



INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Ramiro María
Martín Devasa

PhD Thesis

DEVELOPMENT OF
METHODS FOR THE STUDY
OF BIODIVERSITY BASED ON
MODELLING BIOGEOGRAPHIC
PATTERNS

Santiago de Compostela, 2023



PhD THESIS

**DEVELOPMENT OF METHODS FOR
THE STUDY OF BIODIVERSITY
BASED ON MODELLING
BIOGEOGRAPHIC PATTERNS**

Ramiro María Martín Devasa

INTERNATIONAL DOCTORAL SCHOOL OF THE UNIVERSITY OF SANTIAGO DE COMPOSTELA

DOCTORAL PROGRAMME IN BIODIVERSITY AND PRESERVATION OF THE NATURAL
ENVIRONMENT

SANTIAGO DE COMPOSTELA

2023



DECLARACIÓN DEL AUTOR DE LA TESIS

D. **Ramiro María Martín Devasa**

Título de la tesis: **Development of methods for the study of biodiversity based on modelling biogeographic patterns**

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento y declaro que:

- 1) La tesis abarca los resultados de la elaboración de mi trabajo.
- 2) De ser el caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.
- 3) Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.
- 4) La tesis es la versión definitiva presentada para su defensa y coincide la versión impresa con la presentada en formato electrónico.

Y me comprometo a presentar el Compromiso Documental de Supervisión en el caso que el original no esté depositado en la Escuela.

En **Santiago de Compostela**, Seleccione la fecha.

Firma

AUTORIZACIÓN DEL DIRECTOR / TUTOR DE LA TESIS

**Development of methods for the study of biodiversity based on
modelling biogeographical patterns**

D. Andrés Baselga Fraga

D^a. Carola Gómez-Rodríguez

INFORMAN:

Que la presente tesis, se corresponde con el trabajo realizado por D. Ramiro María Martín Devasa, bajo mi dirección/tutorización, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

De acuerdo con lo indicado en el Reglamento de Estudios de Doctorado, declara también que la presente tesis doctoral es idónea para ser defendida en base a la modalidad de Monográfica con reproducción de publicaciones, en los que la participación del doctorando fue decisiva para su elaboración y las publicaciones se ajustan al Plan de Investigación.

En Santiago de Compostela, de de 2023

Agradecimientos

Primero de todo agradecer a mis directores de tesis (y TFG, y TFM) Andrés y Carola, por todo el esfuerzo y tiempo dedicado a que un estudiante que estaba acabando tercero de carrera pudiera llegar a hacer una tesis y llegar a ser doctor, y lo más importante, poder decir que es científico.

Agradecérselo también a Rosa, que puso su parte en ello.

También a Fabien, que me acogió en la Universidad de Montpellier en tiempos de covid.

A mi familia y en especial a mis padres, por el apoyo y la confianza de todos estos años, las cebollas y los puerros.

A toda la gente del laboratorio, Andrea y Adrián que me enseñaron los primeros pasos, Victoria y Alba que me vieron en los últimos. Olalla, que está ahí para enseñar un par de cosas cuando hace falta.

Y muy, muy especialmente a Sara, por todas las horas, el esfuerzo, la compañía, y básicamente por todo. Porque eres capaz de mucho más de lo que crees.

A Mari, por tantas y tantas horas, y por tantos consejos sobre la vida en Bertairáns.

A Jose, también a Natali, por las horas hablando de ciencia y esas grandes sobremesas de pulpo.

A Erika, porque se puede ser de un lugar aunque hayas nacido y vivas a miles de kilómetros.

A tanta otra gente de la facultad, profesores y compañeros con los que siempre pude contar.

A mis compañeros de carrera (en especial Laura, Emilio y Adrián), fueron muchas horas, muchos trabajos, y también muchas risas.

A mis compañeros de Máster, con los que descubrí el mundo de las matemáticas y la estadística, sin ser nosotros nada de ello (hasta entonces).

También merecen mención mis compañeros de pista y amigos fuera de ella, por las horas, los juegos y las conversaciones sobre...bueno, las conversaciones.

A los amigos de toda la vida, Borja, Ana y demás personas que siempre están, y que sigan estando.

Y a los que he hecho durante esta tesis, Raquel, Sori, Andrea...

A todos los demás, a los que estuvieron, están y estarán, podéis identificaros en estas líneas y recibir, merecidamente, mi más sincero agradecimiento.

Publicaciones derivadas de esta tesis

Artículos

Martín-Devasa, R., Baselga, A. (2018). What caused the disjunct distributions of the *Lachnaia tristigma* species-group (Coleoptera: Chrysomelidae) on the Iberian peninsula? *Coleopterists Bulletin* 72: 421-426

Factor de impacto: 0.795, Cuartil: Q4.

Martín-Devasa, R., Martínez-Santalla, S., Gómez-Rodríguez, C., Crujeiras, R.M., Baselga, A. (2022). Comparing distance-decay parameters: a novel test under pairwise dependence. *Ecological Informatics* 72: 101894

Factor de impacto: 4.498, CiteScore: 5.4, Cuartil: Q1.

Martín-Devasa, R., Martínez-Santalla, S., Gómez-Rodríguez, C., Crujeiras, R.M., Baselga, A. (2022). Species range size shapes distance-decay in community similarity. *Diversity and Distributions* 28: 1348-1357

Factor de impacto: 5.717, CiteScore: 8.2, Cuartil: Q1.

Aportaciones al paquete *betapart* (R) versión 1.6.

Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieux, F., Logez, M., Martínez-Santalla, S., **Martín-Devasa, R.**, Gómez-Rodríguez, C., & Crujeiras, R. M. (2023). *betapart: Partitioning beta diversity into turnover and nestedness* (1.6). <http://CRAN.R-project.org/package=betapart>

- i) Implementación de la función de Gompertz en *decay.model()*.
- ii) Función *zdep*.

INDEX

Abstract in Galician (resumo), Spanish (resumen) and English.....	15
Resumen ampliado en español	17
INTRODUCTION.....	23
OBJECTIVES	31
1. SPECIES RANGE SIZE SHAPES DISTANCE-DECAY IN COMMUNITY SIMILARITY. 35	
ABSTRACT.....	37
1.1 INTRODUCTION	37
1.2 METHODS	41
1.2.1 Data simulation	41
1.2.2 Distance-decay models of simulated data	42
1.2.3 Real data.....	43
1.3 RESULTS.....	43
1.4 DISCUSSION.....	46
2. COMPARING DISTANCE-DECAY PARAMETERS: A NOVEL TEST UNDER PAIRWISE DEPENDENCE	51
ABSTRACT.....	53
2.1 INTRODUCTION	53
2.2 MATERIAL AND METHODS	58
2.2.1 Simulation of virtual biological communities.....	58
2.2.2 Simulation scenarios for hypothesis testing.....	59
2.2.2.1 Null hypothesis.....	59
2.2.2.2 Alternative hypothesis.....	60
2.2.3 Distance-decay models.....	61
2.2.4 zdep : Calibration procedure	61
2.2.5 Application of the zdep to real data.....	62
2.3 RESULTS.....	62
2.3.1 Calibration of the zdep test.....	63
2.3.2 Real data.....	63
2.4 DISCUSSION.....	64
2.5 CONCLUSIONS	67
3. DISPERSAL LIMITATION SHAPES DISTANCE-DECAY PATTERNS OF EUROPEAN SPIDERS AT THE CONTINENTAL SCALE	69

ABSTRACT.....	71
3.1 INTRODUCTION.....	72
3.2 MATERIAL AND METHODS	75
3.2.1 Distribution data.....	75
3.2.2 Community similarity and distance-decay models	75
3.2.3 Relationship between beta diversity patterns and dispersal-related traits of spiders	77
3.3 RESULTS.....	78
3.4 DISCUSSION.....	82
4. INFERRING THE CAUSAL PROCESSES OF GENETIC STRUCTURE FROM THE SHAPE OF THE GENETIC-SPATIAL DISTANCE RELATIONSHIP	87
ABSTRACT.....	89
4.1 INTRODUCTION.....	89
4.2 MATERIAL AND METHODS	92
4.2.1 Genetic and geographic data	92
4.2.2 Model fitting and parameters comparison.....	93
4.3 RESULTS.....	93
4.4 DISCUSSION.....	97
5. WHAT CAUSED THE DISJUNCT DISTRIBUTIONS OF THE <i>LACHNAIA TRISTIGMA</i> SPECIES-GROUP (COLEOPTERA: CHRYSOMELIDAE) ON THE IBERIAN PENINSULA?	101
ABSTRACT.....	103
5.1 INTRODUCTION.....	103
5.2 MATERIAL AND METHODS	104
5.3 RESULTS.....	105
5.4 DISCUSSION.....	108
DISCUSSION	111
CONCLUSIONS	121
REFERENCES.....	125
- Appendix 1.1: Gompertz function -.....	161
- Appendix 1.2: Code for simulations -.....	165
- Appendix 1.3: Supplementary tables Chapter I -	171
- Appendix 2.1: Test calibration -	177
- Appendix 2.2: Supplementary tables Chapter II -	181
- Appendix 2.3: Code for simulation study H_0 -.....	185
- Appendix 2.4: Code for simulation study H_1 -.....	199

- Appendix 3.1: Supplementary figures and tables Chapter III -	213
- Appendix 4.1: Supplementary tables Chapter IV -	219
Table and figure index	223

Abstract in Galician (resumo), Spanish (resumen) and English

Resumo

Os patróns bioxeográficos e macroecolóxicos son cruciais para comprender os procesos que condicionan a estrutura espacial da biodiversidade a diferentes niveis de organización biolóxica. A caracterización das formas funcionais destes patróns con funcións paramétricas e a comparación dos seus parámetros son polo tanto ferramentas importantes para o seu estudo. Esta tese ten como obxectivo desenvolver novas ferramentas para o estudo de dous importantes patróns, o *distance-decay*, tamén chamado decaemento da similitude biolóxica coa distancia, e a relación entre a distancia xenética e espacial, e aplicar estas novas metodoloxías ó estudo da bioxeografía de arañas europeas e á estrutura xenética de poboacións de crisomélidos ibéricos. Os métodos propostos son o uso da función de Gompertz para o modelado do *distance-decay*, un novo estatístico (z_{dep}) para a comparación dos parámetros de modelos axustados con datos con dependencia a pares e a súa expansión ao estudo da relación entre a distancia xenética e espacial. Nestes estudos tamén se revelou a relación entre a forma funcional do *distance-decay* e o tamaño da distribución das especies. Os resultados revelan que a forma funcional do *distance-decay* troca de *power-law* a exponencial negativa e a sigmoidal (axustada pola función de Gompertz) a medida que aumenta o tamaño medio de distribución dos taxa ou a proporción de grandes distribucións. A función de Gompertz e o estatístico z_{dep} mostraron a súa versatilidade para estudar o *distance-decay* e a relación entre distancias xenéticas e espaciais tanto en simulacións coma con datos empíricos. Usando estas técnicas, revelouuse que a estrutura espacial das comunidades de arañas europeas e a estrutura xenética en poboacións de crisomélidos ibéricos están probablemente condicionadas por procesos de limitación á dispersión. En conxunto, esta tese contribúe a ampliar as ferramentas para estudos bioxeográficos e macroecolóxicos con novas metodoloxías, e a afondar no coñecemento dos patróns bioxeográficos de arañas e crisomélidos.

Palabras chave: diversidade beta, biodiversidade, patróns bioxeográficos, *distance-decay*, macroecoloxía.

Resumen

Los patrones biogeográficos y macroecológicos son cruciales para comprender los procesos que condicionan la estructura espacial de la biodiversidad a diferentes niveles de organización biológica. La caracterización de las formas funcionales de estos patrones con funciones paramétricas y la comparación de sus parámetros son por lo tanto herramientas importantes para su estudio. Esta tesis tiene como objetivo desarrollar nuevas herramientas para el estudio de dos importantes patrones, el *distance-decay*, también llamado descenso de la similitud biótica con la distancia, y la relación entre la distancia genética y la espacial, y aplicar estas nuevas metodologías al estudio de la biogeografía de arañas europeas y la estructura genética de poblaciones de crisomélidos ibéricos. Los métodos propuestos son el uso de la función de Gompertz para la modelización del *distance-decay*, un nuevo estadístico (z_{dep}) para la comparación de parámetros de modelos ajustados con datos con dependencia a pares y su expansión al estudio de la relación entre la distancia genética y la espacial. En estos estudios también se evaluó la relación entre la forma funcional del *distance-decay* y el tamaño de la distribución de las especies. Los resultados revelan que la forma funcional del *distance-decay* cambia de *power-law* a exponencial negativa y a sigmoidal (ajustada por la función de Gompertz) a medida que aumenta el tamaño medio

de distribución de los taxa o la proporción de grandes distribuciones. La función de Gompertz y el estadístico Z_{dep} mostraron su versatilidad para el estudio del *distance-decay* y la relación entre la distancia genética y la espacial tanto en simulaciones como con datos empíricos. Usando estas técnicas, se reveló que la estructura espacial de las comunidades de arañas europeas y la estructura genética en poblaciones de crisomélidos ibéricos están condicionadas por procesos de limitación a la dispersión. En conjunto, esta tesis contribuye a ampliar las herramientas para estudios biogeográficos y macroecológicos con nuevas metodologías, y a profundizar en el conocimiento de los patrones biogeográficos de arañas y crisomélidos.

Palabras clave: diversidad beta, biodiversidad, patrones biogeográficos, *distance-decay*, macroecología.

Abstract

Biogeographical and macroecological patterns are crucial to understand the processes shaping the spatial structure of biodiversity at different levels of biological organization. The characterization of the functional shapes of these patterns with parametric functions and the comparison of their parameters are hence important tools for these studies. This thesis aims is to develop new tools for the study of two important patterns, the distance-decay (i.e. decay in biological similarity with distance) and the genetic-spatial distance relationship, and apply these new methodologies to the study of the biogeography of European spiders and the genetic structure of populations of Iberian leaf beetles. The methods introduced here include the use of the Gompertz function for the modelling of distance-decay patterns, a new test statistic (Z_{dep}) for parameters comparison of models fitted with pairwise dependent data and their expansion to the study of the genetic-spatial distance relationship. In these studies, the relationship between the functional form of distance-decay patterns and the species range size was also assessed. The results show that the functional form of distance-decay patterns changes from power-law to negative exponential and to a sigmoidal form (fitted by the Gompertz function) as mean species range size or the proportion of large ranges increase. The Gompertz function and the Z_{dep} statistic showed their versatility to study distance-decay and genetic-spatial relationship patterns in both simulated and empirical scenarios. Using these techniques, it was revealed that the spatial structure of European spider diversity is mostly driven by dispersal limitation, which is also an important factor driving the genetic structure of populations of Iberian leaf beetles. Altogether, this thesis contributes to the toolbox for biogeographical and macroecological studies with new and useful methods and to increase our knowledge of spiders and beetles biogeographical patterns.

Keywords: beta diversity, biodiversity, biogeographical patterns, distance-decay, macroecology.

Resumen ampliado en español

Los patrones biogeográficos y macroecológicos son cruciales para comprender los procesos que condicionan la estructura espacial de la biodiversidad en los diferentes niveles de organización biológica, desde especies hasta comunidades pasando por poblaciones. La biogeografía estudia las razones que determinan la distribución espacial de la biodiversidad, empleando aproximaciones ecológicas, históricas y evolutivas. Por su parte, la macroecología investiga los patrones ecológicos de especies y comunidades a gran escala, para desvelar relaciones entre dichos patrones y los factores abióticos que caracterizan a cada sistema. Existe una amplia gama de ejemplos de patrones macroecológicos descritos en la literatura. Algunos de los patrones macroecológicos más estudiados son el gradiente latitudinal de riqueza de especies, la distribución geográfica de los tamaños de distribución de las especies, la relación rango-abundancia, la relación entre el número de especies y el área (*Species-Area Relationship*, SAR), la relación entre el número de endemismos y el área (*Endemism-Area Relationship*, EAR). No obstante, la presente tesis doctoral se ha enfocado en dos de ellos debido a su especial interés e importancia para la ecología, ya que reflejan como la distancia espacial entre localidades afecta a la similitud entre sus poblaciones y comunidades. El primero de ellos es el patrón de *distance-decay* (descenso de la similitud biológica entre comunidades con el aumento de la distancia genética que las separa) y el segundo el *genetic-spatial distance relationship* (aumento de las distancias genéticas entre poblaciones con el incremento de la distancia espacial que las separa). Específicamente la presente tesis pretende desarrollar herramientas de acceso abierto que permitan el estudio y caracterización de ambos patrones macroecológicos.

Ambos patrones definen el aumento de diferencias entre poblaciones o comunidades con el aumento de la distancia espacial que las separa, aunque modelada de forma inversa. El primero de ellos, el *distance-decay*, se describe como la disminución de la similitud entre comunidades al incrementar la distancia, construyéndose regresando la similitud entre comunidades sobre la distancia espacial que las separa. El segundo de ellos, el *genetic-spatial distance relationship*, se describe como el aumento de las distancias genéticas entre poblaciones de una misma especie con la distancia, construyéndose mediante la regresión de las distancias genéticas entre poblaciones sobre la distancia espacial que las separa. Ambos patrones tienen importantes similitudes, e incluso pueden llegar a considerarse análogos. Desde el punto de vista ecológico ambos dependen de los mismos procesos: dispersión, selección, evolución y procesos estocásticos. Matemáticamente, ambos se construyen regresando una variable de similitud o disimilitud entre localidades sobre la distancia espacial que los separa. La caracterización de estos patrones biogeográficos se suele hacer con funciones paramétricas, que permiten modelar su forma funcional y obtener los parámetros de la función. Además, la comparación de estos parámetros permite arrojar luz sobre el diferente efecto de los procesos biogeográficos en diferentes grupos biológicos, territorios o periodos de tiempo.

Las funciones más utilizadas son la *power-law* y la exponencial negativa, aunque en los estudios de patrones genéticos también se puede utilizar una función lineal. Sin embargo, en los patrones de *distance-decay* y de *genetic-spatial distance relationship* la relación entre disimilitud y la distancia geográfica puede dibujar una forma sigmoideal. Esto ocurre cuando poblaciones o comunidades cercanas son muy similares (porque las especies tienen buena capacidad de dispersión) lo que hace que las disimilitudes entre las poblaciones no aumenten (o las similitudes no decaigan) en distancias espaciales pequeñas, permaneciendo constantes. Desafortunadamente, ni la función *power-law* ni la exponencial negativa pueden ajustar bien esta forma sigmoideal, por lo que contar con una función que ajuste la forma sigmoideal de estos patrones biogeográficos es sumamente importante, para poder modelar bien todas las posibles formas de los patrones biogeográficos. Por otro lado, en la comparación de parámetros

obtenidos de la función que describe el patrón ecológico, es crítico considerar la dependencia a pares (cada valor es el resultado de la comparación entre dos localidades) de los índices de similitud, disimilitud y la distancia espacial. Esta dependencia puede sesgar los resultados de los test que se utilizan para la comparación a pares al incrementar el error de tipo I (rechazo de la hipótesis nula cuando es cierta).

Por todo esto, la presente tesis doctoral tiene como objetivo desarrollar y aplicar nuevas metodologías y herramientas, que permitan estudiar los mecanismos que impulsan la distribución espacial de la diversidad biológica, inferida a partir de los patrones de *distance-decay* y *genetic-spatial distance relationship*. Específicamente, la tesis explora: (i) la utilidad de una función sigmoideal (función de Gompertz) para ajustar el patrón de *distance-decay* y examinar cómo cambia la forma funcional cuando se modifica el tamaño de distribución de las especies; (ii) la propuesta de la prueba estadística z_{dep} para la comparación de los parámetros obtenidos de las funciones; (iii) la aplicación de los métodos descritos en los objetivos anteriores para evaluar los procesos que condujeron al actual patrón espacial de diversidad de las arañas a escala europea y (iv) la aplicación de estos métodos para estudiar la relación de distancia genético-espacial usando especies endémicas de escarabajos ibéricos, junto con sus patrones de distribución. La aplicación de estos métodos al estudio de patrones biogeográficos de invertebrados terrestres ha supuesto un avance sin precedentes, ya que la amplia mayoría de estudios realizados hasta la fecha, han sido llevados a cabo en plantas y vertebrados, creando un sesgo en el conocimiento de estos invertebrados.

El Capítulo I se centra en investigar la relación entre el tamaño de la distribución de las especies y la forma del patrón de *distance-decay*, mediante la aplicación de la función de Gompertz para modelar dicho patrón espacial. Para ello se simulaban comunidades en un espacio virtual, compuestas por 300 especies con tamaños de distribución diferentes. La distribución de cada especie fue representada por un polígono espacial circular de diámetro definido dentro de un rango de valores previamente establecidos. Dichos valores fueron modificados para cada simulación realizada. En relación con el tamaño de las distribuciones, se definieron dos categorías dependiendo de su tamaño, grandes y pequeñas. Para obtener las curvas de *distance-decay*, en primer lugar se crearon puntos de muestreo en el espacio virtual distribuidos de forma aleatoria. Después se calcularon las especies presentes en cada punto de muestreo (matriz de presencia/ausencia) obteniendo la similitud entre ellos utilizando el índice de Simpson. En cuanto a la distancia espacial se calculó como la distancia euclídea entre cada par de puntos de muestreo. Con las matrices de similitud biológica y de distancia espacial se construyeron las curvas de *distance-decay*. Para comprobar qué función se ajustaba mejor a los datos, se ajustaron las curvas de las funciones *power-law*, exponencial negativa y Gompertz. La comparación de ajustes entre modelos se hizo basada en el AIC y pseudo- r^2 . Adicionalmente y con el objetivo de comprobar el correcto funcionamiento de esta función con datos empíricos, se usaron datos de distribución de mamíferos de Sudamérica. Para ello, se definieron cuatro grupos de acuerdo con los tamaños de las distribuciones, basados en los cuartiles de estos tamaños (Q1, Q2, Q3 y Q4) y al igual que con los datos simulados, se comprobó que función (*power-law*, exponencial negativa o Gompertz) presenta mejor ajuste en cada caso (menor AIC y mayor pseudo- r^2).

El Capítulo II tiene el objetivo de desarrollar un test estadístico robusto (z_{dep}) que permita la comparación de parámetros de las funciones utilizadas para ajustar el *distance-decay*. Específicamente, el principal reto que ha solventado el desarrollo de este estadístico ha sido evitar que esté afectado por la dependencia a pares de los datos ecológicos y por ende que aumente su error de tipo I. El estadístico propuesto (z_{dep}), se basa en la estandarización de la diferencia entre

los parámetros comparados. Al tratarse de una estandarización, el test sigue una distribución $N(0,1)$. La clave del correcto funcionamiento está en recoger la dependencia a pares utilizando *site-block resampling* (o remuestreo por bloques) para calcular la varianza de las estimaciones de los parámetros. Para comprobar el funcionamiento del estadístico se llevó a cabo un estudio de simulación. En este estudio se comprobó el error de tipo I (mediante la aproximación al nivel, α , bajo la hipótesis nula) y el error de tipo II (mediante el nivel, $1 - \text{error de tipo II}$, bajo la hipótesis alternativa) del estadístico con datos con dependencia a pares. Utilizando un *framework* similar al del capítulo anterior, se simuló curvas de *distance-decay* y se comprobó cuál era el mejor modelo entre las funciones *power-law*, exponencial negativa y Gompertz. A continuación, se utilizó el estadístico z_{dep} para testar la igualdad entre parámetros de dos curvas seleccionadas (ajustadas con el mismo modelo), obteniendo el correspondiente p-valor. Repitiendo esta operación 1000 veces se obtuvieron vectores de 1000 p-valores para cada escenario simulado (se definieron tres escenarios bajo la hipótesis alternativa y tres bajo la nula, que se diferencian en la mejor función de ajuste a las curvas y en sus diferencias de parámetros), con los que se pudo calcular: la aproximación al nivel (error de tipo I) y la potencia ($1 - \text{error de tipo II}$) del estadístico. La hipótesis nula se simuló tomando dos muestras por bloques (las unidades muestreadas son el conjunto de valores derivados de la comparación de una localidad determinada con las demás) de las matrices de similitud y distancia espacial calculadas, mientras que la nula se simuló definiendo comunidades con composiciones en tamaños de distribución muy dispares. La aplicabilidad y correcto funcionamiento del test (z_{dep}), se evaluó en un caso empírico comparando los parámetros de las curvas de *distance-decay* de diversos órdenes de mamíferos en la península ibérica.

En el Capítulo III se evalúa el efecto relativo del espacio y las variables ambientales en la estructura espacial de la diversidad de arañas europeas, aplicando las herramientas desarrolladas en los dos anteriores capítulos. Para ello se obtuvieron datos de presencia de especies de arañas pertenecientes a 15 familias en los países europeos. Con el objetivo de evitar comparaciones entre territorios con grandes diferencias de tamaño, se seleccionaron los países pertenecientes a Europa continental y con un tamaño mayor a 2.000 km². También se obtuvieron rasgos de la capacidad de dispersión de las diferentes familias para comprobar si los patrones observados se relacionan con la capacidad de dispersión de dichas familias, y por tanto con procesos espaciales. Dado que la historia biogeográfica del norte y el sur de Europa es diferente (en los periodos glaciales el norte estuvo cubierto por glaciares, en los que sobrevivieron pocas especies, mientras que el sur sirvió como refugio glacial a una gran cantidad de especies), los datos se dividieron en dos grupos correspondientes al norte y del sur de Europa. Para evaluar los efectos del espacio y las variables ambientales en la composición de las comunidades de arañas europeas, se calcularon las curvas de *distance-decay* espacial y climático para el norte y el sur de Europa. Para poder compararlas directamente, se estandarizaron las distancias espaciales y climáticas entre 0 y 1. Una vez calculadas las curvas de *distance-decay* con las distancias estandarizadas se compararon los parámetros de los modelos espaciales y climáticos utilizando el test z_{dep} . Para conocer si las diferencias en los patrones biogeográficos entre territorios (norte-sur) se debían a procesos de dispersión, se estudió la relación entre las diferencias de pendientes norte-sur tanto en *distance-decay* espacial como climático y los rasgos de la capacidad de dispersión de las familias. Adicionalmente, para conocer si las diferencias entre comunidades de arañas fueron debidas a la sustitución de especies o a diferencias en riqueza de especies se investigó la composición de las comunidades de arañas entre el norte y sur de Europa.

Una vez desarrollada y aplicada la metodología de *distance-decay*, en el capítulo IV se examinó el uso de la función de Gompertz y del estadístico z_{dep} para el estudio del patrón de *genetic-spatial distance relationship*, utilizando como grupo de estudio las especies endémicas de crisomélidos ibéricos.

Para el desarrollo de este capítulo usamos la base de datos GeneBank para obtener el fragmento *barcode* (cox1-5') de cada especie, junto con su correspondiente localización geográfica. Las especies con menos de 15 secuencias disponibles o presentes en menos de tres localidades fueron excluidas del estudio para que los resultados fueran lo más robustos posible. Una vez construida la base de datos, compuesta por 10 especies de crisomélidos, se colapsaron los haplotipos únicos para cada una de las especies. En base a esos haplotipos únicos se calcularon las distancias genéticas entre poblaciones de la misma especie. Paralelamente, se calcularon las distancias espaciales entre las localidades en las que se colectó cada individuo usando las coordenadas asociadas a cada muestra. Con las matrices de distancia genética y distancia espacial se construyeron las curvas de *genetic-spatial distance relationship*, que relacionaban ambas distancias para cada especie. A continuación, se ajustaron las funciones lineal, *power-law*, exponencial negativa y de Gompertz con el objetivo de testar, si la relación entre distancias genética y espacial era significativa para cada especie. Los ajustes de cada función, como en capítulos anteriores, se compararon utilizando el AIC y pseudo-r². Paralelamente, se construyeron redes de haplotipos para conocer la estructura espacial de su diversidad genética para cada especie de crisomélido. Por último, se evaluó si la presencia de relación entre distancias genética y espaciales está relacionado con la capacidad de dispersión de las especies, utilizando la presencia de alas como *proxy* de la capacidad de dispersión.

En el capítulo V se estudian las causas de la distribución disyunta que presentan las especies del grupo de *Lachnaia tristigma*. El grupo de especies de *L. tristigma* se compone de tres especies: *L. tristigma*, *L. gallaeca* y *L. pseudobaratraea*, todas ellas endemismos de la península ibérica, aunque es posible que *L. tristigma* se encuentre también en el sur de Francia. Estas especies son evolutivamente muy cercanas, sin embargo, sus distribuciones no se solapan. Para conocer las posibles causas de esta distribución se han considerado factores ambientales (nicho climático) y factores históricos. En primer lugar, para evaluar el papel que ha tenido el nicho climático en la desigual distribución de las tres especies, este se caracterizó para cada una de ellas, utilizando seis variables climáticas de precipitación y temperatura. Seguidamente, se redujeron las dimensiones del espacio climático definido por estas seis variables a dos, aplicando un PCA, lo que permitió la comparación de nichos. En segundo lugar, para considerar el papel de los factores históricos (las glaciaciones del Pleistoceno) en la actual distribución de las tres especies, se realizaron dos modelos de distribución de especies (Bioclim y Domain), los cuales permitieron conocer la distribución potencial de las especies en el último máximo glacial. Para ello se proyectaron los nichos climáticos de cada una de las especies bajo las condiciones del último máximo glacial, pudiendo determinar así los lugares en los que las especies pudieron sobrevivir a las glaciaciones. Permitiendo conocer si la existencia de diferentes refugios pudo condicionar la distribución actual de estas especies y explicar sus patrones actuales espaciales de no solapamiento.

Tras los análisis de los Capítulos I y II, los resultados revelaron que la forma funcional del *distance-decay* cambia a medida que aumenta el tamaño medio de distribución de las especies y/o la proporción de distribuciones grandes. A medida que aumentan estos dos factores la función que presenta un mejor ajuste oscila entre una *power-law* (con tamaños de distribución pequeños y una proporción baja de distribuciones grandes), a una exponencial negativa y una Gompertz (con tamaños medios de distribución grandes y/o una elevada proporción de distribuciones grandes). La función de Gompertz resultó ser sumamente útil para ajustar el patrón cuando este dibuja forma sigmoidal, y se presenta como una adición clave al conjunto de herramientas para el análisis del patrón de *distance-decay*. Por su parte, el estadístico z_{dep} demostró una versátil aplicabilidad tanto en el estudio de simulación como con los datos empíricos de vertebrados ibéricos, evidenciando

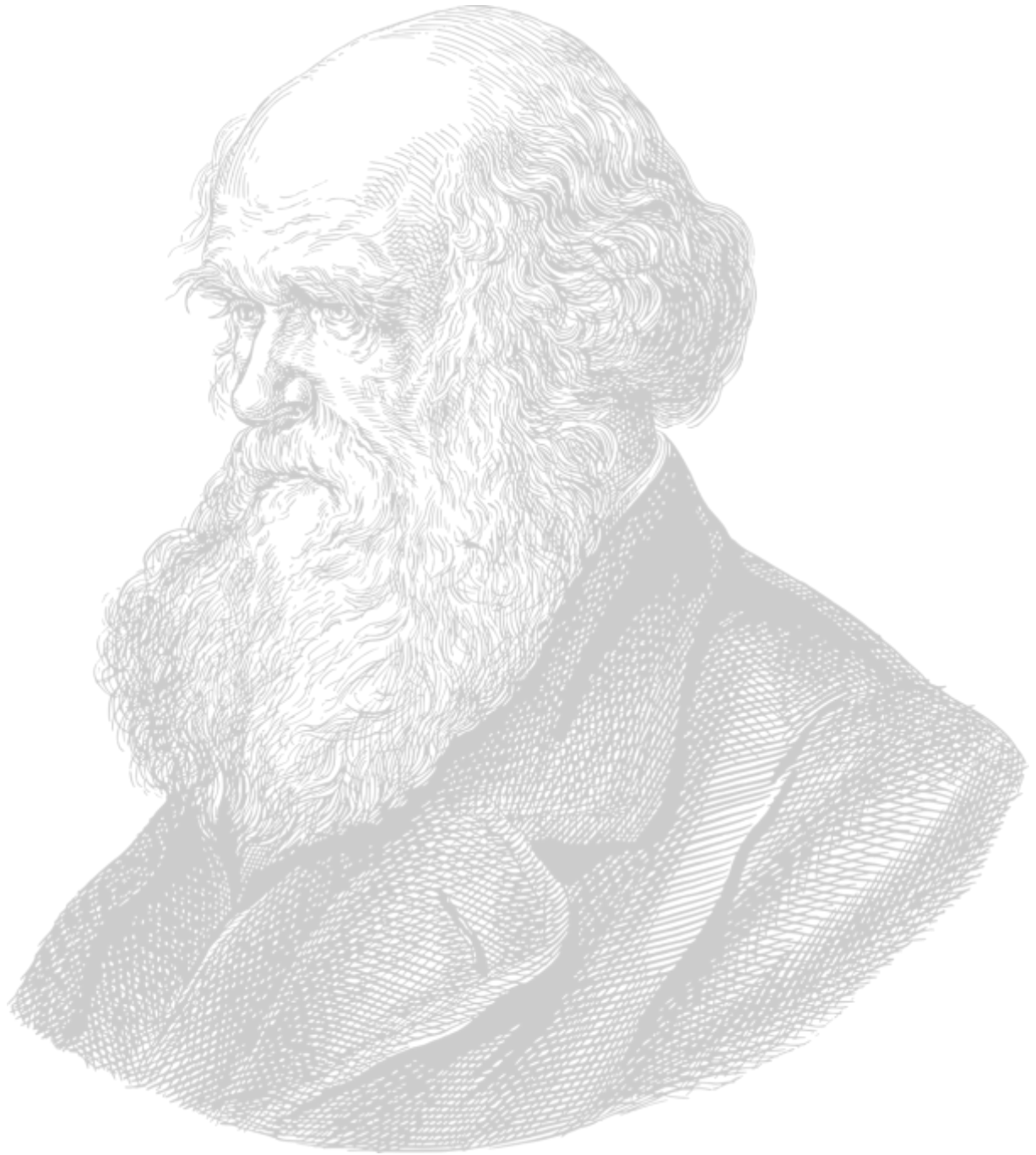
una fiel aproximación al nivel para los tres niveles considerados ($\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$) y una buena potencia (cercana o igual a 1). Ello implica que el estadístico z_{dep} es capaz de detectar diferencias significativas entre parámetros de modelos ajustados con datos con dependencia a pares, sin que su error de tipo I se vea afectado. Por lo que, estos resultados sugieren que el estadístico z_{dep} es una herramienta esencial en los estudios biogeográficos.

Con respecto al estudio aplicado del Capítulo III, la comparación de las curvas de *distance-decay* espaciales y climáticas utilizando distancias estandarizadas, la relación entre los parámetros de *distance-decay* y los rasgos de la capacidad de dispersión de las familias de arañas. Los resultados reflejan que las diferencias en las comunidades de arañas entre el norte y el sur son, principalmente, causadas por diferencias en riqueza de especies. Ello sugiere que la limitación de la dispersión, durante el proceso de colonización postglacial es el principal factor que determina la composición de las comunidades actuales y los patrones biogeográficos de las arañas en la Europa continental.

Con respecto al patrón de *genic-spatial distance relationship*, los resultados muestran que tanto la función de Gompertz como el estadístico z_{dep} tienen gran aplicabilidad en su estudio. Esto se evaluó utilizando crisomélidos endémicos de la península ibérica como caso de estudio en el Capítulo IV. En este estudio, ambos métodos resultaron sumamente útiles para inferir la historia biogeográfica y la capacidad de dispersión a largo plazo de las especies. La función de Gompertz mostró una gran plasticidad, adaptándose bien en la gran mayoría de escenarios en las que la relación entre las distancias genética y espacial fue significativa, y teniendo una especial utilidad en situaciones en las que el patrón toma forma sigmoidal. Más importante aún, los resultados de este capítulo reforzaron la idea de que la caracterización de la forma del *genetic-spatial distance relationship* es una herramienta clave para evaluar la relevancia relativa de los procesos causales que impulsan la estructura espacial de la diversidad genética (es decir, flujo genético frente a deriva genética) de las especies. Entre las especies endémicas ibéricas estudiadas en el capítulo IV, la forma de la relación entre las distancias genética y espacial varió ampliamente, incluso entre especies filogenéticamente cercanas, como *L. gallaeca* y *L. tristigma*, lo que sugiere capacidades de dispersión diferentes y diversas historias de expansión de sus distribuciones. Las diferencias en las dinámicas de distribuciones se exploraron en el Capítulo V, específicamente para el caso de *L. tristigma*, *L. gallaeca* y *L. pseudobarathraea*. Los resultados de este capítulo reflejaron que la evolución en refugios independientes de una posible especie ancestral durante el último máximo glacial, junto con una limitada expansión de sus distribuciones una vez terminadas las glaciaciones son las causas más probables de la configuración actual de las distribuciones disjuntas de estas tres especies.

En resumen, esta tesis propone y aplica métodos novedosos para el estudio de la biodiversidad a partir de la caracterización de patrones biogeográficos. Esta metodología se puede utilizar para modelar patrones a varios niveles de organización biológica y ha servido para profundizar nuestro conocimiento sobre la biogeografía de arañas y escarabajos. Estos nuevos métodos propuestos son una importante aportación al análisis de patrones biogeográficos ya que completan las herramientas disponibles para ello, permitiendo una mejor, más completa y más precisa caracterización. Además, pueden ser también muy útiles para estudios de biogeografía de la conservación y ecología de la conservación, permitiendo inferir los efectos de las actividades humanas en la genética de las poblaciones o la composición de las comunidades.

INTRODUCTION



INTRODUCTION

Understanding the factors affecting the spatial structure of biodiversity at various levels of biological organization is one of the major goals of ecologists (Brown & Maurer, 1989; Gotelli et al., 2009; Lomolino et al., 2017; Slatkin, 1987). From genes to communities and biomes, the knowledge about the processes driving the distribution of genetic variation, species distributions or community assemblages is fundamental to understand the dynamics of natural systems and their responses to a changing world (Ackerly et al., 2010; Richardson & Whittaker, 2010). This is the main focus of biogeography and macroecology. Biogeography studies how and why biodiversity is geographically distributed in space in the way it is, using ecological, historical and evolutionary approaches (Lomolino et al., 2017). Macroecology study large-scale patterns of ecological systems and search for statistical relationships to understand that distribution (Beck et al., 2012; Keith et al., 2012). Some of the macroecological patterns that regularly describe biodiversity are the range-abundance relationship, the range size distribution, the latitudinal gradient in species richness, the Species Area Relationship (SAR, De Candolle, 1855; Lomolino, 2000), the Endemism Area Relationship (EAR), the distance-decay of similarity (Nekola & White, 1999) or the genetic-spatial distance relationship (Hutchison & Templeton, 1999).

The biogeographical and macroecological patterns that describe the distribution of biodiversity, are driven by a series of processes conditioning that different species or genetic variants could live, develop, arrive or simply generate in certain places, but not in others (Vellend, 2010). Although the number of potential processes is very high, the underlying primary processes are just four (Vellend, 2016; Vellend et al., 2014). These four processes are analogous to the major processes considered in population genetics: mutation, dispersal, selection, and drift (Stephens, 2004; Vellend, 2010). Four analogous general processes operate at the community level, linking biodiversity patterns across different levels of biological organization (Vellend, 2010): speciation, dispersal, selection, and drift. Mutation and speciation refer to the appearance of genetic variants or new species, respectively, in a certain place, from where they can move, expand their range and colonise new territories (Losos & Schluter, 2000; Ricklefs, 2008). This colonisation is mediated by dispersal processes that condition the capability of different groups to colonise new territories (Clobert et al., 2012), which mainly depends on the dispersal ability of that groups and the geographic distance separating territories (Lester et al., 2007; Nathan, 2013). When arriving in a new territory, selection processes condition that species can live and develop there or maintain stable populations, depending on the relative fitness of the different species in those territories (Jablonski, 2008; Soininen, 2014). At the genetic level, neutral markers are not affected by natural selection, so they are only used

to study mutation and spatial effects (Holderegger et al., 2006; Kimura, 1986). Non-neutral (adaptive) markers are subject to selection, so their presence in a territory depends on the fitness that different alleles have in the conditions of that territory (Holderegger et al., 2006). The distribution of genetic variants and community composition are also affected by stochastic processes (drift), termed, respectively, genetic or ecological drift (Masel, 2011; Tilman, 2004), although their effects are normally difficult to assess. Working all together, these processes can explain an important part of the spatial configuration of biodiversity. In metacommunity ecology (Leibold et al., 2004), four paradigms have been described, depending especially on the relative importance of the dispersal and selective process in the configuration of metacommunities composition and dynamics. From the neutral paradigm supposing only dispersal effects (Hubbell, 2001) (expressed as patch connectivity) to mass-effects, patch dynamics and species sorting, focus on selection effects (driven by environmental heterogeneity) (Leibold et al., 2004; Leibold & Chase, 2018). In Biogeography, to refer to the selection processes the term species sorting is also used, whereas dispersal limitation is used to refer to dispersal processes (Jablonski, 2008; Rosindell et al., 2011; Soininen, 2014).

At the large spatial scales studied by biogeography and macroecology, it is important to consider that the aforementioned four major processes also take place at large temporal scales, and are thus affected by historical events, that can have lasting effects that are thus detectable in current biodiversity patterns. In Europe, historical events, especially Pleistocene glaciations, have had an important effect on the distribution of biodiversity (Calatayud et al., 2016; Davies et al., 2011; Hewitt, 2003; Stewart et al., 2016; Wilson & Veraguth, 2010). During glaciations, most of the continent was covered by ice sheets, and species survived in glacial refugia, in which conditions were suitable for their survival (Hewitt, 1996, 2000; Taberlet & Cheddadi, 2002). In Europe, Mediterranean peninsulas were the main glacial refugia for most species (Hewitt, 2000), although several extra-Mediterranean refugia have been described (Schmitt & Varga, 2012; Svenning et al., 2008). This had important effects at various levels. At the biogeographical level, glacial refugia became biodiversity hotspots as they retained species that were wiped out from glaciated areas (Petit et al., 2003). Nowadays, former glacial refugia continue to be the ones with the most species richness in the world, in part for that reason (Médail & Quézel, 1999). At the evolutionary level, populations of the same species could be isolated in different refugia. Isolation interrupted gene flow between populations, which evolved independently. This independent evolution led to a genetic divergence between populations that could eventually lead to speciation processes (Carstens & Knowles, 2007; Dufresnes et al., 2020; Hewitt, 1996). When the ice sheet retracted, environmental conditions became suitable for species to live in new territories, so they could colonise these territories and expand their distributions out of the refugia (Hewitt, 1999; Lesbarrères, 2009; Pamilo & Savolainen, 1999). Dispersal processes, and thus the different dispersal ability of different biological groups, played an important role in this post-glacial colonisation. Groups with high dispersal ability could colonise further territories than groups with low dispersal ability (Svenning & Skov, 2007). This caused marked differences in large scale biogeographical patterns of groups with different dispersal abilities (e.g., Gómez-Rodríguez & Baselga, 2018). Thus, the actual configuration of biodiversity and macroecological patterns are the result of the

combined effects of the four major processes (speciation, dispersal, selection, and drift), that are not static in time, but change their relative strengths because of the effects of historical factors (De Candolle, 1820; Douda et al., 2018). The effects of major processes and the influence of glaciations and post-glacial colonisations on macroecological patterns have been studied for a wide variety of biological groups, from plants (O'Connell et al., 2008; Petit et al., 2002; Gugger et al., 2008) to vertebrates (Herman et al., 2014; García-Vázquez et al., 2017; Stone et al., 2002; Burnham et al., 2009). However, the need to overcome the limited knowledge about the distribution, abundance, traits or interactions of most species (Hortal et al., 2015), specially for invertebrates (Cardoso et al., 2011; Diniz-Filho et al., 2010), opens interesting avenues for research as wide differences are expected to arise among different taxa.

Biogeographical and macroecological patterns are very important for the study of biodiversity because the pattern-to-process approach is the core of macroecology (Keith et al., 2012; McGill, 2019), that is, we use patterns to infer the processes behind. Therefore, given the importance that these patterns have for the study of the processes that cause the spatial structure of biodiversity, it is important to be able to characterize them accurately (Connolly et al., 2017). Normally, this characterization is carried out by fitting parametric functions to assess the relationship between biotic variables and their possible explanatory factors (e.g., Dengler, 2009; Harte et al., 2001; Lyons & Willig, 1999; Nekola & McGill, 2014). Parametric functions have several advantages in fitting these types of patterns. They allow for modelling the general shape of the patterns, which is relevant for inferring the causal processes if the shape of the pattern is predicted to vary depending on the relative importance and strength of different processes (Legendre & Legendre, 2012). Another advantage of parametric functions is precisely their quantitative estimation of parameters, which allow quantifying the change in any biological measure (e.g., species richness, compositional similarity or genetic dissimilarity) as a function a certain geographic or environmental factor (e.g., area, latitude, spatial distance or climatic differences). Moreover, these quantitative parameters can then be statistically compared across biological groups, scales, times or biological systems (Morlon et al., 2008; Steinbauer et al., 2012), providing another avenue to improve our inferences about the underlying processes. This is because model parameters are not only a measure of how biodiversity variables change with predictors, but they can be also interpreted in biological terms and be related, for example, to the dispersal ability of species (Qian, 2009; Saito et al., 2015) or niche breadth (Nekola & White, 1999; Stegen & Hurlbert, 2011). Thus, assessing the relationship between the parameters and the species traits related to dispersal ability and ecological niche is also a good method to reveal how dispersal limitation and niche breadth condition spatial diversity distribution in each particular case (Gómez-Rodríguez & Baselga, 2018; Saito et al., 2015).

At the community level, the characterization of the factors causing spatial variation in community composition (i.e. beta diversity) has centred the attention of biogeographers for a long time. Since the introduction of the term beta diversity (R. H. Whittaker, 1960, 1972), several methods and new metrics have been developed for the study of beta diversity (Anderson et al., 2011; Chao et al., 2012; Jost, 2007; Legendre et al., 2005; Tuomisto, 2010a, 2010b). However, different measures reflect different properties, and all give different results (Tuomisto, 2010a, 2010b; Vellend, 2001). The best-known beta diversity metrics are the

Sorensen (Koleff et al., 2003; Sørensen, 1948) and Jaccard (Jaccard, 1912) pairwise dissimilarity indices, that give a value of dissimilarity between communities (ranged between 0 and 1) accounting for unique and shared species between sites based on presence-absence data. For abundance data, the Bray-Curtis dissimilarity index (Bray & Curtis, 1957) is the most used. However, these indices are affected both by species replacement (turnover) and differences in species richness (nestedness) between sites (Baselga, 2010, 2012, 2013; Baselga & Leprieur, 2015). Baselga (2010) proposed the use of the Simpson index (Lennon et al., 2001; Simpson, 1943) to account for species turnover and derived an index for nestedness component of dissimilarity, both for pairwise and multiple-site analyses. The rationale behind is that Simpson dissimilarity is only affected by species replacement, and Simpson and Sorensen dissimilarities are equal in the absence of richness differences. Therefore, their subtraction (Sorensen minus Simpson) provides a measure of the fraction of dissimilarity that can be attributed to nestedness. These indices have also been expanded to compute phylogenetic beta diversity measures and functional beta diversity measures, that account not only for taxonomic differences between sites but also for phylogenetic (Leprieur et al., 2012) and functional (Villéger et al., 2013) similarities between communities. Modelling the relationship between species turnover or nestedness-resultant dissimilarity and spatial distances or environmental differences between sites is very useful to assess the effects of dispersal limitation and species sorting in driving the variation of community composition.

The study of the distance-decay pattern (Nekola & White, 1999), defined as the decrease of compositional similarity between sites with the spatial distance separating them, is a useful method to infer the factors behind community variation (Anderson et al., 2011). Originally, distance-decay models assess the relationship between compositional similarity and spatial distance between sites (Nekola & White, 1999; Morlon et al., 2008), but in a more general way the same framework can be used to analyse the relationship between compositional similarity and any difference between sites, such as climatic distances, differences in altitude and even in a pollution gradient (Goldenberg Vilar et al., 2014; Nekola & White, 1999; Soininen et al., 2007; Tuomisto et al., 2003). This allows for comparing the effects and relative importance of different variables in driving the variation in community composition. In other words, it allows assessing the spatial configuration of beta diversity in a given biological system and identifying the main processes leading to such configuration (Anderson et al., 2011; Graco-Roza et al., 2022; Socolar et al., 2016).

At the genetic level, the genetic-spatial distance relationship is a pattern analogous to the distance-decay, defined as the increase of genetic dissimilarity between individuals of the same species with spatial distance (Hutchison & Templeton, 1999; Slatkin, 1993). Although of opposite signs (distance-decay is the decrease of similarity and the genetic-spatial distance relationship the increase of a dissimilarity), distance-decay and genetic-spatial distance relationship have the same basic mathematical structure and can be produced by the same biological processes working at different levels of biological organization (Vellend, 2010). Biologically, both patterns are mainly shaped by mutation or speciation, dispersal, selection, and drift processes, working at the genetic (genetic-spatial distance relationship) and the community level (distance-decay), respectively. Regarding the causal processes, distance-

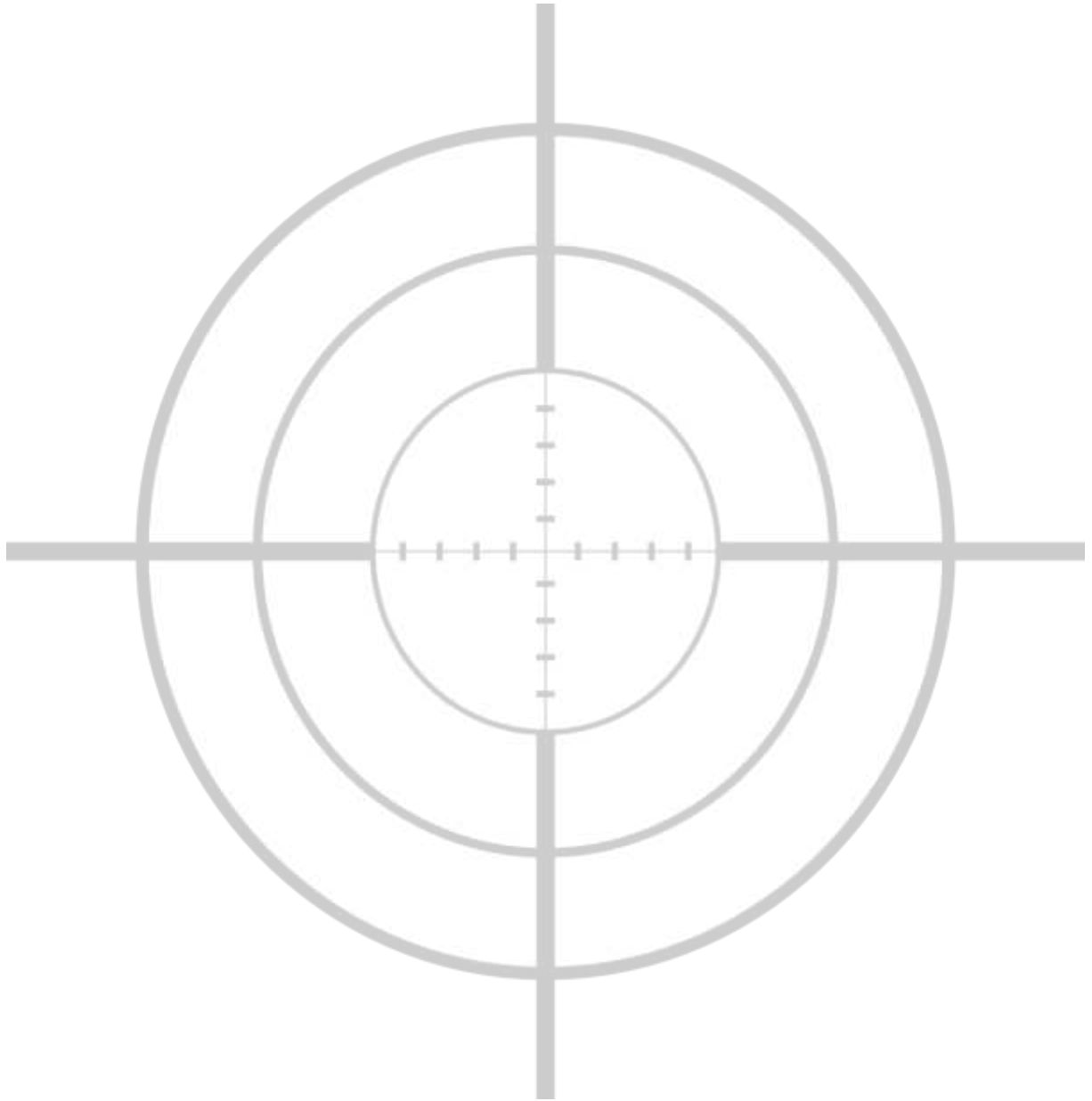
decay analyses may assess the effects of all four major processes in community composition (although for speciation phylogenetic information is needed) (Graham & Fine, 2008). In contrast, the genetic-spatial distance relationship is usually attributed only to dispersal and drift processes (Hutchison & Templeton, 1999), because it is usually studied in neutral genetic markers that are not influenced by selection. When non-neutral markers are considered, the effects of the environment over the spatial structure of genetic variation within species can be assessed, defining a different pattern called Isolation-by-Environment (IBE, Wang & Bradburd, 2014). Likewise, the dispersal limitation caused by the characteristics of physical environment is called Isolation-by-Resistance (IBR, McRae, 2006). Mathematically, both types of models are built regressing (dis)similarity variables between sites over a distance measure separating them. These variables have pairwise dependence as each of their values is obtained by comparing two sites (Diserud & Ødegaard, 2007; Smouse et al., 1986). Thus, any methods used to model these relationships should take this pairwise dependence into account to avoid biased significance tests. The analogies between distance-decay and the genetic-spatial distance relationship allow proposing that the same analytical methods (with the respective adaptations) can be used to study both patterns.

Distance-decay and genetic-spatial distance relationship frequently take linear, power-law or negative exponential shapes (Hutchison & Templeton, 1999, Nekola & McHill, 2014; Gómez-Rodríguez et al., 2018). These models are very useful to study the patterns, although probably do not cover all possible shapes that they can take. The association of these between the shapes and parameters of these relationships and the dispersal ability and ecological traits of different biological groups has been also studied. In general, with low dispersal limitation, the rate of decrease of biological similarity with distance is smaller than with high dispersal limitation, and therefore the slope of the models fitting the relationship is flatter in organisms with good dispersal ability (Qian, 2009; Steinitz et al., 2006). The same occurs with the rate of increase of genetic dissimilarity with spatial distance, which is higher when dispersal limitation is more limited (because gene flow is reduced for separated populations) (Hutchison & Templeton, 1999; Rousset, 1997; Slatkin, 1987). For this reason, an association across taxa between the value of the model parameters (if there are higher or lower) and any surrogate of dispersal limitation, and in general of the potential processes behind the patterns, can be used to infer the relative relevance of different causal processes (Baselga & Gómez-Rodríguez, 2018; Soininen et al., 2007). However, these comparisons are statistically challenging as the pairwise dependence of similarity and distance data can bias the results of significance tests (Legendre, 1993; Smouse et al., 1986).

