Instituto de Investigación
en Recursos Cinegéticos

CSIC - UCLM - JCCM

IREC

Instituto de Investigación
en Recursos Cinegéticos

Víctor Noguerales Rodríguez

Tesis Doctoral

2017

Understanding the processes of diversification along the speciation continuum in a recent evolutionary radiation of grasshoppers

Víctor Noguerales Rodríguez

Tesis Doctoral

Understanding the processes of diversification along the speciation continuum in a recent evolutionary radiation of grasshoppers

---

Procesos de diversificación a lo largo del continuo de especiación en una radiación evolutiva reciente de saltamontes

Víctor Noguerales Rodríguez

TESIS DOCTORAL - PH.D. THESIS

CIUDAD REAL, 2017

Instituto de Investigación en Recursos Cinegéticos (IREC)
(CSIC-UCLM-JCCM)

CITA RECOMENDADA:

Noguerales, V. (2017) *Understanding the processes of diversification along the speciation continuum in a recent evolutionary radiation of grasshoppers.* PhD Thesis. Instituto de Investigación en Recursos Cinegéticos, IREC (CSIC-UCLM-JCCM), Ciudad Real, España.

IMAGEN DE PORTADA:

Chorthippus binotatus binotatus (*C. binotatus*) ♀, Sª de San Vicente (Toledo).

Fotografía tomada por Pedro J. Cordero Tapia.

Understanding the processes of diversification along the speciation continuum in a recent evolutionary radiation of grasshoppers

---

Procesos de diversificación a lo largo del continuo de especiación en una radiación evolutiva reciente de saltamontes

Memoria presentada por el Licenciado Víctor Noguerales Rodríguez

para optar al grado de Doctor con Mención Internacional

por la Universidad de Castilla-La Mancha (UCLM)

El Doctorando	*Vº Bº del Director*	*Vº Bº del Director*
Víctor Noguerales Rodríguez	Dr. Joaquín Ortego Lozano	Prof. Dr. Pedro J. Cordero Tapia
Instituto de Investigación en Recursos Cinegéticos (IREC) CSIC-UCLM-JCCM	Estación Biológica de Doñana (EBD) CSIC	Instituto de Investigación en Recursos Cinegéticos (IREC) CSIC-UCLM-JCCM

Esta Tesis Doctoral ha sido financiada por el Ministerio de Economía y Competitividad (proyectos de investigación: CGL2011-25053, CGL2014-54671-P, CGL2016-80742-R), la Junta de Comunidades de Castilla-La Mancha y el Fondo Social Europeo (proyectos de investigación: POII10-0197-0167, PEII-2014023-P). Durante el desarrollo de mi Tesis Doctoral disfruté de una beca-contrato de Formación de Personal Investigador (FPI) (código de referencia: BES-2012-053741) y de diferentes ayudas para la realización de estancias breves en otros centros de investigaciones (códigos de referencia: EEBB-I-14-08642, EEBB-I-15-09299, EEBB-I-16-10960) concedidas por el Ministerio de Economía y Competitividad.

A mi madre, a mi padre,

a mi hermana,

y a mis abuelas

A Jorge Yepes

“Las ideas no duran mucho. Hay que hacer algo con ellas”

Santiago Ramón y Cajal (1852-1934)

ÍNDICE

RESUMEN 3

SUMMARY 7

INTRODUCCIÓN GENERAL 11

- RADIACIONES EVOLUTIVAS RECIENTES: LABORATORIOS NATURALES PARA EL ESTUDIO DE LA ESPECIACIÓN 14
- MODELO DE ESTUDIO: EL COMPLEJO DE ESPECIES CHORTHIPPUS GRUPO BINOTATUS 18
- JUSTIFICACIÓN Y OBJETIVOS GENERALES 22
- METODOLOGÍA 25
- ORGANIZACIÓN DE LA TESIS DOCTORAL 31

CAPÍTULO I 43
Integrative species delimitation in a recent evolutionary radiation of grasshoppers: evaluating the importance of number of loci and different phenotypic traits

CAPÍTULO II 81
Isolation and characterization of 18 polymorphic microsatellites in the grasshoppers Chorthippus group binotatus (Orthoptera: Acrididae) species complex

CAPÍTULO III 89
Inferring the demographic history of an oligophagous grasshopper: effects of climatic niche stability and host-plant distribution

CAPÍTULO IV 133
Testing the role of ancient and contemporary landscapes on structuring genetic variation in a specialist grasshopper

CAPÍTULO V 171
Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper

CAPÍTULO VI 205
The role of environment and core-margin effects on range-wide phenotypic variation of a montane grasshopper

DISCUSIÓN GENERAL 243

CONCLUSIONES 255

CONCLUSIONS 261

AGRADECIMIENTOS 267

RESUMEN

Un objetivo fundamental de la biología evolutiva es tratar de comprender los procesos que determinan la divergencia genética y fenotípica entre poblaciones, linajes y especies. Las radiaciones evolutivas recientes son un modelo de estudio ideal para abordar estas cuestiones debido a que las señales genéticas que dejan los procesos de divergencia no han sido borradas por el paso del tiempo y pueden ser utilizadas para inferir la historia evolutiva y demográfica de los diferentes taxones y linajes. Esta Tesis Doctoral adopta una aproximación multi-disciplinar para abordar de un modo integrador el estudio de los procesos evolutivos adaptativos y neutrales así como su interacción con el paisaje a lo largo de todo el espectro espacio-temporal en el que tiene lugar la divergencia genética y fenotípica. En particular, el objetivo de esta tesis es comprender la contribución relativa del ambiente, los eventos climáticos y la geografía en la distribución espacial de la variabilidad genética y fenotípica entre especies, linajes y poblaciones usando como modelo de estudio una radiación evolutiva reciente de saltamontes: el complejo de especies *Chorthippus* grupo *binotatus* (Acrididae: Gomphocerinae). Los análisis de delimitación de especies integrando información genómica y morfológica mostraron que el estatus taxonómico de todas las especies y subespecies es equivalente y que, por tanto, el complejo *Chorthippus* grupo *binotatus* estaría conformado por ocho taxones con estatus específico. A nivel específico, el estudio a escala filogeográfica de *Chorthippus binotatus binotatus* mostró que esta especie presenta cuatro grandes linajes mitocondriales que probablemente divergieron en alopatría como consecuencia de su aislamiento por barreras geográficas (estrecho de Gibraltar y Pirineos) y en diferentes refugios climáticos durante el Pleistoceno. La existencia de un linaje propio del centro y oeste de Francia indica la persistencia a largo plazo de la especie en áreas situadas cerca del actual límite septentrional de su rango de distribución, aunque estas poblaciones parecen haber experimentado severos cuellos de botella demográficos y presentan niveles muy reducidos de diversidad genética. A una escala espacial menor, los análisis de genética del paisaje revelaron que el aislamiento por resistencia definido por la complejidad topográfica fue el escenario que mejor explicó la diferenciación genética entre las poblaciones tanto de *C. binotatus binotatus* en el sureste de la península Ibérica como de *C. saulcyi moralesi* en los Pirineos. Aunque las condiciones ambientales no explicaron los patrones espaciales de diferenciación genética de *C. saulcyi moralesi*, la selección divergente mediada por gradientes climáticos sí parece tener un papel importante en los procesos de diferenciación fenotípica de esta especie. En particular, los resultados obtenidos indican que el grado de desarrollo de la tegmina está asociado a las condiciones ambientales locales. Este hecho sugeriría que la variación en las regiones genómicas implicadas en la expresión de este rasgo es el resultado de la respuesta adaptativa a diferentes regímenes climáticos. Por otro lado, aquellas poblaciones

localizadas en los márgenes del rango de distribución de la especie presentaron tegminas proporcionalmente más largas, un rasgo morfológico asociado a la capacidad de dispersión. Estas poblaciones periféricas presentaron una menor diversidad genética neutral que aquellas situadas en las zonas centrales de la distribución, lo que sugiere que las primeras están sometidas a mayores fluctuaciones demográficas y esto a su vez genera una selección direccional hacia fenotipos más dispersivos. En conjunto, los resultados de esta Tesis Doctoral ponen de manifiesto la importancia de combinar datos genéticos, fenotípicos y ambientales con el objetivo de conocer mejor los mecanismos evolutivos que operan a diferentes escalas espacio-temporales y promueven los procesos de divergencia a lo largo del continuo de especiación.

SUMMARY

Understanding the mechanisms that determine genetic and phenotypic divergence among populations, lineages and species is a paramount goal in evolutionary biology. Recent evolutionary radiations constitute a well-suited study system to address the study of such mechanisms because the genetic signatures left by divergence processes have not been erased by time and can be used to infer the evolutionary and demographic trajectories of lineages and species. This PhD thesis employs a multi-disciplinary approach in order to address the study of neutral and adaptive processes, and their links with landscape composition, throughout the whole spatiotemporal spectrum at which genetic and phenotypic divergence takes place. Specifically, the main objective of this PhD thesis is to understand the relative contribution of environment, climatic fluctuations and geography on the spatial patterns of genetic and phenotypic variation at the species, lineage and population levels, using as study model a recent evolution radiation of grasshoppers: the *Chorthippus* group *binotatus* species complex (Acrididae: Gomphocerinae). Our species delimitation analyses integrating genomic and morphological information revealed that the taxonomic status of species and subspecies is statistically equivalent and, hence, the *Chorthippus* group *binotatus* complex is composed of eight different biological entities that merit full species recognition. The study of *Chorthippus binotatus binotatus* at a range-wide phylogeographic scale shows that this taxon presents four main mitochondrial lineages that probably diverged in allopatry as a consequence of its isolation by geographic barriers (*i.e.* the Pyrenees) and in different climatic refugia during the Pleistocene. The existence of a lineage exclusive from western and central France indicates that the species long-term persisted in areas located at the present northern limit of its distribution range, although such populations were likely submitted to severe demographic bottlenecks and currently present very low levels of genetic diversity. At a lower spatial scale, landscape genetic analyses revealed that isolation by resistance defined by topographic complexity was the best-fitting scenario explaining population genetic differentiation of both *C. binotatus binotatus* in southeastern Iberia and *C. saulcyi moralesi* in the Pyrenees. Although environmental conditions did not explain the spatial patterns of genetic differentiation of *C. saulcyi moralesi*, we found that divergent selection mediated by climatic gradients had an important role on phenotypic differentiation processes in this taxon. In particular, our results indicate that the degree of development of forewings is associated to local environmental conditions. This fact suggests that variation in genomic regions involved in the expression of such trait may have resulted from an adaptive response to different climatic regimes. The populations located at the limits of the species distribution range presented proportionally longer forewings, a trait linked with dispersal capability. Such peripheral populations exhibited lower genetic diversity in comparison to those located at the core of the species distribution range, which suggests

that the former are subjected to considerable demographic instability and this leads to directional selection towards more dispersive phenotypes. Overall, the results from this PhD thesis highlight the importance of combining genetic, phenotypic and environmental data in order to better understand the evolutionary mechanisms that are at play at different spatiotemporal scales and drive divergence processes along the speciation continuum.

INTRODUCCIÓN GENERAL

Chorthippus binotatus binotatus (C. binotatus) ♂, Serra de Monchique (Portugal).
Fotografía tomada por Víctor Noguerales Rodríguez.

El estudio de la diversidad biológica ha suscitado desde hace siglos un gran interés en el ser humano. Con el tiempo, este interés ha avanzado más allá de la clasificación sistemática de las distintas formas biológicas y dado lugar a consolidadas disciplinas científicas como son la biología evolutiva y la biogeografía (CUADRO I). El análisis del papel que los mecanismos evolutivos neutrales (*i.e.* deriva genética) y la selección natural (*i.e.* adaptación local) tienen sobre los patrones espaciales de estructura genética y fenotípica resulta esencial para poder dilucidar los procesos que han dado lugar a la formación de linajes y especies y, por lo tanto, para comprender el origen de la enorme diversidad biológica de nuestro planeta (CUADRO I; Smith *et al*. 1997; Coyne y Orr 1998; de Queiroz 2007; Carnaval *et al*. 2009).

La obtención de información procedente de todo el espectro espacio-temporal en el que tienen lugar los procesos de diversificación de los organismos es necesaria para poder preservar la biodiversidad en todos sus niveles de organización, desde los ecosistemas y comunidades hasta los procesos evolutivos intra-específicos. Estos objetivos de conservación adquieren aún mayor trascendencia si tenemos en consideración que una gran proporción de las especies se extinguirán antes de ser descubiertas durante la que ya se considera la sexta extinción en masa (May 2002; ver también Costello *et al*. 2013). La conservación de la biodiversidad es, en todos sus sentidos, uno de los mayores retos globales al que nos enfrentamos. El diseño de estrategias rigurosas dirigidas a la conservación de la biodiversidad se sustenta no solo en el conocimiento exhaustivo de la distribución de las especies, sino también en definir adecuadamente su estatus taxonómico (Padial *et al*. 2010; Fujita *et al*. 2012; Huang y Knowles 2016), descubrir linajes desconocidos y determinar en qué medida éstos presentan trayectorias evolutivas idiosincráticas que merecen ser preservadas (Moritz 2002; Yannic *et al*. 2014). En este contexto cobra especial relevancia el estudio de las interacciones entre el paisaje y los procesos evolutivos, ya que esta aproximación nos permite discriminar cuales son los factores que generan los patrones espaciales de diversidad biológica a nivel de genes, linajes y especies, tanto a escala local como regional (He *et al*. 2013; Andrew *et al*. 2012). Comprender mejor estos procesos evolutivos a diferentes escalas espacio-temporales resulta de gran interés ante el actual escenario de cambio global, ya que nos permitirá predecir las respuestas de las especies y las poblaciones a dichos cambios ambientales inducidos por las actividades humanas (Espindola *et al*. 2014; Moritz y Agudo 2013) e identificar aquellas áreas que potencialmente podrían actuar como refugios (Carnaval *et al*. 2009; Devitt *et al*. 2012; Yannic *et al*. 2014; Prates *et al*. 2016).

CUADRO I

- EL ESTUDIO DE LA SELECCIÓN NATURAL A LO LARGO DE LA HISTORIA -

Cualquiera que haya dedicado algo de tiempo a la observación cuidadosa de la naturaleza, habrá podido percatarse que una de las principales características que la definen es la diversidad de las entidades biológicas que la componen, así como la presencia o ausencia de unas formas u otras en función de la localización analizada. Si llevado por sus inquietudes científicas, ese observador se hubiera finalmente aventurado al estudio pormenorizado de algún grupo concreto de organismos, seguramente habrá sucumbido a la tentación de intentar responder a la cuestión relativa al por qué de aquella inexplicable variación biológica y su caprichoso patrón de distribución geográfico.

Desde la antigüedad, la apreciación de esta diversidad y sus cambios fue objeto de estudio y dedicación por parte de filósofos griegos como Aristóteles (384-322 a.C.) o pensadores taoístas como Zhuangzi (~369-290 a.C.). Sin embargo, no fue hasta el siglo XIX cuando gracias al trabajo de importantes científicos (Lamarck, 1744-1829; von Humboldt, 1769-1859; Darwin, 1809-1882; Wallace, 1823-1913) se empezaron a esbozar las cuestiones más básicas acerca de la diversidad de los organismos, su distribución y su evolución: *¿Cuáles son los procesos que han dado lugar a dicha variación biológica?*

La publicación en 1859 de *El Origen de las Especies* por parte de Darwin supuso el establecimiento del ideario de la biología evolutiva al introducir formalmente que la evolución de las poblaciones a través de las generaciones se debe a un proceso conocido como selección natural. No obstante, la importancia de la selección natural en la evolución de los organismos no siempre ha gozado de la aceptación y reconocimiento por la comunidad científica que posee en la actualidad. No es hasta 1900, con el redescubrimiento de los trabajos sobre herencia genética publicados por Mendel (1822-1884), cuando se reactiva la línea de pensamiento que situaría a la selección natural como el principal motor de cambio de los organismos. Posteriormente, durante las décadas de 1930 y 1940, científicos como Fischer (1890-1962), Haldane (1892-1964), Dobzhansky (1900-1975) y Wright (1889-1988) integran las denominadas Leyes de Mendel, el concepto Darwiniano de selección natural y los avances en genética de poblaciones para dar lugar a la denominada *Síntesis Evolutiva Moderna*. Es en ese momento cuando se asientan las bases del estudio de los procesos evolutivos, al establecerse que las mutaciones son la fuente de la variabilidad genética y que los cambios en las frecuencias alélicas poblacionales entre generaciones son el resultado de la deriva genética, el flujo genético y la selección natural. Años más tarde, Kimura (1924-1994) propone en 1968 la Teoría Neutralista donde atribuye un papel relevante a la deriva genética frente a la selección natural como determinante de la evolución adaptativa.

(continua en la página siguiente)

CUADRO I

- EL ESTUDIO DE LA SELECCIÓN NATURAL A LO LARGO DE LA HISTORIA -

(continuación)

El descubrimiento del ADN, el advenimiento de las técnicas moleculares y el desarrollo de teorías que permiten describir la genealogía de los alelos e inferir el ancestro común más reciente a dos variantes génicas (Teoría de la Coalescencia; Kingman 1982), ha supuesto una revolución a la hora de estudiar los procesos que determinan las trayectorias evolutivas y demográficas de las poblaciones, linajes y especies. Más recientemente, los avances procedentes de otras disciplinas como la biología del desarrollo, la genómica, la epigenética y la ecología están permitiendo que el marco conceptual de la teoría evolutiva haya comenzado a expandirse para dar cabida a otros fenómenos involucrados en mayor o menor grado en el proceso de evolución de las formas biológicas, lo que está dando lugar a la formulación de la *Síntesis Evolutiva Extendida* (para una descripción del avance en este campo, ver Laland y Wray 2014; Laland *et al*. 2015).

RADIACIONES EVOLUTIVAS RECIENTES: LABORATORIOS NATURALES PARA EL ESTUDIO DE LA ESPECIACIÓN

El estudio de radiaciones evolutivas recientes (*i.e.* conjunto de especies que han divergido de un ancestro común recientemente) es especialmente atractivo para investigar los mecanismos que dan lugar a la formación de especies y linajes debido a que las señales genéticas resultantes de dichos procesos no han sido erosionadas por el paso del tiempo (Knowles 2000; Leaché *et al*. 2009; Rundell y Price 2009; Losos y Mahler 2010; Mayer *et al*. 2010). Esta diversificación puede ser el resultado de (i) un proceso de especiación ecológica, donde la disrupción del flujo genético entre poblaciones tiene su origen en la selección divergente en respuesta a factores ambientales o ecológicos, (ii) procesos de divergencia primordialmente neutrales promovidos por aislamiento geográfico (*i.e.* especiación no ecológica), (iii) o una combinación de ambos mecanismos (Graham *et al*. 2004; Sobel *et al*. 2010). La importancia relativa de las distintas barreras de aislamiento postuladas como motores del proceso de especiación es un debate aún muy activo en la comunidad científica (Coyne y Orr 2004; Thorpe *et al*. 2008; Rundell y Price 2009; Butlin *et al*. 2012; ver meta-análisis en Funk *et al*. 2006).

El estudio de los patrones de solapamiento de las áreas de distribución y el espacio ecológico que ocupan las diferentes especies hermanas objeto de estudio puede ayudar a inferir los mecanismos (modos) de especiación (Schluter 2001; Graham *et al*. 2004; Fig. 1). El proceso de especiación puede haberse desarrollado en simpatría como consecuencia de la selección divergente que resulta de la adaptación a diferentes condiciones ambientales o ecológicas (Fig. 1: escenario A), un fenómeno que ocurriría en ausencia de barreras físicas al flujo genético (Schluter 2009; Schluter y Conte 2009). El estudio de la especiación simpátrica ha recibido gran atención por parte de los biólogos evolutivos y su papel en la formación de nuevas especies ha sido aceptada desde un punto de vista tanto teórico (*ej.* Dieckman y Dobeli 1999; Kondrashov y Kondrashov 1999) como empírico (*ej.* Barluenga *et al*. 2006; Ryan *et al*. 2007). En el otro extremo del espectro geográfico-ambiental se encontraría el modo de especiación alopátrica, que predice que las especies han alcanzado su aislamiento reproductivo en condiciones de aislamiento geográfico (Fig. 1). Diferentes eventos geológicos y climáticos han sido documentados como responsables de la fragmentación y expansión/contracción de los rangos de distribución de los organismos, hechos de gran importancia para la formación de nuevas especies debido a su capacidad para generar nuevas oportunidades geográficas y ecológicas (Pepper *et al*. 2011a,b). Un ejemplo de estos eventos serían las fluctuaciones climáticas acontecidas durante el

Pleistoceno (~2,0-0,4 millones de años), las cuales han producido cambios distribucionales, dado lugar a procesos de aislamiento geográfico y promovido la diversificación alopátrica de numerosos grupos de organismos (CUADRO II; Hewitt 2000; Knowles y Richards 2005). En un escenario de vicarianza las especies resultantes pueden ocupar un espacio ambiental similar (Fig. 1: escenario F) o bien presentar diferenciación en sus nichos ecológicos (Fig. 1: escenario C) (Graham *et al*. 2004). En este último caso, la divergencia de nicho ecológico puede haberse generado por mecanismos estocásticos (deriva genética) y/o deterministas como resultado de la especialización hacia algún componente ambiental debido a procesos de selección adaptiva una vez completado el aislamiento reproductivo (Fig. 1: escenarios B-C). Finalmente, en condiciones parapátricas (*i.e.* los rangos de distribución de las especies son contiguos o se solapan levemente) y de ocupación del mismo nicho ecológico, la especiación puede ser el resultado de un proceso inicial de divergencia alopátrica no ecológica al que ha seguido una expansión del área de distribución de una o ambas especies (Fig. 1: escenario E).

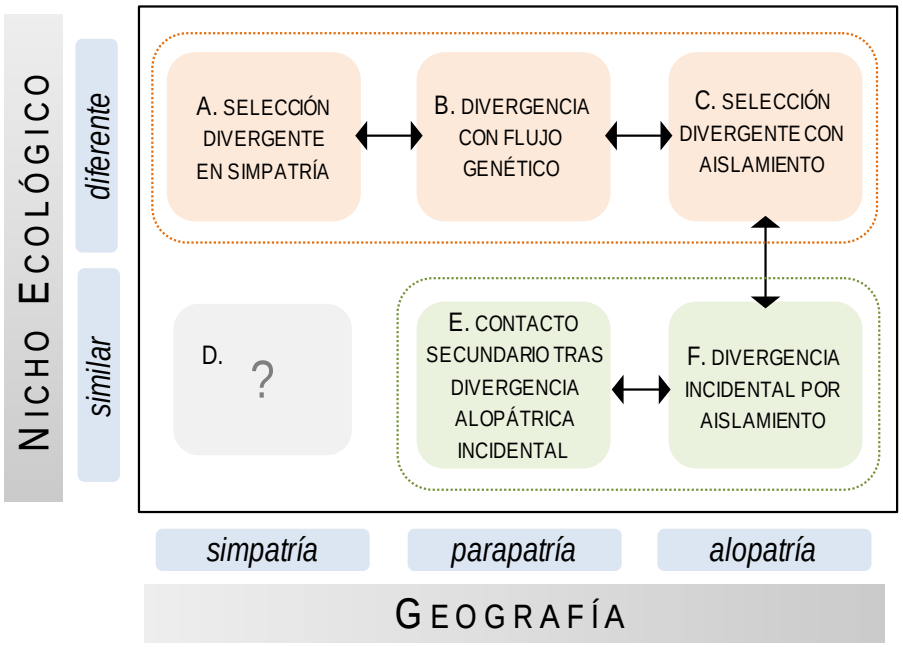


Figura 1 – Resumen de los diferentes mecanismos de especiación en función del escenario geográfico y ecológico/ambiental en el que acontecen los procesos de divergencia. Modificado de Graham *et al*. (2014).

A una escala micro-evolutiva, la diferenciación genética entre poblaciones puede ser mediada por el aislamiento geográfico y la divergencia ecológica, procesos análogos a los anteriormente descritos para explicar los diferentes mecanismos de especiación. A pesar de que la geografía parece jugar un papel primordial en la diferenciación inter-poblacional (aislamiento por distancia, *i.e*. incremento de la diferenciación genética entre poblaciones con el aumento de la distancia geográfica que las separa; en inglés IBD, *isolation-by-distance*; Wright 1943), la divergencia también puede producirse a una escala espacio-temporal reducida cuando la heterogeneidad ambiental y gradientes ecológicos acusados desencadenan selección frente a genotipos no adaptados localmente (Wang *et al*. 2013; Soria-Carrasco *et al*. 2014; Wang y Bradburd 2014). Ambos procesos no son excluyentes y en general participan conjuntamente en el modelado de los patrones de flujo genético entre poblaciones (Shaffer y Wolf 2013; Sexton *et al*. 2014). La creciente atención que se ha dedicado al fenómeno anteriormente descrito, denominado aislamiento por ambiente (IBE, *isolation-by-enviroment, i.e.* incremento de la diferenciación genética entre poblaciones con el aumento de su disimilitud ambiental o ecológica), ha ido paralela al avance conceptual y metodológico de la disciplina acuñada como "genética del paisaje" (*landscape genetics*) (Manel *et al*. 2003; Manel y Holderegger 2013), la cual incorpora la complejidad y heterogeneidad del territorio para identificar barreras y corredores e inferir los patrones espaciales de flujo genético entre-poblaciones (aislamiento por resistencia; IBR, *isolation-by-resistance, i.e.* incremento de la diferenciación genética entre poblaciones a medida que aumenta la distancia entre las mismas calculada considerando la resistencia que diferentes componentes del paisaje ofrecen a la dispersión; McRae 2006; McRae y Beier 2007; McRae *et al.* 2008).

Por todo lo anterior, podemos imaginar el proceso de especiación como un continuo (*speciation continuum*) desde la divergencia completa entre especies y la formación de linajes infra-específicos con una historia evolutiva propia hasta la aparición de diferencias genéticas y fenotípicas sutiles mediadas por fenómenos de adaptación local o aislamiento geográfico de las poblaciones (Fig.1; Nosil *et al*. 2009; Nosil 2012; Safran *et al*. 2016). En este mismo escenario, el objetivo general de esta tesis es abordar de un modo integrador el estudio de los procesos que ocurren a diferentes escalas espacio-temporales para entender mejor la contribución relativa del ambiente y el aislamiento geográfico (*i.e.* selección natural, deriva genética) en la diferenciación genética y fenotípica a lo largo de todo el espectro en el que tiene lugar la especiación (Hendry *et al*. 2009; Butlin *et al*. 2014; Supple *et al*. 2015).

CUADRO II

- EL IMPACTO DE LAS FLUCTUACIONES CLIMÁTICAS DEL PLEISTOCENO -

Las oscilaciones climáticas acontecidas durante el Pleistoceno afectaron extensos territorios de gran parte del mundo (Hewitt 2000, 2004a,b; Allen *et al*. 2008). En Europa, la última glaciación alcanzó el máximo de cobertura de hielo hace aproximadamente 21000 años, retirándose la mayor parte de éste a latitudes más septentrionales hace 8000-15000 años. Los ciclos glaciales y post-glaciales tuvieron un importante efecto sobre la biodiversidad y el actual patrón de distribución de las especies (Hewitt 2000, 2004a,b). El paradigma clásico establece que durante el avance glacial numerosas especies sobrevivieron acantonadas en refugios libres de hielo. La combinación de la deriva genética y la aparición de nuevas mutaciones generaron altos niveles de divergencia genética entre poblaciones aisladas en diferentes refugios. Durante los periodos inter-glaciales, las especies comenzaron su expansión desde sus diferentes refugios, recolonizando nuevos hábitats disponibles y dando lugar a fenómenos de hibridación entre poblaciones previamente aisladas. Las penínsulas situadas en el suroeste del continente europeo (penínsulas Ibérica e Itálica) supusieron un importante refugio para las especies de climas templados y a su vez una "trampa" para su dispersión hacia el norte durante los periodos inter-glaciales debido a las barreras montañosas en sentido este-oeste que las flanquean (Hewitt 1999).

Uno de los ejemplos más conocidos acerca del papel de los ciclos glaciales del Pleistoceno sobre la trayectoria evolutiva de las especies es el caso del saltamontes *Pseudochorthippus paralellus*. Esta especie exhibe un marcado patrón de subdivisión genética asociado a sus refugios en las penínsulas Ibérica, Itálica y Balcánica que ha permitido trazar sus rutas de recolonización post-glacial hacia el resto de Europa (Cooper *et al*. 1995; Lunt *et al*. 1998). Se ha documentado además que la expansión post-glacial del rango de distribución de esta especie ha generado poblaciones híbridas en las zonas de contacto (*ej.* Pirineos o Alpes) entre linajes procedentes de antiguos refugios climáticos (Hewitt 1999; Bridle *et al*. 2001). Sin embargo, aunque la hipótesis clásica del "refugio glacial" parece explicar adecuadamente el patrón espacial de estructuración genética observado en numerosas especies con requerimientos termófilos, el efecto de los ciclos glaciales sobre las especies montanas o adaptadas al frío parece no estar tan claro (Parducci *et al*. 2012). A su vez, la presencia de poblaciones disyuntas en el límite septentrional de las áreas de distribución de ciertas especies sugiere que su origen probablemente no está vinculado a colonizaciones post-glaciales, sino relacionado con su persistencia en refugios norteños hasta ahora desconocidos para un gran número de especies (Stewart y Lister 2001; Stewart *et al*. 2010; Salvi *et al*. 2013). Por todo lo anterior, estas dinámicas climáticas ofrecen un marco excepcional para investigar cómo la composición paisajística presente y pasada han modelado el flujo y estructura genética de las especies en función de sus requerimientos ecológicos y climáticos (Velo-Antón *et al*. 2013; Fordham *et al*. 2014; Yannic *et al*. 2014).

MODELO DE ESTUDIO: EL COMPLEJO DE ESPECIES C*horthippus* GRUPO *binotatus*

El *Chorthippus* grupo *binotatus* es un complejo de especies perteneciente al subgénero *Glyptobothrus* y la familia *Gomphocerinae* cuyo origen se estima que ha tenido lugar hace unos tres millones de años (Mayer *et al*. 2010; Nattier *et al*. 2011; García-Navas *et al*. 2017). La taxonomía de este complejo ha ido variando notablemente a lo largo de los últimos 70 años (CUADRO III), pero sus relaciones filogenéticas y el estatus taxonómico de sus diferentes linajes no se ha resuelto todavía (Defaut 2011, 2015). De acuerdo a estudios morfométricos, Defaut (2011, 2015) establece que este complejo de especies comprende dos especies nominativas y ocho subespecies distribuidas en el suroeste de Europa (España, Francia y Portugal) y norte de África (Marruecos) (Fig. 2). La especie *Chorthippus binotatus* (Charpentier 1825) es la que presenta la mayor área de distribución y engloba tres subespecies que probablemente hayan divergido en alopatría considerando sus actuales distribuciones (*C. b. binotatus* y *C. b. armoricanus* Defaut 2015 en Europa y *C. b. atlasi* Defaut 1987 en Marruecos). Esta especie se alimenta exclusivamente de determinadas especies de leguminosas arbustivas pertenecientes a la tribu *Genisteae* (Talavera *et al*. 2001), con poblaciones asentadas sobre comunidades de estas plantas compuestas generalmente por una o dos especies del mismo o distinto género (Defaut 2011, 2015; Bétard 2016). La subespecie *C. b. armoricanus* ha sido catalogada como amenazada en la Lista Roja de Francia debido al reducido tamaño y alto grado de fragmentación de sus poblaciones (Pratz y Cloupeau 2010; Bétard 2016). La otra especie nominativa, *Chorthippus saulcyi* (Krauss 1888), constituye en sí misma un complicado grupo taxonómico compuesto por cinco subespecies, las cuales son consideradas micro-endemismos con áreas de distribución restringidas a diferentes macizos montañosos europeos (CUADRO III). Mientras *C. s. algoaldensis* Chopard 1952 y *C. s. daimei* (Azam 1893) se distribuyen disyuntamente en el Macizo Central francés y los Alpes Marítimos, respectivamente, las otras tres subespecies (*C. s. moralesi* Uvarov 1954*, C. s. saulcyi* (Krauss 1888) y *C. s. vicdessossi* Defaut 2011) se distribuyen en parapatría/peripatría a lo largo de los Pirineos y el noreste ibérico (Fig. 2).

Las diferentes especies y subespecies que integran el *Chorthippus* grupo *binotatus* exhiben un gradiente de diferenciación en determinados rasgos morfológicos. Encontramos desde taxones con alas bien desarrolladas (macrópteros, como *C. b. atlasi*, *C. b. armoricanus*, *C. b. binotatus*) hasta especies con las alas reducidas (braquípteros, como *C. s. daimei* y *C. s. moralesi*), así como formas

intermedias situadas a lo largo de un eje continuo de diferenciación fenotípica (*C. s. saulcyi* y *C. s. vicdessossi*). Este grupo se configura, por tanto, como un sistema de estudio ideal para analizar la importancia de los factores geográficos y ecológicos en los mecanismos de diversificación que ocurren a diferentes escalas espacio-temporales, es decir, desde los procesos micro-evolutivos que determinan la diferenciación genética/fenotípìca entre poblaciones hasta aquellos que dan lugar a la formación de linajes y especies.

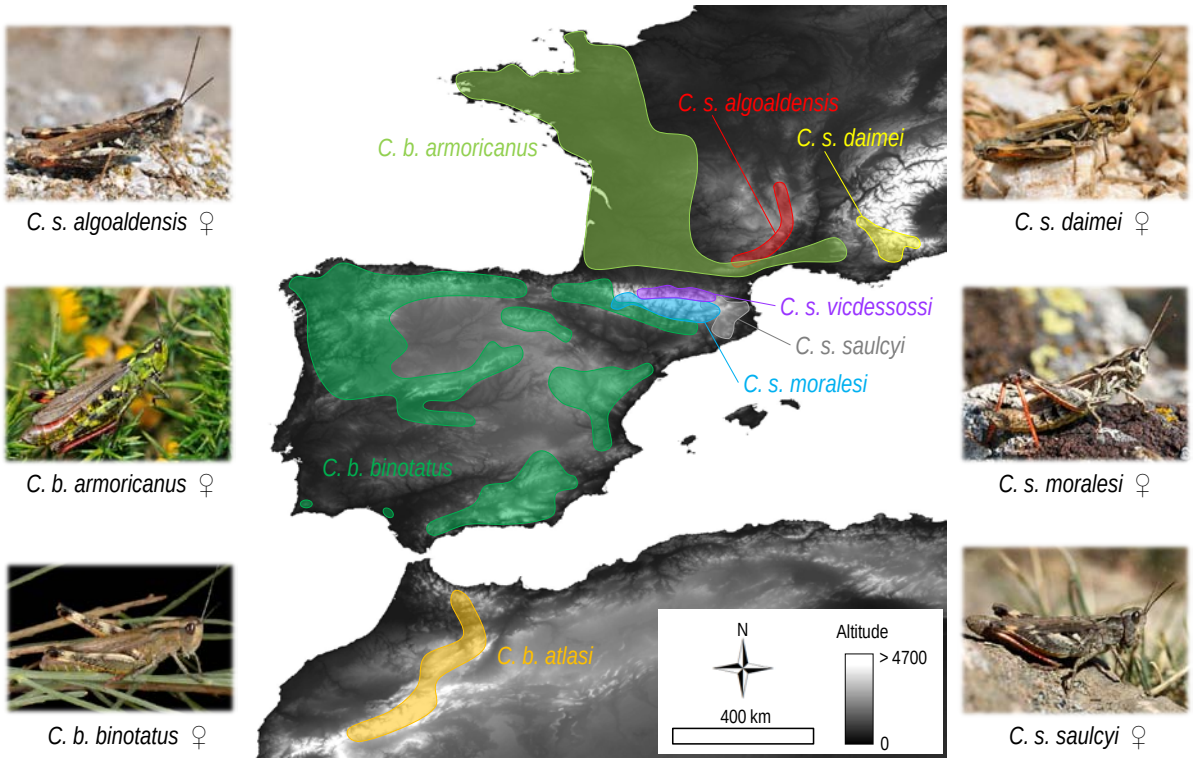


Figure 2 – Distribución aproximada de las distintas especies/subespecies putativas que conforman el complejo *Chorthippus* grupo *binotatus* (*sensu* Defaut 2011, 2015). Se muestran fotografías de algunas de las subespecies donde se pueden apreciar sus diferencias en la longitud de la tegmina. La fotografía del taxón *C. b. binotatus* fue tomada por Pedro J. Cordero, mientras que las fotografías de los restantes taxones fueron tomadas por David Morichón (disponibles en Defaut 2011; Cigliano *et al*. 2017, Orthoptera Species File, *http://orthoptera.speciesfile.org/).*

CUADRO III

- LA CONTROVERTIDA HISTORIA TAXONÓMICA DEL *Chorthippus* GRUPO *binotatus* -

Desde la descripción en 1825 de la especie nominativa que da nombre al grupo, diferentes autores han abordado el estudio de la variación fenotípica que caracteriza al complejo de especies *Chorthippus* grupo *binotatus*. Estos trabajos han dado lugar a sucesivas modificaciones en la taxonomía del grupo originadas por las diferentes interpretaciones otorgadas al grado de solapamiento que presentan sus distintas entidades biológicas en diversos caracteres fenotípicos (ver Defaut 2011 para una revisión exhaustiva).

Chopard describió en 1952 dos nuevos taxones (*algoaldensis* y *reyi*) en el Macizo Central francés considerándolos como miembros infra-especificos de las dos especies nominativas *Chorthippus binotatus* y *Chorthippus saulcyi*, respectivamente (Chopard 1952). Posteriormente, Harz (1975) en su conocida obra *Die Orthopteren Europas* determinó que las subespecies *algoaldensis* y *reyi* eran sinónimas, incluyendo a la subespecie *algoaldensis* dentro de la especie nominativa *C. binotatus*, bajo la que agregó a su vez las seis formas infra-específicas descritas hasta ese momento (*algoaldensis*, *binotatus*, *daimei*, *dilutus*, *moralesi* y *saulcyi*). Más tarde, García y colaboradores establecieron en 1995 que la subespecie *Chorthippus binotatus dilutus* (descrita por Ebner en 1941 en la Sierra de Guadarrama) era efectivamente sinónima de la subespecie *binotatus* basándose en las similitudes de numerosos caracteres morfológicos y perfiles acústicos de sus cantos (García *et al*. 1995). Resulta anecdótico mencionar la descripción de una especie endémica de la Sierra de Gredos y que fue descrita como *Chorthippus ariasi* (Bolívar 1908) basándose en un solo ejemplar recolectado por un colaborador de Bolívar a principios del siglo XX. Sin embargo, el posterior estudio de las anotaciones tomadas por el mismo Bolívar concluye que dicho individuo pertenecería a la especie *Chorthippus binotatus binotatus* (Presa *et al*. 2016), tan abundante en los piornales y erizonales gredenses de *Cytisus oromediterraneus* y *Echinospartum bardanessi*.

La diversidad del complejo de especies continuó incrementándose con el descubrimiento de un nuevo taxón fuera de Europa. En 1987, Defaut descubre en el Atlas Medio marroquí una forma que presenta diferencias morfológicas con la especie nominativa *C. binotatus* y que es descrita como una subespecie de la misma (*C. binotatus atlasi*) (Defaut 1987). Tres años más tarde, Ragge y Reynolds (1998) apoyan el esquema taxonómico propuesto por Harz (1975) pero simplifican aún más la taxonomía del grupo estableciendo que todas las subespecies descritas no son más que variedades locales pertenecientes a una sola especie nominativa ampliamente distribuida (*C. binotatus*). Esta simplificación queda reflejada en el *Catálogo de Ortópteros Caelifera de la Península Ibérica y Baleares* (Presa *et al*. 2007) donde toda la diversidad del grupo queda reducida a una sola especie (*C. binotatus*) sin entidades infra-específicas reconocidas.

(continua en la página siguiente)

CUADRO III

- LA CONTROVERTIDA HISTORIA TAXONÓMICA DEL *Chorthippus* GRUPO *binotatus* -

(continuación)

Recientemente, Defaut ha recuperado el estatus infra-específico para todas las subespecies descritas y aceptadas tras llevar a cabo un exhaustivo estudio que analiza un amplio conjunto de caracteres fenotípicos (morfológicos, biométricos, comportamentales y corológicos) (Defaut 2011). En este trabajo se establece que todas las subespecies estarían organizadas en torno a dos especies nominativas (*C. binotatus* y *C. saulcyi*) (Defaut 2011). A su vez, Defaut (2011, 2015) describe en base a caracteres morfológicos dos nuevas subespecies: *C. b. armoricanus* distribuida por el oeste de Francia y *C. s. vicdessossi* distribuida en los Pirineos franceses pertenecientes a sendas especies nominativas (*C. binotatus* y *C. saulcyi*, respectivamente) (Cigliano *et al*. 2017).

Hasta el momento, la descripción y la determinación del estatus taxonómico de estos linajes se ha realizado en base a datos exclusivamente fenotípicos (Defaut 2011) y empleando criterios poco consensuados y difícilmente reproducibles (Yeates *et al*. 2011). Esta problemática se extiende a cualquier grupo de organismos y ha generado un intenso debate en la literatura especializada acerca de la necesidad de fundamentar la delimitación taxonómica en el uso de diferentes tipos de datos (genéticos y no genéticos) integrados en una aproximación estadística que mediante análisis cuantitativos permita contrastar explícitamente hipótesis sobre los límites de las especies (Wiens 2007; Padial 2010; Yeates *et al*. 2011; Carstens *et al*. 2013). La presente Tesis Doctoral se hace eco de este debate científico e intenta contribuir al avance del campo de la delimitación integrativa de especies mediante el uso de datos genómicos y fenotípicos bajo el marco conceptual desarrollado en una reciente aproximación analítica (Solís-Lemus *et al*. 2015).

JUSTIFICACIÓN Y OBJETIVOS GENERALES

Pocos estudios han abordado de un modo integrador el estudio de los procesos de divergencia genómica a lo largo del continuo de especiación con el propósito de obtener una visión amplia de los mecanismos micro- y macro-evolutivos que determinan la diversidad biológica (Powell *et al*. 2013; Supple *et al*. 2015). Los estudios que han tratado esta cuestión raramente la han abordado desde un punto de vista multi-disciplinar (Fig. 3) que integre diferentes tipos de datos y aproximaciones conceptuales y analíticas (Balkenhol *et al*. 2009; Wang *et al*. 2013; Ruiz-González *et al*. 2015; Ferrer *et al*. 2016). En este sentido, la literatura científica demanda crecientemente la realización de estudios comparativos que examinen múltiples taxones con rasgos específicos y diferentes requerimientos ecológicos con el objetivo de generar respuestas integradoras y más robustas sobre los mecanismos evolutivos involucrados en la divergencia genómica (Hickerson y Cunningham 2005; Ortego *et al*. 2015; Papadopoulou y Knowles 2016).

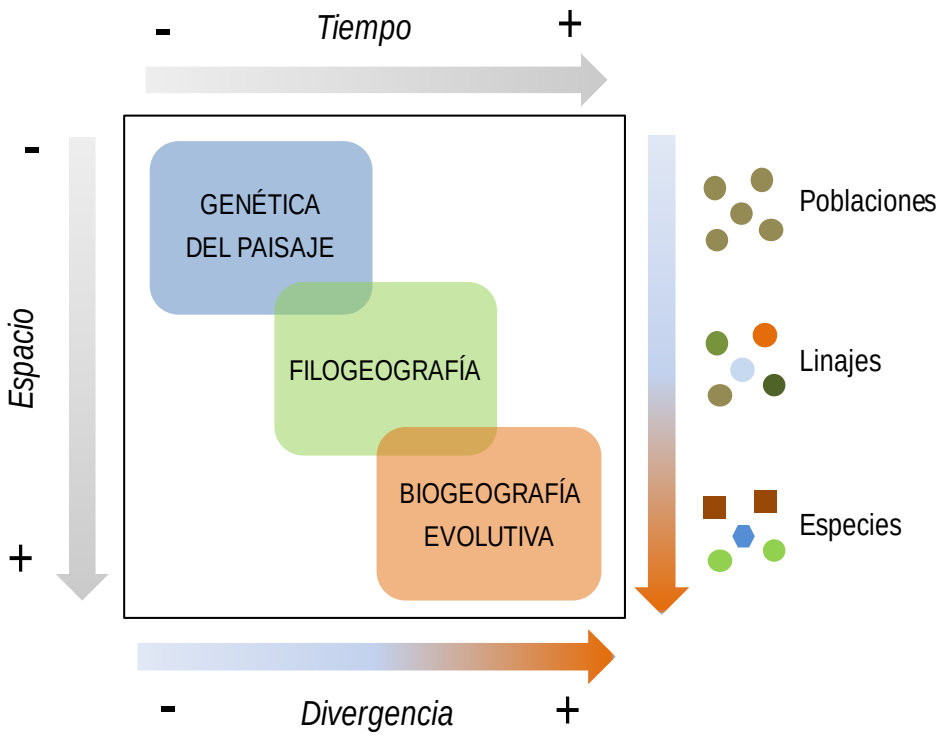


Figura 3 – Esquema de la escala espacio-temporal que abordan las distintas disciplinas sobre cuyas aproximaciones se fundamenta el desarrollo de esta Tesis Doctoral.

Las anteriores motivaciones justifican el desarrollo de la presente Tesis Doctoral, la cual integra conceptos y metodologías procedentes de disciplinas tales como la taxonomía integrativa, biogeografía evolutiva, la filogeografía y la genética del paisaje para abordar el estudio de los mecanismos evolutivos neutrales y/o adaptativos que promueven la divergencia a lo largo del continuo de especiación (Fig. 3). El objetivo general de esta Tesis Doctoral es comprender el papel relativo de los eventos climáticos, la geografía y el ambiente en la distribución de la variabilidad genética y fenotípica entre especies, linajes y poblaciones usando como modelo de estudio una radiación evolutiva reciente de saltamontes. Los cinco objetivos específicos desarrollados en los diferentes capítulos pueden sintetizarse en los siguientes puntos (Fig. 4):

1. Reconstruir las relaciones filogenéticas del complejo de especies *Chorthippus* grupo *binotatus*, delimitar los diferentes linajes y dilucidar el estatus taxonómico de las distintas entidades biológicas mediante la integración de datos genómicos y fenotípicos.

2. Analizar la historia evolutiva de los linajes de un taxón ampliamente distribuido y con requerimientos ecológicos muy restringidos (*Chorthippus binotatus binotatus*). Determinar los factores abióticos/bióticos que modulan la demografía de la especie y que, de modo último, determinan sus patrones espaciales de diversidad genética a diferentes escalas espaciales a lo largo de toda su área de distribución.

3. Examinar la diferenciación y estructuración genética que presentan las poblaciones de *Chorthippus binotatus binotatus* a una escala regional (sureste de la península Ibérica) y analizar el efecto que los cambios temporales de la composición del paisaje (geología y clima) han ejercido sobre los patrones espaciales de flujo y diversidad genética.

4. Cuantificar la contribución relativa de la geografía (*i.e.* procesos neutrales) y el ambiente (*i.e.* selección natural) en la evolución de la divergencia genética de un taxón endémico de Pirineos (*Chorthippus saulcyi moralesi*) de rango de distribución muy restringido, pero cuyas poblaciones están distribuidas a lo largo de abruptos gradientes ambientales.

5. Analizar el papel de la deriva genética y la selección divergente sobre el patrón de diferenciación fenotípica entre las poblaciones de *Chorthippus saulcyi moralesi* a lo largo de su rango de distribución y examinar el efecto de la heterogeneidad ambiental sobre la evolución de rasgos morfológicos involucrados en la capacidad dispersiva.

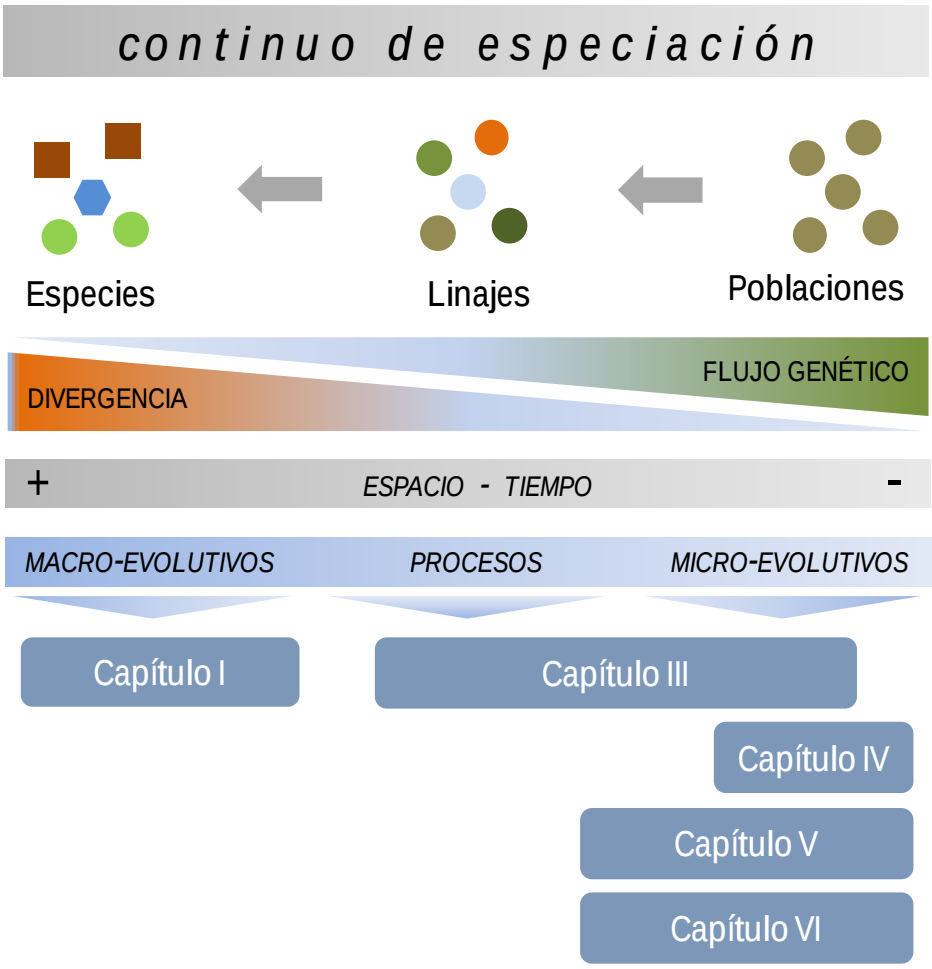


Figura 4 – Representación del concepto del continuo de especiación (*i.e. speciation continuum*). Se muestran tres estadios distintos a lo largo del continuo de especiación y su ubicación lo largo del espectro espacio-temporal que abarcan los procesos evolutivos que promueven la divergencia. De igual modo, se representa el marco espacio-temporal que abordan los diferentes capítulos que componen esta Tesis Doctoral. Los procesos micro-evolutivos serían los responsables de la variación a niveles infra-específicos (linajes, poblaciones) y escalas temporales pequeñas, mientras que los procesos macro-evolutivos afectan a niveles específicos o supra-específicos (especies, géneros, familias) y a escalas temporales más amplias.

METODOLOGÍA

Esta Tesis Doctoral emplea una amplia variedad de tipos de información y aproximaciones analíticas con el objetivo de obtener inferencias robustas acerca de cómo la geografía y la heterogeneidad ambiental determinan los patrones de variación genómica y fenotípica. A continuación, se describe brevemente la metodología empleada para abordar las cuestiones planteadas en los distintos capítulos.

MUESTREO DE ESPECÍMENES

Durante los meses de julio, agosto y septiembre de los años 2012, 2013 y 2014, se realizaron campañas de muestreo en Francia, España, Portugal y Marruecos para colectar los especímenes requeridos para los análisis genéticos y morfológicos. El importante esfuerzo de muestreo desempeñado ha permitido colectar individuos de todas las especies/subespecies abarcando la práctica totalidad de sus rangos de distribución. Los especímenes se conservaron a -20ºC en viales de 2 mL con etanol 96%. La información sobre la localización de las distintas poblaciones fue obtenida a partir de nuestros propios registros, así como de la literatura (*ej*. Llucià-Pomares 2002; Defaut 2011) y bases de datos online (GBIF; *http://www.gbif.es/*). Para algunos taxones (*C. binotatus binotatus* y *C. saulcyi moralesi*) y regiones geográficas concretas (sureste de la península Ibérica y Pirineos) se realizaron muestreos exhaustivos con el objetivo de poder abordar cuestiones a pequeña escala espacial (Capítulos IV, V y VI).

OBTENCIÓN DE DATOS MORFOLÓGICOS

El estudio de la variación morfológica se ha centrado en el análisis de diferentes rasgos morfológicos como son la longitud del fémur posterior, longitud y forma de la tegmina y tamaño y forma del pronoto. La toma de mediciones lineales se realizó mediante un microscopio esteroscópico (Ortego *et al*. 2012). El análisis de la variación de la forma de diferentes estructuras (tegmina y pronoto) se llevó a cabo mediante métodos de morfología geométrica (*geometric morphometrics*; Zelditch *et al*. 2004) basados en la localización de puntos homólogos de referencia (*landmarks*). El análisis de las diferencias inter-poblacionales en la forma de los rasgos se realizó mediante análisis canónicos (*Canonical*

Variates Analysis, CVA). En total, se obtuvieron datos morfológicos de los distintos rasgos para más de 230 individuos pertenecientes a todas las especies/subespecies descritas (Capítulo I y VI).

El grado de divergencia morfológica entre poblaciones fue estimado usando el parámetro P_{ST}, un análogo del parámetro F_{ST} para rasgos cuantitativos (Leinonen *et al*. 2008). El cálculo de los valores P_{ST} se realizó considerando estimas de la heredabilidad de cada rasgo obtenidas de la literatura debido a la dificultad de discernir entre el efecto del componente genético y ambiental sobre la variación genética aditiva (Brommer 2011). Con el objetivo de analizar la robustez de las inferencias obtenidas, se realizaron análisis de sensibilidad para comprender como las estimas de P_{ST} varían en función del parámetro que define la heredabilidad de los rasgos (Wojcik *et al*. 2006) (Capítulo VI).

OBTENCIÓN DE DATOS GENÉTICOS

El ADN de los individuos seleccionados para los análisis genéticos fue extraído y purificado siguiendo un protocolo salino (Aljanabi y Martinez 1997). Se emplearon distintos marcadores moleculares con el propósito de obtener estimas de diferenciación y variabilidad genética a diferentes escalas de resolución. En concreto se usaron microsatélites (secuencias cortas e hipervariables de ADN nuclear no codificante) y un fragmento del gen mitocondrial citocromo oxidasa I (COI). En este punto, cabe señalar que un elemento indispensable en el desarrollo de esta Tesis Doctoral ha sido el aislamiento, optimización y descripción de 18 nuevos marcadores microsatélites para la especie *Chorthippus binotatus binotatus* (Capítulo II).

OBTENCIÓN DE DATOS GENÓMICOS

Se emplearon técnicas de secuenciación masiva con el objetivo de obtener datos genómicos de alta resolución (SNPs, *single nucleotide polymorphisms*, *i.e.* polimorfismos de un solo nucleótido ubicados a lo largo de todo el genoma nuclear) y resolver las relaciones filogenómicas del complejo de especies *Chorthippus* grupo *binotatus*. Las librerías genómicas fueron preparadas siguiendo el método *double-digest RAD sequencing* (ddRADSeq) descrito en Peterson *et al*. (2012). De un modo resumido, el ADN fue digerido con las enzimas de restricción *Eco*RI y *Mse*I y posteriormente ligado a los adaptadores específicos de Illumina y un *barcode* (*i.e.* código de barras) único para cada individuo.

Los fragmentos fueron seleccionados en base a su longitud y aquellos comprendidos entre 475 y 580 pares de bases se amplificaron mediante PCR con una polimerasa de alta fidelidad. El producto resultante de la PCR fue secuenciado en una plataforma Illumina HiSeq2500, la cual proporcionó lecturas de secuencias de 151 pares de bases. El procesamiento bioinformático de los datos genómicos procedentes de Illumina (filtrado, control de calidad y obtención de loci homólogos) se realizó empleando el programa PYRAD (Eaton 2014). Este programa permite la variación en el número de *indels* (*i.e.* inserciones y deleciones) intra- e inter-individuales así como el solapamiento incompleto entre *reads* (*i.e.* lecturas), lo que otorga la posibilidad de recuperar más SNPs entre individuos de linajes o taxones con un cierto grado de divergencia (Capítulo I).

ANÁLISIS FILOGENÓMICOS Y DE DELIMITACIÓN DE ESPECIES EMPLEANDO DATOS ddRADSEQ

Las relaciones filogenómicas del complejo de especies se infirieron mediante métodos de reconstrucción de árboles de especies (*species trees*) basados en modelos de coalescencia. Para ello se utilizaron varios programas desarrollados específicamente para el análisis de datos genómicos (SNPs) como SVDQUARTETS (Chifman y Kubatko 2014) y SNAPP (Bryant *et al*. 2012) (Capítulo I). La delimitación de especies mediante datos genómicos se llevó a cabo empleando métodos basados en modelos de coalescencia. Específicamente se emplearon varios métodos, incluyendo el BAYES FACTOR DELIMITATION (BFD*; Leaché *et al*. 2014) y una nueva versión del programa BAYESIAN PHYLOGENETICS & PHYLOGEOGRAPHY (BPP; Yang y Rannala 2010, 2014) que permite la inferencia del árbol de especies y la delimitación de especies de manera conjunta sin necesidad de requerir una topología previa (Capítulo I). La delimitación integrativa de especies empleando conjuntamente datos genómicos y morfológicos, se realizó aplicando un método publicado recientemente (iBPP; Solís-Lemus *et al*. 2015). Esta aproximación, desarrollada sobre la arquitectura de BPP, integra datos genéticos y fenotípicos bajo un modelo de coalescencia e incorpora un modelo de evolución de rasgos morfológicos bajo un proceso de movimiento Browniano, lo cual es especialmente útil para la definición de los límites taxonómicos de especies en un escenario de especiación temprana (Capítulo I).

ANÁLISIS FILOGENÉTICOS Y DEMOGRÁFICOS EMPLEANDO FRAGMENTOS DE GENES MITOCONDRIALES

Las relaciones filogenéticas a nivel intra-específico y los tiempos de divergencia entre linajes se estimaron con el programa BEAST (Drummond *et al*. 2012) usando un fragmento del gen mitocondrial citocromo oxidasa I (COI). Mediante el mismo programa, se generaron *Bayesian Skyline Plots* (BSP; Drummond *et al*. 2005) para inferir cambios temporales en el tamaño efectivo poblacional. De manera complementaria se realizaron test de Fu (Fu 1997), Tajima (Tajima 1989), R_2 (Ramos-Onsins y Rozas 2002) y distribuciones no coincidentes (*mismatch distributions*; Rogers y Harpending 1992) con el objetivo de detectar señales de cambios demográficos pasados (Capítulo III). La variación genética mitocondrial de las poblaciones y linajes se estimó mediante diferentes parámetros usando el programa DNASP (Librado y Rozas 2009): número de haplotipos (*H*), número de sitios polimórficos (*S*), diversidad haplotípica (H_D) y diversidad nucleotídica (π) (Capítulo III).

ANÁLISIS DE GENÉTICA POBLACIONAL BASADOS EN MICROSATÉLITES

La estructuración genética espacial de las poblaciones fue examinada mediante métodos Bayesianos (STRUCTURE: Pritchard *et al*. 2000; TESS: Chen *et al*. 2007) que asignan estadísticamente cada individuo a un determinado grupo genético en función de su genotipo (Capítulos III, IV y V). Complementariamente a estos métodos Bayesianos, se utilizó un análisis discriminante de componentes principales (*Discriminant Analysis of Principal Components,* DAPC) que permite también inferir la estructura genética poblacional, pero dado que no se fundamenta en modelos genéticos puede ser más eficaz a la hora de detectar patrones espaciales complejos de estructuración genética bajo determinados escenarios evolutivos (Jombart *et al*. 2010) (Capítulo V). Mediante el programa POPULATIONS (Langella 1999) se construyeron árboles filogenéticos basados en microsatélites con el objetivo de visualizar las relaciones genéticas entre las distintas poblaciones (Capítulo IV y V).

