states in ecosystems? *Functional Ecology*, 30(1):88–97.
- Kellogg, C. A. and Griffin, D. W. (2006). Aerobiology and the global transport of desert dust. *Trends in Ecology & Evolution*, 21(11):638–644.
- Kendall, D. G. (1948). On some modes of population growth leading to r. a. fisher’s logarithmic series distribution. *Biometrika*, 35:6–15.
- Klausmeier, C. A., Litchman, E., and Levin, S. A. (2004). Phytoplankton growth and stoichiometry under multiple nutrient limitation. *Limnology and Oceanography*, 49(4part2):1463–1470.
- Konopka, A., Lindemann, S., and Fredrickson, J. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. *The ISME Journal*, 9(7):1488–1495.

- Korhonen, J. J., Soinen, J., and Hillebrand, H. (2010). A quantitative analysis of temporal turnover in aquatic species assemblages across ecosystems. *Ecology*, 91(2):508–517.
- Lahti, L., Salojärvi, J., Salonen, A., Scheffer, M., and de Vos, W. M. (2014). Tipping elements in the human intestinal ecosystem. *Nature Communications*, 5(1):4344.
- Laliberté, E., Legendre, P., and Shipley, B. (2014). FD: Measuring functional diversity from multiple traits, and other tools for functional ecology. *R package*, (1):0–12.
- Langenheder, S. and Székely, A. J. (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *The ISME journal*, 5(7):1086–1094.
- Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9):814–821.
- Lavorel, S. and Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, 16(5):545–556.
- Lee, J. E., Buckley, H. L., Etienne, R. S., and Lear, G. (2013). Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *FEMS Microbiology Ecology*, 86(2):288–302.
- Legendre, P., Dallot, S., and Legendre, L. (1985). Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. *The American Naturalist*, 125(2):257–288.
- Leibold, M. A. and Chase, J. M. (2017). *Metacommunity Ecology, Volume 59*. Princeton University Press.
- Leibold, M. A., Chase, J. M., and Ernest, S. K. (2017). Community assembly and the functioning of ecosystems: how metacommunity processes alter ecosystems attributes. *Ecology*, 98(4):909–919.

- Leigh, E. G. (2007). Neutral theory: a historical perspective. *Journal of Evolutionary Biology*, 20(6):2075–2091.
- Letten, A. D., Ke, P.-J., and Fukami, T. (2017). Linking modern coexistence theory and contemporary niche theory. *Ecological Monographs*, 87(2):161–177.
- Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, 73(6):1943–1967.
- Levins, R. (1969). Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America*, 15(3):237–240.
- Liautaud, K., van Nes, E. H., Barbier, M., Scheffer, M., and Loreau, M. (2019). Superorganisms or loose collections of species? A unifying theory of community patterns along environmental gradients. *Ecology Letters*, 22(8):ele.13289.
- Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., et al. (2015). Determinants of community structure in the global plankton interactome. *Science*, 348(6237):1262073–1262073.
- Lindström, E. S. and Langenheder, S. (2012). Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, 4(1):1–9.
- Lindström, E. S. and Östman, Ö. (2011). The Importance of Dispersal for Bacterial Community Composition and Functioning. *PLoS ONE*, 6(10):e25883.
- Litchman, E., Edwards, K. F., and Klausmeier, C. A. (2015). Microbial resource utilization traits and trade-offs: Implications for community structure, functioning, and biogeochemical impacts at present and in the future.
- Livingston, G., Matias, M., Calcagno, V., Barbera, C., Combe, M., Leibold, M. A., and Mouquet, N. (2012). Competition-colonization dynamics in experimental bacterial metacommunities. *Nature Communications*, 3(1):1234.

- Locey, K. J. (2010). Synthesizing traditional biogeography with microbial ecology: The importance of dormancy. *Journal of Biogeography*, 37(10):1835–1841.
- Loreau, M. (2010). *From populations to ecosystems: Theoretical foundations for a new ecological synthesis*. Princeton University Press.
- Losos, J. B. and Ricklefs, R. E., editors (2010). *The Theory of Island Biogeography Revisited*. Princeton University Press, Princeton, NJ. 476 Pages.
- Lotter, A. F. (1998). The recent eutrophication of Baldeggersee (Switzerland) as assessed by fossil diatom assemblages. *The Holocene*, 8(4):395–405.
- Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., Ackermann, M., et al. (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, 2(6):936–943.
- Lozupone, C. A. and Knight, R. (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences*, 104(27):11436–11440.
- Lynch, M. D. J. and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, 13(4):217–229.
- MacArthur, R. H. and Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, 17:373–387.
- MacArthur, R. H. and Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton.
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., and Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8):2200–2207.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, A. J. a., and Langtimm, C. a. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, 83(8):2248–2255.

- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., and Hines, J. E. (2017). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence: Second Edition*. Elsevier.
- Maestre, F. T., Delgado-Baquerizo, M., Jeffries, T. C., Eldridge, D. J., Ochoa, V., Gozalo, B., Quero, J. L., et al. (2015). Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proceedings of the National Academy of Sciences of the United States of America*, 112(51):15684–9.
- Magurran, A. E. (2008). Diversity Over Time. *Folia Geobotanica*, 43(3):319–327.
- Magurran, A. E. and Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422(6933):714–716.
- Maignien, L., DeForce, E. A., Chafee, M. E., Eren, A. M., and Simmons, S. L. (2014). Ecological succession and stochastic variation in the assembly of *Arabidopsis thaliana* phyllosphere communities. *mBio*, 5(1):e00682–13.
- Margalef, R. (1963). On certain unifying principles in ecology. *The American Naturalist*, 97(897):357–374.
- Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2):102–112.
- Massol, F., Dubart, M., Calcagno, V., Cazelles, K., Jacquet, C., Kéfi, S., and Gravel, D. (2017). Island Biogeography of Food Webs. volume 56, pages 183–262. Academic Press.
- Matthews, B. and Pomati, F. (2012). Reversal in the relationship between species richness and turnover in a phytoplankton community. *Ecology*, 93(11):2435–2447.
- Matulich, K. L., Weihe, C., Allison, S. D., Amend, A. S., Berlemont, R., Goulden, M. L., Kimball, S., et al. (2015). Temporal variation

- overshadows the response of leaf litter microbial communities to simulated global change. *ISME Journal*, 9(11):2477–2489.
- May, F., Gerstner, K., McGlinn, D. J., Xiao, X., and Chase, J. M. (2018). mobsim: An R package for the simulation and measurement of biodiversity across spatial scales. *Methods in Ecology and Evolution*.
- May, F., Huth, A., and Wiegand, T. (2015). Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B*, 282(1802):20141657.
- May, R. M. (2010). Foreword. In Losos, J. B. and Ricklefs, R. E., editors, *The Theory of Island Biogeography Revisited*, pages vii–ix. Princeton University Press, Princeton and Oxford.
- Mayfield, M. M. and Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*.
- Mazar, Y., Cytryn, E., Erel, Y., and Rudich, Y. (2016). Effect of Dust Storms on the Atmospheric Microbiome in the Eastern Mediterranean. *Environmental Science & Technology*, 50(8):4194–4202.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, G. B., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science (New York, N.Y.)*, 331(6014):176–182. 21163965[pmid].
- Mieghem, P. V. and Cator, E. (2012). Epidemics in networks with nodal self-infection and the epidemic threshold. *Physical Review E*, 86.
- Morris, W. F. and Doak, D. F. (2002). *Quantitative conservation biology :theory and practice of population viability analysis /*. Sinauer Associates.
- Morrison-Whittle, P. and Goddard, M. R. (2015). Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. *The ISME Journal*, 9(9):2003–2011.
- Muller, E. E. L. (2019). Determining Microbial Niche Breadth in the Environment for Better Ecosystem Fate Predictions. *mSystems*, 4(3).

- Muller-Landau, H. C. (2010). The tolerance-fecundity trade-off and the maintenance of diversity in seed size. *Proceedings of the National Academy of Sciences of the United States of America*, 107(9):4242–4247.
- Muneepeerakul, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A., and Rodriguez-Iturbe, I. (2008). Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. *Nature*, 453(7192):220–222.
- Murphy, G. E. P. and Romanuk, T. N. (2014). A meta-analysis of declines in local species richness from human disturbances. *Ecology and Evolution*, 4(1):91–103.
- Nadkarni, N. M., Parker, G. G., and Lowman, M. D. (2011). Forest canopy studies as an emerging field of science. *Annals of Forest Science*, 68(2):217–224.
- Naimi, B. and Araújo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39(4):368–375.
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O’Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., et al. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews*, 77(3):342–356.
- Niehus, R., Mitri, S., Fletcher, A. G., and Foster, K. R. (2015). Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nature Communications*, 6(1):8924.
- Niño-García, J. P., Ruiz-González, C., and del Giorgio, P. A. (2016). Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *The ISME Journal*, pages 1–12.
- Nõges, P., Adrian, R., Anneville, O., Arvola, L., Blenckner, T., George, G., Jankowski, T., et al. (2010). The Impact of Variations in the Climate on Seasonal Dynamics of Phytoplankton. In *The Impact of Climate Change on European Lakes*, pages 253–274. Springer

- Netherlands, Dordrecht.
- O'Dwyer, J. P. and Chisholm, R. (2014). A mean field model for competition: From neutral ecology to the Red Queen. *Ecology Letters*, 17(8):961–969.
- Ofteru, I. D., Lunn, M., Curtis, T. P., Wells, G. F., Criddle, C. S., Francis, C. A., and Sloan, W. T. (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proceedings of the National Academy of Sciences of the United States of America*, 107(35):15345–50.
- Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., and O'Hara, R. (2016). Vegan: community ecology package. *R package 2.3-3*, pages Available at: <https://cran.r-project.org/web/packa>.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., et al. (2019). *vegan: Community Ecology Package*. R package version 2.5-4.
- Ontiveros, V. J., Capitán, J. A., Arthur, R., Casamayor, E. O., and Alonso, D. (2019). Colonization and extinction rates estimated from temporal dynamics of ecological communities: The island r package. *Methods in Ecology and Evolution*, 10(7):1108–1117.
- Ortiz-Álvarez, R., Cáliz, J., Camarero, L., and Casamayor, E. O. (2019). Regional community assembly drivers and microbial environmental sources shaping bacterioplankton in an alpine lacustrine district (Pyrenees, Spain). *Environmental Microbiology*, pages 1462–2920.14848.
- Ovaskainen, O. and Cornell, S. J. (2006). Asymptotically exact analysis of stochastic metapopulation dynamics with explicit spatial structure. *Theoretical Population Biology*, 69:13–33.
- Ovaskainen, O. and Hanski, I. (2001). Spatially Structured Metapopulation Models: Global and Local Assessment of Metapopulation Capacity. *Theoretical Population Biology*, 60(4):281–302.
- Page, T., Smith, P. J., Beven, K. J., Jones, I. D., Elliott, J. A., Maberly, S. C., Mackay, E. B., et al. (2018). Adaptive forecasting of

- phytoplankton communities. *Water Research*, 134:74–85.
- Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):637–669.
- Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., et al. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science (New York, N.Y.)*, 355(6332):eaai9214.
- Philippot, L., Andersson, S. G. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., and Hallin, S. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology*, 8(7):523–529.
- Pholchan, M. K., Baptista, J. d. C., Davenport, R. J., Sloan, W. T., and Curtis, T. P. (2013). Microbial community assembly, theory and rare functions. *Frontiers in Microbiology*, 4:68.
- Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in genetics : TIG*, 29(3):170–175.
- Pomati, F., Matthews, B., Jokela, J., Schildknecht, A., and Ibelings, B. W. (2012). Effects of re-oligotrophication and climate warming on plankton richness and community stability in a deep mesotrophic lake. *Oikos*, 121(8):1317–1327.
- Pomati, F., Matthews, B., Seehausen, O., and Ibelings, B. W. (2017). Eutrophication and climate warming alter spatial (depth) co-occurrence patterns of lake phytoplankton assemblages. *Hydrobiologia*, 787(1):375–385.
- Pomati, F., Tellenbach, C., Matthews, B., Venail, P., Ibelings, B. W., and Ptacnik, R. (2015). Challenges and prospects for interpreting long-term phytoplankton diversity changes in Lake Zurich (Switzerland). *Freshwater Biology*, 60(5):1052–1059.
- Preston, F. W. (1948). The Commonness, And Rarity, of Species. *Ecology*, 29(3):254–283.

- Preston, F. W. (1960). Time and Space and the Variation of Species. *Ecology*, 41(4):611–627.
- Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., Green, J. L., et al. (2007). The role of ecological theory in microbial ecology. *Nature reviews. Microbiology*, 5(5):384–392.
- Reche, I., Pulido-Villena, E., Morales-Baquero, R., and Casamayor, E. O. (2005). Does ecosystem size determine aquatic bacterial richness? *Ecology*, 86(7):1715–1722.
- Ridenhour, B. J., Brooker, S. L., Williams, J. E., Van Leuven, J. T., Miller, A. W., Dearing, M. D., and Remien, C. H. (2017). Modeling time-series data from microbial communities. *ISME Journal*, 11(11).
- Ripple, W. J., Wolf, C., Newsome, T. M., Galetti, M., Alamgir, M., Crist, E., Mahmoud, M. I., et al. (2017). World Scientists’ Warning to Humanity: A Second Notice. *BioScience*, 67(12):1026–1028.
- Rogora, M., Frate, L., Carranza, M., Freppaz, M., Stanisci, A., Bertani, I., Bottarin, R., et al. (2018). Assessment of climate change effects on mountain ecosystems through a cross-site analysis in the Alps and Apennines. *Science of The Total Environment*, 624:1429–1442.
- Rosenzweig, M. L. (2001). 14. Preston’s Ergodic Conjecture: The Accumulation of Species in Space and Time. In McKinney, M. L. and Drake, J. A., editors, *Biodiversity Dynamics*. Columbia University Press, New York Chichester, West Sussex.
- Rosindell, J., Hubbell, S. P., and Etienne, R. S. (2011). The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends in Ecology & Evolution*, 26(7):340–348.
- Rosindell, J. and Phillimore, A. B. (2011). A unified model of island biogeography sheds light on the zone of radiation. *Ecology Letters*, 14(6):552–560.
- Ruiz-Gonzalez, C., Niño-García, J. P., and del Giorgio, P. A. (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters*, 18(11):1198–1206.