Most of the biogeographical and macroecological knowledge is based on studies focused on vertebrates and plants (Cardoso et al., 2011). This unbalance causes a taxonomic bias in our knowledge about biogeographical patterns. This, added to other shortfalls in the study of biodiversity (Hortal et al., 2015) as lack of information about the geographic ranges, traits or population dynamics of species, especially affecting invertebrates, stresses the necessity and the interest to deepen the study of the biogeography of invertebrates. Invertebrates, especially arthropods are the most diverse animal group in the world. Only insects have around a million described species (Catalogue of life, 2023), but the estimations compute around 2.6 to 7.2

million species (Stork, 2018). Their high diversity in ecological and dispersal traits (habitats, trophic specialization, thermal tolerances, dispersal ability, etc.) and their wild distribution make them a suitable target group for biogeographical and macroecological studies (Diniz-Filho et al., 2010). In particular, the large variability of traits allows assessing the effects of the different biogeographical and historical processes in shaping the spatial configuration of communities across biological groups with different characteristics (Brousseau et al., 2018). Thus, invertebrates are not only interesting because they have been relatively little studied (Cardoso et al., 2011), but also, and more importantly, because their diversity and variability makes them a good object of study of how current and historical natural mechanisms may cause biodiversity to be distributed as it is.

OBJECTIVES



OBJECTIVES

This thesis aims to develop new methodologies to study the mechanisms driving the spatial distribution of biological diversity as inferred from biogeographical and macroecological patterns. Specifically, it explores (i) the usefulness of a sigmoidal function (Gompertz function) to fit the distance-decay pattern and how its shape change with the species distribution size, (ii) a test statistic for parameters comparison, (iii) the application of these methods to assess the process behind the current large-scale diversity pattern of European spiders and (iv) the extension of these methods to study the genetic-spatial distance relationship using endemic species of Iberian leaf beetles as case-study and the distribution patterns of this group.

The specific objectives of each chapter are:

Chapter I: To explore the relationship between species range size and distance-decay shape and the use of the Gompertz function to model distance-decay patterns.

Chapter II: To develop a robust test statistic (z_{dep}) for distance-decay parameters comparison, not biased by the pairwise dependence of such ecological data.

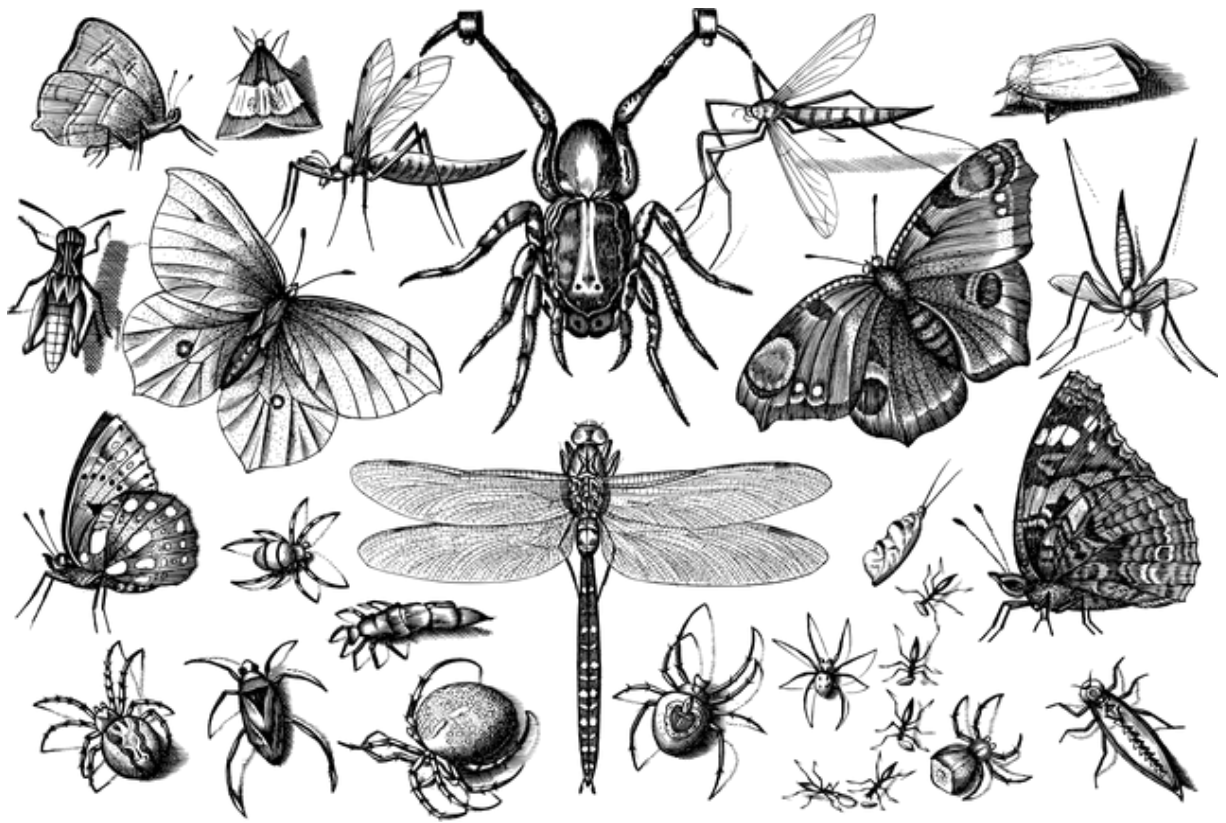
Chapter III: To assess the relevance of spatial and environmental factors shaping the large-scale spatial structure of European spider diversity, applying the methods previously developed.

Chapter IV: To assess the use of the Gompertz function and the z_{dep} statistic for the study of the genetic-spatial distance relationship using Iberian beetle species as target group.

Chapter V: To study the causes of the disjoint distribution of the *Lachnaia tristigma* (Coleoptera: Chrysomelidae) species-group.

Chapter I

SPECIES RANGE SIZE SHAPES DISTANCE-DECAY IN COMMUNITY SIMILARITY



SPECIES RANGE SIZE SHAPES DISTANCE-DECAY IN COMMUNITY SIMILARITY

ABSTRACT

Aim: (i) To assess the dependence between the form of the decrease in biological similarity with distance (distance-decay) and species range size and (ii) to introduce the use of a sigmoidal model, the Gompertz function, as a flexible alternative able to fit distance-decay models under a wide variety of species range sizes.

Location: Applicable worldwide.

Methods: We computed distance-decay curves from simulated communities to assess how the species range sizes shape the functional form of distance-decay patterns (i.e. negative exponential, power-law or sigmoidal [Gompertz] relationships). Simulations were performed using different sample sizes and species detection probabilities. We also used distribution data of South American mammals to explore the relationship between species range size and the distance-decay form in an empirical dataset.

Results: Our simulations showed that the power-law is the best supported model when range sizes tend to be small. An increase in range sizes leads to a negative exponential relationship, taking the shape of a sigmoidal (Gompertz) relationship with the largest range size values. Similar results have been found in the distance-decay pattern of South American mammals. Remarkably, the Gompertz function fits the data reasonably well in all scenarios.

Main conclusions: The functional form of distance-decay patterns depends on a key biogeographical attribute: species range size. This dependence makes it an interesting tool to detect biodiversity threats associated with species range expansion, such as the biotic homogenization of faunas. The Gompertz function is the mathematical model that best accommodates different frequency distributions of species range size and, thus, allows cross-taxa comparison of this biogeographical and ecological pattern.

1.1 INTRODUCTION

The decrease in community similarity with spatial or environmental distance, that is, distance-decay of similarity, is a general property of biological systems that has been studied for a large range of organisms, for example, archaea (Barreto et al., 2014), bacteria (Barreto et al., 2014; Milici et al., 2016), diatoms (Astorga et al., 2012; Wetzel et al., 2012), fungi (Bahram et al.,

2013), plants (König et al., 2017; La Sorte et al., 2008; Nekola & White, 1999; Qian, 2009) and both invertebrate (Saito et al., 2015; Thieltges et al., 2009) and vertebrate animals (Maloney & Munguia, 2011; Qian & Ricklefs, 2012); in different environments, from neotropical forests (Palmer, 2005) to urban environments (La Sorte et al., 2008); and at different spatial scales (Steinbauer et al., 2012). At local or regional scales, distance-decay curves are primarily shaped by the organisms' biological characteristics (e.g., niche requirements or dispersal ability) and their interaction with the environment (Morlon et al., 2008; Steinbauer et al., 2012). The biogeographic characteristics of the study area (such as spatial scale or latitude) also impact the patterns of distance-decay (Nekola & McGill, 2014; Nekola & White, 1999; Soininen et al., 2007). However, the links between distance-decay and key macroecological attributes, such as species range size, have received less attention, even though range size determines to which extent a given species can be found in two different sites. Moreover, species range size also depends on dispersal and niche processes (Brown et al., 1996; Willis, 1922). Thus, organisms with high vagility and wide niches would tend to have larger distributions. In turn, larger distributions increase the probability of different sites having the same species, thus leading to higher similarity in community composition. On the contrary, the distributions of organisms with low vagility and narrow niches are generally smaller and, thus, higher turnover in community composition is expected for these taxa. Therefore, we predict that species range size is a crucial link between the processes controlling species distributions and the shape of the relationship between community similarity and spatial (or environmental) distance.

The assessment of spatial turnover in community composition and, more precisely, its dependence on spatial/environmental distance relies on adequately modelling the shape of the relationship between community similarity and spatial/environmental distance. Non-linear regressions of community similarity (e.g., negative exponential or power-law models) are usually employed, and their parameters are interpreted in biological terms. For example, the intercept estimates the expected community similarity at short distances and the slope quantifies the rate at which communities change with distance (Nekola & White, 1999; Soininen et al., 2007). Therefore, when the same function is used to model distance-decay patterns of different biological groups, these statistical parameters provide a powerful tool to assess how the biological characteristics of different organisms (i.e. dispersal ability or ecological niche) impact the spatial turnover of biological communities (e.g., Gómez-Rodríguez & Baselga, 2018; Soininen et al., 2007). The most frequently used functions are the negative exponential (Nekola & McGill, 2014; Nekola & White, 1999; Preston, 1962; R. H. Whittaker, 1960) and the power-law (Nekola & McGill, 2014). It is not possible to opt for one of these mathematical models on first principles as we lack a theoretical foundation for the form of the relationship (Nekola & McGill, 2014) or how it may be affected by species' attributes, such as range size. For instance, when species have small spatial ranges (relative to the study extent), nearby sites are expected to have very different species composition, causing the similarity to rapidly decay from the shortest distances. In this case, power-law or negative exponential functions would suffice to model distance-decay curves. However, when species have large spatial ranges relative to the study extent, close-by biological communities may be very similar, resulting in a small or null rate of change in community composition at the shortest distances while the

actual decrease in community similarity would start at intermediate distances. In consequence, the similarity among these communities would present a sigmoidal relationship with distance, being more or less constant (and close to 1) at short distances, rather than steeply decaying from the shortest distance. This shape in the relationship between biological similarity and spatial distance cannot be, a priori, captured by negative exponential or power-law models (Fig. 1.1). In other words, when only these two functional forms are assessed, one of them will be identified as the best supported model, but it may still be an inadequate representation of the distance-decay pattern.

To accommodate scenarios in which the distance-decay relationship shows an initial plateau of high similarity, we introduce the use of the Gompertz function (Gompertz, 1825), which is a sigmoidal function (Godeau et al., 2020) that generalizes the negative exponential model,

$$f(x) = ae^{-be^{-cx}}$$

where a is the superior asymptote, b a position parameter, c the rate of change and x a distance metric between sampling sites. This function was designed for growth patterns and is widely used in many fields of biology, like microbiology, cytology or vegetal physiology (Baker et al., 1975; Booth, 1984; Çelekli et al., 2008; Johnsen et al., 2013; Rossi et al., 2003). Brownstein et al. (2012) also used it to fit the increase in biological dissimilarity with spatial distance (i.e. the inverse of a distance-decay relationship), in order to use its intercept as a proxy for randomness in community configuration. One of the main advantages of the Gompertz function is its theoretical versatility because, as a generalization of the negative exponential, it is expected to adequately fit curves with and without an initial plateau. If this is confirmed, distance-decay patterns parameterized with a Gompertz function can be compared across multiple taxa with varied distance-decay shapes. Such cross-taxon comparison is the basis for the identification of species traits related to distance-decay patterns and, hence, the inference of the main processes driving the spatial variation of those biological communities.

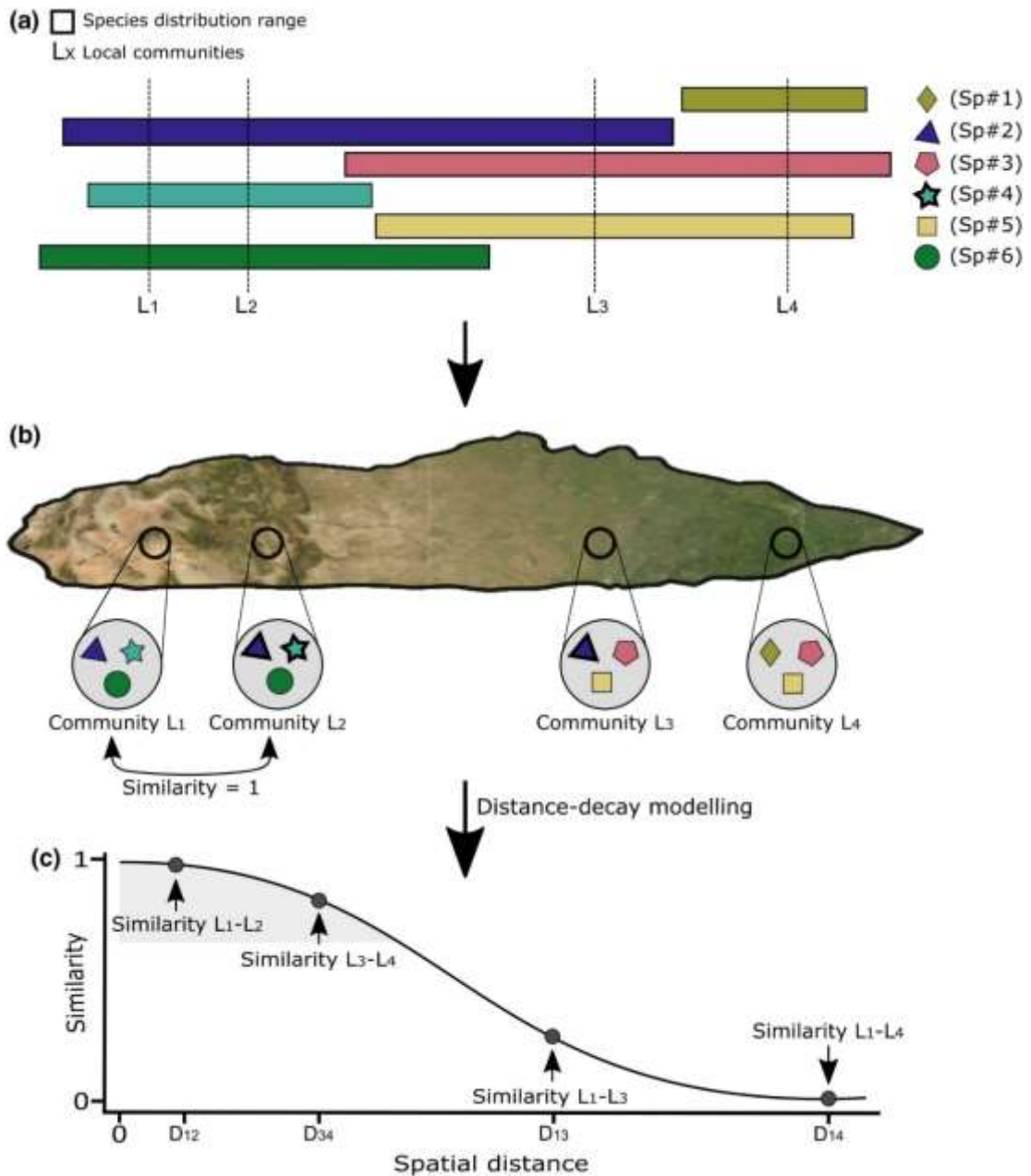


Figure 1.1. Hypothetical example showing how species ranges in a landscape (a) may lead to a spatial configuration of biological communities (b) with a sigmoidal distance-decay relationship (c).

The aim of this paper is to evaluate how species range size affects the distance-decay functional form and which function (i.e. negative exponential, power-law or Gompertz) can better capture the distance-decay relationship under different range size scenarios. We expect the Gompertz model to provide a better fit when all or most species ranges are large in relation to the distance between local communities. To do this, we used simulations to assess the relationship between the functional form of distance-decay patterns and (i) the frequency distribution of species range size (i.e. proportion of spatially restricted vs. widespread species), and (ii) the size of the smallest ranges. We also used data on the distribution of South American

mammals as a case study of how range size influences the form of distance-decay patterns in empirical datasets.

1.2 METHODS

1.2.1 Data simulation

To assess whether species range size affects the functional form of the distance-decay relationship, we compared the fit of the negative exponential, power-law and Gompertz functions in different scenarios of varying (i) proportion of spatially restricted vs. wide-spread species, and (ii) range size of spatially restricted species. We simulated the distribution of 300 species with different range sizes (depending on each scenario) in a virtual landscape and computed the compositional similarities between randomly allocated sampling sites, which correspond to our “virtual communities.” To do so, we first created 300 circular spatial polygons, each representing the range of a species, using the R package “sp” (R. S. Bivand et al., 2013; Pebesma & Bivand, 2005). Each polygon had a different diameter, which was drawn from the frequency distribution of species range size in each simulation scenario. The polygon centre was located at a random point in a virtual landscape of 3000 x 3000 cells. Finally, n sampling sites were randomly allocated in the virtual landscape using the *grf* function of the package “geoR” (Ribeiro and Diggle, 2018). The species composition in each site was computed by identifying the species’ polygons that overlap on that site, using the function *gIntersects* of the package “rgeos” (R. Bivand & Rundel, 2019). This procedure was repeated for different number of sites ($n = 10, 50, 500, 5000$) to assess the effect of sample size on the functional form of distance-decay patterns. Therefore, we obtained four presence/absence matrices of 300 simulated species in $n = 10, 50, 500$ and 5000 sites, respectively. We also assessed the effect of absences within a species’ potential range (e.g., due to imperfect species detection or unsuitable conditions) on the functional form of distance-decay patterns. To do so, the presence/absence of species within its potential range was assigned using a Bernoulli trial with four different probabilities of success ($p = 0.25, p = 0.50$ or $p = 0.75$, representing different degrees of incomplete species detection, and $p = 1.00$, representing perfect detection).

The differences in species range sizes among simulation scenarios were set following two axes of variation: (i) the ratio between spatially restricted and widespread species (ratio-classes 1–3) and (ii) the range size of spatially restricted species (size-classes A–C) (Fig. 1.2). The combination of these two axes of variation leads to nine simulation scenarios. For the first axis of variation, we simulated three ratio-classes with an increasing proportion of widespread species (diameter of widespread species drawn from a uniform distribution $U [3000, 5000]$): ratio-class 1 (10 species with widespread distributions and 290 with spatially restricted distributions); ratio- class 2 (20 widespread species and 280 spatially restricted species); and ratio-class 3 (200 widespread species and 100 spatially restricted species). For the second axis of variation, we simulated spatially restricted species with increasingly larger ranges. Hence, in each size-class the polygon diameters were drawn from different distributions: size-class A with $U [500, 1000]$; size-class B with $U [1000, 1500]$ and size-class C with $U [2000, 2500]$.

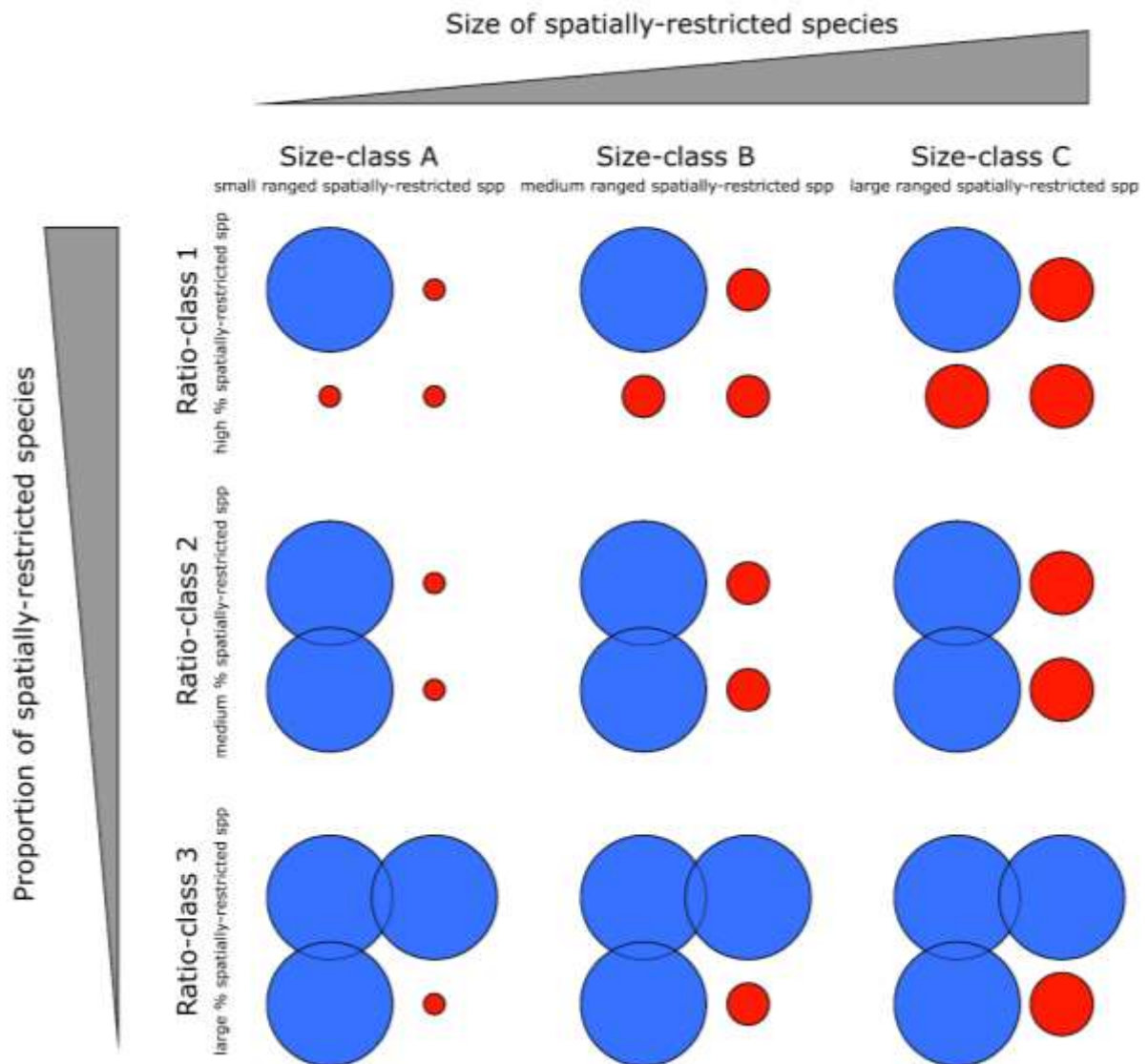


Figure 1.2. Schematic representation of the two axes of variation in simulation scenarios: the proportion of spatially restricted species (ratio-classes 1-3) and the range size of spatially restricted species (size-classes A-C).

1.2.2 Distance-decay models of simulated data

We performed non-linear regressions of pairwise assemblage similarity against spatial distance to fit distance-decay models using three different functions: negative exponential, power-law and Gompertz. We computed the similarity in species composition between pairs of sites with the Simpson pairwise similarity index, $1-\beta\text{sim}$ (Baselga, 2010; Koleff et al., 2003), using the *beta.pair* function of the package “betapart” (Baselga & Orme, 2012). Spatial distances were calculated as the Euclidean distance between pairs of sampling sites. The negative exponential, power-law and Gompertz functions were fitted using the *nlsLM* function of the package “minpack.lm” (Elzhov et al., 2016), which performs a non-linear least squares regression using the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). We

empirically compared the fit of each model based on two independent criteria: AIC and pseudo- r^2 value. For algebraic details on the Gompertz function, please see Appendix 1.1. The code for the simulation of virtual communities and the assessment of distance-decay models is provided in Appendix 1.2.

1.2.3 Real data

To study the relationship between species' range size and the functional form of distance-decay models in empirical data, we used distribution data of South American terrestrial mammals from the IUCN red list (*IUCN Red List of Threatened Species*). We created four subsets of data based on quartiles of species range size (Q1, Q2, Q3 and Q4). Thus, the species with the smallest distribution ranges were included in Q1 while the ones with the largest distribution ranges were included in Q4. For each species, we computed range size as the total area it occupies with the *area* function of the package “raster” (R. J. Hijmans, 2021).

To obtain a presence/absence matrix for each subset, we constructed a $1^\circ \times 1^\circ$ cell grid over South America using the *spsample* function of the package “sp” to define the cells centroids and the function *SpatialPixelsDataFrame* of the same package to obtain the final grid. Then, we computed the species present in each cell with the *gIntersects* function. This function does not work with spatial pixels, so we converted the grid into spatial polygons with the function *Grid2Polygons* of the package “inlmisc” (Fisher, 2020). The $1^\circ \times 1^\circ$ cells with less than five species were excluded from further analyses. Presence/absence tables were used to compute independent distance-decay curves for each quartile of range sizes (Q1, Q2, Q3 and Q4). We used the same procedure as in the simulated data, except for the calculation of spatial distances, which were here computed as the geodesic distance between cell centroids using the *geodist* function of the package “geodist” (Padgham & Sumner, 2020). We finally assessed the best fitting model (power-law, negative exponential or Gompertz) based on the AIC and pseudo- r^2 value, as in the simulation study. To evaluate whether the results are robust to sample size variation, for each quartile dataset (Q1, Q2, Q3 and Q4), we repeated the distance-decay analyses for 100 resamples of 50 sites each.

1.3 RESULTS

We used simulated community data to study how the functional form of the distance-decay curve varies with species range sizes. For the smallest range size and the lowest proportion of widespread species (ratio-class 1 and size-class A, hereinafter scenario 1A), the power-law function presented a better fit (Table 1.1), in a pattern characterized by the rapid decay of initial similarities (results for $n = 50$ and $p = 1$ are shown in Fig. 1.3). An increase in range size and/or the proportion of widespread species (scenarios 1B, 2A and 2B) resulted in biological similarity still decaying from the shortest distances, but with a less steep shape. In these scenarios, the negative exponential function better fitted the data. For the largest range sizes and the highest proportion of widespread species (all scenarios involving either ratio-class 3 or size-class C),

the similarity remained constant at short or intermediate distances, approaching a sigmoidal shape. In these scenarios, the Gompertz function was the one that presented a better fit (Fig. 1.3, Table 1.1). Remarkably, when negative exponential or power-law functions fitted the data best, the performance of the Gompertz function was still relatively good (i.e. with pseudo- r^2 values close to the best fitting model). Sample size and species detection probability had small effects on the relationship between species range size and the best fitting model of the distance-decay curve. Only with a very low sample size ($n = 10$) and/or very low detection probability ($p = 0.25$), the best fitting model may be different to the one found with larger sample sizes and detection probabilities (see Tables S1.1–S1.4 in Appendix 1.3). Importantly, inconsistencies in the identification of the best functional form only occurred when the fit of distance-decay models was very low (i.e. pseudo- $r^2 < 0.1$, Table S1.4 in Appendix 1.3).

Table 1.1. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations of different frequency distribution of species range size, from a higher proportion of spatially restricted species (ratio-class 1) to a smaller proportion (ratio-class 3); and different range size of spatially restricted species, from smaller ranges of spatially restricted species (size-class A) to larger ranges of spatially restricted species (size-class C). Best fitting models are shown in bold.¹

		Size-class A small ranged spatially-restricted species		Size-class B medium ranged spatially-restricted species		Size-class C large ranged spatially- restricted species	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1 high % spatially- restricted species	Power-law	0.67	-2077	0.65	-1339	0.54	-161
	Negative exponential	0.66	-2040	0.88	-2667	0.87	-1697
	Gompertz	0.62	-1912	0.86	-2466	0.94	-2747
Ratio-class 2 medium % spatially- restricted species	Power-law	0.67	-2185	0.72	-1862	0.56	-401
	Negative exponential	0.76	-2547	0.89	-2959	0.88	-2020
	Gompertz	0.74	-2473	0.85	-2598	0.95	-3029
Ratio-class 3 low % spatially- restricted species	Power-law	0.52	-2218	0.57	-2328	0.59	-2016
	Negative exponential	0.78	-3160	0.82	-3384	0.86	-3318
	Gompertz	0.86	-3703	0.87	-3788	0.90	-3774

¹ Results are here shown for simulations of 50 sampling sites ($n = 50$) and perfect detection probability of species ($p = 1$). See Tables S1.1 to S1.4 in Appendix 1.3 for results with different sampling sizes and detection probabilities.

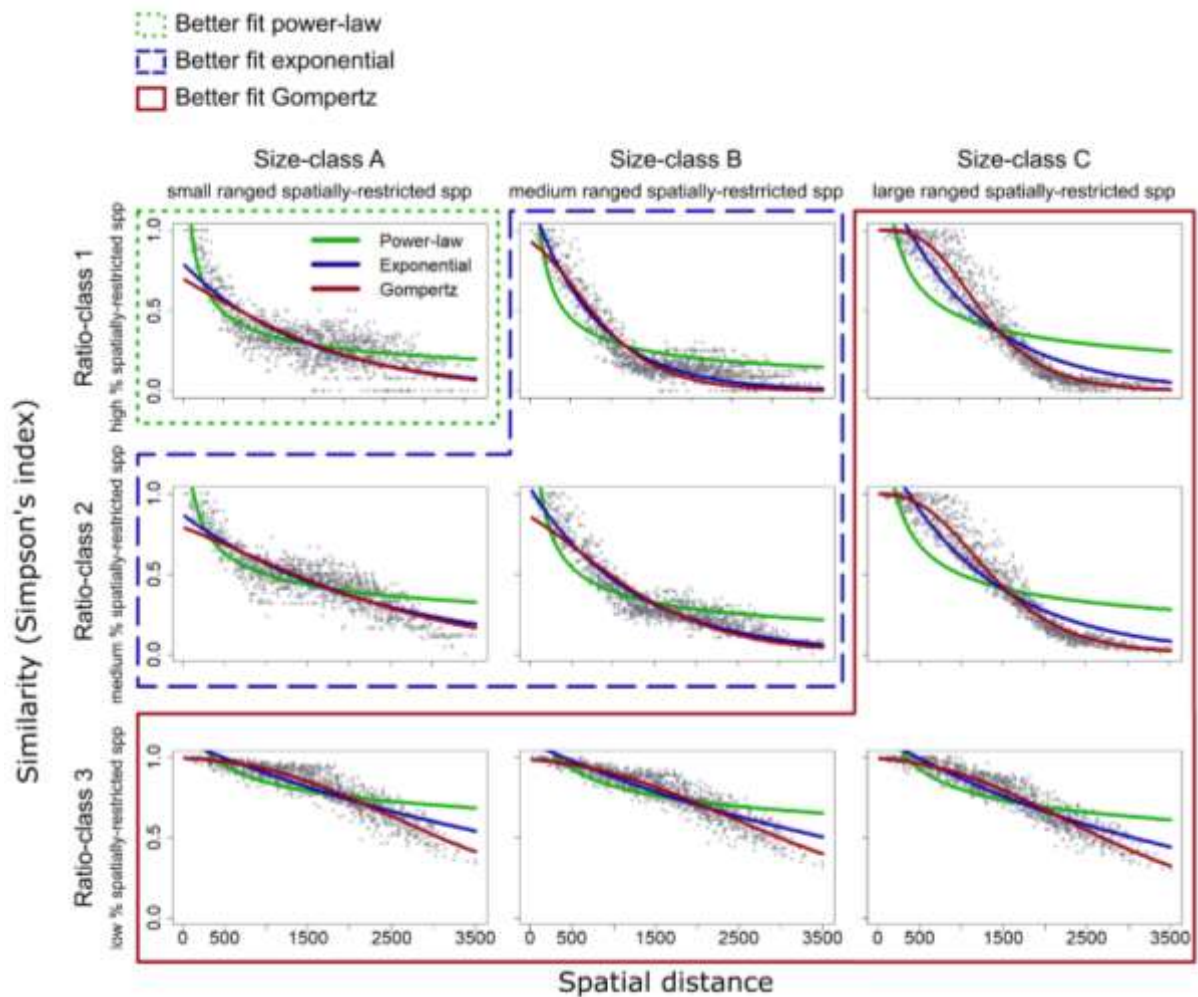


Figure 1.3. Distance-decay models of simulated data under different range size scenarios. Simulated scenarios differ in the proportion of spatially restricted species (ratio-classes 1 to 3) and in the average range size of the smallest size class (size-classes A to C). The adjustment of three different functions is shown: negative exponential, power-law and Gompertz. The model with the best fit for each scenario is also indicated.

Distance-decay patterns in South American terrestrial mammals showed similar shifts in the shape of the distance-decay relationship across the range size classes (Q1-Q4). The negative exponential was the best fitting model for the datasets including the species with smaller range sizes (Q1 and Q2), while the Gompertz function was the best fitting model for the datasets including the species with larger range sizes (Q3 and Q4) (Fig. 1.4, Table 1.2). Similar results were found when the sample size was smaller, as in the analyses of 100 random samples of 50 sites in each dataset. The negative exponential was the best fitting model in 89 and 93 of the 100 samples taken for Q1 and Q2, respectively, while the Gompertz function was the best fitting model in 76 and 68 of the 100 samples taken for Q3 and Q4, respectively.

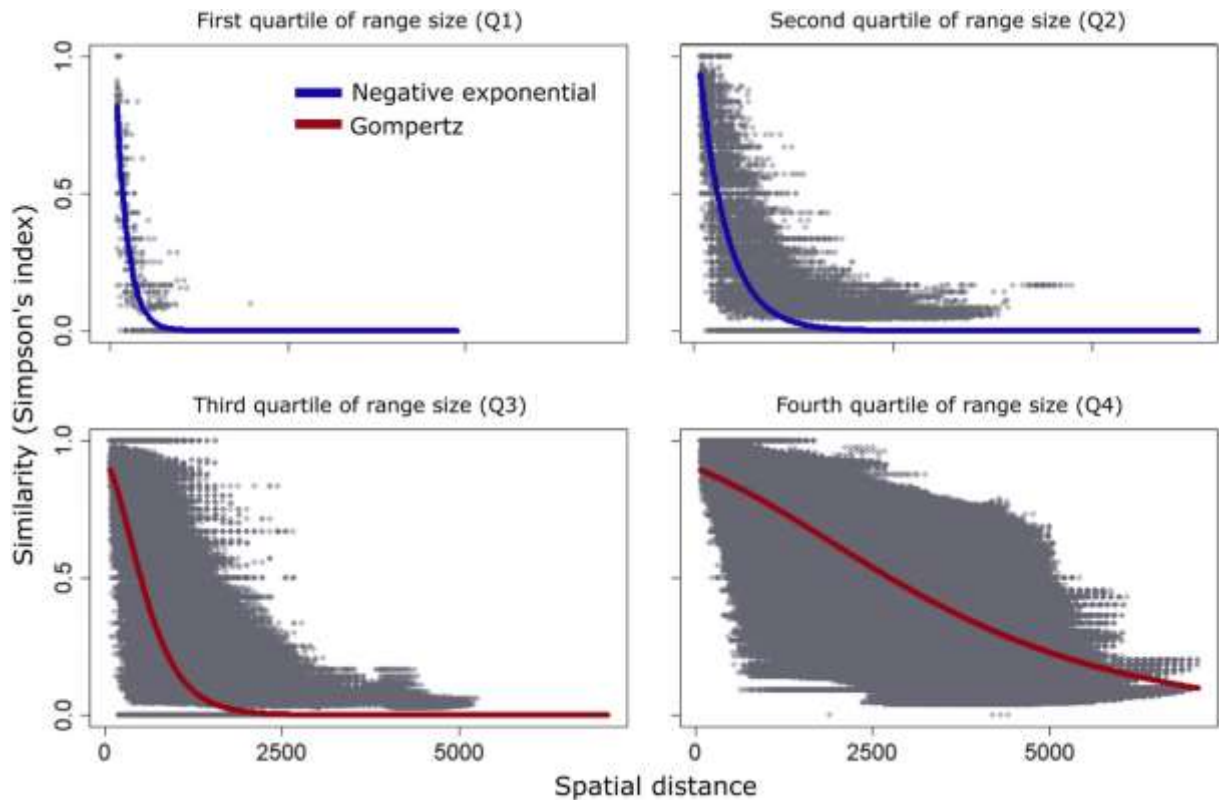


Figure 1.4. Distance-decay models for the four range size groups (quartiles 1 to 4, Q1-Q4) of South American mammals. Note the shift in the shape of the distance-decay relationship from the smallest range size subsets, best fitted by the negative exponential function (Q1-Q2), to the largest size subsets, best fitted by the Gompertz function (Q3-Q4).

Table 1.2. Pseudo-r² and AIC of the power-law, negative exponential and Gompertz models of the distance-decay pattern of South American terrestrial mammals. Best fitting models are shown in bold.

	Power-law		Negative exponential		Gompertz	
	Pseudo-r ²	AIC	Pseudo-r ²	AIC	Pseudo-r ²	AIC
First quartile of range size (Q1)	0.77	-3775	0.80	-3995	0.79	-3971
Second quartile of range size (Q2)	0.65	-84451	0.71	-94365	0.70	-93148
Third quartile of range size (Q3)	0.56	-1389354	0.71	-1792540	0.71	-1801573
Fourth quartile of range size (Q4)	0.41	-560855	0.52	-807862	0.53	-822159

1.4 DISCUSSION

Our results show that a key biogeographical attribute, species range size, drives the functional form of distance-decay patterns. More importantly, when range sizes are large compared with the study extent, distance-decay can take a sigmoidal form that cannot be accurately captured by the most frequently used models: the negative exponential and the power-law. This relationship between the functional form of distance-decay models and the species range size is observed both in simulated and empirical data. The simulated data show that when species ranges are comparatively small, even biological communities at close distance have different

species, leading to a rapid decay of biological similarity from the shortest distances in the form of a power-law relationship. When species ranges are comparatively larger, the initial decay is still present but the decay in similarity is not so steep, being constant and proportional across distances, with the negative exponential being the model that better fits the data. With even larger species ranges, no decay occurs at the shortest distances, with community similarity remaining high and almost constant at short and intermediate distances (i.e. forming an initial plateau of high similarity). This sigmoidal relationship is observed when the distribution range of most species is larger than the spatial distances separating the closest sampling sites. In this scenario, the Gompertz function outperforms the negative exponential and power-law models due to its ability to adapt to this plateau of constant similarity. More importantly, even when species ranges are small and no initial plateau exists, the Gompertz function performs relatively well in all scenarios, showing a performance (in terms of pseudo- r^2) close to the model with the best fit. The results obtained from simulated data were in agreement with those based on an empirical dataset, the distribution data of South American mammals. The negative exponential models showed a better fit of distance-decay patterns in species with smaller ranges (Q1 and Q2), while the Gompertz function was the best fitting model for more widespread species (Q3 and Q4).

Our results reveal that the size of the species ranges (relative to the distances between sampling sites) plays an important role in shaping distance-decay patterns. This inference seems robust to variations in sample size and detection probability. While the most widely used functions in distance-decay analyses (i.e. negative exponential and power-law) work well when species have relatively small ranges, their performance is worse when the distance-decay relationship takes a sigmoidal form associated with the existence of species with widespread distributions. On the contrary, the Gompertz function can accommodate such sigmoidal relationship. More importantly, as a generalization of the negative exponential function, the Gompertz is a versatile function that performs relatively well in all simulation scenarios, independently of the species range size. This is particularly relevant when the goal is to analyse the variation in distance-decay parameters across multiple taxa or biological systems (as in Baselga & Gómez-Rodríguez, 2021; Gómez-Rodríguez & Baselga, 2018; Soininen et al., 2007). In this type of studies, the shape of distance-decay might differ across taxa or systems, so having a general function that performs well under multiple scenarios, as the Gompertz function does, is of crucial importance.

The Gompertz function informs about two important aspects of diversity patterns: species turnover at the shortest distances (position parameter) and the rate of change in community composition (slope). Contrary to other distance-decay models, in which the intercept is a key parameter that informs about similarity at short distances (Morlon et al., 2008; Qian & Ricklefs, 2012; Soininen & Hillebrand, 2007), in the Gompertz model both the extent of the initial plateau and the intercept mathematically depend on the position parameter (b , see Appendix 1.1). The higher the value of the position parameter, the longer the distance at which the initial similarity remains constant before it starts decaying. Thus, the position parameter of this sigmoidal distance-decay pattern informs about the existence and magnitude of turnover at short distances.

Moreover, the ability of the Gompertz function to fit the initial plateau of constant similarity, if it exists, has also important implications in the estimation of the rate at which similarity decays with distance (i.e. the model's slope). It must be highlighted that the slope in a Gompertz model is estimated only over the spatial distances at which similarity actually decays with distance, because the data points in the initial plateau do not bias the estimation of the slope. On the contrary, in convex functions such as the exponential and power-law models, the complete range of spatial distance has an influence on the slope. In other words, the existence of an initial plateau would affect the estimation of the slope, biasing the estimation of the rate at which similarity decays with distance. Thus, the Gompertz model informs about similarity at short distances (i.e. the intercept); how far the initial similarity is maintained, and hence if species turnover occurs at short-intermediate distances (i.e. position parameter); and the rate of similarity decay with the increase in spatial separation among sites (i.e. slope). Additionally, the Gompertz function is bounded between 0 and 1 (see Appendix 1.1), and thus is in accordance with the bounds of biological similarities (also 0 and 1), contrary to the negative exponential and power-law functions, which can take values higher than one at short distances, a result lacking biological meaning.

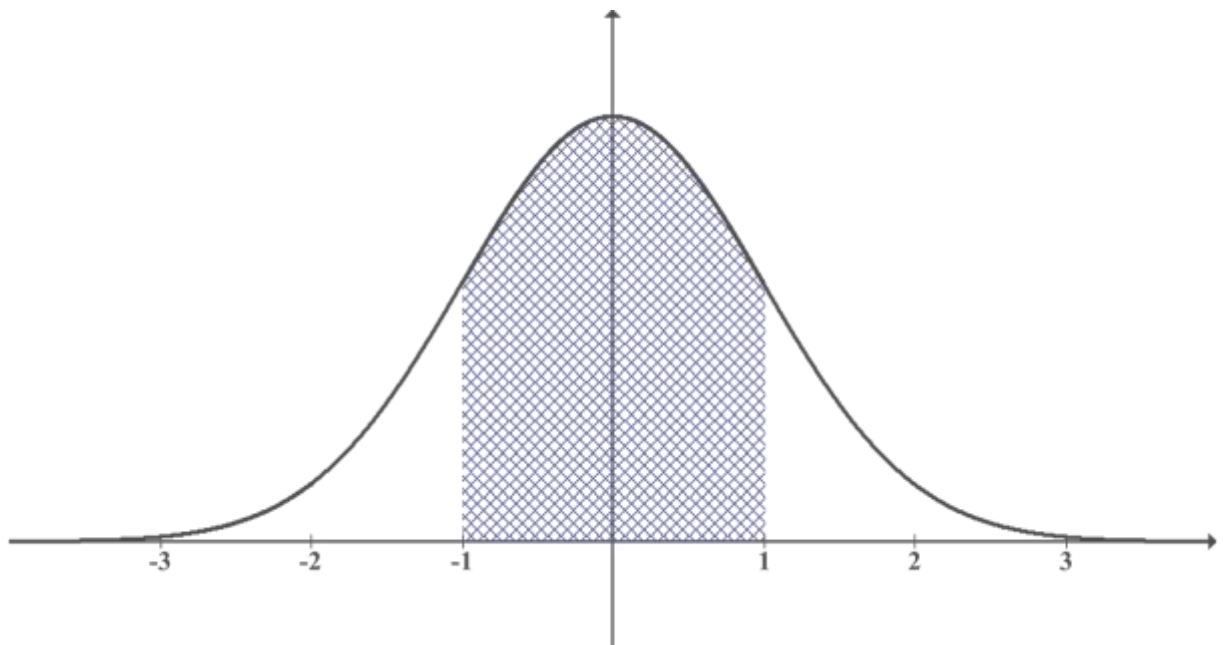
Understanding the link between species range size and how community similarity decreases with spatial distance is relevant for conservation biogeography. Biotic homogenization, that is, “the replacement of local biotas with non-indigenous species, usually introduced by humans” (McKinney & Lockwood, 1999), is one of the major threats for biodiversity (Baiser et al., 2012; Olden et al., 2004; Yang et al., 2021). Biotic homogenization is usually linked to human activities that lead to range expansions of invasive species and range contraction of specialist and endemic species (Clavel et al., 2011; Olden & Poff, 2003). Our results on the relationship between the shape of distance-decay patterns and range size are relevant in the context of biotic homogenization in two ways. First, from a theoretical standpoint, understanding the links between species range size and distance-decay patterns is crucial, because biotic homogenization (i.e. the increase in spatial community similarity over time) is the result of range expansion of some generalist species. Second, from a practical standpoint, accurately measuring the shape of distance-decay patterns and assessing changes in their shapes (from convex to sigmoid functions) can help detecting processes of biotic homogenization and allow quantifying its strength.

In conclusion, species range size is here revealed as a key biogeographical attribute that shapes distance-decay patterns. Previous studies had evidenced the relationship between distance-decay patterns and species biological attributes (Gómez-Rodríguez & Baselga, 2018; Soininen et al., 2017) or the spatial scale of the study (Nekola & McGill, 2014). Our results show that species range size is also a crucial biogeographical driver of the functional form of the distance-decay relationship. From the smallest range sizes to the largest ones, the shape of the distance-decay curve varies from a power-law, negative exponential to a sigmoidal form, which can be accurately fitted with a Gompertz function. Importantly, the Gompertz function provides a general good fit in all situations and, therefore, can be considered a general model

to study distance-decay patterns and the best option to model, study and compare distance-decay patterns across different taxa or biological systems.

Chapter II

COMPARING DISTANCE-DECAY PARAMETERS: A NOVEL TEST UNDER PAIRWISE DEPENDENCE



COMPARING DISTANCE-DECAY PARAMETERS: A NOVEL TEST UNDER PAIRWISE DEPENDENCE

ABSTRACT

Distance-decay models fit parametric functions to assess the relationship between similarity and spatial or environmental distance. Despite the widespread use of distance-decay models in ecology and biogeography, no method has been previously developed and validated to assess the significance of differences between the parameters (i.e. intercept and slope) of two distance-decay models. The pairwise autocorrelation of similarity and spatial distance affects the variance of parameter estimates, precluding the use of ordinary t-tests. Here, we provide a test statistic (z_{dep}) for the equality of parameters between two distance-decay models. The test can be applied, independently, to any of the model parameters (i.e. intercept and/or slopes). The z_{dep} statistic accounts for pairwise dependence, thus avoiding biases associated to the inflation of degrees of freedom, and it is based on the estimation of parameters' variance using site-block resampling. To validate the z_{dep} statistic, type I and type II errors were empirically evaluated through a simulation study. We simulated six scenarios (three under the null and three under the alternative hypothesis) of distance-decay relationships using different functions: negative exponential, power-law or Gompertz function. We applied the z_{dep} statistic and computed the proportion of rejections of the null hypothesis for $\alpha = 0.01, 0.05$ and 0.1 in each scenario. As a case-study, we also compared distance-decay parameters across several groups of Iberian vertebrates (cyprinids, frogs, lizards and snakes, bats, rodents and carnivores). In the simulation study, the z_{dep} statistic showed a good approximation of the nominal level (α , type I error) and a good statistical power ($1 - \text{type II error}$), the later increasing with sample size, as expected. In Iberian vertebrates, we found significant differences between ectotherms and endotherms, but not within these groups except between cyprinids and other ectotherms. The good performance of the z_{dep} statistic makes it the best option to test for differences in parameters obtained from models fitted from data with pairwise dependence, as distance-decay models. It can also be used beyond distance-decay approaches to compare parameters of any other regression models of pairwise dependent data (as genetic distances, for example).

2.1 INTRODUCTION

The distance-decay of similarity (i.e. the decrease of community similarity with spatial/environmental distance) is a ubiquitous macroecological pattern (Graco-Roza et al., 2022; Nekola & White, 1999; Soininen et al., 2007) that results from the interplay between

environmental attributes of the study area and the organisms' dispersal ability and ecological requirements (Morlon et al., 2008; Steinbauer et al., 2012). Distance-decay models are a common approach to analyse the spatial variation of community composition (β -diversity), by assessing the relationship between two pairwise matrices: community similarity and spatial distance. In fact, the shape and slope of such distance-decay relationship can be used to infer the strength of the processes driving the spatial structure of biodiversity patterns, such as environmental filtering and dispersal limitation (Baselga & Gómez-Rodríguez, 2021; Gómez-Rodríguez & Baselga, 2018; Soininen et al., 2007). Moreover, a comparison of distance-decay model parameters (i.e. intercept and slope) across biological groups has proven effective to evidence differences in the role of dispersal limitation within the same geographical context (e.g., Gómez-Rodríguez & Baselga, 2018). However, such comparison of distance-decay parameters is not straightforward because the variance of parameter estimates is affected by the fact that both variables, i.e. community similarity and spatial distance, are pairwise autocorrelated (i.e. similarity and distance values are the result of the comparison of two biological communities), and hence their values cannot be considered independent (Smouse et al., 1986). While the bias introduced by pairwise autocorrelation has been accounted for in some significance tests (e.g., Mantel test), a statistical test for parameter comparison, equivalent to the t-test, has not been developed in a multivariate context.

Various methods have been developed to study the relationship between community similarity and spatial or environmental distances, including semi-variograms (Cressie, 1993), linear regressions (Nekola & White, 1999), or generalised dissimilarity models based on spline regressions (Ferrier et al., 2002, 2007). However, the most widely used methods are (i) the Mantel and partial Mantel tests (Mantel, 1967; Mantel & Valand, 1970; Smouse et al., 1986), which compute the correlation between two distance (or dissimilarity) matrices, and (ii) non-linear regressions, which fit parametric functions, such as the power-law or the negative exponential, to distance-decay data (Nekola & McGill, 2014). Mantel tests have been specifically developed to assess the correlation between distance and/or dissimilarity matrices, and hence are not biased by the structure of distance data. However, its lack of parameters precludes the ecological interpretation of the relationship. On the contrary, in regression-based distance-decay models, the intercept can be interpreted as the expected community similarity at short distances while the slope is the rate at which communities change with distance (Soininen et al., 2007). Therefore, to compare the parameters of two distance-decay models, we need a statistical test that accounts for the pairwise autocorrelation of distance-decay data.

Pairwise autocorrelation arises because both community similarity and spatial or environmental distances are computed by comparing the data observed in two different sites. Thus, each value in a similarity/distance matrix involves two observations (i.e. sites) and each of these observations participates in the computation of more than one of the similarity/distance values (Diserud & Ødegaard, 2007; Smouse et al., 1986). In other words, from N sites we compute $(N \cdot N - 1)/2$ pairwise similarities, so the number of similarity values is larger than the number of independent community observations. This causes the inflation of degrees of freedom (Koenig, 1999) because one value in the similarity matrix does not represent a complete degree of freedom (Dale & Fortin, 2002; Legendre, 1993) (Fig. 2.1). The inflation of

degrees of freedom biases ordinary significance tests, as the F-test, but this bias has been accounted for in tests specifically developed for similarity/distance data. For example, the Mantel test uses permutations to compute the null distribution of the correlation between similarity and distance, thus taking into account pairwise dependence (Koenig, 1999). Beyond correlation, a significance test based on the deviance of non-linear regression models has been recently developed, which uses permutations or site-block resampling to deal with pairwise autocorrelation (Martínez-Santalla et al., 2022). Moreover, and in order to compare the parameters of two distance-decay models, different alternatives have been also used based on permutations or bootstrap (Gómez-Rodríguez & Baselga, 2018; Nekola & White, 1999). However, previous solutions were suboptimal because the resampling procedures did not account for the pairwise dependence of the data. From a practical standpoint, the inflation of degrees of freedom biases the comparison of regression parameters by increasing the rate of false positives (e.g., reporting significant differences in distance-decay parameters when they are not different, inflating Type I error) (Legendre, 1993; Legendre & Fortin, 1989). To overcome this problem, we have designed a specific test to compare the parameters of distance-decay models while controlling for pairwise autocorrelation.

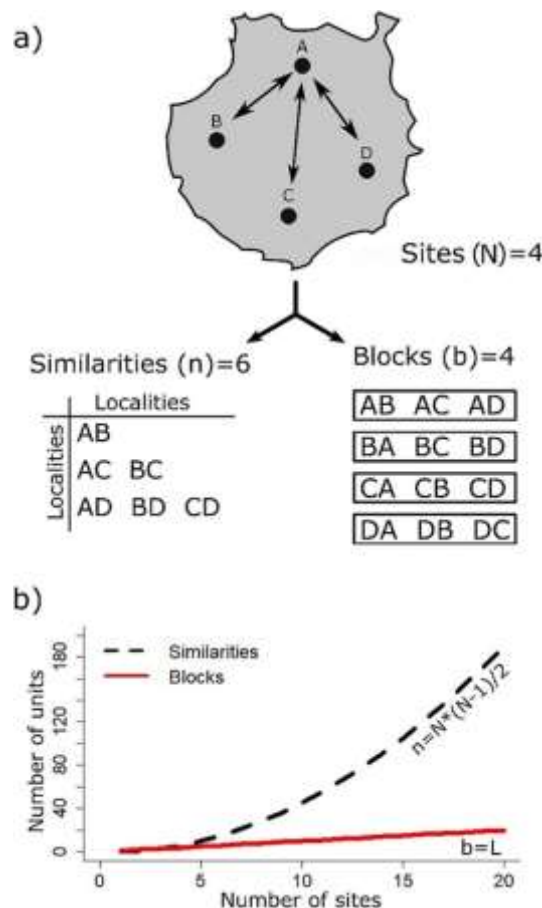


Figure 2.1. a) Example showing how pairwise dependence of similarity/distance values results in pseudoreplication and hence the inflation of degrees of freedom. The effective size of the data ($N = 4$) is less than the number of values in the similarity/distance matrix ($n = 6$). Alternatively, in the site-block construction, the effective size of the data ($N = 4$) and the number or similarity site-blocks ($b = 4$) are the same, thus avoiding the inflation of degrees of freedom, while keeping the pairwise structure of data. b) Increase with sample size of the disparity between the number of sites or site-blocks and the number of similarities/distance values.

Here we introduce and evaluate a novel statistic to independently test for the equality of individual parameters of distance-decay models while accounting for the pairwise dependence of the data. This test statistic (z_{dep}) is based on the standardised difference of the estimated parameters, analogous to a t-test in the sense that the difference between the estimated parameters is divided by the corresponding standard deviation,

$$z_{dep} = \frac{\widehat{\theta}_1 - \widehat{\theta}_2}{\sqrt{\text{var}(\widehat{\theta}_1) + \text{var}(\widehat{\theta}_2) - \text{cov}(\widehat{\theta}_1, \widehat{\theta}_2)}}$$

where θ_1 and θ_2 are the parameters to compare and $\widehat{\theta}_1$ and $\widehat{\theta}_2$ the corresponding sample estimates. Note that the test is applied to each parameter independently, so a test can be run to assess the difference between intercepts of two distance-decay models (θ_{a1} vs. θ_{a2}), and a different test to assess the difference between their slopes (θ_{b1} vs. θ_{b2}). Also note that the test statistic in (1) is based on the asymptotically normal distribution of each parameter estimate, so z_{dep} is a standardization of the estimates difference. In addition, in contrast to an ordinary t-test, the variance in parameter estimation is computed using block-bootstrap techniques (Kunsch, 1989; Liu et al., 1992) to avoid bias due to the pairwise dependence of similarity indices and distance measures. Block-bootstrap techniques are based on defining the sampling unit as a set of values that present a certain degree of correlation, what is called a block, to mimic the dependence structure of the original data (Lahiri, 1999). In our case of pairwise dependent data, we define a block as the set of values obtained from the comparison of one site with all the others, thus capturing their pairwise dependence structure (Martínez-Santalla et al., 2022). In this case, from N sites N blocks are computed (Fig. 2.1) avoiding the inflation of degrees of freedom, as the number of blocks is the same as the number of sites. If we consider the block as the sampling unit for the bootstrap procedure, we ensure that the effective size of the data (N) is equal to the number of observations (N) and, therefore, we avoid inflation-driven biases in the estimation of the variance of parameters.