La cuantificación de la diferenciación genética neutral entre las poblaciones se realizó mediante el cálculo del parámetro F_{ST} usando el programa ARLEQUIN (Excoffier y Lischer 2010). Debido a la frecuencia de alelos nulos que generalmente exhiben los ortópteros, la estimación del parámetro F_{ST} incluyó la corrección por alelos nulos basada en el método ENA e implementada en el programa FREENA (Chapuis y Estoup 2007). Así mismo, se calcularon estimas de diversidad genética mediante el

cálculo de los parámetros de riqueza alélica (A_R) y heterocigosidad (H_O y H_E) empleando los programas HP-RARE (Kalinowski 2005) y ARLEQUIN (Excoffier y Lischer 2010), respectivamente (Capítulos III, IV y VI).

El marco metodológico denominado *Aproximate Bayesian Computation* (ABC; Beaumont 2010) fue empleado para elucidar estadísticamente la historia evolutiva y demográfica de los diferentes linajes y poblaciones. Esta aproximación (i) simula la demografía poblacional y su diversidad genética resultante bajo diferentes escenarios evolutivos (*i.e.* hipótesis) considerando un modelo de coalescencia, (ii) proporciona un valor de apoyo estadístico a cada uno de los escenarios evolutivos y (iii) permite seleccionar aquel que explique más adecuadamente el patrón de variación genética observada (Cornuet *et al*. 2014) (Capítulos III y IV).

MODELADO DE NICHO CLIMÁTICO/ECOLÓGICO

Los modelos de nicho se realizaron con el objetivo de generar hipótesis espacialmente explícitas (*ej*. localización de refugios y corredores) relativas a la dinámica distribucional de los taxones estudiados durante los últimos 21000-120000 años que posteriormente pudieran ser contrastadas con datos genéticos. Los modelos de distribución potencial se realizaron usando un algoritmo de máxima entropía implementado en el programa MAXENT (Phillips *et al*. 2008) y capas de información bioclimática (WorldClim; *www.worldclim.org*). Mediante este mismo programa se generaron reconstrucciones paleo-climáticas de las distribuciones potenciales en el pasado, en concreto durante el último máximo glacial (hace ~21000 años; *i.e. Last Glacial Maximum*, LGM) y el último periodo inter-glacial (hace ~120000 años; *i.e. Last Interglacial,* LIG) (Capítulos III, IV, V y VI). Este procedimiento permite determinar la distribución potencial presente y pasada de una especie en función de sus requerimientos climáticos y estudiar las consecuencias de los cambios de su distribución sobre las rutas de dispersión y los patrones de variación genética. También se generaron modelos de distribución potencial actual y pasada para todas las especies de leguminosas (tribu *Genisteae*; Talavera *et al*. 2001) de las que se alimenta la especie *Chorthippus binotatus binotatus,* con el objetivo de analizar el impacto de la dinámica distribucional de las diferentes plantas nutricias durante los últimos 21000 años sobre la demografía de la especie de ortóptero (Capítulo III). Partiendo de las distribuciones potenciales en el presente y pasado, se calcularon mediante sistemas de información geográfica (ARCGIS; ESRI, Redlands, CA, USA) estimas de la estabilidad climática para cada población

con el propósito de contrastar el efecto de las variaciones climáticas sobre la dinámica demográfica de las mismas y la evolución de los rasgos morfológicos involucrados en la capacidad de dispersión (Capítulos IV, V y VI).

ANÁLISIS DEL EFECTO DE LA GEOGRAFÍA, AMBIENTE Y FLUCTUACIONES CLIMÁTICAS SOBRE LOS PATRONES ESPACIALES DE DIFERENCIACIÓN Y DIVERSIDAD GENÉTICA

El efecto de la heterogeneidad del paisaje sobre las rutas de dispersión y los patrones de variación genética fue abordado desde el marco conceptual de la Teoría de Circuitos (McRae 2006; McRae y Beier 2007). Mediante el programa CIRCUITSCAPE (McRae 2006; McRae y Beier 2007) se calcularon las distancias de resistencia al movimiento inter-poblacional considerando distintas hipótesis espaciales basadas en modelos de distribución potencial, estabilidad climática, reconstrucciones paleo-geográficas y complejidad topográfica contemporánea. Para analizar el efecto del aislamiento por resistencia (*isolation-by-resistance*, IBR), se analizó la correlación entre las distancias de resistencia procedentes de las distintas hipótesis y la diferenciación genética inter-poblacional mediante regresiones múltiples de matrices de distancias (*Multiple Matrix Regression with Randomization*, MMRR; Wang 2013) (Capítulo IV y V). Por otro lado, para contrastar el efecto relativo de la geografía (*isolation-by-distance*, IBD) y el ambiente (*isolation-by-environment*, IBE) sobre el grado de diferenciación genética inter-poblacional se usaron análisis de redundancia basados en distancias (dbRDA; Legendre y Anderson 1999), así como un método recientemente desarrollado de modelado Bayesiano geo-estadístico (SUNDER; Botta *et al*. 2015) (Capítulo V). Finalmente, el efecto de la estabilidad climática/ecológica durante los últimos 21000-120000 años sobre diversidad genética poblacional se analizó mediante modelos lineales generalizados (GLMs; Bates *et al*. 2015) (Capítulos III, IV y VI).

ORGANIZACIÓN DE LA TESIS DOCTORAL

En el Capítulo I se resuelven las relaciones filogenómicas del complejo de especies *Chorthippus* grupo *binotatus* mediante el empleo de datos genómicos (> 30000 SNPs) y diferentes métodos filogenéticos basados en modelos de coalescencia. Así mismo, se combinan datos genómicos y morfológicos bajo un novedoso marco estadístico para delimitar las especies del grupo y determinar su estatus taxonómico. La aproximación empírica empleada permite examinar la capacidad de diferentes métodos de delimitación de especies recientemente desarrollados bajo un amplio abanico de escenarios y posibilidades, proporcionado unas directrices para estudios similares futuros. Así mismo, los resultados obtenidos proporcionan nueva información sobre las relaciones filogenómicas de las especies del grupo y aportan datos relevantes de cara a la actualización del estatus taxonómico de las entidades biológicas inferidas.

En el Capítulo II se aíslan y describen 18 marcadores microsatélites para la subespecie *Chorthippus binotatus binotatus* y se analiza su funcionalidad, características y diversidad genética en todas las restantes subespecies putativas del complejo de especies. Este conjunto de marcadores moleculares constituyen la herramienta utilizada en los siguientes capítulos para investigar la diversidad, estructura genética y trayectoria demográfica de algunas de las especies del complejo a escala de paisaje y filogeográfica.

En el Capítulo III se estudia la filogeografía y la demografía de los diferentes linajes de la subespecie putativa *Chorthippus binotatus binotatus* a lo largo de toda su área de distribución empleando marcadores microsatélites y un fragmento de gen mitocondrial. Adicionalmente, se construyen modelos climáticos de distribución potencial durante el presente y el último máximo glacial (aproximadamente hace 21000 años) tanto para la especie objeto de estudio como para las trece especies de leguminosas arbustivas de los que se nutre. A su vez, se analiza el patrón espacial de diversidad genética nuclear y mitocondrial y se contrasta explícitamente la hipótesis de que los cambios distribucionales de las plantas nutricias durante los últimos 21000 años han modelado la historia demográfica de la especie de saltamontes a escala poblacional y regional. Este capítulo proporciona evidencias sobre el papel de los refugios situados en los límites de las áreas de distribución en la persistencia de las poblaciones a escala local y en los procesos de diversificación de los linajes a escala regional.

En el Capítulo IV se examina a pequeña escala espacial el papel de los cambios paisajísticos a lo largo del tiempo sobre la estructuración genética de la subespecie *Chorthippus binotatus binotatus* en el sureste de la península Ibérica, una región biogeográficamente muy compleja. En este capítulo se examina en qué medida la dinámica geológica del Mioceno (hace ~12,0–7,2 millones de años), la adecuación climática del hábitat en el presente y el último máximo glacial (hace 21000 años) y la complejidad topográfica explican los patrones de diferenciación genética inter-poblacional en el área de estudio. Los resultados obtenidos evidencian un efecto primordial de la topografía como el factor que modula la diferenciación genética entre las poblaciones del área de estudio. Este estudio pone de manifiesto la necesidad de integrar información espacial sobre la estructura paisajística histórica y contemporánea para obtener un conocimiento completo de los factores que determinan la variación genética en organismos que habitan regiones con una historia biogeográfica compleja.

El Capítulo V se centra en el estudio del papel relativo de la geografía y el ambiente en los procesos de diferenciación genética poblacional a una escala de paisaje. Descifrar y cuantificar en qué medida el ambiente determina la divergencia inter-poblacional en paisajes ecológicamente heterogéneos es una cuestión evolutiva importante debido a que este mecanismo puede suponer un primer paso en los procesos de especiación incipiente. Sin embargo, resolver esta pregunta alberga dificultades inherentes debido a la auto-correlación espacial entre la geografía y los factores ecológicos. Para abordar este problema, usamos como modelo de estudio un endemismo pirenaico de distribución restringida (*Chorthippus saulcyi moralesi*) con el objetivo de contrastar diferentes escenarios de aislamiento poblacional (aislamiento por distancia, *isolation-by-distance*, IBD; aislamiento por resistencia, *isolation-by-resistance*, IBR; aislamiento por ambiente, *isolation-by-environment*, IBE) integrando múltiples métodos analíticos. Los resultados obtenidos apuntan a un papel preponderante de los mecanismos neutrales de diferenciación y un impacto limitado de los procesos de divergencia adaptativa. Este capítulo pone en evidencia la necesidad de considerar una amplia gama de potenciales mecanismos de divergencia en un marco metodológico que integre el uso de herramientas estadísticas complementarias de cara a la obtención de inferencias robustas sobre los procesos que promueven la estructuración genética de las poblaciones naturales.

Finalmente, el Capítulo VI complementa al anterior y supone un paso más allá en el estudio de las causas y consecuencias de los mecanismos micro-evolutivos que determinan la divergencia poblacional a pequeña escala espacial. En este capítulo se examina el papel relativo de la selección natural y la deriva genética sobre la diferenciación fenotípica en el endemismo pirenaico *Chorthippus*

saulcyi moralesi. Los resultados obtenidos indican la presencia de procesos de adaptación local a lo largo de gradientes ambientales que se ven reflejados en la diferenciación inter-poblacional en determinados rasgos morfológicos. Paralelamente, se contrasta la hipótesis que predice un mayor desarrollo de rasgos ligados a la capacidad dispersiva en poblaciones periféricas sometidas a una mayor inestabilidad ambiental y demográfica. Los análisis realizados demuestran un empobrecimiento de la diversidad genética y un mayor desarrollo de rasgos fenotípicos relacionados con la dispersión en aquellas poblaciones situadas en los límites de la distribución de la especie donde la inestabilidad climática y demográfica ha sido históricamente más acentuada.

REFERENCIAS

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692-4693.

Allen R, Siegert MJ, Payne AJ (2008) Reconstructing glacier-based climates of LGM Europe and Russia - Part 2: a dataset of LGM precipitation/temperature relations derived from degree-day modelling of palaeo glaciers. *Climate of the Past*, 4, 249-263.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, 21, 2078-2091.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, 32, 818-830.

Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439, 719-723.

Bates D, Maechler M, Bolker BM, Walker SC (2015) Fitting Linear Mixed-Effects Models using *lme4*. *Journal of Statistical Software*, 67, 1-48.

Beaumont MA (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41, 379-406.

Bétard F (2016) Le Criquet des Ajoncs, *Gomphocerippus binotatus* ssp. *armoricanus*, en Vendée et Deux-Sèvres Données nouvelles sur la répartition et l'écologie d'un orthoptère menacé. *Le Naturaliste Vendéen*, 12, 87-95.

Bolívar I (1908) Algunos ortópteros nuevos de España, Marruecos y Canarias. *Boletín de la Real Sociedad Española de Historia Natural*, 8, 317-334.

Botta F, Eriksen C, Fontaine MC, Guillot G (2015) Enhanced computational methods for quantifying the effect of geographic and environmental isolation on genetic differentiation. *Methods in Ecology and Evolution*, 6, 1270-1277.

Bridle JR, Baird SJE, Butlin RK (2001) Spatial structure and habitat variation in a grasshopper hybrid zone. *Evolution*, 55, 1832-1843.

Brommer JE (2011) Whither P_{ST}? The approximation of Q_{ST} by P_{ST} in evolutionary and conservation biology. *Journal of Evolutionary Biology*, 24, 1160-1168.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29, 1917-1932.

Butlin RK, Debelle A, Kerth C*, et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, 27, 27-39.

Butlin RK, Saura M, Charrier G*, et al.* (2014) Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, 68, 935-949.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785-789.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, 22, 4369-4383.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24, 621-631.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, 7, 747-756.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30, 3317-3324.

Chopard L (1952) *Faune de France: 56, Orthoptéroïdes*. Lechevalier, Paris, France.

Cigliano MM, Braun H, Eades DC, Otte D (2017) *Orthoptera Species File*. Version 5.0/5.0. *http://orthoptera.speciesfile.org/*

Cooper SJB, Ibrahim KM, Hewitt GM (1995) Postglacial expansion and genome subdivision in the european grasshopper *Chorthippus parallelus*. *Molecular Ecology*, 4, 49-60.

Cornuet J-M, Pudlo P, Veyssier J*, et al.* (2014) DIYABC v.2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187-1189.

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? *Science*, 339, 413-416.

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 353, 287-305.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland (MA), USA.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, 56, 879-886.

Defaut B (1987) Diagnoses d'Orthoptéroïdes nouveaux ou nouvellement décrits au Maroc. *L'Entomologiste*, 43, 109-112.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (*Caelifera*, *Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Defaut B (2015) *Gomphocerippus binotatus armoricanus*, sous-espèce nouvelle de France (*Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 20, 25-26.

Devitt TJ, Devitt SEC, Hollingsworth BD, McGuire JA, Moritz C (2013) Montane refugia predict population genetic structure in the Large-blotched Ensatina salamander. *Molecular Ecology*, 22, 1650-1665.

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, 400, 354-357.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, 1185-1192.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29, 1969-1973.

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30, 1844-1849.

Espindola A, Carstens BC, Alvarez N (2014) Comparative phylogeography of mutualists and the effect of the host on the genetic structure of its partners. *Biological Journal of the Linnean Society*, 113, 1021-1035.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567.

Ferrer ES, García-Navas V, Bueno-Enciso J*, et al.* (2016) The influence of landscape configuration and environment on population genetic structure in a sedentary passerine: insights from loci located in different genomic regions. *Journal of Evolutionary Biology*, 29, 205-219.

Fordham DA, Brook BW, Moritz C, Nogués-Bravo D (2014) Better forecasts of range dynamics using genetic data. *Trends in Ecology & Evolution*, 29, 436-443.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915-925.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27, 480-488.

Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3209-3213.

García MD, Clemente ME, Presa JJ (1995) Manifestaciones acústicas de *Chorthippus binotatus binotatus* (Charpentier, 1825) (Orthoptera: Acrididae). Su estatus taxonómico y su distribución en la Península Ibérica. *Boletín de la Asociación Española de Entomología*, 19, 229-242.

García-Navas V, Noguerales V, Cordero PJ, Ortego J (2017) Phenotypic disparity in Iberian short-horned grasshoppers (Acrididae): the role of ecology and phylogeny. *BMC Evolutionary Biology*, 17, 109.

Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, 58, 1781-1793.

Harz K (1975) *Die Orthopteren Europas: 2*. W. Junk, La Haye.

He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, 67, 3386-3402.

Hendry AP, Bolnick DI, Berner D, Peichel CL (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, 75, 2000-2036.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68, 87-112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hewitt GM (2004a) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 183-195.

Hewitt GM (2004b) The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology*, 1, 1-16.

Hickerson MJ, Cunningham CW (2005) Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. *Evolution*, 59, 344-360.

Huang J-P, Knowles LL (2016) The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single bayesian approach in Hercules beetles. *Systematic Biology*, 65, 685-699.

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.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5, 187-189.

Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability,* 19, 27-43.

Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54, 1337-1348.

Knowles LL, Richards CL (2005) Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Molecular Ecology*, 14, 4023-4032.

Kondrashov AS, Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature*, 400, 351-354.

Langella O (1999) POPULATIONS. 1.2.31. *http://bioinformatics.org/~typhon/populations/*

Laland KN, Uller T, Feldman MW, *et al*. (2015) The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B-Biological Sciences*, 282, 20151019.

Laland KN, Wray GA (2014) Does evolutionary theory need a rethink? *Nature*, 514, 161-164.

Leaché AD, Koo MS, Spencer CL, *et al*. (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12418-12423.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, 63, 534-542.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69, 1-24.

Leinonen T, O'Hara RB, Cano JM, Merilä J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, 21, 1-17.

Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.

Losos JB and D.L. Mahler DL (2010) Adaptive Radiation: The Interaction of Ecological Opportunity, Adaptation, and Speciation. In: Bell MA, Futuyma DJ, Eanes WF y Levinton JS (Eds*.) Evolution Since Darwin: The First 150 Years*. Sinauer Associates, Sunderland, MA, USA.

Lunt DH, Ibrahim KM, Hewitt GM (1998) MtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity*, 80, 633-641.

Llucià-Pomares D (2002) *Revision of the Orthoptera (Insecta) of Catalonia (Spain).* Monografías S.E.A., vol. 7, Zaragoza, Spain.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, 28, 614-621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18, 189-197.

May RM (2012) The future of biological diversity in a crowed world. *Current Science*, 82, 1325-1331.

Mayer F, Berger D, Gottsberger B, Schulze W (2010) Non-ecological radiations in acoustically communicating grasshoppers? In: Glaubrecht M (Ed.) *Evolution in action. Case studies in adaptive radiation, speciation and the origin of biodiversity*. Springer Verlag, Berlin, Germany.

McRae BH (2006) Isolation by resistance. *Evolution*, 60, 1551-1561.

McRae BH, 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, 19885-19890.

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, 89, 2712-2724.

Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, 51, 238-254.

Moritz C, Agudo R (2013) The future of species under climate change: resilience or decline? *Science*, 341, 504-508.

Nattier R, Robillard T, Amedegnato C*, et al.* (2011) Evolution of acoustic communication in the *Gomphocerinae* (Orthoptera: Caelifera: Acrididae). *Zoologica Scripta*, 40, 479-497.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, 24, 145-156.

Nosil P (2012) *Ecological speciation*. Oxford University Press, New York, USA.

Ortego J, Aguirre MP, Cordero PJ (2012) Genetic and morphological divergence at different spatiotemporal scales in the grasshopper *Mioscirtus wagneri* (Orthoptera: Acrididae). *Journal of Insect Conservation*, 16, 103-110.

Ortego J, García-Navas V, Noguerales V, Cordero PJ (2015) Discordant patterns of genetic and phenotypic differentiation in five grasshopper species codistributed across a microreserve network. *Molecular Ecology*, 24, 5796-5812.

Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, 7, 16.

Papadopoulou A, Knowles LL (2016) Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences of the United States of America,* 113, 8018-8024.

Parducci L, Jorgensen T, Tollefsrud MM*, et al.* (2012) Glacial survival of boreal trees in Northern Scandinavia. *Science*, 335, 1083-1086.

Pepper M, Doughty P, Hutchinson MN, Keogh JS (2011a) Ancient drainages divide cryptic species in Australia's arid zone: morphological and multi-gene evidence for four new species of Beaked Geckos (Rhynchoedura). *Molecular Phylogenetics and Evolution*, 61, 810-822.

Pepper M, Fujita MK, Moritz C, Keogh JS (2011b) Palaeoclimate change drove diversification among isolated mountain refugia in the Australian arid zone. *Molecular Ecology*, 20, 1529-1545.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One*, 7, e37135.

Phillips SJ, Dudik M (2008) Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Powell THQ, Hood GR, Murphy MO*, et al.* (2013) Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). *Evolution*, 67, 2561-2576.

Prates I, Xue AT, Brown JL*, et al.* (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7978-7985.

Pratz J, Cloupeau R (2010) Liste rouge commentée des Orthopéres de la región Centre. *Materiaux Orthopériques et Entomocénotiques*, 15, 17-33.

Presa JJ, García MD, Clemente ME (2007) Catalogue of Orthoptera Caelifera from the Iberian Peninsula and Balearic Islands (Orthoptera: Caelifera). *Journal of Orthoptera Research*, 16, 175- 179.

Presa JJ, García MD, Clemente ME, Barranco Vega P, Correas J, Ferreira S, Hochkirch A, Lemos P, Odé B, Prunier F (2016) *Chorthippus ariasi*. The IUCN Red List of Threatened Species 2016: e.T16084434A75087340. *http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T16084434A75087340.en*

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Ragge DR, Reynolds WJ (1998) *The Songs of the Grasshoppers and Crickets of Western Europe*. Harley Books, Colchester, Essex C06 4 AH, England.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19, 2092-2100.

Rogers AR, Harpending H (1992) Population-growth makes waves in the distribution of pairwise genetic-differences. *Molecular Biology and Evolution*, 9, 552-569.

Ruiz-González A, Cushman SA, Madeira MJ, Randi E, Gómez-Moliner BJ (2015) Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular Ecology*, 24, 5110-5129.

Rundell RJ, Price TD (2009) Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, 24, 394-399.

Ryan PG, Bloomer P, Moloney CL, Grant TJ, Delport W (2007) Ecological speciation in South Atlantic island finches. *Science*, 315, 1420-1423.

Safran RJ, Scordato ESC, Wilkins MR*, et al.* (2016) Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. *Molecular Ecology*, 25, 3865-3883.

Salvi D, James Harris D, Kaliontzopoulou A, Carretero MA, Pinho C (2013) Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. *BMC Evolutionary Biology*, 13, 147.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, 16, 372-380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, 323, 737-741.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 9955-9962.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, 68, 1-15.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, 16, 940-950.

Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. *Science*, 276, 1855-1857.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution*, 64, 295-315.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69, 492-507.

Soria-Carrasco V, Gompert Z, Comeault AA*, et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344, 738-742.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, 16, 608-613.

Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences*, 277, 661-671.

Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA (2015) Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evolutionary Biology*, 15, 204.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.

Talavera S, Aedo C, Castroviejo S, Romero C, Sáez L, Salgueiro FJ, Velayos M (2001) *Flora Ibérica VII/1 Leguminoseae (partim)*. Real Jardín Botánico, CSIC, Madrid, España.

Thorpe RS, Surget-Groba Y, Johansson H (2008) The relative importance of ecology and geographic isolation for speciation in anoles. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 363, 3071-3081.

Velo-Antón G, Parra JL, Parra-Olea G, Zamudio KR (2013) Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, 22, 3261-3278.

Wang IJ (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, 3403-3411.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, 23, 5649-5662.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, 16, 175-182.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, 56, 875-878.

Wojcik AM, Polly PD, Sikorski MD, Wojcik JM (2006) Selection in a cycling population: differential response among skeletal traits. *Evolution*, 60, 1925-1935.

Wright S (1943) Isolation by distance. *Genetics*, 28, 114-138.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9264-9269.

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, 31, 3125-3135.

Yannic G, Pellissier L, Ortego J*, et al.* (2014) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, 4, 132-137.

Yeates DK, Seago A, Nelson L*, et al.* (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, 36, 209-217.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric morphometrics for biologists: a primer*. Elsevier Academic Press, San Diego, CA, USA.

CAPÍTULO I

Integrative species delimitation in a recent evolutionary radiation of grasshoppers: evaluating the importance of number of loci and different phenotypic traits

VÍCTOR NOGUERALES, PEDRO J. CORDERO & JOAQUÍN ORTEGO

(submitted)

Integrative species delimitation in a recent evolutionary radiation of grasshoppers: evaluating the importance of number of loci and different phenotypic traits

Abstract

Although establishing species limits is a primary goal of systematics with important implications for evolutionary, ecological and conservation research, this task remains challenging at both the conceptual (*i.e.* application of different species concepts) and analytical (*i.e.* selection of model-based approaches, integration of different sources of information) levels. Here, we employ a suite of coalescent-based approaches to elucidate the phylogenomic relationships (SVDQUARTETS and SNAPP) and define species limits (BFD*, BPP and iBPP) in a recent evolutionary radiation of grasshoppers (*Chorthippus* group *binotatus* species complex) composed of two species and eight putative subspecies. In particular, we integrated genomic data and different sources of phenotypic information (traditional *vs.* geometric morphometrics) and examined the impacts of different demographic prior combinations, number of loci, and sex-based trait variation on the power and accuracy of the obtained inferences. Although some phylogenomic relationships were not fully resolved, we found unambiguous evidence that all putative lineages were supported as distinct taxa regardless of the different settings or the species delimitation method employed. Our analyses showed that the phylogenetic signal yielded only by morphological datasets was strong enough to recover the eight-species delimitation model under any demographic scenario and, thus, we found no support for the notion that genomic-based inference of species limits overestimates the number of taxa. However, detailed analyses considering specific phenotypic matrices revealed that the support for certain species splits was influenced by differences between sexes on among-taxa trait variation. Overall, our study highlights the power of coalescent-based approaches to delimit taxa boundaries when handling large genomic datasets and indicates that integrating a broad suite of traits for the two sexes could be beneficial when aiming at determining species limits, particularly at early-stages of divergence where conflicting inferences are more prone to appear.

INTRODUCTION

Establishing species boundaries and discovering new taxa is considered one of the paramount goals of systematics (Coyne & Orr 2004; Wiens 2007). Inferring species limits has a great importance on ecological, evolutionary and biodiversity conservation studies of endangered and taxonomically problematic species groups (Agapow *et al*. 2004; Huang & Knowles 2016a; Weir *et al*. 2016). Delimiting and naming species is necessary to ensure that both the taxonomist community and members of our society refer to the same biological entity when using a consensual name (Agapow *et al*. 2004). This

takes a decisive importance under the ongoing biodiversity crisis resulted from the severe impacts of human activities and global change, which are expected to lead numerous species to extinction before they are discovered (Costello *et al*. 2013), including taxa from remote areas but also cryptic species inhabiting well-surveyed regions (Hotaling *et al*. 2006; Weir *et al*. 2016). Despite the relevance of establishing species boundaries, the available approaches are not exempt from limitations and controversial aspects, primarily because they rely on species concepts defined considering different biological properties to recognize taxa (Agapow *et al*. 2004; de Queiroz 2007; Freudenstein *et al*. 2016).

Our ability to distinguish and discover new taxa was strongly triggered by the application of DNA-based species delimitation approaches almost three decades ago (Hebert *et al*. 2003; Yang & Rannala 2017). Yet, molecular species delimitation has been historically limited by the amount of genetic data that can be obtained for non-model organisms using Sanger sequencing (Carstens *et al*. 2013; Caviedes-Solis *et al*. 2015). Recent literature reviews on DNA-based species delimitation indicate that the mean number of loci employed per study was fewer than six loci (Carstens *et al*. 2013: mean ± SD = 4.5 ± 4.4 loci, maximum = 21 loci, *n* = 28; Caviedes-Solis *et al*. 2015: mean ± SD = 5.8 ± 6.5 loci, maximum = 29 loci, *n* = 15). The advent of high-throughput sequencing technology has substantially improved our capability to sample hundreds to thousands of loci from non-model organisms (Emerson *et al*. 2010), opening the door for exploring species-levels of genetic variation at an unprecedented resolution (Leaché *et al*. 2015; Potter *et al*. 2016; Yoder *et al*. 2016) and improving our understanding on the evolutionary mechanisms underlying early-stage speciation processes (Papadopoulou & Knowles 2015; Huang 2016; Weir *et al*. 2016). In parallel with the capability of generating large genomic information, different model-based methods to delimit species have been developed (Knowles & Carstens 2007; O'Meara 2010; Yang & Rannala 2010; Ence & Carstens 2011; Camargo *et al*. 2012; Leaché *et al*. 2014; Jackson *et al*. 2017). In particular, recent species delimitation methods that incorporate the estimation of demographic parameters (divergence time and population size) and accommodate incomplete lineage sorting and gene tree-species tree conflicts have become broadly popular (Yang & Rannala 2010, 2014, 2017; see also Edwards *et al*. 2016). Simulation studies have shown the accuracy and robustness of these methods to infer the correct species delimitation model when using an increasing number of loci under different demographic scenarios (Zhang *et al*. 2011, 2014; Yang & Rannala 2010, 2014, 2017; Jackson *et al*. 2017). Yet, empirical studies exploring the impact of the number of employed loci at the scale yielded by currently available high-throughput sequencing technology remain elusive (for an exception, see Hime *et al*. 2016).

The capability of determining whether a putative taxon represents an independently evolving lineage (*i.e.* species, *sensu* de Queiroz 2007) can strongly differ according to the kind of data used (*e.g.* genetic, ecological, behavior, phenotypic; *e.g.* McKay *et al*. 2013; Edwards & Knowles 2014; Hedin *et al*. 2015; Lamanna *et al*. 2016). However, most currently available advanced methods are exclusively based on molecular data, which can compromise the reliability of the inferences obtained especially when dealing with recent speciation events due to the impact of ancestral polymorphisms or when species divergence occurs with gene flow or is selectively driven (Degnan & Rosenberg 2009; Solís-Lemus *et al*. 2015). For this reason, approaches integrating different kinds of data, such as genetic and phenotypic information, have been traditionally demanded in the literature (Wiens 2007; Padial 2010; Yeates *et al*. 2011; Fujita *et al*. 2012). Recently, a unified statistical framework has been developed to accommodate multisequence DNA data and independent quantitative continuous traits to species delimitation into a Bayesian coalescent model (Solís-Lemus *et al*. 2015). The power of this method to resolve taxonomic problems in different organism groups has been revealed in different empirical studies integrating genetic information and phenotypic data obtained using either traditional approaches (Dornburg *et al*. 2016; Pyron *et al*. 2016; Olave *et al*. 2017) or geometric morphometric analyses (Eberle *et al*. 2016; Huang & Knowles 2016a). However, this approach does not include a model of trait evolution accommodating within-taxa trait variation such as that resulting from sexual dimorphism (Solís-Lemus *et al*. 2015). To date, studies employing this method have eluded the uncertain impact of sexually distinct traits on species delimitation inference by only considering male phenotypes in their analyses (Solís-Lemus *et al*. 2015; Eberle *et al*. 2015; Huang & Knowles 2016a) or ignoring sexually dimorphic traits (Pyron *et al*. 2016). Beyond the importance of evaluating sex-based trait variation on species delimitation, the potential impacts of within-to-between lineage trait variance and of using different number of traits or contrasting sources of phenotypic information have not been yet empirically addressed (Solís-Lemus *et al*. 2015).

The *Chorthippus* group *binotatus* (Orthoptera: Acrididae) is a species complex of grasshoppers belonging to *Gomphocerinae* (Defaut 2011, 2015), the most speciose acridid subfamily (> 1 000 species; Cigliano *et al*. 2017). This species complex is distributed through southwest Europe (France, Spain and Portugal) and northwest Africa (Morocco) and currently comprises two species (*Chorthippus binotatus* Charpentier 1825 and *Chorthippus saulcyi* Krauss 1888) that in turn encompass a total of eight subspecies (*C. b. atlasi*, *C. b. armonicanus*, *C. b. binotatus*, *C. s. algoaldensis*, *C. s. daimei*, *C. s. moralesi*, *C. s. saulcyi*, and *C. s. vicdessossi*) exhibiting slight morphological and ecological differences (Defaut 2011, 2015; Table S1, Fig. 1). While *C. b. atlasi*, *C. b. armonicanus* and

C. b. binotatus present wide distributions, the remaining subspecies are narrow-endemics restricted to different mountain ranges (Fig. 1). The *Chorthippus* group *binotatus* species complex contains either short-winged (*e.g. C. s. daimei* and *C. s. moralesi*) and long-winged taxa (*e.g. C. b. binotatus* and *C. s. saulcyi*) as well as intermediate forms distributed along a continuum of phenotypic differentiation (*C. s. algoaldensis*, *C. s. vicdessossi*; see Fig. S1). Phenotypic differences are also patent for other traits that have been used by traditional taxonomy for describing the different species/subspecies of the group, such as the position of the apical portion of the median area in relation to forewing length and the proportional length of prozone and metazone (Llucià-Pomares 2002; Defaut 2011). Since the description of the main nominal species in the 19's century, the infraspecific taxonomy of the group has undergone important changes on the basis of morphological studies considering the above described traits, but the phylogenetic relationships within this complex remain unresolved (reviewed in Defaut 2011). The different stages of divergence represented within this species complex make it an excellent system for integrating genomic and phenotypic data to test the power and drawbacks of coalescent-based approaches of species delimitation (Shaffer & Thomson 2007).

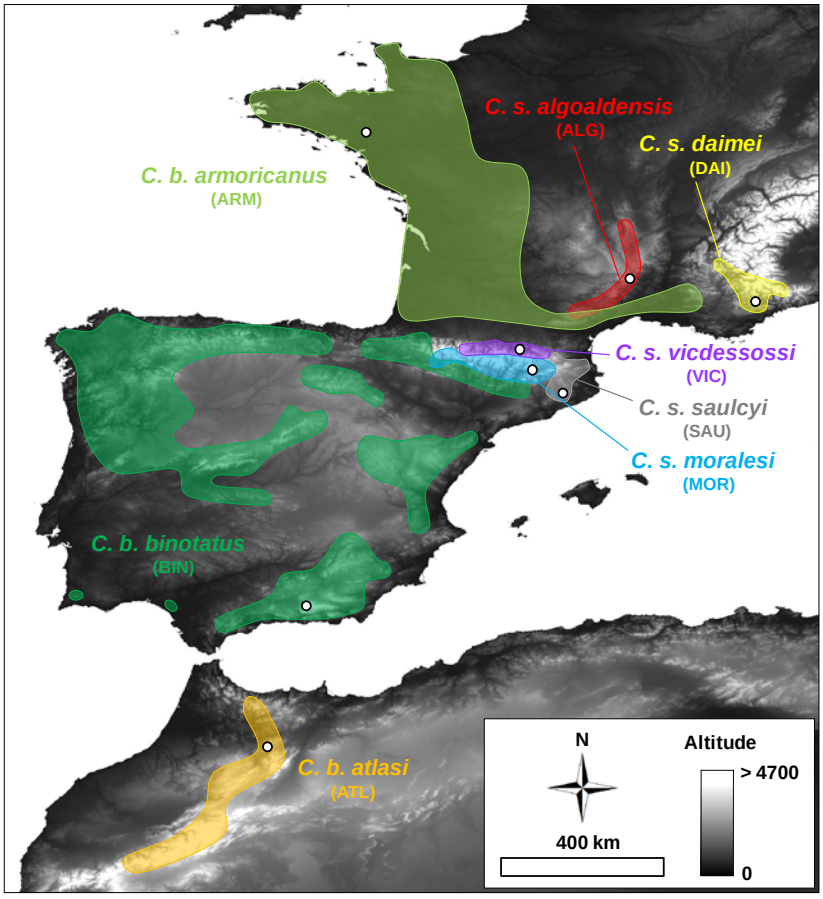


Figure 1 – Map displaying the approximate distribution range of each of the eight subspecies from the *Chorthippus* group *binotatus* species complex. White dots indicate the geographic location of sampling sites for each taxon. Topographic background from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data). Detailed information for each taxon and sampling site is given in Table S1.

Here, we use genomic data obtained via restriction-site-associated DNA sequencing (ddRADSeq; Peterson *et al*. 2012) and a suite of coalescent-based methods of species tree estimation to elucidate the phylogenetic relationships and establishing species limits within the *Chorthippus* group *binotatus* species complex. Firstly, we (i) explore the impact of different demographic priors combinations and number of loci on the power and accuracy of genetic-based species delimitation inference. Afterwards, we (ii) employ a recently developed integrative approach that accommodates genetic and phenotypic data under a coalescent framework (Solís-Lemus *et al*. 2015) to examine and validate whether the obtained inferences are consistent with genetic-based analyses and explore the potential impacts of sex-based trait variation and the number and kind (traditional *vs*. geometric morphometrics) of phenotypic characters employed.

MATERIAL AND METHODS

SAMPLE COLLECTION

Between 2013 and 2014, we sampled the eight taxa (species and subspecies) that constitute the *Chorthippus* group *binotatus* species complex (*sensu* Defaut 2011, 2015; Table S1). In total, we sampled and performed genetic and morphological analyses for 80 individuals from the 8 taxa (5 males and 5 females per taxon) (Table S1). One specimen from *Chorthippus apricarius* Linnaeus 1758, a species also belonging to the subgenus *Glyptrobothrus* (Mayer *et al*. 2010; Nattier *et al*. 2011), was used as outgroup in phylogenetic analyses. We preserved whole specimens in 2 ml vials with 96% ethanol and stored at -20º C until needed for genomic and morphological analyses. Subspecies codes and further information on sampling locations are given in Table S1.

LINEAR AND GEOMETRIC MORPHOMETRIC ANALYSES

We took length measurements of left hind femur, left forewing and its median area, prozone, and pronotum following the procedure described in Noguerales *et al*. (2016). Afterwards, we calculated the ratios between some of these morphological traits previously considered of taxonomic value (Defaut 2011): (i) forewing length relative to femur length (FW/FL), (ii) median area length relative to forewing length (MAL/FL, see Fig. 2) and (iii) prozone length relative to pronotum length (PZ/PR, Fig. 2).

Differences among taxa in these three traits were analyzed separately for each sex using one-way ANOVAs. We took digital images of forewings and pronota in order to characterize shape variation for these two traits using geometric morphometric analyses. We used TPSDIG to digitize ten and eleven homologous landmarks for forewings and pronota, respectively (Rohlf 2015) (Fig. 2). Trait shape was characterized separately for each sex by a procustes fit aligned by principal axes using MORPHOJ v.1.05d (Klingenberg 2011). Allometric effects were examined by carrying out a multivariate regression of shape on centroid size by pooling the dataset by subspecies. Centroid size explained significantly an important proportion of variance of forewing (♂: 9.44%, $P = 0.009$; ♀: 22.63%, $P < 0.001$) and pronotum (♂: 5.21%, $P < 0.043$; ♀: 5.49%, $P < 0.038$) shape. Thus, we calculated new covariance matrices based on the residuals of the multivariate regressions before performing subsequent analyses.

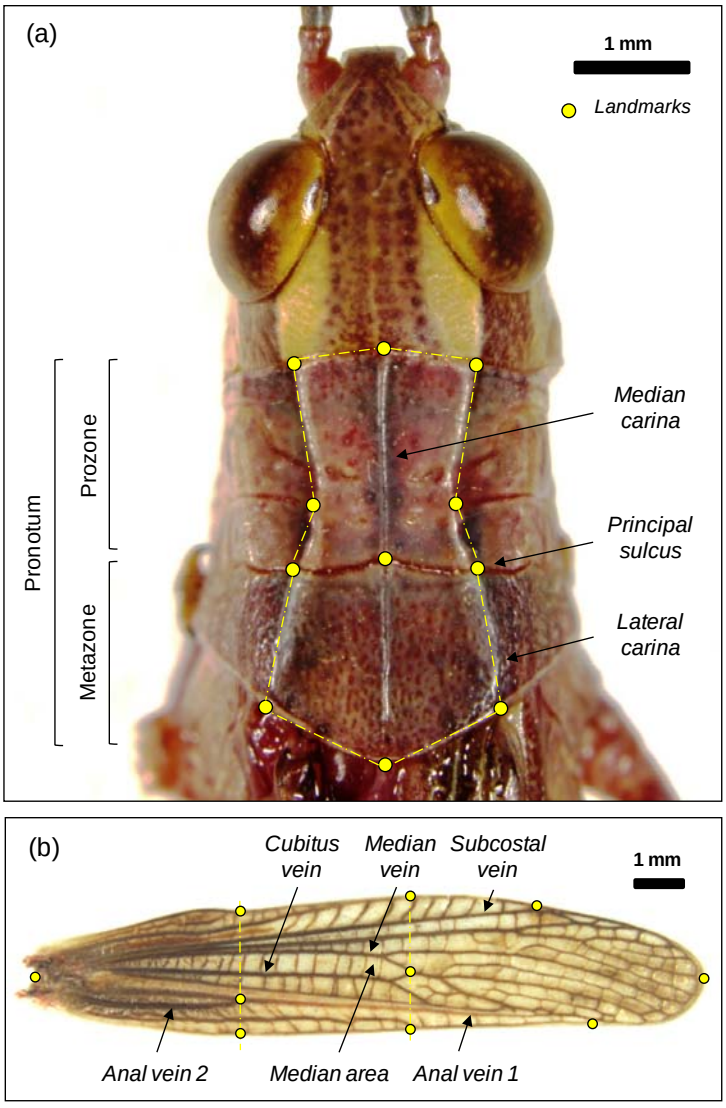


Figure 2 – Positions of the landmarks used to characterize pronotum (panel a) and forewing (panel b) shape in the different subspecies of the *Chorthippus* group *binotatus* species complex. The scale bar was used to standardize landmark distances to the same absolute scale across all images. We also indicate the main traits considered for defining the position of the forewing and pronotum landmarks, as well as the length of prozone and metazone.

We used MORPHOJ to examine shape variation for each trait and sex using a Principal Component Analysis (PCA) on the size-corrected residuals. We retained the first two principal components (PC), which explained a high proportion of forewing (♂: 76.04%; ♀: 82.21%, Fig. S2) and pronotum (♂: 57.67 %; ♀: 62.11%, Fig. S3) shape variation. Principal component scores were used separately by sex and trait in subsequent species delimitation analyses ($PC1_{FW,}$ $PC2_{FW}$ and $PC1_{PR,}$ $PC2_{PR}$). Forewing and pronotum shape variation were visually displayed using thin-plate spline diagrams as implemented in TPSSPLIN (Rohlf 2015). Finally, we conducted a Canonical Variates Analyses (CVA) using MORPHOJ to examine whether shape variation of both traits show significant differences among taxa. We calculated Mahalanobis distances (D_2) between taxa and tested their significance by permutation tests with 10 000 replicates.

LIBRARY PREPARATION AND SEQUENCING

We selected 5 individuals for each taxa (*n* = 40 individuals) and one individual from *C. apricarius* for subsequent genomic analyses. We employed a salt extraction protocol to purify genomic DNA from a hind femur of each specimen (Aljanabi & Martinez 1997). Genomic DNA from each specimen was individually barcoded and processed into a genomic library using the double-digestion restriction-fragment-based procedure (ddRADSeq) described in Peterson *et al*. (2012), with some minor modifications as detailed in Lanier *et al*. (2015) and Massatti & Knowles (2016). Briefly, DNA was double-digested using *Eco*RI and *Mse*I restriction enzymes (New England Biolabs), followed by the ligation of Illumina adaptors and unique 7-base-pair barcodes. Ligation products were pooled, size-selected between 475 and 580 base pairs (bp) using a Pippin Prep (Sage Science) machine, and amplified by iProof™ High-Fidelity DNA Polymerase (BIO-RAD) with 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, Canada).

BIOINFORMATICS PIPELINE

We followed a multi-approach to filter raw sequences and perform data quality controls (for a similar approach see Herrera *et al*. 2015). Firstly, we used the program *process_radtags* distributed as part of the STACKS v.1.35 pipeline (Catchen *et al*. 2013) to demultiplex and quality-filter the sequence reads.

We retained reads with a Phred score > 33, no adaptor contamination, and that had an unambiguous barcode and restriction cut site. Afterwards, raw sequence data quality was checked in FASTQC v.0.11.5 (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*) and sequences were trimmed to 129 bp using SEQTK (Heng Li, *https://github.com/lh3/seqtk*) in order to remove low-quality reads near the 3´ ends. Secondly, reads retained after *process_radtags* were submitted to a further quality-filtering using the program PYRAD v.3.0.66 (Eaton 2014) in order to convert base calls with a Phred score < 20 into *N*s and discard reads with > 2 *N*s (Fig. S4).

As in Takahashi *et al*. (2014), we used PYRAD to cluster retained reads within- and across samples considering different clustering thresholds of sequence similarity (W_{clust} = 0.85 and 0.90). Clusters with a minimum coverage depth < 5 were discarded (*d* = 5). Loci containing one or more heterozygous sites across more than 15% of individuals were excluded (*maxSH* = p.15) because a shared heterozygous site across many samples likely represents clustering of paralogs with a fixed difference rather than a true heterozygous site (Eaton 2014). Following to Hipp *et al*. (2014), the maximum number of polymorphic sites in a final locus was set to 20 (*maxSNPs* = 20). In a final filtering step, we generated three datasets by using different values for the minimum taxon coverage in a given locus discarding loci that were not present in at least 4, 10 or 20 samples (*minCov* = 4, 10 and 20, which represent the 10, 25 and 50% of samples, respectively). This procedure was repeated to generate datasets including or excluding *C. apricarius*, which was used as outgroup in phylogenetic analyses. Finally, following the approach described in Huang (2016), we checked our clustering and filtering output from PYRAD and trimmed our aligned clusters to 110 bp due to a systematic increase in sequence variation after this position (R scripts are available in Huang 2016). The resulting genomic datasets were used in subsequent analyses.

PHYLOGENOMIC INFERENCE - SPECIES TREE ESTIMATION

Several studies have highlighted that phylogenetic methods based on concatenation of genomic data can yield incongruent phylogenetic inferences (*e.g.* Kubatko & Degnan 2007; Song *et al.* 2012; Xi *et al*. 2014; see also Edwards *et al*. 2016; Springer & Gatesy 2016). Therefore, we built phylogenies using two coalescent-based methods of species tree estimation and different datasets of SNPs. Firstly, we generated species trees using SVDQUARTETS (Chifman & Kubatko 2014), a recently developed phylogenetic method implemented in PAUP* v.4.0a152 (Swofford 2002). It has been documented that

this method exhibits a good performance under many simulated conditions in comparison with similar alternative approaches for species tree estimation (Chou *et al*. 2015). SVDQUARTETS uses SNP data to infer phylogenetic relationships between quartets of taxa under the multi-species coalescent and then assembles these quartets into a species tree. Species trees were constructed by exhaustively evaluating all possible quartets from the dataset (a total of 101 270 quartets) and using the QFM quartet assembly algorithm. Uncertainty in relationships was quantified using nonparametric bootstrapping with 100 replicates. For SVDQUARTETS analyses, we analyzed the six SNPs matrices that included *C. apricarius* as outgroup and were obtained by setting different values of clustering thresholds (W_{clust} = 0.85 and 0.90) and minimum taxon coverage in a given locus (*minCov* = 4, 10 and 20). This allowed us to assess the impact of different proportions of missing data and number of loci on the recovered species tree (Takahashi *et al*. 2014; Leaché *et al*. 2015; Huang & Knowles 2016b).

Additionally, we generated a species tree using SNAPP v.1.3.0 (Bryant *et al*. 2012) plug-in for BEAST2 v.2.4.3 (Bouckaert *et al*. 2014). SNAPP uses biallelic SNPs to infer phylogenetic relationships and branch lengths and estimate current and ancestral population sizes. We ran analyses using different gamma prior distributions (gamma, α, β) for the ancestral population size parameter (θ value). The priors used were *G*(2, 200) , *G*(2, 2000) and *G*(2, 20 000), which would define different scenarios ranging from small to a large effective population sizes. The forward (*u*) and reverse (*v*) mutation rates were set to be calculated by SNAPP. We used the log likelihood correction and sampled the coalescent rate. The remaining parameters were left at default values. Due to computational burden, SNAPP analyses were only conducted using the SNPs matrix generated in PYRAD using W_{clust} = 0.90 and *minCov* = 10.

We used the R package *phrynomics* (Barb Banbury, *http://github.com/bbanbury/phrynomics*) to remove non binary and invariant SNPs, to code heterozygotes and to format input files for SNAPP. We ran two independent runs for each prior using different starting seeds for > 2 million generations, sampling every 1 000 steps. We used TRACER 1.6 to check stationarity and convergence of the chains and confirm that effective sampling sizes (ESS) for all parameters were > 200 (Rambaut *et al*. 2014). We removed 10% of trees as burn-in and combined tree and log files for replicated runs using LOGCOMBINER v.2.4.1. We used TREEANNOTATOR v.1.8.3 to obtain maximum credibility trees. The full set of likely species trees was displayed with DENSITREE v.2.2.1 (Bouckaert 2010), which is expected to show fuzziness in parts of the tree due to gene flow or other causes of phylogenetic conflict (*e.g.* Zarza *et al*. 2016).

MOLECULAR SPECIES DELIMITATION

Initially, we tested competing species delimitation hypotheses using the Bayes Factor species delimitation method (BFD*; Leaché *et al*. 2014) as implemented in SNAPP v.1.3.0 (Bryant *et al*. 2012) plug-in for BEAST2 v2.4.3 (Bouckaert *et al*. 2014). This method allows the comparison of alternative species delimitations scenarios in an explicit multi-species coalescent framework by calculating and comparing marginal likelihood estimates (MLE) for each model. We conducted a path sampling analysis of fourteen steps each consisting of 100 000 Markov chains Monte Carlo (MCMC) generations with 10 000 pre-burning generations, sampling each 100 steps using an α-value of 0.3 (see Card *et al*. 2016). These settings were sufficient to ensure convergence and obtain ESS > 200. The Bayes Factor (BF) test statistics (2**ln* (BF)) was calculated, where BF is the difference in MLE between two competing models (*base scenario - alternative scenario*). Six competing species delimitation hypotheses were defined based on current taxonomy (Defaut 2011), geographic distribution of putative species and phylogenetic analyses (see Table 1 for a description of the alternative hypotheses). Given that BFD* analyses are computationally highly demanding, we only ran them using a gamma prior distribution *G*(2, 2000) for the ancestral population size parameter (θ value) (*i.e.* the 'intermediate population size' scenario used for SNAPP analyses). It has been shown that estimates of effective population sizes and divergence times obtained by SNAPP strongly depend on the choice of priors, however, BFD* analyses seen to be robust to prior misspecification (Leaché *et al*. 2014; Rittmeyer & Austin 2015). In addition, SNAPP runs for species tree estimation using alternative gamma priors for θ parameter, such as *G*(2, 200) and *G*(2, 20 000), yielded the same topology (see 'Results' section). The species delimitation hypotheses were tested using the same matrix of SNPs and parameters employed for species tree estimation in SNAPP (see 'Phylogenomic inference - Species tree estimation' section). The estimation of marginal likelihoods of each hypothesis required ~15 days on fourteen Intel Xeon E7 2.8 GHz processors.

Secondly, we delimitated taxa using the BAYESIAN PHYLOGENETICS & PHYLOGEOGRAPHY program (BPP v.3.3; Yang & Rannala 2010, 2014), which has shown to be more accurate than alternative model-based methods for establishing species limits (Camargo *et al*. 2012). BPP analyzes multilocus sequence data of closely-related species under the multi-species coalescent model, employing a reversible-jump MCMC (rjMCMC) to estimate the posterior probability for different delimitation models and species trees. An earlier version of BPP (v.2.x) was reliant on user-specified guide trees and only evaluated those potentials models that were generated by collapsing or failing to collapse nodes on

such predefined topology (option A10; Yang & Rannala 2010; Rannala & Yang 2013; see also Olave *et al*. 2014; Zhang *et al*. 2014). The recently developed BPP version (v.3.x) jointly performs species tree estimation and species delimitation (option A11; Yang & Rannala 2014; Yang 2015). This version includes the nearest-neighbor interchange (NNI) algorithm, which is able to significantly change the topology of an input species tree and circumvents the need of specifying a fixed input guide tree. In this study, we used both guided and unguided implementations of BPP (options A10 and A11, respectively; for a similar approach see Eberle *et al.* 2016; Huang & Knowles 2016a; Koju *et al*. 2017). For guided analyses (option A10), we used as input guide trees the five different topologies yielded by SVDQUARTETS and SNAPP analyses.

We assessed the impact of different demographic scenarios on species delimitation inference considering several combinations of gamma priors for ancestral population size (θ) and root age (τ_0) (Leaché & Fujita 2010). Following Huang & Knowles (2016a), we considered four prior combinations: $\theta = G(1, 10)$, $\tau = G(1, 10)$, which would correspond to a large population size and deep divergence scenario (prior A); $\theta = G(1, 10)$, $\tau = G(2, 2000)$, large population size and recent divergence (prior B); $\theta = G(2, 2000)$, $\tau = G(1, 10)$, small population size and deep divergence (prior C); and $\theta = G(2, 2000)$, $\tau = G(2, 2000)$, small population size and recent divergence (prior D). A Dirichlet prior was assigned to other divergence time parameters (Yang & Rannala 2010). We used a uniform rooted tree-prior on the species tree topology (prior 1). We let the fine-tune settings to be automatically adjusted, allowing swapping rates to range between 0.30 and 0.70 (Yang 2015). Also, our runs were replicated employing the two species delimitation algorithms (0 and 1) to ensure that our results were consistent between different searching algorithms. Fine-tune parameters were adjusted to $\varepsilon = 2$ for algorithm 0, and $\alpha = 2$ and $m = 1$ for algorithm 1 (Yang & Rannala 2010).

Each analysis was run three times to confirm consistency among runs using different starting trees. Two of these runs were always initiated using as starting trees either one-species model (0000, all internal nodes are collapsed) and a fully resolved tree (1111, all internal nodes split) to ensure that the chains were mixing adequately. The third run was started from a randomly selected starting tree considering an intermediate situation (*i.e.* a partially resolved tree). Because BPP could suffer MCMC mixing problems when using large datasets (Yang & Rannala 2010; Rannala & Yang 2013, 2017), we explored the impact of the number of loci employed on species delimitation inference. Accordingly, we performed all guided analysis (option A10) using three genomic datasets consisting of 25, 50 and 200 sequence loci. Moreover, we used two additional larger genomic datasets of 500 and 1 000 loci for

performing jointly species tree estimation and species delimitation (A11 option). We generated these different subsets by randomly selecting loci from a dataset originally containing 32 317 sequence loci. This dataset was generated in PYRAD using W_{clust} = 0.90 and *minCov* = 10 without including outgroup. Considering all the replicates for the different combinations of settings, priors and number of loci, we ran 480 BPP independent runs including both guided and unguided analyses. We ran each analysis for 100 000 generations, sampling every 10 generations (10 000 samples), after a burning of 100 000 generations. Lineages delimited with a posterior probability (PP) of > 0.95 in all analyses were considered to be well supported.

INTEGRATIVE SPECIES DELIMITATION

iBPP v.2.1.2 is a recently developed program intended to delimit species by combining phenotypic and genetic data into a multispecies coalescent model (Solís-Lemus *et al*. 2015). iBPP was built upon the architecture of the early version of BPP v.2.1 (Yang & Rannala 2010; Rannala & Yang 2013) and incorporates models of evolution for continuous quantitative traits under a Brownian motion (BM) process (Solís-Lemus *et al*. 2015). We used non-informative priors for the BM control parameters ($v_0 = 0$; $\kappa_0 = 0$). All iBPP analyses were run considering the same demographic scenarios, settings, tree topologies, number of replicates, and subsets of loci described in the previous section for guided BPP analyses (option A10). To explore the influence of different phenotypic data on species delimitation inference, we ran iBPP using (i) only linear morphological data (FW/FL, PL/FL and PZ/PR), (ii) only geometric morphometrics data ($PC1_{FW,}$ $PC2_{FW}$ and $PC1_{PR,}$ $PC2_{PR}$) and (iii) a combination of the two datasets. As for above guided BPP analyses, all iBPP runs were conducted using three genomic datasets consisting of 25, 50 and 200 sequence loci. Given that Orthoptera show a remarkable sexual dimorphism (Hochkirch & Gröning 2008; Laiolo *et al*. 2013), we performed our analyses separately for each sex in order to assess the impact of sex-based trait variation on species delimitation. Finally, we also replicated all the above described analyses only considering phenotypic data (*i.e.* without genomic data) (Solís-Lemus *et al*. 2015; Dornburg *et al*. 2016; Eberle *et al*. 2016; Huang & Knowles 2016a). Each analysis was run 2-4 times to confirm consistency among runs performed using different random starting trees. In total, we ran 2 880 iBPP independent analyses taking into account all the different replicates and combinations of settings, priors, number of loci and morphological datasets. We ran each analysis for 100 000 generations, sampling every 10 generations (10 000 samples), after a conservative

burning of 300 000 generations. Lineages delimited with PP > 0.95 in all analyses were considered to be well supported.

RESULTS

SEQUENCING AND GENOMIC DATASETS

We obtained 104.29 million sequence reads, of which 88.87 million were retained after different filtering steps. On average, we retained 2.16 (± 0.4 SD) million reads per sample. All individuals were retained in the final analyses with a range of 1.57-3.13 million reads (Fig. S4). Clustering within samples using two values of clustering thresholds (W_{clust} = 0.85 - 0.90) yielded an average of 56 733 (± 7 933 SD) and 67 420 (± 9 309 SD) loci per sample, respectively. After clustering among samples using three different values for the *minCov* parameter, the resulting genomic datasets including outgroup contained 76 966, 24 961, 6 443 SNPs (using W_{clust} = 0.85, and *minCov* = 4, 10 and 20, respectively), and 97 680, 31 706, 7 939 SNPs (using W_{clust} = 0.90, and *minCov* = 4, 10 and 20, respectively).

LINEAR AND GEOMETRIC MORPHOMETRICS ANALYSES

We found that all three linear morphological traits differed significantly among taxa in both sexes (all $P_S < 0.001$) (see Fig. S1). Forewing shape variation analyses showed that individuals from the same subspecies were mostly clustered in the morphospace (Fig. S2). The two putative species (*C. binotatus* and *C. saulcyi*) were well separated in the morphospace, while the different subspecies within them partially overlapped for the two sexes (Fig. S2). Pronotum shape variation analyses revealed a similar clustering pattern between species (Fig. S3), although there was more overlap between putative subspecies, particularly in females (Fig. S3). CVA revealed that Mahalanobis distances (D_2) were significant between all subspecies for the two traits and sexes (Table S2).

PHYLOGENOMIC INFERENCE - SPECIES TREE ESTIMATION

Species tree estimation analyses using SVDQUARTETS provided five slightly different topologies depending on the SNP matrix used (Fig. 3). We found that the main uncertainty was the phylogenetic relationship of the taxon from Massif Central (*C. s. algoaldensis*) with the species into which it is currently assigned (*C. saulcyi*). The *C. s. algoaldensis* subspecies was strongly supported as an external sister lineage of either the *C. binotatus* group or the *C. saulcyi* group depending on the species tree. The phylogenetic relationships within the *C. saulcyi* group also exhibited considerable uncertainty. The lineages from the Pyrenees (*C. s. moralesi* and *C. s. vicdessossi*) and the lineages from northeastern Iberia and the Maritime Alps (*C. s. saulcyi* and *C. s. daimei*, respectively) generally clustered into two different clades, but the split nodes were not always well supported. Conversely, phylogenetic relationships within the putative *C. binotatus* group (*C. b. atlasi*, *C. b. armoricanus* and *C. b. binotatus*) were well-resolved and nodes presented a very high support, either when *C. s. algoaldensis* was included or not as its sister lineage (Fig. 3). The SVDQUARTETS species tree showing the highest overall support across all nodes was the one generated with the SNPs matrix (31 706 SNPs) obtained by setting PYRAD parameters to W_{clust} = 0.90 and *minCov* = 10 (see Fig. 3d). Such species tree inferred two major clades according to the classic taxonomy (*C. binotatus* and *C. saulcyi* groups). The *C. s. algoaldensis* subspecies was well-resolved as an external lineage with respect to the *C. saulcyi* group and the relationships within this group showed high support (see Fig. 3d).

SNAPP recovered a final dataset containing 3 245 biallelic SNPs (Table 1). SNAPP results did not provide a single well-supported topology, confirming the uncertainty detected for the *C. saulcy* group with SVDQUARTETS analyses (Fig. 4a). We used TREE SET ANALYSER as implemented in BEAST2 package to determine the proportion of trees supporting each topology. The most frequent topology (~41%) was similar to that recovered by SVDQUARTETS when analyzing the SNPs matrix obtained with W_{clust} = 0.90 and *minCov* = 10 (Fig. 3d, Fig. 4). The second and third most frequent topologies (~23% and ~20%), which were also similar to that yielded by SVDQUARTETS when using the matrix obtained with W_{clust} = 0.90 and *minCov* = 4, also supported ALG subspecies as an external lineage of the *C. saulcyi* group (see Fig. 3c). Both SNAPP topologies differed on the phylogenetic position of *C. s. saulcyi* and *C. s. daimei* subspecies within the *C. saulcyi* group and consequently the support of such split was very low in the consensus tree (Fig. 4b). SNAPP runs using *G*(2, 200) and *G*(2, 20 000) as priors for θ parameter converged on the same topology (results not shown).

