



UNIVERSITAT DE  
BARCELONA

# Genomic divergence along the continuum of speciation in a recent evolutionary radiation of montane grasshoppers

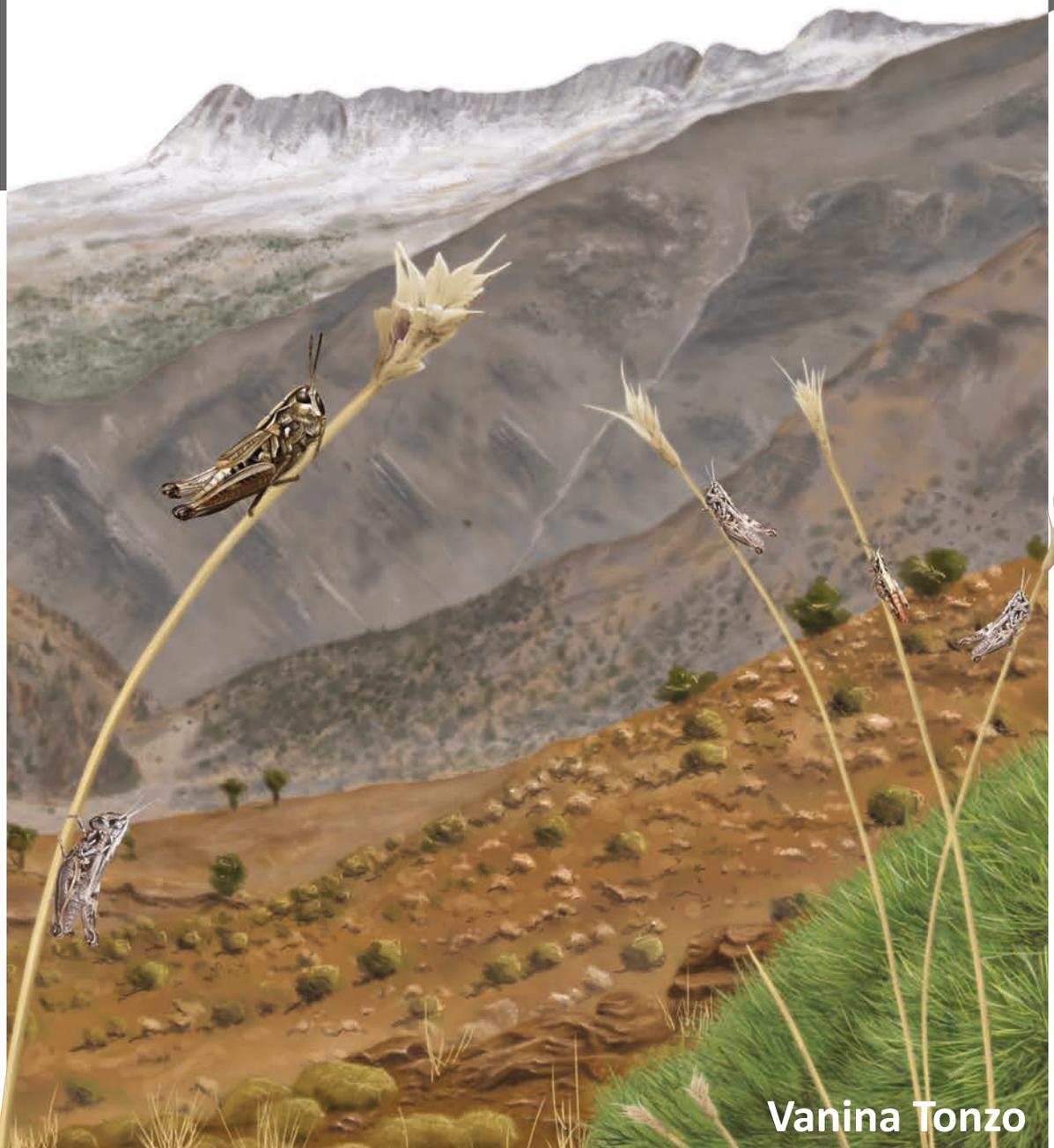
Vanina Tonzo

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

**GENOMIC DIVERGENCE ALONG THE  
CONTINUUM OF SPECIATION IN A RECENT  
EVOLUTIONARY RADIATION OF MONTANE  
GRASSHOPPERS**



**Vanina Tonzo**



Cover and chapters illustrations are an original creation of Amparo Hidalgo Galiana (intellectual property ownership). Any form of reproduction, distribution, public communication or transformation of this property without prior authorization of the authors is forbidden.

IG: @ahg\_ilustracion

Design and layout by Eduardo de la Asunción, José Moral Marín

Portfolio web: <https://uamocbcn.wixsite.com/uamoc>



UNIVERSITY OF BARCELONA  
FACULTY OF BIOLOGY

Doctoral Programme in Biodiversity  
Department of Evolutionary Biology, Ecology and Environmental  
Sciences

# Genomic divergence along the continuum of speciation in a recent evolutionary radiation of montane grasshoppers

DOCTORAL THESIS

Dissertation submitted by

**Vanina Tonzo**

To obtain the PhD degree by the University of Barcelona

Thesis conducted in Doñana Biological Station (CSIC) and  
the University of Barcelona



UNIVERSITAT DE  
BARCELONA

Facultat de Biologia

A handwritten signature in blue ink, appearing to read 'J. Ortego'.

A handwritten signature in blue ink, appearing to read 'A. Papadopoulou'.

Supervisors: Joaquín Ortego and Anna Papadopoulou

A handwritten signature in blue ink, appearing to read 'M. Arnedo'.

Tutor: Miquel Arnedo

A handwritten signature in blue ink, appearing to read 'V. Tonzo'.

Predocctoral student: Vanina Tonzo

Barcelona, 2020



**GENOMIC DIVERGENCE ALONG THE  
CONTINUUM OF SPECIATION IN A RECENT  
EVOLUTIONARY RADIATION OF MONTANE  
GRASSHOPPERS**

**DOCTORAL THESIS**



*A mis padres*

*A Andrés y Sebastián*



During the development of this thesis I was supported by a FPI predoctoral fellowship funded by the Spanish Ministry of Economy and Competitiveness (grant BES-2015-73159). Research was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (grants CGL2014-54671-P and CGL2017-83433-P).



## ACKNOWLEDGMENTS

*I will take the license here to use different languages.*

La canción dice que *“uno vuelve siempre a los viejos sitios donde amo la vida”*. Esta tesis es parte de un largo camino de vuelta a uno de esos sitios...

Y aunque volver ha conllevado algunos sacrificios, también he conocido gente increíble y he vivido momentos inolvidables y yo no puedo más que agradecer a todos los que han hecho esto posible.

Voy a comenzar agradeciendo a los que han sido responsables de mi supervisión durante estos cuatro años y medio, mis directores Joaquín Ortego y Anna Papadopoulou. Les agradezco la oportunidad de cumplir un deseo, que no es la tesis, ni ser doctora sino el de investigar, de preguntarme cosas y buscar respuestas y aprender. Y yo he aprendido muchísimo de ambos. Espero no haberlos decepcionado... (¡Al menos no demasiado!). Antes de que todo se transformara en lo que hoy es, Miquel Arnedo fué quién me abrió la puerta a un grupo de investigación y quien me dió la oportunidad de creer que aún era posible jugar a ser científica. Te lo he dicho muchas veces, pero te lo vuelvo a repetir, estaré agradecida por siempre por tu confianza. A veces no hace falta un doctorado en ciencias para pensar como un científico y yo agradezco profundamente a mi primer mentor, mi papá. Desde que tengo recuerdos me ha transmitido el amor por la naturaleza y me ha enseñado a preguntarme el porqué de todas las cosas. Increíblemente hoy me sigue alentando y cuestionando como el primer día. Mi mamá y mi papá me han regalado creo que lo más preciado, pero también peligroso para el ser humano, la libertad de elegir y de hacer todo lo que yo deseara asumiendo los riesgos, pero sabiendo que ellos están allí junto con mis hermanos, Andrés y Sebastián, para festejar cada logro. Poca gente lo sabe, pero yo ya he sido IP de un grupo donde según mi hermano Andrés, daba muchas ordenes y los tenía explotados! Pobres mis hermanos... Me enorgullece decir que mi familia se expande por el mundo: Azul, Sydney, Melicuccá. Gracias a todos y cada uno por siempre estar allí, por alentarme. Algunos ya no están físicamente, pero forman parte de mi. Agraieixo a la meva família a Catalunya: Pruvi, Cris, Xavi, Greta i Sira per rebre'mi amb els

braços oberts i per tot l'afecte a l'igual que a tot aquest grup de gent diversa i meravellosa: Montse, Albert, Vero, Ivan, Dolors, Juana, Rita, Olivia, Nora, Valentina. Y a mi familia en Sevilla: Luis y Joaquín. Luis, vos has sido un gran apoyo durante mi vida en Sevilla. Guardo en mi corazón tus “*y cuando vuelves a casa con Pumuki*”, cada vez que me iba de viaje. Me hacías sentir que tu casa era mi casa. Y así la sigo sintiendo. Que suerte que Pumuki no sepa leer, no me cabe duda de que sería el lector más crítico.

Si uno tiene mucha suerte en la vida, los amigos también se transforman en familia y yo creo que la tuve. Gracias Edu, Josep, Bárbara, Javi, Ana, Sergi, Enric, Paula, Cris, Laufer y Gemma. Edu, si tuviese que agradecerte todo no me alcanzaría una tesis. Mi amigo del alma, cuantas cosas hemos pasado juntos en estos 12 años y cuántas más vendrán.

En la Estación Biológica de Doñana he recolectado las mejores muestras de personas posibles para el experimento que es la tesis. Maria, Dani, Sara, Ampí, Conxi, Irene, Isa, Rubén, Miguel, Arlo, Victor, Christoph, Joan, Javi, Eneko, Irene y todos los que no estoy nombrando. Me llevo la música, los bailes, la feria, las cenas, la alameda columnas norte, las “calçotadas sevillanas”, las risas, las charlas de café y también las de cerveza. Aquí voy a hacer una mención especial a la Asociación sin fines de lucro FFBP, cuyos miembros no puedo develar por máxima seguridad. Gracias a la locura desmedida y al apoyo de sus miembros... que siempre estén por ahí pululando en el universo de los *K significa potasio*.

Gracias a mis compañeros y amigos de la Universidad de Barcelona: Alba, Marc, Adrià, Luis, Jagoba, Rosa y a mis favourit postdocs Jesús and Martina. I learned so much from you all.

En el tercer año de mi tesis tuve la oportunidad de hacer a wonderful stay at Riesgo Lab in the Natural History Museum of London. It was, as I called, “my blue healing period”. Ana, Sergi, Vassia, Nathan, Nadia, Bruna, Ana Serra, Angelina, Belén, Cristina, Carlos, thank you so much for everything. I had so much fun with you all. Ana, mi amiga y guía científica espiritual y otras cosas más que sólo entre locas entendemos. Gracias por todo.

El periodo de escritura de los capítulos finales quedará en mi memoria no sólo por la tesis en si, sino por la increíble circunstancia de la pandemia generada por el COVID. No sé bien como comenzó todo, solo se que un día no me pude librar más de las continuas interrupciones de dos bonitas moscas vecinas, Maria Alba y Laura. Gracias por las charlas profundas en los recesos de la escritura que algunas veces terminaban en vermouths.

Kele, la vida es curiosa y da muchas vueltas y como una vez me dijiste, yo tuve varias vidas

en una. Agradezco encontrarnos en una de ellas.

Dolors Vinyolas, gracias por tu apoyo y cálida respuesta a cada una de mis preguntas. Quiero dar también, unas gracias especiales a los investigadores que forman parte de mi comité científico en la EBD y la UB: Carles Vilà, Xavi Picó, Marta Riutort, Salvi Carranza y Julie Pujade y a los que, sin ser parte de ningún comité, han mostrado interés en mi trabajo durante la tesis y me han animado a continuar, José A. Godoy, Rosa Fernandez, Cesc Murrià.

Esta tesis no podría haberse llevado a cabo sin la colaboración y apoyo logístico proporcionado por el Laboratorio de Ecología Molecular (LEM-EBD), en especial a su responsable, Ana Piriz y el Laboratorio de Sistemas de Información Geográfica y Teledetección (LAST – EBD) de la Estación Biológica de Doñana. También agradezco al Centro de Supercomputación de Galicia (CESGA) y a la Infraestructura Científico-Técnica Singular de Doñana (ICTS-RBD) por el acceso a los recursos informáticos.

Una vez más, a todos los que me han acompañado en este viaje, gracias...



*¿Por qué no aceptar lo que estaba ocurriendo  
sin pretender explicarlo,  
sin sentar las nociones  
del orden y de desorden?*

Julio Cortázar



## ABSTRACT

Unraveling the proximate processes that have shaped genetic variation of populations and led to lineage diversification and speciation is fundamental to understand the origin of present-day biodiversity patterns at both local and global scales. Cyclical Pleistocene glaciations played a pivotal role in the evolution of mid-latitude montane biotas, leading to distributional shifts that generated multiple opportunities for both allopatric speciation and secondary contact. In this thesis, we integrate genomic, morphological and environmental data and combine a diverse array of analytical procedures to investigate processes of genomic and phenotypic divergence acting at different stages along the speciation continuum. Specifically, this thesis focuses on the subgenus *Dreuxius* (genus *Omocestus*), an Ibero-Maghrebian complex of montane grasshoppers distributed across the main mountain ranges of the region. Phylogenomic and geometric morphometric analyses supported the recent Pleistocene origin (< 1 Ma) of the complex, two reticulation events involving lineages at different stages of the diversification continuum and the phenotypic distinctiveness of most sister taxa (Chapter 1). Moreover, phylogenetic reconstructions did not recover the reciprocal monophyly of taxa from Iberia and northwestern Africa, supporting two overseas Pleistocene migration events between the two continents (Chapter 1). Integrative species delimitation analyses focusing on the Pyrenean endemics *O. navasi* and *O. antigai* did not support their current taxonomic status, pointing to the presence of a single species with little phenotypic variation, a wide climatic niche, and a marked genetic structure explained by limited population connectivity across the abrupt landscapes of the region (Chapter 2). Analyses of inter-specific gene flow in the partially sympatric *O. minutissimus* and *O. uhagonii* rejected the hypothesis of contemporary hybridization but revealed past introgression in the area where the distributions of the two species overlap (Chapters 1 and 3). This supports a scenario of historical gene flow after secondary contact followed by the evolution of reproductive isolation that currently prevents hybridization among sympatric populations (Chapter 3). Demographic inference and testing of alternative models of intraspecific gene flow within each of the narrow-endemics *O. bolivari* and *O. femoralis* supported population genetic admixture during glacial periods and postglacial colonization of sky islands, rather than long-term isolation, as the scenario best explaining the contemporary distribution of genomic variation in the two taxa. The results of this thesis emphasize the key role of range-shifts driven by Pleistocene glacial cycles in promoting not only allopatric

divergence but also secondary contact and genetic admixture among previously isolated gene pools. Overall, this thesis highlights the importance of combining population and phylogenomic approaches to improve our understanding about the processes governing the diversification of montane biotas across evolutionary scales spanning the continuum of speciation, from populations to species.





# **INDEX**

<b>GENERAL INTRODUCTION</b>	<b>3</b>
<b>OBJECTIVES</b>	<b>17</b>
<b>IMPACT AND AUTORSHIP REPORT OF THE PUBLICATIONS</b>	<b>25</b>
<b>CHAPTER I</b>	<b>29</b>
<b>CHAPTER II</b>	<b>65</b>
<b>CHAPTER III</b>	<b>105</b>
<b>CHAPTER IV</b>	<b>141</b>
<b>GENERAL DISCUSSION</b>	<b>175</b>
<b>CONCLUSIONS</b>	<b>185</b>
<b>REFERENCES</b>	<b>191</b>
<b>APPENDICES</b>	<b>205</b>
<b>APPENDIX I</b>	<b>207</b>
<b>APPENDIX II</b>	<b>217</b>
<b>APPENDIX III</b>	<b>235</b>
<b>APPENDIX IV</b>	<b>241</b>
<b>ORIGINAL JOURNAL ARTICLES</b>	<b>257</b>



# GENERAL INTRODUCTION





## General introduction

Evolution generates endless diversity of life on Earth. Organizing living organisms into fundamental units and studying the underlying processes that generate biological diversity are central themes in evolutionary biology. In this thesis we refer to species as separately evolving metapopulation lineages based on the evolutionary species concept advocated by Wiley (1978), Mayden (1997) and De Queiroz (2007). Much progress has been made in comprehending the core evolutionary processes involved in speciation such as mutation, migration, genetic drift and different forms of selection (Coyne & Orr, 2004; Gavrilets, 2000; Schluter, 2000). However, how these processes interact to drive divergence across genomes and ultimately form new species remains surrounded by several fundamental questions (Conflitti, Shields, Murphy, & Currie, 2014; Powell et al., 2013). From a temporal perspective, speciation can be defined as a **continuous and gradual accumulation of genetic differences** that occurs as two lineages diverge from one another on the pathway to **reproductive isolation** (see Table 1) (Mallet, 2008; Powell et al., 2013; Shaw & Mullen, 2014). Indeed, different stages along this continuum of divergence can potentially involve very different processes (Nosil & Feder, 2012). Lineage differentiation can be promoted by barriers linked to **ecological differentiation** (Bolnick & Fitzpatrick, 2007; Mérot et al., 2017), **sexual selection** (Pascoal, Mendrok, Wilson, Hunt, & Bailey, 2017) and/or **geographic isolation** (Boucher, Zimmermann, & Conti, 2016; Lagomarsino, Condamine, Antonelli, Mulch, & Davis, 2016). Some of the barriers developed through

### Glossary

**Adaptive radiation:** Increase in the rate of speciation within a clade driven primarily by biotic factors (i.e., a ‘key innovation’) occurring in sympatry and coupled with ecomorphological divergence.

**Allopatry:** Geographical separation of populations.

**Character displacement:** Pattern of greater differences between a pair of taxa in ecologically related traits when they are found together in sympatry than expected based on differences in allopatry.

**Gene tree:** Evolutionary history of a specific gene across populations/species.

**Hybrid zones:** Area in which genetically distinct populations, lineages or species meet, mate and produce hybrids.

**Incomplete lineage sorting:** Genetic (allelic) variants that fail to coalesce in a common ancestor between lineage splitting events; it is more common when internal branches of the phylogeny are short and/or population sizes are large and often results in gene trees that differ from species trees.

**Introgression:** The incorporation (usually via hybridization and backcrossing) of alleles from one entity (species) into the gene pool of a second, divergent entity (species).

**Parapatry:** Abutting geographical ranges.

**Reinforcement:** Increase in prezygotic isolation between hybridizing populations in response to any type of selection against interspecific matings, regardless of whether hybrids themselves are unfit.

**Species tree:** The branching or reticulating history of species.

**Sympatry:** Overlapping geographical ranges of two entities (species) that encounter each other in at least some places.

any of these processes can be irreversible and lead to complete **reproductive isolation** (Table 1), while others could be transient and allow different rates of hybridization and genetic introgression (Hewitt, 1990; Mallet, 2008; Olave, Avila, Sites, & Morando, 2018). The fact that the different stages of the speciation process do not have obvious boundaries, and transitions among them are not necessarily unidirectional, adds difficulty to the study of how local populations diverge into fully-fledged species (Hendry, Bolnick, Berner, & Peichel, 2009; Smith, Pennell, Dunn, & Edwards, 2020). Moreover, the scarcity or total lack of fossil record for many taxa only allows us to observe the surviving lineages (while enforcing us to ignore the extinct ones), which adds further difficulty when trying to reconstruct the diversification process (Naciri & Linder, 2020).

**Table 1** Simplified summary of reproductive isolation mechanisms that can lead to speciation. Table adapted from Naciri & Linder (2020).

Mechanism	Synonym	Description
Allopatry	Isolation by distance (IBD)	Gene flow reduced due to geographical isolation, in part dependent on the lineage having a narrow niche, so that it is excluded from part of the region, fragmenting the distribution range; results in neutral divergence
Ecology	Isolation by adaptation (IBA) Isolation by environment (IBE)	Hybrids between species adapted to different niches are maladapted to the parental niches
Assortative mating and sexual selection	Temporal, mechanical isolation, behavioural isolation, many premating mechanisms	Mate selection based on shared traits, markers, timing or geography
Post-zygotic divergence		Bateson–Dobzhansky–Muller incompatibilities

## 1 | Radiations as natural laboratories to study speciation

Research on **recent evolutionary radiations** is essential to infer the causes and constrains that shape and maintain biodiversity (Rundell & Price, 2009). Deciphering the processes that influence the progression of genetic and phenotypic divergence is of paramount importance to understand how biotic and abiotic factors impact evolution and infer the tempo and mode of species diversification (Simões et al., 2016). In this context, **recent radiations** offer a superb study system to infer the processes involved in the

establishment of species boundaries and gain insights into the population-level dynamics underpinning species formation. Despite the recognized complexity of these events, radiations have been traditionally classified in two main types based on the driving diversification factors: (i) **adaptive** (see Glossary) and (ii) **non-adaptive radiations**. While adaptive radiations have been intensively studied, non-adaptive radiations (Gittenberger, 1991) have received comparatively less attention (Futuyma, 1998; Mayr, 1942; Simpson, 1953; Vogler & Goldstein, 1997). Furthermore, non-adaptive radiations have been suggested as a control scenario, a prevailing phase (Fitzpatrick, Fordyce, & Gavrillets, 2008; Gavrillets & Losos, 2009; Warren, Cardillo, Rosauer, & Bolnick, 2014) or even, in specific instances, a synergic operator with adaptive diversifications (Losos & Ricklefs, 2009; Simões et al., 2016).

Adaptive and non-adaptive radiations also differ in the spatial context at which the processes of genomic divergence take place. Adaptive radiations generally involve ecological differentiation among sympatric lineages (Givnish, 1997). On the contrary, non-adaptive radiations generally imply lineage divergence with minimal ecological diversification, often resulting in **allopatric** or **parapatric** (see Glossary) taxa that might or might not come into secondary contact (Butlin, Galindo, & Grahame, 2008; Rundell & Price, 2009). Moreover, ecological differentiation in adaptive radiations is usually coupled with the evolution of reproductive isolation barriers developed in most cases while in **sympatry**. In contrast, non-ecological processes of divergence promote a slow and gradual accumulation of reproductive isolation on lineages via genetic drift or as a fortuitous by-product of divergent selection on other traits (Coyne & Orr, 1989; Fitzpatrick, 2002; Sasa, Chippindale, & Johnson, 1998). Thus, many species that diverged in allopatry are not reproductively isolated and continue to form **hybrid zones** when and where they meet (Butlin & Hewitt, 1985; Naciri & Linder, 2020). **Secondary contact** among lineages or species can have different outcomes with important evolutionary consequences depending on the strength of reproductive isolation barriers (Abbott et al., 2013; Mallet, 2005; Mayr, 1963). At one extreme, barriers to gene exchange may breakdown and lead to the collapse of formerly distinct species with extensive admixture of parental genotypes (i.e., **speciation reversal** or **lineage fusion**; e.g., Kearns et al., 2018; Taylor et al., 2006). At the opposite extreme, reduced fitness of hybrids and strong selection against them will favor the rapid evolution of **barriers to gene flow** (i.e., prezygotic isolation), which can ultimately lead to total reproductive isolation and culminate in the completion of the

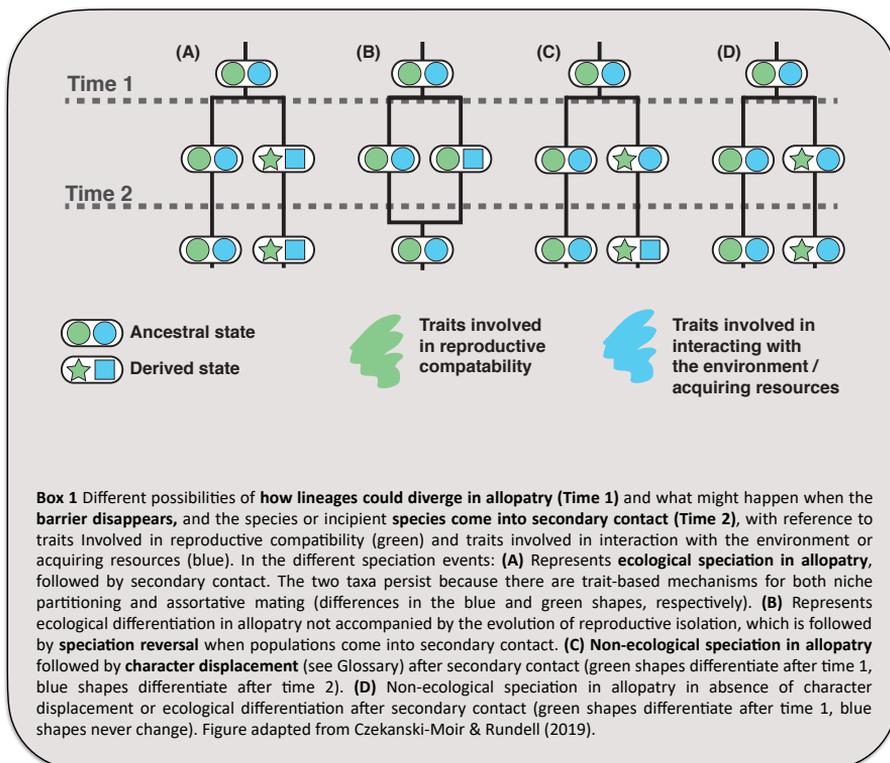
speciation process (i.e., reinforcement of isolation; Dobzhansky, 1937; Servedio & Kirkpatrick, 1997). An intermediate scenario is the formation of **tension zones** with a variable geographical width determined by the equilibrium between interspecific gene flow and selection against hybrids (Fitzpatrick et al., 2009; Smadja & Butlin, 2011; Sobel et al., 2010).

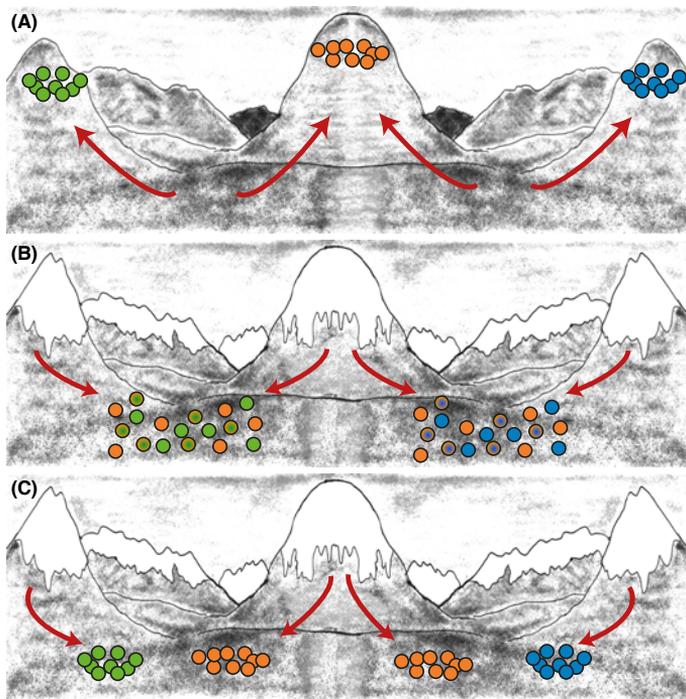
## 2 | Effects of Pleistocene glaciations on the diversification of mid-latitude organisms

The **Quaternary period** (~2.6 Ma) has been dominated by repeated **glacial cycles** that led to important **latitudinal and altitudinal range shifts** in most organisms. **Pleistocene climatic oscillations**, resulting from orbital eccentricity with ~100 ka periodicity cycles (Milankovitch cycles; Hays et al., 1976), caused important alterations in insolation, the recurrent growth and retreat of ice sheets, and sea-level changes (e.g., Ruddiman et al., 1989). Specifically, the **Middle-Late Pleistocene** (130 ka to 10 ka) was characterized by **short-term** glacial cycles and **high-amplitude climatic changes**, with a maximum extension of the ice sheet during the last glacial maximum (**LGM**; 20–18 ka) (Flantua & Hooghiemstra, 2018; Jouzel et al., 2007). Glacial cycles have been considered both to promote **allopatric speciation** (Knowles, 2000) and to **inhibit speciation** through genetic homogenization (Klicka & Zink, 1997; Zink & Slowinski, 1995). Beyond their impact on speciation, Pleistocene glaciations also had an important role in shaping the spatial distribution of genetic variation at both the population and species levels (Hedin, Carlson, & Coyle, 2015; Ribera & Vogler, 2004). However, these impacts varied depending on latitude and regional topography, from **extinction-colonization** dynamics at higher latitudes to elevational shifts and more complex processes of population **fragmentation** and **connectivity** at lower latitudes such as the tropics or **temperate regions** (Gómez & Lund, 2006; Hewitt, 1996). The location and extent of **Pleistocene refugia**, which were essential for the persistence of species during unfavorable climatic periods, largely depended on species-specific niche requirements (Bennett & Provan, 2008; Stewart et al., 2010). Temperate species currently inhabiting low and medium elevation areas generally restricted their distribution to southern refugia during the glacial phases (i.e., **glacial refugia**) and expanded during the interglacial periods (Hewitt, 2000; Teacher, Garner, & Nichols, 2009), while cold-adapted species that currently have fragmented populations at high elevations or are restricted to

high latitudes (i.e., **interglacial refugia**), had much more widespread distributions during glacial periods (Reid et al., 2019; Wang, Jiang, Xie, & Li, 2013).

**Mid-latitude alpine and montane taxa** represent small-scale replicates of cold-adapted species living at high latitudes (Hewitt, 1996; Shepard & Burbrink, 2011). Specifically, interglacial periods forced cold-adapted lineages from mid and low latitude regions to shift their distributions towards high elevations to satisfy their specific habitat and climate niche requirements, leading to range fragmentation and divergence in interglacial refugia (e.g., DeChaine & Martin, 2005; Djamali et al., 2012) (Figure 1A). Conversely, glacial periods led to downslope **migrations** in montane organisms, which likely experienced net range **expansions, colonization** of new suitable habitats, and **secondary contact** and **admixture** among diverged lineages (Hewitt, 1990; Marko & Hart, 2011) (Figure 1B). In some cases, isolation periods (see Box 1) contributed to the enhancement of genetic and phenotypic differences among lineages, leading to allopatric speciation, either adaptively (i.e., due to **divergent natural selection**) or non-adaptively (i.e., due to **genetic drift**) (Czekanski-Moir & Rundell, 2019; Hewitt, 1996, 1999) (Figure 1C). In other cases, reproductive isolation did not evolve while in refugia, and therefore secondary contact during range shifts either resulted in the collapse of formerly distinct lineages (i.e., **speciation reversal**; Kearns et al.,





**Figure 1** Schematic representation of processes of genetic divergence promoted by Pleistocene glacial cycles. (A) Interglacial isolation of populations on mountain tops (interglacial refugia), (B) a scenario of glacial expansion and secondary contact with gene flow, and (C) a scenario of glacial expansion with reproductively isolated lineages (i.e., species).

2018; Maier, Vandergast, Ostoja, Aguilar, & Bohonak, 2019), allowed for **introgressive hybridization** between divergent lineages (e.g., Seixas, Boursot, & Melo-Ferreira, 2018; Wiens, Engstrom, & Chippindale, 2006) or contributed to the completion of the **speciation** process via **reinforcement** of reproductive isolation (Butlin & Hewitt, 1985; Nevado, Contreras-Ortiz, Hughes, & Filatov, 2018).

A special case of high-elevation refugia are **sky islands**, which are geographically isolated mountains surrounded by inhospitable lowlands and characterized by evolutionary dynamics similar to those experienced by their oceanic counterparts (McCormack et al., 2008). Unlike extensive massifs and mountain ranges such as the Pyrenees or the Atlas Mountains that offer a certain degree of connectivity among populations, sky islands often harbor endemic lineages **restricted to them**, with **small population sizes** and **reduced levels of genetic diversity** resulting from extended periods of interglacial isolation. As a result, species and communities restricted to sky islands are at high risk of extinction, especially in the context of ongoing climate warming and the progressive shrink of alpine environments (Atkins et al., 2020; Rubidge et al., 2012).

### 3 | The geographical setting: The Iberian Peninsula and northwestern Africa

Ecologically heterogeneous and/or geographically complex landscapes, such as those characterizing the Mediterranean region, offer contrasting selective environments and barriers to dispersal, the ideal contexts for ecological and allopatric divergence, respectively. **Mountain ranges** have also played a predominant role in modulating the effects of past climatic changes on Mediterranean biodiversity (Médail & Quezel, 1997; Médail & Diadema, 2009). In some instances, they acted as barriers to dispersal, limiting northern colonization routes for some lineages, but also provided refugia for **cold adapted organisms** during **interglacials** (Hewitt, 1996; Schmitt, 2007). The Iberian Peninsula and the Maghreb region present a key geographic position between two continents, are characterized by a complex topography and a wide range of microclimates, and experienced a limited impact of Pleistocene glaciations, aspects that contributed to the diversification of numerous organismal groups and is reflected in their rich biological diversity and high rates of local endemism (Blondel et al., 2010; Médail & Diadema, 2009).

The **main Iberian mountain ranges** (the Pyrenees, the Cantabrian Mountains, and the Iberian, Central and Baetic Systems) and **river basins** (the Ebro, Duero, Tagus, Guadiana and Guadalquivir Basins) formed from the **Upper Cretaceous** to the **Middle Miocene** (~70 to 11 Ma) (Sainz & Faccenna, 2001). As well as, the orogeny of the **Atlas Mountains** in Northwest Africa started ~70 Ma (Stets & Wurster, 1982; Teixell et al., 2007). Despite Europe and North Africa are currently separated by the Mediterranean Sea, emerged lands during the **Messinian Salinity Crisis** (5.96 to 5.33 Ma) favored the migration of terrestrial, flightless, and saltwater-intolerant organisms (Duggen, Hoernle, van den Bogaard, Rüpke, & Phipps Morgan, 2003; Krijgsman, Hilgen, Raffi, Sierro, & Wilson, 1999). The refilling of the Mediterranean Sea and its permanent connection with the Atlantic Ocean marked the end of the Messinian Salinity Crisis (~ 5.3 Ma) and the consequent isolation of most terrestrial biotas on either side of the **Strait of Gibraltar**, which became an important **barrier to gene flow** between southern Europe and North Africa for many organisms (Carranza, Harris, Arnold, Batista, & González de la Vega, 2006; Cosson et al., 2005). However, different studies have revealed **post-Messinian gene flow** across the Strait of Gibraltar for some terrestrial organisms (Agustí et al., 2006; Graciá et al., 2013; Lavergne et al., 2013). This might have been fostered by **eustatic** sea level changes during **Quaternary glacial periods**, which reduced the distance between the two continents

during the coldest **Pleistocene** stages and led to the emergence of islands and shoals that likely favored stepping-stone dispersal (Cosson et al., 2005).

## 4 | The subgenus *Dreuxius* (genus *Omocestus*) species complex

Grasshoppers (Orthoptera: Acrididae) are highly sensitive to environmental alterations and habitat **fragmentation**, as they often present **low dispersal capability** (e.g., brachypterous species) and/or high levels of habitat specificity and **niche conservatism** (e.g., Ortego, García-Navas, Nogueras, & Cordero, 2015; Yadav, Stow, & Dudaniec, 2019) that can trigger **adaptation processes** (Chakrabarty & Schielzeth, 2020; Savolainen, Lascoux, & Merilä, 2013). Although ecological divergence is also likely to have played a role in the diversification of some grasshopper lineages (e.g., Grace et al., 2010), speciation in this group has been mostly linked to processes of population divergence in allopatry (Mayer et al., 2010), which are often the result of range shifts due to climatic oscillations (Knowles, 2000). On the other hand, range expansions due to Pleistocene climatic oscillations or other causes have brought many closely related species into secondary contact, resulting in contemporary or historical **hybridization** in **contact zones** and **introgression** (Kawakami et al., 2007; Rohde, Hau, Weyer, & Hochkirch, 2015). These aspects make grasshoppers an excellent model system for studying the processes underlying genetic divergence and the evolution of reproductive isolation mechanisms at different stages along the speciation continuum (Knowles, 2008; Koot, Morgan-Richards, & Trewick, 2020).

This PhD Thesis focuses on the **Ibero-Maghrebian** subgenus *Dreuxius* Defaut, 1988 (genus *Omocestus* Bolívar, 1878), a complex of predominantly **montane** grasshoppers comprised by nine taxa distributed in the **Iberian Peninsula** (6 taxa) and **northwestern Africa** (3 taxa) (Clemente, García, & Presa, 1991). Nominal species were originally described based on phenotypic traits and courtship behavior (see Box 2 for species pictures and details) (Clemente, García, Arnaldos, Romera, & Presa, 1999; Clemente, García, & Presa, 1991; Puissant, 2008; Ragge & Reynolds, 1998). All taxa within the complex are of diurnal activity, present an **annual life cycle**, are predominantly **graminivorous**, and occupy **open habitats** dominated by cushion and thorny shrub formations (e.g., *Erinacea* sp., *Festuca* sp., *Juniperus* sp., *Thymus* sp.) that they use as



*O. alluaudi* Uvarov, 1927. (Female)  
D: Rif, Middle and High Atlas Mountains in Northwest Africa. W: brachypterous; A: from 2500 to 3400 m.  
Photo credit: V. Tonzo



*O. lecerfi* Chopard, 1937. (Male)  
D: Rif, Middle and High Atlas Mountains in Northwest Africa. W: brachypterous, vestigial hindwings; A: from 1000 to 2500 m. S: *O. maroccanus* Chapman, 1937.  
Photo credit: Jorge Iñiguez Yarra



*O. lepineyi* Chopard, 1937. (Male)  
D: High Atlas Mountains in Northwest Africa. W: brachypterous; A: from 2800 to 3200 m.  
Photo credit: V. Tonzo



*O. bolivari* Chopard, 1939. (Female)  
D: Baetic range. W: brachypterous vestigial hindwings; A: from 1500 to 2900 m. S: *O. casaresi* Chopard, 1936.  
Photo credit: José Ramón Correas



*O. minutissimus* (Brullé, 1832). (Female)  
D: Central System, Baetic range Iberian System and East Pyrenees. W: brachypterous; A: from 0 to 2500 m. S: *O. burri* Uvarov, 1936; *O. knipperi* Harz, 1982; *O. llorentae* Pascual, 1976.  
Photo credit: José Ramón Correas



*O. uhagonii* (Bolívar, 1876). (Female)  
D: Central System. W: brachypterous; A: from 1600 to 2100 m.  
Photo credit: José Ramón Correas



*O. femoralis* Bolívar, 1908. (Male)  
D: Baetic range. W: brachypterous; A: from 1500 to 1900 m. Morphologically similar to *O. navasi*.  
Photo credit: Luis J. González



*O. antigaí* (Bolívar, 1897). (Female)  
D: Pyrenees. W: brachypterous; A: from 1600 to 2200 m. S: *Stenobothrus braelemanni* Azam, 1906. Morphologically similar to *O. navasi*.  
Photo credit: Gilles San Martín

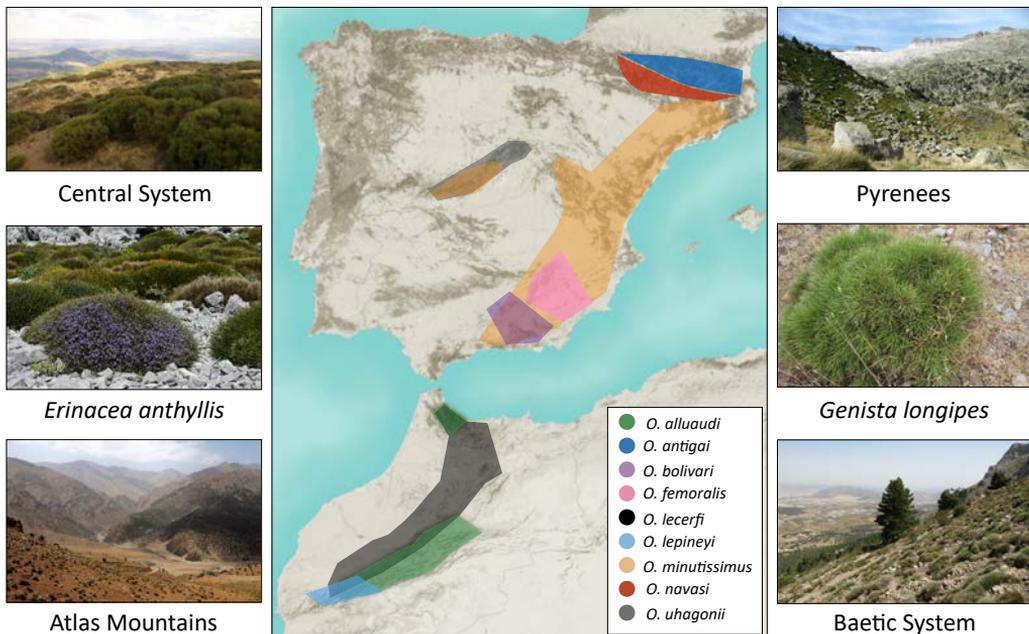


*O. navasi* Bolívar, 1908. (Female)  
D: Pre-Pyrenees. W: brachypterous; A: from 1000 to 1600 m. Morphologically similar to *O. antigaí*.  
Photo credit: Gilles San Martín

**Box 2** Species of the subgenus *Dreuxius* (genus *Omocestus*). **D**, distribution; **A**, altitudinal range; **W**, wing morphology; **S**, synonyms.

refuge (Clemente, García, & Presa, 1990; Gangwere & Morales Agacino, 1970). Most of the species form **isolated populations** at **high elevations** in different mountain systems of the Iberian Peninsula (Pyrenees and Iberian, Central and Baetic Systems) and Northwest Africa (Atlas and Rif Mountains) (Cigliano et al., 2019) (Figure 2). Despite most of the taxa within the complex presenting **allopatric distributions**, there are two exceptions: the Iberian *O. minutissimus* (Brullé 1832) and the Maghrebian *O. lecerfi* Chopard 1937. Both species present comparatively wider elevational and distributional ranges, partially overlapping with the rest of Iberian and Northwest African species of the complex, respectively, and

with which they often form **sympatric populations** (Cigliano et al., 2019; Clemente, García, & Presa, 1991). All species, particularly females, are markedly **brachypterous**, which is expected to extraordinarily limit their dispersal capacity, reduce gene flow even at short spatial scales, contribute to population divergence, and favor processes of speciation due to geographic isolation (Waters, Emerson, Arribas, & McCulloch, 2020; e.g., Huang, Hill, Ortego, & Knowles, 2020). Their similar habitat requirements, little phenotypic differences, and predominantly non-overlapping distributions suggests that species within the complex most likely evolved in allopatry after population fragmentation during Pleistocene glacial cycles. However, secondary contact among some of the species within the complex might have also contributed to historical and/or contemporary post-divergence gene flow. In addition, species distributions span from large suitable and connected areas in the main massifs (e.g., *O. antigai* and *O. navasi* in the Pyrenees) to tiny patches in isolated mountains (e.g., *O. bolivari* and *O. femoralis* in the Baetic Sky Island Archipelago), offering an extraordinary biogeographical scenario to elucidate the consequences of population isolation across different stages along the **speciation continuum** (Figure 3).



**Figure 2** Map showing the distribution of the different species of the subgenus *Dreuxius* (genus *Omocestus*). Photos illustrate the similar habitats occupied by the species in the different mountain ranges and two examples of cushion and thorny shrub formations used by them as refuge. Photo credits: J. Ortego and V. Tonzo.





# OBJECTIVES

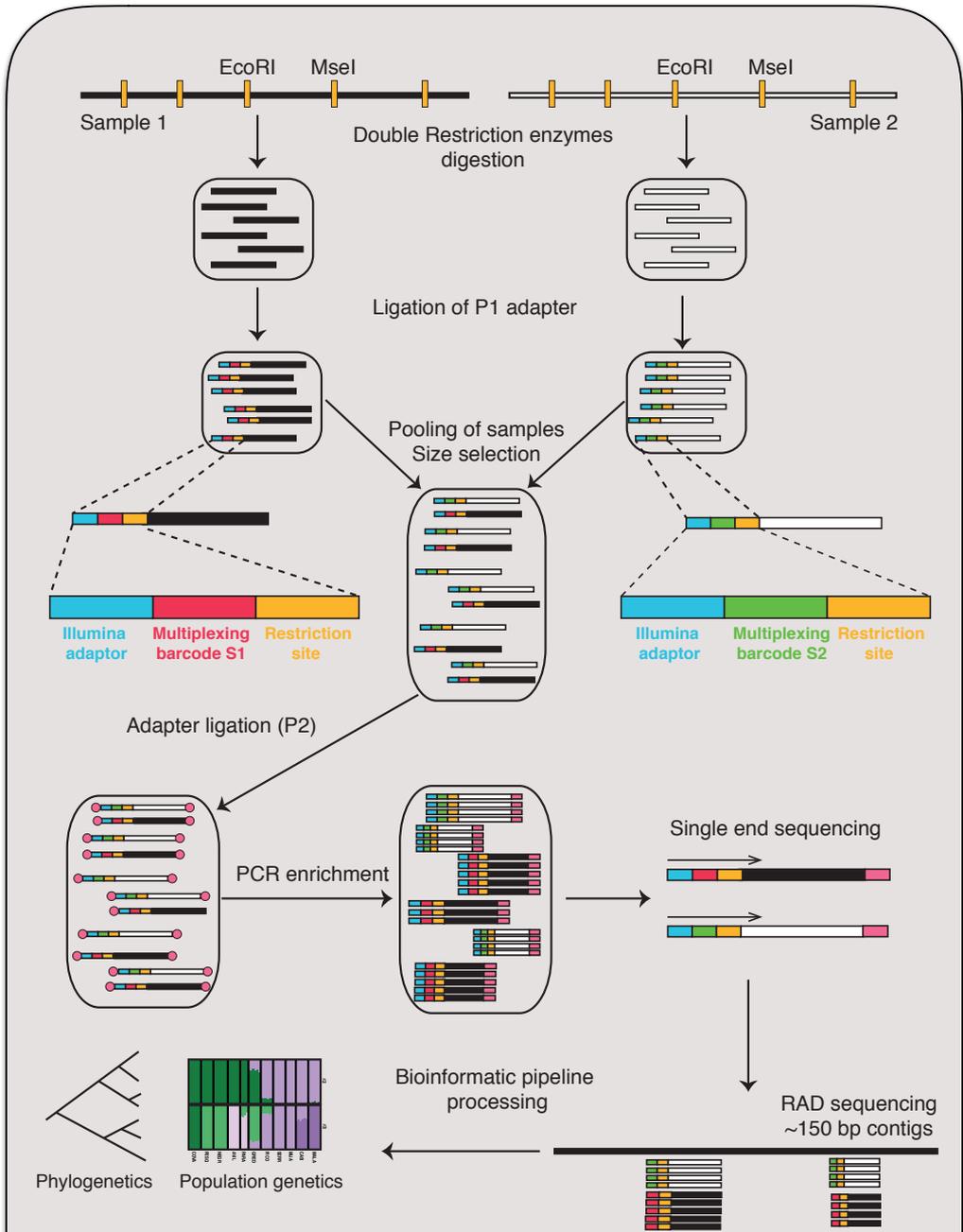


## Objectives

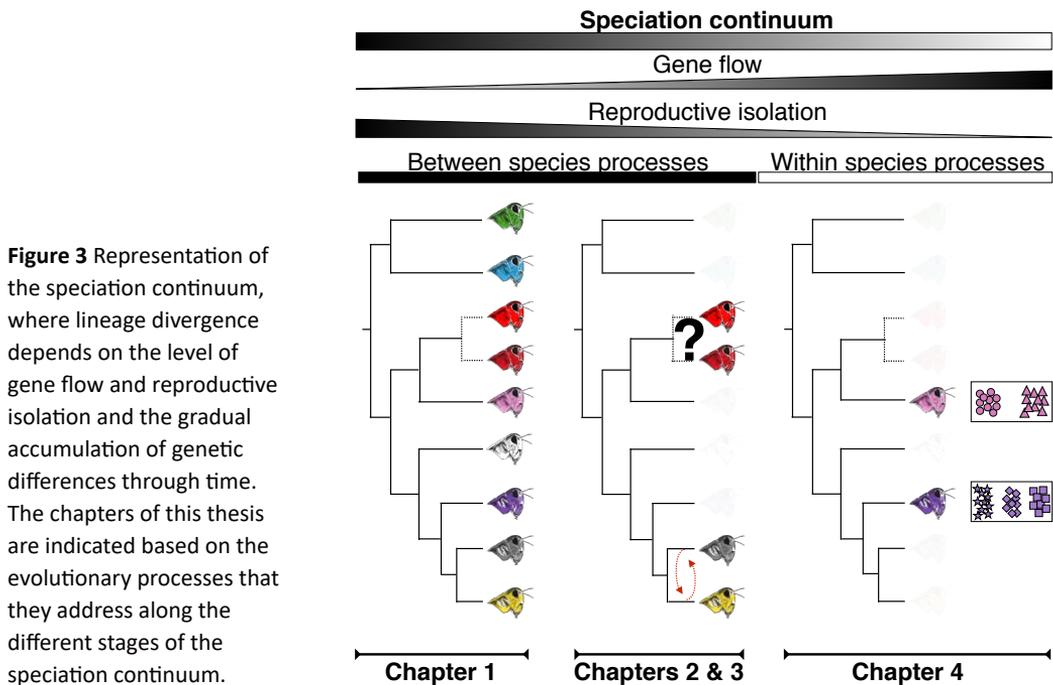
In this thesis, I integrate genomic data obtained via restriction site-associated DNA sequencing (Box 3; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), phenotypic information through geometric morphometric analyses (Bookstein, 1992) and a suite of phylogenetic and population genetic analytical approaches to shed light on the processes responsible for speciation, trait evolution, and spatial patterns of genetic variation in Ibero-Maghrebian grasshoppers of the subgenus *Dreuxius*. In each chapter of this thesis I address the study of evolutionary processes taking place at different stages along the speciation continuum (Figure 3), from the reconstruction of the phylogenetic relationships among the species that conform the *Dreuxius (Omocestus)* subgenus radiation to the inference of the demographic processes that have shaped spatial patterns of population genetic diversity and structure within species. Thus, my PhD dissertation aims to understand the main drivers underlying the diversification processes implicated at different stages of speciation. Specific objectives for each chapter are presented below and shown graphically in Figure 3.

**Chapter 1** In this chapter, we first aim to disentangle the evolutionary relationships and the timing of diversification of the *Dreuxius* species complex by combining high throughput sequence data with novel phylogenetic inference methods able to infer reticulation events and explicitly accounting for gene flow in parameter estimation. In a second step, we employed a geometric morphometric approximation to characterize species' phenotypic variation and evaluate alternative scenarios of trait evolution in a suite of morphological characters involved in sexual selection and/or that have been traditionally used to delineate the different taxa within the species complex.

**Chapter 2** In this chapter we aim to shed light on the taxonomy and evolutionary history of *O. antigai* and *O. navasi*, two Pyrenean species with a controversial taxonomic history. First, we tested the hypothesis that the current taxonomic designation of the complex is supported by genomic and phenotypic data. In particular, we used genomic data and alternative coalescent-based methods to perform species delimitation analyses, reconstruct the phylogenetic relationships among all sampled populations of the two taxa, and determine whether the obtained inferences are compatible with current taxonomic delineation. Second, we assessed the processes underlying spatial patterns of genomic and phenotypic trait variation among populations and tested whether these have been shaped by environmental and/or geographic factors.



**Box 3** Schematic workflow of the **ddRADseq** protocol followed in this thesis. An important component to achieve the main goals in this thesis relies on the use of double-digest RADseq, a restriction site associated DNA high throughput sequencing technique (Peterson et al., 2012). This sequencing approach fulfills the requirements of working with **non-model organisms** (Baxter et al., 2011), provides sufficient **resolution to infer recent evolutionary processes** at both **intraspecific** (Lavretsky et al., 2019) and **shallow phylogenetic** scales (Noguerales, Cordero, & Ortego, 2018), and has proven to be useful to study **contemporary hybridization** (Bartomeus, Molina, Hidalgo-Galiana, & Ortego, 2020) and **discern incomplete lineage sorting from historical introgression** (Saarman & Pogson, 2015).



**Chapter 3** In this chapter we focus on *O. minutissimus* and *O. uhagonii*, two species of the subgenus *Dreuxius* with partially overlapping distributions and that frequently form sympatric populations at high elevations in the Central System Mountains from the Iberian Peninsula. Using this system as a case study, we analyzed spatial patterns of contemporary and past hybridization (or their lack thereof) in order to gain insights into the processes contributing to maintain species boundaries in these two closely related taxa that might have weak or recently evolved reproductive isolation mechanisms. To this end, we used genomic data and a coalescent-based simulation framework to evaluate alternative scenarios of historical and contemporary hybridization and estimate the timing, magnitude and directionality of interspecific gene flow.

**Chapter 4** Southeastern Iberia constitutes an important biodiversity hotspot due to the interplay between a vibrating geological history, extraordinarily complex topography, and limited impact of Quaternary glaciations. At the same time, the Baetic mountain ranges of the region constitute important interglacial refugia for several cold-adapted organisms that currently persist forming severely fragmented populations at high elevations. Here, we used genomic data and a suite of analytical approaches to shed light on the demographic processes underlying contemporary patterns of genetic variation in *O.*

*bolivari* and *O. femoralis*, two narrow-endemic species inhabiting the sky island archipelago of the Baetic System. Specifically, we aim to elucidate the demographic responses of species with similar habitat requirements to historical and contemporary processes of population connectivity and fragmentation in the context of Pleistocene climatic oscillations (i.e., elevational/distributional shifts) and understand the temporal scale at which they have contributed to shape contemporary patterns of genetic variation in montane taxa from temperate regions.





# Impact and autorship report of the publications

Como directores de la tesis doctoral titulada “*Genomic divergence along the continuum of speciation in a recent evolutionary radiation of montane grasshoppers*” realizada por Vanina Faviola Tonzo, presentamos el siguiente informe sobre la contribución de la doctoranda en las publicaciones en coautoría que componen la tesis:

**Capítulo 2.** Tonzo, V., Papadopoulou, A. y Ortego, J (2019) Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology*, 28:3869-3886. <https://doi.org/10.1111/mec.15189>

Contribución de la doctoranda: Participación en la concepción y diseño del estudio y los análisis, muestreos de campo, realización del trabajo de laboratorio (obtención de datos genómicos y morfológicos), análisis de datos, interpretación de los resultados y redacción del manuscrito.

Acerca de la revista: *Molecular Ecology* en el Journal Citation Reports (JRC) de 2018 tiene un índice de impacto de 5.855. Se encuentra en el número 8 de 50 en el área de **biología evolutiva** (1<sup>er</sup> cuartil), 13 de 165 en el área de **ecología** (1<sup>er</sup> cuartil) y en el número 40 de 299 en el área de **bioquímica y biología molecular** (1<sup>o</sup> cuartil).

**Capítulo 3.** Tonzo, V., Papadopoulou, A. y Ortego, J. (2020) Genomic footprints of an old affair: SNP data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. *Molecular Ecology*, en prensa. <https://doi.org/10.1111/mec.15475>

Contribución de la doctoranda: Participación en la concepción y diseño del estudio y los análisis, realización del trabajo de laboratorio (obtención de datos genómicos), análisis de datos, interpretación de los resultados y redacción del manuscrito.

Acerca de la revista: *Molecular Ecology* en el Journal Citation Reports (JRC) de 2018 tiene un índice de impacto de 5.855. Se encuentra en el número 8 de 50 en el área de **biología evolutiva** (1<sup>er</sup> cuartil), 13 de 165 en el área de **ecología** (1<sup>er</sup> cuartil) y en el número 40 de 299 en el área de **bioquímica y biología molecular** (1<sup>o</sup> cuartil).

Sevilla, a 28 de mayo de 2020

Nicosia, a 28 de mayo de 2020



Dr. Joaquín Ortego Lozano  
(director de la Tesis)  
Departamento de Ecología Integrativa  
Estación Biológica de Doñana  
Consejo Superior de Investigaciones Científicas



Dra. Anna Papadopoulou  
(directora de la Tesis)  
Department of Biological Sciences  
University of Cyprus





# CHAPTER I

## RETICULATE EVOLUTIONARY HISTORY IN A RECENT RADIATION OF MONTANE GRASSHOPPERS REVEALED BY GENOMIC DATA





# Reticulate evolutionary history in a recent radiation of montane grasshoppers revealed by genomic data

Vanina Tonzo<sup>1</sup> | Adrià Bellvert<sup>2</sup> | Joaquín Ortego<sup>1</sup>

<sup>1</sup> Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC); Avda. Américo Vespucio 26 – 41092; Seville, Spain

<sup>2</sup> Department of Evolutionary Biology, Ecology and Environmental Sciences, and Biodiversity Research Institute (IRBio), Universitat de Barcelona; Av. Diagonal, 643 – 08028; Barcelona, Spain

## Abstract

Inferring the ecological and evolutionary processes underlying lineage and phenotypic diversification is of paramount importance to shed light on the origin of contemporary patterns of biological diversity. However, reconstructing phylogenetic relationships in recent evolutionary radiations represents a major challenge due to the frequent co-occurrence of incomplete lineage sorting and introgression. In this study, we combined high throughput sequence data (ddRADseq), geometric morphometric information, and novel phylogenetic inference methods that explicitly account for gene flow to infer the evolutionary relationships and the timing and mode of diversification in a complex of Ibero-Maghrebian montane grasshoppers of the subgenus *Dreuxius* (genus *Omocestus*). Our analyses supported the phenotypic distinctiveness of most sister taxa, two events of historical introgression involving lineages at different stages of the diversification continuum, and the recent Pleistocene origin (< 1 Ma) of the complex. Phylogenetic analyses did not recover the reciprocal monophyly of taxa from Iberia and northwestern Africa, supporting overseas migration between the two continents during the Pleistocene. Collectively, these results indicate that periods of isolation and secondary contact linked to Pleistocene glacial cycles likely contributed to both allopatric speciation and post divergence gene flow in the complex. This study exemplifies how the integration of multiple lines of evidence can help to reconstruct complex histories of reticulated evolution and highlights the important role of Quaternary climatic oscillations as a diversification engine in the Ibero-Maghrebian biodiversity hotspot.

**Keywords:** allopatric speciation, introgression, phenotypic divergence, Pleistocene radiations, reticulate evolution

# 1 | Introduction

Recent evolutionary radiations have traditionally received much attention because the signatures of speciation events have not been fully erased by time and, thus, provide the potential to infer processes from fine-scale patterns of genetic and phenotypic variation (Knowles & Chan, 2008; Shaffer & Thomson, 2007; Shaw & Danley, 2003). Phylogenies provide essential tools to infer the processes responsible for speciation, investigate trait evolution, and discern among alternative biogeographic scenarios (Barraclough, Vogler, & Harvey, 1998; Knowles & Chan, 2008). Inferring the mode and timing of speciation is crucial to reconstruct the diversification process and unravel the origin of contemporary patterns of biological diversity. However, reconstructing phylogenetic relationships among recently diverged species can be extremely challenging. One of the main issues is the frequent co-occurrence of incomplete lineage sorting and introgression (Edwards, 2009; Maddison, 1997; Nichols, 2001). Although phylogenetic relationships among species have been typically represented as bifurcating branches (Felsenstein, 2004; Haeckel, 1866), which implicitly assumes that diversification occurred without reticulation (Coyne & Orr, 2004; Mallet, 2007), there are multiples examples of gene flow among independently evolving taxa (Blair et al., 2019; Burbrink & Gehara, 2018; Feder, Egan, & Nosil, 2012; Harrison & Larson, 2014). Thus, failing to account for post-divergence gene flow when estimating evolutionary processes may produce statistical inconsistencies, incorrect phylogenies, inaccurate estimates of key demographic parameters, and wrong biogeographic inferences (Burbrink & Gehara, 2018; Flouri, Jiao, Rannala, & Yang, 2018; Solís-Lemus, Bastide, & Ané, 2017).

Speciation events driven by high amplitude climatic variations in the Late Pleistocene (130 ka to 10 ka), are among the best-known examples of recent diversification processes (Flantua & Hooghiemstra, 2018; Roy et al., 1996). Repeated range expansions and contractions driven by Quaternary glacial cycles have extraordinarily contributed to the diversification of montane and alpine biotas (Hewitt, 1996; Sandel et al., 2011; Shepard & Burbrink, 2008; Wallis, Waters, Upton, & Craw, 2016). Interglacial periods pushed cold-adapted lineages from mid and low latitude regions to shift their distributions towards high elevations to satisfy their specific habitat and climate niche requirements, leading to range fragmentation and divergence in interglacial refugia (e.g., DeChaine & Martin, 2005; Djamali et al., 2012). Conversely, glacial periods forced downslope migrations in montane organisms, which likely experienced net range expansions, colonization of new suitable

habitats in lowlands and secondary contact and admixture among closely related lineages (Excoffier, Foll, & Petit, 2009; Hewitt, 1990; Marko & Hart, 2011). Glacial advances also contributed to allopatric divergence in alpine biotas, particularly those inhabiting extensively glaciated and topographically complex regions where distributional ranges got severely fragmented by ice caps and valley glaciers and populations likely became confined to highly isolated ice-free refugia (Wallis et al., 2016). Isolation periods contributed to genetic and phenotypic differentiation, fueling allopatric adaptive (i.e., divergent natural selection) and non-adaptive (i.e., genetic-drift) lineage divergence and/or reinforcing existing species boundaries (Czekanski-Moir & Rundell, 2019; Hewitt, 1996, 1999). If reproductive isolation did not evolve while in refugia, secondary contact during range shifts resulted in the collapse of formerly distinct lineages (i.e., speciation reversal; Kearns et al., 2018; Maier, Vandergast, Ostojka, Aguilar, & Bohonak, 2019), introgressive hybridization (e.g., Salzburger, Baric, & Sturmbauer, 2002; Schweizer et al., 2019), or even contributed to complete the speciation process via reinforcement of reproductive isolation (Butlin & Hewitt, 1985; Hewitt, 1996; Nevado, Contreras-Ortiz, Hughes, & Filatov, 2018). For these reasons, Pleistocene glacial cycles have been considered to both promote range fragmentation and allopatric speciation (Knowles, 2000) and inhibit speciation through genetic homogenization (Klicka & Zink, 1997; Zink & Slowinski, 1995).

The Iberian Peninsula and Maghreb regions present a rich biodiversity and an alike species composition due to their close geographical proximity, similar climatic and ecological conditions, complex topography, and a geological history that has led to multiple episodes of connectivity and isolation for terrestrial biotas distributed in the two continents (Blondel & Aronson, 2002; Krijgsman, 2002; Meulenkaamp & Sissingh, 2003). As a result, this region is an important center of diversification for numerous organism groups and considered a hotspot for animal and plant biodiversity (Myers, McKelvy, & Burbrink, 2020; Rodríguez-Sánchez, Pérez-Barrales, Ojeda, Vargas, & Arroyo, 2008). The re-opening of the Strait of Gibraltar at the beginning of the Pliocene led to the loss of the last intercontinental land connection established during the desiccation of the Mediterranean Basin in the Messinian Salinity Crisis (Husemann, Schmitt, Zachos, Ulrich, & Habel, 2014; Krijgsman, 2002), a phenomenon representing the starting point for the diversification of many lineages whose distributional ranges resulted fragmented under the new geographic setting (e.g., Faille, Andújar, Fadrique, & Ribera, 2014; Veith, Kosuch, & Vences, 2003). However, empirical evidence has also supported that the shortening of coastline distances

during Pleistocene glacial periods facilitated fauna exchanges and gene flow between southern Europe and North Africa (Agustí, Garcés, & Krijgsman, 2006; Carranza, Harris, Arnold, Batista, & González de la Vega, 2006; Graciá et al., 2013). In this context, resolving the phylogenetic relationships among Ibero-Maghrebian species complexes and estimating their timing of divergence is essential to unravel whether their origin is linked to Pleistocene range expansions/contractions (e.g., Knowles, 2000) and sea-level low stands (e.g., Graciá et al., 2013) or, rather, compatible with a protracted history of diversification dating back to the late Miocene (e.g., Faille et al., 2014; Hidalgo-Galiana & Ribera, 2011).

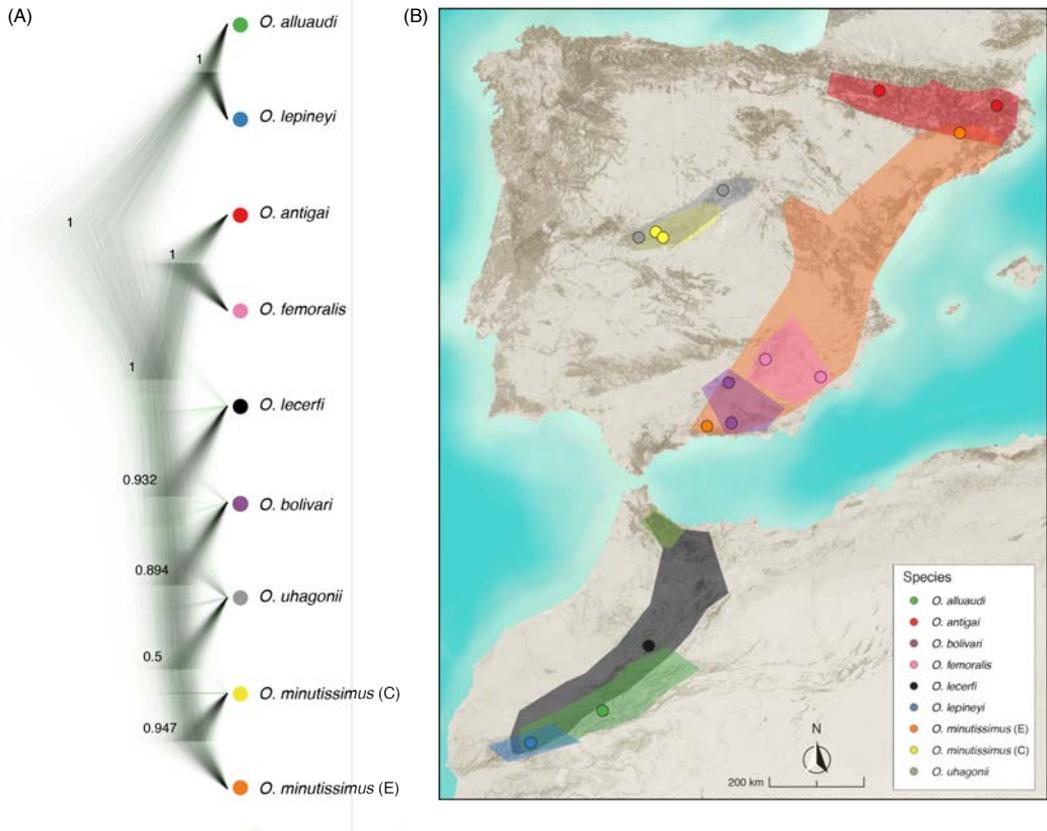
Here we focus on the Ibero-Maghrebian subgenus *Dreuxius* Defaut, 1988 (genus *Omocestus* Bolívar, 1878), a complex of montane grasshoppers currently comprised by eight species distributed in the Iberian Peninsula (5 species) and northwestern Africa (3 species) (Cigliano, Braun, Eades, & Otte, 2019; Tonzo, Papadopoulou, & Ortego, 2019). Most taxa present allopatric distributions and form isolated populations at high elevations in different mountain systems (Tonzo, Papadopoulou, & Ortego, 2019; Tonzo, Papadopoulou, & Ortego, in press; Cigliano, Braun, Eades, & Otte, 2019; Figure 1). The only exceptions are the Iberian *O. minutissimus* (Brullé 1832) and the Maghrebian *O. lecerfi* Chopard 1937, which present wider elevational ranges and spatial distributions partially overlapping with the rest of Iberian and northwestern African species of the complex, respectively, and with which they often form sympatric populations (Cigliano, Braun, Eades, & Otte, 2019; Clemente, García, & Presa, 1990; Tonzo, Papadopoulou, & Ortego, in press). All taxa within the complex are predominantly graminivorous and their distributions are tightly linked to open habitats of cushion and thorny shrub formations (e.g., *Erinacea* sp., *Festuca* sp., *Juniperus* sp., *Thymus* sp.) that they use as refuge (Clemente et al., 1990; Gangwere & Morales Agacino, 1970). Species within this subgenus, particularly females, are markedly brachypterous, which is expected to extraordinarily limit their dispersal capacity, reduce gene flow at short spatial scales and, ultimately, might have contributed to genetic divergence and allopatric speciation (Waters, Emerson, Arribas, & McCulloch, 2020; e.g., Huang, Hill, Ortego, & Knowles, 2020). For these reasons, this transcontinental species complex offers an ideal case study to test alternative biogeographic scenarios underlying the high rates of endemism of the region and gain insights into the proximate processes underlying species formation and patterns of phenotypic variation.

In this study, we integrate high throughput sequence data, geometric morphometrics, and novel phylogenetic inference methods that explicitly account for gene flow to unravel the evolutionary relationships and the timing and mode of diversification in the studied species complex. Specifically, we first generated genomic data for all species within the subgenus *Dreuxius* using a restriction-site-associated DNA sequencing approach (ddRADseq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) and inferred their phylogenetic relationships applying two alternative coalescent-based methods (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012; Yang, 2015) and a maximum pseudolikelihood approach accounting for post-divergence gene flow (Solís-Lemus & Ané, 2016). Second, we estimated species divergence times under the multispecies coalescent (MSC) model (Rannala & Yang, 2003; Yang, 2002) and a new implementation of the MSC model with introgression (MSCi) (Flouri, Jiao, Rannala, & Yang, 2019), and evaluated the potential impact of historical gene flow on demographic parameter estimation and the inferred biogeographic history. Finally, we employed a geometric morphometric approximation (Adams & Otárola-Castillo, 2013) to characterize phenotypic variation at traits of taxonomic relevance and/or putatively linked to reproductive isolation and evaluated whether such variation was shaped by a shared evolutionary history (i.e., Brownian motion under genetic drift) or departed from expectations given the phylogenetic tree, which might be indicative of selective processes acting at different stages of speciation (Gray & McKinnon, 2007; Safran, Scordato, Symes, Rodríguez, & Mendelson, 2013).

## 2 | Materials and Methods

### 2.1 | Species sampling

Between 2011 and 2017, we collected specimens representing all species of the subgenus *Dreuxius* (genus *Omocestus*) (Cigliano, Braun, Eades, & Otte, 2020; Table 1; Figure 1). We considered as independent lineages allopatric populations of *O. minutissimus* from central and eastern Iberia (hereafter, *O. minutissimus* C and *O. minutissimus* E, respectively), as they form distinctive genotypic and phenotypic clusters according to preliminary analyses (Cáliz, 2015; Tonzo, Papadopoulou, & Ortego, in press). Two of the taxa within the complex (*O. navasi* and *O. antigai*) have been recently synonymized on the basis of detailed genomic and phenotypic species delimitation



**Figure 1** (A) Bayesian phylogenetic tree reconstructed with SNAPP. Posterior probabilities of clade support are indicated. (B) Approximate geographical distribution of the different species/lineages of the subgenus *Dreuxius* (shapes) and populations (dots) included in the analyses.

analyses and, thus, they were considered as a single species (*O. antigai*; Cigliano, Braun, Eades, & Otte, 2020; Tonzo, Papadopoulou, & Ortego, 2019). Whenever possible, we collected and analyzed two populations representative of the distribution range of each species/lineage (Table 1; Figure 1). We stored specimens in 2 ml vials with 96% ethanol and preserved them at  $-20^{\circ}\text{C}$  until needed for geometric morphometric and genomic analyses.

## 2.2 I Genomic library preparation and processing

We obtained genomic data for a total of 36 specimens representative of one or two populations per species/lineage (4 individuals per species/lineage in all cases; Table 1). Details on the preparation of ddRADseq libraries (Peterson et al., 2012) are presented in Methods S1. Raw sequences were demultiplexed and pre-processed using STACKS version

**Table 1** Locality and geographical location (latitude, longitude and elevation) for each species and sampled population.

Species	Locality	Latitude	Longitude	Elevation (m)
<i>O. alluaudi</i>	Tizi n' Tirghist (Morocco)	31.7408	-6.3251	2,538
<i>O. antigai</i>	Setcases (Spain)	42.4276	2.2663	2,145
<i>O. antigai</i>	Borau (Spain)	42.6739	-0.5793	1,357
<i>O. bolivari</i>	Sierra de Mágina (Spain)	37.7406	-3.4466	1,900
<i>O. bolivari</i>	Sierra Nevada (Spain)	37.0964	-3.3891	2,446
<i>O. femoralis</i>	Sierra de Espuña (Spain)	37.8652	-1.5712	1,514
<i>O. femoralis</i>	Poyotello (Spain)	38.1195	-2.6165	1,600
<i>O. lecerfi</i>	Col du Zad (Morocco)	32.4531	-5.2413	2,100
<i>O. lepineyi</i>	Jebel Oukaimeden (Morocco)	31.1873	-7.8590	2,870
<i>O. minutissimus</i> (C)	Puerto del Pico (Spain)	40.3458	-5.0143	1,340
<i>O. minutissimus</i> (C)	Puerto de Serranillos (Spain)	40.3067	-4.9467	1,600
<i>O. minutissimus</i> (E)	Sierra de Montsec (Spain)	42.0473	0.7431	1,520
<i>O. minutissimus</i> (E)	Sierra Tejada (Spain)	36.9045	-4.0351	2,040
<i>O. uhagonii</i>	Puerto de Navafria (Spain)	40.9846	-3.8215	1,893
<i>O. uhagonii</i>	Puerto de Peña Negra (Spain)	40.4216	-5.3105	1,885

1.35 (Catchen, Amores, Hohenlohe, Cresko & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores & Cresko, 2013) and assembled using PYRAD version 3.0.66 (Eaton, 2014). Methods S2 provides all details on sequence assembling and data filtering.

## 2.3 | Phylogenomic inference

We estimated species trees using two coalescent-based methods, SNAPP version 1.3 (Bryant et al., 2012) as implemented in BEAST2 version 2.4.3 (Bouckaert et al., 2014) and BPP version 4.2 (Flouri et al., 2018). SNAPP analyses are computationally highly demanding and, for this reason, we only selected two individuals per species (those with the highest number of retained reads; Figure S1), one for each sampled population when two populations were available (i.e., 18 individuals in total). The resulting dataset retained 723 unlinked polymorphic sites. We ran SNAPP analyses for 1,000,000 Markov chain Monte Carlo (MCMC) generations, sampling every 1,000 steps and using as gamma prior distributions for alpha and beta 2 and 2,000 values. The forward (u) and reverse (v) mutation rates were set to be calculated by SNAPP and we left the remaining parameters

at default values. We conducted two independent runs and evaluated convergence with TRACER version 1.6. We removed 10% of trees as burn-in and merged tree and log files from the different runs using LOGCOMBINER version 2.4.1. We used TREEANNOTATOR version 1.8.3 to obtain maximum credibility trees and DENSITREE version 2.2.1 (Bouckaert, 2010) to visualize the posterior distribution of trees.

Complementarily, we ran BPP version 4.2 under module A01 to estimate the species tree (Flouri et al., 2018; Yang, 2015). BPP program is a full-likelihood implementation of the MSC model and uses a reversible-jump Markov chain Monte Carlo (rjMCMC) method to collapse or split nodes in the guide species tree according to node posterior probabilities. We created BPP input files from the ‘loci’ output file from PYRAD using the R scripts `bpp_convert_Ama_sp.r` written by J-P. Huang and available at [https://github.com/airbugs/Dynastes\\_delimitation](https://github.com/airbugs/Dynastes_delimitation) (Huang, 2018). We discarded loci that were not represented in at least one individual per species (i.e., loci with missing species were removed; e.g., Huang et al., 2020). The final dataset retained 333 loci. We considered as prior settings:  $\theta = G(3, 0.002)$  and  $\tau = G(3, 0.004)$ , where  $\theta$  and  $\tau$  refer to the ancestral population sizes and divergence times, respectively. We ran two replicates and used an automatic adjustment of the finetune parameters, allowing swapping rates to range between 0.30 and 0.70 (Yang, 2015). We ran each analysis for 100,000 generations, sampling every 2 generations (10,000 samples), after a burn-in of 50,000 generations. We evaluated convergence of replicates using TRACER version 1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018).

## 2.4 | Phylonetwork reconstruction

Phylogenetic reconstruction without considering the potential occurrence of post-divergence gene flow (i.e., introgressive hybridization) can have severe impacts on the obtained inferences (Burbrink & Gehara, 2018; Olave & Meyer, 2020; Solís-Lemus & Ané, 2016). Although the two phylogenomic inference methods employed (SNAPP and BPP) yielded the same most supported topology (see Results section), unsupported nodes led us to investigate the presence and impact of multiple branches connections using the JULIA package PHYLONETWORKS (Solís-Lemus et al., 2017). This method uses a maximum pseudolikelihood estimator applied to quartet concordance factors (CF) of 4-taxon trees under the coalescent model, incorporating incomplete lineage sorting and reticulation events (Solís-Lemus et al., 2017). The observed CF from the estimated gene trees is then used to estimate a semi-directed species network with estimated reticulation events and

$\gamma$ -values indicating the proportion of ancestral contribution to the hybrid lineage genome.

To estimate individual gene trees for each locus, we followed MAGNET version 0.1.5 pipeline (J. C. Bagley, <http://github.com/justincbagley/MAGNET>). We ran MAGNET pipeline using as input file the aligned DNA sequences from the PYRAD output file '.gphocs'. Specifically, MAGNET first splits each locus contained in the '.gphocs' file into separated phylip-formatted alignment files, and sets up and runs RAXML (Stamatakis 2014) to infer a maximum-likelihood (ML) gene tree for each locus. Prior to obtain the gene trees, we applied TRIMAL version 1.2 (Capella-Gutierrez, Silla-Martinez, & Gabaldon, 2009) to our phylip dataset in order to filter out loci with a high average identity (>0.99 %) across the multisequence alignment and retain only those that are most informative (Bernardes, Dávila, Costa, & Zaverucha, 2007). Then, we used PHYLONETWORKS to read all RAXML gene-trees retained (20,637 trees) and calculate CFs, with all individuals per clade mapped as alleles to species. We used the BPP/SNAPP tree as the starting topology, and tested values for  $h$  (number of reticulations) from 0 to 5, assessing maximum support using a slope heuristic for the increase in likelihood plotted against  $h$  (Solís-Lemus & Ané, 2016). We ran 50 independent runs per  $h$ -value to ensure convergence on a global optimum.

## 2.5 | Divergence time estimation

We ran BPP under module A00 to obtain the posterior distribution of species divergence times ( $\tau$ s) under the multispecies coalescent (MSC) model (Rannala & Yang, 2003; Yang, 2002). A recent implementation of A00 analysis on BPP version 4 allows estimating parameters under the MSC considering past introgression events ( $\varphi$ s) (multispecies coalescent with introgression, MSCi; Burgess & Yang, 2008; Flouri et al., 2019; Rannala & Yang, 2003). To evaluate the impact of introgression events on divergence time estimation, we conducted A00 analyses under both the MSC and MSCi models using as fixed topology i) the one most supported by SNAPP and A01 BPP analyses (MSC model) and ii) the species tree from the most supported phylogenetic network recovered using PHYLONETWORKS (MSCi model). For each analysis, we executed two runs and assigned values for the inverse-gamma priors  $\theta \sim \text{IG}(3, 0.004)$  for all  $\theta$ s and  $\tau \sim \text{IG}(3, 0.004)$  for the age  $\tau_0$  of the root as suggested in Flouri et al., (2019) when no information is available about prior parameters. A total of 50,000 iterations (sample interval of 5) with a burn-in of 10,000 was implemented for each run and convergence was evaluated across replicates

using TRACER (Rambaut et al., 2018). Divergence times were calculated according to the equation  $t = \tau/2\mu$  (e.g., Huang et al., 2020), where  $\tau$  is the divergence in substitutions per site estimated by BPP,  $\mu$  is the per site mutation rate per generation, and  $t$  is the absolute divergence time in years. We assumed a genomic mutation rate of  $2.8 \times 10^{-9}$  per site per generation (Keightley, Ness, Halligan, & Haddrill, 2014) and a one-year generation time (Clemente et al., 1990).

## 2.6 | Geometric morphometric analyses

To characterize phenotypic variation, we chose traits that have been used to delineate taxonomic units in the complex (pronotum; Clemente, García, & Presa, 1991) and associated to courtship behavior (forewing; e.g., Nattier et al., 2011) and reproduction (male genitalia; e.g., Huang et al., 2020) in Orthoptera. We selected 10 individuals from each studied population (5 males and 5 females for each of the two populations per species/lineage, when available) to analyze forewing and pronotum variation and two individuals per population to extract and characterize male genitalia (penis lateral valve shape). To prepare male genitalia, we made a longitudinal cut and peeled back the apex of the abdomen to remove the exoskeleton. Abdominal contents were removed with fine forceps and placed in a Petri dish with 20% KOH for ~2 hours at room temperature to digest connective tissues. After that time, the sclerotized structure of the genitalia became apparent in the materials. We used landmark-based geometric morphometric methods (GMM) to characterize phenotypic variation in the selected traits. We captured digital images of dorsal views of pronota and forewings and of lateral views of male internal genitalia with a Leica MZ16 A stereomicroscope fitted with a DFC 450 camera using the Leica Application Software (LAS) version 3.8 (Leica Microsystems Ltd, Switzerland). We used fixed landmarks to characterize pronotum (9 landmarks) and forewing (11 landmarks) shape and a combination of fixed landmarks (3 landmarks) and semi-landmarks (35 landmarks) to capture the shape of male genitalia. Landmarks were mapped on the images using TPSDIG version 2.2 (Rohlf, 2015) and analyzed as implemented in the R version 3.3.2 (R Core Team, 2018) package GEOMORPH (Adams & Otárola-Castillo, 2013). Semi-landmarks were resampled to be equidistant along their curves and “slid” via minimizing bending energy (Bookstein, 1992; Bookstein et al., 1999; Gunz, Mitteroecker, & Bookstein, 2005). We obtained shape variations for each sex and trait through generalized Procrustes analyses (GPA) (Rohlf, 1999; Rohlf & Slice, 1990) with the package GEOMORPH. Specifically,

we performed GPA to standardize the size and remove the effects of location and rotation of the relative positions of landmarks among specimens using the function *gpagen*. This superimposition method minimizes the sum-of-squared distances between landmarks across samples (Rohlf & Slice, 1990). We used principal components analysis (PCA) of the Procrustes coordinates for each dataset to extract the most explanatory axes of shape variation. To test for shape differences among species, we performed a Procrustes ANOVA using distributions generated from a resampling procedure based on 1,000 iterations in the R package GEOMORPH using the function *procD.lm* (Adams & Collyer, 2018). Significance values (*p*-values) between each pair of species were determined for each sex and trait using the *pairwise* function. To visualize shape differences, we represented the first two principal component axes (the most explicative) in a convex hull for each species and sex, using *ddplyr* function in the R package PLYR (Wickham, François, Henry, & Müller, 2019).

We quantified the phylogenetic signal (i.e., how morphologically similar closely related species are to one another) for each trait using Blomberg et al.'s (2003) *K* under the function *physignal* in GEOMORPH. We used the tree topology most supported by the phylogenetic inference analyses detailed above and performed 1,000 permutations of shape data among the tips of the phylogeny to evaluate statistical significance. A *K*-value of 1 reflects perfect accord with expected patterns of shape variation under Brownian motion, values greater than 1 reflect phylogenetic under-dispersion of shape variation (i.e., close relatives are more similar than expected under Brownian motion), and values less than 1 indicate phylogenetic over-dispersion of shape variation (i.e., close relatives are less similar than expected under Brownian motion).

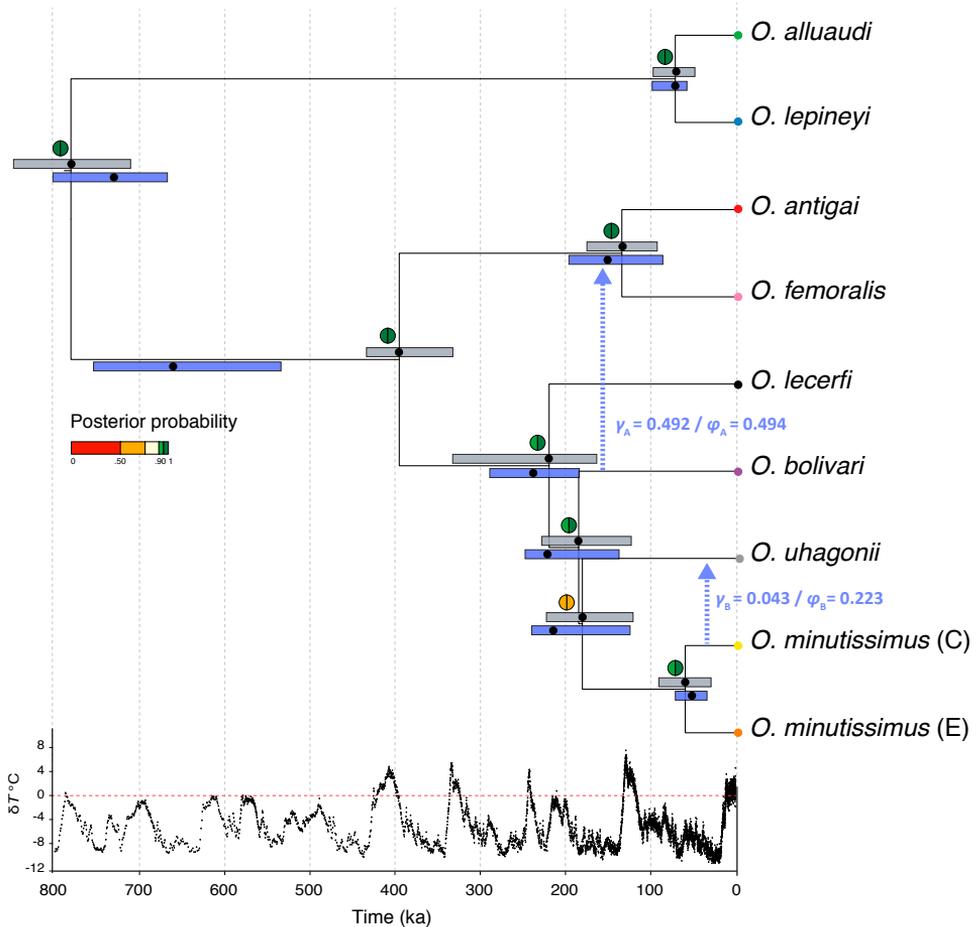
## 3 | Results

### 3.1 | Genomic data

Illumina sequencing returned an average of  $2.89 \times 10^6$  reads per sample. After quality control, an average of  $2.13 \times 10^6$  reads was retained per sample (Figure S1). The genomic datasets obtained with PYRAD (*minCov* = 25%) for the subsets of 18 and 36 individuals retained a total of 50,192 and 21,438 variable loci, respectively.

### 3.2 I Phylogenomic inference

Species trees reconstructed by SNAPP and BPP yielded the same topology and the two analyses only differed in the degree of support for some clades (Figures 1 and 2). These analyses recovered three main monophyletic groups: a Maghrebian clade (*O. alluaudi* and *O. lepineyi*), an Iberian clade (*O. antigai* and *O. femoralis*), and an Ibero-Maghrebian clade (*O. lecerfi*, *O. bolivari*, *O. uhagonii* and *O. minutissimus*). The Maghrebian species



**Figure 2** Species tree including estimates of divergence time and inferred introgression events. The species tree was reconstructed with SNAPP and BPP (option A01) and posterior probabilities of node support are indicated for each analysis in colored semi circles (left: SNAPP; right: BPP). Dots (median) and bars (95% highest posterior density intervals) indicate divergence times estimated by BPP (option A00), colored in grey for standard analysis not considering post-divergence gene flow (MSC model) and blue for analyses accounting for introgression events (MSCi model) inferred using PHYLONETWORKS. Blue arrows indicate inferred introgression events with their corresponding inheritance values ( $\gamma$ ) estimated by PHYLONETWORKS and introgression probabilities ( $\phi$ ) estimated by BPP (not time-scaled). Bottom panel shows temperature anomaly ( $\delta T$  °C) in the Late Quaternary as estimated from the EPICA (European Project for Ice Coring in Antarctica) Dome C ice core (Jouzel et al., 2007).