The site-block resampling procedure we used to estimate the variance of distance-decay model parameters follows the same approach as used by Martínez-Santalla et al. (2022) for a significance test of distance-decay models. In short, resamples of site-blocks (i.e. all similarities or distances involving a given site) are taken with replacement and saved to a similarity block matrix. First, for N sites, N site-blocks are defined by selecting the similarity and distance values involving a given site. These N blocks are saved as columns of two block matrices, one for similarities and another for distances. Second, N block resamples (columns of block matrices) are drawn with replacement, ensuring that the same blocks are sampled for both similarities and distances, and then saved as two resample matrices. Third, resample matrices are initially built with duplicate observations ($N * [N-1]$ values) because each block includes $N - 1$ values. In this step, the resample matrices are downsized to the original size of the similarity and distance matrices ($N * [N - 1] / 2$ values), by sampling $N * [N - 1] / 2$ values from both resample matrices. These values are saved as similarity and distance resample vectors, that

will be used to fit the nonlinear distance-decay models and to estimate the model parameters (θ_1 and θ_2). Fourth, steps 1 to 3 are repeated B times, and B parameter estimates are saved in two vectors, one for θ_1 , and another θ_2 . Fifth, the variance of each parameter, and their covariance, are computed from the vectors of B parameter values.

To evaluate the performance of this approach in the estimation of the parameters' variance and hence the adequacy of the z_{dep} statistic, we conducted a calibration study based on simulations to assess whether the z_{dep} statistic is effective in independently detecting differences in individual distance-decay parameters (i.e. either for the intercepts or for the slopes, or for both). This was done by assessing the proportion of false positives (type I error, i.e. significant p-value when distance-decay curves were simulated with the same dataset, that is, under the null hypothesis of no differences between distance-decay model parameters) and the proportion of false negatives (type II error; i.e. non-significant p-value when simulated distance-decay curves parameters were different). In order to assess if there is a significant difference between two parameters (for example, between two intercepts or between two slopes), and therefore, to conclude if the null hypothesis (e.g., both distance-decay slopes are identical) should be rejected or not, it is necessary to estimate the distribution of the z_{dep} statistic under the null hypothesis, and to obtain the associated p-value. In our case, as the z_{dep} statistic is the standardised difference of parameters, under the null hypothesis, in which θ_1 and θ_2 are assumed to be equal, the statistic z_{dep} follows a $N(0,1)$ distribution. This is because the distribution of the estimators is asymptotically normal, and the linear combination of normal distributions is also normal. In this case, because the z_{dep} statistic is a standardization, it follows a $N(0,1)$.

A non-biased comparison of distance-decay parameters, such as the one provided by the novel z_{dep} test, can be widely applied in biogeographical, ecological and macroecological studies, as evidenced by the variety of studies that have previously sought for analogous analyses. For example, a comparison of distance-decay parameters has been used to assess the differences in distance-decay rates between ecosystems and across scales in fungi (Bahram et al., 2013), between different hosts in parasite communities (Thieltges et al., 2009), between different types of organisms like macro and microinvertebrates (Astorga et al., 2012) or ectotherms and endotherms (Qian & Ricklefs, 2012), between native and non-native species of urban floras (La Sorte et al., 2008), to assess the effects of eutrophication in diatom communities (Goldenberg-Vilar et al., 2014) or to compare how grain size and the study extent affects the form of the distance-decay curves (Steinbauer et al., 2012). Moreover, beta diversity and distance-decay studies are informative for biodiversity conservation strategies and environment management (Gossner et al., 2016; Graco-Roza et al., 2022).

In this paper, we introduce the novel z_{dep} statistic to test for the equality of individual parameters between two distance-decay models and exemplify its use with distribution data of vertebrates in the Iberian Peninsula. The main novelty of this approach is the incorporation of an effective method (i.e. site-block resampling) to avoid the inflation of degrees of freedom in a t-test like statistic, and the performance of a calibration study. First, we performed a

simulation study to analyse the empirical behaviour of the z_{dep} statistic, evaluating its type I and type II errors. For this purpose, we have simulated three scenarios where distance-decay parameters were equal (null hypothesis holds) and three scenarios where the distance-decay parameters were different (alternative hypotheses). This design allows studying the performance of the z_{dep} test with different distance-decay functional forms (i.e. negative exponential, power-law and Gompertz) and in different situations (similar intercept but different slope, different intercept but similar slope, different intercept and slope). Negative exponential and power-law models have been frequently used for distance-decay analyses (Nekola & McGill, 2014), and recently the Gompertz model has been introduced as alternative for sigmoidal distance-decay patterns (Martín-Devasa et al., 2022). Finally, we provide a case-study, applying the z_{dep} statistic to compare distance-decay parameters across several Iberian vertebrate groups (freshwater fishes, frogs, lizards and snakes, bats, rodents and carnivores).

2.2 MATERIAL AND METHODS

2.2.1 Simulation of virtual biological communities

We performed a simulation study to assess the type I and type II errors of the z_{dep} test. We simulated 300 species' ranges ("virtual regional pool") as circular spatial polygons in a virtual landscape of 3000×3000 units. Polygons differed in diameter, as we would expect in species with different dispersal ability and niche breadth, and its centre was allocated randomly into the virtual landscape. As shown by Martín-Devasa et al. (2022a), the shape of the distance-decay patterns varies with (i) the size of species ranges, and (ii) the proportion of restricted vs. widespread species. Therefore, to simulate distance-decay curves with different slopes and intercepts, we varied (i) the diameter of polygons representing species ranges and (ii) the proportion of two range size classes (restricted vs. widespread) in several simulation scenarios (see below). Once virtual species ranges were distributed across the landscape, we randomly allocated sampling sites, each representing a "virtual local community". The presence/absence of species in such virtual local communities was computed from the intersection of species ranges (i.e. circular spatial polygons) at each sampling site. For each virtual landscape, we used two sets of random sampling sites, one set with 25 and the other set with 50 sites, to assess how sample size may affect the type I and type II errors of the test. We performed these simulations using the R packages "sp" (Bivand et al., 2013; Pebesma & Bivand, 2005) and "rgeos" (Bivand & Rundel, 2020). Finally, we measured community similarity between sampling sites using Simpson's pairwise similarity index (Baselga, 2010) with the *beta.pair* function of the R "betapart" package (Baselga & Orme, 2012), as well as the Euclidean spatial distance between them. These similarity and spatial distance matrices are the simulated data used in downstream distance-decay analyses.

2.2.2 Simulation scenarios for hypothesis testing

2.2.2.1 Null hypothesis

We defined our null hypothesis as the equality of a given parameter (intercept or slope) between two distance-decay models. In this case, we simulated the two parameters being equal between both models. To do so, we simulated one virtual regional pool and a single set of virtual local communities, from which we computed one similarity matrix and one spatial distance matrix. To build two distance-decay curves with equal parameters, it is necessary to obtain two bootstrap samples from the same similarity/distance matrices to ensure that both curves followed the same model. Besides, to preserve the pairwise structure of the distance-decay data in the two samples, we performed a site-block resampling of similarity/distance values, with site-blocks defined as the set of values derived from the same observation (Martínez-Santalla et al., 2022). In sum, the two distance-decay models were created by site-block resampling the initial simulated similarities and distances.

Three scenarios were designed to study the performance of the z_{dep} statistic under the null hypothesis:

- Scenario #1: It consisted of 20 widespread species that occupied all the territory, with polygon diameters sampled from a uniform distribution $U [4000, 6000]$, and 280 spatially-restricted species with polygon diameters sampled from a $U [300, 900]$. This scenario is designed to simulate a distance-decay curve with rapid initial decay and incomplete turnover (Fig. 2.2).
- Scenario #2: It consisted of 20 widespread species that did not occupy all the territory, with polygon diameters sampled from a $U [3000, 4000]$, and 280 spatially-restricted species with polygon diameters sampled from a $U [600, 1200]$. This scenario is designed to simulate a distance-decay curve with a decrease of similarity values from the shortest distances, resulting in complete turnover at large distances and therefore values of zero similarity between very distant sites (Fig. 2.2).
- Scenario #3: It consisted of 50 widespread species that did not occupy all the territory, with polygon diameters sampled from a $U [3000, 4000]$, and 250 spatially-restricted species with polygon diameters sampled from a $U [1500, 2000]$. This scenario is designed to simulate a distance-decay curve with high similarity at short spatial distances and a marked decay from medium to large distances (Fig. 2.2).

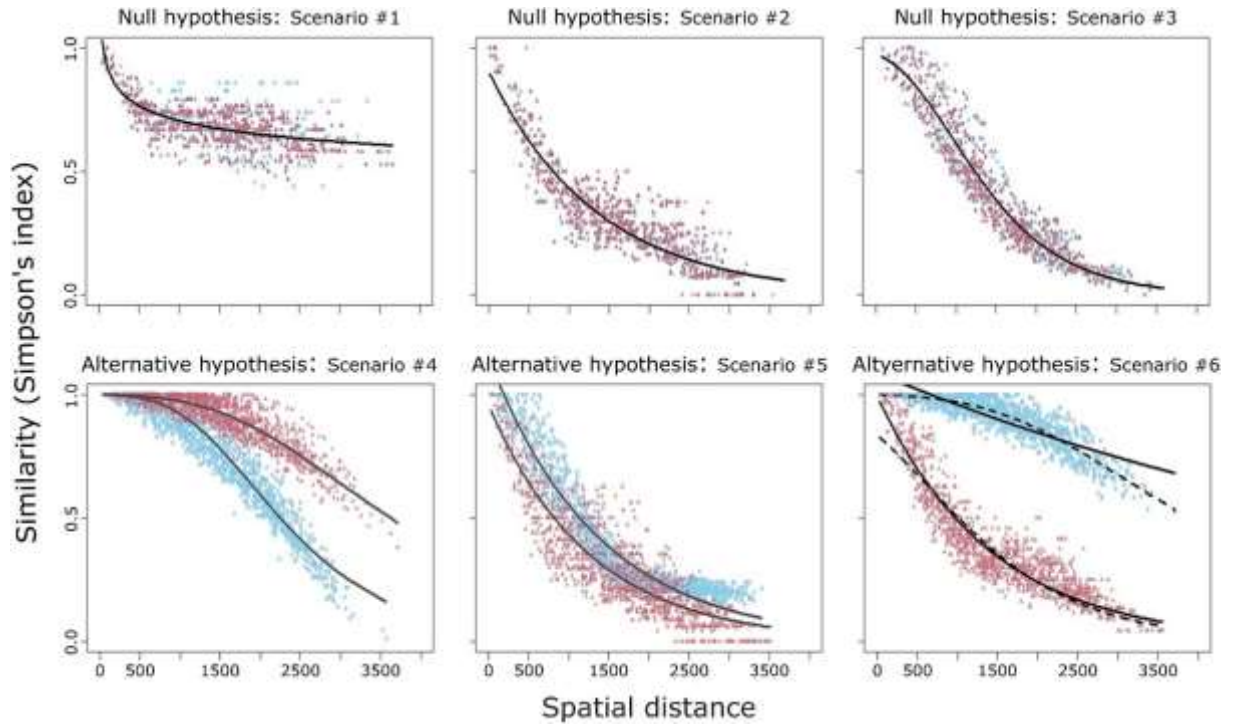


Figure 2.2. Example of distance-decay simulations for the calibration of the z_{dep} statistic under the null (top row) and alternative (bottom row) hypotheses with 50 sampling points. Colours (red and blue) represent the two datasets to compare. For the null hypotheses, each dataset was obtained using site-block resampling from a single similarity and distance matrix. Curves represent the adjustment of the model that best fit the data, as detailed in Table 2.1. In scenario #6, where different models were selected for each dataset, continuous lines represent an exponential model and dashed lines a Gompertz model. Each curve is described by two parameters (θ_a and θ_b , e.g., the intercept and slope of a negative exponential model, respectively), so the z_{dep} test can be used to compare any of them between two datasets (i.e. θ_{a1} vs. θ_{a2} , and/or θ_{b1} vs. θ_{b2}). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2.2.2 Alternative hypothesis

The alternative hypothesis states that the parameters of two distance-decay models (i.e. intercept or slope) are different. In each scenario, we simulated two different virtual communities and, using the same set of sampling sites, we obtained two different similarity matrices and a spatial distance matrix. Thus, we obtained two datasets by combining each similarity matrix with the spatial distance matrix. Then, we computed the distance-decay model for each dataset. Simulating two different datasets, named DS1 and DS2, ensures that their distance-decay curves are also different.

Three scenarios were designed to study the performance of the z_{dep} statistic under the alternative hypothesis:

- Scenario #4: DS1- It consisted of 300 species with polygon diameters sampled from a U [3000, 4000]. DS2- It consisted of 200 widespread species with polygon diameters sampled from a U [4000, 6000] and 100 spatially-restricted species with polygon diameters sampled from a U [3000, 4000]. This scenario is designed to simulate

distance-decay curves with different decay rates and similar, but not identical, initial similarity (Fig. 2.2).

- Scenario #5: DS1- It consisted of 20 widespread species with polygon diameters sampled from a U [4000, 6000] and 280 spatially-restricted species with polygon diameters sampled from a U [1500, 2000]. DS2- It consisted of 20 widespread species with polygon diameters sampled from a U [3000, 4000] and 280 spatially-restricted species with polygon diameters sampled from a U [200,1200]. This scenario is designed to simulate distance-decay curves with similar, but not identical, decay rates and different initial similarity (Fig. 2.2).
- Scenario #6: DS1- It consisted of 200 widespread species with polygon diameters sampled from a U [4000, 6000] and 100 spatially-restricted species with polygon diameters sampled from a U [3000, 4000]. DS2- It consisted of 20 widespread species with polygon diameters sampled from a U [3000, 4000] and 280 spatially-restricted species with polygon diameters sampled from a U [600,1600]. This scenario is designed to simulate distance-decay curves with different initial similarity and decay rates (Fig. 2.2).

2.2.3 Distance-decay models

We performed non-linear regressions of pairwise community similarities over spatial distances to compute distance-decay models. We fit a negative exponential, a power-law and a Gompertz function using the *nlsLM* function of the “minpack.lm” package (Elzhov et al., 2016). We selected the model with the best fit based on their AIC. In the scenarios under the alternative hypothesis, when a different model was selected for each dataset (DS1 and DS2), we compared the parameters in both types of models. Model parameters were the intercept and slope in the case of negative exponential and power-law functions, and the position parameter and slope in the case of the Gompertz function (Martín-Devasa et al., 2022).

2.2.4 z_{dep} : Calibration procedure²

Type I error (rejecting the null hypothesis when it is true) was evaluated by applying the z_{dep} statistic in 1000 simulations of the null hypothesis scenario, thus obtaining 1000 p-values, and computing the proportion of them being less than the nominal values of α (either $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$), i.e. proportion of rejections under the null hypothesis. These α values were chosen because they are the most commonly used significance levels.

The type II error (not rejecting the null hypothesis when it is false) was evaluated with a similar procedure but, in this case, applying the z_{dep} statistic in 1000 simulations of the alternative hypothesis scenario. This provides the empirical proportion of rejections under the alternative hypotheses, i.e. the statistical power of the test (1-type II error). Under the null

² The code used for the simulation study is provided in Appendix 2.3 and 2.4.

hypothesis, the proportion of rejections should be very similar to the significance levels (α), indicating that the type I error is the one assumed for each selected significance level. Under the alternative hypothesis, the test should have a proportion of rejections close to 1 (maximum power), which would indicate that the null hypothesis has been rejected in all cases. The procedure followed for this calibration can be found in Appendix 2.1.

2.2.5 Application of the z_{dep} to real data

We applied the z_{dep} statistic to assess independently the difference in initial similarity (intercept) and in the rate of species turnover (slope) among distance-decay models of several vertebrate orders in the Iberian Peninsula. We obtained the distribution data of freshwater fishes, amphibians, reptiles, and terrestrial mammals from the IUCN red list (<https://www.iucnredlist.org>), accessions 2020 (amphibians, reptiles and mammals) and 2021 (freshwater fishes). Among freshwater fishes, we selected the order Cypriniformes, among amphibians the order Anura (frogs), among reptiles the order Squamata (snakes and lizards) and among mammals the orders Chiroptera (bats), Rodentia (rodents) and Carnivora (carnivores). For each order, we built a presence/absence matrix for the Iberian Peninsula in a $1^\circ \times 1^\circ$ cell grid. The grid was constructed using the “sp” package, with functions *spsample* to place the cell centroids and *SpatialPixelsDataFrame* to obtain the final grid. The community similarity among cells was computed with Simpson's pairwise similarity index, and the spatial distance among them as the geodesic distance between centroids using the *geodist* function of the “geodist” package (Padgham and Sumner, 2020). As in the simulation study, we fitted the negative exponential, power-law and Gompertz functions with the *nlsLM* function and selected the best fitting model based on the AIC statistic.

2.3 RESULTS

To evaluate the type I and type II errors of the z_{dep} test, we performed a simulation study, assessing its performance in three null hypothesis and three alternative hypothesis scenarios. The best fitting models for each distance-decay simulation are shown in Table 2.1. The best supported model only differed between datasets (DS1 and DS2) in scenario #6. Thus, we subdivided this scenario into scenario 6.1, in which we fitted a Gompertz function to both DS1 and DS2 distance-decay curves and, scenario 6.2, in which we fitted an exponential function to both DS1 and DS2 distance-decay curves.

Table 2.1. Models best supported by the data (i.e. smallest AIC value) for each of the simulated distance-decay curves, based on the AIC.

Null hypotheses		Alternative hypotheses		
Scenario	Model	Scenario	Model DS1	Model DS2
1	Power-law	4	Gompertz	Gompertz
2	Negative exponential	5	Negative exponential	Negative exponential
3	Gompertz	6	Negative exponential	Gompertz

2.3.1 Calibration of the z_{dep} test

The z_{dep} test showed good approximation to the nominal levels of $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$ independently of model function (negative exponential, power-law and Gompertz) and sample size ($n = 25$ vs. $n = 50$) (Fig. 2.3, Table S2.1 in Appendix 2.2). This indicates that, under the null hypothesis (equality of parameters), the probability of type I error is the one assumed by the value of α . Under the alternative hypothesis, the test showed low type II error (i.e. the proportion of significant p-values was close to 1). As expected, type II error was slightly larger when the estimated values of the parameters were similar, but not identical (i.e. intercepts in scenario 4 and slopes in scenario 5). In other words, in the scenarios designed to have one parameter very similar and the other quite different (scenarios 4 and 5), the observed type II errors for each parameter reflected this contrast (Fig. 2.3, Table S2.2 in Appendix 2.2). The type II error also decreased with sample size, being lower with $n = 50$ than with $n = 25$.

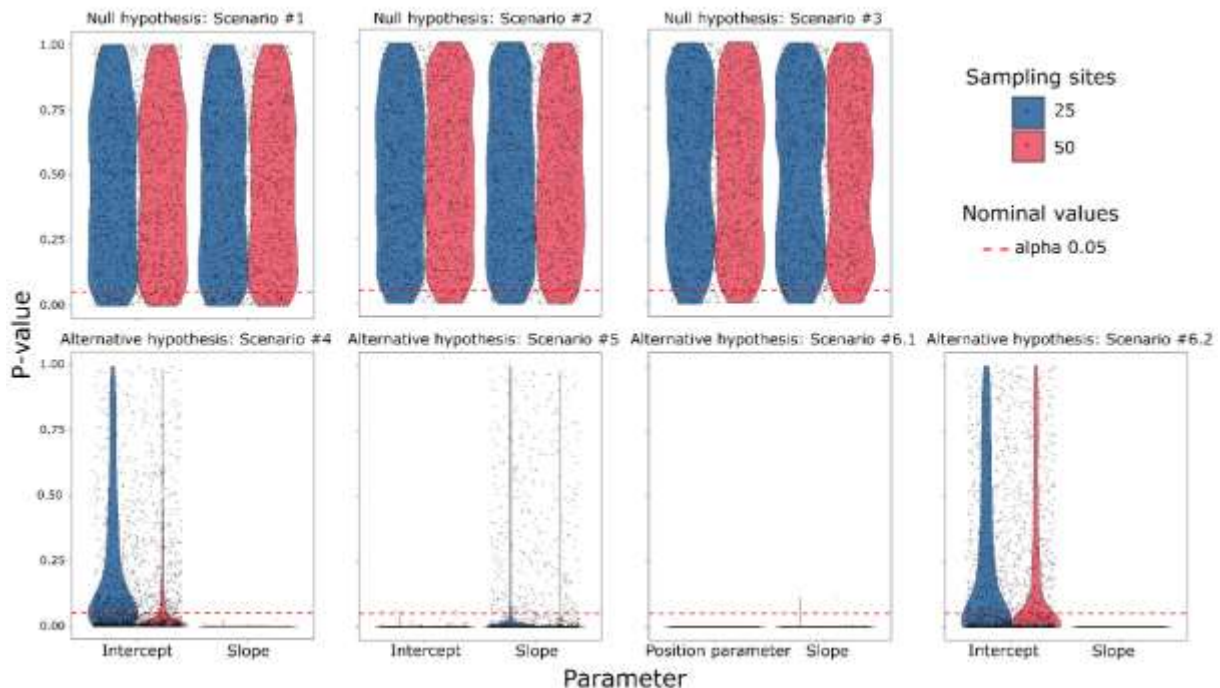


Figure 2.3. Distribution of p-values across 1000 replicates in all simulated scenarios, comparing the results for the two parameters (intercept/position and slope) and with the two sample sizes ($n = 25$ and $n = 50$). Note that under the null hypothesis, the p-values follow a uniform distribution, so the proportion of rejections equals the nominal α level, for any selected value of α . This proves that the test avoids the inflation of degrees of freedom caused by pairwise dependence. Red dotted lines represent $\alpha = 0.05$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3.2 Real data

We compared the distance-decay patterns of Cypriniformes, Anura (frogs), Squamata (lizards and snakes), Chiroptera (bats), Rodentia (rodents) and Carnivora (carnivores) in the Iberian Peninsula by testing the equality of parameters of their distance-decay curves. In all cases, except in the cyprinids, the exponential function was the best fitting model and was thus used for the comparison of distance-decay parameters across taxa. The cyprinids had significantly different intercept and slope than any other group, showing the steepest decay in community similarity (Fig. 2.4). No difference in distance-decay parameters was found

between ectotherm orders (frogs and squamates). Similarly, mammal orders (endotherms) showed all similar intercepts and slopes, with only the carnivores having a marginally significant ($p < 0.1$) flatter slope than bats and rodents (Table 2.2). On the contrary, marked differences were observed between ectotherm and endotherm groups. The frogs and squamates' slopes were significantly different from the ones of bats, rodents and carnivores, being steeper in the ectotherm's groups. However, only bats and frogs showed significant differences in the intercept (Table 2.2). Parameter values can be found in the Appendix 2.2, Table S2.3.

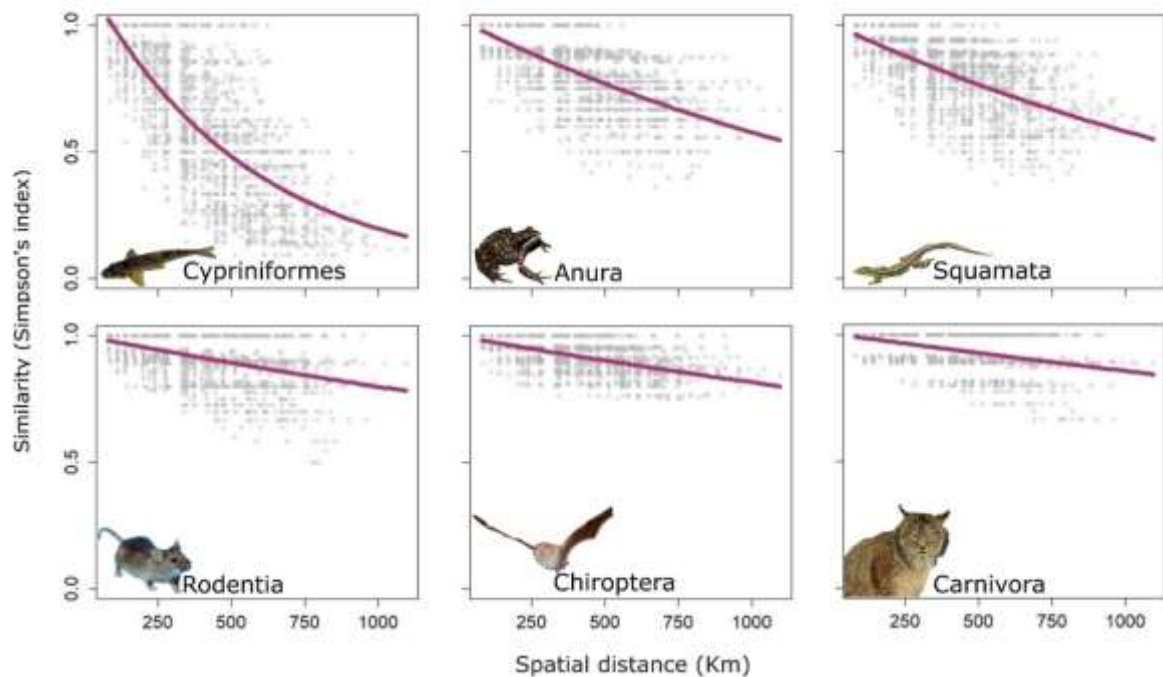


Figure 2.4. Distance-decay models (negative exponential) of Iberian vertebrates' orders (obtained with 63 sampling points). Vertebrate pictures were downloaded from Wikimedia Commons (under Creative Commons licence, authors David Perez, Benny Trapp and Carlosblh).

Table 2.2. P-values of the z_{dep} statistic in the comparison of the intercept and slope between distance-decay models of different groups of Iberian vertebrates (Anu = Anura, Squ = Squamata, Chi = Chiroptera, Rod = Rodentia, Car = Carnivora). Significant values are shown in bold.

	Intercept					Slope				
	Cyp	Anu	Squ	Chi	Rod	Cyp	Anu	Squ	Chi	Rod
Anu	<0.01					Anu	<0.01			
Squ	<0.01	0.34				Squ	<0.01	0.72		
Chi	<0.01	0.02	0.45			Chi	<0.01	<0.01	<0.01	
Rod	<0.01	0.09	0.57	0.88		Rod	<0.01	<0.01	<0.01	0.53
Car	<0.01	0.12	0.84	0.40	0.61	Car	<0.01	<0.01	<0.01	0.07
									0.07	0.08

2.4 DISCUSSION

Our results show the good performance of the z_{dep} statistic for the comparison of parameters between distance-decay models, a widely-used statistical approach in the assessment of biodiversity patterns (Nekola & White, 1999; Morlon et al., 2008; Soininen et al., 2007). To overcome the inflation of degrees of freedom inherent to this type of regression models, the

z_{dep} statistic implements a site-block resampling approach to estimate the variance of model parameters. The site-block resampling ensures that pairwise dependence is preserved in the data while the number of samples used to compute parameter variance (i.e. blocks) is equal to the effective sample size of the original data (i.e. number of sites), avoiding the inflation of degrees of freedom. Site-block resampling has also proven effective in significance tests for the non-linear relationships between community similarity and spatial distance (Martínez-Santalla et al., 2022). Here, we use site-block resampling for a different problem, i.e. a test for the equality of the parameters between two different distance-decay models. The test can be applied, independently, to any of the model parameters (i.e. intercept and/or slopes). Our simulation study under the null hypothesis shows that the type I error of z_{dep} was consistently similar to the nominal α values. Simulations under the alternative hypothesis evidenced that the z_{dep} test also has good statistical power (low type II error). The low Type I error does not seem to be affected by the mathematical function used to fit the curve, so the z_{dep} test can be used to compare parameters of the most frequently used distance-decay models (power-law, negative exponential and Gompertz function). This is an important advantage as the functional form of the distance-decay curve can change depending on the study scale (Nekola & McGill, 2014) and the species range sizes (Martín-Devasa et al., 2022a). As expected, statistical power increases with sample size and decreases when parameter estimates are close, a common behaviour in most tests (Steidl et al., 1997). In the case study with real data, the z_{dep} statistic has also proven useful to reveal differences in the patterns of species turnover with spatial distance among different taxa of Iberian vertebrates.

Previous contributions have evidenced the relevance of comparing the parameters of distance-decay models across biological groups or biogeographic regions in order to better understand the causal processes behind community turnover. Some of these studies have provided different partial solutions to the problem we address in this paper. For example, Goldenberg Vilar et al. (2014) used a two-sample t-test. This approach could be, in principle, suitable to compare regression parameters, but its performance is biased with pairwise dependent data (Rasch et al., 2011). A step forward was introduced by Nekola & White (1999) and Gómez-Rodríguez & Baselga (2018), who used resampling procedures on similarity values to estimate the empirical distribution of either the difference between parameters or the parameters themselves. However, these approaches do not completely solve the problem of Type I error inflation because they do not account for the pairwise dependence of the similarity data during the resampling procedure. In turn, La Sorte & McKinney (2006) developed a resampling method directly on the presence/absence table rather than on the similarity matrix. This method truly accounts for the pairwise dependence of similarity data, but has some important differences with our z_{dep} statistic. First, the observed parameter difference may not be within the bounds of the estimated null distribution, a situation that precludes estimating the probability that the observed parameter difference may be caused by chance (La Sorte & McKinney, 2006). When the value to compare is out of the bounds of the estimated null distribution, this type of permutation tests can result in inaccurate p-value estimations (Onghena & May, 1995; Phipson & Smyth, 2010). This is not a concern for the z_{dep} statistic as its

distribution under the null hypothesis is known to follow a $N(0,1)$ distribution. The second main difference is the calibration test. In this paper we have evaluated the performance of the z_{dep} statistic in terms of type I and type II errors with a simulation study in the specific context of pairwise dependence, obtaining good results. Although permutation tests also tend to have good performances, to our knowledge there is no type I and II error evaluation for La Sorte & McKinney (2006) permutation test, so its approximation to the nominal level (α) and type II error remains unknown. In general, we think that the specific design of the z_{dep} statistic to account for pairwise dependence, its known distribution under the null hypothesis, and the good performance evidenced by our simulation study make it the best option to perform comparisons of parameters between distance-decay models.

The usefulness of the z_{dep} statistic has been exemplified in the study of distance-decay patterns of Iberian vertebrates. We have compared distance-decay parameters among taxonomic groups given that these parameters can be used to understand how the organisms' dispersal and ecological characteristics influence the spatial turnover of biological communities (Gómez-Rodríguez & Baselga, 2018; Morlon et al., 2008; Soininen et al., 2007). Therefore, the steeper distance-decay slope of the cyprinids is probably reflecting the poor dispersal ability of freshwater fishes in comparison with terrestrial vertebrates (Leroy et al., 2019). This is an expected pattern in freshwater fish because drylands and saltwater are barriers that constrain their distribution to rivers and lakes, which can be considered as islands (Dias et al., 2014; Tedesco et al., 2012). As a consequence, dispersal limitation and therefore spatial distance between sites are major factors driving the composition of cyprinids communities (Baselga & Leprieur, 2015; Drakou et al., 2009). Our results also show that, in the Iberian Peninsula, ectotherms have higher turnover rate than endotherms, a pattern that seems to be the general trend across regions (Buckley & Jetz, 2008; Qian & Ricklefs, 2012). This can be explained by the poor dispersal ability of amphibians and reptiles (Crnobrnja-Isailovic, 2007), and/or by their smaller tolerance to environmental changes compared to endotherms (Buckley et al., 2012). In our particular case, the Iberian Peninsula is a biogeographical heterogeneous territory, with regions at both sides of the Eurosiberian / Mediterranean boundary, and multiple mountain ranges creating steep environmental gradients at relatively small spatial scales. As a result, species turnover in lizards and frogs tends to be more marked than in mammals because their community composition is more tightly related to the environmental differences between regions (Sillero et al., 2009). The lack of difference in distance-decay patterns between frogs and lizards, or among mammal orders is probably related to the spatial extent of the study, which may not be large enough to evidence differences when dispersal abilities are largely similar between taxonomic orders.

In general terms, besides the spatial extent of the analysis, the statistical power of the z_{dep} statistic could also be affected by any factor influencing the variance of parameter estimates, as in any other statistical test. For example, smaller grain sizes (e.g., 0.5° cells instead of 1° cells), uneven sampling schemes or low sampling effort might increase the variance of parameters, thus reducing the statistical power of z_{dep} . Regarding the calibration of the Type I error rate, it is important to stress that data simulation under the null hypothesis is far from

trivial. Initially, our intuition was that we could simulate species distributions in a single virtual landscape and use two independent sets of sampling points to compute two distance-decay models (data not shown). However, such procedure yielded a probability of rejection larger than the nominal α value because in fact this approach does not ensure that the distance-decay curves followed the exact same distribution. In other words, the simulations were deviating from the null hypothesis because, even if the species distributions are the same, the patterns of dissimilarity can be different in different regions of the landscape, so using two independent sets of sampling points could result in two different distance-decay curves. To solve this problem, we based our approach to simulate distance-decay data under the null hypothesis on the idea that we needed to generate two models with the same theoretical parameters but different error. By making a resampling of the same model we can ensure that this requirement is fulfilled. With our method we do not know the real parameters, but we are sure that we are working with the null hypothesis because both models are samples of the same one, ensuring that $f_1(x) = \beta + \beta x + \varepsilon_1$, $f_2(x) = \beta + \beta x + \varepsilon_2$. Therefore, we can effectively calibrate Type I error.

Here we have illustrated the use of the z_{dep} statistic to compare distance-decay patterns across taxonomic groups. However, the z_{dep} test can be extended to the study of any pattern resulting from the application of pairwise metrics, for example, the relationship between genetic and spatial distances (Wright, 1943). The functional form of the relationship between genetic and spatial distance can also be modelled with parametric functions, and the slope of this relation is used to infer the organisms' dispersal ability (Chust et al., 2016; Gómez-Rodríguez et al., 2020; Kinlan & Gaines, 2003; Lester et al., 2007). Similarly, the z_{dep} can be also applied to data lacking spatial structure, such as temporal turnover or compositional variation with abiotic gradients. Therefore, the z_{dep} statistic can be extended to the study of numerous diversity patterns at various hierarchical levels of biological organization (e.g., Baselga et al., 2015, 2022a).

2.5 CONCLUSIONS

The good performance of the z_{dep} statistic in the calibration study and its applicability to real biodiversity data illustrated here with the Iberian vertebrates, make it the best option to assess differences in parameters of regression models of pairwise data, such as models of distance-decay of community similarity. The z_{dep} statistic provides accurate results as it accounts for pairwise dependence, which avoids the inflation of degrees of freedom, and thus provides a good approximation to the nominal α level (i.e. good Type I error). Importantly too, the z_{dep} statistic can be applied to any functional form of distance-decay, including non-linear relationships as the negative exponential, power-law, or sigmoidal (i.e. Gompertz). Because the model parameters of these functions have a biological interpretation (Qian, 2009; Saito et al., 2015; Sojininen et al., 2007), assessing their differences across biological groups or biogeographic regions is key for our inferences of the causal processes behind community turnover (Gómez-Rodríguez & Baselga, 2018). In other words, this new method allows assessing whether differences in distance-decay parameters are significant between biological

groups or regions, and this opens the opportunity to study whether the biological attributes of biological groups (e.g., dispersal limitation or ecological requirements) or the characteristics of biogeographic regions drive the differences in distance-decay patterns. This type of inferences, based on comparisons between biological groups or biogeographic regions with contrasting attributes, are crucial for understanding the causal processes behind the spatial turnover of biological communities and, therefore, for our ability to predict how these biodiversity patterns would be affected by future changes in the environmental conditions.

Chapter III

DISPERSAL LIMITATION SHAPES DISTANCE-DECAY PATTERNS OF EUROPEAN SPIDERS AT THE CONTINENTAL SCALE



DISPERSAL LIMITATION SHAPES DISTANCE-DECAY PATTERNS OF EUROPEAN SPIDERS AT THE CONTINENTAL SCALE

ABSTRACT

Aim: To study the relative relevance of dispersal limitation and species sorting as drivers of beta diversity patterns of European spider communities.

Location: Continental Europe.

Time period: Present.

Major taxa studied: Spiders.

Methods: We analysed how distance-decay patterns differ between northern and southern Europe in 15 spider families, using standardized distances to allow a direct comparison between parameters (i.e. slope and intercept) of climatic and spatial distance-decay models. We also assessed north-south differences in parameters of spatial and, independently, climatic distance-decay models, and whether those differences are explained by family-specific traits related to dispersal ability.

Results: Climatic and spatial distance-decay patterns are very similar in northern Europe, where climatic and spatial distances are highly correlated. In contrast, the slopes are steeper in spatial than in climatic distance-decay curves in southern Europe, where climatic and spatial distances are decoupled. Moreover, family traits related to dispersal ability explained well the north-south differences in spatial distance-decay slopes, as well as the amount of nestedness-resultant dissimilarity between southern and northern spider faunas.

Main conclusions: Our results suggest that differences in beta diversity patterns between northern and southern Europe reflect the strength of dispersal limitation in spiders, which varies across families and leads to different degrees of disequilibrium with current climatic conditions depending on the taxon. Moreover, the general decoupling between spatial and climatic distance-decay curves in the South of Europe, where spatial and climatic distances are uncorrelated, suggests that dispersal limitation is the main factor shaping current beta diversity patterns of European spiders at the continental scale.

3.1 INTRODUCTION

Understanding the processes behind the differences in species composition between communities (i.e. beta diversity) is a central question in ecology and biogeography. For this reason, beta diversity patterns have been widely studied across a wide range of territories, scales, times and taxa (e.g., Ávila et al., 2020; Pavlek & Mammola, 2020; Soininen et al., 2007; Steinbauer et al., 2012). Compositional differences between sites are commonly explained by two major processes, dispersal limitation and species sorting (Nekola & White, 1999; Soininen et al., 2007), which are not mutually exclusive (Gravel et al., 2006). Under dispersal limitation, species might not be able to colonize all the areas with suitable conditions because their presence in a site depends on their capacity to get there, so their distributions are not in equilibrium with environmental conditions. As a result, the more spatially distant two sites are, the more dissimilar their communities are as well (Hubbell, 2001; Nekola & White, 1999). In turn, under species sorting, species are present or absent in a site depending on its abiotic characteristics and, in consequence, species distributions are constrained by environmental conditions and at equilibrium with them. Thus, the more different the climatic conditions of two sites are, the more dissimilar their communities (Nekola & White 1999).

The interplay of these two processes (species sorting and dispersal limitation) can lead to perdurable effects of past climatic events on present-day biodiversity patterns (Svenning et al., 2015). Pleistocene glaciations are one of the most important historical climatic events explaining the current distribution of species (Hewitt, 1996, 2000). During this period, most of Europe was covered by ice sheets, and most of the fauna and flora were restricted to glacial refugia, mainly in the Mediterranean region (Hewitt, 1999). When the ice sheets retreated, species were able to expand their ranges and colonize territories in northern latitudes. As a result, biological communities of northern Europe are driven by a post-glacial re-colonization process, which may be still ongoing for some taxa (Gómez-Rodríguez & Baselga, 2018; Svenning & Skov, 2007). Thus, diversity patterns in Europe are expected to be driven by the two aforementioned processes, with varying relative relevance depending on the biological groups: (i) species sorting, when species have recolonized all the areas where climatic conditions are suitable for them, and their distributions have reached equilibrium with climate (Svenning et al., 2011), and (ii) dispersal limitation, when species have not been able to recolonise all the areas where climatic conditions are suitable for them, because their ranges are lagging behind the warming climatic conditions since the last glacial maximum (Svenning & Skov, 2007). This lack of equilibrium between the spatial structure of biodiversity and climate arises due to long-term dispersal limitation in some taxa, which leads to observable legacies of past glaciations in the current structure of biological communities (Gómez-Rodríguez & Baselga 2018).

The effects of long-term dispersal limitation and species sorting in the spatial configuration of European biodiversity have been assessed in a wide range of organisms, especially vertebrates and plants (e.g., Astorga et al., 2012; Leprieur et al., 2009; Saito et al., 2015; Svenning et al., 2011). In invertebrates, these effects are less studied, but previous analyses suggest a major role of dispersal limitation. Specifically, Gómez-Rodríguez & Baselga, (2018)

showed that beta diversity patterns are markedly different between northern and southern Europe in beetle groups with low dispersal ability, evidencing the strong imprint of past climatic events in current diversity patterns. To do this, Gómez-Rodríguez & Baselga, (2018) compared the northern and southern slopes of an ubiquitous macroecological pattern: the decrease of community similarity with spatial distance, also referred to as “distance-decay” (Nekola & McGill, 2014; Nekola & White, 1999). The observed differences in such distance-decay patterns suggested that, in beetle groups with low dispersal ability, most species were restricted to the South of Europe but a few species, i.e. those with good dispersal ability, had been able to reach the North of Europe in a postglacial re-colonization process. As a result, the northern European fauna is composed of species with high dispersal ability, able to colonize distant territories making the species composition between them similar, thus resulting in a flat distance-decay curve (i.e. high community similarity over long spatial distances). In contrast, the dispersal limited species would show narrower distributions and would be restricted to the South of the continent, resulting in steeper distance-decay curves than in the North (Gómez-Rodríguez & Baselga, 2018).

In addition to spatial distance-decay patterns, it is also possible to assess the decay of community similarity with climatic distance (Graco-Roza et al., 2022; Nekola & White, 1999). Analyses of climatic distance-decay relationships aim to assess the effect of species sorting across climatic gradients, in an analogous way as analyses of spatial distance-decay patterns aim to assess the effect of dispersal limitation. Contrasting how well spatial and climatic distances explain community similarity has proven useful to discern the relative relevance of dispersal limitation and species sorting in multiple biological systems (e.g., Baselga & Leprieur, 2015; Leprieur et al., 2009; Qian et al., 2005; Rodriguez-Artigas et al., 2016). However, because there is usually some degree of correlation between spatial and climatic distances, it is necessary to account for their covariation either using variance partitioning (Borcard et al., 1992; Smith & Lundholm, 2010) or alternative approaches. Here, we standardized climatic and spatial distances, to be able to directly compare the slopes of spatial and climatic distance-decay models. Because the slopes of distance-decay models inform about the rate at which species are replaced with distance, using standardized spatial and climatic distance to build the distance-decay models allows directly comparing both rates.

A complementary approach to assess whether diversity patterns may be the result of incomplete postglacial colonization is to assess whether differences between northern and southern European faunas are related to nestedness (i.e. the northern fauna is a subset of the southern fauna) or spatial turnover (i.e. northern and southern faunas are constituted by unique sets of species) in community composition (Baselga, 2010). If northern faunas were nested within southern faunas, it would point to postglacial recolonization processes primarily constrained by dispersal limitation while, on the contrary, spatial turnover would suggest species sorting processes along climatic gradients at continental scales (Dobrovolski et al., 2012; Wu et al., 2017). In other words, if dispersal limitation was the major factor, the northern European fauna would be a subset of the southern European one. Therefore, dissimilarity between southern and northern faunas would be mainly due to nestedness, and groups with poor

dispersal ability would show higher differences, as found in Dobrovolski et al. (2012) for New World vertebrates. In contrast, if species were not dispersal limited but their ranges were constrained by climatic tolerances, we would expect a larger contribution of spatial turnover to dissimilarity between southern and northern faunas, as found by Svenning et al. (2011) for European mammals.

In this paper, we will focus on European spiders. Spiders are one of the most diverse arthropod groups (with ca. 50 000 species described to date, World Spider Catalog, 2023), and are present in most terrestrial ecosystems (Wise, 1993). Their long-distance dispersal ability depends on a method of passive aerial dispersal called ballooning, which is based on being carried by the wind supported by silk threads (Duffey, 1998). This method can be so efficient that some spiders have been found at altitudes up to four kilometres (Freeman, 1946; Glick, 1939) and can travel more than 1000 km (Bell et al., 2005), being one of the first taxa to colonise islands (Edwards & Thornton, 2001; New & Thornton, 1988). Despite this dispersal ability, the European spider fauna is expected to be in disequilibrium with climate, as not all species seem to have recolonized the climatically suitable northern territories (Koponen, 1991). Taxa with strong ballooning propensity might have reached the northern latitudes, but it is not clear how different spider groups have responded to deglaciation (Koponen, 1991) and the degree to which the legacy of past glaciations imprints the spatial structure of present-day assemblages. Ballooning tendency varies among spider families (Bell et al., 2005; Cardoso, Pekár, et al., 2011; Foelix, 2011; Szymkowiak et al., 2007), which also exhibit different ecological and morphological characteristics (e.g., occupied habitat/vertical stratum, web building behaviour or body size) that can condition ballooning propensity as well (Blandenier, 2009; Dean & Sterling, 1985; Larrivé & Buddle, 2011). This variation in dispersal, morphological and ecological traits affects the relative relevance of dispersal limitation and species sorting processes across taxa. The role of both processes in driving spider communities has been studied at small spatial scales (e.g., Baldissera et al., 2012; Carvalho et al., 2011; Fernandez-Fournier & Avilés, 2018; Zhang et al., 2018) but not at a continental scale, with the exception of subterranean spiders (Mammola et al., 2019).

Here we aim to study the role of dispersal limitation and climatic constraints in shaping current beta diversity patterns of European spider fauna at the continental scale. To do so, we compared spatial and climatic distance-decay curves of 15 families in northern and southern Europe to assess (i) how the relationship between compositional similarity and spatial or climatic distances differs between northern and southern Europe. We also assessed, for each spider family, (ii) whether the differences in species composition between northern and southern Europe are more related to nestedness or spatial turnover. We expect nestedness to be predominant if compositional differences are caused by dispersal limitation, whereas turnover will be predominant if the differences are caused by species sorting. Finally, we analysed (iii) whether dispersal-related traits of spider families explained the differences in species composition and in distance-decay parameters between the North and South of Europe. We expect a significant relationship between the differences in distance-decay parameters and

dispersal-related traits of spiders if dispersal limitation is the main process affecting the composition of European spider assemblages.

3.2 MATERIAL AND METHODS

3.2.1 Distribution data

Presence/absence tables for 15 spider families (3075 species in total) in 39 continental European territories were compiled from araneae.nmbe.ch (accession, July 2022). We selected families Agelenidae, Araneidae, Clubionidae, Dictynidae, Gnaphosidae, Hahniidae, Linyphiidae, Liocranidae, Lycosidae, Philodromidae, Pholcidae, Salticidae, Tetragnathidae, Theridiidae and Thomisidae because of their large number of species, which allowed the robust computation of community similarity indices, their wide distribution through continental Europe and their different biological characteristics. The source provides a list of species by country (except for Russia, subdivided into five regions: North, Centre, South, East and West) based on published observations. Thus, most of our territories correspond to European countries, with some remarks. Due to their different biogeography, Islands were excluded from this study. Countries with an area of less than 2 000 km² (Monaco, Liechtenstein, San Marino and Vatican City) were excluded from our dataset to avoid extreme area differences and, for the same reason, we maintained the subdivision of European Russia into five territories (North, Centre, South, East and West). Bosnia-Herzegovina and Croatia were considered a single territory because the latter surrounds the former. Kosovo was excluded from the analyses due to the absence of data on some taxa that may cause bias in computing community similarity. We studied the species-area relationship of the remaining territories to detect if any of them have incomplete inventories. We fitted a linear model to the $\log(\text{richness}) \sim \log(\text{area})$ relationship, and identified influential cases based on Cook's distance (cut-off = 0.15, $F_{1,37}$, 30th percentile (Ayinde et al., 2015)). Moldova and European Turkey stood out (Fig. S3.1 in Appendix 3.1), suggesting their inventories were largely incomplete, so both territories were also excluded from subsequent analyses.

3.2.2 Community similarity and distance-decay models

To assess the effect of dispersal limitation or species sorting across climatic gradients we compared the decay of community similarity with spatial and, independently, with climatic distance for 15 spider families, and for northern and southern Europe separately. For each family, we constructed two presence/absence tables, one for northern (territories with centroids at > 48 degrees) and one for southern (territories with centroids at < 48 degrees) Europe. A latitude of 48 degrees has been chosen as cut-off reference because it is the approximate latitude at which the ice sheets covered the continent during the Last Glacial Maximum (Hughes et al., 2016). Each presence/absence table was used to compute the pairwise similarity between territories using the Simpson's similarity index ($1-\beta_{\text{sim}}$), with the function *beta.pair* of the R package "betapart" (Baselga et al., 2023; Baselga & Orme, 2012). Spatial distances between territories were computed as the geodesic distance between its centroids using the *geodist* function of the package "geodist" (Padgham & Sumner, 2021). To compute climatic distances,

we first extracted the mean value of six climatic variables for each territory from Worldclim (Hijmans et al., 2005): mean annual temperature (Bio1), maximum temperature of the warmest month (Bio5), minimum temperature of the coolest quarter (Bio 6), annual precipitation (Bio12), precipitation of the wettest quarter (Bio16) and precipitation of the driest quarter (Bio17). These variables were submitted to a principal component analysis (PCA) and the first two dimensions (explained variance = 99%) were retained. We computed the climatic distances between territories as the Euclidean distance between their respective PCA scores. We also computed the correlation between spatial and climatic distances, both in northern and southern Europe, with a Mantel test using 1000 permutations. The degree to which spatial and climatic distances are correlated is relevant because, if correlation is low, the relationship between community similarity and spatial or climatic distance could be attributed to dispersal limitation or species sorting, respectively, and it would be independent of the alternative process.

To make the parameters of climatic and spatial distance-decay models directly comparable, we first standardized spatial and climatic distances between 0 and 1. This allows, for example, assessing whether the slopes of spatial or climatic distance-decay models are steeper, pointing to the preponderance of dispersal limitation or species sorting, respectively. Each standardized distance value (d_s) was computed as $d_s = (d_i - \min(d)) / (\max(d) - \min(d))$ where d_i is the original spatial or climatic distance value, and d is the vector of all original spatial or climatic distances. Standardized spatial and climatic distance-decay models were built fitting power-law, negative exponential and Gompertz functions (Martín-Devasa et al., 2022) to the relationship between similarity and standardized spatial/climatic distances with the *nlsLM* function of the package “minpack.lm” (Elzhov et al., 2016). The negative exponential model was selected for subsequent analyses as it showed the lowest AIC values in most cases (Tables S3.1 and S3.2 in Appendix 3.1). The differences between spatial and climatic distance-decay parameters in each spider family were independently assessed for the North and the South with the z_{dep} statistic, a test specifically designed to compare parameters of models fitted with pairwise dependent data (Martín-Devasa et al., 2022b). Additionally, to control for the evolutionary relationships of spider families, we also tested for the difference in the average of intercepts and slopes between standardized spatial and climatic distance-decay models with a phylogenetic paired t-test using the *phyl.pairedttest* of the library “phytools” (Revell, 2012). To this end, a phylogenetic tree for the 15 spider families included in our analyses was obtained by pruning the phylogenetic tree in Macías-Hernández, et al., (2020a), keeping only one species per family. We chose common and widely distributed species (Table S3.3 in Appendix 3.1).

To compute the dissimilarity between northern and southern faunas as a whole, the territory-level presence/absence tables were collapsed into a new table coding the presence/absence of each species in two larger regions, northern (> 48N) and southern Europe (< 48N). This collapsed table was used to compute the pairwise dissimilarity between northern and southern faunas (β_{sor}), and the fractions of dissimilarity related to species turnover (β_{sim}) and nestedness (β_{sne}) (Baselga, 2010). Whether spatial turnover or nestedness-resultant dissimilarity were the dominant fraction was determined using the ratio β_{sim}/β_{sor} (β_{ratio}).

3.2.3 Relationship between beta diversity patterns and dispersal-related traits of spiders

To study whether dispersal-related traits explain the beta diversity patterns of spiders in Europe, we assessed the relationship of these traits with (i) the North-South difference in intercepts of spatial and climatic distance-decay models, (ii) the North-South difference in slopes, and (iii) the values of turnover and nestedness-resultant dissimilarity between the North and South of Europe. Ballooning is the main long-distance dispersal method of spiders, so spider families were classified according to their ballooning tendency in frequent and infrequent ballooners. The assignation of each spider family to these categories was based on studies of ballooning behavior and its proportion in aerial samples of the different families (L. Bishop & Riechert, 1990; Blandenier, 2009; Blandenier & Fürst, 1998; Dean & Sterling, 1985; Greenstone et al., 1987; Pearce et al., 2005). Following Carvalho & Cardoso (2014), families with less than 1% of presence in aerial samples were classified as infrequent ballooners, whereas families with more than 1% of presence in aerial samples were classified as frequent ballooners. We also incorporated into the analysis traits that can affect the ballooning tendency and efficiency: maximum female body size (Dean & Sterling, 1985), vegetation strata (Blandenier, 2009; Platnick, 1976) and silk production (Bonte et al., 2004; Larrivé & Buddle, 2011). We estimated the maximum female body size of each family by obtaining a random sample of 50 species and averaging across species the maximum female body size. For families with less than 50 species, all species were considered. Female body size data was obtained from araneae.nmbe.ch, Esysunin & Sozontov (2016), Koch (1879), Tyschchenko (1965), Marusik & Koponen (1998), Ponomarev (2009), Ledoux (2014) and Macías-Hernández et al. (2020b). We assigned each family a vertical strata category -soil, vegetation or both- based on the classification of Cardoso et al. (2011). Finally, we also classified spider families as web-builders or non-web builders. The spider families' traits can be found in Table 3.1.

Table 3.1. Classification of spider families according to dispersal traits.

	Ballooning	Web-builders	Strata	Max size females
Agelenidae	Infrequent	Yes	Both	9.57
Araneidae	Frequent	Yes	Vegetation	11.70
Clubionidae	Infrequent	No	Vegetation	7.66
Dictynidae	Infrequent	Yes	Both	3.14
Gnaphosidae	Infrequent	No	Soil	7.83
Hahniidae	Infrequent	Yes	Soil	2.66
Linyphiidae	Frequent	Yes	Both	2.80
Liocranidae	Infrequent	No	Soil	5.95
Lycosidae	Frequent	No	Soil	10.03
Philodromidae	Frequent	No	Both	6.62
Pholcidae	Infrequent	Yes	Both	5.50
Salticidae	Frequent	No	Both	6.03
Tetranychidae	Frequent	Yes	Vegetation	9.86
Theridiidae	Frequent	Yes	Both	3.83
Thomisidae	Frequent	No	Both	7.10

To assess whether the differences in spatial and climatic distance-decay parameters between northern and southern Europe are related with the dispersal and ecological traits of spider families we used phylogenetic generalised least square analyses (pgls) with the *pgls* function of the package “caper” (Orme et al., 2018), considering as dependent variable (i) differences in intercepts (northern intercepts – southern intercepts), (ii) differences in slopes (northern slopes – southern slopes), and (iii) dissimilarity in species composition between northern and southern faunas due to (iii.a) turnover (β_{sim}) or (iii.b) nestedness (β_{sne}). Traits in evolutionarily related groups may present phylogenetic autocorrelation that can inflate the type I (false positive) and type II (false negative) errors (Gittleman & Kot, 1990), so we used pgls models to account for phylogenetic autocorrelation as inferred from the pruned phylogenetic tree based on Macías-Hernández et al. (2020a). Sixteen different pgls models were computed considering all possible trait combinations and choosing the model with the lowest AIC.