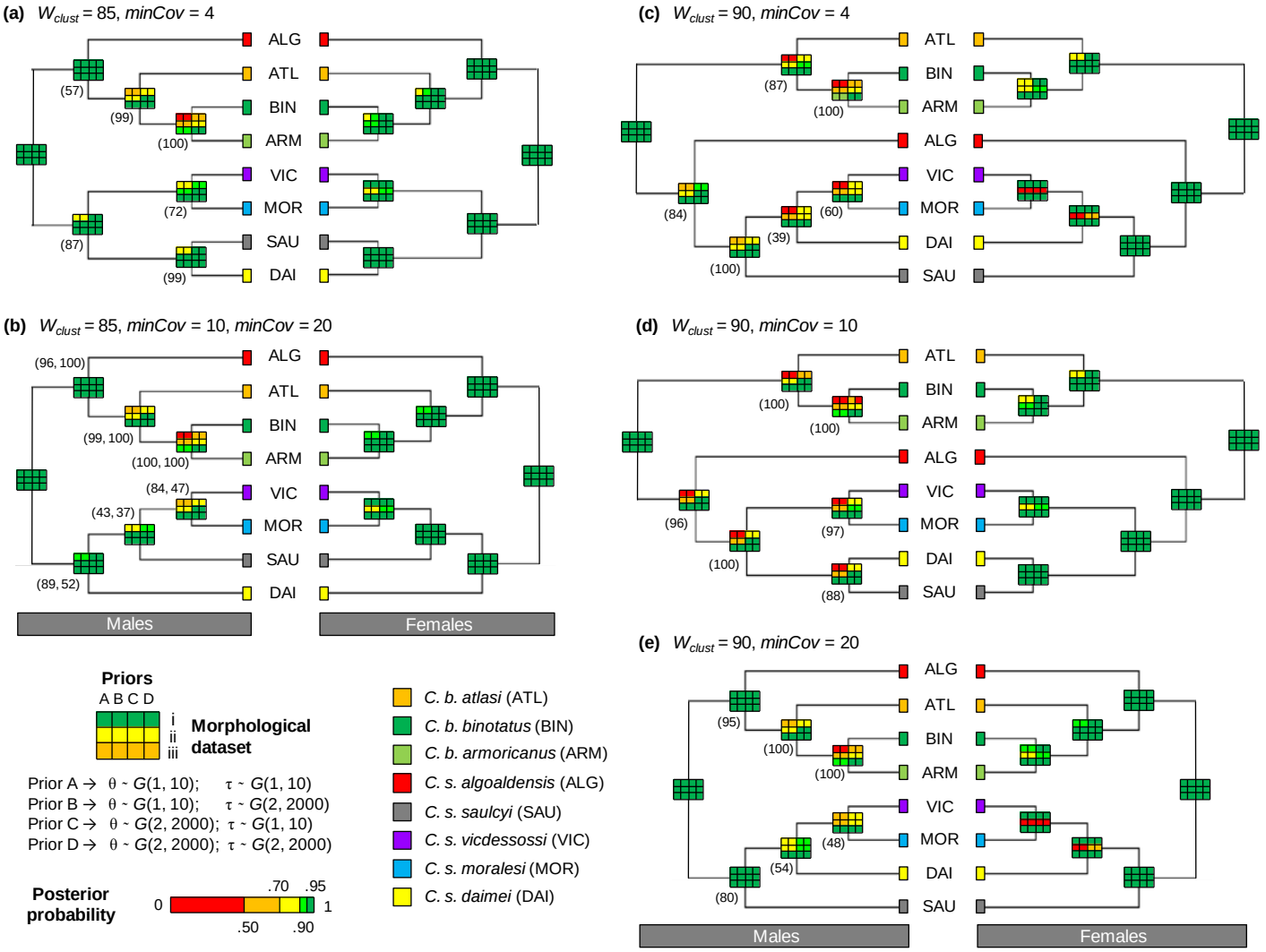


Figure 3 – Mean posterior probabilities (PP) of species delimitation across runs using four gamma prior combinations (gamma, α, β) for ancestral population size (θ) and root age (τ). Panels a-e represent five alternative guide trees recovered from SVDQUARTETS using different genomic datasets generated in PYRAD by setting different values of clustering thresholds (W_{clust} = 0.85 and 0.90) and minimum taxon coverage in a given locus (*minCov* = 4, 10 and 20). Bootstrapping values recovered from SVDQUARTETS are reported for each node. Results of species delimitation are based on analyses only using morphological data in iBPP. These analyses were performed employing three different morphological datasets: (i) only linear morphology, (ii) only geometric morphometrics and (iii) a combined dataset of both matrices. For each guide tree, species delimitation results are shown for males (left) and females (right). Color-coded boxes at each speciation split represent the mean PP for given combinations of demographic priors and morphological datasets (legend at left bottom). Subspecies codes as in Fig. 1 and Table S1.

SPECIES DELIMITATION

The result of the species delimitation analyses with the BFD* method strongly supported an eight species model (H_1, hypothesis 1), so all the lineages were identified as distinct species (Table 1). The second best-supported scenario (H_2, hypothesis 2) considered *C. s. moralesi* and *C.s. vicdessossi* as the same species but received much lower support (Table 1).

Both guided-by tree (option A10) and unguided (option A11) species delimitation analyses using BPP provided always a very high support for divergence (PP = 1.00) of all lineages regardless of guide tree topologies, demographic prior combinations, number of loci or delimitation algorithms. Similarly, integrative species delimitation analyses in iBPP combining genomic and morphological data supported all splitting events (PP = 1.00). This result was consistent irrespective of guide trees, prior combinations, number of loci, delimitation algorithms or morphological matrices based on distinct traits and sexes. Results from iBPP only based on morphological data provided contrasting support for nodes across analyses and were sensitive to sex-specific variation and the morphological matrix. Low support was found for the most recently diverged nodes from either *C. binotatus* or *C. saulcyi* groups (Fig. 3).

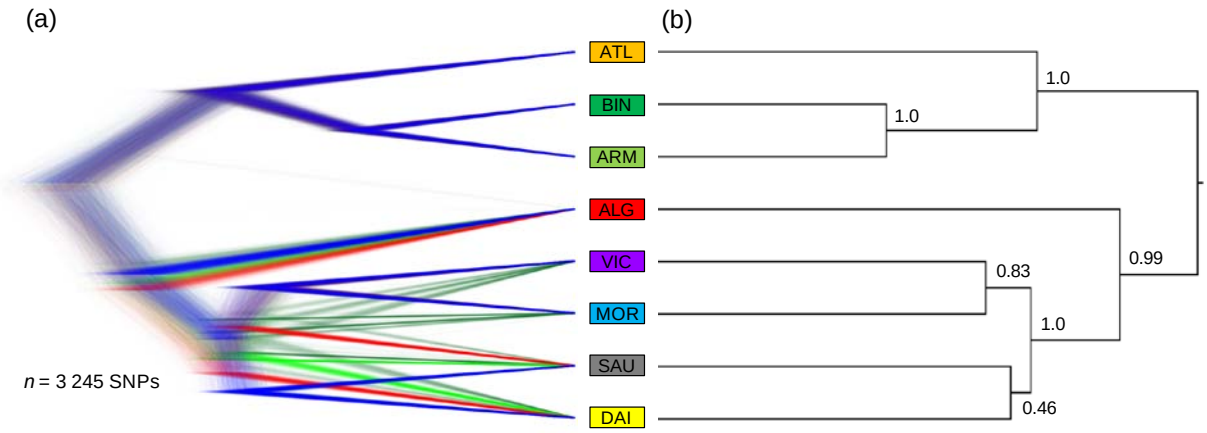


Figure 4 – Species tree inferred by SNAPP using biallelic SNPs (panel a). The original SNP matrix (*n* = 31 700 SNPs) was generated in PYRAD by setting a clustering threshold value of W_{clust} = 0.90 and a minimum taxon coverage value in a given locus of *minCov* = 10. The number of biallelic SNPs recovered by SNAPP is detailed at the left bottom of the figure. The species tree was obtained using a gamma prior distribution (gamma, α, β) of *G*(2, 2000) for the ancestral population size parameter (θ value). Runs performed using *G*(2, 200) and *G*(2, 20000) as priors yielded similar topologies. The first (blue), second (red) and third (green) most supported topologies are shown with different colors. Posterior probabilities for the most supported topology are indicated on the nodes of a maximum credibility tree (panel b). Subspecies codes as in Fig. 1 and Table S1.

Table 1 – 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**ln* (BF)) and their rank. Hypothesis 1 (H_1) was considered as base scenario. Also, we present the number of SNPs recovered by SNAPP from the original biallelic SNPs matrix for testing a given hypothesis. Such original matrix containing 31 700 SNPs was generated in PYRAD without outgroup and by setting a clustering threshold value of W_{clust} = 0.90 and a minimum taxon coverage value in a given locus of *minCov* = 10. Subspecies codes as in Fig. 1 and Table S1.

Species delimitation hypothesis (H_i)	Species	SNPs	MLE	BF	Rank
H_1: (ATL) (BIN) (ARM) (ALG) (SAU) (DAI) (MOR) (VIC)	8	3 245	-29 721.61	-	1
H_2: (ATL) (BIN) (ARM) (ALG) (SAU) (MOR+VIC) (DAI)	7	4 243	-39 643.74	18.40	2
H_3: (ATL) (BIN) (ARM) (ALG) (SAU+MOR+VIC) (DAI)	6	6 412	-57 940.66	20.49	3
H_4: (ATL) (BIN) (ARM) (ALG) (SAU+DAI) (MOR+VIC)	5	9 364	-83 359.68	21.78	4
H_5: (ATL) (BIN, ARM) (ALG) (SAU+DAI+MOR+VIC)	4	11 739	-105 638.43	22.47	5
H_6: (ATL) (BIN+ ARM) (ALG+SAU+DAI+MOR+VIC)	3	16 647	-153 962.18	23.45	6

Female-based morphological data increased the support for closely related lineages compared with male-based morphological data. Indeed, we found that analyses employing the combined morphological dataset (7 traits) provided an increasing support for divergence (Fig. 3). However, when employing male-based morphology, the usage of only geometric morphometrics data (4 traits) tended to provide higher support for divergence than only using data based on linear morphology (3 traits) (Fig. 3). Conversely, linear morphological traits provided an increasing support for divergence than geometric morphometrics data in females (Fig. 3). Analyses based on more conservative prior choices (priors C and D) generally increased support for splitting of the different lineages (Fig. 3).

DISCUSSION

Genome-wide data collected by ddRADSeq and a suite of coalescent-based methods allowed us to infer the phylogenomic relationships among the closely-related taxa forming the *Chorthippus* group *binotatus* species complex. The relationships among some taxa were unresolved or inconsistent across different genomic datasets, stressing the impact of certain settings during ddRADSeq data filtering and assembling on recovering the divergence history of the complex (*e.g.* Takahashi *et al*. 2014). The recently developed iBPP species delimitation framework integrating genetic and phenotypic data (Solís-Lemus *et al*. 2015) provided clear support for an eight-species taxonomy and our sensitivity tests

emphasized the power of such approach for resolving accurately species boundaries regardless of the different settings and topological hypotheses.

PHYLOGENOMIC RECONSTRUCTION

Genome-wide ddRADSeq data have been proven to yield considerable power for identifying species boundaries and clarifying relationships among putative taxa at shallow evolutionary scales (Bryson *et al*. 2016; Yoder *et al*. 2016; Beheregaray *et al*. 2017). However, settings during critical steps in bioinformatic processing of ddRADSeq data can have considerable impacts on sequence homology and matrix size (Rubin *et al*. 2012; Leaché *et al*. 2015; Nieto Montes-de Oca *et al*. 2017). Our results are in agreement with previous studies documenting the effects of different parameters, such as assembly thresholds and minimum number of individuals for recovering a given locus, on phylogenomic inferences (Takahashi *et al*. 2014; Leaché *et al*. 2015; Herrera & Shank 2016). The most robust topology in terms of bootstrap support of nodes was recovered using an inclusive genomic dataset including a high proportion of missing sites, which is in agreement with the findings by previous empirical studies (Rubin *et al*. 2012; Wagner *et al*. 2013). We found that using lower clustering thresholds of sequence similarity decreased around 20% the number of loci in the final genomic dataset, which may lead to a misidentification of orthologous loci and, consequently, generate conflicting topologies even for deep branches and varying branch support values (Takahashi *et al*. 2014). In this sense, our study advocates to consider extensive exploration of different settings during data filtering and assembling when using large genome-wide ddRADSeq data for phylogenetic and demographic inference (Razkin *et al*. 2016; see also Yang 2015; Huang & Knowles 2016b).

The most-supported topology obtained using either SVDQUARTETS or SNAPP yielded two well-resolved clades that correspond to the main nominative species groups (*C. binotatus* and *C. saulcyi*), two lineages whose divergence has been dated around 3 Ma (García-Navas *et al*. 2017). This topology suggests that putative *C. binotatus* and *C. saulcyi* species groups are monophyletic, which is in concordance with the most recent morphology-based taxonomy (Defaut 2011). Notwithstanding, three SVDQUARTETS species trees built using as input a genomic matrix obtained considering the lower clustering threshold, highly supported that the subspecies *C. s. algoaldensis* from Massif Central is a sister lineage of the *C. binotatus* group. This finding is in line with the taxonomic scheme proposed by Chopard (1952), who described the subspecies *algoaldensis* and considered it as a member of the

nominative species *C. binotatus*. Beyond genomic data, the intermediate position of the subspecies *algoaldensis* is supported by both morphological and ecological traits. All our linear and geometric morphometric analyses indicate that this subspecies is placed at an intermediate position along the morphological differentiation axes separating *C. binotatus* and *C. saulcyi* species groups (Fig. S1-S2). While all members from the *C. saulcyi* group exhibit gramineous feeding requirements, the subspecies *algoaldensis* also feeds on scrub-legume species (tribe *Genisteae*), an ecological specialization charactering the *C. binotatus* group (Defaut 2011; V.N. and J. O., personal observation). The phylogenomic relationships among the other putative subspecies of the *C. saulcyi* group (*i.e. C. s. moralesi, C. s. vicdessossi, C. s. saulcyi* and *C. s. daimei*) were not well resolved by any of the analyses employed in this study, which is in agreement with the difficulties already faced by previous taxonomic studies using non-genetic sources of information (Llucià-Pomares 2002; Defaut 2011). The low support for the phylogenetic relationships of these lineages and their similar branch lengths suggest a hard polytomy resulted from a simultaneous split event or, alternatively, the lack of resolution of our genomic dataset for resolving the evolutionary history of this group (Hoelzer & Melnik 1994; Shaffer & McKnight 1996; Campagna *et al*. 2015). The split of the *C. saulcyi* group has been estimated to take place during the Pleistocene (~1.5 Ma), likely driven by glacial cycles (Hewitt 1999), which is congruent with the rapid speciation characterizing the recent radiation of the *Gomphocerinae* subfamily (Mayer *et al*. 2010; Nattier *et al*. 2011, García-Navas *et al*. 2017).

While our analyses revealed considerable uncertainty in the phylogenomic relationships within the *C. saulcyi* group, we found consistently well-resolved branches within the nominative *C. binotatus* group. Phylogenomic analyses also indicate that the recently described French subspecies (*C. b. armoricanus*) and the Iberian one (*C. b. binotatus*) constitute two well-supported lineages (Fig. 4). The large distribution range of *C. binotatus* together with its montane character and narrow feeding requirements could have promoted long-term isolation of populations in different regions and, ultimately, the formation of cryptic evolutionary lineages (Defaut 2011, 2015), which is in agreement with the patterns of genetic structure at nuclear microsatellite markers and mtDNA found in recent phylogeographic studies on this species (Noguerales *et al*. 2017; see Chapter III in this Ph.D. Thesis).

GENOMIC SPECIES DELIMITATION

Results from genetic-based approaches of species delimitation (BFD* and BPP) were congruent and all putative infra-specific lineages were supported as distinct species. Our BPP analyses were robust to the use of alternative guide trees, conversely to that found by other studies documenting an important effect of guide tree misspecification and unreasonable topologies on species delimitation inferences (Leaché & Fujita 2010; Olave *et al*. 2014; Eberle *et al*. 2016; however see Zhang *et al*. 2014). The highest probability species tree recovered by unguided BPP analyses was also congruent with those obtained from other coalescent-based methods (SVDQUARTETS and SNAPP). This fact has been also highlighted by a recent meta-analysis documenting that the unguided BPP implementation yields similar results than analyses guided by reliable trees recovered from other approaches (Caviedes-Solis *et al*. 2015).

Despite literature on systematics is increasingly encouraging the employment of large genome-wide data for taxonomic delimitation (Lemmon & Lemmon 2013; McCormack *et al*. 2013), the accuracy of BPP for establishing species boundaries has been highlighted even when using a few loci (Camargo *et al*. 2012; Caviedes-Solís *et al*. 2015). Accordingly, we obtained virtually identical results using both small (25 loci, 2 750 bp) and large (1 000 loci, 110 000 bp) genomic datasets. A similar result was obtained in a recent empirical study testing the impact of the number of loci on recognizing taxonomic boundaries within stream-dwelling salamanders (Hime *et al*. 2016). Such study documented that accurate species delimitation inference was obtained using a subset of 10 loci from a broader 89 loci dataset generated by parallel tagged amplicon sequencing (Hime *et al*. 2016). Even though it has been suggested that BPP could suffer computational constrains using over 100 loci (Rannala & Yang 2017), we found that it performed well and recovered the true history of species divergence when handling large genomic datasets (1 000 loci, 110 000 bp) (Rannala & Yang 2017; see also Beheregaray *et al*. 2017).

At this point, we must note that our results should be interpreted taking into account an important conceptual issue of multispecies coalescent approaches when delimiting species boundaries. A recent simulation study has brought the attention on the fact that the multispecies coalescent model implemented in BPP may not be able to discriminate between genetic structure resulting from species boundaries or population-level processes (Sukumaran & Knowles 2017). This suggests that a remarkable intra-specific population genetic structure could be misidentified as inter-specific diverging lineages that may not represent full species (Pyron *et al*. 2016; Sukumaran & Knowles 2017). In turn,

such misidentification could be exacerbated when using vast genome-wide data due to its power to detect fine-grain population structuring even at small geographic scales, which could lead to an overinflation of the estimates of species-level diversity (Weir *et al*. 2016). As suggested by Sukumaran & Knowles (2017), the species delimitation inferences resulting from genetic-based multispecies coalescent approaches should be treated as hypothesis to be subsequently tested under an integrative framework incorporating non-genetic sources of information (Schlick-Steiner *et al*. 2010; Yeates *et al*. 2011; Edwards & Knowles 2014).

INTEGRATIVE SPECIES DELIMITATION

To date, integrative approaches of species delimitation have been based on sequential analyses employing genetic and non-genetic sources of data (Andujar *et al*. 2014; Watcher *et al*. 2015). However, recognizing species boundaries under this framework rely on qualitative and comparative analyses that do not provide a statistical confidence parameter to formally test alternative taxa limits hypotheses, hindering the repeatability and objectivity of such methods (Schlick-Steiner *et al*. 2010; Yeates *et al*. 2011). In order to overcome this issue, we validated genomic-based species boundaries inferences using jointly genomic and phenotypic data under a statistically and quantitative unified framework (iBPP, Solís-Lemus *et al*. 2015). In concordance with all foregoing analyses, we found strong support for the presence of eight-species. Our analyses indicate that the analyzed putative subspecies do not represent intermediate stages along the "speciation continuum" but merit full species status given its clear differentiation along genetic and morphological axis of divergence (Nosil *et al*. 2009).

Our iBPP results were consistent regardless the number of employed loci, demographic scenario, morphological dataset or sex-based trait variation, indicating that these factors had little impact on species identification. We found that the phylogenetic signal yielded by morphological datasets used alone (*i.e.* without genomic information) was strong enough to recover the eight-species delimitation model under any demographic scenario (Solís-Lemus *et al*. 2015; Huang & Knowles 2016a). Conversely to our expectations, integrating morphological and genetic information did not result in a fewer number of inferred species and, thus, we were not able to reject the notion that genomic-based inference overestimates the number of taxa (*e.g.* Eberle *et al*. 2016; Pyron *et al*. 2016). However, we found important differences in the properties of the traits considered in our study. Morphological matrices containing traits exhibiting on average a higher between-species variation (λ, *i.e.* phylogenetic

signal), such as forewing length, median plate relative length or the first variation axis of forewing shape (λ > 0.75, for both sexes), provided higher support for species splits (Fig. 3). Forewings are involved in several biological functions in Orthoptera such sound production during courtship, flight and thermoregulation (Thomas *et al*. 2001; Petit *et al*. 2006; Noguerales *et al*. 2016), therefore natural and sexual selection can be responsible of strong among-species variance in this structure (Klingenberg *et al*. 2010). Conversely, prozone length and pronotum shape showed on average low phylogenetic signal (λ = 0.30-0.60) indicating a higher level of phenotypic plasticity in these traits and lower informativeness for recognizing distinct taxa (Solís-Lemus *et al*. 2015). Beyond these differences among traits, we found that the cumulative information contained in an increasing number of traits provided the highest support for all nodes regardless of sex. This indicates that the use of a broad suite of well-known traits exhibiting different degrees of variation could be beneficial when aiming at determining species boundaries, particularly at early-stages of divergence where conflicting inferences are more prone to appear (see Solís-Lemus *et al*. 2015).

CONCLUSIONS

This study brings up the problem of assigning lineages to defined biological entities (species and subspecies), an issue that has been traditionally addressed using a single source of information (morphology or genetics) (Huang & Knowles 2016a). Our findings highlight the power of multispecies coalescent analyses to handle hundreds of loci and the importance of integrating genomic data and multiple sources of phenotypic information from the two sexes to capture subtle patterns of differentiation characterizing early-stages of speciation (Solís-Lemus *et al*. 2015; Rannala & Yang 2017). Future integrative species delimitation approaches should consider developing trait evolution models accommodating phenotypic variation resulted from sexual dimorphism (Solís-Lemus *et al*. 2015) and resolving the proneness of currently available methods to confound divergence patterns promoted by species *vs*. population processes when defining taxa limits (Sukumaran & Knowles 2017).

Acknowledgements

We wish to thank to Bernard Defaut for providing valuable information about sampling locations, Amparo Hidalgo-Galiana for library preparation, Anna Papadopoulou for her valuable help in data analysis and Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing. Logistical support was provided by Laboratorio de Ecología Molecular from Estación Biológica de Doñana (LEM-EBD) and Laboratorio de Genética from Instituto de Investigación en Recursos Cinegéticos (IREC). We also wish to thank to Centro de Supercomputación de Galicia (CESGA) for access to computer resources. This work was supported by Ministerio de Economía y Competitividad and European Social Fund (grant numbers: CGL2011-25053, CGL2014-54671-P, CGL2016-8742-R), Junta de Comunidades de Castilla-La Mancha - European Social Fund (grant numbers: POII10-0197-0167, PEII-2014023-P) and European Regional Development Fund (grant number: UNCM08-1E-018). VN was supported by a FPI pre-doctoral fellowship (fellow number: BES-2012-053741) from Ministerio de Economía y Competitividad. JO was supported by a Ramón y Cajal (fellow number: RYC-2013-12501) and Severo Ochoa (fellow number: SEV-2012-0262) research fellowships.

REFERENCES

Agapow PM, Bininda-Emonds ORP, Crandall KA*, et al.* (2004) The impact of species concept on biodiversity studies. *Quarterly Review of Biology*, 79, 161-179.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692-4693.

Andujar C, Arribas P, Ruiz C, Serrano J, Gómez-Zurita J (2014) Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation of *Mesocarabus* (Coleoptera: Carabidae). *Molecular Ecology*, 23, 4344-4361.

Beheregaray LB, Pfeiffer LV, Attard CRM*, et al.* (2017) Genome-wide data delimits multiple climate-determined species ranges in a widespread Australian fish, the golden perch (*Macquaria ambigua*). *Molecular Phylogenetics and Evolution*, 111, 65-75.

Bouckaert RR (2010) DENSITREE: making sense of sets of phylogenetic trees. *Bioinformatics*, 26, 1372-1373.

Bouckaert RR, Heled J, Kuehnert D*, et al.* (2014) BEAST2: a software platform for Bayesian evolutionary analysis. *Plos Computational Biology*; 10, e1003537.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29, 1917-1932.

Bryson RW, Jr., Savary WE, Zellmer AJ, Bury RB, McCormack JE (2016) Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. *Molecular Ecology*, 25, 3731-3751.

Camargo A, Morando M, Avila LJ, Sites JW, Jr. (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (*Squamata*: *Liolaemidae*). *Evolution*, 66, 2834-2849.

Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ (2015) Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*, 24, 4238-4251.

Card DC, Schield DR, Adams RH*, et al.* (2016) Phylogeographic and population genetic analyses reveal multiple species of *Boa* and independent origins of insular dwarfism. *Molecular Phylogenetics and Evolution*, 102, 104-116.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, 22, 4369-4383.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) STACKS: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124-3140.

Caviedes-Solis IW, Bouzid NM, Banbury BL, Leache AD (2015) Uprooting phylogenetic uncertainty in coalescent species delimitation: a meta-analysis of empirical studies. *Current Zoology*, 61, 866-873.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30, 3317-3324.

Chopard L (1952) *Faune de France: 56, Orthoptéroïdes*. Lechevalier, Paris, France.

Chou J, Gupta A, Yaduvanshi S*, et al.* (2015) A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, 16, S2.

Cigliano MM, Braun H, Eades DC, Otte D (2017) *Orthoptera Species File*. Version 5.0/5.0. *http://orthoptera.speciesfile.org/*

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? *Science*, 339, 413-416.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland (MA), USA.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, 56, 879-886.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (*Caelifera*, *Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Defaut B (2015) *Gomphocerippus binotatus armoricanus*, sous-espèce nouvelle de France (*Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 20, 25-26.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24, 332-340.

Dornburg A, Federman S, Eytan RI, Near TJ (2016) Cryptic species diversity in sub-Antarctic islands: a case study of *Lepidonotothen*. *Molecular Phylogenetics and Evolution*, 104, 32-43.

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30, 1844-1849.

Eberle J, Warnock RCM, Ahrens D (2016) Bayesian species delimitation in *Pleophylla chafers* (*Coleoptera*) - the importance of prior choice and morphology. *BMC Evolutionary Biology*, 16, 94.

Edwards DL, Knowles LL (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B-Biological Sciences*, 281, 20132765.

Edwards SV, Xi Z, Janke A*, et al.* (2016) Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94, 447-462.

Emerson KJ, Merz CR, Catchen JM*, et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 16196-16200.

Ence DD, Carstens BC (2011) SPEDESTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, 11, 473-480.

Freudenstein JV, Broe MB, Folk RA, Sinn BT (2016) Biodiversity and the species concept-lineages are not enough. *Systematic Biology*, in press. DOI: 10.1093/sysbio/syw098

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27, 480-488.

García-Navas V, Noguerales V, Cordero PJ, Ortego J (2017) Phenotypic disparity in Iberian short-horned grasshoppers (Acrididae): the role of ecology and phylogeny. *BMC Evolutionary Biology*, 17, 109.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, 270, 313-321.

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, 3467-3484.

Herrera S, Shank TM (2016) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, 100, 70-79.

Herrera S, Watanabe H, Shank TM (2015) Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. *Molecular Ecology*, 24, 673-689.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68, 87-112.

Hime PM, Hotaling S, Grewelle RE*, et al.* (2016) The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology*, 25, 5959-5974.

Hipp AL, Eaton DAR, Cavender-Bares J*, et al.* (2014) A framework phylogeny of the American oak clade based on sequenced RAD data. *Plos One*, 9, e102272.

Hochkirch A, Gröning J (2008) Sexual size dimorphism in Orthoptera (sens. str.) - a review. *Journal of Orthoptera Research*, 17, 189-196.

Hoelzer GA, Melnick DJ (1994) Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology & Evolution*, 9, 104-107.

Hotaling S, Foley ME, Lawrence NM*, et al.* (2016) Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. *Molecular Ecology*, 25, 2029-2045.

Huang J-P (2016) Parapatric genetic introgression and phenotypic assimilation: testing conditions for introgression between Hercules beetles (Dynastes, Dynastinae). *Molecular Ecology*, 25, 5513-5526.

Huang J-P, Knowles LL (2016a) The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single bayesian approach in Hercules beetles. *Systematic Biology*, 65, 685-699.

Huang H, Knowles LL (2016b) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, 65, 357-365.

Jackson ND, Carstens BC, Morales AE, O'Meara BC (2017) Species delimitation with gene flow. *Systematic Biology*, in press. DOI: 10.1093/sysbio/syw117

Klingenberg CP (2011) MORPHOJ: an integrated software packagefor geometric morphometrics. *Molecular Ecology Resources,* 11, 353-357.

Klingenberg CP, Debat V, Roff DA (2010) Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. *Evolution*, 64, 2935-2951.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, 56, 887-895.

Koju NP, He K, Chalise MK, *et al.* (2017) Multilocus approaches reveal underestimated species diversity and inter-specific gene flow in pikas (*Ochotona*) from southwestern China. *Molecular Phylogenetics and Evolution*, 107, 239-245.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, 56, 17-24.

Laiolo P, Illera JC, Obeso JR (2013) Local climate determines intra- and interspecific variation in sexual size dimorphism in mountain grasshopper communities. *Journal of Evolutionary Biology*, 26, 2171-2183.

Lamanna F, Kirschbaum F, Ernst ARR, *et al.* (2016) Species delimitation and phylogenetic relationships in a genus of African weakly-electric fishes (*Osteoglossiformes*, *Mormyridae*, *Campylomormyrus*). *Molecular Phylogenetics and Evolution*, 101, 8-18.

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, 24, 3688-3705.

Leaché AD, Chavez AS, Jones LN, *et al.* (2015) Phylogenomics of Phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, 7, 706-719.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B-Biological Sciences*, 277, 3071-3077.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, 63, 534-542.

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics,* 44, 99-121.

Llucià-Pomares D (2002) *Revision of the Orthoptera (Insecta) of Catalonia (Spain).* Monografías S.E.A., vol. 7, Zaragoza, Spain.

Massatti R, Knowles LL (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, 3974-3986.

Mayer F, Berger D, Gottsberger B, Schulze W (2010) Non-ecological radiations in acoustically communicating grasshoppers? In: Glaubrecht M (Ed.) *Evolution in action. Case studies in adaptive radiation, speciation and the origin of biodiversity*. Springer Verlag, Berlin, Germany.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66, 526-538.

McKay BD, Mays HL, Jr., Wu Y, *et al.* (2013) An empirical comparison of character-based and coalescent-based approaches to species delimitation in a young avian complex. *Molecular Ecology*, 22, 4943-4957.

Nattier R, Robillard T, Amedegnato C, *et al.* (2011) Evolution of acoustic communication in the *Gomphocerinae* (Orthoptera: Caelifera: Acrididae). *Zoologica Scripta*, 40, 479-497.

Nieto-Montes de Oca A, Barley AJ, Meza-Lazaro RN, *et al.* (2017) Phylogenomics and species delimitation in the knob-scaled lizards of the genus *Xenosaurus* (Squamata: Xenosauridae) using ddRADseq data reveal a substantial underestimation of diversity. *Molecular Phylogenetics and Evolution*, 106, 241-253.

Noguerales V, Cordero PJ, Ortego J (2017) Testing the role of ancient and contemporary landscapes on structuring genetic variation in a specialist grasshopper. *Ecology and Evolution*, 7, 3110-3122.

Noguerales V, García-Navas V, Cordero PJ, 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, 2129-2142.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, 24, 145-156.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, 59, 59-73.

Olave M, Ávila LJ, Sites JW, Jr., Morando M (2017) Hidden diversity within the lizard genus *Liolaemus*: genetic *vs.* morphological divergence in the *L. rothi* complex (*Squamata*: *Liolaeminae*). *Molecular Phylogenetics and Evolution*, 107, 56-63.

Olave M, Solá E, Knowles LL (2014) Upstream analyses create problems with DNA-based species delimitation. *Systematic Biology*, 63, 263-271.

Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, 7, 16.

Papadopoulou A, Knowles LL (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, 1501-1517.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One*, 7, e37135.

Petit D, Picaud F, Elghadraoui L (2006) Geometric morphometrics of Acrididae wings (Orthoptera, Caelifera): sex, stridulation and character. *Annales De La Societe Entomologique De France*, 42, 63-73.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, 25, 1367-1380.

Pyron RA, Hsieh FW, Lemmon AR, Lemmon EM, Hendry CR (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, 937-949.

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) *TRACER* v1.6. *http://beast.bio.ed.ac.uk/Tracer*

Rannala B, Yang Z (2013) Improved Reversible Jump Algorithms for Bayesian Species Delimitation. *Genetics*, 194, 245-253.

Rannala B, Yang Z (2017) Efficient Bayesian species tree inference under the multispecies coalesent. Systematics Biology, in press. DOI: 10.1093/sysbio/syw119

Razkin O, Sonet G, Breugelmans K*, et al.* (2016) Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex *Pyramidula*: the power of RADseq data. *Molecular Phylogenetics and Evolution*, 101, 267-278.

Rittmeyer EN, Austin CC (2015) Combined next-generation sequencing and morphology reveal fine-scale speciation in Crocodile Skinks (Squamata: Scincidae: Tribolonotus). *Molecular Ecology*, 24, 466-483.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *Plos One*, 7, e33394.

Rohlf FJ (2015) *TPS series*. Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA. *http://life.bio.sunysb.edu/morph*

Schlick-Steiner BC, Steiner FM, Seifert B*, et al.* (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, 55, 421-438.

Shaffer HB, McKnight ML (1996) The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. *Evolution*, 50, 417-433.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology*, 56, 896-906.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69, 492-507.

Song S, Liu L, Edwards SV, Wu S (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 14942-14947.

Springer MS, Gatesy J (2016) The gene tree delusion. *Molecular Phylogenetics and Evolution*, 94, 1-33.

Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 1607-16120

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts, USA.

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.

Thomas CD, Bodsworth EJ, Wilson RJ*, et al.* (2001) Ecological and evolutionary processes at expanding range margins. *Nature*, 411, 577-581.

Wachter GA, Muster C, Arthofer W*, et al.* (2015) Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: Megabunus). *Molecular Ecology*, 24, 863-889.

Wagner CE, Keller I, Wittwer S*, et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22, 787-798.

Weir JT, Haddrath O, Robertson HA, Colbourne RM, Baker AJ (2016) Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E5580-E5587.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, 56, 875-878.

Xi Z, Liu L, Rest JS, Davis CC (2014) Coalescent versus concatenation methods and the placement of *Amborella* as sister to Water Lilies. *Systematic Biology*, 63, 919-932.

Yang Z (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology*, 61, 854-865.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9264-9269.

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, 31, 3125-3135.

Yang Z, Rannala B (2017) Bayesian species identificacion under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology*, in press. DOI: 10.1111/mec.14093

Yeates DK, Seago A, Nelson L*, et al.* (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, 36, 209-217.

Yoder AD, Campbell CR, Blanco MB*, et al.* (2016) Geogenetic patterns in mouse lemurs (genus *Microcebus*) reveal the ghosts of Madagascar's forests past. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 8049-8056.

Zarza E, Faircloth BC, Tsai WLE*, et al.* (2016) Hidden histories of gene flow in highland birds revealed with genomic markers. *Molecular Ecology*, 25, 5144-5157.

Zhang C, Rannala B, Yang Z (2014) Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology*, 63, 993-1004.

Zhang C, Zhang D-X, Zhu T, Yang Z (2011) Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, 60, 747-761.

SUPPLEMENTARY MATERIAL

Table S1 – Geographical location, elevation and number of individuals used in ddRADseq analyses (*n*) for each subspecies from the *Chorthippus* group *binotatus* species complex (Defaut 2011, 2015). Subspecies codes as in Fig. 1. Asterisks on locality names indicate type locality where a given subspecies was originally described.

Code	Species/subspecies	*n*	Locality	Country	Latitude	Longitude	Elevation (m.a.s.l.)
ATL	*Chorthippus. binotatus atlasi* Defaut (1987)	5	Bab Bou Idir (Tazekka)*	Morocco	34.07527	-4.12262	1509
BIN	*Chorthippus binotatus binotatus* Charpentier (1825)	5	Campos de Otero (Sierra Nevada)	Spain	37.11013	-3.40512	2314
ARM	*Chorthippus binotatus armoricanus* Defaut (2015)	5	Bodelio (Malansac-Vannes)*	France	47.69921	-2.28028	64
ALG	*Chorthippus. saulcyi algoaldensis* Chopard (1952)	5	Col de Teste Rouge (Montselgues)	France	44.54336	3.99774	955
SAU	*Chorthippus saulcyi saulcyi* Krauss (1888)	5	Turó de l´Home (Montseny)	Spain	41.77240	2.44332	1622
VIC	*Chorthippus saulcyi vicdessossi* Defaut (2011)	5	La Prade (Vicdessos)*	France	42.73964	1.51423	1447
MOR	*Chorthippus saulcyi moralesi* Uvarov (1954)	5	Cortal de Rigat Maia (Err)	France	42.42762	2.05956	1800
DAI	*Chorthippus saulcyi daimei* Azam (1893)	5	Montagne Lanches (La Bastide)	France	43.74340	6.63963	1433
Outgroup	*Chorthippus apricarius* Linnaeus (1758)	1	Tour de Batere (Corsavy)	France	42.50860	2.57638	1430

Table S2 – Results of Canonical Variates Analyses (CVA) testing for forewing and pronotum shape differentiation between the different subspecies of *Chorthippus* group *binotatus*. Mahalanobis distances (D_2) for males are presented below the diagonal and for females above the diagonal. All Mahalanobis distances were significant. Subspecies codes as in Fig. 1 and Table S1.

		ATL	BIN	ARM	ALG	SAU	VIC	MOR	DAI
Forewing	ATL	-	8.23	10.04	18.11	11.34	17.35	26.22	27.66
	BIN	4.27	-	6.08	12.91	6.89	12.05	20.39	22.01
	ARM	4.34	3.23	-	12.43	7.16	11.26	18.30	19.93
	ALG	9.76	7.97	8.47	-	8.18	4.83	10.58	11.40
	SAU	8.07	6.54	7.21	4.28	-	7.45	16.01	17.67
	VIC	8.99	6.91	8.37	5.26	4.82	-	10.07	11.38
	MOR	11.99	8.98	10.03	5.97	6.97	5.21	-	4.54
	DAI	11.94	9.16	9.87	6.01	7.57	5.30	3.22	-
		ATL	BIN	ARM	ALG	SAU	VIC	MOR	DAI
Pronotum	ATL	-	5.20	12.07	9.38	10.28	10.53	10.39	9.27
	BIN	7.49	-	8.99	7.97	8.35	7.88	7.21	6.88
	ARM	5.94	4.67	-	10.31	9.73	8.96	5.81	8.01
	ALG	7.17	5.44	3.88	-	3.59	4.98	7.63	4.97
	SAU	6.55	7.33	4.79	4.69	-	4.54	7.50	5.20
	VIC	11.66	8.92	8.09	7.76	6.92	-	5.80	4.12
	MOR	10.02	5.94	7.18	4.82	7.87	9.05	-	4.98
	DAI	7.56	6.40	5.20	5.55	4.35	4.91	7.51	-

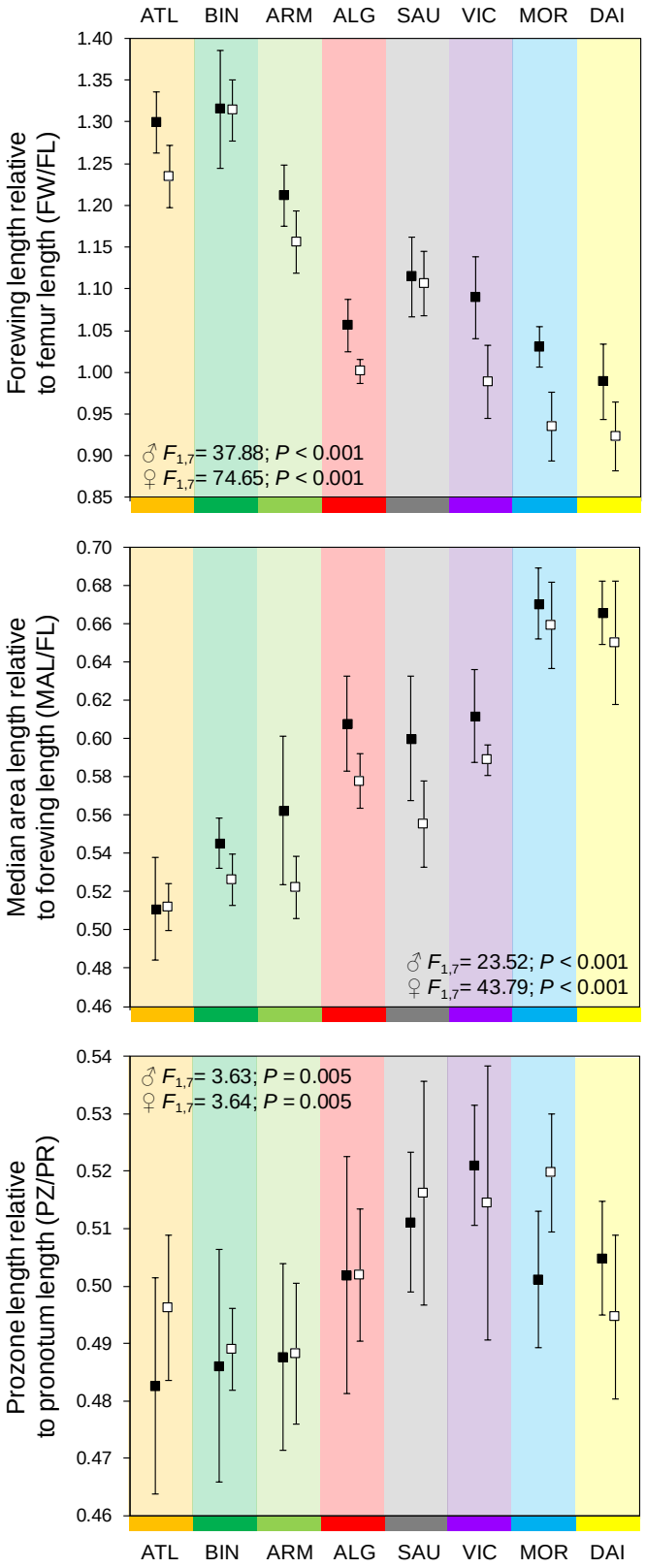


Figure S1 – Mean (± SD) values for linear morphological traits in the different subspecies of *Chorthippus* group *binotatus* (males: dark squares; females: open squares). Results of one-way ANOVAs testing for differences in traits among subspecies are shown for each sex separately. Subspecies codes as in Fig. 1 and Table S1.

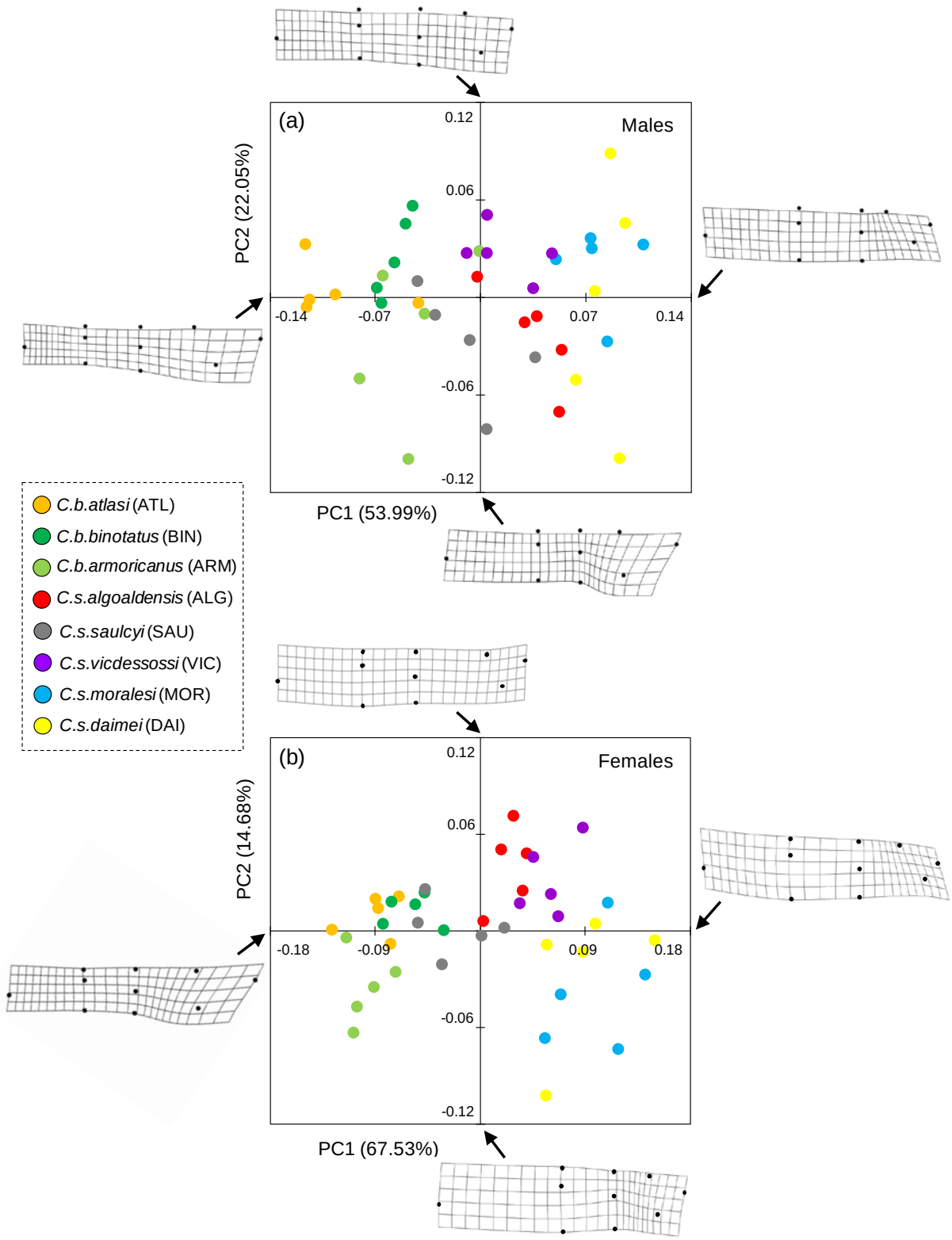


Figure S2 – Forewing shape variation in males (panel a) and females (panel b) for the different subspecies of *Chorthippus* group *binotatus* along the two first principal components (PCs). Thin-plate spline transformation grids show extreme shapes for each PC. Subspecies codes as in Fig. 1 and Table S1.

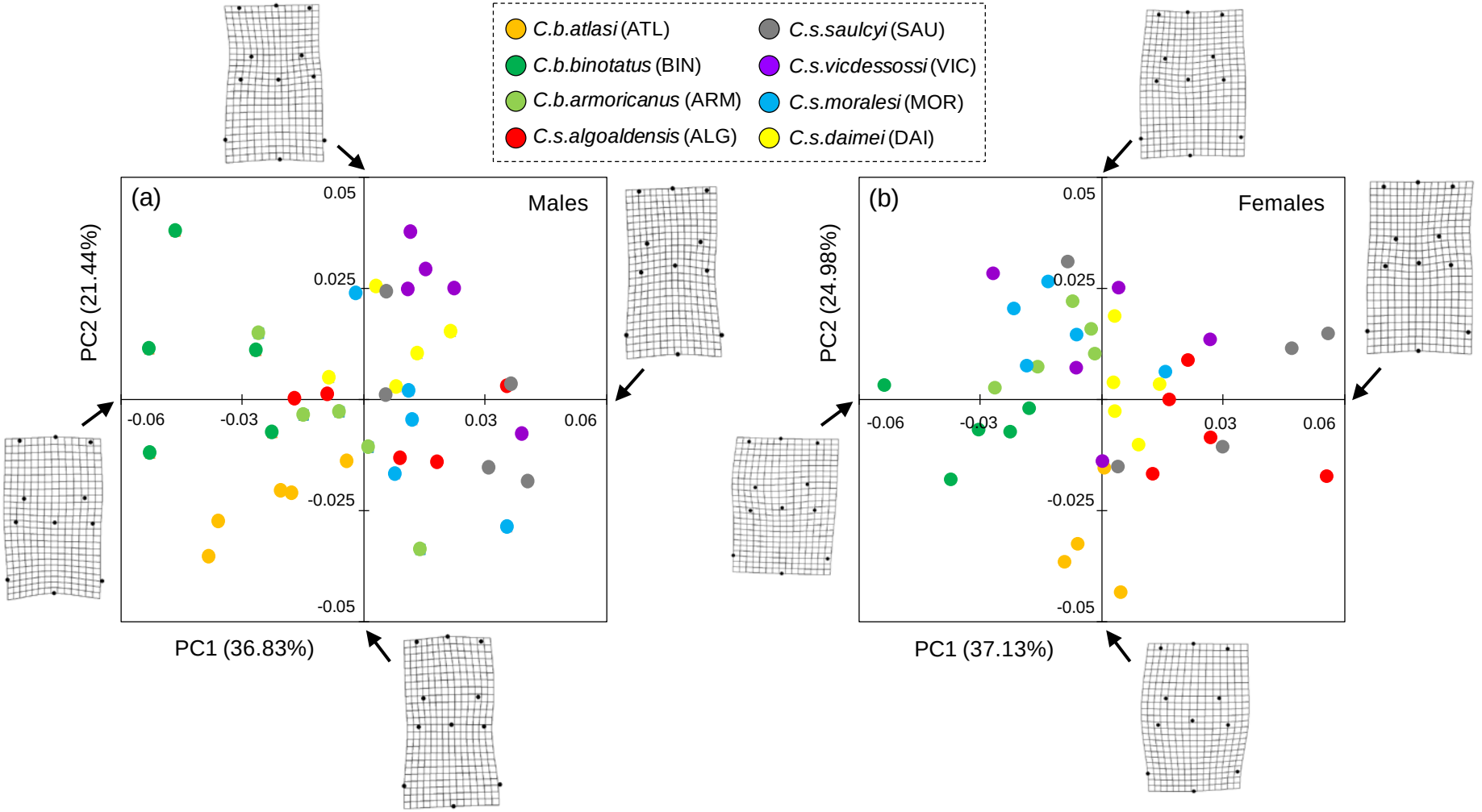


Figure S3 – Pronotum shape variation in males (panel a) and females (panel b) for the different subspecies of *Chorthippus* group *binotatus* along the two first principal components (PCs). Thin-plate spline transformation grids show extreme shapes for each PC. Subspecies codes as in Fig. 1 and Table S1.

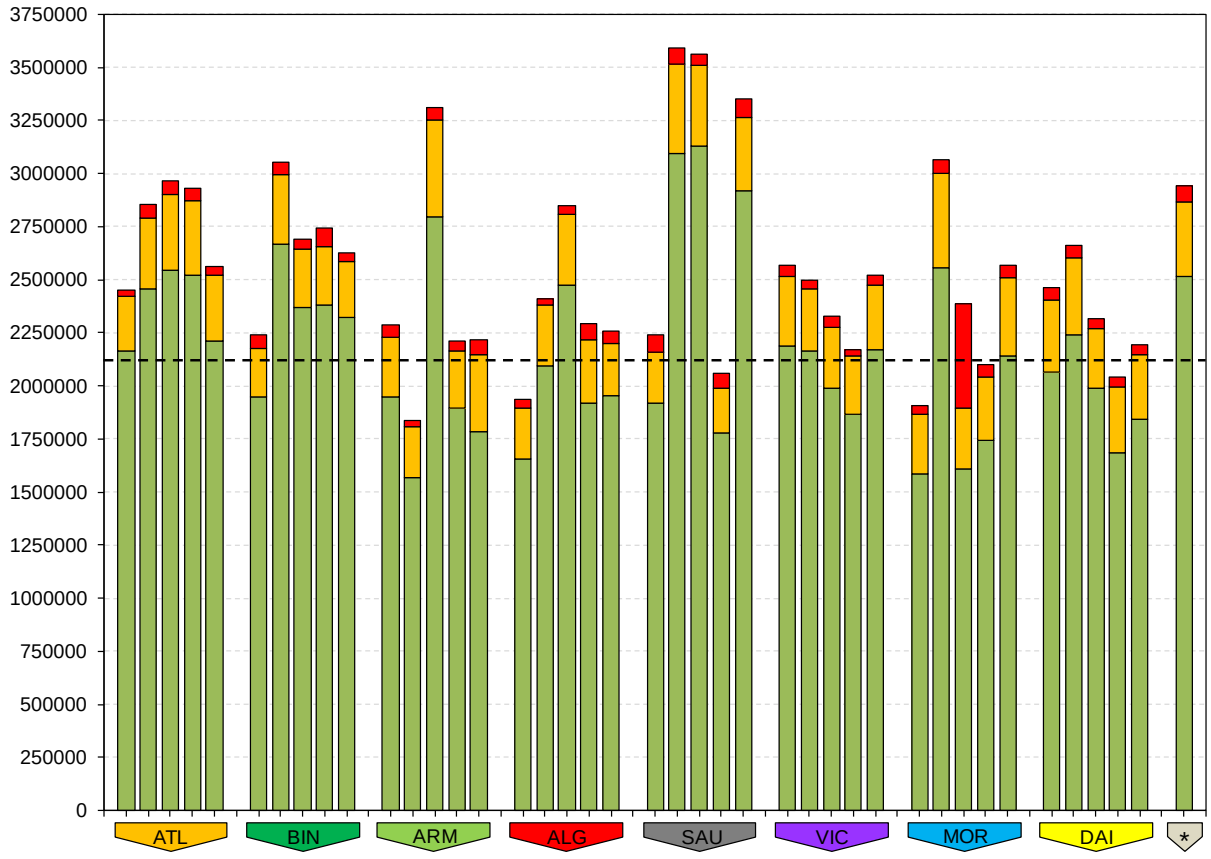


Figure S4 – Number of reads per individual before and after different quality filtering steps in STACKS and PYRAD. The cumulative stacked bars represent the total number of raw reads for each individual. Within each bar the red color represents the reads that were discarded by *process_radtags* as implemented in STACKS due to low quality, adapter contamination or ambiguous barcode. Orange color represents the reads that were discarded by PyRad (step 2) due to low quality or because contained > 2 Ns. Green color represents the total number of retained reads used to identify homologous loci by PYRAD. Dashed horizontal black line shows mean number of reads across all individuals. The specimen from Chorthippus apricarius that was used as outgroup is labelled with an asterisk (*) on the X-axis of the graph. Subspecies codes as in Fig. 1 and Table S1.

CAPÍTULO II

Isolation and characterization of 18 polymorphic microsatellites in the grasshoppers Chorthippus group binotatus (Orthoptera: Acrididae) species complex

VÍCTOR NOGUERALES, CONCHI CÁLIZ-CAMPAL, PEDRO J. CORDERO & JOAQUÍN ORTEGO

Conservation Genetic Resources, 7, 917-944 (2015) (published as Microsatellite Records)

Erratum: *Conservation Genetic Resources*, 8, 85-87 (2016)

Isolation and characterization of 18 polymorphic microsatellites in the *Chorthippus* group *binotatus* (Orthoptera: Acrididae) species complex

Abstract

The *Chorthippus* group *binotatus* (Orthoptera: Acrididae) is a species complex of grasshoppers distributed in southwest Europe and north Africa. They often form small and fragmented populations due to human-driven habitat destruction and/or the patchy distribution of their natural montane habitats. Here, we describe 18 novel polymorphic microsatellite markers obtained from the nominal subspecies *Chorthippus* binotatus *binotatus*. For this subspecies, the number of alleles ranged from 4 to 16 and their observed heterozygosity ranged from 0.21 to 0.94. Most of the loci also amplified and were polymorphic across all species/subspecies of the complex. These markers will be used to investigate the consequences of fragmentation and climate change on the demographic and evolutionary trajectories of these taxa of great conservation concern.

The *Chorthippus* group *binotatus* (Orthoptera: Acridae) is a complex of montane grasshoppers that includes at least two species and seven subspecies distributed through southwest Europe (France, Spain and Portugal) and north of Africa (Morocco) (Defaut 2011). All taxa of this group are oligophagus and feed exclusively on grasses or certain bush legumes (tribe *Genisteae*) (Defaut 2011). These grasshoppers generally form small and highly fragmented populations due to human-driven habitat destruction, which has resulted in the nominal subspecies *C. binotatus binotatus* is included as threatened in the French Red List (Pratz & Cloupeau 2010). Some narrow endemic taxa within this group also show highly fragmented populations due to the patchy distribution of their natural habitats, which are often restricted to small and isolated areas located in different mountain ranges (Defaut 2011). Here, we report the development of 18 polymorphic microsatellite loci from *C. binotatus binotatus* and test their functionality in the remaining species/subspecies of the group.

Genomic libraries were developed by GenoScreen (*www.genoscreen.fr*) using DNA extracted from eight individuals collected in different populations of *C. binotatus binotatus* (Table 1). Libraries were generated using a microsatellite enrichment method and sequenced using a 454 GS-FLX Titanium pyrosequencer (Malausa *et al*. 2011). A total of 398 primer pairs were designed and 66 were tested

Table 1 – Sampling sites of the eight individuals of Chorthippus binotatus binotatus used to construct the genomic library.

Location	Country	Latitude	Longitude
Teruel	Spain	40.32683	-1.21425
Pico Miravete	Spain	39.71325	-5.74611
Pico Almenaras	Spain	38.54352	-2.44000
El Morrón	Spain	37.86513	-1.57113
Plataforma de Gredos	Spain	40.27183	-5.25044
Puerto de San Martín	Spain	40.24580	-6.77755
Pico Almadén	Spain	37.74105	-3.53427
Cueva del Agua	Spain	37.33130	-3.51622

using 20 individuals of *C. binotatus binotatus* collected in Jaén, Spain (Table 2). We discarded 36 loci because they failed to amplify or produced inconsistent and multiple bands in agarose gels. Primers producing products of expected size were labelled with fluorescent dyes (6-FAM, PET, NED or VIC) to allow analysis on an automated DNA sequencer and determination of levels of polymorphism. We selected 18 polymorphic markers and the rest were discarded because they were monomorphic or produced non-resolvable electropherograms. Finally, the 18 polymorphic markers were tested for cross-amplification using 20-22 individuals from one location of each of the species/subspecies within the complex (Table 2).

Amplifications were conducted in 10-µL reaction volumes containing 5 ng of genomic DNA, 1X reaction buffer (EcoStart Reaction-Buffer, Ecogen), 2 mM MgCl2, 0.2 mM of each dNTP, 0.15 µM of each primer and 0.1 U of Taq DNA EcoStart-Polymerase (Ecogen). The PCR programme used was 9 min denaturing at 95ºC followed by 40 cycles of 30 s at 94ºC, 45 s at the annealing temperature (Table 3) and 45 s at 72ºC, ending with a 10 min final elongation stage at 72ºC. Amplification products were run on an ABI 310 Genetic Analyzer (Applied Biosystems) and genotypes were scored using GENEMAPPER 3.7 (Applied Biosystems). Calculation of observed and expected heterozygosities and tests for departure from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were performed in ARLEQUIN 3.5 (Excoffier & Lischer 2010).

Table 2 – Sampling sites of the seven taxa of the *Chorthippus* group *binotatus* species complex for which we tested the characteristics of the 18 microsatellite markers developed from *Chorthippus binotatus binotatus*.