*O. alluaudi* and *O. lepineyi* constituted the most basal and the Iberian and Ibero-Maghrebian clades shared a sister relationship (Figures 1 and 2). SNAPP analyses showed low clade support (i.e., posterior probabilities values  $< 0.95$ ) for internal relationships within the Ibero-Maghrebian clade. Accordingly, the most frequently recovered topology with SNAPP (37.84%) differed from the alternative less supported topologies on the sisterhood relationships among species within this group (Figures 1 and 2). In contrast to SNAPP, posterior probabilities in BPP were consistently high for all clades, except for the split between *O. uhagonii* and the two *O. minutissimus* lineages (Figure 2).

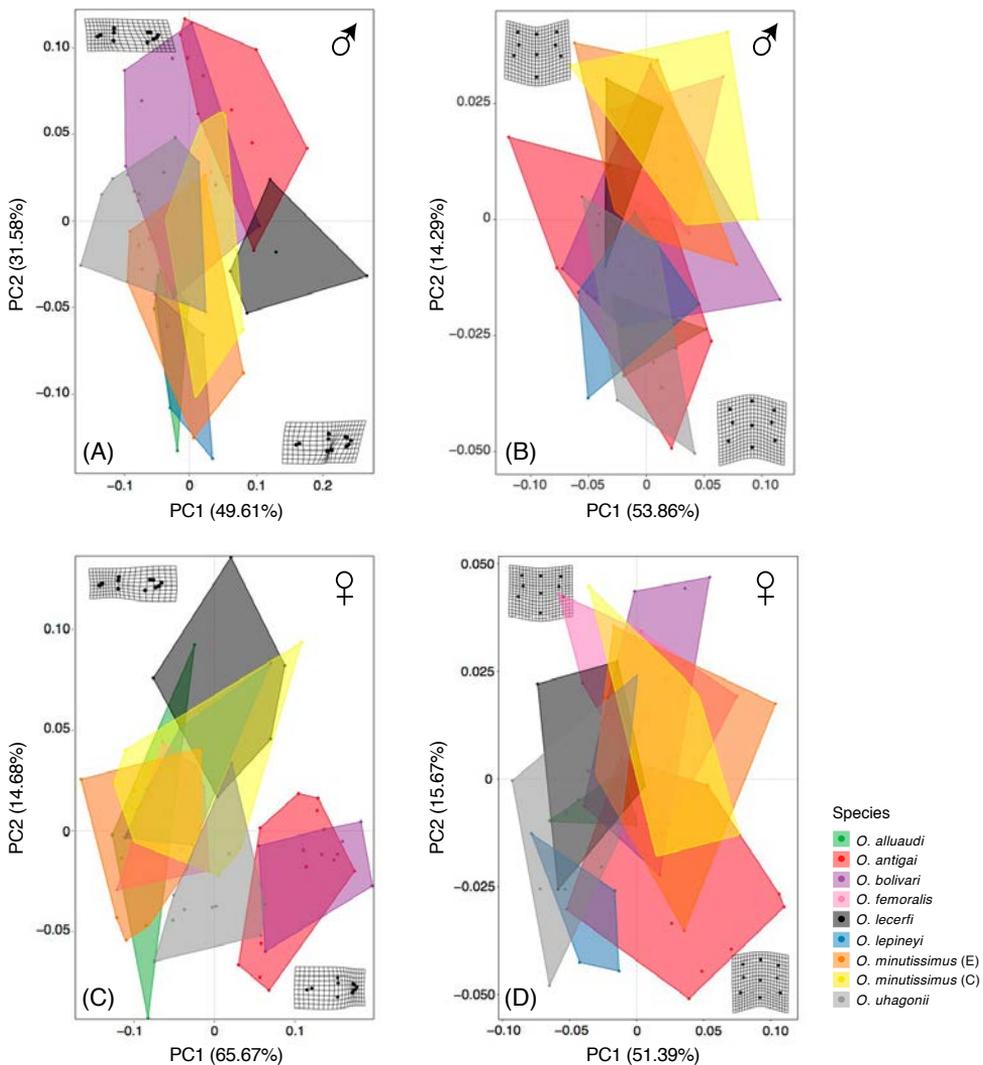
### 3.3 | Phylonetwork reconstruction

PHYLONETWORK analyses revealed that all models involving reticulation events ( $h > 0$ ) fit our data better than models considering strict bifurcating trees ( $h = 0$ ) (Figure S2). The best phylogenetic network inferred by PHYLONETWORKS identified two introgression events ( $h_{max} = 2$ , negative pseudolikelihood = -6.20) and a backbone tree in concordance with the topologies recovered by SNAPP and BPP (Figures 1 and 2). The optimal network supported introgression from *O. minutissimus*-C to the sympatric *O. uhagonii* ( $\gamma_A = 0.043$ ) and from *O. bolivari* to the most recent common ancestor (MRCA) of *O. femoralis* and *O. antigai* ( $\gamma_B = 0.492$ ) (Figure 2).

### 3.4 | Divergence time estimation

Divergence times estimated by BPP both considering (MSCi model) and not considering (MSC model) post-divergence gene flow are summarized in Figure 2. Both analyses supported that the initial split of the Maghrebian clade from the rest of the species took place during the Middle Pleistocene ( $\sim 675$  to 850 ka). Estimates of divergence time between the Iberian and the Ibero-Maghrebian clades was the most important discrepancy between the results yielded by BPP analyses under the MSC and MSCi models. BPP analyses not considering post-divergence gene flow estimated that these two clades split around 340-445 ka. However, divergence times obtained under the MSCi model yielded older estimates, around 540-770 ka. The 95% highest posterior density (HPD) intervals obtained under the two models largely overlapped for the rest of the nodes. Analyses showed that all contemporary species originated in the last  $\sim 200$  ka, during the end of the Middle and the beginning of the Late Pleistocene (Figure 2). Introgression from

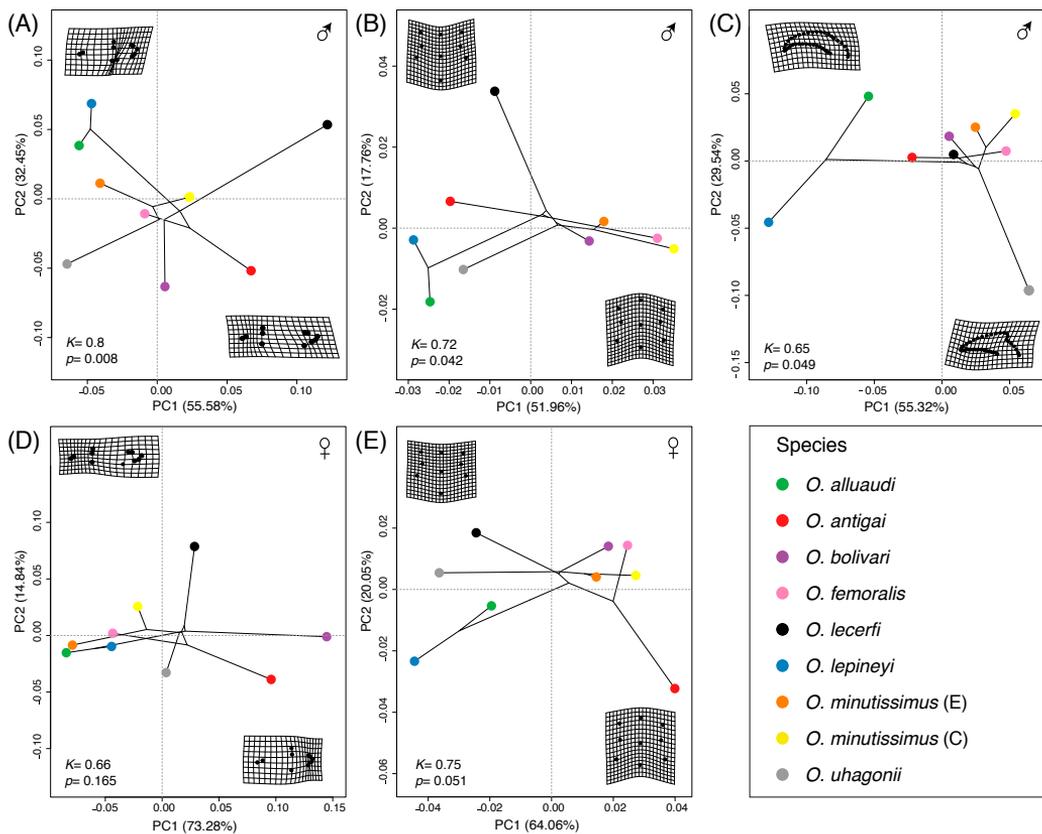
*O. minutissimus* C to *O. uhagonii* took place around 49 ka, whereas introgression from *O. bolivari* to the ancestor of *O. antigai* and *O. femoralis* dated back to 205 ka. The introgression probability ( $\phi$ ; Flouris et al., 2020) estimated by BPP for the introgression event from *O. bolivari* to the ancestor of *O. antigai* and *O. femoralis* was virtually identical ( $\phi_A = 0.494$ ) to the inheritance parameter ( $\gamma$ ; Solís-Lemus & Ane, 2016) estimated by PHYLONETWORK ( $\gamma = 0.492$ ; Figure 2). However, the introgression probability from *O. minutissimus* C to *O. uhagonii* estimated by BPP was much higher ( $\phi_A = 0.223$ ) than the analogous inheritance parameter ( $\gamma = 0.043$ ) yielded by PHYLONETWORK analyses (Figure 2).



**Figure 3** Principal component analyses for (A, C) forewing and (B, D) pronotum size-corrected shape variation in (A, B) males and (C, D) females. Colored convex hull polygons show species/lineage variation and warp grids represent extreme shape variation for the first two principal components.

### 3.5 I Analyses of phenotypic variation

A high proportion of pronotum and forewing shape variation was explained (>65%) by the first two principal components (Figure 3). All analyzed traits significantly differed among species/lineages in both sexes (Table S1). In males, two extreme forms could be distinguished in forewing shape variation: a spindle-like shape (*O. lecerfi*) and an elongated trapezoid shape (*O. uhagonii*) (Figure 3A). In females, species/lineages could be differentiated by rounded (*O. antigai* and *O. bolivari*), sharpened (*O. alluaudi*, *O. lepinyei*, *O. femoralis* and *O. minutissimus* E) and intermediate (*O. uhagonii*, *O. lecerfi*, *O. minutissimus* C) forewing shapes (Figure 3C). Forewing shape in the two sexes was significantly different in most pair-wise species/lineage comparisons (Table S2). Although dorsal pronotal shape



**Figure 4** Phylomorphospaces showing the first two principal components (PC1 and PC2) from a PCA summarizing size-corrected shape variation for (A) forewing, (B) pronotum and (C) genitalia in males and (D) forewing and (E) pronotum in females. Colored dots indicate the different species and warp grids represent extreme shape variations for the first two principal components.

variation in both males and females showed highly significant differences among species/lineages (Table S1), this trait tended to present a higher overlap than forewing shape variation (Figure 3B, D). Accordingly, a fewer number of pair-wise species/lineage comparisons were statistically significant (Table S3). The lower number of samples analyzed for male genitalia made not possible a visual representation of shape variation for this trait. However, results from procrustes ANOVA showed significant differences among species/lineages that were mostly driven by differences between *O. lepineyi* and *O. alluaudi* (hooked shape) and the rest of the species (straight shape) (Table S4).

In males, forewing, pronotum and genitalia shapes exhibited significant phylogenetic signals and  $K$  values  $< 1$  indicated that closely related taxa are less similar in these traits than expected under Brownian motion (Figure 4). The degree of phylogenetic signal varied across male traits, being weaker for male genitalia (Figure 4). In females, forewing and pronotum shapes did not show a significant phylogenetic signal, albeit pronotum shape was marginally non-significant (Figure 4).

## 4 | Discussion

By reconstructing lineage and phenotypic diversification in a complex of montane grasshoppers, our study contributed to shed light on the ecological and evolutionary processes underlying the high rates of local endemism of the Ibero-Maghrebian biodiversity hotspot (Avice & Wollenberg, 1997; Hewitt, 1996). The combination of genomic data and a comprehensive suite of coalescent-based phylogenetic analyses provided strong support for a recent radiation ( $< 1$  Ma) of the subgenus *Dreuxius*, indicating that periods of isolation and secondary contact linked to Pleistocene glacial cycles likely contributed to both allopatric speciation and post divergence gene flow. Geometric morphometric analyses for traits of taxonomic relevance and putatively involved in different components of reproductive isolation (sexual selection, copulation, etc.) supported the phenotypic distinctiveness of most sister taxa within the complex. Moreover, some of the studied traits presented a significantly lower phylogenetic signal than expected under a Brownian motion model of evolution, suggesting that phenotypic variation might have been in part shaped by natural or sexual selection acting at different stages of speciation (Kelly, 2014; Servedio & Boughman, 2017). This research exemplifies how the integration of multiple lines of evidence can help to reconstruct complex histories

of reticulated evolution linked to Late Quaternary climatic changes and highlights the importance of implementing new methodological approaches to deal with post-divergence gene flow, a necessary step toward getting unbiased estimates of key demographic parameters and drawing a more realistic evolutionary portrait of Pleistocene radiations in which incomplete lineage sorting often co-occur with introgressive hybridization (Clemente et al., 1990; Gangwere & Morales Agacino, 1970; Nevado et al., 2018; Wen, Yu, Hahn, & Nakhleh, 2016).

#### 4.1 I Evolutionary biogeography of the species complex

Our phylogenetic reconstructions and estimates of divergence time supported that the diversification of the subgenus *Dreuxius* took place over the last 800 ka, with direct ancestors of extant species tracing back their origins to end of the Middle Pleistocene (<250 Ka) (Figure 2), and most splitting events occurring in a short time span (~200 ka). The earliest split within *Dreuxius* (ca. 800 ka) separated the lineage including the Maghrebian *O. alluaudi* and *O. lepineyi* from the most speciose clade including all Iberian taxa plus the northwestern African *O. lecerfi*. The last clade subsequently split (ca. 400 ka) into two clades, one formed by the Pyrenean *O. antigai* and the Baetican *O. femoralis* and another comprising the rest of Iberian species and *O. lecerfi*. Our genomic data support *O. antigai* and *O. femoralis* as sister taxa and a close relationship between *O. bolivari*, *O. minutissimus* and *O. uhagonii*, which agrees with previous descriptive assessments of species relationships based on morphological and behavioral comparisons (Clemente et al., 1991; Gangwere & Morales Agacino, 1970). Our analyses also supported that the divergence between the two allopatric lineages of *O. minutissimus* is of the same order of magnitude (ca. 60 ka) than that estimated between the sister taxa *O. alluaudi* and *O. lepineyi* (Figure 2). The genotypic and phenotypic distinctiveness of these two lineages (Cáliz, 2015; Tonzo, Papadopoulou, & Ortego, in press) call upon a taxonomic re-assessment of this monotypic taxon that was formerly composed by two distinct taxa: *O. burri* Uvarov, 1936 widely distributed in eastern Iberia and *O. minutissimus* (Brullé, 1832) restricted to the Central System (Clemente et al., 1990 and references therein).

The Pleistocene origin of all clades and lineages within the studied species complex point to the predominant role of Quaternary climatic oscillations in the transcontinental diversification of Ibero-Maghrebian biotas. With the exception of *O. minutissimus*, which is distributed from sea level to alpine areas above the tree line, the rest of taxa within the

complex are montane species restricted to high elevations (>1,300 m) from different mountain ranges. Thus, Late Pleistocene climatic oscillations, when most speciation events within the complex took place, are expected to have contributed to create multiple opportunities for both divergence and post-divergence gene flow through elevational and latitudinal range-shifts (Hewitt, 2000; Knowles, 2000). Despite the genetic and phenotypic distinctiveness of the species within the complex and their current distribution in distant mountain ranges, the habitats occupied show strong similarities across all taxa (Clemente et al., 1990; Clemente et al., 1991; Ragge, 1986; Tonzo, Papadopoulou, & Ortego, in press). This points to allopatric speciation, rather than ecological divergence, as the predominant mechanism of species diversification (Hewitt, 2000; Hewitt, 2004; Mayer, Berger, Gottsberger, & Schulze, 2010; Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998). Topographically complex regions such as Iberia and northwestern Africa offer an ideal biogeographic setting for allopatric speciation, as isolation in valleys during glacial periods (i.e., glacial refugia; Knowles, 2001; Wallis et al., 2016) and confinement in sky-islands during interglacials (i.e., interglacial refugia; Bennett & Provan, 2008; Stewart, Lister, Barnes, & Dalén, 2010) are expected to lead to extended periods of isolation and divergence through genetic drift and/or natural selection under contrasting selective regimes (Djamali et al., 2012; Hewitt, 1996). Furthermore, range shifts might have contributed in some cases to complete the speciation process through the evolution of reproductive isolation in secondary contact zones (i.e., reinforcement; Butlin, 1989, 1998; Hewitt, 2008; Tonzo, Papadopoulou, & Ortego, in press).

Phylogenetic analyses did not recover Maghrebian taxa as a monophyletic clade, supporting two trans-continental colonization events through the Strait of Gibraltar or adjacent areas. Glacial periods reduced the Mediterranean Sea level about 125 m and shortened the distance between northwestern Africa and southern Iberia to less than 5 kilometers, which might have led to the emergence of small islands and shoals and facilitated the exchange of biotas between the two continents during the coldest stages of the Pleistocene (Agustí et al., 2006; Collina-Girard, 2001; Cosson et al., 2005). These results add to the accumulating empirical evidence supporting the migration of numerous organisms across the two continents, either seeking for glacial refugia in North Africa or following post glacial colonization routes to Europe (Graciá et al., 2013; Husemann et al., 2014; Taberlet et al., 1998; Teacher, Garner, & Nichols, 2009).

## 4.2 I A reticulated evolutionary history

Interspecific gene flow and ILS are ubiquitous phenomena in recent evolutionary radiations and, thus, require to be evaluated when inferring phylogenetic relationships and demographic parameters in species complexes of Pleistocene origin (Solís-Lemus & Ané, 2016; Wen, Yu, Zhu, & Nakhleh, 2018; Yu & Nakhleh 2015). Although the monophyly of the three main clades of the subgenus *Dreuxius* was consistently well-supported, internal nodes of the most speciose clade showed weak support in both BPP and SNAPP analyses (Figure 2). Phylogenetic network analyses point to interspecific gene flow, rather than ILS, as the main cause of gene tree conflict (Figure 2). Specifically, we found two events of introgression involving lineages at different stages of the speciation continuum: from *O. bolivari* to the most recent common ancestor of *O. antigai* and *O. femoralis* (ca. 205 ka) and from the lineage of *O. minutissimus* distributed in the Central System to its sympatric counterpart *O. uhagonii* (ca. 49 ka) (Figure 2). As expected, the two introgression events involved taxa from the same continental landmass (i.e., Iberian Peninsula). *Omocestus bolivari* and *O. femoralis* currently present adjacent but non-overlapping distributions in the sky island archipelago of the Baetic System (Figure 1). However, a previous comparative population genetics study on the two species revealed that their distributions likely overlapped during glacial periods, when the two species experienced considerable demographic expansions (V. Tonzo and J. Ortego, Chapter IV). This is expected to have led to secondary contact and might explain the detected signatures of historical gene flow from *O. bolivari* to the common ancestor of *O. femoralis* and *O. antigai*. The very low support for the split between *O. minutissimus* and *O. uhagonii* was explained by historical hybridization between the two taxa in the Central System, where the evolution of reproductive isolation via reinforcement or other mechanisms has been hypothesized to prevent gene flow among contemporary sympatric populations of the two species (Tonzo, Papadopoulou, & Ortego, in press).

Although there is an increasing interest on implementing phylogenomic network approaches to empirical data (Eckert & Carstens, 2008; Pickrell & Pritchard, 2012; Solís-Lemus et al., 2017; Wen et al., 2018; Yu & Nakhleh, 2015), the impact of interspecific gene flow on inferred divergence times has been rarely evaluated (Flouri et al., 2019). We assessed the impact of introgression on the estimated timing of species split and found that, as expected, ignoring interspecific gene flow result in an underestimation of divergence times in some nodes. Specifically, the timing of divergence between *O. antigai*-

*O. femoralis* and the rest of the species was estimated to be ca. 200 ka older when analyses accounted for introgression, whereas historical gene flow between sympatric populations of the more recently diverged *O. minutissimus* and *O. uhagonii* had a little impact on our inferences.

### 4.3 I Phenotypic variation

We found that all the studied phenotypic traits differed among lineages, with most species/lineage pairs presenting significant differences in at least one of them (Tables S1-4). In both sexes, forewing shape tended to show stronger differences among species than pronotum and male genitalia (Figure 3 and Tables S1-4). Forewings are involved in courtship acoustic behavior in grasshoppers (Ronacher, 2019; Vedenina & Mugue, 2011; Von Helversen, Balakrishnan, & Von Helversen, 2004), a character directly implicated in mate attraction and subjected to sexual selection (Oh & Shaw, 2013; Outomuro et al., 2016). Traits under sexual selection can evolve rapidly, accelerating speciation when other forces as ecological adaptations are not so evident or absent (Anderson, 1994; Mendelson & Shaw, 2005; Rundell & Price, 2009). Accordingly, species within the *Dreuxius* species complex show very similar habitat requirements but present distinctive songs (Clemente, García, Arnaldos, Romera, & Presa, 1991, 1999; Ragge, 1986; Reynolds, 1987), which suggests that sexual selection might have played an important role in the completion of the speciation process (e.g., Bridle & Butlin, 2002; Bridle, Saldamando, Koning, & Butlin, 2006) and prevented interbreeding among contemporary sympatric populations in secondary contact zones (Tonzo, Papadopoulou, & Ortego, in press). In the case of male genitalia, we found that interspecific variation was mostly determined by differences between the two earliest diverged clades (Table S4). Although differences among species in this trait must be interpreted with extreme caution due to small sample sizes, the fact that some currently sympatric lineages (e.g., *O. minutissimus* and the rest of Iberian species) share similar male genitalia suggests that reproductive isolation might have been driven by other phenotypic or behavioral traits (e.g., mate selection). Remarkably, species/lineages involved in historical introgression presented significant phenotypic differences for one or more of the studied traits, suggesting that historical hybridization has not led to phenotypic assimilation (Huang, 2016) or that, on the contrary, secondary contact might have contributed to phenotypic divergence through some form of character displacement (Pfennig & Pfennig, 2009). Finally, phylomorphospace analyses and the *K* statistic of

Blomberg et al., (2003) ( $K < 1$  in all cases) indicated that species are less similar at some of the studied morphological traits than expected under a Brownian motion model of evolution (Figure 4). Even when phylogenetic signal alone is not a direct way of elucidating the evolutionary processes responsible for phenotypic diversification, these results also suggests that natural and/or sexual selection might have modulated phenotypic diversification in the complex (Blomberg et al., 2003; Pennell & Harmon, 2013).

#### 4.4 I Conclusions

Our study exemplifies the importance of integrating different sources of information to reconstruct complex biogeographic histories and understand the processes underlying the high rates of local endemism of biodiversity hotspots. Although the retrieved topology, estimates of divergence time (i.e., Pleistocene) and biogeographic inferences did not qualitative change considering or not inter-specific gene flow, past hybridization events are an important component of speciation that must be resolved to shed light on the evolutionary pathways of recent species complexes. Collectively, our analyses demonstrate a very recent origin of the studied radiation ( $< 1$  Ma) and support the permeability of the Strait of Gibraltar to the exchange of low-vagile terrestrial fauna during the Pleistocene (Husemann et al., 2014), rejecting the hypothesis of a protracted history of divergence dating back to ancient southern Europe-northern Africa connections during the Tortonian or the Messinian (e.g., Faille et al., 2014; Hidalgo-Galiana & Ribera, 2011; Ortego, Nogueras, & Cordero, 2017). This points to the important impact of Pleistocene glaciations as a diversification engine in the Ibero-Maghrebian region, which has been often assumed to have been scarcely impacted by Quaternary glaciations due to its low latitude and temperature buffering by the Atlantic Ocean and the Mediterranean Sea (Rodríguez-Sánchez et al., 2008).

#### ACKNOWLEDGEMENTS

We are much indebted to Anna Papadopoulou for her valuable help in study design and useful comments, suggestions and corrections on a first draft of the manuscript. We are also grateful to Amparo Hidalgo-Galiana, Víctor Nogueras, and Pedro J. Cordero for their valuable help during field and laboratory work and Sergio Pereira (The Centre for Applied

Genomics) for Illumina sequencing. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) from Estación Biológica de Doñana. We thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. This study was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (CGL2014-54671-P and CGL2017-83433-P). VT was supported by an FPI predoctoral fellowship (BES-2015-73159) from the Spanish Ministry of Economy and Competitiveness.

## REFERENCES

- Adams, D. C. (2014). A Generalized  $K$  statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Systematic Biology*, *63*(5), 685–697. doi: 10.1093/sysbio/syu030
- Adams, D. C., & Collyer, M. L. (2018). Phylogenetic ANOVA: Group-clade aggregation, biological challenges, and a refined permutation procedure: Phylogenetic permutational ANOVA. *Evolution*, *72*(6), 1204–1215. doi: 10.1111/evo.13492
- Adams, D. C., & Otárola-Castillo, E. (2013). *geomorph*: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, *4*(4), 393–399. doi: 10.1111/2041-210X.12035
- Agustí, J., Garcés, M., & Krijgsman, W. (2006). Evidence for African–Iberian exchanges during the Messinian in the Spanish mammalian record. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *238*(1–4), 5–14. doi: 10.1016/j.palaeo.2006.03.013
- Andersson, M. (1994). *Sexual Selection*. Princeton, USA: Princeton University Press.
- Avise, J. C., & Wollenberg, K. (1997). Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences*, *94*(15), 7748–7755. doi: 10.1073/pnas.94.15.7748
- Barracough, T. G., Vogler, A. P., & Harvey, P. H. (1998). Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1366), 241–249. doi: 10.1098/rstb.1998.0206
- Bennett, K., & Provan, J. (2008). What do we mean by ‘refugia’? *Quaternary Science Reviews*, *27*(27–28), 2449–2455. doi: 10.1016/j.quascirev.2008.08.019
- Bernardes, J. S., Dávila, A. M., Costa, V. S., & Zaverucha, G. (2007). Improving model construction of profile HMMs for remote homology detection through structural alignment. *BMC Bioinformatics*, *8*(1), 435. doi: 10.1186/1471-2105-8-435

- Blair, C., Bryson, R. W., Linkem, C. W., Lazcano, D., Klicka, J., & McCormack, J. E. (2019). Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: Crotalus). *Molecular Ecology Resources*, *19*(2), 349–365. doi: 10.1111/1755-0998.12970
- Blomberg, S. P., Garland Jr, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, *57*(4), 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x
- Blondel, J., Aronson, J., Bodiou, J. Y., & Boeuf, G. (2010). *The Mediterranean Region: Biological Diversity in Space and Time*. Oxford, UK: Oxford University Press.
- Bookstein, F., Stringer, C., Weber, G. W., Arsuaga, J.-L., Slice, D. E., Rohlf, F. J., ... Marcus, L. F. (2003). Comparing frontal cranial profiles in archaic and modern Homo by morphometric analysis. 8.
- Bookstein, F.L. (1992). *Morphometric Tools for Landmark Data*. Cambridge, UK: Cambridge University Press.
- Bouckaert, R. R. (2010). DENSITREE: making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372–1373. doi:10.1093/bioinformatics/btq110
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., . . . Drummond, A. J. (2014). BEAST2: A Software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. doi:10.1371/journal.pcbi.1003537
- Bridle, J. R., Saldamando, C. I., Koning, W., & Butlin, R. K. (2006). Assortative preferences and discrimination by females against hybrid male song in the grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, *19*(4), 1248–1256. doi: 10.1111/j.1420-9101.2006.01080.x
- Bridle, J. R., & Butlin, R. K. (2002). Mating signal variation and bimodality in a mosaic hybrid zone between *Chorthippus* grasshopper species. *Evolution*, *56*(6), 1184–1198. doi: 10.1111/j.0014-3820.2002.tb01431.x
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. doi:10.1093/molbev/mss086
- Burbrink, F. T., & Gehara, M. (2018). The biogeography of deep time phylogenetic reticulation. *Systematic Biology*, *67*(5), 743–755. doi: 10.1093/sysbio/syy019
- Burgess, R., & Yang, Z. (2008). Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Molecular Biology and Evolution*, *25*(9), 1979–1994. doi: 10.1093/molbev/msn148
- Butlin, R. K. (1989). Reinforcement of premating isolation. In D. Otte & J. Endler (Eds.), *Speciation and its Consequences* (pp. 158–179). Sunderland, Massachusetts, USA: Sinauer Associates.

- Butlin, R. K. (1998) What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation? In D. J. Howard & S. H. Berlocher (Eds.), *Endless Forms: Species and Speciation* (pp. 367–378). New York, USA: Oxford University Press.
- Butlin, R. K., & Hewitt, G. M. (1985). A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera: Acrididae): Morphological and electrophoretic characters. *Biological Journal of the Linnean Society*, 26(3), 269–285. doi: 10.1111/j.1095-8312.1985.tb01636.x
- Cáliz, M. C. (2015). *Estructura genética y variabilidad fenotípica en el endemismo ibérico-montano Omocestus minutissimus (Brullé, 1832)*. MSc Thesis. Ciudad Real, Spain: University of Castilla-La Mancha. Retrieved from <http://hdl.handle.net/10578/7219>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldon, T. (2009). TRIMAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973. doi: 10.1093/bioinformatics/btp348
- Carranza, S., Harris, D. J., Arnold, E. N., Batista, V., & González de la Vega, J. P. (2006). Phylogeography of the lacertid lizard, *Psammodromus algirus*, in Iberia and across the Strait of Gibraltar. *Journal of Biogeography*, 33(7), 1279–1288. doi: 10.1111/j.1365-2699.2006.01491.x
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics*, 1(3), 171–182. doi:10.1534/g3.111.000240
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. doi:10.1111/mec.12354
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2019). *Orthoptera Species File*. Version 5.0/5.0. Retrieved from <http://orthoptera.speciesfile.org/>
- Clemente, M. E., García, M. D., & Presa, J. J. (1990). Nuevos datos sobre los acridoidea (Insecta: Orthoptera) del Pirineo y Prepirineo catalano-aragonés. *Butlletí de La Institució Catalana d'Història Natural*, 58, 37–44.
- Clemente, M. E., García, M. D., & Presa, J. J. (1991). Los Gomphocerinae de la península Ibérica: 2. *Omocestus Bolívar*, 1878. (Insecta, Orthoptera, Caelifera). *Graellsia*, 46, 191–246.
- Clemente, M. E., García, M. D., Arnaldos, M. I., Romera, E., & Presa, J. J. (1999). Corroboration of the specific status of *Omocestus antigai* (Bolívar, 1897) and *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae). [Confirmación de las posiciones taxonómicas específicas de *Omocestus antigai* (Bolívar, 1897) y *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae)]. *Boletín de la Real Sociedad Española de Historia Natural Sección Biológica*, 95(3-4), 27–50.
- Collina-Girard, J. (2001). L'Atlantide devant le détroit de Gibraltar ? Mythe et géologie. *Comptes Rendus de l'Académie Des Sciences - Series IIA - Earth and Planetary Science*, 333(4), 233–240. doi: 10.1016/S1251-8050(01)01629-9

- Cosson, J. F., Hutterer, R., Libois, R., Sara, M., Taberlet, P., & Vogel, P. (2005). Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, *14*(4), 1151–1162. doi: 10.1111/j.1365-294X.2005.02476.x
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA, USA: Sinauer Association.
- Czekanski-Moir, J. E., & Rundell, R. J. (2019). The Ecology of nonecological speciation and nonadaptive radiations. *Trends in Ecology & Evolution*, *34*(5), 400–415. doi: 10.1016/j.tree.2019.01.012
- DeChaine, E. G., & Martin, A. P. (2005). Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, *32*(11), 1943–1956. doi: 10.1111/j.1365-2699.2005.01356.x
- Djamali, M., Baumel, A., Brewer, S., Jackson, S. T., Kadereit, J. W., López-Vinyallonga, S., ... Simakova, A. (2012). Ecological implications of *Cousinia* Cass. (Asteraceae) persistence through the last two glacial–interglacial cycles in the continental Middle East for the Irano-Turanian flora. *Review of Palaeobotany and Palynology*, *172*, 10–20. doi: 10.1016/j.revpalbo.2012.01.005
- Eaton, D. A. R. (2014). PYRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. doi:10.1093/bioinformatics/btu121
- Eckert, A., & Carstens, B. (2008). Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics and Evolution*, *49*(3), 832–842. doi: 10.1016/j.ympev.2008.09.008
- Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging?. *Evolution*, *63*(1), 1–19. doi: 10.1111/j.1558-5646.2008.00549.x
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501. doi: 10.1146/annurev.ecolsys.39.110707.173414
- Faille, A., Andújar, C., Fadrique, F., & Ribera, I. (2014). Late Miocene origin of an Ibero-Maghrebian clade of ground beetles with multiple colonizations of the subterranean environment. *Journal of Biogeography*, *41*(10), 1979–1990. doi: 10.1111/jbi.12349
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350. doi: 10.1016/j.tig.2012.03.009
- Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA, USA: Sinauer Associates.
- Flantua, S.G.A., & Hooghiemstra, H. (2018). Historical connectivity and mountain biodiversity. In C. Hoorn, A. Perrigo & A. Antonelli (Eds.), *Mountains, Climate and Biodiversity* (1st ed., pp. 171–185). Oxford, UK: Wiley-Blackwell.
- Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution*, *35*(10), 2585–2593. doi: 10.1093/molbev/msy147

- Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2020). A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, *37*(4), 1211–1223. doi: 10.1093/molbev/msz296
- Gangwere, S. K., & Morales Agacino, E. (1970). The biogeography of iberian orthopteroids. *Miscelánea Zoológica*, *2*(5), 9–75.
- Graciá, E., Giménez, A., Anadón, J. D., Harris, D. J., Fritz, U., & Botella, F. (2013). The uncertainty of Late Pleistocene range expansions in the western Mediterranean: A case study of the colonization of south-eastern Spain by the spur-thighed tortoise, *Testudo graeca*. *Journal of Biogeography*, *40*(2), 323–334. doi: 10.1111/jbi.12012
- Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, *22*(2), 71–79. doi: 10.1016/j.tree.2006.10.005
- Gunz, P., Mitteroecker, P., & Bookstein, F. L. (2005). Semilandmarks in three dimensions. In D. E. Slice (Ed.), *Modern Morphometrics in Physical Anthropology* (pp. 73–98). doi: 10.1007/0-387-27614-9\_3
- Haeckel, E. (1866). *Generelle Morphologie der Organismen*. Berlin, Germany: Georg Reiner.
- Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795–809. doi: 10.1093/jhered/esu033
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. doi: 10.1038/35016000
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1442), 183–195. doi: 10.1098/rstb.2003.1388
- Hewitt, G. M. (1990). Divergence and speciation as viewed from an insect hybrid zone. *Canadian Journal of Zoology*, *68*(8), 1701–1715. doi: 10.1139/z90-251
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247–276. doi: 10.1111/j.1095-8312.1996.tb01434.x
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1–2), 87–112. doi: 10.1111/j.1095-8312.1999.tb01160.x
- Hewitt, G. M. (2008). Speciation, hybrid zones and phylogeography - or seeing genes in space and time: seeing genes in space and time. *Molecular Ecology*, *10*(3), 537–549. doi: 10.1046/j.1365-294x.2001.01202.x
- Hidalgo-Galiana, A., & Ribera, I. (2011). Late Miocene diversification of the genus *Hydrochus* (Coleoptera, Hydrochidae) in the west Mediterranean area. *Molecular Phylogenetics and Evolution*, *59*(2), 377–385. doi: 10.1016/j.ympev.2011.01.018
- Huang, J. P. (2016). Parapatric genetic introgression and phenotypic assimilation: testing conditions for introgression between Hercules beetles (Dynastes, Dynastinae). *Molecular ecology*, *25*(21), 5513–5526. doi: 10.1111/mec.13849

- Huang, J. P., Hill, J. G., Ortego, J., & Knowles, L. L. (2020). Paraphyletic species no more—genomic data resolve a Pleistocene radiation and validate morphological species of the *Melanoplus scudderi* complex (Insecta: Orthoptera). *Systematic Entomology*, in press. doi: 10.1111/syen.12415
- Husemann, M., Schmitt, T., Zachos, F. E., Ulrich, W., & Habel, J. C. (2014). Palaeartic biogeography revisited: evidence for the existence of a North African refugium for Western Palaeartic biota. *Journal of Biogeography*, 41(1), 81–94. doi: 10.1111/jbi.12180
- Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., . . . Wolff, E. W. (2007). Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, 317, 793–796.
- Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., ... Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9(1), 906. doi: 10.1038/s41467-018-03294-w
- Keightley, P. D., Ness, R. W., Halligan, D. L., & Hadrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, 196(1), 313–320. doi: 10.1534/genetics.113.158758
- Kelly, C. D. (2014). Sexual selection, phenotypic variation, and allometry in genitalic and non-genitalic traits in the sexually size-dimorphic stick insect *Micrarchus hystriculeus*: Scaling and selection in a stick insect. *Biological Journal of the Linnean Society*, 113(2), 471–484. doi: 10.1111/bij.12344
- Klicka, J., & Zink, R. M. (1997). The importance of recent ice ages in speciation: A failed paradigm. *Science*, 277(5332), 1666. doi: 10.1126/science.277.5332.1666
- Knowles, L. L. (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54(4), 1337–1348. doi: 10.1111/j.0014-3820.2000.tb00566.x
- Knowles, L. L. (2001). Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, 10: 691–701. doi:10.1046/j.1365-294x.2001.01206.x
- Krijgsman, W. (2002). The Mediterranean: Mare Nostrum of Earth sciences. *Earth and Planetary Science Letters*, 205(1–2), 1–12. doi: 10.1016/S0012-821X(02)01008-7
- Knowles, L. L., & Chan, Y.-H. (2008). Resolving species phylogenies of recent evolutionary radiations 1. *Annals of the Missouri Botanical Garden*, 95(2), 224–231. doi: 10.3417/2006102
- Maier, P. A., Vandergast, A. G., Ostoja, S. M., Aguilar, A., & Bohonak, A. J. (2019). Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). *Evolution*, 73(12), 2476–2496. doi: 10.1111/evo.13868
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic biology*, 46(3), 523–536. doi: 10.1093/sysbio/46.3.523
- Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283. doi: 10.1038/nature05706

- Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, 26(9), 448–456. doi: 10.1016/j.tree.2011.05.007
- Mayer, F., Berger, D., Gottsberger, B., & Schulze, W. (2010). Non-ecological radiations in acoustically communicating grasshoppers? In M. Glaubrecht (Ed.), *Evolution in Action* (pp. 451–464). Berlin, Germany: Springer-Verlag. doi: 10.1007/978-3-642-12425-9\_21
- Mendelson, T. C., & Shaw, K. L. (2005). Rapid speciation in an arthropod. *Nature*, 433(7024), 375–376. doi: 10.1038/433375a
- Meulenkamp, J. E., & Sissingh, W. (2003). Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African–Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 196(1–2), 209–228. doi: 10.1016/S0031-0182(03)00319-5
- Myers, E. A., McKelvy, A. D., & Burbrink, F. T. (2020). Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). *Molecular Ecology*, 29(4), 797–811. doi: 10.1111/mec.15358
- Nattier, R., Robillard, T., Amedegnato, C., Couloux, A., Cruaud, C., & Desutter-Grandcolas, L. (2011). Evolution of acoustic communication in the Gomphocerinae (Orthoptera: Caelifera: Acrididae): Evolution of acoustic communication in the Gomphocerinae. *Zoologica Scripta*, 40(5), 479–497. doi: 10.1111/j.1463-6409.2011.00485.x
- Nevado, B., Contreras-Ortiz, N., Hughes, C., & Filatov, D. A. (2018). Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. *New Phytologist*, 219(2), 779–793. doi: 10.1111/nph.15243
- Nichols, R. (2001). Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, 16(7), 358–364. doi: 10.1016/S0169-5347(01)02203-0
- Oh, K. P., & Shaw, K. L. (2013). Multivariate sexual selection in a rapidly evolving speciation phenotype. *Proceedings of the Royal Society B: Biological Sciences*, 280(1761), 20130482. doi: 10.1098/rspb.2013.0482
- Olave, M., & Meyer, A. (2020). Implementing large genomic SNP datasets in phylogenetic network reconstructions: A case study of particularly rapid radiations of cichlid fish. *Systematic Biology*, in press. doi: 10.1093/sysbio/syaa005
- Ortego, J., Nogueras, V., & Cordero, P. J. (2017). Geographical and ecological drivers of mitonuclear genetic divergence in a Mediterranean grasshopper. *Evolutionary Biology*, 44(4), 505–521. doi: 10.1007/s11692-017-9423-x
- Otomuro, D., Söderquist, L., Nilsson-Örtman, V., Cortázar-Chinarro, M., Lundgren, C., & Johansson, F. (2016). Antagonistic natural and sexual selection on wing shape in a scrambling damselfly. *Evolution*, 70(7), 1582–1595. doi: 10.1111/evo.12951

- Pennell, M. W., & Harmon, L. J. (2013). An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology: Integrative comparative methods. *Annals of the New York Academy of Sciences*, 1289(1), 90–105. doi: 10.1111/nyas.12157
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. doi: 10.1371/journal.pone.0037135
- Pfennig, K. S., & Pfennig, D. W. (2009). Character displacement: Ecological and reproductive responses to a common evolutionary problem. *The Quarterly Review of Biology*, 84(3), 253–276. doi: 10.1086/605079
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967. doi: 10.1371/journal.pgen.1002967
- R Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing.
- Ragge, D. R., & Reynolds, W. J. (1998). The songs of the grasshoppers and crickets of western Europe. Colchester, UK: Harley Books.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using TRACER 1.7. *Systematic Biology*, 67(5), 901–904. doi: 10.1093/sysbio/syy032
- Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, 164(4), 1645–1656.
- Reynolds, W. J. (1986). A description of the song of *Omocestus broelemanni* (Orthoptera: Acrididae) with notes on its taxonomic position. *Journal of Natural History*, 20(1), 111–116. doi:10.1080/00222938600770101
- Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P., & Arroyo, J. (2008). The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews*, 27(23–24), 2100–2117. doi: 10.1016/j.quascirev.2008.08.006
- Rohlf, F. J. (2015). The TPS series of software. *Hystrix*, 26(1), 9–12. doi:10.4404/hystrix-26.1-11264
- Rohlf, F. J., & Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1), 40–59. doi:10.2307/2992207
- Ronacher, B. (2019). Innate releasing mechanisms and fixed action patterns: basic ethological concepts as drivers for neuroethological studies on acoustic communication in Orthoptera. *Journal of Comparative Physiology A*, 205(1), 33–50. doi: 10.1007/s00359-018-01311-3
- Roy, K., Valentine, J. W., Jablonski, D., & Kidwell, S. M. (1996). Scales of climatic variability and time averaging in Pleistocene biotas: Implications for ecology and evolution. *Trends in Ecology & Evolution*, 11(11), 458–463. doi:10.1016/0169-5347(96)10054-9

- Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, *24*(7), 394–399. doi: 10.1016/j.tree.2009.02.007
- Safran, R. J., Scordato, E. S. C., Symes, L. B., Rodríguez, R. L., & Mendelson, T. C. (2013). Contributions of natural and sexual selection to the evolution of premating reproductive isolation: a research agenda. *Trends in Ecology & Evolution*, *28*(11), 643–650. doi: 10.1016/j.tree.2013.08.004
- Salzburger, W., Baric, S., & Sturmbauer, C. (2002). Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, *11*(3), 619–625. doi: 10.1046/j.0962-1083.2001.01438.x
- Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J.-C. (2011). The influence of Late Quaternary climate-change velocity on species endemism. *Science*, *334*(6056), 660–664. doi: 10.1126/science.1210173
- Schweizer, M., Warmuth, V., Alaei Kakhki, N., Aliabadian, M., Förschler, M., Shirihi, H., ... Burri, R. (2019). Parallel plumage colour evolution and introgressive hybridization in wheatears. *Journal of Evolutionary Biology*, *32*(1), 100–110. doi: 10.1111/jeb.13401
- Servedio, M. R., & Boughman, J. W. (2017). The role of sexual selection in local adaptation and speciation. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 85–109. doi: 10.1146/annurev-ecolsys-110316-022905
- Shaffer, H. B., & Thomson, R. C. (2007). Delimiting species in recent radiations. *Systematic Biology*, *56*(6), 896–906. doi: 10.1080/10635150701772563
- Shaw, K. L., & Danley, P. D. (2003). Behavioral genomics and the study of speciation at a porous species boundary. *Zoology*, *106*(4), 261–273. doi: 10.1078/0944-2006-00129
- Shepard, D. B., & Burbrink, F. T. (2008). Lineage diversification and historical demography of a sky island salamander, *Plethodon ouachitae*, from the Interior Highlands. *Molecular Ecology*, *17*(24), 5315–5335. doi: 10.1111/j.1365-294X.2008.03998.x
- Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*, *12*(3), e1005896. doi: 10.1371/journal.pgen.1005896
- Solís-Lemus, C., Bastide, P., & Ané, C. (2017). PHYLONETWORKS: A Package for Phylogenetic Networks. *Molecular Biology and Evolution*, *34*(12), 3292–3298. doi: 10.1093/molbev/msx235
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1682), 661–671. doi: 10.1098/rspb.2009.1272
- Taberlet, P., Fumagalli, L., Wust-Saucy, A., & Cosson, J. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*(4), 453–464. doi: 10.1046/j.1365-294x.1998.00289.x

- Teacher, A. G. F., Garner, T. W. J., & Nichols, R. A. (2009). European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, *102*(5), 490–496. doi: 10.1038/hdy.2008.133
- Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular ecology*, *28*(17), 3869–3886. doi: 10.1111/mec.15189
- Tonzo, V., Papadopoulou, A., & Ortego, J. (in press). Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. *Molecular Ecology*, in press. doi: 10.1111/mec.15475
- Vedenina, V., & Mugue, N. (2011). Speciation in Gomphocerine grasshoppers: Molecular phylogeny versus bioacoustics and courtship behavior. *Journal of Orthoptera Research*, *20*(1), 109–125. doi: 10.1665/034.020.0111
- Veith, M., Kosuch, J., & Vences, M. (2003). Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae). *Molecular Phylogenetics and Evolution*, *26*(2), 310–327. doi: 10.1016/S1055-7903(02)00324-X
- Von Helversen, D., Balakrishnan, R., & Von Helversen, O. (2004). Acoustic communication in a duetting grasshopper: Receiver response variability, male strategies and signal design. *Animal Behaviour*, *68*, 131–144. doi: 10.1016/j.anbehav.2003.10.020
- Wallis, G. P., Waters, J. M., Upton, P., & Craw, D. (2016). Transverse Alpine speciation driven by glaciation. *Trends in Ecology & Evolution*, *31*(12), 916–926. doi: 10.1016/j.tree.2016.08.009
- Waters, J. M., Emerson, B. C., Arribas, P., & McCulloch, G. A. (2020). Dispersal reduction: Causes, genomic mechanisms, and evolutionary consequences. *Trends in Ecology & Evolution*, *35*(6), 512–522. doi: 10.1016/j.tree.2020.01.012
- Wen, D., Yu, Y., Hahn, M. W., & Nakhleh, L. (2016). Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology*, *25*(11), 2361–2372. doi: 10.1111/mec.13544
- Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring phylogenetic networks using PHYLONET. *Systematic Biology*, *67*(4), 735–740. doi: 10.1093/sysbio/syy015
- Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr*: A Grammar of Data Manipulation. R package version 0.8.0.1.
- Yang, Z. (2002). Inference of selection from multiple species alignments. *Current Opinion in Genetics & Development*, *12*(6), 688–694. doi: 10.1016/S0959-437X(02)00348-9
- Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. doi: 10.1093/czoolo/61.5.854

Yu, Y., & Nakhleh, L. (2015). A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*, *16*(S10), S10. doi: 10.1186/1471-2164-16-S10-S10

Zink, R. M., & Slowinski, J. B. (1995). Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. *Proceedings of the National Academy of Sciences*, *92*(13), 5832–5835. doi: 10.1073/pnas.92.13.5832

## **SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article and in Appendix I of this Thesis.





## CHAPTER II

# GENOMIC DATA REVEAL DEEP GENETIC STRUCTURE BUT NO SUPPORT FOR CURRENT TAXONOMIC DESIGNATION IN A GRASSHOPPER SPECIES COMPLEX





# Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex

Vanina Tonzo<sup>1</sup> | Anna Papadopoulou<sup>2</sup> | Joaquín Ortego<sup>1</sup>

<sup>1</sup> Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC); Avda. Américo Vespucio 26 – 41092; Seville, Spain

<sup>2</sup> Department of Biological Sciences, University of Cyprus; 1 Panepistimiou Av. – 2109; Nicosia, Cyprus

## Abstract

Taxonomy has traditionally relied on morphological and ecological traits to interpret and classify biological diversity. Over the last decade, technological advances and conceptual developments in the field of molecular ecology and systematics have eased the generation of genomic data and changed the paradigm of biodiversity analysis. Here we illustrate how traditional taxonomy has led to species designations that are supported neither by high throughput sequencing data nor by the quantitative integration of genomic information with other sources of evidence. Specifically, we focus on *Omocestus antigai* and *O. navasi*, two montane grasshoppers from the Pyrenean region that were originally described based on quantitative phenotypic differences and distinct habitat associations (alpine vs. Mediterranean-montane habitats). To validate current taxonomic designations, test species boundaries, and understand the factors that have contributed to genetic divergence, we obtained phenotypic (geometric morphometrics) and genome-wide SNP data (ddRADSeq) from populations covering the entire known distribution of the two taxa. Coalescent-based phylogenetic reconstructions, integrative Bayesian model-based species delimitation, and landscape genetic analyses revealed that populations assigned to the two taxa show a spatial distribution of genetic variation that do not match with current taxonomic designations and is incompatible with ecological/environmental speciation. Our results support little phenotypic variation among populations and a marked genetic structure that is mostly explained by geographic distances and limited population connectivity across the abrupt landscapes characterizing the study region. Overall, this study highlights the importance of integrative approaches to identify taxonomic units and elucidate the evolutionary history of species.

**Keywords:** ddRADseq, geometric morphometrics, integrative species delimitation, landscape genetics, phylogenomic inference, species delimitation

# 1 I Introduction

Describing biological diversity and understanding the ecological and evolutionary processes that generate it is at the core of molecular ecology and allied disciplines (Avice et al., 1987; Dobzhansky, 1937; Simpson, 1944). In the last decade, this field has experienced a revolution as a result of new technological and conceptual advances that allow biologists to discover new taxa at an unprecedented resolution and gain new insights into the processes leading to species formation (Jackson, Carstens, Morales, & O'Meara, 2017; Sites & Marshall, 2003; Wiens, 2007). These advances are timely, as the current biodiversity crisis – i.e., the accelerated loss of species and the rapid degradation of ecosystems – requires the establishment of efficient conservation policies that often put the focus on red lists of formally described taxa (Agapow et al., 2004; Mace, 2004). Yet, the definition of the fundamental biological unit that conforms a species remains controversial due to many alternative species concepts and the different properties/criteria used to define them (see de Queiroz, 2007). Given the inherent difficulties to test for reproductive isolation (i.e., the biological species concept; Dobzhansky, 1970; Mayr, 1942), especially in allopatric populations, most species descriptions have been traditionally based on qualitative or quantitative phenotypic differences (diagnosable or phenetic species concept; e.g., Nelson & Platnick, 1981; Sokal & Crovello, 1970) or distinct ecological traits (e.g., Jones & Weisrock, 2018 and references therein). However, speciation may not necessarily involve shifts in niche preferences or diagnosable morphological changes and adhering to these two sources of evidence could result in an underestimation of the number of biological species.

The wide use of DNA sequence data in taxonomy (e.g., Hebert, Cywinska, Ball, & DeWaard, 2003; Tautz, Arctander, Minelli, Thomas, & Vogler, 2003) and the application of phylogenetic/coalescent-based species delimitation approaches (e.g., Pons et al., 2006; Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012; Yang & Rannala, 2010) have allowed the discovery of cryptic species and the resolution of many taxonomic uncertainties (e.g., Blair, Bryson, Linkem, Lazcano, Klicka, & McCormack, 2019; Burbrink, Yao, Ingrassi, Bryson, Guiher, & Ruane, 2011; Leavitt, Johnson, Goward, & St Clair, 2011; Niemiller, Near, & Fitzpatrick, 2012). In particular, the incorporation of coalescent theory under an integrative taxonomic framework has increased the statistical rigor of species delimitation and has helped to move away from subjective decisions on the degree of differentiation (e.g., phenotypic, genetic, ecological, etc.) that is needed for considering different lineages

or populations as distinct taxa (Fujita et al., 2012; Jones & Weisrock, 2018; Wiens & Servedio, 2000). In recent years, the advent of high throughput DNA sequencing technology has exponentially increased our capability to obtain large scale genomic data from non-model organisms, removing previous data-related constraints, while promoting further development of new phylogenetic inference methods and model-based approaches of species delimitation (Ekblom & Galindo, 2011; Fujita et al., 2012; van Dijk, Auger, Jaszczyszyn, & Thermes, 2014). In particular, the multispecies coalescent model (MCM) is considered an excellent approach to test alternative hypotheses of lineage divergence (Knowles & Carstens, 2007; Yang & Rannala, 2010) and identify boundaries among recently diverged species using multilocus data (Domingos, Colli, Lemmon, Lemmon, & Beheregaray, 2017; Rannala & Yang, 2013; Yang, 2015). More recent developments have incorporated the possibility of combining diverse sources of information (quantitative traits and genomic data) to objectively identify independently evolving lineages into an integrative analytical framework (Edwards & Knowles, 2014; Fujita et al., 2012; Solís-Lemus, Knowles, & Ané, 2015). However, integrative species delimitation approaches are not exempt of limitations (Sukumaran & Knowles, 2017; Huang, 2018). One of these limitations is the challenge to deal with sexually dimorphic traits (Solís-Lemus et al., 2015), which can have a considerable impact on species delimitation inferences (Chan et al., 2017; Noguerales, Cordero, & Ortego, 2018). More importantly, the finest detection of genetic differentiation brought by the highest resolution of genomic data can lead to a potential confusion of population genetic structure with species boundaries (Carstens, Pelletier, Reid, & Satler, 2013; O'Meara, 2010; Sukumaran & Knowles, 2017). That is, in some situations, the higher power of genomic data to detect population structuring may drive to an artifactual increase in the number of inferred species (Fujita et al., 2012; Huang, 2018; Isaac, Mallet, & Mace, 2004). Taxonomic inflation, in turn, can have a negative impact on subsequent ecological and evolutionary studies, compromise our ability to reach solid conclusions on the origin and dynamics of biodiversity and, ultimately, lead to misguided conservation policies and inadequate management practices (Carstens et al., 2013; O'Meara, 2010; Sukumaran & Knowles, 2017).

As mentioned above, a common source of taxonomic controversy is the historical description of species based on the concurrence of ecological and phenotypic distinctiveness (e.g., Dowle, Morgan-Richards, & Trewick, 2014; Jones & Weisrock, 2018).

These are often presumed to represent cases of ecological speciation, i.e., the formation of new species when divergent natural selection under contrasting environmental conditions (e.g., elevation, salinity, etc.) leads to reproductive isolation (Rundle & Nosil, 2005; Schluter, 2001). Genetic data have often supported classic taxonomic studies separating species based on ecological and phenotypic differences (e.g., Lamichhaney et al., 2015). However, in many other cases the recognized taxonomic units are not consistent with patterns of genome-wide differentiation (e.g., Dowle et al., 2014; Jones & Weisrock, 2018), suggesting that the link between environmental and phenotypic divergence is a consequence of plasticity or divergent selection at a few genes or genomic islands involved in local adaptation processes that have not led to the reduction of gene flow (e.g., Mason & Taylor, 2015; Soria-Carrasco et al., 2014). An intermediate situation along the speciation continuum is the frequently reported association between environmental dissimilarity and phenotypic and genetic differentiation found in many studies (Sexton, Hangartner & Hoffmann, 2014; Shafer & Wolf, 2013; Wang & Bradburd, 2014). This pattern of progressive genetic and phenotypic differentiation among populations experiencing contrasting environmental conditions has been termed isolation-by-ecology (IBE) and is generally interpreted as a signal of ongoing local adaptation processes and incipient speciation (Shafer & Wolf, 2013). Thus, the study of these evolutionary phenomena is not only relevant in the context of resolving taxonomic uncertainties, but it also allows to identify important ecological and evolutionary processes along the speciation continuum that might deserve to be protected (Moritz, 2002).

The grasshopper subgenus *Dreuxius* Defaut, 1988 (genus *Omocestus*, Bolívar, 1878) is a complex of nine recently diversified species distributed in the Iberian Peninsula (6 species) and northwestern Africa (3 species) (Cigliano, Braun, Eades, & Otte, 2019; García-Navas, Noguerales, Cordero, & Ortego, 2017). Most taxa have allopatric distributions and are often isolated at high elevations in different mountain systems (Cigliano et al., 2019). The putative sister species pair *Omocestus antigai* (Bolívar, 1897) and *Omocestus navasi* Bolívar, 1908 are distributed through the Pyrenees, including some populations in the pre-Pyrenees and Catalan Pre-Coastal ranges (Poniatowski, Defaut, Lluçia-Pomares, & Fartmann, 2009). These two species were originally described based on distinct habitat associations and certain quantitative phenotypic traits such as body size and forewing shape (Clemente, García, Arnaldos, Romera, & Presa, 1999; Olmo-Vidal, 2002; Puissant, 2008). The current known ranges of both species follow the west to east orientation of the

Pyrenees, but their populations are mostly allopatric. *Omocestus antigai* is distributed at high elevations (1450-2350 m) associated with alpine or subalpine grasslands above the tree line, whereas *O. navasi* is circumscribed to Mediterranean scrubby habitats at lower elevations (700-1600 m) (Llucia-Pomares, 2002; Olmo-Vidal, 2002; Poniatowski et al., 2009). These strong ecological differences and subtle morphological variation have motivated different authors to synonymize the two taxa (Ragge & Reynolds, 1998; Reynolds, 1986), re-validate their distinct species status (Clemente et al., 1999) and even describe a new subspecific taxon (*O. navasi bellmanni* Puissant 2008) in the last decades (Cigliano et al., 2019). Resolving the taxonomy of the complex also has important implications for conservation and management, as the two putative species are included in the IUCN red list of threatened species but with different assessment categories (Braud et al., 2016a, b). *Omocestus navasi* is classified as “endangered” due to its extremely small estimated area of occupancy (60-200 km<sup>2</sup>) and the continuous decline in the extend and quality of habitats (Braud et al., 2016a). In contrast, the high-elevation grasslands occupied by *O. antigai* are supposed to be less affected by direct human impacts and this taxon was assessed as “vulnerable”, with the main threats identified being the small size and high degree of fragmentation of its populations, overgrazing in some areas, and the decline of the quality of its specific alpine and subalpine habitats due to the progressive abandonment of traditional grassland management practices (Braud et al., 2016b). Overall, this makes this complex an excellent model system to test species boundaries and understand the factors that have contributed to genomic, phenotypic and ecological divergence, which will in turn provide the necessary baseline information to determine the conservation status of the different taxonomic units within the complex and design proper on-ground management practices.

In this study, we generated genomic data using restriction-site-associated DNA sequencing (ddRADseq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) and obtained phenotypic information through geometric morphometric analyses (Bookstein, 1992) to shed light on the taxonomy and evolutionary history of *O. antigai* and *O. navasi*. First, we tested the hypothesis that the current taxonomic designation of the complex is supported by genomic and phenotypic data. In particular, we used genome-wide single nucleotide polymorphisms (SNP) data and alternative coalescent-based methods to infer the phylogenetic relationships among all sampled populations of the two taxa. Then, we compared the taxonomic status of the studied populations with phylogenomic

reconstructions and inferred patterns of spatial genetic structure (Dayrat, 2005; Jones & Weisrock, 2018), while we employed a suite of Bayesian model-based species delimitation approaches, one of them integrating genomic data and quantitative traits, to identify independently-evolving lineages (Leaché, Fujita, Minin, & Bouckaert, 2014; Solís-Lemus et al., 2015; Yang & Rannala, 2010). Second, we tested the hypothesis that ecological-driven divergence is the main process underlying spatial patterns of genomic and phenotypic variation in the study system. In particular, we analyzed whether environmental dissimilarity among populations and processes of local adaptation to different elevational and climatic gradients are the main factors shaping genetic differentiation and phenotypic trait variation in the study system or if, on the contrary, genetic and phenotypic structure is mostly explained by the geographical distance among populations and resistance distances defined by the complex topography of the study area.

## 2 | Methods

### 2.1 | Population sampling

Between 2012 and 2015, we collected specimens of *Omocestus antigai* (7 populations), *O. navasi navasi* (8 populations) and *O. navasi bellmanni* (1 population) from 16 sampling localities (hereafter referred to as populations) (Table 1; Figure 1). Samples represent populations across the entire known distribution range of the three taxa based on the literature (Clemente et al., 1999; Lluçia-Pomares, 2002; Poniowski et al., 2009; Puissant, 2008). We stored specimens in 2 ml vials with 96% ethanol and preserved them at  $-20^{\circ}$  C until needed for geometric morphometric and genomic analyses.

### 2.2 | Geometric morphometric analyses

We selected a maximum of 10 adult individuals from each population (5 males and 5 females if available) and used landmark-based geometric morphometric analyses to characterize phenotypic variation in the complex. We excluded from geometric morphometric analyses five populations (MON, SAR, ERR, QUE, and TUR) for which only a few individuals for one or the two sexes ( $< 3$ ) could be sampled due to low population densities and/or because most specimens presented damaged structures (e.g., broken forewings or ovipositor valves). Populations with available geometric morphometric data

**Table 1** Geographical location, elevation and number of genotyped individuals (*n*) for each sampled population (locality) of the three putative taxa and the outgroup.

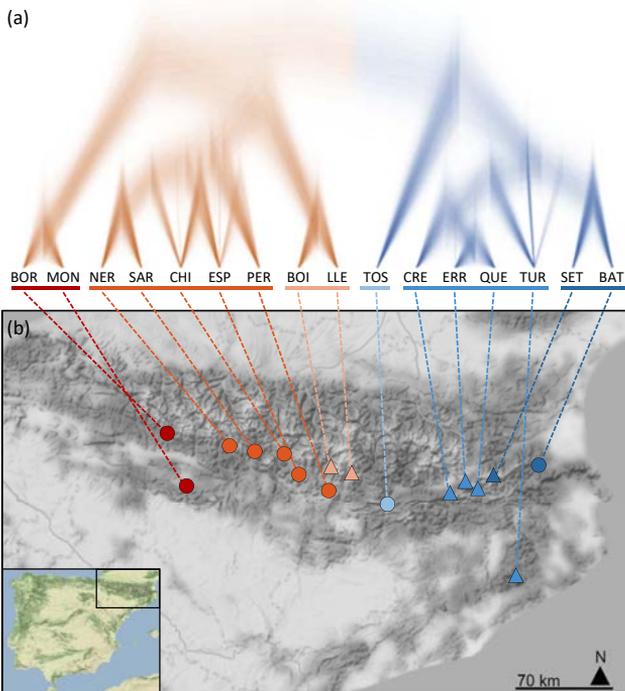
Species	Locality	Code	Latitude	Longitude	Elevation (m)	<i>n</i>
<i>O. antigai</i>	Boí Taüll	BOI	42.47945	0.86732	2,046	5
<i>O. antigai</i>	Llessui	LLE	42.44554	1.03668	1,979	6
<i>O. antigai</i>	Coll de la Creueta	CRE	42.30131	1.99350	1,928	6
<i>O. antigai</i>	Err	ERR	42.38830	2.08812	2,150	3
<i>O. antigai</i>	Queralbs	QUE	42.36627	2.14902	2,070	6
<i>O. antigai</i>	Turó de l'Home	TUR	41.77167	2.44335	1,622	7
<i>O. antigai</i>	Setcases	SET	42.42756	2.26630	2,145	7
<i>O. navasi navasi</i>	Borau	BOR	42.67388	-0.57932	1,357	7
<i>O. navasi navasi</i>	Puerto de Monrepós	MON	42.35067	-0.39762	1,267	5
<i>O. navasi navasi</i>	Nerín	NER	42.59047	0.01062	1,623	7
<i>O. navasi navasi</i>	Saravillo	SAR	42.56089	0.23046	1,313	2
<i>O. navasi navasi</i>	Chía	CHI	42.55418	0.43703	1,691	7
<i>O. navasi navasi</i>	Espés	ESP	42.44307	0.58389	1,361	6
<i>O. navasi navasi</i>	Perves	PER	42.35250	0.83709	1,370	6
<i>O. navasi navasi</i>	Montan de Tost	TOS	42.23790	1.38052	1,182	7
<i>O. navasi bellmanni</i>	Tour de Batère	BAT	42.50785	2.57641	1,430	7
<i>O. femoralis</i>	Sierra de Espuña	ESP	37.86515	-1.57115	1,514	6

included all taxa within the complex (i.e., all species/subspecies) and were representative of all genetic clusters inferred by Bayesian clustering analyses (see Results section). We took digital images of head, pronotum, forewing and ovipositor valves (in females) with a ZEISS stereomicroscope (STEREO DISCOVERY version 8; Carl Zeiss Microscopy GmbH, Germany). These traits correspond to those originally used to distinguish the two putative species (Clemente et al., 1999). The coordinates of landmarks (9-14 landmarks per trait) were mapped on the images using TPSDIG version 2.2 (Rohlf, 2015) and analyzed as implemented in the R version 3.3.2 (R Core Team, 2017) package *Geomorph* (Adams & Otárola-Castillo, 2013). Specifically, we performed generalized Procrustes analyses (GPA) separately for each sex to remove the effects of location, size, and rotation of the relative positions of landmarks among specimens using the function *gpagen*. This superimposition method minimizes the sum-of-squared distances between landmarks across samples (Rohlf & Slice, 1990). To examine shape differences between taxa and among populations, we performed principal component analyses (PCA) on the covariance matrix of aligned Procrustes shape coordinates using the function *plotTangentSpace*. We retained the scores from the two first principal components (PC1 and PC2) for each trait and used multivariate analyses of variance (MANOVA) to investigate if trait shapes were significantly different

between taxa and among sampled populations (e.g., Adams & Otárola-Castillo, 2013; Friedline et al., 2019). Statistical significance of MANOVA models was determined using Wilk's  $\lambda$  as the test statistic and  $\alpha = 0.05$  as a significance threshold. Then, we used one-way analyses of variance (ANOVA) to assess differences between taxa and among sampled populations separately for each phenotypic trait (e.g., Friedline et al., 2019). The first two PCs for each trait were also used for integrative species delimitation analyses using IBPP (Solís-Lemus et al., 2015) and to build matrices of phenotypic differentiation ( $P_{ST}$ ) (see below for details).

### 2.3 I Genomic library preparation and processing

We obtained genomic data for a subset of 94 specimens representative from the 16 sampled populations of *O. antigai*, *O. navasi navasi*, and *O. navasi bellmanni* (~6 individuals per population, 3 males and 3 females when available; Table 1). Additionally, we genotyped six individuals of *O. femoralis* Bolívar 1908 (Table 1), a species also belonging to the subgenus *Dreuxius* (Cigliano et al., 2019) and that was used as an outgroup in some phylogenetic and species delimitation analyses (see below for details). Details on the preparation of ddRADseq libraries (Peterson et al., 2012) are presented in



**Figure 1** (a) Bayesian phylogenetic tree reconstructed with SNAPP (node support presented in Figure 6) and (b) map showing the distribution of the sampled populations of *Omocestus antigai* (triangles) and *O. navasi* (dots) in the Pyrenees. Colors of triangles and dots on the map indicate the main genetic cluster at which populations were assigned according to STRUCTURE analyses for  $K = 6$ . Population codes as in Table 1.

Methods S1. Raw sequences were demultiplexed and pre-processed using STACKS version 1.35 (Catchen, Amores, Hohenlohe, Cresko & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores & Cresko, 2013) and assembled using PYRAD version 3.0.66 (Eaton, 2014). Methods S2 provides all details on sequence assembling and data filtering.

## 2.4 I Population genetic structure

We employed four complementary approaches to exhaustively explore spatial patterns of genetic structure in our study system, including: (i) classic STRUCTURE analyses (Pritchard, Stephens, & Donnelly, 2000); (ii) FASTSTRUCTURE (Raj, Stephens, & Pritchard, 2014); (iii) the recently developed spatial method implemented in the R program CONSTRUCT to infer patterns of genetic structure after controlling for the geographic distance separating the sampled populations (Bradburd, Coop, & Ralph, 2018), given the strong signal of isolation by distance in our study system (see Section 3.6); and (iv) a principal component analysis (PCA) (Jombart, 2008). Further details on these analyses are presented in Methods S3.

## 2.5 I Phylogenetic inference

We estimated species trees using two coalescent-based methods, SNAPP version 1.3 (Bryant et al., 2012) as implemented in BEAST2 version 2.4.3 (Bouckaert et al., 2014) and SVDQUARTETS (Chifman & Kubatko, 2014) as implemented in PAUP\* v. 4.0a152 (Swofford, 2002). SNAPP analyses are computationally highly demanding and, for this reason, we only selected one individual (the one with the highest number of retained reads; Figure S1) from each of the sixteen sampled populations. The resulting dataset contained 16 individuals and retained 858 polymorphic sites. We ran SNAPP analyses for 1,000,000 Markov chain Monte Carlo (MCMC) generations, sampling every 1,000 steps, and using different gamma prior distributions for *alpha* and *beta* (2, 200; 2, 2,000; 2, 20,000). The forward (*u*) and reverse (*v*) mutation rates were set to be calculated by SNAPP and we left the remaining parameters at default values. We conducted two independent runs and evaluated convergence with TRACER version 1.6. We removed 10% of trees as burn-in and merged tree and log files from the different runs using LOGCOMBINER version 2.4.1. We used TREEANNOTATOR version 1.8.3 to obtain maximum credibility trees, TREESETANALYSER version 2.4.1 to identify species trees that were contained in the 95% highest posterior density (HPD), and DENSITREE version 2.2.1 (Bouckaert, 2010) to visualize the posterior

distribution of trees.

SVDQUARTETS uses SNP data to infer phylogenetic relationships between quartets of taxa under the multispecies coalescent (MSC) model and then assembles these quartets into a species tree (Chifman & Kubatko, 2014). We performed SVDQUARTETS analyses including all individuals and populations and setting *O. femoralis* as an outgroup. We constructed species trees by exhaustively evaluating all possible quartets (i.e. every combination of four tips was examined) from the entire genomic dataset (63,860 unlinked SNPs) and used 100 nonparametric bootstrapping replicates to assess branch support (Chifman & Kubatko, 2014).

**Table 2** Results of BFD\* analyses testing the support of competing species delimitation hypotheses. The table shows the clustering scheme defining each alternative species delimitation hypothesis ( $H_i$ ). For each hypothesis, we show marginal likelihood estimates (MLE), their Bayes factors (calculated as  $2 \times \ln \text{BF}$ ) and their rank. Population codes as in Table 1.

Species delimitation hypothesis ( $H_i$ )	Species	MLE	$2 \times \ln \text{BF}$	Rank
$H_1$ : (BAT+PER+SAR+CHI+TOS+NER+ESP+MON+BOR+CRE+ERR+BOI+LLE+QUE+TUR+SET)	1	-4,794.85	$8.21 \times 10^2$	2
$H_2$ : (BOI+LLE+CRE+ERR+QUE+TUR+SET) (PER+SAR+CHI+NER+ESP+MON+BOR+BAT+TOS)	2	-5,601.22	$2.43 \times 10^3$	5
$H_3$ : (BAT) (PER+SAR+CHI+TOS+NER+ESP+MON+BOR) (CRE+ERR+BOI+LLE+QUE+TUR+SET)	3	-5,384.07	$2.00 \times 10^3$	4
$H_4$ : (BAT+TOS+CRE+ERR+QUE+TUR+SET) (PER+SAR+CHI+NER+ESP+MON+BOR+BOI+LLE)	2	-5,221.06	$1.67 \times 10^3$	3
$H_5$ : (BAT+SET) (TOS) (CRE+ERR+QUE+TUR) (PER+SAR+CHI+NER+ESP) (MON+BOR) (BOI+LLE)	6	-4,384.48	-	1

## 2.6 I Species delimitation

We applied three Bayesian coalescent-based species delimitation approaches to determine the number of independently evolving lineages, two of them based only on molecular data (BPP; Yang & Rannala, 2010; BFD\*; Leaché et al., 2014) and the third one integrating molecular and phenotypic data (IBPP; Solís-Lemus et al., 2015).

We used the BFD\* method implemented in SNAPP (Leaché et al., 2014) to contrast five

competing species delimitation hypotheses, including a single-species model and four alternative multi-species models informed by the current taxonomy as well as the phylogenetic and Bayesian clustering analyses (Table 2). This method allows the comparison of alternative species delimitation scenarios in an explicit MSC framework by calculating and comparing marginal likelihood estimates (MLE) for each model. Specifically, our species delimitation hypotheses included: i) the hypothesis of a single species ( $H_1$ ); ii) the current taxonomy of two species ( $H_2$ : *O. antigai* and *O. navasi*); iii) the current taxonomy but considering the subspecies *O. navasi bellmanni* as a separate species ( $H_3$ : three species); iv) a model considering as distinct species the two main genetic clusters (east-west split) revealed by phylogenetic and Bayesian clustering analyses ( $H_4$ ); v) a model considering six species, corresponding to the six genetic clusters inferred by Bayesian clustering analyses in STRUCTURE ( $H_5$ ) (Table 2). We conducted a path sampling analysis of 8 steps each consisting of 100,000 MCMC iterations (after a pre-burn-in of 10,000 iterations), sampling each 100 steps, and using an  $\alpha$  value of 0.3. These settings were sufficient to ensure convergence and obtain ESS > 200. The Bayes factor (BF) test statistic ( $2 \times \ln \text{BF}$ ) was calculated, where BF is the difference in MLE between two competing models (base scenario – alternative scenario). These analyses were performed using a matrix with no missing data (393 SNPs; see Leaché, McElroy, and Trinh 2018 for a similar approach) and including two individuals from each sampled population plus two individuals from the outgroup *O. femoralis*, to be able to test for the single-species model.

BPP version 3.3 can perform species delimitation analyses using a fixed input guide tree specified by the user (option A10, guided analysis) or estimating a species tree (option A11, unguided analysis) (Yang, 2015; Yang & Rannala, 2014). We used both options and for guided analyses we fitted as input trees the topologies yielded by SNAPP and SVDQUARTETS (see Results section). This allowed us to assess the potential impact that alternative tree topologies had on our species delimitation inferences (Leaché & Fujita, 2010). To make BPP analyses computationally tractable, we only included two individuals from each sampled population (same as used for BFD\* analyses) and used a subset of loci from our total genomic dataset (e.g., Huang, 2018; Noguerales et al., 2018; Rancilhac et al., 2019). Specifically, we selected loci with five or more variable sites, as these provide more power in species delimitation analyses than less variable loci (Huang, 2018). To evaluate the impact that the number of employed loci had on our species delimitation inferences, we used three random subsets of loci (200, 500 and 1,000 loci). According to previous studies,

these numbers of loci provide considerable power for species delimitation analyses (e.g., Huang, 2018; Noguerales et al., 2018). We created BPP input files from the .loci output file from PYRAD and randomly selected the different subsets of loci with the specified minimum number of variable sites ( $\geq 5$ ) using two R scripts (*bpp\_convert\_Ama\_sp.r* and *numvarfrombppfile.r*, respectively) written by J-P. Huang and available at [https://github.com/airbugs/Dynastes\\_delimitation](https://github.com/airbugs/Dynastes_delimitation) (Huang, 2018). We analyzed the impact of different demographic scenarios corresponding to different prior combinations of gamma distribution as in Huang and Knowles (2016). We considered four parameter settings:  $\theta = G(1, 10)$  and  $\tau = G(1, 10)$ ,  $\theta = G(1, 10)$  and  $\tau = G(2, 2,000)$ ,  $\theta = G(2, 2,000)$  and  $\tau = G(1, 10)$ , and  $\theta = G(2, 2,000)$  and  $\tau = G(2, 2,000)$ , where  $\theta$  and  $\tau$  refer to the ancestral population sizes and divergence times, respectively. We ran four replicates for each combination. We used an automatic adjustment of the finetune parameters, allowing swapping rates to range between 0.30 and 0.70 (Yang, 2015). We ran each analysis for 100,000 generations, sampling every 10 generations (10,000 samples), after a burn in of 100,000 generations.

The integrative IBPP approach was built upon the architecture of the early version of BPP version 2.1.2 (Rannala & Yang, 2013; Yang & Rannala, 2010) and incorporates models of evolution for continuous quantitative traits under a Brownian motion process (Solís-Lemus et al., 2015). We ran IBPP analyses separately for each sex, incorporating genetic data as well as geometric morphometric data (PC1 and PC2) of three traits for males and four traits for females (see Section 2.2). As in BPP, we ran three independent analyses using as guide trees the topologies yielded by BPP (obtained using option A01; Yang, 2015), SNAPP and SVDQUARTETS, used four prior combinations for gamma distribution and same settings for all other parameter values, and ran four independent replicates for each scenario. These analyses were performed only for populations with available phenotypic data (i.e., excluding MON, SAR, ERR, QUE, and TUR), using a random subset of 500 loci with at least five variable sites, and including five individuals/population (rather than two individuals/population as in BPP analyses) in order to increase the accuracy of estimates of the phenotypic variation of populations. We ran IBPP analyses for 250,000 generations, sampling every 10 generations (25,000 samples), after a conservative burn in of 300,000 generations. Finally, we repeated these analyses considering only phenotypic data for each sex. Posterior probability (PP) of both BPP and IBPP models was considered well supported when  $PP > 0.95$ .

## 2.7 I Landscape genetic and phenotypic analyses

We analyzed three potential factors that could have shaped genetic ( $F_{ST}$ ) and phenotypic ( $P_{ST}$ ) differentiation among populations: (1) geographical distance; (2) environmental heterogeneity; (3) and isolation-by-resistance defined by topographic complexity. We estimated genetic differentiation between each pair of sampled populations calculating pairwise  $F_{ST}$  values in ARLEQUIN version 3.5 (Excoffier, Laval, & Schneider, 2005). Similarly, we estimated phenotypic differentiation between each pair of sampled populations calculating pairwise  $P_{ST}$  values for the first two PCs of each trait using the R package *Pstat* (Silva & Silva, 2018). We calculated the geographical distance between sampled populations using GEOGRAPHIC DISTANCE MATRIX GENERATOR version 1.2.3 ([http://biodiversityinformatics.amnh.org/open\\_source/gdmg](http://biodiversityinformatics.amnh.org/open_source/gdmg)). To estimate environmental dissimilarity between each pair of sampled populations, we extracted for each one the values of the 19 present-day bioclimatic variables available in WORLDCLIM, downloaded at 30 arc-sec (c. 1 km) resolution. Then, we used the “*rda*” function in the R package *vegan* version 2.4-4 (Oksanen et al., 2017) to perform a principal component analysis (PCA) and obtain for each population the PC scores of the first two PCs, which explained the 80 % and 16 % of the environmental variance, respectively. Afterward, we calculated environmental dissimilarity between each pair of populations using Euclidean distances for the obtained PC scores using the “*dist*” function in R. To investigate the role of topographic complexity, we calculated the slope for each cell from a 90 m resolution digital elevation model from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data) and the final raster layer was transformed to 30 arc-sec (c. 1 km) resolution using QGIS version 2.8. Then, based on this layer, we used CIRCUITSCAPE version 4.0 (McRae, 2006; McRae & Beier, 2007) to calculate a matrix of resistance distances between all pairs of populations considering an eight-neighbor cell connection scheme.

We used Multiple Matrix Regressions with Randomization (MMRR) as implemented in R (Wang, 2013) to analyze: (i) pairwise population genetic differentiation ( $F_{ST}$ ) in relation with geographical, resistance and environmental distances, and (ii) pairwise population phenotypic differentiation ( $P_{ST}$ ) in relation with genetic differentiation ( $F_{ST}$ ) and geographical, resistance and environmental distances. Models for all dependent variables were initially constructed with all explanatory terms fitted (i.e., full models) and final models were selected using a backward-stepwise procedure, by progressively removing non-significant variables (starting with the least significant ones) until all retained terms

within the model were significant. Then, we tested the significance of the rejected terms against this model to ensure that no additional variable reached significance. The result was the minimal most adequate model for explaining the variability in the response variable, where only the significant explanatory terms were retained. As we tested a high number of phenotypic traits (eight traits for females and six for males) against the same subset of independent variables, we applied a Benjamini-Hochberg false discovery rate correction to adjust  $p$ -values for multiple statistical tests and calculate the corresponding  $q$ -values using the  $p.adjust$  function in R.

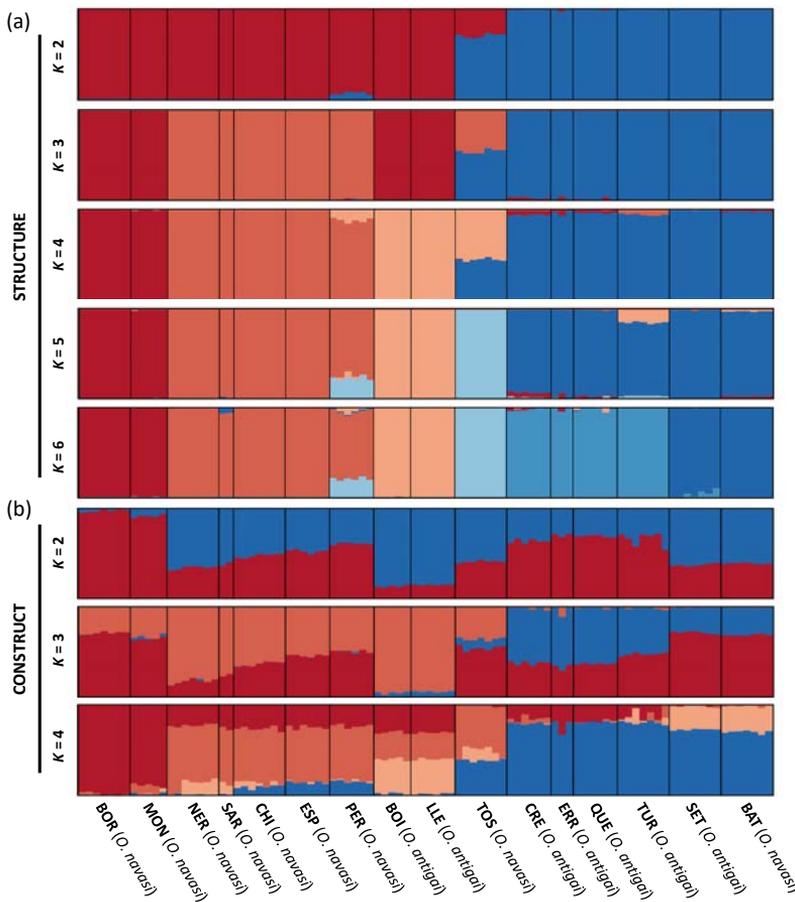
## 3 | Results

### 3.1 | Geometric morphometric analyses

The two first principal component scores, PC1 and PC2, for each trait explained most variation in shapes for pronotum (♀: 87%; ♂: 76%), head (♀: 90%; ♂: 91%), forewing (♀: 65%; ♂: 77%) and ovipositor valves (♀: 95%). Principal component plots for each trait and sex are shown in Figures S2 and S3. MANOVAs based on PC scores for all traits showed significant differences among sampled populations (♀: Wilk's  $\lambda = 0.035$ ,  $F_{80, 192.50} = 1.674$ ,  $p = 0.002$ ; ♂: Wilk's  $\lambda = 0.047$ ,  $F_{54, 142.27} = 2.16$ ,  $p < 0.001$ ) but not between the two taxa (♀: Wilk's  $\lambda = 0.940$ ,  $F_{8, 38} = 0.305$ ,  $p = 0.959$ ; ♂: Wilk's  $\lambda = 0.745$ ,  $F_{6, 35} = 2.00$ ,  $p = 0.092$ ). One-way ANOVAs revealed that multivariate differentiation among populations was driven by shape variation of forewing (PC2:  $F_{10, 40} = 4.86$ ;  $p < 0.001$ ) and ovipositor valves (PC1:  $F_{10, 42} = 2.25$ ;  $p = 0.03$ ) in females and head (PC2:  $F_{10, 37} = 2.22$ ;  $p = 0.039$ ), pronotum (PC1:  $F_{10, 40} = 3.08$ ;  $p = 0.005$ ; PC2:  $F_{10, 40} = 2.47$ ;  $p = 0.021$ ) and forewing (PC1:  $F_{10, 36} = 2.75$ ;  $p = 0.013$ ; PC2:  $F_{10, 36} = 2.33$ ;  $p = 0.031$ ) in males.

### 3.2 | Genomic data

Illumina sequencing returned an average of  $2.19 \times 10^6$  reads per sample. After quality control, an average of  $2.11 \times 10^6$  reads were retained per sample (Figure S1). The total number of unlinked SNPs retained in the final dataset obtained with PYRAD for all individuals and  $minCov = 25\%$  was 64,365 SNPs (63,860 SNPs when the outgroup *O. femoralis* was included).



**Figure 2** Genetic assignment of individuals based on the results of (a) STRUCTURE (from  $K = 2$  to  $K = 6$ ) and (b) CONSTRUCT (from  $K = 2$  to  $K = 4$ ). Individuals are partitioned into  $K$  colored segments representing the probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1.

### 3.3 I Population genetic structure

STRUCTURE analyses yielded an ‘optimal’ clustering value for  $K = 2$  according to the  $\Delta K$  criterion, but log probabilities of the data ( $\text{LnPr}(X|K)$ ) steadily increased up to  $K = 6$  (Figure S4a). The inferred genetic groups were consistent with the geographic distribution of the populations but not with their taxonomic assignment (Figure 2; Figure S5). STRUCTURE analyses showed a hierarchical genetic structure, with a first split between eastern and western populations for  $K = 2$  and subsequent subdivision of these two main population groups into other genetic clusters at smaller geographical scales from  $K = 3$  to  $K = 6$ . Most individuals and populations were assigned to a single genetic cluster, with admixture limited to populations in putative contact zones between genetic clusters.

Maximum-likelihood scores from FASTSTRUCTURE analyses peaked for  $K = 3$  and, according to the  $\Delta K$  criterion the ‘optimal’ clustering solution was  $K = 2$  (Figure S4b). Model complexity that maximizes marginal likelihood was equal to three in 22 replicates, equal to

four in two replicates and equal to five in one replicate, and the number of model components used to explain structure in the data was equal to four in 18 replicates and equal to seven in five replicates. FASTSTRUCTURE results mostly mirrored those obtained with STRUCTURE, but they also presented some differences (Figure S6): (i) For  $K = 3$ , STRUCTURE included in the same genetic cluster the populations BOR-MON and BOI-LLE. However, these two pairs of populations were assigned to different genetic clusters by FASTSTRUCTURE, which makes more geographical sense; (ii) FASTSTRUCTURE analyses for  $K > 5$  did not split populations CRE-ERR-QUE from TUR-SET-BAT as done by STRUCTURE for  $K = 6$ ; (iii) With the exception of TOS, no other population showed any signature of genetic admixture in FASTSTRUCTURE analyses (Figure S6). Assignment values to additional genetic clusters in FASTSTRUCTURE for analyses with  $K > 5$  were extremely low in all cases ( $q < 0.0001$ ) and, thus, their respective bar plots were virtually identical to those obtained for  $K = 5$  (for a similar result, see Baiz, Tucker, & Cortés-Ortiz, 2019).