- Russell, G. J., Diamond, J. M., Pimm, S. L., and Reed, T. M. (1995). A Century of Turnover: Community Dynamics at Three Timescales. *The Journal of Animal Ecology*, 64(5):628.
- Rühland, K., Paterson, A. M., and Smol, J. P. (2008). Hemispheric-scale patterns of climate-related shifts in planktonic diatoms from North American and European lakes. *Global Change Biology*, 14(11):2740–2754.
- Santos, A. M. C., Field, R., and Ricklefs, R. E. (2016). New directions in island biogeography. *Global Ecology and Biogeography*, 25(7):751–768.
- Satake, A., Kawagoe, T., Saburi, Y., Chiba, Y., Sakurai, G., and Kudoh, H. (2013). Forecasting flowering phenology under climate warming by modelling the regulatory dynamics of flowering-time genes. *Nature Communications*, 4(1):2303.
- Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press.
- Scheffer, M. and Carpenter, S. R. (2003). Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology and Evolution*, 18(12):648–656.
- Schlegel, H. G. and Jannasch, H. W. (2013). Prokaryotes and Their Habitats. In *The Prokaryotes*, pages 39–80. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Shade, A., Caporaso, J. G., Handelsman, J., Knight, R., and Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. *The ISME journal*, 7(8):1493–506.
- Shade, A. and Gilbert, J. A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology*, 23(6):335–340.
- Shade, A., Jones, S. E., Gregory Caporaso, J., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J. A. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio*, 5(4).

- Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Bürgmann, H., Huber, D. H., et al. (2012). Fundamentals of Microbial Community Resistance and Resilience. *Frontiers in Microbiology*, 3:417.
- Shurin, J. B., Amarasekare, P., Chase, J. M., Holt, R. D., Hoopes, M. F., and Leibold, M. A. (2004). Alternative stable states and regional community structure. *Journal of Theoretical Biology*, 227(3):359–368.
- Siepielski, A. M., Hung, K.-L., Bein, E. E. B., and McPeck, M. A. (2010). Experimental evidence for neutral community dynamics governing an insect assemblage. *Ecology*, 91(3):847–857.
- Simberloff, D. (1969). Experimental zoogeography on islands: A model for insular colonization. *Ecology*, 50:296–314.
- Simberloff, D. (2004). Community Ecology: Is It Time to Move On? *The American Naturalist*, 163(6):787–799.
- Simberloff, D. and Wilson, E. O. (1969). Experimental zoogeography on islands: The colonization of empty islands. *Ecology*, 50:278–296.
- Simberloff, D. S. (1974). Equilibrium Theory of Island Biogeography and Ecology. *Annual Review of Ecology and Systematics*, 5(1):161–182.
- Sinclair, L., Ijaz, U. Z., Jensen, L. J., Coolen, M. J., Gubry-Rangin, C., Chroňáková, A., Oulas, A., et al. (2016). Seqenv : linking sequences to environments through text mining. *PeerJ*, 4:e2690.
- Singh, B. K., Bardgett, R. D., Smith, P., and Reay, D. S. (2010). Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nature Reviews Microbiology*, 8(11):779–790.
- Sloan, W. T., Lunn, M., Woodcock, S., Head, I. M., Nee, S., and Curtis, T. P. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology*, 8(4):732–740.
- Sloan, W. T., Woodcock, S., Lunn, M., Head, I. M., and Curtis, T. P. (2007). Modeling Taxa-Abundance Distributions in Microbial Communities using Environmental Sequence Data. *Microbial Ecology*, 53(3):443–455.

- Smith, D. J., Griffin, D. W., McPeters, R. D., Ward, P. D., and Schuerger, A. C. (2011). Microbial survival in the stratosphere and implications for global dispersal. *Aerobiologia*, 27(4):319–332.
- Soininen, J. (2012). Macroecology of unicellular organisms - patterns and processes.
- Soininen, J., Korhonen, J. J., Karhu, J., and Vetterli, A. (2011). Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. *Limnology and Oceanography*, 56(2):508–520.
- Solé, R. V., Alonso, D., and McKane, A. J. (2002). Self-organized instability in complex ecosystems. *Philosophical Transactions of the Royal Society, London. Series B*, 357:667–681.
- Solé, R. V., Alonso, D., and Saldaña, J. (2004). Habitat fragmentation and biodiversity collapse in neutral communities. *Ecological Complexity*, 1(1):65–75.
- Sommer, U. (1985). Seasonal Succession of Phytoplankton in Lake Constance. *BioScience*, 35(6):351–357.
- Souffreau, C., Pecceu, B., Denis, C., Rummens, K., and De Meester, L. (2014). An experimental analysis of species sorting and mass effects in freshwater bacterioplankton. *Freshwater Biology*.
- Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal*, 6(9):1653–1664.
- Stroud, J. T., Bush, M. R., Ladd, M. C., Nowicki, R. J., Shantz, A. A., and Sweatman, J. (2015). Is a community still a community? Reviewing definitions of key terms in community ecology. *Ecology and Evolution*, 5(21):4757–4765.
- Subirats, J., Triadó-Margarit, X., Mandaric, L., Acuña, V., Balcázar, J. L., Sabater, S., and Borrego, C. M. (2017). Wastewater pollution differently affects the antibiotic resistance gene pool and biofilm bacterial communities across streambed compartments. *Molecular Ecology*, 26(20):5567–5581.

- Suding, K. N., Lavorel, S., Chapin, F. S., Cornelissen, J. H. C., Díaz, S., Garnier, E., Goldberg, D., et al. (2008). Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants. *Global Change Biology*, 14(5):1125–1140.
- Supp, S. R. and Ernest, S. K. M. (2014). Species-level and community-level responses to disturbance: a cross-community analysis. *Ecology*, 95(7):1717–1723.
- Sutherland, W. J., Freckleton, R. P., Godfray, H. C. J., Beissinger, S. R., Benton, T., Cameron, D. D., Carmel, Y., et al. (2013). Identification of 100 fundamental ecological questions. *Journal of Ecology*, 101(1):58–67.
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., et al. (2017). A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature*, 551(7681):457–463.
- Tilman, D. (1977). Resource Competition between Plankton Algae: An Experimental and Theoretical Approach. *Ecology*, 58(2):338–348.
- Tilman, D. (1994). Competition and Biodiversity in Spatially Structured Habitats. *Ecology*, 75(1):2–16.
- Tonkin, J. D., Bogan, M. T., Bonada, N., Rios-Touma, B., and Lytle, D. A. (2017). Seasonality and predictability shape temporal species diversity. *Ecology*, 98(5):1201–1216.
- Triadó-Margarit, X., Caliz, J., Reche, I., and Casamayor, E. O. (2019). High similarity in bacterial bioaerosol compositions between the free troposphere and atmospheric depositions collected at high-elevation mountains. *Atmospheric Environment*, 203:79–86.
- Triadó-Margarit, X., Veillette, M., Duchaine, C., Talbot, M., Amato, F., Minguillón, M. C., Martins, V., et al. (2017). Bioaerosols in the Barcelona subway system. *Indoor Air*, 27(3):564–575.
- Trolle, D., Elliott, J. A., Mooij, W. M., Janse, J. H., Bolding, K., Hamilton, D. P., and Jeppesen, E. (2014). Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling & Software*, 61:371–379.

- Tucker, C. M. and Fukami, T. (2014). Environmental variability counteracts priority effects to facilitate species coexistence: evidence from nectar microbes. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778):20132637.
- Tuljapurkar, S. (1997). Stochastic Matrix Models. In *Structured-Population Models in Marine, Terrestrial, and Freshwater Systems*, pages 59–87. Springer US, Boston, MA.
- Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Pe’er, G., Singer, A., Bridle, J. R., et al. (2016). Improving the forecast for biodiversity under climate change. *Science (New York, N.Y.)*, 353(6304):aad8466.
- Valente, L. M., Phillimore, A. B., and Etienne, R. S. (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galápagos islands. *Ecology Letters*, 18(8):844–852.
- Van Der Gast, C. J., Ager, D., and Lilley, A. K. (2008). Temporal scaling of bacterial taxa is influenced by both stochastic and deterministic ecological factors. *Environmental Microbiology*, 10(6):1411–1418.
- van der Gast, C. J., Walker, A. W., Stressmann, F. A., Rogers, G. B., Scott, P., Daniels, T. W., Carroll, M. P., et al. (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME Journal*, 5(5):780–791.
- Van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., Jeppesen, E., et al. (2007). The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51):20404–9.
- van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605.
- van der Plas, F., Janzen, T., Ordonez, A., Fokkema, W., Reinders, J., Etienne, R. S., and Olf, H. (2015). A new modeling approach estimates the relative importance of different community assembly processes. *Ecology*, 96(6):1502–1515.

- van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*. North-Holland.
- van Leuken, J. P. G., Swart, A. N., Droogers, P., van Pul, A., Heederik, D., and Havelaar, A. H. (2016). Climate change effects on airborne pathogenic bioaerosol concentrations: a scenario analysis. *Aerobiologia*, 32(4):607–617.
- Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22(2):252–260.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly review of biology*, 85(2):183–206.
- Vellend, M. (2016). *The Theory of Ecological Communities*. Princeton University Press.
- Vellend, M., Baeten, L., Myers-Smith, I. H., Elmendorf, S. C., Beauséjour, R., Brown, C. D., De Frenne, P., et al. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences of the United States of America*, 110(48):19456–9.
- Volkov, I., Banavar, J. R., Hubbell, S. P., and Maritan, A. (2009). Inferring species interactions in tropical forests. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33):13854–9.
- Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, 10(12):828–840.
- Wade, M. J., Harmand, J., Benyahia, B., Bouchez, T., Chaillou, S., Cloez, B., Godon, J. J., et al. (2016). Perspectives in mathematical modelling for microbial ecology. *Ecological Modelling*, 321:64–74.
- Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J. C., He, J., et al. (2013). Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *The ISME journal*, 7(7):1310–1321.
- Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., et al. (2015). Islands as model systems

- in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters*, 18(2):200–217.
- Weil, T., De Filippo, C., Albanese, D., Donati, C., Pindo, M., Pavarini, L., Carotenuto, F., et al. (2017). Legal immigrants: invasion of alien microbial communities during winter occurring desert dust storms. *Microbiome*, 5(1):32.
- Werner, E. E. and McPeck, M. A. (1994). Direct and Indirect Effects of Predators on Two Anuran Species along an Environmental Gradient. *Ecology*, 75(5):1368–1382.
- Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science (New York, N.Y.)*, 301(5635):976–8.
- White, E. P., Adler, P. B., Lauenroth, W. K., Gill, R. A., Greenberg, D., Kaufman, D. M., Rassweiler, A., et al. (2006). A comparison of the species-time relationship across ecosystems and taxonomic groups. *Oikos*, 112(1):185–195.
- Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., et al. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME Journal*.
- Wilson, E. O. and Simberloff, D. (1969). Experimental zoogeography on islands: Defaunation and monitoring techniques. *Ecology*, 50:267–278.
- Winder, M. and Schindler, D. E. (2004). Climate change uncouples trophic interactions in an aquatic ecosystem. *Ecology*, 85(8):2100–2106.
- Wolfe, D. A., Champ, M. A., Flemer, D. A., and Mearns, A. J. (1987). Long-Term Biological Data Sets: Their Role in Research, Monitoring, and Management of Estuarine and Coastal Marine Systems. *Estuaries*, 10(3):181.
- Woodcock, S., Van Der Gast, C. J., Bell, T., Lunn, M., Curtis, T. P., Head, I. M., and Sloan, W. T. (2007). Neutral assembly of bacterial

- communities. *FEMS Microbiology Ecology*, 62(2):171–180.
- Yawata, Y., Cordero, O. X., Menolascina, F., Hehemann, J.-H., Polz, M. F., and Stocker, R. (2014). Competition-dispersal tradeoff ecologically differentiates recently speciated marine bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 111(15):5622–5627.
- Zachmann, L., Moffet, C., and Adler, P. (2010). Mapped quadrats in sagebrush steppe: long-term data for analyzing demographic rates and plant–plant interactions. *Ecology*, 91(11):3427.
- Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, 44(1-2):109–123.
- Zhou, J. and Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews*, 81(4):e00002–17.
- Zhu, Y. G., Gillings, M., Simonet, P., Stekel, D., Banwart, S., and Penuelas, J. (2017). Microbial mass movements. *Science*, 357(6356):1099–1100.

Colonization and Extinction Rates estimated from Temporal Dynamics of Ecological Communities: The island R Package

A.1 Revisiting Island Biogeography Theory

The research program that grew around IBT (MacArthur and Wilson, 1967) has influenced a range of ecological sub-disciplines including conservation biology, biodiversity research, and biogeography (Cassey et al., 2006; Losos and Ricklefs, 2010; Warren et al., 2015). It inspired Stephen Hubbell to develop the unified neutral theory of biodiversity and biogeography (Hubbell, 2001), a formal framework to derive several community assembly patterns from immigration, birth and death processes affecting individuals. More recently, some authors have successfully extended IBT to deal with trophic interactions (Gravel et al., 2011; Massol et al., 2017). However, such extensions are mainly intended to study the stationary end point of community assembly, and/or focus on rather conceptual and theoretical questions. Despite half of century of sustained interest in IBT, our ability to explore dynamic patterns as communities assemble through time is still rather limited. Mature mathematical tools to test the temporal predictions of IBT and confront these models with data should promote critical theoretical progress (Losos and Ricklefs, 2010).

The seminal presentation of MacArthur and Wilson’s theory is formulated as an equilibrium theory (MacArthur and Wilson, 1963), introduced as the Equilibrium Theory of Island Biogeography (ETIB) (Simberloff, 1974). The theory is used to explain the link between species diversity across islands and islands characteristics mainly their degree of isolation and size. The number of species in an island results from a dynamic equilibrium between colonizations from a mainland and local extinctions. This number corresponds to the the stationary point of the famous equation:

$$\frac{dS}{dt} = c\left(1 - \frac{S}{S_P}\right)S - eS \quad (\text{A.1})$$

where S is the number of species in an island at a given time, S_P is the potential number of colonizer species, that is, the richness on the mainland (or species pool), and c and e are species colonization and extinction rates, respectively. As time progresses, the number of species in the island stabilizes around S^* :

$$S^* = S_P \frac{c}{c + e} \quad (\text{A.2})$$

As a consequence, species richness on an island at stationarity may decrease as colonization decreases due to greater distances from the mainland, or increase in larger islands because they harbor larger populations that, in turn, have lower extinction risks. An initial emphasis on this dynamic equilibrium may be the reason why the theory was first called the ETIB (Simberloff, 1974) even though the temporal dynamic aspects of the theory were clearly inherent to it from the very beginning (Losos and Ricklefs, 2010; Simberloff, 1969).

The two fundamental assumptions of the theory can be summarized as follows:

Species independence Species undergo colonization and extinction dynamics independently of each other.

Species equivalence Colonization and extinction rates are the same for all species within the regional species pool.