Species/subspecies	Sample size	Location	Country	Latitude	Longitude
C. binotatus binotatus	20	Peña Blanca (Sierra Mágina)	Spain	37.74416	-3.54888
C. binotatus atlasi	22	Bab Bou-Idir (Tazzeka)	Morocco	34.07527	-4.12250
C. saulcyi saulcyi	20	Turo de l´Home (Montseny)	Spain	41.77222	2.44333
C. saulcyi moralesi	20	Hoyo de San Miguel (Saravillo)	Spain	42.56083	0.23027
C. saulcyi vicdessossi	20	La Prade (Vicdessos)	France	42.73944	1.51416
C. saulcyi algoaldensis	20	Col de Teste Rouge (Montselgues)	France	44.54333	3.99750
C. saulcyi daimei	20	Col de Peone (Peone)	France	44.11694	6.96000

We found no evidence of genotypic linkage disequilibrium at any pair of loci. For the nominal subspecies, five loci significantly deviated from HWE and MICRO-CHECKER analyses (Van Oosterhout *et al*. 2004) indicated that this could be due to the presence of null alleles (Table 4). The number of alleles per locus (N_A) ranged from 4 to 16 and observed (H_O) and expected (H_E) heterozygosity per locus ranged from 0.21 to 0.94 and from 0.66 to 0.92, respectively. Most of the loci also amplified and were polymorphic in the other taxa of the species complex (Table 4). Overall, the panel of microsatellites obtained constitutes a basic genetic tool to study the genetic diversity, demographic changes and genetic structure in the different species and subspecies of the *C. binotatus* group, many of them of great conservation concern.

Table 3 – Characteristics of 18 microsatellite markers developed from *Chorthippus binotatus binotatus*. For each locus, we list the primer pair, the repeat motif, the annealing temperature (T_a) and allele size range in base pairs (bp).

Locus	GenBank accession no.	Primer sequence (5'-3')	Repeat motif	T_a (º C)	Allele size range (bp)
Cbin02	KP792777	F: CAGCAACCTGAAGCCTGTT	$(TGT)_{12}$	55	204-315
		R: GACATGTCGGATTGGACC			
Cbin05	KP792778	F: AGAATGAGCCACAAATTAAAACA	$(CA)_{12}$	60	260-332
		R: CAAACTTCATATTGCAACCCC			
Cbin07	KP792779	F: CAAGTGGACAAACTCGAGCA	$(GA)_{12}$	55	150-246
		R: TCAGAGTGACGGCAATATACAA			
Cbin08	KP792780	F: TCTGCACATTAATTTTGATAGGG	$(TC)_{12}$	55	99-291
		R: CGATGTGTACATAGAGAAATCTAGTGA			
Cbin12	KP792781	F: AGAGTAGCGTGGAGAGCTGC	$(CCA)_{13}$	60	60-153
		R: CACTGTATCATGCTGAAGGCA			
Cbin15	KP792782	F: TGAATGGGCTTTCAAAGAGAA	$(CA)_{14}$	60	95-191
		R: ACAGCAGCCAATCTGGAGTT			
Cbin16	KP792783	F: TTGCTTCGTAGCTTATCGGTG	$(CA)_{15}$	60	240-382
		R: GCACCGGAGACTTGAAACAG			
Cbin27	KP792784	F: GTTCGGTTGACCGATGTTC	$(CA)_{19}$	50	103-221
		R: AACGTTCAGTTCGATTATCCG			
Cbin31	KP792785	F: AACAAGTTTGGACGTTTCGC	$(AC)_{21}$	60	62-166
		R: ATATCGAACCACGATTCGGC			
Cbin33	KP792786	F: CACTTTGAATGATTAATCTCCTGATTT	$(GT)_{22}$	55	90-192
		R: CGTCAAACGTAGTGAAGTTAGTAGG			
Cbin36	KP792787	F: AGAGATTCCAGAGCTATGCTGG	$(AC)_{23}$	60	72-254
		R: GCTGTAACACCACGACGGAC			
Cbin48	KP792788	F: TTCAGAGATGAAGAGGCTTGC	$(ACA)_{10}$	55	109-193
		R: CGGCTTTCTAGCATTTGTGC			
Cbin50	KP792789	F: AGGCAGGAGCTTAGAAACGTC	$(TCT)_{10}$	55	162-247
		R: ATCATCTGGCGAGCAAAAGT			
Cbin56	KP792790	F: AACAGCCGCCGAGATAAAG	$(CT)_{10}$	55	178-234
		R:CTGGACAATTCAACCGAAGG			
Cbin57	KP792791	F: CGGAGAGCACAAACACAAAA	$(TC)_{10}$	57	64-138
		R: GCAGTGGTGTGCAATAATCG			
Cbin59	KP792792	F: TAAATTACCTCCCCTCAATCCA	$(ATC)_{10}$	55	132-237
		R: TCAATTCGAACTGGACAGTCA			
Cbin62	KP792793	F: CCTCTCGTTGTATACACTCCCA	$(TG)_{10}$	52	244-358
		R: GGCCCTGGTCCAATAAGAAT			
Cbin66	KP792794	F: AAGTCCCCTCTACATCCCCT	$(GA)_9$	55	217-391
		R: CTCGAGCTTCCATCAGCTTC			

Table 4 – Data on polymorphism for the 18 microsatellite markers developed from *Chorthippus binotatus binotatus*. Data are provided for 20-22 genotyped individuals from one location for each species/subspecies (see Table 2).

	C. binotatus. binotatus			*C. binotatus atlasi*			*C. saulcyi saulcyi*			*C. saulcyi moralesi*			*C. saulcyi vicdessossi*			*C. saulcyi algoaldensis*			*C. saulcyi daimei*		
Locus	N_A	H_O	H_E	N_A	H_O	H_E	N_A	H_O	H_E	N_A	H_O	H_E	N_A	H_O	H_E	N_A	H_O	H_E	N_A	H_O	H_E
Cbin02	6	0.35	0.68	2	0.04	0.04	8	0.40*	0.82	6	0.35*	0.73	8	0.33*	0.84	1**	0.00	0.00	5	0.11*	0.55
Cbin05	9	0.83	0.88	2	0.09	0.08	8	0.80	0.80	7	0.40	0.50	5	0.65	0.68	7	0.73	0.82	8	0.50	0.47
Cbin07	6	0.64	0.75	9	0.59	0.70	11	0.47*	0.84	5	0.30	0.45	4	0.33	0.49	2	0.31	0.34	4	0.16*	0.42
Cbin08	14	0.72*	0.91	11	0.18*	0.89	14	0.25*	0.92	15	0.38*	0.92	6	0.06*	0.64	16	0.75*	0.93	16	0.50*	0.90
Cbin12	4	0.38	0.51	5	0.62	0.72	4	0.10	0.14	2	0.00	0.09	4	0.35	0.61	4	0.10	0.14	3	0.10	0.18
Cbin15	9	0.94	0.77	18	0.90	0.92	14	0.63*	0.89	7	0.45*	0.73	7	0.69	0.76	3	0.82*	0.61	4	0.89*	0.65
Cbin16	11	0.21*	0.90	14	0.61*	0.92	6	0.00*	0.90	5	0.33	0.84	2	0.99	0.99	5	0.40	0.84	5	0.00*	0.71
Cbin27	11	0.50*	0.89	2	0.00*	0.35	18	0.50*	0.95	18	0.75	0.92	11	0.80	0.84	20	0.95	0.95	22	0.90	0.96
Cbin31	10	0.50	0.71	12	0.93	0.89	13	0.35*	0.85	12	0.57	0.81	11	0.37*	0.90	10	0.60*	0.88	21	0.55*	0.96
Cbin33	16	0.66	0.91	16	0.80	0.93	16	0.89	0.94	11	0.70	0.87	13	0.56	0.77	7	0.55	0.61	11	0.84	0.82
Cbin36	13	0.31*	0.92	15	0.63*	0.92	20	0.76*	0.95	19	0.31*	0.96	14	0.26*	0.93	9	0.35*	0.73	25	0.78	0.97
Cbin48	8	0.83	0.87	8	0.73	0.73	5	0.36	0.59	2	0.06	0.06	7	0.15*	0.47	5	0.80	0.75	6	0.26*	0.51
Cbin50	9	0.83	0.87	7	0.55	0.78	7	0.14*	0.76	2	0.14	0.53	6	0.28*	0.85	8	0.47*	0.77	7	0.30*	0.84
Cbin56	10	0.50*	0.80	10	0.38*	0.78	8	0.31*	0.84	6	0.15*	0.57	7	0.45	0.64	5	0.38	0.53	9	0.21*	0.84
Cbin57	6	0.44	0.78	5	0.23*	0.50	7	0.40*	0.85	3	0.16*	0.50	4	0.25*	0.61	5	0.40*	0.78	5	0.18*	0.75
Cbin59	11	0.88	0.89	10	0.57	0.66	11	0.63	0.83	7	0.90	0.82	8	0.50	0.59	10	0.80	0.88	11	0.80	0.89
Cbin62	8	0.35	0.66	2	0.05	0.05	3	0.00*	0.80	7	0.75	0.79	8	0.50	0.85	6	0.16*	0.75	10	0.33*	0.94
Cbin66	11	0.58	0.86	5	0.63	0.67	12	0.31*	0.76	10	0.66	0.85	7	0.40*	0.64	6	0.75	0.80	11	0.20*	0.77

For each locus and each subspecies, we list the number of observed alleles (N_A) and observed (H_O) and expected (H_E) heterozygosity.

** Locus showing significant deviation from Hardy-Weinberg equilibrium after sequential Bonferroni correction for α=0.05.*

*** Locus Cbin02 was monomorphic in the analysed population for the subspecies Chorthippus saulcyi algoaldensis.*

Acknowledgements

This work received financial support from the projects POII10-0197-0167 and CGL2011-25053. During this work V.N. was supported by a FPI pre-doctoral fellowship (BES-2012-053741). J.O. was supported by a Severo Ochoa (EBD) (SEV-2012-0262) research contract.

REFERENCES

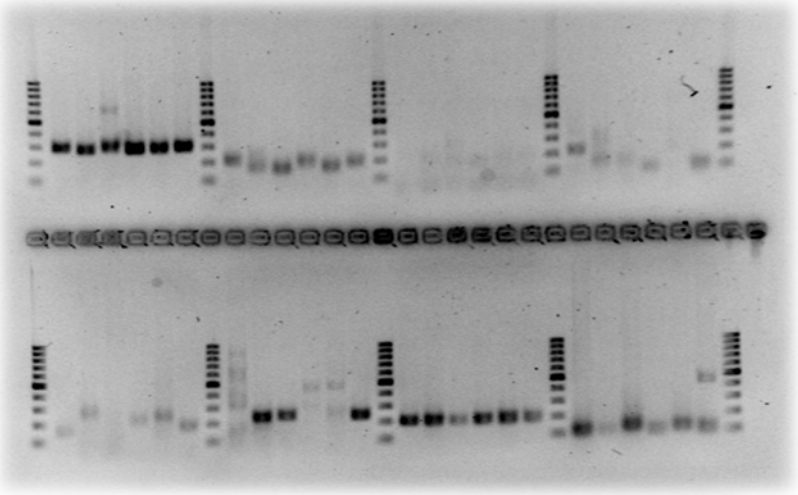
Defaut B (2011) Révision préliminaire des *Chorthippus* du groupe *binotatus* (Charpentier, 1825) (Caelifera, Acrididae, Gomphocerinae). *Materiaux Orthopériques et Entomocénotiques*, 16, 17-54.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* 10, 564-567.

Malausa T, Gilles A, Meglécz E, *et al*. (2011) High throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNAlibraries. *Molecular Ecology Resources*, 11, 638-644.

Platz J, Cloupeau R (2010) Liste rouge commentée des Orthopéres de la región Centre. *Materiaux Orthopériques et Entomocénotiques*, 15, 17-33.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes,* 4, 535-538.



CAPÍTULO III

Inferring the demographic history of an oligophagous grasshopper: effects of climatic niche stability and host-plant distribution

VÍCTOR NOGUERALES, PEDRO J. CORDERO & JOAQUÍN ORTEGO

(submitted)

Inferring the demographic history of an oligophagous grasshopper: effects of climatic niche stability and host-plant distribution

Abstract

Understanding the consequences of past environmental changes on the abiotic and biotic components of the landscape and deciphering their impacts on the demographic trajectories of species is a major issue in evolutionary biogeography. In this study, we combine nuclear and mitochondrial genetic data to study the phylogeographical structure and lineage-specific demographic histories of the scrub-legume grasshopper (*Chorthippus binotatus binotatus*), a montane taxon distributed in the Iberian Peninsula and France that exclusively feeds on certain scrub-legume species. Genetic data and paleo-distribution modelling indicate the presence of four main lineages that seem to have diverged in allopatry and long-term persisted in Iberian and French refugia since the Mid Pleistocene. Comparisons of different demographic hypotheses in an Approximate Bayesian Computation (ABC) framework supported a population bottleneck in the northwestern French clade and paleo-distribution modelling indicate that the populations of this lineage have experienced more severe environmental fluctuations during the last 21 000 years than those from the Iberian Peninsula. Accordingly, we found that nuclear genetic diversity of the populations of scrub-legume grasshopper is positively associated with local stability of suitable habitats defined by both Pleistocene climate changes and historical distributional shifts of host-plant species. Overall, our study highlights the importance of integrating the potential effects of abiotic (*i.e.* climate and geography) and biotic components (*i.e.* inter-specific interactions) into the study of the evolutionary and demographic history of specialist taxa with narrow ecological requirements.

INTRODUCTION

Quaternary climatic fluctuations have reshaped the distribution of worldwide biotas and impacted the demographic trajectories of most organisms (Hewitt 2004a, 2011). In Europe, the prevailing paradigm establishes that most species responded to cooling phases by southern distributional shifts and survival in glacial refugia predominantly located in Mediterranean peninsulas that maintained the source populations from which northern areas were re-colonized during interglacial periods (Hewitt 1999; Schmitt 2007). Many empirical (Hewitt 2004b; Cornille *et al*. 2013) and theoretical studies (Excoffier & Ray 2008; Arenas *et al*. 2012) have analyzed the genetic footprints of such distributional shifts and demonstrated the validity of the expansion-contraction model to explain patterns of genetic diversity and structure in many organisms (Taberlet *et al*. 1998; Provan & Bennet 2008). The refugia hypothesis has

been invoked as an excellent conceptual framework to understand why populations from areas experiencing long-term climatic stability harbor a higher and unique genetic diversity in comparison with those from regions submitted to more climate instability (Carnaval *et al*. 2009; Yannic *et al*. 2014). Nevertheless, most research on this topic has focused on temperate species and the specific predictions of the expansion-contraction model may not be valid for montane or cold-adapted species (Galbreath *et al*. 2009; Ricanova *et al*. 2013). Species adapted to cool environments are likely to have experienced range expansions and higher connectivity among populations in glacial periods and persisted forming isolated populations in interglacial refugia during warming phases (Willis & van Andel 2004; Galbreath *et al*. 2009). Accordingly, a growing body of literature is yielding strong evidence about many species that survived in cryptic refugia located in the northernmost areas of their present-day distribution (Stewart & Lister 2001; Stewart *et al*. 2010). In comparison with southern refugia, northern refugia are less predictable through time and space and, therefore, the impacts of climate-driven changes in landscape composition on the demography of cool-adapted species are still poorly understood (Vega *et al*. 2010; Parducci *et al*. 2012).

Beyond the importance of past environmental fluctuations, the distribution and abundance of a species are often shaped by many other ecological processes that are not necessarily captured by its climatic niche envelop (Mouritsen & Poulin 2002; Hampe 2004). This is the case of many specialist organisms whose demography is expected to be determined not only by changes in the spatial distribution of climatically suitable areas but also by range shifts experienced by the hosts on which they depend for feeding or development (*e.g.* Tsai & Manos 2010; Cangi *et al*. 2013). Climate-based distributional shifts of host species can have a considerable impact on specialist taxa, particularly if differences in species-specific environmental tolerances result in the responses to climate change of the latter inferred on the basis of bioclimatic niche models are largely uncoupled from those experienced by their hosts (Tsai & Manos 2010; Borer *et al*. 2012). This can lead to some species are not able to persist in regions falling within its optimal climate niche if their host-taxa are absent and, conversely, they may maintain viable populations in climatically sub-optimal areas as long as they sustain stable or abundant host populations (Jackson & Overpeck 2000; Criscione & Blouin 2004). Under this scenario, it would be expected that population genetic connectivity and persistence are shaped by the combined stability of suitable habitats for both the focal taxon and its hosts and that the impact of the latter is more pronounced in specialist species exclusively dependent on one or a few host taxa (Laukkanen *et al*. 2014). A biologically realistic approach would be to integrate information on the climate niche of the focal species (as a proxy of fundamental niche concept) and well-understood ecological relationships

with host taxa (as a proxy of realized niche concept) (Hutchinson 1957; Wharton & Kriticos 2004; Freeman & Mason 2015). This approach can potentially offer a more comprehensive picture about the demographic history of a species and how it is being shaped by the abiotic and biotic components of the landscape (Jackson & Overpeck 2000; Svenning *et al*. 2011). A number of studies have addressed the potential negative impacts of ongoing climate change on biodiversity as a consequence of spatiotemporal mismatches in distributional shifts between host-plants and their specialist herbivores (*e.g.* Schweiger *et al*. 2008). However, studies looking backward in time to analyze the impacts of Pleistocene distributional shifts of host-plants on the population dynamics of their associated phytophagous taxa are almost absent (for an exception see Tsai & Manos 2010).

The scrub-legume grasshopper, *Chorthippus binotatus binotatus* (Charpentier 1825) (Orthoptera: Acrididae), is a well-suited study system to analyze the influence of Quaternary climatic fluctuations on the demographic history of a phytophagous species presenting narrow ecological requirements. This taxon is distributed throughout France, Spain and Portugal (Defaut 2011) and occupies primarily montane habitats between 900 and 2 500 m.a.s.l., often forming highly isolated populations separated by unsuitable low elevation areas (Noguerales *et al*. 2017). The scrub-legume grasshopper can be also exceptionally found at low elevations (even at sea level) near to the Atlantic coast, where it forms small populations in highly isolated patches of suitable habitat (Picaud *et al*. 2003; Pratz & Cloupeau 2010). Quaternary climatic fluctuations are likely to have strongly impacted the genetic structure of this species, whose populations are expected to have persisted in high elevation and cold refugia during interglacials and experienced regional expansions during glacial periods. The scrub-legume grasshopper is an oligophagous taxon linked to certain scrub-legume communities on which it depends for feeding and refuge (Llucià-Pomares 2002; Picaud *et al*. 2002; Defaut 2011). Specifically, this grasshopper exclusively feeds on thirteen scrub-legume species belonging to very closely related genera (*Cytisus*, *Echinospartium*, *Erinacea*, *Genista* and *Ulex*) from the tribe *Genisteae* (Llucià-Pomares 2002; Defaut 2011 and references therein; V.N., P.J.C. and J.O., pers. obs.). Plant species of this tribe can be found up to 2 500 m.a.s.l. and exhibit important differences in their specific ecological requirements although they share a general preference for montane and cool habitats. Nowadays, some of these scrub species are widely distributed taxa, such as *Cytisus scoparius* or *Ulex europaeus,* whose ranges span entire continental Europe, whereas other species are narrow endemics exclusively distributed in certain mountain ranges (*e.g. Echinospartium bardanessi*, *E. boissieri*, *E. horridum* or *Genista versicolor*) (Talavera *et al*. 2001). For these reasons, it is expected that the past demography of the scrub-legume grasshopper has been shaped by spatiotemporal

changes in both the abiotic (*i.e*. climate regimes, geography) and biotic (*i.e.* host-plant distributional shifts) components that jointly define the ecological requirements of the species.

Here, we employ an integrative approach aimed to disentangle the evolutionary history and past demography of scrub-legume grasshopper across its entire distribution range. To this end, we combined nuclear and mitochondrial genetic data with paleo-distribution modelling to determine the stability of climatically suitable areas for both the grasshopper and its host-plant species and infer lineage-specific demographic histories from the genetic footprints left by past population size changes and distributional shifts. First, we used phylogenetic and Bayesian clustering analyses to determine the phylogeographic genetic structure of the scrub-legume grasshopper, estimate divergence times among the main lineages, and identify areas of long-term population persistence (*i.e.* refugia). Specifically, we tested whether northern populations are the outcome of range expansions from the Iberian Peninsula during favorable periods (*i.e.* "southern refugium and expansion" hypothesis) or if they conform lineages that have evolved *in situ* and long-term persisted into micro-refugia (*i.e.* "cryptic northern refugium" hypothesis) (Hickersson & Cunningham 2005; Parducci *et al*. 2012). Second, given that different demographic events (such as range expansions or bottlenecks) can leave similar signatures on patterns of genetic diversity and structure (Falush *et al*. 2016), we used an Approximate Bayesian Computation (ABC) framework to formally compare different demographic scenarios. Finally, we tested whether the spatial distribution of genetic diversity in the scrub-legume grasshopper is explained by the geographic location of stable areas sustaining longstanding host-plant communities (Tsai & Manos 2010; Laukkanen *et al*. 2014) and/or climatically suitable habitats for our focal species (Carnaval *et al*. 2009; Yannic *et al*. 2014).

MATERIAL AND METHODS

POPULATION SAMPLING

Between 2012 and 2014, we collected 794 individuals from 55 populations of scrub-legume grasshopper, *Chorthippus binotatus binotatus* (Charpentier 1825). These populations span the entire European distribution range of the species (~900 000 km^2; see Fig.S1) according to our own surveys and occurrence-data available in the literature (Llucià-Pomares 2002; Defaut 2011). Population codes and further information on sampling locations are given in Table S1.

MICROSATELLITE DATA AND ESTIMATES OF GENETIC DIVERSITY

We employed a salt extraction protocol to purify genomic DNA from a hind leg of each specimen (Aljanabi & Martinez 1997). All specimens were genotyped at 18 polymorphic microsatellites markers whose characteristics and PCR cycling conditions are described in Basiita *et al*. (2016). We performed PCR amplifications and genotyping as described in Ortego *et al*. (2015a). We tested for deviations from Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and the presence of null alleles following the procedure described in Noguerales *et al*. (2016a). Six loci (Cbin08, Cbin16, Cbin36, Cbin50, Cbin56 and Cbin57) were discarded from all downstream analyses because of HW disequilibrium in all populations and the presence of null alleles. We did not find evidence for linkage disequilibrium between any pair of loci in any sampling population after sequential Bonferroni corrections (Rice 1989).

We estimated nuclear genetic diversity for populations with more than nine genotyped individuals. As estimates of population genetic diversity, we used allelic richness (A_R) standardized for sample size and expected heterozygosity (H_E), calculated as implemented in HP-RARE (Kalinowski 2005) and ARLEQUIN 3.5 (Excoffier & Lischer 2010), respectively. For illustrative purposes, population genetic diversity was displayed in a map by conducting a spatial interpolation of A_R and H_E values using the Inverse Distance Weight (IDW) function available in ARCGIS 10.3 (ESRI, Redlands, CA, USA).

MITOCHONDRIAL DNA DATA AND ESTIMATES OF GENETIC DIVERSITY

For a subset of the collected specimens (Table S1), we amplified a fragment of the mitochondrial cytochrome oxidase subunit I (COI) using the universal primers LCO1490 and HCO2198 (Folmer *et al*. 1994). PCR reactions were performed in 15 µl volumes, with the same reagents and PCR program than for microsatellite markers, but using an annealing temperature of 50ºC. Amplified products were commercially purified and sequenced (Macrogen, South Korea). We edited, removed the primers and trimmed all the sequences to the same length (552 bp) using SEQUENCHER 4.10.1 (GeneCodes Corporation, Ann Arbor, MI, USA). The sequences were aligned using CLUSTALW web-service (*www.genome.jp/tools/clustalw/*). None of the sequences had premature stop codons and were deposited in GENBANK with accession numbers KY709453-KY709671.

For sampling locations with more than 4 sequenced individuals for COI, we used DNASP 5.10 (Librado & Rozas 2009) to calculate different population genetic diversity indices including number of haplotypes (*H*), number of polymorphic sites (*S*), haplotype diversity (H_D) and nucleotide diversity (π). In order to visualize the geographic distribution of mtDNA genetic diversity, we also performed a spatial interpolation of population H_D and π values as described for microsatellite makers.

GENETIC STRUCTURE ANALYSES

For the microsatellite dataset, we inferred genetic structure using Bayesian clustering analyses in STRUCTURE 2.3.3 (Pritchard *et al*. 2000; Falush *et al*. 2003). We considered correlated allele frequencies and an admixture model without prior information on population origin. We performed 10 independent runs for each value of K = 1-25 with a burn-in period of $2x10^5$ steps and a run length of $1x10^6$ Markov chain Monte Carlo (MCMC) cycles. The number of populations best fitting the dataset was defined using log probabilities [Pr(X|*K*)] (Pritchard *et al*. 2000) and the Δ*K* method (Evanno *et al*. 2005). We used the Greedy algorithm in the program CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) to align replicated runs and average individual assignment probabilities for the most likely *K* values. Afterwards, we produced bar plots displaying probabilities of individual membership using DISTRUCT 1.1 (Rosenberg 2004). Complementarily, we visualized genetic relationships among STRUCTURE clusters for the most likely *K* values by constructing a neighbour-joining (NJ) tree based on net nucleotide distances (Pritchard *et al*. 2010) using NEIGHBOR program as implemented in PHYLIP 3.695 software (Felsentein 2013).

For mtDNA, we assessed population genetic structure using a spatial analysis of molecular variance as implemented by SAMOVA 2.0 (Dupanloup *et al*. 2002). This method employs a simulated annealing procedure to identify the optimal grouping option (*K*) for the data by maximizing the among-group component (F_{CT}) of the overall genetic variance. We conducted 10 independent runs for each value of K = 1-12, using default parameters and 500 simulated annealing processes. Mean F_{CT} values and their standard deviation were plotted to determine the most likely population clustering solution (Dupanloup *et al*. 2002). We only included in this analysis those populations with more than four sequenced individuals for COI.

We visualized spatial patterns of nuclear and mtDNA genetic structure by interpolating probabilities of population membership to each inferred genetic cluster from STRUCTURE and SAMOVA

analyses using the '*maps*' function from POPSUTILITIES (Jay *et al*. 2012; François 2016) in R 3.2.3 (R Core Team 2015).

PHYLOGENETIC ANALYSES AND DIVERGENCE TIMES

We employed BEAST 1.8.0 (Drummond *et al*. 2012) to obtain a combined estimation of an ultrametric phylogenetic tree and divergence times using sequence data for the COI gene fragment. Analyses were conducted applying a HKY+I model of sequence evolution, which was selected as the best-fitting nucleotide substitution model for our dataset via the Bayesian Information Criteria (BIC) as implemented in JMODELTEST2 (Darriba *et al*. 2012). We assumed a normal distributed substitution rate of 0.0169 (± 0.0019 SD) per site per million years for COI gene (~1.69% per million year divergence rate; Papadopoulou *et al*. 2010). We ran several analyses considering different clocks and demographic models (Table S2). The best-fitting clock and demographic model to our dataset was determined via Akaike's information criterion (AIC) through Markov chain Monte Carlo (AICM; Baele *et al*. 2012) with 100 bootstraps in TRACER. Each analysis was run with two independent MCMC chains of 100 million generations (sampling every 10 000 generations and discarding the first 10% as burn-in) and we used TRACER 1.5 to examine stationarity and convergence of the chains and confirm that effective sampling sizes (ESS) for all parameters were > 200. We used LOGCOMBINER 1.8.0 to discard 10% of trees as burn-in and combine tree files from replicated runs. We used TREEANNOTATOR 1.8.0 to obtain maximum credibility trees and FIGTREE 1.4.2 to visualize final trees. Complementarily, we used POPART software (Leigh & Bryant 2015) to produce a statistical parsimony network and infer the phylogenetic relationships among COI haplotypes.

DEMOGRAPHIC ANALYSES

We inferred past demographic changes for the main COI clades using three different approaches. First, we calculated Fu´s F_S (Fu 1997), Tajima´s D (Tajima 1989) and R_2 statistics (Ramos-Onsins & Rozas 2002) and tested their significance by generating 10 000 coalescent simulations in DNASP software. For neutrally evolving loci, significant and negative (F_S and D) or positive (R_2) deviations of these indexes from zero are interpreted as past population expansions. Second, for each lineage we calculated mismatch distributions of pairwise nucleotide differences (Rogers & Harpending 1992) under the sudden

expansion model by 1000 bootstrap replicates using ARLEQUIN. The goodness-of-fit between observed and expected distributions was tested by calculating sum of squared deviations (SSD; Scheneider & Excoffier 1999) and Harpening´s Raggedness index (H_{rg}; Harpending 1994) in ARLEQUIN. A unimodal mismatch distribution and small and non-significant SSD and H_{rg} indexes are indicative of demographic expansions. Finally, we inferred changes in effective population sizes (N_e) over time for each lineage using the coalescent-based Bayesian skyline method (BSP) (Drummond *et al*. 2005) as implemented in BEAST 1.8.0. We used a piecewise constant skyline model and a number of 10 groups for every lineage (only five groups were considered for northwestern (NW)/southeastern (SE) France lineage due to its small sample size) (Ho & Shapiro 2011). Time and population size were calibrated assuming a strict clock prior (as supported by exploratory runs) and the substitution rate was the same considered for phylogenetic analyses. The best-fitting nucleotide substitution model for each dataset (*i.e.* lineage) was determined using JMODELTEST2 (Table S3). Number of independent replicates for each dataset, MCMC chains lengths and examination of stationarity, convergence and ESS values was conducted as described in the previous section.

TESTING DEMOGRAPHIC SCENARIOS: APPROXIMATE BAYESIAN COMPUTATION (ABC)

We used an Approximate Bayesian Computation (ABC) framework to compare different plausible scenarios of population divergence and past demography in the scrub-legume grasshopper (Beaumont 2010). The topology of the different tested scenarios was designed and informed with phylogenetic and clustering analyses (Fig. 1-2). We considered different scenarios of population bottlenecks that were designed on the basis of observed spatial patterns of genetic diversity and climate niche distribution models for either the scrub-legume grasshopper and its host-plant species (see next section) (Fig. 2; Fig. S3). To simplify the analyses and reduce computational demands, we defined five main population groups by pooling sampling sites according to their geographic location and the results obtained from phylogenetic and clustering analyses (*e.g.* Inoue *et al*. 2015; Tsuda *et al*. 2015). Our five population groups were northwestern/southeastern France (NWSEF), eastern France and Pyrenees (EF), northern Iberia (NI), eastern Iberia (EI) and southern Iberia (SI) (see Fig 1c; Fig. 2).

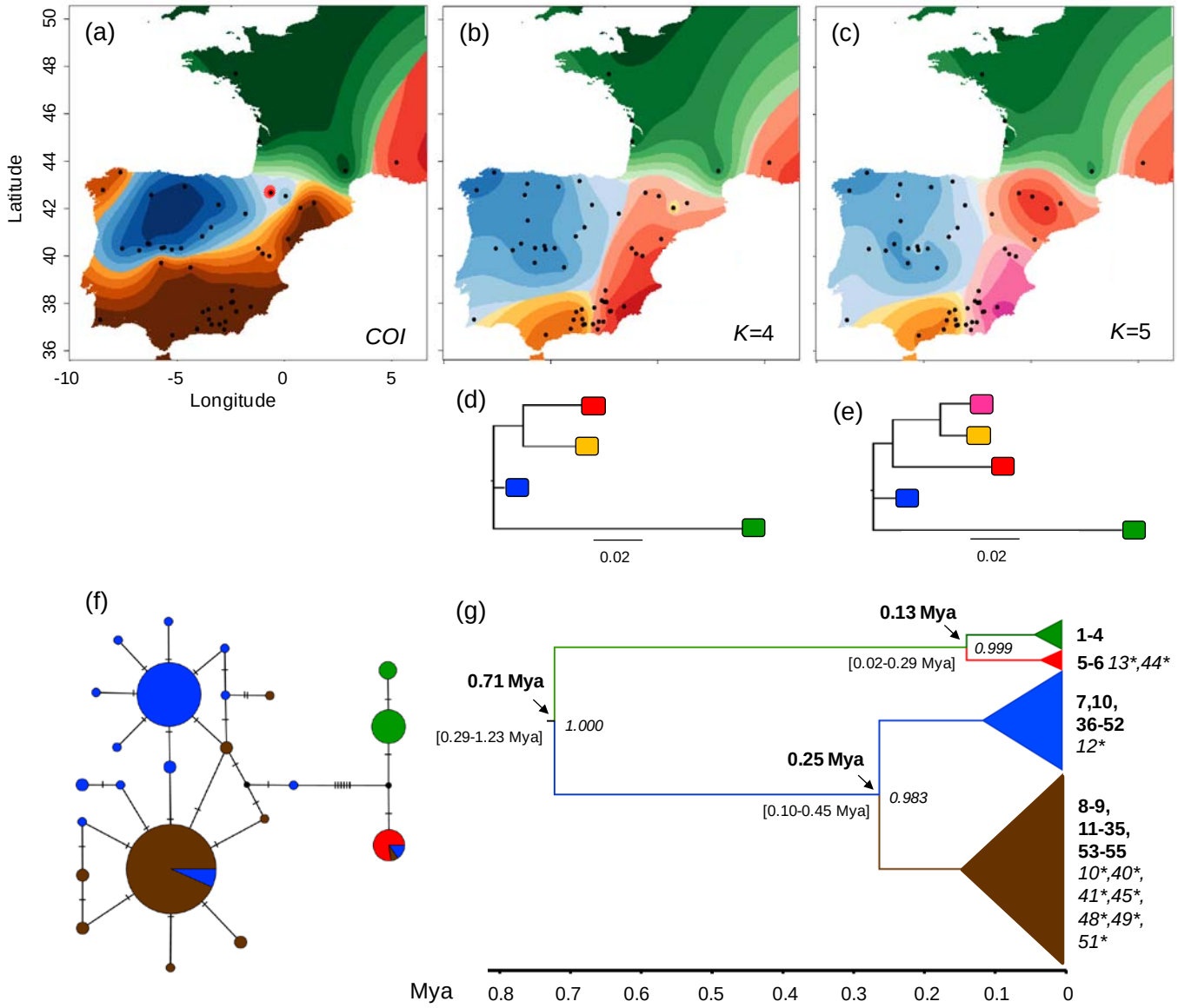


Figure 1 – Results of clustering and phylogenetic analyses for the scrub-legume grasshopper across its distribution range. The top panels represent the geographic distribution of the genetic groups inferred by (a) SAMOVA for mitochondrial cytochrome oxidase subunit I (COI) gene fragment and (b, c) by STRUCTURE for nuclear microsatellite markers. Black dots in the maps indicate the location of sampling sites. Panels (d, e) show neighbor-joining trees based on net nucleotide distances among genetic clusters inferred by STRUCTURE for K = 4 and K = 5, respectively. Panel (f) shows a statistical parsimony network inferred from COI haplotypes. Circle size is proportional to the relative frequency of a given haplotype. Panel (g) represents a maximum clade credibility tree for COI with estimated ages of divergence (mean and lower and upper 95% highest posterior density) for each main node. Bayesian posterior probability for each main clade is indicated at the right of the node. Tip labels in bold type indicate the code of the populations represented in each collapsed clade. Populations codes in italics and with an asterisk indicate those populations assigned primarily to another phylogroup but that present some individuals in that phylogroup. Population codes are described in Table S1. Colour codes for the main genetic groups are maintained across all panels.

Up to a maximum of 40 individuals representative of all putative populations were randomly selected per group (*n* = 200). The first scenario (scenario 1), considered as null model, consisted of a simultaneous split of all these five groups. The second scenario (scenario 2) predicts an old split between an ancestral French clade and the ancestral Iberian one. Afterwards, Iberian populations (northern, eastern and southern) diverged simultaneously from each other, and French populations split originating northwestern/southeastern and eastern French groups. The third scenario (scenario 3) is similar to the second one, but predicts a hierarchical divergence of the Iberia groups at two different times: an older split of northern Iberian populations from the rest of the Iberian groups followed by a more recent split of eastern and south Iberian populations (see Fig. 2). Because of genetic diversity estimates for French and Pyrenean populations were consistently lower than those obtained for Iberian ones, we considered four different demographic hypotheses for scenarios 2 and 3: (A) no change in effective populations sizes, considered as the null demographic model, (B) a population bottleneck in the NWSEF group, (C) a population bottleneck in the EF group, and (D) population bottlenecks in both NWSEF and EF groups (see Fig. 2).

We conducted all the computations in DIYABC 2.0.4 combining nuclear and mitochondrial markers (Cornuet *et al*. 2014). To avoid biases in parameter estimates, we selected the subset of seven microsatellites markers with lower frequency of null alleles, estimated using the Expectation Maximization (EM) algorithm implemented in the program FREENA (Chapuis & Estoup 2007). We ran one million of simulated datasets per scenario assuming a 1:1 female to male sex ratio. We used a generalized mutation model (GSM) and no single nucleotide indels for the microsatellite data and a HKY model of sequence evolution for COI. Summary statistics (SS) for microsatellite markers and mtDNA marker are detailed in Table S4. Information from previous studies (Noguerales *et al*. 2017) and a pre-evaluation of scenarios and prior distributions were employed to adjust the priors of effective population sizes (N_e) and timing of divergence (t) to their most appropriate values (see Table S4), assuming a uniform prior probability distribution for them. Prior values of N_e for groups with an assumed population bottleneck was set to 10-500 000 individuals, which was 50% smaller than N_e priors for groups with constant population sizes (*i.e.* no bottleneck assumed). Selection of the most probable scenario, confidence in scenario choice (type I and type II errors), model checking and estimation of the posterior distribution of all parameters under the best supported model were performed as detailed in Ortego *et al*. (2015b).

ECOLOGICAL NICHE MODELLING

We modeled the potential distribution of the scrub-legume grasshopper to assess whether habitat suitability and stability shape its geographic patterns of genetic diversity. We reconstructed the present and past distribution of the scrub-legume grasshopper considering three different subsets of environmental layers: (i) bioclimatic layers available in WorldClim (Hijmans *et al*. 2005); (ii) bioclimatic layers plus a layer of host-plant species richness obtained from climate niche models built for each host-plant species (see below); and (iii) considering only the layer of host-plant species richness.

We employed the maximum entropy presence-only algorithm implemented in MAXENT 3.3.3 (Phillips *et al*. 2006; Phillips & Dudik 2008) to build species-specific ecological niche models (ENMs) for each host-plant species using the 19 bioclimatic variables available in WorldClim at 30 arc-sec resolution (Hijmans *et al*. 2005). It has been demonstrated that overparameterized models could perform better than underparameterized ones and thus we initially included all 19 bioclimatic variables in MAXENT model building (Warren & Seifer 2011; Freeman & Mason 2015). Maximum iterations were set to 5 000 to ensure model convergence. Models for the different host-plants were built using occurrence data available in the Global Biodiversity Information Facility (GBIF; Table S5). All records

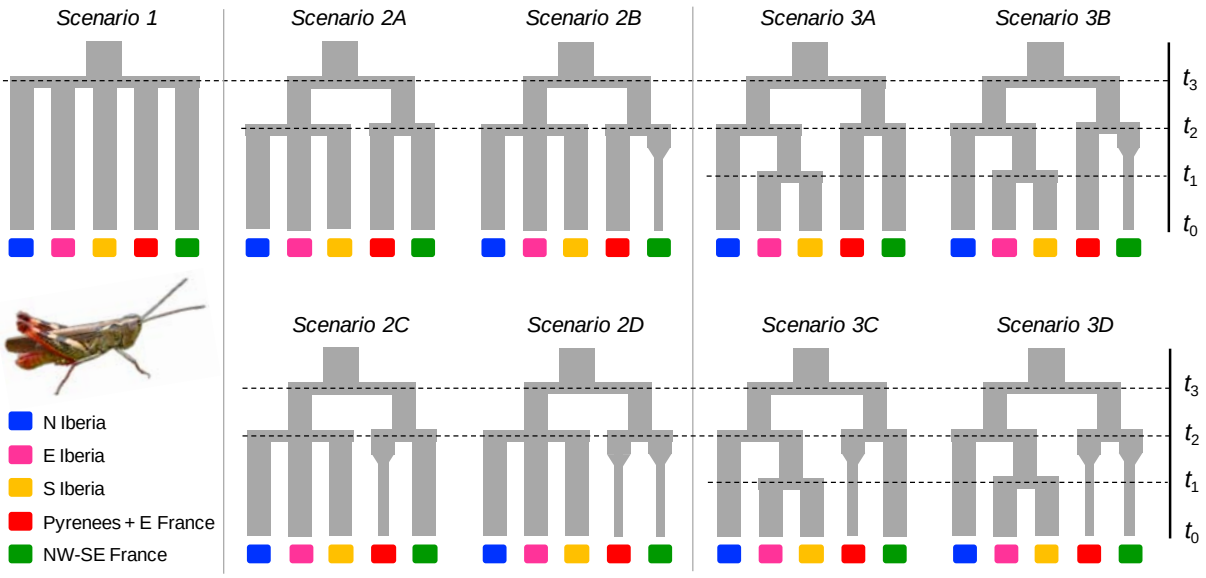


Figure 2 – Different scenarios of population divergence and past demography for the scrub-legume grasshopper compared using an Approximate Bayesian Computation (ABC) approach ($t_{\#}$ represents time in number of generations). Colours of genetic groups are in accordance to the genetic clusters inferred by STRUCTURE for K = 5 (see Fig. 1c.).

were spatially checked to exclude species misidentifications, geo-referencing errors and duplicate locations that fell into the same map pixel. Records in areas where the different species are considered non-native based on the literature were also discarded (Talavera *et al*. 2001; Anthos 2016). We limited the geographic extent of the climate layers to an area approximately 20% larger than the known distribution range of each host-plant species following suggestions from Anderson & Raza (2010). Model performance was assessed using the area under the receiver operating characteristic curve (AUC) estimated from test data after averaging over 10 cross-validation replicate runs. To obtain the distribution of the species during the Last Glacial Maximum (LGM, *c*. 21 Kya), we projected contemporary species-climate relationships to the LGM using the Community Climate System Model (CCSM3; Collins *et al*. 2006) from the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2; Braconnot *et al*. 2007). Layers for the LGM were downloaded from WorldClim at 2.5 arc-min and interpolated to 30 arc-sec resolution. The "fade by clamping" option in MAXENT was applied to the past predictions in order to reduce unreliable extrapolation onto extensive areas with environmental conditions not encountered during model training (Phillips *et al*. 2006). In order to obtain a proxy of overall host-plant species richness and summarize climate niche models of all host-plant species, we converted logistic output maps into binary maps (presence = 1; absence = 0) using threshold values for occurrence based on the maximum training sensitivity plus specificity (MTSS) obtained for each host-plant species (Liu *et al*. 2005). Afterwards, we summed all binary maps of the thirteen host-plant species separately for current and LGM periods.

For the scrub-legume grasshopper, the three subsets of models described above were built using a total of 287 occurrence locations obtained from our own sampling, the GBIF database, and the literature (Llucià-Pomares 2002; Defaut 2011). Niche models were built and projections were performed as described above for host-plants. We assessed the relative support of each MAXENT model using the sample-size adjusted Akaike's information criterion (AIC_C) calculated as implemented in ENMTOOLS (Warren *et al*. 2010).

Stability maps for each MAXENT model of the scrub-legume grasshopper were generated by averaging current and LGM maps. All raster calculations were conducted using RASTER package (Hijmans & van Etten 2016) in R. Finally, we used ARCGIS 10.3 and a buffer of 10 km^2 around sampling locations to extract values of climate/habitat suitability stability for each studied population and the models based on (i) only bioclimatic variables (C_{STA}), (ii) bioclimatic variables and host-plant species richness ($C\text{-}HP_{STA}$) and (iii) only host-plant species richness (HP_{STA}).

GENETIC DIVERSITY AND ESTIMATES OF HABITAT STABILITY

We used generalized linear models (GLMs) and an information-theoretic model selection approach (Burnham & Anderson 2002) to test the association between nuclear (A_R, H_E) and mtDNA (H_D and π) genetic diversity estimates and stability of climate/habitat suitability for the scrub-legume grasshopper obtained from the three different MAXENT models described in the previous section. GLMs were constructed considering a Gaussian error distribution and an identity link function using R package LME4 (Bates *et al*. 2015). Longitude and latitude were included as covariates to take in account possible geographical clines of genetic diversity (Guo 2012; Eckert *et al*. 2008). GLMs were fitted considering potential bias resulted from model overfitting and spurious relationships between dependent and independent variables because of the high collinearity between the different estimates of climate stability (Burnham & Anderson 2002; Arnold 2010). Accordingly, we constructed the different combinations of GLMs only including one of the estimates of climate stability at a time.

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 sample size for each studied population. We analyzed the goodness of fit of the models using AIC_C values (Burnham & Anderson 2002; *e.g.* Noguerales *et al*. 2015). We ranked models according to their AIC_C values using the R package MUMIN (Barton 2015) and those models with $\Delta AIC_C \leq 2$ were considered to have similar empirical support than the best model (*i.e.* that with the lowest AIC_C; Burnham & Anderson 2002). The Akaike weight (ω_i) was calculated for each model, which represents the relative probability that a given model will be the best among those considered. For each best ranked models ($\Delta AIC_C \leq 2$), we calculated separately the 95% confidence intervals (CI) of their estimators. The effect of variables was considered significant if the 95% confidence intervals (CI) of their estimators excluded zero (Burnham & Anderson 2002). If the null model (*i.e.* without predictors) was included within the best ranked models ($\Delta AIC_C \leq 2$), we considered that any remaining variables included in other equivalent models ($\Delta AIC_C \leq 2$) did not have a significant effect on the dependent variable.

RESULTS

SPATIAL PATTERNS OF GENETIC DIVERSITY AND STRUCTURE

Nuclear genetic diversity ranged from 3.54 to 5.98 for A_R, and from 0.59 to 0.85 for H_E. Genetic diversity at nuclear markers was spatially homogenous throughout Iberian populations, which exhibited the highest values, whereas French and Pyrenean populations showed the lowest levels of genetic variability (Fig. 3). At mitochondrial level, approximately the 50 % of the populations exhibited no genetic variation. In populations that showed intra-population genetic variation at mtDNA, genetic diversity estimates ranged from 0.28 to 0.83 for H_D, and from 0.0005 to 0.0108 for π. Populations from east and northwestern Iberia presented the highest levels of mtDNA genetic diversity (Fig. 3).

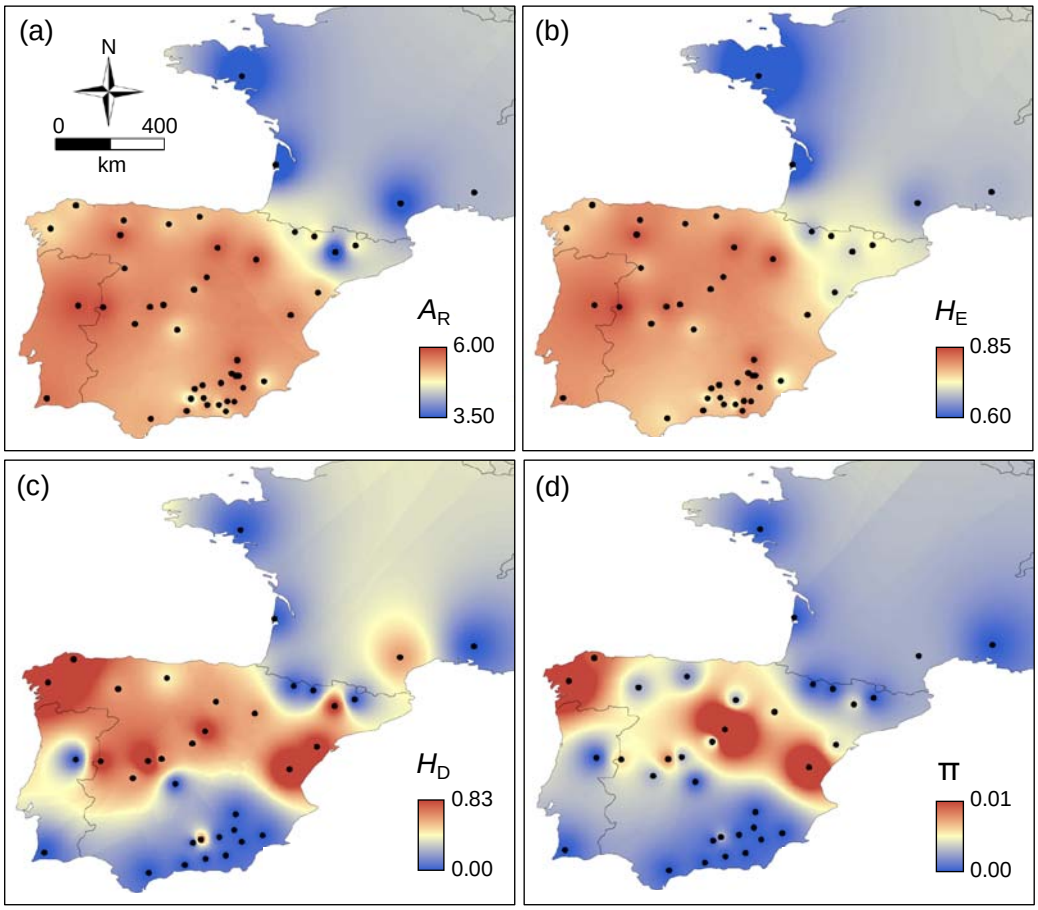


Figure 3 – Maps displaying spatial patterns of genetic diversity at nuclear (A_R, allelic richness, panel (a); H_E, expected heterozygosity, panel (b)) and mtDNA (H_D, haplotype diversity, panel (c); and π, nucleotide diversity, panel (d)) markers across the distribution range of the scrub legume grasshopper. Black dots indicate the location of sampling sites.

STRUCTURE analyses showed a steadily increase of log probabilities [Ln Pr (X|*K*)] from *K* = 2 to *K* = 5 (Fig. S1b). According to the Δ*K* method, the best-supported number of genetic clusters was *K* = 4 (Fig. S1c). When we considered *K* = 4, a first cluster included all the populations located in northwestern France and also the population 'Col de Fontfroide' from the Massif Central in southeastern France (Fig. 1b; Fig. S1d). A second cluster included the easternmost French population ('Gordes') and the Pyrenean and eastern Iberian populations. The populations from south Iberia were included in a third cluster and a fourth cluster was primarily composed of populations from central and north Iberia (Fig. 1b; Fig. S1d). The above described clustering pattern was similar when we considered *K* = 5, however the easternmost French population ('Gordes') and the Pyrenean populations were included in a new cluster differentiated from eastern Iberian populations (Fig. 1c). Neighbour-joining trees considering both grouping solutions revealed that the NW-SE France cluster was the most genetically differentiated (Fig. 1d-e). STRUCTURE analyses showed that the *F*-value, which represents the amount of drift for each cluster from a common ancestral population, was 0.24 for the NW-SE France cluster whereas this index ranged from 0.04 to 0.06 in the remaining genetic clusters.

SAMOVA analyses on mtDNA showed that mean F_{CT} values increased from *K* = 2 to *K* = 5, and reached a plateau for *K* = 7 (Fig. S1a). The most conservative clustering solution (*K* = 4) grouped the populations from northwestern France and the population 'Col de Fontfroide' from southeastern France. A second group was composed of the easternmost French ('Gordes') and westernmost Pyrenean ('Borau') populations. The east, south-central and northwestern Iberian populations composed a third group. Finally, north-central Iberian populations were included in the fourth cluster (Fig. 1a; Table S1). These four SAMOVA groups were consistent with the main clades inferred in phylogenetic analyses (see next section and Fig. 1f-g). The clustering solution based on *K* = 5 yielded a new cluster that was composed by a single population ('Puerto de la Quesera') from Central Iberia.

PHYLOGENETIC ANALYSES AND DIVERGENCE TIMES

Phylogenetic reconstruction in BEAST was conducted using a strict clock and a constant demographic model, the one that best fitted our data according to the model selection procedure implemented in TRACER (Table S2). The phylogenetic tree revealed four well-supported phylo-groups, which were equivalent to those inferred by SAMOVA (Fig. 1g). Most individuals from the same population were grouped in a single phylo-group. However, Iberian phylo-groups were not reciprocally monophyletic and

exhibited certain degree of admixture (*i.e.* populations presented individuals assigned to different clades) (Fig. 1g). Divergence times indicated a split of French and Iberian populations around the Mid Pleistocene (~0.71 Mya), whereas divergence within Iberian and French clades took place more recently, around the Late Pleistocene (~0.25 and ~0.13 Mya, respectively) (Fig. 1g). Nevertheless, divergence times must be considered cautiously given that their respective confidence intervals are broad (Fig. 1g). According to the phylogenetic reconstruction in BEAST, network genealogy revealed a total of 23 different haplotypes that clustered in four main groups exhibiting a highly congruent geographical distribution (Fig. 1f).

DEMOGRAPHIC ANALYSES

Northern and south/southeastern Iberia lineages presented significant negative Fu´s F_s and Tajima´s D and significant positive R_2 statistics, which is interpreted as past population expansions (Table S3). All tests for Fu´s F_s, Tajima´s D and R_2 statistics failed to reject the null hypothesis of constant population size for the French lineages (Table S3). Pairwise mismatch distributions were unimodal for all lineages and test of goodness-of-fit between observed and expected distributions were not significant, which suggests that all lineages have undergone demographic expansions (in all cases for SSD and H_{rg}, P-values > 0.16) (Fig. S2). Finally, BSP analyses showed a more complex demographic history and confirmed that northern and southern/eastern Iberia lineages experienced a demographic expansion that began around 20 000 years ago followed by a subtle negative trend in N_e during the most recent time interval (Fig. S2). Conversely, skyline plots for NW-SE France and all lineages from France and the Pyrenees revealed that their effective population sizes have remained steadily small during the last 50 000 years (Fig. S2).

TESTING DEMOGRAPHIC SCENARIOS: APPROXIMATE BAYESIAN COMPUTATION (ABC)

ABC analyses indicated that “scenario 3B” of population divergence and past demography had the highest posterior probability and its 95% confidence intervals did not overlap with those obtained for others scenarios (Table 1). This scenario was defined by a population bottleneck for the NW-SE French group and a hierarchical divergence scheme for the Iberian groups.

Table 1 – Posterior probability for each of the nine tested scenarios of population divergence and past demography in the scrub-legume grasshopper and 95% confidence intervals (CI) based on the weighted polychotomous logistic regression approach for approximate Bayesian Computation Analyses (ABC). Type I and II errors for the best supported scenario (in bold) are indicated.

Scenario	Posterior Probability	95% CI	Type I error	Type II error
1	0.0179	[0.0153 - 0.0205]		
2A	0.0299	[0.0272 - 0.0325]		
2B	0.0771	[0.0733 - 0.0809]		
2C	0.0189	[0.0164 - 0.0215]		
2D	0.0467	[0.0435 - 0.0498]		
3A	0.1640	[0.1587 - 0.1693]		
3B	**0.3792**	**[0.3718 - 0.3866]**	**0.246**	**0.220**
3C	0.0828	[0.0792 - 0.0865]		
3D	0.1835	[0.1777 - 0.1893]		

For the best-supported scenario, we found that observed data fell within simulated data and all SS used for model checking showed *P*-values > 0.2, suggesting good model fit. Overall type I and type II errors were relatively low (< 0.25, Table 1) and RMAEs exhibited values lower than 0.25 in most parameters, indicating a good reliability of their estimates (Table 2). Assuming a constant mutation rate, the posterior estimates of N_e for the NW-SE French group were much lower (~2-3 fold) than those obtained for the remaining groups (Table 2). Considering the one-year generation time of scrub-legume grasshopper (Defaut 2011), the French ancestral group was estimated to diverge from the ancestral Iberian group ~600 Kya (t_3). Subsequently, a split within French and Iberian populations took place ~230 Kya (t_2), whereas the most recent divergence happened between the eastern and southern Iberian groups, ~60 Kya (t_1). These estimates of divergence times are highly consistent with those inferred by phylogenetic reconstruction in BEAST (Fig. 1g).

ECOLOGICAL NICHE MODELLING

Species-specific niche models for host-plants generally presented high AUC values (average AUC across species: 0.85; range: 0.621-0.983), indicating overall good model performance (Table S5). Host-plants showing the lowest AUC values (< 0.75) correspond to taxa widely distributed throughout Europe and North Africa (*C. scoparius, G. scorpius* and *U. europaeus*; Table S5). For the scrub-legume grasshopper, niche models presented AUC values ranging from 0.788 to 0.867 (Table S5).

Table 2 – Posterior parameter estimates (median and 95% confidence intervals) for the best supported scenario of population divergence and past demography (scenario 3B; see Fig. 2). Estimates are based on 1% of simulated datasets closest to the observed values. Relative median absolute errors (RMAE) based on 500 pseudo-observed data sets are also indicated for each parameter.

Parameter	Median	$q_{0.025}$	$q_{0.975}$	RMAE
$N_{NWSEF-B}$	283 000	127 000	459 000	0.230
N_{EF}	767 000	421 000	980 000	0.206
N_{NI}	747 000	428 000	967 000	0.213
N_{EI}	640 000	252 000	969 000	0.232
N_{SI}	682 000	291 000	971 000	0.233
N_F	658 000	212 000	983 000	0.278
N_I	688 000	120 000	984 000	0.403
N_{ESI}	563 000	556 000	978 000	0.352
N_X	465 000	26 700	969 000	0.376
t_1	62 900	6 830	279 000	0.408
t_2	230 000	62 300	670 000	0.224
t_3	603 000	188 000	978 000	0.146
μ_{nDNA}	5.46×10^{-6}	2.24×10^{-6}	1.68×10^{-5}	0.392
μ_{mtDNA} (COI)	2.12×10^{-9}	1.21 x 10-9	4.30×10^{-9}	0.269

$N_{NWSEF-B}$, effective population size of the northwestern and southeastern France group (assuming a population bottleneck event)
N_{EF}, effective population size of the eastern France and Pyrenees group
N_{NI}, effective population size of the northern Iberia group
N_{EI}, effective population size of the eastern Iberia group
N_{SI}, effective population size of the southern Iberia group
N_F, effective population size of the ancestral France group
N_I, effective population size of the ancestral Iberia group
N_{ESI}, effective population size of the ancestral south and eastern Iberia group
N_X, effective population size of the ancestral population
t_1, time (in generations = years) to the most recent divergence event
t_2, time (in generations = years) to the intermediate divergence event
t_2, time (in generations = years) to the most ancient divergence event
μ_{nDNA}, mean mutation rate for microsatellites markers
μ_{mtDNA}, mean mutation rate per site per generation for the mitochondrial marker (COI)

The most supported model (*i.e.* with the lowest AIC_C) for the scrub-legume grasshopper was the one only based on bioclimatic variables (Table S5). However, host-plant species richness was the most important predictor when included together with the remaining bioclimatic predictors (*i.e.* it provided the highest gain when was used in isolation and decreased the gain the most when was omitted; Table S5). According to the model of lowest AIC_C, the current distribution of the scrub-legume grasshopper is strongly fragmented and restricted to the main Iberian and French mountain ranges, northwestern Iberia, and the Atlantic French coast (Fig. S3a). Projections into the LGM indicated that the distribution of the scrub-legume grasshopper was more widespread and its populations exhibited higher

connectivity during this period than in the present (Fig. S3a). We observed that most sampling locations were located in regions with high climate suitability stability. Nonetheless, populations from northern France were located in areas characterized by low climate suitability both in the present and during the LGM (Fig. S3c). The model including host-plant species richness and bioclimatic layers (ΔAIC_C = 10.03; Table S5) exhibited areas with lower suitability in western Iberian during LGM compared to the lowest AIC_C model (Fig. 3b-e). Models constructed using a subset of less correlated bioclimatic variables (r < 0.90) provided qualitatively analogous results (not shown). Inspection of predicted distributions of host-plants and scrub-legume grasshopper confirmed that niche models reported distribution patterns coherent with their respective observed distributions.

GENETIC DIVERSITY AND ESTIMATES OF HABITAT STABILITY

Model selection for nuclear genetic diversity indicated that latitude, longitude and C_{STA} had a significant effect on A_R (*i.e.* 95% CIs of such predictors did not cross zero). Similarly, we found that latitude, longitude, C_{STA} and also $C\text{-}HP_{STA}$ had a significant effect on H_E (Table 3; Table S6). Latitude and longitude were negatively associated with A_R and H_E, whereas C_{STA} and $C\text{-}HP_{STA}$ had a positive effect on genetic diversity estimates (Fig. 4). Model selection for mitochondrial genetic diversity showed that the null model was among the best ranked models ($\Delta AIC_C \leq 2$) for both H_D and π, indicating that no predictor had a significant effect (Table S6).