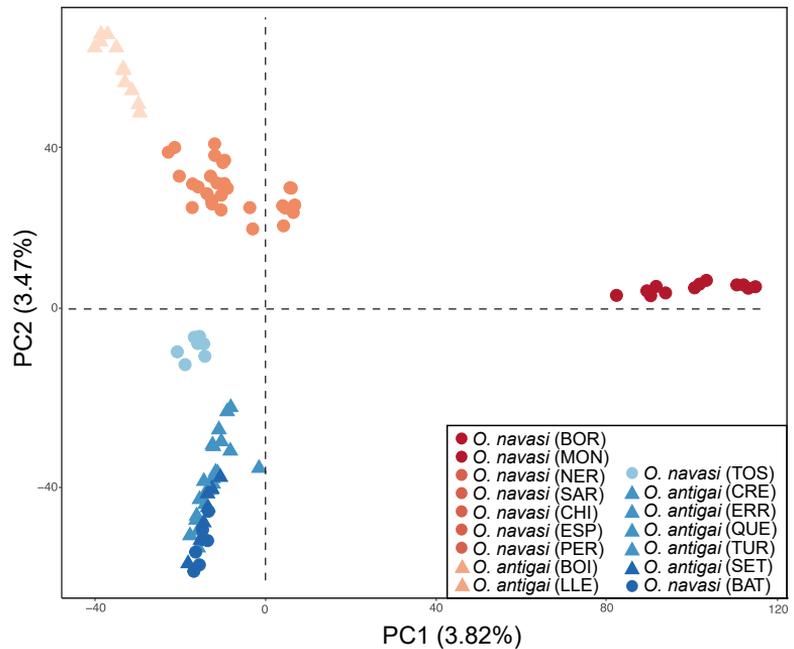
The spatial model of CONSTRUCT showed better model fit at every value of  $K$  than the non-spatial model (Figure S7a). Spatial models had strong statistical support up to  $K = 3-5$  (Figure S7a), but layers beyond  $K = 3$  contributed relatively little to total covariance (Figure S7b). CONSTRUCT analyses showed strong genetic admixture for all  $K$  values (e.g. Whelan et al., 2019) and only genetic assignments for  $K = 3-4$  were partially compatible with the presence of the two main genetic clusters revealed by STRUCTURE and FASTSTRUCTURE (Figure 2; Figure S8).

In agreement with previous analyses, PCA grouped populations according to their geographic location and main genetic clusters rather than by described species (Figure 3). PC1 separated the westernmost populations BOR-MON from the rest of the populations, whereas PC2 separated the rest of western populations from eastern populations and placed the admixed population of TOS at an intermediate position.

### 3.4 I Phylogenetic inference

Population trees reconstructed in BPP, SNAPP, and SVDQUARTETS yielded similar topologies that only differed in the inferred relationships among some nearby populations belonging to the same genetic clusters (Figures 1a and 4). For the west genetic group the three methods inferred different relationships for the populations NER-SAR-CHI-ESP-PER (Figure 4), which were grouped within the same genetic cluster at the lower hierarchical level by STRUCTURE and FASTSTRUCTURE analyses (Figure 2) and clumped together in PCA

**Figure 3** Principal component analysis (PCAs) of genetic variation for *Omocestus antigai* (triangles) and *O. navasi* (dots). Colors of triangles and dots indicate the main genetic cluster at which individuals were assigned according to STRUCTURE analyses for  $K = 6$ . Population codes as in Table 1.

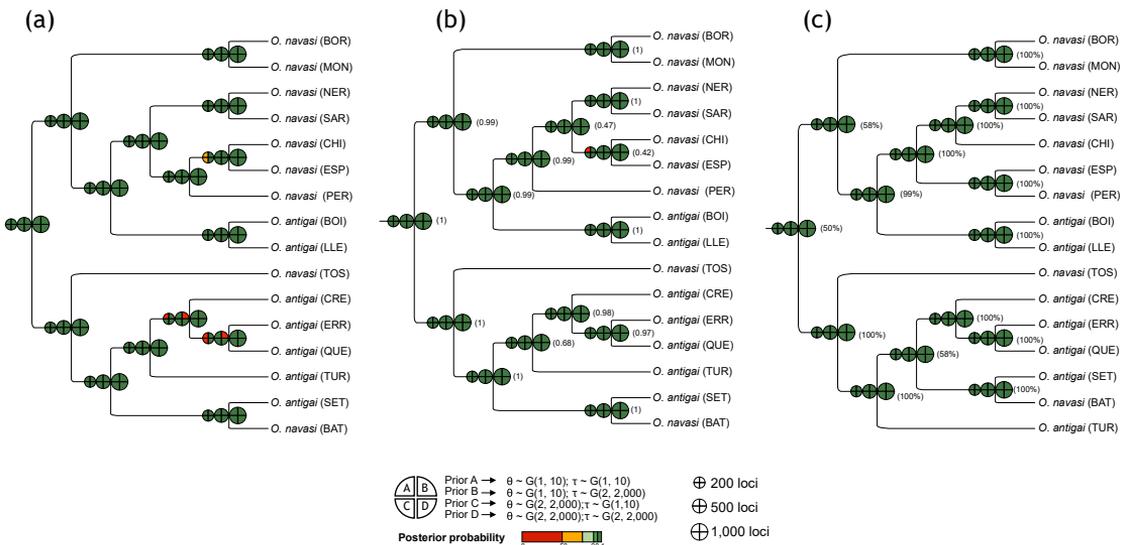


(Figure 3). For the east genetic group, SNAPP and BPP analyses produced the same topology and placed TUR as a sister clade of CRE-ERR-QUE, whereas SVDQUARTETS placed TUR as external group with respect to the five easternmost populations CRE-ERR-QUE-SET-BAT. As expected, the relationship among these populations also showed considerable uncertainty in terms of low node support (i.e. posterior probabilities for SNAPP and bootstrapping values for SVDQUARTETS) (Figure 4). This is also evidenced by the SNAPP analyses, which showed considerable fuzziness in some parts of the tree probably due to gene flow among nearby populations belonging to the same genetic clusters (Figure 1a). In general, the relationships among populations inferred by all phylogenetic analyses are in good agreement with the hierarchical genetic structure yielded by STRUCTURE, FASTSTRUCTURE, and PCA and indicate no support (i.e. polyphyly) for the separation of the named taxa (Figures 1).

### 3.5 I Species delimitation

Species delimitation analyses with the BFD\* method strongly supported a six species model ( $H_5$ ), which considered populations assigned to each genetic cluster identified by Bayesian clustering analyses in STRUCTURE as a distinct species (Table 2). The second most-supported scenario ( $H_1$ ) was the one considering one species, but received much lower support (Table 2). BPP analyses fully supported the presence of 13 species across all tested

topologies (BPP, SNAPP and SVDQUARTETS), prior combinations for gamma distribution, and number of loci (Figure 4). All sampled populations were consistently supported as distinct species when using the subset of 1,000 loci and only some nearby populations belonging to the same genetic clusters (CHI-ESP and CRE-ERR-QUE) were not supported as distinct taxa under a few prior combinations and specific guide trees for the subsets of 200 and 500 loci (Figure 4). IBPP analyses supported all populations as different species using either female or male phenotypic data (Figure 5). IBPP analyses based only on morphological data showed inconsistent results depending on sex and the tested demographic scenario, but tended to support the split into different species for the 3-4 most external nodes (Figure 5). A validation test randomizing phenotypic data across individuals yielded almost an identical result (Figure S9), indicating that the influence of phenotypic traits in IBPP analyses is probably overridden by the high amount of genomic data. Remarkably, IBPP analyses only based on randomized morphological datasets (i.e. without genomic data) also tended to support the split into different species for some external nodes, although the results strongly varied between sexes and across tested topologies and demographic priors (Figure S9).



**Figure 4** Results of species delimitation analyses in BPP. Analyses were performed using three alternative topologies (a: BPP; b: SNAPP; c: SVDQUARTETS), three different subsets of loci (200, 500, and 1000 loci; two individuals/population), and four gamma prior combinations (gamma,  $\alpha$ ,  $\beta$ ) for ancestral population size ( $\theta$ ) and root age ( $\tau$ ). Colored boxes at each node represent the mean posterior probability (PP) for different combinations of demographic priors (legend at bottom). Node support in terms of posterior probabilities (for SNAPP tree) and bootstrapping values (for SVDQUARTETS tree) is indicated in parentheses for each node. Population codes as in Table 1.

### 3.6 | Landscape genetic and phenotypic analyses

MMRR analyses indicated that genetic differentiation ( $F_{ST}$ ) was significantly correlated with both geographic distances ( $p = 0.001$ ) and resistance distances defined by topographic complexity ( $p = 0.004$ ) (Table 3). However, environmental dissimilarity did not show a significant relationship with genetic differentiation and this variable was excluded from the final model ( $p = 0.374$ ; Table 3). MMRR analyses for phenotypic differentiation ( $P_{ST}$ ) showed that no trait was significantly correlated in any sex with geographical distances, resistance distances defined by topographic complexity, environmental dissimilarity or genetic differentiation after applying Benjamini-Hochberg false discovery rate corrections for multiple testing (all  $q$ -values  $> 0.05$ ) (Table 4).

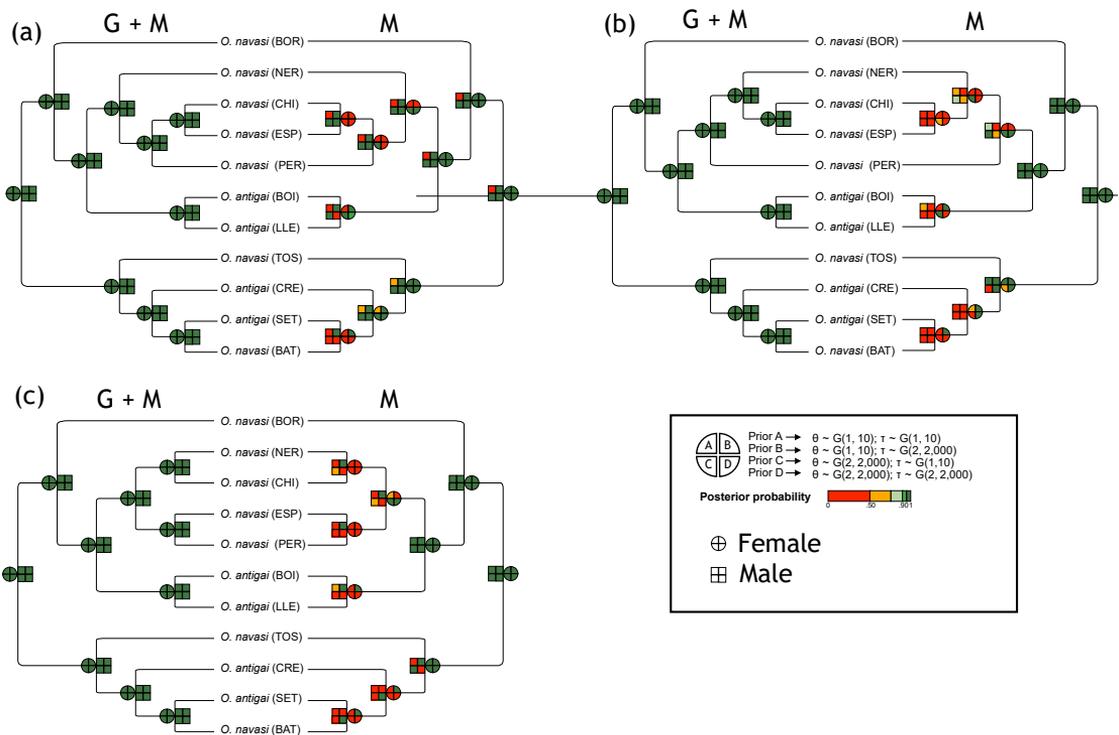
## 4 | Discussion

Collectively, our analyses rejected the current hypothesis of two species and indicated that populations assigned to *O. antigai* and *O. navasi* show a distribution of genetic variation that (i) does not match with their respective taxonomic designation and (ii) is incompatible with ecological/environmental speciation. Our results show the presence of two main genetic groups corresponding to an east-west split analogous to that found in many other Pyrenean taxa (Wallis, Waters, Upton, & Craw, 2016) and a marked genetic differentiation at local spatial scales reflecting limited population connectivity across the abrupt landscapes characterizing the study region (e.g., Noguerales, Cordero, & Ortego, 2016a).

### 4.1 | Integrative species delimitation

As in many other species complexes (e.g., Leaché et al., 2018) the taxonomy of *O. antigai* and *O. navasi* has changed through time in conjunction with the discussion about the validity of the different criteria considered to establish the boundaries between the two putative taxa (Clemente et al., 1999; Ragge & Reynolds, 1998). After the two species were described by Bolívar in the transition between the XIX and XX centuries (Bolívar, 1897; Bolívar, 1908) and Clemente et al. (1990) confirmed their taxonomic distinction, Reynolds (1986) and Ragge & Reynolds (1998) attributed their subtle morphological diversity to the effect of population geographic isolation, denied differences in acoustic communication (i.e., courtship songs), and recommended the synonymy of *O. navasi* to *O.*

*antigai*. However, new acoustic and biometrical analyses by Clemente et al. (1999) resurrected the species *O. navasi* and the two taxa are now considered valid species (Cigliano et al., 2019) and treated as such in conservation assessment programs (Braud et al., 2016a, b; Hochkirch et al., 2016). Up to now, no molecular-based analysis had been performed to test the monophyly of the two taxa and determine whether their specific habitat associations and slight phenotypic differences have a genetic basis in line with a scenario of ecological speciation. Using genomic and phenotypic data and a comprehensive suite of phylogenomic and Bayesian clustering analyses, our results indicate no support for the current taxonomy. Phylogenomic and genetic structure analyses indicate that populations split in two main genetic groups (east/west) that do not match current taxonomic designations and, accordingly, populations sampled in the two distinct habitats (alpine vs. Mediterranean/montane) supposedly occupied by the two putative species are phylogenetically interspersed (Figure 1).



**Figure 5** Results of species delimitation analyses in IBPP. Analyses were performed separately for each sex both combining genomic (500 loci and 5 individuals/population) and phenotypic data (trees on the left) and only using phenotypic data (trees on the right). All analyses were performed using three alternative topologies (a: BPP; b: SNAPP; c: SVDQUARTETS) and four gamma prior combinations (gamma,  $\alpha$ ,  $\beta$ ) for ancestral population size ( $\theta$ ) and root age ( $\tau$ ). Colored boxes at each node represent the mean posterior probability (PP) for different combinations of demographic priors (legend at bottom right). Population codes as in Table 1.

Despite phylogenomic reconstructions not recovering the monophyly of the two nominated taxa, molecular-based species delimitation analyses using BPP and BFD\* identified the presence of several plausible species. Independently of the number of loci used, BPP analyses fully supported the presence of 13 species across all tested topologies and prior combinations for gamma distribution. In most cases, each of the 17 sampled populations was supported as a distinct species and only some nearby populations separated by less than 20 km and belonging to the same genetic clusters (CHI-ESP and CRE-ERR-QUE) were not supported as distinct taxa under a few prior combinations and specific topologies. It is also remarkable that many nearby populations (BOR-MON, NER-SAR, BOILLE and SET-BAT) assigned to the same genetic groups by PCA and Bayesian clustering analyses (STRUCTURE, FASTSTRUCTURE and CONSTRUCT) were consistently delimited as distinct species (Figure 4). Accordingly, the species delimitation model most supported by BFD\* analyses was the one considering the highest number of species, one per genetic cluster inferred by STRUCTURE (see also Fig. 2b in Leaché et al., 2018). Recent empirical and theoretical studies have shown the limitations of Bayesian species delimitation approaches based on the multi-species coalescent (MSC) model (Rannala & Yang, 2013; Yang & Rannala, 2010), pointing out that such methods are not able to statistically distinguish genetic structure due to population isolation from true species boundaries and tend to over-split genetically differentiated populations rather than capturing species divergence (Huang, 2018; Leaché et al., 2018; Leaché, Zhu, Rannala, & Yang, 2019; Sukumaran & Knowles, 2017). The problem of species overestimation can be particularly exacerbated when using vast genome-wide data due to the high power of a large number of markers (hundreds to tens of thousands) to detect fine-grain population structure (Hime et al., 2016; Noguerales et al., 2018; Sukumaran & Knowles, 2017). In this line, recent empirical studies have found that geographically continuous intraspecific populations presenting genotypic differences simply resulted from isolation by distance can be delimited as

**Table 3** Multiple Matrix Regression with Randomization (MMRR) for pairwise population genetic differentiation ( $F_{ST}$ ) in relation with geographical, resistance and environmental distances.

Variable	Coefficient	<i>t</i>	<i>p</i> – Value
Retained terms			
Intercept	-0.108	-3.079	0.991
Geographical distance	0.255	6.841	0.001
Resistance distance (slope)	0.007	5.278	0.004
Rejected term			
Environmental distance	-	-1.061	0.374

distinct species when using hundreds of loci (Huang, 2018; Pyron, Hsieh, Lemmon, Lemmon, & Hendry, 2016).

Recent studies have suggested that species delimitation approaches integrating genomic data with other sources of information (ecological, morphological, ethological, etc.) could help to mitigate the problem of confounding population structure with species limits (Pyron et al., 2016; Sukumaran & Knowles, 2017). Although most phenotypic traits analyzed only presented very subtle differences among some pairs of populations and did not differ between currently recognized taxa (see Figures S2 and S3), our IBPP analyses supported most populations as distinct species and a validation test randomizing phenotypic data across individuals yielded an almost identical result. Thus, incorporating morphological traits has no or very little influence on the outcome of IBPP analyses and the impact of phenotypic traits seems to be overridden by the high amount of genetic data. Our analyses suggest that the species delimitation approach implemented in IBPP does not really help to reduce the number of delineated species, indicating that integrating phenotypic and genetic data into this framework cannot prevent species overestimation (Noguerales et al., 2018; Sukumaran & Knowles, 2017). In sum, the strong genetic structure in our study system and the high number of loci employed are likely to have extraordinarily inflated the number of species identified by BPP and BFD\* and integrative analyses implemented in IBPP does not seem to help to ameliorate this problem.

Beyond the taxonomic inflation issue linked with the fact that the multispecies coalescent model frequently diagnoses genetic structure rather than species, different intersecting lines of evidence suggest that the multiple delineated entities do not meet the requirements to be considered distinct taxa according to alternative contemporary species concepts (see Table 1 in de Queiroz, 2007). These lines of evidence include: (i) the presence of largely overlapping phenotypes across populations, taxa, and genetic clusters for all studied traits (Figure S2 and S3; *phenetic species concept*); (ii) No evidence for environmental-driven divergence and the presence of phylogenetically interspersed high and low elevation populations (*ecological + phylogenetic species concepts*); (iii) Genetic clusters are always allopatric and admixture at contact zones happens even between the two most diverging eastern and western genetic groups (*genotypic cluster species concept*), which is indicative of lack of intrinsic reproductive isolation (*biological species concept*) (de Queiroz, 2007). Although the strong signal for isolation-by-distance might be potentially compatible with a rapid stepping-stone speciation process (e.g. VanderWerf,

Young, Yeung, & Carlon, 2010), such pattern most plausibly reflects population connectivity (Hutchison & Templeton, 1999; Slatkin, 1993). Thus, in our opinion, the studied populations should be treated – at least by now – as a single species taxon with strong genetic structure, which is not incompatible with the possibility of designating the main genetic clusters (e.g. east-west genetic groups) as intraspecific evolutionary

	Females		Males	
	<i>t</i>	<i>q</i>	<i>t</i>	<i>q</i>
Head – PC1				
Geographical distance	0.448	0.919	-0.048	0.980
Resistance distance (slope)	0.989	0.791	-1.106	0.853
Environmental distance	-1.699	0.791	-2.993	0.288
$F_{ST}$	0.125	0.950	-0.052	0.980
Head – PC2				
Geographical distance	1.882	0.686	1.223	0.864
Resistance distance (slope)	-1.047	0.791	-0.429	0.973
Environmental distance	-1.423	0.791	0.135	0.980
$F_{ST}$	2.781	0.400	1.735	0.840
Pronotum – PC1				
Geographical distance	-1.264	0.791	-0.858	0.973
Resistance distance (slope)	-0.416	0.919	0.376	0.973
Environmental distance	0.963	0.791	1.713	0.853
$F_{ST}$	-0.960	0.791	0.587	0.973
Pronotum – PC2				
Geographical distance	0.868	0.919	0.715	0.973
Resistance distance (slope)	-0.267	0.919	0.318	0.973
Environmental distance	-1.240	0.791	-0.054	0.980
$F_{ST}$	-0.735	0.919	1.152	0.853
Forewing – PC1				
Geographical distance	2.775	0.405	0.695	0.973
Resistance distance (slope)	-0.190	0.919	-0.386	0.973
Environmental distance	-2.197	0.686	-2.455	0.288
$F_{ST}$	2.695	0.520	-0.031	0.980
Forewing – PC2				
Geographical distance	-0.222	0.919	-1.471	0.840
Resistance distance (slope)	0.388	0.919	2.185	0.496
Environmental distance	3.159	0.400	0.782	0.973
$F_{ST}$	1.110	0.791	2.045	0.720
Ovopositor valve – PC1				
Geographical distance	-0.282	0.919	–	–
Resistance distance (slope)	0.210	0.919	–	–
Environmental distance	1.614	0.686	–	–
$F_{ST}$	-0.408	0.919	–	–
Ovopositor valve – PC2				
Geographical distance	-0.257	0.919	–	–
Resistance distance (slope)	0.637	0.919	–	–
Environmental distance	1.401	0.791	–	–
$F_{ST}$	0.617	0.919	–	–

**Table 4** Multiple Matrix Regressions with Randomization (MMRR) for pairwise population phenotypic differentiation ( $P_{ST}$ ) in relation with genetic differentiation ( $F_{ST}$ ) and geographical, resistance and environmental distances. No independent variable was significant and retained into final models. Table shows *q*-values after applying Benjamini-Hochberg false discovery rate corrections of *p*-values to adjust for multiple statistical tests

significant units (Moritz, 2002) that should be considered in future conservation and management strategies (Braud et al., 2016a, b).

## 4.2 I Factors structuring genomic and phenotypic variation

Phylogenomic reconstructions and spatial patterns of genetic structure inferred from classic Bayesian clustering analyses supported that genetic variation within the complex is hierarchically structured in congruence with the geographical distribution of the sampled populations. However, the strong signal of isolation-by-distance ( $r = 0.71$ ,  $p < 0.0001$ ; Table 3) probably explains why the spatial analyses in CONSTRUCT blur most genetic structure and do not even clearly recover the west/east split revealed by phylogenetic reconstructions and all other analyses. Our study system presents a high degree of genetic differentiation (mean  $F_{ST} = 0.35$ ) and extraordinarily high estimates of genetic drift after divergence for all genetic clusters inferred by STRUCTURE analyses ( $F$ -value  $> 0.5$ ; Pritchard et al., 2000). Also, there is a remarkable congruence between the hierarchical genetic structure inferred by non-spatial clustering analyses (STRUCTURE, FASTSTRUCTURE) and both phylogenetic reconstructions (SNAPP, SVDQUARTETS, and BPP) and PCA, the latter being a method free of the assumptions made by classic clustering analyses (Jombart, Devillard, & Balloux, 2010). All these lines of evidence suggest that results from CONSTRUCT make little biological sense in our specific study system and indicate that the strong genetic structure revealed by all other analyses is genuine and resulted from historical processes (i.e. strong isolation) rather than being an artifactual consequence of gradual genetic differentiation (i.e. migration and genetic drift equilibrium) merely driven by geographical distance (Bradburd et al., 2018).

In agreement with STRUCTURE and FASTSTRUCTURE results for  $K = 2$ , all phylogenetic analyses showed a basal split in two clades corresponding to populations located east and west of an imaginary line located around the Segre river ( $\sim 19^{\circ}\text{E}$ ). Similar east-west genetic splits have been reported for many other plant and animal species from the Pyrenean region (e.g., Alvarez-Presas, Mateos, Vila-Farre, Sluys, & Riutort, 2012; Bidegaray-Batista et al., 2016; Charrier, Dupont, Pornon, & Escaravage, 2014; Mila, Carranza, Guillaume, & Clobert, 2010; Valbuena-Ureña et al., 2018). This east-west split might be well explained by detailed geological reconstructions documenting heavily glaciated areas in the central part of the range and the likely presence of ice-free refugia at its longitudinal extremes (Ehlers, Gibbard, & Hugues, 2011) that might have allowed montane/alpine species to

survive glaciations and recolonize high elevations during interglacial periods (Charrier et al., 2014; Valbuena-Ureña et al., 2018). Overall, these results are consistent with the transverse breaks proposed by Wallis *et al.* (2016) for numerous alpine taxa and support the idea that the fragmentation of ancestral distributions during glacial periods and the presence of Pleistocene refugia along mountain ranges has played a key role on the diversification not only of lowland species but also of montane and alpine organisms.

Genomic data also revealed strong genetic differentiation and a deep genetic structure at fine spatial scales, with up to three genetic clusters comprised within each of the two main west and east genetic groups (Figure 2). The fact that almost all sampled populations have been assigned to a unique genetic cluster, with only a few cases of genetic admixture between nearby populations, suggest a strong effect of geographic isolation (Wang, 2013). Only populations TOS and PER (and into a lesser extent SAR), located in the contact zone between the west and east genetic groups presented signatures of genetic admixture, evidencing ongoing gene flow between them (Figures 1 and 2). It is noteworthy that other populations located at a similar longitude but at higher elevations showed no evidence of genetic admixture (e.g., BOI, LLE), which suggests a higher isolation of alpine populations and increased gene flow through the less abrupt landscapes characterizing the Pyrenean foothills. Accordingly, landscape genetic analyses indicated that genetic differentiation was explained by both the geographical distance among populations and resistance distances defined by topographic roughness (Table 3). These results might reflect the low dispersal capability of the studied taxon, males being brachypterous and females micropterous, and are comparable to those obtained by previous studies showing the impact of steep slopes and complex landscapes on structuring genetic variation in montane/alpine grasshoppers (Noguerales, Cordero, & Ortego, 2016a). Genetic structure analyses showed that the split of the different populations at local/regional scales followed a longitudinal cline rather than a segregation of alpine and Mediterranean-montane populations (e.g., SET and BAT), indicating no support for either ecologically driven divergence or the taxonomic separation between the supposedly Mediterranean *O. navasi* and the alpine *O. antigai*. This was corroborated by our landscape genetic analyses, which revealed no effect of environmental dissimilarity on structuring genetic variation in the complex (i.e., isolation-by-environment; Shafer & Wolf, 2013).

Our geometric morphometric analyses showed that the subtle phenotypic differences found among populations were not explained by genetic differentiation, geographical

distances or environmental dissimilarity. Overall, this suggests that the weak phenotypic differences found among populations are not a consequence of genetic drift or environmental-driven selection (Keller, Alexander, Holderegger, & Edwards, 2013; Leinonen, Cano, Makinen, & Merila, 2006; Leinonen, O'Hara, Cano, & Merila, 2008) and might be explained by ecological and evolutionary aspects not considered in this study such as sexual selection, predation risk, microhabitat structure or adaptations to different feeding resources (e.g., Ingley, Billman, Belk, & Johnson, 2014; Laiolo, Illera, & Obeso, 2013; Nogueras, Cordero, & Ortego, 2016b).

### 4.3 I Conclusions

This study exemplifies the problems associated with species validation tests involving recently diverged allopatric taxa and highlights the importance of integrating different sources of information to delimit species that have been described solely by morphology or ecological distinctiveness (e.g., Jones & Weisrock, 2018). Phylogenetic inferences, Bayesian species delimitation analyses, and phenotypic and ecological data did not support the current taxonomic status of the complex and indicate that *O. navasi* and *O. antigai* must be synonymized into a unique taxon: *O. antigai* (Bolívar, 1897) (for a list of synonyms, see Table S1). Our study also illustrates the implications that incorrect taxonomic designations can have for species conservation and management. The recent assessment of the conservation status of European grasshoppers has assigned the two taxa to different IUCN Red List categories (Hochkirch et al., 2016), considering *O. navasi* as a “endangered” species (Braud et al., 2016a) and *O. antigai* as “vulnerable” (Braud et al., 2016b). In the view of our results, the conservation status of the complex requires total reconsideration in future IUCN Red List assessments. Given that the range and population sizes of *O. antigai* are larger and the ecological and habitat requirements much wider than previously thought, the conservation status of the taxon should be probably downlisted to “near threatened” (Braud et al., 2016a, b). However, the presence of two marked genetic groups and many genetically sub-structured populations should be also considered in future conservation plans as potential evolutionary significant units that probably deserve to be protected and managed independently (Moritz, 2002). Overall, our analyses point to the presence of a single species characterized by a strong genetic structure, little phenotypic variation, and a wide environmental niche. Future studies performing demographic reconstructions and spatiotemporally explicit landscape genetic analyses

that integrate available information on past glacier extent (Ehlers et al., 2011) and inferred distributional shifts of the species linked to Pleistocene glacial cycles could greatly help to further understand the historical processes that have shaped the spatial distribution of genetic variation within and among the current populations (e.g., Massatti & Knowles, 2016).

## ACKNOWLEDGEMENTS

We are grateful to Amparo Hidalgo-Galiana for her support during the preparation of genomic libraries, Víctor Noguerales and Pedro J. Cordero for their help during fieldwork, Francisco Rodríguez-Sánchez for providing valuable advice in spatial data analyses and Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing. We also thank three anonymous referees for helpful and constructive comments on an earlier version of this article. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) and Laboratorio de Sistemas de Información Geográfica y Teledetección (LAST-EBD) from Estación Biológica de Doñana. We also thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. This work was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (CGL2014-54671-P and CGL2017-83433-P). VT was supported by an FPI predoctoral fellowship (BES-2015-73159) from Ministerio de Economía y Competitividad. During this work, AP and JO were supported by a Severo Ochoa (SEV-2012-0262) and a Ramón y Cajal (RYC-2013-12501) research fellowship, respectively.

## AUTHOR CONTRIBUTIONS

VT, AP and JO conceived and designed the study and analyses. All authors collected the samples. VT performed the lab work and analysed the data guided by AP and JO. VT wrote the manuscript with help of JO and inputs from AP.

## REFERENCES

- Adams, D. C., & Otarola-Castillo, E. (2013). *Geomorph*: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4), 393-399. doi:10.1111/2041-210x.12035
- Agapow, P. M., Bininda-Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C., & Purvis, A. (2004). The impact of species concept on biodiversity studies. *Quarterly Review of Biology*, 79(2), 161-179. doi:10.1086/383542
- Alvarez-Presas, M., Mateos, E., Vila-Farre, M., Sluys, R., & Riutort, M. (2012). Evidence for the persistence of the land planarian species *Microplana terrestris* (Muller, 1774) (Platyhelminthes, Tricladida) in microrefugia during the Last Glacial Maximum in the northern section of the Iberian Peninsula. *Molecular Phylogenetics and Evolution*, 64(3), 491-499. doi:10.1016/j.ympev.2012.05.001
- Avice, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., . . . Saunders, N. C. (1987). Intraspecific phylogeography - The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18, 489-522. doi:10.1146/annurev.ecolsys.18.1.489
- Baiz, M. D., Tucker, P. K., & Cortes-Ortiz, L. (2019). Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular Ecology*, 28(5), 1056-1069. doi:10.1111/mec.14966
- Bidegaray-Batista, L., Sanchez-Gracia, A., Santulli, G., Maiorano, L., Guisan, A., Vogler, A. P., & Arnedo, M. A. (2016). Imprints of multiple glacial refugia in the Pyrenees revealed by phylogeography and palaeodistribution modelling of an endemic spider. *Molecular Ecology*, 25(9), 2046-2064. doi:10.1111/mec.13585
- Blair, C., Bryson, R. W., Linkem, C. W., Lazcano, D., Klicka, J., & McCormack, J. E. (2019). Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: Crotalus). *Molecular Ecology Resources*, 19(2), 349-365. doi:10.1111/1755-0998.12970
- Bookstein, F.L. (1992). *Morphometric tools for landmark data*. Cambridge, UK: Cambridge University Press.
- Bolívar, I. (1897). Catálogo sinóptico de los ortópteros de la fauna ibérica. *Annaes de Ciencias Naturaes*, 4(4): 203-232.
- Bolívar, I. (1908). Algunos ortopteros nuevos de España, Marruecos y Canarias. *Boletin de la Real Sociedad Española de Historia Natural*, 8, 317-334.
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., . . . Drummond, A. J. (2014). BEAST2: A Software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537. doi:10.1371/journal.pcbi.1003537
- Bouckaert, R. R. (2010). DENSITREE: making sense of sets of phylogenetic trees. *Bioinformatics*, 26(10), 1372-1373. doi:10.1093/bioinformatics/btq110
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1), 33-52. doi:10.1534/genetics.118.301333

- Braud, Y., Hochkirch, A., Fartmann, T., Presa, J.J., Monnerat, C., Roesti, C., ... Dusoulier, F. (2016a). *Omocestus antigai*: The IUCN Red List of Threatened Species 2016 (Publication no. e.T16084614A70647955. <http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T16084614A70647955.en>). Retrieved from <https://www.iucnredlist.org/species/16084614/70647955>
- Braud, Y., Hochkirch, A., Presa, J.J., Roesti, C., Rutschmann, F., Monnerat, C., ... Dusoulier, F. (2016b). *Omocestus navasi*: The IUCN Red List of Threatened Species 2016 (Publication no. e.T69641425A69641429. <http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T69641425A69641429.en>). Retrieved from <https://www.iucnredlist.org/species/69641425/69641429>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), 1917-1932. doi:10.1093/molbev/mss086
- Burbrink, F. T., Yao, H., Ingrasci, M., Bryson, R. W., Guiher, T. J., & Ruane, S. (2011). Speciation at the Mogollon Rim in the Arizona Mountain Kingsnake (*Lampropeltis pyromelana*). *Molecular Phylogenetics and Evolution*, 60(3), 445-454. doi:10.1016/j.ympev.2011.05.009
- Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, 22(17), 4369-4383. doi:10.1111/mec.12413
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140. doi:10.1111/mec.12354
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genetics*, 1(3), 171-182. doi:10.1534/g3.111.000240
- Clemente, M. E., García, M. D., Arnaldos, M. I., Romera, E., & Presa, J. J. (1999). Corroboration of the specific status of *Omocestus antigai* (Bolívar, 1897) and *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae). [Confirmación de las posiciones taxonómicas específicas de *Omocestus antigai* (Bolívar, 1897) y *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae)]. *Boletín de la Real Sociedad Española de Historia Natural Sección Biológica*, 95(3-4), 27-50.
- Clemente, M. E., García, M. D., & Presa, J. J. (1990). Nuevos datos sobre los Acridoidea (Insecta: Orthoptera) del Pirineo y prepirineo Catalano-Aragonés. *Butlletí de la Institució Catalana d'Història Natural*, 58, 37-44.
- Chan, K. O., Alexander, A. M., Grismer, L. L., Su, Y. C., Grismer, J. L., Quah, E. S. H., & Brown, R. M. (2017). Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. *Molecular Ecology*, 26(20), 5435-5450. doi:10.1111/mec.14296
- Charrier, O., Dupont, P., Pornon, A., & Escaravage, N. (2014). Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the Pyrenees. *PLoS One*, 9(3), e92976. doi:10.1371/journal.pone.0092976

- Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317-3324. doi:10.1093/bioinformatics/btu530
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2019). *Orthoptera Species File*. Version 5.0/5.0. Retrieved from <http://orthoptera.speciesfile.org/>
- Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85(3), 407-415. doi:10.1111/j.1095-8312.2005.00503.x
- de Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56(6), 879-886. doi:10.1080/10635150701701083
- Dobzhansky, T. (1937). Genetic nature of species differences. *American Naturalist*, 71, 404-420. doi:10.1086/280726
- Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. New York, NY: Columbia University Press.
- Domingos, F., Colli, G. R., Lemmon, A., Lemmon, E. M., & Beheregaray, L. B. (2017). In the shadows: Phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: Tropidurus). *Molecular Phylogenetics and Evolution*, 107, 455-465. doi:10.1016/j.ympev.2016.12.009
- Dowle, E. J., Morgan-Richards, M., & Trewick, S. A. (2014). Morphological differentiation despite gene flow in an endangered grasshopper. *BMC Evolutionary Biology*, 14, 216. doi:10.1186/s12862-014-0216-x
- Eaton, D. A. R. (2014). PYRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844-1849. doi:10.1093/bioinformatics/btu121
- Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B-Biological Sciences*, 281, 20132765. doi:10.1098/rspb.2013.2765
- Ehlers, J., Gibbard, P. L., & Hugues, P. D. (2011). *Quaternary glaciations – Extent and chronology: A closer look* (1st ed.). Amsterdam, The Netherlands: Elsevier.
- Eklblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1-15. doi:10.1038/hdy.2010.152
- Excoffier, L., Laval, G., & Schneider, S. (2005). ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47-50.
- Friedline, C. J., Faske, T. M., Lind, B. M., Hobson, E. M., Parry, D., Dyer, R. J. ... Eckert, A. J. (2019) Evolutionary genomics of gypsy moth populations sampled along a latitudinal gradient. *Molecular Ecology*, 28(9), 2206-2223. doi.org/10.1111/mec.15069
- Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27(9), 480-488. doi:10.1016/j.tree.2012.04.012

- García-Navas, V., Noguerales, V., Cordero, P. J., & Ortego, J. (2017). Ecological drivers of body size evolution and sexual size dimorphism in short-horned grasshoppers (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, 30(8), 1592-1608. doi:10.1111/jeb.13131
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, 270(1512), 313-321. doi:10.1098/rspb.2002.2218
- Hime, P. M., Hotaling, S., Grewelle, R. E., O'Neill, E. M., Voss, S. R., Shaffer, H. B., & Weisrock, D. W. (2016). The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology*, 25(23), 5959-5974. doi:10.1111/mec.13883
- Hochkirch, A., Nieto, A., García-Criado, M., Cálix, M., Braud, Y., Buzzetti, F. M., ... Tumbrinck, J. (2016). *European red list of grasshoppers, crickets and bush-crickets*. Luxembourg, Luxembourg: Publications Office of the European Union.
- Huang, J. P. (2018). What have been and what can be delimited as species using molecular data under the multi-species coalescent model? A case study using *Hercules* beetles (Dynastes: Dynastidae). *Insect Systematics and Diversity*, 2(2), 1-10. doi:10.1093/isd/ixy003
- Huang, J. P., & Knowles, L. L. (2016). The species versus subspecies conundrum: Quantitative delimitation from integrating multiple data types within a single Bayesian approach in hercules beetles. *Systematic Biology*, 65(4), 685-699. doi:10.1093/sysbio/syv119
- Hutchison, D. W., & Templeton, A. R. (1999). Correlation of 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.
- Ingle, S. J., Billman, E. J., Belk, M. C., & Johnson, J. B. (2014). Morphological divergence driven by predation environment within and between species of *Brachyrhaphis* fishes. *PLoS One*, 9(2), e90274. doi:10.1371/journal.pone.0090274
- Isaac, N. J. B., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, 19(9), 464-469. doi:10.1016/j.tree.2004.06.004
- Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, 66(5), 799-812. doi:10.1093/sysbio/syw117
- Jombart, T. (2008). *Adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405. doi:10.1093/bioinformatics/btn129
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. doi:10.1186/1471-2156-11-94
- Jones, K. S., & Weisrock, D. W. (2018). Genomic data reject the hypothesis of sympatric ecological speciation in a clade of *Desmognathus* salamanders. *Evolution*, 72(11), 2378-2393. doi:10.1111/evo.13606

- Keller, I., Alexander, J. M., Holderegger, R., & Edwards, P. J. (2013). Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *Journal of Evolutionary Biology*, *26*(12), 2527-2543. doi:10.1111/jeb.12255
- Knowles, L. L., & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, *56*(6), 887-895. doi:10.1080/10635150701701091
- Laiolo, P., Illera, J. C., & Obeso, J. R. (2013). Local climate determines intra- and interspecific variation in sexual size dimorphism in mountain grasshopper communities. *Journal of Evolutionary Biology*, *26*(10), 2171-2183. doi:10.1111/jeb.12213
- Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martínez-Barrio, A., . . . Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371-375. doi:10.1038/nature14181
- Leinonen, T., Cano, J. M., Makinen, H., & Merila, J. (2006). Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*, *19*(6), 1803-1812. doi:10.1111/j.1420-9101.2006.01182.x
- Leinonen, T., O'Hara, R. B., Cano, J. M., & Merila, J. (2008). Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, *21*(1), 1-17. doi:10.1111/j.1420-9101.2007.01445.x
- Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B-Biological Sciences*, *277*(1697), 3071-3077. doi:10.1098/rspb.2010.0662
- Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, *63*(4), 534-542. doi:10.1093/sysbio/syu018
- Leaché, A. D., McElroy, M. T., & Trinh, A. (2018). A genomic evaluation of taxonomic trends through time in coast horned lizards (genus *Phrynosoma*). *Molecular Ecology*, *27*(13), 2884-2895. doi:10.1111/mec.14715
- Leaché, A. D., Zhu, T. Q., Rannala, B., & Yang, Z. H. (2019). The spectre of too many species. *Systematic Biology*, *68*(1), 168-181. doi:10.1093/sysbio/syy051
- Leavitt, S. D., Johnson, L. A., Goward, T., & St Clair, L. L. (2011). Species delimitation in taxonomically difficult lichen-forming fungi: An example from morphologically and chemically diverse *Xanthoparmelia* (Parmeliaceae) in North America. *Molecular Phylogenetics and Evolution*, *60*(3), 317-332. doi:10.1016/j.ympev.2011.05.012
- Llucia-Pomares, D. (2002). Revision of the Orthoptera (Insecta) of Catalonia (Spain). [Revisión de los ortópteros (Insecta: Orthoptera) de Cataluña (España)]. *Monografías de la Sociedad Entomológica Aragonesa*, *7*, 1-226.
- Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *359*(1444), 711-719. doi:10.1098/rstb.2003.1454

- Mason, N. A., & Taylor, S. A. (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, 24(12), 3009-3025. doi:10.1111/mec.13140
- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, 25(16), 3974-3986. doi:10.1111/mec.13735
- Mayr, E. (1942). *Systematics and the Origin of Species*. New York, NY: Columbia University Press.
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8), 1551-1561. doi:10.1111/j.0014-3820.2006.tb00500.x
- McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), 19885-19890. doi:10.1073/pnas.0706568104
- Milá, B., Carranza, S., Guillaume, O., & Clobert, J. (2010). Marked genetic structuring and extreme dispersal limitation in the Pyrenean brook newt *Calotriton asper* (Amphibia: Salamandridae) revealed by genome-wide AFLP but not mtDNA. *Molecular Ecology*, 19(1), 108-120. doi:10.1111/j.1365-294X.2009.04441.x
- Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, 51(2), 238-254.
- Nelson, G., & Platnick, N. (1981). *Systematics and Biogeography*. New York, NY: Columbia University Press.
- Niemiller, M. L., Near, T. J., & Fitzpatrick, B. M. (2012). Delimiting species using multilocus data: Diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution*, 66(3), 846-866. doi:10.1111/j.1558-5646.2011.01480.x
- Noguerales, V., Cordero, P. J., & Ortego, J. (2016a). Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, 16, 96. doi:10.1186/s12862-016-0663-7
- Noguerales, V., García-Navas, V., Cordero, P. J., & Ortego, J. (2016b). The role of environment and core-margin effects on range-wide phenotypic variation in a montane grasshopper. *Journal of Evolutionary Biology*, 29(11), 2129-2142. doi:10.1111/jeb.12915
- Noguerales, V., Cordero, P. J., & Ortego, J. (2018). Integrating genomic and phenotypic data to evaluate alternative phylogenetic and species delimitation hypotheses in a recent evolutionary radiation of grasshoppers. *Molecular Ecology*, 27(5), 1229-1244. doi:10.1111/mec.14504
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ...Wagner, H. (2017). *vegan*: Community Ecology Package. R package version 2.4-4. Retrieved from <http://cran.r-project.org/package=vegan>
- Olmo-Vidal, J.M. (2002). *Atlas Dels Ortòpters de Catalunya i Llibre Vermell 2002 / Llagostes, Saltamartins, Grills, Someretes /Atlas de Los Ortópteros de Cataluña y Libro Rojo 2002 / Atlas of the Otthoptera of*

- Catalonia and Read Data Book 2002*. Barcelona, Spain: Generalitat de Catalunya, Departament de Medi Ambient i Habitatge.
- O'Meara, B. C. (2010). New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, *59*(1), 59-73. doi:10.1093/sysbio/syp077
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. doi:10.1371/journal.pone.0037135
- Poniatowski, D., Defaut, B., Lluçia-Pomares, D., & Fartmann, T. (2009). The Orthoptera fauna of the Pyrenean region - a field guide. *Articulata Beiheft*, *14*, 1-143.
- Pons, J., Barraclough, T. G., Gómez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., . . . Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*(4), 595-609.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.
- Puissant, S. (2008). The presence of *Omocestus navasi* Bolívar, 1908 in France with description of a new subspecies (Caelifera, Acrididae, Gomphocerinae). [Sur la presence en France d'*Omocestus navasi* Bolívar, 1908, avec description d'une nouvelle sous-espece (Caelifera, Acrididae, Gomphocerinae).]. *Materiaux Orthopteriques et Entomocenotiques*, *13*, 69-73.
- Pyron, R. A., Hsieh, F. W., Lemmon, A. R., Lemmon, E. M., & Hendry, C. R. (2016). Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). *Zoological Journal of the Linnean Society*, *177*(4), 937-949. doi:10.1111/zoj.12392
- R Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing.
- Ragge, D. R., & Reynolds, W. J. (1998). *The songs of the grasshoppers and crickets of western Europe*. Colchester, UK: Harley Books.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573-589. doi:10.1534/genetics.114.164350
- Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M. R., Elmer, K. R., Vences, M., & Steinfarz, S. (2019). Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis. *Molecular Phylogenetics and Evolution*, *133*, 189-197. doi:10.1016/j.ympev.2019.01.003
- Rannala, B., & Yang, Z. H. (2013). Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, *194*(1), 245-253. doi:10.1534/genetics.112.149039
- Reynolds, W. J. (1986). A description of the song of *Omocestus broelemanni* (Orthoptera: Acrididae) with notes on its taxonomic position. *Journal of Natural History*, *20*(1), 111-116. doi:10.1080/00222938600770101

- Rohlf, F. J. (2015). The TPS series of software. *Hystrix-Italian Journal of Mammalogy*, 26(1), 9-12. doi:10.4404/hystrix-26.1-11264
- Rohlf, F. J., & Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1), 40-59. doi:10.2307/2992207
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336-352. doi:10.1111/j.1461-0248.2004.00715.x
- Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7), 372-380.
- Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which patterns of gene flow is most common? *Evolution*, 68(1), 1-15. doi:10.1111/evo.12258
- Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, 16(7), 940-950. doi:10.1111/ele.12120
- Da Silva, S.B., & Da Silva, A. (2018). *Pstat*: An R Package to assess population differentiation in phenotypic traits. *The R Journal*, 10(1), 447-454.
- Simpson, G. G. (1944). *Tempo and Mode in Evolution*. New York, NY: Columbia University Press.
- Sites, J. W., & Marshall, J. C. (2003). Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, 18(9), 462-470. doi:10.1016/s0169-5347(03)00184-8
- Slatkin, M. (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, 47(1), 264-279.
- Sokal, R. R., & Crovello, T. J. (1970). Biological species concept - A critical evaluation. *American Naturalist*, 104(936), 127-153. doi:10.1086/282646
- Solis-Lemus, C., Knowles, L. L., & Ane, C. (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69(2), 492-507. doi:10.1111/evo.12582
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., . . . Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(6185), 738-742. doi:10.1126/science.1252136
- Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, 114(7), 1607-1612. doi:10.1073/pnas.1607921114
- Swofford, D. L. (2002). PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sunderland, MA: Sinauer Associates.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., & Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, 18(2), 70-74. doi:10.1016/s0169-5347(02)00041-1
- Valbuena-Ureña, E., Oromi, N., Soler-Membrives, A., Carranza, S., Amat, F., Camarasa, S., . . . Steinfartz, S. (2018). Jailed in the mountains: Genetic diversity and structure of an endemic newt species across the Pyrenees. *PLoS One*, 13(8), e0200214. doi:10.1371/journal.pone.0200214

- VanderWerf, E. A., Young, L. C., Yeung, N. W., & Carlon, D. B. (2010). Stepping stone speciation in Hawaii's flycatchers: molecular divergence supports new island endemics within the elepaio. *Conservation Genetics*, *11*(4), 1283-1298. doi:10.1007/s10592-009-9958-1
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, *30*(9), 418-426. doi:10.1016/j.tig.2014.07.001
- Wallis, G. P., Waters, J. M., Upton, P., & Craw, D. (2016). Transverse alpine speciation driven by glaciation. *Trends in Ecology & Evolution*, *31*(12), 916-926. doi:10.1016/j.tree.2016.08.009
- Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, *67*(12), 3403-3411. doi:10.1111/evo.12134
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649-5662. doi:10.1111/mec.12938
- Whelan, N. V., Galaska, M. P., Siple, B. N., Weber, J. M., Johnson, P. D., Halanych, K. M., & Helms B. S. (2019). Riverscape genetic variation, migration patterns, and morphological variation of the threatened round rocksnail, *Leptoxis ampla*. *Molecular Ecology*, *28*(7), 1593-1610. doi:10.1111/mec.15032
- Wiens, J. J. (2007). Species delimitation: New approaches for discovering diversity. *Systematic Biology*, *56*(6), 875-878. doi:10.1080/10635150701748506
- Wiens, J. J., & Servedio, M. R. (2000). Species delimitation in systematics: inferring diagnostic differences between species. *Proceedings of the Royal Society B-Biological Sciences*, *267*(1444), 631-636. doi:10.1098/rspb.2000.1049
- Yang, Z. H. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854-865.
- Yang, Z. H., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9264-9269. doi:10.1073/pnas.0913022107

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article and in Appendix II of this Thesis.

SCAN ME TO SEE ARTICLE ONLINE



Tonzo, V., Papadopoulou, A., & Ortego, J. (2019).

Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex.

*Molecular Ecology*, 28(17), (pp. 3869-3886).



## CHAPTER III

# GENOMIC FOOTPRINTS OF AN OLD AFFAIR: SINGLE NUCLEOTIDE POLYMORPHISM DATA REVEAL HISTORICAL HYBRIDITATION AND THE SUBSEQUENT EVOLUTION OF REPRODUCTIVE BARRIERS IN TWO RECENTLY DIVERGED GRASSHOPPERS WITH PARTLY OVERLAPPING DISTRIBUTIONS





# Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions.

Vanina Tonzo<sup>1</sup> | Anna Papadopoulou<sup>2</sup> | Joaquín Ortego<sup>1</sup>

<sup>1</sup> Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC); Avda. Américo Vespucio 26 – 41092; Seville, Spain

<sup>2</sup> Department of Biological Sciences, University of Cyprus; 1 Panepistimiou Av. – 2109; Nicosia, Cyprus

## Abstract

Secondary contact in close relatives can result in hybridization and the admixture of previously isolated gene pools. However, after an initial period of hybridization, reproductive isolation can evolve through different processes and lead to the interruption of gene flow and the completion of the speciation process. *Omocestus minutissimus* and *O. uhagonii* are two closely related grasshoppers with partially overlapping distributions in the Central System mountains of the Iberian Peninsula. To analyse spatial patterns of historical and/or contemporary hybridization between these two taxa and understand how species boundaries are maintained in the region of secondary contact, we sampled sympatric and allopatric populations of the two species and obtained genome-wide SNP data using a restriction site-associated DNA sequencing approach. We used Bayesian clustering analyses to test the hypothesis of contemporary hybridization in sympatric populations and employed a suite of phylogenomic approaches and a coalescent-based simulation framework to evaluate alternative hypothetical scenarios of interspecific gene flow. Our analyses rejected the hypothesis of contemporary hybridization but revealed past introgression in the area where the distributions of the two species overlap. Overall, these results point to a scenario of historical gene flow after secondary contact followed by the evolution of reproductive isolation that currently prevents hybridization among sympatric populations.

**Keywords:** coalescent-based simulations, ddRAD-seq, hybridization, introgression, reproductive isolation

# 1 I INTRODUCTION

Elucidating the processes that generate and maintain species diversity is a main ambition of evolutionary research (Graham, Ron, Santos, Schneider, & Moritz, 2004; Grant, Grant, Markert, Keller, & Petren, 2004; Fitzpatrick, Fordyce, & Gavrillets, 2009). The formation and persistence of species often depends on the evolution of reproductive isolation mechanisms that prevent interbreeding with other recently diverged taxa or closely related lineages (Marques, Draper, Riofrio, & Naranjo, 2014; Soltis & Soltis, 2009). In the context of allopatric populations, long-term geographic isolation can facilitate the development of reproductive barriers through genetic drift or divergent selection pressures, which can ultimately lead to population divergence and speciation (Hoskin, Higgie, McDonald, & Moritz, 2005; Maguilla, Escudero, Hipp, & Luceno, 2017; Schenk, Kontur, Wilson, Noble, & Derryberry, 2018). However, lineages or species often come into secondary contact and, if barriers to gene flow are lacking or incomplete, hybridization will take place. This phenomenon can have different outcomes with important evolutionary consequences (Abbott et al., 2013; Mallet, 2005; Mayr, 1963). At one extreme, barriers to gene exchange may break down upon secondary contact and lead to the collapse of formerly distinct species in a hybrid swarm characterized by extensive admixture of parental genotypes (i.e., speciation reversal or lineage fusion; Kearns et al., 2018; Taylor et al., 2006). At the opposite extreme, reduced fitness of hybrids and strong selection against them will favour the rapid evolution of barriers to gene flow (i.e., prezygotic isolation), which can ultimately lead to total reproductive isolation and culminate in the completion of the speciation process (i.e., reinforcement of isolation; Butlin, 1995; Dobzhansky, 1937; Lemmon & Juenger, 2017; Servedio & Kirkpatrick, 1997). An intermediate scenario is the formation of tension zones with a variable geographical width determined by the equilibrium between dispersal and selection against hybrids (Barton & Hewitt, 1985; Key, 1968).

Hybrid zones have been defined as “windows on the evolutionary processes” (Harrison, 1990; Hewitt, 1988) and their study through space (Barton & Hewitt, 1989; Howard, Waring, Tibbets, & Gregory, 1993) and time (Britch, Cain, & Howard, 2001; Buggs, 2007) have provided important insights into the evolution of reproductive isolation and the formation of new species. Past climate changes, such as Pleistocene glacial cycles, have resulted in recurrent range expansions and contractions in many organisms, often putting into geographic contact closely related species and lineages that had remained

geographically isolated for large periods of time. Alpine and montane species represent a paradigmatic example of species experiencing dramatic distributional shifts in response to Pleistocene glacial cycles, descending to lower altitudes and expanding their distributions in cold periods and shrinking their ranges to high elevation areas during interglacials (Schmitt, 2009; Seddon, Santucci, Reeve, & Hewitt, 2001; Tzedakis, Emerson, & Hewitt, 2013). Under these conditions, closely related species that evolved in isolation during interglacial periods can recurrently come into secondary contact and hybridize. In other cases, interbreeding species share a large portion of their respective ranges but only hybridize in certain areas with specific environmental conditions or scattered patches where they meet, forming a mosaic hybrid structure rather than a well-defined cline limited to narrow contact zones (Barton & Hewitt, 1985). Independently of their nature and origin, contact zones offer the opportunity to study in real time the process of reproductive isolation or, if completed, to obtain indirect evidence about when (tempo) and how (modes) it might have evolved. However, testing alternative hypotheses about the evolution of reproductive isolation can be extremely challenging, making necessary the integration of multiple sources of analytical evidence (e.g., phylogenetics and population genetics) and the development of model-based approaches considering the biogeographical context of gene flow (Payseur & Rieseberg, 2016). The high power of genomic data to resolve historical events of hybridization and test complex scenarios of gene flow in virtually any organismal model and biogeographical setting has exponentially increased our capacity to quantify the magnitude and timing of interspecific gene flow and distinguish among alternative demographic scenarios (e.g., de Manuel et al., 2016; Lohse, Clarke, Ritchie, & Etges, 2015; Ortego, Gugger, & Sork, 2018).

Here, we use as a study system two recently diverged grasshopper species with partly overlapping distributions to illustrate the potential of integrating different analytical approaches for inferring the tempo and mode of evolution of reproductive isolation (or its lack thereof) and gain insights into the speciation process. Grasshoppers (Orthoptera: Caelifera) are an interesting system to study hybridization and its evolutionary consequences, as many species have very recently evolved in allopatry (Mayer, Berger, Gottsberger, & Schulze, 2010; Ragge & Reynolds, 1998) and present incomplete barriers to gene flow (e.g., Saldamando, Tatsuta, & Butlin, 2005). In turn, their distributions often overlap across large geographical areas (Hill, 2015), show a mosaic distribution (Rohde et al., 2017) or have recurrently come into secondary contact as a consequence of range

shifts driven by past climate changes (Bridle, Baird, & Butlin, 2001), providing ideal biogeographic scenarios for the study of hybridization and the evolution of reproductive isolation (Butlin, Ritchie, & Hewitt, 1991; Virdee & Hewitt, 1994). In this study we focus on two grasshopper species of the subgenus *Dreuxius* (genus *Omocestus*), a species complex comprised of nine taxa distributed in the Iberian Peninsula and Northwestern Africa (Cigliano, Braun, Eades, & Otte, 2019; García-Navas, Nogueras, Cordero, & Ortego, 2017). Most species of this complex are distributed in allopatry, isolated at high elevation in different mountain ranges. One exception are the taxa *Omocestus minutissimus* (Brullé 1832) and *O. uhagonii* (Bolivar 1876), which show partially overlapping distributions in the Central System Mountains from the Iberian Peninsula. As the rest of species of the subgenus, both taxa are brachypterous, present a low dispersal capacity, and have a similar annual life cycle, with an adult breeding phase from the end of July to the beginning of October (Clemente, Garcia, & Presa, 1991). The two species are predominantly graminivorous and occupy open habitats tightly linked to cushion and thorny shrub formations that they use as refuge (Clemente et al., 1991). However, both species differ on the extent of their distributions and elevational ranges. While *O. minutissimus* presents a wider distribution, with patchy populations distributed in eastern and central Iberia from sea level to 2,500 m of elevation, *O. uhagonii* is restricted to the Central System and altitudes over 1,600-1,800 m (Figure 1). The two taxa partially co-occur across the distribution range of *O. uhagonii*, with several sympatric populations at high elevations in which adult individuals of the two species co-exist at high numbers in the same microhabitats (J. Ortego, personal observation). Therefore, this system provides an interesting case study to analyse the presence of contemporary and past hybridization and understand the maintenance of species boundaries in two closely related taxa that might have weak or recently evolved reproductive isolation mechanisms.

We extensively sampled sympatric and allopatric populations of *O. minutissimus* and *O. uhagonii* across their respective distribution ranges and genotyped them via restriction-site-associated DNA sequencing (ddRAD-seq; (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) to infer historical and contemporary interspecific gene flow and elucidate the evolutionary outcomes of such processes. Specifically, we first used Bayesian clustering analyses to determine the genetic ancestry of individuals and test the hypothesis of contemporary hybridization in sympatric populations of the two taxa. Second, we employed a suite of phylogenomic approaches and a coalescent-based simulation

framework to evaluate alternative scenarios of historical hybridization and estimate the timing, magnitude and directionality of interspecific gene flow. Our genomic data rejected the hypothesis of contemporary hybridization but revealed the presence of past genetic introgression, pointing to a scenario of historical hybridization after secondary contact followed by the evolution of reproductive barriers that nowadays prevent gene flow among sympatric populations of the two species.

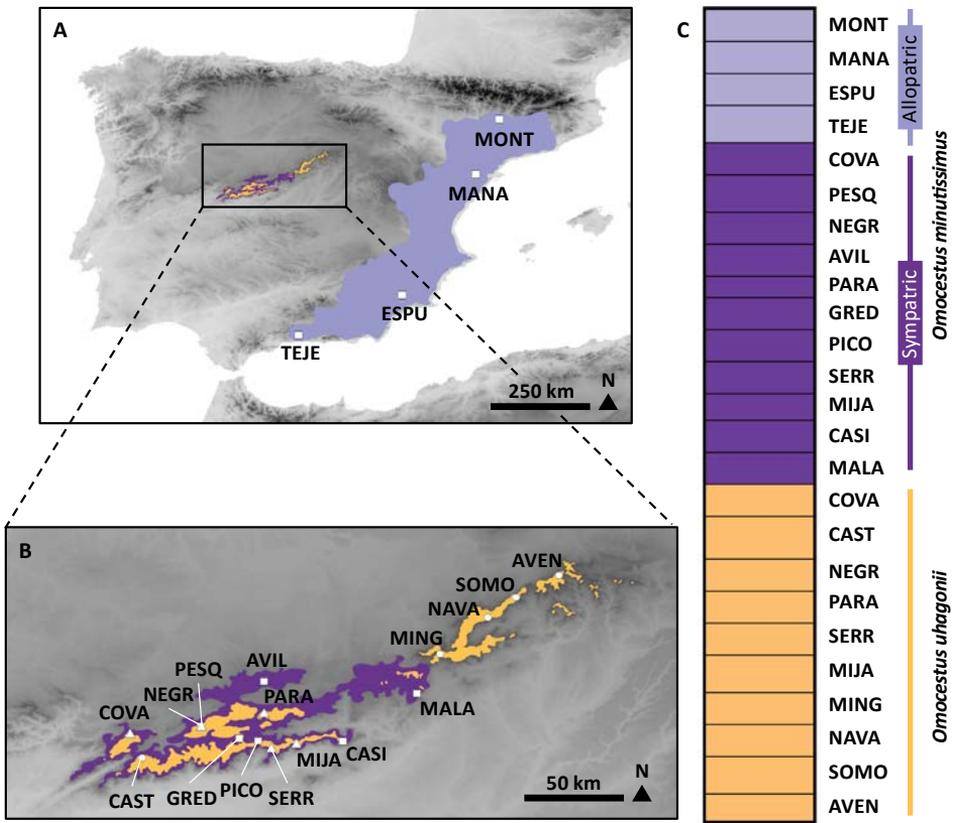
## 2 | MATERIALS AND METHODS

### 2.1 | Population sampling

Between 2011 and 2015, we sampled *Omocestus minutissimus* (88 individuals, 15 populations) and *O. uhagonii* (64 individuals, 10 populations) from a total of 25 populations that cover the entire distribution range of the two taxa (Table S1; Figure 1). Eleven sampling populations of *O. minutissimus* are located in central Iberia and partly overlap with the distribution range of *O. uhagonii* (hereafter referred as sympatric populations). Among these populations of *O. minutissimus*, five were strictly sympatric with *O. uhagonii* and the two species were collected from the same localities (Table S1; Figure 1). The rest of the sampled populations of *O. minutissimus* are located in eastern Iberia and separated >300 km from the nearest population of *O. uhagonii* (hereafter referred as allopatric populations). We stored specimens in 2 ml vials with 96% ethanol and preserved them at  $-20^{\circ}$  C until needed for DNA extraction. Detailed information on sampling populations is presented in Table S1.

### 2.2 | Genomic library preparation and data processing

We used NucleoSpin Tissue kits (Macherey-Nagel, Düren, Germany) to extract and purify total genomic DNA from a hind leg of each individual. We processed genomic DNA in house following the double-digest restriction-site associated DNA procedure (ddRADseq) described in Peterson et al. (2012), with some minor modifications detailed in Lanier, Massatti, He, Olson, & Knowles (2015). Briefly, we digested DNA with the restriction enzymes EcoRI and MseI (New England Biolabs, Ipswich, MA, USA) and ligated Illumina adaptors including unique 7-base-pair barcodes to the digested fragments of each individual. We pooled ligation products into four different libraries, size selected for



**Figure 1** Biogeographical setting of the study system. (A-B) Maps show the sampled populations and the distribution range of the two studied taxa based on our own species records (*O. minutissimus*: purple areas and squares; *O. uhagonii*: light orange areas and dots). *Omocestus minutissimus* presents a partially overlapping distribution with *O. uhagonii* in the Central System (deep purple) and allopatric populations (light purple) in eastern Iberia. Triangles indicate sampling localities where the two species were found living in sympatry. (C) Genetic assignment of individuals based on the results of FASTSTRUCTURE. Individuals are partitioned into  $K$  coloured segments representing the probability of belonging to the cluster with that colour and thin vertical black lines separate individuals from different populations. Population codes as in Table S1.

fragments between 475 and 580 bp using a Pippin Prep machine (Sage Science, Beverly, MA, USA), and amplified them by PCR with 10-12 cycles using the iProof™ High-Fidelity DNA Polymerase (BIO-RAD, Veenendaal, The Netherlands). We sequenced the libraries in single-read 151-bp lanes on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, ON, Canada).

We demultiplexed raw sequences using *process\_radtags*, a program distributed as part of the STACKS pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). We only retained reads with Phred scores  $\geq 10$  (using a sliding window of 15%), no adaptor

contamination, and unambiguous barcode and restriction cut sites. We checked read quality in FASTQC v.0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed sequences to 130 bp using SEQTK script (Heng Li, <https://github.com/lh3/seqtk>) in order to remove low-quality reads near the 3' ends. As an additional quality-filtering step, we used PYRAD v.3.0.66 (Eaton, 2014) to convert base calls with a Phred score <20 into Ns and discard reads with >2 Ns. We assembled retained reads into *de novo* loci using PYRAD, considering parameter values for clustering threshold of sequence similarity ( $W_{CLUST} = 0.85$ ), minimum coverage depth ( $d = 5$ ), maximum number of individuals with shared heterozygous sites ( $maxSH = p.10$ ), and maximum number of polymorphic sites in a final locus ( $maxSNPs = 20$ ) based on suggestions from the literature (Eaton, 2014; Eaton & Ree, 2013; Takahashi, Nagata, & Sota, 2014). Finally, we generated final datasets for subsequent analyses discarding loci that were not present in at least ~25 % of the samples ( $minCov = \sim 25\%$ ).

## 2.3 I Genetic structure and hybrid identification

We identified hybrids and introgressed individuals between *O. minutissimus* and *O. uhagonii* using the Bayesian clustering methods implemented in the programs FASTSTRUCTURE v.1.0 (Raj, Stephens, & Pritchard, 2014) and STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000). First, we used the highly efficient algorithm implemented in FASTSTRUCTURE to analyse the entire dataset including all populations. Second, we used both FASTSTRUCTURE and classic STRUCTURE to perform more detailed analyses focused on the Central System, the region where the ranges of the two taxa partly overlap, with some populations living in sympatry and, thus, the two species currently have the opportunity to hybridize. We ran FASTSTRUCTURE analyses using a simple prior, considering a convergence criterion of  $1 \times 10^{-7}$  and conducting 25 independent runs for each value of  $K$  (from  $K = 1$  to  $K = 10$ ). Following Raj et al. (2014), we used the *chooseK.py* script to assess model complexity by estimating the metrics  $K_{\phi}^*$ , the value of  $K$  that maximizes log-marginal likelihood lower bound (LLBO) of the data, and  $K_{\epsilon}^*$ , the smallest number of model components explaining at least 99% of cumulative ancestry contribution. We plotted individual co-ancestry coefficients for the most likely  $K$  value using DISTRUCT v.1.1 (Rosenberg, 2004). We ran STRUCTURE assuming correlated allele frequencies and admixture, and without using prior population information (Hubisz, Falush, Stephens, & Pritchard, 2009; Pritchard et al., 2000). We conducted 15 independent runs for each value

of  $K$  (from  $K = 1$  to  $K = 10$ ) to estimate the optimal number of genetic clusters with 200,000 MCMC cycles, following a burn-in step of 100,000 iterations. We used STRUCTURE HARVESTER (Earl & vonHoldt, 2012) to assess the number of genetic clusters that best describes our data according to log probabilities of the data ( $\text{LnPr}(X|K)$ ) for each value of  $K$  (Pritchard et al., 2000) and the  $\Delta K$  method (Evanno, Regnaut, & Goudet, 2005). We used CLUMPP v.1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same  $K$  value (Jakobsson & Rosenberg, 2007) and DISTRUCT to visualize as bar plots the individual's probabilities of population membership.

## 2.4 I Phylogenomic analyses and inference of historical hybridization

To determine the presence of historical hybridization (i.e., introgression), we employed four-taxon ABBA/BABA tests based on the  $D$ -statistic (Durand, Patterson, Reich, & Slatkin, 2011) and TREEMIX analyses (Pickrell & Pritchard, 2012). Assuming that the sister taxa P1 and P2 diverged from P3 and an outgroup species O, the  $D$ -statistic is used to test the null hypothesis of no introgression ( $D = 0$ ) between P3 and P1 or P2.  $D$  values significantly different from zero indicate gene flow between P1 and P3 ( $D < 0$ ) or between P2 and P3 ( $D > 0$ ). We assigned sympatric populations of *O. minutissimus* to P1 (64 individuals, from localities COVA, PESQ, NEGR, AVIL, PARA, GRED, PICO, SERR, MIJA, CASI, and MALA; hereafter, MS), allopatric populations of *O. minutissimus* to P2 (24 individuals, from localities MONT, PORT, ESPU, and TEJE; hereafter, MA), and populations of *O. uhagonii* to P3 (64 individuals from all sampling localities for this taxon; hereafter, US). Note that the sympatric and allopatric populations of *O. minutissimus* are located in the central and west portions of the species distribution range, respectively, and correspond to the two well-defined genetic clusters identified by Bayesian clustering analyses for this taxon (see Results section and Figure 1). We used as the outgroup (O) the taxon *O. antigai* (Bolívar, 1897), a species also belonging to the subgenus *Dreuxius* (Cigliano et al., 2019; García-Navas et al., 2017). Specifically, we used sequences from 94 individuals of this species available at NCBI Sequence Read Archive (SRA) under BioProject PRJNA543714 (Tonzo, Papadopoulou, & Ortego, 2019). We performed ABBA/BABA tests in PYRAD and used 1,000 bootstrap replicates to obtain the standard deviation of the  $D$ -statistic (Eaton & Ree, 2013; e.g., Huang, 2016).

We also analysed the potential presence of introgression and determined the direction

of gene flow with TREEMIX v.1.12 (Pickrell & Pritchard, 2012). We used TREEMIX to construct a tree-based model of population genetic relationships and infer events of genetic admixture using SNP frequency data. TREEMIX fits a population graph (i.e., a phylogenetic tree that incorporates admixture) on the basis of allele frequencies and a Gaussian approximation to genetic drift, allowing patterns of splits and mixtures in multiple populations to be inferred. To perform TREEMIX analyses, we pooled populations into the same three groups used to run ABBA/BABA tests. In a first step, we estimated a maximum-likelihood tree rooted with *O. antigai*. Then, we tested the existence of a range of migration events ( $m = 0$  to 5, with three replicated runs each) and calculated the proportion of the variance in population covariances explained by the population graph with different numbers of admixture events to determine the model best fitting the data (e.g., Gompert et al., 2014). We assumed the independence of all SNPs and used a window size of one SNP ( $k = 1$ ; e.g., Vera, Díez-del-Molino, & García-Marín, 2016).

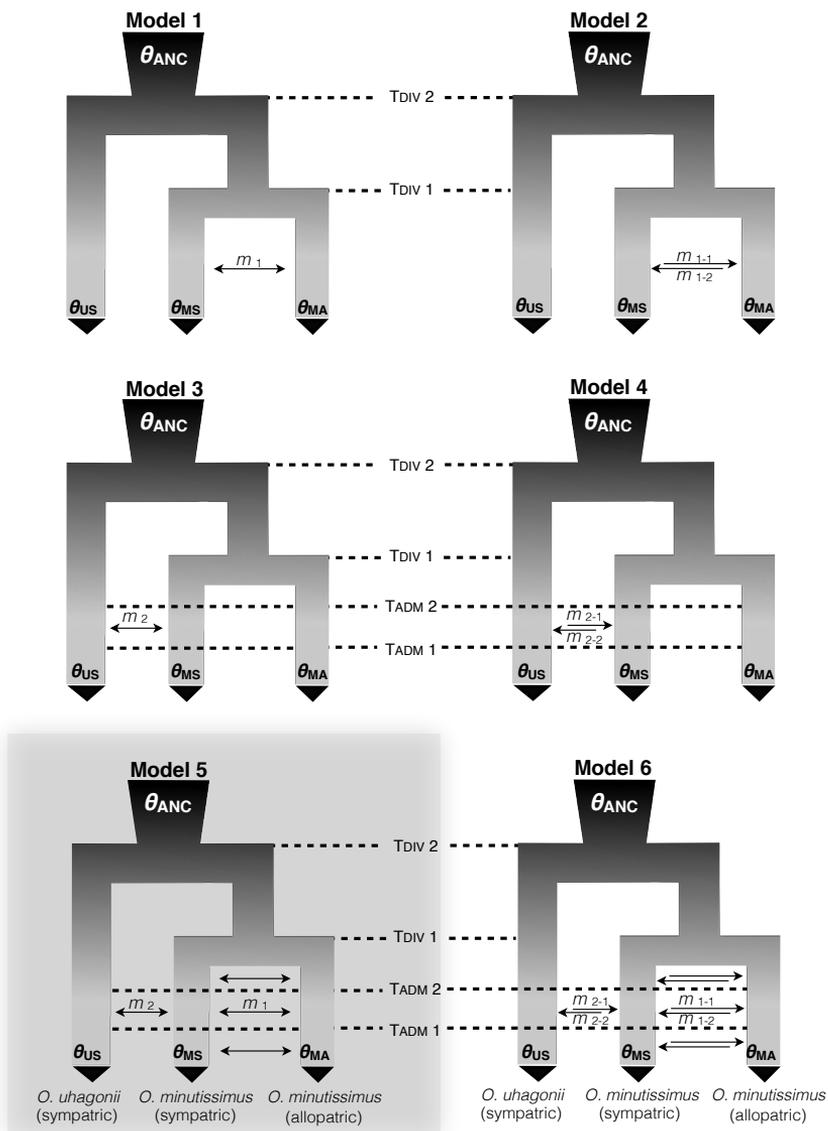
## 2.5 I Testing alternative models of gene flow

As TREEMIX only models migration as discrete events and does not consider continuous gene flow (Pickrell & Pritchard, 2012) we applied the coalescent-based modelling approach implemented in FASTSIMCOAL2 (Excoffier, Dupanloup, Huerta-Sanchez, Sousa, & Foll, 2013) to statistically test the relative fit of more complex historical demographic models to our genomic data. We used FASTSIMCOAL2 and the site frequency spectrum (SFS) (Excoffier et al., 2013) to test seven alternative models of gene flow (Figure 2). These models considered the same three population groups used in ABBA/BABA tests and TREEMIX analyses. All scenarios considered an early split between the two species followed by the divergence between sympatric (MS) and allopatric (MA) populations of *O. minutissimus*. The tested scenarios considered (i) total absence of post-divergence gene flow (Model 0, not shown in Figure 2); (ii) gene flow only between the two population groups of *O. minutissimus* (i.e., absence of interspecific gene flow) (Models 1-2); (iii) historical gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* (i.e., interspecific gene flow) but absence of post-divergence gene flow between the two population groups of *O. minutissimus* (Models 3-4); (iv) gene flow between the two population groups of *O. minutissimus* and historical gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* (Models 5-6) (Figure 2). These scenarios were tested considering both symmetric (Models 1, 3 and 5) and asymmetric (Models 2, 4

and 6) gene flow (Figure 2). It must be noted that detailed Bayesian clustering analyses performed across multiple populations of the two taxa in the region where their distribution ranges overlap did not show any evidence of ongoing interspecific gene flow (see Results section) and, for this reason, we did not consider models incorporating contemporary interspecific gene flow.

We calculated a folded joint SFS considering a single SNP per locus to avoid the effects of linkage disequilibrium. Because we did not include invariable sites in the SFS, we fixed the effective population size for one population group (US) to enable the estimation of other parameters in FASTSIMCOAL2 (e.g., Lanier et al., 2015; Papadopoulou & Knowles, 2015). The effective population size ( $N_e$ ) fixed in the models was calculated from the level of nucleotide diversity ( $\pi$ ) and estimates of mutation rate per site per generation ( $\mu$ ), according to the equation  $N_e = \pi/4\mu$  (Lynch & Conery, 2003). Nucleotide diversity for US ( $\pi = 0.005$ ) was estimated from polymorphic and non-polymorphic loci using DNASP v.6.11.01 (Librado & Rozas, 2009) and the *.allele* file generated by PYRAD. We considered the average mutation rate per site per generation of  $2.80 \times 10^{-9}$  estimated for *Drosophila melanogaster* (Keightley, Ness, Halligan, & Haddrill, 2014). To remove all missing data for the calculation of the joint SFS and minimize errors with allele frequency estimates, each population group was down sampled to 25% of individuals (32, 12 and 32 genes for MS, MA and US, respectively) using a custom Python script written by Isaac Overcast and available at GitHub (<https://github.com/isaacovercast/easySFS>). The final SFS contained 2,071 variable SNPs. Each of the seven models was run 100 replicated times using the computing resources provided by CESGA (Galician Supercomputer Center, Spain) considering 100,000–250,000 simulations for the calculation of the composite likelihood, 10–40 expectation-conditional maximization (ECM) cycles, and a stopping criterion of 0.001 (Papadopoulou & Knowles, 2015). We used an information-theoretic model selection approach based on the Akaike's information criterion (AIC) to determine the probability of each model given the observed data (Burnham & Anderson, 2002; e.g., Abascal et al., 2016; Thome & Carstens, 2016). After the maximum likelihood was estimated for each model in every replicate, we calculated the AIC scores as detailed in Thome & Carstens (2016). AIC values for each model were rescaled ( $\Delta AIC$ ) calculating the difference between the AIC value of each model and the minimum AIC obtained among all competing models (i.e., the best model has  $\Delta AIC = 0$ ). Point estimates of the different demographic parameters for the best-supported model were selected from the run with the highest

maximum composite likelihood. Finally, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier et al., 2013; e.g., Papadopoulou & Knowles, 2015).



**Figure 2** Alternative migration models tested using FASTSIMCOAL2. Parameters include ancestral ( $\theta_{ANC}$ ) and contemporary ( $\theta_{US}$ ,  $\theta_{MS}$ ,  $\theta_{MA}$ ) effective population sizes, timing of population split ( $T_{DIV}$ ) and admixture ( $T_{ADM}$ ), and migration rates ( $m$ ) between different pairs of populations. Grey background highlights the most supported model.

## 3 I RESULTS

### 3.1 I Genomic data

Illumina sequencing returned an average of  $2.74 \times 10^6$  reads per sample. After quality control, an average of  $2.33 \times 10^6$  reads per sample was retained (Figure S1). The data sets obtained with PYRAD for all populations and only those from the Central System retained 15,219 and 20,350 unlinked SNPs, respectively.

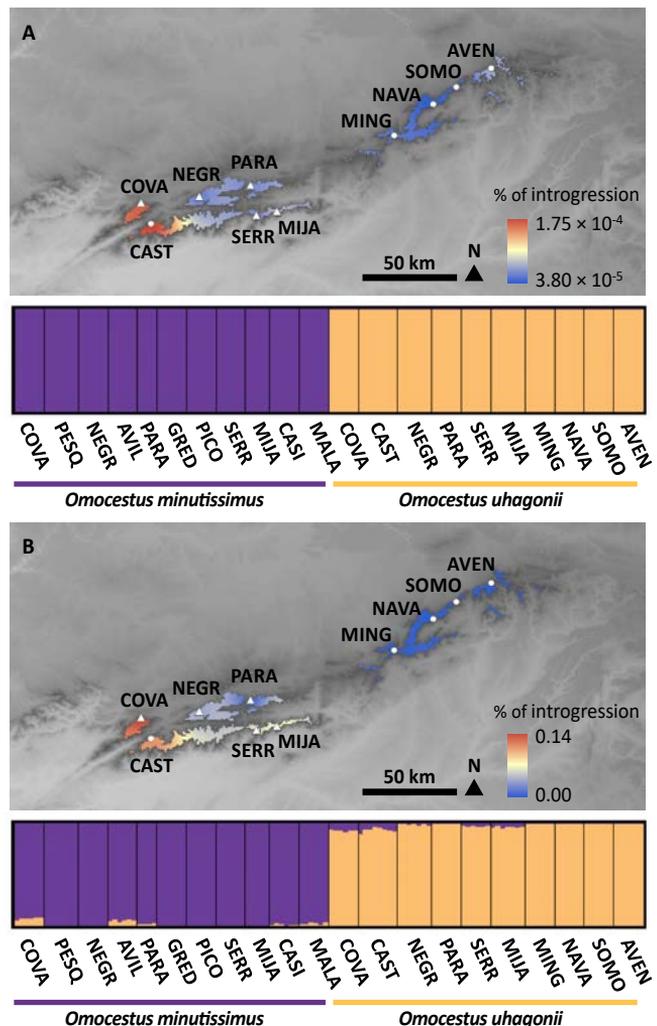
### 3.2 I Genetic structure and hybrid identification

The model complexity value that maximized the marginal likelihood in FASTSTRUCTURE analyses for the dataset including all populations was  $K = 3$  across all replicates and the number of model components used to explain structure in the data was  $K = 3$  in 15 replicates and  $K = 4$  in 10 replicates. FASTSTRUCTURE analyses for  $K = 3$  split populations of *O. uhagonii* and sympatric and allopatric populations of *O. minutissimus* in different genetic clusters in which all individuals showed high probabilities of population membership ( $q > 0.99$ ; Figure 1C). Assignment values to additional genetic clusters in FASTSTRUCTURE for analyses with  $K > 4$  were extremely low in all cases ( $q < 0.001$ ) and, thus, their respective bar plots were virtually identical to those obtained for  $K = 3$  (for a similar result, see Baiz, Tucker, & Cortes-Ortiz, 2019; Tonzo et al., 2019).

The model complexity value that maximized the marginal likelihood in FASTSTRUCTURE analyses of the Central System populations was equal to  $K = 2$  in all replicates and the number of model components used to explain structure in the data was  $K = 2$  in 16 replicates and  $K = 3$  in nine replicates. FASTSTRUCTURE analyses for  $K = 2$  split populations of the two species in different genetic clusters and all individuals showed high probabilities of population membership ( $q > 0.99$ ; Figure 3A). Again, assignment values to additional genetic clusters in FASTSTRUCTURE for analyses with  $K > 2$  were extremely low in all cases ( $q < 0.001$ ) and, thus, their respective bar plots were virtually identical to those obtained for  $K = 2$ . Classic STRUCTURE analyses also yielded an 'optimal' clustering value for  $K = 2$  according to the  $\Delta K$  criterion (Figure S2). The two inferred genetic groups supported a clear separation of the two species (Figure 3B). STRUCTURE analyses showed that populations of *O. uhagonii* from the eastern Central System present no signal of introgression from *O. minutissimus*. However, several populations of *O. minutissimus* and *O. uhagonii* from the western Central System, where the distribution of the two taxa overlap and five sampling

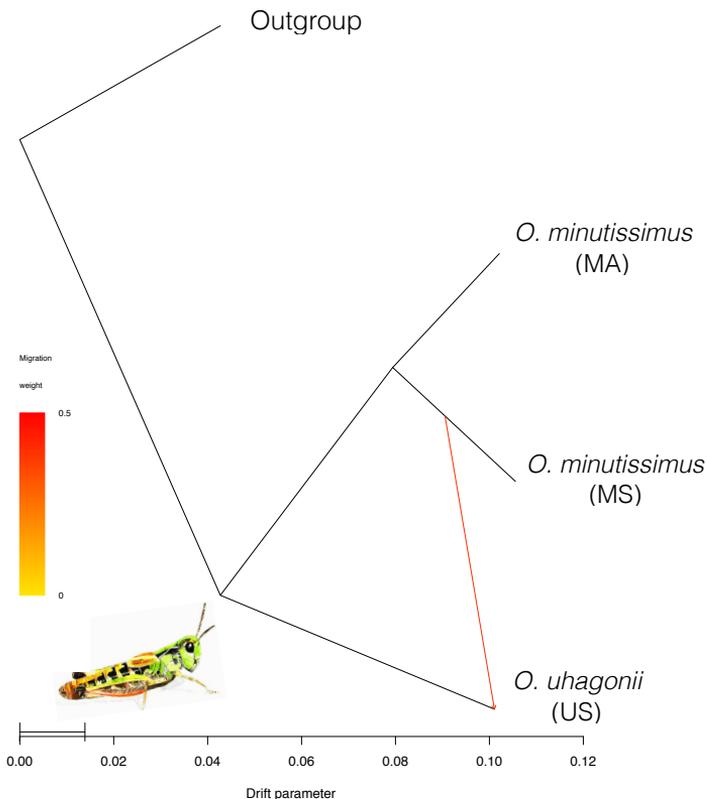
localities present sympatric populations, showed signals of reciprocal genetic introgression (Figure 3B). These results suggest that FASTSTRUCTURE is less likely to reveal small proportions of admixed ancestry in comparison with STRUCTURE, which has been also suggested in previous studies (Stift, Kolar, & Meirmans, 2019; Tonzon et al., 2019).

STRUCTURE analyses indicate that populations of *O. uhagonii* and *O. minutissimus* present significant differences in the degree of introgression from the other species (one-way ANOVAs, introgression of *O. minutissimus* into *O. uhagonii*:  $F_{9,54} = 48.86$ ,  $P < 0.001$ ; introgression of *O. uhagonii* into *O. minutissimus*:  $F_{10,53} = 96.68$ ,  $P < 0.001$ ; Figure 3B). Although visually imperceptible in the FASTSTRUCTURE bar plot (Figure 3A), probabilities of population membership inferred by this software also revealed that populations of both *O. uhagonii* and *O. minutissimus* present significant differences in the degree of introgression



**Figure 3** Genetic assignment of *O. minutissimus* and *O. uhagonii* from the Central System based on the results of (A) FASTSTRUCTURE and (B) STRUCTURE. Individuals are partitioned into  $K$  coloured segments representing the probability of belonging to the cluster with that colour and thin vertical black lines separate individuals from different populations. Maps display the distribution of *O. uhagonii* and the probability of assignment of the populations of this species (spatial interpolation) to the genetic cluster of *O. minutissimus* (i.e., the degree of introgression from *O. minutissimus* to *O. uhagonii*). Population codes as in Table S1.

from the other species (one-way ANOVAs, introgression of *O. minutissimus* into *O. uhagonii*:  $F_{9,54} = 14.01$ ,  $P < 0.001$ ; introgression of *O. uhagonii* into *O. minutissimus*:  $F_{10,53} = 3.19$ ,  $P = 0.003$ ; Figure 3A). The degree of introgression from *O. minutissimus* into *O. uhagonii* estimated by either FASTSTRUCTURE or STRUCTURE significantly decreased with longitude (FASTSTRUCTURE:  $F_{1,8} = 6.99$ ,  $P = 0.030$ ; STRUCTURE:  $F_{1,8} = 16.22$ ,  $P = 0.004$ ; see maps in Figure 3), but did not significantly differ between currently sympatric and allopatric populations (FASTSTRUCTURE:  $F_{1,8} = 0.06$ ,  $P = 0.81$ ; STRUCTURE:  $F_{1,8} = 1.64$ ,  $P = 0.236$ ). In contrast, the degree of introgression from *O. uhagonii* into *O. minutissimus* estimated by either FASTSTRUCTURE or STRUCTURE did not significantly decrease with longitude (FASTSTRUCTURE:  $F_{1,9} = 1.12$ ,  $P = 0.317$ ; STRUCTURE:  $F_{1,9} = 0.41$ ,  $P = 0.538$ ) or differ between currently sympatric and allopatric populations (FASTSTRUCTURE:  $F_{1,9} = 0.39$ ,  $P = 0.550$ ; STRUCTURE:  $F_{1,9} = 0.03$ ,  $P = 0.863$ ). For illustrative purposes, we displayed on a map the probabilities of assignment of the populations of *O. uhagonii* to the genetic cluster of *O. minutissimus* (i.e., the degree of introgression from *O. minutissimus* into *O. uhagonii*) by conducting a spatial interpolation using the Inverse Distance Weight (IDW) function available in ARCGIS v.10.5 (ESRI, Redlands, CA, USA) (Figure 3).



**Figure 4** Maximum-likelihood tree inferred with TREEMIX for *O. uhagonii* (US) and sympatric (MS) and allopatric (MA) populations of *O. minutissimus*. The direction of gene flow (from MS to US) for the most likely migration event ( $m = 1$ ) inferred is represented with an arrow colored according to the percentage of alleles (weight) originating from the source.

### 3.3 I Phylogenomic analyses and inference of historical hybridization

The test of introgression based on the  $D$ -statistic supported post-divergence gene flow between sympatric populations of *O. minutissimus* and *O. uhagonii* (BABA = 220.59; ABBA = 162.38;  $D$ -statistic = -0.152;  $Z = 3.39$ ;  $P < 0.001$ ). Accordingly, TREEMIX analyses supported a single migration event (Figure S3) corresponding to admixture between sympatric populations of *O. minutissimus* and *O. uhagonii* (Figure 4).