3.3 RESULTS

We first measured the correlation between spatial and climatic distances to know whether any relationship between community similarity and spatial or climatic distance could be attributed to dispersal limitation or climatic sorting, respectively, independently of the alternative process. Correlation between spatial and climatic distance was high in the North (Mantel test, Pearson $r = 0.72$, $p < 0.01$), but low in the South (Pearson $r = 0.14$, $p = 0.17$). In southern Europe, spatial distance explained a relevant fraction of variation in community similarity in most families (pseudo- $r^2 > 0.3$ in 9 families, see Fig. 3.1), while climatic distance explained a negligible fraction (pseudo- $r^2 < 0.07$ in all families, Fig. 3.1). On average, the intercept was significantly

lower (phylogenetic paired t-test, $t_{12} = 3.33$, $p < 0.01$) for climatic (0.78 ± 0.08) than for spatial models (0.86 ± 0.06), and mean slope was also significantly flatter ($t_{12} = 6.68$, $p < 0.01$) for climatic (-0.08 ± 0.07) than for spatial models (-0.53 ± 0.29). When comparisons were made within each family using the z_{dep} statistic, the intercepts were significantly higher for spatial curves in 10 out of 15 families, and slopes were significantly steeper for spatial than for climatic-decay curves in 12 out of 15 families (Table S3.4 in Appendix 3.1). All parameters are shown in table S3.5 in Appendix 3.1. In Northern Europe, spatial distance explained a small fraction of variation in community similarity (pseudo- $r^2 > 0.3$ in only three families, see Fig. 3.2), and climatic distance explained a similar fraction of variation in most families (difference in pseudo- r^2 between climatic and spatial models ≤ 0.1 in all families but two, Fig. 3.2). In the North, the average intercept was significantly lower (phylogenetic paired t-test $t_{12} = 3.25$, $p < 0.01$) for climatic (0.94 ± 0.04) than for spatial models (0.96 ± 0.03), and mean slopes were not significantly different (phylogenetic paired t-test $t_{12} = 1.70$, $p < 0.11$) between climatic (-0.18 ± 0.12) and spatial models (-0.23 ± 0.10). When comparisons were made within each family, spatial and climatic intercepts and slopes were significantly different only for one of the 15 families (Table S3.4 in Appendix 3.1). Taken altogether, these results evidenced that spatial and climatic distance-decay models are decoupled in the South of Europe, but not in the North.

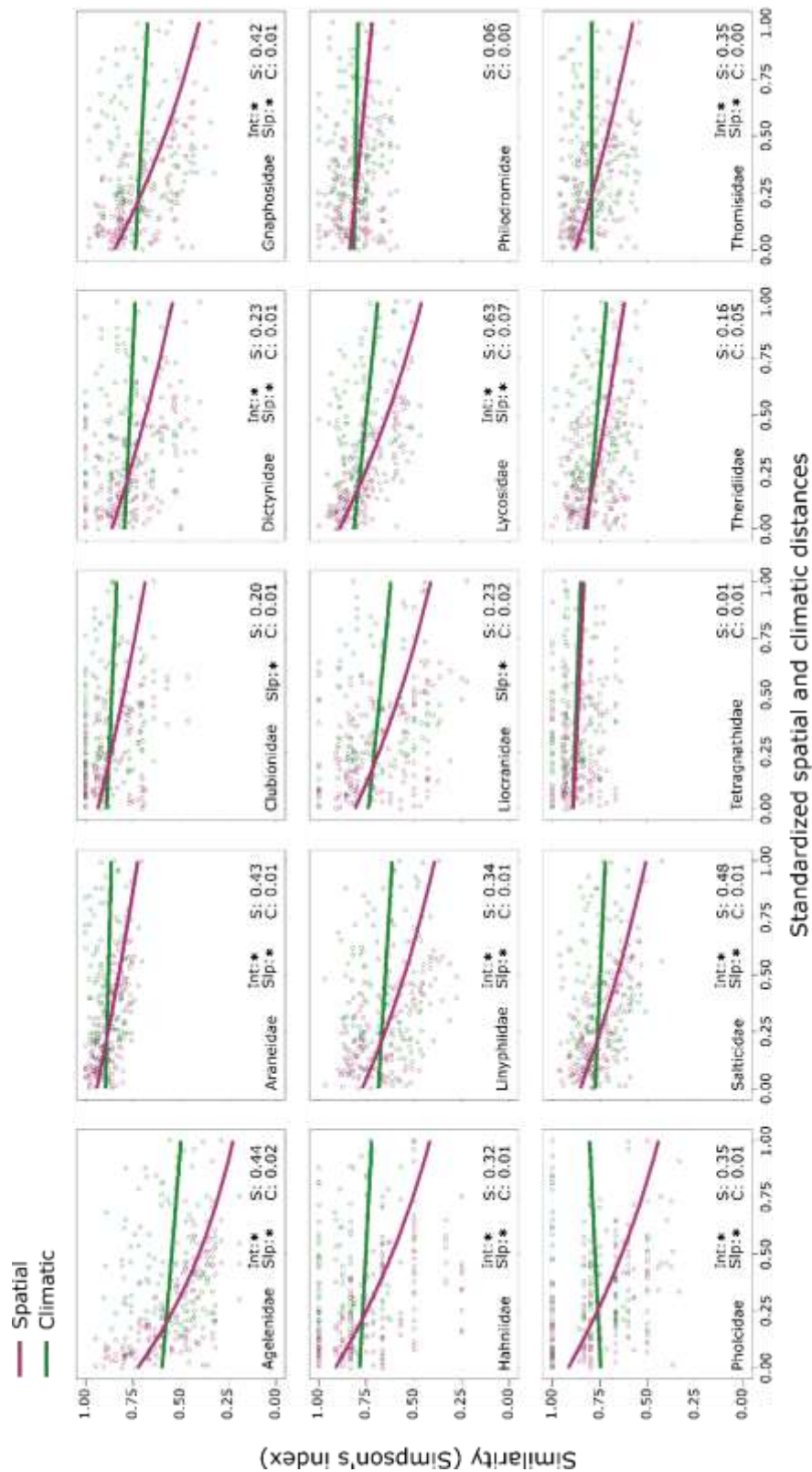


Figure 3.1. Distance-decay models (fitted with a negative exponential function) using standardized spatial and climatic distances for 15 spider families in southern Europe. Spatial and climatic distances were scaled to [0,1] range to allow a direct comparison of models' parameters. S and C indicate the pseudo-r2 of the spatial and climatic distance-decay models, respectively. Int: * indicate significant differences between intercepts and Slp:* indicate significant differences between slopes.

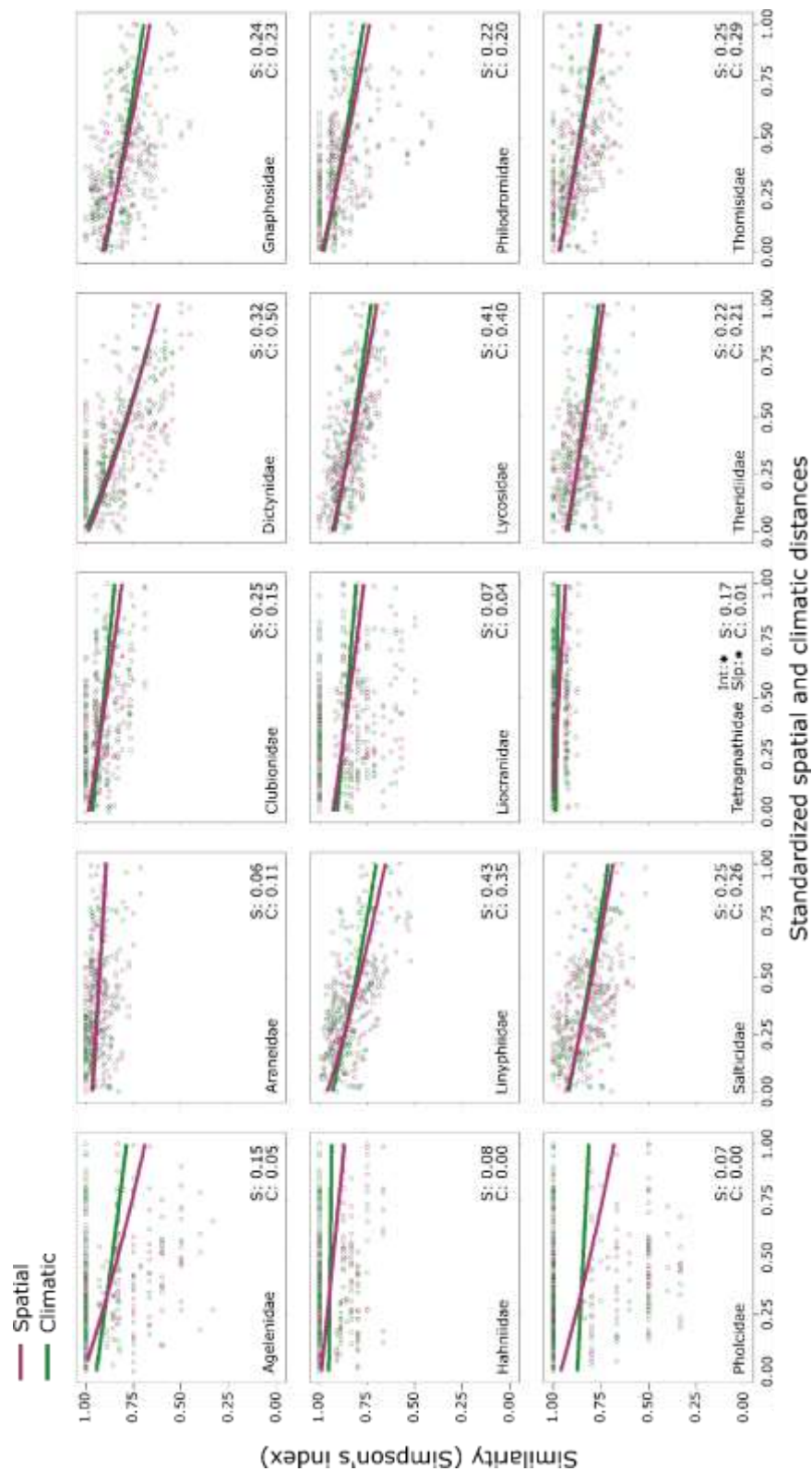


Figure 3.2. Distance-decay models (fitted with a negative exponential function) using standardized spatial and climatic distances for 15 spider families in northern Europe. Spatial and climatic distances were scaled to [0,1] range to allow a direct comparison of models' parameters. S and C indicate the pseudo-r² of the spatial and climatic distance-decay models, respectively. Int: * indicate significant differences between intercepts and SIp:* indicate significant differences between slopes.

When the relationship between the North-South difference in parameters of spatial distance-decay models and family traits was assessed, we found that southern slopes were steeper than northern slopes and that North-South differences in slopes were well explained by all the spiders' dispersal traits, as the best model included ballooning (negative effect), vegetation strata (negative effect, less difference to higher strata), silk production (positive effect) and maximum female size (positive effect) of the spider families ($R^2 = 0.71$, $F_{5,9} = 4.35$, $p = 0.03$). In turn, no predictor significantly explained the North-South difference in intercepts. Regarding climatic distance-decay models, we found that intercepts were higher in the north than in the south, and that North-South differences in intercept were well explained by the model containing ballooning (negative effect), vegetation strata (negative effect), and maximum female size (positive effect) of the spider families ($R^2 = 0.63$, $F_{4,10} = 4.33$, $p = 0.03$). However, North-South differences in climatic distance-decay slopes were not significantly related to any of the predictors.

When the dissimilarity between the northern and southern European spider faunas was assessed, nestedness-resultant dissimilarity (β_{sne}) was the larger component of compositional differences between northern and southern Europe for 10 of the 15 families studied ($\beta_{ratio} < 0.5$). In contrast, species turnover was the largest component in the families Clubionidae, Dictynidae, Linyphiidae, Lycisodae and Thomisidae ($\beta_{ratio} > 0.5$) (Table 3.2). The amount of nestedness-resultant dissimilarity between southern and northern Europe was well explained by dispersal traits of spider families. The best model included ballooning tendency (negative effect), maximum female body size (positive effect) and vegetation strata (negative effect) of spider families ($R^2 = 0.68$, $F_{4,10} = 5.38$, $p = 0.01$). In turn, no significant relationship was found between any of the dispersal traits and species turnover between North and South.

Table 3.2. Values of β_{sim} , β_{sne} , and β_{ratio} for each spider family.

	sim	sne	β_{ratio}		sim	sne	β_{ratio}
Agelenidae	0.118	0.573	0.170	Lycosidae	0.220	0.142	0.607
Araneidae	0.063	0.132	0.321	Philodromidae	0.111	0.233	0.323
Clubionidae	0.114	0.081	0.587	Pholcidae	0.083	0.393	0.175
Dictynidae	0.190	0.116	0.622	Salitricidae	0.114	0.292	0.281
Gnaphosidae	0.143	0.347	0.292	Tetragnathidae	0.050	0.190	0.208
Hahniidae	0.063	0.223	0.219	Theridiidae	0.093	0.268	0.258
Liocranidae	0.100	0.329	0.233	Thomisidae	0.207	0.193	0.517
Linyphiidae	0.294	0.152	0.659				

3.4 DISCUSSION

Our results show that dispersal limitation is the main factor shaping present-day beta diversity patterns of European spiders at the continental scale. We can draw this conclusion because in southern Europe (i) spatial and climatic distances are not correlated, (ii) community similarity is well explained by spatial distance but not by climatic distance, and (iii) slopes are much steeper in spatial than climatic distance-decays. In turn, in northern Europe the explanatory

power of spatial and climatic distances is similar, as expected from the high correlation of spatial and climatic distances. Therefore, in northern Europe we cannot easily infer the causal processes behind the observed distance-decay patterns. We could speculate that either (i) climatic constraints are more relevant in northern Europe due to the prevalence of harsher climatic conditions or, alternatively, (ii) that the same dispersal limitation observed in southern Europe is behind the observed distance-decay patterns in the North, and climatic distance explains the pattern just because it is correlated to spatial distance. Finally, dispersal related traits of spider families explain (i) the North-South difference between slopes of spatial distance-decay curves, and (ii) the amount of nestedness-resultant dissimilarity between northern and southern faunas. Moreover, the relationships between differences in slope and the morphological and ecological traits were in agreement with the predictions derived from their hypothesized link with dispersal (except silk production). Taking all together, our results point to a major role of dispersal limitation and postglacial recolonization lags in the continental scale biogeographic patterns of spiders, as previously found for beetles (Gómez-Rodríguez & Baselga, 2018).

Previous studies have revealed that community composition in spiders generally depends on the combination of dispersal and environmental (climate and vegetation structure) constraints (Ávila et al., 2020; Barton et al., 2017; Carvalho et al., 2011; Mammola et al., 2019; Rodríguez-Artigas et al., 2016; Tonkin et al., 2016; Q. Zhang et al., 2018), but the importance of each process seems to change with scale. At small spatial scales, species sorting has been frequently suggested as the most important process driving spider community composition due to the strong relationship between environmental variables and spider communities (Bowden & Buddle, 2010; Finch et al., 2008; Jiménez-Valverde et al., 2010; Jiménez-Valverde & Lobo, 2007), but dispersal limitation has also been shown to affect the spatial structure of spider communities, for example, in dune and cave systems (Bonte et al., 2004; Carvalho et al., 2011; Mammola et al., 2019). At the continental scale, the preponderance of dispersal limitation as a driver of differences in species composition is here revealed by multiple results, as argued above. These lines of evidence include macroecological patterns based purely in the distribution of species (as the contrast between climatic and spatial distance-decay models), but also clear relationships between these distributional patterns and organismal traits related to dispersal (ballooning, vegetation stratum, body size or silk production). For example, ballooning was negatively associated with North-South differences in distance-decay slopes, suggesting that smaller differences in slope were linked to families with higher dispersal ability. Marked North-South differences in the slopes of spatial distance-decay curves have been previously interpreted as the result of dispersal limitation in a previous paper on European beetles (Gómez-Rodríguez & Baselga, 2018). The rationale behind is that, under strong dispersal limitation, most species have not been able to re-colonise the North of Europe, so these species are restricted to southern territories. This leads to steep distance-decay curves in southern Europe. In contrast, the few species able to colonise the northern regions are good dispersers which tend to occupy larger regions, leading to flatter distance-decay slopes in northern Europe. In European spider communities, we have observed the same contrast between northern and southern slopes and, as for beetles, the differences between slopes are well explained by

dispersal-related traits. The relationship between these traits and the differences in climatic-decay intercepts between north and south might just reflect the general smaller similarities in southern than in northern Europe. This inference can be reached because in southern Europe similarity depends on spatial distance but is unrelated to climatic distance, and spatial and climatic distances are mostly decoupled. This suggests that lower intercepts in climatic distance-decay models in the south are just the result of spatially distant pairs of countries (with low similarity) that are climatically similar (small climatic distances).

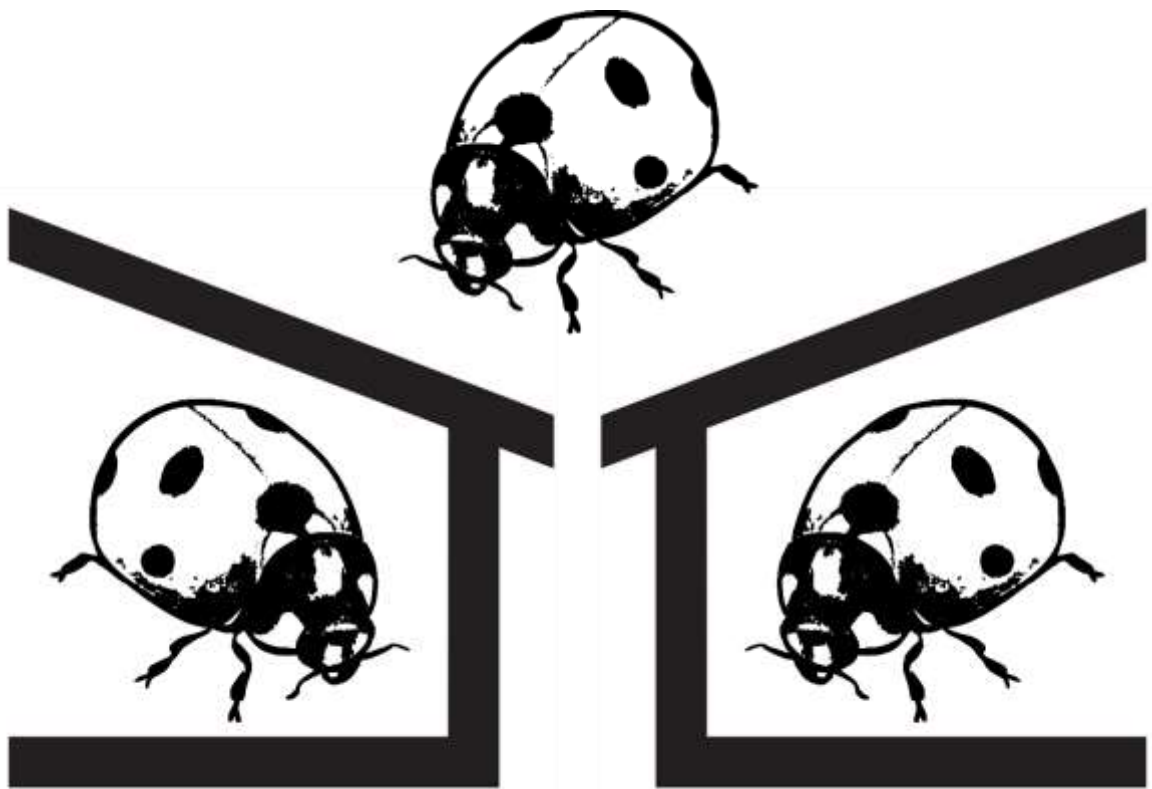
Besides contrasting distance-decay patterns, dispersal limitation and incomplete post-glacial colonisation also lead to northern faunas being a subset of southern ones, as has been also observed in a variety of taxa (e.g., Boston et al., 2015; Csapó et al., 2020; Orlov et al., 2007), including spiders in northern Canada (Loboda & Buddle, 2018), and subterranean spiders in Europe (Mammola et al., 2019). It could be argued that environmental tolerances in a climatic gradient may also produce a nested structure if only organisms with high tolerance could colonise the extreme climates of northern Europe (Ulrich et al., 2009). Previous work has already pointed out that temperature is an important variable driving spider alpha and beta diversity (Finch et al., 2008; Jiménez-Valverde et al., 2010), especially at large scales (Carvalho et al., 2011). However, the relationship between spider dispersal traits and the nestedness-resultant dissimilarity between North and South points out that nestedness may be primarily caused by differences in dispersal ability among families. It should be stressed that the dominant role of dispersal limitation is here being discussed for the spider faunas as a whole, but species sorting could still be especially relevant for some specific spider families. For example, several families as the sheet weaver spiders (Linyphiidae), wolf spiders (Lycosidae) and crab spiders (Thomisidae) present relatively high North-South species replacement (high β ratio), suggesting that they are better dispersers and closer to equilibrium with climatic conditions. These families, particularly Linyphiidae, tend to be dominant in northern territories (Koponen, 1991), especially in boreal and arctic areas (Dahl et al., 2018; Marusik & Koponen, 2002), so a deep study of the processes underlying this high North-South species replacement in these particular families still deserves future work, as previously suggested by Koponen (1991).

In conclusion, our results show that although spiders are usually considered to have good dispersal ability, at the continental scale dispersal limitation appears to be the main process shaping European spider communities. This does not negate the fact that some spider species have outstanding dispersal abilities and can use ballooning to travel large distances (as shown in Bell et al., 2005). Those species are probably the ones that have been able to colonise northern Europe since the last glacial maximum. However, a large proportion of species seem to have limited dispersal ability, and hence their distribution is still restricted to southern Europe. This emerges as contrasting beta diversity patterns between the northern and southern regions, a legacy of the incomplete recolonization of northern Europe, as also shown for other taxa (Baselga et al., 2012; Gómez-Rodríguez & Baselga, 2018; Pinkert et al., 2018; Svenning et al., 2015; Svenning & Skov, 2007; Willner et al., 2009). An incomplete recolonization of northern regions implies that species distributions are far from equilibrium with climatic conditions,

which poses important challenges for biodiversity in a context of climate change (Lenoir et al., 2020).

Chapter IV

INFERRING THE CAUSAL PROCESSES OF GENETIC STRUCTURE FROM THE SHAPE OF THE GENETIC-SPATIAL DISTANCE RELATIONSHIP



INFERRING THE CAUSAL PROCESSES OF GENETIC STRUCTURE FROM THE SHAPE OF THE GENETIC-SPATIAL DISTANCE RELATIONSHIP

ABSTRACT

A widely known biogeographic pattern is the tendency of genetic distance to increase with spatial distance among populations of the same species (i.e. the genetic-spatial distance relationship). However, the strength and shape of this relationships may differ among species and regions as it is related to genetic drift and gene flow, which ultimately depends on the species' dispersal ability. From a theoretical standpoint, the shape of this relationship could be linear or asymptotical and, thus, could be also modelled by functions such as power-law or negative exponential. Additionally, here we postulate the need to account for an additional shape in the relationship: a sigmoidal relationship arising when gene flow is weak or absent at large distances but very high at the shortest distances. Thus, genetic distance would remain constant when populations are close, then it would tend to increase with distance until gene flow among populations is minimal. To model such pattern of genetic structure, we introduce the use of the Gompertz function, which has been recently shown to successfully capture the analogous pattern at the species-level: the distance-decay of biotic similarity. Here, we study the differences in the shape of the genetic-spatial distance relationship pattern among closely-related species, using Iberian leaf beetles as case-study. We test which model function (i.e. linear, power-law, negative exponential and Gompertz) better fits such relationship and use the z_{dep} statistic to test for differences in model parameters. The Gompertz function has shown its usefulness and versatility to model the genetic-spatial distance relationship of Iberian leaf beetles. It can fit various forms of the pattern, from almost linear to sigmoidal shapes and even cases in which lineages seem to have evolved independently. In Iberian leaf beetles, the existence of a significant genetic-spatial distance relationship seems to be related with the dispersal ability of species. Thus, the Gompertz function is a good alternative to fit and study the genetic-spatial distances relationship pattern.

4.1 INTRODUCTION

Understanding the spatial structure of biodiversity and its driving processes is a main goal in biogeography and macroecology. A widely used approach is to model how biodiversity changes along spatial or environmental gradients, hence using mathematical functions to describe the

relationship between a biological attribute and spatial/abiotic variables (e.g., Connor & McCoy, 1979; Hillebrand, 2004; Nekola & McGill, 2014). At the community level, one of the most studied biodiversity patterns is the decay of community similarity with spatial distance i.e., distance-decay (Nekola & White, 1999), which is usually modelled using negative exponential or power-law functions (Nekola & McGill, 2014). Although these models traditionally describe the decay of community similarity with spatial distance, they can also be expressed as the

inverse pattern: an increase of community dissimilarity with spatial distance, because dissimilarity is just $1 - \text{similarity}$. An analogous pattern arises at the genetic level, given that genetic distance between populations tends to increase with the spatial distance, the so-called Isolation-by-distance (Hutchison & Templeton, 1999; Wright, 1943). This pattern is the result of dispersal limitation (Slatkin, 1993), as it reduces gene flow between populations as a function of distance. However, as shown by Hutchison and Templeton (1999), the genetic-spatial distance relationship should not be expected to be linear, as it can also take asymptotic shapes depending on the relative relevance of gene flow and drift. Therefore, modelling approaches should account for the variety of functional forms that can provide the best model fit of this relationship.

The genetic-spatial distance relationship can take different shapes depending on the degree of gene flow and genetic drift (Hutchison & Templeton, 1999). Without any restriction to gene flow (i.e., no dispersal limitation), genetic distance depends on drift and is independent of spatial distance, so there is no significant relationship between them. Under dispersal limitation, genetic distance increases with spatial distance as gene flow decreases with spatial distance, a pattern that can be modelled with a linear function. However, when the spatial extent is large, we may observe an asymptotic relationship, with an increase of genetic distance along the shortest distances coupled with a flat relationship at the largest distances, where populations are too far apart for gene flow to be maintained (Hutchison & Templeton, 1999). Therefore, at these large spatial distances, genetic distance is not driven by gene flow. In this scenario, the relationship between genetic and spatial distance is thus better fitted with asymptotic functions such as power-law or negative exponential, rather than linear models. It should be noted that all these scenarios assume an increase of genetic distance from the shortest distances. However, it is also possible that high levels of gene flow at the shortest distances may also lead to a lack of relationship between genetic and spatial distance. Thus, the spatial-genetic relationship may potentially take a sigmoidal form, with two asymptotes instead of one and, in this case, neither power-law nor negative exponential functions can properly fit the form of the curves. Aiming to accommodate the potential varying functional forms, the same type of models used to fit distance-decay patterns have also been used to model the genetic-spatial distance relationship (e.g., Gómez-Rodríguez et al., 2020). However, to our knowledge, no study has attempted to model a potential sigmoidal relationship, such as the Gompertz function, which has proven a highly versatile and useful functional form to model distance-decay at the community level (Martín-Devasa et al., 2022).

The assessment of the genetic-spatial distance relationship with different parametric functions allows testing hypothesis about the relevance of gene flow and drift in the spatial

genetic structure (Gómez-Rodríguez et al., 2020). In addition, the parameters of the function used to fit the data (intercept and slope in most cases) can be used to compare such structure among species and, more importantly, can be interpreted in biological terms (Rousset, 1997). Intercepts inform about genetic difference at the shortest distance, whereas slopes are the rate at which genetic difference increases with spatial distance, which is directly related to gene flow and thus to the dispersal ability of the species (Chust et al., 2016; Gómez-Rodríguez et al., 2020; Lester et al., 2007). For example, in a study of marine species Chust et al., (2016) found significative higher IBD slopes (and distance-decay slopes) in species with poor dispersal ability than in species with high dispersal ability.

Genetic-spatial distance models and distance-decay models have the same basic structure: a variable with pairwise dependence (similarity or dissimilarity between sites) with values between 0 and 1 is regressed over spatial distances separating those sites, which is also a pairwise measure itself. Given this analogy, mathematical tools developed to model distance-decay can be also useful to study the genetic-spatial distance relationship. Recently, in addition to the negative exponential and power-law functions, the Gompertz function (Gompertz, 1825) was introduced as an alternative to fit distance-decay patterns (Martín-Devasa et al., 2022a). The Gompertz function is a sigmoidal function (Godeau et al., 2020) normally used to study growing patterns in fields such as microbiology or cytology (Baker et al., 1975; Booth, 1984; Çelekli et al., 2008). In the case of distance-decay models, it captures patterns where similarities remain constant over short distances and start to decay at medium or large distances (Martín-Devasa et al., 2022). Thus, a modified Gompertz function needs to be defined between 0 and 1 given that similarity only takes values between 0 and 1 (Martín-Devasa et al., 2022a). In an the genetic-spatial distance model, the most interesting parameters are slope, which can be related to species dispersal ability (Gómez-Rodríguez et al., 2020), and the asymptote, related to the absence of gene flow at certain spatial distances (Hutchison & Templeton, 1999). Both parameters are defined in the Gompertz function in addition to a position parameter that controls the length of the initial plateau of constant low genetic distances.

Beetles are a good model group for ecological studies because of their variety in ecological characteristics, distribution sizes and dispersal abilities (Fountain-Jones et al., 2015). In the same way that varying dispersal abilities and ecological characteristics are related to contrasting distance-decay patterns at the community level (Gómez-Rodríguez & Baselga 2018), we expect that different beetle species will show different spatial structures of intraspecific genetic variation and thus different shapes in the genetic-spatial distance relationship. Previous studies have assessed the spatial genetic structure in European beetles, finding a positive genetic-spatial distance relationship in some cases (Desender & Verdyck, 2001; Garnier et al., 2004; Oleksa et al., 2013), although a study of ca. 400 species in Central Europe found that only a small fraction of the species showed a significant relationship (Ottati et al., 2022). In the cited study, the main factors shaping the populations genetic structure were dispersal ability and historical factors such as Pleistocene glaciations, that could result in independent evolution of populations in separated refugia, in combination with the species ecological traits.

In this work, we aim to assess the variation in shape of the genetic-spatial distance relationship, using Iberian endemic beetles as case-study. With this aim we compare the performance of the linear, negative exponential, power-law and Gompertz functions. Due to its sigmoidal form and double asymptote, we expect the Gompertz function to perform especially well for genetic-spatial distance relationships with asymptotic shapes, in which dissimilarities and spatial distances are independent at a certain distance due to the absence of gene flow, and when dissimilarities do not increase at short distances due to a high gene flow between populations. In addition, we performed a comparison of parameters between species with significant genetic-spatial distance relationships to assess the differences in the rate of increase of genetic differences with spatial distance between them. We also built haplotype networks to assess the spatial structure of genetic variation in each beetle species studied and detect potential phylogenetic breaks or clusters. We expect that some species may show strong genetic-distance decay patterns given that the Iberian Peninsula contained multiple refugia at small scale (Gómez & Lunt, 2007; Schmitt, 2007). For beetles, the presence of different glacial refugia in the Iberian Peninsula has already been shown studying the *Lachnaia tristigma* (Lacordaire, 1848) (Coleoptera: Chrysomelidae) species-group (*L. tristigma*, *L. pseudobarathraea* (Daniel & Daniel, 1898) and *L. gallaeca* Baselga & R.García, 2007) (Martín-Devasa & Baselga, 2018).

4.2 MATERIAL AND METHODS

4.2.1 Genetic and geographic data

Sequences of the barcode fragment (i.e. *cox1-5'*) for 17 Iberian endemic leaf beetle species were downloaded from GenBank (species list and accession numbers in Table S4.1 in Appendix 4.1). We excluded sequences without geographic reference. All remaining sequences belong to individuals collected for a previous study on leaf beetle communities (Baselga et al., 2015), from which geographical references of the sampling localities were also obtained. Species with sequences in less than three sites or with less than 15 sequences were not considered. Our final set of species includes 10 species: *Calomicrus suturalis* (Joannis, 1866), *Cryptocephalus cantabricus* Franz, 1958, *Gastrophysa janthina* (Suffrian, 1851), *Gonioctena leprieuri* (Pic, 1911), *Lachnaia gallaeca*, *Lachnaia tristigma*, *Luperus sulphuripes* Graells, 1858, *Oedionychus cinctus* (Fabricius, 1781), *Psylliodes gougeleti* Allard, 1859 and *Smaragdina reyi* (Brisout, 1866). To assess the relationship between genetic distance between populations and spatial distance, sequences were aligned with transAlign (Bininda-Emonds, 2005), with the final alignments having a length between 654 and 655 positions, depending on the species. As various individuals can have the same *cox1-5'* sequence, identical sequences were collapsed into a single one (unique haplotypes) using the *haplotype* function of the package “pegas” (Paradis, 2010). With these haplotypes, each sequence is represented by only one copy in the sites where present. For each species, genetic distances were computed as the proportion of different nucleotide positions between haplotypes, using the function *dist.alignment* of the package “seqinr” (Charif & Lobry, 2007). Spatial distances between localities were computed as the geodesic distances between their geographical coordinates (latitude and longitude) using

the function *geodist* of the package “geodist” (Padgham & Sumner, 2021). To assess whether dispersal ability influenced the genetic-spatial distance relationship, we classified species into two categories according to the presence or absence of wings (Table 4.2), as this trait is directly related to dispersal ability.

4.2.2 Model fitting and parameters comparison

To model the genetic-spatial distance relationship for each species, we used linear regression to fit linear models, and non-linear regressions to fit power-law, negative exponential and Gompertz functions. Non-linear models were fitted with the *nlsLM* function of package “minpack.lm” (Elzhov et al., 2016). Model performance was evaluated with AIC and goodness of fit (pseudo-r²). Model significance was assessed using a permutation test (Martínez-Santalla et al., 2022). To assess the differences in the spatial genetic structure among species, we tested the equality in the rate of increase of genetic distance with spatial distance between them comparing the slopes of the best-fitting models using the z_{dep} statistic (Martín-Devasa et al., 2022b). The z_{dep} statistic was designed for parameter comparison in models fitted with pairwise dependent data, as (dis)similarity index and spatial distances. This parameter comparison only considered species in which the genetic-spatial distance relationship was significant. To assess the role of dispersal ability affecting the pattern, we evaluated, descriptively, if significant genetic-spatial distance relationships are more related with apterous species (low dispersal ability) or not. Additionally, to graphically represent the spatial pattern of genetic variation in Iberian beetles, we built a haplotype network for each species using the function *haploNet* of the package “pegas” (Paradis, 2010).

4.3 RESULTS

We compiled 446 *cox1*-5' sequences of the 10 species studied. In the species alignments, the percentage of sequences covering all positions varied between 86% and 100%. In 5 species out of 10 there was no significant relationship between genetic and spatial distance ($p > 0.05$, Table 4.2) (Fig. 4.1a, b, c, e, j). In the remaining five species, we found a positive relationship between spatial distance and genetic distances ($p < 0.05$, Table 4.1) (Fig. 4.1a, b, c, e, j) that was captured by all the models fitted (linear, power-law, negative exponential and Gompertz). Based on the AIC, the power-law showed the best performance in one species (*L. tristigma*), and the Gompertz function showed the best performance in two species (*L. sulphuripes* and *P. gougeleti*). The Gompertz was as good as the linear and negative exponential functions ($\Delta AIC < 2$) in another two species (*G. janthina* and *O. cinctus*). In all cases, the Gompertz function (pseudo-r² between 0.05 and 0.88, Table 4.1), presented equal or better goodness of fit than the negative exponential (pseudo-r² between 0.05 and 0.78, Table 4.1) which also showed good performance. Therefore, to compare parameters of genetic-spatial distance models of different species we used Gompertz models for all of them.

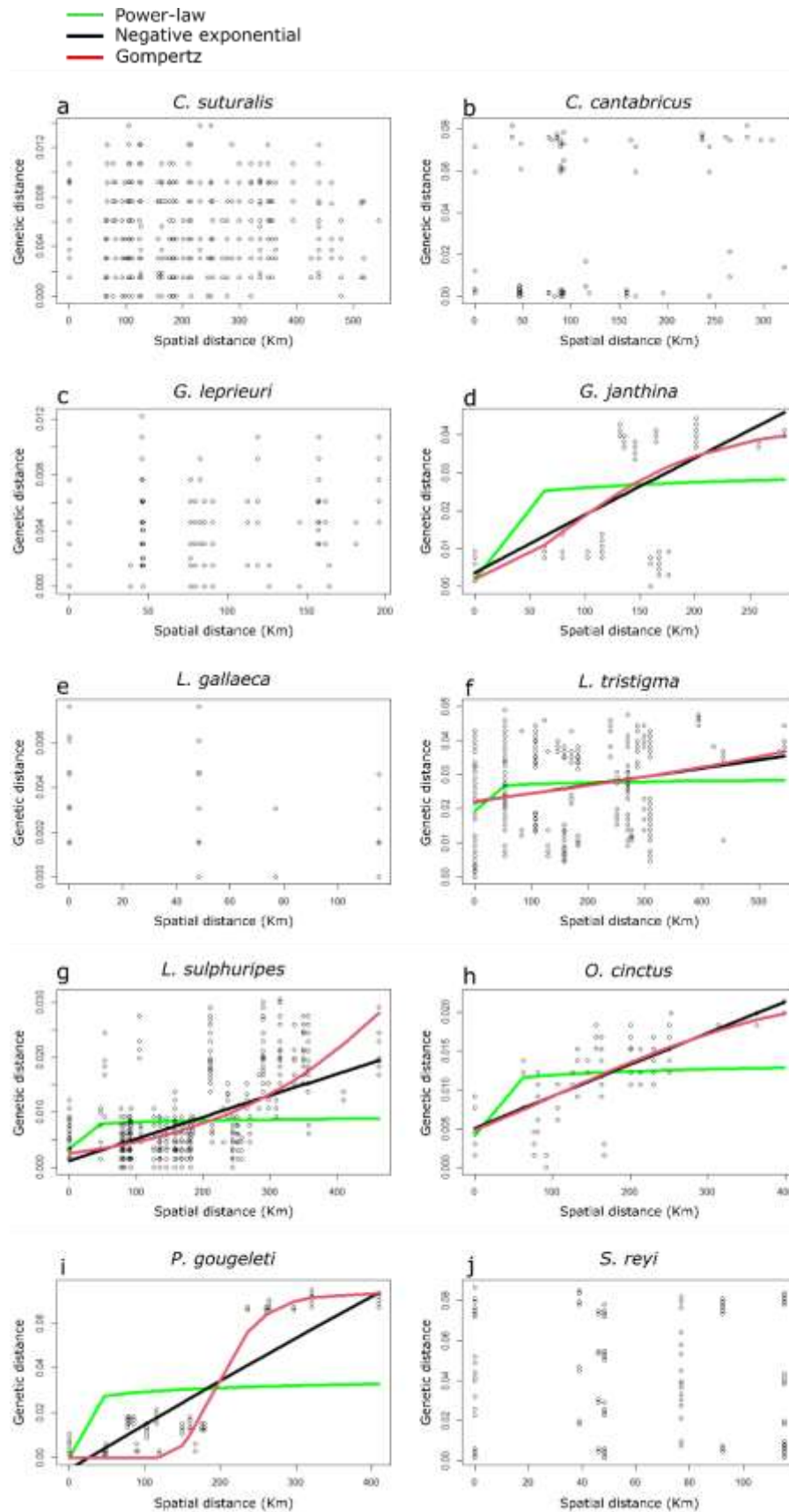


Figure 4.1. Relationship between genetic and spatial distance for all the Iberian beetle species considered in this study. Fitting curves are only showed for when significant ($p < 0.05$).

Table 4.1. AIC, pseudo-r² and p-value for the linear, power-law, negative exponential and Gompertz functions used to model the genetic-spatial distance relationship in each Iberian beetle species. Best fitting models in cases when genetic-spatial distance relationship was significative are shown in bold.

	Model	AIC	Pseudo-r ²	p-value
<i>C. suturalis</i>	Linear	-6209.9	<0.01	0.19
	Power-law	-6207.9	<0.01	0.46
	Exponential	-6209.9	<0.01	0.42
	Gompertz	-6208.2	<0.01	0.44
<i>C. cantabricus</i>	Linear	-356.7	0.09	0.08
	Power-law	-350.9	0.03	0.13
	Exponential	-356.7	0.09	0.10
	Gompertz	-354.8	0.09	0.09
<i>G. leprieuri</i>	Linear	-1498.6	0.05	0.10
	Power-law	-1492.9	0.02	0.18
	Exponential	-1498.6	0.05	0.19
	Gompertz	-1497.5	0.05	0.12
<i>G. janthina</i>	Linear	-451.3	0.39	0.01
	Power-law	-432.8	0.23	0.01
	Exponential	-451.4	0.39	0.01
	Gompertz	-450.8	0.40	0.01
<i>L. gallaeca</i>	Linear	-1740.0	0.09	0.92
	Power-law	-1732.3	0.04	0.27
	Exponential	-1740.0	0.09	0.13
	Gompertz	-1740.5	0.10	0.15
<i>L. tristigma</i>	Linear	-2898.8	0.05	0.01
	Power-law	-2901.7	0.06	0.01
	Exponential	-2898.8	0.05	0.01
	Gompertz	-2897.0	0.05	0.01
<i>L. sulphuripes</i>	Linear	-5966.9	0.40	0.01
	Power-law	-5627.7	0.06	0.01
	Exponential	-5965.7	0.39	0.01
	Gompertz	-6066.6	0.46	0.01
<i>O. cinctus</i>	Linear	-761.2	0.53	0.01
	Power-law	-719.2	0.26	0.01
	Exponential	-761.3	0.53	0.01
	Gompertz	-759.9	0.53	0.01
<i>P. gougeleti</i>	Linear	-1256.2	0.79	0.01
	Power-law	-975.2	0.18	0.01
	Exponential	-1252.9	0.78	0.01
	Gompertz	-1390.8	0.88	0.01
<i>S. reyi</i>	Linear	-498.6	<0.01	0.31
	Power-law	-498.6	<0.01	0.87
	Exponential	-498.6	<0.01	0.87
	Gompertz	-496.6	<0.01	0.93

Although the low sample size precludes a formal statistical analysis, our results showed that the genetic-spatial distance relationship was significant in the three apterous species, but not significant in 5 out 7 winged species (Table 4.2). Although the Gompertz function was the one showing the better fit to the data (Table 4.1), the shape and strength of this relationship differed among species. Some species, such as *L. sulphuripes* or *O. cinctus*, showed a gradual increase of the genetic distance with spatial distance (with asymptote) (Fig. 4.1f, g, h). Others, such as *G. janthina* or *P. gougeleti* showed a break that reveals separated populations having high genetic distances, whereas close populations are genetically homogeneous (Fig. 4.1d, i). In both species and in *L. sulphuripes* it can be observed that genetic distance remains constant at short and medium distances, which, among all the functions considered, only the Gompertz function can accommodate.

Table 4.2. Association between the dispersal ability of species (as inferred from the presence of wings), and the existence of significant genetic-spatial distance relationship.

	Apterous	Significant
<i>C. suturalis</i>	No	No
<i>C. cantabricus</i>	No	No
<i>G. leprieuri</i>	No	No
<i>G. janthina</i>	Yes	Yes
<i>L. gallaeca</i>	No	No
<i>L. tristigma</i>	No	Yes
<i>L. sulphuripes</i>	No	Yes
<i>O. cinctus</i>	Yes	Yes
<i>P. gougeleti</i>	Yes	Yes
<i>S. reyi</i>	No	No

The haplotype networks revealed that different spatial structures of genetic diversity arise depending on whether the genetic-spatial distance relationship is significant or not. The species with significant genetic-spatial distance relationships showed a marked spatial structure of haplotypes in which most haplotypes are exclusive of a single locality (Fig. 4.2a). On the contrary, haplotype networks of species without significant genetic-spatial distance relationship showed weak spatial genetic structure, with the same haplotype occurring in several localities (Fig. 4.2b). The test of the equality of parameters showed that all differences in slope were significant (p -value < 0.05) except the one between *G. janthina* and *P. gougeleti* (p -value = 0.56). The species with higher slope was *P. gougeleti* (0.026) followed by *G. janthina* (0.014), *O. cinctus* (0.005), *L. sulphuripes* (7.98×10^{-4}) and the steepest slope was the one of *L. tristigma* (9.13×10^{-5}).

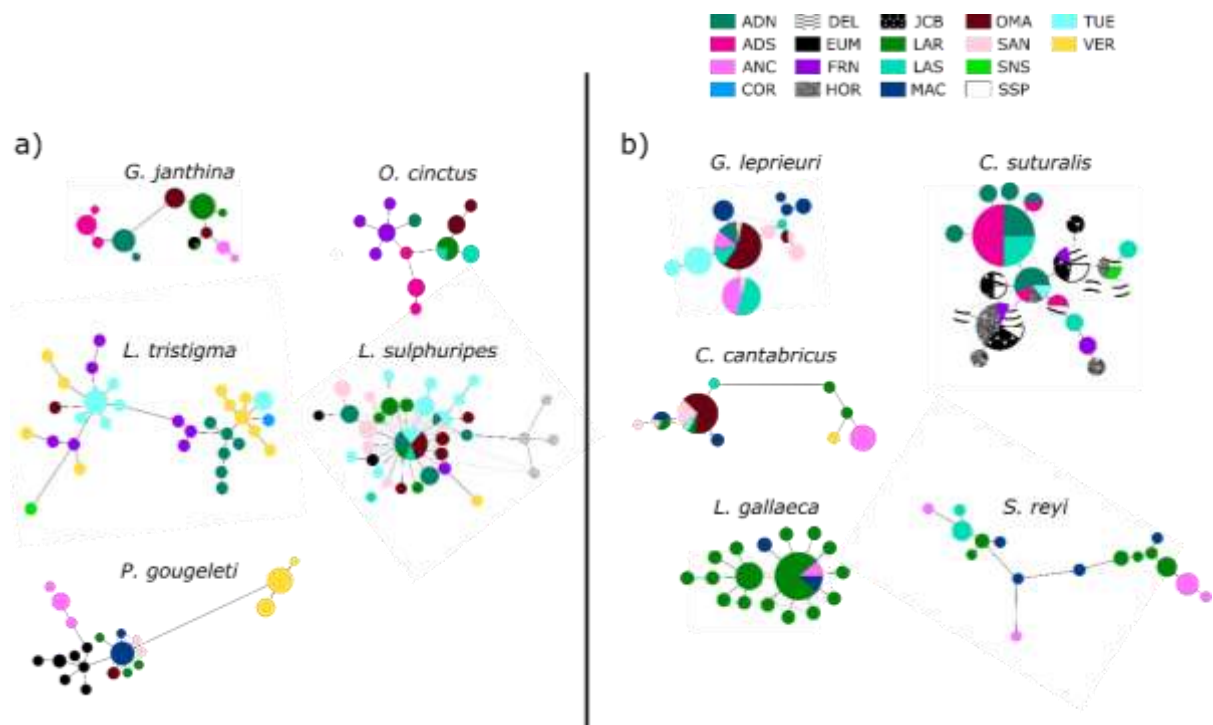


Figure 4.2. Haplotype networks for the 10 species of Iberian leaf beetles considered in this study: a) Species with a significant genetic-spatial distance relationship. b) Species with no significant genetic-spatial distance.

4.4 DISCUSSION

Our results evidenced large differences in the shape of the genetic-spatial distance relationship in Iberian leaf beetles, from a lack of relationship to positive relationships that ranged from linear to asymptotic. This variety of shapes evidenced the need for a complete toolbox of alternative functions that may capture these relationships (i.e. linear, asymptotic, or sigmoidal), given that the shape of the genetic-spatial distance relationship can be used to assess the relative relevance of the causal processes driving the spatial genetic structure (i.e. gene flow vs. drift, see Hutchison & Templeton (1999)). Here we have provided an empirical evaluation of the fit of multiple candidate functions (linear, power-law, negative exponential and Gompertz functions) for spatial patterns of genetic variation and, in particular, we show that the Gompertz function is a versatile option that can accommodate different shapes, as it has been previously reported for other macroecological patterns (i.e. distance-decay, Martín-Devasa et al., 2022a).

One of the main advantages of the Gompertz function is its ability to accommodate different shapes (from almost linear to sigmoidal), as evidenced by the fact that it was consistently selected amongst the models with the best fit to the data for most species. This is particularly relevant if we aim to search for differences in the genetic-spatial relationships between species, because model's parameters can be compared only if they are fitted with the same function. Thus, using a statistical model that fits the data well and that can be applied in most cases allows the inference of differences in the strength of processes driving the spatial pattern of genetic diversity (e.g. weaker dispersers are predicted to have steeper slopes, (Lester

et al., 2007)). This versatility arises because the Gompertz is a generalization of the negative exponential function (Eghwerido et al., 2020). As expected, its sigmoidal form is an advantage compared with the negative exponential and power-law functions as it is capable to accommodate patterns with constant genetic distances at short distances. This initial plateau appears when gen flow is high between close sites but limited or even absent at larger spatial distances. This pattern cannot be accommodated by the linear, negative exponential or power-law functions. In contrast, when the initial plateau does not exist because genetic distance steadily increases from the closest spatial distances, the Gompertz function has shown good performance in all cases where a significant genetic-spatial distance relationship was detected (i.e. isolation-by-distance). In other words, the Gompertz function has shown high versatility in fitting most shapes of the genetic-spatial distance relationship, making it a good alternative to the linear, negative exponential and power-law functions, which have also proven useful to fit the pattern (Gómez-Rodríguez et al., 2020), but have relatively less versatility. For example, in the negative exponential model the superior asymptote is always 1, and thus does not reflect the value at which genetic distance becomes decoupled from spatial distance due to the absence of gene flow. The power-law function can actually fit this asymptote, but it cannot accommodate an initial plateau of constant low genetic distance due to high gene flow. Finally, the linear model cannot accommodate either an initial plateau or the superior asymptote.

In Iberian endemic leaf beetles, a significant genetic-spatial distance relationship (i.e., isolation-by-distance) was observed in the 3 apterous species, but only in 2 out of 7 winged species. This suggests that dispersal ability is an important factor shaping the spatial genetic structure of Iberian endemic leaf beetles. These results are in line with previous studies that also pointed to dispersal limitation as an important factor causing isolation-by-distance in other European and Iberian beetles (Garnier et al., 2004; Oleksa et al., 2013; Rizzo et al., 2017). However, in most cases, dispersal ability is not the only factor shaping the genetic structure of species. Historical biogeographical events, especially Pleistocene glaciations, also play an important role (Hewitt, 1996), with post-glacial colonization events from different glacial refugia having likely affected the relationship between spatial and genetic distances in some beetle species (Ottati et al., 2022; Samarasekera et al., 2012). Populations isolated in different glacial refugia evolve independently during that period, thus accumulating genetic differences between them. During post-glacial colonisation, populations can have a secondary contact, overlapping their distributions (Widmer & Lexer, 2001). In this case, no significant relationship between genetic and spatial distance can be detected as genetic variance is homogeneous in space, but different genetic groups may be detected coexisting in the same localities (e.g., Austin, 2002; Conroy & Cook, 2000; Taberlet et al., 1995). This could be the case of *C. cantabricus*, and is evidenced by the two horizontal groups of points in the genetic to spatial distance plot (Fig. 1b), with a clear gap in genetic distances (but not in spatial distances). A similar pattern is observed in *L. tristigma* and *S. reyi*. Although having a very weak relationship between genetic and spatial distance, the haplotype networks reveal two different haplotype groups, but without a clear spatial configuration. The existence of several glacial refugia for the *L. tristigma* species-group in the Iberian Peninsula (Martín-Devasa & Baselga, 2018) supports this hypothesis. Coincidentally, the three species (*L. tristigma*, *S. reyi* and *C. cantabricus*) have

the highest intraspecific genetic differences among all species studied. To this respect, the comparisons between *L. galleca* and *L. tristigma* are particularly interesting because although being phylogenetically close present disjoint distributions (Baselga & Ruiz-García, 2007) and different genetic patterns, which causes may deserve a particular study.

The gaps in genetic distances observed in *G. janthina* and *P. gougeleti* can be probably attributed to the same cause, i.e., two populations (showed in the haplotype networks) evolving independently in different glacial refugia. However, in contrast to the previous species, *G. janthina* and *P. gougeleti* are apterous species, so they are probably more dispersal limited. This would have caused slow range expansion during post-glacial (Svenning & Skov, 2007), and therefore the isolation of populations for a longer period of time. In this scenario, populations that are far apart are expected to have high genetic distances, whereas close populations would be genetically more homogeneous (Peterson & Denno, 1998). In this case, a significant relationship between genetic and spatial distances is detected, and that particular shape is well fitted by the Gompertz function. The absence of significant differences between both curves' slopes, also pointed to a congruent biogeographical history, compatible with stability periods (Marske et al., 2020). In the remaining species, the scenario seems to be simpler. In *C. suturalis*, *G. leprieuri* and *L. gallaeca* there is no spatial structure of genetic variation nor clues of separated glacial refugia. In *L. sulphuripes* and *O. cinctus* the absence of gaps in the genetic to spatial distance plots or clear clusters in haplotype networks point to an isolation-by-distance scenario, in which dispersal limitation would be the main factor behind the significant relationship between genetic and spatial distance in these species.

In conclusion, the Gompertz completes the set of functions needed to accommodate the variety of potential shapes of the genetic-spatial distance relationship. The application of Gompertz models allows identifying scenarios in which a significant relationship between genetic and spatial distance exists, but derives from the existence of historically isolated populations with limited range expansion, and not from a continuous effect of limited dispersal in which the amount of gene flow depends on spatial distance. Moreover, the great versatility to fit well multiple shapes makes it a good candidate when a single function must be used for different species in order to compare the function's parameters. Our case study on the relationship between genetic and spatial distances in Iberian endemic leaf beetles points to the importance of dispersal ability and long term isolation associated to historical events in shaping the spatial genetic structure of this group. In general, its theoretical and practical advantages make the Gompertz function a great alternative to consider in studies of the spatial genetic structure across species.

Chapter V

WHAT CAUSED THE DISJUNCT DISTRIBUTIONS OF THE *LACHNAIA* *TRISTIGMA* SPECIES-GROUP (COLEOPTERA: CHRYSOMELIDAE) ON THE IBERIAN PENINSULA?



WHAT CAUSED THE DISJUNCT DISTRIBUTIONS OF THE *LACHNAIA TRISTIGMA* SPECIES-GROUP (COLEOPTERA: CHRYSOMELIDAE) ON THE IBERIAN PENINSULA?

ABSTRACT

Non-overlapping distributions of closely related species can be explained by two mechanisms, either niche divergence or vicariance in the absence of niche differentiation. We assess which of these mechanisms is the most likely cause of the disjunct distributions of three closely related species of *Lachnaia* Dejean, 1836 (Coleoptera: Chrysomelidae) on the Iberian Peninsula. We estimated the three species' climatic niches, compared them using multivariate analysis, and projected their potential distributions under climatic conditions during the Last Glacial Maximum (LGM) around 20,000 years ago. *Lachnaia tristigma* (Lacordaire, 1848) and *Lachnaia pseudobarathraea* (Daniel & Daniel, 1898) hardly differ in their climatic niches. The temperature ranges are similar among the three species, but *Lachnaia gallaeca* Baselga & Ruíz-García (2007) occupies a fraction of the environmental space with higher precipitation. The three climatic niches present a broad overlap, which does not suggest speciation events driven by niche differentiation. In turn, the projection of their climatic envelopes to the LGM conditions shows that their potential distributions are compatible with the existence of different glacial refugia for each species. Therefore, we hypothesize that the contraction of species ranges during LGM and subsequent vicariance could have been a major mechanism behind the disjunct distributions that we observe in the present.

5.1 INTRODUCTION

The distribution range of a species can be defined as the portion of geographical space in which one species is present and interacts in a non-ephemeral way with the ecosystem (Zunino & Zullini, 2003). Knowledge of the distribution range of species and the factors that determine their shape and size is an essential tool for assessing species invasions, assessing the impact of climate, land use, and other environmental changes on species distributions, supporting conservation planning and reserve selection, and even testing biogeographical, ecological, and evolutionary hypotheses (Guisan & Thuiller, 2005).