DISCUSSION

Integrating nuclear and mitochondrial markers with a suite of analytical approaches (including phylogenetic Bayesian reconstructions, ecological niche modelling and explicit hypothesis testing using an ABC framework), allowed us to conclude that the scrub-legume grasshopper presents a striking phylogeographic structure resulted from long-term isolation in Iberian and French refugia during Pleistocene glacial cycles. The variability through time of habitat suitability impacted the demographic history of the northernmost populations of the species, a process that reduced local/regional levels of genetic diversity and left detectable signatures of population bottlenecks. Our analyses also support that such process of genetic erosion was explained by habitat suitability stability defined by climate and the

Table 3 – Generalized linear models (GLMS) for nuclear genetic diversity of scrub-legume grasshopper (A_R, allelic richness standardized for sample size; H_E, expected heterozygosity) testing the effects of (i) stability of climate suitability (C_{STA}) based on a MAXENT model built using only climatic variables and (ii) stability of habitat suitability ($C\text{-}HP_{STA}$) based on a MAXENT model built using climatic variables and host-plant species richness. The stability of habitat suitability (HP_{STA}) based on a MAXENT model built using only host-plant species richness was not included in any of the best ranked equivalent models ($\Delta AIC_C \leq 2$; for model selection results, see Table S6). Latitude (Lat) and longitude (Lon) were also included as covariates in the models. For each model we indicate AIC_C, sample-size adjusted Akaike's information criterion (AIC) value. Predictors excluding the value 0 in their 95% confidence intervals (CI) are indicated in bold type and their effects were considered significant.

Models	Predictors	Estimate ± SE	Lower 95% CI	Upper 95% CI	AIC_C
A_R	Lat	-0.0684 ± 0.0260	-0.1218	-0.0151	64.50
	Lon	-0.0684 ± 0.0273	-0.1157	-0.0137	
	C_{STA}	0.7204 ± 0.3433	0.0475	1.3934	
A_R	Lat	-0.0636 ± 0.0293	-0.1212	-0.0060	65.44
	Lon	-0.0730 ± 0.0260	-0.1240	-0.0220	
	$C\text{-}HP_{STA}$	0.6492 ± 0.3490	-0.0347	1.3333	
H_E	Lat	-0.0054 ± 0.0024	-0.0102	-0.0006	-156.60
	Lon	-0.0060 ± 0.0023	-0.0106	-0.0014	
	C_{STA}	0.0817 ± 0.0310	0.0209	0.1425	
H_E	Lat	-0.0048 ± 0.0026	-0.0100	0.0004	-155.43
	Lon	-0.0070 ± 0.0023	-0.0116	-0.0023	
	$C\text{-}HP_{STA}$	0.0755 ± 0.0316	0.0135	0.1375	

spatial distribution of host-plant species richness, indicating the importance of integrating both biotic and abiotic factors into the study of the demographic history of specialist taxa with narrow ecological requirements.

PAST DISTRIBUTION AND SPATIAL GENETIC STRUCTURE

Analyses of mtDNA showed that populations of the scrub-legume grasshopper split into two major phylogeographic lineages from France and Iberia that probably diverged during the Mid-Pleistocene, around 700 Kya. The long branches defining French and Iberian clades in the phylogenetic tree and their geographic coherence suggest a scenario of allopatric divergence that was probably mediated by geographic barriers (*i.e.* the Pyrenees) and long-term lineage persistence in different refugia (Cooper & Hewitt 1993; Cooper *et al*. 1995). This result contrasts with phylogeographic patterns observed for many temperate-adapted taxa whose present-day northern populations are the result of post-glacial

recolonizations from southern Mediterranean peninsulas (Iberian, Italian and Balkan) during warming periods (Taberlet *et al*. 1998; Schmitt 2007). The most recent divergence event within Iberian and French lineages probably took place around the Late Pleistocene (*c*. 250 and 130 Kya, respectively) and could have been also promoted by isolation in local refugia during interglacial periods (Petit *et al*. 1999; Snyder 2016). Thus, the global phylogeographic structure of the species seems to have been mostly driven by the presence of major geographic barriers and temporal changes in the spatial

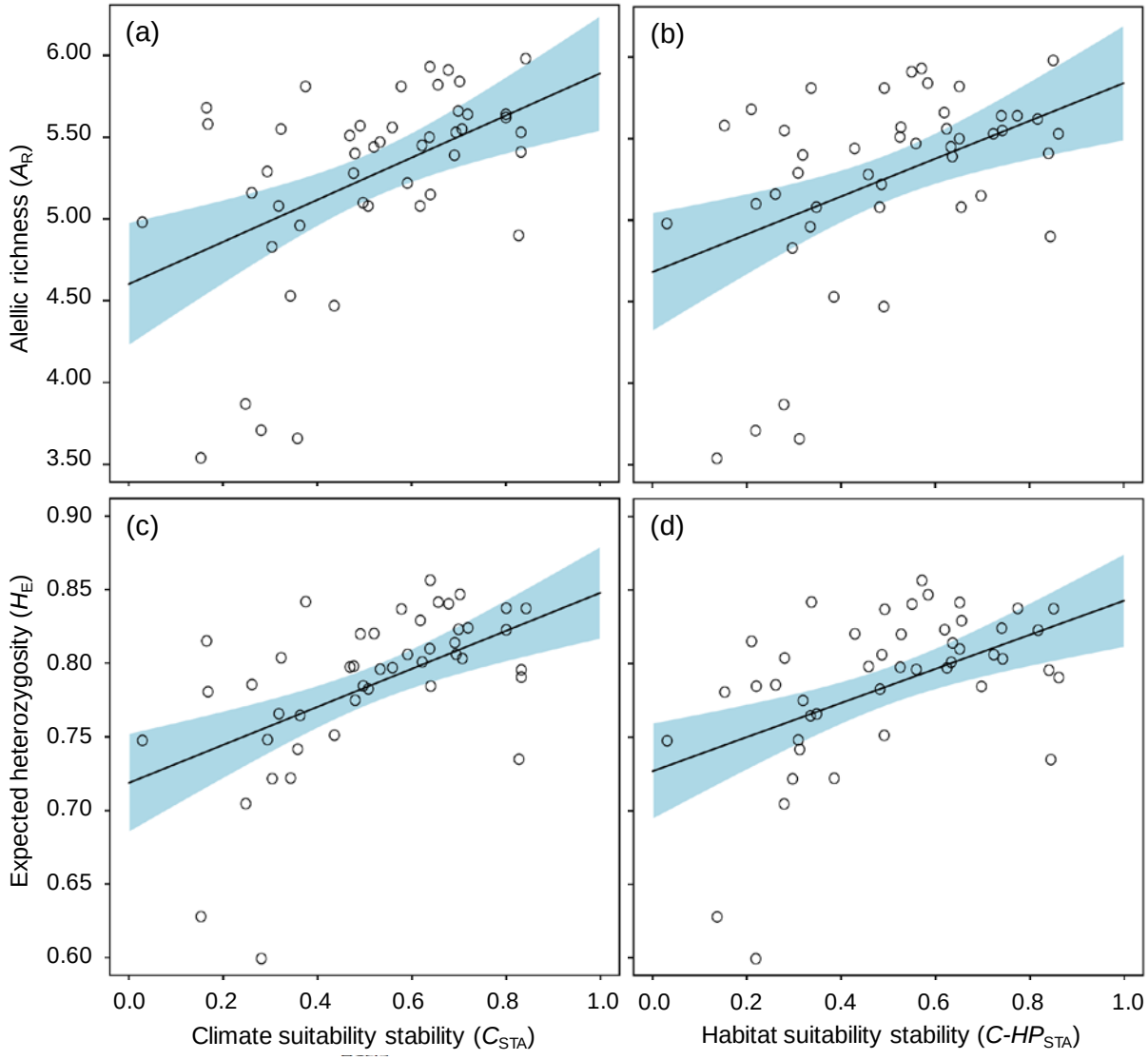


Figure 4 – Scatterplots representing the relationship between (a, b) allelic richness (A_R) and (c, d) expected heterozygosity (H_E) of scrub-legume grasshopper in relation to (a, c) stability of climate suitability (C_{STA}) based on a MAXENT model built using only climatic variables and (b, d) stability of habitat suitability ($C\text{-}HP_{STA}$) based on a MAXENT model built using climatic variables and host-plant species richness. Regression lines and 95% confidence intervals are shown.

distribution of suitable habitats mediated by Pleistocene glacial cycles. Assuming niche conservatism (Peterson 1999; Nogués-Bravo 2009), our ENMs suggest that during interglacials the populations of scrub-legume grasshopper were likely pushed up to higher elevations in response to uphill distributional shifts of its cold-adapted host-plants species, which may be responsible of more recent population fragmentation at local/regional scales (Noguerales *et al*. 2017). Accordingly, STRUCTURE analyses showed a finer genetic structure at regional scales that divided the mitochondrial lineage from east and south Iberia into two different genetic clusters. ABC analyses suggest that these genetic clusters diverged from a common ancestor after the Last Interglacial, likely during a period defined by short warming pulses that took place during the Late Pleistocene (Petit *et al*. 1999). Overall, the combined effect of the complex topography of the Iberian Peninsula and climate-driven distributional shifts of host-plants could have acted together to reduce or disrupt gene flow among populations of scrub-legume grasshopper at a regional scale (Noguerales *et al*. 2017).

The observed pattern of multiple refugia throughout the current distribution range of the species has been also documented in many other plants and animals in southwestern Europe and is in good agreement with the "refugia-within-refugia" concept (Gómez & Lunt 2007; Abellán & Svenning 2014). Strong genetic subdivision promoted by population isolation in different Pleistocene Iberian refugia has also been reported for the meadow grasshopper (*Pseudochorthippus paralellus*) (Hewitt 1996; Cooper & Hewitt 1993). However, contrary to the general pattern of population isolation in northern Mediterranean peninsulas, the existence of a French lineage distributed from the Atlantic coast to Massif Central suggests long-term persistence of our study species outside of the Iberian Peninsula in a cryptic northern refugium (Provan & Bennet 2008; Stewart *et al*. 2010; Salvi *et al*. 2013). The existence of a single and large refugium encompassing from Massif Central to Brittany seems unlikely according to our climatic reconstructions and different hypotheses could explain the origin and the demography of the NW-SE France lineage (Schmitt & Varga 2012). The fact that our study species feeds on several montane or cool-adapted scrub legume species could have facilitated *in situ* population persistence in different micro-refugia during unfavorable periods in the northernmost areas (Parducci *et al*. 2012). Accordingly, our ENMs suggest a patchy distribution of climatically suitable areas for either the scrub-legume grasshopper or some of its host-plant species since the LGM. Alternatively, the higher mtDNA genetic diversity of populations from Massif Central in southeast France ('Col de Fontfroide') in comparison with those located along the Atlantic coast may also suggest a recent range expansion from a putative refugium located in this region (Gutierrez-Jiménez *et al*. 2016). As suggested by ENMs, the more continuous distribution of suitable habitat during the LGM suggests that scrub-legume

grasshopper could have dispersed from the Massif Central to the northwest across central France during favorable periods (Gutierrez-Jiménez *et al*. 2016). This hypothesis is in agreement with the recent idea that the Massif Central was an important reservoir of stable populations and the geographic origin of range expansions for some taxa during the Pleistocene (Schmitt & Varga 2012; Ursenbacher *et al*. 2015).

HABITAT STABILITY AND SPECIES DEMOGRAPHIC HISTORY

The inference of refugia has traditionally relied on species-specific responses to climate (Ashcroft 2010; Keppel *et al*. 2012), which are expected to be especially relevant for ectotherm organisms whose niches are often shaped by temperature tolerances (Noguerales *et al*. 2016b). However, refugium is a broad concept that should be defined not only taking into account the environmental tolerances and responses of species to climatic oscillations but, ideally, also considering their ecological interactions with other taxa (Ashcroft 2010; Stewart *et al*. 2010). The findings from our different analyses revealed an important role of environmental stability on the demographic history of the scrub-legume grasshopper at both population and lineage scales. Demographic reconstructions and ABC analyses supported a population bottleneck in the NW-SE France lineage after it split from its sister Pyrenean and Mediterranean French clade, a process that likely resulted in strong genetic drift and erased its levels of genetic variation (Qu *et al*. 2014). Accordingly, STRUCTURE analyses reported that the amount of drift after divergence (*F*-value) was 0.24 for the NW-SE French cluster, whereas all other clusters presented between four- and fivefold lower estimates for this statistic (*F*-values ranging from 0.04 to 0.06) (for a similar approach see Harter *et al*. 2004). This, together with the severity of the inferred population bottleneck (N_e of the NW-SE French lineage was reduced by ~50% with respect to its ancestral N_e), lend also support to a demographic scenario probably defined by geographically isolated populations inhabiting fragmented patches of suitable habitat that sustain small effective population sizes. In this vein, whereas ecological suitability characterizing the northernmost distribution area of the species make feasible that it has maintained viable populations of scrub-legume grasshopper through time, such populations have probably experienced remarkable demographic fluctuations (*i.e.* bottlenecks and extinctions-recolonization processes) that ultimately eroded their genetic variability (Yannic *et al*. 2014; Bidegaray-Batista *et al*. 2016). In contrast, Iberian populations show signatures of past demographic expansions that could have been driven by the higher availability of suitable habitats during the LGM period, as

suggested by our demographic and ENMs analyses. In addition, most Iberian populations have been located in environmentally more stable and connected suitable areas during the last 21 000 years, which has probably contributed to maintain larger effective population sizes and higher levels of nuclear genetic variation (Inoue *et al*. 2015). In this regard, we found a positive association between nuclear genetic diversity (H_E and A_R) and local stability of suitable habitats since the LGM, a result that is in agreement with previous empirical studies addressing the consequences of environmental stability at local/regional scales on the demographic history of different taxa (Qu *et al*. 2014; Faye *et al*. 2016; Gutierrez-Jiménez *et al*. 2016). Furthermore, nuclear genetic diversity was significantly higher at lower latitudes following the so-called "southern richness *vs*. northern purity" pattern of genetic variation (Hewitt 2004b). This spatial trend of intra-specific genetic diversity has been attributed to two non-mutually exclusive processes including, on the one hand, strong geographic genetic structure due to allopatric divergence in different southern refugia followed by admixture among differentiated lineages and, on the other hand, the different magnitude of demographic fluctuations mediated by geographical variation in the impacts of climatic oscillations (Hewitt 2000; Gómez & Lunt 2007; Canestrelli *et al*. 2006).

Despite the spatial pattern of mtDNA genetic diversity was partially similar to that inferred from nuclear markers, we did not detect a significant association between estimates of mitochondrial genetic diversity (H_D and π) and stability of climate suitability. This could be due to the observed signature of mitochondrial genetic variation was primarily shaped by historical factors predating the temporal scale of our estimates of habitat suitability stability. However, we cannot rule out that the high proportion of populations exhibiting no mitochondrial genetic variability could have weakened the statistical power of our analyses.

CONCLUSIONS

Overall, our study highlights the importance of integrating the potential effects of abiotic (*i.e.* climate and geography) and biotic components (*i.e.* inter-specific interactions) into the study of the evolutionary history of specialist taxa with narrow ecological requirements. Harnessing different genetic markers and analytical tools, our integrative approach was able to disentangle numerous cryptic aspects of the demographic trajectories of the scrub legume grasshopper across its entire distribution range. Different

lines of evidence pointed to the presence of at least one northern refugium out of the Iberian Peninsula and revealed that the stability of suitable habitats defined by both climate and host-plant species richness had important consequences on the spatial patterns of population genetic variability of the species. Future research combining high-throughput sequencing and spatially-explicit testing of biologically informed models that integrate interspecific interactions could provide a more comprehensive framework to elucidate the factors structuring genetic variation in natural populations (Massatti & Knowles 2016).

Acknowledgements

We thank to Conchi Cáliz for her advice during lab genetic work and Bernard Defaut for providing valuable information about sampling locations. We also wish to thank to Centro de Supercomputación de Galicia (CESGA) for access to computer resources. VN was supported by a FPI pre-doctoral fellowship (BES-2012-053741) from Ministerio de Economía y Competitividad. JO was supported by Severo Ochoa (SEV-2012-0262) and Ramón y Cajal (RYC-2013-12501) research fellowships. This work received financial support from research grants CGL2011-25053 and CGL2014-54671-P (Ministerio de Economía y Competitividad and European Social Fund), POII10-0197-0167 and PEII-2014023-P (Junta de Comunidades de Castilla-La Mancha and European Social Fund) and UNCM08-1E-018 (European Regional Development Fund).

REFERENCES

Abellán P, Svenning J-C (2014) Refugia within refugia - patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, 113, 13-28.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692-4693.

Anderson RP, Raza A (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, 37, 1378-1393.

Anthos (2016) *Sistema de información de las plantas de España*. Real Jardín Botánico, CSIC-Fundación Biodiversidad. *https://www.anthos.es*

Arenas M, Ray N, Currat M, Excoffier L (2012) Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution*, 29, 207-218.

Arnold TW (2010) Uninformative parameters and model selection using Akaike's Information criterion. *Journal of Wildlife Management*, 74, 1175-1178.

Ashcroft MB (2010) Identifying refugia from climate change. *Journal of Biogeography*, 37, 1407-1413.

Baele G, Lemey P, Bedford T*, et al.* (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, 29, 2157-2167.

Barton K (2015) *MuMIn: multimodel Inference*. R Package, version 1.15.6. *https://cran.r-project.org/web/packages/MuMIn/index.html*

Basiita RK, Henrich Bruggemann J, Cai N*, et al.* (2016) Erratum to: Microsatellite records for volume 7, issue 4. *Conservation Genetics Resources*, 8, 85-87.

Bates D, Maechler M, Bolker BM, Walker SC (2015) Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67, 1-48.

Beaumont MA (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41, 379-406.

Bidegaray-Batista L, Sánchez-Gracia A, Santulli G*, et al.* (2016) Imprints of multiple glacial refugia in the Pyrenees revealed by phylogeography and palaeodistribution modelling of an endemic spider. *Molecular Ecology*, 25, 2046-2064.

Borer M, Arrigo N, Buerki S, Naisbit RE, Álvarez N (2012) Climate oscillations and species interactions: large-scale congruence but regional differences in the phylogeographic structures of an alpine plant and its monophagous insect. *Journal of Biogeography*, 39, 1487-1498.

Braconnot P, Otto-Bliesner B, Harrison S*, et al.* (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, 261-277.

Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, USA.

Canestrelli D, Cimmaruta R, Costantini V, Nascetti G (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad *Bombina pachypus*, with implications for conservation. *Molecular Ecology*, 15, 3741-3754.

Cangi N, Horak IG, Apanaskevich DA*, et al.* (2013) The influence of interspecific competition and host preference on the phylogeography of two African Ixodid tick species. *Plos One*, 8, e796930.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785-789.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24, 621-631.

Collins WD, Bitz CM, Blackmon ML*, et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, 19, 2122-2143.

Cooper SJB, Hewitt GM (1993) Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Insect Molecular Biology*, 2, 185-194.

Cooper SJB, Ibrahim KM, Hewitt GM (1995) Postglacial expansion and genome subdivision in the european grasshopper *Chorthippus parallelus*. *Molecular Ecology*, 4, 49-60.

Cornille A, Giraud T, Bellard C*, et al.* (2013) Postglacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology*, 22, 2249-2263.

Cornuet J-M, Pudlo P, Veyssier J*, et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187-1189.

Criscione CD, Blouin MS (2004) Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. *Evolution*, 58, 198-202.

Darriba D, Taboada GL, Doallo R, Posada D (2012) JMODELTEST2: more models, new heuristics and parallel computing. *Nature Methods*, 9, 772-772.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (Caelifera, Acrididae, Gomphocerinae). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, 1185-1192.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution*, 29, 1969-1973.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, 11, 2571-2581.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170-1188.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23, 347-351.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.

Falush D, van Dorp L, Lawson D (2016) A tutorial on how (not) over-interpret STRUCTURE/ADMIXTURE barplots. *bioRxiv*, DOI: 10.1101/066431. *http://biorxiv.org/content/early/2016/07/28/066431*

Faye A, Deblauwe V, Mariac C*, et al.* (2016) Phylogeography of the genus *Podococcus* (Palmae/Arecaceae) in Central African rain forests: climate stability predicts unique genetic diversity. *Molecular Phylogenetics and Evolution*, 105, 126-138.

Felsenstein J (2013) PHYLIP (Phylogeny Inference Package) version 3.695. Department of Genome Sciences, University of Washington, Seattle. *http://evolution.genetics.washington.edu/phylip.html*

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294-299.

François O (2016) *Running Structure-like population genetic analyses with R*. R Tutorials in Population Genetics, U. Grenoble-Alpes. *http://membres-timc.imag.fr/Olivier.Francois/tutoRstructure.pdf*

Freeman BG, Mason NA (2015) The geographic distribution of a tropical montane bird is limited by a tree: acorn woodpeckers (*Melanerpes formicivorus*) and Colombian Oaks (*Quercus humboldtii*) in the Northern Andes. *Plos One*, 10, e0128675.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915-925.

Galbreath KE, Hafner DJ, Zamudio KR (2009) When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (American Pika, *Ochotona princeps*). *Evolution*, 63, 2848-2863.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N (Eds.) *Phylogeography in southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity*). Springer Verlag, Dordrecht, The Netherlands.

Guo Q (2012) Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology*, 21, 5396-5403.

Gutierrez-Rodríguez J, Barbosa AM, Martínez-Solano I (2016) Present and past climatic effects on the current distribution and genetic diversity of the Iberian spadefoot toad (*Pelobates cultripes*): an integrative approach. *Journal of Biogeography*, 44, 245-258.

Hampe A (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, 13, 469-471.

Harpending HC (1994) Signature of ancient population-growth in a low-resolution mitochondrial-DNA mismatch distribution. *Human Biology*, 66, 591-600.

Harter AV, Gardner KA, Falush D*, et al.* (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature*, 430, 201-205.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68, 87-112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hewitt GM (2004a) The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology*, 1, 1-16.

Hewitt GM (2004b) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 183-195.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, 139, 617-638.

Hickerson MJ, Cunningham CW (2005) Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. *Evolution*, 59, 344-360.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Hijmans RJ, van Etten J (2012) RASTER*: geographic analysis and modeling with raster data*. R package version 2.5-8. *http://CRAN.R-project.org/package=raster*

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, 11, 423-434.

Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantative Biology*, 22, 415–427.

Inoue K, Lang BK, Berg DJ (2015) Past climate change drives current genetic structure of an endangered freshwater mussel species. *Molecular Ecology*, 24, 1910-1926.

Jackson ST, Overpeck JT (2000) Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, 26, 194-220.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801-1806.

Jay F, Manel S, Alvarez N*, et al.* (2012) Forecasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology*, 21, 2354-2368.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5, 187-189.

Keppel G, Van Niel KP, Wardell-Johnson GW*, et al.* (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, 21, 393-404.

Laukkanen L, Mutikainen P, Muola A, Leimu R (2014) Plant-species diversity correlates with genetic variation of an oligophagous seed predator. *Plos One*, 9, e94105.

Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110-1116.

Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.

Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.

Llucià-Pomares D (2002) *Revision of the Orthoptera (Insecta) of Catalonia (Spain).* Monografías S.E.A., vol. 7, Zaragoza, Spain.

Massatti R, Knowles LL (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, 3974-3986.

Mouritsen KN, Poulin R (2002) Parasitism, climate oscillations and the structure of natural communities. *Oikos*, 97, 462-468.

Noguerales V, Cordero PJ, Ortego J (2016a) Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, 16, 96.

Noguerales V, García-Navas V, Cordero PJ, 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, 2129-2142.

Noguerales V, Traba J, Mata C, Morales M (2015) Winter habitat selection and partitioning in two sympatric farmland small mammals: *Apodemus sylvaticus* and *Mus spretus*. *Revue d'écologie (La Terre et la Vie)*, 70, 70-82.

Noguerales V, Cordero PJ, Ortego J (2017) Testing the role of ancient and contemporary landscapes on structuring genetic variation in a specialist grasshopper. *Ecology and Evolution*, 7, 3110-3122.

Nogués-Bravo D (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, 18, 521-531.

Ortego J, Aguirre MP, Noguerales V, Cordero PJ (2015a) Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications*, 8, 621-632.

Ortego J, Noguerales V, Gugger PF, Sork VL (2015b) Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology*, 24, 6188-6208.

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the Mid-Aegean trench calibration. *Molecular Biology and Evolution*, 27, 1659-1672.

Parducci L, Jorgensen T, Tollefsrud MM*, et al.* (2012) Glacial survival of boreal trees in Northern Scandinavia. *Science*, 335, 1083-1086.

Peterson AT, Soberon J, Sánchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science*, 285, 1265-1267.

Petit JR, Jouzel J, Raynaud D*, et al.* (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399, 429-436.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

Phillips SJ, Dudik M (2008) Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Picaud F, Bonnet E, Gloaguen V, Petit D (2003) Decision making for food choice by grasshoppers (Orthoptera : Acrididae): Comparison between a specialist species on a shrubby legume and three graminivorous species. *Environmental Entomology*, 32, 680-688.

Picaud F, Gloaguen V, Petit D (2002) Mechanistic aspects to feeding preferences in *Chorthippus binotatus* (Acrididae, gomphocerinae). *Journal of Insect Behavior*, 15, 513-526.

Platz J, Cloupeau R (2010) Liste rouge commentée des Orthopéres de la región Centre. *Materiaux Orthopériques et Entomocénotiques,* 15, 17-33.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Pritchard JK, Wen X, Falush D (2010) Documentation for STRUCTURE software: version 2.3. *http://pritchardlab.stanford.edu/structure.html*

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, 23, 564-571.

Qu Y, Ericson PGP, Quan Q*, et al.* (2014) Long-term isolation and stability explain high genetic diversity in the Eastern Himalaya. *Molecular Ecology*, 23, 705-720.

R Core Team (2015). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org/*

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19, 2092-2100.

Ricanova S, Koshev Y, Rican O*, et al.* (2013) Multilocus phylogeography of the European ground squirrel: cryptic interglacial refugia of continental climate in Europe. *Molecular Ecology*, 22, 4256-4269.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223-225.

Rogers AR, Harpending H (1992) Population-growth makes waves in the distribution of pairwise genetic-differences. *Molecular Biology and Evolution*, 9, 552-569.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137-138.

Salvi D, James Harris D, Kaliontzopoulou A, Carretero MA, Pinho C (2013) Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. *BMC Evolutionary Biology*, 13, 147.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4, 11.

Schmitt T, Varga Z (2012) Extra-Mediterranean refugia: the rule and not the exception? *Frontiers in Zoology*, 9, 22.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: application to human mitochondrial DNA. *Genetics*, 152, 1079-1089.

Schweiger O, Settele J, Kudrna O, Klotz S, Kuhn I (2008) Climate change can cause spatial mismatch of trophically interacting species. *Ecology*, 89, 3472-3479.

Snyder CW (2016) Evolution of global temperature over the past two million years. *Nature*, 538, 226-228.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, 16, 608-613.

Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences*, 277, 661-671.

Svenning J-C, Flojgaard C, Marske KA, Nogués-Bravo D, Normand S (2011) Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, 30, 2930-2947.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7, 453-464.

Talavera S, Aedo C, Castroviejo S, Romero C, Sáez L, Salgueiro FJ, Velayos M (2001) *Flora Ibérica VII/1 Leguminoseae (partim)*. Real Jardín Botánico, CSIC, Madrid, Spain.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.

Tsai Y-HE, Manos PS (2010) Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana*. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17035-17040.

Tsuda Y, Nakao K, Ide Y, Tsumura Y (2015) The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. *Molecular Ecology*, 24, 1403-1418.

Ursenbacher S, Guillon M, Cubizolle H*, et al.* (2015) Postglacial recolonization in a cold climate specialist in western Europe: patterns of genetic diversity in the adder (*Vipera berus*) support the central-marginal hypothesis. *Molecular Ecology*, 24, 3639-3651.

Vega R, Flojgaard C, Lira-Noriega A*, et al.* (2010) Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography*, 33, 260-271.

Warren DL, Glor RE, Turelli M (2010) ENMTOOLS: a toolbox for comparative studies of environmental niche models. *Ecography*, 33, 607-611.

Warren DL, Seifert SN (2011) Ecological niche modeling in MAXENT: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335-342.

Wharton TN, Kriticos DJ (2004) The fundamental and realized niche of the Monterey Pine aphid, *Essigela californica* (Essig) (Hemiptera: Aphididae): implications for managing softwood plantations in Australia. *Diversity and Distributions*, 10, 253-262.

Willis KJ, van Andel TH (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews*, 23, 2369-2387.

Yannic G, Pellissier L, Ortego J*, et al.* (2014) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, 4, 132-137.

SUPPLEMENTARY MATERIAL

Table S1 – Geographical location and elevation of the studied populations of scrub-legume grasshopper (*Chorthippus binotatus binotatus*), number of individuals analysed for nuclear microsatellite markers (μsat) and mitochondrial cytochrome oxidase subunit I (COI) gene fragment, and population groups inferred by SAMOVA analyses.

Code	Locality	Country	Latitude	Longitude	Elevation (m.a.s.l)	μsat	COI	SAMOVA
1	Bodelio	France	47.6992	-2.2803	64	20	5	1
2	Le Porge	France	44.8431	-1.1875	7	15	7	1
3	Le Phare de la Coubre	France	45.7344	-1.2194	31	3	2	-
4	Col de Fontfroide	France	43.5990	2.8397	950	12	5	1
5	Gordes	France	43.9527	5.2256	461	16	5	2
6	Borau	Spain	42.6778	-0.5695	1534	20	5	2
7	Buerba	Spain	42.5344	0.0569	1136	20	5	3
8	Sierra del Montsec	Spain	42.0361	0.7399	1249	9	5	4
9	Montant de Tost	Spain	42.2386	1.3804	1182	11	6	4
10	Macizo del Moncayo	Spain	41.7892	-1.8201	1700	16	5	3
11	Coll Manado	Spain	40.7173	0.1676	1243	20	5	4
12	Sierra del Toro	Spain	40.0047	-0.7063	1000	20	5	4
13	Teruel	Spain	40.3268	-1.2142	1091	6	2	-
14	Pico Javalambre	Spain	40.1015	-1.0120	1886	3	2	-
15	Sierra Espuña	Spain	37.8652	-1.5713	1514	18	6	4
16	Pico Almenaras	Spain	38.5435	-2.4402	1655	19	6	4
17	Sierra de Cabras (Fte. Carrasca)	Spain	38.0455	-2.3961	1526	20	-	-
18	Sierra de Cabras (Pico Cabras)	Spain	38.0626	-2.3999	2038	8	2	-
19	Puerto del Pinar	Spain	38.0463	-2.4851	1662	20	5	4
20	Poyotello	Spain	38.1195	-2.6165	1600	20	-	-
21	Sierra de María (Bco. Molina)	Spain	37.6621	-2.2467	1670	20	8	4
22	Sierra de María (Portal Chico)	Spain	37.6764	-2.1831	1860	5	-	
23	Sierra Cazorla	Spain	37.8106	-2.9626	1772	20	5	4
24	Sierra Filabres	Spain	37.2197	-2.5307	2117	20	-	-
25	Sierra de Baza	Spain	37.2273	-2.7521	1876	20	5	4
26	Sierra de Gador	Spain	36.9015	-2.8039	2017	11	2	-
27	Sierra Nevada (Puerto Ragua)	Spain	37.1165	-3.0260	2100	20	2	-
28	Sierra Nevada (Cdo. Ruquino)	Spain	37.0925	-3.4814	1750	4	-	-
29	Sierra Nevada (Campos Otero)	Spain	37.1101	-3.4051	2314	20	5	4
30	Sierra Arana	Spain	37.3313	-3.5162	1736	15	2	-
31	Sierra Mágina	Spain	37.7442	-3.5490	1325	23	4	4
32	La Pandera	Spain	37.6369	-3.8059	1512	20	5	4
33	Sierra de Parapanda	Spain	37.3055	-3.9297	1578	20	-	-
34	Pico Maroma	Spain	36.9187	-4.0668	1595	20	5	4

(Table S1 continues in the next page)

(continuation of Table S1)

Code	Locality	Country	Latitude	Longitude	Elevation (m.a.s.l)	μsat	COI	SAMOVA
35	Serranía de Ronda	Spain	36.6640	-5.2298	958	20	5	4
36	Sierra de la Demanda	Spain	42.1653	-3.0730	1679	13	5	3
37	Portillo Lunada	Spain	43.1709	-3.6547	1350	20	-	-
38	Cardaño de Arriba	Spain	42.9314	-4.6481	1307	12	7	3
39	Lagos de Saliencia	Spain	43.0550	-6.1018	1600	10	-	-
40	Puerto de Gañidoira	Spain	43.5384	-7.6477	521	14	6	3
41	Vedra	Spain	42.7865	-8.4668	237	12	4	3
42	Astorga	Spain	42.5791	-6.2138	1094	11	5	4
43	Arribes del Duero	Spain	41.5007	-6.0795	718	10	-	-
44	Puerto de la Quesera	Spain	41.2158	-3.4204	1730	12	5	3
45	Puerto de la Morcuera	Spain	40.8301	-3.8305	1790	12	5	3
46	Puerto Peña Negra	Spain	40.4222	-5.2974	1907	4	-	-
47	Puerto de Mijares	Spain	40.3321	-4.8158	1609	12	5	3
48	Plataforma de Gredos	Spain	40.2718	-5.2504	2020	14	6	3
49	Canalizo	Spain	40.3391	-5.7297	1830	7	2	-
50	Monsagro	Spain	40.5248	-6.3820	887	4	3	-
51	Puerto de San Martin	Spain	40.2458	-6.7776	1031	12	6	3
52	Serra da Estrela	Portugal	40.3173	-7.5684	1621	14	5	3
53	Puerto del Lanchar	Spain	39.5214	-4.3769	1068	14	5	4
54	Pico Miravete	Spain	39.7132	-5.7461	835	13	5	4

Table S2 – Likelihoods and AICM values of different clock and demographic models used to infer an ultrametric tree and estimate divergence times for cytochrome oxidase subunit I (COI) sequences in BEAST.

Clock	Tree prior	Likelihood (mean ± SD)	AICM (mean ± SD)
Strict	Constant	-1001.33 ± 8.70	2154.17 ± 0.18
Strict	Exponential growth	-1002.12 ± 9.00	2166.57 ± 0.21
Uncorrelated relaxed lognormal	Constant	-989.48 ± 9.57	2182.16 ± 0.18
Uncorrelated relaxed lognormal	Exponential growth	-999.49 ± 9.97	2197.99 ± 0.43

Table S3 – Mitochondrial genetic diversity estimates, results of demographic expansion tests, and nucleotide substitution models used in Bayesian skyline analyses for the major lineages of the scrub-legume grasshopper. Number of individuals (N), number of polymorphic sites (*S*), number of haplotypes (*H*), haplotype diversity (H_D), nucleotide diversity (π), Fús F_s, Tajima´s *D* and Ramos-Onsins and Rozas R_2 statistic. Asterisks indicate significant deviations (F_s, *D* and R_2) from the null hypothesis of constant population size.

Lineage	N	*S*	*H*	H_D	π	F_s	*D*	R_2	Skyline model
Northwestern and southeastern France	19	1	2	0.351	0.0006	0.758	0.417	0.175	HKY
Eastern France and Pyrenees	13	0	1	0.000	0.0000	-	-	-	-
All France and Pyrenees[1]	32	3	3	0.619	0.0022	2.251	1.479	0.950	HKY
Northern Iberia	63	8	9	0.267	0.0007	-8.877*	-1.944*	0.044*	HKY
South and southeastern Iberia	124	10	11	0.269	0.0007	-11.428*	-1.950*	0.026*	HKY
All populations[2]	219	21	23	0.699	0.0064	-3.941	0.026	0.084	-

[1]*This group includes both "Northwestern and southeastern France" and "Eastern France and Pyrenees" lineages.*

[2]*This group includes all Iberian and French populations.*

Table S4 – Prior distributions of demographic and mutation model parameters used in Approximate Bayesian Computation (ABC) analyses testing different scenarios of population divergence and past demography for the scrub-legume grasshopper. Graphical representations of all tested scenarios are presented in Fig. 2.

Priors for the demographic parameters	
N_{NWSEF}	UN ~ [10 – 1 000 000]
$N_{NWSEF-B}$	UN ~ [10 – 500 000]
N_{EF}	UN ~ [10 – 1 000 000]
N_{EF-B}	UN ~ [10 – 500 000]
N_{NI}	UN ~ [10 – 1 000 000]
N_{EI}	UN ~ [10 – 1 000 000]
N_{SI}	UN ~ [10 – 1 000 000]
N_F	UN ~ [10 – 1 000 000]
N_I	UN ~ [10 – 1 000 000]
N_{ESI}	UN ~ [10 – 1 000 000]
N_X	UN ~ [10 – 1 000 000]
t_1	UN ~ [10 – 1 000 000]
t_2	UN ~ [10 – 1 000 000]
t_3	UN ~ [10 – 1 000 000]
Constraint on parameter *t*	$t_2 > t_1$; $t_3 > t_2$; $t_3 > t_1$
Constraint on parameters N_e	$N_{NF-B} < N_F$; $N_{SF-B} < N_F$
Priors for the mutation model for microsatellites markers	
Mean mutation rate (μ)	LU ~ [1.0 x 10^{-7} – 1.0 x 10^{-3}]
Individual locus mutation rate	GAM ~ [1.0 x 10^{-7} – 1.0 x 10^{-2}, 5.0 x 10^{-4}, 2]
Mean coefficient (*p*)	GAM ~ [1.0 x 10^{-1} – 1.0, 0.5, 2]
Individual locus coefficient (*p*)	GAM ~ [1.0 x 10^{-2} – 1.0, 0.5, 2]
Priors for the mutation model for mitochondrial marker (COI)	
Mean mutation rate (μ) per site per generation	LU ~ [1.0 x 10^{-9} – 1.0 x 10^{-6}]
Individual locus mutation rate	GAM ~ [1.0 x 10^{-9} – 1.0 x 10^{-6}, 5.0 x 10^{-8}, 2]
Mean coefficient (*k*) C/T	UN ~ [0.05 – 30]
Individual locus coefficient (*k*) C/T	GAM ~ [0.05 – 30, 10, 2]
Summary statistics (SS) for microsatellite markers	
Mean number of alleles for each population	
Mean genic diversity for each population	
Mean allele size variance for each population	
Mean Garza-Williamson's *M* for each population	
F_{ST} for each pair of populations	
Summary statistics (SS) for mitochondrial marker (COI)	
Number of haplotypes for each population	
Number of segregation sites for each population	
Tajima´s *D* for each population	
Mean of number of the rarest nucleotide at segregation sites for each population	
Number of haplotypes for each pair of populations	
Mean pairwise difference	
F_{ST} for each pair of populations	

(Table S4 continues in the next page)

(continuation of Table S4)

N_{NWSEF}, effective population size of the northwestern and southeastern France group
$N_{NWSEF\text{-}B}$, effective population size of the northwestern and southeastern France group (assuming a population bottleneck event)
N_{EF}, effective population size of the eastern France and Pyrenees group
$N_{EF\text{-}B}$, effective population size of the eastern France and Pyrenees group (assuming a population bottleneck event)
N_{NI}, effective population size of the northern Iberia group
N_{EI}, effective population size of the eastern Iberia group
N_{SI}, effective population size of the southern Iberia group
N_{F}, effective population size of the ancestral France group
N_{I}, effective population size of the ancestral Iberia group
N_{ESI}, effective population size of the ancestral south and eastern Iberia group
N_{X}, effective population size of the most ancestral population
t_1, time (in generations = years) to the most recent divergence event
t_2, time (in generations = years) to the intermediate divergence event
t_2, time (in generations = years) to the most ancient divergence event
UN, uniform distribution with two parameters [min-max]
LU, Log-uniform distribution with two parameters [min-max]
GAM, gamma distribution with four parameters [min-max, mean, shape]

Table S5 – Summary of ecological niche modeling (ENM) for the scrub-legume grasshopper and its host-plant species. Models for the scrub-legume grasshopper were built using (i) only climatic variables, (ii) climatic variables and host-plant species richness and (iii) only host-plant species richness. We report the number of occurrence points (n) employed in MAXENT model building; model performance assessed by the area under the receiver operating characteristic curve (AUC). The three ENMs for the scrub-legume grasshopper were compared by sample-size adjusted Akaike's information criterion value (AIC_c); ΔAIC_c, difference in AIC_c value from that of the most supported model; $perm_{50\%}$, variables(s) with the highest permutation contribution to the model (>50% based on the permutation importance statistic); $Jack_{ISOL}$, the variable with the highest gain when used in isolation; $Jack_{OMIT}$, the variable that decreases the gain the most when it is omitted. For the host-plants species-specific ENMs, we also report the logistic values for species occurrence based on the maximum training sensitivity plus specificity (MTSS) threshold. Maps of host-plant species richness were obtained by converting the logistic output of species-specific ENM maps into binary maps (presence = 1; absence = 0) using thresholds of species occurrence based on the MTSS values and then summing all binary maps of all plant species for each period (current and Last Glacial Maximum, LGM, *c*. 21 Kya).

Model	n	AUC_{TEST} (mean ± SD)	Parameters	AIC_c	ΔAIC_c	MTSS	$perm_{50\%}$	$Jack_{ISOL}$	$Jack_{OMIT}$
Scrub-legume grasshopper									
(i) 19 WorldClim variables	287	0.864 ± 0.035	98	7581.53	0.00	-	bio0 + bio14 + bio17+ bio11	bio8	bio17
(ii) 19 WorldClim variables + Host-plant species richness (HP_{rich})	287	0.867 ± 0.034	107	7591.57	10.03	-	bio14 + bio17 + bio0+ bio13	HP_{rich}	HP_{rich}
(iii) Host-plant species richness (HP_{rich})	287	0.788 ± 0.035	12	7917.58	336.05	-	HP_{rich}	HP_{rich}	HP_{rich}
Host-plant species									
Cytisus oromediterraneus	1 002	0.912 ± 0.005	-	-	-	0.209	bio0 + bio11 + bio3	bio5	bio3
Cytisus scoparius	35 669	0.621 ± 0.003	-	-	-	0.462	bio3 + bio9 + bio10 + bio17	bio10	bio3
Echinospartium bardanessi	61	0.959 ± 0.057	-	-	-	0.183	bio0 + bio13	bio14	bio6
Echinospartium boisseiri	488	0.955 ± 0.011	-	-	-	0.113	bio13	bio7	bio13
Echinospartium horridum	208	0.963 ± 0.010	-	-	-	0.188	bio11 + bio18	bio14	bio18
Erinacea anthyllis	1 500	0.921 ± 0.005	-	-	-	0.317	bio17 + bio12 + bio11	bio13	bio14
Genista cinerea	1 600	0.910 ± 0.009	-	-	-	0.293	bio13 + bio16 + bio11 + bio1	bio1	bio13
Genista scorpius	1 092	0.737 ± 0.006	-	-	-	0.340	bio14 + bio2 + bio13 + bio12 + bio18	bio13	bio14
Genista versicolor	288	0.983 ± 0.005	-	-	-	0.077	bio13	bio7	bo13
Ulex europaeus	16 874	0.714 ± 0.006	-	-	-	0.366	bio3 + bio9 + bio10	bio8	bio3
Ulex gallii	3 965	0.827 ± 0.005	-	-	-	0.367	bio3	bio3	bio3
Ulex minor	3 280	0.809 ± 0.008	-	-	-	0.380	bio13 + bio2 + bio4 + bio3	bio4	bio2
Ulex parviflorus	6 559	0.795 ± 0.005	-	-	-	0.383	bio17 + bio11 + bio1+ bio14	bio17	bio1

Table S6 – Model selection to analyze the association between genetic diversity of scrub-legume grasshopper (nuclear: A_R, allelic richness standardized for sample size; H_E, expected heterozygosity; mitochondrial: H_D, haplotype diversity; π, nucleotide diversity) and (i) stability of climate suitability (C_{STA}) based on a MAXENT model built using only climatic variables, (ii) stability of habitat suitability ($C\text{-}HP_{STA}$) based on a MAXENT model built using climatic variables and host-plant species richness and (iii) stability of habitat suitability (HP_{STA}) based on a MAXENT model built using only host-plant species richness. Latitude (Lat) and longitude (Lon) were also included as covariates in the models. Details are given for the best ranked equivalent models ($\Delta AIC_C \leq 2$). For each model we indicate *K*, number of parameters in the model; AIC_C, sample-size adjusted Akaike's information criterion (AIC) value; ΔAIC_C, difference in AIC_C value from that of the most supported model; ω_i, AIC_C weight.

Nuclear						Mitochondrial					
Model	Parameters	*K*	AIC_C	ΔAIC_C	ω_i	Model	Parameters	*K*	AIC_C	ΔAIC_C	ω_i
A_R						H_D					
1	Lat + Lon + C_{STA}	3	64.50	0.00	0.616	1	Lat + Lon	2	17.40	0.00	0.472
2	Lat + Lon + $C\text{-}HP_{STA}$	3	65.44	0.94	0.384	2	Lat	1	18.38	0.98	0.289
						3	Null model	0	18.76	1.36	0.239
H_E						π					
1	Lat + Lon + C_{STA}	3	-156.60	0.00	0.642	1	Null model	0	-340.10	0.00	0.542
2	Lat + Lon + $C\text{-}HP_{STA}$	3	-155.43	1.17	0.358	2	Lat	1	-338.54	1.56	0.248
						3	Lon	1	-338.20	1.90	0.210

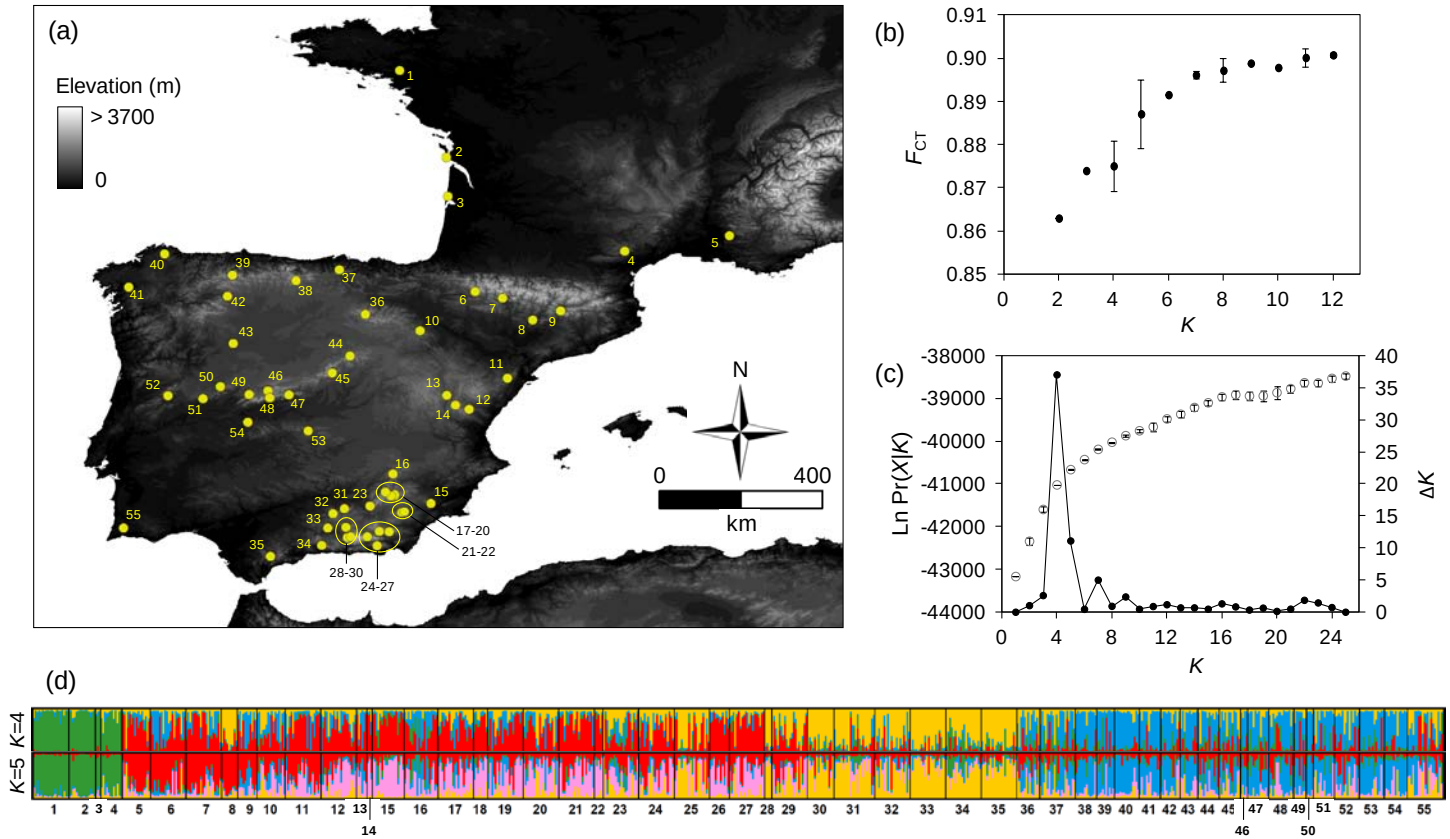


Figure S1 – Panel (a) shows the geographic location of sampling sites of scrub-legume grasshopper across the species distribution range. Panel (b) shows the genetic differentiation between populations (fixation index, F_{CT}) for mitochondrial cytochrome oxidase subunit I (COI) gene fragment as a function of the user-defined number of population groups (*K*) inferred over 10 runs using SAMOVA. Panel (c) represents the results of Bayesian clustering analyses in STRUCTURE for nuclear microsatellite markers. Mean (± SD) log probability of the data [Ln Pr (X|*K*)] over 10 runs (left axis, open dots and error bars) for each *K*-value. The magnitude of Δ*K* as a function of *K* determines the best-supported number of clusters (*K* = 4) in STRUCTURE analyses (right axis and black dots). (d) Results of genetic assignments of 794 individuals of scrub-legume grasshopper based on the Bayesian methods implemented in STRUCTURE for *K* = 4 and *K* = 5. Each individual corresponds to a vertical bar, which is partitioned into *K*-coloured segments that represent the individual's probability of belonging to the cluster with that colour. Black lines separate individuals from different populations. Population codes and further information on sampling sites are given in Table S1.

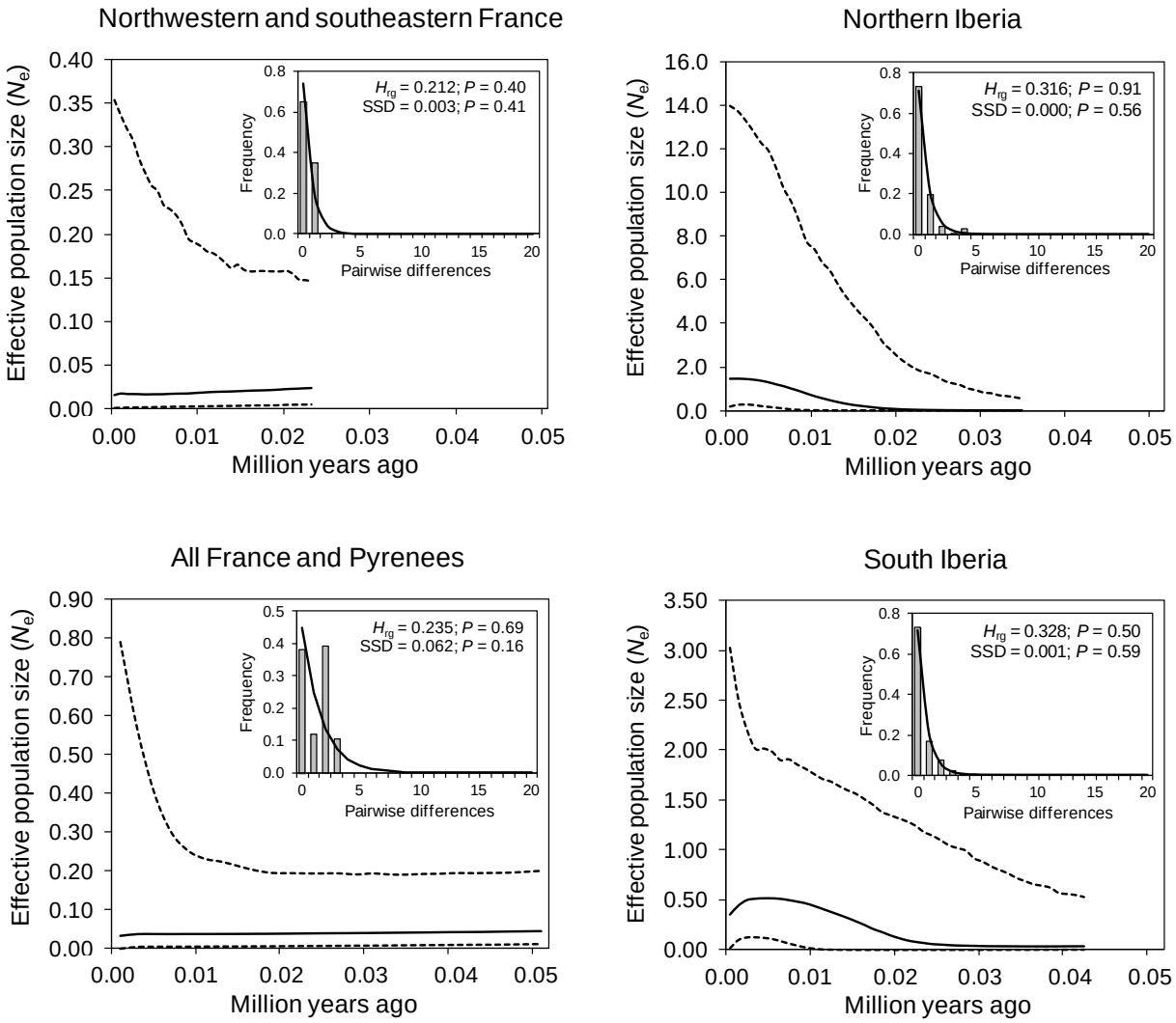


Figure S2 – Demographic history of the major lineages of scrub-legume grasshopper according to Bayesian skyline plots (BSP) and mismatch distribution analyses. Analyses for the lineage from eastern France and the Pyrenees were not performed because this lineage consists only of one haplotype (see Table S3). BSP represent historical demographic trends of each lineage during the last 0.05 million years; the median (solid lines) and 95% highest posterior density (dashed lines) values of the $\log_{10}$ of the effective population size (N_e) are represented. Insets represent the pairwise mismatch distribution for each lineage; histograms represent the observed frequency of pairwise nucleotide differences and solid lines represent the expected distribution under the sudden expansion model. Results of goodness-of-fit tests between observed and expected distributions were performed using the raggedness index (H_{rg}) and sum of square deviations (SSD) and the results for each lineage are reported inside mismatch distribution graphs.

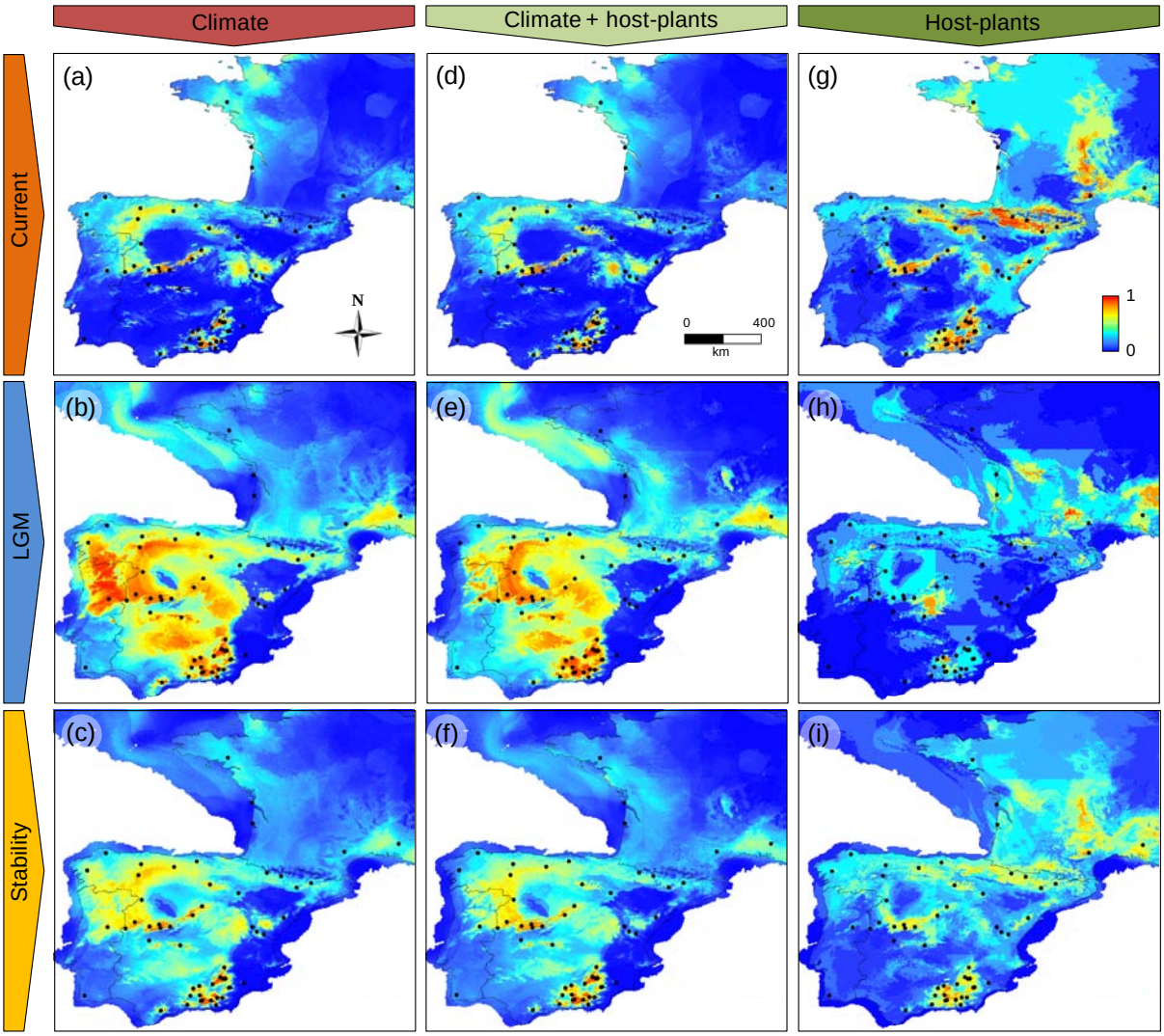


Figure S3 – Ecological niche models (ENM) of the scrub-legume grasshopper built in MAXENT using only climatic variables (panels a, b, c), using climatic variables and host-plant species richness (panels d, e, f), and using only host-plant species richness (panels g, h, i). The maps show the habitat suitability for the scrub-legume grasshopper during the present (panels a, d, g) and the Last Glacial Maximum (LGM, *c*. 21 Kya) (panels b, e, h). Maps at the bottom show habitat suitability stability since the LGM for the three different models (panels c, f, i). Colour scale refers to climate/habitat suitability based on the MAXENT logistic output, with increasingly warmer colours indicating increasing suitability for the scrub-legume grasshopper. Spatial data of host-plant species richness were obtained by converting logistic output of species-specific ENM maps into binary maps (presence = 1; absence = 0) using thresholds of species occurrence based on the maximum training sensitivity plus specificity (MTSS) values and then summing all binary maps of all plant species for each period.