### 3.4 I Testing alternative models of gene flow

FASTSIMCOAL2 analyses showed that the most supported model (Model 5;  $\Delta AIC = 0$ ; Table 1) was the one considering symmetric interspecific gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* during a given period of time ( $T_{ADM1}$  and  $T_{ADM2}$ ) and symmetric gene flow between sympatric (central) and allopatric (eastern) populations of *O. minutissimus*. Remarkably, models only considering interspecific gene flow during a given period of time tended to be slightly more supported ( $\Delta AIC > 0.6$ ) than models only considering intraspecific gene flow between sympatric and allopatric populations of *O. minutissimus* (Table 1). Point estimates of demographic parameters under the best fitting model are presented in Table 2. Considering that the studied taxa are univoltine (i.e., 1-year generation time), FASTSIMCOAL2 analyses inferred that the separation of the two species ( $T_{DIV2}$ ) and the split of eastern and central populations of *O. minutissimus* ( $T_{DIV1}$ ) took place during the Early Pleistocene (Calabrian age) (Table 2). Gene flow between sympatric populations of *O. minutissimus* and *O. uhagonii* was estimated to happen during a period of time expanding  $\sim 15$  ka in the Late Pleistocene (Tarantian age) (Table 2). Note, however, that although confidence intervals around point estimates of most parameters were reasonably tight, there was considerable uncertainty around estimates for the two-time parameters delimiting the period of interspecific gene flow (i.e.,  $T_{ADM1}$  and  $T_{ADM2}$ ; Table 2). Finally, migration rates between sympatric populations of *O. uhagonii* and *O. minutissimus* ( $m_2$ ) did not significantly differ from those estimated between sympatric (western) and allopatric (eastern) populations of *O. minutissimus* ( $m_1$ ) (i.e., 95% CIs of  $m_1$  and  $m_2$  overlapped; Table 2). This indicates that historical interspecific gene flow was of the same order of magnitude as intraspecific gene flow between the two allopatric genetic clusters of *O. minutissimus*.

## 4 | DISCUSSION

Hybridization is a phenomenon that has been extensively documented in contact zones where the distributions of closely related species with weak reproductive barriers meet (e.g., Folk, Soltis, Soltis, & Guralnick, 2018; Gugger & Cavender-Bares, 2013; Nadeau et al., 2013; Ortego, Gugger, Riordan, & Sork, 2014). Inferring events of past interspecific gene flow has important implications to understand the evolutionary history of organisms (e.g., humans; Prüfer et al., 2014; Wall et al., 2013), yet, detecting the footprints of such processes is challenging (Payseur & Rieseberg, 2016; e.g., Eaton, Hipp, González-Rodríguez, & Cavender-Bares, 2015; Ortego et al., 2018). Here, by combining a suite of phylogenomic and population genetic tools and extensive population sampling across the

**Table 1** Comparison of alternative migration models (detailed in Figure 2) tested using FASTSIMCOAL2. For each model, the table shows the maximum likelihood estimate of the model ( $\log_{10}L$ ), the number of parameters ( $k$ ), the Akaike's information criterion score (AIC), the difference in AIC value of each model from that of the strongest model ( $\Delta AIC$ ), and AIC weight ( $\omega_i$ ). Best-supported model ( $\Delta AIC < 2$ ) is indicated in bold.

Model	$k$	$\log_{10}L$	AIC	$\Delta AIC$	$\omega_i$
Model 0	6	-3,539.32	7,090.64	90.01	0.00
Model 1	7	-3,518.85	7,051.71	51.08	0.00
Model 2	8	-3,518.42	7,052.84	52.21	0.00
Model 3	9	-3,515.49	7,048.98	48.34	0.00
Model 4	10	-3,515.54	7,051.09	50.45	0.00
<b>Model 5</b>	<b>10</b>	<b>-3,490.32</b>	<b>7,000.63</b>	<b>0.00</b>	<b>0.86</b>
Model 6	12	-3,490.09	7,004.19	3.56	0.14

entire distribution of our two focal species, including currently sympatric and allopatric populations, we found no evidence for contemporary hybridization. However, we did detect signals of past introgression in the geographical region where the distribution range of the two taxa currently overlap.

### 4.1 | Absence of contemporary interspecific gene flow

Bayesian clustering analyses showed a clear genotypic differentiation of the two species and further revealed the presence of two well-defined genetic clusters within *O. minutissimus*, corresponding with the populations of this taxon located in eastern

(allopatric) and central (sympatric) Iberia (Figure 1). Detailed analyses across 21 populations from central Iberia where the distribution of the two species partially overlap and, thus, they currently have the opportunity to hybridize, showed no evidence of ongoing interspecific gene flow (i.e., F1 or first generation backcrosses). However, these analyses also revealed footprints of reciprocal introgression in the westernmost portion of the Central System, the area where the two species present overlapping distributions and some populations even co-occur. Although the degree of introgression did not statistically differ between currently sympatric and allopatric populations of the two species in the Central System, it was not spatially homogeneous. On the one hand, the proportion of genetic introgression significantly differed across populations of the two species. On the other hand, the degree of introgression from *O. minutissimus* into *O. uhagonii* increased westwards and the populations from the easternmost portion of the distribution range of this species, where *O. minutissimus* is not currently present, showed negligible signals of past hybridization (Figure 3). These results indicate spatial heterogeneity in the levels of introgression, suggesting that the magnitude and/or timing of historical hybridization differed among populations of the two species in the Central System (e.g., de Manuel et al., 2016; Ortego et al., 2018; Wall et al., 2013). The degree of introgression was consistently small in all populations of both species (STRUCTURE: <9 %; FASTRUCTURE: <0.02 %) and similar across individuals within populations. Thus, although our sample sizes are modest (128 individuals) and we cannot categorically discard that the two species sporadically hybridize, the observed patterns of introgression indicate that contemporary populations are at genotypic equilibrium (i.e., backgrounds of introgression are similar across all individuals within a given population) and suggest that hypothetical contemporary hybridization, if it even happens, is unlikely to have transcended F1 hybrids at least in the last generations. This end is also supported by the fact that the five populations where the two species currently co-occur do not show higher levels of introgression than nearby allopatric populations, which points to the fact that the observed patterns of genetic introgression reflect historical rather than contemporary interspecific gene flow.

## 4.2 I Inferring historical hybridization

Both the phylogenomic analyses in TREEMIX and the *D*-statistic test yielded results compatible with those inferred by Bayesian clustering analyses and supported the

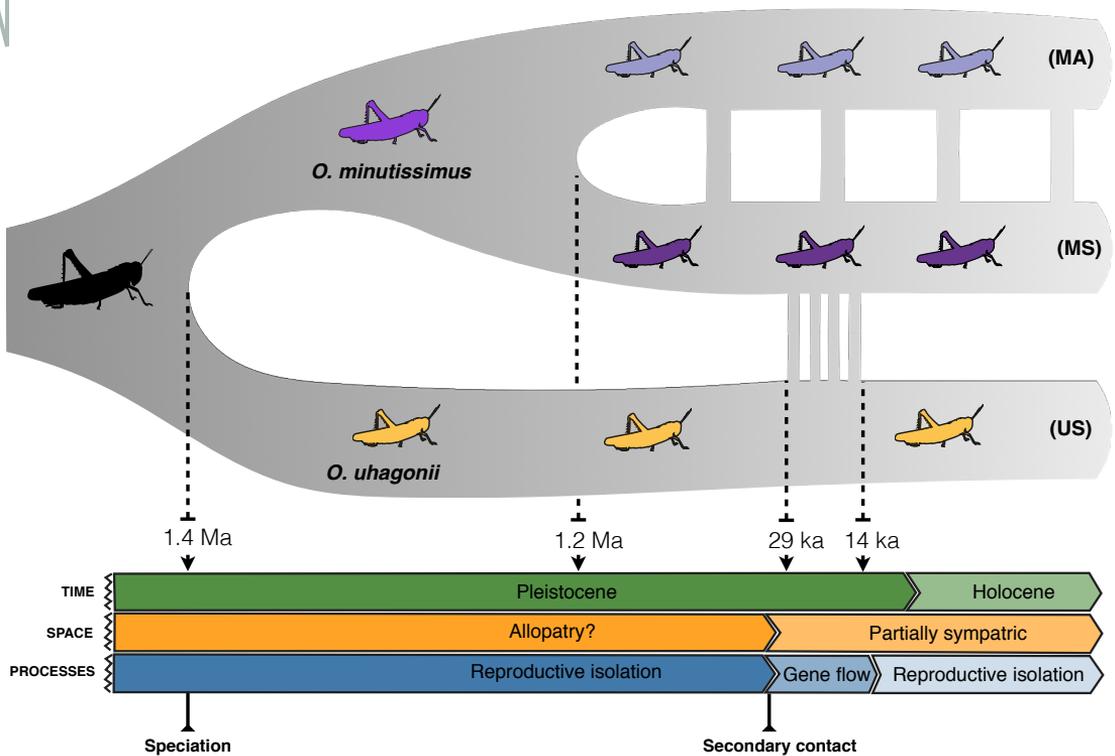
hypothesis of historical hybridization between *O. uhagonii* and geographically overlapping populations of *O. minutissimus*. TREEMIX identifies gene flow in the context of competing hypotheses, taking into account the full sampled phylogeny when inferring admixture and introgression events (Pickrell & Pritchard, 2012). Even when we allowed several admixture edges for the two species, TREEMIX only inferred one event of introgression involving the sympatric populations of the two species. In agreement with the TREEMIX results, the *D*-statistic analysis detected the same pattern of interspecific gene flow. Bringing the TREEMIX and the *D*-statistic results together support a scenario of historical hybridization and enhance the interpretation of the low levels of genetic introgression revealed by clustering analyses in currently coexisting populations.

Coalescent-based analyses in FASTSIMCOAL2, which provides detailed estimation of demographic parameters, further supported a scenario of past interspecific gene flow over a strictly bifurcating evolutionary history. Specifically, the most supported model was the one considering interspecific gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* during a given period of time and gene flow between sympatric (central) and allopatric (eastern) populations of *O. minutissimus*. The preferred model traced back the divergence of *O. uhagonii* and *O. minutissimus* to the Early Pleistocene (~1.4 Ma) and this event was followed shortly after by the split of eastern and central lineages of *O. minutissimus* (~1.2 Ma). These divergence times agree with the Pleistocene diversification observed across most clades within the highly speciose acridid subfamily Gomphocerinae (Song et al., 2015). Additionally, the split time estimates are congruent with the crown age (< 3 Ma) inferred for the recent radiation of the subgenus *Dreixius* based on mitochondrial DNA (García-Navas et al., 2017). The results of coalescent analyses also indicate that gene flow between *O. uhagonii* and *O. minutissimus* in the area of geographical overlap happened during a limited amount of time (~15,000 years) around the last glacial maximum (14-29 ka). During this period the two species engaged in gene flow at the same rate as the one estimated between the two lineages of *O. minutissimus*, indicating that interspecific gene flow, albeit low, was comparatively remarkable. The estimated timing of interspecific gene flow is compatible with the likely expansion of *O. uhagonii* to lower elevations during glacial periods. The shift to lower elevations might have put this species into extensive geographic contact with the more ubiquitous *O. minutissimus*, which presents many populations in foothills and valley bottoms where *O. uhagonii* is not present nowadays. It must be noted, however, that confidence intervals around point estimates for

the time of interspecific gene flow are ample (Table 2), particularly for the onset of this period ( $T_{ADM2}$ ), and thus these results must be interpreted with extreme caution. Uncertainty in these estimates could in part be driven by heterogeneity in the timing and extent of hybridization among the different populations in the sympatric area, as suggested by the significant differences in the signal of genetic introgression observed among populations (see previous section).

### 4.3 I Inferred evolutionary scenario

Our genomic analyses point to a scenario in which *O. uhagonii* and ancestral populations of *O. minutissimus* likely diverged in allopatry followed by the split of *O. minutissimus* into two lineages, one of which came into secondary contact and hybridized with *O. uhagonii* during a limited period of time (Figure 5). Despite the fact that nowadays the two species have ample opportunity to hybridize (i.e., a large proportion their respective ranges currently overlap and several populations even co-occur in the Central System), our analyses did not find any evidence of contemporary hybridization, which suggests that reproductive isolation likely evolved after secondary contact and historical gene flow (Figure 5). Speciation of Gomphocerinae and other grasshoppers has been generally linked to allopatric divergence (Mayer et al. 2010), a process that in the specific case of montane/alpine species of Pleistocene origin was probably caused by the extensive fragmentation of ancestral populations driven by Quaternary climatic oscillations (e.g., Huang, Hill, Ortego, & Knowles, 2020; Knowles, 2000; Scattolini, Confalonieri, Lira-Noriega, Petrokovsky, & Cigliano, 2018). Different lines of evidence also point to allopatric divergence as the most plausible mode of speciation for *O. uhagonii* and *O. minutissimus*. First, the allopatric lineage of *O. minutissimus* from eastern Iberia present a much larger distribution (Figure 1) and significantly higher levels of genomic diversity (one-way ANOVA:  $F_{1,13} = 12.05$ ,  $P = 0.004$ ; Table S1) and effective population sizes (non-overlapping 95% CIs for  $N_e$  estimates in FASTSIMCOAL2 analyses; Table 2) than the sympatric lineage of *O. minutissimus* from the Central System. This supports that *O. minutissimus* most likely originated in eastern Iberia and subsequently colonized the Central System, where it came into secondary contact with *O. uhagonii*. Second, the ecological niches of *O. minutissimus* and *O. uhagonii* are very similar and the two species present graminivorous feeding habits and same microhabitat preferences (Clemente et al., 1991; J. Ortego, personal observation). This points to considerable niche conservatism, typical of allopatric



**Figure 5** Schematic representation of the events documented in this study and the inferred biological processes. These correspond to the best-fit demographic model (Model 5) for *O. uhagonii* (US) and sympatric (MS) and allopatric (MA) populations of *O. minutissimus*. Vertical bars connecting MS and US represent historical gene flow. Note that geological reference time is not scaled and only point estimates inferred by FASTSIMCOAL2 are presented to simplify visualization.

speciation, and rejects sympatric speciation via disruptive ecological selection (e.g., Grace, Wisely, Brown, Dowell, & Joern, 2010). Finally, our analyses showed no evidence of gene flow between *O. uhagonii* and ancestral populations of *O. minutissimus* (Figure 4), indicating that the two species likely evolved in allopatry and only exchanged gene flow after secondary contact in the region where their distribution ranges overlap (see also Sankararaman et al., 2014; Sankararaman, Patterson, Li, Paabo, & Reich, 2012).

The footprints of historical introgression among currently sympatric populations of *O. uhagonii* and *O. minutissimus* and the lack of evidence for contemporary gene flow lead us to hypothesize an evolutionary scenario in which reproductive isolation evolved after historical hybridization in the area where the distribution ranges of the two taxa currently overlap. The observed differences in the levels of introgression among sympatric

populations suggest that barriers to gene flow might have evolved multiple times or, alternatively, could reflect heterogeneity in the proportion of the genome of the other species retained after the interruption of interspecific gene flow due to differences among the studied populations in their demographic histories (e.g., bottlenecks; Amorim et al., 2017; Lawson, Van Dorp, & Falush, 2018; Quilodran, Nussberger, Montoya-Burgos, & Currat, 2019) or spatial variation in the strength of hypothetical purifying selection acting against introgressed alleles (Juric, Aeschbacher, & Coop, 2016; Petr, Paabo, Kelso, & Vernot, 2019). Alternative processes might have led to the evolution of reproductive isolation after secondary contact and hybridization. It has been frequently documented that interspecific gene flow can increase phenotypic and genomic divergence via the evolution of reproductive isolation and character displacement (Garner, Goulet, Farnitano, Molina-Henao, & Hopkins, 2018; Hopkins, Levin, & Rausher, 2012; Pfennig & Pfennig, 2009). One of the many potential costs of hybridization are the ecological and genetic dysfunctions of hybrid offspring, which can reduce their fitness and drive to reinforcement (Ortiz-Barrientos, Counterman, & Noor, 2004). In the reinforcement process, enhanced prezygotic isolation is favoured in sympatry in response to postzygotic isolation due to a strong selection against hybrids (Butlin, 1995; Coyne & Orr, 2004; Servedio & Noor, 2003).

**Table 2** Parameters inferred from coalescent simulations with FASTSIMCOAL2 under the most supported demographic model (Model 5). Table shows point estimates and lower and upper 95% confidence intervals. Note that the effective population size of *O. uhagonii* is not presented in this table because it was fixed in FASTSIMCOAL2 analyses to enable the estimation of other parameters (see the Materials and Methods section for further details).  $\theta$ , mutation-scaled effective population sizes;  $T_{DIV}$  and  $T_{ADM}$ , timing of population divergence and admixture, respectively (given in number of generations);  $m$ , migration rates per generation. Each specific parameter is illustrated in Figure 2.

Parameter	Point estimate	Lower Bound	Upper Bound
$\theta_{ANC}$	702,036	376,562	832,461
$\theta_{MS}$	813,263	686,748	890,000
$\theta_{MA}$	1,556,235	1,293,951	1,675,243
$T_{DIV1}$	1,199,191	922,762	1,380,088
$T_{DIV2}$	1,382,958	1,278,828	1,670,203
$T_{ADM1}$	14,090	1,632	23,626
$T_{ADM2}$	29,040	50,260	750,959
$m_1$	$1.28 \times 10^{-07}$	$9.02 \times 10^{-08}$	$1.82 \times 10^{-07}$
$m_2$	$2.41 \times 10^{-07}$	$2.14 \times 10^{-08}$	$1.37 \times 10^{-07}$

Selection for prezygotic isolation leads, in turn, to more divergent phenotypes and reproductive behaviours between species in sympatry than in allopatry (Moran, Zhou, Catchen, & Fuller, 2018). Thus, one possibility is that secondary contact and hybridization promoted the evolution of reproductive isolation via reinforcement, probably after an initial balance between dispersal and selection against hybrids in historical tension zones during which the two species experienced genetic exchange and introgression (Barton & Hewitt, 1985). An alternative explanation is that reproductive isolation evolved in geographical isolation as a consequence of genetic drift or as a fortuitous by-product of divergent selection on other traits (Coyne & Orr, 1989; Fitzpatrick, 2002; Sasa, Chippindale, & Johnson, 1998). The mosaic distribution of the two species in the Central System, with the presence of several sympatric populations but also large areas where the two species do not occur (e.g., eastern Central System and foothills; Figure 1), might have also provided ample opportunity for the evolution of reproductive isolation in geographically separated populations (Fitzpatrick, 2002).

#### **4.4 I Conclusions and future directions**

The results of this study add to the growing body of evidence supporting that speciation-with-gene-flow is more prevalent in nature than formerly acknowledged (Nosil, 2008; Pinho & Hey, 2010; Roux et al., 2016). Our study system is very well-suited to study the proximate mechanisms (e.g., reinforcement vs. genetic drift) that might have led to the evolution of reproductive isolation. Future research should focus on analysing mating preferences, phenotypic differentiation and reproductive character displacement (song, courtship behaviour, genitalic structures, etc.) between currently sympatric and allopatric populations of the two species (e.g., Butlin et al., 1991; Hollander, Smadja, Butlin, & Reid, 2013), determining mating success and the viability of offspring through experimental hybridization attempts in the laboratory (e.g., Coyne & Orr, 1989; Hoskin et al., 2005; Saldamando et al., 2005) and, applying whole genome or transcriptome sequencing data to detect potential genomic signatures of reinforcement and/or identify loci that might be involved in reproductive isolation (Garner et al., 2018; Hopkins et al., 2012; Roda, Mendes, Hahn, & Hopkins, 2017).

## ACKNOWLEDGEMENTS

We are grateful to Amparo Hidalgo-Galiana for her valuable help during laboratory work and the grasshopper illustration. We also would like to thank to Víctor Noguerales, Conchi Cáliz, and Pedro J. Cordero for their help during field and laboratory work, Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing, and two anonymous referees for constructive comments on an earlier version of this article. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) and Laboratorio de Sistemas de Información Geográfica y Teledetección (LAST-EBD) from Estación Biológica de Doñana. We also thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. This study was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (CGL2014-54671-P and CGL2017-83433-P). VT was supported by an FPI predoctoral fellowship (BES-2015-73159) from Spanish Ministry of Economy and Competitiveness. During this work, AP and JO were supported by a Severo Ochoa (SEV-2012-0262) and a Ramón y Cajal (RYC-2013-12501) research fellowship, respectively.

## AUTHOR CONTRIBUTIONS

V.T., A.P., and J.O. conceived and designed the study and analyses. J.O. collected the samples. V.T. performed the laboratory work and analysed the data guided by J.O. V.T. wrote the manuscript with help of J.O., and inputs from A.P.

## REFERENCES

- Abascal, F., Corvelo, A., Cruz, F., Villanueva-Canas, J. L., Vlasova, A., Marcet-Houben, M., . . . Godoy, J. A. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17, 251. doi:10.1186/s13059-016-1090-1
- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., . . . Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229-246. doi:10.1111/j.1420-9101.2012.02599.x

- Amorim, C. E. G., Hofer, T., Ray, N., Foll, M., Ruiz-Linares, A., & Excoffier, L. (2017). Long-distance dispersal suppresses introgression of local alleles during range expansions. *Heredity*, *118*(2), 135-142. doi:10.1038/hdy.2016.68
- Baiz, M. D., Tucker, P. K., & Cortes-Ortiz, L. (2019). Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular Ecology*, *28*(5), 1056-1069. doi:10.1111/mec.14966
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113-148. doi:10.1146/annurev.ecolsys.16.1.113
- Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497-503. doi:10.1038/341497a0
- Bridle, J. R., Baird, S. J. E., & Butlin, R. K. (2001). Spatial structure and habitat variation in a grasshopper hybrid zone. *Evolution*, *55*(9), 1832-1843. doi:10.1554/0014-3820(2001)055[1832:ssahvi]2.0.co;2
- Britch, S. C., Cain, M. L., & Howard, D. J. (2001). Spatio-temporal dynamics of the *Allonemobius fasciatus*-*A. socius* mosaic hybrid zone: a 14-year perspective. *Molecular Ecology*, *10*(3), 627-638. doi:10.1046/j.1365-294x.2001.01215.x
- Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*(3), 301-312. doi:10.1038/sj.hdy.6800997
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, USA: Springer.
- Butlin, R. K. (1995). Reinforcement - An idea evolving. *Trends in Ecology & Evolution*, *10*(11), 432-434. doi:10.1016/s0169-5347(00)89173-9
- Butlin, R. K., Ritchie, M. G., & Hewitt, G. M. (1991). Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera, Acrididae). *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *334*(1271), 297-308. doi:10.1098/rstb.1991.0119
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. doi:10.1111/mec.12354
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2019). *Orthoptera Species File*. Version 5.0/5.0. Retrieved from <http://orthoptera.speciesfile.org/>
- Clemente, M. E., Garcia, M. D., & Presa, J. J. (1991). Los Gomphocerinae de la península Iberica: 2. *Omocestus* Bolivar, 1878. (Insecta, Orthoptera, Caelifera). *Graellsia*, *46*, 191-246.
- Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*, *43*(2), 362-381. doi:10.2307/2409213
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Association.

- de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V. C., Desai, T., Prado-Martinez, J., . . . Marques-Bonet, T. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, *354*(6311), 477-481. doi:10.1126/science.aag2602
- Dobzhansky, T. (1937). Genetic nature of species differences. *American Naturalist*, *71*, 404-420. doi:10.1086/280726
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239-2252. doi:10.1093/molbev/msr048
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361. doi:10.1007/s12686-011-9548-7
- Eaton, D. A. R. (2014). PYRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844-1849. doi:10.1093/bioinformatics/btu121
- Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, *69*(10), 2587-2601. doi:10.1111/evo.12758
- Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis: Orobanchaceae*). *Systematic Biology*, *62*(5), 689-706. doi:10.1093/sysbio/syt032
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. doi:10.1371/journal.pgen.1003905
- Fitzpatrick, B. M. (2002). Molecular correlates of reproductive isolation. *Evolution*, *56*(1), 191-198. doi:10.1111/j.0014-3820.2002.tb00860.x
- Fitzpatrick, B. M., Fordyce, J. A., & Gavrillets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, *22*(11), 2342-2347. doi:10.1111/j.1420-9101.2009.01833.x
- Folk, R. A., Soltis, P. S., Soltis, D. E., & Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*, *105*(3), 364-375. doi:10.1002/ajb2.1018
- García-Navas, V., Noguerales, V., Cordero, P. J., & Ortego, J. (2017). Ecological drivers of body size evolution and sexual size dimorphism in short-horned grasshoppers (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, *30*(8), 1592-1608. doi:10.1111/jeb.13131
- Garner, A. G., Goulet, B. E., Farnitano, M. C., Molina-Henao, Y. F., & Hopkins, R. (2018). Genomic signatures of reinforcement. *Genes*, *9*(4), 191. doi:10.3390/genes9040191

- Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*(18), 4555-4573. doi:10.1111/mec.12811
- Grace, T., Wisely, S. M., Brown, S. J., Dowell, F. E., & Joern, A. (2010). Divergent host plant adaptation drives the evolution of sexual isolation in the grasshopper *Hesperotettix viridis* (Orthoptera: Acrididae) in the absence of reinforcement. *Biological Journal of the Linnean Society*, *100*(4), 866-878.
- Graham, C. H., Ron, S. R., Santos, J. C., Schneider, C. J., & Moritz, C. (2004). Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, *58*(8), 1781-1793. doi:10.1554/03-274
- Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F., & Petren, K. (2004). Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, *58*(7), 1588-1599.
- Gugger, P. F., & Cavender-Bares, J. (2013). Molecular and morphological support for a Florida origin of the Cuban oak. *Journal of Biogeography*, *40*(4), 632-645. doi:10.1111/j.1365-2699.2011.02610.x
- Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford Surveys of Evolutionary Biology*, *7*, 69-128.
- Hewitt, G. M. (1988). Hybrid zones - Natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, *3*(7), 158-167. doi:10.1016/0169-5347(88)90033-x
- Hill, J. G. (2015). Revision and biogeography of the *Melanoplus scudderii* species group (Orthoptera: Acrididae: Melanoplinae) with a description of 21 new species and establishment of the *Carnegiei* and *Davisi* species groups. *Transactions of the American Entomological Society*, *141*(2), 252-350. doi:10.3157/061.141.0201
- Hollander, J., Smadja, C. M., Butlin, R. K., & Reid, D. G. (2013). Genital divergence in sympatric sister snails. *Journal of Evolutionary Biology*, *26*(1), 210-215. doi:10.1111/jeb.12029
- Hopkins, R., Levin, D. A., & Rausher, M. D. (2012). Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution*, *66*(2), 469-485. doi:10.1111/j.1558-5646.2011.01452.x
- Hoskin, C. J., Higgie, M., McDonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid allopatric speciation. *Nature*, *437*(7063), 1353-1356. doi:10.1038/nature04004
- Howard, D. J., Waring, G. L., Tibbets, C. A., & Gregory, P. G. (1993). Survival of hybrids in a mosaic hybrid zone. *Evolution*, *47*(3), 789-800. doi:10.2307/2410184
- Huang, J. P. (2016). Parapatric genetic introgression and phenotypic assimilation: testing conditions for introgression between Hercules beetles (Dynastes, Dynastinae). *Molecular Ecology*, *25*(21), 5513-5526. doi:10.1111/mec.13849
- Huang, J. P., Hill, J. G., Ortego, J., & Knowles, L. L. (2020) Paraphyletic species no more - genomic data resolve a Pleistocene radiation and validate morphological species of the *Melanoplus scudderii* complex (Insecta: Orthoptera). *Systematic Entomology*, in press. doi:10.1111/syen.12415

- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332. doi:10.1111/j.1755-0998.2009.02591.x
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806. doi:10.1093/bioinformatics/btm233
- Juric, I., Aeschbacher, S., & Coop, G. (2016). The strength of selection against Neanderthal introgression. *PLoS Genetics*, 12(11), e1006340. doi:10.1371/journal.pgen.1006340
- Kearns, A. M., Restani, M., Szabo, I., Schroder-Nielsen, A., Kim, J. A., Richardson, H. M., . . . Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9, 906. doi:10.1038/s41467-018-03294-w
- Keightley, P. D., Ness, R. W., Halligan, D. L., & Hadrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, 196(1), 313-320. doi:10.1534/genetics.113.158758
- Key, K. H. L. (1968). Concept of stasipatric speciation. *Systematic Zoology*, 17(1), 14-22. doi:10.2307/2412391
- Knowles, L. L. (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54(4), 1337-1348.
- Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in collared pikas (*Ochotona collaris*). *Molecular Ecology*, 24(14), 3688-3705. doi:10.1111/mec.13270
- Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9, 3258. doi:10.1038/s41467-018-05257-7
- Lemmon, E. M., & Juenger, T. E. (2017). Geographic variation in hybridization across a reinforcement contact zone of chorus frogs (*Pseudacris*). *Ecology and Evolution*, 7(22), 9485-9502. doi:10.1002/ece3.3443
- Librado, P., & Rozas, J. (2009). DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452. doi:10.1093/bioinformatics/btp187
- Lohse, K., Clarke, M., Ritchie, M. G., & Etges, W. J. (2015). Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution*, 69(5), 1178-1190. doi:10.1111/evo.12650
- Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science*, 302(5649), 1401-1404. doi:10.1126/science.1089370
- Maguilla, E., Escudero, M., Hipp, A. L., & Luceno, M. (2017). Allopatric speciation despite historical gene flow: Divergence and hybridization in *Carex furva* and *C. lucennoiberica* (Cyperaceae) inferred from plastid and nuclear RAD-seq data. *Molecular Ecology*, 26(20), 5646-5662. doi:10.1111/mec.14253
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229-237.

- Marques, I., Draper, D., Riofrio, L., & Naranjo, C. (2014). Multiple hybridization events, polyploidy and low postmating isolation entangle the evolution of neotropical species of *Epidendrum* (Orchidaceae). *BMC Evolutionary Biology*, *14*, 20. doi:10.1186/1471-2148-14-20
- Mayer, F., Berger, D., Gottsberger, B., & Schulze, W. (2010). Non-ecological radiations in acoustically communicating grasshoppers? In M. Glaubrecht (Ed.), *Evolution in Action* (pp. 451-464). Berlin, Germany: Springer-Verlag.
- Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, USA: Harvard University Press.
- Moran, R. L., Zhou, M. C., Catchen, J. M., & Fuller, R. C. (2018). Hybridization and postzygotic isolation promote reinforcement of male mating preferences in a diverse group of fishes with traditional sex roles. *Ecology and Evolution*, *8*(18), 9282-9294. doi:10.1002/ece3.4434
- Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., . . . Jiggins, C. D. (2013). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, *22*(3), 814-826. doi:10.1111/j.1365-294X.2012.05730.x
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*(9), 2103-2106. doi:10.1111/j.1365-294X.2008.03715.x
- Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography*, *41*(10), 1895-1908. doi:10.1111/jbi.12334
- Ortego, J., Gugger, P. F., & Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist*, *218*(2), 804-818. doi:10.1111/nph.14951
- Ortiz-Barrientos, D., Counterman, B. A., & Noor, M. A. F. (2004). The genetics of speciation by reinforcement. *PLoS Biology*, *2*(12), 2256-2263. doi:10.1371/journal.pbio.0020416
- Papadopoulou, A., & Knowles, L. L. (2015). Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution*, *69*(6), 1501-1517. doi:10.1111/evo.12667
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337-2360. doi:10.1111/mec.13557
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. doi:10.1371/journal.pone.0037135
- Petr, M., Paabo, S., Kelso, J., & Vernet, B. (2019). Limits of long-term selection against Neandertal introgression. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(5), 1639-1644. doi:10.1073/pnas.1814338116
- Pfennig, K. S., & Pfennig, D. W. (2009). Character displacement: Ecological and reproductive responses to a common evolutionary problem. *Quarterly Review of Biology*, *84*(3), 253-276. doi:10.1086/605079

- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. doi:10.1371/journal.pgen.1002967
- Pinho, C., & Hey, J. (2010). Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 215-230.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., . . . Paabo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43-49. doi:10.1038/nature12886
- Quilodran, C. S., Nussberger, B., Montoya-Burgos, J. I., & Currat, M. (2019). Hybridization and introgression during density-dependent range expansion: European wildcats as a case study. *Evolution*, *73*(4), 750-761. doi:10.1111/evo.13704
- Ragge, D. R., & Reynolds, W. J. (1998). *The Songs of the Grasshoppers and Crickets of Western Europe*. Essex, UK: Harley Books.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573-589. doi:10.1534/genetics.114.164350
- Roda, F., Mendes, F. K., Hahn, M. W., & Hopkins, R. (2017). Genomic evidence of gene flow during reinforcement in Texas Phlox. *Molecular Ecology*, *26*(8), 2317-2330. doi:10.1111/mec.14041
- Rohde, K., Hau, Y., Kranz, N., Weinberger, J., Elle, O., & Hochkirch, A. (2017). Climatic effects on population declines of a rare wetland species and the role of spatial and temporal isolation as barriers to hybridization. *Functional Ecology*, *31*(6), 1262-1274. doi:10.1111/1365-2435.12834
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x
- Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, *14*(12), e2000234. doi:10.1371/journal.pbio.2000234
- Saldamando, C. I., Tatsuta, H., & Butlin, R. K. (2005). Hybrids between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae) do not show endogenous postzygotic isolation. *Biological Journal of the Linnean Society*, *84*(2), 195-203. doi:10.1111/j.1095-8312.2005.000424.x
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Paabo, S., . . . Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354-357. doi:10.1038/nature12961
- Sankararaman, S., Patterson, N., Li, H., Paabo, S., & Reich, D. (2012). The date of interbreeding between Neandertals and modern humans. *PLoS Genetics*, *8*(10), e1002947. doi:10.1371/journal.pgen.1002947

- Sasa, M. M., Chippindale, P. T., & Johnson, N. A. (1998). Patterns of postzygotic isolation in frogs. *Evolution*, 52(6), 1811-1820. doi:10.1111/j.1558-5646.1998.tb02258.x
- Scattolini, M. C., Confalonieri, V., Lira-Noriega, A., Pietrokovsky, S., & Cigliano, M. M. (2018). Diversification mechanisms in the Andean grasshopper genus *Orotettix* (Orthoptera: Acrididae): ecological niches and evolutionary history. *Biological Journal of the Linnean Society*, 123(4), 697-711. doi:10.1093/biolinnean/bly008
- Schenk, J. J., Kontur, S., Wilson, H., Noble, M., & Derryberry, E. (2018). Allopatric speciation drives diversification of ecological specialists on sandhills. *International Journal of Plant Sciences*, 179(4), 325-339. doi:10.1086/697073
- Schmitt, T. (2009). Biogeographical and evolutionary importance of the European high mountain systems. *Frontiers in Zoology*, 6, 9. doi:10.1186/1742-9994-6-9
- Seddon, J. M., Santucci, F., Reeve, N. J., & Hewitt, G. M. (2001). DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology*, 10(9), 2187-2198. doi:10.1046/j.0962-1083.2001.01357.x
- Servedio, M. R., & Kirkpatrick, M. (1997). The effects of gene flow on reinforcement. *Evolution*, 51(6), 1764-1772. doi:10.2307/2410999
- Servedio, M. R., & Noor, M. A. F. (2003). The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology Evolution and Systematics*, 34, 339-364. doi:10.1146/annurev.ecolsys.34.011802.132412
- Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, 60, 561-588. doi:10.1146/annurev.arplant.043008.092039
- Song, H. J., Amedegnato, C., Cigliano, M. M., Desutter-Grandcolas, L., Heads, S. W., Huang, Y., . . . Whiting, M. F. (2015). 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. *Cladistics*, 31(6), 621-651. doi:10.1111/cla.12116
- Stift, M., Kolar, F., & Meirmans, P. G. (2019). STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity*, 123(4), 429-441. doi:10.1038/s41437-019-0247-6
- Takahashi, T., Nagata, N., & Sota, T. (2014). Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution*, 80, 137-144. doi:10.1016/j.ympev.2014.07.016
- Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2), 343-355. doi:10.1111/j.1365-294X.2005.02794.x
- Thome, M. T. C., & Carstens, B. C. (2016). Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8010-8017. doi:10.1073/pnas.1601064113

- Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology*, *28*(17), 3869-3886. doi:10.1111/mec.15189
- Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, *28*(12), 696-704. doi:10.1016/j.tree.2013.09.001
- Vera, M., Díez-del-Molino, D., & García-Marín, J. L. (2016). Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). *Molecular Ecology*, *25*(5), 1089-1105. doi:10.1111/mec.13545
- Virdee, S. R., & Hewitt, G. M. (1994). Clines for hybrid dysfunction in a grasshopper hybrid zone. *Evolution*, *48*(2), 392-407. doi:10.2307/2410100
- Wachter, G. A., Papadopoulou, A., Muster, C., Arthofer, W., Knowles, L. L., Steiner, F. M., & Schlick-Steiner, B. C. (2016). Glacial refugia, recolonization patterns and diversification forces in Alpine-endemic *Megabunus* harvestmen. *Molecular Ecology*, *25*(12), 2904-2919. doi:10.1111/mec.13634
- Wall, J. D., Yang, M. A., Jay, F., Kim, S. K., Durand, E. Y., Stevison, L. S., . . . Slatkin, M. (2013). Higher levels of Neanderthal ancestry in east Asians than in Europeans. *Genetics*, *194*(1), 199-209. doi:10.1534/genetics.112.148213

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article and in Appendix III of this Thesis.



SCAN ME TO SEE ARTICLE ONLINE



Tonzo, V., Papadopoulou, A., & Ortego, J. (2020).

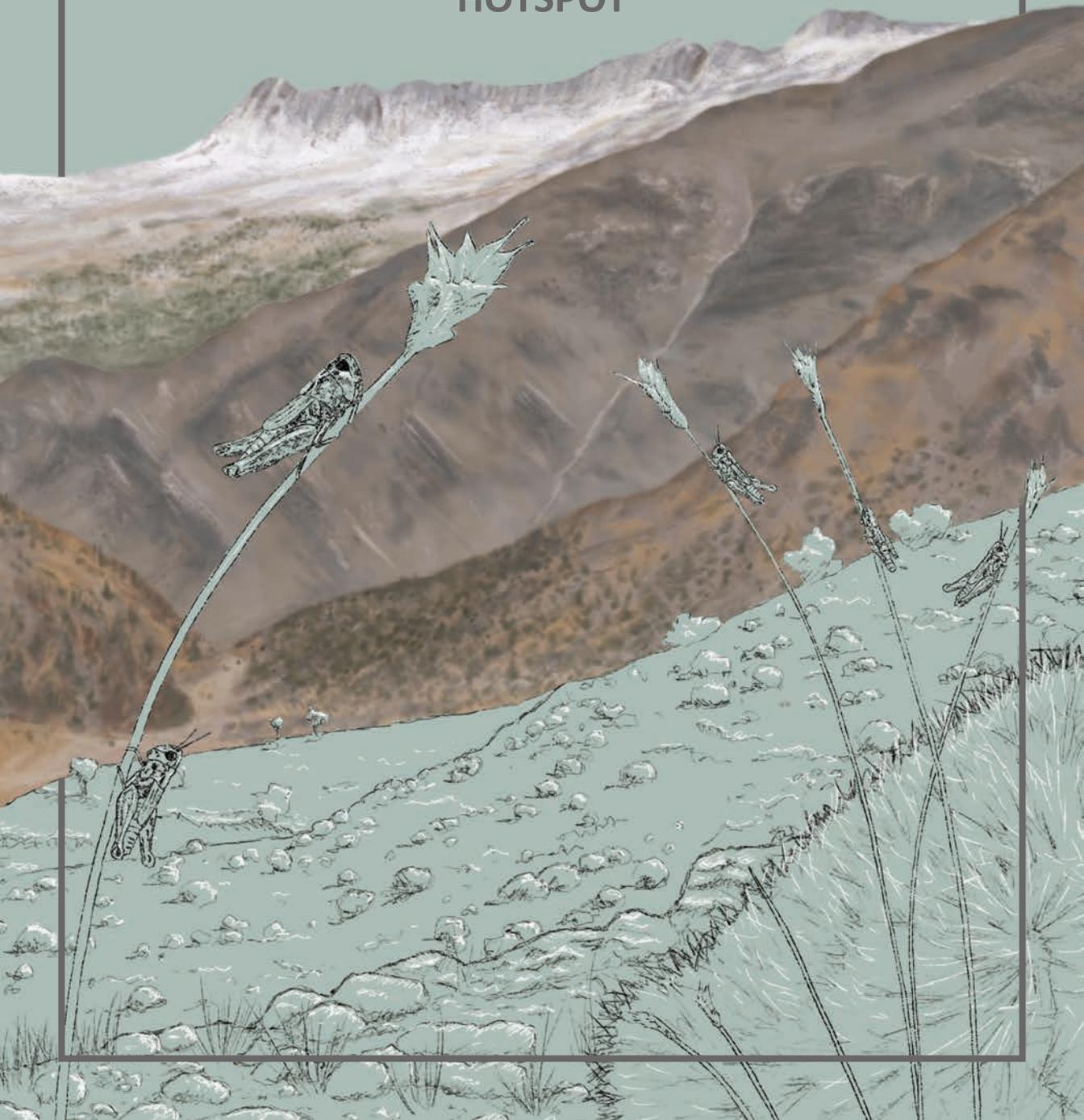
Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions.

*Molecular Ecology. (In Press).*



## CHAPTER IV

# GLACIAL CONNECTIVITY AND CURRENT POPULATION FRAGMENTATION IN SKY-ISLANDS EXPLAIN THE CONTEMPORARY DISTRIBUTION OF GENOMIC VARIATION IN TWO NARROW-ENDEMIC MONTANE GRASSHOPPERS FROM A BIODIVERSITY HOTSPOT





# Glacial connectivity and current population fragmentation in sky-islands explain the contemporary distribution of genomic variation in two narrow-endemic montane grasshoppers from a biodiversity hotspot

Vanina Tonzo<sup>1</sup> | Joaquín Ortego<sup>1</sup>

<sup>1</sup> Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC); Avda. Américo Vespucio 26 – 41092; Seville, Spain

## Abstract

Montane biotas from mid-latitudes are likely to have experienced dramatic distributional shifts in response to Quaternary glacial cycles, alternating periods of isolation on mountain tops during interglacials with phases of demographic expansion and population connectivity in glacial stages. Here, we use genomic data and a suite of analytical approaches to infer the demographic processes that have shaped contemporary patterns of genetic variation in *Omocestus bolivari* and *O. femoralis*, two narrow-endemic Iberian grasshoppers inhabiting the sky island archipelago of the Baetic System. Inference of spatial patterns of genetic structure, environmental niche modelling, and statistical evaluation of alternative demographic models within an Approximate Bayesian Computation framework collectively supported genetic admixture during glacial periods and postglacial colonization of sky islands, rather than long-term population isolation, as the scenario best explaining the contemporary distribution of genomic variation in the two focal taxa. Although the genetic makeup of contemporary populations has been shaped to a great extent by the historical process of colonization of mountain tops, consistent with the expanding-contracting sky-island archipelago model, our analyses revealed that postglacial range fragmentation and isolation have also contributed to reduce within-population levels of genetic diversity. Our study exemplifies the potential of integrating genomic and environmental niche modelling data across biological and spatial replicates to determine whether organisms with similar habitat requirements have experienced concerted/idiosyncratic responses to Quaternary climatic oscillations, which can ultimately help to reach more general conclusions about the vulnerability of mountain biodiversity hotspots to ongoing climate warming.

**Keywords:** Approximate Bayesian Computation, ddRAD-seq, demographic history, environmental niche modelling, landscape genetics, Pleistocene glaciations

## 1 | Introduction

Distributional shifts in response to Quaternary climatic oscillations had a dramatic impact on biogeographical patterns of species diversity, abundance and local endemism (Hewitt, 1996; Sandel et al., 2011). These cycles, in particular the high-amplitude climatic changes characterizing the Middle-Late Pleistocene (Jouzel et al., 2007), have also shaped the distribution and spatial patterns of genetic variation in many organisms (Hewitt, 2000; Ribera & Vogler, 2004). However, multiple studies have documented considerable heterogeneity across regions and taxa in the demographic consequences of Pleistocene glacial cycles (Cooper, Ibrahim, & Hewitt, 1995; Hewitt, 2000; Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998). On the one hand, the impact of Pleistocene glaciations strongly depended on latitude and regional topography, with extinction-recolonization dynamics at higher latitudes (i.e., “southern richness to northern purity” paradigm; Hewitt, 1996) and elevational shifts and more complex processes of population fragmentation and connectivity at lower latitudes such as the tropics or temperate regions (e.g., “refugia within refugia” concept; Gómez and Lund 2006). On the other hand, the way organisms respond to climate changes strongly depend on species-specific niche requirements and life-history traits, which define favorable/unfavorable climatic periods (glacials or interglacials; Bennett & Provan, 2008) and their capacity to deal with population fragmentation (e.g., micro-habitat preferences, dispersal capacity, etc.; e.g., Massatti & Knowles, 2016; Papadopoulou & Knowles, 2015a; Paz, Ibanez, Lips, & Crawford, 2015). These aspects determined the location and extension of Pleistocene refugia, which have played a predominant role on species’ persistence during unfavorable climatic periods and acted as source populations from which species expanded their ranges at the onset of more favorable conditions (Bennett & Provan, 2008; Hawks, Hunley, Lee, & Wolpoff, 2000; Stewart, Lister, Barnes, & Dalen, 2010). Temperate species currently inhabiting low elevation areas generally restricted their distributions to southern refugia during glacial phases (i.e., glacial refugia) and expanded during interglacial periods whereas cold-adapted species, nowadays presenting fragmented populations at high elevations/latitudes (i.e., interglacial refugia), had much more widespread distributions in glacial stages (Hewitt, 2000; Reid et al., 2019).

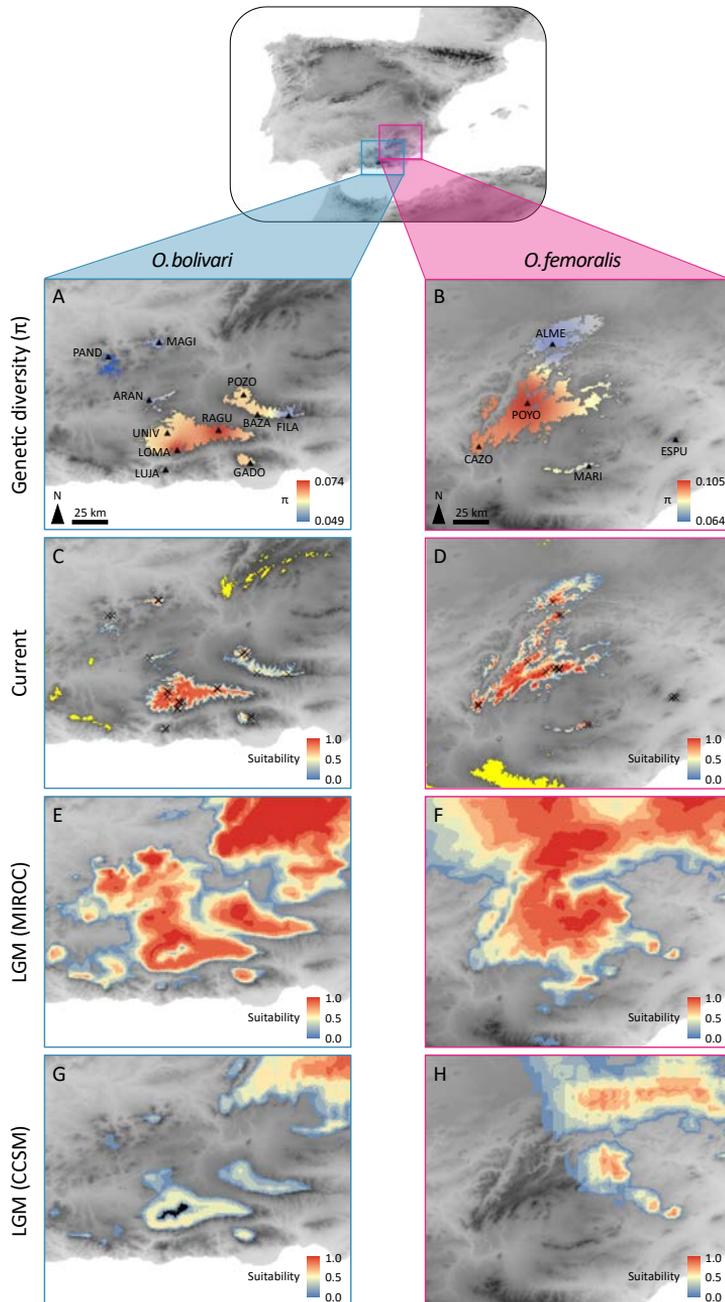
Mid- and low-latitude alpine and montane taxa represent small-scale replicates of cold-adapted species living at high latitudes (Hewitt, 1996; 2000). Sky islands are a paradigmatic example of interglacial refugia in which alpine and montane organisms currently form

highly isolated populations in mountain ranges separated from each other by intervening valleys with unsuitable environmental conditions (e.g., DeChaine & Martin, 2005; Knowles & Massatti, 2017; McCormack, Bowen, & Smith, 2008; Mouret et al., 2011). Changing climatic conditions cause suitable habitats to expand or shrink along elevational gradients, leading to alternating periods of population isolation and connectivity (DeChaine & Martin, 2005; Knowles, 2000). Periods of isolation on mountain tops are expected to lead to demographic bottlenecks, genetic drift and divergence among populations, while periods of connectivity allow for dispersal and gene flow among formerly isolated populations (DeChaine & Martin, 2005). Species restricted to sky islands often show unique patterns of population genetic structure resulted from climate-induced distributional shifts (Knowles & Alvarado-Serrano, 2010; Mouret et al., 2011), processes that in some cases have led to lineage diversification and speciation (Knowles, 2000; Knowles & Massatti, 2017). Therefore, montane habitats are often important hotspots of intraspecific genetic diversity, species richness and local endemism, providing an excellent study system for understanding how climatic changes associated with Pleistocene glacial cycles impacted species distributions, demography and genetic diversification (Mairal et al., 2017; Mastretta-Yanes et al., 2018). Their study takes particular relevance if we consider that species restricted to sky-islands often show small population sizes, harbor reduced levels of local genetic diversity and are highly vulnerable to extinction due to ongoing climate warming and the progressive shrink of alpine ecosystems (Atkins et al., 2020; Rubidge et al., 2012).

Southeastern Iberia constitutes an important biodiversity hotspot due to the interplay among a vibrating geological history (Meulenkamp & Sissingh, 2003), an extraordinarily complex topography (Braga, Martin, & Quesada, 2003), and a limited impact of Quaternary glaciations (Hughes & Woodward, 2017; Palacios, Gomez-Ortiz, Andres, Salvador, & Oliva, 2016). The presence of vast areas free of permanent ice during the coldest stages of the Pleistocene made this region an important glacial refugium for warm-temperate taxa (Gómez & Lunt, 2007; Hewitt, 2011; Stewart et al., 2010). At the same time, the mountain ranges of the region (Baetic system; > 3,000 m) constitute important interglacial refugia for several cold-adapted organisms that currently persist in severely fragmented populations at high elevation patches of suitable habitat and contribute disproportionately to the extraordinary rates of local endemism of the area (Mota, Pérez-García, Jimenez, Amate, & Peñas, 2002; Peñas, Perez-García, & Mota, 2005). This offers an ideal biogeographical

setting to jointly study the genetic consequences of long-term population fragmentation in sky islands and evaluate the temporal scale at which past distributional shifts have shaped contemporary patterns of genetic variation in alpine organisms from mid latitudes.

Here, we use genomic data obtained via restriction site-associated DNA sequencing (ddRAD seq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) and a suite of analytical approaches to shed light on the demographic processes underlying contemporary patterns of genetic variation in the grasshoppers *Omocestus bolivari* Chopard, 1939 and *O. femoralis* Bolívar, 1908, two narrow-endemic Iberian species inhabiting high elevation areas in the sky island archipelago of the Baetic System (Presa et al., 2016a, b). The two taxa present adjacent but not overlapping distributions, both occupying severely fragmented open habitats of thorny shrub formations (e.g., *Erinacea* sp., *Festuca* sp., *Juniperus* sp.) located at the upper (>1,500 m) thermoclimatic belts of the Mediterranean region. The distribution range of the two species include sky islands of varying size, ranging from large ones in the main massifs of the region (*O. bolivari*: Sierra Nevada and Sierra de Baza-Filabres; *O. femoralis*: Sierra de Cazorla) to tiny patches of suitable habitat located in isolated mountains with maximum elevations close to the lower altitudinal limit of the species (Figure 1). By using these two species of uniform ecological and life-history traits as independent replicates, we exemplify how the integration of genomic and environmental niche modelling (ENM) data within a comparative framework can help to elucidate the responses of species with similar habitat requirements to processes of population fragmentation/connectivity driven by Pleistocene climatic oscillations and understand the temporal scale at which they contributed to shape their contemporary patterns of genetic variation (Knowles & Massatti, 2017; Pan, Hulber, Willner, & Schneeweiss, 2020). In a first step, (i) we tested whether observed patterns of genomic variation in the two biological replicates reflect the signals of historical population connectivity (i.e., gene flow during glacial periods) and subsequent colonization and population isolation in sky islands (i.e., genetic drift during interglacial periods) or, rather, if post-glacial population fragmentation and genetic drift after the Holocene have entirely monopolized the genetic makeup of contemporary populations. Specifically, we quantified genetic structure and coupled ecological niche models (ENMs) and a spatiotemporally explicit simulation approach based on coalescent theory to mimic complex demographic processes under a suite of scenarios representing historical gene flow and post glacial colonization in sky islands vs. contemporary population isolation (Currat, Ray, & Excoffier,



**Figure 1** (A-B) Spatial patterns of genetic diversity and (C-H) environmental niche models (ENM) for *Omocestus bolivari* and *O. femoralis*. Panels A-B show genetic diversity ( $\pi$ ) of the studied populations (triangles, population codes as in Table 1) interpolated across the current distribution of each species as predicted by their respective ENMs. Panels C-H show the projection of ENMs for (C-D) present and last glacial maximum (LGM) conditions under the (E-F) CCSM4 and (G-H) MIROC-ESM general atmospheric circulation models. Maps for current distributions show records (crosses) used to build the ENMs and dark yellow color indicate areas predicted as suitable out of the known limits of species' ranges (i.e., over-predictions, not shown in panels A-B). All maps show suitable areas above the maximum training sensitivity plus specificity (MTSS) logistic threshold of MAXENT. Grey background represents elevation, with darker areas corresponding to higher altitudes.

2004; Ray, Currat, Foll, & Excoffier, 2010; e.g., Knowles & Massatti, 2017). The expectations under these alternative scenarios were tested and validated against observed genomic data within an Approximate Bayesian Computation (ABC) framework (e.g., Estoup et al., 2010; He, Edwards, & Knowles, 2013; Neuenschwander et al., 2008). In a second step, (ii) we analyzed the impact of sky-island size and the spatial location of contemporary populations on their demographic fate and levels of genetic variation. To this end, we tested the hypothesis that genetic diversity is positively associated with the geographic centrality of populations and habitat patch-size and used demographic reconstructions based on genomic data to determine whether changes in effective population size ( $N_e$ ) over time differed between sky-islands of contrasting size and degree of peripheral location (Duncan, Crespi, Mattheus, & Rissler, 2015; Lira-Noriega & Manthey, 2014).

## 2 | Materials and Methods

### 2.1 | Population sampling

Between 2011 and 2016, we sampled eleven populations of *Omocestus bolivari* and five populations of *O. femoralis* (Table 1; Figure 1). According to our extensive surveys in southeastern Iberia and available records in the literature (Prunier, 2014 and references therein), the sampled populations cover the entire known distribution range of the two species (Figure 1). We stored specimens in 2 ml vials with 96 % ethanol and preserved them at  $-20^{\circ}\text{C}$  until needed for genomic analyses.

### 2.2 | Genomic library preparation and processing

We used NucleoSpin Tissue kits (Macherey-Nagel, Düren, Germany) to extract and purify genomic DNA from the hind femur of each individual. We processed genomic DNA into four genomic libraries using the double-digestion restriction-fragment-based procedure (ddRADseq) described in Peterson et al. (2012) (see Tonzo, Papadopoulou, & Ortego, 2019 for details) and used the different programs distributed as part of the STACKS v. 1.35 pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to filter and assemble our sequences into *de novo* loci and call genotypes (for details, see Supplementary Methods S1).

**Table 1** Locality, code, number of genotyped individuals ( $n$ ), latitude, longitude, elevation and genetic diversity ( $\pi$ , nucleotide diversity) for each studied population of *Omocestus bolivari* and *O. femoralis*.

Species	Locality	Code	$n$	Latitude	Longitude	Elevation	$\pi$
<i>O. bolivari</i>	Sierra de La Pandera	PAND	8	37.636927	-3.805924	1,510	0.0488
<i>O. bolivari</i>	Sierra Mágina	MAGI	8	37.740555	-3.446550	1,930	0.0573
<i>O. bolivari</i>	Sierra de Arana	ARAN	7	37.331304	-3.516227	1,736	0.0640
<i>O. bolivari</i>	Albergue Universitario	UNIV	6	37.096398	-3.389056	2,446	0.0726
<i>O. bolivari</i>	Pozo de la Nieve	POZO	6	37.368102	-2.849309	2,040	0.0739
<i>O. bolivari</i>	Sierra de Baza	BAZA	8	37.227338	-2.752014	1,876	0.0685
<i>O. bolivari</i>	Sierra de Los Filabres	FILA	7	37.219673	-2.530702	2,117	0.0637
<i>O. bolivari</i>	Puerto de La Ragua	RAGU	8	37.116477	-3.026024	2,100	0.0724
<i>O. bolivari</i>	Loma de Piedra Blanca	LOMA	8	36.975003	-3.319273	2,258	0.0718
<i>O. bolivari</i>	Sierra de Lújar	LUJA	8	36.836947	-3.399291	1,750	0.0621
<i>O. bolivari</i>	Sierra de Gádor	GADO	8	36.884610	-2.803412	2,077	0.0646
<i>O. femoralis</i>	Pico Almenara	ALME	7	38.543520	-2.440236	1,620	0.0796
<i>O. femoralis</i>	Sierra de Cazorla	CAZO	8	37.815379	-2.959905	1,814	0.0994
<i>O. femoralis</i>	Poyotello	POYO	8	38.119515	-2.616541	1,600	0.1050
<i>O. femoralis</i>	Sierra de María	MARI	5	37.676456	-2.183102	1,882	0.0884
<i>O. femoralis</i>	Sierra de Espuña	ESPU	8	37.865166	-1.571250	1,514	0.0639

## 2.3 | Population genetic structure

We analyzed population genetic structure of the two species using the Bayesian Markov chain Monte Carlo clustering method implemented in the program STRUCTURE v. 2.3.3 (Pritchard, Stephens, & Donnelly, 2000). We ran STRUCTURE assuming correlated allele frequencies and admixture without using prior population information. We conducted 15 independent runs with 200,000 MCMC cycles, following a burn-in step of 100,000 iterations, for each of the different possible  $K$  genetic clusters. We retained the ten runs having the highest likelihood for each value of  $K$  and inferred the number of populations best fitting the dataset using log probabilities [ $\Pr(X|K)$ ] (Pritchard et al., 2000) and the  $\Delta K$  method (Evanno, Regnaut, & Goudet, 2005), as implemented in STRUCTURE HARVESTER v. 0.9.93 (Earl & vonHoldt, 2012). We used CLUMPP v. 1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same  $K$ -value (Jakobsson & Rosenberg, 2007) and DISTRUCT v. 1.1 (Rosenberg, 2004) to visualize individual's probabilities of population membership as bar plots. Complementary to STRUCTURE analyses, we performed principal component analyses (PCA) as implemented in the R package *adegenet* (Jombart, 2008). Before running the PCA, we scaled and centered allele frequencies and replaced missing data with mean allele frequencies using the *scaleGen* function as recommended by Jombart (2008).

## 2.4 | Environmental niche modelling

We built environmental niche models (ENM) to predict the geographic distribution of climatically suitable habitats for *O. bolivari* and *O. femoralis* both in the present and during the last glacial maximum (LGM, 21 ka). Then, we used this information to create maps of environmental suitability in these two time periods and generate alternative demographic models as detailed in the next section (e.g., He et al., 2013; Massatti & Knowles, 2016). To build ENMs, we used the maximum entropy algorithm implemented in MAXENT v.3.3.3 (Phillips, Anderson, & Schapire, 2006; Phillips & Dudik, 2008) and the 19 bioclimatic variables from the WORLDCLIM dataset (<http://www.worldclim.org/>) interpolated to 30-arcsec resolution (~1 km<sup>2</sup> cell size) (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). To generate climate suitability maps during the LGM, we projected the ENM onto LGM bioclimatic conditions derived from the CCSM4 (Community Climate System Model; Braconnot et al., 2007) and the MIROC-ESM (Model of Interdisciplinary Research on Climate; Hasumi & Emori, 2004) general atmospheric circulation models. We built ENMs using our own species occurrence data and records available in the literature (Prunier, 2014 and references therein). Prior to modelling, we mapped and examined all records to identify and exclude those having obvious geo-referencing errors. After data filtering and excluding duplicates (i.e., records falling within the same grid cell), we retained a total of 26 occurrence records for *O. bolivari* and 16 records for *O. femoralis*. We used the R package ENMEVAL (Muscarella et al., 2014) and the Akaike's Information Criterion corrected for small sample size (AICc) (Warren & Seifert, 2011) to conduct parameter tuning and determine the optimal feature class (FC) and regularization multiplier (RM) settings for MAXENT. We tested a total of 248 models of varying complexity by combining a range of regularization multipliers (RM) (from 0 to 15 in increments of 0.5) with eight different feature classes (FC) combinations (L, LQ, LQP, H, T, LQH, LQHP, LQHPT, where L = linear, Q = quadratic, H = hinge, P = product and T = threshold) and followed the approach detailed in González-Serna, Cordero, & Ortego (2019) for variable selection.

## 2.5 | Testing alternative demographic models

We used the integrative distributional, demographic and coalescent (iDDC) approach (He et al., 2013) and an Approximate Bayesian Computation (ABC) framework (Beaumont, Zhang, & Balding, 2002; Csilléry, Blum, Gaggiotti, & Francois, 2010) to test alternative scenarios representing different hypotheses about how landscape heterogeneity (or its

lack thereof) and colonization from glacial refugia vs. contemporary population isolation in sky islands explain the spatial distribution of genomic variation in our two focal species. This approach is described in He et al. (2013) and consists of three main steps: (i) constructing alternative demographic models representing different hypotheses about the processes shaping spatial patterns of genetic structure and diversity; (ii) running demographic and genetic simulations under each model using the software SPLATCHE2 (see Ray et al., 2010); (iii) evaluating the fit of observed genomic data (i.e., empirical genomic data obtained after genotyping sampled populations) to the genetic expectations under each model, identifying the most probable model/s, and estimating demographic parameters (e.g., He et al., 2013; González-Serna et al., 2019). Below we describe the most relevant aspects of this approach and present an extended version with all the details in Supplementary Methods S2.

**Constructing demographic models.** We generated two main sets of models that differ in the hypothetical demographic processes that have shaped spatial patterns of contemporary genetic variation (Table 2):

i) Colonization of sky islands from glacial refugia (Models A, B, D and E). The first two models (Models A and B) are dynamic models (*sensu* He et al., 2013) incorporating the colonization process from hypothetical glacial refugia and distributional shifts resulted from the interaction between the species bioclimatic envelope and Pleistocene glacial cycles (e.g., He et al., 2013; Massatti & Knowles, 2016). In these scenarios, carrying capacities change over time according to climatic suitability maps obtained from projections of the ENM to the present and the LGM bioclimatic conditions under the MIROC-ESM (Model A) and CCSM (Model B) general atmospheric circulation models (see section *Environmental niche modelling*). These models considered landscapes from three consecutive time periods (LGM, intermediate, current) reflecting temporal shifts in the spatial distribution of environmentally suitable areas for the species in response to climate changes since the LGM (e.g., He et al., 2013; Massatti & Knowles, 2016). Forward demographic simulations under this model initialized 21 ka BP from hypothesized refugial populations (i.e., source populations, each one with an effective population size of  $N_{ANC}$ ) that were located at every habitat patch predicted as suitable for each focal species during the LGM according to MIROC-ESM (Model A) and CCSM (Model B) projections. Specifically, suitable habitat patches during the LGM were identified as those cells with a probability of presence of the species above the maximum training sensitivity plus specificity (MTSS)

logistic threshold for occurrence from MAXENT (Liu, Berry, Dawson, & Pearson, 2005). Finally, we generated two more models (Models D and E) analogous to the previous ones but in which carrying capacities ( $k$ ) are homogeneous across space and through time. These static models (*sensu* He et al., 2013) are equivalent to a flat landscape or an isolation-by-distance model and only differ among them in the location of ancestral populations, which were based on the patches of suitable habitat identified under LGM-MIROC (Model D) and LGM-CCSM (Model E) bioclimatic conditions.

ii) Population isolation in sky islands (Models C and F). The first model of this set (Model C) is a static model representing genetic drift associated with the isolation of populations in sky-islands according to the current geographical configuration of suitable habitats. In this model, carrying capacities do not change over time and are defined by an environmental suitability layer obtained from the projection of the ENM to the current bioclimatic conditions. Forward demographic simulations under this model initialized 7 ka BP from hypothesized refugial populations located in every patch of habitat predicted as suitable during present time according to the MTSS logistic threshold for occurrence from MAXENT. Thus, this model hypothesizes that the spatial distribution of contemporary genetic variation reflects population fragmentation and isolation in sky islands since the Mid-Holocene (e.g., Knowles & Massatti, 2017). As done for the first set of models, we also generated a scenario (Model F) analogous to the previous one but in which carrying capacities ( $k$ ) are homogeneous across space (i.e., equivalent to an isolation-by-distance model; He et al., 2013).

**Model choice and parameter estimation.** We used an Approximate Bayesian computation (ABC) framework to perform model selection and parameter estimation, as implemented in ABCTOOLBOX programs (TRANSFORMER and ABCESTIMATOR) and R scripts (*findPLS*) (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010). We used the R package *pls* v.2.6-0 (Mevik & Wehrens, 2007) and the *findPLS* script to extract partial least squares (PLS) components with Box-Cox transformation from the summary statistics of the first 10,000 simulations for each model (Boulesteix & Strimmer, 2007; Wegmann et al., 2010). The first five PLSs extracted from the summary statistics were used for ABC analyses, as the root-mean-squared error (RMSE) of the three demographic parameters employed ( $K_{\text{MAX}}$ ,  $m$ ,  $N_{\text{ANC}}$ ; see Methods S2 for details) for the two species did not decrease significantly with additional PLSs (Figure S1). We used the linear combinations of summary statistics obtained from the first 10,000 simulations for each model to transform all

datasets (observed empirical and simulated datasets) with the program TRANSFORMER (Wegmann et al., 2010). For each model, we retained the 1,000 simulations (0.5%) closest to observed empirical data and used them to approximate marginal densities and posterior distributions of the parameters with a postsampling regression adjustment using the ABC-GLM (general linear model) procedure detailed in Leuenberger & Wegmann (2010) and implemented in ABCESTIMATOR (see also Csilléry et al., 2010). We used Bayes factors (BF) for model selection (Jeffreys, 1961; Kass & Raftery, 1995).

**Model validation.** To evaluate the ability of each model to generate the empirical data, we calculated the Wegmann's  $p$ -value from the 1,000 retained simulations (Wegmann et al., 2010). We also assessed the potential for a parameter to be correctly estimated by computing the proportion of parameter variance that was explained (i.e., the coefficient of determination,  $R^2$ ) by the retained PLSs (Neuenschwander et al., 2008). For the most supported model for each species, we determined the accuracy of parameter estimation using a total of 1,000 pseudo-observation datasets (PODs) generated from prior distributions of the parameters. If the estimation of the parameters is unbiased, posterior quantiles of the parameters obtained from PODs should be uniformly distributed (Wegmann et al., 2010). As with the empirical data, we calculated the posterior quantiles of true parameters for each pseudo run based on the posterior distribution of the regression-adjusted 1,000 simulations closest to each pseudo-observation.

## 2.6 I Genetic diversity and historical changes in effective population size

To further explore the processes determining the spatial distribution of genetic variation in the two species we (i) tested the association between genetic diversity and the geographic centrality and habitat patch-size of the studied populations, and (ii) used genomic data to reconstruct changes in effective population size ( $N_e$ ) over time. We employed linear regressions in SPSS v. 26.0 to analyze whether genetic diversity (nucleotide diversity,  $\pi$ ) is associated with the distance of each population to the species' distribution centroid and contemporary suitable habitat patch-size. Given that the precision of genetic diversity estimates may differ among populations due to differences in sample sizes, we used a weighted least square (WLS) method where weight equals the number of genotyped individuals for each population (Table 1). We calculated nucleotide diversity for each population using the program *populations* from STACKS. The centroid of species

distribution was calculated in ARCMAP v.10.3 on the basis of polygons (patches) of suitable habitat defined by grid cells with a probability of presence of the focal species above the maximum training sensitivity plus specificity (MTSS) logistic threshold for occurrence from MAXENT (Liu et al., 2005). Similarly, patch-size for each study population was calculated considering the area (in km<sup>2</sup>) of continuous suitable habitat defined as above on the basis of raster maps obtained from projecting ENMs to contemporary bioclimatic conditions.

We inferred changes in effective population sizes ( $N_e$ ) over time for each studied population using STAIRWAY PLOT (Liu & Fu, 2015), a flexible multi-epoch model approach based on the site frequency spectrum (SFS) that does not require a predefined demographic model for estimating past demographic histories. To compute a folded SFS for each population, we ran STACKS in order to obtain a VCF file retaining only loci represented in at least 50% of the individuals of the population. To remove all missing data for the calculation of the SFS and minimize errors with allele frequency estimates, each population was down-sampled to five individuals using a custom Python script written by Qixin He and available on Dryad (Papadopoulou & Knowles, 2015b). We ran STAIRWAY PLOT considering a 1-year generation time, assuming the mutation rate per site per generation of  $2.8 \times 10^{-9}$  estimated for *Drosophila melanogaster* (Keightley, Ness, Halligan, & Haddrill, 2014), and performing 200 bootstrap replicates to estimate 95% confidence intervals.

## 3 | Results

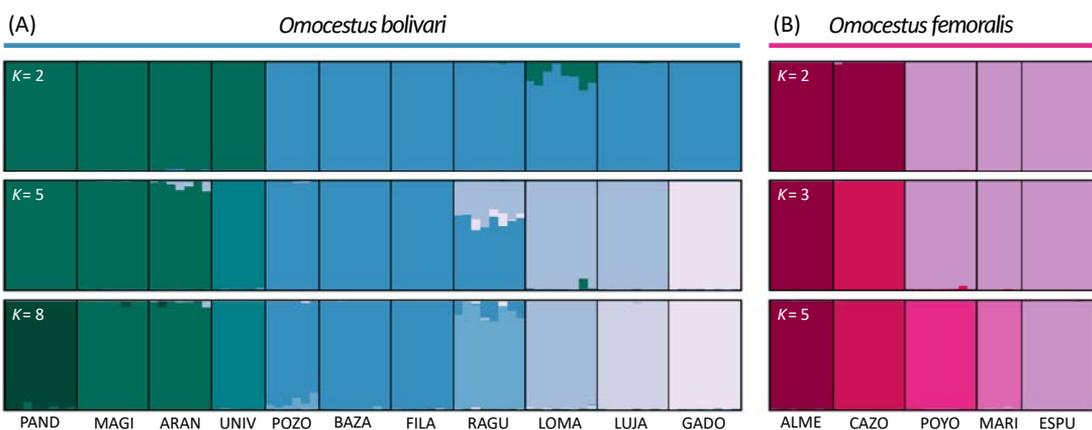
### 3.1 | Genomic data

After quality filtering, 182,588,301 and 74,830,407 reads were retained across all genotyped individuals of *O. bolivari* and *O. femoralis*, respectively (Figure S2). Final exported data sets obtained with STACKS after removing loci that did not meet the different filtering requirements contained 2,262 SNPs for *O. bolivari* and 14,116 SNPs for *O. femoralis*.

### 3.2 | Population genetic structure

STRUCTURE analyses showed that log probabilities of the data ( $\text{LnPr}(X|K)$ ) reached a plateau for  $K = 8$  in *O. bolivari* and  $K = 5$  in *O. femoralis* (Figure S3). The  $\Delta K$  criterion indicated an “optimal” clustering solution for  $K = 5$  in *O. bolivari* and  $K = 3$  in *O. femoralis*

(Figure S3). The two species presented a strong genetic structure, with most (*O. bolivari*) or all (*O. femoralis*) populations being assigned to unique genetic clusters with little or no genetic admixture among them (Figure 2 and S4). In the two species,  $K = 2$  separated northwestern (*O. bolivari*: PAND, MAGI, ARAN, and UNIV; *O. femoralis*: ALME and CAZO) from southeastern populations (*O. bolivari*: POZO, BAZA, FILA, RAGU, LUMA, LUJA, and GADO; *O. femoralis*: POYO, MARI, and ESPU). These two main clusters split hierarchically at increasingly higher  $K$  values, generally grouping the different populations in accordance with their geographical location. Remarkably, populations located in the same “sky island” did not always cluster together at different hierarchical levels. For instance, in *O. bolivari*, population UNIV grouped with northwestern populations (PAND, MAGI, and ARAN) rather than with other populations (LOMA and RAGU) located within its same mountain range (i.e., the main Sierra Nevada massif; Figure 1a and 2b). In the same line, population POYO from *O. femoralis* clustered together with the highly isolated southeastern populations (MARI and ESPU) rather than with the two other populations (ALME and CAZO) located within the Sierra de Cazorla massif and with which it is expected to be currently more connected (Figure 1b and 2b). Principal component analyses for both species yielded analogous results to those obtained with STRUCTURE, separating along PC1 northwestern from southeastern populations and grouping populations according to their geographical location (Figure S5).



**Figure 2** Genetic assignment of individuals based on the results of STRUCTURE for (A) *Omocestus bolivari* ( $K = 2$ ,  $K = 5$  and  $K = 8$ ) and (B) *O. femoralis* ( $K = 2$ ,  $K = 3$  and  $K = 5$ ). Individuals are partitioned into  $K$  colored segments representing the probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1.

### 3.3 I Environmental niche modelling

We generated final ENMs using the feature class (FC) combination (*O. bolivari* = LQ; *O. femoralis* = LQ) and regularization multiplier (*O. bolivari* = 6.0; *O. femoralis* = 2.0) that minimized AICc across the set of models tested with ENMEVAL for each species. After removing highly correlated variables ( $r > 0.9$ ) and those with a zero percent contribution, our models retained four bioclimatic variables for *O. bolivari* (sorted by percent contribution, BIO8: 36.8 %; BIO7: 34.8 %; BIO11: 16.5 %; BIO2: 11.9 %) and eight variables for *O. femoralis* (BIO11: 37.9 %; BIO2: 33.9 %; BIO7: 13.6 %; BIO8: 7.4 %; BIO19: 4.3 %; BIO15: 1.3 %; BIO12: 1.2 %; BIO18: 0.4 %). Inspection of predicted distributions for the present confirmed that ENMs yielded distribution patterns coherent with the observed current distribution of the two species (Figure 1c-d). However, current distribution maps for the two species also identified as suitable some high elevation areas outside their respective distribution ranges. In the two species, over-predicted areas included partially (*O. bolivari*) or totally (*O. femoralis*) the distribution range of each other, which is not surprising given their similar habitat, elevation, and environmental requirements. Projections of present-day climate niche envelopes to LGM climatic conditions under two general atmospheric circulation models (CCSM4 and MIROC-ESM) suggest that the two species have experienced important distributional shifts in response to Pleistocene glaciations (Figure 1e-h). As expected for montane-alpine organisms, the distribution of the two species expanded into lowland areas during the LGM (Figure 1e-h). However, projections under the two circulation models also presented important differences. Populations of the two species were predicted to be much more connected during the LGM under the MIROC-ESM than under the CCSM model. According to MIROC-ESM, most currently isolated populations of the two species (with the exception of Sierra de Lújar and Sierra de Gádor in *O. bolivari*) became connected by corridors of suitable habitat during the LGM. However, in *O. bolivari*, most contemporary sky islands were also predicted to be isolated from each other during the LGM according to the CCSM model (with the exception of Sierra de Arana and Sierra Nevada; Figure 1g). In the case of *O. femoralis*, the southwesternmost portion of its current distribution was predicted as unsuitable under the CCSM model, suggesting that this species might have experienced important distributional changes beyond elevational shifts (Figure 1h).

### 3.4 I Testing alternative demographic models

Based on marginal densities calculated from the 0.5% of retained simulations, the best ranked models in the two species were those incorporating the colonization process from hypothetical glacial refugia (Models A, B, D and E; Table 2). In contrast, models of isolation in contemporary sky islands had considerably lower marginal densities and a difference in Bayes factors  $> 10^6$  with the most supported model (Table 2), indicating strong relative support for colonization models in the two species (Jeffreys, 1961; Kass & Raftery, 1995). Moreover, only colonization models were able to simulate genetic data comparable with empirical data (Models A and B in *O. bolivari* and Models B and E in *O. femoralis*), unlike the isolation models in which there was a substantial difference between the likelihoods of the simulated data compared with the empirical data (Wegmann's  $p$ -values  $< 0.001$  in all cases; Table 2). Specifically, for *O. bolivari*, dynamic colonization models incorporating changes in habitat suitability over time under both the CCSM and MIROC projections (Models A and B) were those most supported by our empirical data and small Bayes factors

**Table 2** Statistics from the ABC procedure used for evaluating the relative support of each demographic model in *Omocestus bolivari* and *O. femoralis*. A higher marginal density correspond to a higher model support and a high Wegmann's  $p$ -value indicates that the model is able to generate data in agreement with empirical data. Bayes factors represent the degree of relative support for the model with the highest marginal density over the other models.  $R^2$  is the coefficient of determination from a regression between each demographic parameter ( $K_{MAX}$ ,  $m$ ,  $N_{ANC}$ ) and the five partial least squares (PLS) extracted from all summary statistics.

Model	Ancestral sources	Layer/s	Marginal density	Wegmann's $p$ -value	Bayes factor	$R^2$			
						$K_{MAX}$	$m$	$N_{ANC}$	
<i>O. bolivari</i>									
A	Colonization	LGM <sub>MIROC</sub>	LGM <sub>MIROC</sub> → Current	$9.14 \times 10^{-07}$	0.368	2.09	0.78	0.61	0.92
B	Colonization	LGM <sub>CCSM</sub>	LGM <sub>CCSM</sub> → Current	$1.91 \times 10^{-06}$	0.174	—	0.73	0.61	0.91
C	Isolation	Current	Current	$1.52 \times 10^{-13}$	<0.001	$1.26 \times 10^{07}$	0.79	0.66	0.91
D	Colonization	LGM <sub>MIROC</sub>	Flat landscape	$2.76 \times 10^{-09}$	0.006	$6.93 \times 10^{02}$	0.69	0.68	0.93
E	Colonization	LGM <sub>CCSM</sub>	Flat landscape	$9.15 \times 10^{-08}$	0.012	$2.09 \times 10^{01}$	0.70	0.70	0.91
F	Isolation	Current	Flat landscape	$3.67 \times 10^{-13}$	<0.001	$5.22 \times 10^{06}$	0.69	0.72	0.90
<i>O. femoralis</i>									
A	Colonization	LGM <sub>MIROC</sub>	LGM <sub>MIROC</sub> → Current	$1.26 \times 10^{-06}$	0.026	$1.28 \times 10^{01}$	0.36	0.68	0.96
B	Colonization	LGM <sub>CCSM</sub>	LGM <sub>CCSM</sub> → Current	$8.82 \times 10^{-06}$	0.195	1.83	0.28	0.66	0.93
C	Isolation	Current	Current	$1.00 \times 10^{-100}$	<0.001	$1.61 \times 10^{95}$	0.41	0.59	0.92
D	Colonization	LGM <sub>MIROC</sub>	Flat landscape	$6.35 \times 10^{-09}$	0.001	$2.54 \times 10^{03}$	0.24	0.69	0.96
E	Colonization	LGM <sub>CCSM</sub>	Flat landscape	$1.61 \times 10^{-05}$	0.239	—	0.19	0.73	0.96
F	Isolation	Current	Flat landscape	$1.00 \times 10^{-100}$	<0.001	$1.61 \times 10^{95}$	0.19	0.66	0.91

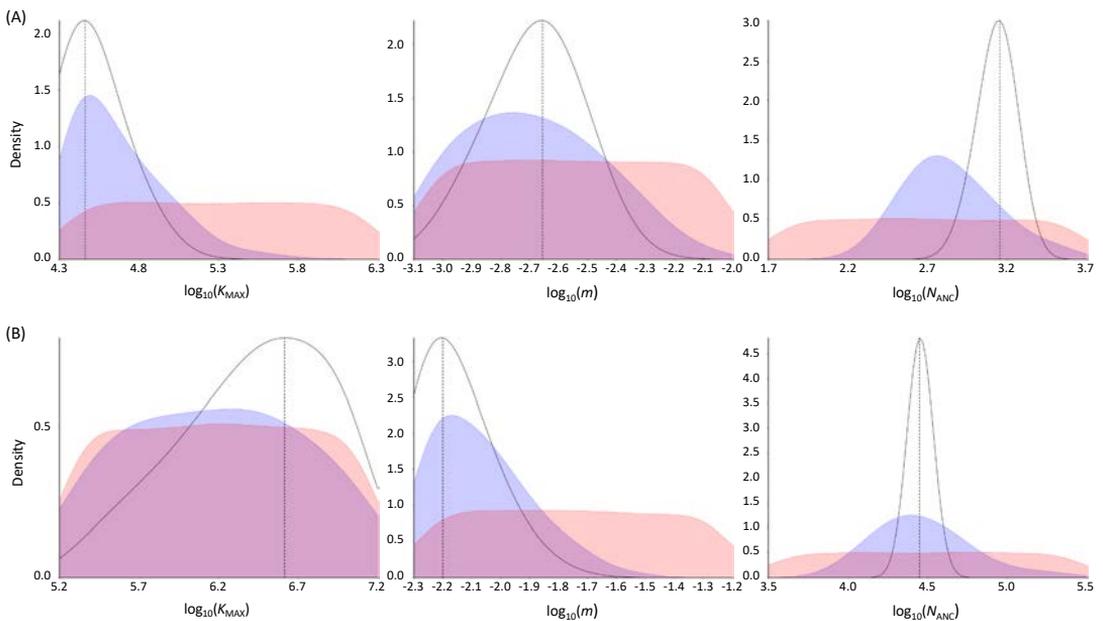
$K_{MAX}$ , carrying capacity of the deme with highest suitability (100%)

$m$ , migration rate per deme per generation

$N_{ANC}$ , ancestral population size

(BF < 3; Table 2) suggest that they are statistically indistinguishable (Kass & Raftery, 1995). In *O. femoralis*, both dynamic and static (i.e., flat landscape) colonization models based on CCSM projections (Models B and E) were the most supported by our empirical data and statistically indistinguishable (BF < 2; Table 2). The dynamic colonization model based on MIROC bioclimatic conditions also had a low Bayes factor (< 20), but its low Wegmann's  $p$ -value (< 0.03) indicates that it was not capable of generating data compatible with our empirical data (Table 2).

Posterior distributions of parameters under the most probable models were considerably distinct from the prior, indicating that the simulated data contained information relevant to estimating the parameters (Figure 3). Comparison of the posterior distributions before and after the ABC-GLM also showed the improvement that this procedure had on parameter estimates (Figure 3). The posterior distribution of  $K_{MAX}$  in *O.*

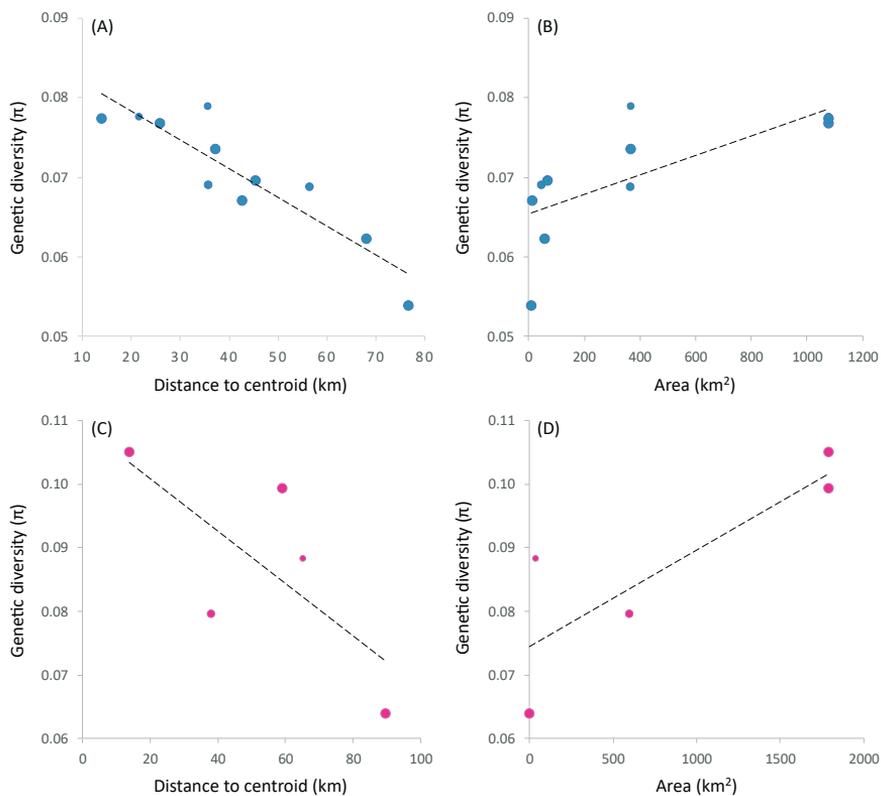


**Figure 3** Posterior distribution (solid black line) and mode (vertical dotted black line) of parameter estimates ( $K_{MAX}$ ,  $m$ ,  $N_{ANC}$ ) for the most supported model for (A) *Omocestus bolivari* and (B) *O. femoralis* based on a general linear model (GLM) regression adjustment of the 1,000 retained simulations (0.5%) closest to empirical data. The comparison of posterior distributions before (blue shading) and after (solid black line) the ABC-GLM shows the improvement that this procedure had on parameter estimates. The comparison of prior (red shading) and posterior (solid black line and blue shading) distributions demonstrates that the data contained information relevant to estimating the parameters.  $K_{MAX}$ , carrying capacity of the deme with highest suitability (100%);  $m$ , migration rate per deme per generation;  $N_{ANC}$ , ancestral population size.

*femoralis* was flat, indicating uncertainty in the estimation of this parameter (Figure 3B). Accordingly, the coefficients of determination ( $R^2$ ) from a multiple regression between each demographic parameter and the five retained PLS indicated that the employed summary statistics had a moderately high potential to correctly estimate all the parameters except  $K_{MAX}$  in *O. femoralis* (Table 2). Posterior quantiles of  $m$  in *O. bolivari*,  $K_{MAX}$  in *O. femoralis* and  $N_{ANC}$  in both taxa significantly deviated from a uniform distribution, indicating a potential bias in the estimation of these parameters (Figure S6).

### 3.5 | Genetic diversity and historical changes in effective population sizes

Visual inspection of the spatial distribution of genetic variation in the two species evidenced that peripheral populations and those located in smaller “sky islands” tended to



**Figure 4** Relationship between the genetic diversity ( $\pi$ ) of populations and (A, C) distance to centroid of species' distribution and (B, D) area of climatically suitable habitat for (A, B) *Omocestus bolivari* and (C, D) *O. femoralis*. Regression lines are indicated and dot size is proportional to sample size for each studied population.

present lower levels of genetic diversity (Figure 1a, b). Accordingly, genetic diversity ( $\pi$ ) in populations of *O. bolivari* was negatively associated with population centrality ( $r = -0.913$ ,  $t = -6.72$ ,  $P < 0.001$ ; Figure 4a) and showed a positive relationship with patch size ( $r = 0.733$ ,  $t = 3.23$ ,  $P = 0.010$ ; Figure 4b). In the case of *O. femoralis*, genetic diversity was positively associated with patch size ( $r = 0.878$ ,  $t = 3.17$ ,  $P = 0.050$ ; Figure 4d) but the negative relationship with population centrality did not reach statistical significance ( $r = 0.743$ ,  $t = -1.92$ ,  $P = 0.150$ ; Figure 4c). STAIRWAY PLOT analyses for *O. bolivari* revealed similar results across most of the populations, with an abrupt demographic expansion led by the advance of the glaciation till its maximum level around 21 ka BP (Figure 5a, b). Afterwards, populations experienced a decline in  $N_e$  until recent times (Figure 5a, b). One exception was the highly isolated populations PAND, ARAN, and GADO, which presented a general demographic stability during the last 100 ka (Figure 5b). In the case of *O. femoralis*, most populations showed idiosyncratic demographic histories, which ranged from expansions during the last glacial period followed by reductions in  $N_e$  at the onset of the Holocene (ALME and ESPU) to marked bottlenecks around the LGM (MARI) (Figure 5c, d).