One topic intersecting the aforementioned biogeographical, ecological, and evolutionary processes is the analysis of the factors that lead sister species to have disjunct distributions. There are several mechanisms that can lead to closely related species to have non-overlapping

distributions. One of them is that closely related species can have different ecological niches (Leibold & McPeck, 2006). This would imply that one species could not live in the same territories as the others because the conditions in that territory are not appropriate for their life and development. In consequence, the distribution range of the species cannot overlap. Alternatively, speciation events can occur even in the absence of niche differentiation via vicariance (Wiley & Lieberman, 2011). In this scenario, one species occupies a territory and a barrier appears, so that the distribution range splits and two or more populations are now isolated. This causes independent evolution of the separated populations that eventually leads to allopatric speciation, but if selective forces are not operating on ecological niches of the vicariant populations, these may remain identical (Wiens & Graham, 2005).

In this study, we aim to assess whether the disjunct distribution ranges of three closely related species of *Lachnaia* Dejean, 1836 (Coleoptera: Chrysomelidae) are associated with different climatic niches or not. The first scenario would point to a process of niche differentiation during or after speciation, while the second case would suggest a process of allopatric speciation in the absence of selection on the climatic niche of the new species. The genus *Lachnaia* is distributed in the Western Palaearctic region, especially in the western Mediterranean area (Iberian Peninsula and France), but a few species reach the northern Afrotropical region (Codina Padilla, 1958). Eight species of *Lachnaia* occur on the Iberian Peninsula (Petitpierre, 2000). Three of them are included in the *Lachnaia tristigma* species-group: *Lachnaia gallaeca* Baselga and Ruiz-García, *Lachnaia pseudobarathraea* Daniel and Daniel, and *Lachnaia tristigma* (Lacordaire). These three species have extremely similar external morphologies and can only be diagnosed by the size and shape of the median lobe

of the aedeagus, the size of the sclerites of the endophallus, and the size of the spermathecal duct (Baselga & Ruiz-García, 2007).

5.2 MATERIAL AND METHODS

This study is primarily based on the data published by Baselga & Ruiz-García (2007), who reviewed the collection of the Museo Nacional de Ciencias Naturales, Madrid, Spain and established the current understanding of species delimitations based on male and female genitalia. In addition, multiple leaf beetle assemblages were sampled across the Iberian Peninsula in a subsequent study (Baselga et al., 2015), rendering new records of *L. gallaeca* and *L. tristigma*. The sex of the collected specimens and the coordinates of their localities (Military Grid Reference System format) were recorded. Detailed information on the procedure can be found in Baselga & Ruiz-García (2007). As a result, our dataset included the taxonomic and geographical information of 292 specimens belonging to the three species of the *Lachnaia tristigma* species-group.

To reduce the geographical biases, we conducted a bibliographical review to gather additional records of *L. tristigma* in Spain. Thus, we made two estimates of the ecological niches, one with no taxonomic uncertainty but some geographical biases (hereinafter, core dataset) and another more spatially complete but with some degree of taxonomic uncertainty

due to potential misidentifications of *L. tristigma* and *L. pseudobarathraea* (hereinafter, extended dataset).

Maps of the Iberian Peninsula were created in R (R Development Core Team 2016) using a combination of packages “maps” (Becker, Wilks, Brownrigg, et al., 2016), “mapdata” (Becker et al., 2016), “maptools” (Bivand & Lewin-Koh, 2017), and “scales” (Wickham, 2016). Subsequently, the points where the different species were found (core and extended datasets) were superimposed using the package “rgdal” (Bivand et al., 2016). We estimated the climatic niche for the three species as the n-dimensional hyper-volume (n = number of variables considered) that each species occupies in the climatic hyperspace. To do this, bioclimatic data for the Iberian Peninsula were downloaded from the website wordclim.org (Hijmans et al., 2017). Next, the information of six bioclimatic factors was obtained for all points where each species was recorded in the core and extended datasets: average annual temperature; maximum temperature of the warmest month; minimum temperature of the coldest month; annual precipitation; precipitation of the wettest quarter; and precipitation of the driest quarter. These six variables were submitted to a Principal Components Analysis (PCA) with varimax rotation (function *principal* in package “psych”, Revelle, 2017) in order to represent the climatic niches in two dimensions instead of six, allowing a simpler representation of the degree of overlap of species climatic niches.

Finally, we performed two envelope models for the climatic niche of the three species, using the functions *bioclim* and *domain* (package “dismo”, Hijmans et al., 2017). Both models, Bioclim (Busby, 1991) and Domain (Carpenter et al., 1993), are strictly a presence-only method, thus allowing us to approximate the potential distributions of clades rather than their realized distributions (Jiménez-Valverde et al., 2008). To do this, the probabilistic model outputs were converted into presence-absence values using as the cut-off the quantile 0.05 of the presences, projecting their potential distributions under the climatic conditions of Last Glacial Maximum (LGM, around 20,000 years ago) (Braconnot et al., 2007).

5.3 RESULTS

The geographic distributions of *L. gallaeca*, *L. pseudobarathraea*, and *L. tristigma* on the Iberian Peninsula based on the core dataset (no taxonomic uncertainty, but some spatial biases) are shown in Fig. 5.1a. The distribution ranges based on the extended dataset (higher spatial completeness, some degree of taxonomic uncertainty) is shown in Fig. 5.1b. The known distribution ranges jointly occupy almost all of the Iberian Peninsula. *Lachnaia gallaeca* is located exclusively in the inland mountains of Galicia (northwestern Spain), occupying a narrow range in the provinces of Lugo and Ourense. Some localities are extremely close to the Portuguese border, although no specimen has been yet collected in Portugal. *Lachnaia pseudobarathraea* is located in the southeastern portion of the Iberian Peninsula, in an area larger than that occupied by *L. gallaeca*. Its distribution is spindle-shaped, flanked to the south by the Mediterranean Sea and to the north by the Subbaetic and Prebaetic mountain ranges. *Lachnaia tristigma* is distributed throughout the central part of the Iberian Peninsula, reaching

the Atlantic coast in the South-west (including southern Portugal) and entering France across the Pyrenees in the Northeast.

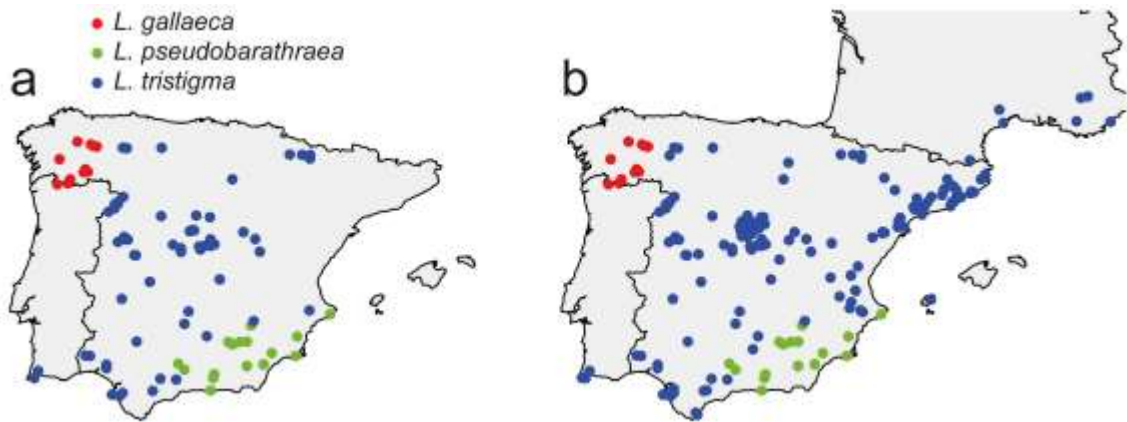


Figure 5.1. Known records for *Lachnaia gallaeca* (red dots), *Lachnaia pseudobarathraea* (green dots), and *Lachnaia tristigma* (blue dots), using the core (a) and extended (b) datasets.

The two axes of the climatic PCA explained 93% of the variance in the core dataset. For the extended dataset, the variance explained by two axes was 87%. Therefore, climatic niches are represented in two dimensions (Fig. 5.2), the first one correlated with the precipitation variables ($r = 0.96$ with annual precipitation, $r = 0.98$ with precipitation of the wettest quarter), the second one correlated with the temperature variables ($r = 0.94$ with average annual temperature, $r = 0.96$ with the minimum temperature of the coldest month). *Lachnaia gallaeca* occupies a fraction of the environmental space characterized by higher precipitation values than the other species. However, the temperature ranges are similar among the three species, although *L. pseudobarathraea* and especially *L. tristigma* live in a wider range of temperatures. The climatic niches of *L. tristigma* and *L. pseudobarathraea* widely overlap, although the climatic range occupied by *L. tristigma* is wider (therefore, the niche of *L. pseudobarathraea* is a subset of that of *L. tristigma*). Extreme precipitation values for *L. tristigma* overlap with the precipitation range of *L. gallaeca*, unlike their geographic distributions.

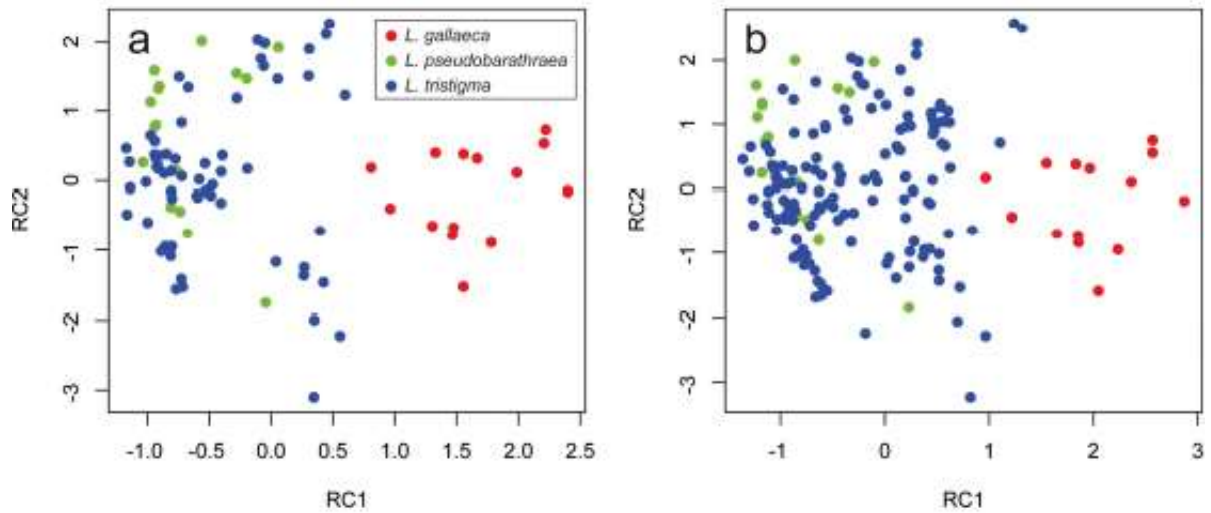


Figure 5.2. Climatic niches of *Lachnaia gallaeca* (red dots), *Lachnaia pseudobarathraea* (green dots), and *Lachnaia tristigma* (blue dots), as represented in two Principal Components Analysis axes (first axis (RC1) is correlated with precipitation; second axis (RC2) is correlated with temperature) for the core (a) and extended (b) datasets.

The potential distributions of the three *Lachnaia* species under LGM climatic conditions are shown in Fig. 5.3. This projection shows that *L. gallaeca* was likely restricted to glacial refuge in western Iberian Peninsula, completely isolated from the other two species. The potential distributions *L. tristigma* and *L. pseudobrathraea* during the LGM overlap in some regions of southern Spain, but the projected distribution range of *L. tristigma* was much wider.

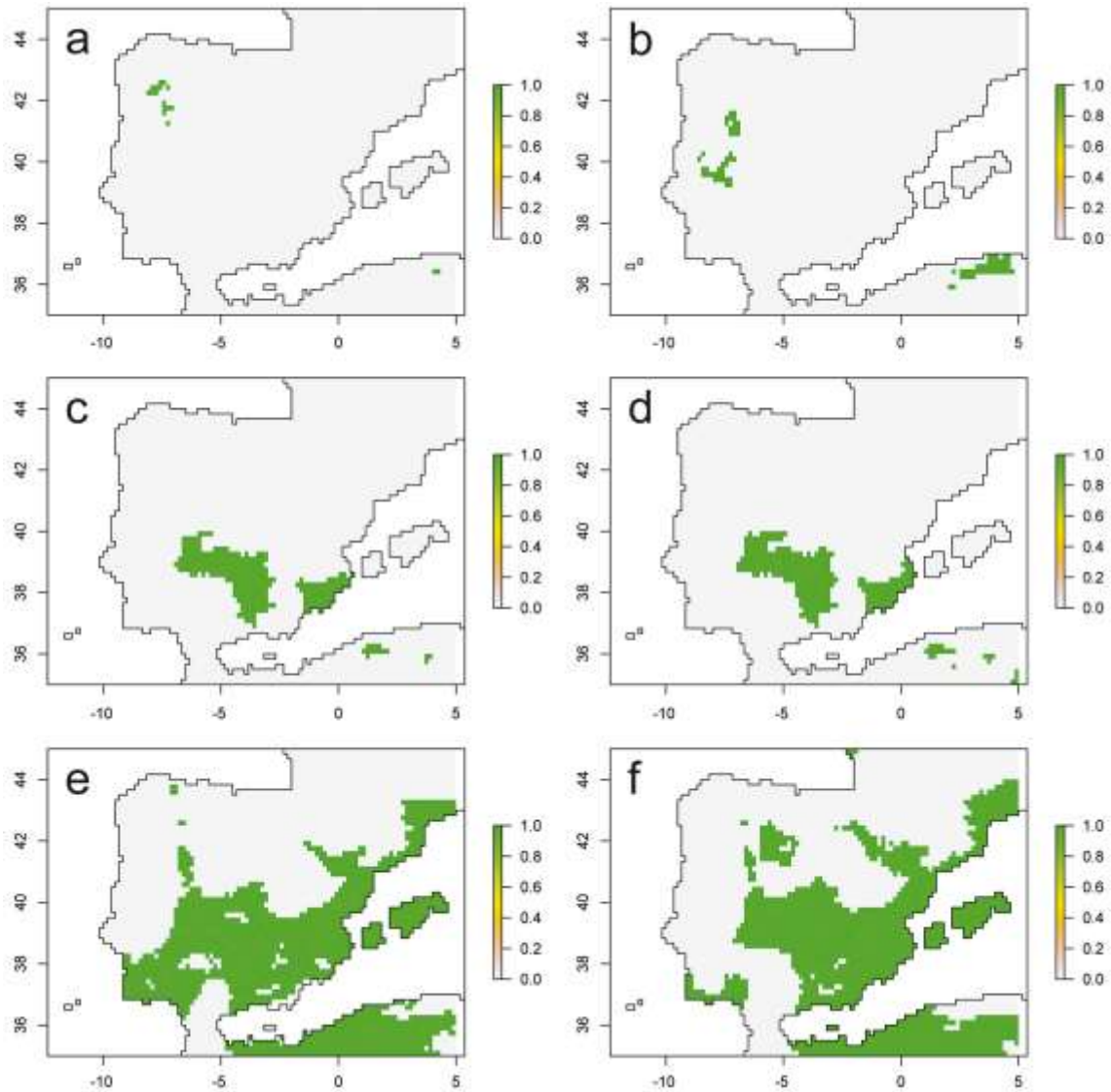


Figure 5.3. Potential distributions of *Lachnaia gallaeca* (a, b), *Lachnaia pseudobarathraea* (c, d), and *Lachnaia tristigma* (e, f) during the Last Glacial Maximum as estimated from Bioclim (a, c, e) and Domain (b, d, f) models.

5.4 DISCUSSION

The geographic ranges of *L. gallaeca*, *L. pseudobarathraea*, and *L. tristigma* occupy distinctly defined regions of the Iberian Peninsula and show no overlap. One hypothesis to explain the three disjunct distributions could be that the species occupy different climatic niches. This would imply that each species could not live in the same territories as the others because the conditions in that territory are not appropriate for their development. Figure 5.2 shows that *L. tristigma* and *L. pseudobarathraea* have very similar climatic niches (the latter being a subset of the former). *Lachnaia gallaeca* occupies a different part of the climatic space, characterized by higher precipitation values (although still with some overlap with *L. tristigma*) but the same range of temperatures as for the other two species. Based on this, the broad overlap among the

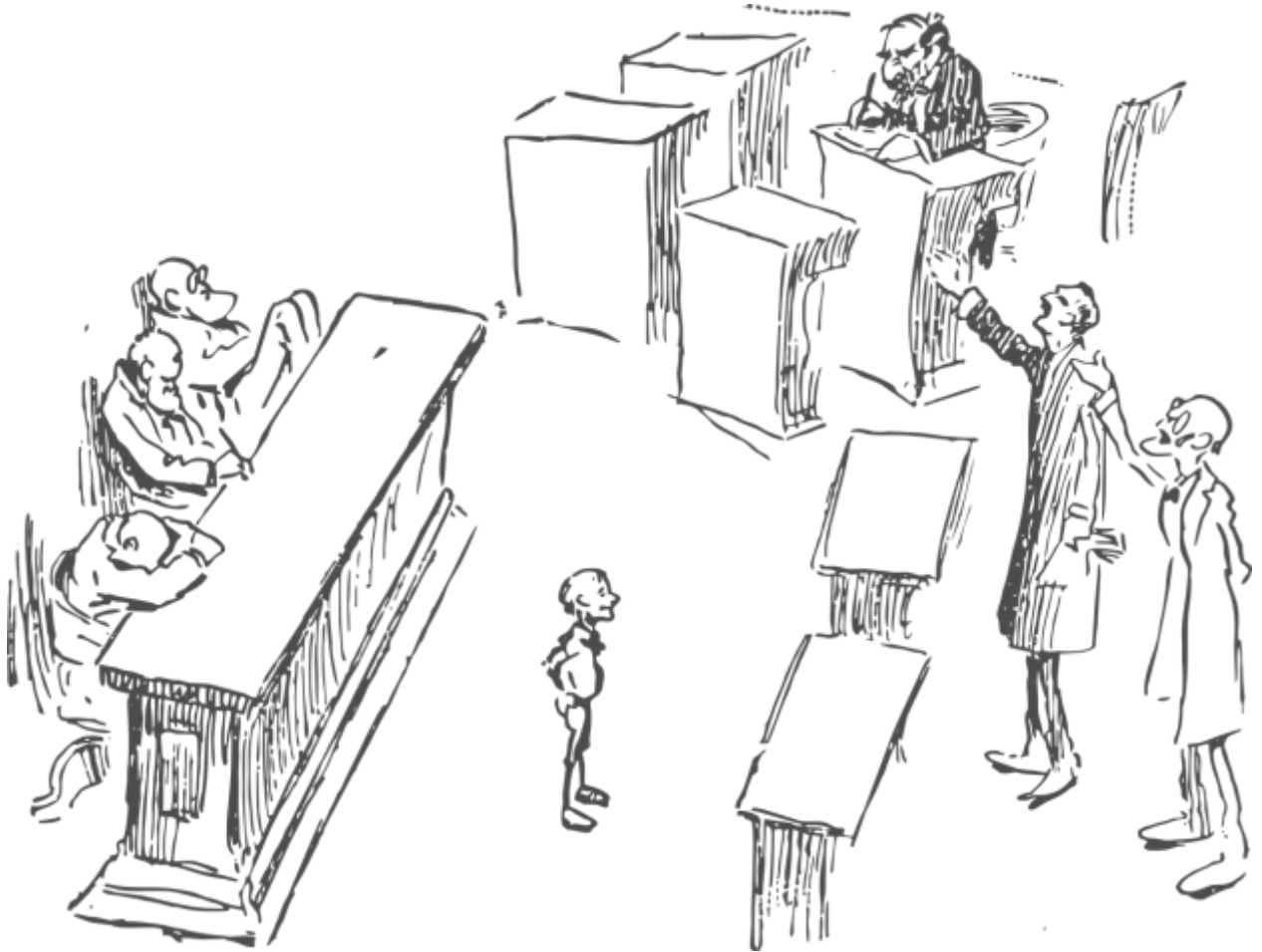
three climatic niches does not support the idea of a speciation event linked to niche differentiation.

The alternative hypothesis is a speciation process via vicariance, with no niche differentiation involved. According to this hypothesis, several populations of an ancestral species would have been isolated in different locations and started a speciation process resulting in the species that can be delimited today. Range fragmentation can be caused by multiple processes of different nature, including topographic barriers. In the Iberian Peninsula, a plausible explanation would be the extreme climatic changes that occurred during the Pleistocene glaciations (from 2.5 million years to 10,000 years ago) (Pisias & Moore, 1981). Those cooling episodes could have caused the contraction of a species' population distributions and, eventually, the fragmentation of ranges in different refugia (Hewitt, 2000). This episode of contraction would be analogous to that described by Wiens (2004) and Baselga et al. (2011), but with fragmentation occurring in the lowlands during cold periods. In the case of the *Lachnaia tristigma* species-group, separate populations of an ancestral species residing on the Iberian Peninsula would have remained in different glacial refugia during these episodes (Provan & Bennett, 2008). Therefore, these populations would have been isolated, and, eventually, they would have undergone a process of allopatric speciation, resulting in the three species that we know now. This hypothesis is supported by our projection of the three potential distributions under LGM climatic conditions (Fig. 5.3), which are compatible with the existence of separated refugia: one in the northwest; another in the southeastern portion of the peninsula; and one in the central plateau or even the northeastern Iberian Mediterranean coast.

For *L. gallaeca*, the potential distribution during the LGM is close to the species' present distribution. Although the overlap is slight, it is plausible that *L. gallaeca* was isolated in northern Portugal during LGM, and the species began to expand its geographic distribution when the climatic conditions became warmer, arriving in northwestern Spain. It is even possible that the species will be detected in Portugal in the future, matching its LGM distribution from where the species could have expanded towards the Galician mountain ranges. The potential distribution of *L. pseudobarathraea* during the LGM occupies part of southern and southeastern Iberian Peninsula. The present distribution range of this species is located in southeastern Spain, broadly overlapping the potential distribution in the LGM. The same occurs for *L. tristigma*; its present distribution and potential distribution in the LGM broadly overlap.

Taken together, these results support the allopatric speciation hypothesis and suggest that the present-day disjunct distributions of *L. gallaeca*, *L. pseudobarathraea*, and *L. tristigma* are largely the consequence of speciation via fragmentation during Pleistocene glacial episodes. Once the last glaciation had retracted and the climatic conditions were favorable for the species, there was no longer a barrier that prevented their dispersion. Thus, each of the three differentiated evolutionary lineages could colonize adjacent territories (Stewart et al., 2009). The present-day distributions do not show evidence of this, as they remain disjunct despite the broad niche overlap, suggesting that dispersal is limited in these species, despite their flying ability.

DISCUSSION



DISCUSSION

Throughout the five chapters of this thesis, new methods for the study of the spatial structure of biological diversity were developed, explored and applied to the study of spiders and beetles biogeography in Europe and the Iberian Peninsula. The first three chapters are mainly focused on the distance-decay pattern. Chapter I assesses how the shape of the pattern changes with different species distribution sizes in a community, and introduces the use of the Gompertz function as an alternative to fit the distance-decay pattern when its shape is sigmoidal. Using simulations, it was revealed that distance-decay shape changes when the mean species range size or the proportion of large distributions increases, so the best fitting function varies from the power-law to the negative exponential and to the Gompertz function. Therefore, the Gompertz function is revealed as a key addition to the toolkit for the analyses of distance-decay patterns. Chapter II introduces a test statistic (z_{dep}) specifically designed to test for the equality of parameters of distance-decay models, or any other models fitted with pairwise dependent data. In this chapter, the good performance of the z_{dep} statistic has been demonstrated with a simulation study, and its usefulness has been exemplified with empirical data. In Chapter III the aforementioned novel methods were used to study of distance-decay patterns of European spiders. The comparison of spatial and climatic distance-decay curves using standardised distances, and the relationship between distance-decay parameters and dispersal traits of spider families revealed dispersal limitation during post-glacial colonisation as the main factor shaping the turnover patterns across continental Europe in spider communities. The last two chapters expand the application of the newly developed methods to alternative fields of community ecology. The new methods are adapted and used for the study of the genetic-spatial distance relationship, and the results obtained show that these models are useful to infer the biogeographic history and the long term dispersal ability of the species. This was exemplified using Iberian endemic leaf beetles as a case-study. In Chapter IV the shape of the genetic-spatial distance relationship of 10 species of Iberian endemic leaf beetles was assessed, by fitting linear, power-law, exponential and Gompertz functions. The Gompertz function showed great plasticity, adapting well in situations in which the pattern was detected, and having a special usefulness in situations in which the pattern had a sigmoidal shape. More importantly, the results of this chapter revealed that characterising the shape of the genetic-spatial distance relationship can be used to assess the relative relevance of the causal processes driving the spatial structure of genetic diversity (i.e. gene flow vs. drift). Among the Iberian endemic species, the shape varied widely, even between phylogenetically close species, as *Lachnaia gallaeca* and *L. tristigma*, suggesting contrasting dispersal abilities and different histories of range expansion. In Chapter V the potential drivers of the distributions of the

L. tristigma species-group were assessed. Independent glacial refugia and limited expansion from them has been revealed as the most probable cause of the current configuration of the disjoint distributions of the three *Lachnaia* species in the *L. tristigma*-group.

Altogether, this thesis is a humble contribution to the methods for the study of biodiversity patterns and to the macroecology of spiders and beetles. From a methodological point of view, this thesis introduces the use of the Gompertz function to the pool of statistical functions used to fit distance-based biogeographic models, and the z_{dep} statistic as a useful tool for parameter comparison between models. From an applied point of view, it uses these methods to deepen the knowledge on the biogeographical patterns of European spiders and the spatial genetic structure of Iberian leaf beetles. The development of novel methods for the study of biogeographical patterns that can be used at different levels of biological organization is one of the highlights of this thesis.

The study of the relationship between species range size and the functional form of the distance-decay pattern, the use of the Gompertz function to fit biogeographical patterns at different levels of biological organization and the proposal and development of the z_{dep} statistic for parameters comparison under pairwise dependence are relevant contributions of this thesis to the methodological approaches for biogeographical and macroecological studies. From a broad perspective, these studies may have theoretical and methodological implications and wide applicability. From a theoretical perspective, it is of particular relevance the identification of species range size as a factor affecting patterns of species replacement with distance and thus the shape of distance-decay relationships. Related to this, it is also especially interesting the fact that biogeographical patterns at different levels of biological organization can take sigmoidal shapes, which is crucial to better understanding and studying the processes that can shape spatial configuration of biodiversity and is key to developing models that can handle that sigmoidal shape, as the Gompertz function. The spatial organization of biological groups at different levels of biological organization use to follow Tobler's first law of geography (Tobler, 1970). In a few words, this law postulates that everything is related to everything, but closer elements, in this case communities or populations, are more related (more similar) than those far away (Miller, 2004; Tobler, 1970). This is the relation on which distance-decay and genetic-spatial distance relationship patterns are based (Kimura & Weiss, 1964; Nekola & White, 1999; Rocchini, 2007; Zhao et al., 2015).

Sigmoidal shapes appear when close elements are very similar or even equal, causing constant high similarities at short distances (little growth), and further elements are very dissimilar or different, causing constant low similarities at large distances, and therefore an asymptote (Godeau et al., 2020). The results in Chapter I related the emergence of a sigmoidal shape in the distance-decay pattern to large species range sizes. It can be theorized that in a population genetics context sigmoidal shapes could be related to large ranges of genetic variants or haplotypes, in both cases relative to the study extent. Large range sizes cause the same species or haplotypes to be present in close localities, so their community or genetic composition is similar (Figure 1.1 illustrates the concept graphically). The sigmoidal shape of

the Gompertz function fits particularly well these situations where similarity does not decay at short distances, which happened when species disperse well and can reach further territories and when in population gene flow can be maintained at a certain distance. Moreover, the Gompertz function is very useful for fitting distance-decay models because it can fit reasonably well all possible shapes (although not being the best model for all of them). The same advantage has been shown for fitting the genetic-spatial distance relationship, which is a considerable advantage. This is other evidence of its usefulness to fit both patterns at different levels of biological organization.

Throughout this thesis (and specially remarked in Chapter IV) the analogies between patterns at the community and population levels have been stressed. The fact that patterns that emerge at both levels of biological organization (genetic-spatial relationship at the population level and distance-decay at the community level) are caused and affected by the same general processes (dispersal, selection, evolution and drift, (Vellend, 2010, 2016), and both have similar responses (Chust et al., 2016) to them is another reason that can justify the use of a common frame. The main advantage of this is, as has been proposed in this thesis, that the same methodological framework can be applied to the study of both, distance-decay and genetic-spatial relationship patterns. This general framework can bring closer the community ecology and genetic population perspectives, at least from a methodological point of view, which facilitates the collaboration between disciplines. Clearly, much more work is needed in order to create a formal general framework of multi-hierarchical biodiversity studies (if that is possible), but the studies contained here could be a little step in the process, in addition to previous ones (Baselga et al., 2013, 2015, 2022; Múrria et al., 2017; Vellend, 2003; Vellend & Geber, 2005). Either way, the analogies between patterns and the possible use of a shared methodology allow to development of common methods for the study of both patterns.

Methodologically, the introduction of the Gompertz function to fit distance-decay and genetic-spatial distance relationships when they take sigmoidal forms, and the z_{dep} statistic to compare the function parameters (both implemented in the latest version of the package “betapart” (version 1.6, Baselga et al., 2023) are contributions that enrich the methodological toolbox for studying biogeographical and macroecological patterns. Since macroecology (and most part biogeography) focuses on the study of patterns (Beck et al., 2012; Keith et al., 2012) the more abundant the tools that can be to characterize and compare the patterns, the most accurate, complete and easy the inference of their causal processes can be. Moreover, knowing the relationship between the shapes of distance-decay patterns and species range, these shapes can be the first insight into the process shaping spatial variation of diversity in a region and the dispersal ability of the studied groups. This is because the range sizes of species and haplotypes distributions mainly depend on the dispersal ability of species and their niche breath (in the case of neutral loci only dispersal processes) (Eckert et al., 2008; Gaston, 1996; Kirkpatrick & Barton, 1997). High dispersal ability and wide niche breath normally imply large range sizes, whereas low dispersal ability and narrow niches tend to cause small ranges (Gaston, 1996; Gaston & Fuller, 2009). Thus, the shape of the distance-decay patterns (or the genetic-spatial distance relationship) could be a first clue of the characteristics of the studied organisms and

the processes causing that particular shape. This is clear for community composition, as has been studied in Chapter I (Martín-Devasa et al., 2022a), and a similar approach can be used for genetic differences between populations, as shown in Chapter IV (expanding the framework of Hutchison & Templeton, 1999).

Regarding its applicability, the most obvious application of these methods is in studies assessing the causes of the spatial structure of the genetic or taxonomic diversity of different groups, as has been done in chapters III and IV with European spiders and Iberian leaf beetles. This is a central question in biogeography and macroecology (Keith et al., 2012; Lomolino et al., 2017), which has been less addressed for invertebrates than for vertebrates or plants. Both studies empirically exemplify the wide applicability of the Gompertz function and the z_{dep} statistic, as it can be applied for distance-decay and genetic-spatial distance relationship patterns (due to its analogies as has been extensively explained along this thesis) and various spatial scales (continental and national). Moreover, results as those obtained in Chapter IV for the genetic-spatial distance relationship patterns of Iberian leaf beetles, assessed with the Gompertz function, open new questions as those of Chapter V (in which the causes of the different distribution ranges of close species are investigated).

Focusing on the inferences enhanced by the novel methods introduced in this thesis, it has been revealed that in both European spiders and Iberian leaf beetles, dispersal limitation and Pleistocene glaciations are the main factors affecting the spatial structure of community composition and genetic variance respectively (Chapters III, IV and V). This result is in concordance with previous studies (e.g., Bonte et al., 2004; Carvalho et al., 2011; Rizzo et al., 2017). This is another example that the same biogeographical processes can influence patterns at different hierarchical levels (Baselga et al., 2013, 2015, 2022; Chust et al., 2016; Vellend, 2010; Vellend & Geber, 2005), which opens the opportunity to assess the application of the proposed methodology in empirical studies. In fact, one of the most immediate observation is that both the Gompertz function and the z_{dep} statistic performed as well as expected in applied studies.

Here, the wide analogies between biogeographical patterns at population and community levels have been explored and a common methodology successfully applied to their study. But is also necessary to take into account some considerations when comparing spatial and environmental effects across hierarchical levels of biological organization. In community ecology, biogeography and macroecology, although some species can respond to dispersal limitation and others be more affected by environmental filtering, the analysis of dispersal and selection effects can be done with the same set of community data, and the fraction of changes in species composition caused by each factor can be assessed using, for example, variation partitioning analyses (Smith & Lundholm, 2010). However, at the genetic level, there is a dichotomy between neutral and non-neutral variation (Holderegger et al., 2006; Kirk & Freeland, 2011). This is because neutral genes do not affect the fitness of individuals and thus are not affected by selection, only by dispersal and genetic drift. In turn, non-neutral genes affect fitness, and therefore are susceptible to natural selection processes. This implies that to study dispersal

processes affecting the spatial configuration of genetic variability in a species, a concrete set of genes has to be selected (usually neutral genes), whereas for study selection processes a different set of genes is needed (non-neutral genes) (Kirk & Freeland, 2011; Nosil et al., 2009).

This difference between community and population genetics studies is relevant, as said before, when comparing dispersal and selection effects in both fields. Is important to remark that, as long as the patterns are caused by ecological and historical biogeographical processes (therefore take different shapes depending on species niche breadth and dispersal ability), and build with pairwise depended data (similarity or dissimilarity measures regressed over spatial or environmental distances between sites) the methods developed here (i.e. Gompertz function to capture sigmoidal shapes and the z_{dep} statistic for parameters comparison) are perfectly applicable. The difficulties may come when trying to directly compare the effects of dispersal and selective processes in shaping the genetic-spatial distance relationship pattern. This is mainly because a gene that could be affecting fitness for a specific environmental selective pressure, and therefore susceptible to selection process has to be found (Kirk & Freeland, 2011; Meirmans, 2012), although methods to detect loci under selection have been developed (Eckert et al., 2010; Excoffier et al., 2009; Foll & Gaggiotti, 2008; Joost et al., 2007). Finally, to compare the relative relevance of dispersal and selection, both neutral and non-neutral loci have to be taken into account (Kirk & Freeland, 2011; Orsini et al., 2013). In this case, due to the absence of gene flow and the effect of genetic drift, both populations may suffer a genetic divergence also in neutral loci (Wang & Bradburd, 2014), but now genetic differences would be related to environmental distances and not whit spatial distances, although the relationship tend to be less clear than with non-neutral loci (Nosil et al., 2009; Nosil & Crespi, 2004). Thus, by studying neutral markers both dispersal and selective processes could be assessed, as long as selective processes affect gene flow (Nosil et al., 2008, 2009). Despite this particularity of the population genetics, and even working only with neutral markers or having non-neutral markers, a similar method used in Chapter III, based on standardized distances may be useful to compare dispersal and selective effects in the spatial structure of genetic variation configuration within a specie. Thus a common framework for studying biogeographical patterns at the population and community level can still be devised.

Apart from theoretical discussions, these tools have also great applicability in conservation and human impact studies, taking part in conservation biogeography. Conservation biogeography is based on the use of biogeographical tools for management and biodiversity conservation (Richardson & Whittaker, 2010; Whittaker et al., 2005). Macroecology can also be useful for these tasks because not all the processes that condition biodiversity take place at a small scale, and macroecological tools and perspectives can also be useful for conservation studies (Santini et al., 2021). In this way, the study of macroecological patterns as distance-decay and genetic-spatial distance relationship is very useful for understanding the response of biodiversity to different perturbations, which makes it a great tool for assessing the effects of pollution, global warming or invasive species in the structure of biodiversity at various levels (populations, communities) and scales (from local systems to continents and even globally). The first example of these applications has been briefly discussed in Chapter I when referring

to biological homogenization. Biological homogenization is one of the main problems caused by invasive species (McKinney & Lockwood, 1999; Olden & Poff, 2003), derived in most cases from human activities (e.g., Katsanevakis et al., 2014; McKinney, 2006; Pyšek et al., 2010). Invasions are primarily due to a range expansion of invasive species, which is reflected in the shape of distance-decay curves and therefore can be detected through their characterization. The effect of invasive species in a community can also be assessed by comparing species turnover with distance i.e. distance-decay curves before and after the invasion, this way can be studied how is affected and if there is biotic homogenization. These comparisons can be made using the z_{dep} statistic to evaluate if there are significative differences between the parameters of the models for the distance-decay curves before and after an invasion. However, the usefulness of these methods goes beyond this particular example and can be applied also to genetic diversity. When working with genetic data, the same method can be used to assess the colonization or range expansion of alleles or haplotypes, as contact can change the shape of the genetic-spatial distance relationship curve. Again, the z_{dep} statistic can be applied to these questions.

Another major application of these methods is the evaluation of the response of a community as a consequence of pollution or changes in land use. Human activities (e.g., industry, transport) produce different classes of pollutants that may cause biodiversity loss and changes in community composition and its spatial structure (McNeely, 1992; Tang et al., 2021; Tovar-Sánchez et al., 2018). Other activities, such as intensive agriculture have a great impact on land use, even changing completely the local environment depauperating natural systems and also affecting biodiversity (e.g., Karp et al., 2012; Lal et al., 1988; Tsiafouli et al., 2015). Pollution can frequently present spatial correlation caused, for example, by a concentration gradient from the source (in point source pollution) as the concentration of a certain chemical substance is greater near the source and decrease as the distance from the source increases, or by the different characteristics of the environment between places (in point and non-point source pollution), for example precipitation or vegetation cover, that leads to different pollutants concentrations (Escobedo & Nowak, 2009; Wang et al., 2022; Zhang et al., 2022a). This spatial autocorrelation (derived from the dispersal of the pollutants) makes biogeographical approaches the best way to study the impact of pollution on biodiversity. Beta diversity analysis, and by extension distance-decay curves, can be used to evaluate the change in species composition (species turnover) between sites with different concentrations of certain pollutants (e.g., Barbato et al., 2016; She et al., 2023; Zhang et al., 2022b). In this analysis, the Gompertz function can be very useful as community composition may not change with high concentrations of a chemical and therefore a plateau can be formed and distance-decay can take a sigmoidal form. In this case, the concentration at which species composition starts to change can be assessed. Another approach, is the comparison of curves before and after a pollution event or between polluted and non-polluted systems (Yan et al., 1996). Assessing differences in distance-decay before and after pollution is very useful for evaluating the effects of pollution in the community. Again, the Gompertz function has to be taken into account when modelling both curves and the z_{dep} is crucial to perform such comparison.

The same analysis can be done at genetic level to the genetic-spatial distance relationship. Both, adaptative and neutral genetic variation are important for conservation as adaptative genes respond to environmental changes and neutral genes inform about the dispersal processes among populations (Orsini et al., 2013). The insights into the causes of genetic variation within a species have a lot of interest in conservation biology and biogeography. Intraspecific adaptative genetic variation is key for natural selection and the response of the species (and therefore populations, communities and ecosystems) to a different perturbations (Oliver et al., 2015), whereas neutral genetic variation informs about demographic and colonisation history of populations (Rand, 1996). Human activities can also have genomic effects in natural populations. For example, some pollutants, as heavy metals can cause mutations at a certain concentration, in addition to change community composition (Azarbad et al., 2013; Zhang et al., 2022b). In response for a perturbation (pollution, changes in land use, etc.) the genetic structure of natural populations can also change. There are some famous experiments showing how pollution can modify the genetic structure of a population, especially in non-neutral genes subject to the selective pressure that pollution supposes, probably being the best-known example the industrial melanism of the moth *Biston betularia* (Bishop, 1972; Lees & Creed, 1975). In a similar way as distance-decay can be used to assess the impact of human activities on the spatial structure of communities, genetic-spatial distance relationship curves can be used to assess the impact of human disturbances on the spatial structure of genetic variation of species. This way, the effects of pollution gradients in genetic similarity among populations can be evaluated, and comparisons of genetic spatial structure of populations, after and before a perturbation event can be easily done. Moreover, if sufficient data are available, the effects of human activities on community composition and intraspecific genetic variance can be assessed simultaneously as analogous patterns and a common methodological approach is used for its study (although having all the data that these studies would need is very complicated) (Baselga et al., 2013, 2015; Vellend, 2003; Vellend & Geber, 2005).

In sum, this thesis provides novel methods and apply them to the study of biodiversity based on the characterization of biogeographical patterns. This methodology can be used to model patterns at various levels of biological organization and has served to deepen our knowledge on the biogeography of spiders and beetles. In addition, the new methods could be very useful for studies of conservation biology and environmental impact, which will be very important in a world that is increasingly changing due to human activity.

CONCLUSIONS

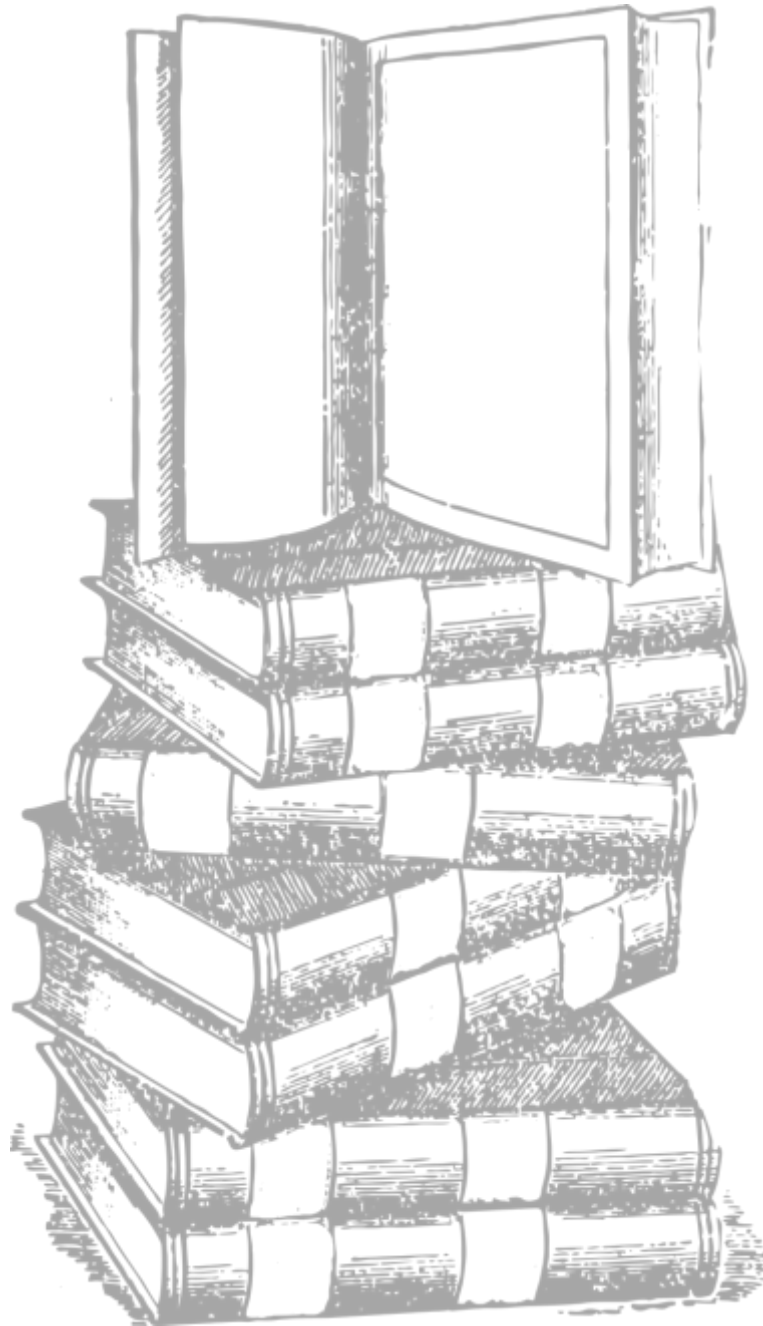


CONCLUSIONS

1. The modelling of two of the most well-known diversity patterns, the distance-decay of community similarity and the genetic-spatial distance relationship, should account for potential sigmoidal relationships that cannot be accurately captured by the most frequently used functions: linear, negative exponential and power-law.
2. The Gompertz function is a versatile function that can accommodate different shapes of distance-decay and genetic-spatial distance relationships, thus allowing the comparison among different taxa and regions.
3. The functional form of distance-decay patterns depends on the species' range sizes, as evidenced both with simulated and empirical data. As species' ranges increase in size, the function better fitting the relationship changes from power-law to negative exponential and to Gompertz, being the latter also able to capture the other two forms.
4. The z_{dep} statistic developed here is a robust statistic and should be used to compare model parameters under pairwise autocorrelation, as in distance-decay models or genetic-spatial distance relationships. It can be applied to any of the recommended functional forms for this type of patterns: linear, power-law, negative exponential and Gompertz.
5. Climatic and spatial distance-decay patterns of European spiders are largely congruent in the north of Europe but not in the South, where climatic and spatial distances are decoupled. In the South of Europe, the slopes are steeper in spatial than in climatic distance-decay curves, in accordance with the existence of relatively-close but isolated multiple glacial refugia (e.g. Iberian or Italian peninsulas).
6. Family traits related to dispersal ability explained the north-south difference the slope of spatial distance-decay patterns of European spiders, thus suggesting that their diversity patterns are shaped by incomplete postglacial limitation due to dispersal limitation in some taxa.
7. The spatial genetic structure of Iberian leaf beetles varies among species, with some species showing a lack of relationship between genetic and spatial distance while others show linear, asymptotic or sigmoidal relationships. This variety of shapes evidences the need for a complete toolbox of alternative functions that may capture the spatial-genetic relationship.

8. The Gompertz function successfully captures sigmoidal forms of the genetic-spatial distance relationships, arising when gene flow is weak or absent at large distances but very high at the shortest distances, with a linear increase of genetic distance with spatial distance in-between. Such a relationship would suggest the existence of historically isolated populations with limited range expansion. Moreover, the Gompertz function can also accommodate linear and asymptotic relationship, making it a versatile function for a wide range of gene flow/drift scenarios.
9. Dispersal ability and long-term isolation seem to be important factors shaping the spatial genetic structure of Iberian endemic leaf beetles, as evidenced by significant spatial-distance relationship being observed mostly in apterous species.
10. Independent glacial refugia and limited expansion from them has been revealed as the most probable cause of the current configuration of the disjoint distributions of the three *Lachnaia* species in the *L. tristigma*-group.
11. Methodological advances for the modelling of biodiversity patterns, i.e. the Gompertz function and the z_{dep} statistic, should be readily available for users and hence implemented as accessible software, such as the “betapart” library in R.

REFERENCES



REFERENCES

- Ackerly, D. D., Loarie, S. R., Cornwell, W. K., Weiss, S. B., Hamilton, H., Branciforte, R., & Kraft, N. J. B. (2010). The geography of climate change: Implications for conservation biogeography. *Diversity and Distributions*, 16(3), 476–487.
<https://doi.org/10.1111/j.1472-4642.2010.00654.x>
- Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C., & Swenson, N. G. (2011). Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist. *Ecology Letters*, 14(1), 19–28.
<https://doi.org/10.1111/j.1461-0248.2010.01552.x>
- Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Distance decay of similarity in freshwater communities: Do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*, 21(3), 365–375. <https://doi.org/10.1111/j.1466-8238.2011.00681.x>
- Austin, J. (2002). Cryptic lineages in a small frog: The post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Molecular Phylogenetics and Evolution*, 25(2), 316–329. [https://doi.org/10.1016/S1055-7903\(02\)00260-9](https://doi.org/10.1016/S1055-7903(02)00260-9)
- Ávila, A. C., Pires, M. M., Rodrigues, E. N. L., Costi, J. A. R., Stenert, C., & Maltchik, L. (2020). Drivers of the beta diversity of spider assemblages in southern Brazilian temporary wetlands. *Ecological Entomology*, 45(3), 466–475.
<https://doi.org/10.1111/een.12816>
- Ayinde, K., Lukman, A. F., & Arowolo, O. (2015). Robust Regression Diagnostics of Influential Observations in Linear Regression Model. *Open Journal of Statistics*, 05(04), 273–283. <https://doi.org/10.4236/ojs.2015.54029>
- Azarbad, H., Niklińska, M., van Gestel, C. A. M., van Straalen, N. M., Röling, W. F. M., & Laskowski, R. (2013). Microbial community structure and functioning along metal

- pollution gradients: Microbial structure and function along pollution gradients. *Environmental Toxicology and Chemistry*, 32(9), 1992–2002.
<https://doi.org/10.1002/etc.2269>
- Bahram, M., Kõljalg, U., Courty, P.-E., Diédhiou, A. G., Kjølner, R., Põlme, S., Ryberg, M., Veldre, V., & Tedersoo, L. (2013). The distance decay of similarity in communities of ectomycorrhizal fungi in different ecosystems and scales. *Journal of Ecology*, 101(5), 1335–1344. <https://doi.org/10.1111/1365-2745.12120>
- Baiser, B., Olden, J. D., Record, S., Lockwood, J. L., & McKinney, M. L. (2012). Pattern and process of biotic homogenization in the New Pangaea. *Proceedings of the Royal Society B: Biological Sciences*, 279(1748), 4772–4777.
<https://doi.org/10.1098/rspb.2012.1651>
- Baker, C., Horrocks, R., & Goering, C. (1975). Use of the Gompertz function for predicting corn leaf area. *Transactions of the ASAE*, 18(2), 323–326.
- Baldissera, R., Rodrigues, E. N. L., & Hartz, S. M. (2012). Metacommunity Composition of Web-Spiders in a Fragmented Neotropical Forest: Relative Importance of Environmental and Spatial Effects. *PLoS ONE*, 7(10), e48099.
<https://doi.org/10.1371/journal.pone.0048099>
- Bánki, O., Roskov, Y., Döring, M., Ower, G., Vandepitte, L., Hobern, D., Remsen, D., Schalk, P., DeWalt, R. E., Keping, M., Miller, J., Orrell, T., Aalbu, R., Abbott, J., Adlard, R., Adriaenssens, E. M., Aedo, C., Aesch, E., Akkari, N., ... Legume Phylogeny Working Group (LPWG). (2023). *Catalogue of Life Checklist* (Version 2023-03-09). Catalogue of Life. <https://doi.org/10.48580/dfrt>
- Barbato, M., Mapelli, F., Magagnini, M., Chouaia, B., Armeni, M., Marasco, R., Crotti, E., Daffonchio, D., & Borin, S. (2016). Hydrocarbon pollutants shape bacterial community assembly of harbor sediments. *Marine Pollution Bulletin*, 104(1–2), 211–220. <https://doi.org/10.1016/j.marpolbul.2016.01.029>
- Barreto, D. P., Conrad, R., Klose, M., Claus, P., & Enrich-Prast, A. (2014). Distance-decay and taxa-area relationships for bacteria, archaea and methanogenic archaea in a tropical lake sediment. *PLoS ONE*, 9(10), e110128.
<https://doi.org/10.1371/journal.pone.0110128>

- Barton, P. S., Evans, M. J., Foster, C. N., Cunningham, S. A., & Manning, A. D. (2017). Environmental and spatial drivers of spider diversity at contrasting microhabitats. *Austral Ecology*, 42(6), 700–710. <https://doi.org/10.1111/aec.12488>
- Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19(1), 134–143. <https://doi.org/10.1111/j.1466-8238.2009.00490.x>
- Baselga, A. (2012). The relationship between species replacement, dissimilarity derived from nestedness, and nestedness: Species replacement and nestedness. *Global Ecology and Biogeography*, 21(12), 1223–1232. <https://doi.org/10.1111/j.1466-8238.2011.00756.x>
- Baselga, A. (2013). Separating the two components of abundance-based dissimilarity: Balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, 4(6), 552–557. <https://doi.org/10.1111/2041-210X.12029>
- Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., & Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, 4(1), 1892. <https://doi.org/10.1038/ncomms2881>
- Baselga, A., & Gómez-Rodríguez, C. (2021). Assessing the equilibrium between assemblage composition and climate: A directional distance-decay approach. *Journal of Animal Ecology*, 90(8), 1906–1918. <https://doi.org/10.1111/1365-2656.13509>
- Baselga, A., Gómez-Rodríguez, C., Araújo, M. B., Castro-Insua, A., Arenas, M., Posada, D., & Vogler, A. P. (2022). Joint analysis of species and genetic variation to quantify the role of dispersal and environmental constraints in community turnover. *Ecography*, 2022(5). <https://doi.org/10.1111/ecog.05808>
- Baselga, A., Gómez-Rodríguez, C., & Vogler, A. P. (2015). Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Global Ecology and Biogeography*, 24(8), 873–882. <https://doi.org/10.1111/geb.12322>
- Baselga, A., & Leprieur, F. (2015). Comparing methods to separate components of beta diversity. *Methods in Ecology and Evolution*, 6(9), 1069–1079. <https://doi.org/10.1111/2041-210X.12388>
- Baselga, A., Lobo, J. M., Svenning, J.-C., Aragón, P., & Araújo, M. B. (2012). Dispersal ability modulates the strength of the latitudinal richness gradient in European beetles.