CAPÍTULO IV

Testing the role of ancient and contemporary landscapes on structuring genetic variation in a specialist grasshopper

VÍCTOR NOGUERALES, PEDRO J. CORDERO & JOAQUÍN ORTEGO

Ecology and Evolution, 7, 3110-3122 (2017)

Testing the role of ancient and contemporary landscapes on structuring genetic variation in a specialist grasshopper

Abstract

Understanding the processes underlying spatial patterns of genetic diversity and structure of natural populations is a central topic in evolutionary biogeography. In this study, we combine data on ancient and contemporary landscape composition to get a comprehensive view of the factors shaping genetic variation across the populations of the scrub-legume grasshopper (*Chorthippus binotatus binotatus*) from the biogeographically complex region of southeast Iberia. First, we examined geographic patterns of genetic structure and employed an Approximate Bayesian Computation (ABC) approach to compare different plausible scenarios of population divergence. Second, we used a landscape genetic framework to test for the effects of i) Late Miocene paleogeography, ii) Pleistocene climate fluctuations, and iii) contemporary topographic complexity on the spatial patterns of population genetic differentiation. Genetic structure and ABC analyses supported the presence of three genetic clusters and a sequential west to east splitting model that predated the Last Glacial Maximum (LGM, *c*. 21 Kya). Landscape genetic analyses revealed that population genetic differentiation was primarily shaped by contemporary topographic complexity, but was not explained by any paleogeographic scenario or resistance distances based on climate suitability in the present or during the LGM. Overall, this study emphasizes the need of integrating information on ancient and contemporary landscape composition to get a comprehensive view of their relative importance to explain spatial patterns of genetic variation in organisms inhabiting regions with complex biogeographical histories.

INTRODUCTION

Understanding the mechanisms that shape spatial patterns of genetic diversity and structure is a central topic in evolutionary biogeography (Peterman *et al*. 2014; Yannic *et al*. 2014; Habel *et al*. 2015). Present day landscape configuration and the geographic distribution of suitable habitats, jointly with species-specific ecological characteristics, define contemporary inter-population dispersal and realized gene flow (Edwards *et al*. 2012; Castillo *et al*. 2014). However, past climate changes and ancient geological events have also greatly altered the spatial configuration of corridors and barriers to gene flow (Pepper *et al*. 2008; He *et al*. 2013). Such temporal shifts in landscape structure and dispersal routes have often left genetic signatures in contemporary populations that are useful to track back in time their past demographic trajectories (He *et al*. 2013; Lanier *et al*. 2015). Thus, the study of how present-day and past landscape composition have impacted gene flow is necessary to get a

comprehensive view of the processes underlying spatial patterns of genetic diversity and structure of natural populations, which can ultimately help to predict their responses to ongoing or future environmental changes (Fordham *et al*. 2014; Yannic *et al*. 2014).

Quaternary climatic fluctuations, characterized by cold glacial stages alternated with warm interglacial periods, have strongly influenced the demography of many organisms during the past two million years (*c*. 2-0.04 Mya) (Hewitt 2000). During glacial periods, the distribution ranges of most species from temperate zones contracted and their populations persisted in refugia located at lower elevations or latitudes (Homburg *et al*. 2013; Qu *et al*. 2014). Conversely, the populations from cool-adapted species expanded during glacial periods and shrank during interglacials (Canestrelli & Nascetti 2008). Under any scenario, populations from regions subjected to major climate changes experience fluctuating demographic dynamics that, ultimately, are expected to reduce their effective population sizes, increase genetic drift, and erode local levels of genetic diversity (Carnaval *et al*. 2009; Brown & Knowles 2012; Yannic *et al*. 2014). However, populations from climatically unstable areas can recurrently go extinct and be re-colonized by immigrants from multiple source populations, which can increase local levels of genetic diversity via admixture (Petit *et al*. 2003; Ortego *et al*. 2015b). Thus, the stability of climatically suitable habitats can impact patterns of genetic diversity and admixture in opposite directions, a possibility that has been generally overlooked (Ortego *et al*. 2015b). Beyond Quaternary climatic fluctuations and contemporary landscape features, much older paleogeological events such as up-lifting of mountain ranges or the emergence of islands and sea corridors are also considered important factors responsible of geographic patterns of genetic differentiation in many taxa (Papadopoulou *et al*. 2009; Mastretta-Yanés *et al*. 2015; Ceccarelli *et al*. 2016). Although ancient geological changes are known to underlie the spatial patterns of genetic divergence found in several organisms (Ortego *et al*. 2009; Abellán *et al*. 2012; Cheng *et al*. 2016; Opell *et al*. 2016), in many other cases the genetic signals left by paleogeological events are expected to have been totally or partially eroded as a result of gene flow promoted by subsequent landscape changes (Pepper *et al*. 2008; Graham *et al*. 2015). Thus, examining landscape configuration at different time periods can help to better understand the mechanisms by which intraspecific genetic diversity and differentiation arise and are maintained (Reilly *et al*. 2015).

The mountainous area of southeast Iberia has undergone remarkable geological changes that have shaped the complex biogeographic history of the region. Geological reconstructions based on stratigraphic and sedimentary data show that the emergence of mountain chains in the Tortonian

(*c*. 12 Mya) configured a mosaic of islands (hereafter Betic Islands) at the confluence of European and African continental platforms. The rotation of the Betic Islands towards the Iberian Peninsula, in combination with sedimentation processes, resulted in their fusion to the continent and the configuration of a continuous emerged landscape that is currently conformed by the Prebetic, Penibetic and Subbetic mountain ranges of southeast Iberia (Braga *et al*. 2003, 2010; Martín *et al*. 2009). Furthermore, this area is also the southernmost limit of the influence of Quaternary glaciations (*c*. 2-0.04 Mya) in Europe, during which vast portions of land were free of permanent ice at elevations below 2 500 m.a.s.l. (Hughes & Woodward 2008) and constituted an important refugium for biota from temperate habitats (Hewitt 2000). The magnitude and complexity of these paleogeological and climate events are considered the most important engines of diversification and genetic structuring of many taxa in the region (Fromhage *et al*. 2004; Andújar *et al*. 2012; Faille *et al*. 2014). For all these reasons, southeast Iberia is an ideal template for testing the combined effects of ancient and more contemporary climate and landscape changes on spatial patterns of genetic diversity and structure of local populations (Faille *et al*. 2014).

The scrub-legume grasshopper (*Chorthippus binotatus binotatus* Charpentier 1825) (Orthoptera: Acrididae) is a winged Orthoptera with a one-year generation time (Fig. 1; Defaut 2011). This species is primarily distributed in montane regions from southwest Europe, including France and the Iberian Peninsula (Defaut 2011). The scrub-legume grasshopper is an oligophagous species that exclusively feeds on some scrub-legume taxa from the tribe *Genisteae* (Defaut 2011). In southeast Iberia, the host-plants (primarily *Erinacea anthyllis* and, more occasionally, *Echinospartum boissieri*, *Genista versicolor* and *Ulex parviflorus*) form scattered vegetation patches located at moderate to high elevations (> 1200 m.a.s.l.). This fact restricts the distribution of the scrub-legume grasshopper to the different mountain ranges of the region (Prebetic, Penibetic and Subbetic systems) (Defaut 2011; Table S1). Thus, the narrow ecological requirements of the scrub-legume grasshopper, the patchy distribution of its host-plants, and the limited dispersal abilities of the species (low flying capacity; V.N., P.J.C and J.O., pers. obs.), have resulted in most of its populations from southeast Iberia being currently highly fragmented and separated by extensive lowlands of unsuitable habitats (Defaut 2011).

Here, we used the scrub-legume grasshopper as model system to analyse the contribution of contemporary (present-day topography and distribution of climatically suitable habitats) and historical (paleoclimate-based distribution of suitable habitats and Late Miocene paleogeography) factors on shaping spatial patterns of genetic diversity and structure across the populations of the species from

southeast Iberia. In particular, we first (i) examined geographic patterns of genetic structure and employed an Approximate Bayesian Computation (ABC) framework to compare different plausible scenarios of population divergence (Beaumont 2010; Cornuet *et al*. 2014). Second, we (ii) applied circuit theory to test whether observed patterns of genetic differentiation are explained by a comprehensive suite of isolation-by-resistance (IBR) scenarios (McRae 2006; McRae & Beier 2007), including paleogeography at different time periods since Late Miocene (*c*. 12.0-7.0 Mya; Martín *et al*. 2009), current and Last Glacial Maximum (LGM, *c*. 21 Kya) climate suitability and stability, and contemporary topographic complexity. Finally, we (iii) tested the hypothesis predicting more genetic diversity in populations from areas with high past and present climate suitability and stability since the LGM.

MATERIAL AND METHODS

POPULATION SAMPLING

In 2012 and 2013, we collected 354 individuals from 19 populations of scrub-legume grasshopper from southeast Iberia (~80 000 km^2) (Table S1; Fig. 2-4). Our sampling included populations from all mountain ranges in the region (Prebetic, Penibetic and Subbetic ranges) and covered the entire elevation range of the scrub-legume grasshopper in the study area (958-2 314 m.a.s.l.; Table S1).

Figure 1 – Scrub-legume grasshopper (*Chorthippus binotatus binotatus*), the study organism. The photography shows a male specimen on a legume host-plant of the genus *Ulex* (tribe *Genisteae*). Photography by Víctor Noguerales.

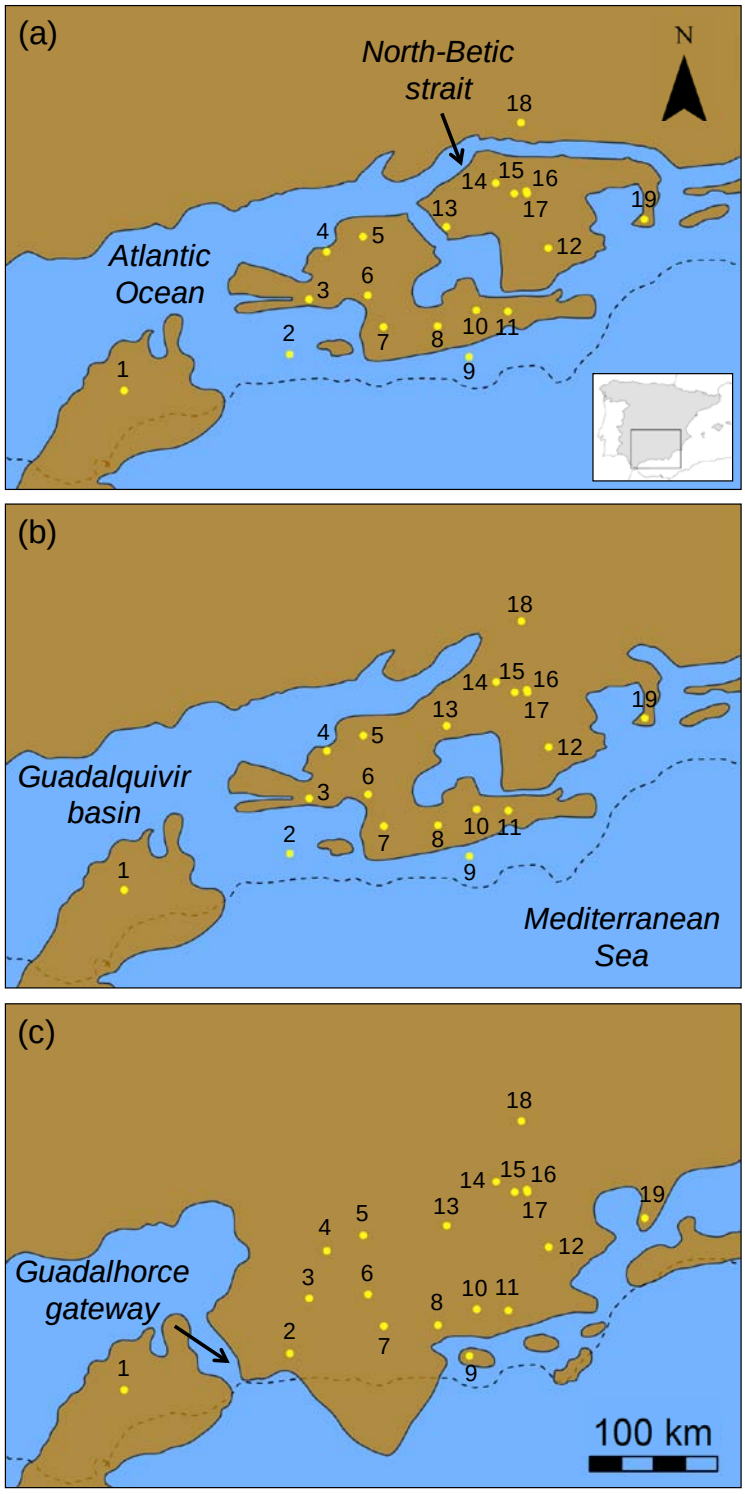


Figure 2 – Paleogeographic maps showing the spatial configuration of emerged lands in the study area during the (a) Early Tortonian (*c.* 12.0-11.6 Mya), (b) Late Tortonian (*c.* 8.0-7.3 Mya) and (c) Earliest Messinian (*c.* 7.2-7.0 Mya) according to Martín *et al.* (2009). Yellow dots indicate the location of sampled populations (number codes as in Table S1). Dashed lines represent continental limits in the present. Inset map from panel a) shows the location of our study area within the Iberian Peninsula.

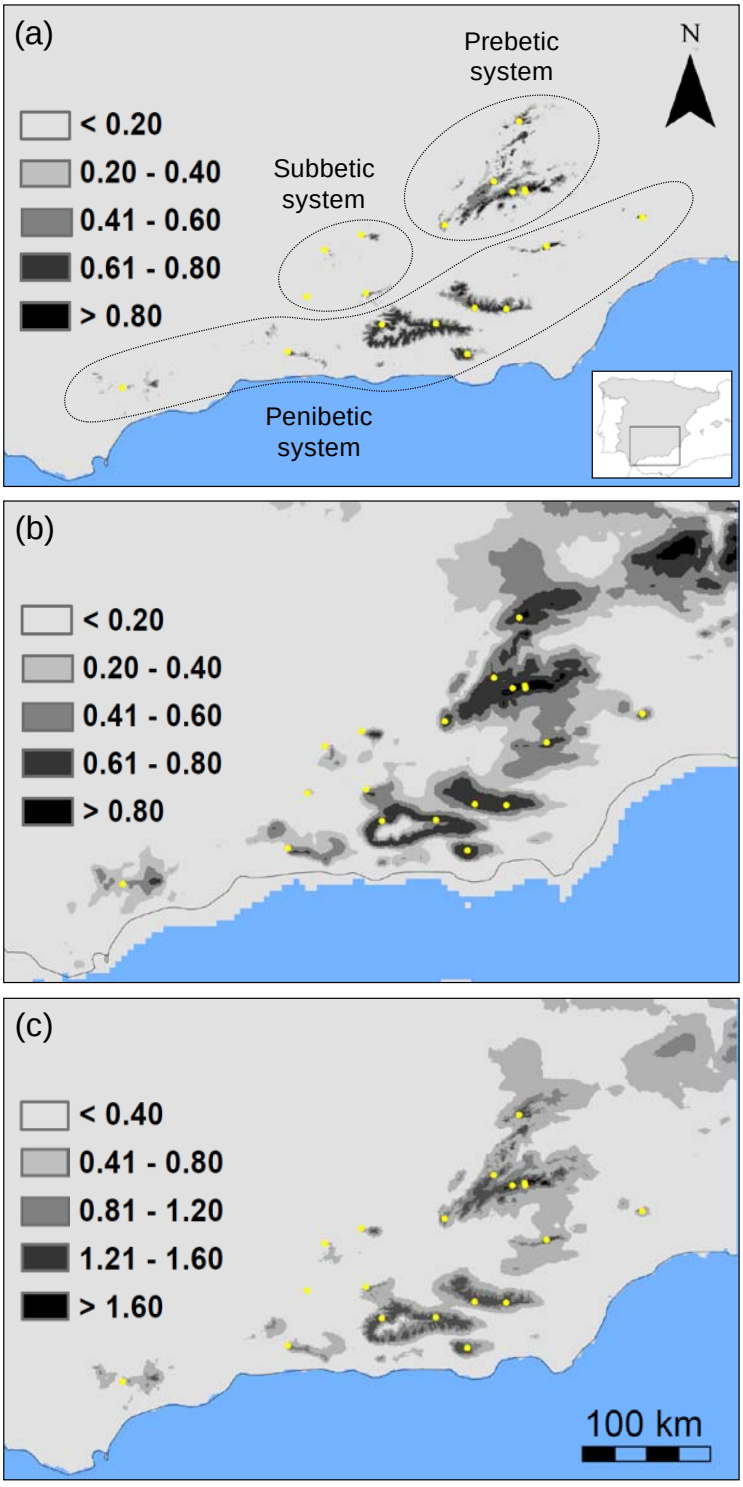


Figure 3 – Climate niche modelling for the scrub-legume grasshopper in southeast Iberia for (a) the present and (b) the Last Glacial Maximum (LGM, *c*. 21 Kya). Panel (c) shows climate stability estimated as the sum of pixel values of current and LGM climate suitability maps. The LGM maps represent the average climate suitability index of the projections obtained from CCSM and MIROC climate models. Grey scales refer to climate suitability (range: 0-1) and climate stability (range: 0-2), with increasingly darker shades of grey indicating increasing climate suitability and stability. Inset map from panel a) shows the location of our study area within the Iberian Peninsula.

This allowed us to sample populations from different habitats such as alpine and Mediterranean scrub-legume formations. Specimens were collected using a butterfly net and the whole body was preserved in 2 mL vials with 96% ethanol and stored at -20º C until needed for DNA extraction. Our sampling was performed under licenses from the 'Junta de Comunidades de Castilla-La Mancha', 'Junta de Andalucía' and 'Gobierno de la Región de Murcia'. Population codes and more information on sampling sites are presented in Table S1.

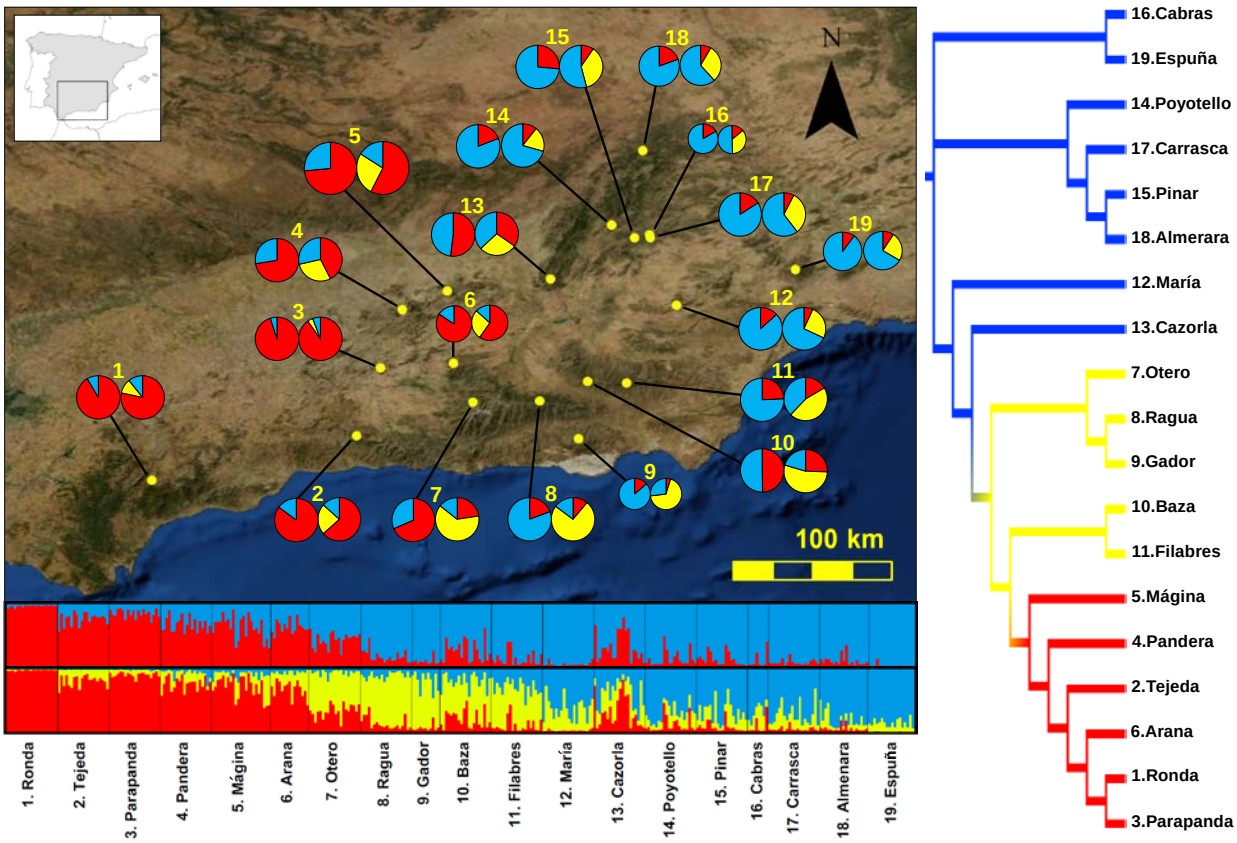


Figure 4 – Sampling sites of scrub-legume grasshoppers and genetic structure based on Bayesian clustering analyses. Pie charts on the map represent the genetic assignments for each sampling population according to STRUCTURE analyses. For each population, left and right pie charts represent the admixture proportions considering *K* = 2 and *K* = 3, respectively. Circle size is proportional to the number of genotyped individuals in each population. Code numbers are described in Table S1. On the bottom, barplots represent the assignment of individuals to each genetic group according to TESS analyses considering *K* = 2 (top) and *K* = 3 (bottom). Each individual corresponds to a vertical bar, which is partitioned into *K*-coloured segments that represent the individual's probability of belonging to the cluster with that colour. Vertical black lines separate individuals from different populations. On the right, neighbour-joining tree based on Cavalli-Sforza and Edwards chord distances. Colours are according to STRUCTURE analyses based on *K* = 3. Inset map shows the location of our study area within the Iberian Peninsula.

MICROSATELLITE GENOTYPING AND BASIC GENETIC STATISTICS

We extracted genomic DNA from a hind-leg of each individual using a salt extraction protocol (Aljanabi & Martinez 1997). Each individual was genotyped at eighteen species-specific microsatellites markers (Basiita *et al*. 2016). All microsatellite markers were polymorphic in all populations and the most common alleles were shared across all populations. We performed PCRs and genotyping following the procedure described in Ortego *et al*. (2015a) and Basiita *et al*. (2016). We tested for deviations from Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and the presence of null alleles as described in Noguerales *et al*. (2016). Two loci (Cbin16 and Cbin36) were discarded from all downstream analyses because of HW disequilibrium in all populations and the presence of null alleles. We did not find evidence for linkage disequilibrium between any pair of loci in any sampling population after sequential Bonferroni corrections (Rice 1989).

ANALYSES OF GENETIC STRUCTURE

We estimated population genetic differentiation calculating F_{ST}-values between all pairs of sampling populations. Significance of genetic differentiation between all pairs of populations was tested with Fisher´s exact tests after 10 000 permutations using ARLEQUIN 3.5 (Excoffier & Lischer 2010). *P*-values were corrected using a sequential Bonferroni adjustment (Rice 1989). Due to the frequent presence of null alleles in Orthoptera (Keller *et al*. 2013a), we also calculated pairwise F_{ST}-values corrected for null alleles (F_{ST}NA) using the so-called ENA-method implemented in the program FREENA (Chapuis & Estoup 2007).

We inferred genetic structure using Bayesian clustering analyses in STRUCTURE 2.3.3 (Pritchard *et al*. 2000; Falush *et al*. 2003). We considered correlated allele frequencies and an admixture model without prior information on population origin. We performed 10 independent runs for each value of assumed number of genetic clusters (K = 1-12) with a burn-in period of 200 000 steps and a run length of 1 000 000 Markov chain Monte Carlo (MCMC) cycles. The number of genetic clusters (K) best fitting the data set was defined using log probabilities [Pr(X|K)] (Pritchard *et al*. 2000) and the ΔK method (Evanno *et al*. 2005). We used the Greedy algorithm in the program CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) to align replicated runs and average individual assignment probabilities for the most

likely *K* values. Finally, we used DISTRUCT 1.1 (Rosenberg 2004) to produce bar plots displaying probabilities of individual membership to each inferred genetic cluster.

We also examined the spatial genetic structure considering geographical coordinates of sampling sites as *a priori* information in the Bayesian clustering method implemented in TESS 2.3.1 (Chen *et al*. 2007; Durand *et al*. 2009). We used the conditional autoregressive (CAR) Gaussian model of admixture with a linear trend surface, updating the spatial interaction parameter (ψ), initially set to the default value 0.99. The variance term (initially set to 1) permitted to update during the course of runs. CAR model was chosen in order to avoid overestimation of the most likely *K* in presence of genetic clines (François & Durand 2010; Guillot 2009). We ran 20 independent replicates for each value of *K* = 2-12 using 50 000 sweeps of which 10 000 were used as burn-in period. The best-supported number of genetic clusters (*K*) was estimated using the deviance information criterion (DIC) values and stabilization of the Q-matrix of posterior probabilities (Chen *et al*. 2007; Gao *et al*. 2011). For each K_{MAX}-value considered, we conducted 180 additional replicate runs up to a total of 200 replicates. We used the 10 runs with the lowest DIC values to align and average individual assignment probabilities with CLUMPP before being represented using DISTRUCT as indicated above for STRUCTURE analyses.

Complementarily, we constructed a phylogenetic tree to visualize the genetic relationships between all populations. We used the program POPULATIONS 1.2.31 (Langella 1999) to obtain a neighbour-joining tree based on pairwise Cavalli-Sforza and Edwards (D_c) genetic distances (Cavalli-Sforza & Edwards 1967). Finally, we carried out analyses of molecular variance (AMOVAS) to examine the partitioning of the genetic variation among and within regions and populations as defined by five population grouping hypotheses. Populations were pooled according to their historical location in the three different paleogeographical Late Miocene scenarios (see Fig. 2 and section “Landscape genetic analyses”) and their current location in the main mountain ranges of the region (Prebetic, Penibetic and Subbetic systems; see Table S1 and Fig. 3a). Additionally, we tested the grouping scheme used for Approximate Bayesian Computation (ABC) analyses (see next section). AMOVAS were performed in ARLEQUIN 3.5 (Excoffier & Lischer 2010) and the significance of the variance components was tested using 10 000 permutations of the original data.

APPROXIMATE BAYESIAN COMPUTATION (ABC)

In order to infer the evolutionary and demographic history of the scrub-legume grasshopper in the region, we compared four plausible scenarios of population divergence using an Approximate Bayesian Computation (ABC) approach (Beaumont 2010). To simplify the analyses, we defined three main groups (groups A, B and C) of populations by pooling sampling sites according to their geographic location and the results from AMOVAS (Table S2) and Bayesian clustering analyses (STRUCTURE and TESS) (*e.g.* Tsuda *et al*. 2015). Note that although K = 2 was the most supported clustering solution for both STRUCTURE and TESS analyses, K = 3 revealed further hierarchical genetic substructure with geographic coherence (see section "Results" and Fig. S1). In group A, we included populations 1-6 (western populations); B, 7-12 (south-eastern populations) and group C, 13-19 (north-eastern populations) (see Fig. 4-5). The topology of each scenario was designed considering the connectivity of populations according to Bayesian clustering analyses (Fig. 4; Fig. S1c). The scenarios tested were: (i) *Scenario I*, *null model*: the three groups diverged simultaneously; (ii) *Scenario II*, *sequential splitting model from west to east*: Group A split from Group B and C at t_2, and these two groups subsequently split at t_1; (iii) *Scenario III*, *sequential splitting model from east to west*: Group C split from Group A and B at t_2, and these two groups split at t_1; (iv) *Scenario IV*, *splitting model from central to peripheral populations*: Group B split from Group A and C at t_2, and these two groups subsequently split at t_1 (Fig. 5).

We conducted all the computations using DIYABC 2.0.4 (Cornuet *et al*. 2014). We generated three millions of simulated datasets per scenario considering a generalized mutation model (GSM) and no single nucleotide indels (Table S3). The summary statistics (SS) used in ABC analyses are

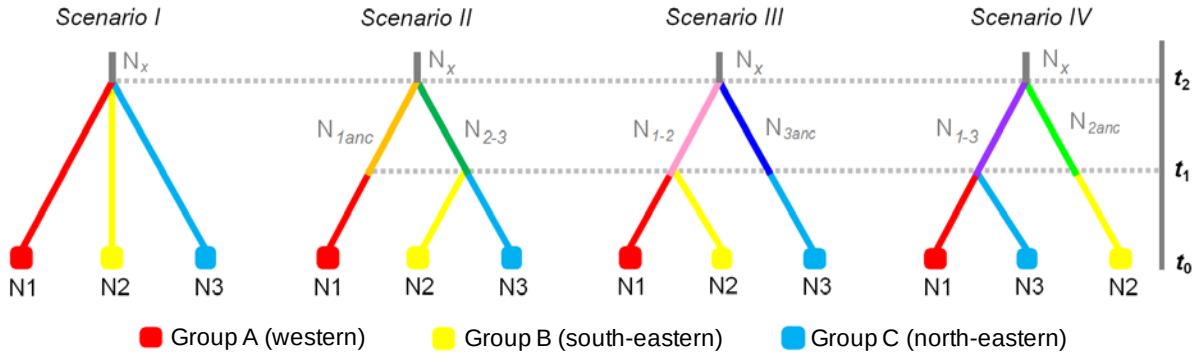


Figure 5 – Scenarios compared using an Approximate Bayesian Computation (ABC) approach ($t_{\#}$ represents time in number of generations; $N_{\#}$ represents effective population sizes during each time period)

described in Table S3. We performed pre-evaluation of scenarios and prior distributions in DIYABC to adjust the priors of N_e and t to their most appropriate values (see Table S3), assuming a uniform prior probability distribution for them. To avoid biases in parameter estimates, we selected the subset of seven microsatellites markers with lower frequency of null alleles as estimated in the program FREENA. Selection of the most probable scenario, confidence in scenario choice (type I and II errors), model checking and estimation of the posterior distribution of all parameters under the best supported model were performed as described in Ortego *et al*. (2015c).

LANDSCAPE GENETIC ANALYSES

We applied circuit theory (McRae 2006; McRae & Beier 2007) and a multiple matrix regression with randomization (MMRR) approach (Wang 2013) to examine the relative contribution of a suite of isolation-by-resistance (IBR) scenarios to explain patterns of genetic differentiation in our study populations. Specifically, we tested nine different hypothetical scenarios of population connectivity, which included (i) three paleogeographic scenarios defined by the spatial configuration of emerged lands at different time periods (Early Tortonian, Late Tortonian, and Earliest Messinian); (ii) three scenarios based on the distribution of climatically suitable habitats since the Last Glacial Maximum (LGM) (current climate suitability, LGM climate suitability, and climate suitability stability since the LGM); (iii) a scenario of population connectivity defined by contemporary topographic complexity; (iv) an isolation-by-distance (IBD) scenario representing the geographical distance between each pair of populations. Below we describe in detail the methods followed to generate these scenarios and test their relative contribution to contemporary patterns of genetic differentiation.

Paleogeographic scenarios. To test the possible effect of the complex geological history of the study region on contemporary patterns of genetic differentiation, we considered three paleogeographic scenarios: Early Tortonian (*c*. 12.0-11.6 Mya), Late Tortonian (*c*. 8.0-7.3 Mya) and Earliest Messinian (*c*. 7.2-7.0 Mya) (see Fig. 2). We used ARCGIS 10.0 (ESRI, Redlands, CA, USA) to create vector layers for emerged lands based on geological models from Martín *et al*. (2009). Then, we transformed vectors layers for each scenario into raster maps with a 30 arc-sec (*c*. 1 km) resolution that were finally used as

inputs in CIRCUITSCAPE (McRae 2006; McRae & Beier 2007) (see below for details on CIRCUITSCAPE analyses).

Climatic suitability scenarios. We modelled the potential climate distribution of scrub-legume grasshopper at different time periods to investigate whether the spatial distribution of climatically suitability habitats are relevant factors shaping observed patterns of genetic differentiation in the study populations. For this purpose, we built a climate niche model (CNM) using the maximum entropy presence-only algorithm implemented in MAXENT 3.3.3 (Phillips *et al*. 2006; Phillips & Dudik 2008) based on current climate. We used a total of 85 occurrence points obtained from the Global Biodiversity Information Facility (GBIF), literature (Defaut 2011) and our own sampling. To construct the models, we used the 19 bioclimatic variables available in WorldClim and downloaded at 30 arc-sec (*c*. 1 km) resolution (Hijmans *et al*. 2005). Variables retained in the final models were selected following several complementary criteria (Vega *et al*. 2010). At first, we used ENMTOOLS (Warren *et al*. 2010) to examine colinearity among variables, in order to retain a single layer among those with a high Pearson correlation coefficient ($r > 0.85$). Then, we used the Jackknife of regularized training gain procedure implemented in MAXENT to retain the variables with the maximum contribution to the model. We discarded the worst and highest correlated predictors among the whole set of variables, conducted a new model with the remaining variables and repeated this backward process until the final model only retained the best explanatory and less correlated variables (Vega *et al*. 2010). Model evaluation statistics were produced from 10 cross-validation replicate model runs.

To obtain the distribution of scrub-legume grasshopper during the Last Glacial Maximum (LGM, *c*. 21 Kya), we projected contemporary species-climate relationships to the LGM using two atmospheric circulation models: the Community Climate System Model (CCSM3; Collins *et al*. 2006) and the Model for Interdisciplinary Research on Climate (MIROC 3.2; Hasumi & Emori 2004) from the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2; Braconnot *et al*. 2007). LGM layers were downloaded from WorldClim at 2.5 arc-min and interpolated to 30 arc-sec resolution. To reduce the level of uncertainty arising from different past projections, we averaged climate suitability scores from projections based on CCSM and MIROC models to obtain a consensus LGM map of climatically suitable areas. In addition, we summed current and LGM climate suitability layers to generate a map of climate suitability stability, with pixel values ranging from 0 (minimum climate suitability in both periods)

to 2 (maximum climate suitability in both periods). All GIS calculations were conducted in ARCGIS 10.0. Finally, current, LGM and stability climate suitability raster maps were used as inputs in CIRCUITSCAPE to calculate IBR distance matrices (see below for details).

Contemporary topographic complexity scenario. We investigated the role of contemporary topographic complexity (TC) as a potential factor shaping patterns of genetic differentiation in our study populations. We calculated the surface ratio index for each cell from a present-day digital elevation model using DEM SURFACE TOOLS (Jenness 2013) in ARCGIS 10.0. Surface ratio is an index of topographic complexity, with values close to one indicating flat areas and values higher than one indicating a more abrupt relief with deeper slopes (Jenness 2004). Calculations were conducted on a 90 m resolution digital elevation model from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data; *http://srtm.csi.cgiar.org/*). Although no information is available on the dispersal distance and home range of the study species, the high resolution of the digital elevation model is expected to capture well the topographic complexity relevant for a medium-size grasshopper with a suspected low dispersal ability. The final raster map was transformed to 30 arc-sec (*c*. 1 km) resolution and used as input in CIRCUITSCAPE (see below for details).

CIRCUITSCAPE analyses. We used CIRCUITSCAPE 4.0 (McRae 2006; McRae & Beier 2007) to calculate resistance distance matrices between all pairs of populations considering an eight-neighbour cell connection scheme. The raster layers generated for the nine different scenarios of population connectivity were used as inputs in CIRCUITSCAPE. For the three paleogeographic scenarios, the raster layers included two element classes: '*emerged land*' and '*sea water*'. We considered that 'sea water' was the main landscape feature limiting the dispersal of terrestrial fauna in the study area during these periods. We generated different IBR scenarios assigning different resistance values to '*sea water*' (10, 50, 100, 500, 1000, 10 000), an approach that allowed us to identify the optimal ratio of landscape resistance between both landscape elements that best fit our data on genetic differentiation (*e.g.* Andrew *et al*. 2012; Ortego *et al*. 2015a). To test the effect of isolation-by-distance (IBD), we calculated pairwise resistance distances on a completely 'flat' landscape based on a raster layer in which all cells had an equal value (conductance = 1). This IBD resistance model is expected to yield

similar results than a matrix of Euclidean geographical distances but it is more appropriate for comparison with others competing models also generated with CIRCUITSCAPE (Velo-Antón *et al*. 2013).

Statistical analyses. We used a multiple matrix regression with randomization (MMRR) approach to examine the relative contribution of all isolation-by-resistance (IBR) and isolation-by-distance (IBD) scenarios to explain patterns of genetic differentiation in our study populations (Wang 2013). We tested the two matrices of genetic differentiation (F_{ST} and F_{ST}NA) against all pairwise resistance distance matrices representing the nine different IBR/IBD scenarios. We used a backward procedure to select final models, eliminating non-significant variables from an initial full model including all explanatory predictors. We tested the significance of the remaining variables again until no additional term reached significance (Ortego *et al*. 2015b; Noguerales *et al*. 2016).

ANALYSES OF GENETIC DIVERSITY AND ADMIXTURE

Allelic richness (A_R) standardized for sample size was calculated for each population using HP-RARE (Kalinowski 2005). We estimated the genetic admixture of populations using a genetic admixture index (G_{ADMIX}) obtained from the probabilities of population membership to each genetic cluster inferred by STRUCTURE analyses (Ortego *et al*. 2015b). This index was designed to standardize the degree of genetic admixture across populations with different probabilities of membership to different genetic clusters (Ortego *et al*. 2015b) and its advantages and potential caveats are those inherent to STRUCTURE analyses (Pritchard *et al*. 2000; Falush *et al*. 2003). G_{ADMIX} ranges from 0 (indicating no admixture, *i.e.* genetically pure populations assigned to a single genetic cluster) to 1 (indicating maximum admixture, *i.e.* genetically admixed populations with an equal probability of membership to each inferred genetic cluster). We used generalized linear models (GLMs, using a Gaussian error distribution and an identity link function) and an information-theoretic model selection approach to analyze A_R and G_{ADMIX} (Burnham & Anderson 2002). Models for A_R included as independent variables current climate suitability (HS_{CUR}), LGM climate suitability (HS_{LGM}), and climate suitability stability (HS_{STA}). Models for G_{ADMIX} included HS_{STA} as independent variable. Longitude and latitude were included as additional covariates in models for both A_R and G_{ADMIX} to take in account possible geographical clines of genetic diversity and admixture (*e.g*. Guo 2012). We calculated average HS_{CUR},

HS_{LGM} and HS_{STA} with ARCMAP 10.0 at different spatial scales using buffers of 1 km^2, 10 km^2 and 100 km^2 around sampling locations. Given that the precision of A_R and G_{ADMIX} estimates may differ among populations due to differences in sample sizes, we used a weighted least square (WLS) method where weight equals the sample size for each studied population. GLMs were built in the R package LME4 (Bates *et al*. 2015; R Core Team 2015) and model selection and averaging were performed using the R package MUMIN (Barton 2015) as detailed in Noguerales *et al*. (2015) and Ortego *et al*. (2015a).

RESULTS

POPULATION GENETIC STRUCTURE

We found that most pairs of populations were genetically differentiated. In particular, 139 of 171 pairwise F_{ST}-values (~81%) were significantly higher than zero after sequential Bonferroni correction (Table S4). Significant pairwise F_{ST}-values ranged from 0.025 to 0.164 whereas pairwise F_{ST}NA-values were slightly lower and ranged from 0.008 to 0.144. Pairwise F_{ST} and F_{ST}NA-values were highly correlated (Mantel r = 0.993; P < 0.001). The populations from Ronda, Parapanda and Tejeda, located at the westernmost portion of the study area, exhibited the highest levels of genetic differentiation with the rest of populations. Analyses in STRUCTURE showed a best-supported number of clusters for K = 2 according to the ΔK method. The first cluster included the western populations whereas the second cluster included the remaining populations located in the east part of the study area. However, log probabilities [Ln Pr (X|K)] steadily increased from K = 2 to K = 5 (Fig. S1a). Individual assignment probabilities to a certain genetic cluster were moderately high up to K = 5 and the spatial distribution of genetic variation exhibited geographical consistency, but most populations showed a considerable degree of genetic admixture (Fig. 4; Fig. S1c). Genetic clustering analyses in TESS resulted in an optimal K = 4 according to the DIC criterion (Fig. S1b), but one of the inferred clusters represented a 'ghost cluster' with no individual assigned to it (see Chen *et al*. 2007; Guillot *et al*. 2005). When K = 3 was considered, the first cluster included the western populations, the second cluster included the south-eastern populations and the third cluster included the north-eastern populations of the study area (Fig. 4). TESS and STRUCTURE analyses yielded similar results for K = 2 and K = 3 (Fig. S1c). The result of the neighbour-joining tree based on Cavalli-Sforza and Edwards chord distances (D_c) was also congruent with the results from Bayesian clustering analyses (Fig. 4). Finally, AMOVA analyses indicated most genetic variance was attributed to differences within populations (> 90 %, for all grouping

hypotheses; Table S2). The population grouping hypothesis that explained the highest percentage of total variation attributed to differences among groups was the one used for ABC analyses (Table S2).

APPROXIMATE BAYESIAN COMPUTATION (ABC)

The scenario considering sequential population divergence from west to east (scenario II) had the highest posterior probability based on both direct and logistic regression-based estimates, and its 95% confidence interval did not overlap with those obtained for others scenarios that showed much lower support (Table 1). Observed data fell within simulated data (all summary statistics *P*s > 0.2) for scenario II, suggesting good model fit. Type I and II errors were 0.394 and 0.372 respectively and RMAE values were moderate in most cases (Table 1).

Considering the one-year generation time of scrub-legume grasshopper (Defaut 2011), the western genetic group (group A) diverged from the south-eastern and north-eastern groups (Groups B and C, respectively) ~215 000 years ago (t_2) (95% CI: 67 600-342 000 years ago), whereas these two groups split ~42 000 years ago (t_1) (95% CI: 5 540-165 000 years ago) (Table 2). Assuming a constant mutation rate, the posterior estimates of effective populations sizes (N_e) indicated no important demographic changes after the different splitting events (Table 2).

CLIMATE NICHE MODELLING

The variables included in the final climate niche model were temperature seasonality (*Bio4*), mean temperature of the driest quarter (*Bio9*), precipitation of the driest month (*Bio14*) and precipitation of the warmest quarter (*Bio18*). This model had a very high value of Area Under the Curve (AUC; 0.982 ± 0.007 SD), indicating overall good performance. The predicted distribution of scrub-legume grasshopper in the present (Fig. 3a) is consistent with its observed fragmented distribution. The distribution of the species was more extensive and populations were better connected during the LGM than at present time (Fig. 3b). However, populations located in the western portion of our study area (corresponding with Group A in ABC analyses) have remained highly isolated during both the LGM and the present.

Table 1 – Posterior probability for each of the four tested scenarios and 95% confidence intervals (CI) based on the weigthed polychotomous logistic regression approach for Approximate Bayesian Computation (ABC) analyses. Type I and type II errors for the best supported scenario (in bold) are indicated.

Scenario	Posterior probability	95% CI	Type I error	Type II error
I	0.0230	[0.0222 - 0.0253]		
II	**0.9618**	**[0.9595 - 0.9640]**	0.394	0.372
III	0.0059	[0.0054 - 0.0064]		
IV	0.0085	[0.0078 - 0.0093]		

LANDSCAPE GENETIC ANALYSES

Tejeda and Ronda populations are currently located in areas that were not likely to form emerged lands during Early and Late Tortonian and for this reason they were excluded from landscape genetic analyses. Thus, we performed MMRR analyses using the 17 populations presumably located on permanently emerged lands since the late Miocene in order to make our landscape genetic analyses comparable across all tested scenarios and time periods. Considering these 17 populations, only resistance distances based on contemporary topographic complexity and isolation-by-distance (IBD) were significantly associated with genetic differentiation (*P*s < 0.006) (Table 3). However, only topographic complexity was retained into the final model (β = 0.826, t = 7.56, P = 0.004) (Fig. 6). Analyses based on F_{ST}NA gave similar results (Table 3), but models had slightly lower values of r^2. Analyses considering all populations (n = 19) yielded qualitatively analogous results (data not shown).

ANALYSES OF GENETIC DIVERSITY AND ADMIXTURE

$G_{ADMIX\ [K=i,\ 5]}$ indexes obtained considering different K values were highly correlated among them (all r > 0.598, all *P*s < 0.007). Population genetic admixture based on any $G_{ADMIX\ [K=i,\ 5]}$ index and A_R were also correlated (all r > 0.451, all *P*s < 0.050). Model selection results showed that A_R was not significantly associated with longitude, latitude or HS_{CUR}, HS_{LGM} or HS_{STA} at any analysed spatial scale (all unconditional 95% CIs of the predictors crossed zero; Table S5). Likewise, population genetic admixture (based on any $G_{ADMIX\ [K=i,\ 5]}$ index) was not significantly associated with longitude, latitude or HS_{STA} at any analysed spatial scale (all unconditional 95% CIs of the predictors crossed zero; Table S6).

Table 2 – Posterior parameter estimates (median and 95% confidence intervals) for the best supported scenario (scenario II, see Fig. 5). Estimates are based on 1% of simulated datasets closest to the observed values. Relative median absolute errors (RMAE) based on 500 pseudo-observed data sets are also indicated for each parameter.

Parameter	Median	$q_{0.025}$	$q_{0.975}$	RMAE
N1	520 000	194 000	736 000	0.279
N2	407 000	113 000	715 000	0.246
N3	600 000	277 000	740 000	0.258
N_{1anc}	303 000	39 800	690 000	0.394
$N_{2\text{-}3}$	345 000	46 600	695 000	0.372
N_x	57 000	24 800	541 000	0.384
t_1	42 700	5 540	165 000	0.404
t_2	215 000	67 600	342 000	0.213
μ	7.75×10^{-6}	4.24×10^{-6}	$2.85\ 10^{-5}$	0.357

N1, effective population size of group A
N2, effective population size of group B
N3, effective population size of group C
N_{1anc}, effective population size of the ancestral group A
$N_{2\text{-}3}$, effective population size of the ancestral group B-C
N_x, effective population size of the most ancestral population
t_1, time (in generations = years) to the most recent divergence event
t_2, time (in generations = years) to the most ancient divergence event (see scenarios in Fig. 5)
μ , mean mutation rate

Table 3 – Results of univariate matrix regressions with randomization (MMRR) for genetic differentiation [F_{ST} and F_{ST} corrected for null alleles (F_{ST}NA)] in relation with different isolation-by-resistance (IBR) scenarios: geographical distance (IBD), contemporary topographic complexity (TC), current climate suitability (HS_{CUR}), Last Glacial Maximum climate suitability (HS_{LGM}), climate suitability stability (HS_{STA}), and three paleogeographical models (Early Tortonian, *c*. 12.0-11.6 Mya; Late Tortonian, *c*. 8.0-7.3 Mya; Earliest Messinian, *c*. 7.2-7.0 Mya). For paleogeographical models, we considered high resistance values for sea water (= 100) and low for emerged lands (= 1). Table shows the results based on the 17 populations presumably located on permanently emerged lands since the Late Miocene.

	F_{ST}				F_{ST} NA			
Model	r^2	β	*t*	*P*	r^2	β	*t*	*P*
IBD	0.298	0.824	7.543	0.004	0.273	0.811	7.097	0.011
TC	0.299	0.826	7.561	0.006	0.274	0.814	7.118	0.021
HS_{CUR}	0.031	0.163	2.084	0.264	0.055	0.224	2.815	0.086
HS_{LGM}	0.220	0.446	6.16	0.051	0.194	0.429	5.662	0.064
HS_{STA}	0.210	0.432	5.978	0.053	0.185	0.417	5.521	0.062
Early Tortonian	0.115	0.320	4.180	0.063	0.096	0.301	3.780	0.055
Late Tortonian	0.101	0.296	3.886	0.066	0.084	0.278	3.522	0.052
Earliest Messinian	0.088	0.278	3.612	0.065	0.072	0.258	3.237	0.062

DISCUSSION

Assuming ecological niche stability through time, our niche model revealed a moderate shift in the distribution of climatically suitable habitats for the scrub-legume grasshopper in southeast Iberia during the last 21 000 years (Nogués-Bravo 2009). Climate niche modelling indicated that the potential distribution of the species is more fragmented in the present than during the LGM, a pattern congruent with the increased population connectivity during glacial periods inferred for many other montane species from temperate regions (Blanco-Pastor *et al*. 2013; Velo-Antón *et al*. 2013). Continuous climatically suitable habitats connected the southern foothills of the Prebetic mountain range and the eastern portion of the Penibetic system during the LGM, an area where the species is not present today as verified by our own surveys. Our climate niche model also outlined the isolation of the western populations since LGM, which has probably contributed to shape their strong genetic differentiation with the rest of the populations within the study area. The isolation of the western populations could have occurred during the Last Interglacial (LIG) period (*c*. 120-140 Kya), when the scrub-legume grasshopper probably showed a distribution similar to that in the present.

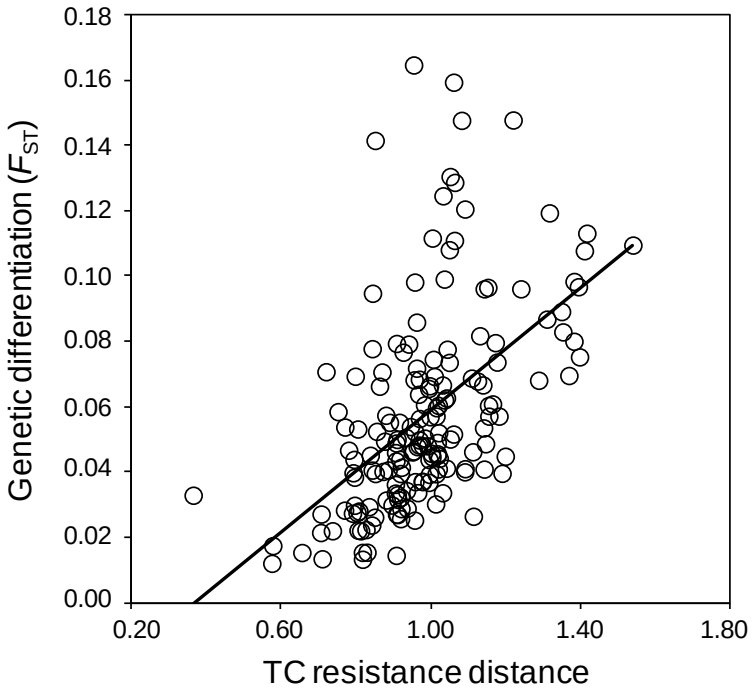


Figure 6 – Relationship between genetic differentiation (F_{ST}) and resistance distances calculated using CIRCUITSCAPE on the basis of contemporary topographic complexity (TC).

We did not find support for the hypothesis predicting more genetic diversity in populations from areas with higher past and current climate suitability and stability since the LGM. Unlike other organisms from temperate regions whose distributions were confined to isolated refugia during glacial or interglacial periods (Homburg *et al*. 2013), the climate niche model for the scrub-legume grasshopper indicate a considerable stability of climate suitability since LGM in most of the study area. Thus, the lack of major range shifts and geographically restricted climate refugia may explain why genetic diversity and admixture are decoupled from local stability of climatically suitable habitats (Petit *et al*. 2003; Yannic *et al*. 2014; Ortego *et al*. 2015b).

Populations of scrub-legume grasshopper showed higher levels of genetic differentiation than those observed at similar spatial scales in generalist grasshoppers (Wiesner *et al*. 2011; Keller *et al*. 2013a,b; Blanchet *et al*. 2012a,b; Ortego *et al*. 2015a), but lower than in populations of specialist Orthoptera inhabiting highly fragmented habitats (Streiff *et al*. 2006; Ortego *et al*. 2012). The highest pairwise F_{ST} values were found in comparisons involving the western and easternmost populations of the Penibetic system, which are located at the extremes of the species distribution range in the study area and that seem to have remained poorly connected according to our climate niche model (Fig. 3). Despite the relatively small size of the study area (the most distant populations are separated by less than 400 km), the neighbour-joining tree and Bayesian clustering analyses revealed a moderate degree of spatial genetic structure. The three genetic clusters found in our study populations spanned different mountain systems: western Penibetic and Subbetic (western cluster), central Penibetic (south-eastern cluster), and easternmost Penibetic and Prebetic ranges (north-eastern cluster). Similar geographical patterns of genetic structure have been found in other co-distributed organisms from the region, which has been interpreted as a result of long-term isolation during the Pleistocene of populations inhabiting different mountain ranges (Albert *et al*. 2007; Dias *et al*. 2015). Our Bayesian clustering analyses also showed a relatively high degree of genetic admixture that was more evident in populations located at the core of our study area (southern Prebetic and eastern Penibetic systems), which may reflect their higher connectivity with the rest of the populations. ABC analyses supported a sequential splitting model in which population divergence took place from the western to the eastern portion of the study area (scenario II). The western populations diverged from the remaining populations ~215 Kya, while the south-eastern and north-eastern populations split ~42.7 Kya. This result suggests that the most ancient divergence event likely took place during the interglacial period between the Riss and Mindel glaciations, whereas the most recent divergence event could have occurred during a short warm period between the LGM (*c*. 21 Kya) and the LIG (*c*. 120-140 Kya) (Siddall *et al*. 2003). Both warm periods

were characterized by a generalized expansion of deciduous and Mediterranean forests and the contraction of the vegetation adapted to dry and cool climates (Sánchez Goñi *et al*. 2005). The distribution of shrub-like vegetation probably retracted to higher elevations during interglacial periods, which may have promoted the isolation and progressive differentiation of the populations of scrub-legume grasshopper. However, we must note that our estimates of divergence times should be interpreted with extreme caution given that the confidence intervals for both t_2 (95% CI: 67 600-342 000 years ago) and t_1 (95% CI: 5 540-165 000 years ago) were very broad. In addition, it must be also considered that DIYABC does not accommodate gene flow after divergence (Cornuet *et al*. 2014), which is expected to underestimate divergence times (Tsuda *et al*. 2015; Ortego *et al*. 2015c).

Our landscape genetic analyses showed that resistance distances based on contemporary topographic complexity provided the best model fit, indicating that gene flow is primarily shaped by physical features of the landscape (Wang 2012). This suggests that the contemporary topographic complexity of southeast Iberia is playing a more important role in determining dispersal of scrub-legume grasshoppers than the distribution of climatically suitable habitat patches. A similar result was found for the Morales grasshopper (*Chorthippus saulcyi moralesi*), an endemic taxon from the Pyrenees belonging to the same species complex than the scrub legume grasshopper (Noguerales *et al*. 2016). This is in line with several studies on other taxa finding that topography is one of the main predictors of genetic differentiation in terrestrial organisms, a fact that has been linked to the greater energetic expenditure associated with dispersal across abrupt reliefs (Castillo *et al*. 2014; Velo-Antón *et al*. 2013; Benham & Witt 2016; Hemp *et al*. 2016). We did not find a significant relationship between genetic differentiation and resistance distances based on any climate suitability or paleogeographic scenario. The lack of association between genetic differentiation and resistance distances based on climate suitability could be explained by the fact that the observed patterns of genetic differentiation were shaped by environmental factors predating the LGM period as suggested by our ABC analyses. An alternative explanation for such lack of association could be that our climate niche model is not capturing well the microhabitat structure that defines the spatial configuration of corridors/barriers to dispersal in our study species (Noguerales *et al*. 2016). Different factors could be behind the lack of association between the spatial genetic structure and ancient geological changes: (i) the genetic signature eventually left during the Tortonian as consequence of isolation processes could have been progressively eroded over time due to post-Messinian population connectivity and gene flow; (ii) the Betic islands could have remained non-colonized by the scrub-legume grasshopper until the Earliest Messinian (*c*. 7 Mya) due to the large distance to the mainland and/or the absence of optimal habitats

given the warmer climate and the forest-like vegetation prevailing during the Late/Middle Miocene (*c*. 23-14 Mya) (Jiménez-Moreno *et al*. 2010). Finally, (iii) the high mutation rates of microsatellite markers that make them adequate to track recent and ongoing processes of genetic differentiation are expected to reduce their power to capture signatures of old demographic events such as those driven by ancient geological changes (Zellmer & Knowles 2009; Anderson *et al*. 2010; Wang 2010; Velo-Antón *et al*. 2013).

CONCLUSIONS

Overall, this study shows that contemporary topographic complexity is the main landscape factor predicting spatial patterns of population genetic differentiation in the scrub-legume grasshopper. Future studies analysing the complete distribution range of the species and considering current and past distributional data for all host plant taxa can provide new insights into the consequences of ancient and contemporary landscape changes on the evolutionary and demographic history of this specialist grasshopper. Our study emphasizes the need of integrating spatial data of ancient and contemporary landscape composition to get a comprehensive view of their relative importance in explaining genetic variation of organisms inhabiting regions with a complex biogeographical history.

Acknowledgements

We wish to thank Conchi Cáliz for her advice during lab genetic work and Bernard Defaut for providing valuable information about sampling locations. Three anonymous referees provided valuable comments on an earlier draft of this manuscript. VN was supported by a FPI pre-doctoral fellowship (BES-2012-053741) from Ministerio de Economía y Competitividad. JO was supported by Severo Ochoa (SEV-2012-0262) and Ramón y Cajal (RYC-2013-12501) research fellowships. This work received financial support from research grants CGL2011-25053 and CGL2014-54671-P (Ministerio de Economía y Competitividad and European Social Fund), POII10-0197-0167 and PEII-2014023-P (Junta de Comunidades de Castilla-La Mancha and European Social Fund) and UNCM08-1E-018 (European Regional Development Fund).

REFERENCES

Abellán P, Arribas P, Svenning JC (2012) Geological habitat template overrides late Quaternary climate change as a determinant of range dynamics and phylogeography in some habitat-specialist water beetles. *Journal of Biogeography*, 39, 970-983.

Albert EM, Zardoya R, García-París M (2007) Phylogeographical and speciation patterns in subterranean worm lizards of the genus *Blanus* (Amphisbaenia: Blanidae). *Molecular Ecology*, 16, 1519-1531.

Aljanabi SM, Martínez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692-4693.

Anderson CD, Epperson BK, Fortin MJ, *et al*. (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, 19, 3565-3575.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, 21, 2078-2091.

Andújar C, Gómez-Zurita J, Rasplus JY, Serrano J (2012) Molecular systematics and evolution of the subgenus *Mesocarabus* Thomson, 1875 (Coleoptera: Carabidae: Carabus), based on mitochondrial and nuclear DNA. *Zoological Journal of the Linnean Society*, 166, 787-804.

Barton K (2015) *MuMIn: multimodel inference*. R Package, version 1.15.6. *https://cran.r-project.org/web/packages/MuMIn/index.html*

Basiita RK, Bruggemann JH, Cai N, *et al*. (2016) Erratum to: Microsatellite records for volume 7, issue 4. *Conservation Genetic Resources*, 8, 85-87.

Bates D, Maechler M, Bolker BM, Walker SC (2015) Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67, 1-48.

Beaumont MA (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41, 379-406.

Benham PM, Witt CC (2016) The dual role of Andean topography in primary divergence: functional and neutral variation among populations of the hummingbird, *Metallura tyrianthina*. *BMC Evolutionary Biology*, 16, 22.

Blanchet E, Lecoq M, Sword GA, *et al*. (2012) A comparative analysis of fine-scale genetic structure in three closely related syntopic species of the grasshopper genus *Calliptamus*. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 90, 31-41.

Blanchet E, Lecoq M, Sword GA, *et al*. (2012) Population structures of three *Calliptamus* spp. (Orthoptera: Acrididae) across the Western Mediterranean Basin. *European Journal of Entomology*, 109, 445-455.

Blanco-Pastor JL, Fernández-Mazuecos M, Vargas P (2013) Past and future demographic dynamics of alpine species: limited genetic consequences despite dramatic range contraction in a plant from the Spanish Sierra Nevada. *Molecular Ecology*, 22, 4177-4195.

Braconnot P, Otto-Bliesner B, Harrison S, *et al*. (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, 261-277.

Braga JC, Martín JM, Aguirre J, *et al*. (2010) Middle-Miocene (Serravallian) temperate carbonates in a seaway connecting the Atlantic Ocean and the Mediterranean Sea (North Betic Strait, S Spain). *Sedimentary Geology*, 225, 19-33.

Braga JC, Martín JM, Quesada C (2003) Patterns and average rates of late Neogene-Recent uplift of the Betic Cordillera, SE Spain. *Geomorphology*, 50, 3-26.

Brown JL, Knowles LL (2012) Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology*, 21, 3757-3775.

Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, USA.

Canestrelli D, Nascetti G (2008) Phylogeography of the pool frog *Rana* (*Pelophylax*) lessonae in the Italian peninsula and Sicily: multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *Journal of Biogeography*, 35, 1923-1936.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785-789.

Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology*, 23, 843-856.

Cavalli-Sforza L, Edwards AWF (1967) Phylogenetic analyses: models and estimation procedures. *Evolution*, 21, 550-570.

Ceccarelli FS, Ojanguren-Affilastro AA, Ramírez MJ, Ochoa JA, Mattoni CI, Prendini L (2016) Andean uplift drives diversification of the bothriurid scorpion genus *Brachistostenus*. *Journal of Biogeography*, 43, 1942-1954.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24, 621-631

Chen C, Durand E, Forbes F, Francois O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, 7, 747-756.

Cheng R, Jiang N, Yang X, Xue D, Liu S, Han H (2016) The influence of geological movements on the population differentiation of *Biston panterinaria* (Lepidoptera: Geometridae). *Journal of Biogeography*, 43, 691-702.

Collins WD, Bitz CM, Blackmon ML, *et al*. (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, 19, 2122-2143.

Cornuet JM, Pudlo P, Veyssier J, *et al*. (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187-1189.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (Caelifera, Acrididae, Gomphocerinae). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Dias G, Beltrán JF, Tejedo M, *et al*. (2015) Limited gene flow and high genetic diversity in the threatened Betic midwife toad (*Alytes dickhilleni*): evolutionary and conservation implications. *Conservation Genetics*, 16, 459-476.

Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, 26, 1963-1973.

Edwards DL, Keogh JS, Knowles LL (2012) Effects of vicariant barriers, habitat stability, population isolation and environmental features on species divergence in the south-western Australian coastal reptile community. *Molecular Ecology*, 21, 3809-3822.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567.

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, 1979-1990.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.

Fordham DA, Brook BW, Moritz C, Nogués-Bravo D (2014) Better forecast of range dynamics using genetic data. *Trends in Ecology and Evolution*, 29, 436-443.

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, 10, 773-784.

Fromhage L, Vences M, Veith M (2004) Testing alternative vicariance scenarios in Western Mediterranean discoglossid frogs. *Molecular Phylogenetics and Evolution*, 31, 308-322.

Gao H, Bryc K, Bustamante CD (2011) On identifying the optimal number of population clusters via the deviance information criterion. *Plos One*, 6, e21014.

Graham MR, Hendrixson BE, Hamilton CA, Bond JE (2015) Miocene extensional tectonics explain ancient patterns of diversification among turret-building tarantulas (*Aphonopelma mojave* group) in the Mojave and Sonoran deserts. *Journal of Biogeography*, 42, 1052-1065.

Guillot G (2009) On the inference of spatial structure from population genetics data. *Bioinformatics*, 25, 1796-1801.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, *170*, *1261-1280.*

Guo Q (2012) Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology*, 21, 5396-5403.

Habel JC, Zachos FE, Dapporto L, *et al*. (2015) Population genetics revisited - towards a multidisciplinary research field. *Biological Journal of the Linnean Society*, 115, 1-12.

Hasumi H, Emori S (2004) *K-1 coupled GCM (MIROC) description*. K-1 Technical Report Nº1. Center for Climate System Research, University of Tokyo, Tokyo, Japan.

He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, 67, 3386-3402.

Hemp C, Grzywacz B, Warchalowska-Sliwa E, Hemp A (2016) Topography and climatic fluctuations boosting speciation: biogeography and a molecular phylogeny of the East African genera *Afroanthracites* Hemp & Ingrisch and *Afroagraecia* Ingrisch & Hemp (Orthoptera, Tettigoniidae, Conocephalinae, Agraeciini). *Organisms Diversity and Evolution*, 16, 211-223.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Homburg K, Drees C, Gossner MM, Rakosy L, Vrezec A, Assmann T (2013) Multiple glacial refugia of the low-dispersal ground beetle *Carabus irregularis*: molecular data support predictions of species distribution models. *Plos One*, 8, e61185.

Hughes PD, Woodward JC (2008) Timing of glaciation in the Mediterranean mountains during the last cold stage. *Journal of Quaternary Science*, 23, 575-588.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801-1806.

Jenness JS (2004) Calculating landscape surface area from digital elevation models. *Wildlife Society Bulletin*, 32, 829-839.

Jenness J (2013) DEM SURFACE TOOLS. Jenness Enterprises. *http://www.jennessent.com/arcgis/surface_area.htm*

Jiménez-Moreno G, Fauquette S, Suc JP (2010) Miocene to Pliocene vegetation reconstruction and climate estimates in the Iberian Peninsula from pollen data. *Review of Palaeobotany and Palynology*, 162, 403-415.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5, 187-189.

Keller D, Holderegger R, van Strien MJ (2013a) Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology*, 22, 2467-2482.

Keller D, van Strien MJ, Herrmann M, *et al*. (2013b) Is functional connectivity in common grasshopper species affected by fragmentation in an agricultural landscape? *Agriculture Ecosystems and Environment*, 175, 39-46.

Langella O (1999) POPULATIONS 1.2.31. *http://bioinformatics.org/~typhon/populations/*

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, 24, 3688-3705.

Martín JM, Braga JC, Aguirre J, Puga-Bernabeú A (2009) History and evolution of the North-Betic Strait (Prebetic Zone, Betic Cordillera): a narrow, early Tortonian, tidal-dominated, Atlantic-Mediterranean marine passage. *Sedimentary Geology*, 216, 80-90.

Mastretta-Yanes A, Moreno-Letelier A, Pinero D, Jorgensen TH, Emerson BC (2015) Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, 42, 1586-1600.

McRae BH (2006) Isolation by resistance. *Evolution*, 60, 1551-1561.

McRae BH, 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, 19885-19890.

Noguerales V, Cordero PJ, Ortego J (2016) Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, 16, 96.

Noguerales V, Traba J, Mata C, Morales M (2015) Winter habitat selection and partitioning in two sympatric farmland small mammals: *Apodemus sylvaticus* and *Mus spretus*. *Revue d'écologie (La Terre et la Vie)*, 70, 70-82.

Nogués-Bravo D (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, 18, 521-531.

Opell BD, Helweg SG, Kiser KM (2016) Phylogeography of Australian and New Zealand spray zone spiders (Anyphaenidae: *Amaurobioides*): Moa´s Ark loses a few more passengers. *Biological Journal of Linnean Society*, 118, 959-969.

Ortego J, Aguirre MP, Cordero PJ (2012) Landscape genetics of a specialized grasshopper inhabiting highly fragmented habitats: a role for spatial scale. *Diversity and Distributions*, 18, 481-492.

Ortego J, Aguirre M.P, Noguerales V, Cordero PJ (2015a) Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications*, 8, 621-632.

Ortego J, Bonal R, Cordero PJ, Aparicio JM (2009) Phylogeography of the Iberian populations of *Mioscirtus wagneri* (Orthoptera: Acrididae), a specialized grasshopper inhabiting highly fragmented hypersaline environments. *Biological Journal of the Linnean Society*, 97, 623-633.

Ortego J, Gugger PF, Sork VL (2015b) Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography*, 42, 328-338.

Ortego J, Noguerales V, Gugger PF, Sork VL (2015c) Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology*, 24, 6188-6208.

Papadopoulou A, Anastasiou I, Keskin B, Vogler AP (2009) Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. *Molecular Ecology*, 18, 2503-2517.

Pepper M, Doughty P, Arculus R, Keogh JS (2008) Landforms predict phylogenetic structure on one of the world's most ancient surfaces. *BMC Evolutionary Biology*, 8, 152.

Peterman WE, Connette GM, Semlitsch RD, Eggert LS (2014) Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, 23, 2402-2413.

Petit RJ, Aguinagalde I, de Beaulieu JL, *et al*. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, 300, 1563-1565.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

Phillips SJ, Dudik M (2008) Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Qu Y, Ericson PGP, Quan Q, *et al*. (2014) Long-term isolation and stability explain high genetic diversity in the Eastern Himalaya. *Molecular Ecology*, 23, 705-720.

R Core Team (2015) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org/*

Reilly SB, Corl A, Wake DB (2015) An integrative approach to phylogeography: investigating the effects of ancient seaways, climate, and historical geology on multi-locus phylogeographic boundaries of the Arboreal Salamander (*Aneides lugubris*). *BMC Evolutionary Biology*, 15, 241.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223-225.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137-138.

Sánchez Goñi MF, Loutre MF, Crucifx M, *et al*. (2005). Increasing vegetation and climate gradient in Western Europe over the Last Glacial Inception (122-110 ka): data-model comparison. *Earth and Planetary Science Letters*, 231, 111-130.

Siddall M, Rohling EJ, Almogi-Labin A, *et al*. (2003) Sea-level fluctuations during the last glacial cycle. *Nature*, 423, 853-858.

Streiff R, Audiot P, Foucart A, Lecoq M, Rasplus JY (2006) Genetic survey of two endangered grasshopper subspecies, *Prionotropis hystrix rhodanica* and *Prionotropis hystrix azami* (Orthoptera, Pamphagidae): within- and between-population dynamics at the regional scale. *Conservation Genetics*, 7, 331-344.

Tsuda Y, Nakao K, Ide Y, Tsumura Y (2015) The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. *Molecular Ecology*, 24, 1403-1418.

Vega R, Flojgaard C, Lira-Noriega A, Nakazawa Y, Svenning JC, Searle JB (2010) Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography*, 33, 260-271.

Velo-Antón G, Parra JL, Parra-Olea G, Zamudio KR (2013) Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, 22, 3261-3278.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, 19, 2605-2608.

Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, 18, 1033-1041.

Wang IJ (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, 3403-3411.

Warren DL, Glor RE, Turelli M (2010) ENMTOOLS: a toolbox for comparative studies of environmental niche models. *Ecography*, 33, 607-611.

Wiesner KR, Loxdale HD, Koehler G, Schneider ARR, Tiedemann R, Weisser, WW (2011) Patterns of local and regional genetic structuring in the meadow grasshopper, *Chorthippus parallelus* (Orthoptera: Acrididae), in Central Germany revealed using microsatellite markers. *Biological Journal of the Linnean Society*, 103, 875-890.

Yannic G, Pellissier L, Ortego J, *et al*. (2014) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, 4, 132-137.

Zellmer AJ, Knowles LL (2009) Disentangling the effects of historic *vs*. contemporary landscape structure on population genetic divergence. *Molecular Ecology*, 18, 3593-3602.

SUPPLEMENTARY MATERIAL

Table S1 – Geographical location, number of genotyped individuals (*N*) and population genetic variability (allelic richness standardized for sample size, A_R) for the studied populations of the scrub-legume grasshopper in southeast Iberia.

ID	Locality	Code	Mountain range	Latitude	Longitude	Elevation (m.a.s.l.)	*N*	A_R
1	Ronda	RON	Penibetic	36.663990	-5.229845	958	20	5.39
2	Tejeda	TEJ	Penibetic	36.918736	-4.066843	1595	20	5.65
3	Parapanda	PAR	Subbetic	37.305478	-3.929667	1578	20	4.55
4	Pandera	PAN	Subbetic	37.636927	-3.805924	1512	20	5.56
5	Mágina	MAG	Subbetic	37.744182	-3.549024	1325	23	5.61
6	Arana	ARA	Subbetic	37.331304	-3.516227	1736	15	5.16
7	Otero	OTE	Penibetic	37.110127	-3.405121	2314	20	5.45
8	Ragua	RAG	Penibetic	37.116477	-3.026024	2100	20	4.99
9	Gador	GAD	Penibetic	36.901460	-2.803908	2017	11	5.17
10	Baza	BAZ	Penibetic	37.227329	-2.752059	1876	20	5.50
11	Filabres	FIL	Penibetic	37.219673	-2.530702	2117	20	5.55
12	María	MAR	Penibetic	37.662071	-2.246718	1670	20	5.70
13	Cazorla	CAZ	Prebetic	37.810585	-2.962646	1772	20	5.64
14	Poyotello	POY	Prebetic	38.119515	-2.616541	1600	20	5.53
15	Pinar	PIN	Prebetic	38.046300	-2.485148	1662	20	5.99
16	Cabras	CAB	Prebetic	38.062612	-2.399850	2038	8	5.07
17	Carrasca	CAR	Prebetic	38.045469	-2.396126	1526	20	5.55
18	Almenara	ALM	Prebetic	38.543520	-2.440236	1655	19	5.78
19	Espuña	ESP	Penibetic	37.865166	-1.571250	1514	18	5.08

Table S2 – Results of AMOVAs used to test different population grouping hypotheses for scrub-legume grasshopper in southeast Iberia. NS = Not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Hypothesis	Hierarchical level	Sum of squares	Variance components	% variance
Early Tortonian	Among groups (four groups)	89.06	0.094	1.53*
(17 populations)	Among populations within groups	221.43	0.297	4.80***
	Within populations	3655.45	5.811	93.67***
Late Tortonian	Among groups (two groups)	31.36	0.160	2.54**
(17 populations)	Among populations within groups	279.12	0.338	5.36***
	Within populations	3655.45	5.811	92.09***
Earliest Messinian	Among groups (three groups)	34.16	0.001	0.02 NS
(19 populations)	Among populations within groups	310.49	0.358	5.79***
	Within populations	4011.97	5.822	94.19***
Mountain ranges	Among groups (three groups)	77.96	0.097	1.57**
(19 populations)	Among populations within groups	266.69	0.293	4.72***
	Within populations	4011.97	5.822	93.71***
ABC grouping	Among groups (three groups)	126.09	0.208	3.33***
(19 populations)	Among populations within groups	218.56	0.211	3.39***
	Within populations	4011.97	5.822	93.27***

Table S3 – Prior distributions of demographic and mutation model parameters for the four scenarios tested using Approximate Bayesian Computation (ABC) in DIYABC. Graphical representations of scenario topologies are presented in Fig. 5.

Priors for the demographic parameters	
N1	UN~ [10 – 750 000]
N2	UN~ [10 – 750 000]
N3	UN~ [10 – 750 000]
$N_{1\text{-}2}$	UN~ [10 – 750 000]
$N_{1\text{-}3}$	UN~ [10 – 750 000]
$N_{2\text{-}3}$	UN~ [10 – 750 000]
N_{1anc}	UN~ [10 – 750 000]
N_{2anc}	UN~ [10 – 750 000]
N_{3anc}	UN~ [10 – 750 000]
N_x	UN~ [10 – 1 500 000]
t_1	UN~ [10 – 350 000]
t_2	UN~ [10 – 350 000]
Constraint on parameter	$t_2 > t_1$
Priors for the mutation model for microsatellites	
Mean mutation rate (μ)	LU~ [1.0×10^{-7} – 1.0×10^{-3}]
Individual locus mutation rate	GAM ~ [1.0×10^{-7} – 1.0×10^{-2}, 1.0×10^{-4}, 2]
Mean coefficient (p)	GAM ~ [1.0×10^{-1} – 1.0, 0.5, 2]
Individual locus coefficient (p)	GAM ~ [1.0×10^{-2} – 1.0, 0.5, 2]
Summary statistics (SS) used in ABC analyses	
Mean number of alleles	
Mean genetic diversity	
Mean allele size variance	
Mean Garza-Williamson's (M) for each population	
F_{ST} for each pair of populations	

N1, effective population size of group A
N2, effective population size of group B
N3, effective population size of group C
$N_{1\text{-}2}$, effective population size of group A under scenario III
$N_{1\text{-}3}$, effective population size of the ancestral group A-C under scenario IV
$N_{2\text{-}3}$, effective population size of the ancestral group B-C under scenario II
N_{1anc}, effective population size of the ancestral group A under scenario II
N_{2anc}, effective population size of the ancestral group B under scenario IV
N_{3anc}, effective population size of the ancestral group C under scenario III
N_x, effective population size of the most ancestral population under every scenario
t_1, time (in generations=years) to the most recent divergence event under the scenarios II, III or IV
t_2, time (in generations=years) to the most ancient divergence event under scenarios I, II, III or IV (see scenarios in Fig. 5)
UN, uniform distribution with two parameters [min-max]
LU, Log-uniform distribution with two parameters [min-max]
GAM, gamma distribution with four parameters [min-max, mean, shape]

Table S4 – Genetic differentiation between all populations of the scrub-legume grasshopper in southeast Iberia. Pairwise F_{ST} values and F_{ST} values corrected for null alleles are presented below and above the diagonal, respectively. Significant pairwise F_{ST} values after sequential Bonferroni correction are indicated in bold. Population codes are described in Table S1.

	RON	TEJ	PAR	PAN	MAG	ARA	OTE	RAG	GAD	BAZ	FIL	MAR	CAZ	POY	PIN	CAB	CAR	ALM	ESP
RON	-	0.016	0.036	0.026	0.028	0.027	0.078	0.094	0.098	0.060	0.076	0.083	0.049	0.062	0.049	0.076	0.063	0.055	0.097
TEJ	0.026	-	0.040	0.006	0.015	0.016	0.060	0.080	0.079	0.029	0.054	0.060	0.032	0.047	0.035	0.036	0.048	0.037	0.076
PAR	**0.046**	**0.053**	-	0.044	0.064	0.051	0.122	0.144	0.144	0.094	0.117	0.126	0.084	0.100	0.093	0.108	0.099	0.104	0.136
PAN	**0.041**	0.014	**0.054**	-	0.012	0.016	0.052	0.062	0.067	0.022	0.040	0.044	0.025	0.033	0.027	0.029	0.037	0.025	0.063
MAG	**0.039**	**0.025**	**0.078**	0.022	-	0.024	0.055	0.060	0.055	0.024	0.037	0.050	0.027	0.038	0.031	0.034	0.037	0.026	0.071
ARA	**0.045**	**0.033**	**0.069**	0.030	**0.040**	-	0.055	0.073	0.065	0.033	0.054	0.056	0.036	0.041	0.040	0.041	0.046	0.039	0.079
OTE	**0.096**	**0.077**	**0.141**	**0.070**	**0.066**	**0.071**	-	0.033	0.025	0.034	0.033	0.026	0.034	0.035	0.033	0.051	0.043	0.041	0.049
RAG	**0.119**	**0.099**	**0.164**	**0.079**	**0.079**	**0.095**	**0.044**	-	0.022	0.039	0.030	0.036	0.040	0.050	0.038	0.048	0.048	0.051	0.055
GAD	**0.113**	**0.096**	**0.159**	**0.077**	**0.069**	**0.086**	**0.041**	0.028	-	0.036	0.015	0.020	0.027	0.032	0.025	0.032	0.026	0.033	0.035
BAZ	**0.083**	**0.040**	**0.111**	**0.037**	**0.034**	**0.049**	**0.055**	**0.058**	**0.047**	-	0.014	0.026	0.037	0.037	0.034	0.021	0.034	0.025	0.066
FIL	**0.097**	**0.067**	**0.130**	**0.052**	**0.048**	**0.072**	**0.046**	**0.045**	0.022	0.021	-	0.018	0.028	0.024	0.027	0.024	0.019	0.027	0.051
MAR	**0.108**	**0.079**	**0.148**	**0.063**	**0.065**	**0.074**	**0.039**	**0.054**	0.037	**0.040**	**0.029**	-	0.015	0.016	0.016	0.023	0.017	0.020	0.020
CAZ	**0.068**	**0.050**	**0.098**	**0.043**	**0.040**	**0.057**	**0.046**	**0.049**	0.037	**0.052**	**0.041**	**0.031**	-	0.015	0.003	0.023	0.015	0.014	0.024
POY	**0.089**	**0.068**	**0.124**	**0.050**	**0.050**	**0.068**	**0.048**	**0.068**	**0.045**	**0.055**	**0.032**	0.024	0.028	-	0.005	0.013	0.007	0.013	0.027
PIN	**0.069**	**0.053**	**0.108**	**0.045**	**0.046**	**0.060**	**0.046**	**0.052**	0.030	**0.050**	**0.036**	**0.027**	0.015	0.015	-	0.009	0.008	0.004	0.021
CAB	**0.098**	**0.057**	**0.129**	0.045	**0.050**	**0.057**	**0.060**	**0.056**	0.044	0.029	0.031	0.038	0.039	0.027	0.017	-	0.023	0.006	0.024
CAR	**0.080**	**0.060**	**0.111**	**0.049**	**0.048**	**0.066**	**0.057**	**0.064**	**0.041**	**0.044**	**0.027**	**0.027**	**0.026**	0.013	0.012	0.033	-	0.011	0.032
ALM	**0.075**	**0.057**	**0.120**	**0.041**	**0.039**	**0.062**	**0.051**	**0.067**	**0.041**	**0.044**	**0.037**	**0.033**	0.027	0.022	0.013	0.022	0.015	-	0.045
ESP	**0.109**	**0.087**	**0.148**	**0.074**	**0.082**	**0.096**	**0.061**	**0.069**	**0.049**	**0.073**	**0.060**	**0.030**	**0.034**	**0.034**	**0.029**	**0.026**	**0.039**	**0.048**	-

Table S5 – Generalized linear models (GLM) testing the association between allelic richness (A_R, standardized for sample size) in populations of the scrub-legume grasshopper and current climate suitability (HS_{CUR}), LGM climate suitability (HS_{LGM}), and climate suitability stability (HS_{STA}) estimated at three different spatial scales (1, 10 and 100 km^2) around sampling localities. Longitude and latitude were also included as covariates in the models. Details about model selection are given for the best ranked equivalent models ($\Delta AIC_c \leq 2$). For each model we indicate *K*, number of parameters in the model; AIC_c, corrected Akaike's information criterion (AIC) value; ΔAIC_c, difference in AIC_c value from that of the strongest model; ωi, AIC_c weight. Model averaging was performed for the best ranked equivalent models ($\Delta AIC_c \leq 2$) in order to obtain parameter estimates and unconditional standard errors (S.E.). The relative importance of each predictor is indicated ($\sum \omega_i$, sum of Akaike weights of models with $\Delta AIC_c \leq 2$ in which the predictor was present).

Model	Model parameters	*K*	AIC_c	ΔAIC_c	ω_i	Estimate ± SE	$\sum \omega_i$	Lower 95% CI	Upper 95% CI
A_R (1 km^2)									
1	Latitude	1	16.90	0.00	0.418	0.10 ± 0.15	0.42	-0.20	0.41
2	Null model	0	16.91	0.01	0.416				
3	HS_{LGM}	1	18.76	1.86	0.165	0.05 ± 0.19	0.17	-0.31	0.42
A_R (10 km^2)									
1	Latitude	1	16.90	0.00	0.300	0.07 ± 0.42	0.30	-0.20	0.35
2	Null model	0	16.91	0.01	0.299				
3	HS_{STA}	1	18.47	1.57	0.137	0.02 ± 0.09	0.14	-0.15	0.20
4	HS_{LGM}	1	18.51	1.61	0.134	0.04 ± 0.17	0.13	-0.28	0.38
5	HS_{CUR}	1	18.58	1.68	0.130	0.04 ± 0.16	0.13	-0.28	0.37
A_R (100 km^2)									
1	Latitude	1	16.90	0.00	0.269	0.06 ± 0.13	0.27	-0.20	0.33
2	Null model	0	16.91	0.01	0.268				
3	HS_{STA}	1	17.47	0.57	0.203	0.03 ± 0.11	0.20	-0.18	0.25
4	HS_{LGM}	1	18.03	1.13	0.153	0.09 ± 0.24	0.15	-0.38	0.58
5	HS_{CUR}	1	18.74	1.84	0.107	0.04 ± 0.18	0.11	-0.31	0.40

Table S6 – Generalized linear models (GLM) testing the association between the genetic admixture (G_{ADMIX}) in populations of the scrub-legume grasshopper and climate suitability stability (HS_{STA}) estimated at three different spatial scales (1, 10 and 100 km^2) around sampling localities. Genetic admixture was estimated on the basis of STRUCTURE analyses for K = 2 (left) and K = 3 (right).Longitude and latitude were also included as covariates in the models. Details about model selection are given for the best ranked equivalent models ($\Delta AIC_c \leq 2$). For each model we indicate K, number of parameters in the model; AIC_c, corrected Akaike's information criterion (AIC) value; ΔAIC_c, difference in AIC_c value from that of the strongest model; ωi, AIC_c weight. Model averaging was performed for the best ranked equivalent models ($\Delta AIC_c \leq 2$) in order to obtain parameter estimates and unconditional standard errors (S.E.). The relative importance of each predictor is indicated ($\sum\omega_i$, sum of Akaike weights of models with $\Delta AIC_c \leq 2$ in which the predictor was present). GLMs built using G_{ADMIX} calculated for K= 4-5 provided similar results (data not shown).

Model	Model parameters	K	AIC_c	ΔAIC_c	ω_i	Estimate ± SE	$\sum\omega_i$	Lower 95% CI	Upper 95% CI
$G_{ADMIX\ [K=2]}$ (1 km^2)									
1	Null model	0	8.60	0.00	0.711				
2	HS_{STA}	1	10.41	1.81	0.289	0.04 ± 0.01	0.29	-0.15	0.23
$G_{ADMIX\ [K=2]}$ (10 km^2)									
1	Null model	0	8.60	0.00	0.596				
2	HS_{STA}	1	9.38	0.78	0.404	0.07 ± 0.12	0.40	-0.17	0.32
$G_{ADMIX\ [K=2]}$ (100 km^2)									
1	Null model	0	8.60	0.00	0.643				
2	HS_{STA}	1	9.78	1.18	0.357	0.06 ± 0.12	0.36	-0.18	0.31

Model	Model parameters	K	AIC_c	ΔAIC_c	ω_i	Estimate ± SE	$\sum\omega_i$	Lower 95% CI	Upper 95% CI
$G_{ADMIX\ [K=3]}$ (1 km^2)									
1	Null model	0	3.90	0.00	0.550				
2	Latitude	1	5.52	1.62	0.245	0.02 ± 0.07	0.24	-0.11	0.17
3	Longitude	1	5.88	1.98	0.205	0.01 ± 0.03	0.20	-0.06	0.08
$G_{ADMIX\ [K=3]}$ (10 km^2)									
1	Null model	0	3.90	0.00	0.550				
2	Latitude	1	5.52	1.62	0.245	0.03 ± 0.08	0.24	-0.12	0.18
3	Longitude	1	5.88	1.98	0.205	0.01 ± 0.04	0.20	-0.06	0.09
$G_{ADMIX\ [K=3]}$ (100 km^2)									
1	Null model	0	3.90	0.00	0.550				
2	Latitude	1	5.52	1.62	0.245	0.03 ± 0.08	0.24	-0.12	0.18
3	Longitude	1	5.88	1.98	0.205	0.01 ± 0.04	0.20	-0.06	0.09

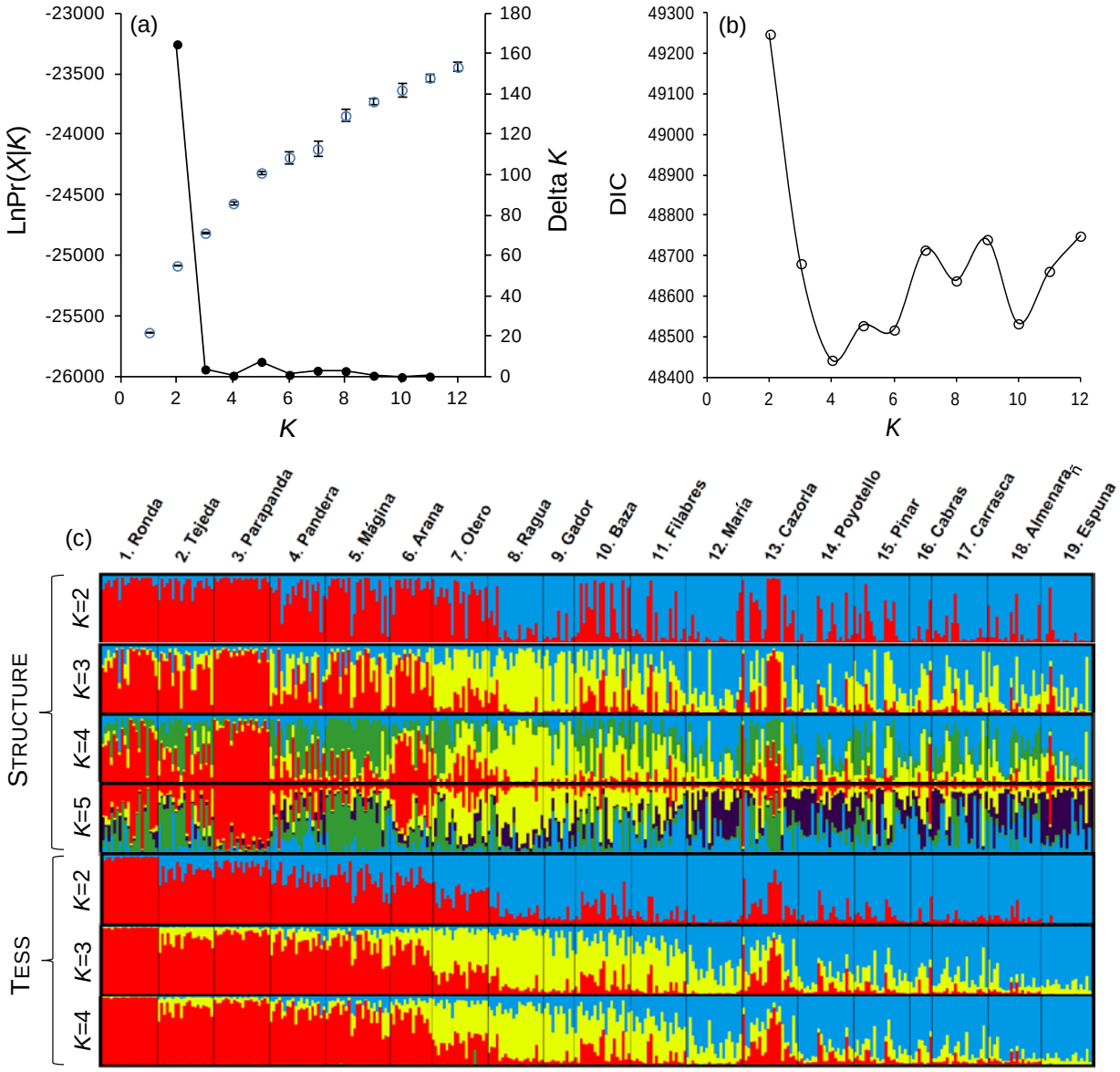


Figure S1 – (a) Results of Bayesian clustering analyses in STRUCTURE. Mean (± SD) log probability of the data [Ln Pr (*X*|*K*)] over 10 runs (left axis, open dots and error bars) for each *K*-value. The magnitude of Δ*K* as a function of *K* determines the best-supported number of clusters (*K* = 2) in STRUCTURE analyses (right axis and black dots). (b) Results of Bayesian clustering analyses in TESS. Mean DIC value over 20 runs for each *K*-value in TESS analyses. The minimum value of DIC before the first increase or stabilization indicates the best-supported number of clusters. (c) Results of genetic assignments of 354 individuals of scrub-legume grasshopper from southeast Iberia based on the Bayesian methods implemented in the program STRUCTURE and TESS for different numbers of genetic clusters (*K*). Each individual corresponds to a vertical bar, which is partitioned into *K*-coloured segments that represent the individual's probability of belonging to the cluster with that colour. Black lines separate individuals from different populations.

CAPÍTULO V

Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper

VÍCTOR NOGUERALES, PEDRO J. CORDERO & JOAQUÍN ORTEGO

BMC Evolutionary Biology, 16, 96 (2016)

Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper

Abstract

Understanding the underlying processes shaping spatial patterns of genetic structure in free-ranging organisms is a central topic in evolutionary biology. Here, we aim to disentangle the relative importance of neutral (*i.e.* genetic drift) and local adaptation (*i.e.* ecological divergence) processes in the evolution of spatial genetic structure of the Morales grasshopper (*Chorthippus saulcyi moralesi*), a narrow-endemic taxon restricted to the Central Pyrenees. More specifically, we analysed range-wide patterns of genetic structure and tested whether they were shaped by geography (isolation-by-distance, IBD), topographic complexity and present and past habitat suitability models (isolation-by-resistance, IBR), and environmental dissimilarity (isolation-by-environment, IBE). Different clustering analyses revealed a deep genetic structure that was best explained by IBR based on topographic complexity. Our analyses did not reveal a significant role of IBE, a fact that may be due to low environmental variation among populations and/or consequence of other ecological factors not considered in this study are involved in local adaptation processes. IBR scenarios informed by current and past climate distribution models did not show either a significant impact on genetic differentiation after controlling for the effects of topographic complexity, which may indicate that they are not capturing well microhabitat structure in the present or the genetic signal left by dispersal routes defined by habitat corridors in the past. Overall, these results indicate that spatial patterns of genetic variation in our study system are primarily explained by neutral divergence and migration-drift equilibrium due to limited dispersal across abrupt reliefs, whereas environmental variation or spatial heterogeneity in habitat suitability associated with the complex topography of the region had no significant effect on genetic discontinuities after controlling for geography. Our study highlights the importance of considering a comprehensive suite of potential isolating mechanisms and analytical approaches in order to get robust inferences on the processes promoting genetic divergence of natural populations.

INTRODUCTION

Understanding the factors structuring genetic variation in natural populations is a paramount topic in evolutionary biology (Carnaval *et al*. 2009; Yannic *et al*. 2014). The genetic structure of populations is primarily determined by inter-population dispersal rates and realized gene flow, which in turn are shaped by geography, environment, historical processes and, more frequently, their combined effects (Lee & Mitchell-Olds 2011; Shaffer & Wolf 2013; Wang *et al*. 2013). The isolation-by-distance (IBD) model predicts that genetic differentiation increases with Euclidean geographic distance because of limited dispersal and genetic drift (Wright 1943; Slatkin 1993). Even though this classic model explains spatial

patterns of genetic differentiation in a wide range of organisms (Wright 1943; but see review in Jenkins *et al*. 2010), it does not consider more sophisticated information than straight-line geographic distances and assumes landscape homogeneity, an unrealistic scenario for most natural systems (Manel *et al*. 2003; Manel & Holderegger 2013). Recently, the emergence of landscape genetics has explicitly incorporated landscape complexity into the study of evolutionary processes (Manel *et al*. 2003; Manel & Holderegger 2013), bringing new approaches that take into account the ability of organisms to disperse across different landscape features according with the resistance that they offer to movement (*i.e*. isolation-by-resistance, IBR; McRae 2006; McRae & Beier 2007; *e.g*. Ruiz-González *et al*. 2015; Ortego *et al*. 2015a).

Beyond geography and the spatial configuration of connecting corridors and isolating barriers, environment can also play a major role in shaping spatial patterns of genetic differentiation (Shaffer & Wolf 2013; Wang & Bradburd 2014). This occurs when populations inhabiting ecologically dissimilar habitats experience limited gene flow due to the low performance of immigrants arriving to areas where they may not be locally adapted or as consequence of the reluctance of individuals to cross or establish in unfamiliar habitats (*i.e.* isolation-by-environment, IBE; Nosil *et al*. 2005; Nosil 2012; Wang & Bradburd 2014). Recent research has revealed the ubiquity of IBE patterns, highlighting the importance of ecological factors in shaping genetic structure of natural populations (see meta-analyses in Shaffer & Wolf 2013; Sexton *et al*. 2014). The contribution of environment relative to geography in explaining spatial patterns of genetic variation may vary depending on species characteristics and ecological features of the natural systems (Funk *et al*. 2006; Sexton *et al*. 2014). Indeed, different isolating mechanisms are not mutually exclusive and gene flow is often influenced by a combination of geographical and ecological factors (Crispo *et al*. 2006; Edwards *et al*. 2012; Wang *et al*. 2013). Given that geography and environment are often highly correlated, disentangling their relative impact on population genetic differentiation harbors inherent analytical difficulties (for a discussion about eco-spatial autocorrelation, see Meirmars 2012) that have promoted the progressive development of more robust and accurate statistical methods (Balkenhol *et al*. 2009; Wang 2013; Botta *et al*. 2015; Kierepka & Latch 2015; Forester *et al*. 2016). However, only a few studies have jointly considered the relative effects of geography, landscape composition and environmental heterogeneity on either landscape-level (Ferrer *et al*. 2016; Munshi-South *et al*. 2016) or range-wide patterns of genetic structure (Wang *et al*. 2013).

The Pyrenean Morales grasshopper (*Chorthippus saulcyi moralesi*) (Orthoptera: Acrididae) is a narrow-endemic subspecies belonging to the *Chorthippus binotatus* group, exclusively distributed in central Aragón and Catalonia Pyrenean mountains (for taxonomic status and detailed description, see Defaut 2011). It is a winged grasshopper primarily feeding on gramineous herbs (Defaut 2011) and patchily distributed across a gradient of habitats, including submediterranean shrub formations, mesophile grasslands, montane shrubby vegetation and subalpine open grasslands located at elevations ranging between 1 100 and 2 400 m.a.s.l. (Llucià-Pomares 2002; Defaut 2011). Its distribution range is restricted to a narrow longitudinal axis with an east-west orientation characterized by a gradient of precipitation and temperature, from Atlantic to Mediterranean climate regimes (Llucià-Pomares 2002). The Pyrenees constitute the natural northern border of the Iberian Peninsula and present a high topographic and environmental complexity, rich biodiversity and considerable number of endemic species (García-Barros *et al*. 2002). These mountains experienced dramatic climate fluctuations during the Pleistocene (Jiménez-Sánchez *et al*. 2013), which are expected to have strongly influenced the demographic history and altered the distribution of many organisms of the region (Hewitt 2000). Despite the wide variety of habitats and altitudinal ranges occupied by the Pyrenean Morales grasshopper, populations at elevations lower than 1 400 m and above 2 100 m are anecdotal (Llucià-Pomares 2002; Defaut 2011). This suggests that valley bottoms and mountain tops, together with the complex topographic complexity of the region, may act as barriers to dispersal in this species (Wang 2012; Castillo *et al*. 2014). Thus, our study system has a great potential to examine the relative role of geography, environmental heterogeneity and present and past configuration of suitable habitats (*i.e*. corridors and barriers to dispersal) in shaping patterns of genetic differentiation throughout the entire distribution range of a narrowly distributed taxon (Sobel *et al*. 2010). We first analyzed spatial patterns of genetic structure and then employed a suite of complementary statistical approaches to test three different plausible scenarios of population differentiation (see Fig. 1 for a summary of the workflow employed in this study). In particular, we used multiple matrix regressions with randomization (MMRR; Wang 2013), distance-based redundancy analyses (dbRDA; Legendre & Anderson 1999) and geostatistical modelling based on Bayesian inference (Botta *et al*. 2015) to test whether the spatial pattern of genetic differentiation in the Pyrenean Morales grasshopper is explained by (i) geographic distances (IBD), (ii) resistance distances based on topographic complexity and current and past (last glacial maximum and last interglacial) climate suitability (IBR); and (iii) altitudinal and environmental dissimilarity between populations (IBE) (Fig. 1). If geography, topography or corridors/barriers defined by habitat suitability are identified as the major drivers of genetic structure, then migration-drift

equilibrium and neutral divergence can be regarded as the main evolutionary force shaping genetic discontinuities (Wang *et al*. 2013). A predominant or independent significant effect of environment on disrupting gene among populations would point to a role of ecological divergence and local adaptation processes in the evolution of spatial genetic structure (Pflüeger & Balkenhol 2014).

MATERIAL AND METHODS

POPULATION SAMPLING

Between 2012 and 2014, we collected 202 individuals from 11 populations of the Pyrenean Morales grasshopper. We aimed to sample populations throughout the entire distribution range of the species (~7 000 km^2; Fig. 2a) based on occurrence-data available in the literature (Llucià-Pomares 2002; Defaut 2011) and our own prospection of areas with potentially suitable habitats. The Morales grasshopper is primarily distributed in the south side of the Pyrenees (Spanish Pyrenees) and a small area located in the northeastern part of these mountains (French Pyrenees). The latter portion of the species distribution was intensively prospected during our field work but we only found a single population in the area (Err, France; Table 1). Overall, we were able to collect individuals from populations located across almost the entire climatic and elevation gradient occupied by the species (Table 1; Fig. S1), including populations from a wide range of habitats (from submediterranean shrub formations to subalpine grasslands) and differing in up to ~900 m of elevation (Table 1). We preserved specimens in 2 ml vials with 96% ethanol and stored at -20º C until DNA extraction. Population codes and more information on sampling sites are in Table 1. Specimens were collected in public lands under license from 'Gobierno de Aragón', 'Generalitat de Catalunya' and 'Ordesa y Monte Perdido National Park'.

MICROSATELLITE GENOTYPING AND BASIC GENETIC STATISTICS

We extracted genomic DNA from a hind-leg of each individual using a salt extraction protocol (Aljanabi & Martinez 1997). We amplified and genotyped each individual using the 18 microsatellites markers described in (Basiita *et al*. 2016). We performed PCR amplifications following the procedure described in Ortego *et al*. (2015a), run PCR products on an ABI 310 Genetic Analyzer (Applied Biosystems, Foster

City, CA, USA) and scored genotypes using GENEMAPPER 3.7 (Applied Biosystems, Foster City, CA, USA). We tested for deviations from Hardy-Weinberg Equilibrium (HWE) using exact tests (Guo & Thompson 1992) based on 900 000 Markov chain iterations and 100 000 dememorization steps as implemented in the program ARLEQUIN 3.5 (Excoffier & Lischer 2010). We discarded two loci from all downstream analyses because of heterozygosity departure from HWE in all populations, probably due to the presence of null alleles according to MICRO-CHECKER analyses (Van Oosterhout *et al*. 2004). We also used ARLEQUIN to test for linkage disequilibrium using a likelihood-ratio statistic, whose distribution was generated with 10 000 permutations. We did not find evidence for linkage disequilibrium between any pair of loci in any sampling population after sequential Bonferroni corrections (Rice 1989).

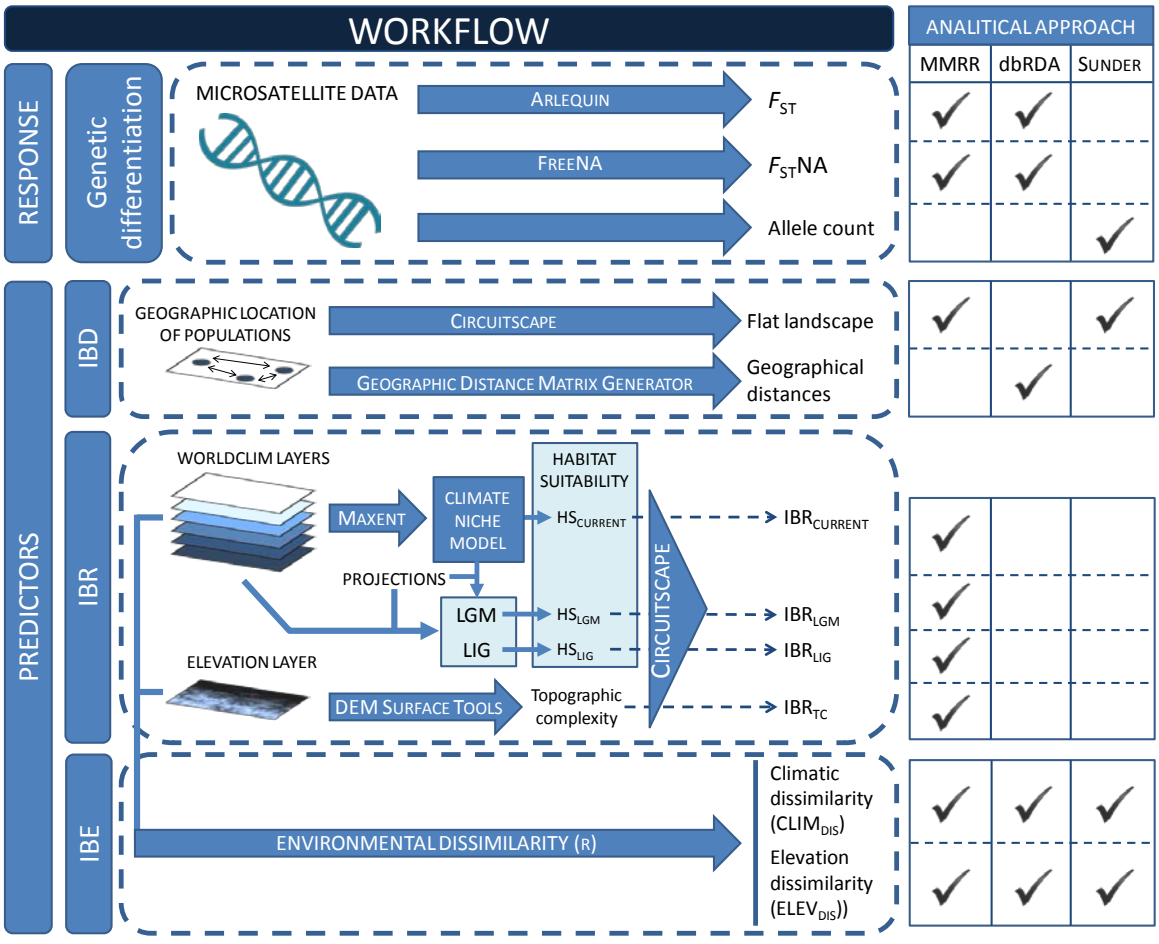


Figure 1 – Workflow summarizing the methodological approach employed in this study to analyze the relative contribution of isolation by distance (IBD), isolation by resistance (IBR), and isolation by environment (IBE) in structuring genetic variation in the Pyrenean Morales grasshopper. The response variables and predictors considered for each analytical approach (MMRR, dbRDA, and SUNDER) are indicated. HS: Habitat suitability; LGM: Last Glacial Maximum; LIG: Last Inter Glacial; MMRR: multiple matrix regression with randomization; dbRDA: distance-based redundancy analysis.

Table 1 – Geographical location, elevation and number of genotyped individuals (*n*) for the studied populations of the Pyrenean Morales grasshopper.

Locality	Code	Latitude	Longitude	Elevation (m.a.s.l)	*n*
Torla	TOR	42.63860	-0.08196	1822	20
Nerín	NER	42.59047	0.01062	1623	20
Saravillo	SAR	42.56089	0.23046	1313	20
Chía	CHI	42.56833	0.41708	1900	20
Aspes	ASP	42.44307	0.58389	1361	20
Boi	BOI	42.47945	0.86732	2046	12
Perves	PER	42.35250	0.83709	1370	20
Carmeniu	CAR	42.37243	1.33766	1140	20
Err	ERR	42.42762	2.05956	1800	19
Creueta	CRE	42.30131	1.99350	1928	11
Rasos	RAS	42.14165	1.76461	1840	20

GENETIC STRUCTURE ANALYSES

We estimated population genetic differentiation calculating F_{ST}-values between all pairs of sampling populations and testing their significance with Fisher´s exact test after 10 000 permutations using ARLEQUIN. We corrected *P*-values using a sequential Bonferroni adjustment (Rice 1989). Due to the high frequency of null alleles in Orthoptera (*e.g.* Chapuis *et al*. 2008; Ortego *et al*. 2015a), we also calculated pairwise F_{ST}-values corrected for null alleles (F_{ST}NA) using the so-called ENA-method implemented in the program FREENA (Chapuis & Estoup 2007).

We inferred genetic structure using Bayesian clustering analyses in STRUCTURE 2.3.3 (Pritchard *et al*. 2000; Falush *et al*. 2003). We performed an iterative approach to assess the hierarchical genetic structure that could underlie broad genetic clustering patterns identified by STRUCTURE analyses including all populations (for a similar approach, see Papadopoulou & Knowles 2015). After an initial global analysis including all populations, we analyzed subsequent subsets of the data corresponding to the respective genetic clusters identified in the previous hierarchical level. For all analyses, we considered correlated allele frequencies and an admixture model without prior information on population origin. We performed 10 independent runs for each value of *K* (1 to 12 for the complete dataset; and 1 to *n*+*x* for reduced datasets of *n* populations, being *x* a number to achieve at least three Δ*K* values) with a burn-in period of 200 000 steps and a run length of 1 000 000 Markov chain Monte Carlo (MCMC) cycles. We estimated the best-supported number of genetic clusters with the log probability of the data [Ln Pr (X|*K*)] (Falush *et al*. 2003) and the Δ*K* method (Evanno *et al*. 2005). We used the '*full search*'

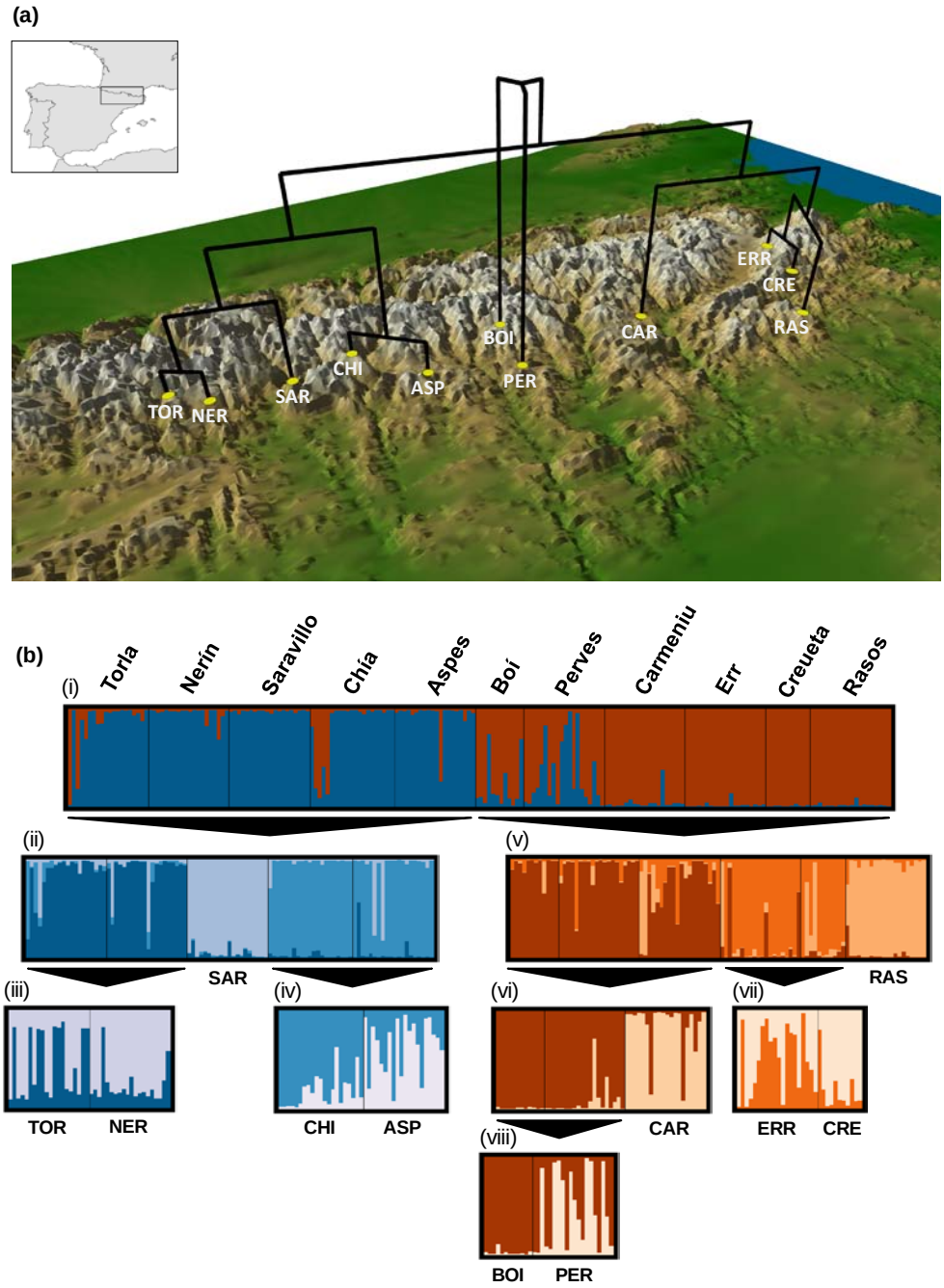


Figure 2 – Population genetic structure in the Pyrenean Morales grasshopper. Panel (a) shows sampling sites of the species and phylogenetic relationships among the 11 populations inferred from a neighbor-joining (NJ) tree based on Cavalli-Sforza and Edwards chord distances (D_c). The tree was plotted on a topographic map of the Pyrenees using the software GENGIS (Parks *et al*. 2009). We downloaded topographic data from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data; *http://srtm.csi.cgiar.org/*) as 90 m resolution digital elevation model and subsequently transformed to 30 arc-sec (*c*. 1 km) resolution for representation. Panel (b) represents the results of genetic assignment of 202 individuals of the Pyrenean Morales grasshopper based on the Bayesian method implemented in STRUCTURE. We performed hierarchical analyses for subsets of populations considering the most probable *K*-value inferred at the previous hierarchical level (Fig. S2). Each individual corresponds to a vertical bar partitioned into *K*-colored segments that represent the individual's probability of belonging to the cluster with that color. Black lines separate individuals from different populations. Population codes as in Table 1.

algorithm in the program CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) to align replicated runs and average individual assignment probabilities for the most likely *K*-value. Finally, we used DISTRUCT 1.1 (Rosenberg 2004) to produce bar plots displaying probabilities of individual membership to each inferred genetic cluster.

Complementary to the Bayesian clustering method, we used a discriminant analysis of principal components (DAPC) to identify clusters of genetically related individuals (Jombart *et al*. 2010). Although both clustering methods exhibit similarities (*e.g.* they do not require *a priori* delimitation of populations), several important differences exist in their analytical approaches. STRUCTURE suffers from the assumption of HWE and gametic disequilibrium (Pritchard *et al*. 2000) and typically fails to detect isolation-by-distance (IBD) (Jombart *et al*. 2010) and hierarchical patterns of spatial genetic structure (Evanno *et al*. 2005). However, the multivariate analyses implemented in DAPC do not lay on the assumptions of STRUCTURE Bayesian analyses and could be more efficient to detect complex patterns of genetic differentiation (Jombart *et al*. 2010). DAPC is a methodological approach that requires data transformation using a principal component analysis (PCA) as a prior step to a discriminant analysis (DA). DA partitions genetic variation maximizing differences between clusters while minimizing within-cluster variation. We implemented DAPC analysis in R 3.1.2 using the package ADEGENET (Jombart *et al*. 2010; R Core Team 2015). At first, we used the '*find.cluster*' function using all available principal components (PCs) to determine the best-supported number of genetic clusters using the Bayesian Information Criterion (BIC). The '*find.cluster*' function runs successive *K*-means clustering with increasing number of clusters (*K*) and provides a BIC value for each simulated *K*-value (*i.e. K*-value with the lowest BIC is the 'optimal' number of clusters). Secondarily, we determined the optimal number of PCs for the DAPC by cross-validation using the '*xvalDapc'* function with 100 replicates. We selected the number of PCs associated with the lowest 'root mean squared error' (RMSE) value. We ran DAPC using all the available discriminant functions and calculated the assignment probability of individuals to each cluster, which were represented with barplots using DISTRUCT.

We constructed a phylogenetic tree to evaluate the genetic relationships among all populations. We used the program POPULATIONS 1.2.31 (Langella 1999) to obtain a neighbor-joining (NJ) tree based on pairwise Cavalli-Sforza and Edwards (D_c) genetic distances (Cavalli-Sforza & Edwards 1967). This genetic distance is the most accurate to yield the correct tree topology for microsatellite markers under a variety of evolutionary scenarios without making assumptions in relation to mutation rates or constant population sizes (Takezaki & Nei 1996).

CLIMATE NICHE MODELLING

We used climate niche models (CNMs) at different time periods to investigate whether current and past climate suitability are relevant factors shaping observed patterns of genetic differentiation among populations of the Pyrenean Morales grasshopper. For this purpose, we built a CNM using the 'maximum entropy presence-only' algorithm implemented in MAXENT 3.3.3 (Phillips *et al*. 2006; Phillips & Dudik 2008) based on current climate layers and using 50 cross-validation replicate model runs. We used a total of 47 occurrence points obtained from the literature (Llucià-Pomares 2002; Defaut 2011) and our own sampling. To construct the models, we used the 19 present-day bioclimatic variables available in WorldClim (*http://www.worldclim.org/*) and downloaded at 30 arc-sec (*c*. 1 km) resolution (Hijmans *et al*. 2005). To obtain the distribution of the Pyrenean Morales grasshopper in the Last Glacial Maximum (LGM, *c*. 21 Kya) and the Last Interglacial (LIG, *c*. 120-140 Kya), we projected contemporary species-climate relationships to these periods. The LGM layers were based on the Community Climate System Model (CCSM3; Collins *et al*. 2006) from the Palaeoclimate Modelling Intercomparison Project Phase II (PMIP2; Braconnot *et al*. 2007). We downloaded LGM layers from WorldClim at 2.5 arc-min and interpolated to 30 arc-sec resolutions. LIG layers were based on (Otto-Bliesner *et al*. 2006) and downloaded from WorldClim at 30 arc-sec resolution. According to the suggestions from (Anderson & Raza 2010), we limited the geographic extent of the climate layers to an area approximately 20% larger than the known distribution range of the species in order to avoid model over-fitting. We used multivariate environmental similarity surfaces (MESS) calculation to address the problems derived from projecting the current distribution into novel climates (*i.e.* LGM and LIG periods) (Alvarado-Serrano & Knowles 2014). We used this approach to identify and discard climate layers with areas where the predictions should be treated with caution, due to the variables are outside the range present in the training data (for more details see Elith *et al*. 2010). We carried out MESS analyses iteratively excluding one variable in each step until discarding all out of range LGM and LIG variables compared to present-day variables. Finally, we checked the reliability of our past and current climate models following two approaches. At first, we developed a new current climate model using the variables with greater 5% importance (as selected by Jackknife of regularized training gain procedure) and we compared its similarity with the current model generated by MESS analyses. Second, we developed a new model including only the most informative variable (*Bio1*) and we compared its LGM and LIG projections with those obtained by MESS analyses (for a similar approach, see Massatti & Knowles 2014; Lanier *et al*. 2015). All the output maps from the models were visualized using threshold values based on maximum training sensitivity plus specificity (MTSS).

TOPOGRAPHIC COMPLEXITY

In order to investigate the role of topographic complexity (TC) as a potential factor shaping patterns of genetic differentiation, we calculated the surface ratio index for each cell from a digital elevation model using DEM SURFACE TOOLS (Jenness 2013) in ARCGIS 10.0 (ESRI, Redlands, CA, USA). Surface ratio is an index of topographic complexity, with values close to one indicating flat areas and values higher than one indicating an abrupt relief and deep slopes (Jenness 2004). We made calculations on a 90 m resolution digital elevation model from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data, *http://srtm.csi.cgiar.org/*) and the final layer was transformed to 30 arc-sec (*c*. 1 km) resolution for subsequent analyses. Additionally, we used the digital elevation model to calculate a matrix of differences in elevation between each pair of studied populations (*i.e.* an elevation dissimilarity matrix).

ENVIRONMENTAL CHARACTERIZATION OF POPULATIONS

In order to analyze the potential role of environment as a driving factor of genetic differentiation, we characterized the environmental space of each population using a principal component analysis (PCA) with 'varimax' rotation applied to the values of the 19 present-day bioclimatic variables from WorldClim extracted from sampling sites, occurrence points used in MAXENT and 1 000 randomly distributed points in the study area. This procedure allowed us to capture the environmental variation of the study area and avoid potential bias resulting from just considering environmental conditions from the sampling sites. Then, we obtained for each population the PC scores of the first three PCs, which explained the 73.18%, 10.92% and 8.38% respectively of the environmental variance (Fig. S1). Finally, we calculated environmental dissimilarity between each pair of populations using Euclidean distances for the obtained PC scores using the '*dist*' function in R. We performed PCA analysis using IBM SPSS 21.0 (IBM Coorp., Armonk, NY, USA).

ISOLATION BY RESISTANCE MATRICES

We applied circuit theory to model gene flow between populations and test the effects of different landscape resistance scenarios (IBR) on observed patterns of genetic differentiation (McRae & Beier

2007; McRae *et al*. 2008). We used CIRCUITSCAPE 4.0 (McRae 2006) to calculate resistance distance matrices between all pairs of populations considering an eight-neighbor cell connection scheme. We obtained different IBR distance matrices considering as inputs in CIRCUITSCAPE the following raster layers: current, LGM and LIG climate niche suitability and topographic complexity. We assigned pixel values of climate niche suitability maps as conductance values and pixel values of topographic complexity layers as resistance values. We also used CIRCUITSCAPE to test for the effect of isolation-by-distance (IBD) by calculating pairwise resistance distances on a completely 'flat' landscape based on a raster layer in which all cells had an equal value (conductance = 1). This IBD resistance model yields similar results than a matrix of Euclidean geographical distances, but it is more appropriate for comparison with others competing models also generated with CIRCUITSCAPE (Jha & Kremen 2013; Velo-Antón *et al*. 2013).