Given these two radical assumptions, Hanski (2001) showed that the ETIB can be derived from a more general framework that considers the occupancy probability of a species in the island whose temporal evolution is given by:

$$\frac{dp}{dt} = c(1 - p) - ep \quad (\text{A.3})$$

where p is the probability of a given species being present at time t . This is so because the expected number of independent species on a given island is given by the sum of probabilities, or occupancies, of different potential species occurring on the island. In addition, for a given initial condition, Eq. (A.3) can be solved exactly. For instance, if the species is absent at time zero, then the solution for the species occupancy probability at time t reads:

$$p(t) = \frac{c}{e + c}(1 - \exp(-(e + c)t)) \quad (\text{A.4})$$

This solution is the probability of a species being present in the community after some time t , given that the species was initially absent. Therefore, it can be interpreted as a transition probability of a Markov chain model where time has been discretized at equal time intervals of size t (see Eqs (A.5) and (A.6)). This is the central idea of package `island`. Because `island` functions estimate colonization and extinction rates rather than probabilities per time step, they can deal with either regular or very sparse, uneven sampling schemes with the same ease. As a consequence, unlike typical Markov chain models, such as those developed using similar approaches (Cazelles et al., 2016b; Fiske and Chandler, 2011), `island` package methods do not rely on equal time steps as a necessary requirement for parameter maximum-likelihood estimation and model selection. In addition, if we consider a species pool of S_P species each characterized by a pair of colonization

and extinction rates, and governed independently by Eq. (A.3), we obtain a community-level model that can be considered the simplest stochastic formulation of the colonization-extinction model underlying IBT (Alonso et al., 2015; Simberloff, 1969; Simberloff and Wilson, 1969; Wilson and Simberloff, 1969), which represents an entire, and arguably, the simplest Gleasonian view of community assembly.

To sum up, under the assumption of species independence, the description of community dynamics reduces to a simple one-dimensional model (Alonso et al., 2015). Package `island` implements the stochastic discrete-time version of the model and gives the temporal evolution of the occupancy probability, p , over discrete time steps:

$$\begin{pmatrix} 1-p \\ p \end{pmatrix}_{t+\Delta t} = \begin{bmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{bmatrix} \begin{pmatrix} 1-p \\ p \end{pmatrix}_t, \quad (\text{A.5})$$

where Δt is usually chosen to be 1 in suitable time units. We can consistently derive all four entries of the transition matrix in terms of the colonization and extinction rates, c and e , of a particular species (see Eq. (A.4) and Alonso et al. (2015)):

$$\begin{cases} T_{10}(\Delta t|c, e) = \frac{c}{e+c}(1 - \exp(-(e+c)\Delta t)), \\ T_{01}(\Delta t|c, e) = \frac{e}{e+c}(1 - \exp(-(e+c)\Delta t)), \end{cases} \quad (\text{A.6})$$

T_{10} (T_{01}) being the probability of (not) finding certain species at time t given that it was absent (present) at time $t = 0$. In addition, conditional probabilities T_{00} and T_{11} satisfy $T_{00} = 1 - T_{10}$ and $T_{11} = 1 - T_{01}$, respectively. Notice that these transition probabilities are all functions of Δt and model parameters rates, c and e .

Data should be provided as an R data frame where columns represent community configurations at increasing (not necessarily equally-spaced) time values and each row describes the temporal evolution of

the presence/absence state of a given species in the system:

$$M = \begin{bmatrix} & t_1 & t_2 & t_3 & t_4 & t_5 & \dots & t_N \\ 1: & 1 & 0 & 1 & 1 & 1 & \dots & 0 \\ 2: & 1 & 1 & 1 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ S_P: & 0 & 1 & 0 & 1 & 1 & \dots & 0 \end{bmatrix}, \quad (\text{A.7})$$

where S_P is the total number of potential species ("species pool" richness).

The central point in likelihood estimation is the ability to calculate the probability of observing a set of data under given model assumptions and parameter values. If we assume that the data matrix above has been generated by random colonization-extinction processes under the species independence assumption, it is easy to calculate the likelihood of observing the actual temporal evolution of each of the rows. For instance, if species i is characterized by a colonization-extinction pair (c_i, e_i) and a presence/absence vector $\mathbf{n}_i = (n_i(t_1), \dots, n_i(t_N))$ from times t_1 to t_N , then the probability of observing the temporal presence/absence sequence in the i -th row of matrix (A.7) is

$$P(\mathbf{n}_i | c_i, e_i) = \prod_{j=1}^{N-1} T_{n_i(t_{j+1}), n_i(t_j)}(t_{j+1} - t_j | c_i, e_i), \quad (\text{A.8})$$

and the likelihood for the whole data matrix, M , is

$$P(M | \{(c_i, e_i), i = 1 \dots, S_P\}) = \prod_{i=1}^{S_P} P(\mathbf{n}_i | c_i, e_i). \quad (\text{A.9})$$

Under the species equivalence assumption, the last likelihood function depends only on two model parameters, the same colonization-extinction pair for all species in the system. Eq. (A.9) is central in package `island` for perfect detectability. The package also implements a similar expression accounting for imperfect detectability, which depends, instead, on four model parameters (MacKenzie et al., 2003).

The characteristic time of ecological communities

B.1 Estimation of colonization and extinction rates

Alonso *et al.* (2015) showed that the fundamental equation of Island Biogeography Theory (TIB), assuming *species equivalence* and *species independence*, can be solved for a single species. *Species independence* means that the presence or absence of a species is not influenced by the presence or absence of the other species, and *species equivalence* means that all the species studied have the same response to the different factors affecting their state, which is probable in what it is traditionally called *horizontal communities*. Species equivalence implies that all species in TIB are described by aggregated colonization and extinction rates, c and e , which take uniform values for any species. Under these assumptions, the fundamental equation of TIB can be written as follows: let $T_{10}(t|c, e)$ denote the probability that a particular species is present in the community at time t given that it was initially absent, given the colonization and extinction rates. Therefore T_{10} satisfies the probability balance differential equation

$$\frac{dT_{10}}{dt} = cT_{00} - eT_{10}, \quad (\text{B.1})$$

where $T_{00}(t|c, e)$ is defined as the probability that the species was initially absent and remains so at time t . Note that, because of normalization, $T_{00} + T_{10} = 1$, hence the equation can be fully solved to

yield

$$\begin{aligned} T_{10}(t|c, e) &= \frac{c}{c+e} [1 - \exp(-(e+c)t)], \\ T_{00}(t|c, e) &= 1 - \frac{c}{c+e} [1 - \exp(-(e+c)t)]. \end{aligned} \tag{B.2}$$

Similarly, the probability $T_{01}(t)$ that a species initially present is not observed at time t , together with the probability $T_{11}(t)$ that a species initially present remains in the community at time t , can be expressed as

$$\begin{aligned} T_{01}(t|c, e) &= \frac{e}{c+e} [1 - \exp(-(e+c)t)], \\ T_{11}(t|c, e) &= 1 - \frac{e}{c+e} [1 - \exp(-(e+c)t)]. \end{aligned} \tag{B.3}$$

These probabilities actually define a two-state (0 for absent and 1 for present) discrete-time Markov chain for each species that can be used to predict the community configuration vector $\mathbf{n}(t) = (n_1(t), n_2(t), \dots, n_{S_p}(t))$ of the community after a time t has elapsed, where $n_i \in \{0, 1\}$. With these probabilities, we can estimate the likelihood of observing a sequence of presences and absences for a single species. Conversely, given a presence-absence matrix M obtained by sampling the community at uniformly spaced times (i.e., a species \times samples matrix), thanks to species independence we can estimate community colonization and extinction rates by means of the following likelihood function:

$$P(M|c, e) = T_{00}^{N_{00}} T_{10}^{N_{10}} T_{01}^{N_{01}} T_{11}^{N_{11}}, \tag{B.4}$$

N_{01} representing the number of times that we have observed species changing from state 1 (present) to state 0 (absent), and so on.

We used different methods to find the maximum likelihood estimators (m.l.e.) for colonization and extinction rates. If we are dealing with *regular sampling schemes*, i.e. equally spaced temporal samples, Δt being the time separation between samples, we used an analytically exact method that is based on the maximum likelihood estimators for transition probabilities, $T_{10} = N_{10}/(N_{10} + N_{00})$ and $T_{01} = N_{01}/(N_{01} + N_{11})$,

as in Alonso et al. (2015): in this reference, it is shown that m.l.e. estimators for c and e can be calculated in terms of N_{ij} and Δt by solving the system

$$\begin{aligned} \frac{e}{c} &= \frac{N_{01}(N_{10} + N_{00})}{N_{10}(N_{01} + N_{11})}, \\ e + c &= -\frac{1}{\Delta t} \log \left(\frac{N_{00}N_{11} - N_{01}N_{10}}{(N_{10} + N_{00})(N_{01} + N_{11})} \right). \end{aligned} \quad (\text{B.5})$$

However, if communities are sampled at not equally-spaced times (i.e., we have *irregular sampling schemes*), the likelihood function is a bit more complicated—but amenable for some analytical treatment, however. Let t_0, t_1, \dots, t_τ be the irregularly-sampled times, so the species \times times presence-absence matrix M is an $S_P \times (\tau + 1)$ matrix. Now focus on an arbitrary species i ($1 \leq i \leq S_P$). Let $T(n_i(t_{j+1}) \leftarrow n_i(t_j) | c, e)$ denote the transition probability from state $n_i(t_j)$ at time t_j to state $n_i(t_{j+1})$ at time t_{j+1} for that species, given the rates c, e . As for the regular sampling schemes, the probabilities can be expressed as

$$\begin{aligned} T_{00}(\Delta t_j | c, e) &= 1 - \frac{c}{c + e} [1 - \exp(-(e + c)\Delta t_j)], \\ T_{10}(\Delta t_j | c, e) &= \frac{c}{c + e} [1 - \exp(-(e + c)\Delta t_j)], \\ T_{01}(\Delta t_j | c, e) &= \frac{e}{c + e} [1 - \exp(-(e + c)\Delta t_j)], \\ T_{11}(\Delta t_j | c, e) &= 1 - \frac{e}{c + e} [1 - \exp(-(e + c)\Delta t_j)], \end{aligned} \quad (\text{B.6})$$

where $\Delta t_j = t_{j+1} - t_j$. Note here that the temporal evolution of the model is determined by $c + e$. The likelihood of the configurations for species i across sampling times can be written as

$$P(n_i(t_0), \dots, n_i(t_\tau) | c, e) = \prod_{j=0}^{\tau-1} T(n_i(t_{j+1}) \leftarrow n_i(t_j) | c, e), \quad (\text{B.7})$$

and, because of species independence, the likelihood of matrix M is

$$P(M | c, e) = \prod_{i=1}^{S_P} \prod_{j=0}^{\tau-1} T(n_i(t_{j+1}) \leftarrow n_i(t_j) | c, e). \quad (\text{B.8})$$

We aim to optimize this likelihood in order to obtain estimates for colonization and extinction rates c and e . For this purpose, we have devised two methods, a *heuristic* search or a *semianalytical* method based on the gradient of the likelihood function. Once we have obtained our estimates, we can also find their confidence intervals using two different methods: a *stepwise* procedure and a *Hessian*-based binary search. Both procedures find the values at which the likelihood function varies 1.96 units for each of the rates.

The semianalytical method to find m.l.e. for colonization and extinction rates is based on the gradient of the log-likelihood function. It can be expressed in a compact form as follows:

$$\begin{aligned} \log P(M|c, e) = & \sum_{j=0}^{\tau-1} \left[N_{00}^j \log T_{00}(\Delta t_j|c, e) + N_{10}^j \log T_{10}(\Delta t_j|c, e) \right. \\ & \left. + N_{01}^j \log T_{01}(\Delta t_j|c, e) + N_{11}^j \log T_{11}(\Delta t_j|c, e) \right], \quad (\text{B.9}) \end{aligned}$$

where N_{00}^j was defined as the number of $0 \leftarrow 0$ transitions, across all species, when $t_{j+1} \leftarrow t_j$, and likewise for N_{10}^j , N_{01}^j and N_{11}^j . These four numbers, as well as the τ time intervals Δt_j , are known from the data. We simply calculated the critical points solving the nonlinear equations obtained by equating to zero the gradient of the log-likelihood. We reproduce here the two partial derivatives,

$$\begin{aligned} \frac{\partial \log P}{\partial c} = & -\frac{1}{e+c} \sum_{j=0}^{\tau-1} \left\{ \left(\frac{N_{00}^j}{T_{00}^j} - \frac{N_{10}^j}{T_{10}^j} \right) \left[c\lambda_j \Delta t_j + \frac{e}{e+c} (1 - \lambda_j) \right] \right. \\ & \left. + \left(\frac{N_{11}^j}{T_{11}^j} - \frac{N_{01}^j}{T_{01}^j} \right) \left[e\lambda_j \Delta t_j - \frac{e}{e+c} (1 - \lambda_j) \right] \right\}, \quad (\text{B.10}) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \log P}{\partial e} = & -\frac{1}{e+c} \sum_{j=0}^{\tau-1} \left\{ \left(\frac{N_{00}^j}{T_{00}^j} - \frac{N_{10}^j}{T_{10}^j} \right) \left[c\lambda_j \Delta t_j - \frac{c}{e+c} (1 - \lambda_j) \right] \right. \\ & \left. + \left(\frac{N_{11}^j}{T_{11}^j} - \frac{N_{01}^j}{T_{01}^j} \right) \left[e\lambda_j \Delta t_j + \frac{c}{e+c} (1 - \lambda_j) \right] \right\}. \quad (\text{B.11}) \end{aligned}$$

In these two expressions above we have defined $\lambda_j = \exp(-(e + c)\Delta t_j)$ and denoted $T_{00}^j = T_{00}(\Delta t_j | c, e)$, and so on. For each dataset (matrix M), we found the roots of the gradient using function *multroot* of the R package *rootSolve*. These are the critical points of the likelihood function. Thus, we can encounter a local maximum, local minimum, or a saddle point, and in the case of a local maximum, these values are the m.l.e. of colonization and extinction rates, c and e .

To obtain confidence intervals for the estimates of c and e , we developed two methods, a stepwise procedure and a Hessian-based binary search. We defined the confidence intervals as the values of c or e at which the likelihood function varies 1.96 log-units. The stepwise procedure starts at the m.l.e. for each rate and proceeds stepwise, adding (or subtracting) a small, user-defined quantity to them. When the likelihood function reaches a difference of 1.96 log-units, the algorithm has found the confidence interval. The Hessian-based procedure uses that, for a multivariate Gaussian random vector, the Hessian of the log-likelihood exactly coincides with minus the inverse of the covariance matrix, so the amplitudes of the confidence intervals (standard deviations) can be calculated as the square roots of the diagonal elements of the inverse negative Hessian matrix. In general (for arbitrarily distributed random variables), standard deviations for each m.l.e. can be approximated this way and yield good approximations to true deviations when their values are small. We used the standard deviations so estimated as an initial guess, and then refined the estimates with a binary search to find the exact value at which the likelihood function has a difference with the m.l.e. of 1.96 log-units. Explicit expressions for the entries of the Hessian matrix can be computed in terms of the number of transitions N^j of each class, the time intervals Δt_j , the transition probabilities T^j , and the rates c and e . Expressions are too convoluted to be reproduced here, in any case.

We end this section showing the simulated species richness as a function of time for two classic datasets, namely: Farne Island birds

and *Neotoma* deciduous forest birds, see Figure B.1. As we found for the Florida keys arthropods community, the observed richness lies within the 95% confidence interval of the simulations (see Figure 4.1).