## 4 | Discussion

Our study contributed to a better understanding of the demographic processes underlying spatial patterns of genetic variation in cold-adapted biotas currently forming severely fragmented populations in sky-island archipelagos from temperate regions (Hewitt, 1996; Knowles & Massatti, 2017; Perrigo, Hoorn, & Antonelli, 2020). Specifically, the combination of population genomic data, niche modelling and spatiotemporally explicit coalescent-based simulations supported population connectivity during glacial periods followed by colonization and population isolation in sky islands as the most supported scenario explaining contemporary patterns of population genetic diversity and structure in the two focal taxa. Although our analyses indicate that post-glacial fragmentation and genetic drift have not blurred the genomic signatures left by historical patterns of population connectivity, the lower levels of genetic diversity in peripheral populations confined to small patches of suitable habitat (i.e., small-sized sky islands) evidence the genetic consequences of long-term isolation. This study exemplifies how the integration of multiple lines of evidence provided by a comprehensive suite of analytical approaches applied to two independent spatial and biological replicates can help to identify concerted or idiosyncratic evolutionary and demographic histories of species with

similar habitat requirements and, ultimately, reach more general conclusions about the way organisms and whole communities have responded to Quaternary climatic oscillations. These aspects are of pivotal importance to understand the dynamics of montane and alpine ecosystems from mid and low latitude regions, which not only harbor high rates of local endemism but have been also identified among the most vulnerable to extinction due to ongoing climate warming and the progressive shrink of their habitats (Kidane, Steinbauer, & Beierkuhnlein, 2019; Perrigo et al., 2020).

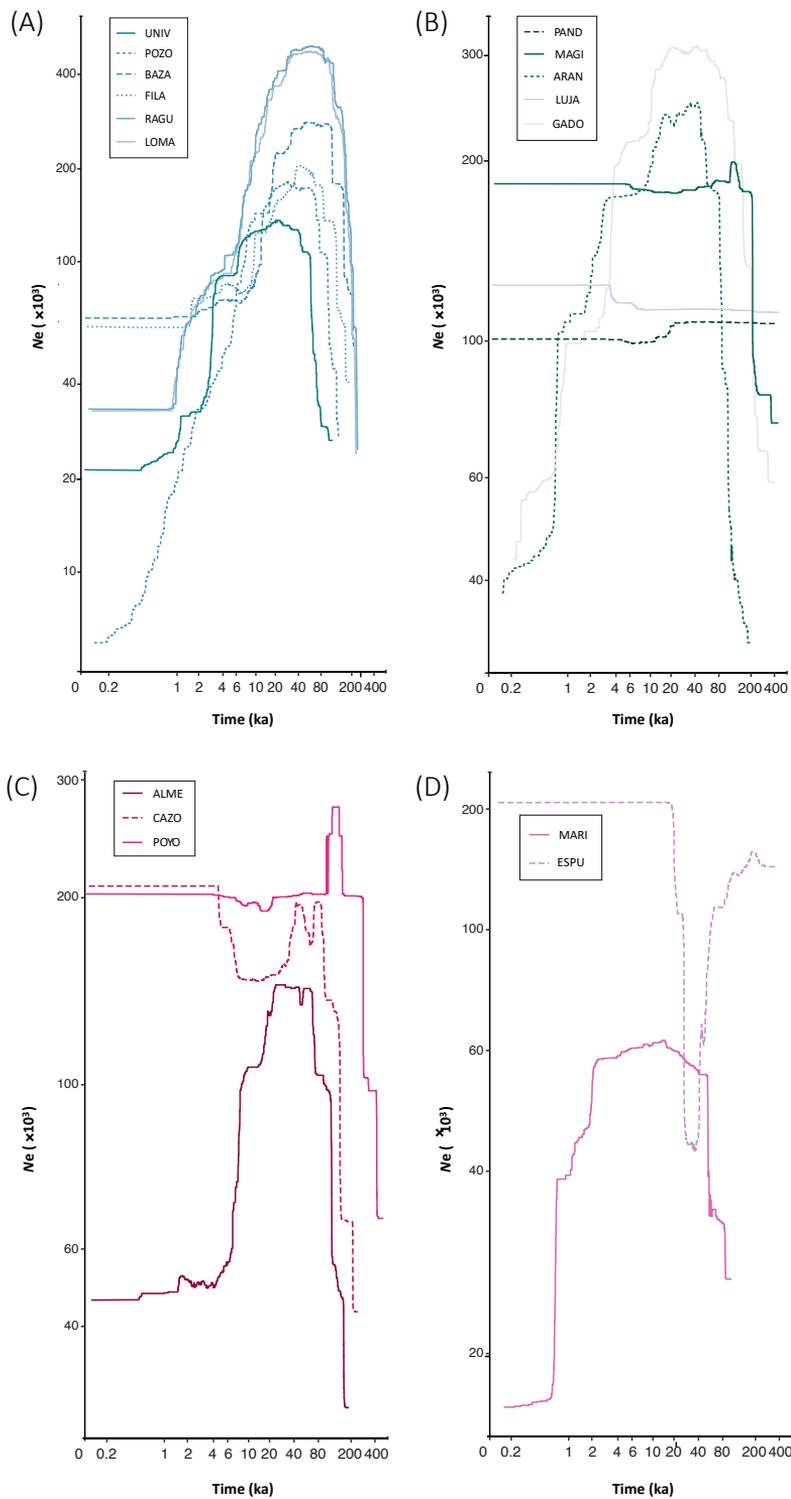
#### 4.1 I Marked genetic structure in sky islands

The contemporary genetic structure of *O. bolivari* and *O. femoralis* exhibited many similarities, with both taxa showing a marked northwest-southeast pattern of genetic differentiation and most genetic clusters identified at the different hierarchical levels presenting a very low degree of genetic admixture (Figure 2). The northwest-southeast pattern of differentiation likely reflects the predominant connectivity and origin of source populations that led to the colonization of sky islands from the more widespread lowland distributions presented by the two species during glacial periods, as supported by projections of environmental niche models to LGM bioclimatic conditions under alternative general circulation models (Figure 1). Remarkably, the two taxa presented populations that, being located on the same mountain range, are assigned to different genetic clusters, which, in turn, are shared with populations located in different sky-islands. This is exemplified by the genetic assignment of UNIV and POYO populations from *O. bolivari* and *O. femoralis*, respectively. Population UNIV is situated in a slope facing north in Sierra Nevada massif but shares a common ancestry with the northernmost populations of the species rather than with the other two populations (LOMA and RAGU) located within the same mountain range (Figure 2a). Similarly, the large Sierra de Cazorla massif encompasses populations CAZO, ALME and POYO, but the latter clusters with the highly isolated and southernmost populations MARI and ESPU (Figure 2b). This indicates that patterns of genetic structure at different hierarchical levels are in some instances more congruent with the geographical arrangement of populations than with their location in a specific sky island. Projections of ENMs to LGM bioclimatic conditions point to the connection of these populations during cold periods, indicating that apparently incongruent patterns of genetic structure likely reflect the identity of ancestral glacial-source populations rather than the spatial configuration of contemporary suitable habitats

and hypothetical corridors to gene flow. The low dispersal capacity of the brachypterous focal taxa (Clemente, García, & Presa, 1991), the considerable topographic complexity (i.e., ridges and steep slopes) within the main massifs (> 3,400 m in Sierra Nevada) of the region, and the limited ability of grasshoppers to move across abrupt landscapes (e.g., González-Serna et al., 2019; Noguerales, Cordero, & Ortego, 2016; Tonzo et al., 2019) might collectively explain the deep genetic structure and low or absent genetic admixture among populations located within the same sky island even if these are predicted to be nowadays connected by corridors of suitable habitat according to ENMs.

## 4.2 I Post-glacial colonization of sky-islands

Model selection under an ABC framework revealed that the spatially-explicit demographic scenarios most supported by empirical data were those in which the spatial distribution of contemporary genomic variation was shaped by processes of post-glacial colonization of sky islands from source populations defined by the spatial location of environmentally suitable areas during the last ice age (Knowles & Massatti, 2017; Mouret et al., 2011). This is in agreement with the classical pattern of glacial expansion and interglacial contraction suggested for montane species from temperate regions (e.g., DeChaine & Martin, 2005). In contrast, models of isolation in sky islands since the onset of the Holocene had a very low statistical support, indicating that genetic drift after post-glacial fragmentation cannot solely explain the genetic makeup of contemporary populations. Although scenarios based on colonization of sky-islands were the only ones able to simulate genomic data compatible with observed data, our approach was not able to distinguish between dynamic models considering changes in landscape heterogeneity since the LGM and static models considering a diffusion scenario of expansion from glacial source populations (see also Knowles & Massatti, 2017). These results are surprising considering that the two focal taxa have strict habitat preferences for high elevation areas and unsuitable lowlands are expected to act as impassable barriers to dispersal, as evidenced by distribution models (Figure 1) and strong genetic structure of contemporary populations (Figure 2). Simple Mantel tests indicate that genetic differentiation ( $F_{ST}$ ) is positively correlated with geographical distances among populations in *O. bolivari* ( $r = 0.57$ ,  $P < 0.001$ ,  $n = 11$ ) and marginally in *O. femoralis* ( $r = 0.60$ ,  $P = 0.058$ ,  $n = 5$ ). Processes of genetic admixture and drift mediated by population expansion and contraction during Quaternary glacial cycles is expected to have resulted in this pattern of isolation-by-



**Figure 5** Demographic history of each studied population of (A, B) *O. bolivari* and (C, D) *O. femoralis* inferred using STAIRWAY PLOT. Lines show the median estimate of effective population size ( $N_e$ ) over time for populations located in (A, C) large and (B, D) small sky islands. Population codes shown in the legend of each panel are described in Table 1.

distance (Hutchison & Templeton, 1999; Slatkin, 1993), which is similar to the expected outcome of migration-genetic drift equilibrium under a static, flat landscape scenario (He et al., 2013; Knowles & Massatti, 2017). Thus, admixture among nearby populations is likely to have reduced the power of our analyses to discriminate between dynamic (Models A-B) and static (Models D-E) scenarios of postglacial colonization, particularly in *O. femoralis* for which only five populations were available for analyses (Table 1).

The two general circulation palaeoclimatic models supported the expansion of the two focal taxa toward lowland areas during the LGM, but their respective populations were predicted to be much more connected under the MIROC-ESM than under the CCSM model (Figure 1). Although differences between palaeoclimatic models in predicting past species distributions is a frequent outcome of most studies (e.g., Ramírez-Barahona & Eguiarte, 2014; Wachter et al., 2016), the relative support for contrasting predictions is very rarely validated using independent sources of information (Nogués-Bravo, 2009). Such differences seemed to be captured by our model-based approach, which revealed that glacial-source populations identified for the two focal taxa on the basis of projections of ENMs to bioclimatic conditions under the CCSM general circulation model tended to provide a better fit to our genomic data than those obtained based on MIROC projections ( $BF > 2$  in all cases; Table 2). Although differences in the relative likelihood for scenarios based on the two palaeoclimatic models are small and must be interpreted with extreme caution, our results highlight the potential of spatiotemporally explicit coalescent-based simulations to refine hypotheses about the location of glacial refugia (Bemmels, Knowles, & Dick, 2019) and validate distributional shifts inferred from ENMs that cannot usually be contrasted with independent sources of information (e.g., fossil records; Davis, McGuire, & Orcutt, 2014; Fordham, Brook, Moritz, & Nogués-Bravo, 2014; Metcalf et al., 2014).

### 4.3 | Genetic diversity and demographic reconstructions

The spatial distribution of genetic variation in the two species evidenced that peripheral populations and those located in smaller sky islands tend to present comparatively lower levels of genetic diversity (Figure 1). These findings are consistent with the center-periphery model, which predicts that populations located on the species range edges become progressively smaller, are spatially more isolated, and harbor lower levels of genetic diversity than populations situated at the core of the distribution (e.g., Guo 2012; Lira-Noriega & Manthey, 2014). In contrast, genomic-based demographic

reconstructions indicate that, in general terms, most populations presented larger effective population sizes during the last glacial period and experienced demographic declines at the onset of the Holocene (Figure 5), in agreement with palaeoclimatic-based reconstructions of species distributions (Figure 1). Inference of changes of  $N_e$  through time did not reveal obvious differences between populations located in large vs. small sky islands, suggesting that contemporary populations, and their ancestral sources, reacted in a similar fashion to Quaternary climatic oscillations independently of their current geographical location. This is not surprising given the small range of the two focal taxa (Presa et al., 2016a, b), with most distant populations separated by <150 km, and the likely similar regional impact of Pleistocene glacial cycles across the entire distribution of the species. Also, it must be considered that, according to our own field observations, small habitat patches can often sustain high local densities of the species, which might contribute to maintain effective population sizes above a certain threshold and avoid strong genetic drift even in populations currently confined to tiny sky islands. Collectively, these results indicate that although the genetic makeup of contemporary populations has been shaped in a great extent by historical processes of genetic admixture (glacial periods) and drift during the retreat to inter-glacial refugia (e.g., Knowles & Massatti, 2017), contemporary isolation has also contributed to erode the levels of genetic diversity of populations persisting in peripheral sky islands of small size (Guo, 2012; Lira-Noriega & Manthey, 2014; Rubidge et al., 2012).

#### 4.4 | Conclusions

By considering two biological replicates from the sky island archipelago of the Baetic System, our study highlights the potential of integrating different sources of information to infer the evolutionary and demographic processes shaping spatial patterns of genetic variation in cold-adapted faunas from temperate regions. Our spatiotemporally explicit simulations and testing of alternative demographic scenarios supported population expansion during cold periods and subsequent postglacial colonization and isolation in sky islands as the most likely explanation for the current distribution of genetic variation in both *O. bolivari* and *O. femoralis*. Global warming is expected to threaten the persistence of alpine and cold adapted organisms from mid and low latitude regions by reducing the availability of suitable habitats confined to mountain tops (Moritz & Agudo, 2013; Rubidge et al., 2012). The similar demographic responses inferred for the two focal taxa to past

climate changes suggest that they and other co-distributed organisms with similar life-history traits will probably present concerted responses to ongoing global change. Thus, our results can help to establish unified conservation strategies aimed at preserving the extraordinary rates of local endemism of this and other mountain biodiversity hotspots, from a community-level perspective (Rahbek et al., 2019).

## ACKNOWLEDGEMENTS

We are much indebted to Anna Papadopoulou for her valuable help in study design and useful comments, suggestions and corrections on a first draft of the manuscript. We also wish to thank to Pedro J. Cordero and Víctor Noguerales for their help during fieldwork and providing samples from some populations, Amparo Hidalgo for support during laboratory work, Maria Lucena for her insightful comments and stimulating discussion on the interpretation of some of the demographic analyses and Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) and Laboratorio de Sistemas de Información Geográfica y Teledetección (LAST-EBD) from Estación Biológica de Doñana. We also thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. This study was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (CGL2014-54671-P and CGL2017-83433-P). VT was supported by an FPI predoctoral fellowship (BES-2015-73159) from Ministerio de Economía y Competitividad.

## AUTHOR CONTRIBUTIONS

VT and JO conceived and designed the study, collected the samples, and analyzed the data. VT performed the laboratory work. JO and VT wrote the manuscript.

## REFERENCES

- Atkins, Z. S., Amor, M. D., Clemann, N., Chapple, D. G., While, G. M., Gardner, M. G., . . . Robert, K. A. (2020). Allopatric divergence drives the genetic structuring of an endangered alpine endemic lizard with a sky-island distribution. *Animal Conservation*, *23*(1), 104-118. doi:10.1111/acv.12519
- Beaumont, M. A., Zhang, W. Y., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025-2035.
- Bemmels, J. B., Knowles, L. L., & Dick, C. W. (2019). Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(17), 8431-8436. doi:10.1073/pnas.1901656116
- Bennett, K. D., & Provan, J. (2008). What do we mean by refugia? *Quaternary Science Reviews*, *27*(27-28), 2449-2455. doi:10.1016/j.quascirev.2008.08.019
- Boulesteix, A. L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, *8*(1), 32-44. doi:10.1093/bib/bb1016
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J. Y., Abe-Ouchi, A., . . . 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. doi:10.5194/cp-3-261-2007
- Braga, J. C., Martin, J. M., & Quesada, C. (2003). Patterns and average rates of late Neogene-Recent uplift of the Betic Cordillera, SE Spain. *Geomorphology*, *50*(1-3), 3-26. doi:10.1016/s0169-555x(02)00205-2
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. doi:10.1111/mec.12354
- Clemente, M. E., García, M. D., & Presa, J. J. (1991). Los Gomphocerinae de la península Iberica: 2. *Omocestus* Bolivar, 1878. (Insecta, Orthoptera, Caelifera). *Graellsia*, *46*, 191-246.
- Cooper, S. J. B., Ibrahim, K. M., & Hewitt, G. M. (1995). Postglacial expansion and genome subdivision in the European grasshopper *Chorthippus parallelus*. *Molecular Ecology*, *4*(1), 49-60. doi:10.1111/j.1365-294X.1995.tb00191.x
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & Francois, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*(7), 410-418. doi:10.1016/j.tree.2010.04.001
- Currat, M., Ray, N., & Excoffier, L. (2004). SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, *4*(1), 139-142. doi:10.1046/j.1471-8286.2003.00582.x
- Davis, E. B., McGuire, J. L., & Orcutt, J. D. (2014). Ecological niche models of mammalian glacial refugia show consistent bias. *Ecography*, *37*(11), 1133-1138. doi:10.1111/ecog.01294
- DeChaine, E. G., & Martin, A. P. (2005). Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, *32*(11), 1943-1956. doi:10.1111/j.1365-2699.2005.01356.x

- Duncan, S. I., Crespi, E. J., Mattheus, N. M., & Rissler, L. J. (2015). History matters more when explaining genetic diversity within the context of the core-periphery hypothesis. *Molecular Ecology*, *24*(16), 4323-4336. doi:10.1111/mec.13315
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361. doi:10.1007/s12686-011-9548-7
- Estoup, A., Baird, S. J. E., Ray, N., Currat, M., Cornuet, J. M., Santos, F., . . . Excoffier, L. (2010). Combining genetic, historical and geographical data to reconstruct the dynamics of bioinvasions: application to the cane toad *Bufo marinus*. *Molecular Ecology Resources*, *10*(5), 886-901. doi:10.1111/j.1755-0998.2010.02882.x
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Fordham, D. A., Brook, B. W., Moritz, C., & Noguez-Bravo, D. (2014). Better forecasts of range dynamics using genetic data. *Trends in Ecology & Evolution*, *29*(8), 436-443. doi:10.1016/j.tree.2014.05.007
- Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss & N. Ferrand (Eds.), *Phylogeography of southern European refugia* (pp. 155-188). Dordrecht, The Netherlands: Springer. doi.org/10.1007/1-4020-4904-8
- González-Serna, M. J., Cordero, P. J., & Ortego, J. (2019). Spatiotemporally explicit demographic modelling supports a joint effect of historical barriers to dispersal and contemporary landscape composition on structuring genomic variation in a red-listed grasshopper. *Molecular Ecology*, *28*(9), 2155-2172. doi:10.1111/mec.15086
- Guo, Q. F. (2012). Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology*, *21*(22), 5396-5403. doi:10.1111/mec.12012
- Hasumi, H., & Emori, S. (2004). *K-1 Coupled GCM (MIROC) Description*. Tokyo, Japan: Center for Climate System Research (CCSR), University of Tokyo; National Institute for Environmental Studies (NIES); Frontier Research Center for Global Change (FRCGC).
- Hawks, J., Hunley, K., Lee, S. H., & Wolpoff, M. (2000). Population bottlenecks and Pleistocene human evolution. *Molecular Biology and Evolution*, *17*(1), 2-22.
- He, Q. X., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, *67*(12), 3386-3402. doi:10.1111/evo.12159
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247-276.
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907-913.

- Hewitt, G.M. (2011). Mediterranean Peninsulas: The evolution of hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity Hotspots* (pp. 123–147). Berlin, Heidelberg: Springer.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. doi:10.1002/joc.1276
- Hughes, P. D., & Woodward, J. C. (2017). Quaternary glaciation in the Mediterranean mountains: a new synthesis. In P. D. Hughes & J. C. Woodward (Eds.), *Quaternary Glaciation in the Mediterranean Mountains* (Vol. 433, pp. 1–23). Bath: Geological Soc Publishing House.
- Hutchison, D. W., & Templeton, A. R. (1999). Correlation of 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.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. doi:10.1093/bioinformatics/btm233
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.
- Jombart, T. (2008). *Adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi:10.1093/bioinformatics/btn129
- Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., . . . Wolff, E. W. (2007). Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, *317*(5839), 793–796. doi:10.1126/science.1141038
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.1080/01621459.1995.10476572
- Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, *196*(1), 313–320. doi:10.1534/genetics.113.158758
- Kidane, Y. O., Steinbauer, M. J., & Beierkuhnlein, C. (2019). Dead end for endemic plant species? A biodiversity hotspot under pressure. *Global Ecology and Conservation*, *19*, e00670. doi:10.1016/j.gecco.2019.e00670
- Knowles, L. L. (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, *54*(4), 1337–1348.
- Knowles, L. L., & Alvarado-Serrano, D. F. (2010). Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, *19*(17), 3727–3745. doi:10.1111/j.1365-294X.2010.04702.x
- Knowles, L. L., & Massatti, R. (2017). Distributional shifts - not geographic isolation - as a probable driver of montane species divergence. *Ecography*, *40*(12), 1475–1485. doi:10.1111/ecog.02893

- Leuenberger, C., & Wegmann, D. (2010). Bayesian computation and model selection without likelihoods. *Genetics*, *184*(1), 243-252. doi:10.1534/genetics.109.109058
- Lira-Noriega, A., & Manthey, J. D. (2014). Relationship of genetic diversity and niche centrality: A survey and analysis. *Evolution*, *68*(4), 1082-1093. doi:10.1111/evo.12343
- Liu, C. R., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*(3), 385-393. doi:10.1111/j.0906-7590.2005.03957.x
- Liu, X. M., & Fu, Y. X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, *47*(5), 555-U172. doi:10.1038/ng.3254
- Mairal, M., Sanmartin, I., Herrero, A., Pokorny, L., Vargas, P., Aldasoro, J. J., & Alarcon, M. (2017). Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afromontane biodiversity hotspot. *Scientific Reports*, *7*, 45749. doi:10.1038/srep45749
- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, *25*(16), 3974-3986. doi:10.1111/mec.13735
- Mastretta-Yanes, A., Xue, A. T., Moreno-Letelier, A., Jorgensen, T. H., Alvarez, N., Pinero, D., & Emerson, B. C. (2018). Long-term insitu persistence of biodiversity in tropical sky islands revealed by landscape genomics. *Molecular Ecology*, *27*(2), 432-448. doi:10.1111/mec.14461
- McCormack, J. E., Bowen, B. S., & Smith, T. B. (2008). Integrating paleoecology and genetics of bird populations in two sky island archipelagos. *BMC Biology*, *6*, 28. doi:10.1186/1741-7007-6-28
- Metcalfe, J. L., Prost, S., Nogues-Bravo, D., DeChaine, E. G., Anderson, C., Batra, P., . . . Guralnick, R. P. (2014). Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study. *Proceedings of the Royal Society B-Biological Sciences*, *281*(1777), 20132782. doi:10.1098/rspb.2013.2782
- Meulenkamp, J. E., & Sissingh, W. (2003). Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African-Eurasian convergent plate boundary zone. *Palaeogeography Palaeoclimatology Palaeoecology*, *196*(1-2), 209-228. doi:10.1016/s0031-0182(03)00319-5
- Mevik, B. H., & Wehrens, R. (2007). The *pls* package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, *18*(2), 1-23.
- Moritz, C., & Agudo, R. (2013). The future of species under climate change: Resilience or decline? *Science*, *341*(6145), 504-508. doi:10.1126/science.1237190
- Mota, J. F., Pérez-García, F. J., Jiménez, M. L., Amate, J. J., & Peñas, J. (2002). Phytogeographical relationships among high mountain areas in the Baetic Ranges (South Spain). *Global Ecology and Biogeography*, *11*(6), 497-504. doi:10.1046/j.1466-822X.2002.00312.x

- Mouret, V., Guillaumet, A., Cheylan, M., Pottier, G., Ferchaud, A. L., & Crochet, P. A. (2011). The legacy of ice ages in mountain species: post-glacial colonization of mountain tops rather than current range fragmentation determines mitochondrial genetic diversity in an endemic Pyrenean rock lizard. *Journal of Biogeography*, *38*(9), 1717-1731. doi:10.1111/j.1365-2699.2011.02514.x
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198-1205. doi:10.1111/2041-210x.12261
- Neuenschwander, S., Lurgiader, C. R., Ray, N., Currat, M., Vonlanthen, P., & Excoffier, L. (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, *17*(3), 757-772.
- Noguerales, V., Cordero, P. J., & Ortego, J. (2016). Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, *16*, 96. doi:10.1186/s12862-016-0663-7
- Nogués-Bravo, D. (2009). Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, *18*(5), 521-531. doi:10.1111/j.1466-8238.2009.00476.x
- Palacios, D., Gómez-Ortiz, A., Andres, N., Salvador, F., & Oliva, M. (2016). Timing and new geomorphologic evidence of the last deglaciation stages in Sierra Nevada (southern Spain). *Quaternary Science Reviews*, *150*, 110-129. doi:10.1016/j.quascirev.2016.08.012
- Pan, D., Hulber, K., Willner, W., & Schneeweiss, G. M. (2020). An explicit test of Pleistocene survival in peripheral versus nunatak refugia in two high mountain plant species. *Molecular Ecology*, *29*(1), 172-183. doi:10.1111/mec.15316
- Papadopoulou, A., & Knowles, L. L. (2015a). Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology*, *24*(16), 4252-4268. doi:10.1111/mec.13305
- Papadopoulou, A., & Knowles, L. L. (2015b). Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution*, *69*(6), 1501-1517. doi:10.1111/evo.12667
- Paz, A., Ibañez, R., Lips, K. R., & Crawford, A. J. (2015). Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. *Molecular Ecology*, *24*(14), 3723-3737. doi:10.1111/mec.13275
- Peñas, J., Pérez-García, F. J., & Mota, J. F. (2005). Patterns of endemic plants and biogeography of the Baetic high mountains (south Spain). *Acta Botanica Gallica*, *152*(3), 347-360. doi:10.1080/12538078.2005.10515494
- Perrigo, A., Hoorn, C., & Antonelli, A. (2020). Why mountains matter for biodiversity. *Journal of Biogeography*, *47*(2), 315-325. doi:10.1111/jbi.13731

- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. doi:10.1371/journal.pone.0037135
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231-259. doi:10.1016/j.ecolmodel.2005.03.026
- Phillips, S. J., & Dudik, M. (2008). Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31(2), 161-175. doi:10.1111/j.0906-7590.2008.5203.x
- Presa, J. J., García, M., Clemente, M., Barranco Vega, P., Correas, J., Ferreira, S., ... Prunier, F. (2016a). *Omocestus femoralis*: The IUCN Red List of Threatened Species (Publication no. <https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T16084447A75088712.en>). Retrieved from <https://www.iucnredlist.org/species/16084447/75088712>
- Presa, J. J., García, M., Clemente, M., Barranco Vega, P., Correas, J., Ferreira, S., ... Prunier, F. (2016b). *Omocestus bolivari*: The IUCN Red List of Threatened Species (Publication no. <https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T16084452A75088604.en>). Retrieved from <https://www.iucnredlist.org/species/16084452/75088604>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Prunier, F. (2014). *Omocestus bolivari* Chopard, 1939 (Orthoptera: Acrididae), new species for the province of Jaen. [*Omocestus bolivari* Chopard, 1939 (Orthoptera: Acrididae), especie nueva para la provincia de Jaen.]. *Zoologica Baetica*, 25, 75-78.
- Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., . . . Fjeldsa, J. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science*, 365(6458), 1114-1119. doi:10.1126/science.aax0151
- Ramírez-Barahona, S., & Eguarte, L. E. (2014). Changes in the distribution of cloud forests during the last glacial predict the patterns of genetic diversity and demographic history of the tree fern *Alsophila firma* (Cyatheaceae). *Journal of Biogeography*, 41(12), 2396-2407. doi:10.1111/jbi.12396
- Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, 26(23), 2993-2994. doi:10.1093/bioinformatics/btq579
- Reid, B. N., Kass, J. M., Wollney, S., Jensen, E. L., Russello, M. A., Viola, E. M., . . . Naro-Maciel, E. (2019). Disentangling the genetic effects of refugial isolation and range expansion in a trans-continentally distributed species. *Heredity*, 122(4), 441-457. doi:10.1038/s41437-018-0135-5
- Ribera, I., & Vogler, A. P. (2004). Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, 13(1), 179-193.
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x

- Rubidge, E. M., Patton, J. L., Lim, M., Burton, A. C., Brashares, J. S., & Moritz, C. (2012). Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nature Climate Change*, *2*(4), 285-288. doi:10.1038/nclimate1415
- Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J. C. (2011). The influence of Late Quaternary climate-change velocity on species endemism. *Science*, *334*(6056), 660-664. doi:10.1126/science.1210173
- Slatkin, M. (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, *47*(1), 264-279.
- Stewart, J. R., Lister, A. M., Barnes, I., & Dalen, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences*, *277*(1682), 661-671. doi:10.1098/rspb.2009.1272
- Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*(4), 453-464.
- Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology*, *28*(17), 3869-3886. doi:10.1111/mec.15189
- Wachter, G. A., Papadopoulou, A., Muster, C., Arthofer, W., Knowles, L. L., Steiner, F. M., & Schlick-Steiner, B. C. (2016). Glacial refugia, recolonization patterns and diversification forces in Alpine-endemic *Megabunus* harvestmen. *Molecular Ecology*, *25*(12), 2904-2919. doi:10.1111/mec.13634
- Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in MAXENT: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335-342. doi:10.1890/10-1171.1
- Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCTOOLBOX: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*, 116. doi:10.1186/1471-2105-11-116

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article and in Appendix IV of this Thesis.



# GENERAL DISCUSSION





## General discussion

This thesis exemplifies the importance of integrating multiple sources of information (genomic, phenotypic and environmental data) and analytical approaches from different disciplines (phylogenomic, population genetics, species distribution modelling) to infer diversification processes across the different stages of the **continuum of speciation**, from cladogenesis and lineage formation to more recent phenomena shaping the contemporary spatial distribution of genomic variation within species. The results obtained in this thesis also represent the first major contribution to the understanding of the systematics and evolution of the **Ibero-Maghrebian** grasshopper subgenus *Dreuxius* (genus *Omocestus*), which provides important insights into the processes underlying the high biodiversity and rates of local endemism in montane biotas from this **Mediterranean biodiversity hotspot** (Krijgsman, 2002; Medail & Quezel, 1999). Although this thesis primarily has an evolutionary standpoint, the findings obtained also have important implications for biodiversity conservation. Among others, these include the resolution of old taxonomic problems involving red-listed species of conservation concern. Important insights are also provided into the processes determining the long-term persistence of species forming highly fragmented populations in sky-island habitats that are expected to be especially vulnerable to the consequences of ongoing global warming (Rubidge et al., 2012).

The obtained results respond to the main objectives originally contemplated in this thesis and provide a solid starting point to further deepen into many other relevant questions, such as those related with the evolution of local adaptations to contrasting environmental conditions or the study of the genomic basis of reproductive isolation (e.g., Garner, Goulet, Farnitano, Molina-Henao, & Hopkins, 2018; Yeaman et al., 2016). The specific results of this thesis have been discussed in detail in each chapter. In this section, I focus on summarizing and discussing some of the most relevant findings with the main purpose of providing a global and integrative vision of the different processes that shaped genomic and phenotypic variation across different evolutionary scales along the speciation continuum of the studied species complex.

## 1 | Evolutionary biogeography of the species complex

The results of phylogenomic reconstructions and estimates of divergence time inferred in Chapter 1 supported that the diversification of the subgenus *Dreuxius* took place over the last **800 ka**. Phylogenomic analyses revealed three main cladogenetic events and indicate that all contemporary species have a very **recent origin** (<250 ka), supporting **rapid speciation** starting at the end of the Middle Pleistocene (Figure 1A) (Knowles, 2008; Pouchon et al., 2018). First, the Maghrebian lineages (*O. alluaudi* and *O. lepineyi*) split (ca. 800 ka) from the clade including all Iberian taxa and the north African *O. lecerfi*. This second group differentiated into two main clades (ca. 400 ka), one formed by the Pyrenean *O. antigai* and the Baetican *O. femoralis* and another comprising the rest of Iberian species plus the Maghrebian *O. lecerfi*. Note that the former taxa *O. antigai* and *O. navasi* were synonymized in Chapter 2 and considered as a single species in the phylogenetic reconstructions of the complex. Surprisingly, our analyses **did not recover reciprocal monophyly of north African and Iberian taxa**, supporting **two trans-continental colonization** events through the Strait of Gibraltar or adjacent areas. In agreement with previous descriptive assessments of species relationships based on morphological and behavioural comparisons (Gangwere & Morales Agacino, 1970), our analyses supported the sisterhood of *O. antigai* and *O. femoralis* and a close relationship between *O. bolivari*, *O. minutissimus* and *O. uhagonii*. With the exception of *O. minutissimus*, all taxa of the complex are strict montane species restricted to high elevations (>1,300 m) in different mountain ranges. Considering the Pleistocene origin and the current species distribution, the results of this chapter point to the **Pleistocene glacial cycles** and the **high amplitude climatic variations** characterizing the **Late Pleistocene** as the main drivers of **lineage divergence** through elevational and latitudinal range-shifts (Avice, Walker, & Johns, 1998; Knowles, 2000; Schmitt, 2007). Glacial periods also reduced the Mediterranean Sea level and considerably shortened the distance between Northwest Africa and southern Iberia to less than 5 km, which might have led to the emergence of small islands and shoals and facilitated the exchange of biotas between the two continents during the coldest stages of the Pleistocene (Agustí, Garcés, & Krijgsman, 2006; Collina-Girard, 2001; Cosson et al., 2005). These results add to the accumulating empirical evidence supporting the migration of numerous organisms across the two continents, either seeking for glacial refugia in Northwest Africa (Husemann, Schmitt, Zachos, Ulrich, & Habel, 2014) or following post-glacial colonization routes to Europe (Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998;

Teacher, Garner, & Nichols, 2009). Habitat preferences and environmental niches show strong similarities across all species (Clemente, García, & Presa, 1991; Ragge, 1986), which points to allopatric speciation, rather than ecological divergence, as the predominant mechanism of diversification (He et al., 2019; Hewitt, 2004; Taberlet et al., 1998). Topographically complex regions, such as the Iberian Peninsula and Northwest Africa, offer an ideal biogeographic setting for allopatric speciation. Isolation in valleys during glacial periods (Wallis, Waters, Upton, & Craw, 2016) and confinement in sky-islands during interglacials (i.e., interglacial refugia; Bennett & Provan, 2008; Stewart, Lister, Barnes, & Dalén, 2010) are expected to have led to extended periods of isolation and divergence through genetic drift and/or divergent selection pressures (Djamali et al., 2012; Hewitt, 1996).

In the particular case of the Pyrenean taxa *O. antigai* and *O. navasi* studied in Chapter 2, subtle phenotypic diversity likely led to the controversial differentiation of these two species, which were described by Bolívar (1897, 1908) and their taxonomic status was confirmed later by Clemente et al. (1999). These authors largely based the distinction between the two putative species on their specific habitat associations: *Omocestus antigai* was distributed at high elevations (1,600–2,200 m) and associated with alpine or subalpine grasslands above the tree line, whereas *O. navasi* was circumscribed to Mediterranean scrubby habitats at lower elevations (1000–1,600 m; Lluçia-Pomares, 2002; Olmo-Vidal, 2002; Poniatowski et al., 2009). Using genomic and phenotypic data, our species delimitation analyses indicate no support for the current taxonomic designation. On the one hand, geometric morphometric analyses did not show consistent **phenotypic differentiation** between the two putative taxa. On the other hand, phylogenomic and genetic structure analyses indicated that sampled populations of both species split in two main genetic groups corresponding to the East and West Pyrenees (Figure 1B), which is analogous to what has been found in many other Pyrenean taxa (Wallis et al., 2016). As a consequence, populations distributed in distinct habitats (alpine vs. Mediterranean/montane) supposedly occupied by the two putative species are phylogenetically interspersed. Genetic clustering analyses revealed a longitudinal cline of genetic structure rather than a segregation of alpine and Mediterranean-montane populations, providing no support for **ecologically driven divergence**, neither for the **taxonomic separation** between the supposedly Mediterranean *O. navasi* and the alpine *O. antigai*.

## 2 | The role of Pleistocene glaciations on structuring population genomic variation in the subgenus *Dreuxius*

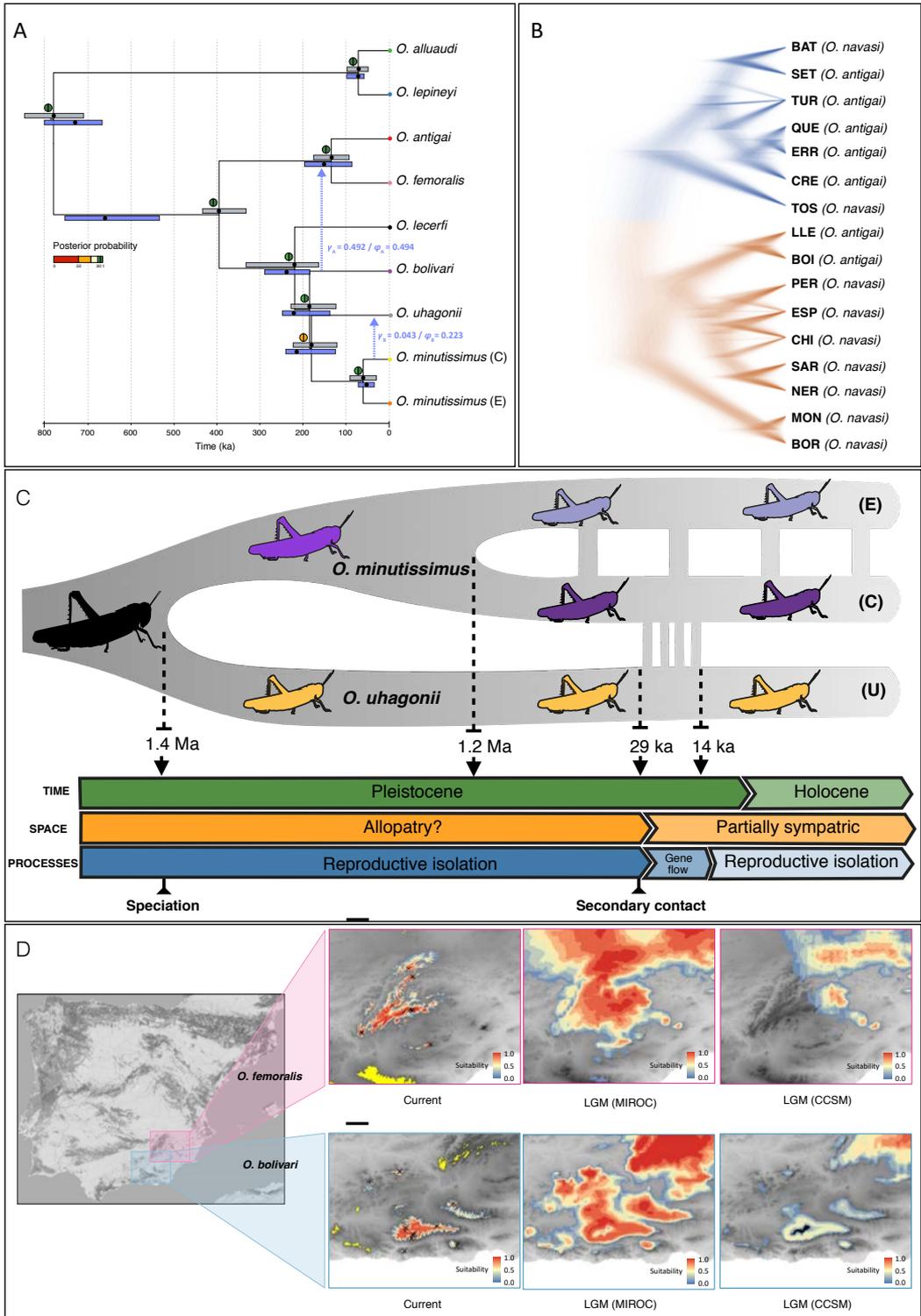
In Chapter 4 we combined genomic data, niche modelling and spatiotemporally explicit coalescent-based simulations to investigate the consequences of range-shifts driven by Pleistocene glaciations on the demography of *O. bolivari* and *O. femoralis*, two species presenting adjacent but non-overlapping distributions in the sky island archipelago of the Baetic System (Figure 1D). Both species showed a marked northwest-southeast pattern of genetic differentiation that reflects the **connectivity** and origin of **source populations** that later led to the **colonization of high elevation interglacial refugia**. Accordingly, our model-based approach supported that contemporary patterns of genetic diversity and structure in both species are best explained by a scenario considering population expansion to lowland areas during glacial periods followed by colonization of high elevation refugia and isolation in sky islands during interglacial periods. Furthermore, the spatial distribution of genetic variation in the two species (see Chapter 4, Figure 1A, B) evidenced that peripheral populations and those located in smaller sky islands tend to present comparatively lower levels of genetic diversity. These results are consistent with the **center-periphery model**, which predicts that populations located on the species range edges become progressively smaller, are spatially more isolated, and harbor lower levels of genetic diversity than populations situated at the core of the distribution (e.g., Guo, 2012; Lira-Noriega & Manthey, 2014).

Although the study of the effects of Quaternary glaciations was not the main objective of Chapter 2, the east-west genetic split in the population structure of *O. antigai* likely reflects the presence of **glacial refugia** at both longitudinal extremes of the Pyrenees. These putative refugia might have allowed the species to survive glaciations, as the highest elevations and central massifs were, according to geological reconstructions, heavily glaciated during the coldest periods and could be only recolonized during the interglacials (Charrier, Dupont, Pornon, & Escaravage, 2014; Valbuena-Ureña et al., 2018). Thus, two very different situations are exemplified in the studied species complex: contemporary isolation in sky-islands and demographic expansions and admixture in glacial periods in the meridional *O. femoralis* and *O. bolivari* vs. population fragmentation during glacial periods and recolonization of suitable habitats during interglacials in the septentrional *O. antigai*.

### 3 | A reticulate evolutionary history

Different lines of evidence support **allopatric divergence** as the predominant mode of speciation in the studied complex (Mayer, Berger, Gottsberger, & Schulze, 2010), which was probably fuelled by the extensive fragmentation of ancestral populations driven by **Quaternary climatic oscillations** (e.g., Huang, Hill, Ortego, & Knowles, 2020; Knowles, 2000; Scattolini, Confalonieri, Lira-Noriega, Pietrokovsky, & Cigliano, 2018). However, different studies have also documented that **range shifts** during **Pleistocene climatic oscillations** have led to **secondary contact**, resulting in the collapse of formerly distinct lineages (i.e., **speciation reversal**; Kearns et al., 2018; Maier et al., 2019), **introgressive hybridization** (Hewitt, 2008; Seixas et al., 2018), or even contributed in some cases to the completion of the **speciation** process through the evolution of reproductive isolation (i.e., reinforcement; Butlin & Hewitt, 1985; Hewitt, 1996; Nevado et al., 2018). Our analyses clearly supported that **post-divergence gene flow** is also an important component of the evolutionary portrait of the *Dreuxius* species complex and might have had an important role in the **speciation process** and the evolution of **reproductive isolation mechanisms**.

Phylogenetic network analyses in Chapter 1 revealed **signatures of historical gene flow** from *O. bolivari* to the common ancestor of *O. femoralis* and *O. antigai* and from *O. minutissimus* to *O. uhagonii* (Figure 1A) that likely explain the weak support of internal nodes when inferring phylogenetic relationships in the studied species complex. Thus, phylogenomic analyses supported that some species within the complex have experienced **post-divergence gene flow**, likely as consequence of secondary contact resulting from range shifts driven by Pleistocene glacial cycles. In Chapter 3 we evaluated in more detail alternative hypothetical scenarios of interspecific gene flow between *O. minutissimus* and *O. uhagonii*, two taxa showing partially overlapping distributions in the Central System Mountains from the Iberian Peninsula. The results obtained from phylogenomic and coalescent-based simulation analyses point to a scenario in which *O. uhagonii* and ancestral populations of *O. minutissimus* likely diverged in allopatry followed by the split of *O. minutissimus* into two lineages, one of which came into secondary contact and hybridized with *O. uhagonii* during a limited period of time (Figure 1C). Our analyses did not find any evidence of contemporary hybridization, which suggests that reproductive isolation likely evolved after secondary contact and **historical gene flow**. It has been frequently documented that interspecific gene flow can increase phenotypic and genomic divergence via the evolution of **reproductive isolation** and **character displacement**



(Garner et al., 2018; Hopkins, Levin, & Rausher, 2012; Pfennig & Pfennig, 2009). One of the many potential costs of hybridization are the ecological and genetic dysfunctions of hybrid offspring, which can **reduce their fitness** and drive to **reinforcement** (Ortiz-Barrientos, Counterman, & Noor, 2004; Servedio & Noor, 2003). Thus, one possibility is that secondary contact and hybridization promoted the evolution of reproductive isolation via reinforcement, probably after an initial balance between dispersal and selection against hybrids in historical **tension zones** during which the two species experienced **genetic exchange** and **introgression** (Barton & Hewitt, 1985).

Future research should focus on analysing **mating preferences**, **phenotypic differentiation** and **reproductive character displacement** (song, courtship behaviour, genital structures, etc.) among species involved and not involved in historical hybridization (e.g., Butlin & Ritchie, 1991; Hollander, Smadja, Butlin, & Reid, 2013) and on generating whole genome or transcriptome sequencing data to detect potential genomic **signatures of reproductive isolation** processes (e.g., reinforcement) that might have contributed to the completion of the speciation process (Gagnaire, Pavey, Normandeau, & Bernatchez, 2013; Garner et al., 2018).

**Figure 1** Main results obtained in this thesis. (A) Species tree (see Glossary) including estimates of divergence time and inferred introgression events. The species tree was reconstructed with SNAPP and BPP and posterior probabilities of node support are indicated for each analysis in colored semi circles (left: SNAPP; right: BPP). Dots (median) and bars (95% highest posterior density intervals) indicate divergence times estimated by BPP, colored in grey for standard analysis not considering post-divergence gene flow (MSC model) and blue for analyses accounting for introgression events (MSCi model) inferred using PHYLONETWORKS. Blue arrows indicate inferred introgression events with their corresponding inheritance values ( $\nu$ ) estimated by PHYLONETWORKS and introgression probabilities ( $\varphi$ ) estimated by BPP (not time-scaled). Lineages of *Omocestus minutissimus* C and E correspond to the distribution of the species in the Central System and eastern Iberian, respectively, and are also detailed in panel C. (B) Bayesian phylogenetic tree reconstructed with SNAPP of sampled populations of *O. antigai* and *O. navasi* in the Pyrenees. Population codes as in Table 1 of Chapter 2; (C) Best-fit demographic model inferred using FASTSIMCOAL2 for *O. uhagonii* and sympatric (C) and allopatric (E) populations of *O. minutissimus*. Vertical bars connecting the two species represent historical gene flow. Note that geological reference time is not scaled and only point estimates inferred by FASTSIMCOAL2 are presented to simplify visualization. The inferred biological processes are indicated. (D) Environmental niche models (ENM) for *O. bolivari* and *O. femoralis*. Insets show the projections of ENMs for current and last glacial maximum (LGM) conditions under the CCSM4 and MIROC-ESM general atmospheric circulation models. Dark yellow color for current maps indicate areas predicted as suitable out of the known limits of species' ranges (i.e., over-predictions). Maps for current distributions show records (crosses) used to build ENMs. Grey background represents elevation, with darker areas corresponding to higher altitudes.



## CONCLUSIONS



## Conclusions

- 1** Phylogenomic reconstructions and estimates of divergence time supported a recent diversification (< 1 Ma) of the *Dreuxius* species complex and point to Pleistocene glacial cycles as the most likely driver of speciation.
- 2** The inferred phylogenetic relationships and divergence time estimates supported post-Messinian dispersal between the Iberian Peninsula and Northwest Africa through the Strait of Gibraltar, probably favoured by glacial eustasy during the Pleistocene.
- 3** Phylogenetic network analyses revealed two historical events of interspecific gene flow involving Iberian lineages: one reticulation edge connecting *O. bolivari* and the ancestor of *O. antigai* and *O. femoralis* and another involving the partially sympatric *O. minutissimus* and *O. uhagonii*.
- 4** These results support speciation-with-gene-flow and emphasize the importance of considering reticulation events in phylogenetic reconstructions of recent evolutionary radiations.
- 5** The integration of genomic, phenotypic and ecological data did not support the current taxonomic distinctiveness between the Pyrenean taxa *O. navasi* and *O. antigai*, indicating the presence of a single species characterized by a strong genetic structure, little phenotypic variation, and a wide environmental niche.
- 6** *Omocestus navasi* and *O. antigai* must be synonymized into a unique taxon: *O. antigai* (Bolívar, 1897). The range and population sizes of *O. antigai* are larger and the ecological and habitat requirements much wider than previously thought and, thus, the conservation status of the species should be reconsidered in future IUCN Red List assessments.
- 7** *Omocestus antigai* presents a hierarchical genetic structure, with a deep split separating eastern and western populations. Genetic differentiation among populations was mainly explained by their geographical distances and topographic roughness, which points to the important role of the abrupt Pyrenean landscapes in structuring genomic variation in this species.
- 8** Phylogenomic inference, bayesian clustering analyses, and coalescent-based model testing revealed historical hybridization between *O. uhagonii* and sympatric populations of *O. minutissimus* in the Central System, but did not support the presence of contemporary gene flow. Allopatric populations of *O. minutissimus* from eastern Iberia showed no

signatures of introgressive hybridization.

**9** The inferred patterns of interspecific gene flow suggest that *O. uhagonii* and *O. minutissimus* likely diverged in allopatry, came into secondary contact, hybridized during a limited period of time and, subsequently, evolved reproductive isolation mechanisms that currently prevent gene flow among sympatric populations of the two species.

**10** Spatiotemporally explicit simulations and testing of alternative demographic scenarios for *O. bolivari* and *O. femoralis* supported population expansions during Pleistocene cold periods and subsequent postglacial colonization and isolation in sky islands as the most likely explanation for the current distribution of genetic variation in both species.

**11** Post-glacial range fragmentation and isolation in sky-islands has contributed to reduced levels of within-population genetic diversity in *O. bolivari* and *O. femoralis* through the disruption of inter-population gene flow.





## REFERENCES



## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. doi: 10.1111/j.1420-9101.2012.02599.x
- Agustí, J., Garcés, M., & Krijgsman, W. (2006). Evidence for African–Iberian exchanges during the Messinian in the Spanish mammalian record. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *238*(1–4), 5–14. doi: 10.1016/j.palaeo.2006.03.013
- Atkins, Z. S., Clemann, N., Chapple, D. G., Edwards, A. M., Sinsch, U., Hantzschmann, A. M., ... Robert, K. A. (2020). Demographic and life history variation in two sky-island populations of an endangered alpine lizard. *Journal of Zoology*, *310*(1), 34–44. doi: 10.1111/jzo.12728
- Avise, J. C., Walker, D., & Johns, G. C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1407), 1707–1712. doi: 10.1098/rspb.1998.0492
- Bartomeus, I., Molina, F. P., Hidalgo-Galiana, A., & Ortego, J. (2020). Safeguarding the genetic integrity of native pollinators requires stronger regulations on commercial lines. *Ecological Solutions and Evidence*, in press.
- Baxter, S. W., Davey, J. W., Johnston, J. S., Shelton, A. M., Heckel, D. G., Jiggins, C. D., & Blaxter, M. L. (2011). Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE*, *6*(4), e19315. doi: 10.1371/journal.pone.0019315
- Bennett, K., & Provan, J. (2008). What do we mean by ‘refugia’? *Quaternary Science Reviews*, *27*(27–28), 2449–2455. doi: 10.1016/j.quascirev.2008.08.019
- Blondel, J., Aronson, J., Bodiou, J. Y., & Boeuf, G. (2010). *The Mediterranean Region: Biological Diversity in Space and Time*. Oxford, UK: Oxford University Press.
- Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 459–487. doi: 10.1146/annurev.ecolsys.38.091206.095804
- Bookstein, F.L. (1992). *Morphometric Tools for Landmark Data*. Cambridge, UK: Cambridge University Press.
- Boucher, F. C., Zimmermann, N. E., & Conti, E. (2016). Allopatric speciation with little niche divergence is common among alpine Primulaceae. *Journal of Biogeography*, *43*(3), 591–602. doi: 10.1111/jbi.12652
- Butlin, R. K., & Hewitt, G. M. (1985). A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera: Acrididae): morphological and electrophoretic characters. *Biological Journal of the Linnean Society*, *26*(3), 269–285. doi: 10.1111/j.1095-8312.1985.tb01636.x
- Butlin, R. K., & Ritchie, M. G. (1991). Variation in female mate preference across a grasshopper hybrid zone. *Journal of Evolutionary Biology*, *4*(2), 227–240. doi: 10.1046/j.1420-9101.1991.4020227.x

- Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506), 2997–3007. doi: 10.1098/rstb.2008.0076
- Carranza, S., Harris, D. J., Arnold, E. N., Batista, V., & Gonzalez de la Vega, J. P. (2006). Phylogeography of the lacertid lizard, *Psammodromus algirus*, in Iberia and across the Strait of Gibraltar. *Journal of Biogeography*, 33(7), 1279–1288. doi: 10.1111/j.1365-2699.2006.01491.x
- Chakrabarty, A., & Schielzeth, H. (2020). Comparative analysis of the multivariate genetic architecture of morphological traits in three species of Gomphocerine grasshoppers. *Heredity*, 124(2), 367–382. doi: 10.1038/s41437-019-0276-1
- Charrier, O., Dupont, P., Pornon, A., & Escaravage, N. (2014). Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the Pyrenees. *PLoS ONE*, 9(3), e92976. doi: 10.1371/journal.pone.0092976
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2019). *Orthoptera Species File*. Version 5.0/5.0. Retrieved from <http://orthoptera.speciesfile.org/>
- Clemente, M. E., García, M. D., & Presa, J. J. (1990). Nuevos datos sobre los acridoidea (Insecta: Orthoptera) del Pirineo y Prepirineo catalano-aragones. *Butlletí de La Institució Catalana d'Història Natural*, 58, 37–44.
- Clemente, M. E., García, M. D., & Presa, J. J. (1991). Los Gomphocerinae de la península Ibérica: 2. *Omocestus Bolívar*, 1878. (Insecta, Orthoptera, Caelifera). *Graellsia*, 46, 191–246.
- Clemente, M. E., García, M. D., Arnaldos, M. I., Romera, E., & Presa, J. J. (1999). Corroboration of the specific status of *Omocestus antigai* (Bolívar, 1897) and *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae). [Confirmación de las posiciones taxonómicas específicas de *Omocestus antigai* (Bolívar, 1897) y *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae)]. *Boletín de la Real Sociedad Española de Historia Natural Sección Biológica*, 95(3–4), 27–50.
- Collina-Girard, J. (2001). L'Atlantide devant le détroit de Gibraltar? Mythe et géologie. *Comptes Rendus de l'Académie Des Sciences - Series IIA - Earth and Planetary Science*, 333(4), 233–240. doi: 10.1016/S1251-8050(01)01629-9
- Conflitti, I. M., Shields, G. F., Murphy, R. W., & Currie, D. C. (2014). The speciation continuum: Population structure, gene flow, and maternal ancestry in the *Simulium arcticum* complex (Diptera: Simuliidae). *Molecular Phylogenetics and Evolution*, 78, 43–55. doi: 10.1016/j.ympev.2014.05.001
- Cosson, J.-F., Hutterer, R., Libois, R., Sarà, M., Taberlet, P., & Vogel, P. (2005). Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, 14(4), 1151–1162. doi: 10.1111/j.1365-294X.2005.02476.x
- Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*, 43(2), 362–381. doi:10.2307/2409213

- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA, USA: Sinauer Association.
- Czekanski-Moir, J. E., & Rundell, R. J. (2019). The ecology of nonecological speciation and nonadaptive radiations. *Trends in Ecology & Evolution*, *34*(5), 400–415. doi: 10.1016/j.tree.2019.01.012
- De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. doi:10.1080/10635150701701083
- DeChaine, E. G., & Martin, A. P. (2005). Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, *32*(11), 1943–1956. doi: 10.1111/j.1365-2699.2005.01356.x
- Djamali, M., Baumel, A., Brewer, S., Jackson, S. T., Kadereit, J. W., Lopez-Vinyallonga, S., ... Simakova, A. (2012). Ecological implications of *Cousinia* Cass. (Asteraceae) persistence through the last two glacial-interglacial cycles in the continental Middle East for the Irano-Turanian flora. *Review of Palaeobotany and Palynology*, *172*, 10–20. doi:10.1016/j.revpalbo.2012.01.005
- Dobzhansky, T. (1937). Genetic nature of species differences. *American Naturalist*, *71*, 404–420. doi:10.1086/280726
- Duggen, S., Hoernle, K., van den Bogaard, P., Rüpke, L., & Phipps Morgan, J. (2003). Deep roots of the Messinian salinity crisis. *Nature*, *422*(6932), 602–606. doi: 10.1038/nature01553
- Fitzpatrick, B. M. (2002). Molecular correlates of reproductive isolation. *Evolution*, *56*(1), 191–198. doi:10.1111/j.0014-3820.2002.tb00860.x
- Fitzpatrick, B. M., Fordyce, J. A., & Gavrillets, S. (2008). What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*, *21*(6), 1452–1459. doi: 10.1111/j.1420-9101.2008.01611.x
- Fitzpatrick, B. M., Fordyce, J. A., & Gavrillets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, *22*(11), 2342–2347. doi: 10.1111/j.1420-9101.2009.01833.x
- Flantua, S. G., & Hooghiemstra, H. (2018). Historical connectivity and mountain biodiversity. In C. Hoorn, A. Perrigo, & A. Antonelli (Eds.), *Mountains, Climate, and Biodiversity* (pp. 171–185). Oxford, UK: Wiley-Blackwell.
- Futuyma, D. (1998). *Evolutionary Biology 3rd edition*. Sunderland, Massachusetts, USA: Sinauer.
- Gagnaire, P. A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, *67*(9), 2483–2497. doi: 10.1111/evo.12075
- Gangwere, S. K., & Morales Agacino, E. (1970). The biogeography of Iberian orthopteroids. *Miscelánea Zoológica*, *2*(5), 9–75.
- Garner, A., Goulet, B., Farnitano, M., Molina-Henao, Y., & Hopkins, R. (2018). Genomic signatures of reinforcement. *Genes*, *9*(4), 191. doi: 10.3390/genes9040191
- Gavrillets, S. (2000). Rapid evolution of reproductive barriers driven by sexual conflict. *Nature*, *403*(6772), 886–889. doi: 10.1038/35002564

- Gittenberger, E. (1991). What about non-adaptive radiation? *Biological Journal of the Linnean Society*, 43(4), 263–272. doi: 10.1111/j.1095-8312.1991.tb00598.x
- Givnish, T. J., & Sytsma, K. J. (Eds.). (1997). *Molecular Evolution and Adaptive Radiation*. Cambridge, UK; New York, USA: Cambridge University Press.
- Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss & N. Ferrand (Eds.), *Phylogeography of Southern European Refugia* (pp. 155–188). Dordrecht, The Netherlands: Springer. doi: 10.1007/1-4020-4904-8
- Grace, T., Wisely, S. M., Brown, S. J., Dowell, F. E., & Joern, A. (2010). Divergent host plant adaptation drives the evolution of sexual isolation in the grasshopper *Hesperotettix viridis* (Orthoptera: Acrididae) in the absence of reinforcement. *Biological Journal of the Linnean Society*, 100(4), 866–878. doi: 10.1111/j.1095-8312.2010.01458.x
- Graciá, E., Giménez, A., Anadón, J. D., Harris, D. J., Fritz, U., & Botella, F. (2013). The uncertainty of Late Pleistocene range expansions in the western Mediterranean: a case study of the colonization of south-eastern Spain by the spur-thighed tortoise, *Testudo graeca*. *Journal of Biogeography*, 40(2), 323–334. doi: 10.1111/jbi.12012
- Guo, Q. (2012). Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology*, 21(22), 5396–5403. doi: 10.1111/mec.12012
- Hays, J. D., Imbrie, J., & Shackleton, N. J. (1976). Variations in the Earth's orbit: Pacemaker of the ice ages. *Science*, 194(4270), 1121–1132. doi: 10.1126/science.194.4270.1121
- He, K., Gutiérrez, E. E., Heming, N. M., Koepfli, K., Wan, T., He, S., ... Jiang, X. (2019). Cryptic phylogeographic history sheds light on the generation of species diversity in sky-island mountains. *Journal of Biogeography*, 46(10), 2232–2247. doi: 10.1111/jbi.13664
- Hedin, M., Carlson, D., & Coyle, F. (2015). Sky island diversification meets the multispecies coalescent - divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Molecular Ecology*, 24(13), 3467–3484. doi: 10.1111/mec.13248
- Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, 75(8), 2000–2036. doi: 10.1111/j.1095-8649.2009.02419.x
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907–913. doi: 10.1038/35016000
- Hewitt, G. M. (1990). Divergence and speciation as viewed from an insect hybrid zone. *Canadian Journal of Zoology*, 68(8), 1701–1715. doi: 10.1139/z90-251
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58(3), 247–276. doi: 10.1111/j.1095-8312.1996.tb01434.x
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68(1–2), 87–112. doi: 10.1111/j.1095-8312.1999.tb01160.x

- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1442), 183–195. doi: 10.1098/rstb.2003.1388
- Hewitt, G. M. (2008). Speciation, hybrid zones and phylogeography - or seeing genes in space and time: seeing genes in space and time. *Molecular Ecology*, 10(3), 537–549. doi: 10.1046/j.1365-294x.2001.01202.x
- Hollander, J., Smadja, C. M., Butlin, R. K., & Reid, D. G. (2013). Genital divergence in sympatric sister snails. *Journal of Evolutionary Biology*, 26(1), 210–215. doi: 10.1111/jeb.12029
- Hopkins, R., Levin, D. A., & Rausher, M. D. (2012). Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution*, 66(2), 469–485. doi: 10.1111/j.1558-5646.2011.01452.x
- Huang, J., Hill, J. G., Ortego, J., & Knowles, L. L. (2020). Paraphyletic species no more – genomic data resolve a Pleistocene radiation and validate morphological species of the *Melanoplus scudderi* complex (Insecta: Orthoptera). *Systematic Entomology*, in press. doi: 10.1111/syen.12415
- Husemann, M., Schmitt, T., Zachos, F. E., Ulrich, W., & Habel, J. C. (2014). Palaeartic biogeography revisited: evidence for the existence of a North African refugium for Western Palaeartic biota. *Journal of Biogeography*, 41(1), 81–94. doi: 10.1111/jbi.12180
- Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., ... Wolff, E. W. (2007). Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, 317(5839), 793–796. doi: 10.1126/science.1141038
- Kawakami, T., Butlin, R. K., Adams, M., Saint, K. M., Paull, D. J., & Cooper, S. J. B. (2007). Differential gene flow of mitochondrial and nuclear DNA markers among chromosomal races of Australian morabine grasshoppers (*Vandiemenella*, *viatica* species group). *Molecular Ecology*, 16(23), 5044–5056. doi: 10.1111/j.1365-294X.2007.03572.x
- Kearns, A. M., Restani, M., Szabo, I., Schröder-Nielsen, A., Kim, J. A., Richardson, H. M., ... Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9(1), 906. doi: 10.1038/s41467-018-03294-w
- Klicka, J., & Zink, R. M. (1997). The importance of recent ice ages in speciation: A failed paradigm. *Science*, 277(5332), 1666. doi: 10.1126/science.277.5332.1666
- Knowles, L. L. (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54(4), 1337–1348. doi: 10.1111/j.0014-3820.2000.tb00566.x
- Knowles, L. L. (2008). Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, 10(3), 691–701. doi: 10.1046/j.1365-294x.2001.01206.x
- Koot, E. M., Morgan-Richards, M., & Trewick, S. A. (2020). An alpine grasshopper radiation older than the mountains, on Kā Tiritiri o te Moana (Southern Alps) of Aotearoa (New Zealand). *Molecular Phylogenetics and Evolution*, 147, 106783. doi: 10.1016/j.ympev.2020.106783

- Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J., & Wilson, D. S. (1999). Chronology, causes and progression of the Messinian salinity crisis. *Nature*, *400*(6745), 652–655. doi: 10.1038/23231
- Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A., & Davis, C. C. (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (*Campanulaceae*). *New Phytologist*, *210*(4), 1430–1442. doi: 10.1111/nph.13920
- Lavergne, S., Hampe, A., & Arroyo, J. (2013). In and out of Africa: how did the Strait of Gibraltar affect plant species migration and local diversification?. *Journal of Biogeography*, *40*(1), 24–36. doi: 10.1111/j.1365-2699.2012.02769.x
- Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*(10), 2594–2609. doi: 10.1111/mec.15091
- Lira-Noriega, A., & Manthey, J. D. (2014). Relationship of genetic diversity and niche centrality: A survey and analysis. *Evolution*, *68*(4), 1082–1093. doi: 10.1111/evo.12343
- Lluçà-Pomares, D. (2002). Revision of the Orthoptera (Insecta) of Catalonia (Spain). [Revisión de los ortópteros (Insecta: Orthoptera) de Cataluña (España)]. *Monografías de la Sociedad Entomológica Aragonesa*, *7*, 1–226.
- Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature*, *457*(7231), 830–836. doi: 10.1038/nature07893
- Maier, P. A., Vandergast, A. G., Ostoja, S. M., Aguilar, A., & Bohonak, A. J. (2019). Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). *Evolution*, *73*(12), 2476–2496. doi: 10.1111/evo.13868
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. doi: 10.1016/j.tree.2005.02.010
- Mallet, J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 2971–2986. doi: 10.1098/rstb.2008.0081
- Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, *26*(9), 448–456. doi: 10.1016/j.tree.2011.05.007
- Mayden, R. L. (1997). A hierarchy of species concepts: the denouement in the saga of the species problem. In M. F. Claridge, H. A. Dawah, & M. R. Wilson (Eds.), *Species: The Units of Diversity*, 381–423. London, UK: Chapman & Hall.
- Mayer, F., Berger, D., Gottsberger, B., & Schulze, W. (2010). Non-ecological radiations in acoustically communicating grasshoppers? In M. Glaubrecht (Ed.), *Evolution in Action* (pp. 451–464). Berlin, Germany: Springer-Verlag.
- Mayr, E. (1942). *Systematics and the Origin of Species*. New York, NY, USA: Columbia University Press.

- Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, Massachusetts, USA. Harvard University Press.
- McCormack, J. E., Bowen, B. S., & Smith, T. B. (2008). Integrating paleoecology and genetics of bird populations in two sky island archipelagos. *BMC Biology*, 6, 28. doi:10.1186/1741-7007-6-28
- Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36(7), 1333–1345. doi:10.1111/j.1365-2699.2008.02051.x
- Médail, F., & Quezel, P. (1999). Biodiversity hotspots in the Mediterranean Basin: Setting global conservation priorities. *Conservation Biology*, 13(6), 1510–1513. doi: 10.1046/j.1523-1739.1999.98467.x
- Merot, C., Salazar, C., Merrill, R. M., Jiggins, C. D., & Joron, M. (2017). What shapes the continuum of reproductive isolation? Lessons from *Heliconius* butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 284(1856), 20170335. doi: 10.1098/rspb.2017.0335
- Naciri, Y., & Linder, H. P. (2020). The genetics of evolutionary radiations: Evolutionary radiations. *Biological Reviews*, in press. doi: 10.1111/brv.12598
- Nevado, B., Contreras-Ortiz, N., Hughes, C., & Filatov, D. A. (2018). Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. *New Phytologist*, 219(2), 779–793. doi: 10.1111/nph.15243
- Noguerales, V., Cordero, P. J., & Ortego, J. (2018). Integrating genomic and phenotypic data to evaluate alternative phylogenetic and species delimitation hypotheses in a recent evolutionary radiation of grasshoppers. *Molecular Ecology*, 27(5), 1229–1244. doi: 10.1111/mec.14504
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 332–342. doi: 10.1098/rstb.2011.0263
- Olave, M., Ávila, L. J., Sites, J. W., & Morando, M. (2018). Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*. *Journal of Evolutionary Biology*, 31(6), 893–903. doi: 10.1111/jeb.13273
- Olmo-Vidal, J.M. (2002). Atlas Dels Ortòpters de Catalunya i Llibre Vermell 2002 / Llagostes, Saltamartins, Grills, Someretes /Atlas de Los Ortópteros de Cataluña y Libro Rojo 2002 / Atlas of the Orthoptera of Catalonia and Red Book 2002. Barcelona, Spain: Generalitat de Catalunya, Departament de Medi Ambient i Habitatge.
- Ortego, J., García-Navas, V., Noguerales, V., & Cordero, P. J. (2015). Discordant patterns of genetic and phenotypic differentiation in five grasshopper species codistributed across a microreserve network. *Molecular Ecology*, 24(23), 5796–5812. doi: 10.1111/mec.13426
- Ortiz-Barrientos, D., Counterman, B. A., & Noor, M. A. (2004). The genetics of speciation by reinforcement. *PLoS Biology*, 2(12), 2256–2263. doi: 10.1371/journal.pbio.0020416
- Pascoal, S., Mendrok, M., Wilson, A. J., Hunt, J., & Bailey, N. W. (2017). Sexual selection and population divergence II. Divergence in different sexual traits and signal modalities in field crickets (*Teleogryllus oceanicus*). *Evolution*, 71(6), 1614–1626. doi: 10.1111/evo.13239

- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135. doi: 10.1371/journal.pone.0037135
- Pfennig, K. S., & Pfennig, D. W. (2009). Character displacement: Ecological and reproductive responses to a common evolutionary problem. *The Quarterly Review of Biology*, *84*(3), 253–276. doi: 10.1086/605079
- Poniatowski, D., Defaut, B., Llucà-Pomares, D., & Fartmann, T. (2009). The Orthoptera fauna of the Pyrenean region - a field guide. *Articulata Beiheft*, *14*, 1–143.
- Pouchon, C., Fernández, A., Nassar, J. M., Boyer, F., Aubert, S., Lavergne, S., & Mavárez, J. (2018). Phylogenomic analysis of the explosive adaptive radiation of the *Espeletia* complex (Asteraceae) in the Tropical Andes. *Systematic Biology*, *67*(6), 1041–1060. doi: 10.1093/sysbio/syy022
- Puissant, S. (2008). Sur la présence en France d'*Omocestus navasi* Bolivar, 1908, avec description d'une nouvelle sous-espèce (Caelifera, Acrididae, Gomphocerinae). *Matériaux Orthoptériques et Entomocénologiques*, *13*, 69–73.
- Powell, T. H., Hood, G. R., Murphy, M. O., Heilveil, J. S., Berlocher, S. H., Nosil, P., & Feder, J. L. (2013). Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). *Evolution*, *67*(9), 2561–2576. doi: 10.1111/evo.12209
- Ragge, D. R. (1986). The songs of the western European grasshoppers of the genus *Omocestus* in relation to their taxonomy (Orthoptera: Acrididae). *Bulletin of the British Museum (Natural History). Entomology*, *53*(4), 213–249.
- Ragge, D. R., & Reynolds, W. J. (1998). *The Songs of the Grasshoppers and Crickets of Western Europe*. Colchester, UK: Harley Books.
- Reid, B. N., Kass, J. M., Wollney, S., Jensen, E. L., Russello, M. A., Viola, E. M., ... Naro-Maciel, E. (2019). Disentangling the genetic effects of refugial isolation and range expansion in a trans-continentally distributed species. *Heredity*, *122*(4), 441–457. doi: 10.1038/s41437-018-0135-5
- Ribera, I., & Vogler, A. P. (2004). Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, *13*(1), 179–193. doi: 10.1046/j.1365-294X.2003.02035.x
- Rohde, K., Hau, Y., Weyer, J., & Hochkirch, A. (2015). Wide prevalence of hybridization in two sympatric grasshopper species may be shaped by their relative abundances. *BMC Evolutionary Biology*, *15*(1), 191. doi: 10.1186/s12862-015-0460-8
- Rubidge, E. M., Patton, J. L., Lim, M., Burton, A. C., Brashares, J. S., & Moritz, C. (2012). Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nature Climate Change*, *2*(4), 285–288. doi: 10.1038/nclimate1415
- Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M., & Backman, J. (1989). Pleistocene evolution: Northern hemisphere ice sheets and North Atlantic Ocean. *Paleoceanography*, *4*(4), 353–412. doi: 10.1029/PA004i004p00353

- Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, 24(7), 394–399. doi: 10.1016/j.tree.2009.02.007
- Saarman, N. P., & Pogson, G. H. (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology*, 24(18), 4723–4738. doi: 10.1111/mec.13340
- Sainz, A. C., & Faccenna, C. (2001). Tertiary compressional deformation of the Iberian plate. *Terra Nova*, 13(4), 281–288.
- Sasa, M. M., Chippindale, P. T., & Johnson, N. A. (1998). Patterns of postzygotic isolation in frogs. *Evolution*, 52(6), 1811–1820. doi:10.1111/j.1558-5646.1998.tb02258.x
- Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11), 807–820. doi: 10.1038/nrg3522
- Scattolini, M. C., Confalonieri, V., Lira-Noriega, A., Pietrokovsky, S., & Cigliano, M. M. (2018). Diversification mechanisms in the Andean grasshopper genus *Orotettix* (Orthoptera: Acrididae): Ecological niches and evolutionary history. *Biological Journal of the Linnean Society*, 123(4), 697–711. doi: 10.1093/biolinnean/bly008
- Schluter, D. (2000) *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.
- Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4(1), 11. doi: 10.1186/1742-9994-4-11
- Seixas, F. A., Boursot, P., & Melo-Ferreira, J. (2018). The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biology*, 19(1), 91. doi: 10.1186/s13059-018-1471-8
- Servedio, M. R., & Kirkpatrick, M. (1997). The effects of gene flow on reinforcement. *Evolution*, 51(6), 1764–1772. doi: 10.1111/j.1558-5646.1997.tb05100.x
- Servedio, M. R., & Noor, M. A. (2003). The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 339–364. doi: 10.1146/annurev.ecolsys.34.011802.132412
- Shaw, K. L., & Mullen, S. P. (2014). Speciation continuum. *Journal of Heredity*, 105(S1), 741–742. doi: 10.1093/jhered/esu060
- Shepard, D. B., & Burbrink, F. T. (2011). Local-scale environmental variation generates highly divergent lineages associated with stream drainages in a terrestrial salamander, *Plethodon caddoensis*. *Molecular Phylogenetics and Evolution*, 59(2), 399–411. doi: 10.1016/j.ympev.2011.03.007
- Simões, M., Breitreuz, L., Alvarado, M., Baca, S., Cooper, J. C., Heins, L., ... & Lieberman, B. S. (2016). The evolving theory of evolutionary radiations. *Trends in Ecology & Evolution*, 31(1), 27–34. doi: 10.1016/j.tree.2015.10.007
- Simpson, G. G. (1953). *The Major Features of Evolution*. New York, USA: Columbia Univ. Press.

- Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular ecology*, 20(24), 5123–5140. doi: 10.1111/j.1365-294X.2011.05350.x
- Smith, S. D., Pennell, M. W., Dunn, C. W., & Edwards, S. V. (2020). Phylogenetics is the new genetics (for most of biodiversity). *Trends in Ecology & Evolution*, in press. doi: 10.1016/j.tree.2020.01.005
- Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, 64(2), 295–315. doi: 10.1111/j.1558-5646.2009.00877.x
- Stets, J., and P. Wurster. (1982). Atlas and Atlantic — Structural Relations. In U. von Rad, K. Hinz, M. Sarnthein, & E. Seibold (Eds.), *Geology of the Northwest African continental margin* (pp. 69–85). Berlin, Germany: Springer.
- Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, 277(1682), 661–671. doi: 10.1098/rspb.2009.1272
- Taberlet, P., Fumagalli, L., Wust-Saucy, A., & Cosson, J. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7(4), 453–464. doi: 10.1046/j.1365-294x.1998.00289.x
- Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2), 343–355. doi: 10.1111/j.1365-294X.2005.02794.x
- Teacher, A. G. F., Garner, T. W. J., & Nichols, R. A. (2009). European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, 102(5), 490–496. doi: 10.1038/hdy.2008.133
- Teixell, A., Ayarza, P., Tesón, E., Babault, J., Alvarez-Lobato, F., Charroud, M., ... & Arboleya, M. L. (2007). Geodinámica de las cordilleras del Alto y Medio Atlas: síntesis de los conocimientos actuales. *Revista de la Sociedad Geológica de España*, 20(3–4), 333–350.
- Valbuena-Ureña, E., Oromi, N., Soler-Membrives, A., Carranza, S., Amat, F., Camarasa, S., ... & Schmeller, D. S. (2018). Jailed in the mountains: Genetic diversity and structure of an endemic newt species across the Pyrenees. *PLoS ONE*, 13(8), e0200214. doi: 10.1371/journal.pone.0200214
- Vogler, A., & Goldstein, P. (1997). Adaptive radiation and taxon cycles in North American tiger beetles: A cladistic perspective. In T. J. Givnish & K. J. Sytsma (Eds.), *Molecular Evolution and Adaptive Radiation* (pp. 353–373). Cambridge, UK: Cambridge University Press.
- Wallis, G. P., Waters, J. M., Upton, P., & Craw, D. (2016). Transverse alpine speciation driven by glaciation. *Trends in Ecology & Evolution*, 31(12), 916–926. doi: 10.1016/j.tree.2016.08.009
- Wang, B., Jiang, J., Xie, F., & Li, C. (2013). Phylogeographic patterns of mtDNA variation revealed multiple glacial refugia for the frog species *Feirana taihangnica* endemic to the Qinling Mountains. *Journal of Molecular Evolution*, 76(3), 112–128. doi: 10.1007/s00239-013-9544-5

- Warren, D. L., Cardillo, M., Rosauer, D. F., & Bolnick, D. I. (2014). Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution*, *29*(10), 572–580. doi: 10.1016/j.tree.2014.08.003
- Waters, J. M., Emerson, B. C., Arribas, P., & McCulloch, G. A. (2020). Dispersal reduction: causes, genomic mechanisms, and evolutionary consequences. *Trends in Ecology & Evolution*, *35*(6), 512–522. doi: 10.1016/j.tree.2020.01.012
- Wiens, J. J., Engstrom, T. N., & Chippindale, P. T. (2006). Rapid diversification, incomplete isolation, and the “speciation clock” in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution*, *60*(12), 2585–2603. doi: 10.1111/j.0014-3820.2006.tb01892.x
- Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Systematic Zoology*, *27*(1), 17–26. doi: 10.2307/2412809
- Yadav, S., Stow, A. J., & Dudaniec, R. Y. (2019). Detection of environmental and morphological adaptation despite high landscape genetic connectivity in a pest grasshopper (*Phaulacridium vittatum*). *Molecular Ecology*, *28*(14), 3395–3412. doi: 10.1111/mec.15146
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., ... Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, *353*(6306), 1431–1433. doi:10.1126/science.aaf7812
- Zink, R. M., & Slowinski, J. B. (1995). Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. *Proceedings of the National Academy of Sciences*, *92*(13), 5832–5835. doi: 10.1073/pnas.92.13.5832



# APPENDICES



## Appendix I

### Supplemental Information for:

# Reticulate evolutionary history in a recent radiation of montane grasshoppers revealed by genomic data

Vanina Tonzo | Adrià Bellvert | Joaquín Ortego

#### Contents:

#### Supplemental methods

**Methods S1** Genomic library preparation

**Methods S2** Processing of genomic data

#### Supplemental figures

**Figure S1** Number of reads per individual before and after different quality filtering steps by PYRAD and STACKS

**Figure S2** Pseudolikelihood score as a function of the number of reticulation events

#### Supplemental tables

**Table S1** ANOVA analyses testing differences among species/lineages in shape variation for the different studied traits

**Table S2** Pairwise species/lineage comparisons for male and female forewing shape

**Table S3** Pairwise species/lineage comparisons for male and female pronotum shape

**Table S4** Pairwise species/lineage comparisons for male genitalia shape

#### Supplemental references

## Supplemental methods

### Methods S1 Genomic library preparation

We used Nucleo Spin Tissue kits (Macherey-Nagel, Düren, Germany) to extract and purify total genomic DNA from a hind leg of each individual. Genomic DNA was individually barcoded and processed in house using the double-digest restriction-site associated DNA procedure (ddRADseq) described in Peterson et al., (2012), with minor modifications as detailed in Lanier et al., (2015) and Massatti & Knowles (2016). Briefly, DNA was doubly digested with EcoRI and MseI restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique 7-base-pair barcodes. Ligation products were pooled into a library, size selected for fragments between 475 and 580 bp using a Pippin prep machine (Sage Science, Beverly, MA, USA), and amplified by iProof™ High-Fidelity DNA Polymerase (BIO-RAD, Veenendaal, The Netherlands) with 10-12 cycles. Single-read 151-bp sequencing was performed on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, ON, Canada).

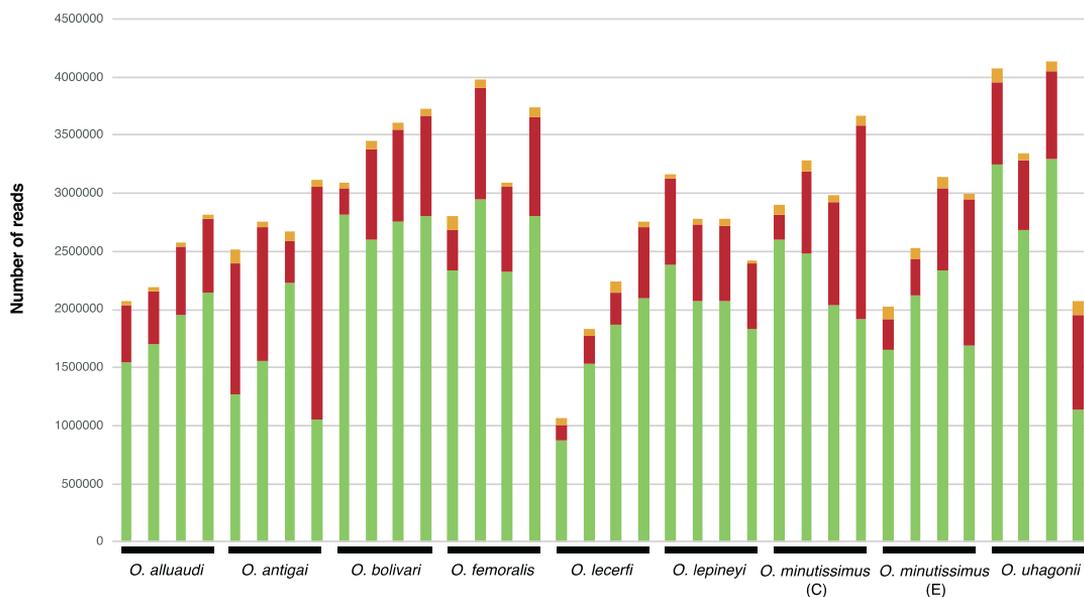
### Methods S2 Processing of genomic data

Raw sequences were demultiplexed using *process\_radtags*, a program distributed as part of the STACKS pipeline (Catchen, Amores, Hohenlohe, Cresko & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores & Cresko, 2013). Only reads with Phred scores  $\geq 10$  (using a sliding window of 15%), no adaptor contamination, and unambiguous barcode and restriction cut sites were retained. Read quality was checked in FASTQC version 0.11.5 (A. Simon, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and sequences were trimmed to 130 bp using SEQTK (L. Heng, <https://github.com/lh3/seqtk>) to remove low-quality reads near the 3' ends. The filtered and trimmed reads were assembled into de novo loci using PYRAD version 3.0.66 (Eaton, 2014). An additional quality-filtering step was performed with PYRAD to convert base calls with a Phred score  $< 20$  into Ns and discard reads with  $> 2$  Ns. Parameter values for clustering threshold of sequence similarity ( $W_{\text{CLUST}} = 0.85$ ), minimum coverage depth ( $d = 5$ ), maximum individuals with shared heterozygous sites ( $\text{maxSH} = p.10$ ), and maximum number of polymorphic sites in a final locus ( $\text{maxSNPs} = 20$ ) were selected based on suggestions from the literature (Eaton, 2014; Eaton & Ree, 2013; Takahashi, Nagata, & Sota, 2014). Final datasets for subsequent analyses were

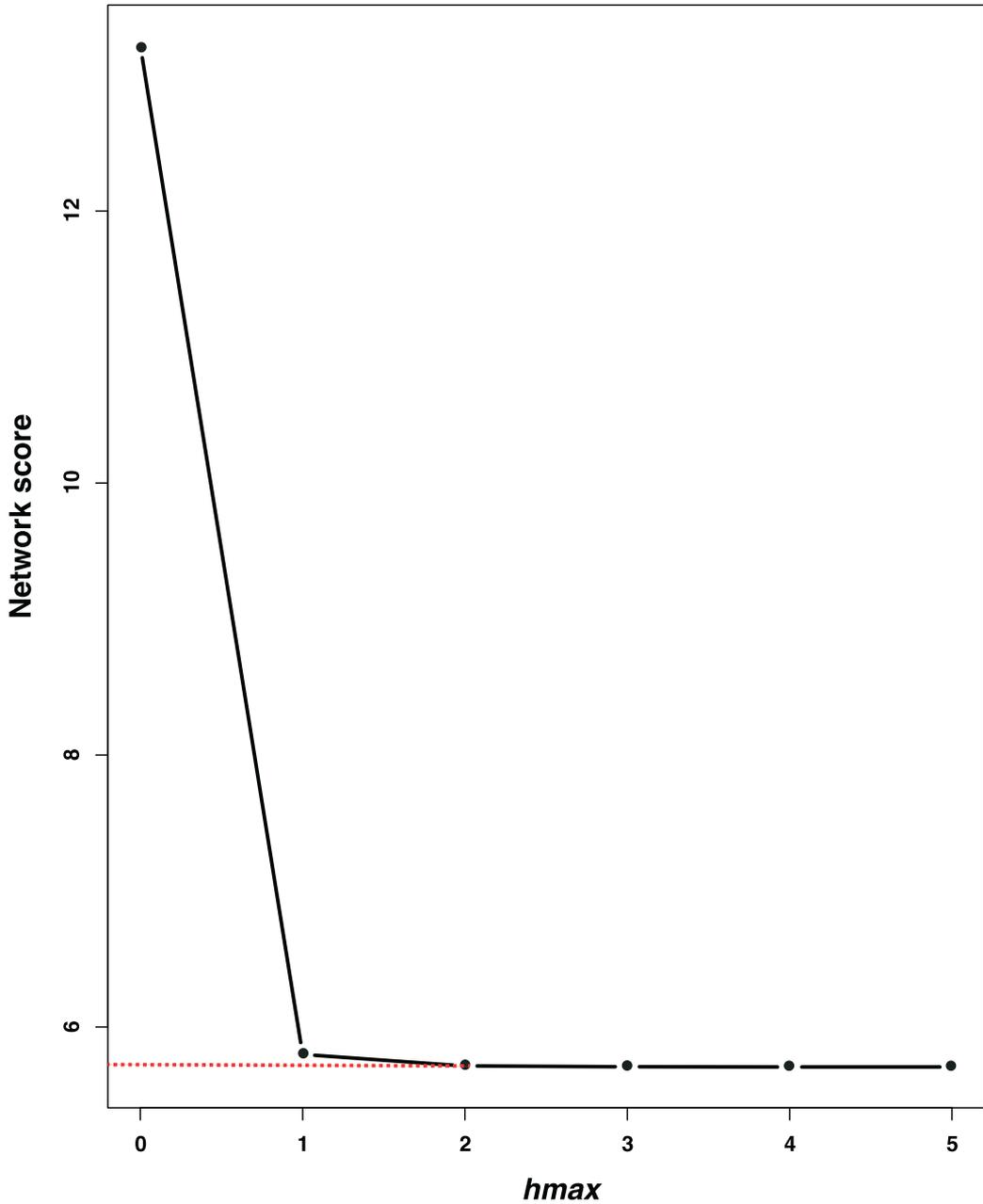
generated discarding loci that were not present in at least ~25 % of the samples.