- Global Ecology and Biogeography*, 21(11), 1106–1113.
<https://doi.org/10.1111/j.1466-8238.2011.00753.x>
- Baselga, A., & Orme, C. D. L. (2012). betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution*, 3(5), 808–812. <https://doi.org/10.1111/j.2041-210X.2012.00224.x>
- Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieux, F., Logez, M., Martínez-Santalla, S., Martín-Devasa, R., Gómez-Rodríguez, C., & Crujeiras, R. M. (2023). *betapart: Partitioning beta diversity into turnover and nestedness* (1.6). <http://CRAN.R-project.org/package=betapart>
- Baselga, A., Recuero, E., Parra-Olea, G., & García-París, M. (2011). Phylogenetic patterns in zopherine beetles are related to ecological niche width and dispersal limitation. *Molecular Ecology*, 20(23), 5060–5073. <https://doi.org/10.1111/j.1365-294X.2011.05342.x>
- Baselga, A., & Ruiz-García, J. (2007). Revision of the *Lachnaia tristigma* (Lacordaire, 1848) species-group (Coleoptera: Chrysomelidae) and description of a new species. *Zootaxa*, 1630, 39–46.
- Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M., & Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*, 35(8), 673–683. <https://doi.org/10.1111/j.1600-0587.2012.07364.x>
- Becker, R. A., Wilks, A. R., & Brownrigg, R. (2016). *mapdata: Extra Map Databases*. <https://CRAN.R-project.org/package=mapdata>
- Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P., & Deckmyn, A. (2016). *maps: Draw Geographical Maps*. <https://CRAN.R-project.org/package=maps>
- Bell, J. R., Bohan, D. A., Shaw, E. M., & Weyman, G. S. (2005). Ballooning dispersal using silk: World fauna, phylogenies, genetics and models. *Bulletin of Entomological Research*, 95(2), 69–114. <https://doi.org/10.1079/BER2004350>
- Bininda-Emonds, O. R. (2005). transAlign: Using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics*, 6(1), 156. <https://doi.org/10.1186/1471-2105-6-156>

- Bishop, J. A. (1972). An experimental study of the cline of industrial melanism in *Biston betularia* (L.) (Lepidoptera) between urban Liverpool and rural North Wales. *The Journal of Animal Ecology*, 41(1), 209. <https://doi.org/10.2307/3513>
- Bishop, L., & Riechert, S. E. (1990). Spider colonization of agroecosystems: Mode and source. *Environmental Entomology*, 19(6), 1738–1745.
- Bivand, R., Keitt, T., & Rowlingson, B. (2016). *rgdal: Bindings for the ‘Geospatial’ Data Abstraction Library*. <https://CRAN.R-project.org/package=rgdal>
- Bivand, R., & Lewin-Koh, N. (2017). *maptools: Tools for Handling Spatial Objects*. <https://CRAN.R-project.org/package=maptools>
- Bivand, R., & Rundel, C. (2020). *Rgeos: Interface to Geometry Engine—Open Source (‘GEOS’)*. <https://CRAN.R-project.org/package=rgeos>
- Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis with R, Second edition*. Springer, NY. <https://asdar-book.org/>
- Blandenier, G. (2009). Ballooning of spiders (Araneae) in Switzerland: General results from an eleven-year Survey. *Arachnology*, 14(7), 308–316. <https://doi.org/10.13156/arac.2009.14.7.308>
- Blandenier, G., & Fürst, P. A. (1998). Ballooning spiders caught by a suction trap in agricultural landscape in Switzerland. *Proceedings of the 17th European Colloquium of Arachnology*, 177–186.
- Bonte, D., Baert, L., Lens, L., & Maelfait, J.-P. (2004). Effects of aerial dispersal, habitat specialisation, and landscape structure on spider distribution across fragmented grey dunes. *Ecography*, 27, 343–349.
- Booth, H. (1984). Transforming Gompertz’s function for fertility analysis: The development of a standard for the relational Gompertz function. *Population Studies*, 38(3), 495–506. <https://www.jstor.org/stable/2174137>
- Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73(3), 1045–1055.
- Boston, E. S. M., Ian Montgomery, W., Hynes, R., & Prodöhl, P. A. (2015). New insights on postglacial colonization in western Europe: The phylogeography of the Leisler’s bat (*Nyctalus leisleri*). *Proceedings of the Royal Society B: Biological Sciences*, 282(1804), 20142605. <https://doi.org/10.1098/rspb.2014.2605>

- Bowden, J. J., & Buddle, C. M. (2010). Spider assemblages across elevational and latitudinal gradients in the Yukon territory, Canada. *Arctic*, 63(3), 261–272.
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, Th., Hewitt, C. D., Kageyama, M., Kitoh, A., Laîné, A., Loutre, M.-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S. L., ... Zhao, Y. (2007). Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: Experiments and large-scale features. *Climate of the Past*, 3(2), 261–277. <https://doi.org/10.5194/cp-3-261-2007>
- Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325–349. <https://doi.org/10.2307/1942268>
- Brousseau, P.-M., Gravel, D., & Handa, I. T. (2018). On the development of a predictive functional trait approach for studying terrestrial arthropods. *Journal of Animal Ecology*, 87(5), 1209–1220. <https://doi.org/10.1111/1365-2656.12834>
- Brown, J. H., & Maurer, B. A. (1989). Macroecology: The division of food and space among species on continents. *Science*, 243(4895), 1145–1150.
- Brown, J. H., Stevens, G. C., & Kaufman, D. M. (1996). The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, 27(1), 597–623.
- Buckley, L. B., Hurlbert, A. H., & Jetz, W. (2012). Broad-scale ecological implications of ectothermy and endothermy in changing environments. *Global Ecology and Biogeography*, 21(9), 873–885. <https://doi.org/10.1111/j.1466-8238.2011.00737.x>
- Buckley, L. B., & Jetz, W. (2008). Linking global turnover of species and environments. *Proceedings of the National Academy of Sciences*, 105(46), 17836–17841. <https://doi.org/10.1073/pnas.0803524105>
- Busby. (1991). *BIOCLIM – a bioclimate analysis and prediction system [pp. 64–68]*. In: *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (C. R. Margules and M. P. Austin, editors). Commonwealth Scientific and Industrial Research.
- Calatayud, J., Hortal, J., Medina, N. G., Turin, H., Bernard, R., Casale, A., Ortuño, V. M., Penev, L., & Rodríguez, M. Á. (2016). Glaciations, deciduous forests, water availability and current geographical patterns in the diversity of European *Carabus*

- species. *Journal of Biogeography*, 43(12), 2343–2353.
<https://doi.org/10.1111/jbi.12811>
- Cardoso, P., Erwin, T. L., Borges, P. A. V., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, 144(11), 2647–2655. <https://doi.org/10.1016/j.biocon.2011.07.024>
- Cardoso, P., Pekár, S., Jocqué, R., & Coddington, J. A. (2011). Global patterns of guild composition and functional diversity of spiders. *PLoS ONE*, 6(6), e21710.
<https://doi.org/10.1371/journal.pone.0021710>
- Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: A flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2(6), 667–680. <https://doi.org/10.1007/BF00051966>
- Carstens, B. C., & Knowles, L. L. (2007). Shifting distributions and speciation: Species divergence during rapid climate change. *Molecular Ecology*, 16(3), 619–627.
<https://doi.org/10.1111/j.1365-294X.2006.03167.x>
- Carvalho, J. C., & Cardoso, P. (2014). Drivers of beta diversity in Macaronesian spiders in relation to dispersal ability. *Journal of Biogeography*, 41(10), 1859–1870.
<https://doi.org/10.1111/jbi.12348>
- Carvalho, J. C., Cardoso, P., Crespo, L. C., Henriques, S., Carvalho, R., & Gomes, P. (2011). Determinants of beta diversity of spiders in coastal dunes along a gradient of mediterraneity. *Diversity and Distributions*, 17(2), 225–234.
<https://doi.org/10.1111/j.1472-4642.2010.00731.x>
- Çelekli, A., Balcı, M., & Bozkurt, H. (2008). Modelling of *Scenedesmus obliquus*; function of nutrients with modified Gompertz model. *Bioresource Technology*, 99(18), 8742–8747. <https://doi.org/10.1016/j.biortech.2008.04.028>
- Chao, A., Chiu, C.-H., & Hsieh, T. C. (2012). Proposing a resolution to debates on diversity partitioning. *Ecology*, 93(9), 2037–2051. <https://doi.org/10.1890/11-1817.1>
- Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In U. Bastolla, M. Porto, H. E. Roman, & M. Vendruscolo (Eds.), *Structural Approaches to Sequence Evolution* (1st ed., pp. 207–232). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-540-35306-5_10

- Chust, G., Villarino, E., Chenuil, A., Irigoien, X., Bizsel, N., Bode, A., Broms, C., Claus, S., Fernández de Puellas, M. L., Fonda-Umani, S., Hoarau, G., Mazzocchi, M. G., Mozetič, P., Vandepitte, L., Veríssimo, H., Zervoudaki, S., & Borja, A. (2016). Dispersal similarly shapes both population genetics and community patterns in the marine realm. *Scientific Reports*, 6(1), 28730. <https://doi.org/10.1038/srep28730>
- Clavel, J., Julliard, R., & Devictor, V. (2011). Worldwide decline of specialist species: Toward a global functional homogenization? *Frontiers in Ecology and the Environment*, 9(4), 222–228. <https://doi.org/10.1890/080216>
- Clobert, J., Baguette, M., Benton, T. G., & Bullock, J. M. (Eds.). (2012). *Dispersal Ecology and Evolution* (1st ed). Oxford University Press.
- Connolly, S. R., Keith, S. A., Colwell, R. K., & Rahbek, C. (2017). Process, mechanism, and modeling in macroecology. *Trends in Ecology & Evolution*, 32(11), 835–844. <https://doi.org/10.1016/j.tree.2017.08.011>
- Connor, E. F., & McCoy, E. D. (1979). The statistics and biology of the species-area relationship. *The American Naturalist*, 113(6), 791–833.
- Conroy, C. J., & Cook, J. A. (2000). Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Molecular Ecology*, 9(2), 165–175. <https://doi.org/10.1046/j.1365-294x.2000.00846.x>
- Cressie, N. A. (1993). *Statistics for Spatial Data* (1st ed). John Wiley & Sons.
- Crnobrnja-Isailovic, J. (2007). Cross-section of a refugium: Genetic diversity of amphibian and reptile populations in the Balkans. In *Phylogeography of southern European refugia* (pp. 327–337). Springer.
- Csapó, H., Krzywoźniak, P., Grabowski, M., Wattier, R., Bącela-Spychalska, K., Mamos, T., Jelić, M., & Rewicz, T. (2020). Successful post-glacial colonization of Europe by single lineage of freshwater amphipod from its Pannonian Plio-Pleistocene diversification hotspot. *Scientific Reports*, 10(1), 18695. <https://doi.org/10.1038/s41598-020-75568-7>
- Dahl, M. T., Yoccoz, N. G., Aakra, K., & Coulson, S. J. (2018). The Araneae of Svalbard: The relationships between specific environmental factors and spider assemblages in the High Arctic. *Polar Biology*, 41(5), 839–853. <https://doi.org/10.1007/s00300-017-2247-4>

- Dale, M. R. T., & Fortin, M.-J. (2002). Spatial autocorrelation and statistical test in ecology. *Écoscience*, 9(2), 162–167. <http://dx.doi.org/10.1080/11956860.2002.11682702>
- Davies, T. J., Buckley, L. B., Grenyer, R., & Gittleman, J. L. (2011). The influence of past and present climate on the biogeography of modern mammal diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1577), 2526–2535. <https://doi.org/10.1098/rstb.2011.0018>
- De Candolle, A. (1855). *Géographie Botanique Raisonnée; ou, Exposition des Faits Principaux et des lois Concernant la Distribution Géographique des Plantes de L'époque Actuelle*. (1st ed, Vol. 2).
- Dean, D. A., & Sterling, W. L. (1985). Size and phenology of ballooning spiders at two locations in Eastern Texas. *Journal of Arachnology*, 13, 111–120.
- Dengler, J. (2009). Which function describes the species-area relationship best? A review and empirical evaluation. *Journal of Biogeography*, 36(4), 728–744. <https://doi.org/10.1111/j.1365-2699.2008.02038.x>
- Desender, K., & Verdyck, P. (2001). Geographic scaling and genetic differentiation in two highly mobile European saltmarsh beetles. *Belg. J. Zool.*, 131(1), 31–42.
- Dias, M. S., Oberdorff, T., Hugueny, B., Leprieux, F., Jézéquel, C., Cornu, J.-F., Brosse, S., Grenouillet, G., & Tedesco, P. A. (2014). Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology Letters*, 17(9), 1130–1140. <https://doi.org/10.1111/ele.12319>
- Diniz-Filho, J. A. F., De Marco Jr, P., & Hawkins, B. A. (2010). Defying the curse of ignorance: Perspectives in insect macroecology and conservation biogeography. *Insect Conservation and Diversity*. <https://doi.org/10.1111/j.1752-4598.2010.00091.x>
- Diserud, O. H., & Ødegaard, F. (2007). A multiple-site similarity measure. *Biology Letters*, 3(1), 20–22. <https://doi.org/10.1098/rsbl.2006.0553>
- Dobrovolski, R., Melo, A. S., Cassemiro, F. A. S., & Diniz-Filho, J. A. F. (2012). Climatic history and dispersal ability explain the relative importance of turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 21(2), 191–197. <https://doi.org/10.1111/j.1466-8238.2011.00671.x>
- Drakou, E. G., Bobori, D. C., Kallimanis, A. S., Mazaris, A. D., Sgardelis, S. P., & Pantis, J. D. (2009). Freshwater fish community structured more by dispersal limitation than by

- environmental heterogeneity. *Ecology of Freshwater Fish*, 18(3), 369–379.
<https://doi.org/10.1111/j.1600-0633.2009.00354.x>
- Duffey, E. (1998). Aerial dispersal in spiders. *Proceedings of the 17th European Colloquium of Arachnology*, 187–191.
- Dufresnes, C., Nicieza, A. G., Litvinchuk, S. N., Rodrigues, N., Jeffries, D. L., Vences, M., Perrin, N., & Martínez-Solano, Í. (2020). Are glacial refugia hotspots of speciation and cytonuclear discordances? Answers from the genomic phylogeography of Spanish common frogs. *Molecular Ecology*, 29(5), 986–1000.
<https://doi.org/10.1111/mec.15368>
- Eckert, A. J., van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-Martínez, S. C., & Neale, D. B. (2010). Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics*, 185(3), 969–982.
<https://doi.org/10.1534/genetics.110.115543>
- Eckert, C. G., Samis, K. E., & Loughheed, S. C. (2008). Genetic variation across species' geographical ranges: The central–marginal hypothesis and beyond. *Molecular Ecology*, 17(5), 1170–1188. <https://doi.org/10.1111/j.1365-294X.2007.03659.x>
- Edwards, J., & Thornton, I. W. B. (2001). Colonization of an island volcano, Long Island, Papua New Guinea, and an emergent island, Motmot, in its caldera lake. VI. The pioneer arthropod community of Motmot. *Journal of Biogeography*, 28, 1379–1388.
- Eghwerido, J. T., Nzei, L., & Adubisi, O. D. (2020). The Gompertz extended generalized exponential distribution: Properties and applications. *Communications Faculty Of Science University of Ankara Series A1 Mathematics and Statistics*, 739–753.
<https://doi.org/10.31801/cfsuasmas.602930>
- Elzhov, T. V., Mullen, K. M., Spiess, A.-N., & Bolker, B. (2016). *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*. <https://CRAN.R-project.org/package=minpack.lm>
- Escobedo, F. J., & Nowak, D. J. (2009). Spatial heterogeneity and air pollution removal by an urban forest. *Landscape and Urban Planning*, 90(3–4), 102–110.
<https://doi.org/10.1016/j.landurbplan.2008.10.021>

- Esyunin, S. L., & Sozontov, A. N. (2016). On a new Eurasian species of *Dictyna Sundevall*, 1833 (Aranei: Dictynidae), with taxonomic notes on poorly known Palearctic Dictyna species. *Arthropoda Selecta*, 25(2), 199–206.
- Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4), 285–298. <https://doi.org/10.1038/hdy.2009.74>
- Fernandez-Fournier, P., & Avilés, L. (2018). Environmental filtering and dispersal as drivers of metacommunity composition: Complex spider webs as habitat patches. *Ecosphere*, 9(2), e02101. <https://doi.org/10.1002/ecs2.2101>
- Ferrier, S., Drielsma, M., Manion, G., & Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, 11, 2309–2338.
- Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13(3), 252–264. <https://doi.org/10.1111/j.1472-4642.2007.00341.x>
- Finch, O.-D., Blick, T., & Schuldt, A. (2008). Macroecological patterns of spider species richness across Europe. *Biodiversity and Conservation*, 17(12), 2849–2868. <https://doi.org/10.1007/s10531-008-9400-x>
- Fisher, J. C. (2020). *inlmisc—Miscellaneous Functions for the U.S. Geological Survey Idaho National Laboratory Project Office*. <https://CRAN.R-project.org/package=inlmisc>
- Foelix, R. F. (2011). *Biology of Spiders* (3rd ed). Oxford University Press.
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics*, 180(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Fountain-Jones, N. M., Baker, S. C., & Jordan, G. J. (2015). Moving beyond the guild concept: Developing a practical functional trait framework for terrestrial beetles. *Ecological Entomology*, 40(1), 1–13. <https://doi.org/10.1111/een.12158>
- Freeman, J. A. (1946). The distribution of spiders and mites up to 300 ft. In the air. *Journal of Animal Ecology*, 15(1), 69–74.
- Garnier, S., Alibert, P., Audiot, P., Prieur, B., & Rasplus, J.-Y. (2004). Isolation by distance and sharp discontinuities in gene frequencies: Implications for the phylogeography of

- an alpine insect species, *Carabus solieri*. *Molecular Ecology*, 13(7), 1883–1897.
<https://doi.org/10.1111/j.1365-294X.2004.02212.x>
- Gaston, K. J. (1996). Species-range-size distributions: Patterns, mechanisms and implications. *Trends in Ecology & Evolution*, 11(5), 197–201. [https://doi.org/10.1016/0169-5347\(96\)10027-6](https://doi.org/10.1016/0169-5347(96)10027-6)
- Gaston, K. J., & Fuller, R. A. (2009). The sizes of species' geographic ranges. *Journal of Applied Ecology*, 46(1), 1–9. <https://doi.org/10.1111/j.1365-2664.2008.01596.x>
- Glick, P. A. (1939). The distribution of insects, spiders and mites in the air. *Technical Bulletin, United States Department of Agriculture*, 673.
- Gloor, D., Blick, T., Nentwig, W., Kropf, C., & Hänggi, A. (2010). *Spiders of Europe*.
<https://doi.org/10.24436/1>
- Gloor, D., Nentwig, W., Blick, T., & Kropf, C. (2017). *World Spider Catalog*.
<https://doi.org/10.24436/2>
- Godeau, U., Bouget, C., Piffady, J., Pozzi, T., & Gosselin, F. (2020). Lack of definition of mathematical terms in ecology: The case of the sigmoid class of functions in macroecology. *Ecology and Evolution*, ece3.7016. <https://doi.org/10.1002/ece3.7016>
- Goldenberg Vilar, A., van Dam, H., van Loon, E. E., Vonk, J. A., van Der Geest, H. G., & Admiraal, W. (2014). Eutrophication decreases distance decay of similarity in diatom communities. *Freshwater Biology*, 59(7), 1522–1531.
<https://doi.org/10.1111/fwb.12363>
- Gómez-Rodríguez, C., & Baselga, A. (2018). Variation among European beetle taxa in patterns of distance decay of similarity suggests a major role of dispersal processes. *Ecography*, 41(11), 1825–1834. <https://doi.org/10.1111/ecog.03693>
- Gómez-Rodríguez, C., Miller, K. E., Castillejo, J., Iglesias-Piñeiro, J., & Baselga, A. (2020). Disparate dispersal limitation in *Geomalacus* slugs unveiled by the shape and slope of the genetic–spatial distance relationship. *Ecography*, 43(8), 1229–1240.
<https://doi.org/10.1111/ecog.05142>
- Gossner, M. M., Lewinsohn, T. M., Kahl, T., Grassein, F., Boch, S., Prati, D., Birkhofer, K., Renner, S. C., Sikorski, J., Wubet, T., Arndt, H., Baumgartner, V., Blaser, S., Blüthgen, N., Börschig, C., Buscot, F., Diekötter, T., Jorge, L. R., Jung, K., ... Allan, E. (2016). Land-use intensification causes multitrophic homogenization of grassland communities. *Nature*, 540(7632), 266–269. <https://doi.org/10.1038/nature20575>

- Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., Currie, D. J., Dunn, R. R., Graves, G. R., Green, J. L., Grytnes, J.-A., Jiang, Y.-H., Jetz, W., Kathleen Lyons, S., McCain, C. M., Magurran, A. E., Rahbek, C., Rangel, T. F. L. V. B., Soberón, J., ... Willig, M. R. (2009). Patterns and causes of species richness: A general simulation model for macroecology. *Ecology Letters*, 12(9), 873–886. <https://doi.org/10.1111/j.1461-0248.2009.01353.x>
- Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A. T. R., Alahuhta, J., Altman, J., Angiolini, C., Aroviita, J., Attorre, F., Baastrop-Spohr, L., Barrera-Alba, J. J., Belmaker, J., Biurrun, I., Bonari, G., Bruelheide, H., Burrascano, S., Carboni, M., Cardoso, P., Carvalho, J. C., ... Soininen, J. (2022). Distance decay 2.0 – A global synthesis of taxonomic and functional turnover in ecological communities. *Global Ecology and Biogeography*, 31(7), 1399–1421. <https://doi.org/10.1111/geb.13513>
- Graham, C. H., & Fine, P. V. A. (2008). Phylogenetic beta diversity: Linking ecological and evolutionary processes across space in time. *Ecology Letters*, 11(12), 1265–1277. <https://doi.org/10.1111/j.1461-0248.2008.01256.x>
- Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006). Reconciling niche and neutrality: The continuum hypothesis. *Ecology Letters*, 9(4), 399–409. <https://doi.org/10.1111/j.1461-0248.2006.00884.x>
- Greenstone, M. H., Morgan, C. E., & Hultsch, A.-L. (1987). Ballooning spiders in Missouri, USA, and New South Wales, Australia: Family and mass distributions. *Journal of Arachnology*, 15, 163–170.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Harte, J., Blackburn, T., & Ostling, A. (2001). Self-Similarity and the relationship between abundance and range size. *The American Naturalist*, 157(4), 374–386.
- Hewitt, G. (2003). Ice ages. In *Evolution on Planet Earth* (pp. 339–361). Elsevier. <https://doi.org/10.1016/B978-012598655-7/50045-8>
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247–276.
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68, 87–112.

- Hewitt, G. M. (2000). *The genetic legacy of the Quaternary ice ages*. 405, 907–913.
- Hijmans, R., Cameron, S., Parra, J. L., Jones, P., Jarvis, A., & Richardson, K. (2017). *WorldClim—Global Climate Data | Free climate data for ecological modeling and GIS*. Worldclim.org
- Hijmans, R. J. (2021). *raster: Geographic Data Analysis and Modeling*. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *dismo: Species Distribution Modeling*. <https://CRAN.R-project.org/package=dismo>
- Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, 163(2), 192–211. <https://doi.org/10.1086/381004>
- Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: Implications for landscape genetics. *Landscape Ecology*, 21(6), 797–807. <https://doi.org/10.1007/s10980-005-5245-9>
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography* (1st ed). Princeton University Press. <https://doi.org/10.1515/9781400837526>
- Hutchison, D. W., & Templeton, A. R. (1999). Correlation of the pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, 53(6), 1898–1914.
- Jablonski, D. (2008). Species selection: Theory and data. *Annual Review of Ecology, Evolution, and Systematics*, 39(1), 501–524. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173510>
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, 11(2), 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Jiménez-Valverde, A., Baselga, A., Melic, A., & Txasko, N. (2010). Climate and regional beta-diversity gradients in spiders: Dispersal capacity has nothing to say? *Insect Conservation and Diversity*, 3(1), 51–60. <https://doi.org/10.1111/j.1752-4598.2009.00067.x>

- Jiménez-Valverde, A., & Lobo, J. M. (2007). Determinants of local spider (Araneidae and Thomisidae) species richness on a regional scale: Climate and altitude vs habitat structure. *Ecological Entomology*, 32, 113–122.
- Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6), 885–890. <https://doi.org/10.1111/j.1472-4642.2008.00496.x>
- Johnsen, A. R., Binning, P. J., Aamand, J., Badawi, N., & Rosenbom, A. E. (2013). The Gompertz function can coherently describe microbial mineralization of growth-sustaining pesticides. *Environmental Science & Technology*, 130711162351002. <https://doi.org/10.1021/es400861v>
- Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16(18), 3955–3969. <https://doi.org/10.1111/j.1365-294X.2007.03442.x>
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10), 2427–2439. <https://doi.org/10.1890/06-1736.1>
- Karp, D. S., Rominger, A. J., Zook, J., Ranganathan, J., Ehrlich, P. R., & Daily, G. C. (2012). Intensive agriculture erodes β -diversity at large scales. *Ecology Letters*, 15(9), 963–970. <https://doi.org/10.1111/j.1461-0248.2012.01815.x>
- Katsanevakis, S., Coll, M., Piroddi, C., Steenbeek, J., Ben Rais Lasram, F., Zenetos, A., & Cardoso, A. C. (2014). Invading the Mediterranean Sea: Biodiversity patterns shaped by human activities. *Frontiers in Marine Science*, 1. <https://doi.org/10.3389/fmars.2014.00032>
- Keith, S. A., Webb, T. J., Böhning-Gaese, K., Connolly, S. R., Dulvy, N. K., Eigenbrod, F., Jones, K. E., Price, T., Redding, D. W., Owens, I. P. F., & Isaac, N. J. B. (2012). What is macroecology? *Biology Letters*, 8(6), 904–906. <https://doi.org/10.1098/rsbl.2012.0672>
- Kimura, M. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 320, 343–354.
- Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4), 561–576. <https://doi.org/10.1093/genetics/49.4.561>

- Kinlan, B. P., & Gaines, S. D. (2003). Propagule dispersal in marine and terrestrial environments: A community perspective. *Ecology*, 84(8), 2007–2020.
- Kirk, H., & Freeland, J. R. (2011). Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, 12(6), 3966–3988. <https://doi.org/10.3390/ijms12063966>
- Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a Species' Range. *The American Naturalist*, 150(1), 1–23. <https://doi.org/10.1086/286054>
- Koch, L. (1879). Arachniden aus Sibirien und Novaja Semlja, eingesammelt von der schwedischen expedition im Jahre 1875. *Kongliga Svenska Vetenskaps-Akademiens Handlingar*, 16(5), 1–136.
- Koenig, W. (1999). Spatial autocorrelation of ecological phenomena. *Tree*, 14(1), 22–26.
- Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, 72(3), 367–382. <https://doi.org/10.1046/j.1365-2656.2003.00710.x>
- König, C., Weigelt, P., & Kreft, H. (2017). Dissecting global turnover in vascular plants. *Global Ecology and Biogeography*, 26(2), 228–242. <https://doi.org/10.1111/geb.12536>
- Koponen, S. (1991). On the biogeography and faunistics of European spiders: Latitude, altitude and insularity. *XIIIe Colloque Européen d'Arachnologie. Bulletin de La Societe Neuchateloise Des Sciences Naturelles.*, 116, 141–152.
- Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 1217–1241.
- La Sorte, F. A., & McKinney, M. L. (2006). Compositional similarity and the distribution of geographical range size for assemblages of native and non-native species in urban floras. *Diversity and Distributions*, 12, 679–686. <https://doi.org/10.1111/j.1366-9516.2006.00276.x>
- La Sorte, F. A., McKinney, M. L., Pyšek, P., Klotz, S., Rapson, G. L., Celesti-Grapow, L., & Thompson, K. (2008). Distance decay of similarity among European urban floras: The impact of anthropogenic activities on β diversity. *Global Ecology and Biogeography*, 17(3), 363–371. <https://doi.org/10.1111/j.1466-8238.2007.00369.x>
- Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1), 386–404.

- Lal, R., Miller, F. P., & Logan, T. J. (1988). Are intensive agricultural practices environmentally and ethically sound? *Journal of Agricultural Ethics*, 1(3), 193–210. <https://doi.org/10.1007/BF01833409>
- Larrivé, M., & Buddle, C. M. (2011). Ballooning propensity of canopy and understorey spiders in a mature temperate hardwood forest. *Ecological Entomology*, 36(2), 144–151. <https://doi.org/10.1111/j.1365-2311.2010.01255.x>
- Ledoux, J.-C. (2014). Les Hahniidae de la faune française (Araneae). *Revue Arachnologique*, 1, 29–40.
- Lees, D. R., & Creed, E. R. (1975). Industrial Melanism in *Biston betularia*: The Role of Selective Predation. *The Journal of Animal Ecology*, 44(1), 67. <https://doi.org/10.2307/3852>
- Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74(6), 1659–1673.
- Legendre, P., Borcard, D., & Peres-Neto, P. R. (2005). Analyzing beta diversity: Partitioning the spatial variation of community composition data. *Ecological Monographs*, 75(4), 435–450. <https://doi.org/10.1890/05-0549>
- Legendre, P., & Fortin, M.-J. (1989). Spatial pattern an ecological analysis. *Vegetatio*, 80, 107–138.
- Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (3rd ed). Elsevier.
- Leibold, M. A., & Chase, J. M. (2018). *Metacommunity Ecology*. Princeton University Press. <https://doi.org/10.1515/9781400889068>
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology: The metacommunity concept. *Ecology Letters*, 7(7), 601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
- Leibold, M. A., & McPeck, M. A. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, 87(6), 1399–1410. [https://doi.org/10.1890/0012-9658\(2006\)87\[1399:COTNAN\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[1399:COTNAN]2.0.CO;2)
- Lennon, J. J., Koleff, P., Greenwood, J. J. D., & Gaston, K. J. (2001). The geographical structure of British bird distributions: Diversity, spatial turnover and scale: *Spatial*

- structure of assemblages. *Journal of Animal Ecology*, 70(6), 966–979.
<https://doi.org/10.1046/j.0021-8790.2001.00563.x>
- Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Muriénne, J., & Grenouillet, G. (2020). Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*, 4(8), 1044–1059. <https://doi.org/10.1038/s41559-020-1198-2>
- Leprieur, F., Albouy, C., De Bortoli, J., Cowman, P. F., Bellwood, D. R., & Mouillot, D. (2012). Quantifying phylogenetic beta diversity: Distinguishing between ‘true’ turnover of lineages and phylogenetic diversity gradients. *PLoS ONE*, 7(8), e42760. <https://doi.org/10.1371/journal.pone.0042760>
- Leprieur, F., Olden, J. D., Lek, S., & Brosse, S. (2009). Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. *Journal of Biogeography*, 36(10), 1899–1912. <https://doi.org/10.1111/j.1365-2699.2009.02107.x>
- Leroy, B., Dias, M. S., Giraud, E., Hugueny, B., Jézéquel, C., Leprieur, F., Oberdorff, T., & Tedesco, P. A. (2019). Global biogeographical regions of freshwater fish species. *Journal of Biogeography*, 46(11), 2407–2419. <https://doi.org/10.1111/jbi.13674>
- Lesbarrères, D. (2009). Post-glacial phylogeography: New insight into an old story: the post-glacial recolonization of European biota. *Heredity*, 102(3), 213–213. <https://doi.org/10.1038/hdy.2008.121>
- Lester, S. E., Ruttenberg, B. I., Gaines, S. D., & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, 10(8), 745–758. <https://doi.org/10.1111/j.1461-0248.2007.01070.x>
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2), 164–168. <https://doi.org/10.1090/qam/10666>
- Liu, R. Y., Singh, K., & others. (1992). Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the Limits of Bootstrap*, 225, 248.
- Loboda, S., & Buddle, C. M. (2018). Small to large-scale patterns of ground-dwelling spider (Araneae) diversity across northern Canada. *Facets*, 3(1), 880–895. <https://doi.org/10.1139/facets-2018-0007>
- Lomolino, M. V. (2000). Ecology’s most general, yet protean 1 pattern: The species-area relationship. *Journal of Biogeography*, 27(1), 17–26. <https://doi.org/10.1046/j.1365-2699.2000.00377.x>

- Lomolino, M. V., Riddle, B. R., & Whittaker, R. J. (2017). *Biogeography: Biological Diversity Across Space and Time* (5th ed). Sinauer Associates, Inc., Publishers.
- Losos, J. B., & Schluter, D. (2000). Analysis of an evolutionary species–area relationship. *Nature*, 408(6814), 847–850. <https://doi.org/10.1038/35048558>
- Lyons, K., & Willig, M. R. (1999). A hemispheric assesment of scale dependence in latitudinal gradients of species richness. *Ecology*, 80(8), 2483–2491.
- Macías-Hernández, N., Domènech, M., Cardoso, P., Emerson, B. C., Borges, P. A. V., Lozano-Fernandez, J., Paulo, O. S., Vieira, A., Enguídanos, A., Rigal, F., Amorim, I. R., & Arnedo, M. A. (2020a). Building a robust, densely-sampled spider tree of life for ecosystem research. *Diversity*, 12(8), 288. <https://doi.org/10.3390/d12080288>
- Macías-Hernández, N., Ramos, C., Domènech, M., Febles, S., Santos, I., Arnedo, M., Borges, P., Emerson, B., & Cardoso, P. (2020b). A database of functional traits for spiders from native forests of the Iberian Peninsula and Macaronesia. *Biodiversity Data Journal*, 8, e49159. <https://doi.org/10.3897/BDJ.8.e49159>
- Maloney, K. O., & Munguia, P. (2011). Distance decay of similarity in temperate aquatic communities: Effects of environmental transition zones, distance measure, and life histories. *Ecography*, 34(2), 287–295. <https://doi.org/10.1111/j.1600-0587.2010.06518.x>
- Mammola, S., Cardoso, P., Angyal, D., Balázs, G., Blick, T., Brustel, H., Carter, J., Čurčić, S., Danflous, S., Dányi, L., Déjean, S., Deltchev, C., Elverici, M., Fernández, J., Gasparo, F., Komnenov, M., Komposch, C., Kováč, L., Kunt, K. B., ... Isaia, M. (2019). Local- versus broad-scale environmental drivers of continental β -diversity patterns in subterranean spider communities across Europe. *Proceedings of the Royal Society B: Biological Sciences*, 286(1914), 20191579. <https://doi.org/10.1098/rspb.2019.1579>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209–220.
- Mantel, N., & Valand, R. (1970). A technique of nonparametric multivariate analysis. *Biometrics*, 26(3), 547–558.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441. JSTOR. <http://www.jstor.org/stable/2098941>

- Marske, K. A., Thomaz, A. T., & Knowles, L. L. (2020). Dispersal barriers and opportunities drive multiple levels of phylogeographic concordance in the Southern Alps of New Zealand. *Molecular Ecology*, 29(23), 4665–4679. <https://doi.org/10.1111/mec.15655>
- Martín-Devasa, R., & Baselga, A. (2018). What caused the disjunct distributions of the *Lachnaia tristigma* species-group (Coleoptera: Chrysomelidae) on the Iberian Peninsula? *Coleopterists Bulletin*, 72(3), 421–426.
- Martín-Devasa, R., Martínez-Santalla, S., Gómez-Rodríguez, C., Crujeiras, R. M., & Baselga, A. (2022a). Species range size shapes distance-decay in community similarity. *Diversity and Distributions*, 28(7), 1348–1357. <https://doi.org/10.1111/ddi.13550>
- Martín-Devasa, R., Martínez-Santalla, S., Gómez-Rodríguez, C., Crujeiras, R. M., & Baselga, A. (2022b). Comparing distance-decay parameters: A novel test under pairwise dependence. *Ecological Informatics*, 72, 101894. <https://doi.org/10.1016/j.ecoinf.2022.101894>
- Martínez-Santalla, S., Martín-Devasa, R., Gómez-Rodríguez, C., Crujeiras, R. M., & Baselga, A. (2022). Assessing the nonlinear decay of community similarity: Permutation and site-block resampling significance tests. *Journal of Biogeography*, jbi.14351. <https://doi.org/10.1111/jbi.14351>
- Marusik, Y. M., & Koponen, S. (1998). New and little known spiders of the subfamily Dictyninae (Araneae: Dictynidae) from South Siberia. *Entomological Problems*, 29(2), 79–86.
- Marusik, Y. M., & Koponen, S. (2002). Diversity of spiders in boreal and arctic zones. *The Journal of Arachnology*, 30, 205–210.
- Masel, J. (2011). Genetic drift. *Current Biology*, 21(20), 837–838.
- McGill, B. J. (2019). The what, how and why of doing macroecology. *Global Ecology and Biogeography*, 28(1), 6–17. <https://doi.org/10.1111/geb.12855>
- McKinney, M. L. (2006). Urbanization as a major cause of biotic homogenization. *Biological Conservation*, 127(3), 247–260. <https://doi.org/10.1016/j.biocon.2005.09.005>
- McKinney, M. L., & Lockwood, J. L. (1999). Biotic homogenization: A few winners replacing many losers in the next mass extinction. *Trends in Ecology & Evolution*, 14(11), 450–453. [https://doi.org/10.1016/S0169-5347\(99\)01679-1](https://doi.org/10.1016/S0169-5347(99)01679-1)
- McNeely, J. A. (1992). The sinking ark: Pollution and the worldwide loss of biodiversity. *Biodiversity and Conservation*, 1(1), 2–18. <https://doi.org/10.1007/BF00700247>

- McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8), 1551–1561.
<https://doi.org/10.1111/j.0014-3820.2006.tb00500.x>
- Médail, F., & Quézel, P. (1999). Biodiversity hotspots in the Mediterranean Basin: Setting global conservation priorities. *Conservation Biology*, 13(6), 1510–1513.
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839–2846. <https://doi.org/10.1111/j.1365-294X.2012.05578.x>
- Milici, M., Tomasch, J., Wos-Oxley, M. L., Decelle, J., Jáuregui, R., Wang, H., Deng, Z.-L., Plumeier, I., Giebel, H.-A., Badewien, T. H., Wurst, M., Pieper, D. H., Simon, M., & Wagner-Döbler, I. (2016). Bacterioplankton biogeography of the Atlantic Ocean: A case study of the distance-decay relationship. *Frontiers in Microbiology*, 7.
<https://doi.org/10.3389/fmicb.2016.00590>
- Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2), 284–289. <https://doi.org/10.1111/j.1467-8306.2004.09402005.x>
- Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R., & Green, J. L. (2008). A general framework for the distance-decay of similarity in ecological communities. *Ecology Letters*, 11(9), 904–917.
<https://doi.org/10.1111/j.1461-0248.2008.01202.x>
- Múrria, C., Bonada, N., Vellend, M., Zamora-Muñoz, C., Alba-Tercedor, J., Sainz-Cantero, C. E., Garrido, J., Acosta, R., El Alami, M., Barquín, J., Derka, T., Álvarez-Cabria, M., Sáinz-Bariain, M., Filipe, A. F., & Vogler, A. P. (2017). Local environment rather than past climate determines community composition of mountain stream macroinvertebrates across Europe. *Molecular Ecology*, 26(21), 6085–6099.
<https://doi.org/10.1111/mec.14346>
- Nathan, R. (2013). Dispersal Biogeography. In *Encyclopedia of Biodiversity* (pp. 539–561). Elsevier. <https://doi.org/10.1016/B978-0-12-384719-5.00033-2>
- Nekola, J. C., & McGill, B. J. (2014). Scale dependency in the functional form of the distance decay relationship. *Ecography*, 37(4), 309–320. <https://doi.org/10.1111/j.1600-0587.2013.00407.x>
- Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, 26(4), 867–878. <https://doi.org/10.1046/j.1365-2699.1999.00305.x>

- New, T. R., & Thornton, I. W. B. (1988). A pre-vegetation population of crickets subsisting on allochthonous aeolian debris on Anak Krakatau. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 322(1211), 481–485.
<https://doi.org/10.1098/rstb.1988.0140>
- Nosil, P., & Crespi, B. J. (2004). Does gene flow constrain adaptative divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution*, 58(1), 102–112. <https://doi.org/10.1111/j.0014-3820.2004.tb01577.x>
- Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walkink-stick ecotypes: ‘Isolation by Adaptation’ and multiple roles for divergent selection. *Evolution*, 62(2), 316–336. <https://doi.org/10.1111/j.1558-5646.2007.00299.x>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402.
<https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Olden, J. D., LeRoy Poff, N., Douglas, M. R., Douglas, M. E., & Fausch, K. D. (2004). Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution*, 19(1), 18–24. <https://doi.org/10.1016/j.tree.2003.09.010>
- Olden, J. D., & Poff, N. L. (2003). Toward a mechanistic understanding and prediction of biotic homogenization. *The American Naturalist*, 162(4), 442–460.
<https://doi.org/10.1086/378212>
- Oleksa, A., Chybicki, I. J., Gawroński, R., Svensson, G. P., & Burczyk, J. (2013). Isolation by distance in saproxylic beetles may increase with niche specialization. *Journal of Insect Conservation*, 17(2), 219–233. <https://doi.org/10.1007/s10841-012-9499-7>
- Oliver, T. H., Heard, M. S., Isaac, N. J. B., Roy, D. B., Procter, D., Eigenbrod, F., Freckleton, R., Hector, A., Orme, C. D. L., Petchey, O. L., Proença, V., Raffaelli, D., Suttle, K. B., Mace, G. M., Martín-López, B., Woodcock, B. A., & Bullock, J. M. (2015). Biodiversity and resilience of ecosystem functions. *Trends in Ecology & Evolution*, 30(11), 673–684. <https://doi.org/10.1016/j.tree.2015.08.009>
- Onghena, P., & May, R. B. (1995). Pitfalls in computing and interpreting randomization test p values: A commentary on Chen and Dunlap. *Behavior Research Methods, Instruments & Computers*, 27(3), 408–411.

- Orlov, V. N., Kozlovsky, A. I., Okulova, N. M., & Balakirev, A. E. (2007). Postglacial recolonisation of European Russia by the common shrew *Sorex araneus*. *Russian Journal of Theriology*, 6(1), 097–104. <https://doi.org/10.15298/rusjtheriol.06.1.13>
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2018). *caper: Comparative Analyses of Phylogenetics and Evolution in R*. <https://CRAN.R-project.org/package=caper>
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999. <https://doi.org/10.1111/mec.12561>
- Ottati, S., Eberle, J., Rulik, B., Köhler, F., & Ahrens, D. (2022). From DNA barcodes to ecology: Meta-analysis of central European beetles reveal link with species ecology but also to data pattern and gaps. *Ecology and Evolution*, 12(12). <https://doi.org/10.1002/ece3.9650>
- Padgham, M., & Sumner, M. D. (2020). *geodist: Fast, Dependency-Free Geodesic Distance Calculations*. <https://CRAN.R-project.org/package=geodist>
- Palmer, M. W. (2005). Distance decay in an old-growth neotropical forest. *Journal of Vegetation Science*, 16(2), 161–166. <https://doi.org/10.1111/j.1654-1103.2005.tb02351.x>
- Pamilo, P., & Savolainen, O. (1999). Post-glacial colonization, drift, local selection and conservation value of populations: A northern perspective. *Hereditas*, 130(3), 229–238. <https://doi.org/10.1111/j.1601-5223.1999.00229.x>
- Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Pavlek, M., & Mammola, S. (2020). Niche-based processes explaining the distributions of closely related subterranean spiders. *Journal of Biogeography*, 48(1), 118–133. <https://doi.org/10.1111/jbi.13987>
- Pearce, S., Zalucki, M. P., & Hassan, E. (2005). Spider ballooning in soybean and non-crop areas of southeast Queensland. *Agriculture, Ecosystems & Environment*, 105(1–2), 273–281. <https://doi.org/10.1016/j.agee.2004.02.010>

- Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2), 9–13. <https://CRAN.R-project.org/doc/Rnews/>
- Peterson, M. A., & Denno, R. F. (1998). Life-History Strategies and the Genetic Structure of Phytophagous Insect Populations. In S. Mopper & S. Y. Strauss (Eds.), *Genetic Structure and Local Adaptation in Natural Insect Populations* (pp. 263–322). Springer US. https://doi.org/10.1007/978-1-4757-0902-5_12
- Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S., & Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, 300(5625), 1563–1565. <https://doi.org/10.1126/science.1083264>
- Petitpierre, E. (2000). *Coleoptera, Chrysomelidae I. Fauna Ibérica*. Museo Nacional de Ciencias Naturales.
- Phipson, B., & Smyth, G. K. (2010). Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, 9(1). <https://doi.org/10.2202/1544-6115.1585>
- Pinkert, S., Dijkstra, K.-D. B., Zeuss, D., Reudenbach, C., Brandl, R., & Hof, C. (2018). Evolutionary processes, dispersal limitation and climatic history shape current diversity patterns of European dragonflies. *Ecography*, 41(5), 795–804. <https://doi.org/10.1111/ecog.03137>
- Pisias, N. G., & Moore, T. C. (1981). The evolution of Pleistocene climate: A time series approach. *Earth and Planetary Science Letters*, 52(2), 450–458. [https://doi.org/10.1016/0012-821X\(81\)90197-7](https://doi.org/10.1016/0012-821X(81)90197-7)
- Ponomarev, A. V. (2009). New speceis and finds of spiders (Aranei) from the south of Russia and Western Kazakhstan. *Caucasian Entomological Bulletin*, 5(2), 143–146.
- Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part II. *Ecology*, 43(3), 410–432. <https://doi.org/10.2307/1933371>
- Provan, J., & Bennett, K. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, 23(10), 564–571. <https://doi.org/10.1016/j.tree.2008.06.010>
- Pyšek, P., Jarošík, V., Hulme, P. E., Kühn, I., Wild, J., Arianoutsou, M., Bacher, S., Chiron, F., Didžiulis, V., Essl, F., Genovesi, P., Gherardi, F., Hejda, M., Kark, S., Lambdon, P. W., Desprez-Loustau, M.-L., Nentwig, W., Pergl, J., Pobljšaj, K., ... Winter, M.

- (2010). Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences*, 107(27), 12157–12162. <https://doi.org/10.1073/pnas.1002314107>
- Qian, H. (2009). Beta diversity in relation to dispersal ability for vascular plants in North America. *Global Ecology and Biogeography*, 18(3), 327–332. <https://doi.org/10.1111/j.1466-8238.2009.00450.x>
- Qian, H., & Ricklefs, R. E. (2012). Disentangling the effects of geographic distance and environmental dissimilarity on global patterns of species turnover. *Global Ecology and Biogeography*, 21(3), 341–351. <https://doi.org/10.1111/j.1466-8238.2011.00672.x>
- Qian, H., Ricklefs, R. E., & White, P. S. (2005). Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecology Letters*, 8(1), 15–22. <https://doi.org/10.1111/j.1461-0248.2004.00682.x>
- Rand, D. M. (1996). Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. *Conservation Biology*, 10(2), 665–671. <https://doi.org/10.1046/j.1523-1739.1996.10020665.x>
- Rasch, D., Kubinger, K. D., & Moder, K. (2011). The two-sample t test: Pre-testing its assumptions does not pay off. *Statistical Papers*, 52(1), 219–231. <https://doi.org/10.1007/s00362-009-0224-x>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things): *phytools: R package. Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Revelle, W. (2017). *psych: Procedures for Psychological, Psychometric, and Personality Research*. <https://CRAN.R-project.org/package=psych>
- Richardson, D. M., & Whittaker, R. J. (2010). Conservation biogeography - foundations, concepts and challenges: Conservation biogeography: foundations, concepts and challenges. *Diversity and Distributions*, 16(3), 313–320. <https://doi.org/10.1111/j.1472-4642.2010.00660.x>
- Ricklefs, R. E. (2008). Disintegration of the ecological community. *The American Naturalist*, 172(6), 741–750. <https://doi.org/10.1086/593002>
- Rizzo, V., Sánchez-Fernández, D., Alonso, R., Pastor, J., & Ribera, I. (2017). Substratum karstificability, dispersal and genetic structure in a strictly subterranean beetle. *Journal of Biogeography*, 44(11), 2527–2538. <https://doi.org/10.1111/jbi.13074>

- Rocchini, D. (2007). Distance decay in spectral space in analysing ecosystem β -diversity. *International Journal of Remote Sensing*, 28(11), 2635–2644.
<https://doi.org/10.1080/01431160600954712>
- Rodriguez-Artigas, S. M., Ballester, R., & Corronca, J. A. (2016). Factors that influence the beta-diversity of spider communities in northwestern Argentinean Grasslands. *PeerJ*, 4, e1946. <https://doi.org/10.7717/peerj.1946>
- Rosindell, J., Hubbell, S. P., & Etienne, R. S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution*, 26(7), 340–348. <https://doi.org/10.1016/j.tree.2011.03.024>
- Rossi, S., Deslauriers, A., & Morin, H. (2003). Application of the Gompertz equation for the study of xylem cell development. *Dendrochronologia*, 21(1), 33–39.
<https://doi.org/10.1078/1125-7865-00034>
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F -statistics under isolation by distance. *Genetics*, 145(4), 1219–1228.
<https://doi.org/10.1093/genetics/145.4.1219>
- Saito, V. S., Soininen, J., Fonseca-Gessner, A. A., & Siqueira, T. (2015). Dispersal traits drive the phylogenetic distance decay of similarity in Neotropical stream metacommunities. *Journal of Biogeography*, 42(11), 2101–2111. <https://doi.org/10.1111/jbi.12577>
- Samarasekera, G. D. N., Bartell, N. V., Lindgren, B. S., Cooke, J. E. K., Davis, C. S., James, P. M. A., Coltman, D. W., Mock, K. E., & Murray, B. W. (2012). Spatial genetic structure of the mountain pine beetle (*Dendroctonus ponderosae*) outbreak in western Canada: Historical patterns and contemporary dispersal. *Molecular Ecology*, 21(12), 2931–2948. <https://doi.org/10.1111/j.1365-294X.2012.05587.x>
- Santini, L., Antão, L. H., Jung, M., Benítez-López, A., Rapacciuolo, G., Di Marco, M., Jones, F. A. M., Haghkerdar, J. M., & González-Suárez, M. (2021). The interface between Macroecology and Conservation: Existing links and untapped opportunities. *Frontiers of Biogeography*, 13(4). <https://doi.org/10.21425/F5FBG53025>
- Schmitt, T., & Varga, Z. (2012). Extra-Mediterranean refugia: The rule and not the exception? *Frontiers in Zoology*, 9(1), 22. <https://doi.org/10.1186/1742-9994-9-22>
- She, Z., Pan, X., Yue, Z., Shi, X., Gao, Y., Wang, S., Chuai, X., & Wang, J. (2023). Contrasting prokaryotic and eukaryotic community assembly and species coexistence

- in acid mine drainage-polluted waters. *Science of The Total Environment*, 856, 158954. <https://doi.org/10.1016/j.scitotenv.2022.158954>
- Sillero, Brito, Skidmore, & Toxopeus. (2009). Biogeographical patterns derived from remote sensing variables: The amphibians and reptiles of the Iberian Peninsula. *Amphibia-Reptilia*, 30(2), 185–206. <https://doi.org/10.1163/156853809788201207>
- Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*, 241(1), 1–31. <https://doi.org/10.2475/ajs.241.1.1>
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787–792.
- Slatkin, M. (1993). Isolation by Distance in equilibrium and non-equilibrium populations. *Evolution*, 47(1), 264–279. <https://doi.org/10.1111/j.1558-5646.1993.tb01215.x>
- Smith, T. W., & Lundholm, J. T. (2010). Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography*, 33(4), 648–655. <https://doi.org/10.1111/j.1600-0587.2009.06105.x>
- Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology*, 35, 6.
- Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, 31(1), 67–80. <https://doi.org/10.1016/j.tree.2015.11.005>
- Soininen, J. (2014). A quantitative analysis of species sorting across organisms and ecosystems. *Ecology*, 95(12), 3284–3292. <https://doi.org/10.1890/13-2228.1>
- Soininen, J., Heino, J., & Wang, J. (2017). A meta-analysis of nestedness and turnover components of beta diversity across organisms and ecosystems. *Global Ecology and Biogeography*, 27(1), 96–109. <https://doi.org/10.1111/geb.12660>
- Soininen, J., McDonald, R., & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, 30(1), 3–12. <https://doi.org/10.1111/j.0906-7590.2007.04817.x>
- Sørensen, T. A. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, 5, 1–34.