B.2 Community composition patterns

Jaccard index allowed us to compare sites using presence-absence data, and it is defined as follows:

$$\mathcal{J}_{ij} = \frac{C}{A + B + C}, \quad (\text{B.12})$$

where A is the number of species present at time i and not at time j , B the number of species present at time j and not at i , and C the number of species present at both times. Jaccard's index also lets us compare the same site at different times, and its expected value can be estimated with our colonization and extinction dynamics and its associated Markov chain.

Then, if we considered i and j different samples taken at the same site at times $t_i = 0$ and $t_j = t$, the expected number of species present at both times, C , can be calculated as the product of T_{11} , the transition probability for a species being present in both i and j , S_P , the number of species in the pool, and p_i , the proportion of species present at t_i . Besides, the average number of species present at time i but absent at time j , A , can be obtained as the product of T_{01} , the transition probability for species being present in i and not j , S_P and p_i , whereas B can be estimated as the product of T_{10} , the transition probability of a species being absent from i and present in j , and S_P minus the product of S_P and p_i . If the random variables C and $A + B + C$ were independent, we could calculate the expected value of the ratio $C/(A + B + C)$ as

$$\text{E} \left[\frac{C}{A + B + C} \right] = \text{E}[C] \text{E} \left[\frac{1}{A + B + C} \right]. \quad (\text{B.13})$$

Ignoring correlations and approximating $\text{E} \left[\frac{1}{A+B+C} \right] \approx \text{E} [A + B + C]^{-1}$ —which is not necessarily true in general—we obtain

the following approximation for the Jaccard index in terms of p_i and transition probabilities,

$$\mathcal{J}_i(t) \approx \frac{p_i T_{11}}{p_i T_{01} + (1 - p_i) T_{10} + p_i T_{11}}. \quad (\text{B.14})$$

Using that $T_{11} + T_{01} = 1$, Eq. (B.14) can be expressed as

$$\mathcal{J}_i(t) \approx \frac{p_i T_{11}}{p_i + (1 - p_i) T_{10}}. \quad (\text{B.15})$$

However, note that (B.14) is not the expectation of the Jaccard index, as such expected value would imply summing over all values of the Jaccard index given the probability of each possible configuration of the community. However, the law of large numbers probably reduce that hypothetical expression to equation (B.14) and substituting the rates (B.2) and (B.3) leads into the expressions for $\mathcal{J}_i(t)$ reproduced in the main text (Box 2). If we compare any given community with the same community when time goes to infinity, we can calculate the Jaccard Index as follows:

$$\tilde{\mathcal{J}}_i = \lim_{\Delta t \rightarrow \infty} \mathcal{J}_i \approx \frac{p_i c}{p_i e + c}. \quad (\text{B.16})$$

If the initial proportion of species coincides with that of equilibrium, $p_i = \frac{c}{e+c}$, our approximation for the Jaccard index simplifies to

$$\mathcal{J}^*(t) \approx \frac{c + e \exp(-(e + c)t)}{c + 2e - e \exp(-(e + c)t)}, \quad (\text{B.17})$$

which converges to $\tilde{\mathcal{J}}^* = \frac{c}{c+2e}$ as $t \rightarrow \infty$. We now estimate the time t , in units of the characteristic time $T_c = \frac{1}{c+e}$, that takes the system to reach the asymptotic Jaccard index value. Let $t = \xi T_c = \frac{\xi}{c+e}$ and let us fix the relative error $|\mathcal{J}^*(t) - \tilde{\mathcal{J}}^*|/\tilde{\mathcal{J}}^*$ equal to a certain amount χ . From

$$\frac{|\mathcal{J}^*(t) - \tilde{\mathcal{J}}^*|}{\tilde{\mathcal{J}}^*} = \chi \quad (\text{B.18})$$

we can solve for $x = \exp(-(e + c)t) = e^{-\xi}$ to get

$$e^{-\xi} = \frac{c(c + 2e)\chi}{2e(c + e) + ce\chi} \approx \frac{\chi(1 + 2z)}{2z(1 + z)} \quad (\text{B.19})$$

where we used that $\chi \ll 1$ and defined $z = e/c$. Taking logarithms at both sides, we get

$$\xi = -\log \chi - \log \frac{1+2z}{2z(1+z)} = 4.61 - \log \frac{1+2z}{2z(1+z)} \quad (\text{B.20})$$

for a small relative error $\chi = 0.01$. Now we solve for the values of z that yield $\log \frac{1+2z}{2z(1+z)} = \pm 1$, which are $z = 0.217$ and $z = 2.31$, respectively. This means that, if $0.22 \leq \frac{e}{c} \leq 2.31$, then the ratio $\frac{t}{T_c}$ is bounded by 3.61 and 5.61. Therefore, for a wide range of extinction to colonization ratios, we expect that the time t the system takes to reach the asymptotic Jaccard index is about 4 or 5 characteristic times, as stated in the main text.

We have defined an alternative measure to the characteristic time T_c , the *half-relaxation time*, which corresponds to the time that takes the system to have a decrease in similarity, measured with the Jaccard index, that is half the decrease that it encounters as time tends to infinity. To find an expression for the half-relaxation time, we have to solve for $\Delta t = T_r$ in the following equation:

$$\frac{1}{2} \left(1 - \underbrace{\frac{p_i c}{p_i e + c}}_{\lim_{t \rightarrow \infty} \mathcal{J}_i(t)} \right) = \frac{p_i [c + e \exp(-(e+c)T_r)]}{\underbrace{p_i [e + c \exp(-(e+c)T_r)] + c [1 - \exp(-(e+c)T_r)]}_{\mathcal{J}_i(T_r)}}. \quad (\text{B.21})$$

After some algebra, it reduces to the following expressions, starting or not with an initial p_i equal to that of the equilibrium:

$$T_r \approx T_c \log \frac{c + (c + 2e)p_i}{c + ep_i}, \quad (\text{B.22})$$

$$T_r^* \approx T_c \log \frac{2c + 3e}{c + 2e}. \quad (\text{B.23})$$

In Figure B.2, we simulated colonization and extinction dynamics for a hundred simulations of a model with $T_c = 20$ and occupancy at equilibrium equal to 0.1. Starting from an initial occupancy of 0.5, we have calculated the Jaccard index over a hundred timesteps for each of these realizations. Besides, we calculated our theoretical approximation

(blue line) for the expected Jaccard index [cf. Eq. (B.15)], and the half-saturation time for this set of model parameters.

Finally, we have simulated a sequence of models with T_c ranging from 10 to 100 and three different occupancies at the equilibrium, 0.1, 0.5, and 0.9. For each of these models, we obtained 200 realizations and averaged their evolution for the Jaccard index with time. We show in Figure B.3 the relation between T_c and *half-relaxation time*, that we found theoretically [colored lines, Eqs. (B.22) and (B.23)] and empirically (black points) the *half-relaxation time* from the averaged dynamics for each of these models.

B.3 Properties of c , e , and T_c as estimators

The main properties that determine a good estimator are *bias*, *consistency*, and *efficiency*. *Bias* is the difference between the true parameter and the mean of the distribution of the estimator. An estimator with no difference is considered unbiased. *Consistency* is the property of an estimator of being near the value of the parameter as sample size increases. Finally, *efficiency* refers to the size of the distribution of the estimator, being efficient an estimator with small support.

To study bias, we simulated a 1000 species \times 1001 samples dataset with equal colonization and extinction rates ($c = e = 0.1$) and a characteristic time $T_c = 5$. From this matrix, we obtained *true* colonization and extinction rates, because these can vary from true ones as it is a stochastic process. We subsampled 1000 times this matrix to have only 100 transitions each time and recalculated colonization and extinction rates. We used differences of *true* and estimated c , e , and T_c to quantify bias using skewness and the D'Agostino test, which yielded non-significant biases for each of the estimators (c skew = 0.0270, p-value *n.s.*; e skew = -0.0101 , p-value *n.s.*; T_c skew = 0.0642, p-value *n.s.*). Following the same procedure for unbalanced values of the rates ($c = 0.02$, $e = 0.18$ or $c = 0.18$, $e = 0.02$) yielded the same absence of

bias.

We also studied the *consistency* of the estimators. We simulated 1000 replicates of a 1000 species \times 1001 samples matrix following a specified colonization and extinction dynamics with $T_c = 5$. We calculated the *true* parameters for each one, and then we subsampled them to obtain 10, 20, 50, 100, 200, and 500 transitions and estimate c , e , and T_c . Then we calculated the bias between true and estimated parameters along this sequence. Figure B.4 shows that, as we increase the number of transitions, which can be interpreted as increasing the sample size, the estimated parameters get closer to the true values, decreasing their dispersion. This is also a sign of an increased *efficiency*.

B.4 Sources of error in T_c

We have identified three possible sources of error when estimating c , e , and T_c . The first one and probably the most evident is the lack of conformity of the studied communities to the assumptions of the model, namely, *independence* and *equivalence*. However, we want to stress that our model should be regarded as a null model, acting as a mean-field approximation to the actual dynamics (that may include many other processes).

The second one is related to how thoroughly we have sampled our community. We have found that the estimates may not be precise when the *relative sampling frequency* is below ~ 0.3 . Figure B.5 shows the effect of low sampling frequencies on T_c . We simulated 30 different matrices of 1000 species \times 1000 samples, each one with a different pair of colonization and extinction but a characteristic time of 10, and we subsampled to achieve relative sampling frequencies from 0.1 to 10. From the subsampled matrices we estimated their characteristic time, that seemed to vary wildly under a relative sampling frequency $\nu_s \approx 0.3$. A procedure to discern if our estimates are reliable would be to undersample our data to recalculate the estimates; if those are

consistent with the previous estimates, these may be acceptable. If the estimates change, we might be sampling below T_c .

The last possible source of error that we have identified occurs when we subsample our initial data, and we find a linear relationship between mean sampling time and T_c . As can be shown trivially from (B.2) and (B.3), in the regular sampling scheme T_c depends on the transition probabilities T_{01} and T_{10} . Together they determine the degree of lag-1 autocorrelation, ρ , in our data (following Tuljapurkar 1997):

$$T_c = \frac{1}{c + e} = -\frac{\Delta t}{\log(1 - T_{01} - T_{10})} = -\frac{\Delta t}{\log \rho}. \quad (\text{B.24})$$

Thus, in order to find a linear relationship between sampling time and T_c we need a ρ constant at different sampling times. We have examined data from the left hand of individual M3 (Caporaso et al., 2011). We found an estimate of $T_c = 2.05$ days, with a mean Δt of 1.23 days. In order to examine the robustness of the estimates, we subsampled the data, increasing Δt to 1.5, 1.8, and 2.5. Lag-1 autocorrelation was measured row by row, adopting the convention of considering the correlation of a series formed only by ones or zeros as having a value of 1. We found that autocorrelation remained constant, thus making T_c to follow a linear relationship with Δt . Also, we explored the autocorrelation of random strings of ones and zeros and the sequences produced by stochastic realizations of the dynamics corresponding to the estimated parameters for that dataset (Figure B.6). The same procedure was followed to examine data of island E1 from the arthropods dataset (Simberloff and Wilson, 1969). The subsampled data showed less autocorrelation than the original data, and the model simulations were congruent with it. The difference between the observed autocorrelation and the stochastic realizations allow us to conclude that when this difference is high, the estimates should be considered with caution. The reason behind this mismatch may correspond to the subsampled data being close to a random matrix with the proportion of presence/absence of the original data, and it

might indicate that we need to sample more to find correct estimates of the dynamics.

B.5 Supplementary figures

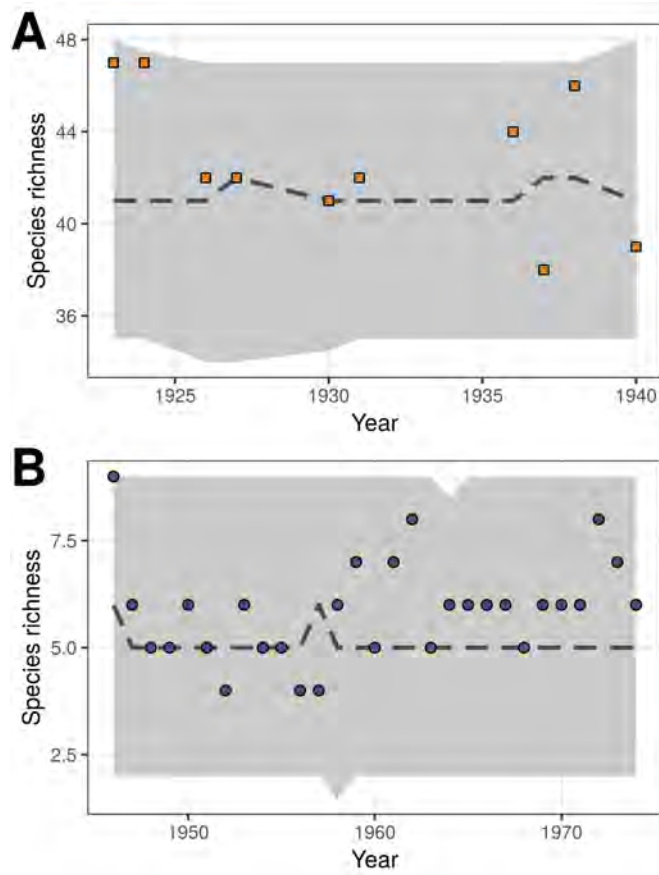


Figure B.1: **Temporal dynamics of two classic communities.** A) Deciduous forest birds at Neotoma, B) island birds in the Farne Islands. Shaded areas correspond to the 95% confidence interval obtained from model simulations, whereas the dashed line refers to the median of the simulations.

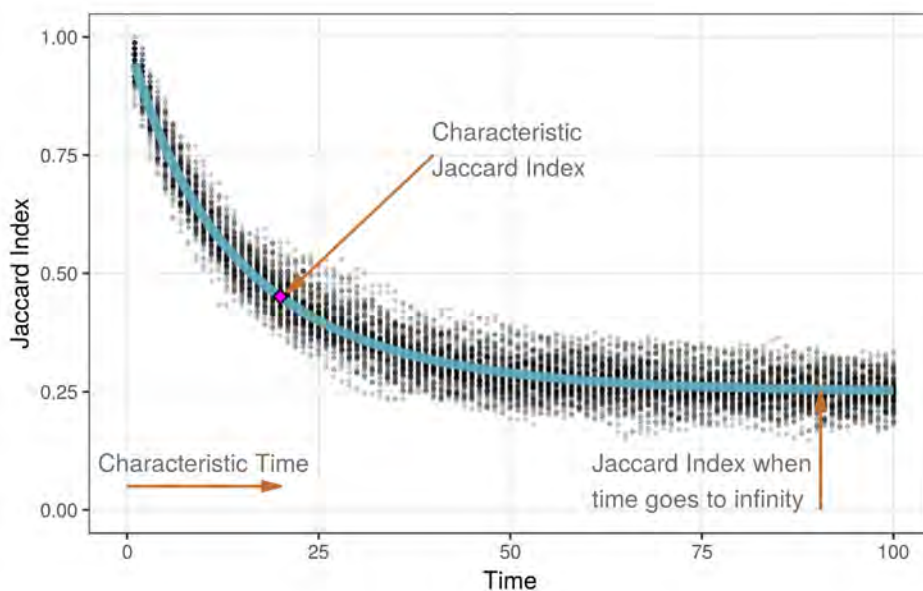


Figure B.2: **Evolution of the Jaccard index with time under simulations.** The blue line shows the expected value for the Jaccard index using our theoretical approximation. The magenta diamond indicates the characteristic Jaccard index, that is, the value of the Jaccard index between the initial community and the one present after a characteristic time (20 units in this case). Approximately after four times T_C , the Jaccard index stabilizes.