## Supplemental figures

**Figure S1** Number of reads per individual before and after different quality filtering steps by PYRAD and STACKS. The cumulative stacked bars represent the total number of raw reads for each individual. Dark yellow color represents the reads that were discarded by *process\_radtags* in STACKS due to low quality, adapter contamination or ambiguous barcode. Red color represents the reads that were discarded during step 2 in PYRAD after filtering out reads that did not comply with the quality criteria (reads with >2 sites with a Phred quality score < 20 were discarded). Green color represents the total number of retained reads used to identify homologous loci.



**Figure S2** Plot of pseudolikelihood score as a function of the number of reticulation events to estimate network complexity in PHYLONETWORKS. Here, the slope heuristic suggests the presence of two hybrid nodes ( $hmax = 2$ ).



## Supplemental tables

**Table S1** Results of ANOVA analyses testing differences among species/lineages in shape variation for the different studied traits. *Z*, effect sizes based on a *F* distribution; *P*, significance level.

<b>Trait</b>	<b><i>Z</i></b>	<b><i>P</i></b>
Male forewing	6.724	0.001
Female forewing	7.177	0.001
Male pronotum	3.631	0.001
Female pronotum	5.226	0.001
Male genitalia	1.969	0.031

**Tables S2** Pairwise species/lineage comparisons between means of Procrustes shape for male (above the diagonal) and female (below the diagonal) forewing. Table shows *P*-values.

	<i>O. alluaudi</i>	<i>O. antigai</i>	<i>O. bolivari</i>	<i>O. femoralis</i>	<i>O. lecerf</i>	<i>O. lepineyi</i>	<i>O. minutissimus</i> (E)	<i>O. minutissimus</i> (C)
<i>O. alluaudi</i>	-	0.002	0.004	0.187	0.003	0.766	0.501	0.032
<i>O. antigai</i>	0.001	-	0.007	0.015	0.002	0.001	0.001	0.004
<i>O. bolivari</i>	0.001	0.098	-	0.06	0.001	0.002	0.012	0.05
<i>O. femoralis</i>	0.508	0.001	0.001	-	0.001	0.058	0.237	0.182
<i>O. lecerfi</i>	0.012	0.012	0.004	0.042	-	0.002	0.001	0.007
<i>O. lepineyi</i>	0.423	0.003	0.001	0.409	0.039	-	0.172	0.01
<i>O. minutissimus</i> (E)	0.767	0.001	0.001	0.389	0.005	0.11	-	0.075
<i>O. minutissimus</i> (C)	0.131	0.001	0.001	0.468	0.101	0.121	0.088	-
<i>O. uhagonii</i>	0.058	0.008	0.001	0.086	0.021	0.088	0.034	0.056

**Tables S3** Pairwise species/lineage comparisons between means of Procrustes shape for male (above the diagonal) and female (below the diagonal) pronotum. Table shows *P*-values.

	<i>O. alluaudi</i>	<i>O. antigai</i>	<i>O. bolivari</i>	<i>O. femoralis</i>	<i>O. lecerf</i>	<i>O. lepineyi</i>	<i>O. minutissimus</i> (E)	<i>O. minutissimus</i> (C)	<i>O. uhagonii</i>
<i>O. alluaudi</i>	-	0.111	0.052	0.024	0.087	0.747	0.058	0.011	0.273
<i>O. antigai</i>	0.003	-	0.093	0.007	0.112	0.15	0.042	0.001	0.296
<i>O. bolivari</i>	0.071	0.012	-	0.252	0.069	0.054	0.388	0.152	0.111
<i>O. femoralis</i>	0.056	0.006	0.66	-	0.042	0.015	0.813	0.377	0.014
<i>O. lecerfi</i>	0.463	0.003	0.14	0.055	-	0.243	0.12	0.022	0.053
<i>O. lepineyi</i>	0.422	0.001	0.003	0.001	0.144	-	0.066	0.017	0.249
<i>O. minutissimus</i> (E)	0.163	0.019	0.648	0.655	0.118	0.009	-	0.231	0.043
<i>O. minutissimus</i> (C)	0.055	0.04	0.414	0.837	0.042	0.001	0.631	-	0.005
<i>O. uhagonii</i>	0.102	0.001	0.006	0.001	0.583	0.146	0.004	0.002	-

**Tables S4** Pairwise species/lineage comparisons between means of Procrustes shape for male genitalia. Table shows *P*-values.

	<i>O. alluaudi</i>	<i>O. antigai</i>	<i>O. bolivari</i>	<i>O. femoralis</i>	<i>O. lecerfi</i>	<i>O. lepineyi</i>	<i>O. minutissimus</i> (E)	<i>O. minutissimus</i> (C)	<i>O. uhagonii</i>
<i>O. alluaudi</i>	-	0.678	0.756	0.17	0.575	0.147	0.366	0.22	0.01
<i>O. antigai</i>		-	0.931	0.475	0.96	0.146	0.743	0.446	0.104
<i>O. bolivari</i>			-	0.792	0.998	0.059	0.988	0.898	0.132
<i>O. femoralis</i>				-	0.878	0.004	0.649	0.937	0.21
<i>O. lecerfi</i>					-	0.061	0.976	0.831	0.227
<i>O. lepineyi</i>						-	0.012	0.003	0.003
<i>O. minutissimus</i> (E)							-	0.808	0.105
<i>O. minutissimus</i> (C)								-	0.107
<i>O. uhagonii</i>									-

## Supplemental references

- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genetics*, *1*(3), 171-182. doi:10.1534/g3.111.000240
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. doi:10.1111/mec.12354
- Eaton, D. A. R. (2014). PYRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844-1849. doi:10.1093/bioinformatics/btu121
- Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis: Orobanchaceae*). *Systematic Biology*, *62*(5), 689-706. doi:10.1093/sysbio/syt032
- Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in collared pikas (*Ochotona collaris*). *Molecular Ecology*, *24*(14), 3688-3705. doi:10.1111/mec.13270
- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, *25*(16), 3974-3986. doi:10.1111/mec.13735
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. doi:10.1371/journal.pone.0037135
- Takahashi, T., Nagata, N., & Sota, T. (2014). Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution*, *80*, 137-144. doi:10.1016/j.ympev.2014.07.016





## Appendix II

### Supplemental Information for:

# Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex

Vanina Tonzo | Anna Papadopoulou | Joaquín Ortego

#### Contents:

#### Supplemental methods

**Methods S1** Genomic library preparation

**Methods S2** Processing of genomic data

**Methods S3** Analyses of population genetic structure

#### Supplemental figures

**Figure S1** Number of reads per individual before and after different quality filtering steps by PYRAD and STACKS.

**Figure S2** Shape variation along the two first principal components for four phenotypic traits in females.

**Figure S3** Shape variation along the two first principal components for three phenotypic traits in males.

**Figure S4** Log probability of the data (STRUCTURE), marginal likelihoods (FASTSTRUCTURE), and the magnitude of  $\Delta K$  for Bayesian clustering analyses.

**Figure S5** Genetic assignment of individuals based on the results of STRUCTURE from  $K = 2$  to  $K = 10$ .

**Figure S6** Genetic assignment of individuals based on the results of FASTSTRUCTURE from  $K = 2$  to  $K = 5$ .

**Figure S7** Cross-validation results and layer contributions for CONSTRUCT analyses.

**Figure S8** Genetic assignment of individuals based on the results of CONSTRUCT from  $K = 2$  to  $K = 10$ .

**Figure S9** Results of species delimitation analyses in IBPP using a dataset in which phenotypic information was randomized across individuals.

#### Supplemental table

**Table S1** List of synonyms for *Omocestus antigai* (Bolívar, 1897).

#### Supplemental references

## Supplemental methods

### Methods S1 Genomic library preparation

We used Nucleo Spin Tissue kits (Macherey-Nagel, Düren, Germany) to extract and purify total genomic DNA from a hind leg of each individual. Genomic DNA was individually barcoded and processed in house using the double-digest restriction-site associated DNA procedure (ddRADseq) described in Peterson et al., (2012), with minor modifications as detailed in Lanier et al., (2015) and Massatti & Knowles (2016). Briefly, DNA was doubly digested with EcoRI and MseI restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique 7-base-pair barcodes. Ligation products were pooled into two libraries of 48 samples each, size selected for fragments between 475 and 580 bp using a Pippin prep machine (Sage Science, Beverly, MA, USA), and amplified by iProof<sup>TM</sup> High-Fidelity DNA Polymerase (BIO-RAD, Veenendaal, The Netherlands) with 10-12 cycles. Single-read 151-bp sequencing was performed on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, ON, Canada).

### Methods S2 Processing of genomic data

Raw sequences were demultiplexed using *process\_radtags*, a program distributed as part of the STACKS pipeline (Catchen, Amores, Hohenlohe, Cresko & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores & Cresko, 2013). Only reads with Phred scores  $\geq 10$  (using a sliding window of 15%), no adaptor contamination, and unambiguous barcode and restriction cut sites were retained. Read quality was checked in FASTQC version 0.11.5 (A. Simon, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and sequences were trimmed to 130 bp using SEQTK (L. Heng, <https://github.com/lh3/seqtk>) to remove low-quality reads near the 3' ends. The filtered and trimmed reads were assembled into *de novo* loci using PYRAD version 3.0.66 (Eaton, 2014). An additional quality-filtering step was performed with PYRAD to convert base calls with a Phred score  $< 20$  into Ns and discard reads with  $> 2$  Ns. Parameter values for clustering threshold of sequence similarity ( $W_{CLUST} = 0.90$ ), minimum coverage depth ( $d = 5$ ), maximum individuals with shared heterozygous sites ( $maxSH = p.10$ ), and maximum number of polymorphic sites in a final locus ( $maxSNPs = 20$ ) were selected based on suggestions from the literature (Eaton, 2014; Eaton & Ree, 2013; Takahashi, Nagata, & Sota, 2014). Final datasets for subsequent analyses were

generated discarding loci that were not present in at least ~25 % of the samples ( $minCov = 24$ ).

### Methods S3 Analyses of population genetic structure

We ran STRUCTURE version 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) for a random subset of 10,000 unlinked SNPs, assuming correlated allele frequencies and admixture, and without using prior population information (Hubisz, Falush, Stephens, & Pritchard, 2009). We conducted 20 independent runs for each value of  $K$  (from  $K = 1$  to  $K = 10$ ) to estimate the optimal number of genetic clusters with 200,000 MCMC cycles, following a burn-in step of 100,000 iterations. STRUCTURE HARVESTER (Earl & vonHoldt, 2012) was used to assess the number of genetic clusters that best describes our data according to log probabilities of the data ( $\ln Pr(X|K)$ ) for each value of  $K$  (Pritchard et al., 2000) and the  $\Delta K$  method (Evanno, Regnaut, & Goudet, 2005). We used CLUMPP version 1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same  $K$  value (Jakobsson & Rosenberg, 2007) and DISTRUCT version 1.1 (Rosenberg, 2004) to visualize as bar plots the individual's probabilities of population membership.

FASTSTRUCTURE (Raj, Stephens, & Pritchard, 2014) implements an algorithm able to efficiently analyze large SNP datasets. We ran FASTSTRUCTURE on our entire genomic dataset (64,365 unlinked SNPs) using a simple prior, considering a convergence criterion of  $1 \times 10^{-7}$  and conducting 25 independent runs for each value of  $K$  (from  $K = 1$  to  $K = 10$ ). Following Raj et al. (2014), we used the *chooseK.py* script to assess model complexity by estimating the metrics  $K_{\phi}^{*c}$ , the value of  $K$  that maximizes log-marginal likelihood lower bound (LLBO) of the data, and  $K_{e}^{*}$ , the smallest number of model components explaining at least 99% of cumulative ancestry contribution. We also assessed the number of genetic clusters that best describes our data calculating  $\Delta K$  based on the log-marginal likelihoods obtained for each value of  $K$  (Evanno et al., 2005). We plotted individual co-ancestry coefficients for the most likely  $K$  values using DISTRUCT.

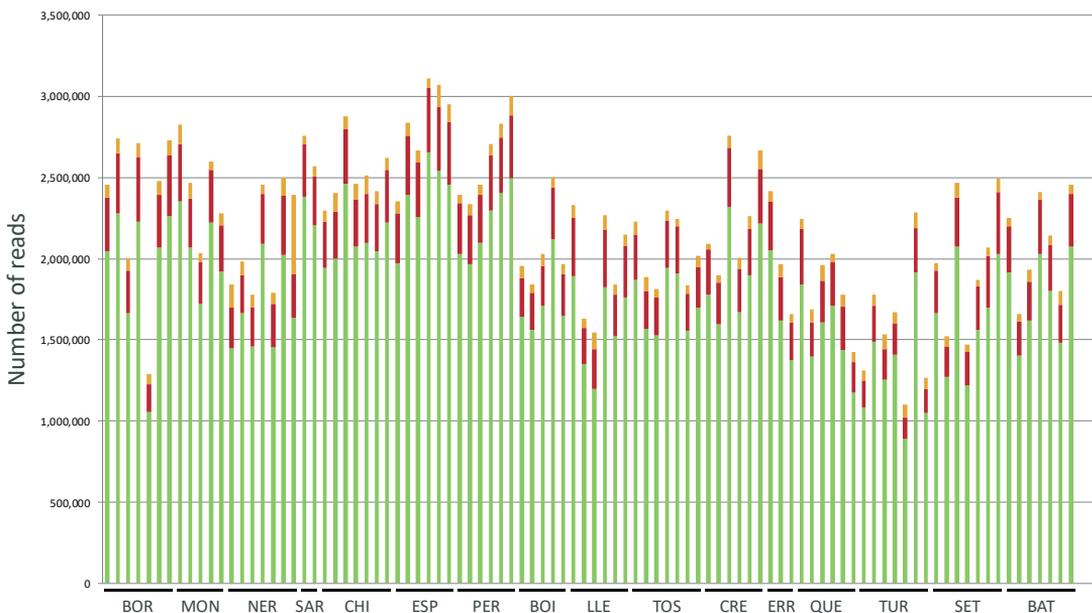
Given the strong signal of isolation by distance in our study system (see Section 3.6), we used the spatial model implemented in R package CONSTRUCT version 1.0.3 to infer patterns of genetic structure after controlling for the geographic distance separating the sampled populations and determine whether genetic differentiation is a consequence of continuous (i.e. isolation by distance) or discrete (e.g. separation by geographic barriers,

speciation, etc.) processes (see Bradburd, Coop, & Ralph, 2018). Given that this approach is known to be sensitive to missing data (Bradburd et al., 2018; Barley et al., 2019), we ran PYRAD using a minimum sample coverage of 75% (*minCov* = 70) to obtain a new dataset (7,929 unlinked SNPs) in which samples and loci had on average <15% missing data (e.g. Barley et al., 2019). We ran CONSTRUCT analyses from  $K = 1$  to  $K = 10$  with 5,000 iterations and visually checked for convergence using trace plots (Bradburd et al., 2018). We used a 5-fold cross-validation approach to examine predictive accuracy across the range of tested  $K$  values and determine the best-fit number of genetic clusters under both the spatial and non-spatial models (Bradburd et al., 2018; e.g. Barley et al., 2019). As done for STRUCTURE and FASTSTRUCTURE, we plotted individual co-ancestry coefficients for the different  $K$  values using DISTRUCT.

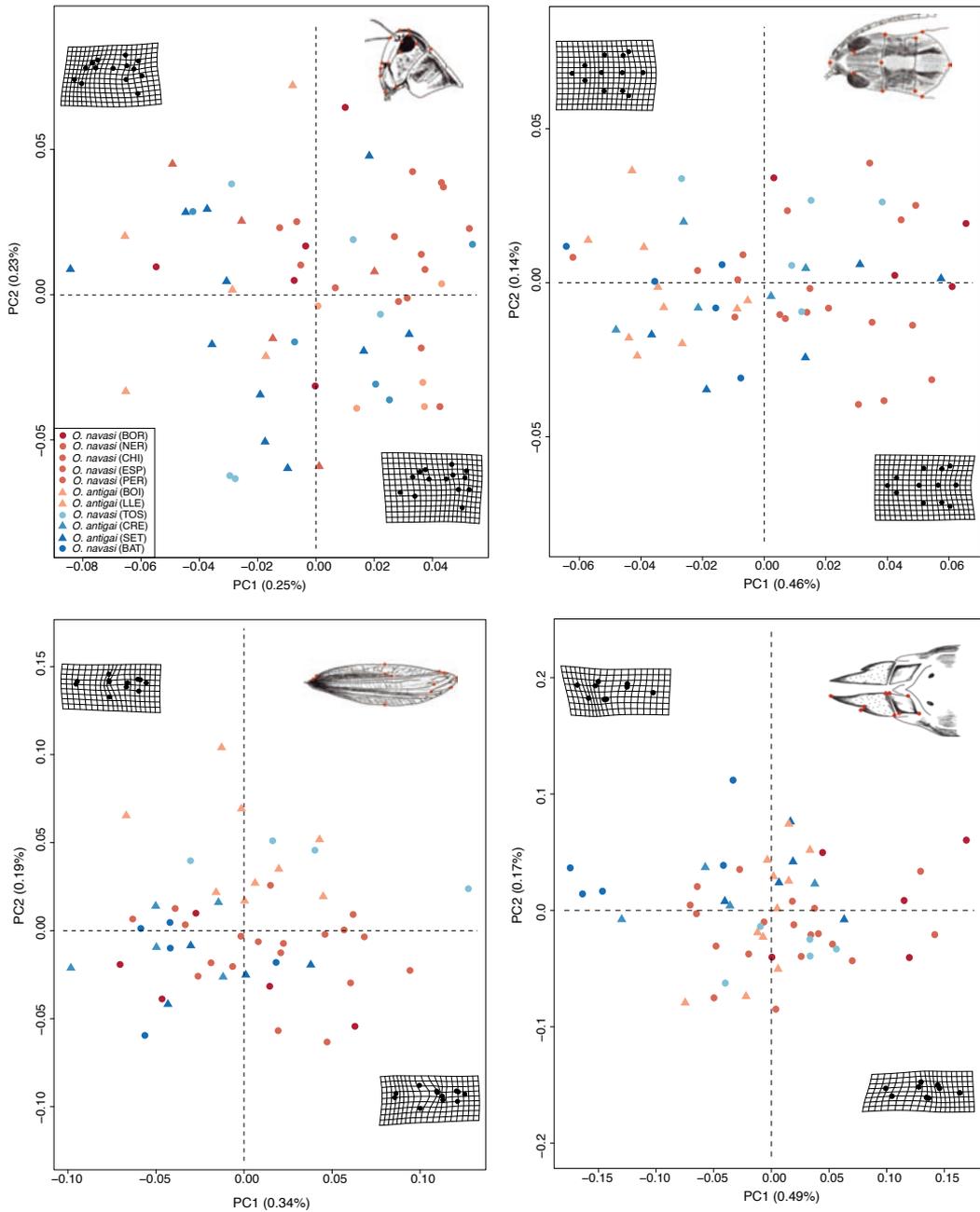
Finally, we used the entire genomic dataset (64,365 unlinked SNPs and all genotyped individuals) to perform a principal component analysis (PCA) as implemented in the R package *adeigenet* (Jombart, 2008). Before running the PCA, we scaled and centered allele frequencies and replaced missing data with mean allele frequencies using the *scaleGen* function as recommended by Jombart (2008).

## Supplemental figures

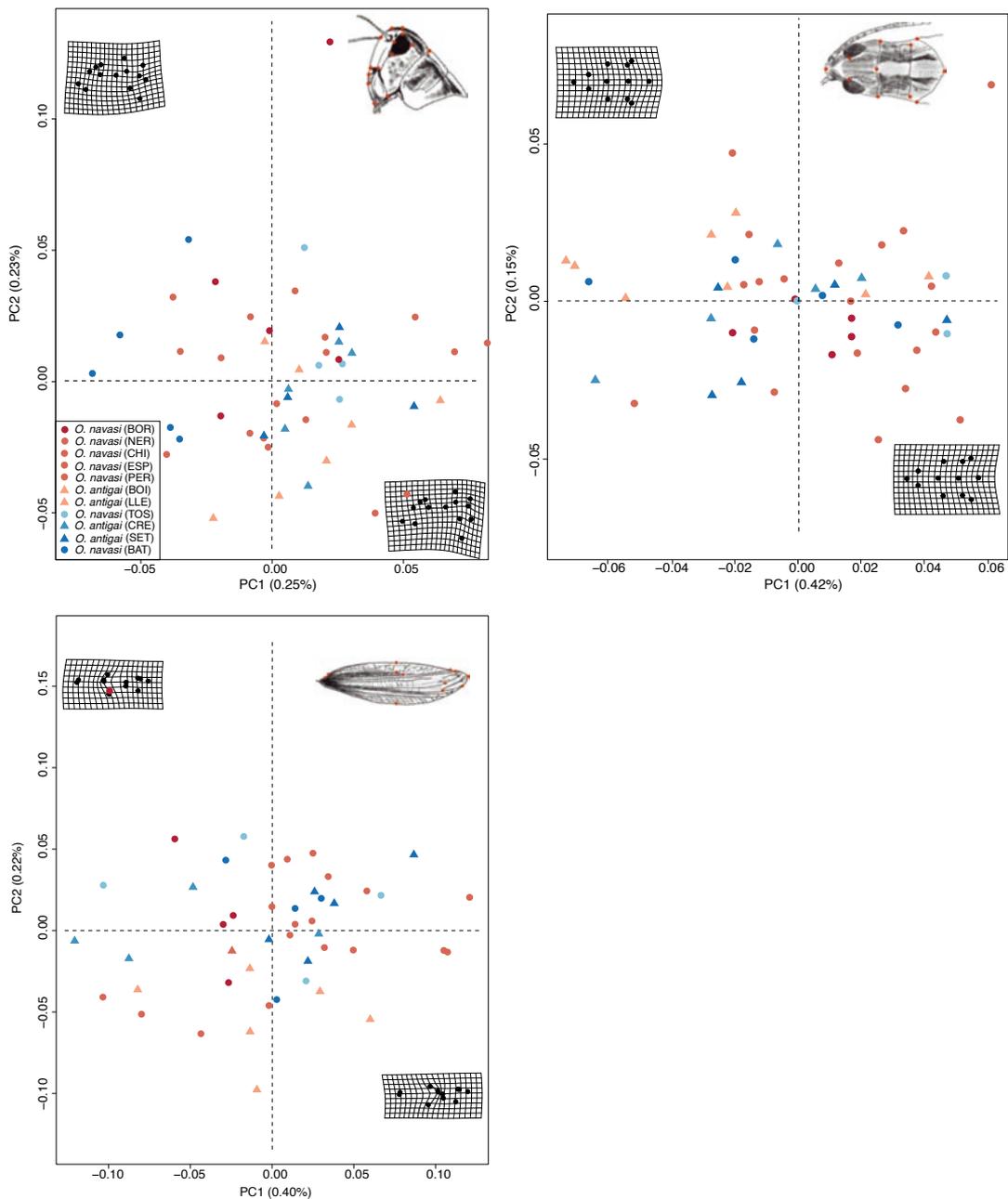
**Figure S1** Number of reads per individual before and after different quality filtering steps by PYRAD and STACKS. The cumulative stacked bars represent the total number of raw reads for each individual. Dark yellow color represents the reads that were discarded by *process\_radtags* in STACKS due to low quality, adapter contamination or ambiguous barcode. Red color represents the reads that were discarded during *step 2* in PYRAD after filtering out reads that did not comply with the quality criteria (reads with >2 sites with a Phred quality score < 20 were discarded). Green color represents the total number of retained reads used to identify homologous loci. Population codes as in Table 1.



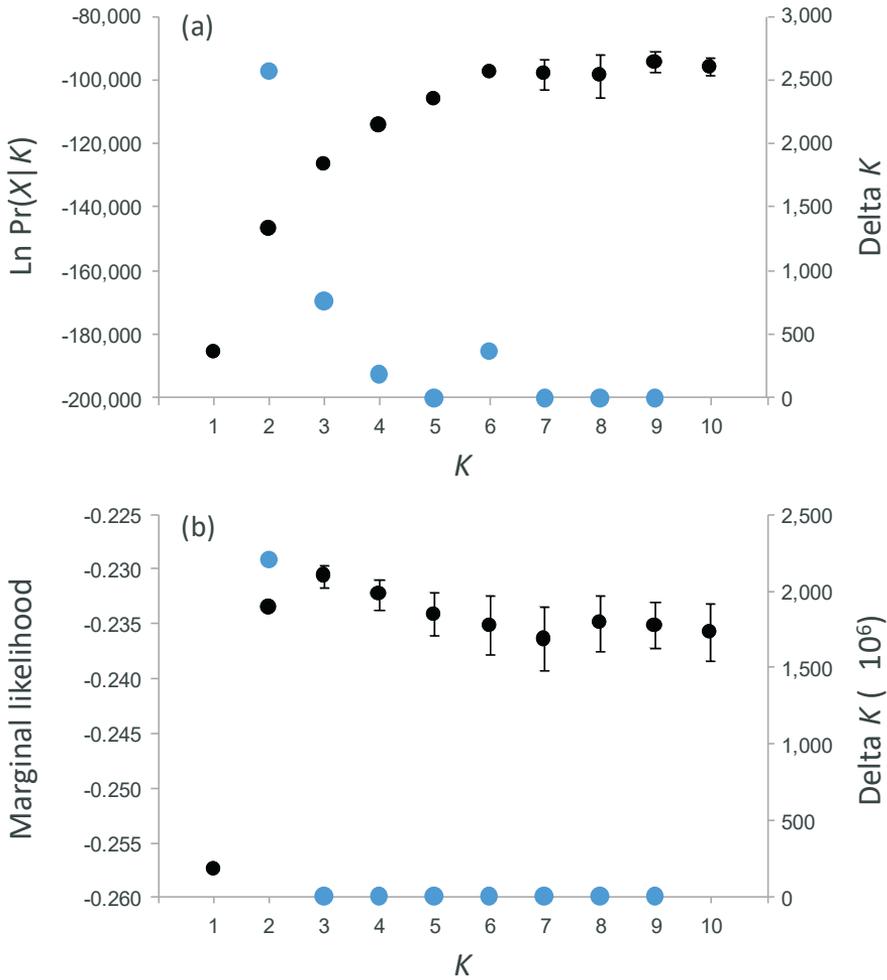
**Figure S2** Shape variation along the two first principal components (PCs) for four phenotypic traits (a: head; b: pronotum; c: forewing; d: ovopositor valve) in females of *Omocestus antigai* (triangles) and *O. navasi* (dots). Thin-plate spline transformation grids show extreme shapes for each trait. Colors of triangles and dots indicate the main genetic cluster at which individuals were assigned according to STRUCTURE analyses for  $K = 6$ . Population codes as in Table 1.



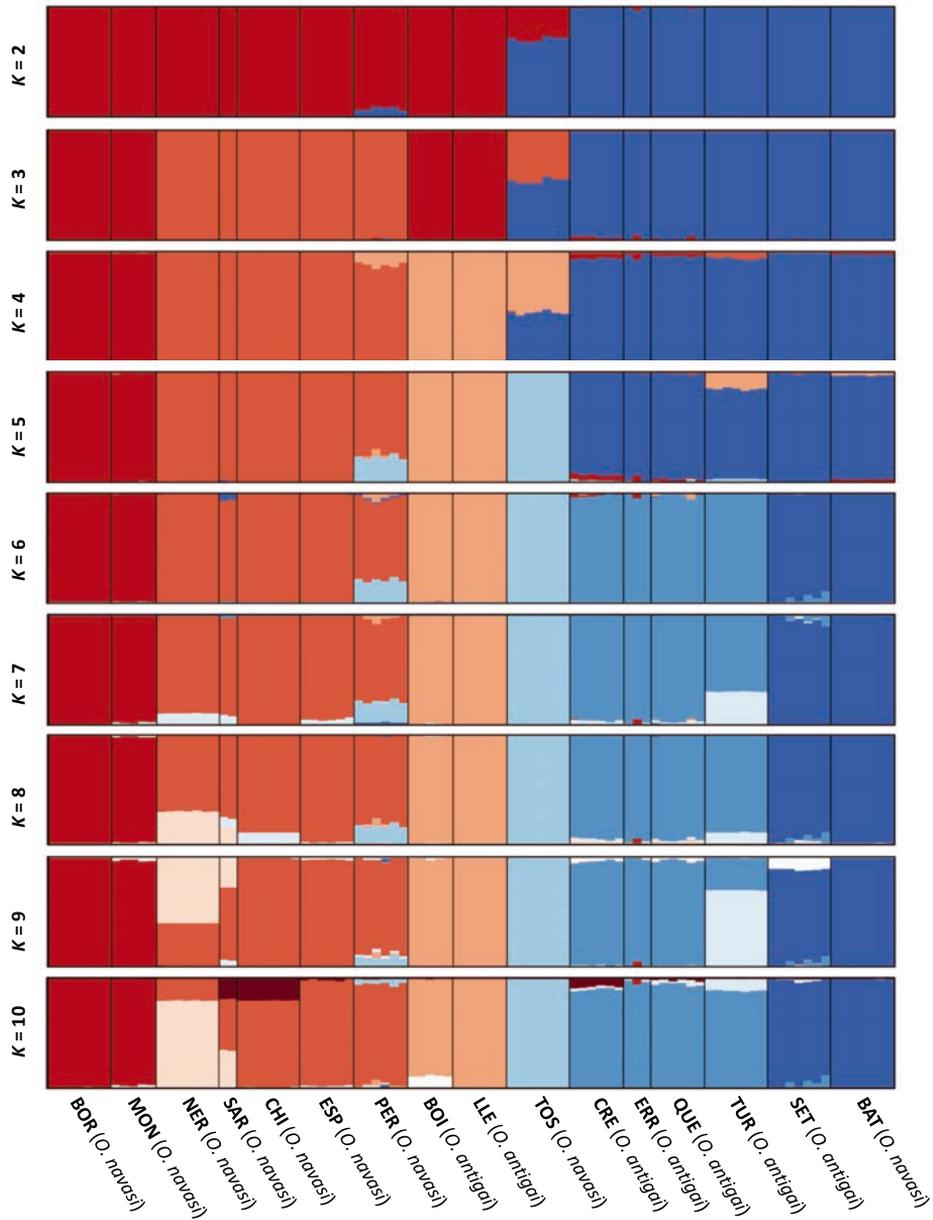
**Figure S3** Shape variation along the two first principal components (PCs) for three phenotypic traits (a: head; b: pronotum; c: forewing) in males of *Omocestus antigai* (triangles) and *O. navasi* (dots). Thin-plate spline transformation grids show extreme shapes for each trait. Colors of triangles and dots indicate the main genetic cluster at which individuals were assigned according to STRUCTURE analyses for  $K = 6$ . Population codes as in Table 1.



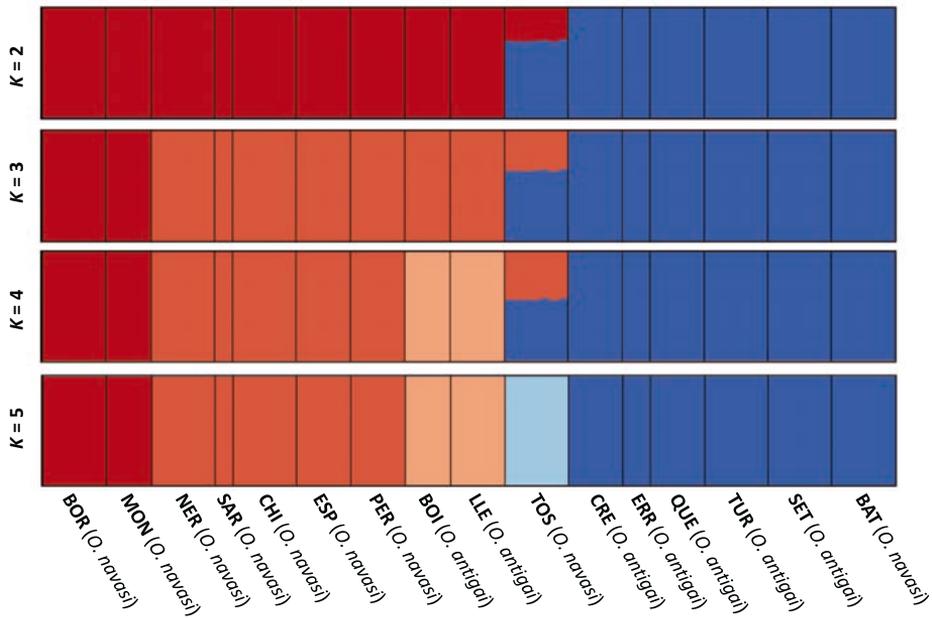
**Figure S4** Left axes, black dots and error bars show for each value of  $K$  the mean ( $\pm$ SD) (a) log probability of the data ( $\text{LnPr}(X|K)$ ) over 10 runs of STRUCTURE and (b) marginal likelihood over 25 runs of FASTSTRUCTURE. Right axes and blue dots show the magnitude of  $\Delta K$ .



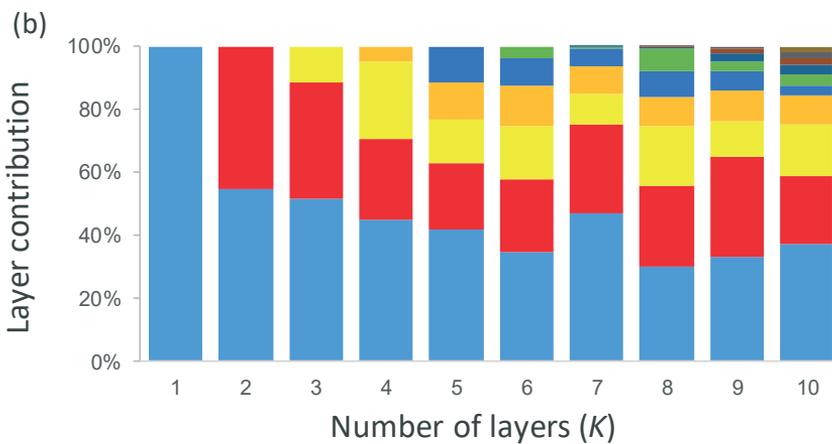
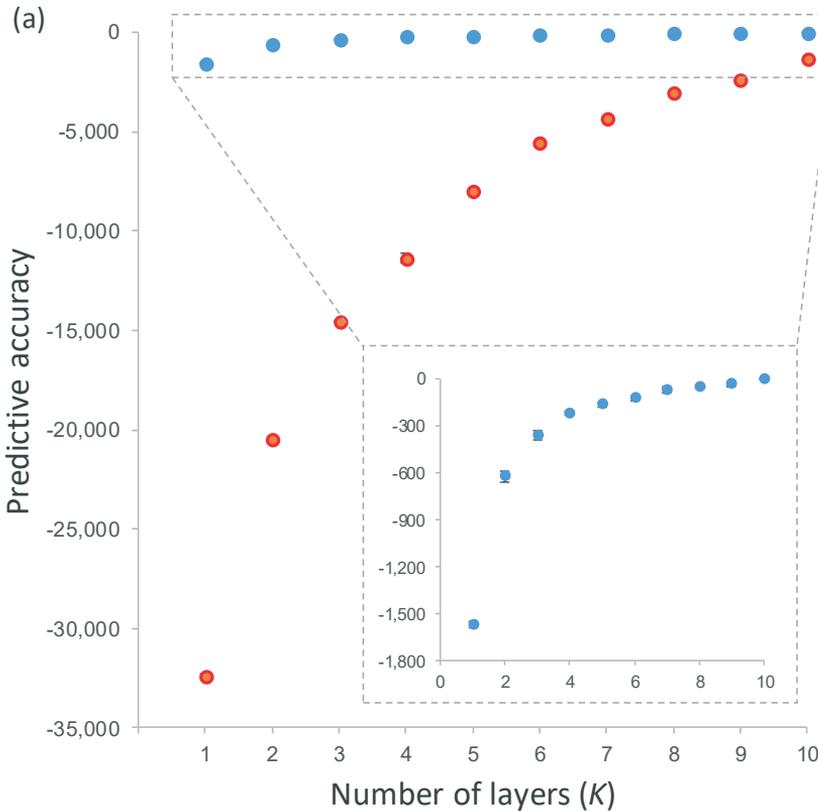
**Figure S5** Genetic assignment of individuals based on the results of STRUCTURE from  $K = 2$  to  $K = 10$ . Individuals are partitioned into  $K$  colored segments representing the probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1.



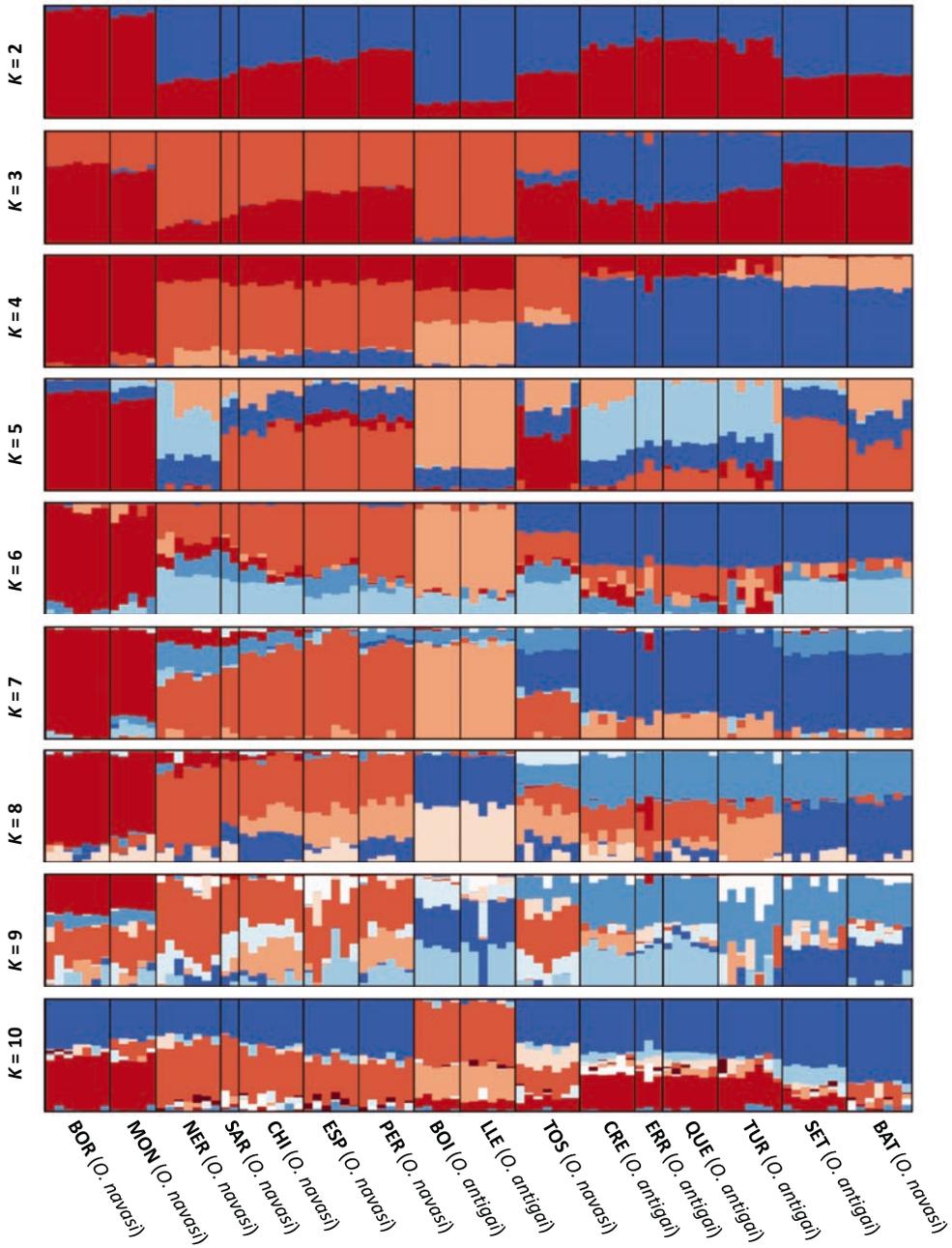
**Figure S6** Genetic assignment of individuals based on the results of FASTSTRUCTURE from  $K = 2$  to  $K = 5$ . Individuals are partitioned into  $K$  colored segments representing the probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1.



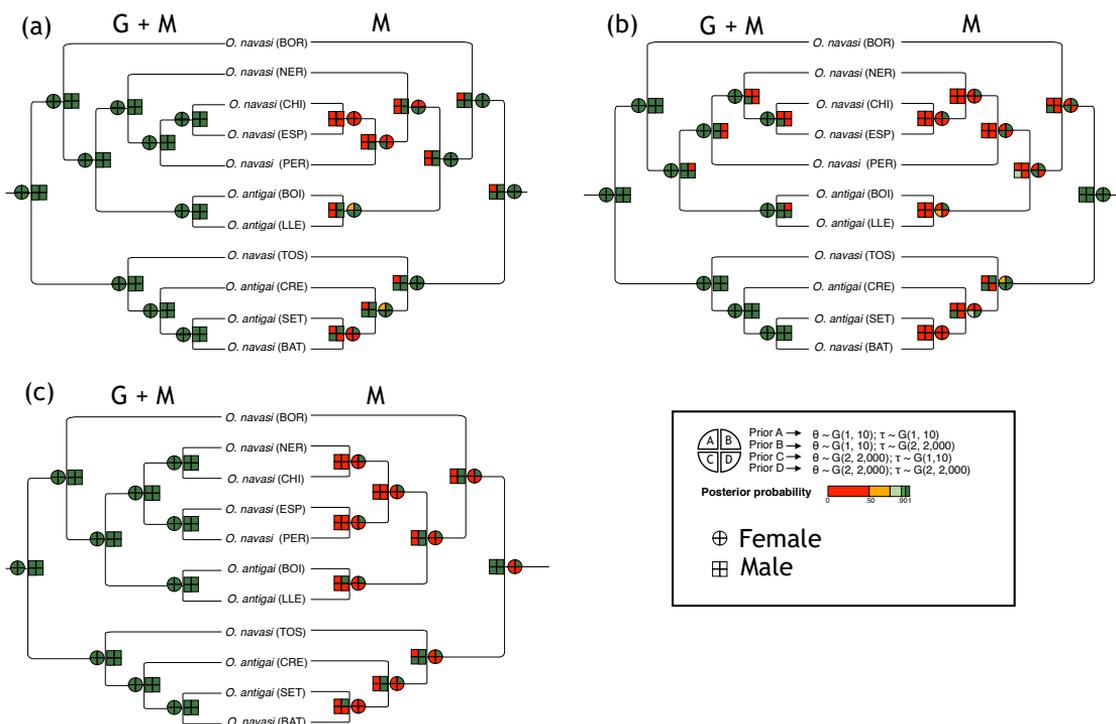
**Figure S7** (a) Cross-validation results for data simulated under  $K = 1-10$  comparing spatial (blue) and non-spatial (red) CONSTRUCT models. Plot shows mean predictive accuracy and 95% confidence intervals for each value of  $K$ . The inset plot zooms cross-validation results for the spatial model; (b) layer/cluster contributions (i.e., how much each layer/cluster contributes to total covariance) for spatial models run from  $K = 1$  to  $K = 10$ .



**Figure S8** Genetic assignment of individuals based on the results of CONSTRUCT spatial models from  $K = 2$  to  $K = 10$ . Individuals are partitioned into  $K$  colored segments representing the probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1.



**Figure S9** Results of species delimitation analyses in IBPP based on matrices in which phenotypic data were randomized across individuals. Analyses were performed separately for each sex both combining genomic (500 loci and 5 individuals/population) and phenotypic data (trees on the left) and only using phenotypic data (trees on the right). All analyses were performed using three alternative topologies (a: BPP; b: SNAPP; c: SVDQUARTETS) and four gamma prior combinations (gamma,  $\alpha$ ,  $\beta$ ) for ancestral population size ( $\theta$ ) and root age ( $\tau$ ). Colored boxes at each node represent the mean posterior probability (PP) for different combinations of demographic priors (legend at bottom right). Population codes as in Table 1.



## Supplemental table

**Table S1** List of synonyms for *Omocestus antigai* (Bolívar, 1897)

List of synonyms
------------------

<i>Omocestus antigai</i> (Bolívar, 1897)
--

= <i>Stenobothrus antigai</i> (Bolívar, 1897)
---

= <i>Stenobothrus brölemanni</i> (Azam, 1906)
---

= <i>Stenobothrus broelemanni</i> (Azam, 1906)
--

= <i>Omocestus navasi</i> Bolívar, 1908
---

= <i>Omocestus navasi bellmanni</i> Puissant, 2008
--

## Supplemental references

- Barley, A. J., de Oca, A. N. M., Reeder, T. W., Manriquez-Moran, N. L., Monroy, J. C. A., Hernández-Gallegos, O., & Thomson, R. C. (2019). Complex patterns of hybridization and introgression across evolutionary timescales in Mexican whiptail lizards (*Aspidoscelis*). *Molecular Phylogenetics and Evolution*, *132*, 284-295. doi:10.1016/j.ympev.2018.12.016
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, *210*(1), 33-52. doi:10.1534/genetics.118.301333
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. doi:10.1111/mec.12354
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genetics*, *1*(3), 171-182. doi:10.1534/g3.111.000240
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361. doi:10.1007/s12686-011-9548-7
- Eaton, D. A. R. (2014). PYRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844-1849. doi:10.1093/bioinformatics/btu121
- Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis: Orobanchaceae*). *Systematic Biology*, *62*(5), 689-706. doi:10.1093/sysbio/syt032
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322-1332. doi:10.1111/j.1755-0998.2009.02591.x
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801-1806. doi:10.1093/bioinformatics/btm233
- Jombart, T. (2008). *Adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405. doi:10.1093/bioinformatics/btn129
- Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: glacial signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, *24*(14), 3688-3705. doi:10.1111/mec.13270

- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, *25*(16), 3974-3986. doi:10.1111/mec.13735
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. doi:10.1371/journal.pone.0037135
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573-589. doi:10.1534/genetics.114.164350
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x
- Takahashi, T., Nagata, N., & Sota, T. (2014). Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution*, *80*, 137-144. doi:10.1016/j.ympev.2014.07.016





## Appendix III

### Supplemental Information for:

# Genomic footprints of an old affair: SNP data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions

Vanina Tonzo | Anna Papadopoulou | Joaquín Ortego

#### Contents:

#### Supplemental figures

**Figure S1** Number of reads per individual before and after different quality filtering steps by PYRAD and STACKS.

**Figure S2** Log probability of the data and the magnitude of  $\Delta K$  for Bayesian clustering analyses in STRUCTURE.

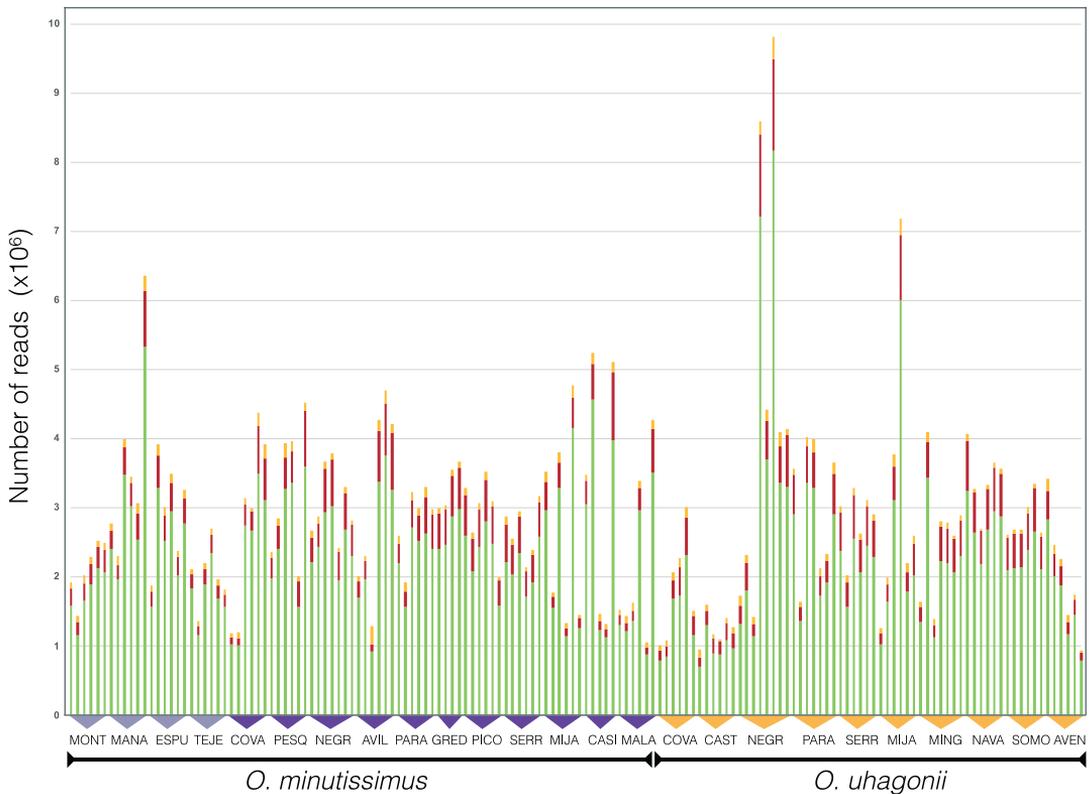
**Figure S3** Summary of population graph model fits with TREEMIX.

#### Supplemental table

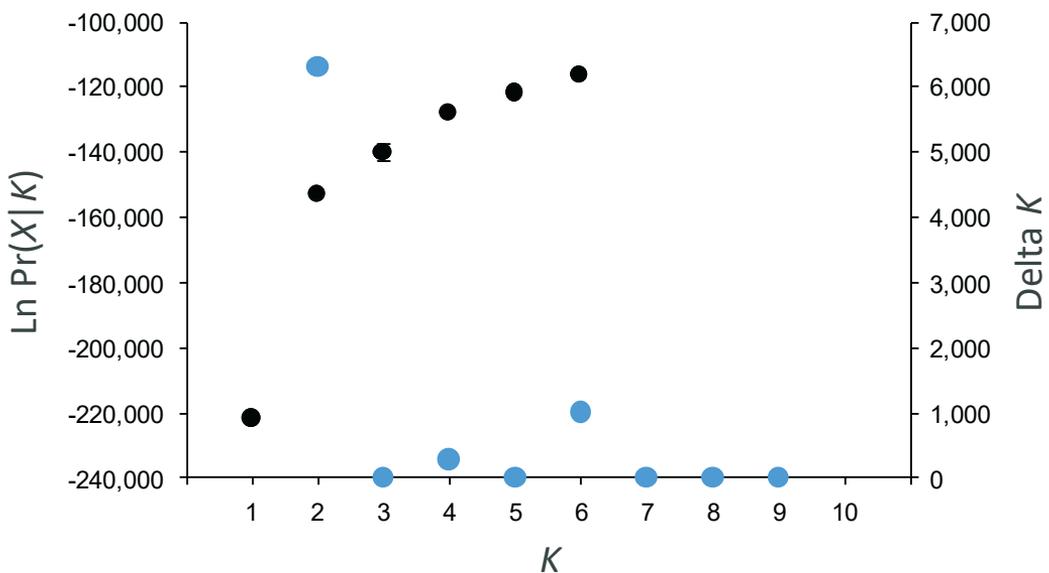
**Table S1** Geographical location and number of genotyped individuals for each studied population.

## Supplemental figures

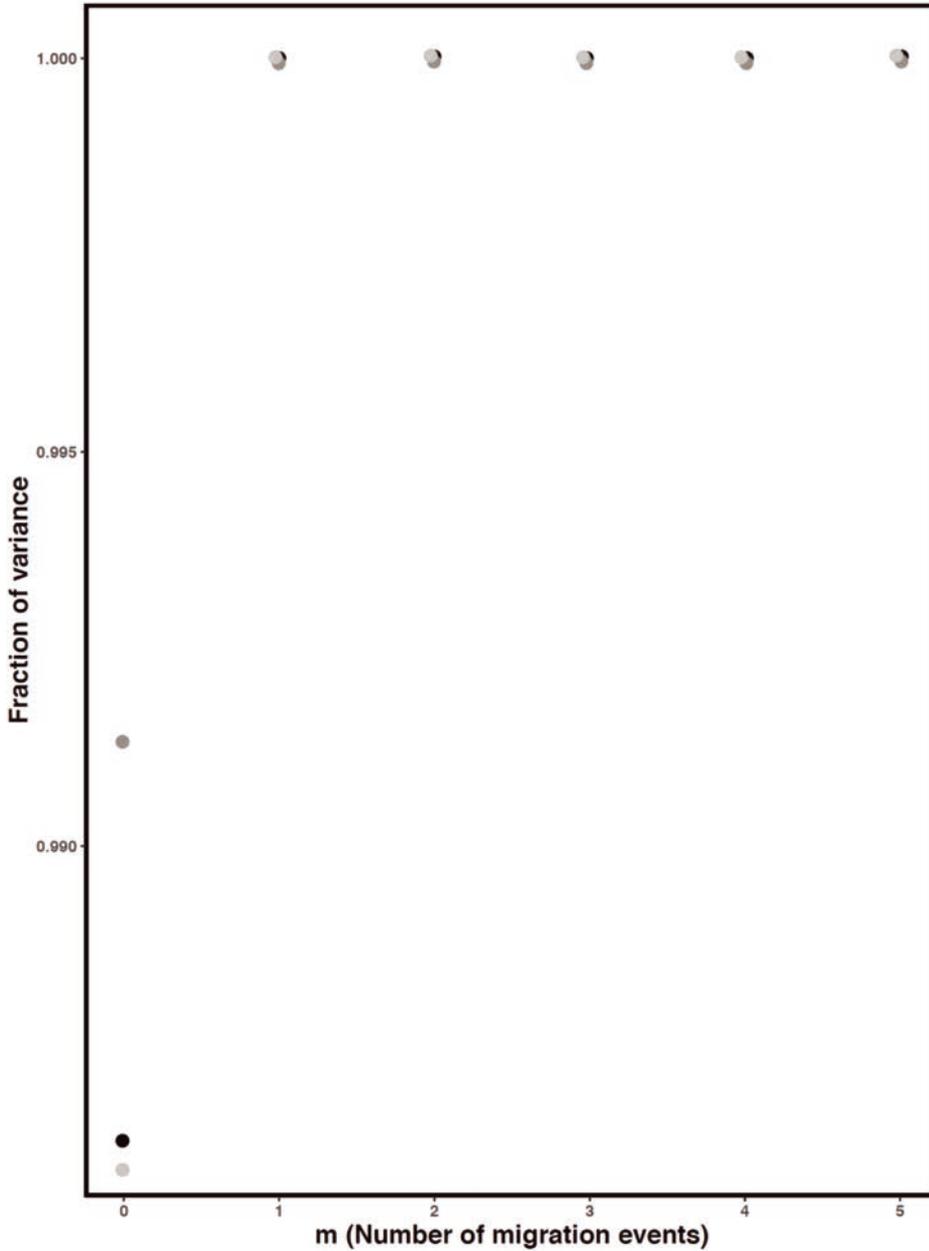
**Figure S1** Number of reads per individual before and after different quality filtering steps by PYRAD and STACKS. The cumulative stacked bars represent the total number of raw reads for each individual. Dark yellow color represents the reads that were discarded by *process\_radtags* in STACKS due to low quality, adapter contamination or ambiguous barcode. Red color represents the reads that were discarded during *step 2* in PYRAD after filtering out reads that did not comply with the quality criteria (reads with >2 sites with a Phred quality score < 20 were discarded). Green color represents the total number of retained reads used to identify homologous loci. Population codes as in Table S1.



**Figure S2** Mean ( $\pm$ SD) log probability of the data ( $\text{Ln Pr}(X|K)$ ) over 10 runs of STRUCTURE (left axes, black dots and error bars) for each value of  $K$  and the magnitude of  $\Delta K$  (right axes, blue dots). Log probability of the data is only shown from  $K = 1$  to  $K = 6$ , as runs for  $K > 6$  had disproportionately small values ( $< -350,000$ ).



**Figure S3** Summary of population graph model fits with TREEMIX. The figure shows the percent variance in genetic relatedness among populations explained by models with different number of migration events ( $m$ ) (from 0 to 5) over three independent runs.



## Supplemental table

**Table S1** Geographical location, elevation and number of genotyped individuals (*n*) for each studied population of *Omocestus minutissimus* and *O. uhagonii*.

Species	Locality	Code	Latitude	Longitude	Elevation (m)
<i>O. minutissimus</i>	Serra de Montsec	MONT	42.04734	0.74305	1,520
<i>O. minutissimus</i>	Coll de Manado	MANA	40.73023	0.18451	1,200
<i>O. minutissimus</i>	Sierra de Espuña	ESPU	37.86454	-1.57133	1,520
<i>O. minutissimus</i>	Sierra de Tejada	TEJE	36.90451	-4.03515	2,040
<i>O. minutissimus</i>	La Covatilla	COVA	40.38435	-5.68139	1,725
<i>O. minutissimus</i>	Pesquera	PESQ	40.43148	-5.31657	1,520
<i>O. minutissimus</i>	Puerto de Peña Negra	NEGR	40.42164	-5.31054	1,885
<i>O. minutissimus</i>	Sierra de Ávila	AVIL	40.65511	-4.98368	1,680
<i>O. minutissimus</i>	Sierra de la Paramera	PARA	40.49152	-4.98348	1,705
<i>O. minutissimus</i>	Parador de Gredos	GRED	40.35627	-5.11176	1,618
<i>O. minutissimus</i>	Puerto del Pico	PICO	40.32216	-5.01139	1,420
<i>O. minutissimus</i>	Puerto de Serranillos	SERR	40.30671	-4.94667	1,600
<i>O. minutissimus</i>	Puerto de Mijares	MIJA	40.33145	-4.81575	1,600
<i>O. minutissimus</i>	Puerto de Casillas	CASI	40.34351	-4.57407	1,480
<i>O. minutissimus</i>	Puerto de Malagón	MALA	40.59208	-4.19012	1,600
<i>O. uhagonii</i>	La Covatilla	COVA	40.38435	-5.68139	1,725
<i>O. uhagonii</i>	Puerto de Castilla	CAST	40.25709	-5.61588	1,780
<i>O. uhagonii</i>	Puerto de Peña Negra	NEGR	40.42164	-5.31054	1,885
<i>O. uhagonii</i>	Sierra de la Paramera	PARA	40.49152	-4.98348	1,705
<i>O. uhagonii</i>	Puerto de Serranillos	SERR	40.30671	-4.94667	1,600
<i>O. uhagonii</i>	Puerto de Mijares	MIJA	40.33145	-4.81575	1,600
<i>O. uhagonii</i>	Cerro de Minguete	MING	40.79388	-4.06943	1,994
<i>O. uhagonii</i>	Puerto de Navafría	NAVA	40.98462	-3.82151	1,920
<i>O. uhagonii</i>	Sierra de Somosierra	SOMO	41.08867	-3.67476	1,796
<i>O. uhagonii</i>	Cerro del Aventadero	AVEN	41.20336	-3.45290	1,965



## Appendix IV

### Supporting Information for

# Glacial connectivity and current population fragmentation in sky-islands explain the contemporary distribution of genomic variation in two narrow-endemic montane grasshoppers from a biodiversity hotspot

Vanina Tonzo | Joaquín Ortego

#### Contents:

##### Supplemental methods

**Methods S1** Processing of genomic data

**Methods S2** Testing alternative demographic models

##### Supplemental figures

**Figure S1** Root Mean Square Error (RMSE) of parameter estimates

**Figure S2** Number of reads before and after different quality filtering steps by STACKS.

**Figure S3** Log probability of the data and the magnitude of  $\Delta K$  for Bayesian clustering analyses in STRUCTURE.

**Figure S4** Genetic assignment of individuals based on the results of Bayesian clustering analyses in STRUCTURE.

**Figure S5** Principal component analysis (PCAs) of genetic variation

**Figure S6** Distribution of posterior quantiles of true parameter values based on pseudo-observed data sets

##### Supplemental references

## Supplemental methods

### Methods S1 Processing of genomic data

We used the different programs distributed as part of the STACKS v. 1.35 pipeline (*process\_radtags*, *ustacks*, *cstacks*, *sstacks*, and *populations*) to assemble our sequences into *de novo* loci and call genotypes (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Hohenlohe et al., 2010). We demultiplexed and filtered reads for overall quality using the program *process\_radtags*, retaining reads with a Phred score > 10 (using a sliding window of 15%), no adaptor contamination, and that had an unambiguous barcode and restriction cut site. We screened raw reads for quality with FASTQC v. 0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed all sequences to 129-bp using seqtk (Heng Li, <https://github.com/lh3/seqtk>) in order to remove low-quality reads near the 3' ends. We assembled filtered reads *de novo* into putative loci with the program *ustacks*. We set the minimum stack depth ( $m$ ) to three and allowed a maximum distance of two nucleotide mismatches ( $M$ ) to group reads into a "stack". We used the "removal" ( $r$ ) and "deleveraging" ( $d$ ) algorithms to eliminate highly repetitive stacks and resolve over-merged loci, respectively. We identified single nucleotide polymorphisms (SNPs) at each locus and called genotypes using a multinomial-based likelihood model that accounts for sequencing errors, with the upper bound of the error rate ( $\epsilon$ ) set to 0.2 (Catchen et al., 2011; Catchen et al., 2013; Hohenlohe et al., 2010). Then, we built a catalogue of loci using the *cstacks* program, with loci recognized as homologous across individuals if the number of nucleotide mismatches between consensus sequences ( $n$ ) was  $\leq 2$ . Finally, we matched each individual data against this catalogue using the program *sstacks* and exported output files in different formats for subsequent analyses using the program *populations*. We used the "blacklist" option to remove outlier loci (i.e., putatively under selection) identified using BAYESCAN v.2.1 (Foll & Gaggiotti, 2008) and the hierarchical and non-hierarchical island models (FDIST method; Excoffier, Hofer, & Foll, 2009) implemented in ARLEQUIN v. 3.5.2.2 (Excoffier & Lischer, 2010) (see Ortego, Gugger, & Sork, 2018 for details). For all downstream analyses, we exported only the first SNP per RAD locus and retained loci with a minimum stack depth  $\geq 5$  ( $m = 5$ ), a minimum minor allele frequency (MAF)  $\geq 0.01$  ( $min\_maf = 0.01$ ) and that were represented in all populations ( $p = 11$  for *O. bolivari* and  $p = 5$  for *O. femoralis*) and the 50% of the individuals

within each population ( $r = 0.5$ ).

## **Methods S2** Testing alternative demographic models

We used the integrative distributional, demographic and coalescent (iDDC) approach (He, Edwards, & Knowles, 2013) and an Approximate Bayesian Computation (ABC) framework (Beaumont, Zhang, & Balding, 2002; Csilléry, Blum, Gaggiotti, & Francois, 2010) to test alternative scenarios representing different hypotheses about how landscape heterogeneity (or its lack thereof) and colonization from glacial refugia vs. contemporary population isolation in sky islands explain the spatial distribution of genomic variation in our two focal species. This approach is described in He et al. (2013) and consists of three main steps: (i) constructing alternative demographic models representing different hypotheses about the processes shaping spatial patterns of genetic structure and diversity; (ii) running demographic and genetic simulations under each model using the software SPLATCHE2 (see Ray, Currat, Foll, & Excoffier, 2010); (iii) evaluating the fit of observed genomic data (i.e., empirical genomic data obtained after genotyping sampled populations) to the genetic expectations under each model, identifying the most probable model/s, and estimating demographic parameters (e.g., González-Serna, Cordero, & Ortego, 2019; He et al., 2013). Below we describe the most relevant aspects of this approach.

**Constructing demographic models.** We generated two main sets of models that differ in the hypothetical demographic processes that have shaped spatial patterns of contemporary genetic variation (Table 2):

i) Colonization of sky islands from glacial refugia (Models A, B, D and E). The first two models (Models A and B) are dynamic models (*sensu* He et al., 2013) incorporating the colonization process from hypothetical glacial refugia and distributional shifts resulted from the interaction between the species bioclimatic envelope and Pleistocene glacial cycles (e.g., He et al., 2013; Massatti & Knowles, 2016). In these scenarios, carrying capacities change over time according to climatic suitability maps obtained from projections of the ENM to the present and the LGM bioclimatic conditions under the MIROC-ESM (Model A) and CCSM (Model B) general atmospheric circulation models (see section *Environmental niche modelling*). These models considered landscapes from three consecutive time periods (LGM, intermediate, current) reflecting temporal shifts in the

spatial distribution of environmentally suitable areas for the species in response to climate changes since the LGM (e.g., He et al., 2013; Massatti & Knowles, 2016). As done in previous studies, carrying capacities were scaled proportionally to logistic climatic suitability scores obtained from the ENM for each time period (e.g., He et al., 2013; Knowles & Massatti, 2017; Massatti & Knowles, 2016). Thus, we assumed that the carrying capacity for each grid cell was proportional to the estimated probability of presence of the species in that grid cell. As SPLATCHE2 requires a single raster file with positive integer numbers, we used ARCMAP v.10.3 to categorize cell values from logistic ENM maps (ranging continuously from 0 to 1) under each time period into 20 bins of equal magnitude (i.e., intervals of 0.05) (e.g., Bemmels, Title, Ortego, & Knowles, 2016). Then, we ran a custom Python script written by Q. He (deposited in Dryad; Bemmels et al., 2016) to convert the maps from the different time periods into a single raster map in which each category represents a unique combination of climatic suitability bins across the three time periods (e.g., Bemmels et al., 2016; He et al., 2013; Massatti & Knowles, 2016). Climatic suitability bins corresponding to each of the three periods (LGM, intermediate, current) were applied to one-third of the total number of simulated generations. Forward demographic simulations under this model initialized 21 ka BP from hypothesized refugial populations (i.e., source populations, each one with an effective population size of  $N_{ANC}$ ) that were located at every habitat patch predicted as suitable for each focal species during the LGM according to MIROC-ESM (Model A) and CCSM (Model B) projections. Specifically, suitable habitat patches during the LGM were identified as those cells with a probability of presence of the species above the maximum training sensitivity plus specificity (MTSS) logistic threshold for occurrence from MAXENT (Liu, Berry, Dawson, & Pearson, 2005). We allowed source population overflow, so that all individuals exceeding the carrying capacity of initial populations (i.e., demes) spread around neighboring cells (Ray et al., 2010). Finally, we generated two more models (Models D and E) analogous to the previous ones but in which carrying capacities ( $k$ ) are homogeneous across space and through time. These static models (*sensu* He et al., 2013) are equivalent to a flat landscape or an isolation-by-distance model and only differ among them in the location of ancestral populations, which were based on the patches of suitable habitat identified under LGM-MIROC (Model D) and LGM-CCSM (Model E) bioclimatic conditions.

ii) Population isolation in sky islands (Models C and F). The first model of this set (Model C) is a static model representing genetic drift associated with the isolation of populations

in sky-islands according to the current geographical configuration of suitable habitats. In this model, carrying capacities do not change over time and are defined by an environmental suitability layer obtained from the projection of the ENM to the current bioclimatic conditions. Forward demographic simulations under this model initialized 7 ka BP from hypothesized refugial populations located in every patch of habitat predicted as suitable during present time according to the MTSS logistic threshold for occurrence from MAXENT. Thus, this model hypothesizes that the spatial distribution of contemporary genetic variation reflects population fragmentation and isolation in sky islands since the Mid-Holocene (e.g., Knowles & Massatti, 2017). As done for the first set of models, we also generated a scenario (Model F) analogous to the previous one but in which carrying capacities ( $k$ ) are homogeneous across space (i.e., equivalent to an isolation-by-distance model; He et al., 2013).

**Demographic and genetic simulations.** We used SPLATCHE2 to perform forward-in-time demographic simulations followed by backward-in-time genetic (coalescent) simulations under each model (see Ray et al., 2010), which are expected to produce contrasting patterns of genetic variation due to differences among scenarios in the location of ancestral populations and the way carrying capacities vary across the landscape and through time (see Massatti & Knowles, 2016). To have a computationally tractable number of cells for demographic simulations, we statistically downscaled cell sizes to 2-arcminute ( $\sim 4 \text{ km}^2$ ) (e.g., Bemmels et al., 2016; Massatti & Knowles, 2016). Despite of the studied species having a generation time of one year, we scaled it by a factor of 15 to make simulations computationally tractable (e.g., Massatti & Knowles, 2016), which results in a total of 1,400 generations from the LGM to present (21 ka) and 467 generations from the mid-Holocene to present (7 ka). Because of this scaling, any biological interpretation of absolute values of population genetic parameters would need to be adjusted accordingly (Massatti & Knowles, 2016). For each model, we ran 200,000 simulations using the uniform priors for the three demographic parameters of the spatially explicit coalescent: carrying capacity of the deme with highest suitability ( $K_{\text{max}}$ ; range of  $\log(K_{\text{max}})$  for *O. bolivari*: 4.3, 6.3; range of  $\log(K_{\text{max}})$  for *O. femoralis*: 5.2, 7.2), migration rate per deme per generation ( $m$ ; range of  $\log(m)$  for *O. bolivari*: -3.1, 2.0; range of  $\log(m)$  for *O. femoralis*: -2.3, -1.2), and ancestral population size ( $N_{\text{ANC}}$ ; range of  $\log(N_{\text{ANC}})$  for *O. bolivari*: 1.7, 3.7; range of  $\log(N_{\text{ANC}})$  for *O. femoralis*: 3.5, 5.5). Before setting the final prior values used in the simulations for each species, we tested a broad range of priors in pilot runs in order to identify those that

result in the colonization of the landscape and generate genetic data within the range of observed empirical data. Following each time-forward demographic simulation, a spatially-explicit time-backward coalescent model informed by the deme-specific demographic parameters ( $K$ ,  $m$  and  $N_{\text{ANC}}$ ) was used to generate genetic data (Currat, Ray, & Excoffier, 2004; Ray et al., 2010). We ran an independent coalescent process to trace the genealogy for each locus from the present to the onset of population expansion from ancestral source populations and beyond, until alleles coalesced in a single ancestral population of size  $N_{\text{ANC}}$ . We set a maximum of  $10^6$  generations for providing ample time for coalescence. To make simulations computationally tractable, we randomly selected 1,250 loci. Simulated datasets were sampled from the same geographical locations (grid cells) from which the empirical genomic data were obtained (Table 1) and consisted of the same number of loci, number of individuals, and amount and pattern of missing data as in the observed empirical dataset (see Massatti & Knowles, 2016). We used ARLSUMSTAT v.3.5.2 (Excoffier & Lischer, 2010) to calculate summary statistics (SS) for simulated datasets under each model, including mean heterozygosity across loci for each population and across populations ( $H$ ), number of segregating sites for each population and across populations ( $S$ ), and pairwise population  $F_{\text{ST}}$  values (79 SS for the eleven populations of *O. bolivari* and 22 SS for the five populations of *O. femoralis*). The same SS were extracted from observed empirical data for each species. We ran all simulations on the high-performance computing cluster from Centro de Supercomputación de Galicia (CESGA, Spain). Simulations required  $\sim 7,000$  CPU hours per model.

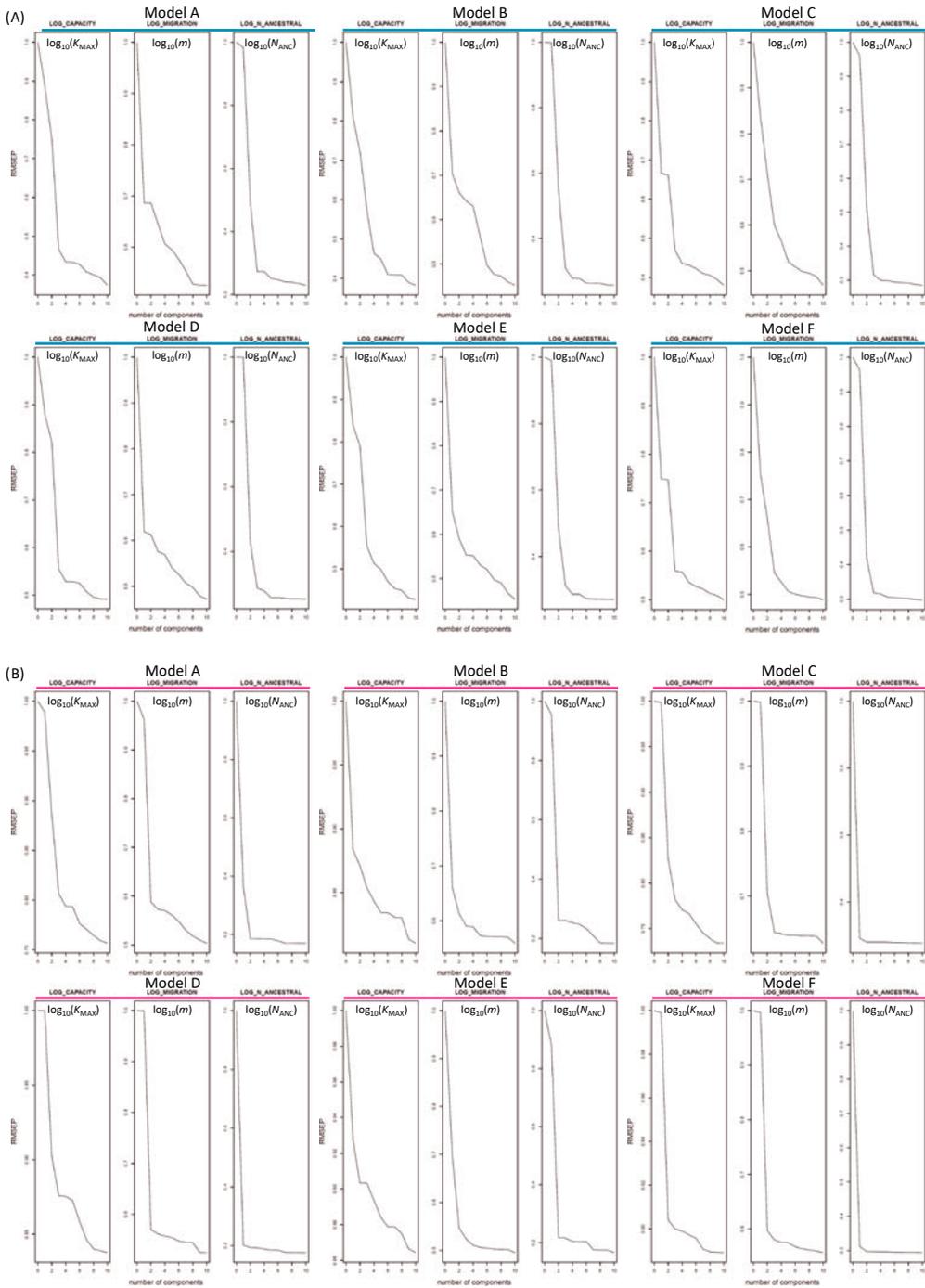
**Model choice and parameter estimation.** We used an Approximate Bayesian computation (ABC) framework to perform model selection and parameter estimation (for an overview of ABC, see Beaumont et al., 2002), as implemented in ABCTOOLBOX programs (TRANSFORMER and ABCESTIMATOR) and R scripts (*findPLS*) (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010). In order to account for correlations between summary statistics and reduce the “curse of dimensionality” associated with using a large number of statistics (Boulesteix & Strimmer, 2007), we used the R package *pls* v.2.6-0 (Mevik & Wehrens, 2007) and the *findPLS* script to extract partial least squares (PLS) components with Box-Cox transformation from the summary statistics of the first 10,000 simulations for each model (Boulesteix & Strimmer, 2007; Wegmann et al., 2010). The first five PLSs extracted from the summary statistics were used for ABC analyses, as the root-mean-squared error (RMSE) of the three demographic parameters ( $K_{\text{MAX}}$ ,  $m$ ,  $N_{\text{ANC}}$ ) for the

two species did not decrease significantly with additional PLSs (Figure S1). We used the linear combinations of summary statistics obtained from the first 10,000 simulations for each model to transform all datasets (observed empirical and simulated datasets) with the program TRANSFORMER (for details about this procedure, see Wegmann et al., 2010). For each model, we retained the 1,000 simulations (0.5%) closest to observed empirical data and used them to approximate marginal densities and posterior distributions of the parameters with a postsampling regression adjustment using the ABC-GLM (general linear model) procedure detailed in Leuenberger & Wegmann (2010) and implemented in ABCESTIMATOR (see also Csilléry et al., 2010). We used Bayes factors (BF) for model selection, defined as the ratio between marginal densities of the model with the highest marginal density and the alternative model (Jeffreys, 1961). The higher the ratio is, the more supported the first model is. A BF > 20 indicates strong relative support for the first model, while those >150 indicate very strong support (Jeffreys, 1961; Kass & Raftery, 1995; Leuenberger & Wegmann, 2010).

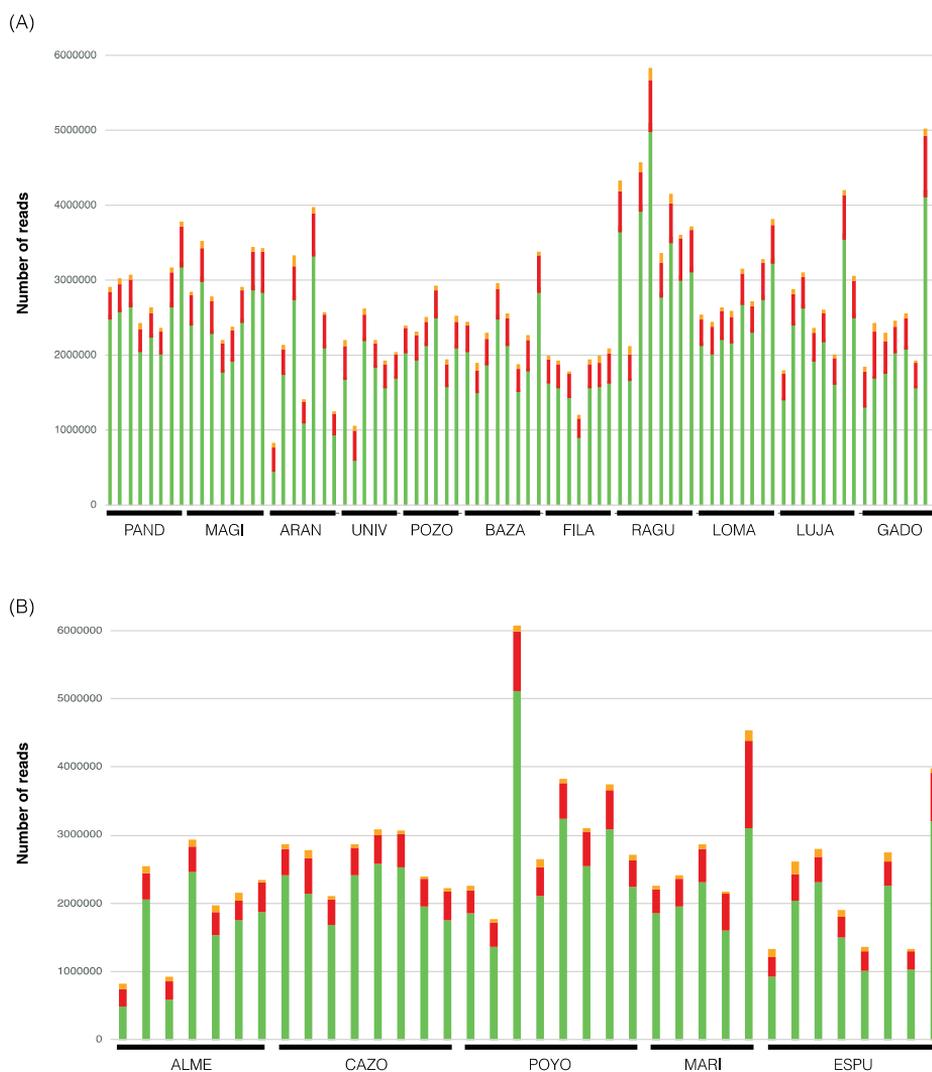
**Model validation.** To evaluate the ability of each model to generate the empirical data, we calculated the Wegmann's  $p$ -value from the 1,000 retained simulations (Wegmann et al., 2010). The Wegmann's  $p$ -value is calculated as the fraction of the retained simulations with a smaller or equal likelihood than the empirical data, with low values indicating that a model is highly unlikely (Wegmann et al., 2010). We also assessed the potential for a parameter to be correctly estimated by computing the proportion of parameter variance that was explained (i.e., the coefficient of determination,  $R^2$ ) by the retained PLSs (Neuenschwander et al., 2008). For the most supported model for each species, we determined the accuracy of parameter estimation using a total of 1,000 pseudo-observation datasets (PODs) generated from prior distributions of the parameters. If the estimation of the parameters is unbiased, posterior quantiles of the parameters obtained from PODs should be uniformly distributed (Cook, Gelman, & Rubin, 2006; Wegmann et al., 2010). As with the empirical data, we calculated the posterior quantiles of true parameters for each pseudo run based on the posterior distribution of the regression-adjusted 1,000 simulations closest to each pseudo-observation.

## Supplemental figures

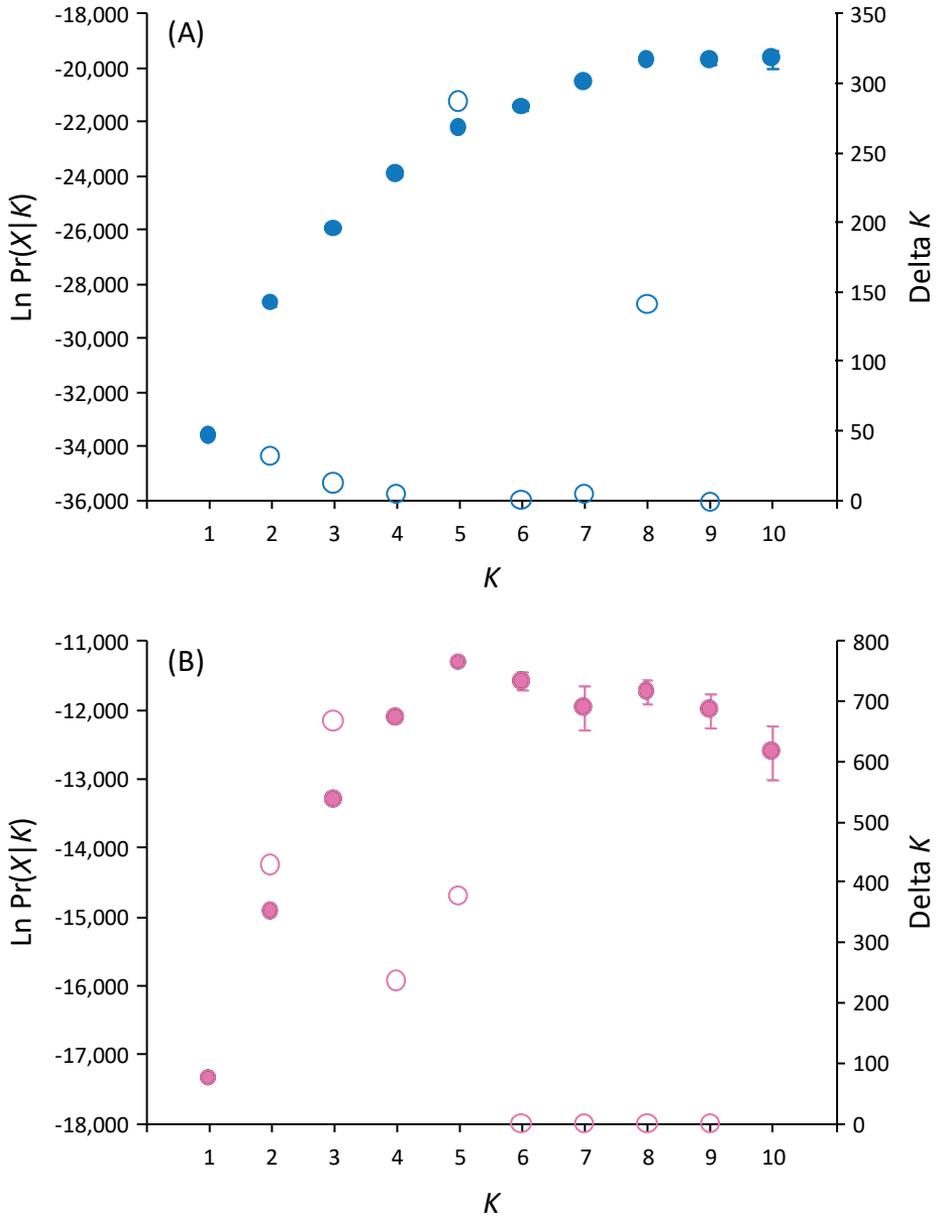
**Figure S1** Root mean square error (RMSE) of parameter estimates against the number of partial least squares (PLS) components under six demographic models for (A) *Omocestus bolivari* and (B) *O. femoralis*.



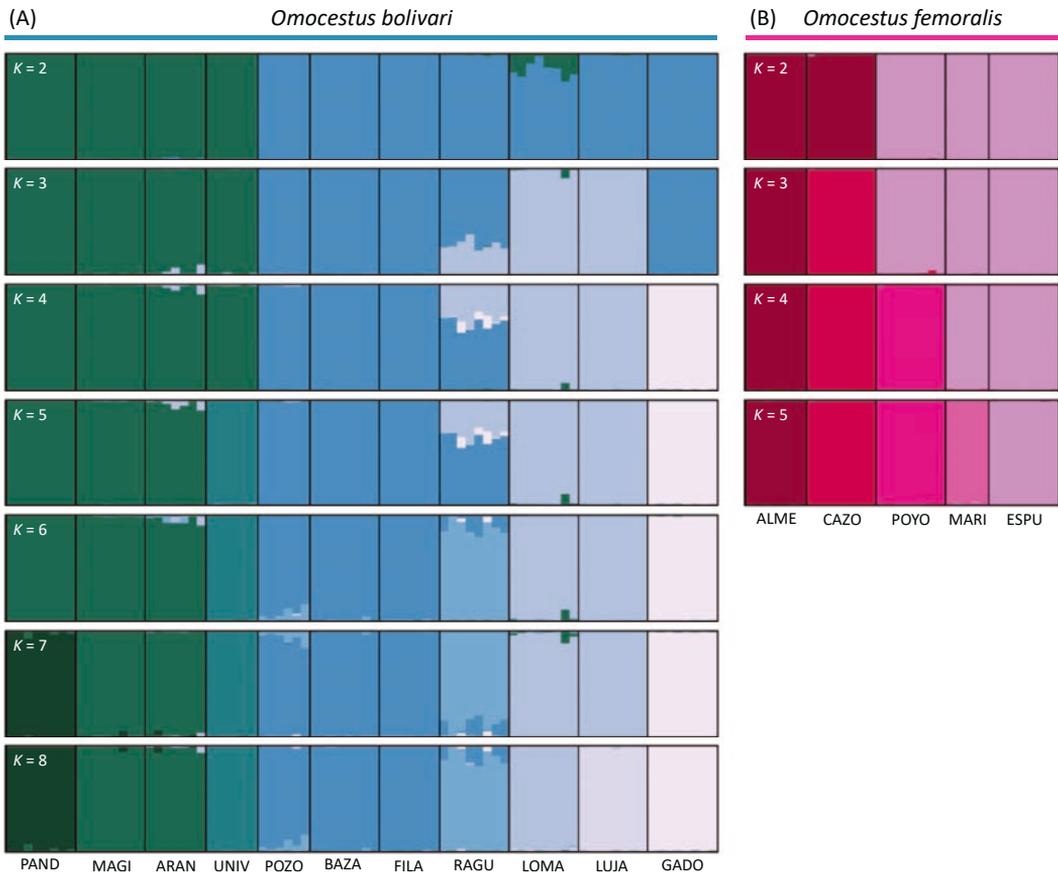
**Figure S2** Number of reads per individual before and after different quality filtering steps by STACKS for (A) *Omocestus bolivari* and (B) *O. femoralis*. The total height of the bars represents the total number of raw reads obtained for each individual. Within each bar, the dark red color represents the reads that were discarded by *process\_radtags* due to low quality, adapter contamination or ambiguous barcode and orange color represents the reads that were discarded by *ustacks* after filtering out repetitive elements and reads that did not comply the different criteria required to create a “stack”. Green color represents the number of retained reads used to identify homologous loci. Population codes as in Table 1.