- Stegen, J. C., & Hurlbert, A. H. (2011). Inferring ecological processes from taxonomic, phylogenetic and functional trait β -diversity. *PLoS ONE*, 6(6), e20906. <https://doi.org/10.1371/journal.pone.0020906>
- Steidl, R. J., Hayes, J. P., & Schaubert, E. (1997). Statistical power analysis in wildlife research. *The Journal of Wildlife Management*, 61(2), 270–279.
- Steinbauer, M. J., Dolos, K., Reineking, B., & Beierkuhnlein, C. (2012). Current measures for distance decay in similarity of species composition are influenced by study extent and grain size. *Global Ecology and Biogeography*, 21(12), 1203–1212. <https://doi.org/10.1111/j.1466-8238.2012.00772.x>
- Steinitz, O., Heller, J., Tsoar, A., Rotem, D., & Kadmon, R. (2006). Environment, dispersal and patterns of species similarity. *Journal of Biogeography*, 33(6), 1044–1054. <https://doi.org/10.1111/j.1365-2699.2006.01473.x>
- Stephens, C. (2004). Selection, drift, and the “forces” of evolution. *Philosophy of Science*, 71(4), 550–570. <https://doi.org/10.1086/423751>
- Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2009). Refugia revisited: Individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, 277(1682), 661–671. <https://doi.org/10.1098/rspb.2009.1272>
- Stewart, L., Alsos, I. G., Bay, C., Breen, A. L., Brochmann, C., Boulanger-Lapointe, N., Broennimann, O., Bültmann, H., Bøcher, P. K., Damgaard, C., Daniëls, F. J. A., Ehrich, D., Eidesen, P. B., Guisan, A., Jónsdóttir, I. S., Lenoir, J., le Roux, P. C., Lévesque, E., Luoto, M., ... Pellissier, L. (2016). The regional species richness and genetic diversity of Arctic vegetation reflect both past glaciations and current climate: Glaciation effect on arctic vegetation diversity. *Global Ecology and Biogeography*, 25(4), 430–442. <https://doi.org/10.1111/geb.12424>
- Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, 63(1), 31–45. <https://doi.org/10.1146/annurev-ento-020117-043348>
- Svenning, J.-C., Eiserhardt, W. L., Normand, S., Ordonez, A., & Sandel, B. (2015). The Influence of Paleoclimate on present-day patterns in biodiversity and ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 551–572. <https://doi.org/10.1146/annurev-ecolsys-112414-054314>

- Svenning, J.-C., Fløjgaard, C., & Baselga, A. (2011). Climate, history and neutrality as drivers of mammal beta diversity in Europe: Insights from multiscale deconstruction: Deconstructing mammal beta diversity. *Journal of Animal Ecology*, 80(2), 393–402. <https://doi.org/10.1111/j.1365-2656.2010.01771.x>
- Svenning, J.-C., Normand, S., & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: Insights from species distribution modelling. *Journal of Ecology*, 96(6), 1117–1127. <https://doi.org/10.1111/j.1365-2745.2008.01422.x>
- Svenning, J.-C., & Skov, F. (2007). Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, 10(6), 453–460. <https://doi.org/10.1111/j.1461-0248.2007.01038.x>
- Szymkowiak, P., Górski, G., & Bajerlein, D. (2007). Passive dispersal in arachnids. *Biological Letters*, 44(2), 75–101.
- Taberlet, P., & Cheddadi, R. (2002). Quaternary refugia and persistence of biodiversity. *Science*, 297(5589), 2009–2010.
- Taberlet, P., Swenson, J. E., Sandegren, F., & Bjärvall, A. (1995). Localization of a contact zone between two highly divergent mitochondrial DNA Lineages of the brown bear *Ursus arctos* in Scandinavia. *Conservation Biology*, 9(5), 1255–1261. <https://doi.org/10.1046/j.1523-1739.1995.951255.x>
- Tang, F. H. M., Lenzen, M., McBratney, A., & Maggi, F. (2021). Risk of pesticide pollution at the global scale. *Nature Geoscience*, 14(4), 206–210. <https://doi.org/10.1038/s41561-021-00712-5>
- Tedesco, P. A., Leprieux, F., Hugueny, B., Brosse, S., Dürr, H. H., Beauchard, O., Busson, F., & Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism: Global riverine fish endemism patterns. *Global Ecology and Biogeography*, 21(10), 977–987. <https://doi.org/10.1111/j.1466-8238.2011.00749.x>
- Thieltges, D. W., Ferguson, M. A. D., Jones, C. S., Krakau, M., de Montaudouin, X., Noble, L. R., Reise, K., & Poulin, R. (2009). Distance decay of similarity among parasite communities of three marine invertebrate hosts. *Oecologia*, 160(1), 163–173. <https://doi.org/10.1007/s00442-009-1276-2>
- Tilman, D. (2004). Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the*

- National Academy of Sciences*, 101(30), 10854–10861.
<https://doi.org/10.1073/pnas.0403458101>
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234. <https://doi.org/10.2307/143141>
- Tonkin, J. D., Stoll, S., Jähnig, S. C., & Haase, P. (2016). Contrasting metacommunity structure and beta diversity in an aquatic-floodplain system. *Oikos*, 125(5), 686–697. <https://doi.org/10.1111/oik.02717>
- Tovar-Sánchez, E., Hernández-Plata, I., Martínez, M. S., Valencia-Cuevas, L., & Galante, P. M. (2018). Heavy Metal Pollution as a Biodiversity Threat. In H. E.-D. M. Saleh & R. F. Aglan (Eds.), *Heavy Metals*. InTech. <https://doi.org/10.5772/intechopen.74052>
- Tsiafouli, M. A., Thébault, E., Sgardelis, S. P., de Ruiter, P. C., van der Putten, W. H., Birkhofer, K., Hemerik, L., de Vries, F. T., Bardgett, R. D., Brady, M. V., Bjornlund, L., Jørgensen, H. B., Christensen, S., Hertefeldt, T. D., Hotes, S., Gera Hol, W. H., Frouz, J., Liiri, M., Mortimer, S. R., ... Hedlund, K. (2015). Intensive agriculture reduces soil biodiversity across Europe. *Global Change Biology*, 21(2), 973–985. <https://doi.org/10.1111/gcb.12752>
- Tuomisto, H. (2010a). A diversity of beta diversities: Straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, 33(1), 2–22. <https://doi.org/10.1111/j.1600-0587.2009.05880.x>
- Tuomisto, H. (2010b). A diversity of beta diversities: Straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography*, 33(1), 23–45. <https://doi.org/10.1111/j.1600-0587.2009.06148.x>
- Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of Western Amazonian forests. *Science*, 299(5604), 241–244. <https://doi.org/10.1126/science.1078037>
- Tyschchenko, B. (1965). A new genus and new species of spiders (Aranei) of Kazjtán. ..Э и Т ОМОЛОГИЧЕСКОЕ ОБОЗРЕНИЕ, 44(3), 696–704.
- Ulrich, W., Almeida-Neto, M., & Gotelli, N. J. (2009). A consumer's guide to nestedness analysis. *Oikos*, 118(1), 3–17. <https://doi.org/10.1111/j.1600-0706.2008.17053.x>
- Vellend, M. (2001). Do commonly used indices of β -diversity measure species turnover? *Journal of Vegetation Science*, 12(4), 545–552. <https://doi.org/10.2307/3237006>

- Vellend, M. (2003). Island biogeography of genes and species. *The American Naturalist*, 162(3), 358–365. <https://doi.org/10.1086/377189>
- 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., & Geber, M. A. (2005). Connections between species diversity and genetic diversity: Species diversity and genetic diversity. *Ecology Letters*, 8(7), 767–781. <https://doi.org/10.1111/j.1461-0248.2005.00775.x>
- Vellend, M., Srivastava, D. S., Anderson, K. M., Brown, C. D., Jankowski, J. E., Kleynhans, E. J., Kraft, N. J. B., Letaw, A. D., Macdonald, A. A. M., Maclean, J. E., Myers-Smith, I. H., Norris, A. R., & Xue, X. (2014). Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos*, 123(12), 1420–1430. <https://doi.org/10.1111/oik.01493>
- Villéger, S., Grenouillet, G., & Brosse, S. (2013). Decomposing functional β -diversity reveals that low functional β -diversity is driven by low functional turnover in European fish assemblages. *Global Ecology and Biogeography*, 22(6), 671–681. <https://doi.org/10.1111/geb.12021>
- Wang, H., Chen, Z., & Zhang, P. (2022). Spatial autocorrelation and temporal convergence of PM2.5 concentrations in Chinese cities. *International Journal of Environmental Research and Public Health*, 19(21), 13942. <https://doi.org/10.3390/ijerph192113942>
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649–5662. <https://doi.org/10.1111/mec.12938>
- Wetzel, C. E., Bicudo, D. de C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., & Bini, L. M. (2012). Distance decay of similarity in neotropical diatom communities. *PLoS ONE*, 7(9), e45071. <https://doi.org/10.1371/journal.pone.0045071>
- Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3), 279–338.
- Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 21(2–3), 213–251. <https://doi.org/10.2307/1218190>
- Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation Biogeography: Assessment and prospect. *Diversity and Distributions*, 11(1), 3–23. <https://doi.org/10.1111/j.1366-9516.2005.00143.x>

- Wickham, H. (2016). *scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>
- Widmer, A., & Lexer, C. (2001). Glacial refugia: Sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution*, 16(6), 267–269. [https://doi.org/10.1016/S0169-5347\(01\)02163-2](https://doi.org/10.1016/S0169-5347(01)02163-2)
- Wiens, J. J. (2004). Speciation and ecology revised: Phylogenetic niche conservatism and the origin of species. *Evolution*, 58(1), 193–197. <https://doi.org/10.1111/j.0014-3820.2004.tb01586.x>
- Wiens, J. J., & Graham, C. H. (2005). Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, 36(1), 519–539. <https://doi.org/10.1146/annurev.ecolsys.36.102803.095431>
- Wiley, E. O., & Lieberman, B. S. (2011). *Phylogenetics: The Theory of Phylogenetic Systematics* (2nd ed). Wiley-Blackwell.
- Willis, J. C. (1922). *Age and Area: A Study in Geographical Distribution and Origin of Species*. CUP Archive.
- Willner, W., Di Pietro, R., & Bergmeier, E. (2009). Phytogeographical evidence for post-glacial dispersal limitation of European beech forest species. *Ecography*, 32(6), 1011–1018. <https://doi.org/10.1111/j.1600-0587.2009.05957.x>
- Wilson, A. B., & Eigenmann Veraguth, I. (2010). The impact of Pleistocene glaciation across the range of a widespread European coastal species. *Molecular Ecology*, 19(20), 4535–4553. <https://doi.org/10.1111/j.1365-294X.2010.04811.x>
- Wise, D. H. (1993). *Spiders in Ecological Webs*. Cambridge University Press.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114–138.
- Wu, L., Si, X., Didham, R. K., Ge, D., & Ding, P. (2017). Dispersal modality determines the relative partitioning of beta diversity in spider assemblages on subtropical land-bridge islands. *Journal of Biogeography*, 44(9), 2121–2131. <https://doi.org/10.1111/jbi.13007>
- Yan, N. D., Keller, W., Somers, K. M., Pawson, T. W., & Girard, R. E. (1996). Recovery of crustacean zooplankton communities from acid and metal contamination: Comparing manipulated and reference lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(6), 1301–1327. <https://doi.org/10.1139/f96-065>
- Yang, Q., Weigelt, P., Fristoe, T. S., Zhang, Z., Kreft, H., Stein, A., Seebens, H., Dawson, W., Essl, F., König, C., Lenzner, B., Pergl, J., Pouteau, R., Pyšek, P., Winter, M.,

- Ebel, A. L., Fuentes, N., Giehl, E. L. H., Kartesz, J., ... van Kleunen, M. (2021). The global loss of floristic uniqueness. *Nature Communications*, 12(1), 7290. <https://doi.org/10.1038/s41467-021-27603-y>
- Zhang, M., Zhang, T., Zhou, L., Lou, W., Zeng, W., Liu, T., Yin, H., Liu, H., Liu, X., Mathivanan, K., Praburaman, L., & Meng, D. (2022b). Soil microbial community assembly model in response to heavy metal pollution. *Environmental Research*, 213, 113576. <https://doi.org/10.1016/j.envres.2022.113576>
- Zhang, Q., He, D., Wu, H., Shi, W., & Chen, C. (2018). Local-scale determinants of arboreal spider beta diversity in a temperate forest: Roles of tree architecture, spatial distance, and dispersal capacity. *PeerJ*, 6, e5596. <https://doi.org/10.7717/peerj.5596>
- Zhang, Y., Rashid, A., Guo, S., Jing, Y., Zeng, Q., Li, Y., Adyari, B., Yang, J., Tang, L., Yu, C.-P., & Sun, Q. (2022a). Spatial autocorrelation and temporal variation of contaminants of emerging concern in a typical urbanizing river. *Water Research*, 212, 118120. <https://doi.org/10.1016/j.watres.2022.118120>
- Zhao, Z., Li, S., Liu, J., Peng, J., & Wang, Y. (2015). The distance decay of similarity in climate variation and vegetation dynamics. *Environmental Earth Sciences*, 73(8), 4659–4670. <https://doi.org/10.1007/s12665-014-3751-2>
- Zunino, M., & Zullini, A. (2003). *Biogeografía, la Dimensión Espacial de la Evolución* (1st ed). Fondo de Cultura Económica.

- Appendix 1.1: Gompertz function -

Appendix 1.1: Gompertz function

Distance-decay patterns were modelled as the inverse of the Gompertz function in order to have a negative slope model. Besides, given that the parameter a controls the inferior asymptote of the Gompertz function (inferior asymptote = $1 - a$), we forced the value of this parameter to 1. This makes the bottom asymptote 0 and allows the existence of completely different biological communities, an essential requirement for a distance-decay fitting function. Therefore, the Gompertz function can accommodate the most used (dis)similarity indices' values, which range from zero (completely different assemblages) to one (completely similar assemblages). Notably, this is a major drawback of negative exponential and power-law models, as they cannot be bounded between 0 and 1. Both functions have an inferior asymptote in 0, but none of them has an upper bound of 1 and therefore their intercepts can reach higher values, which would have no biological interpretation.

The Gompertz model used to model distance-decay relationships in this study is:

$$y = 1 - e^{-be^{-cx}} + \varepsilon, \quad (1)$$

where b is a position parameter, c decay rate, x spatial distance and ε a random error. Assuming that distances can take any positive value, it can be proved that the Gompertz function (eq. 1) is bounded in $[0, 1]$ for all possible values of the spatial and environmental distances computing its superior and inferior limits in the interval $[0, \infty]$.

$$\begin{aligned} \lim_{x \rightarrow 0} 1 - e^{-be^{-cx}} &= 1 - e^{-b} \text{ (intercept)} & \lim_{b \rightarrow \infty} 1 - e^{-b} &= 1 \\ \lim_{x \rightarrow \infty} 1 - e^{-be^{-cx}} &= 0 \end{aligned}$$

- Appendix 1.2: Code for simulations -

Appendix 1.2: Code for simulations

```
# - . # - . # - . # - . # - . # - . # - . # - . # - . #
# SIMULATIONS TO ASSESS THE RELATIONSHIP BETWEEN SPECIES RANGE SIZE AND
# DISTANCE-DECAY FUNCTIONAL FORM
# - . # - . # - . # - . # - . # - . # - . # - . # - . #

# Supplementary R script for manuscript: Martin-Devasa R, Martinez-Santalla
# S, Gomez-Rodriguez R, Crujeiras RM, Baselga A: Species range size shapes
# distance-decay in community similarity.

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# SESSION PREPARATION
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

#Load libraries

library(sp)
library(geoR)
library(rgeos)
library(betapart)
library(minpack.lm)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# SIMULATION STUDY: GOAL
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# To assess the functional form of the best fitting model in distance-decay
# curves of community similarity and spatial distance when species with
# markedly different range sizes are studied.
# Different simulation scenarios are considered to account for differences
# in range size among species as well as differences in sample size and
# probability of species detection. In the case of range size, we simulated
# (i) species with different range sizes and also (ii) different proportions
# of range sizes (widespread vs. spatially-restricted species) in the
# regional pool.
# To generate distance-decay data, we simulated a virtual regional pool and
# a set of virtual local communities (sampling sites). This provides a
# biological similarity matrix and a spatial distance matrix.
# Then, we fitted a power-law, a negative exponential and a Gompertz model
# to the data in order to assess which model best fits the distance-decay
# curves in each simulation scenario.

## . . . . .
## GENERATION OF SPECIES RANGES IN A VIRTUAL SCENARIO

## Virtual scenario properties:
# We created a 3000*3000 units virtual landscape where we placed 300
# circular polygons representing virtual species ranges.
# landscape = 3000 # Virtual landscape dimensions
# species = 300 # Number of species (species ranges are represented as
# circular polygons)

# Number of virtual sampling sites
nsites = c(10, 50, 500, 5000)
```

```

# Probability of species detection
prob_detec = 1 # Analyses are done for perfect species detection
               (prob_detec = 1) and three different levels of incomplete species detection
               (prob_detec = 0.75, 0.5 and 0.25)

## Generation of species ranges (as circular polygons):
# We simulated species with different range sizes and also different
# proportions of range sizes (widespread vs. spatially-restricted species),
# which is the key to create the main different simulation scenarios.
# This simulation study is exemplified for ratio-class 1, Size-class A in
# Martin-Devasa et al. 2022
    nu.w = 10 # Number of widespread species
    min.w = 3000; max.w = 5000 # Minimum and maximum polygon diameters of
    widespread species
    min.s = 500; max.s = 1000 # Minimum and maximum polygon diameters of
    spatially-restricted species

# For other scenarios, change the number of widespread species, the number
# of spatially-restricted species and/or the polygon diameters as follows:
# Ratio-classes: For ratio-class 2 nu.w = 20; for ratio-class 3 nu.w =
200
# Size-classes: For size-class B min.s = 1000, max.s = 1500; for size-
class C min.s = 2000, max.s = 2500

## . . . . .
## DATA STORAGE

# Generation of empty matrices, to store the AIC and pseudo-r2 values of
# each distance-decay model (power-law, negative exponential and Gompertz)
# Data is stored for different sample sizes: n = 10; n = 50; n = 500 and n
= 5000
    aic = matrix(ncol = length(nsites), nrow = 3, dimnames = list(c("Power-
    law", "Negative exponential", "Gompertz"), c("10", "50", "500",
    "5000")))
    pseudo_r2 = matrix(ncol = length(nsites), nrow = 3, dimnames =
    list(c("Power-law", "Negative exponential", "Gompertz"), c("10", "50",
    "500", "5000")))

# Seed
    set.seed(123456)

## . . . . .
# A. SIMULATION OF DISTANCE-DECAY DATA

    ### A.1. VIRTUAL SPECIES POLYGONS: (i) Central coordinates and diameters
    # Widespread species: Central coordinates (prt_gx, prt_gy) and radius
    (rad_g) of the polygons representing species ranges
    prt_gx <- sample(seq(0, landscape, by = 0.01), size = nu.w, replace = T)
# Random sampling of spatial coordinate along the x-axis of the simulation
landscape (size = 3000 x 3000)
    prt_gy <- sample(seq(0, landscape, by = 0.01), size = nu.w, replace = T)
# Random sampling of spatial coordinate along the y-axis of the simulation
landscape (size = 3000 x 3000)

    rad_g <- sample(seq(min.w/2,max.w/2), size = length(prt_gx), replace =
T) # Random sampling of radius values from the uniform distribution of
polygon diameters for widespread species (radius = diameter/2)

```

```

# Spatially-restricted species: Central coordinates (prt_mx, prt_my) and
radius (rad_m) of the polygons representing species ranges
prt_mx <- sample(seq(0, landscape, by = 0.01), size = species - nu.w,
replace = T) # Random sampling of spatial coordinate along the x-axis
of the simulation landscape (size = 3000 x 3000)
prt_my <- sample(seq(0, landscape, by = 0.01), size = species - nu.w,
replace = T) # Random sampling of spatial coordinate along the y-axis
of the simulation landscape (size = 3000 x 3000)

rad_m <- sample(seq(min.s/2,max.s/2), size = length(prt_mx), replace =
T) # Random sampling of radius values from the uniform
distribution of polygon diameters for spatially-restricted species (radius
= diameter/2)

# Merging of the vectors with polygon parameters
ejex <- c(prt_gx, prt_mx); ejey <- c(prt_gy, prt_my) # Central
coordinates of all species polygons
radio <- c(rad_g, rad_m) # Radius of all
species polygons

### (ii) Transformation of central coordinates and diameters into
spatial polygons
# Vector with 1000 values between 0 and 2*pi. These values are needed in
the formula to build circles
t <- seq(0, 2*pi, length.out = 1000)

lis=list() # List to store polygons
for (p in 1:species){ # Loop to create circular polygons from central
coordinates and radius information

  x <- ejex[p] + cos(t)*radio[p] # Definition of circles (x axis)
  y <- ejey[p] + sin(t)*radio[p] # Definition of circles (y axis)

  dep <- SpatialPoints(cbind(x, y)) # Convert circle points to spatial
points
  pol <- Polygon(dep) # Convert spatial points to polygon
  pols <- Polygons(list(pol), p) # Storage of new circular polygon
  lis[[p]] <- pols # Storage polygons as list (needed
for the SpatialPolygons function)

}

# Conversion of polygon list into a spatial polygons object
SPlgs <- SpatialPolygons(lis, proj4string = CRS("+proj=longlat
+datum=WGS84")) # Add projection

### A.2. PRESENCE/ABSENCE OF SPECIES IN VIRTUAL SAMPLING SITES

# Sampling sites are randomly allocated in the virtual landscape. The
species polygons intersecting each sampling site represent a local
community.

# A loop is built to generate data with different sample sizes (i.e.
different number of virtual local communities)

```

```

## VIRTUAL SAMPLING SITES: Generation of virtual sampling sites within
the landscape

grc <- grf(n = nsites[i], grid = "irreg", cov.pars =
c(sigma=1,phi=4000), nsim=1, xlims = c(0, landscape), ylims = c(0,
landscape), cov.model = "exponential", messages = F) # Generation of
spatial coordinates within the virtual landscape (using an autocorrelated
normal variable)
pon <- grc$coord # Coordinates of virtual sampling sites
poin <- SpatialPoints(pon, proj4string = CRS("+proj=longlat
+datum=WGS84")) # Transform coordinates into spatial points

## PRESENCE/ABSENCE DATA
# Assessment of the species present in each sampling site
Results <- gIntersects(SPlgs, poin, byid = TRUE)

# Conversion of the logical (true/false) matrix into a 1 (present) and
0 (absence) matrix
Resul <- Results * rbinom(length(Results), 1, probab_detec)

### A.3 SIMILARITY AND DISTANCE MATRICES FOR DISTANCE-DECAY ANALYSES
## Similarity matrix. Computed with the Simpson's similarity index
ydis <- 1 - beta.pair(Resul)$beta.sim

## Spatial distance matrix. Computed as the euclidean distance between
sampling points
xdis <- dist(poin@coords, method = "euclidean")

## . . . . .
# B. DISTANCE-DECAY MODELS AND EVALUATION

# Fit distance-decay models
mod_pl<-nlsLM(ydis ~ a*xdis^-b, start = list(a = 1, b = 0), trace = F)
# Non-linear regression of biological similarity with spatial distance
(power-law model)
mod_ex<-nlsLM(ydis ~ a*exp(-b*xdis), start=list(a = 1, b = 0), trace =
F) # Non-linear regression of biological similarity with
spatial distance (negative exponential model)
mod_go<-nlsLM(ydis ~ 1-(exp(-b*exp(-c*xdis))), start = list(b = 1, c =
0), trace = F) # Non-linear regression of biological similarity with
spatial distance (Gompertz model)

#Evaluate the best fitting model
# AIC
aic[,i] <- c(AIC(mod_pl), AIC(mod_ex), AIC(mod_go))

# pseudo-r2
null_m<-lm(ydis ~ 1)
r2_pl <- 1-deviance(mod_pl)/deviance(null_m)
r2_ex <- 1-deviance(mod_ex)/deviance(null_m)
r2_go <- 1-deviance(mod_go)/deviance(null_m)
pseudo_r2[,i] <- c(r2_pl, r2_ex, r2_go)

}

# Matrices with the models evaluation results
aic
pseudo_r2

```

- Appendix 1.3: Supplementary tables Chapter I -

Appendix 1.3: Supplementary tables Chapter I

Table S1.1. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations of different frequency distribution of range sizes (from a higher proportion of spatially-restricted species (ratio-class 1) to a smaller proportion (ratio-class 3)), and different range size of spatially-restricted species (from smaller ranges of spatially-restricted species (size-class A) to a larger ranges of spatially-restricted species (size-class C)), and species detection probability = 1. Best fitting models are shown in bold.

		Size-class A							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.85	-117	0.67	-2077	0.63	-2.3E+05	0.64	-2.4E+07
	Negative exponential	0.71	-89	0.66	-2040	0.56	-2.1E+05	0.57	-2.4E+07
	Gompertz	0.66	-82	0.62	-1912	0.53	-2.0E+05	0.54	-2.4E+07
Ratio-class 2	Power-law	0.67	-78	0.67	-2185	0.56	-2.0E+05	0.59	-2.1E+07
	Negative exponential	0.72	-85	0.76	-2547	0.60	-2.1E+05	0.64	-2.3E+07
	Gompertz	0.72	-85	0.74	-2473	0.59	-2.1E+05	0.62	-2.2E+07
Ratio-class 3	Power-law	0.58	-88	0.52	-2218	0.49	-2.0E+05	0.49	-2.1E+07
	Negative exponential	0.80	-121	0.78	-3160	0.75	-2.9E+05	0.75	-3.0E+07
	Gompertz	0.88	-143	0.86	-3703	0.85	-3.5E+05	0.85	-3.6E+07
		Size-class B							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.85	-91	0.65	-1339	0.65	-1.6E+05	0.64	-1.5E+07
	Negative exponential	0.81	-81	0.88	-2667	0.86	-2.7E+05	0.86	-2.7E+07
	Gompertz	0.75	-68	0.86	-2466	0.83	-2.5E+05	0.84	-2.5E+07
Ratio-class 2	Power-law	0.89	-100	0.72	-1862	0.70	-1.9E+05	0.70	-1.9E+07
	Negative exponential	0.92	-113	0.89	-2959	0.84	-2.7E+05	0.85	-2.8E+07
	Gompertz	0.87	-94	0.85	-2598	0.80	-2.4E+05	0.81	-2.5E+07
Ratio-class 3	Power-law	0.65	-95	0.57	-2328	0.54	-2.1E+05	0.54	-2.2E+07
	Negative exponential	0.84	-131	0.82	-3384	0.78	-3.0E+05	0.78	-3.1E+07
	Gompertz	0.88	-145	0.87	-3788	0.85	-3.5E+05	0.85	-3.6E+07
		Size-class C							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.68	-26	0.54	-161	0.53	-2.8E+04	0.53	-3.2E+06
	Negative exponential	0.89	-73	0.87	-1697	0.86	-1.7E+05	0.86	-1.8E+07
	Gompertz	0.96	-114	0.94	-2747	0.92	-2.6E+05	0.93	-2.7E+07
Ratio-class 2	Power-law	0.73	-33	0.56	-401	0.55	-4.4E+04	0.55	-4.6E+06
	Negative exponential	0.92	-90	0.88	-2020	0.87	-2.0E+05	0.87	-2.1E+07
	Gompertz	0.98	-155	0.95	-3029	0.94	-2.9E+05	0.94	-2.9E+07
Ratio-class 3	Power-law	0.65	-80	0.59	-2016	0.56	-1.8E+05	0.56	-1.9E+07
	Negative exponential	0.86	-119	0.86	-3318	0.83	-3.0E+05	0.83	-3.1E+07
	Gompertz	0.89	-133	0.90	-3774	0.89	-3.5E+05	0.89	-3.6E+07

Table S1.2. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations of different frequency distribution of range sizes (from a higher proportion of spatially-restricted species (ratio-class 1) to a smaller proportion (ratio-class 3)), and different range size of spatially-restricted species (from smaller ranges of spatially-restricted species (size-class A) to a larger ranges of spatially-restricted species (size-class C)), and species detection probability = 0.75. Best fitting models are shown in bold.

		Size-class A							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.70	-89	0.42	-1687	0.42	-1.9E+05	0.42	-1.9E+07
	Negative exponential	0.58	-74	0.36	-1555	0.36	-1.8E+05	0.37	-1.8E+07
	Gompertz	0.55	-71	0.34	-1530	0.35	-1.8E+05	0.36	-1.8E+07
Ratio-class 2	Power-law	0.45	-59	0.40	-1512	0.42	-1.9E+05	0.41	-1.8E+07
	Negative exponential	0.49	-62	0.40	-1517	0.46	-2.0E+05	0.44	-1.9E+07
	Gompertz	0.49	-62	0.39	-1503	0.45	-2.0E+05	0.44	-1.9E+07
Ratio-class 3	Power-law	0.51	-101	0.41	-2312	0.43	-2.5E+05	0.44	-2.5E+07
	Negative exponential	0.73	-127	0.65	-2969	0.66	-3.1E+05	0.66	-3.1E+07
	Gompertz	0.76	-133	0.69	-3101	0.69	-3.2E+05	0.70	-3.3E+07
		Size-class B							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.88	-110	0.67	-2118	0.59	-1.9E+05	0.58	-1.9E+07
	Negative exponential	0.88	-109	0.81	-2771	0.79	-2.8E+05	0.78	-2.7E+07
	Gompertz	0.85	-100	0.79	-2682	0.78	-2.7E+05	0.77	-2.6E+07
Ratio-class 2	Power-law	0.80	-100	0.62	-2037	0.62	-2.1E+05	0.62	-2.1E+07
	Negative exponential	0.86	-117	0.70	-2336	0.76	-2.7E+05	0.76	-2.7E+07
	Gompertz	0.85	-114	0.68	-2253	0.74	-2.6E+05	0.74	-2.6E+07
Ratio-class 3	Power-law	0.58	-108	0.44	-2370	0.47	-2.6E+05	0.48	-2.6E+07
	Negative exponential	0.78	-137	0.68	-3036	0.69	-3.2E+05	0.70	-3.3E+07
	Gompertz	0.80	-142	0.70	-3148	0.72	-3.4E+05	0.72	-3.4E+07
		Size-class C							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.66	-46	0.53	-943	0.52	-9.9E+04	0.52	-9.6E+06
	Negative exponential	0.85	-85	0.82	-2135	0.84	-2.3E+05	0.84	-2.3E+07
	Gompertz	0.90	-100	0.87	-2501	0.88	-2.7E+05	0.88	-2.7E+07
Ratio-class 2	Power-law	0.72	-59	0.53	-984	0.54	-1.1E+05	0.53	-1.1E+07
	Negative exponential	0.91	-108	0.82	-2147	0.85	-2.5E+05	0.85	-2.5E+07
	Gompertz	0.94	-130	0.87	-2515	0.89	-2.9E+05	0.89	-2.9E+07
Ratio-class 3	Power-law	0.61	-97	0.49	-2218	0.52	-2.4E+05	0.52	-2.4E+07
	Negative exponential	0.81	-129	0.76	-3126	0.76	-3.3E+05	0.77	-3.3E+07
	Gompertz	0.83	-135	0.79	-3294	0.79	-3.4E+05	0.79	-3.5E+07

Table S1.3. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations of different frequency distribution of range sizes (from a higher proportion of spatially-restricted species (ratio-class 1) to a smaller proportion (ratio-class 3)), and different range size of spatially-restricted species (from smaller ranges of spatially-restricted species (size-class A) to a larger ranges of spatially-restricted species (size-class C)), and species detection probability = 0.5. Best fitting models are shown in bold.

		Size-class A							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.47	-72	0.22	-1323	0.19	-1.4E+05	0.20	-1.5E+07
	Negative exponential	0.46	-71	0.20	-1290	0.17	-1.4E+05	0.17	-1.4E+07
	Gompertz	0.45	-71	0.19	-1282	0.17	-1.4E+05	0.17	-1.4E+07
Ratio-class 2	Power-law	0.24	-65	0.23	-1563	0.22	-1.5E+05	0.21	-1.5E+07
	Negative exponential	0.20	-64	0.25	-1581	0.23	-1.5E+05	0.22	-1.5E+07
	Gompertz	0.20	-63	0.24	-1578	0.23	-1.5E+05	0.22	-1.5E+07
Ratio-class 3	Power-law	0.23	-118	0.31	-2668	0.32	-2.8E+05	0.32	-2.8E+07
	Negative exponential	0.42	-130	0.50	-3051	0.49	-3.1E+05	0.48	-3.1E+07
	Gompertz	0.43	-131	0.51	-3077	0.50	-3.2E+05	0.49	-3.2E+07
		Size-class B							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.55	-72	0.43	-2039	0.44	-2.1E+05	0.45	-2.2E+07
	Negative exponential	0.57	-74	0.54	-2287	0.59	-2.5E+05	0.60	-2.6E+07
	Gompertz	0.57	-74	0.53	-2276	0.58	-2.5E+05	0.60	-2.6E+07
Ratio-class 2	Power-law	0.77	-107	0.46	-2279	0.45	-2.2E+05	0.45	-2.3E+07
	Negative exponential	0.68	-94	0.52	-2423	0.56	-2.5E+05	0.55	-2.5E+07
	Gompertz	0.65	-90	0.51	-2400	0.55	-2.5E+05	0.54	-2.5E+07
Ratio-class 3	Power-law	0.28	-122	0.34	-2743	0.36	-2.9E+05	0.35	-2.9E+07
	Negative exponential	0.46	-134	0.52	-3123	0.52	-3.2E+05	0.51	-3.2E+07
	Gompertz	0.47	-135	0.52	-3146	0.53	-3.3E+05	0.52	-3.3E+07
		Size-class C							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.67	-75	0.48	-1719	0.48	-1.7E+05	0.48	-1.7E+07
	Negative exponential	0.86	-115	0.76	-2646	0.77	-2.7E+05	0.77	-2.8E+07
	Gompertz	0.89	-123	0.78	-2770	0.79	-2.9E+05	0.79	-2.9E+07
Ratio-class 2	Power-law	0.72	-84	0.51	-1847	0.49	-1.9E+05	0.49	-1.9E+07
	Negative exponential	0.86	-116	0.76	-2748	0.77	-2.9E+05	0.77	-2.9E+07
	Gompertz	0.88	-122	0.78	-2855	0.79	-3.0E+05	0.79	-3.0E+07
Ratio-class 3	Power-law	0.36	-115	0.41	-2723	0.42	-2.8E+05	0.42	-2.9E+07
	Negative exponential	0.57	-133	0.62	-3258	0.61	-3.3E+05	0.61	-3.3E+07
	Gompertz	0.59	-135	0.63	-3294	0.63	-3.4E+05	0.62	-3.4E+07

Table S1.4. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations of different frequency distribution of range sizes (from a higher proportion of spatially-restricted species (ratio-class 1) to a smaller proportion (ratio-class 3)), and different range size of spatially-restricted species (from smaller ranges of spatially-restricted species (size-class A) to a larger ranges of spatially-restricted species (size-class C)), and species detection probability = 0.25. Best fitting models are shown in bold.

		Size-class A							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.02	-19	0.12	-1233	0.05	-9.5E+04	0.05	-9.8E+06
	Negative exponential	0.03	-19	0.09	-1188	0.05	-9.5E+04	0.04	-9.7E+06
	Gompertz	0.03	-19	0.08	-1186	0.05	-9.5E+04	0.04	-9.7E+06
Ratio-class 2	Power-law	0.00	10	0.08	-1257	0.05	-1.0E+05	0.05	-1.1E+07
	Negative exponential	0.00	10	0.08	-1263	0.05	-1.1E+05	0.05	-1.1E+07
	Gompertz	0.00	10	0.08	-1262	0.05	-1.1E+05	0.05	-1.1E+07
Ratio-class 3	Power-law	0.07	-80	0.10	-2740	0.12	-2.8E+05	0.12	-2.8E+07
	Negative exponential	0.12	-83	0.17	-2834	0.17	-2.9E+05	0.17	-2.9E+07
	Gompertz	0.12	-83	0.17	-2837	0.17	-2.9E+05	0.17	-2.9E+07
		Size-class B							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.09	-57	0.27	-2359	0.19	-2.2E+05	0.18	-2.2E+07
	Negative exponential	0.08	-57	0.32	-2435	0.25	-2.2E+05	0.25	-2.3E+07
	Gompertz	0.08	-56	0.32	-2434	0.25	-2.2E+05	0.25	-2.3E+07
Ratio-class 2	Power-law	0.05	-87	0.17	-2105	0.16	-2.1E+05	0.17	-2.1E+07
	Negative exponential	0.03	-86	0.19	-2137	0.20	-2.1E+05	0.21	-2.2E+07
	Gompertz	0.03	-86	0.19	-2135	0.20	-2.1E+05	0.20	-2.1E+07
Ratio-class 3	Power-law	0.06	-81	0.11	-2840	0.13	-2.9E+05	0.13	-2.9E+07
	Negative exponential	0.11	-83	0.17	-2926	0.19	-3.0E+05	0.19	-3.0E+07
	Gompertz	0.11	-83	0.17	-2928	0.19	-3.0E+05	0.19	-3.0E+07
		Size-class C							
		n = 10		n = 50		n = 500		n = 5000	
		Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC	Pseudo- r^2	AIC
Ratio-class 1	Power-law	0.33	-97	0.36	-2564	0.32	-2.6E+05	0.32	-2.6E+07
	Negative exponential	0.45	-106	0.54	-2977	0.51	-3.0E+05	0.51	-3.0E+07
	Gompertz	0.46	-106	0.55	-2995	0.51	-3.0E+05	0.52	-3.0E+07
Ratio-class 2	Power-law	0.31	-78	0.34	-2651	0.31	-2.6E+05	0.31	-2.6E+07
	Negative exponential	0.36	-82	0.50	-3003	0.48	-2.9E+05	0.49	-3.0E+07
	Gompertz	0.36	-82	0.51	-3017	0.49	-2.9E+05	0.50	-3.0E+07
Ratio-class 3	Power-law	0.08	-82	0.17	-3002	0.18	-3.1E+05	0.18	-3.1E+07
	Negative exponential	0.15	-85	0.25	-3123	0.27	-3.2E+05	0.26	-3.2E+07
	Gompertz	0.15	-86	0.25	-3126	0.27	-3.2E+05	0.27	-3.2E+07

- Appendix 2.1: Test calibration -

Appendix 2.1: Test calibration

The performance of the z_{dep} statistic was evaluated using simulated distance-decay data to build null and alternative hypothesis scenarios in which we assessed which model function (negative exponential, power-law or Gompertz) was the best-supported to describe the relationship between community similarity and spatial distance. Once the best-supported model was fitted to the data, we used the z_{dep} statistic to perform parameters comparison. The complete procedure is as follows:

1. Simulate two datasets (S1, D1) and (S2, D2), where S and D are the community similarity and spatial distance matrices, respectively, in a null or alternative hypothesis scenario.
2. Fit the best-supported model (Table 2.1) to estimate the distance-decay parameters of each curve: $\widehat{\theta}_{a1}$, $\widehat{\theta}_{b1}$ and $\widehat{\theta}_{a2}$, $\widehat{\theta}_{b2}$, respectively. The first subindex stands for the parameter (a = intercept or position parameter, and b = slope), while the second subindex stands for the two models that are being compared (model 1 or model 2).
3. Obtain the value of the z_{dep} statistic (Eq. 1) for the comparisons θ_{a1} vs. θ_{a2} , and θ_{b1} vs. θ_{b2} . To compute the variance in the parameters' estimations, required for obtaining the denominator of the test statistic, we used 300 resamples of a site-block resampling.
4. Repeat steps 1-3 1000 times, obtaining 1000 values of the test statistic, namely $z_{dep.k}$, with $k=1, \dots, 1000$.
5. Compute the corresponding 1000 p-values as $2 \cdot P(|z| > |z_{dep.k}|)$, being $z_{dep.k}$ the observed value of the statistic in each of the k samples computed in step 4, and z a $N(0,1)$ distribution.
6. Calculate the proportion of null hypotheses rejections in each of the k samples with a nominal level of $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$.

- Appendix 2.2: Supplementary tables Chapter II -

Appendix 2.2: Supplementary tables Chapter II

Table S2.1. Results for the calibration of the z_{dep} test under the null hypothesis. Proportion of null hypothesis (i.e. equality of parameters) rejections is shown for $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$, for two different sample sizes (number of sites $N = 25$ and $N = 50$) and for the different simulation scenarios described in the main text. Model parameters were the intercept and slope in the case of the negative exponential and the power-law functions, and the position parameter and slope in the case of the Gompertz function (marked with *).

N = 25						
	Intercept / position parameter			Slope		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Scenario #1	0.124	0.073	0.016	0.129	0.081	0.016
Scenario #2	0.104	0.056	0.014	0.111	0.057	0.011
Scenario #3*	0.091	0.05	0.016	0.105	0.05	0.012
N = 50						
	Intercept / position parameter			Slope		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Scenario #1	0.091	0.05	0.007	0.1	0.054	0.013
Scenario #2	0.099	0.043	0.009	0.104	0.054	0.014
Scenario #3*	0.097	0.046	0.008	0.1	0.045	0.008

Table S2.2. Results for the calibration of the z_{dep} test under the alternative hypothesis. The proportion of hypothesis rejections for the equality of parameters is shown for $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$, for two different sample sizes (number of sites $N = 25$ and $N = 50$) and for the different simulation scenarios described in the main text. Model parameters were the intercept and slope in the case of the negative exponential and the power-law functions, and the position parameter and slope in the case of the Gompertz function (marked with *).

N = 25						
	Intercept / position parameter			Slope		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Scenario #4*	0.457	0.296	0.079	1	1	0.998
Scenario #5	1	0.996	0.987	0.788	0.741	0.659
Scenario #6.1*	1	1	1	0.999	0.998	0.998
Scenario #6.2	0.448	0.378	0.238	1	1	1
N = 50						
	Intercept / position parameter			Slope		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Scenario #4*	0.778	0.697	0.480	1	1	1
Scenario #5	1	1	0.999	0.864	0.847	0.8
Scenario #6.1*	1	1	1	1	1	1
Scenario #6.2	0.596	0.546	0.431	1	1	1

Table S2.3. Parameters estimation for each Iberian vertebrate order.

Order	Intercept	Slope
Cypriniformes	1.18	1.16e-3
Anura	1.03	5.77e-4
Squamata	1.01	5.53e-4
Chiroptera	1.00	2.01e-4
Rodentia	1.00	2.24e-4
Carnivora	1.01	1.16e-4

- Appendix 2.3: Code for simulation study H_0 -

Appendix 2.3: Code for simulation study H_0

```
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #
# SIMULATION STUDY OF THE Zdep STATISTIC FOR REGRESSION PARAMETERS
##COMPARISON UNDER PAIRWISE DEPENDENCE
# NULL HYPOTHESIS
# - . # - . # - . # - . # - . # - . # - . # - . # - . #

# Supplementary R script for manuscript: Martin-Devasa R, Martinez-Santalla
S, Gomez-Rodriguez R, Crujeiras RM, Baselga A: Comparing distance-decay
parameters: a novel test under pairwise dependence.

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# SESSION PREPARATION
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

#Load libraries

library(sp)
library(geoR)
library(rgeos)
library(betapart)
library(minpack.lm)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# SIMULATION STUDY: GOAL
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Simulation study to assess the behaviour of the Zdep statistic under the
null hypothesis (equality of parameters between two distance-decay models).
Simulation steps are as follows:
# 1. We simulated one virtual regional pool and a single set of virtual
local communities to obtain one biological similarity matrix and one
spatial distance matrix.
# 2. From these initial matrices, we obtained two different datasets
(similarity + spatial distance) using site-block sampling in order to keep
the pairwise dependence structure of the data.
# These datasets follow the same mathematical model and therefore have
the same parameters (i.e. conform with the null hypothesis).
# 3. We fitted a distance-decay model to each dataset in order to compute
the estimated parameters (i.e. intercept and slope).
# 4. We repeated (n = 300) steps 2 and 3 in order to compute the variance
of parameter estimation using site-block resampling.
# 5. We computed the Zdep statistic using the variance estimates from step
4.

# 6. To calibrate the Zdep test, we repeated (n = 1000) steps 1 to 5 in
order to obtain 1000 values of the statistic and compute the proportion of
rejections of the null hypothesis.

# . . . . .
## SIMULATION CONDITIONS

## Number of resamples taken to compute the variance of parameter
estimation
B = 300
```

```

# Empty vectors, to store resampled regression parameters (intercept and
slope)
  int_v1 <- rep(0, B) # Intercept values of distance-decay of dataset#1
(resampled)
  slp_v1 <- rep(0, B) # Slope values of distance-decay of dataset#1
(resampled)
  int_v2 <- rep(0, B) # Intercept values of distance-decay of dataset#2
(resampled)
  slp_v2 <- rep(0, B) # Slope values of distance-decay of dataset#2
(resampled)

## Number of simulations conducted to assess type I error (Zdep test
calibration)
  M = 1000

# Empty vectors, to store Zdep statistic values in test calibration
  test_int <- rep(0, M) # Zdep values for the comparison of
intercept/position parameters
  test_slp <- rep(0, M) # Zdep values for the comparison of slopes

# Empty vectors, to store Zdep p-values in test calibration
  pval_int <- rep(0, M) # p-values of the Zdep comparison of
intercept/position parameters
  pval_slp <- rep(0, M) # p-values of the Zdep comparison of slopes

  in_dif2<-rep(0, M)
  sl_dif2<-rep(0, M)
## .....
## GENERATION OF SPECIES RANGES IN A VIRTUAL SCENARIO

## Virtual scenario properties:
# We created a 3000*3000 units virtual landscape where we placed 300
circular polygons representing virtual species ranges.
  landscape = 3000 # Virtual landscape dimensions
  species = 300 # Number of species (species ranges are represented as
circular polygons)

# Number of virtual sampling sites
  nsites = 25 # Analyses in Martín-Devasa et al. (2022) are done for two
different sample sizes: nsites = 25 and nsites = 50

## Generation of species ranges:
# We simulated species with different range sizes and also different
proportions of range sizes (widespread vs. spatially-restricted species),
which is the key to create the different simulation scenarios.
# This simulation study is exemplified for scenario #1 in Martín-Devasa et
al. 2022
  nu.w = 20 # Number of widespread species
  min.w = 4000; max.w = 6000 # Minimum and maximum diameters of polygons
of widespread species
  min.s = 300; max.s = 900 # Minimum and maximum diameters of polygons
of spatially-restricted species
  # The number of spatially-restricted species is computed as species -
nu.w. Thus, the value of this parameter is not defined in the script.
# For other scenarios, change the number of widespread species, the number
of spatially-restricted species and/or the polygon diameters as follows:

```

[illegible]

```

rad_m <- sample(seq(min.s/2,max.s/2), size = length(prt_mx), replace =
T) # Random sampling of radius values from the uniform
distribution of polygon diameters for spatially-restricted species (radius
= diameter/2)

# Merging of the vectors with polygon parameters
x_axis <- c(prt_gx, prt_mx); y_axis <- c(prt_gy, prt_my) # Central
coordinates of all species polygons
radius <- c(rad_g, rad_m) # Radius of all
species polygons

### (ii) Transformation of central coordinates and diameters into
spatial polygons
# Vector with 1000 sequential values between 0 and 2*pi. These values
are needed in the formula to build circles
t <- seq(0, 2*pi, length.out = 1000)

lis=list() # list to store polygons
for (p in 1:species){ # Loop to create circular polygons from central
coordinates and radius information

  x <- x_axis[p] + cos(t)*radius[p] # Definition of circles (x axis)
  y <- y_axis[p] + sin(t)*radius[p] # Definition of circles (y axis)

  dep <- SpatialPoints(cbind(x, y)) # Convert circle points to spatial
points
  pol <- Polygon(dep) # Convert spatial points to polygon
  pols <- Polygons(list(pol), p) # Storage of new circular polygon
  lis[[p]] <- pols # Storage polygons as list (need
for the SpatialPolygons function)

}

# Conversion of polygon list into a spatial polygons object
SPlgs <- SpatialPolygons(lis, proj4string = CRS("+proj=utm
+datum=WGS84")) # Add projection

### A.2. PRESENCE/ABSENCE OF SPECIES IN VIRTUAL SAMPLING SITES

## VIRTUAL SAMPLING SITES: Generation of virtual sampling sites within
the landscape

pon <- cbind(long=runif(nsites, 0, landscape),lat=runif(nsites, 0,
landscape)) # Coordinates of virtual sampling sites
poin <- SpatialPoints(pon, proj4string = CRS("+proj=utm +datum=WGS84"))
# Transform coordinates into spatial points

## PRESENCE/ABSENCE DATA
# Assessment of the species present in each sampling site
Results <- gIntersects(SPlgs, poin, byid = TRUE)

# Conversion of the logical (true/false) matrix into a 1 (present) and 0
(absence) matrix
Resul <- Results*1

```

```

## SIMILARITY MATRIX. Computed with the Simpson's similarity index
ydis <- 1-beta.pair(Resul)$beta.sim

## SPATIAL DISTANCE MATRIX. Computed as the euclidean distance between
sampling sites
xdis <- dist(poin@coords, method = "euclidean")

## . . . . .
. . . . .
# B. SITE-BLOCK SAMPLING TO GENERATE TWO DATASETS FOLLOWING THE SAME
MATHEMATICAL MODEL

# To generate two datasets following the same mathematical model, we
site-block sampled the matrices simulated in the previous step.
# The site-block sampling approach considers as sampling unit all the
similarity/distance values derived from the same site. This ensures that
samples obtained from these similarity/distance matrices preserve the
pairwise dependence of the data.

### B.1. GENERATION OF SITE-BLOCKS (i.e. site-block = all pairwise
similarities/distances involving one particular site)
xdis.matrix <- as.matrix(xdis)
# Convert triangular matrix into square matrix
ydis.matrix <- as.matrix(ydis)
# Convert triangular matrix into square matrix

# Generate site-blocks (1st dataset)
blocks.xdis1 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites)
# Empty matrix, to store site-block spatial distances
blocks.ydis1 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites)
# Empty matrix, to store site-block community similarities

for (i in 1:nsites){

  blocks.xdis1[,i] <- xdis.matrix[i,] [-i] # Store block corresponding
to i site while excluding the diagonal value
  blocks.ydis1[,i] <- ydis.matrix[i,] [-i] # Store block corresponding
to i site while excluding the diagonal value
}

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~
. ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~

# Generate site-blocks (2nd dataset)

blocks.xdis2 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites)
# Empty matrix, to store site-block spatial distances
blocks.ydis2 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites)
# Empty matrix, to store site-block community similarities

for (i in 1:nsites){

  blocks.xdis2[,i] <- xdis.matrix[i,] [-i] # Store block corresponding
to i site while excluding the diagonal value
  blocks.ydis2[,i] <- ydis.matrix[i,] [-i] # Store block corresponding
to i site while excluding the diagonal value
}

```

```

### B.2. SITE-BLOCK SAMPLES OF THE SIMILARITY AND SPATIAL DISTANCE
MATRICES
# Site-block sample: The matrices with site-blocks are sampled ensuring
that the same cases are selected in both matrices
b.i.sample1 <- sample(x = 1:nsites, size=nsites, replace=T)
# Index vector for sampling with replacement the first sample
b.i.sample2 <- sample(x = 1:nsites, size=nsites, replace=T)
# Index vector for sampling with replacement the second sample

b.xdis.sample1 <- blocks.xdis1[, b.i.sample1]
# Sampling (with replacement) of blocks of spatial distance values (First
sample)
b.ydis.sample1 <- blocks.ydis1[, b.i.sample1]
# Sampling (with replacement) of blocks of community similarity values
(First sample)

b.xdis.sample2 <- blocks.xdis2[, b.i.sample2]
# Sampling (with replacement) of blocks of spatial distance values (Second
sample)
b.ydis.sample2 <- blocks.ydis2[, b.i.sample2]
# Sampling (with replacement) of blocks of community similarity values
(Second sample)

# Downsizing of sampled site-blocks: To preserve the size of the
observed distance/similarity matrices, only the number of cases in the
original distance/similarity matrix is randomly selected
boot.index1 <- sample(1:length(b.ydis.sample1), size = nsites*(nsites -
1)/2) # Index vector for downsizing of sampled site-blocks (First
sample)
boot.b.ydis1 <- b.ydis.sample1[boot.index1]
# Downsize of site-blocks of community similarity (First sample)
boot.b.xdis1 <- b.xdis.sample1[boot.index1]

boot.index2 <- sample(1:length(b.ydis.sample2), size = nsites*(nsites -
1)/2) # Index vector for downsizing of sampled site-blocks (Second
sample)
boot.b.ydis2 <- b.ydis.sample2[boot.index2]
# Downsize of site-blocks of community similarity (Second sample)
boot.b.xdis2 <- b.xdis.sample2[boot.index2]

## . . . . .
. . . . .
# C. DISTANCE-DECAY MODELS

# A distance-decay function is fitted independently for the first and
the second sample.
# Parameters of these distance-decay models will be later compared with
the Zdep test based on the variance of these parameter.

### C.1 COMPUTE DISTANCE-DECAY PARAMETERS (intercept/position parameter
AND slope)

# First sample (boot.b.xdis1, boot.b.ydis1)

```

```

if (model == "power-law"){

```

```

    mod1 <- nlsLM(boot.b.ydis1 ~ a*(boot.b.xdis1^-d), start = list(a = 1,
d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (power-law model)
    int_1 <- coefficients(mod1)[1] # intercept value
    slp_1 <- coefficients(mod1)[2] # slope value

    } else if (model == "exponential"){
    mod1 <- nlsLM(boot.b.ydis1 ~ a*exp(boot.b.xdis1*-d), start = list(a =
1, d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (negative exponential model)
    int_1 <- coefficients(mod1)[1] # intercept value
    slp_1 <- coefficients(mod1)[2] # slope value

    } else if (model == "gompertz"){
    mod1 <- nlsLM(boot.b.ydis1 ~ 1-(exp(-a*exp(boot.b.xdis1*-d))), start =
list(a = 1, d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (Gompertz model)
    int_1 <- coefficients(mod1)[1] # position parameter value
    slp_1 <- coefficients(mod1)[2] # slope value
    }

    # Second sample (boot.b.xdis2, boot.bydis2)

    if (model == "power-law"){
    mod2 <- nlsLM(boot.b.ydis2 ~ a*(boot.b.xdis2^-d), start = list(a = 1,
d = 0), trace=F) # Non-linear regression of biological
similarity with spatial distance (power-law model)
    int_2 <- coefficients(mod2)[1] # intercept value
    slp_2 <- coefficients(mod2)[2] # slope value

    } else if (model == "exponential"){
    mod2 <- nlsLM(boot.b.ydis2 ~ a*exp(boot.b.xdis2*-d), start = list(a =
1, d = 0), trace=F) # Non-linear regression of biological
similarity with spatial distance (negative exponential model)
    int_2 <- coefficients(mod2)[1] # intercept value
    slp_2 <- coefficients(mod2)[2] # slope value

    } else if (model == "gompertz"){
    mod2 <- nlsLM(boot.b.ydis2 ~ 1-(exp(-a*exp(boot.b.xdis2*-d))), start =
list(a = 1, d = 0), trace=F) # Non-linear regression of biological
similarity with spatial distance (Gompertz model)
    int_2 <- coefficients(mod2)[1] # position parameter value
    slp_2 <- coefficients(mod2)[2] # slope value
    }

    ## . . . . .
    . . . . .
    # D. Zdep STATISTIC FOR PARAMETER COMPARISON

    ### D.1.SITE-BLOCK RESAMPLING
    # Parameter variance of distance-decay models is estimated with a block-
site resample approach (number of resamples = 300)
    # The site-block resample approach defines the block (set of values
derived of the same observation) as the sampling unit to maintain the
dependence structure of the similarity and spatial distance data

    for (j in 1:B) {

```

```

# Block-resample with replacement: The matrices with site-blocks are
resampled ensuring that the same cases are selected in both matrices
b.i.resample1 <- sample(x = 1:nsites, size=nsites, replace=T)
# Index vector for resampling with replacement (First sample)
b.i.resample2 <- sample(x = 1:nsites, size=nsites, replace=T)
# Index vector for resampling with replacement (Second sample)

b.xdis.resample1 <- b.xdis.sample1[, b.i.resample1]
# Resampling (with replacement) of blocks of spatial distance values (First
sample)
b.ydis.resample1 <- b.ydis.sample1[, b.i.resample1]
# Resampling (with replacement) of blocks of community similarity values
(First sample)

b.xdis.resample2 <- b.xdis.sample2[, b.i.resample2]
# Resampling (with replacement) of blocks of spatial distance values
(Second sample)
b.ydis.resample2 <- b.ydis.sample2[, b.i.resample2]
# Resampling (with replacement) of blocks of community similarity values
(Second sample)

# Downsizing of resampled site-blocks: To preserve the size of the
observed distance/similarity matrices, only the number of cases in the
original distance/similarity matrix is randomly selected
boot.index1.1 <- sample(1:length(b.ydis.resample1), size =
nsites*(nsites - 1)/2) # Index vector for downsizing of resampled site-
blocks (First sample)
boot.b.ysub1 <- b.ydis.resample1[boot.index1.1]
# Downsize of site-blocks of community similarity (First sample)
boot.b.xsub1 <- b.xdis.resample1[boot.index1.1]

boot.index2.2 <- sample(1:length(b.ydis.resample2), size =
nsites*(nsites - 1)/2) # Index vector for downsizing of resampled site-
blocks (Second sample)
boot.b.ysub2 <- b.ydis.resample2[boot.index2.2]
# Downsize of site-blocks of community similarity (Second sample)
boot.b.xsub2 <- b.xdis.resample2[boot.index2.2]

### D.2 DISTANCE-DECAY MODEL AND PARAMETER ESTIMATION FOR EACH RESAMPLE

# First sample

if (model == "power-law"){
  mod1.1 <- nlsLM(boot.b.ysub1 ~ a*(boot.b.xsub1^-d), start = list(a =
1, d = 0), trace = F) # Non-linear regression of
biological similarity with spatial distance (power-law model)
  int_v1[j] <- coefficients(mod1.1)[1] # resampled intercept value
  slp_v1[j] <- coefficients(mod1.1)[2] # resampled slope value

} else if (model == "exponential"){
  mod1.1 <- nlsLM(boot.b.ysub1 ~ a*exp(boot.b.xsub1*-d), start = list(a
= 1, d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (negative exponential model)
  int_v1[j] <- coefficients(mod1.1)[1] # resampled intercept value
  slp_v1[j] <- coefficients(mod1.1)[2] # resampled slope value

} else if (model == "gompertz"){

```

```

mod1.1 <- nlsLM(boot.b.ysub1 ~ 1-(exp(-a*exp(boot.b.xsub1*-d))),
start = list(a = 1, d = 0), trace = F) # Non-linear regression of
biological similarity with spatial distance (Gompertz model)
int_v1[j] <- coefficients(mod1.1)[1] # resampled position parameter
value
slp_v1[j] <- coefficients(mod1.1)[2] # resampled slope value
}

# Second sample

if (model == "power-law"){
mod2.1 <- nlsLM(boot.b.ysub2 ~ a*(boot.b.xsub2^-d), start = list(a =
1, d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (power-law model)
int_v2[j] <- coefficients(mod2.1)[1] # resampled intercept value
slp_v2[j] <- coefficients(mod2.1)[2] # resampled slope value

} else if (model == "exponential"){
mod2.1 <- nlsLM(boot.b.ysub2 ~ a*exp(boot.b.xsub2*-d), start = list(a
= 1,d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (negative exponential model)
int_v2[j] <- coefficients(mod2.1)[1] # resampled intercept value
slp_v2[j] <- coefficients(mod2.1)[2] # resampled slope value

} else if (model == "gompertz"){
mod2.1 <- nlsLM(boot.b.ysub2 ~ 1-(exp(-a*exp(boot.b.xsub2*-d))),
start = list(a = 1, d = 0), trace = F) # Non-linear regression of
biological similarity with spatial distance (Gompertz model)
int_v2[j] <- coefficients(mod2.1)[1] # resampled position parameter
value
slp_v2[j] <- coefficients(mod2.1)[2] # resampled slope value
}
}

### D.3. T_dep STATISTIC COMPUTATION
# Variance of distance-decay parameters, estimated with a block-site
resampling approach (n =300), is used for the Zdep test computation

test_int[k] <- (int_1 - int_2) / sqrt((var(int_v1) + var(int_v2)) - 2 *
cov(int_v1, int_v2))
test_slp[k] <- (slp_1 - slp_2) / sqrt((var(slp_v1) + var(slp_v2)) - 2 *
cov(slp_v1, slp_v2))

in_dif2[k]<-int_1-int_2
sl_dif2[k]<-slp_1-slp_2

message(k)
}

## . . . . .
. . . . .
# E. Zdep CALIBRATION

### P-VALUE OF ZDEP
# If the Zdep value is <= 0 the p-value is computed as 2* the
probability of obtain lower values in a N(0,1)