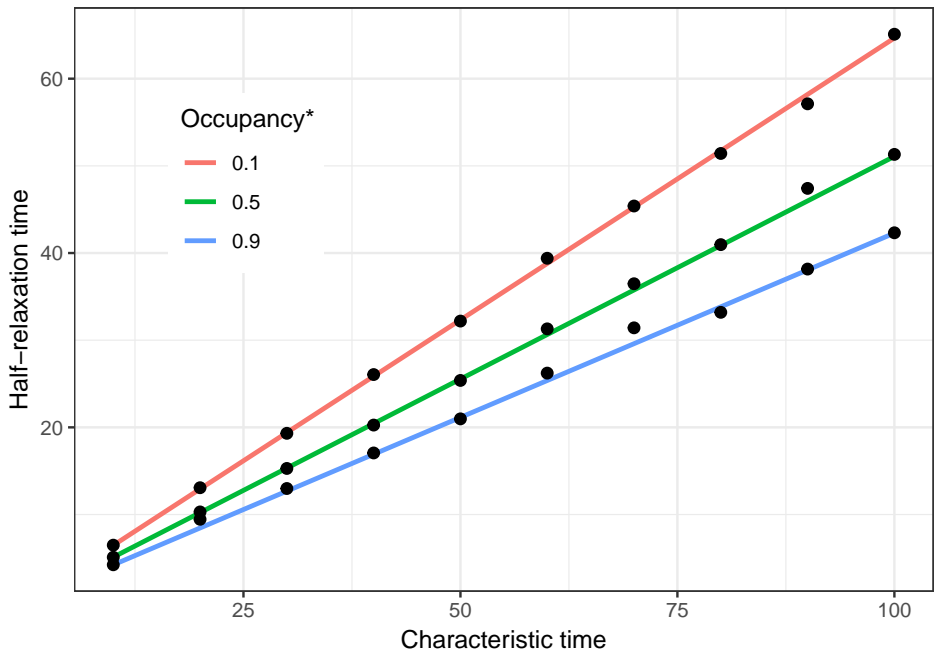


Figure B.3: Relation of the T_r with T_c . For the same characteristic time, the systems that have a higher occupancy at the equilibrium had a lower half-relaxation time, while the systems with a lower occupancy took the most to reach it. As we can see, there is a linear relation between T_c and T_r for equal levels of occupancy.

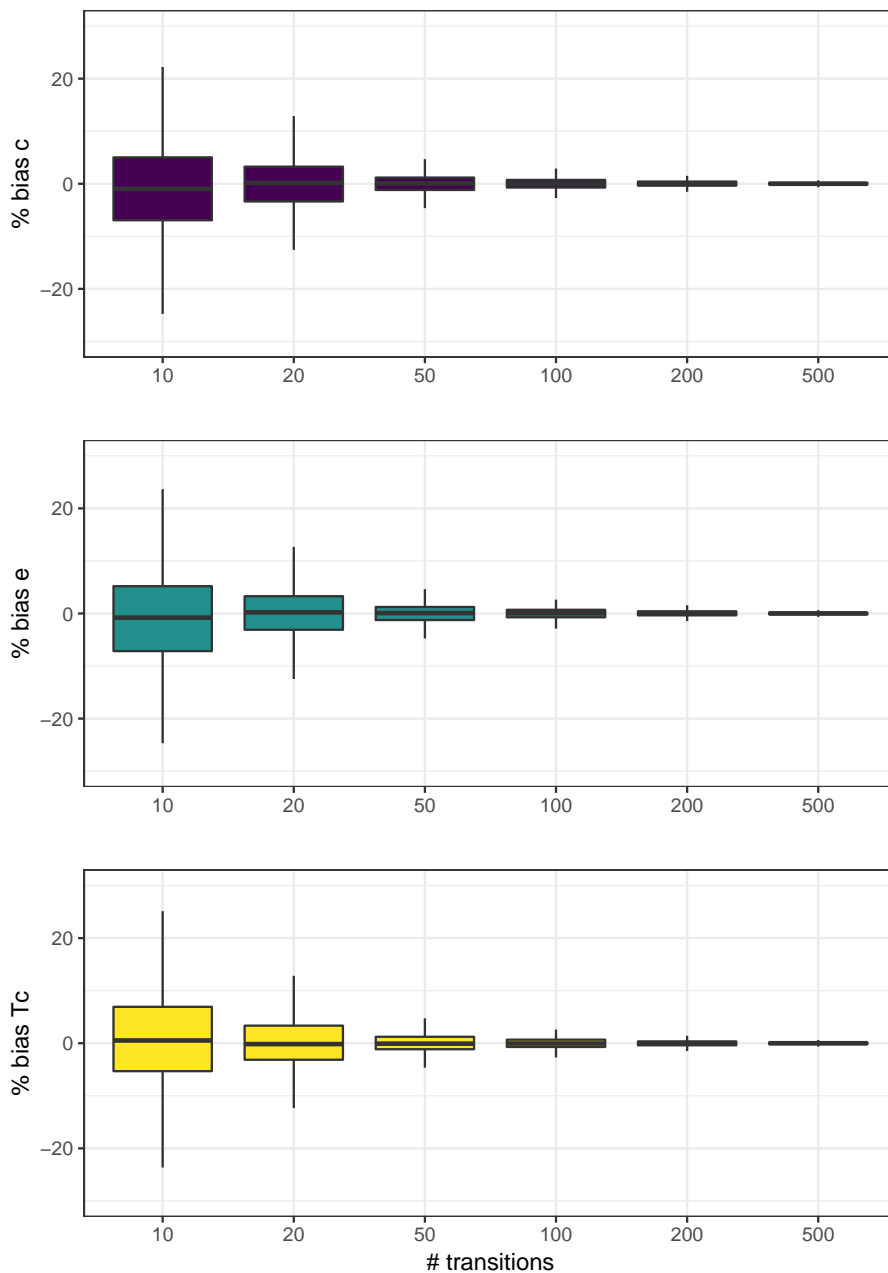


Figure B.4: **Properties of c , e , and T_C as estimators.** We found no bias. Consistency and efficiency increased as sample size (i.e., the number of transitions) increased.

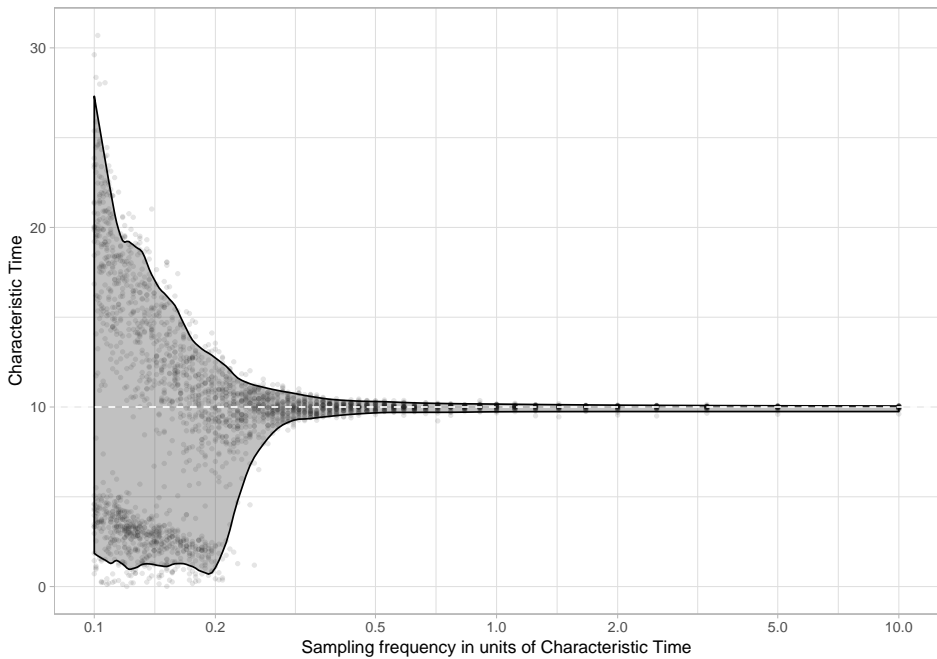


Figure B.5: **Influence of sampling frequency over estimates of characteristic time.** When we sample under a relative sampling frequency of ≈ 0.3 , we can observe that the estimates of the characteristic time, T_c , start to deviate from the true value, in this case, $T_c = 10$. Subsampling is needed in those cases to increase certainty.

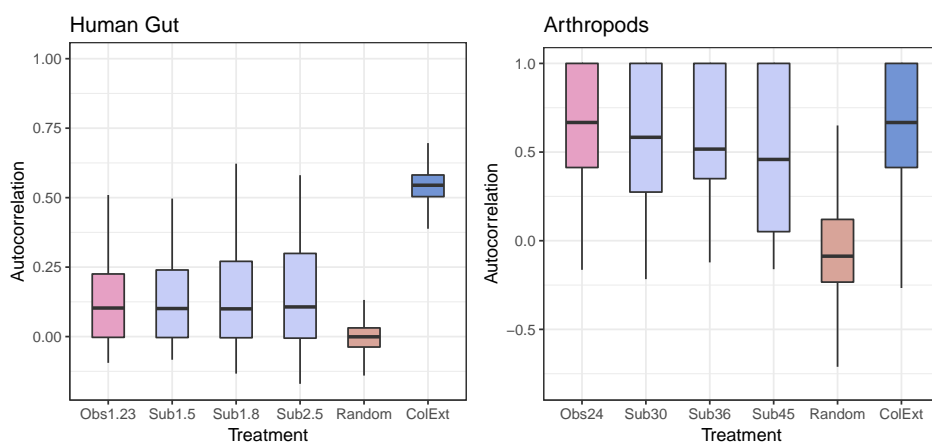


Figure B.6: **Autocorrelation is associated with characteristic time estimation.** Autocorrelation in the case of the data of the left hand of individual M3 seems stable after subsampling, and close to the autocorrelation of random sequences of 0s and 1s. However, in the case of the arthropods of island E1, each subsampling decreased autocorrelation, thus maintaining the estimate of characteristic time. In this last case, we also see that our simulations closely resemble the autocorrelation found for the original data.

General decline in the diversity of the airborne microbiome under future climatic scenarios

C.1 Supplementary tables

Table C.1: **Model selection procedure for bacterial and eukaryal OTUs.** S – single pair of colonization and extinction rates, P – a pair of rates for each phylum, A – a pair of rates for the most abundant groups, E – a pair of rates plus several coefficients indicating the influence of environmental variables, A*E – same but for each of the most abundant groups.

| Model | NLL | Pars | AIC | Δ AIC | AIC _w |
|-----------------|----------|------|----------|--------------|------------------|
| <i>Bacteria</i> | | | | | |
| S | 83682.44 | 2 | 167368.9 | 10230.0 | 0 |
| P | 81519.03 | 74 | 163186.1 | 6047.2 | 0 |
| A | 80215.60 | 26 | 160457.2 | 3318.3 | 0 |
| E | 82859.32 | 9 | 165736.6 | 8597.7 | 0 |
| A * E | 78413.47 | 156 | 157138.9 | 0 | 1 |
| <i>Eukarya</i> | | | | | |
| S | 222073.5 | 2 | 444151.0 | 20715.4 | 0 |
| P | 217973.6 | 24 | 435995.2 | 12559.6 | 0 |
| A | 217182.9 | 28 | 434421.8 | 10986.2 | 0 |
| E | 220856.8 | 8 | 441725.6 | 18290.0 | 0 |
| A * E | 211528.8 | 189 | 423435.6 | 0 | 1 |

Table C.2: **Models and OTU richness for bacterial groups.** Columns represent the model for each group. Values indicate the coefficient that multiplies the corresponding environmental variable. ANC: acid-neutralizing capacity, Cond.: conductivity, DIC: dissolved inorganic carbon, DOC: dissolved organic carbon, Hum.: humidity, Irr.: irradiance, Samp. Eff.: sampling effort, Tmax: maximum temperature, Tmed: mean temperature, Tmin: minimum temperature, c_0 : colonization independent term, e_0 : extinction independent term. Richness indicates OTU richness for each group, and it is not a coefficient included in the model.