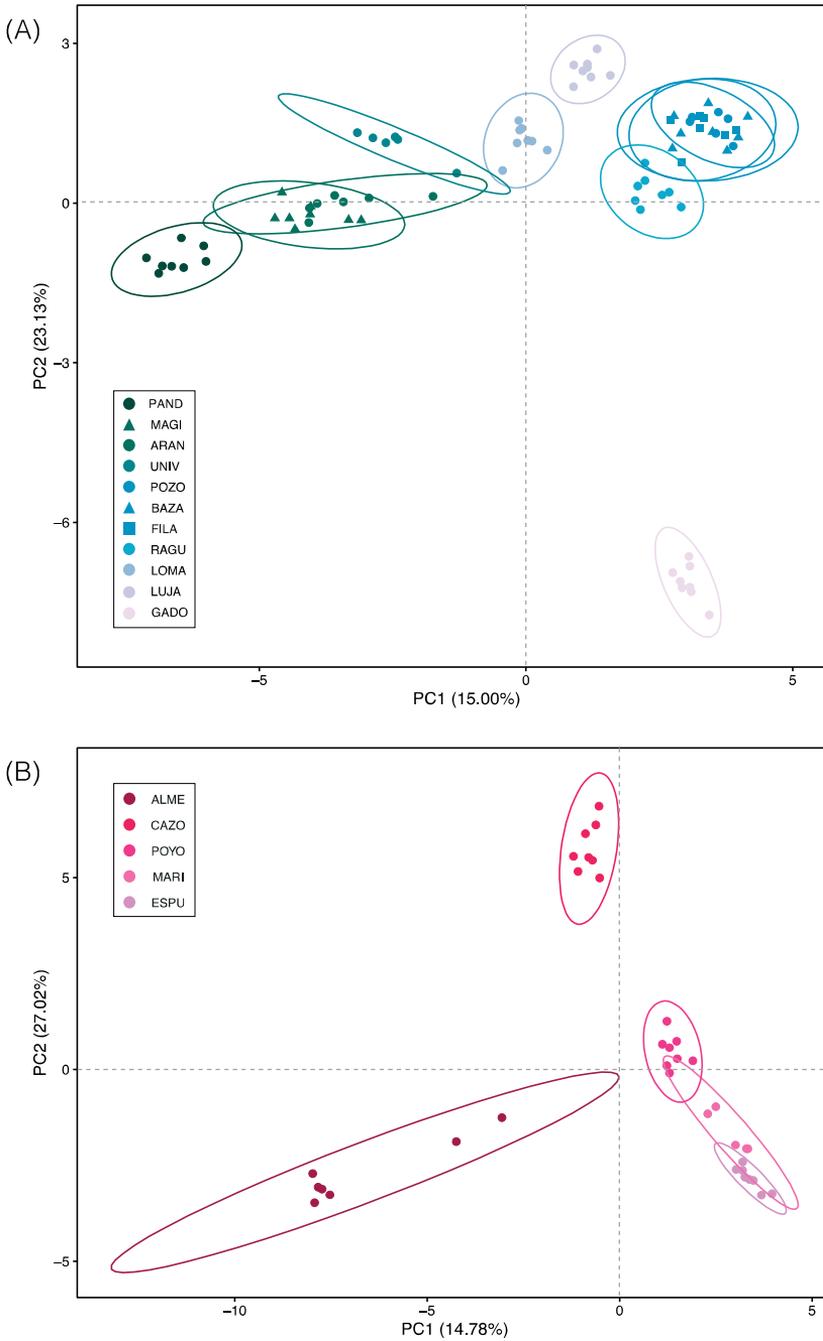
**Figure S3** Results of STRUCTURE analyses for (A) *Omocestus bolivari* and (B) *O. femoralis*. The plots show mean ( $\pm$ SD) log probability of the data ( $\text{Ln Pr}(X|K)$ ) over 10 runs of STRUCTURE (left axes, solid dots and error bars) for each value of  $K$  and the magnitude of  $\Delta K$  (right axes, open dots).



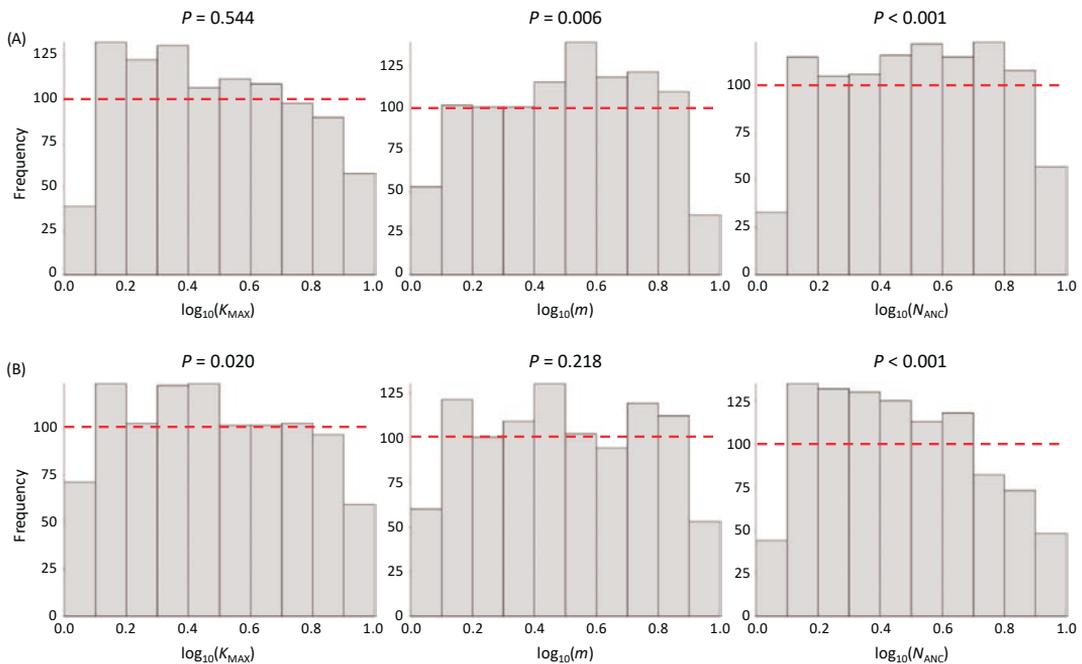
**Figure S4** Genetic assignment of individuals based on the results of STRUCTURE for (A) *Omocestus bolivari* (from  $K = 2$  to  $K = 8$ ) and (B) *O. femoralis* (from  $K = 2$  to  $K = 5$ ). Individuals are partitioned into  $K$  colored segments representing the probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1.



**Figure S5** Principal component analysis (PCAs) of genetic variation for (A) *Omocestus bolivari* and (B) *O. femoralis*. Colors indicate the main genetic cluster at which individuals were assigned according to STRUCTURE analyses for  $K = 8$  (*O. bolivari*) and  $K = 5$  (*O. femoralis*). Population codes as in Table 1.



**Figure S6** Distribution of posterior quantiles of true parameter values from 1,000 pseudo-observed data sets (PODs) used to assess bias in parameter estimation for the most supported model for (A) *Omocestus bolivari* and (B) *O. femoralis*. Posterior quantiles (grey bars) are compared to a uniform distribution (dashed red line) using a Kolmogorov–Smirnov test. Significant  $p$ -values indicate a deviation from a uniform distribution and potential bias in parameter estimation.  $K_{MAX}$ , carrying capacity of the deme with highest suitability (100%);  $m$ , migration rate per deme per generation;  $N_{ANC}$ , ancestral population size.



## Supplemental references

- Beaumont, M. A., Zhang, W. Y., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025-2035.
- Bemmels, J. B., Title, P. O., Ortego, J., & Knowles, L. L. (2016). Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology*, *25*(19), 4889-4906. doi:10.1111/mec.13804
- Boulesteix, A. L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, *8*(1), 32-44. doi:10.1093/bib/bb1016
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. doi:10.1111/mec.12354
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genetics*, *1*(3), 171-182. doi:10.1534/g3.111.000240
- Cook, S. R., Gelman, A., & Rubin, D. B. (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, *15*(3), 675-692. doi:10.1198/106186006x136976
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & Francois, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*(7), 410-418. doi:10.1016/j.tree.2010.04.001
- Curat, M., Ray, N., & Excoffier, L. (2004). SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, *4*(1), 139-142. doi:10.1046/j.1471-8286.2003.00582.x
- Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285-298. doi:10.1038/hdy.2009.74
- Excoffier, L., & Lischer, H. E. L. (2010). ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x
- 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. doi:10.1534/genetics.108.092221
- González-Serna, M. J., Cordero, P. J., & Ortego, J. (2019). Spatiotemporally explicit demographic modelling supports a joint effect of historical barriers to dispersal and contemporary landscape composition on structuring genomic variation in a red-listed grasshopper. *Molecular Ecology*, *28*(9), 2155-2172. doi:10.1111/mec.15086

- He, Q. X., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, *67*(12), 3386-3402. doi:10.1111/evo.12159
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862. doi:10.1371/journal.pgen.1000862
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795. doi:10.1080/01621459.1995.10476572
- Knowles, L. L., & Massatti, R. (2017). Distributional shifts - not geographic isolation - as a probable driver of montane species divergence. *Ecography*, *40*(12), 1475-1485. doi:10.1111/ecog.02893
- Leuenberger, C., & Wegmann, D. (2010). Bayesian computation and model selection without likelihoods. *Genetics*, *184*(1), 243-252. doi:10.1534/genetics.109.109058
- Liu, C. R., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*(3), 385-393. doi:10.1111/j.0906-7590.2005.03957.x
- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, *25*(16), 3974-3986. doi:10.1111/mec.13735
- Mevik, B. H., & Wehrens, R. (2007). The *pls* package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, *18*(2), 1-23.
- Neuenschwander, S., Largiadèr, C. R., Ray, N., Currat, M., Vonlanthen, P., & Excoffier, L. (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, *17*(3), 757-772.
- Ortego, J., Gugger, P. F., & Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist*, *218*(2), 804-818. doi:10.1111/nph.14951
- Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, *26*(23), 2993-2994. doi:10.1093/bioinformatics/btq579
- Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABC TOOLBOX: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*, 116. doi:10.1186/1471-2105-11-116



# ORIGINAL JOURNAL ARTICLES



Received: 12 December 2018 | Revised: 14 July 2019 | Accepted: 15 July 2019

DOI: 10.1111/mec.15189

## ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

# Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex

Vanina Tonzo<sup>1</sup>  | Anna Papadopoulou<sup>2</sup> | Joaquín Ortego<sup>1</sup>

<sup>1</sup>Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC), Seville, Spain

<sup>2</sup>Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus

**Correspondence**

Vanina Tonzo, Estación Biológica de Doñana, EBD-CSIC, Seville, Spain.  
Email: vaninatanzo@gmail.com

**Funding information**

Ministerio de Economía y Competitividad, Grant/Award Number: BES-2015-73159; European Regional Development Fund (ERDF), Grant/Award Number: CGL2014-54671-P and CGL2017-83433-P; Severo Ochoa, Grant/Award Number: SEV-2012-0262; Ramon y Cajal, Grant/Award Number: RYC-2013-12501

**Abstract**

Taxonomy has traditionally relied on morphological and ecological traits to interpret and classify biological diversity. Over the last decade, technological advances and conceptual developments in the field of molecular ecology and systematics have eased the generation of genomic data and changed the paradigm of biodiversity analysis. Here we illustrate how traditional taxonomy has led to species designations that are supported neither by high throughput sequencing data nor by the quantitative integration of genomic information with other sources of evidence. Specifically, we focus on *Omocestus antigai* and *Omocestus navasi*, two montane grasshoppers from the Pyrenean region that were originally described based on quantitative phenotypic differences and distinct habitat associations (alpine vs. Mediterranean-montane habitats). To validate current taxonomic designations, test species boundaries, and understand the factors that have contributed to genetic divergence, we obtained phenotypic (geometric morphometrics) and genome-wide SNP data (ddRADSeq) from populations covering the entire known distribution of the two taxa. Coalescent-based phylogenetic reconstructions, integrative Bayesian model-based species delimitation, and landscape genetic analyses revealed that populations assigned to the two taxa show a spatial distribution of genetic variation that do not match with current taxonomic designations and is incompatible with ecological/environmental speciation. Our results support little phenotypic variation among populations and a marked genetic structure that is mostly explained by geographic distances and limited population connectivity across the abrupt landscapes characterizing the study region. Overall, this study highlights the importance of integrative approaches to identify taxonomic units and elucidate the evolutionary history of species.

**KEYWORDS**

ddRADseq, geometric morphometrics, integrative species delimitation, landscape genetics, phylogenomic inference, species delimitation

## 1 | INTRODUCTION

Describing biological diversity and understanding the ecological and evolutionary processes that generate it is at the core of molecular ecology and allied disciplines (Avice et al., 1987; Dobzhansky, 1937; Simpson, 1944). In the last decade, this field has experienced a revolution as a result of new technological and conceptual advances that allow biologists to discover new taxa at an unprecedented resolution and gain new insights into the processes leading to species formation (Jackson, Carstens, Morales, & O'Meara, 2017; Sites & Marshall, 2003; Wiens, 2007). These advances are timely, as the current biodiversity crisis—i.e., the accelerated loss of species and the rapid degradation of ecosystems—requires the establishment of efficient conservation policies that often put the focus on red lists of formally described taxa (Agapow et al., 2004; Mace, 2004). Yet, the definition of the fundamental biological unit that conforms a species remains controversial due to many alternative species concepts and the different properties/criteria used to define them (see de Queiroz, 2007). Given the inherent difficulties to test for reproductive isolation (i.e., the biological species concept; Dobzhansky, 1970; Mayr, 1942), especially in allopatric populations, most species descriptions have been traditionally based on qualitative or quantitative phenotypic differences (diagnosable or phenetic species concept; e.g., Nelson & Platnick, 1981; Sokal & Crovello, 1970) or distinct ecological traits (e.g., Jones & Weisrock, 2018 and references therein). However, speciation may not necessarily involve shifts in niche preferences or diagnosable morphological changes and adhering to these two sources of evidence could result in an underestimation of the number of biological species.

The wide use of DNA sequence data in taxonomy (e.g., Hebert, Cywinska, Ball, & DeWaard, 2003; Tautz, Arctander, Minelli, Thomas, & Vogler, 2003) and the application of phylogenetic/coalescent-based species delimitation approaches (e.g., Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012; Pons et al., 2006; Yang & Rannala, 2010) have allowed the discovery of cryptic species and the resolution of many taxonomic uncertainties (e.g., Blair et al., 2019; Burbrink et al., 2011; Leavitt, Johnson, Goward, & St Clair, 2011; Niemiller, Near, & Fitzpatrick, 2012). In particular, the incorporation of coalescent theory under an integrative taxonomic framework has increased the statistical rigor of species delimitation and has helped to move away from subjective decisions on the degree of differentiation (e.g., phenotypic, genetic, ecological, etc) that is needed for considering different lineages or populations as distinct taxa (Fujita et al., 2012; Jones & Weisrock, 2018; Wiens & Servedio, 2000). In recent years, the advent of high throughput DNA sequencing technology has exponentially increased our capability to obtain large scale genomic data from non-model organisms, removing previous data-related constraints, while promoting further development of new phylogenetic inference methods and model-based approaches of species delimitation (van Dijk, Auger, Jaszczyszyn, & Thermes, 2014; Ekblom & Galindo, 2011; Fujita et al., 2012). In particular, the multispecies coalescent model (MCM) is considered an excellent approach to test alternative hypotheses of lineage divergence (Knowles & Carstens,

2007; Yang & Rannala, 2010) and identify boundaries among recently diverged species using multilocus data (Domingos, Colli, Lemmon, Lemmon, & Beheregaray, 2017; Rannala & Yang, 2013; Yang, 2015). More recent developments have incorporated the possibility of combining diverse sources of information (quantitative traits and genomic data) to objectively identify independently evolving lineages into an integrative analytical framework (Edwards & Knowles, 2014; Fujita et al., 2012; Solis-Lemus, Knowles, & Ane, 2015). However, integrative species delimitation approaches are not exempt of limitations (Huang, 2018; Sukumaran & Knowles, 2017). One of these limitations is the challenge to deal with sexually dimorphic traits (Solis-Lemus et al., 2015), which can have a considerable impact on species delimitation inferences (Chan et al., 2017; Noguerales, Cordero, & Ortego, 2018). More importantly, the finest detection of genetic differentiation brought by the highest resolution of genomic data can lead to a potential confusion of population genetic structure with species boundaries (Carstens, Pelletier, Reid, & Satler, 2013; O'Meara, 2010; Sukumaran & Knowles, 2017). That is, in some situations, the higher power of genomic data to detect population structuring may drive to an artefactual increase in the number of inferred species (Fujita et al., 2012; Huang, 2018; Isaac, Mallet, & Mace, 2004). Taxonomic inflation, in turn, can have a negative impact on subsequent ecological and evolutionary studies, compromise our ability to reach solid conclusions on the origin and dynamics of biodiversity and, ultimately, lead to misguided conservation policies and inadequate management practices (Carstens et al., 2013; O'Meara, 2010; Sukumaran & Knowles, 2017).

As mentioned above, a common source of taxonomic controversy is the historical description of species based on the concurrence of ecological and phenotypic distinctiveness (e.g., Dowle, Morgan-Richards, & Trewick, 2014; Jones & Weisrock, 2018). These are often presumed to represent cases of ecological speciation, i.e., the formation of new species when divergent natural selection under contrasting environmental conditions (e.g., elevation, salinity, etc) leads to reproductive isolation (Rundle & Nosil, 2005; Schluter, 2001). Genetic data have often supported classic taxonomic studies separating species based on ecological and phenotypic differences (e.g., Lamichhane et al., 2015). However, in many other cases the recognized taxonomic units are not consistent with patterns of genome-wide differentiation (e.g., Dowle et al., 2014; Jones & Weisrock, 2018), suggesting that the link between environmental and phenotypic divergence is a consequence of plasticity or divergent selection at a few genes or genomic islands involved in local adaptation processes that have not led to the reduction of gene flow (e.g., Mason & Taylor, 2015; Soria-Carrasco et al., 2014). An intermediate situation along the speciation continuum is the frequently reported association between environmental dissimilarity and phenotypic and genetic differentiation found in many studies (Sexton, Hangartner, & Hoffmann, 2014; Shafer & Wolf, 2013; Wang & Bradburd, 2014). This pattern of progressive genetic and phenotypic differentiation among populations experiencing contrasting environmental conditions has been termed isolation-by-ecology (IBE) and is generally interpreted as a signal of ongoing local

adaptation processes and incipient speciation (Shafer & Wolf, 2013). Thus, the study of these evolutionary phenomena is not only relevant in the context of resolving taxonomic uncertainties, but it also allows to identify important ecological and evolutionary processes along the speciation continuum that might deserve to be protected (Moritz, 2002).

The grasshopper subgenus *Dreuxius* Defaut, 1988 (genus *Omocestus*, Bolívar, 1878) is a complex of nine recently diversified species distributed in the Iberian Peninsula (six species) and north-western Africa (three species; Cigliano, Braun, Eades, & Otte, 2019; García-Navas, Noguerales, Cordero, & Ortego, 2017). Most taxa have allopatric distributions and are often isolated at high elevations in different mountain systems (Cigliano et al., 2019). The putative sister species pair *Omocestus antigai* (Bolívar, 1897) and *Omocestus navasi* (Bolívar, 1908) are distributed through the Pyrenees, including some populations in the pre-Pyrenees and Catalan Pre-Coastal ranges (Poniatowski, Defaut, Lluçia-Pomares, & Fartmann, 2009). These two species were originally described based on distinct habitat associations and certain quantitative phenotypic traits such as body size and forewing shape (Clemente, García, Arnaldos, Romera, & Presa, 1999; Olmo-Vidal, 2002; Puissant, 2008). The current known ranges of both species follow the west to east orientation of the Pyrenees, but their populations are mostly allopatric. *Omocestus antigai* is distributed at high elevations (1,450–2,350 m) associated with alpine or subalpine grasslands above the tree line, whereas *O. navasi* is circumscribed to Mediterranean scrubby habitats at lower elevations (700–1,600 m; Lluçia-Pomares, 2002; Olmo-Vidal, 2002; Poniatowski et al., 2009). These strong ecological differences and subtle morphological variation have motivated different authors to synonymize the two taxa (Ragge & Reynolds, 1998; Reynolds, 1986), revalidate their distinct species status (Clemente et al., 1999) and even describe a new subspecific taxon (*O. navasi bellmanni*; Puissant, 2008) in the last decades (Cigliano et al., 2019). Resolving the taxonomy of the complex also has important implications for conservation and management, as the two putative species are included in the IUCN red list of threatened species but with different assessment categories (Braud, Hochkirch, Fartmann, et al., 2016; Braud, Hochkirch, Presa, et al., 2016). *Omocestus navasi* is classified as “endangered” due to its extremely small estimated area of occupancy (60–200 km<sup>2</sup>) and the continuous decline in the extend and quality of habitats (Braud, Hochkirch, Fartmann, et al., 2016). In contrast, the high-elevation grasslands occupied by *O. antigai* are supposed to be less affected by direct human impacts and this taxon was assessed as “vulnerable”, with the main threats identified being the small size and high degree of fragmentation of its populations, overgrazing in some areas, and the decline of the quality of its specific alpine and subalpine habitats due to the progressive abandonment of traditional grassland management practices (Braud, Hochkirch, Presa, et al., 2016). Overall, this makes this complex an excellent model system to test species boundaries and understand the factors that have contributed to genomic, phenotypic and ecological divergence, which will in turn provide

the necessary baseline information to determine the conservation status of the different taxonomic units within the complex and design proper on-ground management practices.

In this study, we generated genomic data using restriction-site-associated DNA sequencing (ddRADseq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) and obtained phenotypic information through geometric morphometric analyses (Bookstein, 1992) to shed light on the taxonomy and evolutionary history of *O. antigai* and *O. navasi*. First, we tested the hypothesis that the current taxonomic designation of the complex is supported by genomic and phenotypic data. In particular, we used genome-wide single nucleotide polymorphisms (SNP) data and alternative coalescent-based methods to infer the phylogenetic relationships among all sampled populations of the two taxa. Then, we compared the taxonomic status of the studied populations with phylogenomic reconstructions and inferred patterns of spatial genetic structure (Dayrat, 2005; Jones & Weisrock, 2018), while we employed a suite of Bayesian model-based species delimitation approaches, one of them integrating genomic data and quantitative traits, to identify independently-evolving lineages (Leaché, Fujita, Minin, & Bouckaert, 2014; Solis-Lemus et al., 2015; Yang & Rannala, 2010). Second, we tested the hypothesis that ecological-driven divergence is the main process underlying spatial patterns of genomic and phenotypic variation in the study system. In particular, we analyzed whether environmental dissimilarity among populations and processes of local adaptation to different elevational and climatic gradients are the main factors shaping genetic differentiation and phenotypic trait variation in the study system or if, on the contrary, genetic and phenotypic structure is mostly explained by the geographical distance among populations and resistance distances defined by the complex topography of the study area.

## 2 | MATERIALS AND METHODS

### 2.1 | Population sampling

Between 2012 and 2015, we collected specimens of *Omocestus antigai* (seven populations), *O. navasi navasi* (eight populations) and *O. navasi bellmanni* (one population) from 16 sampling localities (hereafter referred to as populations; Table 1 and Figure 1). Samples represent populations across the entire known distribution range of the three taxa based on the literature (Clemente et al., 1999; Lluçia-Pomares, 2002; Poniatowski et al., 2009; Puissant, 2008). We stored specimens in 2 ml vials with 96% ethanol and preserved them at -20°C until needed for geometric morphometric and genomic analyses.

### 2.2 | Geometric morphometric analyses

We selected a maximum of 10 adult individuals from each population (five males and five females if available) and used landmark-based geometric morphometric analyses to characterize phenotypic variation in the complex. We excluded from geometric

Species	Locality	Code	Latitude	Longitude	Elevation (m)	n
<i>O. antigai</i>	Boí Taüll	BOI	42.47945	0.86732	2,046	5
<i>O. antigai</i>	Llessui	LLE	42.44554	1.03668	1,979	6
<i>O. antigai</i>	Coll de la Creueta	CRE	42.30131	1.99350	1,928	6
<i>O. antigai</i>	Err	ERR	42.38830	2.08812	2,150	3
<i>O. antigai</i>	Queralbs	QUE	42.36627	2.14902	2,070	6
<i>O. antigai</i>	Turó de l'Home	TUR	41.77167	2.44335	1,622	7
<i>O. antigai</i>	Setcases	SET	42.42756	2.26630	2,145	7
<i>O. navasi navasi</i>	Borau	BOR	42.67388	-0.57932	1,357	7
<i>O. navasi navasi</i>	Puerto de Monrepós	MON	42.35067	-0.39762	1,267	5
<i>O. navasi navasi</i>	Nerín	NER	42.59047	0.01062	1,623	7
<i>O. navasi navasi</i>	Saravillo	SAR	42.56089	0.23046	1,313	2
<i>O. navasi navasi</i>	Chía	CHI	42.55418	0.43703	1,691	7
<i>O. navasi navasi</i>	Espés	ESP	42.44307	0.58389	1,361	6
<i>O. navasi navasi</i>	Perves	PER	42.35250	0.83709	1,370	6
<i>O. navasi navasi</i>	Montan de Tost	TOS	42.23790	1.38052	1,182	7
<i>O. navasi bellmanni</i>	Tour de Batère	BAT	42.50785	2.57641	1,430	7
<i>O. femoralis</i>	Sierra de Espuña	ESP	37.86515	-1.57115	1,514	6

**TABLE 1** Geographical location, elevation and number of genotyped individuals (n) for each sampled population (locality) of the three putative taxa and the outgroup

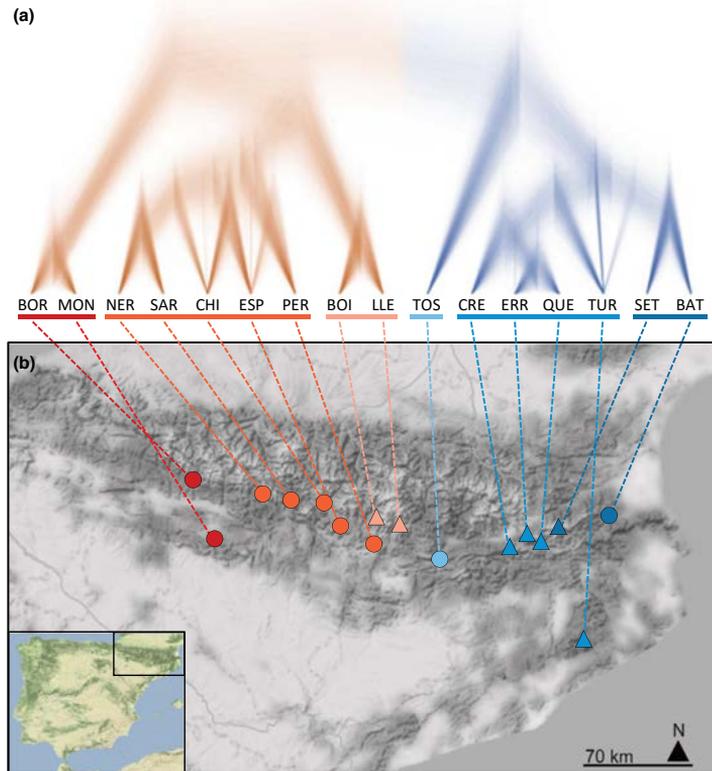
morphometric analyses five populations (MON, SAR, ERR, QUE, and TUR) for which only a few individuals for one or the two sexes (<3) could be sampled due to low population densities and/or because most specimens presented damaged structures (e.g., broken forewings or ovipositor valves). Populations with available geometric morphometric data included all taxa within the complex (i.e., all species/subspecies) and were representative of all genetic clusters inferred by Bayesian clustering analyses (see Section 3). We took digital images of head, pronotum, forewing and ovipositor valves (in females) with a ZEISS stereomicroscope (STEREO DISCOVERY version 8; Carl Zeiss Microscopy GmbH). These traits correspond to those originally used to distinguish the two putative species (Clemente et al., 1999). The coordinates of landmarks (9–14 landmarks per trait) were mapped on the images using TPSDIG version 2.2 (Rohlf, 2015) and analyzed as implemented in the R version 3.3.2 (R Core Team, 2018) package *Geomorph* (Adams & Otárola-Castillo, 2013). Specifically, we performed generalized Procrustes analyses (GPA) separately for each sex to remove the effects of location, size, and rotation of the relative positions of landmarks among specimens using the function *gpagen*. This superimposition method minimizes the sum-of-squared distances between landmarks across samples (Rohlf & Slice, 1990). To examine shape differences between taxa and among populations, we performed principal component analyses (PCA) on the covariance matrix of aligned Procrustes shape coordinates using the function *plotTangentSpace*. We retained the scores from the two first principal components (PC1 and PC2) for each trait and used multivariate analyses of variance (MANOVA) to investigate if trait shapes were significantly different between taxa

and among sampled populations (e.g., Adams & Otárola-Castillo, 2013; Friedline et al., 2019). Statistical significance of MANOVA models was determined using Wilk's  $\lambda$  as the test statistic and  $\alpha = 0.05$  as a significance threshold. Then, we used one-way analyses of variance (ANOVA) to assess differences between taxa and among sampled populations separately for each phenotypic trait (e.g., Friedline et al., 2019). The first two PCs for each trait were also used for integrative species delimitation analyses using iBPP (Solis-Lemus et al., 2015) and to build matrices of phenotypic differentiation ( $P_{ST}$ ; see below for details).

### 2.3 | Genomic library preparation and processing

We obtained genomic data for a subset of 94 specimens representative from the 16 sampled populations of *O. antigai*, *O. navasi navasi*, and *O. navasi bellmanni* (~six individuals per population, three males and three females when available; Table 1). Additionally, we genotyped six individuals of *O. femoralis* Bolívar, 1908 (Table 1), a species also belonging to the subgenus *Dreuxius* (Cigliano et al., 2019) and that was used as an outgroup in some phylogenetic and species delimitation analyses (see below for details). Details on the preparation of ddRADseq libraries (Peterson et al., 2012) are presented in Methods S1. Raw sequences were demultiplexed and preprocessed using STACKS version 1.35 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) and assembled using PYRAD version 3.0.66 (Eaton, 2014). Methods S2 provides all details on sequence assembling and data filtering.

**FIGURE 1** (a) Bayesian phylogenetic tree reconstructed with *SNAPP* (node support presented in Figure 5) and (b) map showing the distribution of the sampled populations of *Omocestus antigai* (triangles) and *Omocestus navasi* (dots) in the Pyrenees. Colours of triangles and dots on the map indicate the main genetic cluster at which populations were assigned according to *STRUCTURE* analyses for  $K = 6$ . Population codes as in Table 1



## 2.4 | Population genetic structure

We employed four complementary approaches to exhaustively explore spatial patterns of genetic structure in our study system, including: (a) classic *STRUCTURE* analyses (Pritchard, Stephens, & Donnelly, 2000); (b) *FASTSTRUCTURE* (Raj, Stephens, & Pritchard, 2014); (c) the recently developed spatial method implemented in the *R* program *CONSTRUCT* to infer patterns of genetic structure after controlling for the geographic distance separating the sampled populations (Bradburd, Coop, & Ralph, 2018), given the strong signal of isolation by distance in our study system (see Section 3.6); and (d) a principal component analysis (PCA; Jombart, 2008). Further details on these analyses are presented in Methods S3.

## 2.5 | Phylogenetic inference

We estimated species trees using two coalescent-based methods, *SNAPP* version 1.3 (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) as implemented in *BEAST2* version 2.4.3 (Bouckaert et al., 2014) and *SVDQUARTETS* (Chifman & Kubatko, 2014) as implemented in *PAUP\** v. 4.0a152 (Swofford, 2002). *SNAPP* analyses are computationally highly demanding and, for this reason, we only

selected one individual (the one with the highest number of retained reads; Figure S1) from each of the 16 sampled populations. The resulting data set contained 16 individuals and retained 858 polymorphic sites. We ran *SNAPP* analyses for 1,000,000 Markov chain Monte Carlo (MCMC) generations, sampling every 1,000 steps, and using different gamma prior distributions for  $\alpha$  and  $\beta$  (2, 200; 2, 2,000; 2, 20,000). The forward ( $u$ ) and reverse ( $v$ ) mutation rates were set to be calculated by *SNAPP* and we left the remaining parameters at default values. We conducted two independent runs and evaluated convergence with *TRACER* version 1.6. We removed 10% of trees as burnin and merged tree and log files from the different runs using *LOGCOMBINER* version 2.4.1. We used *TREEANNOTATOR* version 1.8.3 to obtain maximum credibility trees, *TREESETANALYSERVER* version 2.4.1 to identify species trees that were contained in the 95% highest posterior density (HPD), and *DENSITREE* version 2.2.1 (Bouckaert, 2010) to visualize the posterior distribution of trees.

*SVDQUARTETS* uses SNP data to infer phylogenetic relationships between quartets of taxa under the multispecies coalescent (MSC) model and then assembles these quartets into a species tree (Chifman & Kubatko, 2014). We performed *SVDQUARTETS* analyses including all individuals and populations and setting *O. femoralis* as an outgroup. We constructed species trees by exhaustively evaluating

Species delimitation hypothesis ( $H_i$ )	Species	MLE	$2 \times \ln \text{BF}$	Rank
$H_1$ : (BAT + PER + SAR + CHI + TOS + NER + ESP + MON + BOR + CRE + ERR + BOI + LLE + QUE + TUR + SET)	1	-4,794.85	$8.21 \times 10^2$	2
$H_2$ : (BOI + LLE + CRE + ERR + QUE + TUR + SET) (PER + SAR + CHI + NER + ESP + MON + BOR + BAT + TOS)	2	-5,601.22	$2.43 \times 10^3$	5
$H_3$ : (BAT) (PER + SAR + CHI + TOS + NER + ESP + MON + BOR) (CRE + ERR + BOI + LLE + QUE + TUR + SET)	3	-5,384.07	$2.00 \times 10^3$	4
$H_4$ : (BAT + TOS + CRE + ERR + QUE + TUR + SET) (PER + SAR + CHI + NER + ESP + MON + BOR + BOI + LLE)	2	-5,221.06	$1.67 \times 10^3$	3
$H_5$ : (BAT + SET) (TOS) (CRE + ERR + QUE + TUR) (PER + SAR + CHI + NER + ESP) (MON + BOR) (BOI + LLE)	6	-4,384.48	—	1

**TABLE 2** Results of  $\text{BFD}^*$  analyses testing the support of competing species delimitation hypotheses. The table shows the clustering scheme defining each alternative species delimitation hypothesis ( $H_i$ ). For each hypothesis, we show marginal likelihood estimates (MLE), their Bayes factors (calculated as  $2 \times \ln \text{BF}$ ) and their rank. Population codes as in Table 1

all possible quartets (i.e., every combination of four tips was examined) from the entire genomic data set (63,860 unlinked SNPs) and used 100 nonparametric bootstrapping replicates to assess branch support (Chifman & Kubatko, 2014).

## 2.6 | Species delimitation

We applied three Bayesian coalescent-based species delimitation approaches to determine the number of independently evolving lineages, two of them based only on molecular data (BPP; Yang & Rannala, 2010;  $\text{BFD}^*$ ; Leaché et al., 2014) and the third one integrating molecular and phenotypic data (IBPP; Solis-Lemus et al., 2015).

We used the  $\text{BFD}^*$  method implemented in  $\text{SNAPP}$  (Leaché et al., 2014) to contrast five competing species delimitation hypotheses, including a single-species model and four alternative multi-species models informed by the current taxonomy as well as the phylogenetic and Bayesian clustering analyses (Table 2). This method allows the comparison of alternative species delimitation scenarios in an explicit MSC framework by calculating and comparing marginal likelihood estimates (MLE) for each model. Specifically, our species delimitation hypotheses included: (a) the hypothesis of a single species ( $H_1$ ); (b) the current taxonomy of two species ( $H_2$ : *O. antigai* and *O. navasi*); (c) the current taxonomy but considering the subspecies *O. navasi bellmanni* as a separate species ( $H_3$ : three species); (d) a model considering as distinct species the two main genetic clusters (east-west split) revealed by phylogenetic and Bayesian clustering analyses ( $H_4$ ) and (e) a model considering six species, corresponding to the six genetic clusters inferred by Bayesian clustering analyses in  $\text{STRUCTURE}$  ( $H_5$ ; Table 2). We conducted a path sampling analysis of eight steps each consisting of 100,000 MCMC iterations (after a preburnin of 10,000 iterations), sampling each 100 steps, and using an  $\alpha$  value of 0.3. These settings were sufficient to ensure convergence and obtain  $\text{ESS} > 200$ . The Bayes factor (BF) test statistic ( $2 \times \ln \text{BF}$ ) was calculated, where BF is the difference in MLE between two competing models

(base scenario-alternative scenario). These analyses were performed using a matrix with no missing data (393 SNPs; see Leaché, McElroy, & Trinh, 2018 for a similar approach) and including two individuals from each sampled population plus two individuals from the outgroup *O. femoralis*, to be able to test for the single-species model.

BPP version 3.3 can perform species delimitation analyses using a fixed input guide tree specified by the user (option A10, guided analysis) or estimating a species tree (option A11, unguided analysis; Yang, 2015; Yang & Rannala, 2014). We used both options and for guided analyses we fitted as input trees the topologies yielded by  $\text{SNAPP}$  and  $\text{SVDQUARTETS}$  (see Section 3). This allowed us to assess the potential impact that alternative tree topologies had on our species delimitation inferences (Leaché & Fujita, 2010). To make BPP analyses computationally tractable, we only included two individuals from each sampled population (same as used for  $\text{BFD}^*$  analyses) and used a subset of loci from our total genomic data set (e.g., Huang, 2018; Nogueras et al., 2018; Rancillac et al., 2019). Specifically, we selected loci with five or more variable sites, as these provide more power in species delimitation analyses than less variable loci (Huang, 2018). To evaluate the impact that the number of employed loci had on our species delimitation inferences, we used three random subsets of loci (200, 500 and 1,000 loci). According to previous studies, these numbers of loci provide considerable power for species delimitation analyses (e.g., Huang, 2018; Nogueras et al., 2018). We created BPP input files from the loci output file from  $\text{PYRAD}$  and randomly selected the different subsets of loci with the specified minimum number of variable sites ( $\geq 5$ ) using two R scripts (`bpp_convert_Ama_sp.r` and `numvarfrombppfile.r`, respectively) written by Huang (2018). We analyzed the impact of different demographic scenarios corresponding to different prior combinations of gamma distribution as in Huang and Knowles (2016). We considered four parameter settings:  $\theta = G(1, 10)$  and  $\tau = G(1, 10)$ ,  $\theta = G(1, 10)$  and  $\tau = G(2, 2,000)$ ,  $\theta = G(2, 2,000)$  and  $\tau = G(1, 10)$ , and  $\theta = G(2, 2,000)$  and  $\tau = G(2, 2,000)$ , where  $\theta$  and  $\tau$  refer to the ancestral population

sizes and divergence times, respectively. We ran four replicates for each combination. We used an automatic adjustment of the fine-tune parameters, allowing swapping rates to range between 0.30 and 0.70 (Yang, 2015). We ran each analysis for 100,000 generations, sampling every 10 generations (10,000 samples), after a burnin of 100,000 generations.

The integrative *IBPP* approach was built upon the architecture of the early version of *BPP* version 2.1.2 (Rannala & Yang, 2013; Yang & Rannala, 2010) and incorporates models of evolution for continuous quantitative traits under a Brownian motion process (Solis-Lemus et al., 2015). We ran *IBPP* analyses separately for each sex, incorporating genetic data as well as geometric morphometric data (PC1 and PC2) of three traits for males and four traits for females (see Section 2.2). As in *BPP*, we ran three independent analyses using as guide trees the topologies yielded by *BPP* (obtained using option A01; Yang, 2015), *SNAPP* and *SVDQUARTETS*, used four prior combinations for gamma distribution and same settings for all other parameter values, and ran four independent replicates for each scenario. These analyses were performed only for populations with available phenotypic data (i.e., excluding MON, SAR, ERR, QUE, and TUR), using a random subset of 500 loci with at least five variable sites, and including five individuals/population (rather than two individuals/population as in *BPP* analyses) in order to increase the accuracy of estimates of the phenotypic variation of populations. We ran *IBPP* analyses for 250,000 generations, sampling every 10 generations (25,000 samples), after a conservative burnin of 300,000 generations. Finally, we repeated these analyses considering only phenotypic data for each sex. Posterior probability (PP) of both *BPP* and *IBPP* models was considered well supported when  $PP > 0.95$ .

## 2.7 | Landscape genetic and phenotypic analyses

We analyzed three potential factors that could have shaped genetic ( $F_{ST}$ ) and phenotypic ( $P_{ST}$ ) differentiation among populations: (a) geographical distance; (b) environmental heterogeneity; (c) and isolation-by-resistance defined by topographic complexity. We estimated genetic differentiation between each pair of sampled populations calculating pairwise  $F_{ST}$  values in *ARLEQUIN* version 3.5 (Excoffier, Laval, & Schneider, 2005). Similarly, we estimated phenotypic differentiation between each pair of sampled populations calculating pairwise  $P_{ST}$  values for the first two PCs of each trait using the *R* package *Pstat* (da Silva & da Silva, 2018). We calculated the geographical distance between sampled populations using *GEOGRAPHIC DISTANCE MATRIX GENERATOR* version 1.2.3 ([http://biodiversityinformatics.amnh.org/open\\_source/gdmg](http://biodiversityinformatics.amnh.org/open_source/gdmg)). To estimate environmental dissimilarity between each pair of sampled populations, we extracted for each one the values of the 19 present-day bioclimatic variables available in *WORLDCLIM*, downloaded at 30 arc-s (c. 1 km) resolution. Then, we used the “*rda*” function in the *R* package *vegan* version 2.4-4 (Oksanen et al., 2017) to perform a principal component analysis (PCA) and obtain for each population the PC scores of the first two PCs, which explained the 80% and 16% of the environmental

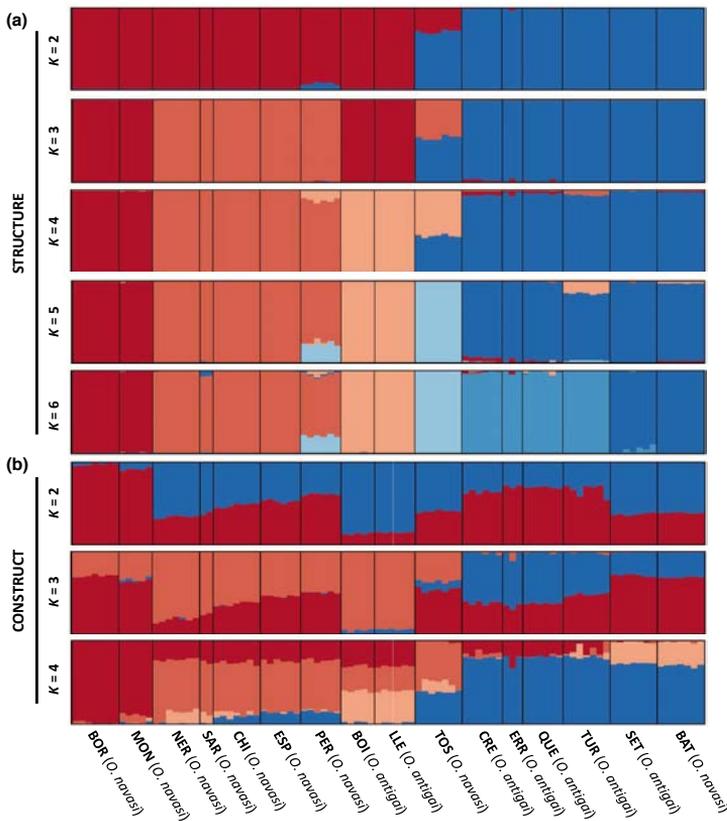
variance, respectively. Afterward, we calculated environmental dissimilarity between each pair of populations using Euclidean distances for the obtained PC scores using the “*dist*” function in *R*. To investigate the role of topographic complexity, we calculated the slope for each cell from a 90 m resolution digital elevation model from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data) and the final raster layer was transformed to 30 arc-s (c. 1 km) resolution using *QGIS* version 2.8. Then, based on this layer, we used *CIRCUITSCAPE* version 4.0 (McRae, 2006; McRae & Beier, 2007) to calculate a matrix of resistance distances between all pairs of populations considering an eight-neighbour cell connection scheme.

We used multiple matrix regressions with randomization (MMRR) as implemented in *R* (Wang, 2013) to analyze: (a) pairwise population genetic differentiation ( $F_{ST}$ ) in relation with geographical, resistance and environmental distances, and (b) pairwise population phenotypic differentiation ( $P_{ST}$ ) in relation with genetic differentiation ( $F_{ST}$ ) and geographical, resistance and environmental distances. Models for all dependent variables were initially constructed with all explanatory terms fitted (i.e., full models) and final models were selected using a backward-stepwise procedure, by progressively removing nonsignificant variables (starting with the least significant ones) until all retained terms within the model were significant. Then, we tested the significance of the rejected terms against this model to ensure that no additional variable reached significance. The result was the minimal most adequate model for explaining the variability in the response variable, where only the significant explanatory terms were retained. As we tested a high number of phenotypic traits (eight traits for females and six for males) against the same subset of independent variables, we applied a Benjamini-Hochberg false discovery rate correction to adjust *p*-values for multiple statistical tests and calculate the corresponding *q*-values using the *p.adjust* function in *R*.

## 3 | RESULTS

### 3.1 | Geometric morphometric analyses

The two first principal component scores, PC1 and PC2, for each trait explained most variation in shapes for pronotum (♀: 87%; ♂: 76%), head (♀: 90%; ♂: 91%), forewing (♀: 65%; ♂: 77%) and ovipositor valves (♀: 95%). Principal component plots for each trait and sex are shown in Figures S2 and S3. MANOVAs based on PC scores for all traits showed significant differences among sampled populations (♀: Wilk's  $\lambda = 0.035$ ,  $F_{80, 192, 50} = 1.674$ ,  $p = .002$ ; ♂: Wilk's  $\lambda = 0.047$ ,  $F_{54, 142, 27} = 2.16$ ,  $p < .001$ ) but not between the two taxa (♀: Wilk's  $\lambda = 0.940$ ,  $F_{8, 38} = 0.305$ ,  $p = .959$ ; ♂: Wilk's  $\lambda = 0.745$ ,  $F_{6, 35} = 2.00$ ,  $p = .092$ ). One-way ANOVAs revealed that multivariate differentiation among populations was driven by shape variation of forewing (PC2:  $F_{10, 40} = 4.86$ ;  $p < .001$ ) and ovipositor valves (PC1:  $F_{10, 42} = 2.25$ ;  $p = .03$ ) in females and head (PC2:  $F_{10, 37} = 2.22$ ;  $p = .039$ ), pronotum (PC1:  $F_{10, 40} = 3.08$ ;  $p = .005$ ; PC2:  $F_{10, 40} = 2.47$ ;  $p = .021$ ) and forewing (PC1:  $F_{10, 36} = 2.75$ ;  $p = .013$ ; PC2:  $F_{10, 36} = 2.33$ ;  $p = .031$ ) in males.



**FIGURE 2** Genetic assignment of individuals based on the results of (a) STRUCTURE (from  $K = 2$  to  $K = 6$ ) and (b) CONSTRUCT (from  $K = 2$  to  $K = 4$ ). Individuals are partitioned into  $K$  coloured segments representing the probability of belonging to the cluster with that colour. Thin vertical black lines separate individuals from different populations. Population codes as in Table 1

### 3.2 | Genomic data

Illumina sequencing returned an average of  $2.19 \times 10^6$  reads per sample. After quality control, an average of  $2.11 \times 10^6$  reads were retained per sample (Figure S1). The total number of unlinked SNPs retained in the final data set obtained with PYRAD for all individuals and  $\text{minCov} = 25\%$  was 64,365 SNPs (63,860 SNPs when the out-group *O. femoralis* was included).

### 3.3 | Population genetic structure

STRUCTURE analyses yielded an “optimal” clustering value for  $K = 2$  according to the  $\Delta K$  criterion, but log probabilities of the data ( $\text{LnPr}(X|K)$ ) steadily increased up to  $K = 6$  (Figure S4a). The inferred genetic groups were consistent with the geographic distribution of the populations but not with their taxonomic assignment (Figure 2 and Figure S5). STRUCTURE analyses showed a hierarchical genetic structure, with a first split between eastern and western populations for  $K = 2$  and subsequent subdivision of these two main population groups into other genetic clusters at smaller geographical scales from  $K = 3$  to  $K = 6$ . Most individuals and populations were assigned to a single genetic cluster, with admixture limited to populations in putative contact zones between genetic clusters.

Maximum-likelihood scores from FASTSTRUCTURE analyses peaked for  $K = 3$  and, according to the  $\Delta K$  criterion the “optimal” clustering solution was  $K = 2$  (Figure S4b). Model complexity that maximizes marginal likelihood was equal to three in 22 replicates, equal to four in two replicates and equal to five in one replicate, and the number of model components used to explain structure in the data was equal to four in 18 replicates and equal to seven in five replicates. FASTSTRUCTURE results mostly mirrored those obtained with STRUCTURE, but they also presented some differences (Figure S6): (a) For  $K = 3$ , STRUCTURE included in the same genetic cluster the populations BOR-MON and BOI-LLE. However, these two pairs of populations were assigned to different genetic clusters by FASTSTRUCTURE, which makes more geographical sense; (b) FASTSTRUCTURE analyses for  $K > 5$  did not split populations CRE-ERR-QUE from TUR-SET-BAT as done by STRUCTURE for  $K = 6$ ; (c) With the exception of TOS, no other population showed any signature of genetic admixture in FASTSTRUCTURE analyses (Figure S6). Assignment values to additional genetic clusters in FASTSTRUCTURE for analyses with  $K > 5$  were extremely low in all cases ( $q < 0.0001$ ) and, thus, their respective bar plots were virtually identical to those obtained for  $K = 5$  (for a similar result, see Baiz, Tucker, & Cortés-Ortiz, 2019).

The spatial model of CONSTRUCT showed better model fit at every value of  $K$  than the non-spatial model (Figure S7a). Spatial models had strong statistical support up to  $K = 3-5$  (Figure S7a), but layers beyond  $K = 3$  contributed relatively little to total covariance (Figure S7b). CONSTRUCT analyses showed strong genetic admixture for all  $K$  values (e.g., Whelan et al., 2019) and only genetic assignments for  $K = 3-4$  were partially compatible with the presence of the two main genetic clusters revealed by STRUCTURE and FASTSTRUCTURE (Figure 2 and Figure S8).

In agreement with previous analyses, PCA grouped populations according to their geographic location and main genetic clusters rather than by described species (Figure 3). PC1 separated the westernmost populations BOR-MON from the rest of the populations, whereas PC2 separated the rest of western populations from eastern populations and placed the admixed population of TOS at an intermediate position.

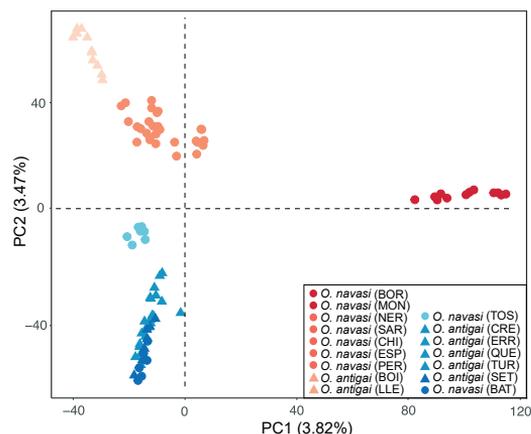
### 3.4 | Phylogenetic inference

Population trees reconstructed in BPP, SNAPP, and SVDQUARTETS yielded similar topologies that only differed in the inferred relationships among some nearby populations belonging to the same genetic clusters (Figures 1a and 4). For the west genetic group the three methods inferred different relationships for the populations NER-SAR-CHI-ESP-PER (Figure 4), which were grouped within the same genetic cluster at the lower hierarchical level by STRUCTURE and FASTSTRUCTURE analyses (Figure 2) and clumped together in PCA (Figure 3). For the east genetic group, SNAPP and BPP analyses produced the same topology and placed TUR as a sister clade of CRE-ERR-QUE, whereas SVDQUARTETS placed TUR as external group with respect to the five easternmost populations CRE-ERR-QUE-SET-BAT. As expected, the

relationship among these populations also showed considerable uncertainty in terms of low node support (i.e. posterior probabilities for SNAPP and bootstrapping values for SVDQUARTETS; Figure 4). This is also evidenced by the SNAPP analyses, which showed considerable fuzziness in some parts of the tree probably due to gene flow among nearby populations belonging to the same genetic clusters (Figure 1a). In general, the relationships among populations inferred by all phylogenetic analyses are in good agreement with the hierarchical genetic structure yielded by STRUCTURE, FASTSTRUCTURE, and PCA and indicate no support (i.e., polyphyly) for the separation of the named taxa (Figure 1).

### 3.5 | Species delimitation

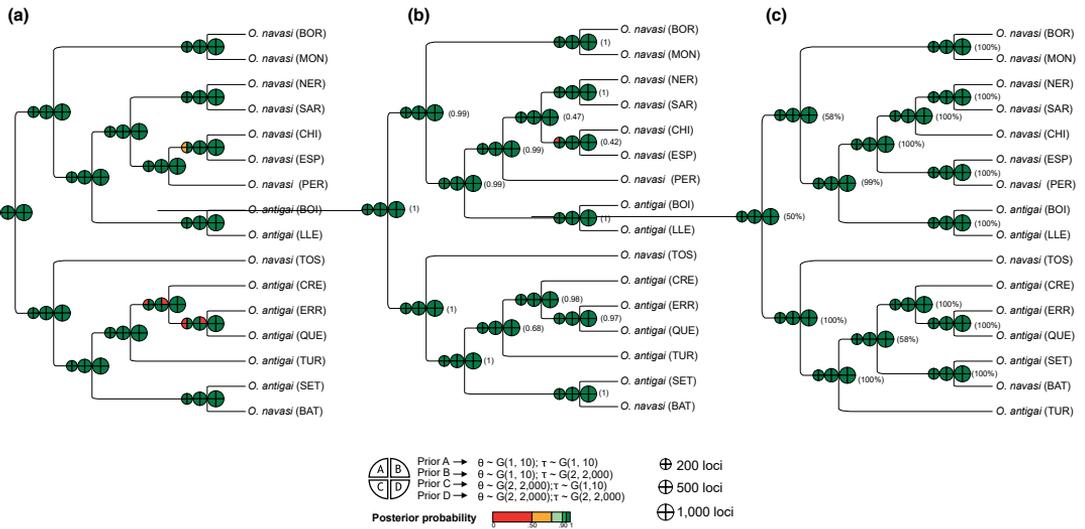
Species delimitation analyses with the  $\text{bFD}^*$  method strongly supported a six species model ( $H_6$ ), which considered populations assigned to each genetic cluster identified by Bayesian clustering analyses in STRUCTURE as a distinct species (Table 2). The second most-supported scenario ( $H_4$ ) was the one considering one species, but received much lower support (Table 2). BPP analyses fully supported the presence of 13 species across all tested topologies (BPP, SNAPP and SVDQUARTETS), prior combinations for gamma distribution, and number of loci (Figure 4). All sampled populations were consistently supported as distinct species when using the subset of 1,000 loci and only some nearby populations belonging to the same genetic clusters (CHI-ESP and CRE-ERR-QUE) were not supported as distinct taxa under a few prior combinations and specific guide trees for the subsets of 200 and 500 loci (Figure 4). IBPP analyses supported all populations as different species using either female or male phenotypic data (Figure 5). IBPP analyses based only on morphological data showed inconsistent results depending on sex and the tested demographic scenario, but tended to support the split into different species for the 3-4 most external nodes (Figure 5). A validation test randomizing phenotypic data across individuals yielded almost an identical result (Figure S9), indicating that the influence of phenotypic traits in IBPP analyses is probably overridden by the high amount of genomic data. Remarkably, IBPP analyses only based on randomized morphological data sets (i.e., without genomic data) also tended to support the split into different species for some external nodes, although the results strongly varied between sexes and across tested topologies and demographic priors (Figure S9).



**FIGURE 3** Principal component analysis (PCAs) of genetic variation for *Omocestus antigai* (triangles) and *Omocestus navasi* (dots). Colours of triangles and dots indicate the main genetic cluster at which individuals were assigned according to STRUCTURE analyses for  $K = 6$ . Population codes as in Table 1

### 3.6 | Landscape genetic and phenotypic analyses

MMRR analyses indicated that genetic differentiation ( $F_{ST}$ ) was significantly correlated with both geographic distances ( $p = .001$ ) and resistance distances defined by topographic complexity ( $p = .004$ ; Table 3). However, environmental dissimilarity did not show a significant relationship with genetic differentiation and this variable was excluded from the final model ( $p = .374$ ; Table 3). MMRR analyses for phenotypic differentiation ( $P_{ST}$ ) showed that no trait was significantly correlated in any sex with geographical distances, resistance distances defined by topographic complexity,



**FIGURE 4** Results of species delimitation analyses in BPP. Analyses were performed using three alternative topologies (a: BPP; b: SNAPP; c: SVDQUARTETS), three different subsets of loci (200, 500, and 1,000 loci; two individuals/population), and four gamma prior combinations (gamma,  $\alpha$ ,  $\beta$ ) for ancestral population size ( $\theta$ ) and root age ( $\tau$ ). Colored boxes at each node represent the mean posterior probability (PP) for different combinations of demographic priors (legend at bottom). Node support in terms of posterior probabilities (for SNAPP tree) and bootstrapping values (for SVDQUARTETS tree) is indicated in parentheses for each node. Population codes as in Table 1

environmental dissimilarity or genetic differentiation after applying Benjamini-Hochberg false discovery rate corrections for multiple testing (all  $q$ -values > 0.05; Table 4).

## 4 | DISCUSSION

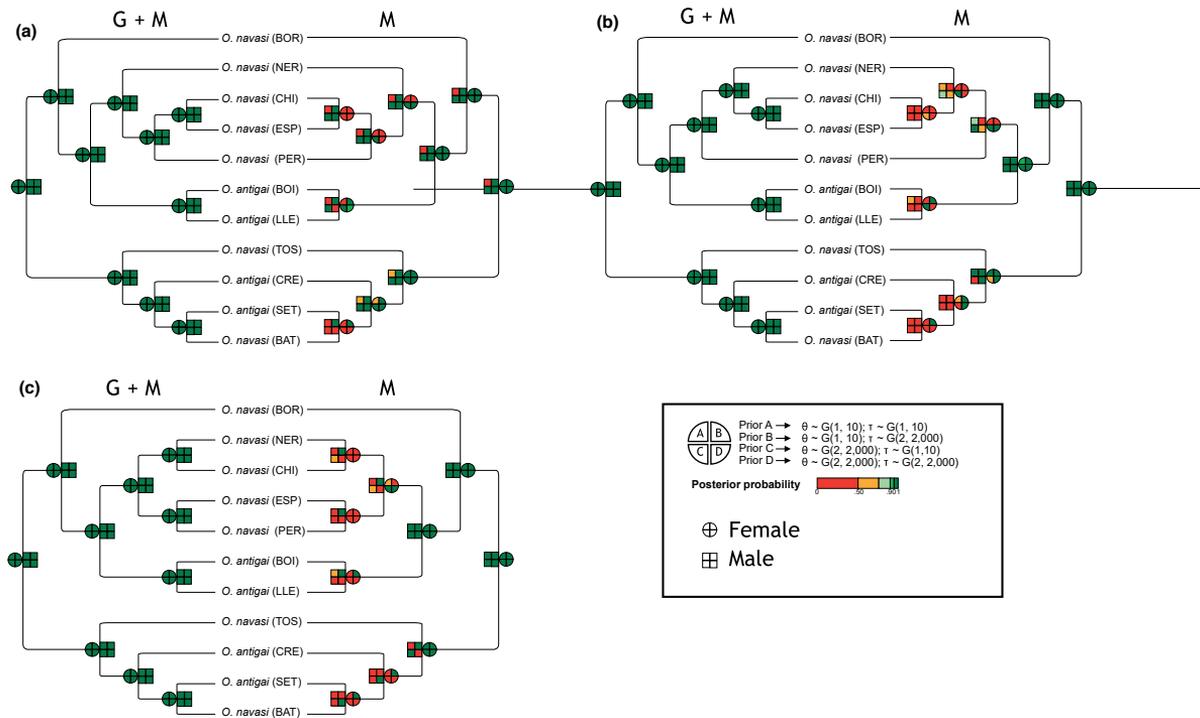
Collectively, our analyses rejected the current hypothesis of two species and indicated that populations assigned to *O. antigai* and *O. navasi* show a distribution of genetic variation that (a) does not match with their respective taxonomic designation and (b) is incompatible with ecological/environmental speciation. Our results show the presence of two main genetic groups corresponding to an east–west split analogous to that found in many other Pyrenean taxa (Wallis, Waters, Upton, & Craw, 2016) and a marked genetic differentiation at local spatial scales reflecting limited population connectivity across the abrupt landscapes characterizing the study region (e.g., Noguerales, Cordero, & Ortego, 2016).

### 4.1 | Integrative species delimitation

As in many other species complexes (e.g., Leaché et al., 2018) the taxonomy of *O. antigai* and *O. navasi* has changed through time in conjunction with the discussion about the validity of the different criteria considered to establish the boundaries between the two putative taxa (Clemente et al., 1999; Ragge & Reynolds, 1998). After the two species were described by Bolívar in the transition between the XIX and XX centuries (Bolívar, 1897, 1908) and

Clemente, García, and Presa (1990) confirmed their taxonomic distinction, Reynolds (1986) and Ragge and Reynolds (1998) attributed their subtle morphological diversity to the effect of population geographic isolation, denied differences in acoustic communication (i.e., courtship songs), and recommended the synonymy of *O. navasi* to *O. antigai*. However, new acoustic and biometrical analyses by Clemente et al. (1999) resurrected the species *O. navasi* and the two taxa are now considered valid species (Cigliano et al., 2019) and treated as such in conservation assessment programs (Braud, Hochkirch, Fartmann, et al., 2016; Braud, Hochkirch, Fartmann, et al., 2016; Hochkirch et al., 2016). Up to now, no molecular-based analysis had been performed to test the monophyly of the two taxa and determine whether their specific habitat associations and slight phenotypic differences have a genetic basis in line with a scenario of ecological speciation. Using genomic and phenotypic data and a comprehensive suite of phylogenomic and Bayesian clustering analyses, our results indicate no support for the current taxonomy. Phylogenomic and genetic structure analyses indicate that populations split in two main genetic groups (east/west) that do not match current taxonomic designations and, accordingly, populations sampled in the two distinct habitats (alpine vs. Mediterranean/montane) supposedly occupied by the two putative species are phylogenetically interspersed (Figure 1).

Despite phylogenomic reconstructions not recovering the monophyly of the two nominated taxa, molecular-based species delimitation analyses using BPP and BFD<sup>+</sup> identified the presence of several plausible species. Independently of the number of loci used, BPP analyses fully supported the presence of 13 species across all tested topologies



**FIGURE 5** Results of species delimitation analyses in IBPP. Analyses were performed separately for each sex both combining genomic (500 loci and five individuals/population) and phenotypic data (trees on the left) and only using phenotypic data (trees on the right). All analyses were performed using three alternative topologies (a: BPP; b: SNAPP; c: SVDQUARTETS) and four gamma prior combinations (gamma,  $\alpha$ ,  $\beta$ ) for ancestral population size ( $\theta$ ) and root age ( $\tau$ ). Coloured boxes at each node represent the mean posterior probability (PP) for different combinations of demographic priors (legend at bottom right). Population codes as in Table 1

and prior combinations for gamma distribution. In most cases, each of the 17 sampled populations was supported as a distinct species and only some nearby populations separated by less than 20 km and belonging to the same genetic clusters (CHI-ESP and CRE-ERR-QUE) were not supported as distinct taxa under a few prior combinations and specific topologies. It is also remarkable that many nearby populations (BOR-MON, NER-SAR, BOI-LLE and SET-BAT) assigned to the same genetic groups by PCA and Bayesian clustering analyses (STRUCTURE, FASTSTRUCTURE and CONSTRUCT) were consistently delimited as

**TABLE 3** Multiple matrix regression with randomization (MMRR) for pairwise population genetic differentiation ( $F_{ST}$ ) in relation with geographical, resistance and environmental distances

Variable	Coefficient	t	p-Value
Retained terms			
Intercept	-0.108	-3.079	.991
Geographical distance	0.255	6.841	.001
Resistance distance (slope)	0.007	5.278	.004
Rejected term			
Environmental distance	—	-1.061	.374

distinct species (Figure 4). Accordingly, the species delimitation model most supported by BFD\* analyses was the one considering the highest number of species, one per genetic cluster inferred by STRUCTURE (see also figure 2b in Leaché et al., 2018). Recent empirical and theoretical studies have shown the limitations of Bayesian species delimitation approaches based on the multi-species coalescent (MSC) model (Rannala & Yang, 2013; Yang & Rannala, 2010), pointing out that such methods are not able to statistically distinguish genetic structure due to population isolation from true species boundaries and tend to over-split genetically differentiated populations rather than capturing species divergence (Huang, 2018; Leaché et al., 2018; Leaché, Zhu, Rannala, & Yang, 2019; Sukumaran & Knowles, 2017). The problem of species overestimation can be particularly exacerbated when using vast genome-wide data due to the high power of a large number of markers (hundreds to tens of thousands) to detect fine-grain population structure (Hime et al., 2016; Noguerales et al., 2018; Sukumaran & Knowles, 2017). In this line, recent empirical studies have found that geographically continuous intraspecific populations presenting genotypic differences simply resulted from isolation by distance can be delimited as distinct species when using hundreds of loci (Huang, 2018; Pyron, Hsieh, Lemmon, Lemmon, & Hendry, 2016).

	Females		Males	
	<i>t</i>	<i>q</i>	<i>t</i>	<i>q</i>
<b>Head—PC1</b>				
Geographical distance	0.448	0.919	-0.048	0.980
Resistance distance (slope)	0.989	0.791	-1.106	0.853
Environmental distance	-1.699	0.791	-2.993	0.288
$F_{ST}$	0.125	0.950	-0.052	0.980
<b>Head—PC2</b>				
Geographical distance	1.882	0.686	1.223	0.864
Resistance distance (slope)	-1.047	0.791	-0.429	0.973
Environmental distance	-1.423	0.791	0.135	0.980
$F_{ST}$	2.781	0.400	1.735	0.840
<b>Pronotum—PC1</b>				
Geographical distance	-1.264	0.791	-0.858	0.973
Resistance distance (slope)	-0.416	0.919	0.376	0.973
Environmental distance	0.963	0.791	1.713	0.853
$F_{ST}$	-0.960	0.791	0.587	0.973
<b>Pronotum—PC2</b>				
Geographical distance	0.868	0.919	0.715	0.973
Resistance distance (slope)	-0.267	0.919	0.318	0.973
Environmental distance	-1.240	0.791	-0.054	0.980
$F_{ST}$	-0.735	0.919	1.152	0.853
<b>Forewing—PC1</b>				
Geographical distance	2.775	0.405	0.695	0.973
Resistance distance (slope)	-0.190	0.919	-0.386	0.973
Environmental distance	-2.197	0.686	-2.455	0.288
$F_{ST}$	2.695	0.520	-0.031	0.980
<b>Forewing—PC2</b>				
Geographical distance	-0.222	0.919	-1.471	0.840
Resistance distance (slope)	0.388	0.919	2.185	0.496
Environmental distance	3.159	0.400	0.782	0.973
$F_{ST}$	1.110	0.791	2.045	0.720
<b>Ovopositor valve—PC1</b>				
Geographical distance	-0.282	0.919	-	-
Resistance distance (slope)	0.210	0.919	-	-
Environmental distance	1.614	0.686	-	-
$F_{ST}$	-0.408	0.919	-	-
<b>Ovopositor valve—PC2</b>				
Geographical distance	-0.257	0.919	-	-
Resistance distance (slope)	0.637	0.919	-	-
Environmental distance	1.401	0.791	-	-
$F_{ST}$	0.617	0.919	-	-

Note: No independent variable was significant and retained into final models. Table shows *q*-values after applying Benjamini-Hochberg false discovery rate corrections of *p*-values to adjust for multiple statistical tests.

**TABLE 4** Multiple matrix regressions with randomization (MMRR) for pairwise population phenotypic differentiation ( $P_{ST}$ ) in relation with genetic differentiation ( $F_{ST}$ ) and geographical, resistance and environmental distances

Recent studies have suggested that species delimitation approaches integrating genomic data with other sources of information (ecological, morphological, ethological, etc) could help to mitigate

the problem of confounding population structure with species limits (Pyron et al., 2016; Sukumaran & Knowles, 2017). Although most phenotypic traits analyzed only presented very subtle differences

among some pairs of populations and did not differ between currently recognized taxa (see Figures S2 and S3), our IBPP analyses supported most populations as distinct species and a validation test randomizing phenotypic data across individuals yielded an almost identical result. Thus, incorporating morphological traits has no or very little influence on the outcome of IBPP analyses and the impact of phenotypic traits seems to be overridden by the high amount of genetic data. Our analyses suggest that the species delimitation approach implemented in IBPP does not really help to reduce the number of delineated species, indicating that integrating phenotypic and genetic data into this framework cannot prevent species overestimation (Noguerales et al., 2018; Sukumaran & Knowles, 2017). In sum, the strong genetic structure in our study system and the high number of loci employed are likely to have extraordinarily inflated the number of species identified by BPP and BFD\* and integrative analyses implemented in IBPP does not seem to help to ameliorate this problem.

Beyond the taxonomic inflation issue linked with the fact that the multispecies coalescent model frequently diagnoses genetic structure rather than species, different intersecting lines of evidence suggest that the multiple delineated entities do not meet the requirements to be considered distinct taxa according to alternative contemporary species concepts (see table 1 in de Queiroz, 2007). These lines of evidence include: (a) The presence of largely overlapping phenotypes across populations, taxa, and genetic clusters for all studied traits (Figure S2 and S3; phenetic species concept); (b) No evidence for environmental-driven divergence and the presence of phylogenetically interspersed high and low elevation populations (ecological + phylogenetic species concepts) and (c) Genetic clusters are always allopatric and admixture at contact zones happens even between the two most diverging eastern and western genetic groups (genotypic cluster species concept), which is indicative of lack of intrinsic reproductive isolation (biological species concept; de Queiroz, 2007). Although the strong signal for isolation-by-distance might be potentially compatible with a rapid stepping-stone speciation process (e.g., VanderWerf, Young, Yeung, & Carlon, 2010), such pattern most plausibly reflects population connectivity (Hutchison & Templeton, 1999; Slatkin, 1993). Thus, in our opinion, the studied populations should be treated—at least by now—as a single species taxon with strong genetic structure, which is not incompatible with the possibility of designating the main genetic clusters (e.g., east–west genetic groups) as intraspecific evolutionary significant units (Moritz, 2002) that should be considered in future conservation and management strategies (Braud, Hochkirch, Fartmann, et al., 2016; Braud, Hochkirch, Fartmann, et al., 2016).

#### 4.2 | Factors structuring genomic and phenotypic variation

Phylogenomic reconstructions and spatial patterns of genetic structure inferred from classic Bayesian clustering analyses supported that genetic variation within the complex is hierarchically structured in congruence with the geographical distribution of the

sampled populations. However, the strong signal of isolation-by-distance ( $r = .71, p < .0001$ ; Table 3) probably explains why the spatial analyses in CONSTRUCT blur most genetic structure and do not even clearly recover the west/east split revealed by phylogenetic reconstructions and all other analyses. Our study system presents a high degree of genetic differentiation (mean  $F_{ST} = 0.35$ ) and extraordinarily high estimates of genetic drift after divergence for all genetic clusters inferred by STRUCTURE analyses ( $F$ -value  $> 0.5$ ; Pritchard et al., 2000). Also, there is a remarkable congruence between the hierarchical genetic structure inferred by nonspatial clustering analyses (STRUCTURE, FASTSTRUCTURE) and both phylogenetic reconstructions (SNAPP, SVDQUARTETS, and BPP) and PCA, the latter being a method free of the assumptions made by classic clustering analyses (Jombart, Devillard, & Balloux, 2010). All these lines of evidence suggest that results from CONSTRUCT make little biological sense in our specific study system and indicate that the strong genetic structure revealed by all other analyses is genuine and resulted from historical processes (i.e., strong isolation) rather than being an artefactual consequence of gradual genetic differentiation (i.e., migration and genetic drift equilibrium) merely driven by geographical distance (Bradburd et al., 2018).

In agreement with STRUCTURE and FASTSTRUCTURE results for  $K = 2$ , all phylogenetic analyses showed a basal split in two clades corresponding to populations located east and west of an imaginary line located around the Segre river ( $\sim 1^\circ\text{E}$ ). Similar east–west genetic splits have been reported for many other plant and animal species from the Pyrenean region (e.g., Alvarez-Presas, Mateos, Vila-Farre, Sluys, & Riutort, 2012; Bidegaray-Batista et al., 2016; Charrier, Dupont, Pornon, & Escaravage, 2014; Milá, Carranza, Guillaume, & Clobert, 2010; Valbuena-Ureña et al., 2018). This east–west split might be well explained by detailed geological reconstructions documenting heavily glaciated areas in the central part of the range and the likely presence of ice-free refugia at its longitudinal extremes (Ehlers, Gibbard, & Hugues, 2011) that might have allowed montane/alpine species to survive glaciations and recolonize high elevations during interglacial periods (Charrier et al., 2014; Valbuena-Ureña et al., 2018). Overall, these results are consistent with the transverse breaks proposed by Wallis et al. (2016) for numerous alpine taxa and support the idea that the fragmentation of ancestral distributions during glacial periods and the presence of Pleistocene refugia along mountain ranges has played a key role on the diversification not only of lowland species but also of montane and alpine organisms.

Genomic data also revealed strong genetic differentiation and a deep genetic structure at fine spatial scales, with up to three genetic clusters comprised within each of the two main west and east genetic groups (Figure 2). The fact that almost all sampled populations have been assigned to a unique genetic cluster, with only a few cases of genetic admixture between nearby populations, suggest a strong effect of geographic isolation (Wang, 2013). Only populations TOS and PER (and into a lesser extent SAR), located in the contact zone between the west and east genetic groups presented signatures of genetic admixture, evidencing ongoing gene flow between them (Figures 1 and 2). It is noteworthy that other populations located at a

similar longitude but at higher elevations showed no evidence of genetic admixture (e.g., BOI, LLE), which suggests a higher isolation of alpine populations and increased gene flow through the less abrupt landscapes characterizing the Pyrenean foothills. Accordingly, landscape genetic analyses indicated that genetic differentiation was explained by both the geographical distance among populations and resistance distances defined by topographic roughness (Table 3). These results might reflect the low dispersal capability of the studied taxon, males being brachypterous and females micropterous, and are comparable to those obtained by previous studies showing the impact of steep slopes and complex landscapes on structuring genetic variation in montane/alpine grasshoppers (Noguerales, Cordero, et al., 2016). Genetic structure analyses showed that the split of the different populations at local/regional scales followed a longitudinal cline rather than a segregation of alpine and Mediterranean-montane populations (e.g., SET and BAT), indicating no support for either ecologically driven divergence or the taxonomic separation between the supposedly Mediterranean *O. navasi* and the alpine *O. antigai*. This was corroborated by our landscape genetic analyses, which revealed no effect of environmental dissimilarity on structuring genetic variation in the complex (i.e., isolation-by-environment; Shafer & Wolf, 2013).

Our geometric morphometric analyses showed that the subtle phenotypic differences found among populations were not explained by genetic differentiation, geographical distances or environmental dissimilarity. Overall, this suggests that the weak phenotypic differences found among populations are not a consequence of genetic drift or environmental-driven selection (Keller, Alexander, Holderegger, & Edwards, 2013; Leinonen, Cano, Makinen, & Merila, 2006; Leinonen, O'Hara, Cano, & Merila, 2008) and might be explained by ecological and evolutionary aspects not considered in this study such as sexual selection, predation risk, microhabitat structure or adaptations to different feeding resources (e.g., Ingley, Billman, Belk, & Johnson, 2014; Laiolo, Illera, & Obeso, 2013; Noguerales, García-Navas, Cordero, & Ortego, 2016).

## 5 | CONCLUSIONS

This study exemplifies the problems associated with species validation tests involving recently diverged allopatric taxa and highlights the importance of integrating different sources of information to delimit species that have been described solely by morphology or ecological distinctiveness (e.g., Jones & Weisrock, 2018). Phylogenetic inferences, Bayesian species delimitation analyses, and phenotypic and ecological data did not support the current taxonomic status of the complex and indicate that *O. navasi* and *O. antigai* must be synonymized into a unique taxon: *O. antigai* (Bolívar, 1897; for a list of synonyms, see Table S1). Our study also illustrates the implications that incorrect taxonomic designations can have for species conservation and management. The recent assessment of the conservation status of European grasshoppers has assigned the two taxa to

different IUCN Red List categories (Hochkirch et al., 2016), considering *O. navasi* as a “endangered” species (Braud, Hochkirch, Fartmann, et al., 2016) and *O. antigai* as “vulnerable” (Braud, Hochkirch, Fartmann, et al., 2016). In the view of our results, the conservation status of the complex requires total reconsideration in future IUCN Red List assessments. Given that the range and population sizes of *O. antigai* are larger and the ecological and habitat requirements much wider than previously thought, the conservation status of the taxon should be probably downlisted to “near threatened” (Braud, Hochkirch, Fartmann, et al., 2016; Braud, Hochkirch, Fartmann, et al., 2016). However, the presence of two marked genetic groups and many genetically structured populations should be also considered in future conservation plans as potential evolutionary significant units that probably deserve to be protected and managed independently (Moritz, 2002). Overall, our analyses point to the presence of a single species characterized by a strong genetic structure, little phenotypic variation, and a wide environmental niche. Future studies performing demographic reconstructions and spatiotemporally explicit landscape genetic analyses that integrate available information on past glacier extent (Ehlers et al., 2011) and inferred distributional shifts of the species linked to Pleistocene glacial cycles could greatly help to further understand the historical processes that have shaped the spatial distribution of genetic variation within and among the current populations (e.g., Massatti & Knowles, 2016).

## ACKNOWLEDGEMENTS

We are grateful to Amparo Hidalgo-Galiana for her support during the preparation of genomic libraries, Víctor Noguerales and Pedro J. Cordero for their help during fieldwork, Francisco Rodríguez-Sánchez for providing valuable advice in spatial data analyses and Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing. We also thank three anonymous referees for helpful and constructive comments on an earlier version of this article. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) and Laboratorio de Sistemas de Información Geográfica y Teledetección (LAST-EBD) from Estación Biológica de Doñana. We also thank to Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. This work was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (CGL2014-54671-P and CGL2017-83433-P). VT was supported by an FPI predoctoral fellowship (BES-2015-73159) from Ministerio de Economía y Competitividad. During this work, AP and JO were supported by a Severo Ochoa (SEV-2012-0262) and a Ramón y Cajal (RYC-2013-12501) research fellowship, respectively.

## AUTHOR CONTRIBUTIONS

V.T., A.P., and J.O. conceived and designed the study and analyses. All authors collected the samples. V.T. performed the laboratory

work and analysed the data guided by A.P., and J.O. V.T. wrote the manuscript with help of J.O., and inputs from A.P.

## DATA AVAILABILITY STATEMENT

Raw Illumina reads have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA543714. Morphometric data and input files for all analyses are available for download from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.249v096>).

## ORCID

Vanina Tonzo  <https://orcid.org/0000-0002-5062-1070>

## REFERENCES

- Adams, D. C., & Otarola-Castillo, E. (2013). *Geomorph*: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4), 393–399. <https://doi.org/10.1111/2041-210x.12035>
- Agapow, P. M., Bininda-Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C., & Purvis, A. (2004). The impact of species concept on biodiversity studies. *Quarterly Review of Biology*, 79(2), 161–179. <https://doi.org/10.1086/383542>
- Alvarez-Presas, M., Mateos, E., Vila-Farre, M., Sluys, R., & Riutort, M. (2012). Evidence for the persistence of the land planarian species *Microplana terrestris* (Muller, 1774) (Platyhelminthes, Tricladida) in microrefugia during the Last Glacial Maximum in the northern section of the Iberian Peninsula. *Molecular Phylogenetics and Evolution*, 64(3), 491–499. <https://doi.org/10.1016/j.ympev.2012.05.001>
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., ... Saunders, N. C. (1987). Intraspecific phylogeography – The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18, 489–522. <https://doi.org/10.1146/annurev.ecolsys.18.1.489>
- Baiz, M. D., Tucker, P. K., & Cortés-Ortiz, L. (2019). Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular Ecology*, 28(5), 1056–1069. <https://doi.org/10.1111/mec.14966>
- Bidegaray-Batista, L., Sanchez-Gracia, A., Santulli, G., Maiorano, L., Guisan, A., Vogler, A. P., & Arnedo, M. A. (2016). Imprints of multiple glacial refugia in the Pyrenees revealed by phylogeography and palaeodistribution modelling of an endemic spider. *Molecular Ecology*, 25(9), 2046–2064. <https://doi.org/10.1111/mec.13585>
- Blair, C., Bryson, R. W., Linkem, C. W., Lazcano, D., Klicka, J., & McCormack, J. E. (2019). Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: Crotalus). *Molecular Ecology Resources*, 19(2), 349–365. <https://doi.org/10.1111/1755-0998.12970>
- Bolívar, I. (1897). Catálogo sinóptico de los ortópteros de la fauna ibérica. *Annaes de Ciencias Naturaes*, 4(4), 203–232.
- Bolívar, I. (1908). Algunos ortópteros nuevos de España, Marruecos y Canarias. *Boletín de La Real Sociedad Española de Historia Natural*, 8, 317–334.
- Bookstein, F. L. (1992). *Morphometric tools for landmark data*. Cambridge, UK: Cambridge University Press.
- Bouckaert, R. R. (2010). DENSITREE: Making sense of sets of phylogenetic trees. *Bioinformatics*, 26(10), 1372–1373. <https://doi.org/10.1093/bioinformatics/btq110>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., ... Drummond, A. J. (2014). BEAST2: A Software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1), 33–52. <https://doi.org/10.1534/genetics.118.301333>
- Braud, Y., Hochkirch, A., Fartmann, T., Presa, J. J., Monnerat, C., Roesti, C., & Dusoulier, F. (2016). *Omocestus antgai*: The IUCN Red List of Threatened Species 2016 (Publication no. e.T16084614A70647955). doi: 10.2305/IUCN.UK.2016-3.RLTS.T16084614A70647955.en. Retrieved from <https://www.iucnredlist.org/species/16084614/70647955>
- Braud, Y., Hochkirch, A., Presa, J. J., Roesti, C., Rutschmann, F., Monnerat, C., & Dusoulier, F. (2016). *Omocestus navasi*: The IUCN Red List of Threatened Species 2016 (Publication no. e.T69641425A69641429). doi: 10.2305/IUCN.UK.2016-3.RLTS.T69641425A69641429.en. Retrieved from <https://www.iucnredlist.org/species/69641425/69641429>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), 1917–1932. <https://doi.org/10.1093/molbev/mss086>
- Burbrink, F. T., Yao, H., Ingrasci, M., Bryson, R. W., Guiher, T. J., & Ruane, S. (2011). Speciation at the Mogollon Rim in the Arizona Mountain Kingsnake (*Lampropeltis pyromelana*). *Molecular Phylogenetics and Evolution*, 60(3), 445–454. <https://doi.org/10.1016/j.ympev.2011.05.009>
- Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, 22(17), 4369–4383. <https://doi.org/10.1111/mec.12413>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chan, K. O., Alexander, A. M., Grismer, L. L., Su, Y. C., Grismer, J. L., Quah, E. S. H., & Brown, R. M. (2017). Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. *Molecular Ecology*, 26(20), 5435–5450. <https://doi.org/10.1111/mec.14296>
- Charrier, O., Dupont, P., Pornon, A., & Escaravage, N. (2014). Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the Pyrenees. *PLoS ONE*, 9(3), e92976. <https://doi.org/10.1371/journal.pone.0092976>
- Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317–3324. <https://doi.org/10.1093/bioinformatics/btu530>
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2019). *Orthoptera Species File*. Version 5.0/5.0. Retrieved from <http://orthoptera.speciesfile.org/>
- Clemente, M. E., García, M. D., Arnaldos, M. I., Romera, E., & Presa, J. J. (1999). Corroboration of the specific status of *Omocestus antgai* (Bolívar, 1897) and *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae). [Confirmación de las posiciones taxonómicas específicas de *Omocestus antgai* (Bolívar, 1897) y *O. navasi* Bolívar, 1908 (Orthoptera, Acrididae)]. *Boletín De La Real Sociedad Española De Historia Natural Sección Biológica*, 95(3–4), 27–50.
- Clemente, M. E., García, M. D., & Presa, J. J. (1990). Nuevos datos sobre los Acridoidea (Insecta: Orthoptera) del Pirineo y prepirineo Catalano-Aragonés. *Butlletí de la Institució Catalana d'història Natural*, 58, 37–44.