```

```

# If the Zdep value is > 0 the p-value is computed as 2* the probability
of obtain greater values in a N(0,1)
# In general, the p-value is computed as 2*(P|N| > |Zdep|)

# For the intercept/position parameter

for (q in 1:length(test_int)) {
  if (test_int[q] <= 0 ) {
    pval_int[q] <- 2*pnorm(test_int[q], mean=0, sd=1, lower.tail = T)
  } else if (test_int[q] > 0){
    pval_int[q] <- 2*pnorm(test_int[q], mean=0, sd=1, lower.tail = F)
  }
}

# For the slope

for (q in 1:length(test_slp)) {
  if (test_slp[q] <= 0) {
    pval_slp[q] <- 2*pnorm(test_slp[q], mean=0, sd=1, lower.tail = T)
  } else if (test_slp[q]>0){
    pval_slp[q] <- 2*pnorm(test_slp[q], mean=0, sd=1, lower.tail = F)
  }
}

# Save the p-value vectors to plot the results.
# For nsites = 25
# pval_int25 <- pval_int
# pval_slp25 <- pval_slp

# For nsites = 50
# pval_int50 <- pval_int
# pval_slp50 <- pval_slp

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# APPROXIMATION TO THE LEVEL (alpha) OF THE Zdep STATISTIC
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# The proportion of rejections with different values of alpha (0.1, 0.05
and 0.01) is assess for data produced in the simulation study

# Intercept/position parameter
mean(pval_int < 0.1); mean(pval_int < 0.05); mean(pval_int < 0.01)      #
Assess the proportion of rejections

# Slope
mean(pval_slp < 0.1); mean(pval_slp < 0.05); mean(pval_slp < 0.01)      #
Assess the proportion of rejections

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# VIOLINPLOTS
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Note that for the boxplot the results of the simulations with nsites = 25
and nsites = 50 are needed
loc<-as.factor(c(rep("25",2000),rep("50",2000)))

```

```

par<-
as.factor(c(rep("Intercept",1000),rep("Slope",1000),rep("Intercept",1000),r
ep("Slope",1000)))
pval<-c(pval_int25,pval_slp25,pval_int50,pval_slp50)

sc<-data.frame(pval,par,loc)

ggplot(sc, aes(x=par, y=val, fill=loc)) +
  geom_violin(scale = "width") + ylim(0,1) +
  scale_fill_manual(values=c("#4477aa", "#ee6677")) +
  theme_test() +
  geom_point(alpha = .4, position = position_jitter(seed = 1),cex=0.4) +
  geom_hline(yintercept=0.05, linetype=c("dashed"), color =c("red"),
size=1)

```


- Appendix 2.4: Code for simulation study H₁ -

Appendix 2.4: Code for simulation study H₁

```
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #
# SIMULATION STUDY OF THE Zdep STATISTIC FOR REGRESSION PARAMETERS
# COMPARISON UNDER PAIRWISE DEPENDENCE
# ALTERNATIVE HYPOTHESIS
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #

# Supplementary R script for manuscript: Martin-Devasa R, Martinez-Santalla
# S, Gomez-Rodriguez R, Crujeiras RM, Baselga A: Comparing distance-decay
# parameters: a novel test under pairwise dependence.

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# SESSION PREPARATION
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

#Load libraries

library(sp)
library(rgeos)
library(betapart)
library(minpack.lm)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# SIMULATION STUDY: GOAL
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Simulation study to assess the behaviour of the Zdep statistic under the
# alternative hypothesis (difference of parameters between two distance-decay
# models). Simulation steps are as follows:
# 1. We simulated two different virtual regional pools and two sets of
# virtual local communities. For each of them, we obtained a biological
# similarity matrix and a spatial distance matrix.
# 2. From these initial matrices, we defined two different datasets (DS1
# and DS2), each of them containing one similarity matrix and the associated
# spatial distance matrix.
# These datasets are independently generated and thus, follow different
# mathematical models.
# All analyses will be independently conducted on DS1 and DS2
# 3. We fitted a distance-decay model to each dataset in order to compute
# the estimated parameters (i.e. intercept and slope).
# 4. We repeated (n = 300) steps 2 and 3 in order to compute the variance
# of parameter estimation using site-block resampling.
# 5. We computed the Zdep statistic using the variance estimates from step
# 4.

# 6. To calibrate the Zdep test, we repeated (n = 1000) steps 1 to 5 in
# order to obtain 1000 values of the statistic and compute the proportion of
# rejections of the hypothesis of equality of parameters.
# Under the alternative hypothesis scenario, we expect the hypothesis of
# "equality of parameters" to be rejected most of the time given that
# datasets have been independently generated and, thus, follow a different
# mathematical model.
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
```

```

## SIMULATION CONDITIONS

## Number of resamples taken to compute the variance of parameter
estimation
  B = 300

# Empty vectors, to store resampled regression parameters (intercept and
slope)
int_v1 <- rep(0, B) # Intercept values of dataset#1 (DS1) (resampled)
slp_v1 <- rep(0, B) # Slope values of dataset#1 (DS1) (resampled)
int_v2 <- rep(0, B) # Intercept values of dataset#2 (DS2) (resampled)
slp_v2 <- rep(0, B) # Slope values of dataset#2 (DS2) (resampled)

## Number of simulations conducted to assess type II error (Zdep test
calibration)
  M = 1000

# Empty vectors, to store Zdep statistic values
test_int <- rep(0, M) # Zdep values for the comparison of
intercept/position parameters
test_slp <- rep(0, M) # Zdep values for the comparison of slopes

# Empty vectors, to store Zdep p-values
pval_int <- rep(0, M) # p-values for the Zdep comparison of
intercepts/position parameters
pval_slp <- rep(0, M) # p-values for the Zdep comparison of slopes

## . . . . .
## GENERATION OF SPECIES RANGES IN A VIRTUAL SCENARIO

## Virtual scenario properties:
# We created a 3000*3000 units virtual landscape where we placed 300
circular polygons representing virtual species ranges.
  landscape = 3000 # Virtual landscape dimensions
  species = 300 # Number of species (species ranges are represented as
circular polygons)

# Number of virtual sampling sites
  nsites = 50 # Analyses in Martín-Devasa et al. (2022) are done for two
different sample sizes: nsites = 25 and nsites = 50

## Generation of species ranges:
# We simulated species with different range sizes and also different
proportions of range sizes (widespread vs. spatially-restricted species),
which is the key to create the different simulation scenarios.
# This simulation study is exemplified for Scenario #6 in Martín-Devasa et
al. 2022
  # For DS1
  nu.w.1 = 200 # Number of widespread species
  min.w1 = 4000; max.w1 = 6000 # Minimum and maximum polygon diameters of
widespread species
  min.s1 = 3000; max.s1 = 4000 # Minimum and maximum polygon diameters of
spatially-restricted species
  # The number of spatially-restricted species is computed as species -
nu.w. Thus, the value of this parameter is not defined in the script.
  # For DS2
  nu.w.2 = 20 # Number of widespread species

```

```

    min.w2 = 3000; max.w2 = 4000 # Minimum and maximum polygon diameters of
widespread species
    min.s2 = 600; max.s2 = 1600 # Minimum and maximum polygon diameters of
spatially-restricted species
    # The number of spatially-restricted species is computed as species -
nu.w. Thus, the value of this parameter is not defined in the script.

# For other scenarios, change the number of widespread species, the number
of spatially-restricted species and/or the range in the polygon diameter as
follows:
## Scenario #4
# For DS1
# nu.w.1 = 300
# min.w1 = 3000; max.w1 = 4000

#For DS2
# nu.w.2 = 200
# min.w2 = 4000; max.w2 = 6000
# min.s2 = 3000; max.s2 = 4000

## Scenario #5
# For DS1
# nu.w.1 = 20
# min.w1 = 4000; max.w1 = 6000
# min.s1 = 1500; max.s1 <- 2000

# For DS2
# nu.w.2 = 20
# min.w2 = 3000; max.w2 = 4000
# min.s2 = 200; max.s2 = 1200

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# STUDY OF THE Zdep UNDER ALTERNATIVE HYPOTHESIS FOR TYPE II ERROR
ASSESSMENT
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Seed
set.seed(1234)

## DISTANCE-DECAY FUNCTION: Select a function to fit the distance-decay
curves. Choose between "power-law", "exponential" or "gompertz"
model="gompertz"
# Note that in the power-law and exponential models the parameters are
the intercept and the slope. In the Gompertz model, the parameters are
position parameter and slope.

## LOOP
# Loop to compute and store Zdep values under the null hypothesis. These n
= 1000 values are used for test calibration.

for (k in 1:M) {

## . . . . .
# A. SIMULATION OF DATA FOR DISTANCE-DECAY ANALYSIS

### A.1. VIRTUAL SPECIES RANGES: (i.i) Central coordiantes and diameters
## FOR DS1

```

```

# Widespread species: Central coordinates (prt_gx, prt_gy) and radius
(rad_g) of the polygons representing species ranges
prt_gx <- sample(seq(0, landscape, by = 0.01), size = nu.w.1, replace =
T) # Random sampling of spatial coordinate along the x-axis of the
simulation landscape (size = 3000 x 3000)
prt_gy <- sample(seq(0, landscape, by = 0.01), size = nu.w.1, replace =
T) # Random sampling of spatial coordinate along the y-axis of the
simulation landscape (size = 3000 x 3000)

rad_g <- sample(seq(min.w1/2, max.w1/2), size = length(prt_gx), replace
= T) # Random sampling of radius values from the uniform distribution of
polygon diameters for widespread species (radius = diameter/2)

# Spatially-restricted species: Central coordinates (prt_mx, prt_my) and
radius (rad_m) of the polygons representing species ranges
prt_mx <- sample(seq(0, landscape, by = 0.01), size = species - nu.w.1,
replace = T) # Random sampling of spatial coordinate along the x-axis of
the simulation landscape (size = 3000 x 3000)
prt_my <- sample(seq(0, landscape, by = 0.01), size = species - nu.w.1,
replace = T) # Random sampling of spatial coordinate along the y-axis of
the simulation landscape (size = 3000 x 3000)

rad_m <- sample(seq(min.s1/2, max.s1/2), size = length(prt_mx), replace
= T) # Random sampling of radius values from the uniform
distribution of polygon diameters for spatially restricted species (radius
= diameter/2)

# Merging of the vectors with polygon parameters
ejex <- c(prt_gx, prt_mx); ejey <- c(prt_gy, prt_my) # Central
coordinates of all species polygons
radio <- c(rad_g, rad_m) # Radius of all
species polygons

### (ii.i) Transformation of central coordinates and diameters into
spatial polygons
# Vector with 1000 values between 0 and 2*pi. These values are needed in
the formula to build circles
t <- seq(0, 2*pi, length.out = 1000)

lis=list() # list to store polygons
for (p in 1:species){ # Loop to create circular polygons from central
coordinates and radius information

  x <- ejex[p] + cos(t) * radio[p] # Definition of circles (x axis)
  y <- ejey[p] + sin(t) * radio[p] # Definition of circles (y axis)

  dep <- SpatialPoints(cbind(x, y)) # Convert circle points to spatial
points
  pol <- Polygon(dep) # Convert spatial points to polygon
  pols <- Polygons(list(pol), p) # Storage of new circular polygon
  lis[[p]] <- pols # Storage polygons as list (need for
the SpatialPolygons function)

}

# Conversion of polygon list into a spatial polygons object
SPlgs <- SpatialPolygons(lis, proj4string = CRS("+proj=utm
+datum=WGS84")) # Add projection

```

```

### THE SAME PROCESS IS REPEATED FOR DS2

### A.1 VIRTUAL SPECIES RANGES: (i.ii) Central coordiantes and diameters

# Widespread species: Central coordinates (prt_gx, prt_gy) and radius
(rad_g) of the polygons representing species ranges
prt_gx2 <- sample(seq(0, landscape, by = 0.01), size = nu.w.2, replace =
T) # Random sampling of spatial coordinate along the x-axis of the
simulation landscape (size = 3000 x 3000)
prt_gy2 <- sample(seq(0, landscape, by = 0.01), size = nu.w.2, replace =
T) # Random sampling of spatial coordinate along the y-axis of the
simulation landscape (size = 3000 x 3000)

rad_g2 <- sample(seq(min.w2/2, max.w2/2), size=length(prt_gx2), replace
= T) # Random sampling of radius values from the uniform distribution of
polygon diameters for widespread species (radius = diameter/2)

# Spatially-restricted species: Central coordinates (prt_mx, prt_my) and
radius (rad_m) of the polygons representing species ranges
prt_mx2 <- sample(seq(0, landscape, by = 0.01), size = species - nu.w.2,
replace = T) # Random sampling of spatial coordinate along the x-axis of
the simulation landscape (size = 3000 x 3000)
prt_my2 <- sample(seq(0, landscape, by = 0.01), size = species - nu.w.2,
replace = T) # Random sampling of spatial coordinate along the y-axis of
the simulation landscape (size = 3000 x 3000)

rad_m2 <- sample(seq(min.s2/2, max.s2/2), size = length(prt_mx2),
replace = T) # Random sampling of radius values from the uniform
distribution of polygon diameters for spatially-restricted species (radius
= diameter/2)

# Merging of the vectors with polygon parameters
ejex2 <- c(prt_gx2, prt_mx2); ejey2 <- c(prt_gy2, prt_my2) # Central
coordinates of all species polygons
radio2 <- c(rad_g2, rad_m2) # Radius of
all species polygons

### (ii.ii) Transformation of central coordiantes and diameters into
spatial polygons

lis2=list() # list to store polygons
for (z in 1:species){ # Loop to create circular polygons from central
coordinates and radius information

  x2 <- ejex2[z] + cos(t) * radio2[z] # Definition of circles (x axis)
  y2 <- ejey2[z] + sin(t) * radio2[z] # Definition of circles (y axis)

  dep2 <- SpatialPoints(cbind(x2, y2)) # Convert circle points to
spatial points
  pol2 <- Polygon(dep2) # Convert spatial points to
polygon
  pols2 <- Polygons(list(pol2), z) # Storage of new circular polygon
  lis2[[z]] <- pols2 # # Storage polygons as list
  (need for the SpatialPolygons function)
}

```

```

# Conversion of polygon list into a spatial polygons object
SPlgs2 <- SpatialPolygons(lis2, proj4string = CRS("+proj=utm
+datum=WGS84")) # Add projection

### A.2. PRESENCE/ABSENCE OF SPECIES IN VIRTUAL SAMPLING SITES

## VIRTUAL SAMPLING SITES: Generation of virtual sampling sites within
the landscape
# FOR DS1

pon <- cbind(lat=runif(nsites, 0, landscape), long=runif(nsites, 0,
landscape))
poin <- SpatialPoints(pon, proj4string = CRS("+proj=utm +datum=WGS84"))
# Transform coordinates into spatial points

## PRESENCE/ABSENCE DATA
# FOR DS1
# Assessment of the species present in each sampling site
Results1 <- gIntersects(SPlgs, poin, byid = TRUE)

# Conversion the logical (true/false) matrix obtained in a 1 (present)
and 0 (absence) matrix
Result1 <- Results1 * 1

# FOR DS2
# Assessment of the species present in each sampling site
Results2 <- gIntersects(SPlgs2, poin, byid = TRUE)

# Conversion the logical (true/false) matrix obtained in a 1 (present)
and 0 (absence) matrix
Result2 <- Results2 * 1

### A.3. SIMILARITY AND DISTANCE MATRICES FOR DISTANCE-DECAY MODELS
# FOR DS1

## SIMILARITY MATRIX. Computed with the Simpson's similarity index
ydis1 <- 1 - beta.pair(Result1)$beta.sim

## SPATIAL DISTANCE MATRIX. Computed as the euclidean distance between
sampling sites
xdis1 <- dist(poin@coords, method = "euclidean")

## DS1 = (xdis1, ydis1)

# FOR DS2

## SIMILARITY MATRIX. Computed with the Simpson's similarity index
ydis2 <- 1 - beta.pair(Result2)$beta.sim

## SPATIAL DISTANCE MATRIX. Computed as the euclidean distance between
sampling sites
xdis2 <- dist(poin@coords, method = "euclidean")

```

```

## DS2 = (xdis2, ydis2)

## . . . . .
. . . . .
# B. DISTANCE-DECAY MODELS

# A distance-decay function is fitted independently for DS1 and DS2.
# Parameters of these distance-decay models will be later compared with
the Zdep test based on the variance of these parameter.

### B.1 COMPUTE DISTANCE-DECAY PARAMETERS (intercept/position parameter
AND slope)

# FOR DS1 (xdis1, ydis1)

if (model == "power-law"){
  mod1 <- nlsLM(ydis1 ~ a*(xdis1^-d), start = list(a = 1, d = 0.1),
trace = F) # Non-linear regression of biological similarity
with spatial distance (power-law model)
  int_1 <- coefficients(mod1)[1] # intercept value
  slp_1 <- coefficients(mod1)[2] # slope value

} else if (model == "exponential"){
  mod1 <- nlsLM(ydis1 ~ a*exp(xdis1*-d), start = list(a = 1, d = 0),
trace = F) # Non-linear regression of biological similarity with
spatial distance (negative exponential model)
  int_1 <- coefficients(mod1)[1] # intercept value
  slp_1 <- coefficients(mod1)[2] # slope value

} else if (model == "gompertz"){
  mod1 <- nlsLM(ydis1 ~ 1-(exp(-a*exp(-d*xdis1))), start = list(a = 1, d
= 0), trace = F) # Non-linear regression of biological similarity with
spatial distance (Gompertz model)
  int_1 <- coefficients(mod1)[1] # position parameter value
  slp_1 <- coefficients(mod1)[2] # slope value
}

# FOR DS2 (xdis2, ydis2)

if (model == "power-law"){
  mod2 <- nlsLM(ydis2 ~ a*(xdis2^-d), start = list(a = 1, d = 0.1),
trace = F) # Non-linear regression of biological similarity
with spatial distance (power-law model)
  int_2 <- coefficients(mod2)[1] # intercept value
  slp_2 <- coefficients(mod2)[2] # slope value

} else if (model == "exponential"){
  mod2 <- nlsLM(ydis2 ~ a*exp(xdis2*-d), start = list(a = 1, d = 0),
trace = F) # Non-linear regression of biological similarity
with spatial distance (negative esponential model)
  int_2 <- coefficients(mod2)[1] # intercept value
  slp_2 <- coefficients(mod2)[2] # slope value

} else if (model == "gompertz"){
  mod2 <- nlsLM(ydis2 ~ 1 - (exp(-a*exp(xdis2*-d))), start = list(a =
1,d = 0), trace = F) # Non-linear regression of biological similarity with
spatial distance (Gompertz model)

```

```

    int_2 <- coefficients(mod2)[1] # position parameter value
    slp_2 <- coefficients(mod2)[2] # slope value
  }

  ## . . . . .
  . . . . .
  # C. Zdep STATISTIC FOR PARAMETER COMPARISON

  ### C.1.SITE-BLOCK RESAMPLING
  # Parameter variance of distance-decay models is estimated with a block-
  site resample approach (number of resamples = 300)
  # The site-block resample approach defines the block (set of values
  derived of the same observation) as the sampling unit to maintain the
  dependence structure of the similarity and spatial distance data

  ## GENERATION OF SITE-BLOCKS (i.e. site-block = all pairwise
  similarities/distances involving one particular site)

  # For DS1
  # Create site-blocks (i.e. site-block = all pairwise
  similarities/distances involving one particular site)
  xdis.matrix1 <- as.matrix(xdis1) #
  Convert triangular matrix into square matrix
  ydis.matrix1 <- as.matrix(ydis1) #
  Convert triangular matrix into square matrix

  blocks.xdis1 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites) #
  Empty matrix, to store site-block spatial distances
  blocks.ydis1 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites) #
  Empty matrix, to store site-block community similarities

  for (i in 1:nsites){

    blocks.xdis1[,i] <- xdis.matrix1[i,][-i] # Store block
    corresponding to i site while excluding the diagonal value
    blocks.ydis1[,i] <- ydis.matrix1[i,][-i] # Store block
    corresponding to i site while excluding the diagonal value
  }

  # For DS2
  # Create site-blocks (i.e. site-block = all pairwise
  similarities/distances involving one particular site)
  xdis.matrix2 <- as.matrix(xdis2) #
  Convert triangular matrices into square matrix
  ydis.matrix2 <- as.matrix(ydis2) #
  Convert triangular matrices into square matrix

  blocks.xdis2 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites) #
  Empty matrix, to store site-block spatial distances
  blocks.ydis2 <- matrix(rep(0, nsites*(nsites - 1)), ncol = nsites) #
  Empty matrix, to store site-block community similarities

```

```

    blocks.xdis2[,i] <- xdis.matrix2[i,][-i]    # Store block
corresponding to i site while excluding the diagonal value
    blocks.ydis2[,i] <- ydis.matrix2[i,][-i]    # Store block
corresponding to i site while excluding the diagonal value
  }

# RESAMPLING

for (i in 1:B){

  # Block-resample with replacement: The matrices with site-blocks are
resampled ensuring that the same cases are selected in both matrices
  b.i.resample1<-sample(x = 1:nsites, size=nsites, replace=T)
# Index vector for resampling with replacement (DS1)
  b.i.resample2<-sample(x = 1:nsites, size=nsites, replace=T)
# Index vector for resampling with replacement (DS2)

  b.xdis.resample1 <- blocks.xdis1[, b.i.resample1]
# Resampling (with replacement) of blocks of spatial distance values (DS1)
  b.ydis.resample1 <- blocks.ydis1[, b.i.resample1]
# Resampling (with replacement) of blocks of community similarity values
(DS1)

  b.xdis.resample2 <- blocks.xdis2[, b.i.resample2]
# Resampling (with replacement) of blocks of spatial distance values (DS2)
  b.ydis.resample2 <- blocks.ydis2[, b.i.resample2]
# Resampling (with replacement) of blocks of community similarity values
(DS2)

  # Downsizing of resampled site-blocks: To preserve the size of the
observed distance/similarity matrices, only the number of cases in the
original distance/similarity matrix is randomly selected
  boot.index1 <- sample(1:length(b.ydis.resample1), size = nsites*(nsites
- 1)/2)    # Index vector for downsizing of resampled site-blocks (DS1)
  boot.b.ydis1 <- b.ydis.resample1[boot.index1]
# Downsize of site-blocks of community similarity (DS1)
  boot.b.xdis1 <- b.xdis.resample1[boot.index1]

  boot.index2 <- sample(1:length(b.ydis.resample2), size = nsites*(nsites
- 1)/2)    # Index vector for downsizing of resampled site-blocks (DS2)
  boot.b.ydis2 <- b.ydis.resample2[boot.index2]
# Downsize of site-blocks of community similarity (DS2)
  boot.b.xdis2 <- b.xdis.resample2[boot.index2]

### C.2 DISTANCE-DECAY MODEL AND PARAMETER ESTIMATION FOR EACH RESAMPLE

# FOR DS1

if (model == "power-law"){
  mod1.1 <- nlsLM(boot.b.ydis1 ~ a*(boot.b.xdis1^-d), start=list(a = 1,
d = 0.1), trace=F)    # Non-linear regression of biological
similarity with spatial distance (power-law model)
  int_v1[i] <- coefficients(mod1.1)[1]    # resampled intercept value
  slp_v1[i] <- coefficients(mod1.1)[2]    # resampled slope value
} else if (model == "exponential"){

```

```

    mod1.1 <- nlsLM(boot.b.ydis1 ~ a*exp(boot.b.xdis1*-d), start = list(a
= 1,d = 0), trace=F) # Non-linear regression of biological
similarity with spatial distance (negative exponential model)
    int_v1[i] <- coefficients(mod1.1)[1] # resampled intercept value
    slp_v1[i] <- coefficients(mod1.1)[2] # resampled slope value

  } else if (model == "gompertz"){
    mod1.1 <- nlsLM(boot.b.ydis1 ~ 1 - (exp(-a*exp(boot.b.xdis1*-d))),
start = list(a = 1, d = 0), trace = F) # Non-linear regression of
biological similarity with spatial distance (Gompertz model)
    int_v1[i] <- coefficients(mod1.1)[1] # resampled position parameter
value
    slp_v1[i] <- coefficients(mod1.1)[2] # resampled slope value
  }

# FOR DS2

  if (model == "power-law"){
    mod2.1 <- nlsLM(boot.b.ydis2 ~ a*(boot.b.xdis2^-d), start = list(a =
1, d = 0.1), trace = F) # Non-linear regression of biological
similarity with spatial distance (power-law model)
    int_v2[i] <- coefficients(mod2.1)[1] # resampled intercept value
    slp_v2[i] <- coefficients(mod2.1)[2] # resampled slope value

  } else if (model == "exponential"){
    mod2.1 <- nlsLM(boot.b.ydis2 ~ a*exp(boot.b.xdis2*-d), start = list(a
= 1, d = 0), trace = F) # Non-linear regression of biological
similarity with spatial distance (negative exponential model)
    int_v2[i] <- coefficients(mod2.1)[1] # resampled intercept value
    slp_v2[i] <- coefficients(mod2.1)[2] # resampled slope value

  } else if (model == "gompertz"){
    mod2.1 <- nlsLM(boot.b.ydis2 ~ 1-(exp(-a*exp(boot.b.xdis2*-c))),
start = list(a = 1, c = 0), trace = F) # Non-linear regression of
biological similarity with spatial distance (Gompertz model)
    int_v2[i] <- coefficients(mod2.1)[1] # resampled position parameter
value
    slp_v2[i] <- coefficients(mod2.1)[2] # resampled slope value
  }

}

### C.3. T_dep STATISTIC COMPUTATION
# Variance of distance-decay parameters, estimated with a block-site
resampling approach (n =300), is used for the Zdep test computation

test_int[k] <- (int_1 - int_2)/ sqrt((var(int_v1) + var(int_v2)) - 2 *
cov(int_v1, int_v2))
test_slp[k] <- (slp_1 - slp_2)/ sqrt((var(slp_v1) + var(slp_v2)) - 2 *
cov(slp_v1, slp_v2))

message(k)
}

```

```

## . . . . .
. . . . .
# D. Zdep CALIBRATION

### P-VALUE OF ZDEP
# If the Zdep value is <= 0 the p-value is computed as 2* the
probability of obtain lower values in a N(0,1)
# If the Zdep value is > 0 the p-value is computed as 2* the probability
of obtain greater values in a N(0,1)
# In general, the p-value is computed as 2*(P|N| > |Zdep|)

# For the intercept/position parameter

for (q in 1:length(test_int)) {
  if (test_int[q]<=0) {
    pval_int[q]<-2*pnorm(test_int[q], mean=0, sd=1, lower.tail =T)
  } else if (test_int[q]>0){
    pval_int[q]<-2*pnorm(test_int[q], mean=0, sd=1, lower.tail =F)
  }
}

# For the slope

for (q in 1:length(test_slp)) {
  if (test_slp[q]<=0) {
    pval_slp[q]<-2*pnorm(test_slp[q], mean=0, sd=1, lower.tail =T)
  } else if (test_slp[q]>0){
    pval_slp[q]<-2*pnorm(test_slp[q], mean=0, sd=1, lower.tail =F)
  }
}

# Save the p-value vectors to plot the results.
# For nsites = 25
# pval_int25 <- pval_int
# pval_slp25 <- pval_slp

# For nsites = 50
# pval_int50 <- pval_int
# pval_slp50 <- pval_slp

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# POWER OF THE Zdep STATISTIC
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# The proportion of rejections with different values of alpha (0.1, 0.05
and 0.01) is assess for data produced in the simulation study

# Intercept/position parameter
mean(pval_int < 0.1); mean(pval_int < 0.05); mean(pval_int < 0.01) #
Assess the proportion of rejections

# Slope
mean(pval_slp < 0.1); mean(pval_slp < 0.05); mean(pval_slp < 0.01) #
Assess the proportion of rejections
~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

```

```

# VIOLINPLOTS
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Note that for the boxplot the results of the simulations with nsites = 25
and nsites = 50 are needed

loc<-as.factor(c(rep("25",2000),rep("50",2000)))
par<-
as.factor(c(rep("Intercept",1000),rep("Slope",1000),rep("Intercept",1000),r
ep("Slope",1000)))
pval<-c(pval_int25,pval_slp25,pval_int50,pval_slp50)

sc<-data.frame(pval,par,loc)

ggplot(sc, aes(x=par, y=val, fill=loc)) +
  geom_violin(scale = "width") + ylim(0,1) +
  scale_fill_manual(values=c("#4477aa", "#ee6677")) +
  theme_test() +
  geom_point(alpha = .4, position = position_jitter(seed = 1),cex=0.4) +
  geom_hline(yintercept=0.05, linetype=c("dashed"), color =c("red"),
size=1)

```

- Appendix 3.1: Supplementary figures and tables Chapter III -

Appendix 3.1: Supplementary figures and tables Chapter III

Figure S3.1. Relationship between the number of species and the area of the territory. a) $\log(\text{richness})$ - $\log(\text{area})$ curve for the studied European territories with the fitted model (red line). b) Cook distances for the studied European territories (cutoff = 0.15) outliers are marked in red.

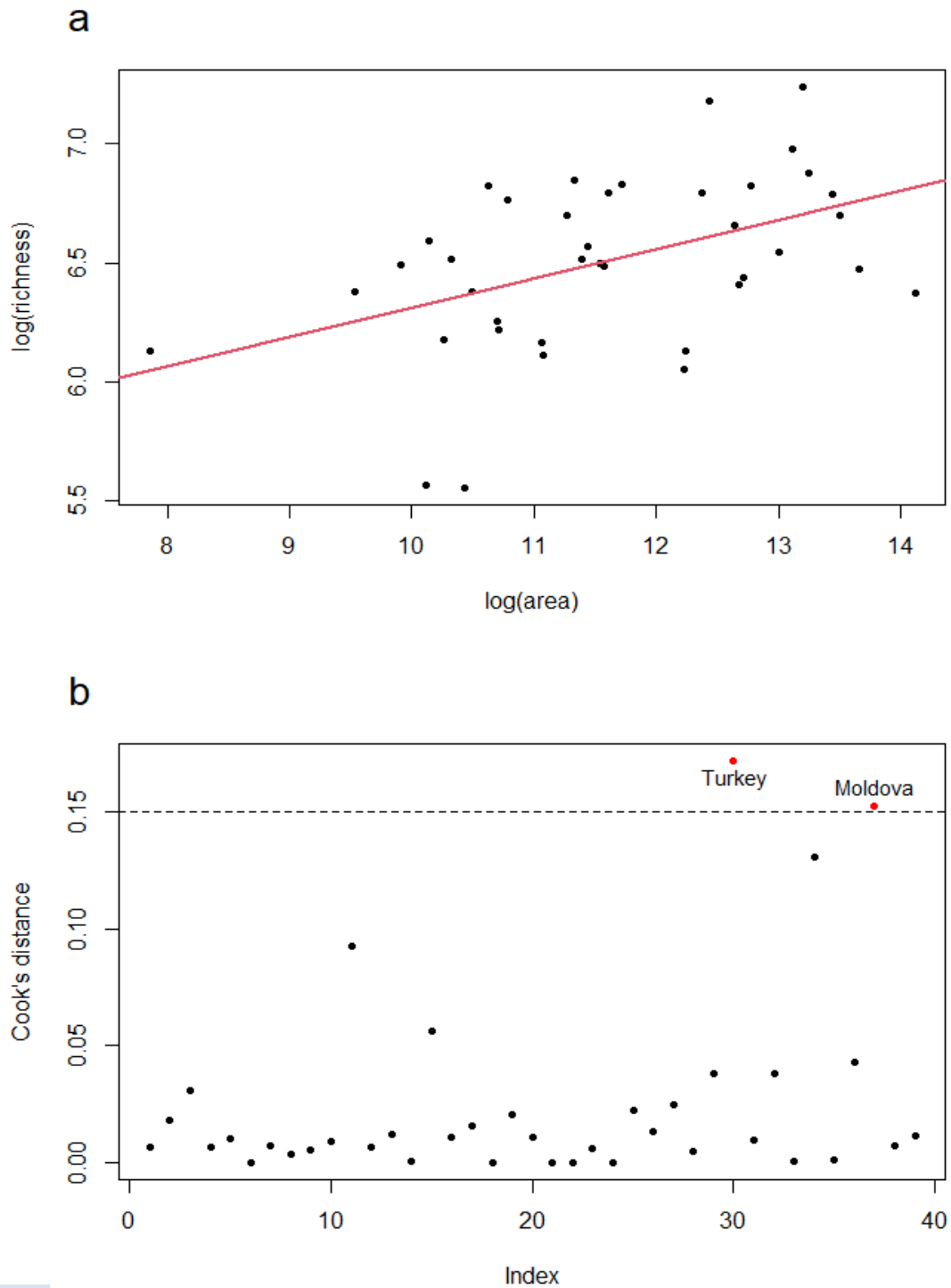


Table S3.1. AIC values of the power-law, negative exponential and Gompertz models for the spatial and climatic distance-decay models of each spider family in southern Europe.

Family	Spatial distance-decay			Climatic distance-decay		
	Power-law	Negative exponential	Gompertz	Power-law	Negative exponential	Gompertz
Agelenidae	-65.5	-163.3	-161.5	-68.1	-86.8	-86.7
Araneidae	-231.6	-411.9	-412.0	-212.4	-337.3	-337.2
Clubionidae	-158.9	-220.5	-224.4	-138.7	-190.9	-190.9
Dictynidae	-116.6	-167.9	-170.5	-97.0	-133.2	-133.2
Gnaphosidae	-4.2	-188.5	-192.4	-86.7	-115.0	-114.9
Hahniidae	-36.1	-99.5	-101.3	-30.7	-48.0	-48.0
Linyphiidae	-110.6	-190.8	-192.5	-107.1	-135.6	-135.6
Liocranidae	-54.5	-100.2	-101.7	-43.6	-67.8	-67.6
Lycosidae	-103.2	-320.1	-317.6	-137.7	-194.0	-194.1
Philodromidae	-206.5	-247.8	-248.5	-186.1	-239.8	-239.8
Pholcidae	-56.2	-134.8	-135.4	-49.0	-76.6	-76.5
Salticidae	-37.3	-302.0	-306.2	-167.9	-214.1	-214.0
Tetragnathidae	61.9	-263.0	-263.0	-182.7	-262.7	-262.6
Theridiidae	-171.1	-230.1	-231.0	-152.2	-213.7	-213.6
Thomisidae	98.2	-268.7	-270.5	-157.1	-209.6	-209.6

Table S3.2. AIC values of the power-law, negative exponential and Gompertz models for the spatial and climatic distance-decay models of each spider family in northern Europe.

Family	Spatial distance-decay			Climatic distance-decay		
	Power-law	Negative exponential	Gompertz	Power-law	Negative exponential	Gompertz
Agelenidae	-90.1	-152.8	-151.80	-102.9	-131.2	-130.8
Araneidae	-359.8	-554.0	-554.18	-401.5	-563.0	-564.4
Clubionidae	-313.7	-494.1	-499.37	-335.0	-471.8	-473.1
Dictynidae	-155.5	-273.8	-274.01	-168.7	-330.4	-337.7
Gnaphosidae	-241.1	-356.5	-355.21	-245.3	-352.6	-352.1
Hahniidae	-279.3	-394.3	-402.73	-279.5	-377.8	-377.8
Linyphiidae	-257.2	-443.2	-443.67	-263.6	-417.3	-421.5
Liocranidae	-175.3	-238.4	-238.61	-188.8	-232.3	-232.9
Lycosidae	-317.7	-531.8	-527.10	-329.3	-527.0	-527.7
Philodromidae	-221.3	-324.7	-326.48	-221.2	-321.4	-319.6
Pholcidae	-11.8	-45.8	-44.12	-12.1	-32.6	-32.4
Salticidae	-251.9	-377.1	-376.89	-280.1	-379.3	-378.0
Tetragnathidae	-411.8	-765.2	-772.91	-412.1	-730.3	-730.2
Theridiidae	-277.4	-420.8	-422.35	-312.9	-476.7	-417.3
Thomisidae	-268.7	-415.0	-409.15	-290.3	-425.7	-423.7

Table S3.3. Selected species for each family to build a family phylogenetic tree by pruning the phylogeny provided by Macías-Hernández et al., (2020a).

Family	Selected species
Agelenidae	<i>Agelena labyrinthica</i> (Clerck, 1757)
Araneidae	<i>Araneus diadematus</i> (Clerck, 1757)
Clubionidae	<i>Clubiona neglecta</i> (Pickard-Cambridge, 1862)
Dictynidae	<i>Brigittea latens</i> (Fabricius, 1775)
Gnaphosidae	<i>Drassyllus praeficus</i> (Koch, 1866)
Hahniidae	<i>Hahnina nava</i> (Blackwall, 1841)
Linyphiidae	<i>Walckenaeria antica</i> (Wider, 1834)
Liocranidae	<i>Scotina celans</i> (Blackwall, 1841)
Lycosidae	<i>Piratula latitans</i> (Blackwall, 1841)
Philodromidae	<i>Philodromus aureoles</i> (Clerck, 1757)
Pholcidae	<i>Pholcus phalangioides</i> (Fuesslin, 1775)
Salticidae	<i>Evarcha arcuata</i> (Clerck, 1757)
Tetragnathidae	<i>Tetragnatha extensa</i> (Linnaeus, 1758)
Theridiidae	<i>Crustulina guttata</i> (Wider, 1834)
Thomisidae	<i>Xysticus cristatus</i> (Clerck, 1757)

Table S3.4. Z-dep and p-values of parameter comparison (i.e. intercept and slope) of the spatial and climatic distance-decay curves of northern and southern Europe. In bold p-values < 0.05.

Family	North		South	
	p-value intercept	p-value slope	p-value intercept	p-value slope
Agelenidae	Z-dep = 1.86 p = 0.07	Z-dep = 1.54 p = 0.13	Z-dep = 2.59 p = 0.01	Z-dep = 6.10 p < 0.01
Araneidae	Z-dep = -0.27 p = 0.76	Z-dep = -0.16 p = 0.87	Z-dep = 2.83 p < 0.01	Z-dep = 5.31 p < 0.01
Clubionidae	Z-dep = 1.35 p = 0.19	Z-dep = 1.48 p = 0.14	Z-dep = 1.69 p = 0.09	Z-dep = 3.14 p < 0.01
Dictynidae	Z-dep = 0.76 p = 0.40	Z-dep = -0.22 p = 0.81	Z-dep = 2.05 p = 0.04	Z-dep = 2.89 p < 0.01
Gnaphosidae	Z-dep = 0.61 p = 0.55	Z-dep = 0.91 p = 0.36	Z-dep = 2.64 p < 0.01	Z-dep = 4.90 p < 0.01
Hahniidae	Z-dep = 1.82 p = 0.06	Z-dep = 1.60 p = 0.10	Z-dep = 2.90 p < 0.01	Z-dep = 4.03 p < 0.01
Linyphiidae	Z-dep = 1.28 p = 0.20	Z-dep = 1.35 p = 0.17	Z-dep = 2.33 p = 0.02	Z-dep = 4.56 p < 0.01
Liocranidae	Z-dep = 0.77 p = 0.44	Z-dep = 1.05 p = 0.29	Z-dep = 1.34 p = 0.18	Z-dep = 2.91 p < 0.01
Lycosidae	Z-dep = 0.81 p = 0.42	Z-dep = 1.15 p = 0.25	Z-dep = 3.34 p < 0.01	Z-dep = 6.77 p < 0.01
Philodromidae	Z-dep = 0.74 p = 0.46	Z-dep = 0.69 p = 0.49	Z-dep = 0.75 p = 0.45	Z-dep = 1.40 p = 0.16
Pholcidae	Z-dep = 1.69 p = 0.09	Z-dep = 1.81 p = 0.07	Z-dep = 3.78 p < 0.01	Z-dep = 5.63 p < 0.01
Salticidae	Z-dep = 0.37 p = 0.71	Z-dep = 0.77 p = 0.44	Z-dep = 3.32 p < 0.01	Z-dep = 5.82 p < 0.01
Tetragnathidae	Z-dep = 3.04 p < 0.01	Z-dep = 2.78 p < 0.01	Z-dep = -0.15 p = 0.88	Z-dep = 0.21 p = 0.83
Theridiidae	Z-dep = 0.50 p = 0.62	Z-dep = 0.95 p = 0.34	Z-dep = 0.54 p = 0.59	Z-dep = 1.64 p = 0.10
Thomisidae	Z-dep = 0.18 p = 0.86	Z-dep = 0.41 p = 0.68	Z-dep = 3.06 p < 0.01	Z-dep = 6.48 p < 0.01

Table S3.5. Parameter values (i.e. intercept and slope) of the spatial and climatic distance-decay curves of northern and southern Europe.

Family	North				South			
	Spatial		Climatic		Spatial		Climatic	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Agelenidae	1.01	0.38	0.95	0.18	0.73	1.17	0.60	0.18
Araneidae	0.96	0.08	0.97	0.08	0.94	0.26	0.89	0.03
Clubionidae	0.99	0.20	0.97	0.13	0.94	0.31	0.89	0.06
Dictynidae	0.99	0.13	1.00	0.47	0.86	0.46	0.79	0.07
Gnaphosidae	0.91	0.32	0.90	0.26	0.85	0.75	0.74	0.09
Hahniidae	0.99	0.13	0.95	0.02	0.91	0.79	0.79	0.08
Linyphiidae	0.96	0.38	0.93	0.28	0.77	0.68	0.69	0.11
Liocranidae	0.93	0.19	0.91	0.12	0.81	0.68	0.74	0.17
Lycosidae	0.93	0.29	0.92	0.24	0.89	0.67	0.82	0.17
Philodromidae	0.99	0.29	0.98	0.24	0.84	0.15	0.82	0.03
Pholcidae	0.96	0.34	0.87	0.07	0.92	0.73	0.75	-0.08
Salticidae	0.93	0.30	0.92	0.26	0.85	0.52	0.77	0.07
Tetragnathidae	1.01	0.07	0.98	0.01	0.89	0.06	0.89	0.05
Theridiidae	0.93	0.24	0.92	0.19	0.83	0.29	0.82	0.13
Thomisidae	0.97	0.26	0.97	0.23	0.88	0.42	0.79	0.00

- Appendix 4.1: Supplementary tables Chapter IV -

Appendix 4.1: Supplementary tables Chapter IV

Table S4.1. Accession numbers for the *cox1-5'* sequences of each of the 17 Iberian leaf beetle species preliminarily considered in this study.

Specie	Accession numbers
<i>Aphthona albertinae</i>	KF656660.1, KF656651.1, KF656650.1, KF656634.1, KF656626.1, KF656617.1, KF656616.1, KF656610.1, KF656609.1, KF656410.1, KF656409.1, KF656383.1, KF656382.1, KF656368.1, KF656367.1, KF656366.1, KF656338.1, KF656337.1, KF656336.1, KF656326.1, KF656325.1, KF656324.1, KF656323.1, KF656322.1, KF656313.1, KF656312.1, KF656311.1
<i>Aphthona melancholica</i>	KP763100.1, KF656295.1, KF656294.1, KF655125.1, KF655124.1, KF655123.1, KF655122.1
<i>Calomicrus suturalis</i>	KF652499.1, KF652504.1, KF652617.1, KF652642.1, KF652656.1, KF652674.1, KF652675.1, KF652676.1, KF652691.1, KF652692.1, KF652693.1, KF652721.1, KF652745.1, KF652798.1, KF652821.1, KF652901.1, KF652902.1, KF653861.1, KF653925.1, KF653926.1, KF653927.1, KF653928.1, KF653929.1, KF653941.1, KF653984.1, KF653985.1, KF654104.1, KF654105.1, KF654107.1, KF654108.1, KF654114.1, KF654140.1, KF654168.1, KF654205.1, KF654231.1, KF654244.1, KF654256.1, KF654314.1, KF654393.1, KF654401.1, KF654410.1, KF654419.1, KF654427.1, KF654446.1, KF654467.1, KF654497.1, KF654511.1, KF654516.1, KF654523.1, KF654528.1, KF654530.1, KF654549.1, KF654573.1, KF654669.1, KF654677.1, KF654681.1, KF654689.1, KF654698.1, KF654792.1, KF654927.1, KF654935.1, KF654952.1, KF654962.1, KF655276.1
<i>Chaetocnema paganettii</i>	KF655221.1, KF655220.1, KF654421.1
<i>Clytra espanoly</i>	KF655796.1, KF655814.1, KF655821.1, KF655822.1, KF655832.1, KF656088.1, KF656094.1, KF656153.1, KF656162.1, KF656179.1, KF656188.1, KF656520.1
<i>Cryptocephalus cantabricus</i>	KF653170.1, KF654692.1, KF655000.1, KF655046.1, KF655047.1, KF655048.1, KF655061.1, KF655062.1, KF655069.1, KF655089.1, KF655111.1, KF655140.1, KF655141.1, KF655354.1, KF655577.1, KF655663.1, KF655680.1, KF655681.1, KF655682.1, KF655847.1, KF655877.1, KF655917.1, KF655918.1, KF655936.1, KF655937.1, KF656104.1, KF656136.1, KF656142.1, KF656192.1, KF656248.1, KF656331.1
<i>Galeruca macchoi</i>	KF656030.1, KF656022.1, KF655826.1
<i>Gonioctena leprieuri</i>	KF654484.1, KF654509.1, KF654510.1, KF654694.1, KF654695.1, KF654696.1, KF654697.1, KF654712.1, KF654728.1, KF654765.1, KF654766.1, KF654784.1, KF654785.1, KF654825.1, KF654826.1, KF654848.1, KF654972.1, KF654973.1, KF654974.1, KF654986.1, KF654996.1, KF655003.1, KF655035.1, KF655054.1, KF655156.1, KF655169.1, KF655178.1, KF655192.1, KF655202.1, KF655217.1, KF655243.1, KF655244.1, KF655256.1, KF655257.1, KF655258.1, KF655298.1, KF655310.1, KF655311.1, KF655320.1, KF655377.1, KF655378.1, KF655379.1, KF655396.1, KF655805.1, KF655816.1, KF655836.1, KF655852.1, KF655866.1, KF655883.1, KF655892.1, KF655910.1, KF655923.1, KF655943.1, KF655952.1, KF655957.1, KF656041.1, KF656042.1, KF656043.1, KF656066.1, KF656119.1, KF656120.1, KF656124.1, KF656128.1, KF656138.1, KF656146.1, KF656150.1, KF656363.1, KF656364.1, KF656423.1, KF656424.1, KF656437.1, KF656438.1, KF656452.1, KF656453.1, KF656473.1, KF656474.1, KF656475.1

Table S4.1 (cont.).

<i>Gastrophysa janthina</i>	KF654267.1, KF654282.1, KF654283.1, KF654389.1, KF654390.1, KF654391.1, KF654398.1, KF654399.1, KF654400.1, KF654406.1, KF654407.1, KF654457.1, KF654458.1, KF654476.1, KF654482.1, KF654495.1, KF654507.1, KF654513.1, KF654547.1, KF654561.1, KF654570.1, KF654580.1, KF654600.1, KF654616.1, KF654659.1, KF654967.1, KF654968.1, KF654969.1, KF654993.1, KF654994.1, KF655001.1, KF655032.1, KF655033.1, KF655112.1, KF655154.1, KF655190.1, KF655200.1, KF6555617.1, KF6555618.1, KF6555619.1, KF6555636.1, KF6555637.1, KF6555638.1, KF6555652.1, KF6555653.1, KF6555654.1, KF6555668.1, KF6555669.1, KF655670.1, KF655671.1, KF655687.1, KF655688.1, KF6556289.1, KF6556376.1, KF6556395.1, KF6556396.1, KF6556564.1, KF6556573.1, KF6556574.1
<i>Hidrothassa fairmairei</i>	KF654791.1, KF656290.1, KF656306.1, KF656307.1, KF656346.1, KF656347.1, KF656348.1, KF656349.1, KF656584.1, KF656593.1
<i>Lachnaia gallaeca</i>	KF655537.1, KF655553.1, KF6555561.1, KF6555562.1, KF6555676.1, KF6555694.1, KF6555695.1, KF6555696.1, KF6555717.1, KF6555718.1, KF6555719.1, KF6555727.1, KF6555728.1, KF6555749.1, KF6555750.1, KF6555751.1, KF6555752.1, KF6555753.1, KF6555768.1, KF6555769.1, KF6555770.1, KF6555771.1, KF6555781.1, KF6555782.1, KF6556121.1, KF6556135.1, KF6556461.1
<i>Lachnaia tristigma</i>	KF652327.1, KF652877.1, KF652896.1, KF652915.1, KF652927.1, KF652946.1, KF652982.1, KF652994.1, KF653004.1, KF653050.1, KF653059.1, KF653084.1, KF653085.1, KF653094.1, KF653212.1, KF653213.1, KF653214.1, KF653220.1, KF653221.1, KF6533586.1, KF654474.1, KF654492.1, KF654504.1, KF654505.1, KF654558.1, KF654655.1, KF655045.1, KF655398.1, KF655399.1, KF655423.1, KF655424.1, KF655425.1, KF655439.1, KF655441.1, KF655456.1, KF655457.1, KF655458.1, KF655485.1, KF655486.1, KF655487.1, KF655488.1
<i>Luperus sulphuripes</i>	KF652720.1, KF652725.1, KF652726.1, KF652727.1, KF652734.1, KF652823.1, KF653056.1, KF654584.1, KF654585.1, KF654586.1, KF654608.1, KF654609.1, KF654610.1, KF654619.1, KF654620.1, KF654660.1, KF654834.1, KF654835.1, KF655121.1, KF655144.1, KF655170.1, KF655179.1, KF655203.1, KF655218.1, KF655219.1, KF655234.1, KF655247.1, KF655278.1, KF655301.1, KF655314.1, KF655323.1, KF655338.1, KF655356.1, KF655435.1, KF655452.1, KF655469.1, KF655481.1, KF655500.1, KF655550.1, KF655572.1, KF6555621.1, KF6555641.1, KF6555642.1, KF6555735.1, KF6555854.1, KF6555868.1, KF6555884.1, KF6555911.1, KF6555924.1, KF6555958.1, KF6555966.1, KF6555974.1, KF6556501.1, KF6556547.1
<i>Oedionychus cinctus</i>	KF652629.1, KF652884.1, KF652925.1, KF652934.1, KF652959.1, KF652981.1, KF654335.1, KF654420.1, KF654471.1, KF654472.1, KF654614.1, KF654776.1, KF654841.1, KF654842.1, KF655030.1, KF655105.1, KF655106.1, KF655626.1, KF655674.1, KF655693.1
<i>Psylliodes gougeleti</i>	KF653112.1, KF653113.1, KF653145.1, KF653146.1, KF653147.1, KF653167.1, KF653168.1, KF653235.1, KF653251.1, KF653252.1, KF653784.1, KF653787.1, KF653788.1, KF653789.1, KF653791.1, KF653792.1, KF653799.1, KF655158.1, KF655160.1, KF6555646.1, KF6555779.1, KF6555793.1, KF6555947.1, KF6555993.1, KF656055.1, KF656056.1, KF656057.1, KF656099.1, KF656107.1, KF656108.1, KF656115.1, KF656116.1, KF656187.1, KF656210.1, KF656211.1, KF656212.1, KF656269.1, KF656590.1, KF656591.1, KF656596.1, KF656629.1, KF656655.1, KF656656.1, KF656657.1, KF656666.1
<i>Psylliodes heydeni</i>	KF656665.1, KF656664.1, KF6566630.1, KF6566602.1, KF655692.1
<i>Smaragdina reyi</i>	KF654721.1, KF654756.1, KF654779.1, KF654893.1, KF655504.1, KF655516.1, KF655538.1, KF655539.1, KF655563.1, KF655564.1, KF6555629.1, KF6555659.1, KF6555660.1, KF6555661.1, KF6555677.1, KF6555678.1, KF656062.1, KF656102.1, KF656103.1, KF656111.1, KF656344.1, KF656358.1, KF656389.1, KF656390.1, KF656430.1, KF656431.1, KF656464.1, KF656465.1

Table and figure index

Table index

Table 1.1. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations.....	44
Table 1.2. Pseudo-r2 and AIC of the power-law, negative exponential and Gompertz models of the distance-decay pattern of South American terrestrial mammals	46
Table 2.1. Models best supported by the data for each of the simulated distance-decay curves, based on the AIC	62
Table 2.2. P-values of the z_{dep} statistic in the comparison of the intercept and slope between distance-decay models of different groups of Iberian vertebrates	64
Table 3.1 Classification of spider families according to dispersal traits	78
Table 3.2. Values of β_{sim} , β_{sne} , and β_{ratio} for each spider family	82
Table 4.1. AIC, pseudo-r2 and p-value for the linear, power-law, negative exponential and Gompertz functions used to model the genetic-spatial distance relationship in each Iberian beetle species.....	95
Table 4.2. Association between the dispersal ability of species and the existence of significant genetic-spatial distance relationship	96
Table S1.1. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations, species detection probability = 1	173
Table S1.2. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations, species detection probability = 0.75	174
Table S1.3. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations, species detection probability = 0.5	175
Table S1.4. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations, species detection probability = 0.25.....	176
Table S2.1. Results for the calibration of the z_{dep} test under the null hypothesis	183
Table S2.2. Results for the calibration of the z_{dep} test under the alternative hypothesis	183
Table S2.3. Parameters estimation for each Iberian vertebrate order	184

Table S3.1. AIC values of the power-law, negative exponential and Gompertz models for the spatial and climatic distance-decay models of each spider family in southern Europe	216
Table S3.2. AIC values of the power-law, negative exponential and Gompertz models for the spatial and climatic distance-decay models of each spider family in northern Europe	216
Table S3.3. Selected species for each family to build a family phylogenetic tree by pruning the phylogeny provided by Macías-Hernández et al., (2020a).....	217
Table S3.4. z_{dep} and p-values of parameter comparison of the spatial and climatic distance-decay curves of northern and southern Europe	217
Table S3.5. Parameter values (i.e. intercept and slope) of the spatial and climatic distance-decay curves of northern and southern Europe	218
Table S4.1. Accession numbers for the cox1-5' sequences of each of the 17 Iberian leaf beetle species preliminarily considered	221

Figure index

Figure 1.1. Hypothetical example of how species ranges in a landscape may lead to a sigmoidal distance-decay relationship.....	40
Figure 1.2. Schematic representation of the two axes of variation in simulation scenarios.....	42
Figure 1.3. Model comparison, based on AIC values, of the power-law, negative exponential and Gompertz functions in distance-decay simulations.....	45
Figure 1.4. Distance-decay models for the four range size groups of South American mammals.....	46
Figure 2.1 Example showing how pairwise dependence of similarity/distance values results in the inflation of degrees of freedom and increase with sample size of the number of site-blocks and the number of similarities/distance values.....	55
Figure 2.2. Example of distance-decay simulations for the calibration of the Z_{dep} statistic.....	60
Figure 2.3. Distribution of p-values across 1000 replicates in all simulated scenarios.....	63
Figure 2.4. Distance-decay models of Iberian vertebrates' orders.....	64
Figure 3.1. Distance-decay models using standardized spatial and climatic distances for 15 spider families in southern Europe.....	80
Figure 3.2. Distance-decay models using standardized spatial and climatic distances for 15 spider families in northern Europe.....	81
Figure 4.1. Relationship between genetic and spatial distance for all the Iberian beetle species considered.....	94
Figure 4.2. Haplotype networks for the 10 species of Iberian leaf beetles considered.....	97
Figure 5.1. Known records for <i>Lachnaia gallaeca</i> , <i>Lachnaia pseudobarathraea</i> , and <i>Lachnaia tristigma</i> , using the core and extended datasets.....	106
Figure 5.2. Climatic niches of <i>Lachnaia gallaeca</i> , <i>Lachnaia pseudobarathraea</i> , and <i>Lachnaia tristigma</i> , as represented in two Principal Components Analysis axes.....	107
Figure 5.3. Potential distributions of <i>Lachnaia gallaeca</i> , <i>Lachnaia pseudobarathraea</i> and <i>Lachnaia tristigma</i> during the Last Glacial Maximum.....	108
Figure S3.1. log(richness)~log (area) curve and Cook distances for the studied European territories.....	215



Understanding the factors affecting the spatial structure of biodiversity at various levels of biological organization is one of the main goals for ecologists. This thesis aims to develop new methodologies to study the mechanisms driving the spatial distribution of biological diversity as inferred from biogeographical and macroecological patterns. It explores the usefulness of a sigmoidal function to fit the distance-decay pattern and how its shape changes with the species distribution size, a test statistic for parameters comparison, the application of these methods to study current large-scale diversity patterns of European spiders and the extension of these methods to study the genetic-spatial distance relationship, using endemic species of Iberian leaf beetles as a case-study.