| | Acidobacteria | Actinobacteria | Alphaproteobacteria | Bacteroidetes | Betaproteobacteria | Cyanobacteria | Deinococcus-Thermus | Deltaproteobacteria | Firmicutes | Gammaaproteobacteria | Bacteria - other | Planctomycetes | Verrucomicrobia |
|---------------------|---------------|----------------|---------------------|---------------|--------------------|---------------|---------------------|---------------------|------------|----------------------|------------------|----------------|-----------------|
| <i>Colonization</i> | | | | | | | | | | | | | |
| ANC | - | 0.00177 | - | - | - | - | - | - | -0.00117 | - | - | -0.00002 | - |
| Cond. | - | - | - | - | - | - | - | - | 0.00639 | - | - | - | - |
| DIC | - | -0.00199 | - | - | - | - | - | - | - | -0.00058 | - | - | - |
| DOC | 0.00047 | - | 0.00127 | 0.00113 | - | 0.00216 | - | - | - | - | 0.00092 | - | - |
| Hum. | - | - | -0.00056 | - | - | - | - | - | -0.00043 | -0.00015 | - | -0.00233 | - |
| Irr. | - | - | -0.00046 | - | -0.00094 | -0.00109 | - | -0.00084 | -0.00061 | -0.00066 | -0.00036 | -0.00086 | - |
| K | - | - | - | - | - | - | 0.00068 | - | -0.00026 | - | - | - | - |
| Mg | - | - | - | -0.00049 | - | - | - | - | -0.00335 | - | - | - | - |
| Na | - | - | - | - | - | - | - | - | - | - | - | - | 0.1222 |
| NO ₃ | - | - | - | - | - | -0.00030 | - | - | -0.00090 | 0.00095 | - | - | - |
| pH | - | - | - | - | - | - | - | - | 0.00019 | - | - | - | - |
| Samp. Eff. | - | -0.00067 | - | -0.00067 | -0.00108 | - | -0.00129 | - | - | -0.00044 | -0.00026 | - | - |
| SO ₄ | - | - | -0.00040 | - | - | - | - | - | - | - | -0.00044 | - | - |
| Tmax | - | - | - | - | -0.00292 | -0.00140 | - | - | - | - | - | - | - |
| Tmed | - | 0.00418 | 0.00059 | 0.00293 | 0.01182 | - | -0.00161 | 0.00430 | - | - | - | - | - |
| Tmin | -0.00164 | -0.00458 | -0.00263 | -0.00374 | -0.00989 | - | - | -0.00539 | - | - | -0.00081 | -0.00134 | - |
| TN | - | -0.00063 | -0.00098 | - | -0.00085 | - | - | - | - | -0.00054 | - | - | - |
| Rain | - | - | - | - | - | 0.00102 | - | - | - | - | - | - | - |
| TP | - | - | 0.00039 | - | 0.00103 | - | 0.00288 | - | - | - | - | - | - |
| c_0 | 0.00449 | 0.00513 | 0.00640 | 0.00447 | 0.00730 | 0.00546 | 0.00595 | 0.00362 | 0.00399 | 0.00324 | 0.00282 | 0.00551 | 0.06843 |
| <i>Extinction</i> | | | | | | | | | | | | | |
| ANC | - | -0.03670 | - | - | - | - | - | - | - | - | - | - | - |
| Ca | - | - | 0.00574 | - | - | - | - | - | -0.05831 | - | - | - | - |
| DIC | 0.02118 | 0.04848 | 0.00680 | - | 0.00709 | - | - | - | 0.07247 | - | - | - | - |
| DOC | - | - | - | - | - | - | - | -0.01873 | - | -0.00482 | - | - | - |
| Hum. | 0.00709 | - | -0.00698 | -0.01137 | - | - | - | - | - | - | - | -0.0702 | - |
| Irr. | 0.02858 | - | - | - | -0.00449 | - | - | - | - | - | - | 0.1350 | 0.88336 |
| K | - | - | - | - | - | - | - | - | 0.02321 | 0.01518 | - | - | - |
| Mg | - | 0.00653 | - | 0.00207 | - | 0.07650 | 0.01047 | - | - | - | 0.03158 | - | - |
| Na | - | -0.00534 | - | - | - | - | - | - | - | - | - | - | 6.21016 |
| NO ₃ | - | - | 0.00429 | - | 0.00476 | -0.02878 | - | - | - | - | - | - | - |
| pH | - | - | -0.00660 | - | -0.00629 | - | - | - | 0.01427 | - | - | 0.0870 | - |
| Samp. Eff. | 0.01161 | - | 0.00318 | - | - | - | - | - | - | -0.01269 | - | - | - |
| SO ₄ | - | - | - | 0.01197 | - | - | - | 0.01166 | - | - | - | - | - |
| Tmax | - | - | 0.01404 | - | - | - | -0.02148 | -0.02513 | - | -0.01680 | 0.04954 | -0.2450 | -1.09316 |
| Tmed | - | - | -0.05201 | -0.05976 | -0.00512 | - | - | - | - | - | -0.06120 | 0.2000 | - |
| Tmin | - | - | 0.03011 | 0.04896 | - | - | - | -0.04544 | - | 0.01089 | - | 0.0599 | 1.06673 |
| TN | -0.00665 | -0.00870 | -0.00731 | - | -0.00471 | - | - | - | -0.00244 | -0.00685 | -0.01052 | - | - |
| Rain | - | - | - | 0.00044 | - | - | - | - | - | - | - | - | - |
| TP | - | - | - | - | - | - | 0.01163 | - | - | - | - | - | - |
| e_0 | 0.09834 | 0.06555 | 0.03869 | 0.07206 | 0.02869 | 0.21179 | 0.05324 | 0.17452 | 0.11938 | 0.06517 | 0.12126 | 0.45600 | 3.96541 |
| Richness | 214 | 330 | 741 | 701 | 100 | 74 | 25 | 201 | 129 | 318 | 519 | 172 | 106 |

Table C.3: **Models and OTU richness for eukaryal groups.** Columns represent the model for each group. Values indicate the coefficient that multiplies the corresponding environmental variable. ANC: acid-neutralizing capacity, Cond.: conductivity, DIC: dissolved inorganic carbon, DOC: dissolved organic carbon, Hum.: humidity, Irr.: irradiance, Samp. Eff.: sampling effort, Tmax: maximum temperature, Tmed: mean temperature, Tmin: minimum temperature, c_0 : colonization independent term, e_0 : extinction independent term. Richness indicates OTU richness for each group, and it is not a coefficient included in the model.

| | SAR - Alveolata | Amoebozoa | Ascomycota | Basidiomycota | Centrolelella | Chloroplastida | Cytridomycota | Dikarya - other | Excavata Discoba | Fungi - other | Opisthokonta - unclass. | Eukarya - other | SAR - Rhizaria | SAR - Stramenopiles |
|---------------------|-----------------|-----------|------------|---------------|---------------|----------------|---------------|-----------------|------------------|---------------|-------------------------|-----------------|----------------|---------------------|
| <i>Colonization</i> | | | | | | | | | | | | | | |
| ANC | - | - | -0.00210 | -0.00776 | - | - | -0.00451 | - | - | - | - | - | - | - |
| Ca | 0.00093 | - | -0.00767 | -0.00627 | - | - | - | - | - | - | - | - | - | - |
| Cl | - | - | -0.00863 | -0.00842 | - | 0.00041 | - | 0.00402 | - | - | - | - | - | - |
| Cond. | - | 0.00131 | 0.01890 | 0.01820 | - | - | - | - | - | - | - | - | - | 0.00007 |
| DIC | - | - | 0.00107 | 0.00298 | - | - | 0.00441 | - | - | - | 0.00150 | - | - | - |
| DOC | 0.00090 | - | -0.00085 | -0.00254 | - | -0.00055 | - | -0.00259 | - | - | - | - | - | - |
| Hum. | 0.00095 | - | - | - | - | 0.00106 | 0.00058 | - | - | 0.00046 | 0.00136 | - | 0.00111 | - |
| Irr. | 0.00065 | - | -0.00194 | -0.00202 | - | 0.00092 | 0.00196 | -0.00224 | 0.00223 | 0.00047 | 0.00084 | - | - | - |
| K | 0.00215 | 0.00094 | 0.00217 | 0.00201 | - | 0.00067 | - | 0.00448 | - | - | - | 0.00085 | 0.00119 | - |
| Mg | -0.00133 | - | - | - | - | - | - | - | - | - | -0.00114 | - | - | - |
| Na | - | - | - | - | - | - | - | -0.00377 | - | 0.00054 | - | -0.00020 | - | 0.00051 |
| NO ₃ | -0.00153 | - | -0.00111 | 0.00009 | - | - | - | - | - | - | - | - | - | - |
| pH | - | - | 0.00127 | 0.00298 | - | - | 0.00073 | - | - | 0.00155 | -0.00093 | 0.00105 | - | 0.00117 |
| SO ₄ | - | -0.00071 | -0.00365 | -0.00381 | - | -0.00017 | - | - | - | - | - | - | - | - |
| Tmax | - | - | - | - | - | - | - | - | - | - | - | - | - | 0.00183 |
| Tmed | 0.00142 | - | -0.00226 | - | - | - | -0.00533 | - | - | - | - | - | - | - |
| Tmin | - | - | -0.00005 | - | - | -0.00058 | 0.00293 | - | - | -0.00122 | - | -0.00057 | - | -0.00381 |
| TN | - | - | -0.00119 | - | - | - | - | -0.00083 | - | 0.00100 | - | - | - | - |
| Rain | - | - | - | - | - | - | - | - | - | -0.00039 | - | - | - | -0.00057 |
| TP | - | - | -0.00185 | -0.00244 | - | - | 0.00305 | -0.00211 | - | - | - | - | - | - |
| c_0 | 0.00547 | 0.00392 | 0.01150 | 0.01010 | 0.43845 | 0.00443 | 0.00809 | 0.00989 | 0.00511 | 0.00692 | 0.00569 | 0.00312 | 0.00565 | 0.00425 |
| <i>Extinction</i> | | | | | | | | | | | | | | |
| ANC | - | - | -0.01390 | - | - | -0.00826 | - | -0.01757 | - | - | - | - | - | - |
| Ca | - | - | 0.02820 | - | - | 0.03728 | - | 0.03900 | - | 0.01133 | - | - | - | - |
| DOC | - | - | 0.01060 | - | - | - | - | 0.01761 | - | - | - | - | -0.01170 | - |
| Hum. | - | -0.02822 | 0.00624 | - | - | 0.01158 | -0.00699 | 0.01119 | - | -0.00509 | 0.00769 | - | - | - |
| Irr. | -0.01550 | -0.02898 | 0.01130 | - | - | 0.01936 | -0.01031 | 0.02149 | -0.01780 | -0.00780 | - | -0.01504 | - | - |
| K | - | - | -0.00590 | - | 7.92365 | - | - | 0.00332 | -0.01387 | -0.00591 | -0.01193 | - | - | - |
| Mg | - | - | - | - | - | - | -0.00668 | - | - | - | - | - | 0.01915 | - |
| Na | - | - | 0.00193 | - | - | - | - | - | - | - | - | - | - | - |
| NO ₃ | - | - | -0.01700 | - | - | -0.00948 | 0.00732 | -0.03476 | - | - | - | - | - | - |
| pH | - | - | - | - | -36.74749 | -0.01690 | - | - | - | - | -0.01184 | - | -0.01049 | - |
| SO ₄ | - | - | - | - | - | -0.00182 | - | - | - | - | - | - | - | - |
| Tmax | - | -0.01876 | - | 0.00076 | - | - | -0.03287 | - | 0.21520 | - | - | - | - | - |
| Tmed | - | - | - | - | - | - | 0.05888 | - | -0.23696 | - | - | - | -0.07541 | 0.00488 |
| Tmin | - | - | - | - | - | - | -0.02160 | - | - | - | - | - | 0.07546 | - |
| TN | - | - | -0.00752 | - | - | -0.00634 | -0.00699 | -0.00715 | - | - | - | - | - | -0.01260 |
| Rain | - | - | - | 0.00858 | - | - | - | - | - | - | - | - | 0.00443 | - |
| TP | - | - | 0.01040 | - | - | 0.00605 | - | - | 0.02637 | - | 0.00843 | - | - | - |
| e_0 | 0.06807 | 0.12609 | 0.05410 | 0.05590 | 61.25877 | 0.04805 | 0.04301 | 0.06212 | 0.10836 | 0.06662 | 0.06219 | 0.07068 | 0.06909 | 0.09310 |
| Richness | 530 | 288 | 1746 | 1800 | 5 | 195 | 204 | 110 | 36 | 204 | 53 | 228 | 247 | 168 |

Table C.4: **Model selection procedure for bacterial and eukaryal pathogens.** S – single pair of colonization and extinction rates, G – a pair of rates for some predefined groups, E – a pair of rates plus several coefficients indicating the influence of environmental variables, G*E – same but for each of the predefined groups.

| Model | NLL | Pars | AIC | Δ AIC | AIC _w |
|-----------------|----------|------|----------|--------------|------------------|
| <i>Bacteria</i> | | | | | |
| S | 3585.02 | 2 | 7174.05 | 109.26 | 0 |
| G | 3555.97 | 8 | 7127.94 | 63.15 | 0 |
| E | 3541.14 | 8 | 7098.27 | 33.48 | 0 |
| G * E | 3512.40 | 20 | 7064.79 | 0 | 1 |
| <i>Eukarya</i> | | | | | |
| S | 12824.24 | 2 | 25652.48 | 332.02 | 0 |
| G | 12806.25 | 10 | 25632.50 | 312.04 | 0 |
| E | 12665.62 | 8 | 25347.24 | 26.78 | 0 |
| G * E | 12624.23 | 36 | 25320.46 | 0 | 1 |

Table C.5: **Models and OTU richness for two pathogen groups.**

| Group | Colonization | Extinction | N |
|--------|--|---|-----|
| Humans | $-2.64e-3 * TP + -4.28e-3 * T_{min} + 2.523e-2$ | $-8.47e-3 * K + 4.86e-3 * TP + -3.78e-3 * TN + 3.15e-3 * Mg + 3.116e-2$ | 32 |
| Plants | $-4.8e-4 * T_{max} + -5.51e-3 * T_{min} + -3.90e-3 * Rain + 3.77e-3 * Samp. Eff. + -1.19e-3 * SO_4 + 3.537e-2$ | $-3.64e-3 * Cl + 3.537e-2$ | 120 |

C.2 Supplementary figures

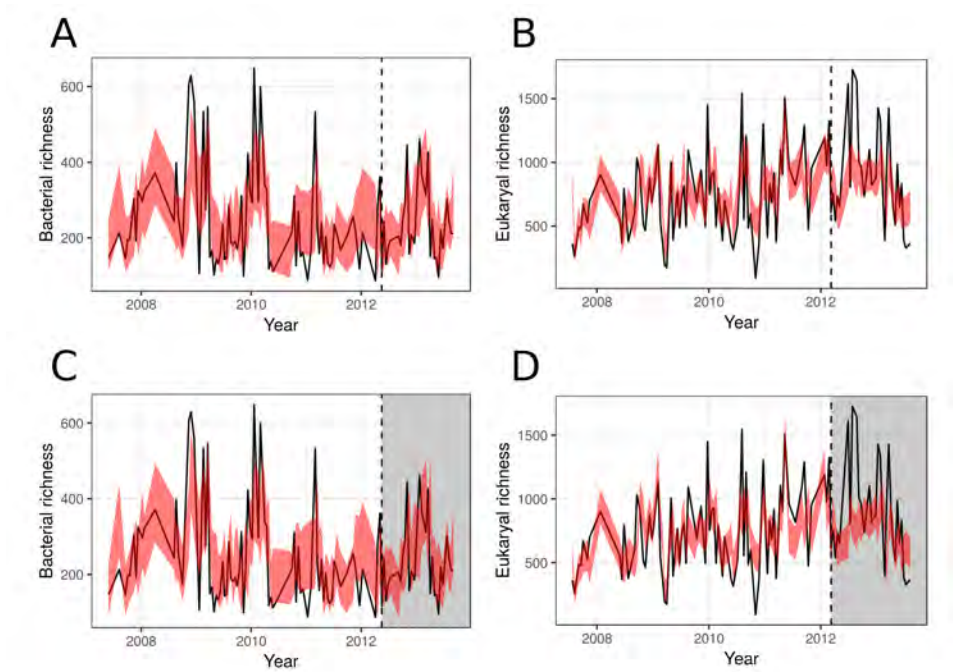


Figure C.1: **Fits for Bacteria and Eukarya.** A) Fit produced with parameters estimated for the whole bacterial data. B) Same but for eukarial data. C) Fit produced with parameters estimated for the first three-quarters of bacterial data. D) Same but for Eukarya. The black line indicates the observed OTU richness, while the red band indicates 95% of the simulations of model A*E. The grey shaded area is the area used for validating the prediction.