- da Silva, S. B., & da Silva, A. (2018). Pstat: An R Package to assess population differentiation in phenotypic traits. *The R Journal*, 10(1), 447–454. <https://doi.org/10.32614/RJ-2018-010>
- Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85(3), 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- de Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56(6), 879–886. <https://doi.org/10.1080/10635150701701083>
- Dobzhansky, T. (1937). Genetic nature of species differences. *American Naturalist*, 71, 404–420. <https://doi.org/10.1086/280726>
- Dobzhansky, T. (1970). *Genetics of the evolutionary process*. New York, NY: Columbia University Press.
- Domingos, F., Colli, G. R., Lemmon, A., Lemmon, E. M., & Beheregaray, L. B. (2017). In the shadows: Phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: Tropidurus). *Molecular Phylogenetics and Evolution*, 107, 455–465. <https://doi.org/10.1016/j.ympev.2016.12.009>
- Dowle, E. J., Morgan-Richards, M., & Treweek, S. A. (2014). Morphological differentiation despite gene flow in an endangered grasshopper. *BMC Evolutionary Biology*, 14, 216. <https://doi.org/10.1186/s12862-014-0216-x>
- Eaton, D. A. R. (2014). PYRAD: Assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: Can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences*, 281, 20132765. <https://doi.org/10.1098/rspb.2013.2765>
- Ehlers, J., Gibbard, P. L., & Huges, P. D. (2011). *Quaternary glaciations - Extent and chronology: A closer look* (1st ed.). Amsterdam, the Netherlands: Elsevier.
- Eklom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1–15. <https://doi.org/10.1038/hdy.2010.152>
- Excoffier, L., Laval, G., & Schneider, S. (2005). ARLEQUIN ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47–50.
- Friedline, C. J., Faske, T. M., Lind, B. M., Hobson, E. M., Parry, D., Dyer, R. J., ... Eckert, A. J. (2019). Evolutionary genomics of gypsy moth populations sampled along a latitudinal gradient. *Molecular Ecology*, 28(9), 2206–2223. <https://doi.org/10.1111/mec.15069>
- Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27(9), 480–488. <https://doi.org/10.1016/j.tree.2012.04.012>
- García-Navas, V., Noguerales, V., Cordero, P. J., & Ortego, J. (2017). Ecological drivers of body size evolution and sexual size dimorphism in short-horned grasshoppers (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, 30(8), 1592–1608. <https://doi.org/10.1111/jeb.13131>
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), 313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Hime, P. M., Hotaling, S., Grewelle, R. E., O'Neill, E. M., Voss, S. R., Shaffer, H. B., & Weisrock, D. W. (2016). The influence of locus number and information content on species delimitation: An empirical test case in an endangered Mexican salamander. *Molecular Ecology*, 25(23), 5959–5974. <https://doi.org/10.1111/mec.13883>
- Hochkirch, A., Nieto, A., García-Criado, M., Cáliz, M., Braud, Y., Buzzetti, F. M., ... Tumbirck, J. (2016). *European red list of grasshoppers, crickets and bush-crickets*. Luxembourg: Publications Office of the European Union.
- Huang, J. P. (2018). What have been and what can be delimited as species using molecular data under the multi-species coalescent model? A case study using *Hercules* beetles (Dynastes: Dynastidae). *Insect Systematics and Diversity*, 2(2), 1–10. <https://doi.org/10.1093/isd/ixy003>
- Huang, J. P., & Knowles, L. L. (2016). The species versus subspecies conundrum: Quantitative delimitation from integrating multiple data types within a single Bayesian approach in *Hercules* beetles. *Systematic Biology*, 65(4), 685–699. <https://doi.org/10.1093/sysbio/syv119>
- Hutchison, D. W., & Templeton, A. R. (1999). Correlation of 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. <https://doi.org/10.1111/j.1558-5646.1999.tb04571.x>
- Ingle, S. J., Billman, E. J., Belk, M. C., & Johnson, J. B. (2014). Morphological divergence driven by predation environment within and between species of *Brachyrhaphis* fishes. *PLoS ONE*, 9(2), e90274. <https://doi.org/10.1371/journal.pone.0090274>
- Isaac, N. J. B., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology & Evolution*, 19(9), 464–469. <https://doi.org/10.1016/j.tree.2004.06.004>
- Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, 66(5), 799–812. <https://doi.org/10.1093/sysbio/syw117>
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. <https://doi.org/10.1186/1471-2156-11-94>
- Jones, K. S., & Weisrock, D. W. (2018). Genomic data reject the hypothesis of sympatric ecological speciation in a clade of *Desmognathus* salamanders. *Evolution*, 72(11), 2378–2393. <https://doi.org/10.1111/evo.13606>
- Keller, I., Alexander, J. M., Holderegger, R., & Edwards, P. J. (2013). Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *Journal of Evolutionary Biology*, 26(12), 2527–2543. <https://doi.org/10.1111/jeb.12255>
- Knowles, L. L., & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, 56(6), 887–895. <https://doi.org/10.1080/10635150701701091>
- Laiolo, P., Illera, J. C., & Obeso, J. R. (2013). Local climate determines intra- and interspecific variation in sexual size dimorphism in mountain grasshopper communities. *Journal of Evolutionary Biology*, 26(10), 2171–2183. <https://doi.org/10.1111/jeb.12213>
- Lamichhane, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martínez-Barrio, A., ... Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375. <https://doi.org/10.1038/nature14181>
- Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences*, 277(1697), 3071–3077. <https://doi.org/10.1098/rspb.2010.0662>
- Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, 63(4), 534–542. <https://doi.org/10.1093/sysbio/syu018>
- Leaché, A. D., McElroy, M. T., & Trinh, A. (2018). A genomic evaluation of taxonomic trends through time in coast horned lizards (genus *Phrynosoma*). *Molecular Ecology*, 27(13), 2884–2895. <https://doi.org/10.1111/mec.14715>
- Leaché, A. D., Zhu, T. Q., Rannala, B., & Yang, Z. H. (2019). The spectre of too many species. *Systematic Biology*, 68(1), 168–181. <https://doi.org/10.1093/sysbio/syy051>
- Leavitt, S. D., Johnson, L. A., Goward, T., & St Clair, L. L. (2011). Species delimitation in taxonomically difficult lichen-forming fungi: An

- example from morphologically and chemically diverse *Xanthoparmelia* (Parmeliaceae) in North America. *Molecular Phylogenetics and Evolution*, 60(3), 317–332. <https://doi.org/10.1016/j.ympev.2011.05.012>
- Leinonen, T., Cano, J. M., Makinen, H., & Merila, J. (2006). Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*, 19(6), 1803–1812. <https://doi.org/10.1111/j.1420-9101.2006.01182.x>
- Leinonen, T., O'Hara, R. B., Cano, J. M., & Merila, J. (2008). Comparative studies of quantitative trait and neutral marker divergence: A meta-analysis. *Journal of Evolutionary Biology*, 21(1), 1–17. <https://doi.org/10.1111/j.1420-9101.2007.01445.x>
- Lucía-Pomares, D. (2002). Revision of the Orthoptera (Insecta) of Catalonia (Spain). [Revisión de los ortópteros (Insecta: Orthoptera) de Cataluña (España)]. *Monografías de la Sociedad Entomológica Aragonesa*, 7, 1–226.
- Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359(1444), 711–719. <https://doi.org/10.1098/rstb.2003.1454>
- Mason, N. A., & Taylor, S. A. (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, 24(12), 3009–3025. <https://doi.org/10.1111/mec.13140>
- Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: Species-specific effects of climate change in alpine sedges. *Molecular Ecology*, 25(16), 3974–3986. <https://doi.org/10.1111/mec.13735>
- Mayr, E. (1942). *Systematics and the origin of species*. New York, NY: Columbia University Press.
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8), 1551–1561. <https://doi.org/10.1111/j.0014-3820.2006.tb00500.x>
- McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), 19885–19890. <https://doi.org/10.1073/pnas.0706568104>
- Milá, B., Carranza, S., Guillaume, O., & Clobert, J. (2010). Marked genetic structuring and extreme dispersal limitation in the Pyrenean brook newt *Calotriton asper* (Amphibia: Salamandridae) revealed by genome-wide AFLP but not mtDNA. *Molecular Ecology*, 19(1), 108–120. <https://doi.org/10.1111/j.1365-294X.2009.04441.x>
- Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, 51(2), 238–254. <https://doi.org/10.1080/10635150252899752>
- Nelson, G., & Platnick, N. (1981). *Systematics and biogeography*. New York, NY: Columbia University Press.
- Niemiller, M. L., Near, T. J., & Fitzpatrick, B. M. (2012). Delimiting species using multilocus data: Diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution*, 66(3), 846–866. <https://doi.org/10.1111/j.1558-5646.2011.01480.x>
- Noguerales, V., Cordero, P. J., & Ortego, J. (2016). Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, 16, 96. <https://doi.org/10.1186/s12862-016-0663-7>
- Noguerales, V., Cordero, P. J., & Ortego, J. (2018). Integrating genomic and phenotypic data to evaluate alternative phylogenetic and species delimitation hypotheses in a recent evolutionary radiation of grasshoppers. *Molecular Ecology*, 27(5), 1229–1244. <https://doi.org/10.1111/mec.14504>
- Noguerales, V., García-Navas, V., Cordero, P. J., & Ortego, J. (2016). The role of environment and core-margin effects on range-wide phenotypic variation in a montane grasshopper. *Journal of Evolutionary Biology*, 29(11), 2129–2142. <https://doi.org/10.1111/jeb.12915>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., ... Wagner, H. (2017). *vegan: Community Ecology Package. R package version 2.4-4*. Retrieved from <http://cran.r-project.org/package=vegan>
- Olmo-Vidal, J. M. (2002). *Atlas Dels Ortòpters de Catalunya i Llibre Vermell 2002/Llagostes, Saltamartins, Grills, Someretes /Atlas de Los Ortópteros de Cataluña y Libro Rojo 2002/Atlas of the Orthoptera of Catalonia and Read Data Book 2002*. Barcelona, Spain: Generalitat de Catalunya, Departament de Medi Ambient i Habitatge.
- O'Meara, B. C. (2010). New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, 59(1), 59–73. <https://doi.org/10.1093/sysbio/syp077>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Poniatowski, D., Defaut, B., Lucía-Pomares, D., & Fartmann, T. (2009). The Orthoptera fauna of the Pyrenean region – A field guide. *Articulata Beiheft*, 14, 1–143.
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., ... Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55(4), 595–609. <https://doi.org/10.1080/10635150600852011>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Puissant, S. (2008). The presence of *Omocestus navasi* Bolívar, 1908 in France with description of a new subspecies (Caelifera, Acrididae, Gomphocerinae). [Sur la présence en France d'*Omocestus navasi* Bolívar, 1908, avec description d'une nouvelle sous-espèce (Caelifera, Acrididae, Gomphocerinae)]. *Materiaux Orthopteriques et Entomocentriques*, 13, 69–73.
- Pyron, R. A., Hsieh, F. W., Lemmon, A. R., Lemmon, E. M., & Hendry, C. R. (2016). Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). *Zoological Journal of the Linnean Society*, 177(4), 937–949. <https://doi.org/10.1111/zooj.12392>
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ragge, D. R., & Reynolds, W. J. (1998). *The songs of the grasshoppers and crickets of western Europe*. Colchester, UK: Harley Books.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M. R., Elmer, K. R., Vences, M., & Steinfarz, S. (2019). Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis. *Molecular Phylogenetics and Evolution*, 133, 189–197. <https://doi.org/10.1016/j.ympev.2019.01.003>
- Rannala, B., & Yang, Z. H. (2013). Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, 194(1), 245–253. <https://doi.org/10.1534/genetics.112.149039>
- Reynolds, W. J. (1986). A description of the song of *Omocestus broelemanni* (Orthoptera: Acrididae) with notes on its taxonomic position. *Journal of Natural History*, 20(1), 111–116. <https://doi.org/10.1080/00222938600770101>
- Rohlf, F. J. (2015). The tps series of software. *Hystrix-Italian Journal of Mammalogy*, 26(1), 9–12. <https://doi.org/10.4404/hystrix-26.1-11264>
- Rohlf, F. J., & Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1), 40–59. <https://doi.org/10.2307/2992207>
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352. <https://doi.org/10.1111/j.1461-0248.2004.00715.x>

- Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7), 372–380. [https://doi.org/10.1016/S0169-5347\(01\)02198-X](https://doi.org/10.1016/S0169-5347(01)02198-X)
- Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which patterns of gene flow is most common? *Evolution*, 68(1), 1–15. <https://doi.org/10.1111/evo.12258>
- Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*, 16(7), 940–950. <https://doi.org/10.1111/ele.12120>
- Simpson, G. G. (1944). *Tempo and mode in evolution*. New York, NY: Columbia University Press.
- Sites, J. W., & Marshall, J. C. (2003). Delimiting species: A Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, 18(9), 462–470. [https://doi.org/10.1016/s0169-5347\(03\)00184-8](https://doi.org/10.1016/s0169-5347(03)00184-8)
- 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>
- Sokal, R. R., & Crovello, T. J. (1970). Biological species concept – A critical evaluation. *American Naturalist*, 104(936), 127–153. <https://doi.org/10.1086/282646>
- Solis-Lemus, C., Knowles, L. L., & Ane, C. (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69(2), 492–507. <https://doi.org/10.1111/evo.12582>
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(6185), 738–742. <https://doi.org/10.1126/science.1252136>
- Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits its structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, 114(7), 1607–1612. <https://doi.org/10.1073/pnas.1607921114>
- Swofford, D. L. (2002). *PAUP\*. Phylogenetic analysis using parsimony (\*and other methods)*. Version 4. Sunderland, MA: Sinauer Associates.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., & Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, 18(2), 70–74. [https://doi.org/10.1016/s0169-5347\(02\)00041-1](https://doi.org/10.1016/s0169-5347(02)00041-1)
- Valbuena-Ureña, E., Oromi, N., Soler-Membrives, A., Carranza, S., Amat, F., Camarasa, S., ... Steinfartz, S. (2018). Jailed in the mountains: Genetic diversity and structure of an endemic newt species across the Pyrenees. *PLoS ONE*, 13(8), e0200214. <https://doi.org/10.1371/journal.pone.0200214>
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418–426. <https://doi.org/10.1016/j.tig.2014.07.001>
- VanderWerf, E. A., Young, L. C., Yeung, N. W., & Carlon, D. B. (2010). Stepping stone speciation in Hawaii's flycatchers: Molecular divergence supports new island endemics within the elepaio. *Conservation Genetics*, 11(4), 1283–1298. <https://doi.org/10.1007/s10592-009-9958-1>
- Wallis, G. P., Waters, J. M., Upton, P., & Craw, D. (2016). Transverse alpine speciation driven by glaciation. *Trends in Ecology & Evolution*, 31(12), 916–926. <https://doi.org/10.1016/j.tree.2016.08.009>
- Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, 67(12), 3403–3411. <https://doi.org/10.1111/evo.12134>
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649–5662. <https://doi.org/10.1111/mec.12938>
- Whelan, N. V., Galaska, M. P., Siple, B. N., Weber, J. M., Johnson, P. D., Halanych, K. M., & Helms, B. S. (2019). Riverscape genetic variation, migration patterns, and morphological variation of the threatened round rocksnail, *Leptoxis amplia*. *Molecular Ecology*, 28(7), 1593–1610. <https://doi.org/10.1111/mec.15032>
- Wiens, J. J. (2007). Species delimitation: New approaches for discovering diversity. *Systematic Biology*, 56(6), 875–878. <https://doi.org/10.1080/10635150701748506>
- Wiens, J. J., & Servedio, M. R. (2000). Species delimitation in systematics: Inferring diagnostic differences between species. *Proceedings of the Royal Society B: Biological Sciences*, 267(1444), 631–636. <https://doi.org/10.1098/rspb.2000.1049>
- Yang, Z. H. (2015). The *BP3* program for species tree estimation and species delimitation. *Current Zoology*, 61(5), 854–865. <https://doi.org/10.1093/czoolo/61.5.854>
- Yang, Z. H., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9264–9269. <https://doi.org/10.1073/pnas.0913022107>
- Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, 31(12), 3125–3135.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Tonzo V, Papadopoulou A, Ortego J. Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Mol Ecol*. 2019;00:1–18. <https://doi.org/10.1111/mec.15189>





Received: 18 December 2019 | Revised: 10 May 2020 | Accepted: 11 May 2020

DOI: 10.1111/mec.15475

## ORIGINAL ARTICLE

MOLECULAR ECOLOGY | WILEY

# Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions

Vanina Tonzo<sup>1</sup>  | Anna Papadopoulou<sup>2</sup> | Joaquín Ortego<sup>1</sup> 

<sup>1</sup>Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC), Seville, Spain

<sup>2</sup>Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus

**Correspondence**

Vanina Tonzo, Estación Biológica de Doñana, EBD-CSIC, Avda. Américo Vespucio 26, E-41092 Seville, Spain  
Email: vaninattonzo@gmail.com

**Funding information**

Spanish Ministry of Economy and Competitiveness; European Regional Development Fund (ERDF), Grant/Award Number: CGL2014-54671-P and CGL2017-83433-P; FPI Predoctoral Fellowship, Grant/Award Number: BES-2015-73159; Severo Ochoa, Grant/Award Number: SEV-2012-0262; Ramón y Cajal, Grant/Award Number: RYC-2013-12501

**Abstract**

Secondary contact in close relatives can result in hybridization and the admixture of previously isolated gene pools. However, after an initial period of hybridization, reproductive isolation can evolve through different processes and lead to the interruption of gene flow and the completion of the speciation process. *Omocestus minutissimus* and *O. uhagonii* are two closely related grasshoppers with partially overlapping distributions in the Central System mountains of the Iberian Peninsula. To analyse spatial patterns of historical and/or contemporary hybridization between these two taxa and understand how species boundaries are maintained in the region of secondary contact, we sampled sympatric and allopatric populations of the two species and obtained genome-wide single nucleotide polymorphism data using a restriction site-associated DNA sequencing approach. We used Bayesian clustering analyses to test the hypothesis of contemporary hybridization in sympatric populations and employed a suite of phylogenomic approaches and a coalescent-based simulation framework to evaluate alternative hypothetical scenarios of interspecific gene flow. Our analyses rejected the hypothesis of contemporary hybridization but revealed past introgression in the area where the distributions of the two species overlap. Overall, these results point to a scenario of historical gene flow after secondary contact followed by the evolution of reproductive isolation that currently prevents hybridization among sympatric populations.

**KEYWORDS**

coalescent-based simulations, ddRAD-seq, hybridization, introgression, reproductive isolation

**1 | INTRODUCTION**

Elucidating the processes that generate and maintain species diversity is a main ambition of evolutionary research (Fitzpatrick, Fordyce, & Gavrilets, 2009; Graham, Ron, Santos, Schneider, & Moritz, 2004; Grant, Grant, Markert, Keller, & Petren, 2004). The formation and persistence of species often depends on the evolution of reproductive isolation mechanisms that prevent interbreeding with other

recently diverged taxa or closely related lineages (Marques, Draper, Riofrio, & Naranjo, 2014; Soltis & Soltis, 2009). In the context of allopatric populations, long-term geographical isolation can facilitate the development of reproductive barriers through genetic drift or divergent selection pressures, which can ultimately lead to population divergence and speciation (Hoskin, Higgie, McDonald, & Moritz, 2005; Maguilla, Escudero, Hipp, & Luceno, 2017; Schenk, Kontur, Wilson, Noble, & Derryberry, 2018). However, lineages or species often

come into secondary contact and, if barriers to gene flow are lacking or incomplete, hybridization will take place. This phenomenon can have different outcomes with important evolutionary consequences (Abbott et al., 2013; Mallet, 2005; Mayr, 1963). At one extreme, barriers to gene exchange may break down upon secondary contact and lead to the collapse of formerly distinct species in a hybrid swarm characterized by extensive admixture of parental genotypes (i.e., speciation reversal or lineage fusion; Kearns et al., 2018; Taylor et al., 2006). At the opposite extreme, reduced fitness of hybrids and strong selection against them will favour the rapid evolution of barriers to gene flow (i.e., prezygotic isolation), which can ultimately lead to total reproductive isolation and culminate in the completion of the speciation process (i.e., reinforcement of isolation; Butlin, 1995; Dobzhansky, 1937; Lemmon & Juenger, 2017; Servedio & Kirkpatrick, 1997). An intermediate scenario is the formation of tension zones with a variable geographical width determined by the equilibrium between dispersal and selection against hybrids (Barton & Hewitt, 1985; Key, 1968).

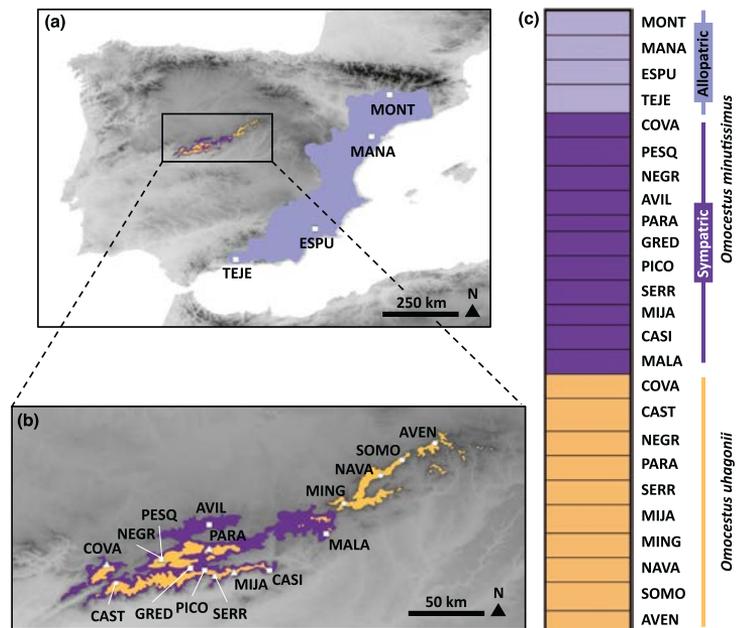
Hybrid zones have been defined as “windows on the evolutionary processes” (Harrison, 1990; Hewitt, 1988) and their study through space (Barton & Hewitt, 1989; Howard, Waring, Tibbets, & Gregory, 1993) and time (Britch, Cain, & Howard, 2001; Buggs, 2007) has provided important insights into the evolution of reproductive isolation and the formation of new species. Past climate changes, such as the glacial cycles of the Pleistocene, have resulted in recurrent range expansions and contractions in many organisms, often putting into geographical contact closely related species and lineages that had remained geographically isolated for large periods of time. Alpine and montane species represent a paradigmatic example of species experiencing dramatic distributional shifts in response to Pleistocene glacial cycles, descending to lower altitudes and expanding their distributions in cold periods and shrinking their ranges to high-elevation areas during interglacials (Schmitt, 2009; Seddon, Santucci, Reeve, & Hewitt, 2001; Tzedakis, Emerson, & Hewitt, 2013). Under these conditions, closely related species that evolved in isolation during interglacial periods can recurrently come into secondary contact and hybridize. In other cases, interbreeding species share a large portion of their respective ranges but only hybridize in certain areas with specific environmental conditions or scattered patches where they meet, forming a mosaic hybrid structure rather than a well-defined cline limited to narrow contact zones (Barton & Hewitt, 1985). Independently of their nature and origin, contact zones offer the opportunity to study in real time the process of reproductive isolation or, if completed, to obtain indirect evidence about when (tempo) and how (modes) it might have evolved. However, testing alternative hypotheses about the evolution of reproductive isolation can be extremely challenging, making it necessary to integrate multiple sources of analytical evidence (e.g., phylogenetics and population genetics) and to develop model-based approaches considering the biogeographical context of gene flow (Payseur & Rieseberg, 2016). The high power of genomic data to resolve historical events of hybridization and test complex scenarios of gene flow in virtually any organismal model

and biogeographical setting has exponentially increased our capacity to quantify the magnitude and timing of interspecific gene flow and distinguish among alternative demographic scenarios (e.g., de Manuel et al., 2016; Lohse, Clarke, Ritchie, & Etges, 2015; Ortego, Gugger, & Sork, 2018).

Here, we use as a study system two recently diverged grasshopper species with partly overlapping distributions to illustrate the potential of integrating different analytical approaches for inferring the tempo and mode of evolution of reproductive isolation (or its lack thereof) and gain insights into the speciation process. Grasshoppers (Orthoptera: Caelifera) are an interesting system to study hybridization and its evolutionary consequences, as many species have very recently evolved in allopatry (Mayer, Berger, Gottsberger, & Schulze, 2010; Ragge & Reynolds, 1998) and present incomplete barriers to gene flow (e.g., Saldamando, Tatsuta, & Butlin, 2005). In turn, their distributions often overlap across large geographical areas (Hill, 2015), show a mosaic distribution (Rohde et al., 2017) or have recurrently come into secondary contact as a consequence of range shifts driven by past climate changes (Bridle, Baird, & Butlin, 2001), providing ideal biogeographical scenarios for the study of hybridization and the evolution of reproductive isolation (Butlin, Ritchie, & Hewitt, 1991; Virdee & Hewitt, 1994). In this study we focus on two grasshopper species of the subgenus *Dreuxius* (genus *Omocestus*), a species complex comprising nine taxa distributed in the Iberian Peninsula and northwest Africa (Cigliano, Braun, Eades, & Otte, 2019; García-Navas, Noguerales, Cordero, & Ortego, 2017). Most species of this complex are distributed in allopatry, isolated at high elevation in different mountain ranges. One exception are the taxa *Omocestus minutissimus* (Brullé 1832) and *O. uhagonii* (Bolivar 1876), which show partially overlapping distributions in the Central System Mountains of the Iberian Peninsula. As with other species of the subgenus, both taxa are brachypterous, present a low dispersal capacity and have a similar annual life cycle, with an adult breeding phase from the end of July to the beginning of October (Clemente, Garcia, & Presa, 1991). The two species are predominantly graminivorous and occupy open habitats tightly linked to cushion and thorny shrub formations that they use as refuge (Clemente et al., 1991). However, both species differ on the extent of their distributions and elevational ranges. While *O. minutissimus* presents a wider distribution, with patchy populations distributed in eastern and central Iberia from sea level to 2,500 m elevation, *O. uhagonii* is restricted to the Central System and altitudes over 1,600–1,800 m (Figure 1). The two taxa partially co-occur across the distribution range of *O. uhagonii*, with several sympatric populations at high elevations in which adult individuals of the two species co-exist at high numbers in the same microhabitats (J. Ortego, personal observation). Therefore, this system provides an interesting case study to analyse the presence of contemporary and past hybridization and understand the maintenance of species boundaries in two closely related taxa that might have weak or recently evolved reproductive isolation mechanisms.

We extensively sampled sympatric and allopatric populations of *O. minutissimus* and *O. uhagonii* across their respective

**FIGURE 1** Biogeographical setting of the study system. (a,b) Maps showing the sampled populations and the distribution range of the two studied taxa based on our own species records (*Omocestus minutissimus*: purple areas and squares; *O. uhagonii*: light orange areas and dots). *O. minutissimus* presents a partially overlapping distribution with *O. uhagonii* in the Central System (deep purple) and allopatric populations (light purple) in eastern Iberia. Triangles indicate sampling localities where the two species were found living in sympatry. (c) Genetic assignment of individuals based on the results of *FASTSTRUCTURE*. Individuals are partitioned into *K* coloured segments representing the probability of belonging to the cluster with that colour and thin vertical black lines separate individuals from different populations. Population codes as in Table S1



distribution ranges and genotyped them via restriction-site-associated DNA sequencing (ddRAD-seq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) to infer historical and contemporary interspecific gene flow and elucidate the evolutionary outcomes of such processes. Specifically, we first used Bayesian clustering analyses to determine the genetic ancestry of individuals and test the hypothesis of contemporary hybridization in sympatric populations of the two taxa. Second, we employed a suite of phylogenomic approaches and a coalescent-based simulation framework to evaluate alternative scenarios of historical hybridization and estimate the timing, magnitude and directionality of interspecific gene flow. Our genomic data rejected the hypothesis of contemporary hybridization but revealed the presence of past genetic introgression, pointing to a scenario of historical hybridization after secondary contact followed by the evolution of reproductive barriers that today prevent gene flow among sympatric populations of the two species.

## 2 | MATERIALS AND METHODS

### 2.1 | Population sampling

Between 2011 and 2015, we sampled *O. minutissimus* (88 individuals, 15 populations) and *O. uhagonii* (64 individuals, 10 populations) from a total of 25 populations that cover the entire distribution range of the two taxa (Table S1; Figure 1). Eleven sampling populations of *O. minutissimus* are located in central Iberia and partly overlap with the distribution range of *O. uhagonii* (hereafter referred as sympatric

populations). Among these populations of *O. minutissimus*, five were strictly sympatric with *O. uhagonii* and the two species were collected from the same localities (Table S1; Figure 1). The rest of the sampled populations of *O. minutissimus* are located in eastern Iberia and separated >300 km from the nearest population of *O. uhagonii* (hereafter referred to as allopatric populations). We stored specimens in 2-ml vials with 96% ethanol and preserved them at  $-20^{\circ}\text{C}$  until needed for DNA extraction. Detailed information on sampling populations is presented in Table S1.

### 2.2 | Genomic library preparation and data processing

We used NucleoSpin Tissue kits (Macherey-Nagel) to extract and purify total genomic DNA from a hind leg of each individual. We processed genomic DNA in house following the ddRAD-seq procedure described in Peterson et al. (2012), with some minor modifications as detailed in Lanier, Massatti, He, Olson, and Knowles (2015). Briefly, we digested DNA with the restriction enzymes *EcoRI* and *MseI* (New England Biolabs) and ligated Illumina adaptors including unique 7-bp barcodes to the digested fragments of each individual. We pooled ligation products into four different libraries, size selected for fragments between 475 and 580 bp using a Pippin Prep machine (Sage Science), and amplified them by polymerase chain reaction (PCR) with 10–12 cycles using the iProof High-Fidelity DNA Polymerase (Bio-Rad). We sequenced the libraries in single-read 151-bp lanes on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Hospital for Sick Children).

We demultiplexed raw sequences using `PROCESS_RADTAGS`, a program distributed as part of the `STACKS` pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). We only retained reads with Phred scores  $\geq 10$  (using a sliding window of 15%), no adaptor contamination and unambiguous barcode and restriction cut sites. We checked read quality in `FASTQC` version 0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed sequences to 130 bp using a `SEQTK` script (Heng Li, <https://github.com/lh3/seqtk>) in order to remove low-quality reads near the 3' ends. As an additional quality-filtering step, we used `PYRAD` version 3.0.66 (Eaton, 2014) to convert base calls with a Phred score  $< 20$  into Ns and discard reads with  $> 2$  Ns. We assembled retained reads into de novo loci using `PYRAD`, considering parameter values for clustering threshold of sequence similarity ( $W_{\text{CLUST}} = 0.85$ ), minimum coverage depth ( $d = 5$ ), maximum number of individuals with shared heterozygous sites ( $\text{maxSH} = p.10$ ) and maximum number of polymorphic sites in a final locus ( $\text{maxSNPs} = 20$ ) based on suggestions from the literature (Eaton, 2014; Eaton & Ree, 2013; Takahashi, Nagata, & Sota, 2014). Finally, we generated final data sets for subsequent analyses discarding loci that were not present in at least  $\sim 25\%$  of the samples ( $\text{minCov} = \sim 25\%$ ).

### 2.3 | Genetic structure and hybrid identification

We identified hybrids and introgressed individuals between *O. minutissimus* and *O. uhagonii* using the Bayesian clustering methods implemented in the programs `FASTSTRUCTURE` version 1.0 (Raj, Stephens, & Pritchard, 2014) and `STRUCTURE` version 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). First, we used the highly efficient algorithm implemented in `FASTSTRUCTURE` to analyse the entire data set including all populations. Second, we used both `FASTSTRUCTURE` and classic `STRUCTURE` to perform more detailed analyses focused on the Central System, the region where the ranges of the two taxa partly overlap, with some populations living in sympatry and, thus, the two species currently have the opportunity to hybridize. We ran `FASTSTRUCTURE` analyses using a simple prior, considering a convergence criterion of  $1 \times 10^{-7}$  and conducting 25 independent runs for each value of  $K$  (from  $K = 1$  to  $K = 10$ ). Following Raj et al. (2014), we used the `chooseK.py` script to assess model complexity by estimating the metrics  $K_{\text{sc}}$ , the value of  $K$  that maximizes the log-marginal likelihood lower bound (LLBO) of the data, and  $K_{\text{c}}$ , the smallest number of model components explaining at least 99% of the cumulative ancestry contribution. We plotted individual co-ancestry coefficients for the most likely  $K$  value using `DISTRUCT` version 1.1 (Rosenberg, 2004). We ran `STRUCTURE` assuming correlated allele frequencies and admixture, and without using prior population information (Hubisz, Falush, Stephens, & Pritchard, 2009; Pritchard et al., 2000). We conducted 15 independent runs for each value of  $K$  (from  $K = 1$  to  $K = 10$ ) to estimate the optimal number of genetic clusters with 200,000 Markov chain Monte Carlo (MCMC) cycles, following a burn-in step of 100,000 iterations. We used `STRUCTURE`

`HARVESTER` (Earl & vonHoldt, 2012) to assess the number of genetic clusters that best describes our data according to log probabilities of the data ( $\text{LnPr}(X|K)$  for each value of  $K$  (Pritchard et al., 2000) and the  $\Delta K$  method (Evanno, Regnaut, & Goudet, 2005). We used `CLUMPP` version 1.1.2 and the Greedy algorithm to align multiple runs of `STRUCTURE` for the same  $K$  value (Jakobsson & Rosenberg, 2007) and `DISTRUCT` to visualize as bar plots the individual's probabilities of population membership.

### 2.4 | Phylogenomic analyses and inference of historical hybridization

To determine the presence of historical hybridization (i.e., introgression), we employed four-taxon ABBA/BABA tests based on the  $D$ -statistic (Durand, Patterson, Reich, & Slatkin, 2011) and `TREEMIX` analyses (Pickrell & Pritchard, 2012). Assuming that the sister taxa P1 and P2 diverged from P3 and an outgroup species O, the  $D$ -statistic is used to test the null hypothesis of no introgression ( $D = 0$ ) between P3 and P1 or P2.  $D$  values significantly different from zero indicate gene flow between P1 and P3 ( $D < 0$ ) or between P2 and P3 ( $D > 0$ ). We assigned sympatric populations of *O. minutissimus* to P1 (64 individuals, from localities COVA, PESQ, NEGR, AVIL, PARA, GRED, PICO, SERR, MIJA, CASI and MALA; hereafter, MS), allopatric populations of *O. minutissimus* to P2 (24 individuals, from localities MONT, PORT, ESPU and TEJE; hereafter, MA) and populations of *O. uhagonii* to P3 (64 individuals from all sampling localities for this taxon; hereafter, US). Note that the sympatric and allopatric populations of *O. minutissimus* are located in the central and west portions of the species distribution range, respectively, and correspond to the two well-defined genetic clusters identified by Bayesian clustering analyses for this taxon (see Section 3 and Figure 1). We used as the outgroup (O) the taxon *O. antigai* (Bolívar, 1897), a species also belonging to the subgenus *Dreuxius* (Cigliano et al., 2019; García-Navas et al., 2017). Specifically, we used sequences from 94 individuals of this species available at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA543714 (Tonzo, Papadopoulou, & Ortego, 2019). We performed ABBA/BABA tests in `PYRAD` and used 1,000 bootstrap replicates to obtain the standard deviation of the  $D$ -statistic (Eaton & Ree, 2013; e.g., Huang, 2016).

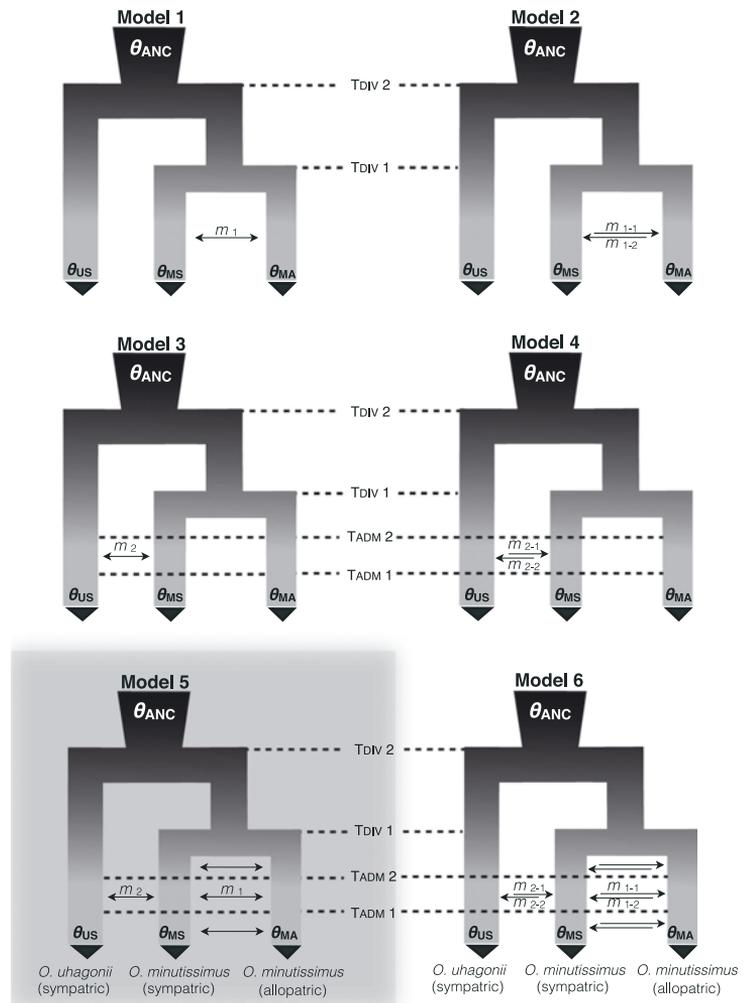
We also analysed the potential presence of introgression and determined the direction of gene flow with `TREEMIX` version 1.12 (Pickrell & Pritchard, 2012). We used `TREEMIX` to construct a tree-based model of population genetic relationships and infer events of genetic admixture using single nucleotide polymorphism (SNP) frequency data. `TREEMIX` fits a population graph (i.e., a phylogenetic tree that incorporates admixture) on the basis of allele frequencies and a Gaussian approximation to genetic drift, allowing patterns of splits and mixtures in multiple populations to be inferred. To perform `TREEMIX` analyses, we pooled populations into the same three groups used to run ABBA/BABA tests. In a first step, we estimated a maximum-likelihood tree rooted with *O. antigai*. Then, we tested

the existence of a range of migration events ( $m = 0-5$ , with three replicated runs each) and calculated the proportion of the variance in population covariances explained by the population graph with different numbers of admixture events to determine the model best fitting the data (e.g., Gompert et al., 2014). We assumed the independence of all SNPs and used a window size of one SNP ( $k = 1$ ; e.g., Vera, Díez-del-Molino, & García-Marín, 2016).

## 2.5 | Testing alternative models of gene flow

AS TREEMIX only models migration as discrete events and does not consider continuous gene flow (Pickrell & Pritchard, 2012), we applied the coalescent-based modelling approach implemented in FASTSIMCOAL2 (Excoffier, Dupanloup, Huerta-Sanchez, Sousa, & Foll, 2013)

to statistically test the relative fit of more complex historical demographic models to our genomic data. We used FASTSIMCOAL2 and the site frequency spectrum (SFS; Excoffier et al., 2013) to test seven alternative models of gene flow (Figure 2). These models considered the same three population groups used in ABBA/BABA tests and TREEMIX analyses. All scenarios considered an early split between the two species followed by the divergence between sympatric (MS) and allopatric (MA) populations of *O. minutissimus*. The tested scenarios considered (a) total absence of post-divergence gene flow (Model 0, not shown in Figure 2); (b) gene flow only between the two population groups of *O. minutissimus* (i.e., absence of interspecific gene flow; Models 1 and 2); (c) historical gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* (i.e., interspecific gene flow) but absence of post-divergence gene flow between the two population groups of *O. minutissimus* (Models 3 and 4); and (d) gene



**FIGURE 2** Alternative migration models tested using FASTSIMCOAL2. Parameters include ancestral ( $\theta_{ANC}$ ) and contemporary ( $\theta_{US}$ ,  $\theta_{MS}$ ,  $\theta_{MA}$ ) effective population sizes, timing of population split ( $T_{DIV}$ ) and admixture ( $T_{ADM}$ ), and migration rates ( $m$ ) between different pairs of populations. Grey background highlights the best-supported model

flow between the two population groups of *O. minutissimus* and historical gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* (Models 5 and 6) (Figure 2). These scenarios were tested considering both symmetric (Models 1, 3 and 5) and asymmetric (Models 2, 4 and 6) gene flow (Figure 2). Note that detailed Bayesian clustering analyses performed across multiple populations of the two taxa in the region where their distribution ranges overlap did not show any evidence of ongoing interspecific gene flow (see Section 3) and, for this reason, we did not consider models incorporating contemporary interspecific gene flow.

We calculated a folded joint SFS considering a single SNP per locus to avoid the effects of linkage disequilibrium. Because we did not include invariable sites in the SFS, we fixed the effective population size for one population group (US) to enable the estimation of other parameters in FASTSIMCOAL2 (e.g., Lanier et al., 2015; Papadopoulou & Knowles, 2015). The effective population size ( $N_e$ ) fixed in the models was calculated from the level of nucleotide diversity ( $\pi$ ) and estimates of mutation rate per site per generation ( $\mu$ ), according to the equation  $N_e = \pi/4\mu$  (Lynch & Conery, 2003). Nucleotide diversity for US ( $\pi = 0.005$ ) was estimated from polymorphic and nonpolymorphic loci using DNASP version 6.11.01 (Librado & Rozas, 2009) and the *allele* file generated by PYRAD. We considered the average mutation rate per site per generation of  $2.80 \times 10^{-9}$  estimated for *Drosophila melanogaster* (Keightley, Ness, Halligan, & Haddrill, 2014). To remove all missing data for the calculation of the joint SFS and minimize errors with allele frequency estimates, each population group was down-sampled to 25% of individuals (32, 12 and 32 genes for MS, MA and US, respectively) using a custom Python script written by Isaac Overcast and available at GitHub (<https://github.com/isaacovercast/easySFS>). The final SFS contained 2,071 variable SNPs. Each of the seven models was run 100 replicated times using the computing resources provided by CESGA (Galician Supercomputer Center) considering 100,000–250,000 simulations for the calculation of the composite likelihood, 10–40 expectation-conditional maximization (ECM) cycles, and a stopping criterion of 0.001 (Papadopoulou & Knowles, 2015). We used an information-theoretic model selection approach based on Akaike's information criterion (AIC) to determine the probability of each model given the observed data (Burnham & Anderson, 2002; e.g., Abascal et al., 2016; Thome & Carstens, 2016). After the maximum likelihood was estimated for each model in every replicate, we calculated the AIC scores as detailed in Thome and Carstens (2016). AIC values for each model were rescaled ( $\Delta$ AIC) calculating the difference between the AIC value of each model and the minimum AIC obtained among all competing models (i.e., the best model has  $\Delta$ AIC = 0). Point estimates of the different demographic parameters for the best-supported model were selected from the run with the highest maximum composite likelihood. Finally, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier et al., 2013; e.g., Papadopoulou & Knowles, 2015).

### 3 | RESULTS

#### 3.1 | Genomic data

Illumina sequencing returned an average of  $2.74 \times 10^6$  reads per sample. After quality control, an average of  $2.33 \times 10^6$  reads per sample was retained (Figure S1). The data sets obtained with PYRAD for all populations and only those from the Central System retained 15,219 and 20,350 unlinked SNPs, respectively.

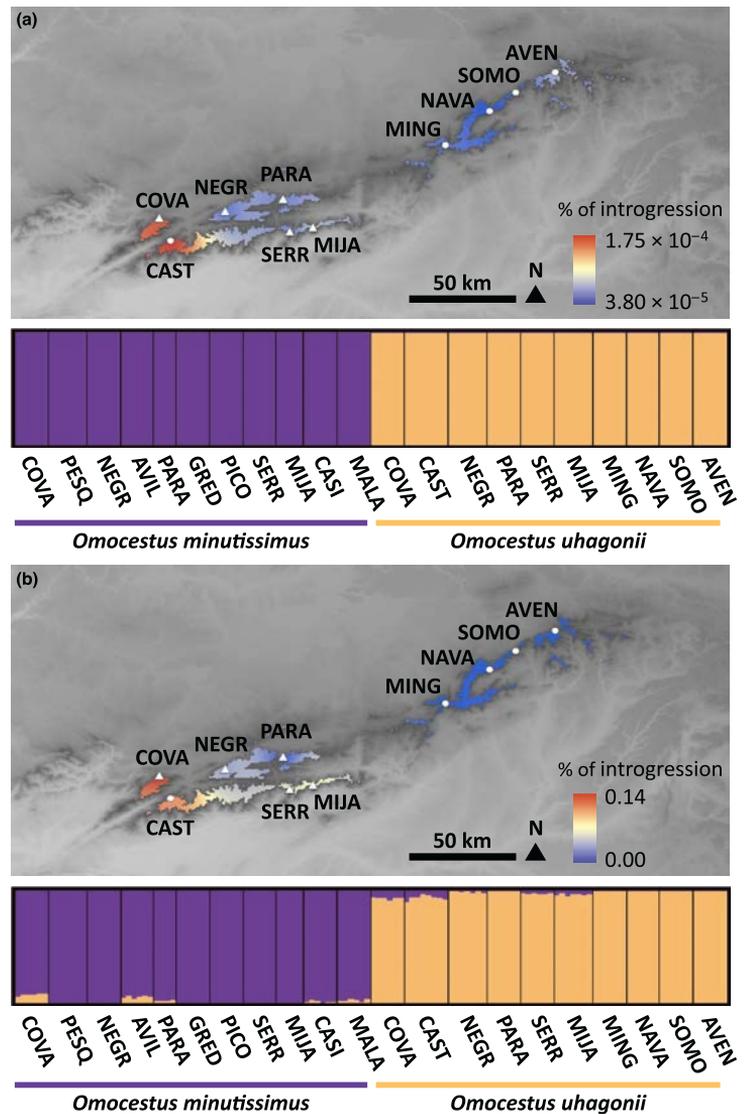
#### 3.2 | Genetic structure and hybrid identification

The model complexity value that maximized the marginal likelihood in FASTSTRUCTURE analyses for the data set including all populations was  $K = 3$  across all replicates and the number of model components used to explain structure in the data was  $K = 3$  in 15 replicates and  $K = 4$  in 10 replicates. FASTSTRUCTURE analyses for  $K = 3$  split populations of *O. uhagonii* and sympatric and allopatric populations of *O. minutissimus* into different genetic clusters in which all individuals showed high probabilities of population membership ( $q > 0.99$ ; Figure 1c). Assignment values to additional genetic clusters in FASTSTRUCTURE for analyses with  $K > 4$  were extremely low in all cases ( $q < 0.001$ ) and, thus, their respective bar plots were virtually identical to those obtained for  $K = 3$  (for a similar result, see Baiz, Tucker, & Cortes-Ortiz, 2019; Tonzo et al., 2019).

The model complexity value that maximized the marginal likelihood in FASTSTRUCTURE analyses of the Central System populations was equal to  $K = 2$  in all replicates and the number of model components used to explain structure in the data was  $K = 2$  in 16 replicates and  $K = 3$  in nine replicates. FASTSTRUCTURE analyses for  $K = 2$  split populations of the two species in different genetic clusters and all individuals showed high probabilities of population membership ( $q > 0.99$ ; Figure 3a). Again, assignment values to additional genetic clusters in FASTSTRUCTURE for analyses with  $K > 2$  were extremely low in all cases ( $q < 0.001$ ) and, thus, their respective bar plots were virtually identical to those obtained for  $K = 2$ . Classic STRUCTURE analyses also yielded an "optimal" clustering value for  $K = 2$  according to the  $\Delta K$  criterion (Figure S2). The two inferred genetic groups supported a clear separation of the two species (Figure 3b). STRUCTURE analyses showed that populations of *O. uhagonii* from the eastern Central System present no signal of introgression from *O. minutissimus*. However, several populations of *O. minutissimus* and *O. uhagonii* from the western Central System, where the distribution of the two taxa overlap and five sampling localities present sympatric populations, showed signals of reciprocal genetic introgression (Figure 3b). These results suggest that FASTSTRUCTURE is less likely to reveal small proportions of admixed ancestry in comparison with STRUCTURE, which has been also suggested in previous studies (Stift, Kolar, & Meirns, 2019; Tonzo et al., 2019).

STRUCTURE analyses indicate that populations of *O. uhagonii* and *O. minutissimus* present significant differences in the degree of introgression from the other species (one-way ANOVAs, introgression

**FIGURE 3** Genetic assignment of *Omocestus minutissimus* and *O. uhagonii* from the Central System based on the results of (a) FASTSTRUCTURE and (b) STRUCTURE. Individuals are partitioned into  $K$  coloured segments representing the probability of belonging to the cluster with that colour and thin vertical black lines separate individuals from different populations. Maps display the distribution of *O. uhagonii* and the probability of assignment of the populations of this species (spatial interpolation) to the genetic cluster of *O. minutissimus* (i.e., the degree of introgression from *O. minutissimus* to *O. uhagonii*). Population codes as in Table S1



of *O. minutissimus* into *O. uhagonii*:  $F_{9,54} = 48.86$ ,  $p < .001$ ; introgression of *O. uhagonii* into *O. minutissimus*:  $F_{10,53} = 96.68$ ,  $p < .001$ ; Figure 3b). Although visually imperceptible in the FASTSTRUCTURE bar plot (Figure 3a), probabilities of population membership inferred by this software also revealed that populations of both *O. uhagonii* and *O. minutissimus* present significant differences in the degree of introgression from the other species (one-way ANOVAs, introgression of *O. minutissimus* into *O. uhagonii*:  $F_{9,54} = 14.01$ ,  $p < .001$ ; introgression of *O. uhagonii* into *O. minutissimus*:  $F_{10,53} = 3.19$ ,  $p = .003$ ; Figure 3a). The degree of introgression from *O. minutissimus* into *O. uhagonii* estimated by either FASTSTRUCTURE or STRUCTURE decreased significantly with longitude (FASTSTRUCTURE:  $F_{1,8} = 6.99$ ,  $p = .030$ ; STRUCTURE:

$F_{1,8} = 16.22$ ,  $p = .004$ ; see maps in Figure 3), but did not differ significantly between currently sympatric and allopatric populations (FASTSTRUCTURE:  $F_{1,8} = 0.06$ ,  $p = .81$ ; STRUCTURE:  $F_{1,8} = 1.64$ ,  $p = .236$ ). In contrast, the degree of introgression from *O. uhagonii* into *O. minutissimus* estimated by either FASTSTRUCTURE or STRUCTURE did not decrease significantly with longitude (FASTSTRUCTURE:  $F_{1,9} = 1.12$ ,  $p = .317$ ; STRUCTURE:  $F_{1,9} = 0.41$ ,  $p = .538$ ) or differ between currently sympatric and allopatric populations (FASTSTRUCTURE:  $F_{1,9} = 0.39$ ,  $p = .550$ ; STRUCTURE:  $F_{1,9} = 0.03$ ,  $p = .863$ ). For illustrative purposes, we displayed on a map the probabilities of assignment of the populations of *O. uhagonii* to the genetic cluster of *O. minutissimus* (i.e., the degree of introgression from *O. minutissimus* into *O. uhagonii*) by conducting a spatial

interpolation using the Inverse Distance Weight (IDW) function available in ARCGIS version 10.5 (ESRI; Figure 3).

### 3.3 | Phylogenomic analyses and inference of historical hybridization

The test of introgression based on the  $D$ -statistic supported post-divergence gene flow between sympatric populations of *O. minutissimus* and *O. uhagonii* (BABA = 220.59; ABBA = 162.38;  $D$ -statistic =  $-0.152$ ;  $Z = 3.39$ ;  $p < .001$ ). Accordingly, TREEMIX analyses supported a single migration event (Figure S3) corresponding to admixture between sympatric populations of *O. minutissimus* and *O. uhagonii* (Figure 4).

### 3.4 | Testing alternative models of gene flow

FASTSIMCOAL2 analyses showed that the best supported model (Model 5;  $\Delta AIC = 0$ ; Table 1) was the one considering symmetric interspecific gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* during a given period of time ( $T_{ADM1}$  and  $T_{ADM2}$ ) and symmetric gene flow between sympatric (central) and allopatric (eastern) populations of *O. minutissimus*. Remarkably, models only considering interspecific gene flow during a given period of time

tended to be slightly more well supported ( $\Delta AIC > 0.6$ ) than models only considering intraspecific gene flow between sympatric and allopatric populations of *O. minutissimus* (Table 1). Point estimates of demographic parameters under the best fitting model are presented in Table 2. Considering that the studied taxa are univoltine (i.e., 1-year generation time), FASTSIMCOAL2 analyses inferred that the separation of the two species ( $T_{DIV2}$ ) and the split of eastern and central populations of *O. minutissimus* ( $T_{DIV1}$ ) took place during the Early Pleistocene (Calabrian age; Table 2). Gene flow between sympatric populations of *O. minutissimus* and *O. uhagonii* was estimated to have occurred during a period of  $\sim 15$  ka in the Late Pleistocene (Tarantian age; Table 2). Note, however, that although confidence intervals around point estimates of most parameters were reasonably

**TABLE 1** Comparison of alternative migration models (detailed in Figure 2) tested using FASTSIMCOAL2

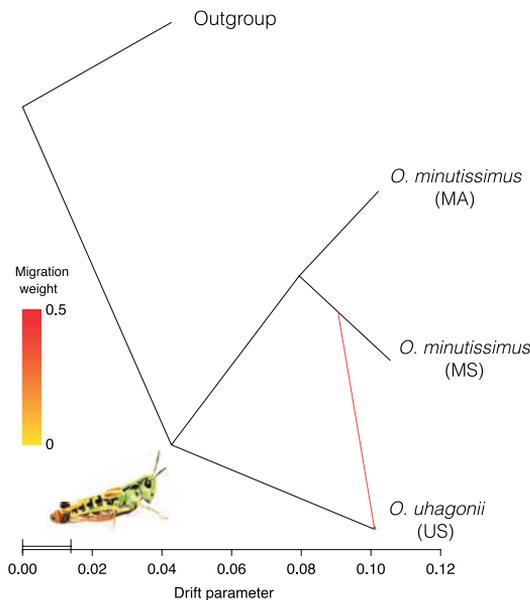
Model	$k$	$\log_{10}L$	AIC	$\Delta AIC$	$\omega_i$
Model 0	6	-3,539.32	7,090.64	90.01	0.00
Model 1	7	-3,518.85	7,051.71	51.08	0.00
Model 2	8	-3,518.42	7,052.84	52.21	0.00
Model 3	9	-3,515.49	7,048.98	48.34	0.00
Model 4	10	-3,515.54	7,051.09	50.45	0.00
Model 5	<b>10</b>	<b>-3,490.32</b>	<b>7,000.63</b>	<b>0.00</b>	<b>0.86</b>
Model 6	12	-3,490.09	7,004.19	3.56	0.14

For each model, the table shows the maximum likelihood estimate of the model ( $\log_{10}L$ ), the number of parameters ( $k$ ), Akaike's information criterion score (AIC), the difference in AIC value of each model from that of the strongest model ( $\Delta AIC$ ) and AIC weight ( $\omega_i$ ). The best-supported model ( $\Delta AIC < 2$ ) is indicated in bold.

**TABLE 2** Parameters inferred from coalescent simulations with FASTSIMCOAL2 under the best-supported demographic model (Model 5)

Parameter	Point estimate	Lower bound	Upper bound
$\theta_{ANC}$	702,036	376,562	832,461
$\theta_{MS}$	813,263	686,748	890,000
$\theta_{MA}$	1,556,235	1,293,951	1,675,243
$T_{DIV1}$	1,199,191	922,762	1,380,088
$T_{DIV2}$	1,382,958	1,278,828	1,670,203
$T_{ADM1}$	14,090	1,632	23,626
$T_{ADM2}$	29,040	50,260	750,959
$m_1$	$1.28 \times 10^{-7}$	$9.02 \times 10^{-8}$	$1.82 \times 10^{-7}$
$m_2$	$2.41 \times 10^{-7}$	$2.14 \times 10^{-8}$	$1.37 \times 10^{-7}$

The table shows point estimates and lower and upper 95% confidence intervals. Note that the effective population size of *Omocestus uhagonii* is not presented because it was fixed in FASTSIMCOAL2 analyses to enable the estimation of other parameters (see the Materials and Methods section for further details).  $\theta$ , mutation-scaled effective population sizes;  $T_{DIV}$  and  $T_{ADM}$ , timing of population divergence and admixture, respectively (given in number of generations);  $m$ , migration rates per generation. Each specific parameter is illustrated in Figure 2.



**FIGURE 4** Maximum-likelihood tree inferred with TREEMIX for *Omocestus uhagonii* (US) and sympatric (MS) and allopatric (MA) populations of *O. minutissimus*. The direction of gene flow (from MS to US) for the most likely migration event ( $m = 1$ ) inferred is represented with an arrow coloured according to the percentage of alleles (weight) originating from the source

tight, there was considerable uncertainty around estimates for the two time parameters delimiting the period of interspecific gene flow (i.e.,  $T_{ADM1}$  and  $T_{ADM2}$ ; Table 2). Finally, migration rates between sympatric populations of *O. uhagonii* and *O. minutissimus* ( $m_2$ ) did not differ significantly from those estimated between sympatric (western) and allopatric (eastern) populations of *O. minutissimus* ( $m_1$ ) (i.e., 95% confidence intervals [CIs] of  $m_1$  and  $m_2$  overlapped; Table 2). This indicates that historical interspecific gene flow was of the same order of magnitude as intraspecific gene flow between the two allopatric genetic clusters of *O. minutissimus*.

## 4 | DISCUSSION

Hybridization has been extensively documented in contact zones where the distributions of closely related species with weak reproductive barriers meet (e.g., Folk, Soltis, Soltis, & Guralnick, 2018; Gugger & Cavender-Bares, 2013; Nadeau et al., 2013; Ortego, Gugger, Riordan, & Sork, 2014). Inferring events of past interspecific gene flow has important implications to understand the evolutionary history of organisms (e.g., humans; Prüfer et al., 2014; Wall et al., 2013), yet detecting the footprints of such processes is challenging (Payseur & Rieseberg, 2016; e.g., Eaton, Hipp, González-Rodríguez, & Cavender-Bares, 2015; Ortego et al., 2018). Here, by combining a suite of phylogenomic and population genetic tools and extensive population sampling across the entire distribution of our two focal species, including currently sympatric and allopatric populations, we found no evidence for contemporary hybridization. However, we did detect signals of past introgression in the geographical region where the distribution range of the two taxa currently overlap.

### 4.1 | Absence of contemporary interspecific gene flow

Bayesian clustering analyses showed a clear genotypic differentiation of the two species and further revealed the presence of two well-defined genetic clusters within *O. minutissimus*, corresponding to the populations of this taxon located in eastern (allopatric) and central (sympatric) Iberia (Figure 1). Detailed analyses across 21 populations from central Iberia where the distribution of the two species partially overlap and, thus, where they currently have the opportunity to hybridize, showed no evidence of ongoing interspecific gene flow (i.e.,  $F_1$  or first generation backcrosses). However, these analyses also revealed footprints of reciprocal introgression in the westernmost portion of the Central System, the area where the two species present overlapping distributions and some populations even co-occur. Although the degree of introgression did not differ statistically between currently sympatric and allopatric populations of the two species in the Central System, it was not spatially homogeneous. On the one hand, the proportion of genetic introgression differed significantly across populations of the

two species. On the other hand, the degree of introgression from *O. minutissimus* into *O. uhagonii* increased westwards and the populations from the easternmost portion of the distribution range of this species, where *O. minutissimus* is not currently present, showed negligible signals of past hybridization (Figure 3). These results indicate spatial heterogeneity in the levels of introgression, suggesting that the magnitude and/or timing of historical hybridization differed among populations of the two species in the Central System (e.g., de Manuel et al., 2016; Ortego et al., 2018; Wall et al., 2013). The degree of introgression was consistently small in all populations of both species (STRUCTURE: <9%; FASTSTRUCTURE: <0.02%) and similar across individuals within populations. Thus, although our sample sizes are modest (128 individuals) and we cannot categorically discard that the two species sporadically hybridize, the observed patterns of introgression indicate that contemporary populations are at genotypic equilibrium (i.e., backgrounds of introgression are similar across all individuals within a given population) and suggest that hypothetical contemporary hybridization, if it even happens, is unlikely to have transcended  $F_1$  hybrids at least in the last generations. This result is also supported by the fact that the five populations where the two species currently co-occur do not show higher levels of introgression than nearby allopatric populations, which points to the fact that the observed patterns of genetic introgression reflect historical rather than contemporary interspecific gene flow.

### 4.2 | Inferring historical hybridization

Both the phylogenomic analyses in TREEMIX and the  $D$ -statistic test yielded results compatible with those inferred by Bayesian clustering analyses and supported the hypothesis of historical hybridization between *O. uhagonii* and geographically overlapping populations of *O. minutissimus*. TREEMIX identifies gene flow in the context of competing hypotheses, taking into account the full sampled phylogeny when inferring admixture and introgression events (Pickrell & Pritchard, 2012). Even when we allowed several admixture edges for the two species, TREEMIX only inferred one event of introgression involving the sympatric populations of the two species. In agreement with the TREEMIX results, the  $D$ -statistic analysis detected the same pattern of interspecific gene flow. Bringing the TREEMIX and the  $D$ -statistic results together supports a scenario of historical hybridization and enhances the interpretation of the low levels of genetic introgression revealed by clustering analyses in currently coexisting populations.

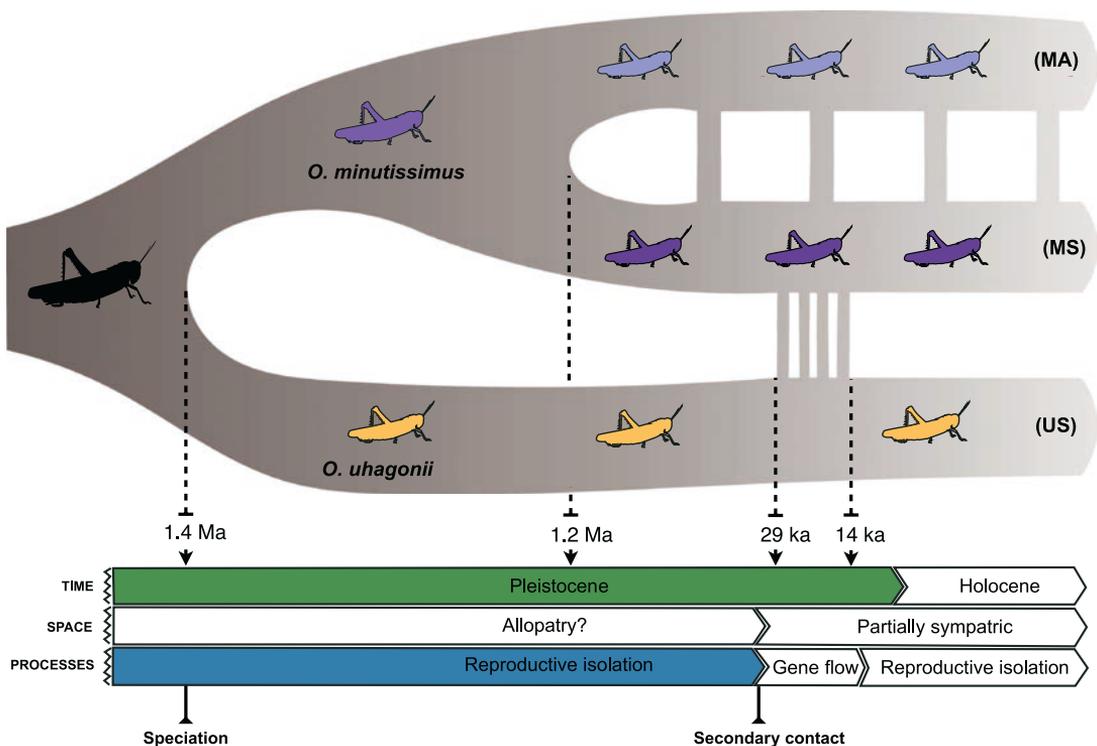
Coalescent-based analyses in FASTSIMCOAL2, which provides detailed estimation of demographic parameters, further supported a scenario of past interspecific gene flow over a strictly bifurcating evolutionary history. Specifically, the most well-supported model was the one considering interspecific gene flow between sympatric populations of *O. uhagonii* and *O. minutissimus* during a given period of time and gene flow between sympatric (central) and allopatric (eastern) populations of *O. minutissimus*. The preferred model traced back the divergence of *O. uhagonii* and *O. minutissimus* to the

Early Pleistocene (~1.4 million years ago [Ma]) and this event was followed shortly after by the split of eastern and central lineages of *O. minutissimus* (~1.2 Ma). These divergence times agree with the Pleistocene diversification observed across most clades within the highly speciose acridid subfamily Gomphocerinae (Song et al., 2015). Additionally, the split time estimates are congruent with the crown age (<3 Ma) inferred for the recent radiation of the subgenus *Dreixius* based on mitochondrial DNA (García-Navas et al., 2017). The results of coalescent analyses also indicate that gene flow between *O. uhagonii* and *O. minutissimus* in the area of geographical overlap happened during a limited amount of time (~15,000 years) around the last glacial maximum (14–29 thousand years ago). During this period the two species engaged in gene flow at the same rate as the one estimated between the two lineages of *O. minutissimus*, indicating that interspecific gene flow, albeit low, was comparatively remarkable. The estimated timing of interspecific gene flow is compatible with the probable expansion of *O. uhagonii* to lower elevations during glacial periods. The shift to lower elevations might have put this species into extensive geographical contact with the more ubiquitous *O. minutissimus*, which presents many populations in foothills and valley bottoms where *O. uhagonii* is not present today. It must

be noted, however, that confidence intervals around point estimates for the time of interspecific gene flow are wide (Table 2), particularly for the onset of this period ( $T_{ADM2}$ ), and thus these results must be interpreted with extreme caution. Uncertainty in these estimates could in part be driven by heterogeneity in the timing and extent of hybridization among the different populations in the sympatric area, as suggested by the significant differences in the signal of genetic introgression observed among populations (see previous section).

### 4.3 | Inferred evolutionary scenario

Our genomic analyses point to a scenario in which *O. uhagonii* and ancestral populations of *O. minutissimus* probably diverged in allopatry followed by the split of *O. minutissimus* into two lineages, one of which came into secondary contact and hybridized with *O. uhagonii* during a limited period of time (Figure 5). Despite the fact that today the two species have ample opportunity to hybridize (i.e., a large proportion of their respective ranges currently overlap and several populations even co-occur in the Central System), our analyses did not find any evidence of contemporary hybridization,



**FIGURE 5** Schematic representation of the events documented in this study and the inferred biological processes. These correspond to the best-fitting demographic model (Model 5) for *Omocestus uhagonii* (US) and sympatric (MS) and allopatric (MA) populations of *O. minutissimus*. Vertical bars connecting MS and US represent historical gene flow. Note that geological reference time is not scaled and only point estimates inferred by FASTSIMCOAL2 are presented to simplify visualization

which suggests that reproductive isolation probably evolved after secondary contact and historical gene flow (Figure 5). Speciation of Gomphocerinae and other grasshoppers has been generally linked to allopatric divergence (Mayer et al., 2010), a process that in the specific case of montane/alpine species of Pleistocene origin was probably caused by the extensive fragmentation of ancestral populations driven by Quaternary climatic oscillations (e.g., Huang, Hill, Ortego, & Knowles, 2020; Knowles, 2000; Scattoni, Confalonieri, Lira-Noriega, Pietrovsky, & Cigliano, 2018). Different lines of evidence also point to allopatric divergence as the most plausible mode of speciation for *O. uhagonii* and *O. minutissimus*. First, the allopatric lineage of *O. minutissimus* from eastern Iberia presents a much larger distribution (Figure 1) and significantly higher levels of genomic diversity (one-way ANOVA:  $F_{1,13} = 12.05$ ,  $p = .004$ ; Table S1) and effective population sizes (nonoverlapping 95% CIs for  $N_e$  estimates in FASTSIMCOAL2 analyses; Table 2) than the sympatric lineage of *O. minutissimus* from the Central System. This supports the view that *O. minutissimus* probably originated in eastern Iberia and subsequently colonized the Central System, where it came into secondary contact with *O. uhagonii*. Second, the ecological niches of *O. minutissimus* and *O. uhagonii* are very similar and the two species present graminivorous feeding habits and the same microhabitat preferences (Clemente et al., 1991; J. Ortego, personal observation). This points to considerable niche conservatism, typical of allopatric speciation, and rejects sympatric speciation via disruptive ecological selection (e.g., Grace, Wisely, Brown, Dowell, & Joern, 2010). Finally, our analyses showed no evidence of gene flow between *O. uhagonii* and ancestral populations of *O. minutissimus* (Figure 4), indicating that the two species probably evolved in allopatry and only exchanged gene flow after secondary contact in the region where their distribution ranges overlap (see also Sankararaman et al., 2014; Sankararaman, Patterson, Li, Paabo, & Reich, 2012).

The footprints of historical introgression among currently sympatric populations of *O. uhagonii* and *O. minutissimus* and the lack of evidence for contemporary gene flow lead us to hypothesize an evolutionary scenario in which reproductive isolation evolved after historical hybridization in the area where the distribution ranges of the two taxa currently overlap. The observed differences in the levels of introgression among sympatric populations suggest that barriers to gene flow might have evolved multiple times or, alternatively, could reflect heterogeneity in the proportion of the genome of the other species retained after the interruption of interspecific gene flow due to differences among the studied populations in their demographic histories (e.g., bottlenecks; Amorim et al., 2017; Lawson, Van Dorp, & Falush, 2018; Quilodran, Nussberger, Montoya-Burgos, & Currat, 2019) or spatial variation in the strength of hypothetical purifying selection acting against introgressed alleles (Juric, Aeschbacher, & Coop, 2016; Petr, Paabo, Kelso, & Vernot, 2019). Alternative processes might have led to the evolution of reproductive isolation after secondary contact and hybridization. It has been frequently documented that interspecific gene flow can increase phenotypic and genomic divergence via the evolution of reproductive isolation and character displacement (Garner, Goulet, Farnitano,

Molina-Henao, & Hopkins, 2018; Hopkins, Levin, & Rausher, 2012; Pfennig & Pfennig, 2009). One of the many potential costs of hybridization is the ecological and genetic dysfunctions of hybrid offspring, which can reduce their fitness and drive reinforcement (Ortiz-Barrientos, Counterman, & Noor, 2004). In the reinforcement process, enhanced prezygotic isolation is favoured in sympatry in response to postzygotic isolation due to a strong selection against hybrids (Butlin, 1995; Coyne & Orr, 2004; Servedio & Noor, 2003). Selection for prezygotic isolation leads, in turn, to more divergent phenotypes and reproductive behaviours between species in sympatry than in allopatry (Moran, Zhou, Catchen, & Fuller, 2018). Thus, one possibility is that secondary contact and hybridization promoted the evolution of reproductive isolation via reinforcement, probably after an initial balance between dispersal and selection against hybrids in historical tension zones during which the two species experienced genetic exchange and introgression (Barton & Hewitt, 1985). An alternative explanation is that reproductive isolation evolved in geographical isolation as a consequence of genetic drift or as a fortuitous by-product of divergent selection on other traits (Coyne & Orr, 1989; Fitzpatrick, 2002; Sasa, Chippindale, & Johnson, 1998). The mosaic distribution of the two species in the Central System, with the presence of several sympatric populations but also large areas where the two species do not occur (e.g., eastern Central System and foothills; Figure 1), might have also provided ample opportunity for the evolution of reproductive isolation in geographically separated populations (Fitzpatrick, 2002).

## 5 | CONCLUSIONS AND FUTURE DIRECTIONS

The results of this study add to the growing body of evidence supporting that speciation-with-gene-flow is more prevalent in nature than formerly acknowledged (Nosil, 2008; Pinho & Hey, 2010; Roux et al., 2016). Our study system is very well suited to study the proximate mechanisms (e.g., reinforcement vs. genetic drift) that might have led to the evolution of reproductive isolation. Future research should focus on analysing mating preferences, phenotypic differentiation and reproductive character displacement (song, courtship behaviour, genitalic structures, etc.) between currently sympatric and allopatric populations of the two species (e.g., Butlin et al., 1991; Hollander, Smadja, Butlin, & Reid, 2013), determining mating success and the viability of offspring through experimental hybridization attempts in the laboratory (e.g., Coyne & Orr, 1989; Hoskin et al., 2005; Saldamando et al., 2005), and applying whole genome or transcriptome sequencing data to detect potential genomic signatures of reinforcement and/or identify loci that might be involved in reproductive isolation (Garner et al., 2018; Hopkins et al., 2012; Roda, Mendes, Hahn, & Hopkins, 2017).

### ACKNOWLEDGEMENTS

We are grateful to Amparo Hidalgo-Galiana for her valuable help during laboratory work and the grasshopper illustration. We also

thank to Víctor Noguerales, Conchi Cáliz and Pedro J. Cordero for their help during field and laboratory work, Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing, and two anonymous referees for constructive comments on an earlier version of the manuscript. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) and Laboratorio de Sistemas de Información Geográfica y Teledetección (LAST-EBD) from Estación Biológica de Doñana. We also thank the Centro de Supercomputación de Galicia (CESGA) and Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for access to computer resources. This study was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) (CGL2014-54671-P and CGL2017-83433-P). V.T. was supported by an FPI Predoctoral Fellowship (BES-2015-73159) from the Spanish Ministry of Economy and Competitiveness. During this work, A.P. and J.O. were supported by a Severo Ochoa (SEV-2012-0262) and a Ramón y Cajal (RYC-2013-12501) research fellowship, respectively.

#### AUTHOR CONTRIBUTIONS

V.T., A.P. and J.O. conceived and designed the study and analyses. J.O. collected the samples. V.T. performed the laboratory work and analysed the data guided by J.O. V.T. wrote the manuscript with help of J.O., and with inputs from A.P.

#### DATA AVAILABILITY STATEMENT

Raw Illumina reads have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA543714. Input files for all analyses are available for download on Figshare (<https://doi.org/10.6084/m9.figshare.12251600>).

#### ORCID

Vanina Tonzo  <https://orcid.org/0000-0002-5062-1070>

Joaquín Ortego  <https://orcid.org/0000-0003-2709-429X>

#### REFERENCES

- Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M., ... Godoy, J. A. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17, 251. <https://doi.org/10.1186/s13059-016-1090-1>
- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Amorim, C. E. G., Hofer, T., Ray, N., Föll, M., Ruiz-Linares, A., & Excoffier, L. (2017). Long-distance dispersal suppresses introgression of local alleles during range expansions. *Heredity*, 118(2), 135–142. <https://doi.org/10.1038/hdy.2016.68>
- Baiz, M. D., Tucker, P. K., & Cortes-Ortiz, L. (2019). Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular Ecology*, 28(5), 1056–1069. <https://doi.org/10.1111/mec.14966>
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16, 113–148. <https://doi.org/10.1146/annurev.ecolsys.16.1.113>
- Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, 341(6242), 497–503. <https://doi.org/10.1038/341497a0>
- Bridle, J. R., Baird, S. J. E., & Butlin, R. K. (2001). Spatial structure and habitat variation in a grasshopper hybrid zone. *Evolution*, 55(9), 1832–1843. [https://doi.org/10.1554/0014-3820\(2001\)055\[1832:sahvij\]2.0.co;2](https://doi.org/10.1554/0014-3820(2001)055[1832:sahvij]2.0.co;2)
- Britch, S. C., Cain, M. L., & Howard, D. J. (2001). Spatio-temporal dynamics of the *Allonemobius fasciatus*-*A. socius* mosaic hybrid zone: A 14-year perspective. *Molecular Ecology*, 10(3), 627–638. <https://doi.org/10.1046/j.1365-294x.2001.01215.x>
- Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, 99(3), 301–312. <https://doi.org/10.1038/sj.hdy.6800997>
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer.
- Butlin, R. K. (1995). Reinforcement – an idea evolving. *Trends in Ecology & Evolution*, 10(11), 432–434. [https://doi.org/10.1016/s0169-5347\(00\)89173-9](https://doi.org/10.1016/s0169-5347(00)89173-9)
- Butlin, R. K., Ritchie, M. G., & Hewitt, G. M. (1991). Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera, Acrididae). *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 334(1271), 297–308. <https://doi.org/10.1098/rstb.1991.0119>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Cigliano, M. M., Braun, H., Eades, D. C., & Otte, D. (2019). *Orthoptera Species File*. Version 5.0/5.0. Retrieved from <http://orthoptera.speciesfile.org/>
- Clemente, M. E., Garcia, M. D., & Presa, J. J. (1991). Los Gomphocerinae de la península Ibérica: 2. *Omocestus* Bolívar, 1878. (Insecta, Orthoptera, Caelifera). *Graellsia*, 46, 191–246.
- Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*, 43(2), 362–381. <https://doi.org/10.2307/2409213>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Association.
- de Manuel, M., Kuhlwilms, M., Frandsen, P., Sousa, V. C., Desai, T., Prado-Martinez, J., ... Marques-Bonet, T. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, 354(6311), 477–481. <https://doi.org/10.1126/science.aag2602>
- Dobzhansky, T. (1937). Genetic nature of species differences. *American Naturalist*, 71, 404–420. <https://doi.org/10.1086/280726>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eaton, D. A. R. (2014). PYRAD: Assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, 69(10), 2587–2601. <https://doi.org/10.1111/evo.12758>
- Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: *Orobanchaceae*). *Systematic Biology*, 62(5), 689–706. <https://doi.org/10.1093/sysbio/syt032>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A

- simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9(10), e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Fitzpatrick, B. M. (2002). Molecular correlates of reproductive isolation. *Evolution*, 56(1), 191–198. <https://doi.org/10.1111/j.0014-3820.2002.tb00860.x>
- Fitzpatrick, B. M., Fordyce, J. A., & Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, 22(11), 2342–2347. <https://doi.org/10.1111/j.1420-9101.2009.01833.x>
- Folk, R. A., Soltis, P. S., Soltis, D. E., & Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*, 105(3), 364–375. <https://doi.org/10.1002/ajb2.1018>
- García-Navas, V., Noguerales, V., Cordero, P. J., & Ortego, J. (2017). Ecological drivers of body size evolution and sexual size dimorphism in short-horned grasshoppers (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, 30(8), 1592–1608. <https://doi.org/10.1111/jeb.13131>
- Garner, A. G., Goulet, B. E., Farnitano, M. C., Molina-Henao, Y. F., & Hopkins, R. (2018). Genomic signatures of reinforcement. *Genes*, 9(4), 191. <https://doi.org/10.3390/genes9040191>
- Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, 23(18), 4555–4573. <https://doi.org/10.1111/mec.12811>
- Grace, T., Wisely, S. M., Brown, S. J., Dowell, F. E., & Joern, A. (2010). Divergent host plant adaptation drives the evolution of sexual isolation in the grasshopper *Hesperotettix viridis* (Orthoptera: Acrididae) in the absence of reinforcement. *Biological Journal of the Linnean Society*, 100(4), 866–878. <https://doi.org/10.1111/j.1095-8312.2010.01458.x>
- Graham, C. H., Ron, S. R., Santos, J. C., Schneider, C. J., & Moritz, C. (2004). Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, 58(8), 1781–1793. <https://doi.org/10.1155/03-274>
- Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F., & Petren, K. (2004). Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, 58(7), 1588–1599. <https://doi.org/10.1111/j.0014-3820.2004.tb01738.x>
- Gugger, P. F., & Cavender-Bares, J. (2013). Molecular and morphological support for a Florida origin of the Cuban oak. *Journal of Biogeography*, 40(4), 632–645. <https://doi.org/10.1111/j.1365-2699.2011.02610.x>
- Harrison, R. G. (1990). Hybrid zones: Windows on evolutionary process. *Oxford Surveys of Evolutionary Biology*, 7, 69–128.
- Hewitt, G. M. (1988). Hybrid zones - Natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, 3(7), 158–167. [https://doi.org/10.1016/0169-5347\(88\)90033-x](https://doi.org/10.1016/0169-5347(88)90033-x)
- Hill, J. G. (2015). Revision and biogeography of the *Melanoplus scudderi* species group (Orthoptera: Acrididae: Melanopliinae) with a description of 21 new species and establishment of the *Carnegiei* and *Davisi* species groups. *Transactions of the American Entomological Society*, 141(2), 252–350. <https://doi.org/10.3157/061.141.0201>
- Hollander, J., Smadja, C. M., Butlin, R. K., & Reid, D. G. (2013). Genital divergence in sympatric sister snails. *Journal of Evolutionary Biology*, 26(1), 210–215. <https://doi.org/10.1111/jeb.12029>
- Hopkins, R., Levin, D. A., & Rausher, M. D. (2012). Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution*, 66(2), 469–485. <https://doi.org/10.1111/j.1558-5646.2011.01452.x>
- Hoskin, C. J., Higgie, M., McDonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid allopatric speciation. *Nature*, 437(7063), 1353–1356. <https://doi.org/10.1038/nature04004>
- Howard, D. J., Waring, G. L., Tibbets, C. A., & Gregory, P. G. (1993). Survival of hybrids in a mosaic hybrid zone. *Evolution*, 47(3), 789–800. <https://doi.org/10.2307/2410184>
- Huang, J. P. (2016). Parapatric genetic introgression and phenotypic assimilation: Testing conditions for introgression between Hercules beetles (Dynastes, Dynastinae). *Molecular Ecology*, 25(21), 5513–5526. <https://doi.org/10.1111/mec.13849>
- Huang, J. P., Hill, J. G., Ortego, J., & Knowles, L. L. (2020). Paraphyletic species no more - genomic data resolve a Pleistocene radiation and validate morphological species of the *Melanoplus scudderi* complex (Insecta: Orthoptera). *Systematic Entomology*, in press. <https://doi.org/10.1111/syen.12415>
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322–1332. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>
- Jakobsson, M., & Rosenber, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Juric, I., Aeschbacher, S., & Coop, G. (2016). The strength of selection against Neanderthal introgression. *PLoS Genetics*, 12(11), e1006340. <https://doi.org/10.1371/journal.pgen.1006340>
- Kearns, A. M., Restani, M., Szabo, I., Schröder-Nielsen, A., Kim, J. A., Richardson, H. M., ... Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9, 906. <https://doi.org/10.1038/s41467-018-03294-w>
- Keightley, P. D., Ness, R. W., Halligan, D. L., & Hadrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, 196(1), 313–320. <https://doi.org/10.1534/genetics.113.158758>
- Key, K. H. L. (1968). Concept of stasipatric speciation. *Systematic Zoology*, 17(1), 14–22. <https://doi.org/10.2307/2412391>
- Knowles, L. L. (2000). Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54(4), 1337–1348. <https://doi.org/10.1111/j.0014-3820.2000.tb00566.x>
- Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: Glaciation signatures on contemporary patterns of genomic variation in collared pikas (*Ochotona collaris*). *Molecular Ecology*, 24(14), 3688–3705. <https://doi.org/10.1111/mec.13270>
- Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9, 3258. <https://doi.org/10.1038/s41467-018-05257-7>
- Lemmon, E. M., & Juenger, T. E. (2017). Geographic variation in hybridization across a reinforcement contact zone of chorus frogs (*Pseudacris*). *Ecology and Evolution*, 7(22), 9485–9502. <https://doi.org/10.1002/ece3.3443>
- Librado, P., & Rozas, J. (2009). DNASP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Lohse, K., Clarke, M., Ritchie, M. G., & Etges, W. J. (2015). Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution*, 69(5), 1178–1190. <https://doi.org/10.1111/evo.12650>
- Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science*, 302(5649), 1401–1404. <https://doi.org/10.1126/science.1089370>
- Maguilla, E., Escudero, M., Hipp, A. L., & Luceno, M. (2017). Allopatric speciation despite historical gene flow: Divergence and hybridization

- in *Carex furva* and *C. lucennoiberica* (Cyperaceae) inferred from plastid and nuclear RAD-seq data. *Molecular Ecology*, 26(20), 5646–5662. <https://doi.org/10.1111/mec.14253>
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229–237.
- Marques, I., Draper, D., Riofrio, L., & Naranjo, C. (2014). Multiple hybridization events, polyploidy and low postmating isolation entangle the evolution of neotropical species of *Epidendrum* (Orchidaceae). *BMC Evolutionary Biology*, 14, 20. <https://doi.org/10.1186/1471-2148-14-20>
- Mayer, F., Berger, D., Gottsberger, B., & Schulze, W. (2010). Non-ecological radiations in acoustically communicating grasshoppers? In M. Glaubrecht (Ed.), *Evolution in action* (pp. 451–464). Berlin, Germany: Springer-Verlag.
- Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.
- Moran, R. L., Zhou, M. C., Catchen, J. M., & Fuller, R. C. (2018). Hybridization and postzygotic isolation promote reinforcement of male mating preferences in a diverse group of fishes with traditional sex roles. *Ecology and Evolution*, 8(18), 9282–9294. <https://doi.org/10.1002/ece3.4434>
- Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., ... Jiggins, C. D. (2013). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, 22(3), 814–826. <https://doi.org/10.1111/j.1365-294X.2012.05730.x>
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, 17(9), 2103–2106. <https://doi.org/10.1111/j.1365-294X.2008.03715.x>
- Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography*, 41(10), 1895–1908. <https://doi.org/10.1111/jbi.12334>
- Ortego, J., Gugger, P. F., & Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist*, 218(2), 804–818. <https://doi.org/10.1111/nph.14951>
- Ortiz-Barrientos, D., Counterman, B. A., & Noor, M. A. F. (2004). The genetics of speciation by reinforcement. *PLoS Biology*, 2(12), 2256–2263. <https://doi.org/10.1371/journal.pbio.0020416>
- Papadopoulou, A., & Knowles, L. L. (2015). Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution*, 69(6), 1501–1517. <https://doi.org/10.1111/evo.12667>
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, 25(11), 2337–2360. <https://doi.org/10.1111/mec.13557>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Petr, M., Paabo, S., Kelso, J., & Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. *Proceedings of the National Academy of Sciences of the United States of America*, 116(5), 1639–1644. <https://doi.org/10.1073/pnas.1814338116>
- Pfennig, K. S., & Pfennig, D. W. (2009). Character displacement: Ecological and reproductive responses to a common evolutionary problem. *Quarterly Review of Biology*, 84(3), 253–276. <https://doi.org/10.1086/605079>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Pinho, C., & Hey, J. (2010). Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics*, 41, 215–230. <https://doi.org/10.1146/annurev-ecolsys-102209-144644>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., ... Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481), 43–49. <https://doi.org/10.1038/nature12886>
- Quilodran, C. S., Nussberger, B., Montoya-Burgos, J. I., & Currat, M. (2019). Hybridization and introgression during density-dependent range expansion: European wildcats as a case study. *Evolution*, 73(4), 750–761. <https://doi.org/10.1111/evo.13704>
- Ragge, D. R., & Reynolds, W. J. (1998). *The songs of the grasshoppers and crickets of Western Europe*. Essex, UK: Harley Books.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Roda, F., Mendes, F. K., Hahn, M. W., & Hopkins, R. (2017). Genomic evidence of gene flow during reinforcement in Texas Phlox. *Molecular Ecology*, 26(8), 2317–2330. <https://doi.org/10.1111/mec.14041>
- Rohde, K., Hau, Y., Kranz, N., Weinberger, J., Elle, O., & Hochkirch, A. (2017). Climatic effects on population declines of a rare wetland species and the role of spatial and temporal isolation as barriers to hybridization. *Functional Ecology*, 31(6), 1262–1274. <https://doi.org/10.1111/1365-2435.12834>
- Rosenberg, N. A. (2004). DISTRICT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Roux, C., Fraise, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14(12), e2000234. <https://doi.org/10.1371/journal.pbio.2000234>
- Saldamando, C. I., Tatsuta, H., & Butlin, R. K. (2005). Hybrids between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae) do not show endogenous postzygotic isolation. *Biological Journal of the Linnean Society*, 84(2), 195–203. <https://doi.org/10.1111/j.1095-8312.2005.000424.x>
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., ... Reich, D. (2014). The genomic landscape of Neandertal ancestry in present-day humans. *Nature*, 507(7492), 354–357. <https://doi.org/10.1038/nature12961>
- Sankararaman, S., Patterson, N., Li, H., Paabo, S., & Reich, D. (2012). The date of interbreeding between Neandertals and modern humans. *PLoS Genetics*, 8(10), e1002947. <https://doi.org/10.1371/journal.pgen.1002947>
- Sasa, M. M., Chippindale, P. T., & Johnson, N. A. (1998). Patterns of postzygotic isolation in frogs. *Evolution*, 52(6), 1811–1820. <https://doi.org/10.1111/j.1558-5646.1998.tb02258.x>
- Scattolini, M. C., Confalonieri, V., Lira-Noriega, A., Pietrovskiy, S., & Cigliano, M. M. (2018). Diversification mechanisms in the Andean grasshopper genus *Orotettix* (Orthoptera: Acrididae): Ecological niches and evolutionary history. *Biological Journal of the Linnean Society*, 123(4), 697–711. <https://doi.org/10.1093/biolinnean/bly008>
- Schenk, J. J., Kontur, S., Wilson, H., Noble, M., & Derryberry, E. (2018). Allopatric speciation drives diversification of ecological specialists on sandhills. *International Journal of Plant Sciences*, 179(4), 325–339. <https://doi.org/10.1086/697073>
- Schmitt, T. (2009). Biogeographical and evolutionary importance of the European high mountain systems. *Frontiers in Zoology*, 6, 9. <https://doi.org/10.1186/1742-9994-6-9>
- Seddon, J. M., Santucci, F., Reeve, N. J., & Hewitt, G. M. (2001). DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology*, 10(9), 2187–2198. <https://doi.org/10.1046/j.0962-1083.2001.01357.x>

- Servedio, M. R., & Kirkpatrick, M. (1997). The effects of gene flow on reinforcement. *Evolution*, 51(6), 1764–1772. <https://doi.org/10.2307/2410999>
- Servedio, M. R., & Noor, M. A. F. (2003). The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology Evolution and Systematics*, 34, 339–364. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132412>
- Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, 60, 561–588. <https://doi.org/10.1146/annurev.arplant.043008.092039>
- Song, H., Amédégato, C., Cigliano, M. M., Desutter-Grandcolas, L., Heads, S. W., Huang, Y., ... Whiting, M. F. (2015). 300 million years of diversification: Elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. *Cladistics*, 31(6), 621–651. <https://doi.org/10.1111/cda.12116>
- Stift, M., Kolar, F., & Meirns, P. G. (2019). STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity*, 123(4), 429–441. <https://doi.org/10.1038/s41437-019-0247-6>
- Takahashi, T., Nagata, N., & Sota, T. (2014). Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution*, 80, 137–144. <https://doi.org/10.1016/j.ympev.2014.07.016>
- Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2), 343–355. <https://doi.org/10.1111/j.1365-294X.2005.02794.x>
- Thome, M. T. C., & Carstens, B. C. (2016). Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8010–8017. <https://doi.org/10.1073/pnas.1601064113>
- Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology*, 28(17), 3869–3886. <https://doi.org/10.1111/mec.15189>
- Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, 28(12), 696–704. <https://doi.org/10.1016/j.tree.2013.09.001>
- Vera, M., Díez-del-Molino, D., & García-Marín, J. L. (2016). Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). *Molecular Ecology*, 25(5), 1089–1105. <https://doi.org/10.1111/mec.13545>
- Virdee, S. R., & Hewitt, G. M. (1994). Clines for hybrid dysfunction in a grasshopper hybrid zone. *Evolution*, 48(2), 392–407. <https://doi.org/10.2307/2410100>
- Wall, J. D., Yang, M. A., Jay, F., Kim, S. K., Durand, E. Y., Stevison, L. S., ... Slatkin, M. (2013). Higher levels of Neanderthal ancestry in east Asians than in Europeans. *Genetics*, 194(1), 199–209. <https://doi.org/10.1534/genetics.112.148213>

### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Tonzo V, Papadopoulou A, Ortego J. Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. *Mol Ecol*. 2020;00:1–15. <https://doi.org/10.1111/mec.15475>