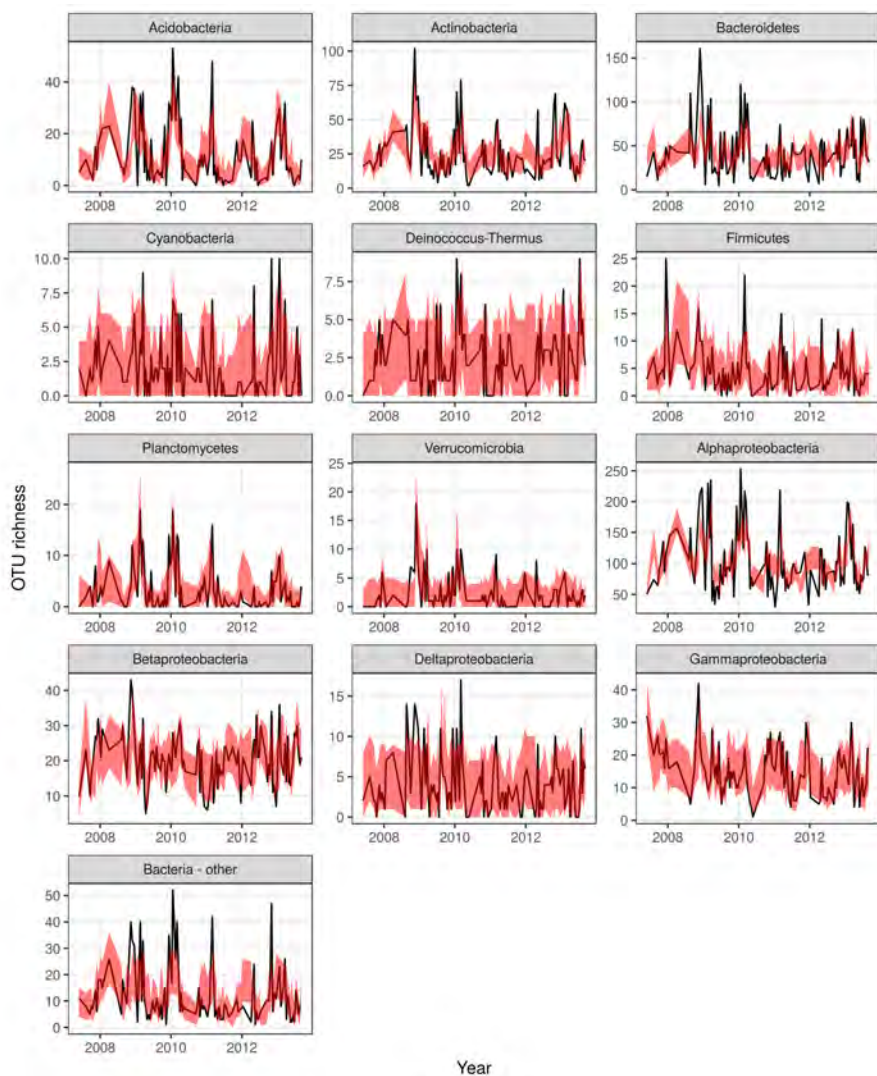


Figure C.2: **Fit for bacterial groups.** The black line indicates the observed OTU richness, while the red band indicates 95% of the simulations.

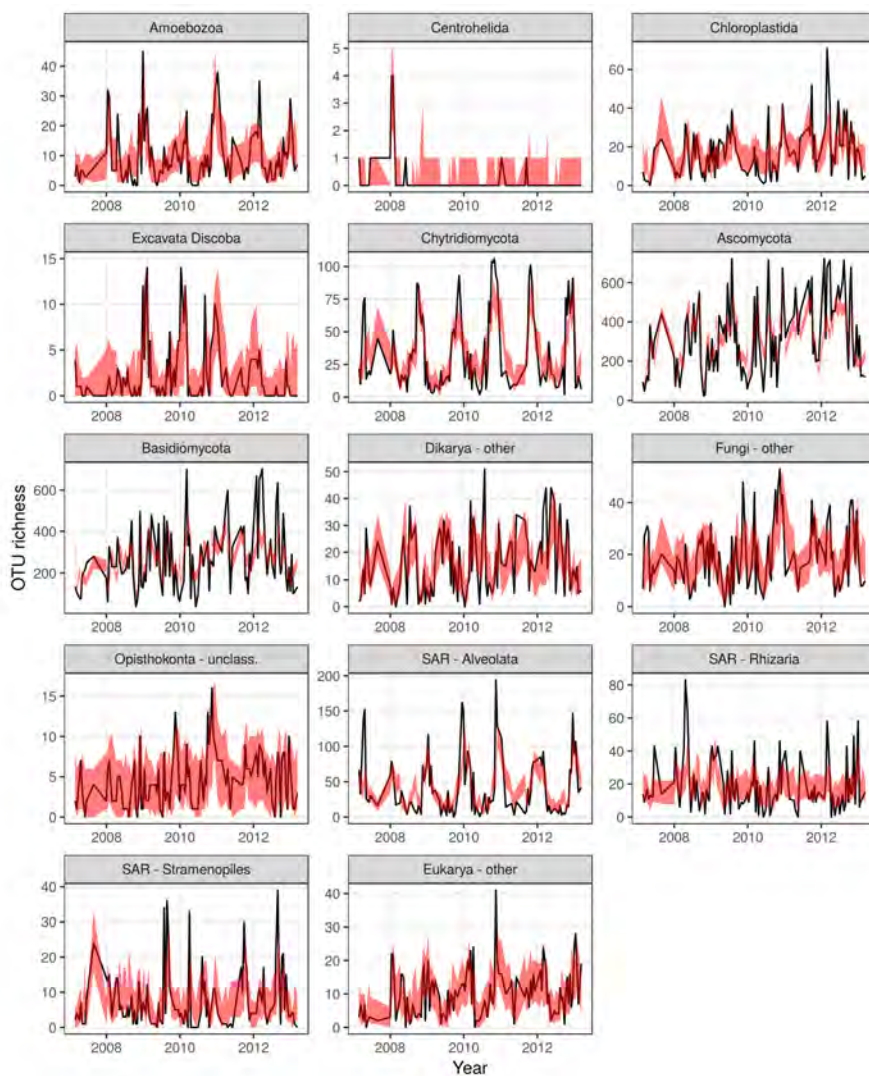


Figure C.3: **Fit for eukaryal groups.** The black line indicates the observed OTU richness, while the red band indicates 95% of the simulations.

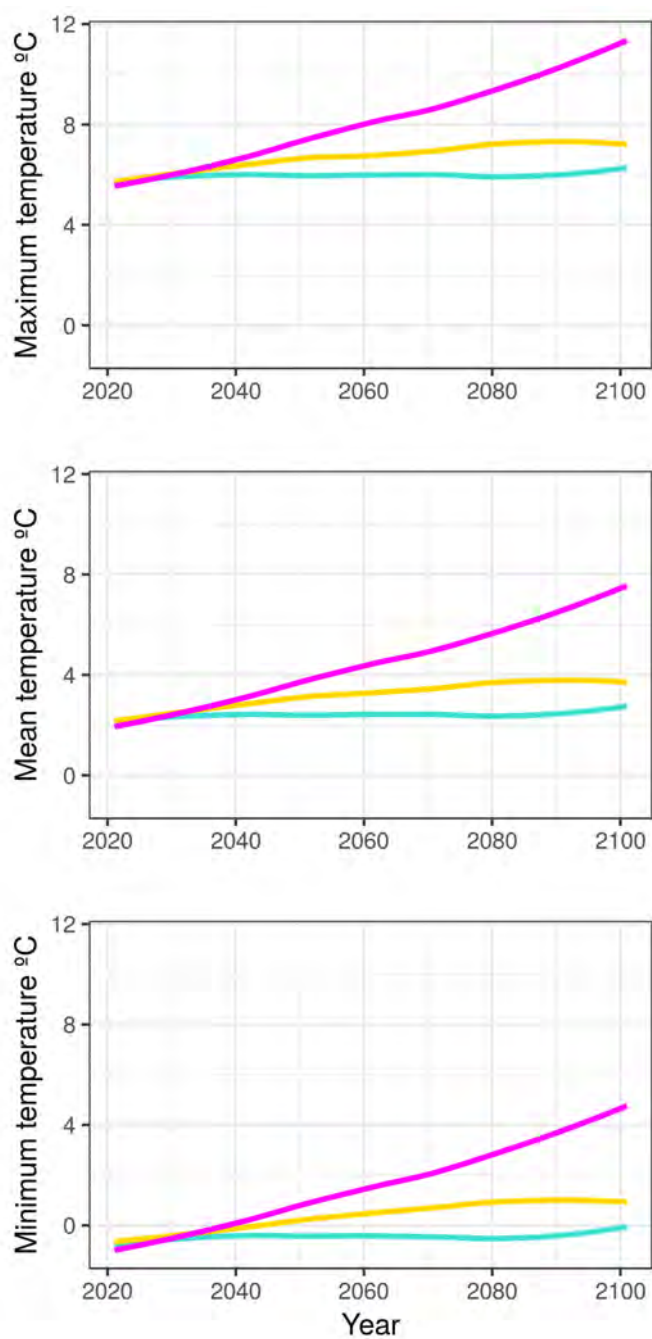


Figure C.4: Predicted increase in temperatures by the ensemble model in the period 2021 – 2100.

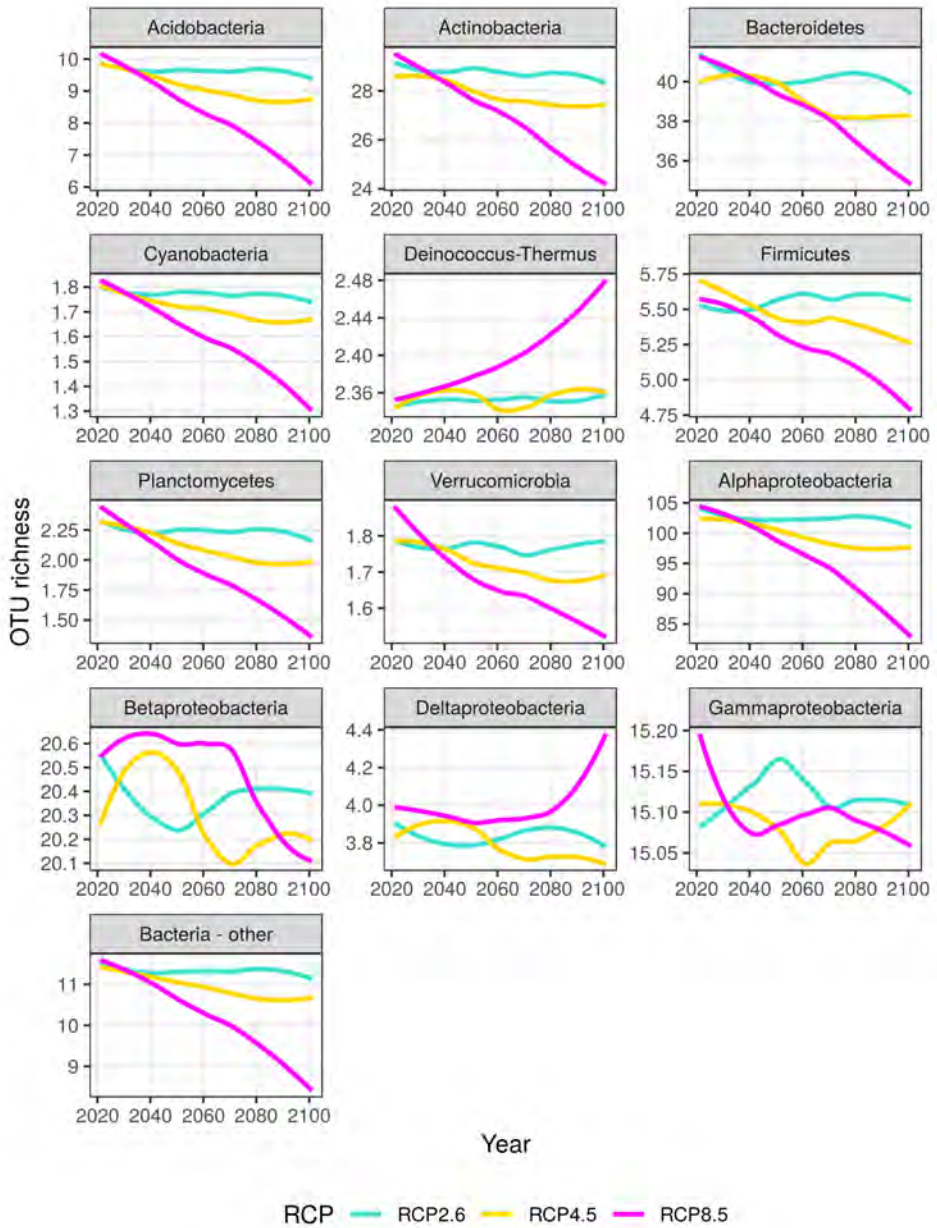


Figure C.5: Prediction for bacterial groups.

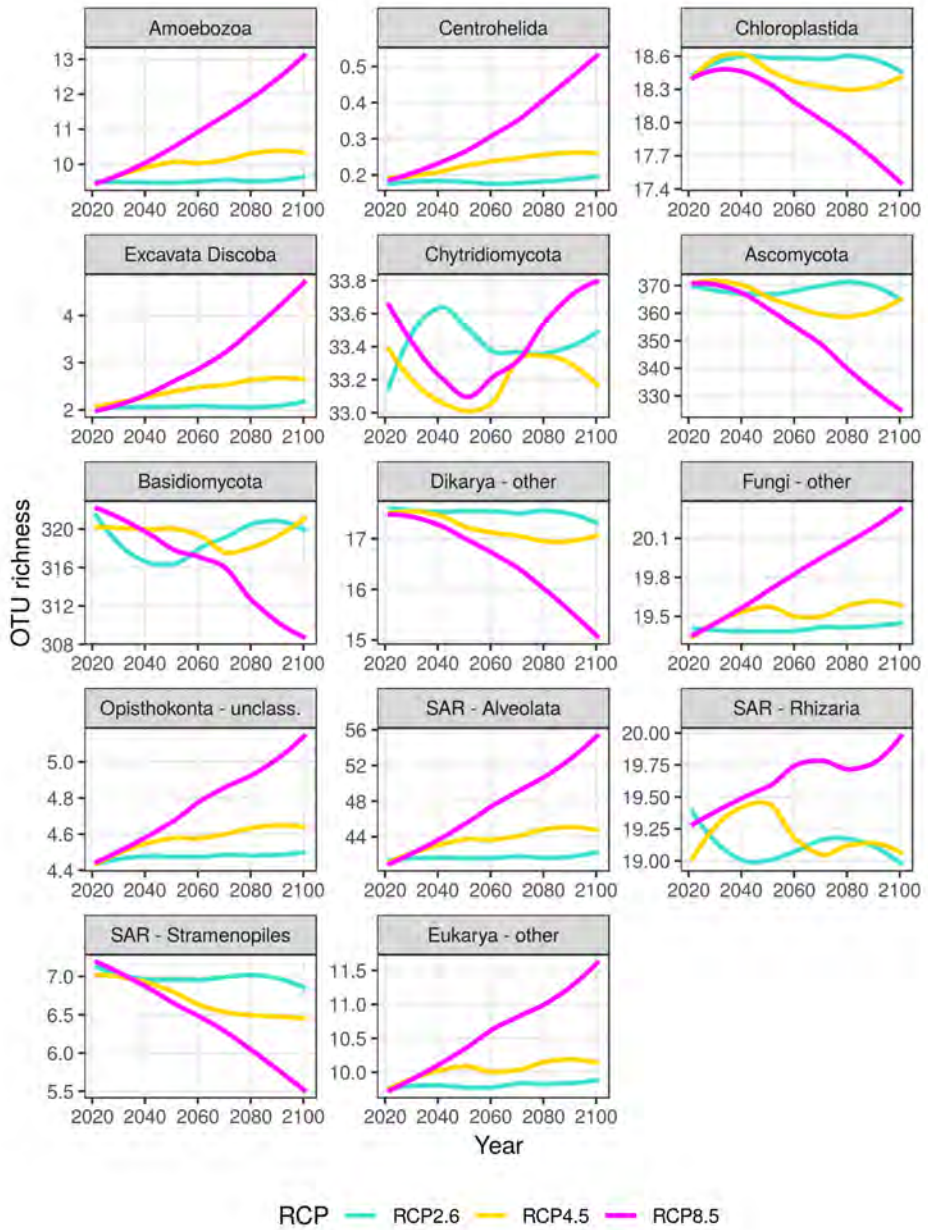


Figure C.6: Prediction for eukaryal groups.

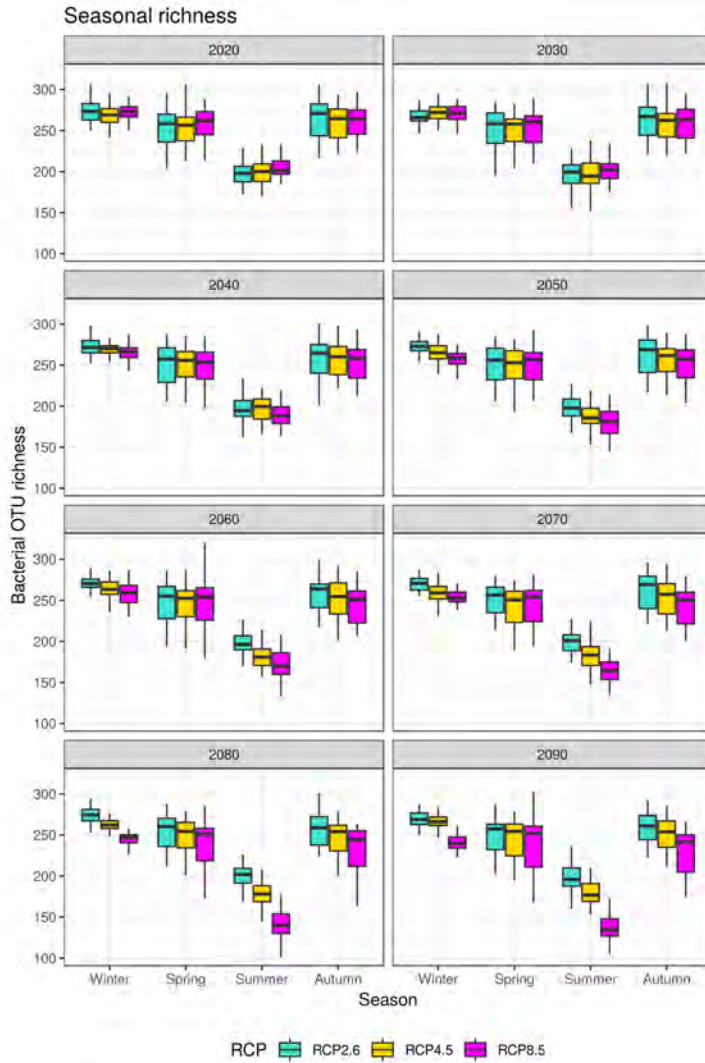


Figure C.7: **Decadal prediction for seasonal bacterial OTU richness.** Each panel corresponds to ten years, i.e., 2020 shows to the period from 2021 to 2030. We have considered seasons as groups of three months; for example, Winter comprises the months of January, February, and March.

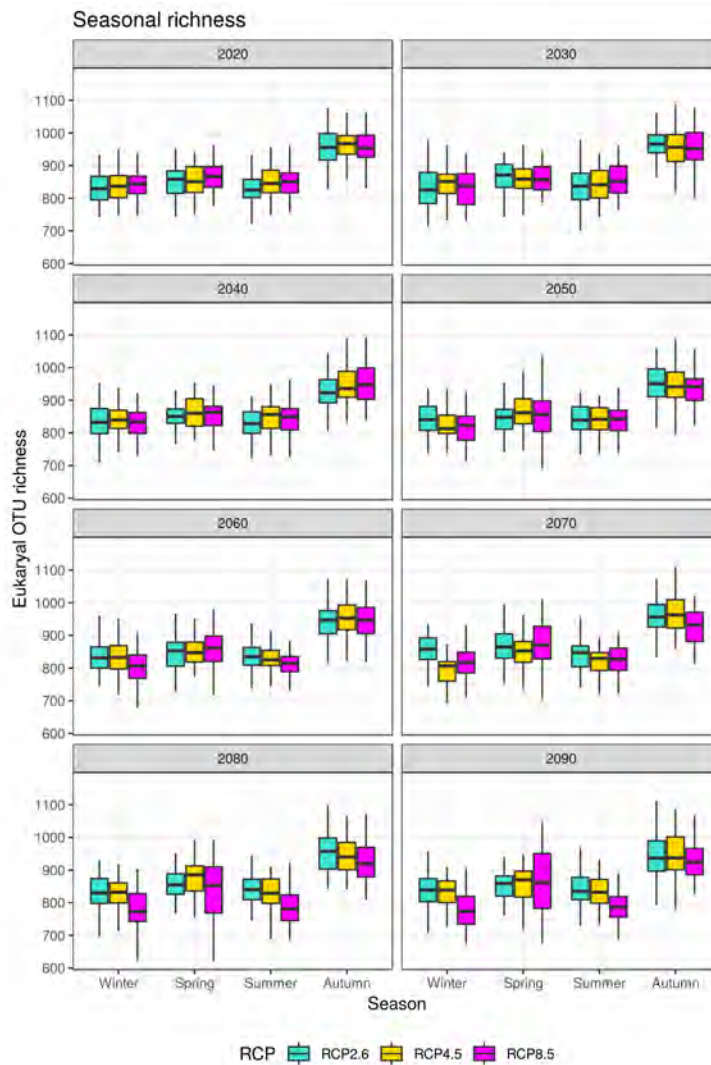


Figure C.8: **Decadal prediction for seasonal eukaryal OTU richness.** Each panel corresponds to ten years, i.e., 2020 shows to the period from 2021 to 2030. We have considered seasons as groups of three months; for example, Winter comprises the months of January, February, and March.

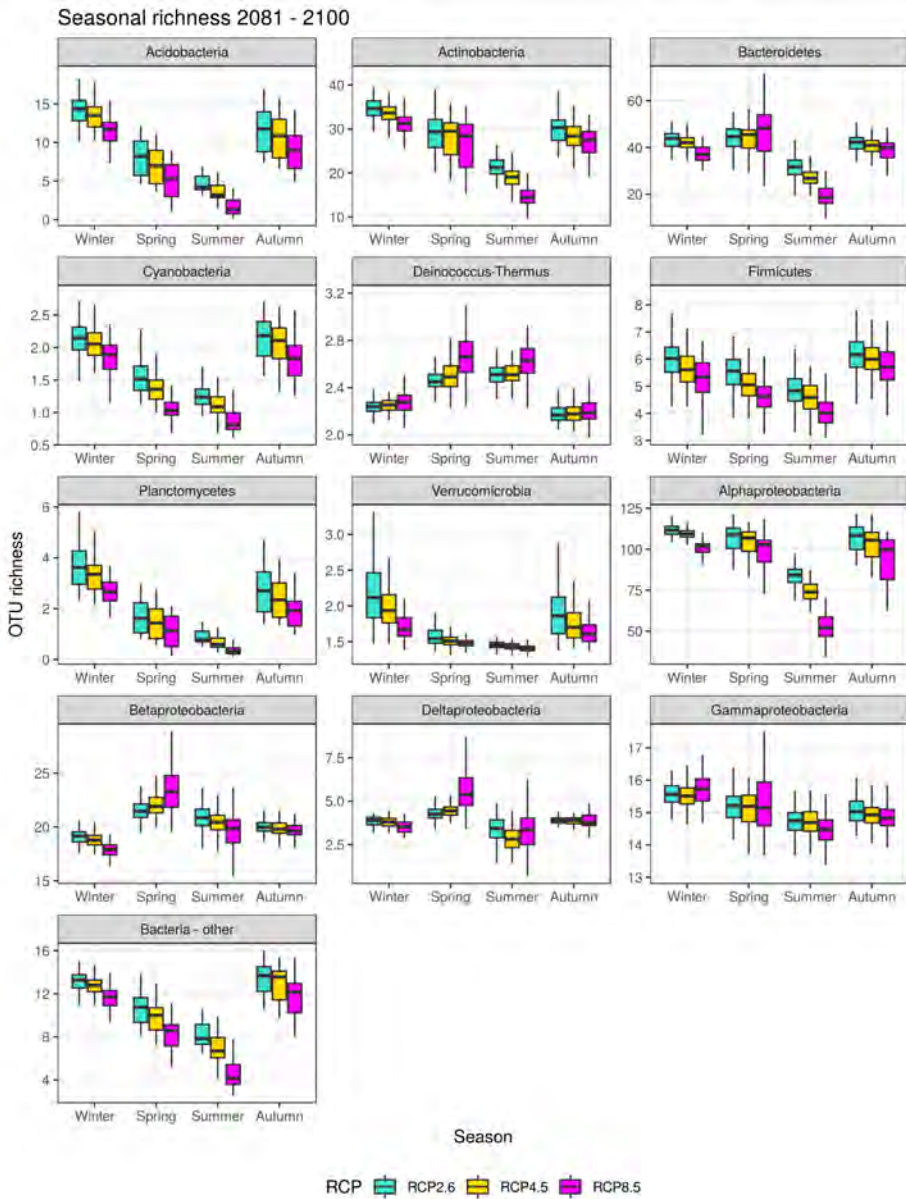


Figure C.9: **Bacterial groups prediction of seasonal OTU richness in the period 2081 – 2100.** We have considered seasons as groups of three months; for example, Winter comprises the months of January, February, and March.

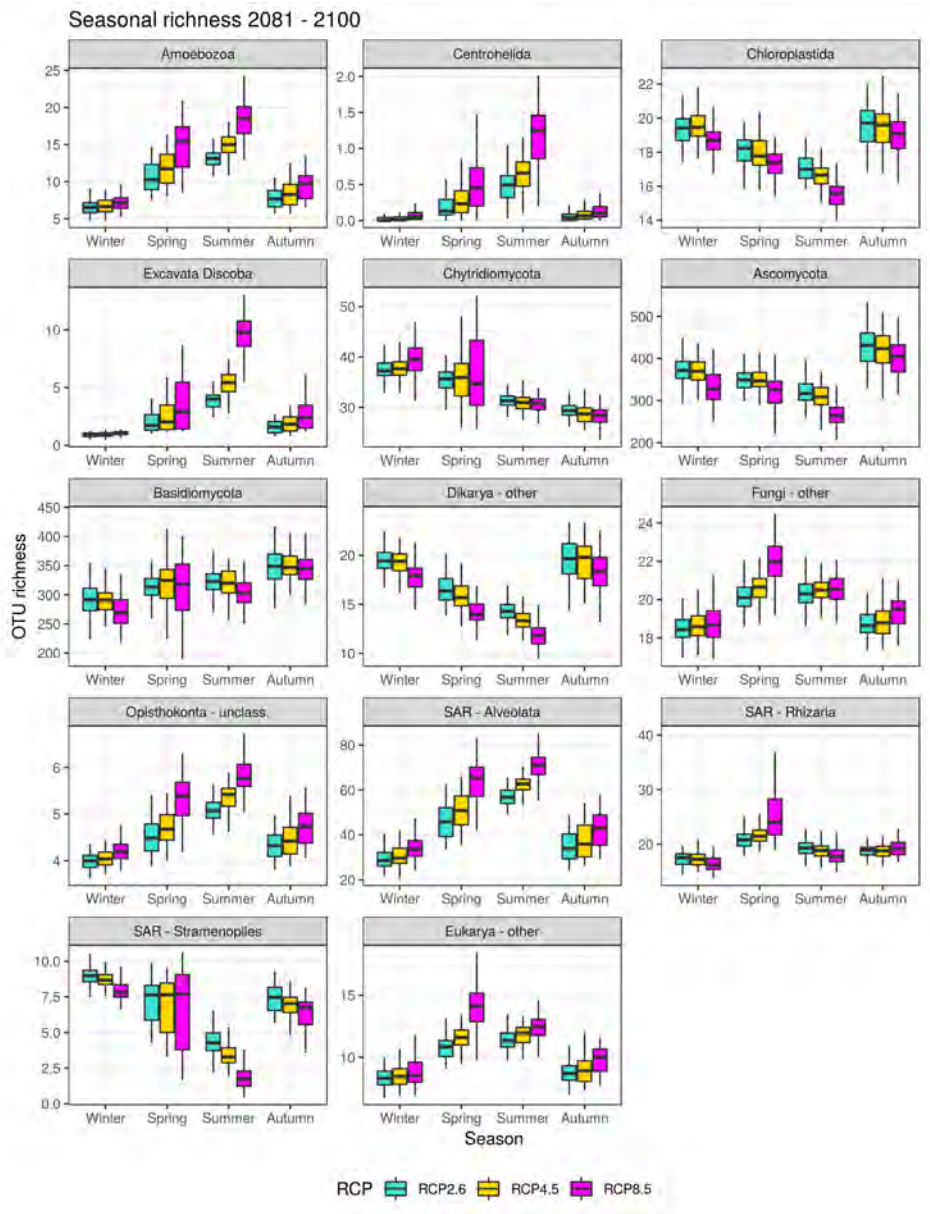


Figure C.10: **Eukaryal groups prediction of seasonal OTU richness in the period 2081 – 2100.** We have considered seasons as groups of three months; for example, Winter comprises the months of January, February, and March.