MULTIPLE MATRIX REGRESSION WITH RANDOMIZATION

All IBR matrices were tested against genetic distance matrices (pairwise F_{ST} and F_{ST}NA-values) using multiple matrix regressions with randomization (MMRR; Wang 2013) as implemented in R. In these models, we also included elevation ($ELEV_{DIS}$, see "Topographic complexity" section) and climate dissimilarity ($CLIM_{DIS}$, see "Environmental characterization of populations" section) distance matrices in order to test a possible pattern of IBE. We used a backward procedure to select final models, removing non-significant variables from an initial full model including all explanatory predictors. We tested the significance of the remaining variables again until no additional term reached significance (*e.g.* Ortego *et al*. 2015b).

DISTANCE-BASED REDUNDANCY ANALYSIS

Complementary to MMRR analyses, we tested the relationship between genetic differentiation, geography and environment using distance-based redundancy analyses (dbRDA; Legendre & Anderson 1999). This approach is based on a multivariate multiple regression and estimates the percentage of genetic variation explained by a given predictor or set of predictors. We performed dbRDA using the '*capscale*' function in the package VEGAN (Oksanen *et al*. 2015) as implemented in R. The genetic distance matrix (pairwise F_{ST} or F_{ST}NA-values) was tested against the following variables: (i) geographic

distances (IBD), (ii) elevation and (iii) population´s PC scores of the first three axes from the PCA performed on the environmental data (see "Environmental characterization of populations" section). Euclidean geographical distances between sampled populations were calculated using GEOGRAPHIC DISTANCE MATRIX GENERATOR 1.2.3 (Erst 2015). Geographic distances were tested after transforming the Euclidean geographical distance matrix to a continuous rectangular vector by principal coordinates analyses (PCoA) using the '*pcnm*' function in the package VEGAN. Significance of the predictors was assessed using multivariate *F*-statistics with 9 999 permutations using the '*anova.cca*' function included in the package VEGAN. We first analyzed the relationship between the genetic distance matrix and each variable separately (marginal test) and then we performed a partial dbRDA (conditional test) for each variable while controlling for the influence of geography (fitted as covariate).

GEOSTATISTICAL SIMULATIONS AND BAYESIAN INFERENCE

We quantified the relative effects of geography and environment on genetic differentiation using SUNDER (Botta *et al*. 2015), a novel geostatistical method modelling covariance in allele frequencies between populations as a decreasing function of geographical and ecological distances (Botta *et al*. 2015; see also Bradburd *et al*. 2013). SUNDER uses a Bayesian framework with a MCMC algorithm to estimate the magnitude of the effects of these variables and implements a model selection procedure by cross-validation to assess which sub-model (*e.g.* with or without the effect of environment) best fits to the data. Using the multinomial model, we ran SUNDER with 10 million of iterations for each data set, sampling every 1 000 iterations. We set to update in the MCMC iterations all parameters of the vector θ (α, variance of allele frequencies; β_G and β_E, magnitude of the effect of geography and environment respectively on genetic covariance; γ, smoothness of spatial variation of allele frequencies; δ, variation in the allele frequency of a population departing from the other populations). We also set their initial state and upper bounds of their Dirichlet prior distributions following suggestions in (Botta *et al*. 2015). We visually checked trace plots for parameters to assure convergence. We used the 10% of our data set (sites × locus) as validation set during the cross-validation procedure. We performed SUNDER analyses using as environmental matrices the elevation dissimilarity matrix ($ELEV_{DIS}$) and the climate dissimilarity distance matrix ($CLIM_{DIS}$), which we separately tested against IBR distance matrix based on a completely 'flat' landscape (IBD). Before analyses, we standardized all distance matrices by their respective standard deviations.

RESULTS

CLIMATE NICHE MODELLING

We constructed past and present-day final climate niche models using six bioclimatic variables: annual mean temperature (*Bio1*), mean temperature of the coldest quarter (*Bio11*), annual precipitation (*Bio12*), precipitation of the driest month (*Bio*14), precipitation seasonality (*Bio15*), and precipitation of the warmest quarter (*Bio18*). This model had a very high value of area under the receiving operator characteristics curve (AUC = 0.919 ± 0.067 SD), indicating overall good performance. The predicted habitat suitability area in the present was consistent with the current distribution of the species, but some areas in the northern and west side of the Pyrenees where the Morales grasshopper has not been recorded were also predicted to be suitable for the species (Fig. 3a). Projections of the present-day climate niche envelope to the LGM and LIG suggesting that the Pyrenean Morales grasshopper has experienced important distributional shifts during the Pleistocene. During the LGM, areas above 1 800-2 000 m.a.s.l. resulted unsuitable for the species and its potential distribution range expanded to areas of lower altitude across the western and northern side of the Pyrenees (Fig. 3b). Conversely, the potential distribution range of the species during the LIG expanded to higher altitudes but its peripheral geographical limits were similar than in the present (Fig. 3c).

POPULATION GENETIC STRUCTURE

Pairwise F_{ST}-values ranged from 0.021 to 0.216 and 53 of 55 comparisons were significant after sequential Bonferroni correction (Table S1). Pairwise F_{ST}NA-values were lower than F_{ST}-values and ranged from 0.014 to 0.148. Pairwise F_{ST}-values were highly correlated with pairwise F_{ST}NA-values (Mantel r = 0.941; P < 0.001). STRUCTURE analyses on all populations showed that the best-supported number of clusters was K = 2 according to the ΔK method (Fig. 2b; Fig. S2). These initial analyses detected a strong correspondence between the inferred genetic clusters and their geographic location, even when a broader range of K-values (K = 2-7) was evaluated. Subsequent hierarchical analyses on different subsets of populations detected further genetic structuring (Fig. 2b). Individual assignment probabilities to a certain genetic cluster were generally high and the distribution of the hierarchical genetic structure exhibited congruence with the geographical location of the studied populations. Consecutive hierarchical analyses revealed that each population constituted a single cluster, although

many pairs of populations showed a considerable degree of genetic admixture (Fig. 2b). DAPC analyses identified also a deep spatial genetic structure in concordance with the geographic location of populations. The minimum BIC values were obtained for *K* = 3-5 (Fig. 4a). Considering *K* = 4 (the *K*-value with the lowest BIC value), DAPC partitioned all individuals in western (TOR-NER-SAR populations), central-west (CHI-ASP), central-east (BOI-PER) and eastern (CAR-ERR-CRE and RAS) groups (Fig. 5). Discriminant functions based on DAPC analyses correctly assigned most individuals to the genetic cluster where they were assigned *a priori* by *K*-means analyses used to infer the best-supported clustering solution (Jombart *et al*. 2010) (Fig. 4c). The low overlapping of the genetic clusters on the ordination plot indicated high degree of differentiation between them (Fig. 4b). When *K* = 3 and *K* = 5 were considered and compared with *K* = 4, DAPC revealed a hierarchical distribution of genetic variation similar to that identified by STRUCTURE (Fig. 2b; Fig. 5). The result of the NJ tree based on D_c genetic distances showed that populations were included into monophyletic groups geographically clustered according to the hierarchical structure inferred by STRUCTURE and DAPC analyses (Fig. 2a-b; Fig. 5).

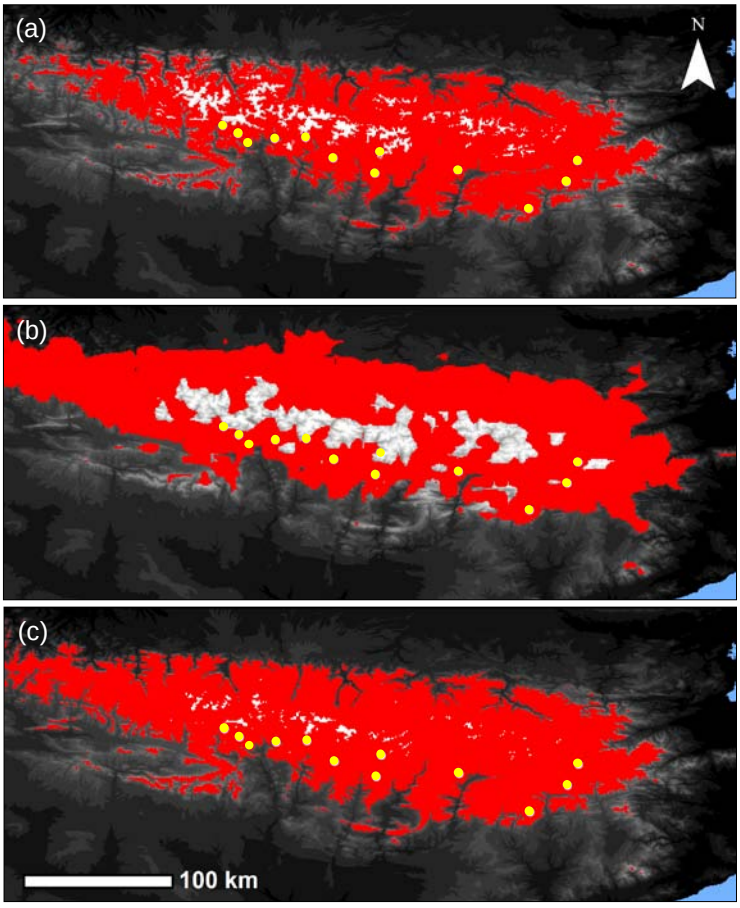


Figure 3 – Climate niche modelling of the Pyrenean Morales grasshopper for (a) the present, (b) the Last Glacial Maximum (LGM, *c*. 21 Kya) and (c) the Last Interglacial (LIG; *c*. 120-140 Kya). Climatically suitable areas defined using the maximum training sensitivity plus specificity threshold (MTSS) is in red. The topography is in the background map, in which whiter colors indicate higher elevations and darker colors indicate lower elevations (range from 0 to 3 404 m.a.s.l.). Yellow dots represent the eleven sampling sites.

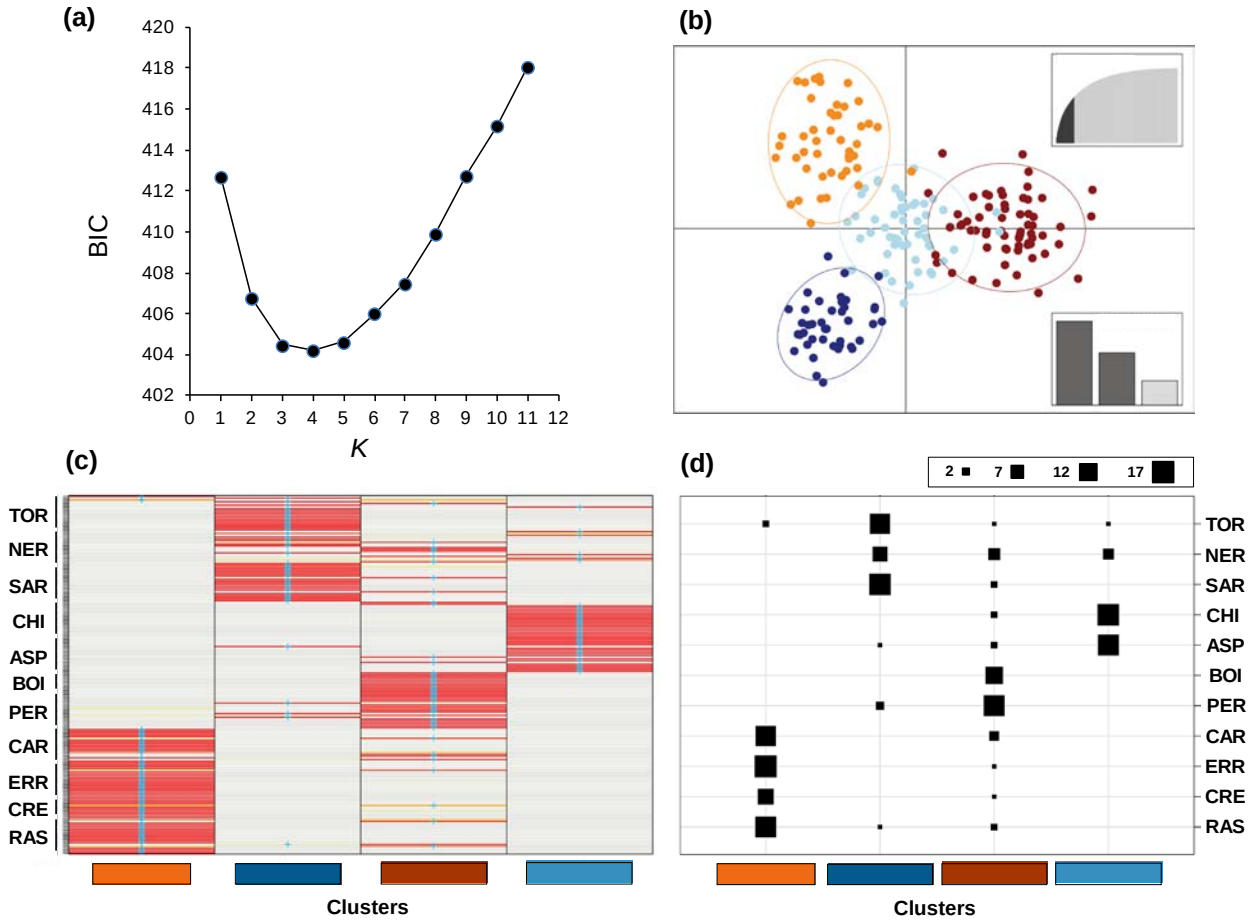


Figure 4 – Summary of the results of discriminant analysis of principal components (DAPC). Panel (a) represents the Bayesian Information Criterion (BIC) for each value of *K*. The minimum value of BIC before the first increase or stabilization indicates the best-supported number of genetic clusters (*K* = 4 in this case). Panel (b) represents an ordination plot for the first two discriminant axes. Each dot represents one individual and colors and inertia ellipses indicate their assignment to one of the four genetic clusters inferred by DAPC. The up-right graph inset displays the variance explained by the principal component axes used for DAPC (in dark grey). The bottom-right inset displays in relative magnitude the variance explained by the two discriminant axes plotted (in dark grey). Panel (c) represents whether the individuals (rows) were correctly assigned (based on discriminant functions) to the genetic cluster where they were included *a priori* (columns) by *K*-means analyses used to infer the best-supported clustering solution. Colors represent membership probabilities to each genetic cluster (red = 1, orange = 0.75, yellow = 0.25, white = 0) and blue crosses indicate the cluster where the individuals were originally assigned by *K*-means analyses. Generally, the DAPC classification of individuals is consistent with their assignment to the clusters originally identified by *K*-means analyses (*i.e.* blue crosses are on red rectangles). Panel (d) shows the number of individuals from each population (rows) assigned to each of the four inferred genetic clusters (columns). The size of black squares is proportional to the number of individuals assigned to the different clusters (up-right legend). Population codes are described in Table 1.

MULTIPLE MATRIX REGRESSION WITH RANDOMIZATION

Genetic differentiation (F_{ST}) was positively associated with geographic distance (*i.e.* IBD, resistance distances based on a completely 'flat' landscape), topographic complexity (IBR_{TC}) and current ($IBR_{CURRENT}$), LGM (IBR_{LGM}) and LIG (IBR_{LIG}) habitat resistance distances when these variables were included alone into different models (all *P*s < 0.012). Indeed, Mantel tests showed that all these variables were highly inter-correlated (Table S2). Climatic dissimilarity ($CLIM_{DIS}$) was also correlated with all other variables, but with comparatively lower Mantel *r* values, whereas elevation dissimilarity ($ELEV_{DIS}$) was only significantly correlated with $CLIM_{DIS}$ (Table S2). Univariate models including IBD and IBR_{TC} provided the highest and most similar coefficients of determination (r^2) (Table 2). However, only IBR_{TC} was included into the final model after the backward selection procedure (Fig. 6). The rest of variables did not remain significant when they were tested against topographic complexity (all *P*s > 0.13). Analyses based on F_{ST}NA yielded similar results, but models had generally lower values of coefficient of determination (r^2) (Table 2). Our results remained similar after sequential Bonferroni correction for multiple testing (Rice 1989).

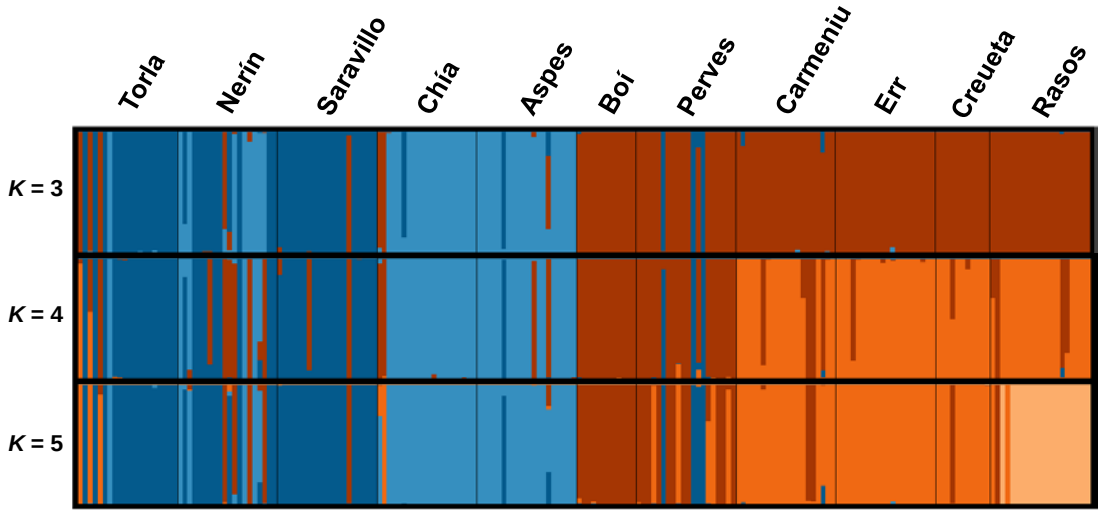


Figure 5 – Results of clustering analyses for 202 individuals of the Pyrenean Morales grasshopper for different numbers of genetic clusters (*K*) based on a discriminant analysis of principal components (DAPC). Each individual corresponds to a vertical bar partitioned into *K*-colored segments that represent the individual's probability of belonging to the cluster with that color. Black lines separate individuals from different populations. Cluster colors in *K* = 4 barplot correspond to those of Fig. 4.

Table 2 – Results of univariate matrix regressions with randomization for genetic differentiation (F_{ST} and F_{ST}NA-values corrected for null alleles) among eleven populations of the Pyrenean Morales grasshopper in relation with elevation ($ELEV_{DIS}$) and climatic ($CLIM_{DIS}$) dissimilarity and five isolation by resistance (IBR) scenarios: IBD, isolation by distance (*i.e.* equal resistance to all pixel values, equivalent to geographical distance); IBR_{TC}, topographic complexity; $IBR_{CURRENT}$, current habitat suitability; IBR_{LGM}, Last Glacial Maximum habitat suitability and IBR_{LIG}, Last Interglacial habitat suitability. Predictors with $P < 0.05$ in bold.

	F_{ST}				F_{ST} NA			
	r^2	β	*t*	*P*	r^2	β	*t*	*P*
IBD	0.390	0.92	5.82	0.001	0.332	0.83	5.13	0.001
IBR_{TC}	0.391	0.92	5.83	0.001	0.335	0.83	5.16	0.001
$IBR_{CURRENT}$	0.235	0.47	4.03	0.005	0.198	0.41	3.61	0.006
IBR_{LGM}	0.162	0.42	3.20	0.012	0.259	0.52	4.30	0.002
IBR_{LIG}	0.308	0.56	4.85	0.001	0.251	0.49	4.21	0.002
$ELEV_{DIS}$	0.008	-0.07	-0.64	0.509	0.013	-0.09	-0.83	0.397
$CLIM_{DIS}$	0.069	0.25	1.98	0.075	0.041	0.18	1.51	0.132

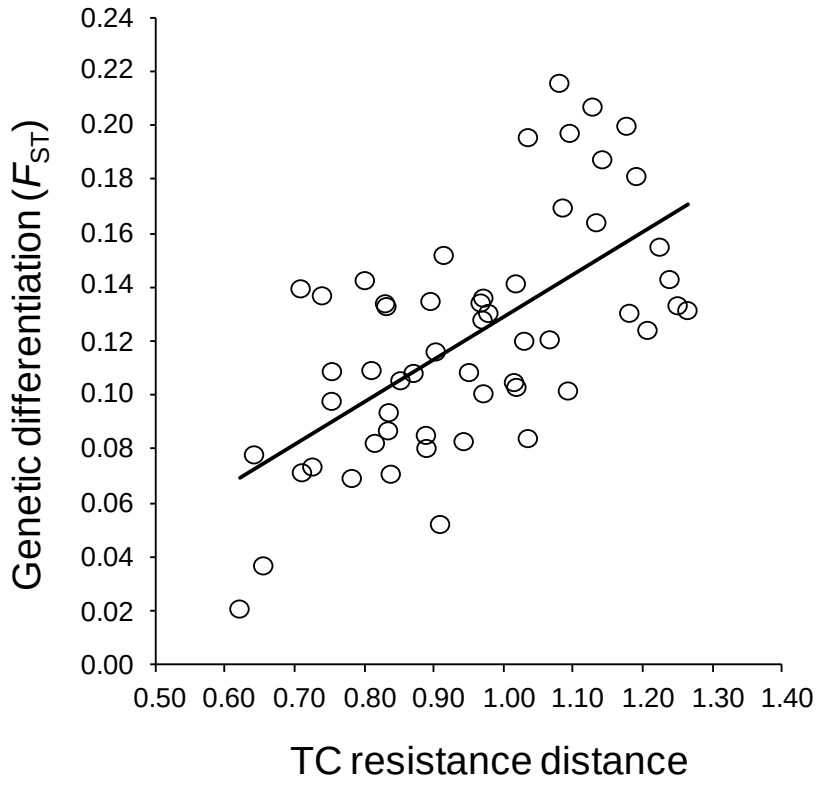


Figure 6 – Relationship between genetic differentiation (F_{ST}) and topographic complexity (TC) resistance distances (IBR_{TC}) (calculated using CIRCUITSCAPE) in eleven populations of Pyrenean Morales grasshopper. Regression line is shown.

DISTANCE-BASED REDUNDANCY ANALYSIS

Marginal tests showed a significant association between genetic differentiation (F_{ST} and F_{ST}NA-values) and geography and one environmental predictor (climate PC2) that explained 46.80-53.66% and 40.12-45.70% of genetic variation, respectively (Table 3). Climate PC2 was explained by a pool of bioclimatic variables related to annual temperature variation (*Bio*2, *Bio*3 and *Bio*7; Table S3). However, the environmental predictor (climate PC2) remained not significant after accounting for the influence of geography in the conditional test (Table 3). Our results remained similar after sequential Bonferroni correction (Rice 1989). Analyses based on F_{ST}NA-values (Table 3) or performed with PCA considering only the six bioclimatic layers employed to build the climate niche model in MAXENT provided similar results (data not shown).

GEOSTATISTICAL SIMULATIONS AND BAYESIAN INFERENCE

SUNDER analyses indicated that the models that best fit to the genetic data were those exclusively including the geographic component (Table 4). The Bayesian posterior estimates of the parameter β_G (representing the magnitude of the effect of geography) were smaller than β_E (representing the magnitude of the effect of environment), indicating a primordial effect of geographical distances and an absence of isolation by climate or elevation on genetic differentiation (Table 4).

DISCUSSION

Despite its small distribution range (< 200 km between the most distant populations), the Pyrenean Morales grasshopper exhibits a strong genetic structure congruent with the geographical location of its different populations. Our microsatellite-based clustering and phylogenetic analyses revealed a strong hierarchical structure and a relatively low degree of inter-group genetic admixture. In most cases, the distinct genetic clusters corresponded to single populations, which clustered in the hierarchically superior genetic group according to the main mountain chains of the region. In comparison with other Mediterranean orthoptera, the Pyrenean Morales grasshopper has a population genetic structure as remarkable as that found at a wider spatial scale in *Mioscirtus wagneri* (global F_{ST} ~0.19) and much

Table 3 – Results of distance-based redundancy analyses (dbRDA) testing the effects of geography, climate and elevation on genetic differentiation among eleven populations of the Pyrenean Morales grasshopper. Geography is tested after transforming the Euclidean geographical distance matrix to a continuous rectangular vector by principal coordinates analyses (PCoA). In marginal tests, we tested each predictor separately, while in conditional (partial) tests geography was always included as covariate. The proportion of the multivariate genetic variation explained (% var) by a given predictor or set of predictors is indicated. Predictors with $P < 0.05$ in bold.

Marginal tests				Conditional tests			
Variable	*F*	*P*	% var	Variable	*F*	*P*	% var
F_{ST}							
Geography	10.42	0.001	53.66				
Elevation	0.54	0.651	5.75	Elevation	0.48	0.844	2.64
Climate PC1	0.67	0.581	7.00	Climate PC1	0.42	0.883	2.35
Climate PC2	7.57	0.002	45.70	Climate PC2	0.62	0.720	3.35
Climate PC3	1.59	0.194	15.06	Climate PC3	2.02	0.084	9.37
F_{ST} NA							
Geography	7.91	0.001	46.80				
Elevation	0.48	0.747	5.14	Elevation	0.43	0.871	2.76
Climate PC1	0.61	0.661	6.40	Climate PC1	0.27	0.961	1.79
Climate PC2	6.02	0.005	40.12	Climate PC2	0.70	0.665	4.33
Climate PC3	1.84	0.144	17.05	Climate PC3	2.13	0.073	11.22

Table 4 – Results of Bayesian inference and model selection in SUNDER testing the relative effects of geographical and environmental variables on genetic differentiation among eleven populations of the Pyrenean Morales grasshopper. We separately tested the environmental variables [elevation ($ELEV_{DIS}$) and climatic ($CLIM_{DIS}$) dissimilarity matrices] against an IBD resistance distance matrix (*i.e.* equal conductance to all pixel values, equivalent to geographic distance). 'G' corresponds to models only considering geography, 'E' corresponds to models only considering the environmental variable, and 'G + E' corresponds to models considering both of them. For all runs, we show the likelihood of each model based on the validation dataset and the values of β_i parameter. The β_i parameter quantifies the magnitude of the effect of each variable on genetic covariance (small values correspond to a strong decreasing of the genetic covariance with increasing geographical or environmental distance, *i.e*. small values indicate an important effect of such variable). The most likely model for each comparison is in bold.

		Likelihood and (β_i) for each model		
		G	E	G + E
Environmental variable	$CLIM_{DIS}$	-469.43 **(β_G = 14.95)**	-471.44 (β_E = 25.44)	-469.78 (β_G = 25.50) (β_E = 21.31)
	$ELEV_{DIS}$	-460.11 **(β_G = 23.01)**	-461.24 (β_E = 31.08)	-460.63 (β_G = 17.45) (β_E = 31.11)

higher than that shown for *Ramburiella hispanica* at a similar spatial scale (global F_{ST} ~ 0.02), two species presenting a patchy distribution restricted to highly isolated and fragmented habitats (Ortego *et al*. 2010, 2012, 2015a). The genetic structure found in the Pyrenean Morales grasshopper is in accordance with isolation driven by geographical factors, in which the presence of deep barriers disrupting gene flow between populations primarily explained levels of genetic differentiation (Lanier *et al*. 2015). Despite climate warming during LIG and present has probably led to upward altitudinal shifts in the species distribution, its populations have presented a continuous distribution and exhibited a similar connectivity during both cold and warm periods characterizing the last 120 Kya. Our niche models suggest that only a few contemporary populations probably went extinct during the LGM, but this might have had a little impact on global patterns of genetic structure if nearby populations, with a similar genetic makeup, colonized present-day suitable habitat patches. The absence of the species in areas identified as climatically suitable and stable during the last 120 Kya according to our niche models (*i.e.* large areas in the peripheral northern and western portion of the Pyrenees) could be linked to historical events (such as extinctions) predating the temporal scale addressed in our study (Rizzo *et al*. 2013). It could also depend on constraints of our climate models not capturing other important abiotic or biotic interactions that may contribute to the distribution of the species in those areas (Hampe 2004). Thus, our analyses suggest that the strong genetic structure found across the species distribution range has not arisen as consequence of long-term isolation driven by Pleistocene climatic oscillations that shaped range-wide patterns of genetic structure in many other taxa from temperate regions (Hewitt 2000, 2004; Knowles & Richards 2005).

Our different landscape genetic approaches confirmed that neutral divergence resulted from the isolating effects of topography primarily drove the deep patterns of population genetic differentiation observed in the Morales grasshopper. The resistance model based on topographic complexity was the best fit to our data, indicating that physical features defining the abrupt landscape characterizing the Pyrenees shaped genetic differentiation. In particular, deep valleys with a north-south orientation and slopes that are generally steep and create canyons and ridges on the landscape crisscross the central and eastern portion of the Pyrenees inhabited by Morales grasshopper. Hence, these topographic features could become impassable barriers and restrict gene flow as has been previously documented for other species with limited dispersal capacities and inhabiting regions of remarkable topographical roughness (Murphy *et al*. 2010; Wang 2012; Castillo *et al*. 2014; Benham & Witt 2016). The remaining analyzed landscape factors, such as resistance-based distances informed by current and past climate niche models, did not show either a significant association with genetic differentiation after controlling for

the influence of topographic complexity or geographical distances. The lower explanatory power of CNM-based resistance distances may be related with the fact that they are not capturing well microhabitat structure, which has been found to be highly relevant in determining the distribution and demography of grasshoppers (Reinhardt *et al*. 2005; San Martín y Gómez & Van Dyck 2012; Keller *et al*. 2013; Gauffre *et al*. 2015). The short generation time of the studied species (= 1 year) may have also resulted in contemporary patterns of genetic differentiation are not capturing the genetic signal left by dispersal routes defined by habitat corridors during the past. This fact contrasts with patterns found in species with longer generation times and that are likely to show a time lag in their response to changing climatic conditions (Ortego *et al*. 2015b; He *et al*. 2013).

We found no support for ecological divergence and local adaptation processes have contributed to population genetic differentiation and the three different employed approaches (MMRR analysis using dissimilarity matrices, dbRDA analysis using raw variables and Bayesian inference) confirmed the consistency of this result. We did not find support either for altitude as an isolating mechanism despite elevation gradients have been previously found to be a significant driver of genetic and phenotypic variation in grasshoppers and many other organisms (Roff & Mousseau 2005; Laiolo *et al*. 2013). Our results contrasts with other studies that have documented an important role of environment on genetic differentiation in many taxa (Shaffer & Wolf 2013; Sexton *et al*. 2014), including some insects such as grasshoppers (Grace *et al*. 2010; Hernández-Teixidor *et al*. 2014), walking sticks (Nosil *et al*. 2008) or beetles (Funk *et al*. 2011). This discrepancy may be in part due to these studies considered species exhibiting narrow feeding preferences and analyzed ecological dissimilarity in terms of host-plant associations, an aspect that may have a higher impact on genetic divergence than the climate or elevation gradients considered in our study (Nosil *et al*. 2005; Grace *et al*. 2010; Funk *et al*. 2011; Hernández-Teixidor *et al*. 2014). The meta-analysis by (Sexton *et al*. 2014) showed that isolation-by-ecology is more frequent than IBD in insects, particularly in species with strong patterns of genetic structure. Considering the high degree of genetic differentiation among our study populations, we can discard the hypothesis that a high level of gene flow has counteracted the potential disruption of gene flow driven by local adaptation processes mediated by environmental heterogeneity (Nosil 2012). Hence, the lack of effects of environment on gene flow could be due to different reasons, including low environmental variation among sampling sites (Wang *et al*. 2013) or local adaptation driven by other ecological factors not considered in this study (*e.g.* distinct host plants or habitat structure; Grace *et al*. 2010).

CONCLUSIONS

This study emphasizes the importance of examining jointly different scenarios of population isolation to understand their contribution to the spatial distribution of genetic variation across a species range. Our analyses evidence the importance of topographic complexity in determining patterns of genetic differentiation, indicating that limited dispersal and drift, due to scarce population connectivity, is shaping the genetic structure found in our study system (*e.g*. Knowles & Richards 2005). Further research harnessing high-throughput sequencing will provide a better understanding about the potential association between loci under selection and different ecological factors, which may help to identify genomic regions involved in local adaptation processes (Wang & Bradburd 2014, Manthey & Moyle 2015). Exploring the relationship between environmental features and genetic and phenotypic patterns of variation could also provide insights about the potential interplay of evolutionary and ecological processes in shaping range-wide patterns of genetic differentiation (Wang & Summers 2010; Sistrom *et al*. 2012).

Acknowledgements

We wish to thank Bernard Defaut for providing valuable information about sampling locations. Two anonymous referees provided useful and valuable comments on an earlier draft of this manuscript. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). VN was supported by a FPI pre-doctoral scholarship (BES-2012-053741) from Ministerio de Economía y Competitividad. JO was supported by Severo Ochoa (SEV-2012-0262) and Ramón y Cajal (RYC-2013-12501) research fellowships. This work received financial support from research grants CGL2011-25053 (Ministerio de Ciencia e Innovación and European Social Fund), POII10-0197-0167, PEII-2014-023-P (Junta de Comunidades de Castilla-La Mancha and European Social Fund) and UNCM08-1E-018 (European Regional Development Fund).

Availability of data and materials

Nuclear microsatellite data are available in the LabArchives repository (*http://dx.doi.org/10.6070/H4QC01JB*). All other data supporting the results of this article are included within the article and its supplementary material.

REFERENCES

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692-4693.

Alvarado-Serrano DF, Knowles LL (2014) Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular Ecology Resources*, 14, 233-248.

Anderson RP, Raza A (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, 37, 1378-1393.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, 32, 818-830.

Basiita RK, Bruggemann JH, Cai N, *et al*. (2016) Erratum to: Microsatellite records for volume 7, issue 4. *Conservation Genetic Resources,* 8, 85-87.

Benham PM, Witt CC (2016) The dual role of Andean topography in primary divergence: functional and neutral variation among populations of the hummingbird, *Metallura tyrianthina*. *BMC Evolutionary Biology*, 16, 22.

Botta F, Eriksen C, Fontaine MC, Guillot G (2015) Enhanced computational methods for quantifying the effect of geographic and environmental isolation on genetic differentiation. *Methods in Ecology and Evolution*, 6, 1270-1277.

Braconnot P, Otto-Bliesner B, Harrison S*, et al.* (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, 261-277.

Bradburd GS, Ralph PL, Coop GM (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, 67, 3258-3273.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785-789.

Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology*, 23, 843-856.

Cavalli-Sforza L, Edwards AWF (1967) Phylogenetic analyses: models and estimation procedures. *Evolution*, 21, 550-570.

Collins WD, Bitz CM, Blackmon ML*, et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, 19, 2122-2143.

Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, 15, 49-62.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24, 621-631.

Chapuis MP, Lecoq M, Michalakis Y*, et al.* (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology*, 17, 3640-3653.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (*Caelifera*, *Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Erst PJ (2015) GEOGRAPHIC DISTANCE MATRIX GENERATOR, version 1.2.3. American Museum of Natural History, Center for Biodiversity and Conservation. *http://biodiversityinformatics.amnh.org/open_source/gdmg*

Edwards DL, Keogh JS, Knowles LL (2012) Effects of vicariant barriers, habitat stability, population isolation and environmental features on species divergence in the south-western Australian coastal reptile community. *Molecular Ecology*, 21, 3809-3822.

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330-342.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.

Ferrer ES, García-Navas V, Bueno-Enciso J*, et al.* (2016) The influence of landscape configuration and environment on population genetic structure in a sedentary passerine: insights from loci located in different genomic regions. *Journal of Evolutionary Biology*, 29, 205-219.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, 25, 104-120.

Funk DJ, Egan SP, Nosil P (2011) Isolation by adaptation in Neochlamisus leaf beetles: host-related selection promotes neutral genomic divergence. *Molecular Ecology*, 20, 4671-4682.

Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3209-3213.

García-Barros E, Gurrea P, Luciañez MJ*, et al.* (2002) Parsimony analysis of endemicity and its application to animal and plant geographical distributions in the Ibero-Balearic region (western Mediterranean). *Journal of Biogeography*, 29, 109-124.

Gauffre B, Mallez S, Chapuis M-P*, et al.* (2015) Spatial heterogeneity in landscape structure influences dispersal and genetic structure: empirical evidence from a grasshopper in an agricultural landscape. *Molecular Ecology*, 24, 1713-1728.

Grace T, Wisely SM, Brown SJ, Dowell FE, 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, 866-878.

Guo SW, Thompson EA (1992) A Monte-Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics*, 51, 1111-1126.

Hampe A (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, 13, 469-471.

He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, 67, 3386-3402.

Hernández-Teixidor D, López H, Nogales M*, et al.* (2014) Genetic, morphological, and dietary changes associated with novel habitat colonisation in the Canary Island endemic grasshopper *Acrostira bellamyi*. *Ecological Entomology*, 39, 703-715.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 183-195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801-1806.

Jenkins DG, Carey M, Czerniewska J*, et al.* (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, 33, 315-320.

Jenness JS (2004) Calculating landscape surface area from digital elevation models. *Wildlife Society Bulletin*, 32, 829-839.

Jenness J (2013) DEM SURFACE TOOLS. Jenness Enterprises. *http://www.jennessent.com/arcgis/surface_area.htm*

Jha S, Kremen C (2013) Urban land use limits regional bumble bee gene flow. *Molecular Ecology*, 22, 2483-2495.

Jiménez-Sanchez M, Rodríguez-Rodríguez L, García-Ruiz JM*, et al.* (2013) A review of glacial geomorphology and chronology in northern Spain: timing and regional variability during the last glacial cycle. *Geomorphology*, 196, 50-64.

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.

Keller D, Holderegger R, van Strien MJ (2013) Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology*, 22, 2467-2482.

Kierepka EM, Latch EK (2015) Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources*, 15, 512-525.

Knowles LL, Richards CL (2005) Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Molecular Ecology*, 14, 4023-4032.

Laiolo P, Illera JC, Obeso JR (2013) Local climate determines intra- and interspecific variation in sexual size dimorphism in mountain grasshopper communities. *Journal of Evolutionary Biology*, 26, 2171-2183.

Langella O (1999) POPULATIONS 1.2.31 software. *http://bioinformatics.org/populations/*

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, 24, 3688-3705.

Lee C-R, Mitchell-Olds T (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, 20, 4631-4642.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69, 1-24.

Llucià-Pomares D (2002) *Revision of the Orthoptera (Insecta) of Catalonia (Spain). Monografias S.E.A.*, vol. 7, Sociedad Aragonesa Aragonesa, Zaragoza, Spain.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, 28, 614-621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18, 189-197.

Manthey JD, Moyle RG (2015) Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*, 24, 3628-3638.

Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, 68, 2833-2846.

McRae BH (2006) Isolation by resistance. *Evolution*, 60, 1551-1561.

McRae BH, 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, 19885-19890.

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, 89, 2712-2724.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, 21, 2839-2846.

Munshi-South J, Zolnik CP, Harris S (2016) Population genomics of the Anthropocene: urbanization is negatively associated with genome-wide variation in whitefooted mouse populations. *Evolutionary Applications*, 9, 546–564.

Murphy MA, Dezzani R, Pilliod DS, Storfer A (2010) Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, 19, 3634-3649.

Nosil P (2012) *Ecological speciation*. Oxford University Press, New York, USA.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, 62, 316-336.

Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, 59, 705-719.

Oksanen J, Blanchet FG, Kindt R, *et al*. (2015) VEGAN: community ecology package. R Package Version 2.3–1. *http://r-forge.r-project.org/projects/vegan*

Ortego J, Aguirre MP, Cordero PJ (2010) Population genetics of *Mioscirtus wagneri,* a grasshopper showing a highly fragmented distribution. *Molecular Ecology*, 19, 472-483.

Ortego J, Aguirre MP, Cordero PJ (2012) Genetic and morphological divergence at different spatiotemporal scales in the grasshopper *Mioscirtus wagneri* (Orthoptera: Acrididae). *Journal of Insect Conservation*, 16, 103-110.

Ortego J, Aguirre MP, Noguerales V, Cordero PJ (2015a) Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications*, 8, 621-632.

Ortego J, Gugger PF, Sork VL (2015b) Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography*, 42, 328-338.

Otto-Bliesner BL, Marsha SJ, Overpeck JT*, et al.* (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science*, 311, 1751-1753.

Papadopoulou A, Knowles LL (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, 1501-1517.

Parks DH, Porter M, Churcher S, *et al*. (2009) GENGIS: a geospatial information system for genomic data. *Genome Re*search, 19, 1896–1904.

Pflüeger FJ, Balkenhol N (2014) A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology*, 23, 2146-2156.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

Phillips SJ, Dudik M (2008) Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

R Core Team (2015) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org/*

Reinhardt K, Kohler G, Maas S, Detzel P (2005) Low dispersal ability and habitat specificity promote extinctions in rare but not in widespread species: the Orthoptera of Germany. *Ecography*, 28, 593-602.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223-225.

Rizzo V, Comas J, Fadrique F, Fresneda J, Ribera I (2013) Early Pliocene range expansion of a clade of subterranean Pyrenean beetles. *Journal of Biogeography*, 40, 1861-1873.

Roff DA, Mousseau T (2005) The evolution of the phenotypic covariance matrix: evidence for selection and drift in Melanoplus. *Journal of Evolutionary Biology*, 18, 1104-1114.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137-138.

Ruiz-González A, Cushman SA, Madeira JM, Randi E, Gómez-Moliner BJ (2015) Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular Ecology*, 24, 5110-5129.

San Martín y Gómez G, Van Dyck H (2012) Ecotypic differentiation between urban and rural populations of the grasshopper *Chorthippus brunneus* relative to climate and habitat fragmentation. *Oecologia*, 169, 125-133.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, 68, 1-15.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, 16, 940-950.

Sistrom M, Edwards DL, Donnellan S, Hutchinson M (2012) Morphological differentiation correlates with ecological but not with genetic divergence in a Gehyra gecko. *Journal of Evolutionary Biology*, 25, 647-660.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, 47, 264-279.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution*, 64, 295-315.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144, 389-399.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535-538.

Velo-Antón G, Parra JL, Parra-Olea G, Zamudio KR (2013) Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, 22, 3261-3278.

Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, 18, 1033-1041.

Wang IJ (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, 3403-3411.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, 23, 5649-5662.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, 16, 175-182.

Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, 19, 447-458.

Wright S (1943) Isolation by distance. *Genetics*, 28, 114-138.

Yannic G, Pellissier L, Ortego J*, et al.* (2014) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, 4, 132-137.

SUPPLEMENTARY MATERIAL

Table S1 – Genetic differentiation between eleven populations of the Pyrenean Morales grasshopper (*Chorthippus saulcyi moralesi*). We present pairwise F_{ST} values below the diagonal and F_{ST}NA values corrected for null alleles above the diagonal. FREENA software used to calculate F_{ST}NA values does not implement statistical tests of population genetic differentiation and these were only performed for F_{ST} values in ARLEQUIN.F_{ST} values in bold are statistically significant after sequential Bonferroni correction. Population codes as in Table 1.

	TOR	NER	SAR	CHI	ASP	BOI	PER	CAR	ERR	CRE	RAS
TOR	-	0.014	0.065	0.080	0.087	0.111	0.081	0.075	0.107	0.104	0.096
NER	0.021	-	0.057	0.059	0.076	0.095	0.066	0.076	0.103	0.103	0.098
SAR	0.070	0.074	-	0.104	0.114	0.127	0.083	0.105	0.136	0.139	0.127
CHI	0.106	0.083	0.140	-	0.043	0.114	0.105	0.094	0.136	0.141	0.120
ASP	0.116	0.108	0.143	0.072	-	0.105	0.110	0.103	0.148	0.141	0.134
BOI	0.131	0.109	0.135	0.133	0.109	-	0.055	0.085	0.109	0.093	0.112
PER	0.101	0.083	0.085	0.134	0.137	0.078	-	0.065	0.069	0.073	0.106
CAR	0.102	0.121	0.142	0.135	0.152	0.087	0.071	-	0.037	0.041	0.083
ERR	0.132	0.143	0.181	0.188	0.197	0.120	0.084	0.052	-	0.025	0.099
CRE	0.134	0.155	0.200	0.207	0.216	0.105	0.103	0.081	0.037	-	0.075
RAS	0.124	0.131	0.164	0.170	0.196	0.136	0.128	0.094	0.110	0.098	-

Table S2 – Results of Mantel tests analyzing the relationship between the different distance matrices (predictors) used to evaluate the factors associated with population genetic differentiation in the Pyrenean Morales grasshopper (*Chorthippus saulcyi moralesi*): elevation ($ELEV_{DIS}$) and climatic ($CLIM_{DIS}$) dissimilarity and five isolation by resistance (IBR) scenarios: IBD, isolation by distance (*i.e.* equal resistance to all pixel values, equivalent to geographical distance); IBR_{TC}, topographic complexity; $IBR_{CURRENT}$, current habitat suitability; IBR_{LGM}, Last Glacial Maximum habitat suitability and IBR_{LIG}, Last Interglacial habitat suitability. We present Mantel *r* values below the diagonal and *P*-values for each comparison above the diagonal. Significant *P*-values after sequential Bonferroni correction in bold.

	IBD	IBR_{TC}	$IBR_{CURRENT}$	IBR_{LGM}	IBR_{LIG}	$ELEV_{DIS}$	$CLIM_{DIS}$
IBD	-	0.001	0.001	0.001	0.001	0.875	0.001
IBR_{TC}	0.990	-	0.001	0.001	0.001	0.877	0.001
$IBR_{CURRENT}$	0.924	0.930	-	0.001	0.001	0.946	0.001
IBR_{LGM}	0.620	0.625	0.629	-	0.002	0.648	0.003
IBR_{LIG}	0.976	0.977	0.962	0.584	-	0.882	0.001
$ELEV_{DIS}$	-0.136	-0.139	-0.184	-0.072	-0.147	-	0.004
$CLIM_{DIS}$	0.641	0.642	0.691	0.474	0.669	0.479	-

Table S3 – Results of principal component analysis (PCA) applied to the values of the 19 present day bioclimatic variables obtained from the WorldClim dataset. We report factor loadings for the first three principal components (PC) and the 19 bioclimatic variables. Bold type indicates variables with factor loadings higher than 0.7.

	PC1	PC2	PC3
Bio1	**0.920**	0.207	0.268
Bio2	0.271	**0.952**	0.059
Bio3	-0.058	**0.885**	0.328
Bio4	0.656	0.398	-0.522
Bio5	**0.909**	0.358	0.146
Bio6	**0.920**	0.026	0.323
Bio7	0.538	**0.788**	-0.213
Bio8	**0.901**	0.091	-0.157
Bio9	0.246	0.213	**0.800**
Bio10	**0.939**	0.218	0.196
Bio11	**0.911**	0.136	0.333
Bio12	**-0.899**	-0.412	0.066
Bio13	**-0.814**	-0.512	0.145
Bio14	**-0.922**	-0.336	-0.109
Bio15	**0.874**	-0.020	-0.005
Bio16	**-0.860**	-0.449	0.102
Bio17	**-0.924**	-0.349	0.013
Bio18	**-0.816**	-0.475	-0.153
Bio19	**-0.898**	-0.232	0.318

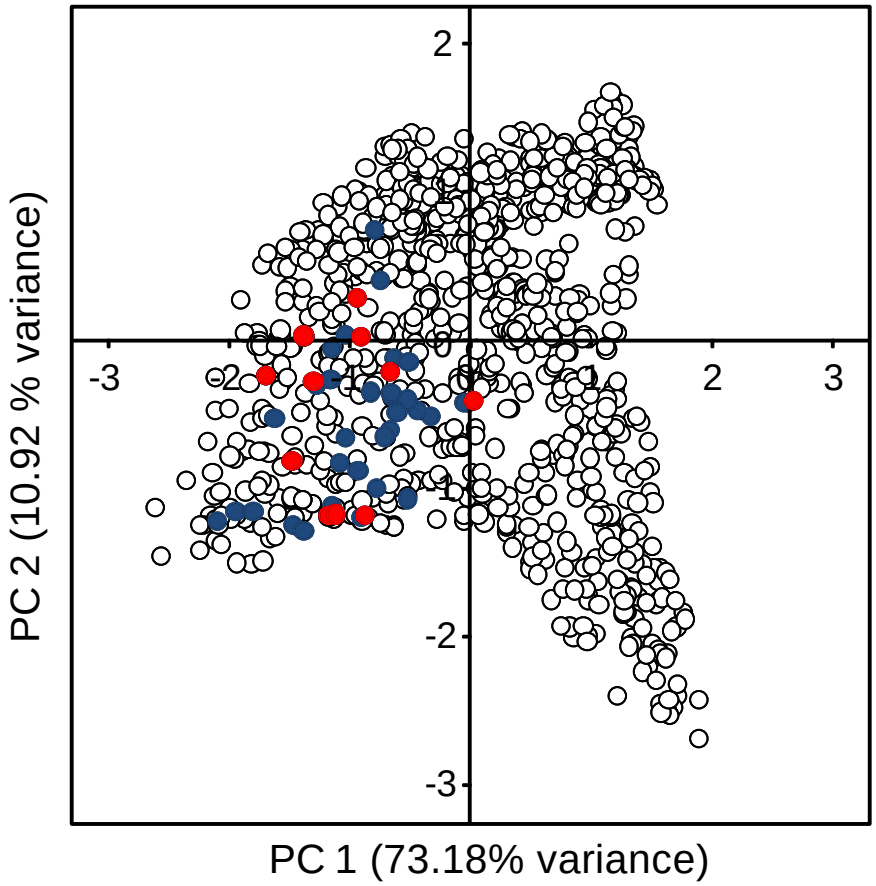


Figure S1 – The position in the environmental space (first two principal components of a PCA based on the 19 bioclimatic variables from the WorldClim dataset) of all known populations of the Pyrenean Morales grasshopper (*Chorthippus saulcyi moralesi*). Sampling sites used in genetic analyses, occurrence points used to build the Climate Niche Models (CNM) in MAXENT and the 1 000 random points used to perform the PCA are represented with red, blue and white open dots, respectively. Population codes as in Table 1.

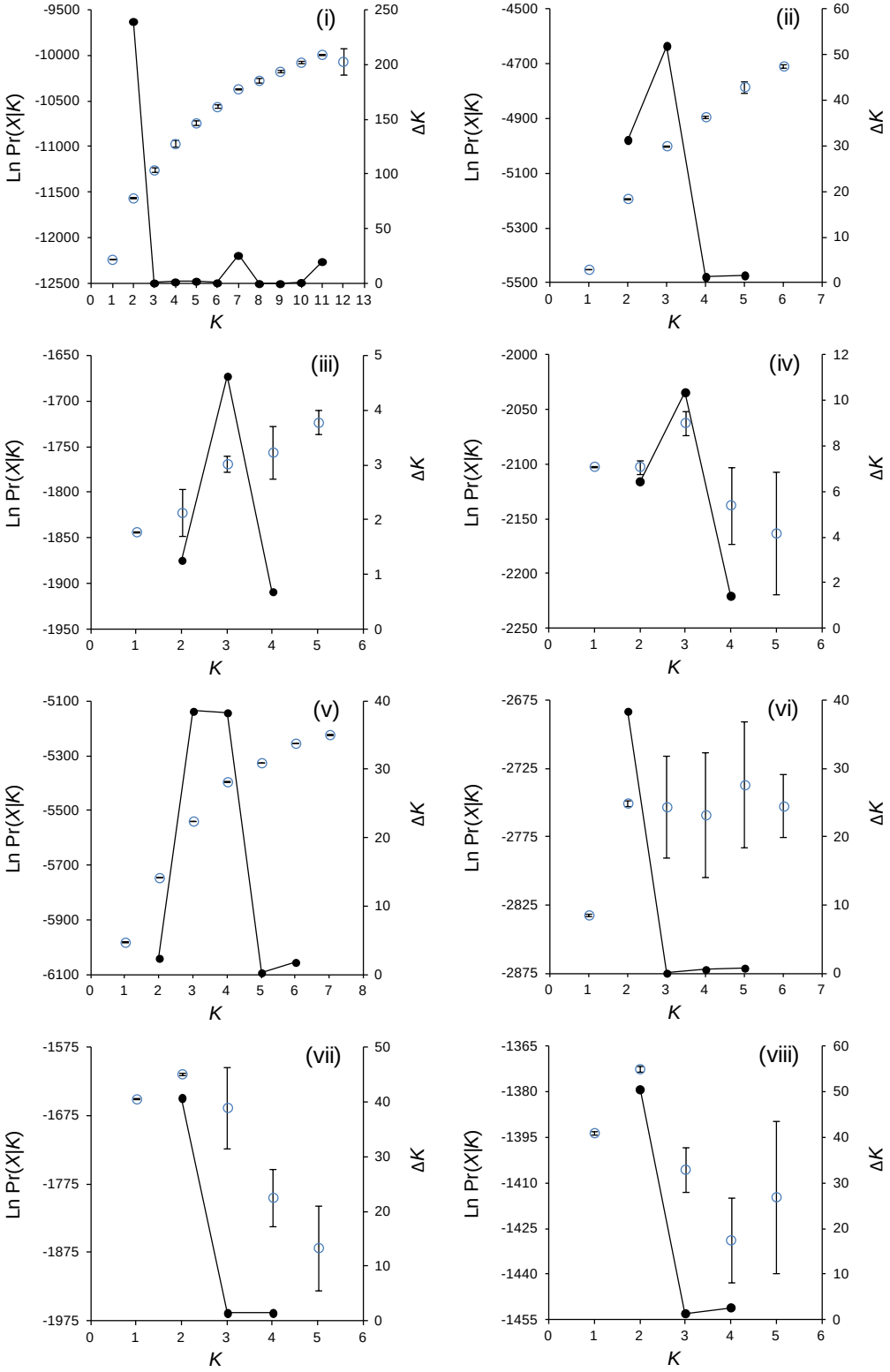


Figure S2 – Results of Bayesian clustering analyses in STRUCTURE to determine the best-supported number of clusters in hierarchical analyses including (i) all populations or (ii - viii) different subsets of them (see Fig. 2b for correspondence between panel codes and population codes). Each plot shows the mean (± SD) log probability of the data [Ln Pr (*X*|*K*)] over 10 runs (left axis, open dots and error bars) for each *K*-value. The magnitude of Δ*K* as a function of *K* determines the best-supported number of clusters in STRUCTURE analyses (right axis, black dots and continuous line).

CAPÍTULO VI

The role of environment and core-margin effects on range-wide phenotypic variation of a montane grasshopper

VÍCTOR NOGUERALES, VICENTE GARCÍA-NAVAS, PEDRO J. CORDERO & JOAQUÍN ORTEGO

Journal of Evolutionary Biology, 29, 2129-2142 (2016)

The role of environment and core-margin effects on range-wide phenotypic variation of a montane grasshopper

Abstract

The integration of genetic information with ecological and phenotypic data constitutes an effective approach to gain insight into the mechanisms determining interpopulation variability and the evolutionary processes underlying local adaptation and incipient speciation. Here, we use the Pyrenean Morales grasshopper (*Chorthippus saulcyi moralesi*) as study system to (i) analyze the relative role of genetic drift and selection on range-wide patterns of phenotypic differentiation and (ii) identify the potential selective agents (environment, elevation) responsible for variation. We also test the hypothesis that (iii) the development of dispersal-related traits is associated with different parameters related to population persistence/turnover, including habitat suitability stability over the last 120 000 years, distance to the species distribution core, and population genetic variability. Our results indicate that selection shaped phenotypic differentiation across all the studied morphological traits (body size, forewing length and shape). Subsequent analyses revealed that among-population differentiation in forewing length was significantly explained by a temperature gradient, suggesting an adaptive response to thermoregulation or flight performance under contrasting temperature regimes. We found support for our hypothesis predicting a positive association between the distance to the species distribution core and the development of dispersal-related morphology, which suggests increased dispersal capability in populations located at range edges that, in turn, exhibit lower levels of genetic variability. Overall, our results indicate that range-wide patterns of phenotypic variation are partially explained by adaptation in response to local environmental conditions and differences in habitat persistence between core and peripheral populations.

INTRODUCTION

Disentangling the relative contribution of natural selection and random genetic drift on phenotypic diversity is of great importance for understanding the mechanisms shaping intraspecific variation and incipient speciation processes (Leinonen *et al*. 2006, 2008; Oneal & Knowles 2013; Bertrand *et al*. 2016). Phenotypic divergence can arise as consequence of purely stochastic processes such as bottlenecks or founder effects, which can lead to genetic drift in genomic regions involved in trait expression (Lande 1976; Zhan *et al*. 2005). However, numerous studies have found local adaptation as the main evolutionary force responsible for phenotypic differentiation in natural populations (Kekkonen *et al*. 2012; Oneal & Knowles 2013; Ortego *et al*. 2015b; see reviews in Merilä & Crnokrak 2001; Leinonen *et al*. 2008) and considerable research has been devoted to identify the ecological conditions

under which this phenomenon arises (Nosil & Crespi 2004; Räsänen & Hendry 2008; Schluter 2000). Theoretical models have shown that local adaptation can occur even in the face of high gene flow when environmental heterogeneity results in spatially and temporally contrasting selection pressures (Merilä & Crnokrak 2001; Räsänen & Hendry 2008; see also Edelaar *et al*. 2012; Edelaar & Bolnick 2012 and references therein). Although several studies have empirically demonstrated that adaptive differentiation is possible in presence of realized gene flow (García-Navas *et al*. 2014; Egan *et al*. 2015), it is more likely to occur when realized dispersal does not counteract the effects of local selection (Nosil & Crespi 2004; Räsänen & Hendry 2008). Even though, the different mechanisms potentially resulting in phenotypic divergence are not mutually exclusive and, in fact, both deterministic (local adaptation) and stochastic (random genetic drift) processes can act in concert when shaping phenotypic variation in natural populations (Fornel *et al*. 2010).

Numerous organisms exhibit considerable morphological variation across their distribution range, which suggests that local adaptation processes in response to spatially varying evolutionary pressures are at play (Tregenza *et al*. 2000; Levy & Nufio 2015). Accordingly, a broad plethora of selective agents has been identified as drivers of phenotypic divergence and life-history trait variation in natural populations, including environmental factors such as temperature (San Martín y Gómez & Van Dyck 2012; Wojcieszek & Simmons 2012; Laiolo & Obeso 2015) and elevation (Berner *et al*. 2004; Keller *et al*. 2013; Laiolo *et al*. 2013). Landscape dynamics and spatiotemporal changes in habitat suitability and fragmentation can also have a considerable impact on phenotypic variation (Thomas *et al*. 2001; Hanski *et al*. 2004; Dytham 2009; Berggren *et al*. 2012). In this sense, theoretical and empirical studies suggest that populations at expanding range margins (*i.e.* peripheral populations) or unstable habitats should experience selection towards phenotypes with a higher dispersal capability (Hughes *et al*. 2007; Dytham 2009; Hill *et al*. 2011) due to increased local extinction rates and reduced mate availability associated with low habitat persistence and small population sizes (Denno *et al*. 1991; Denno 1994). Beyond selection, colonizers are also likely to constitute an unrepresentative sample of individuals from the origin population. On this regard, several studies have shown that dispersing individuals are larger, which may be due to individuals with such phenotype are more likely to be successful immigrants (Zera & Denno 1997; Debeffe *et al*. 2012; San Martín y Gómez & Van Dyck 2012). This dispersive morph can get fixed quickly due to the strong genetic drift characterizing recently founded populations (O'Riain *et al*. 1996; Hampe & Petit 2005; Calabuig *et al*. 2010; Fountain *et al*. 2016). Most of the studies addressing this question have focused on recent range expansions (Hill *et al*. 2011) or the persistence of habitats at short temporal scales (Heidinger *et al*. 2010; Berggren *et al*.

2012; see however Denno *et al*. 1991), but the effects of long-term habitat stability (*e.g.* related to Quaternary climate fluctuations and range shifts; Hewitt 2000) on dispersal-related traits has received little attention.

In this study, we use the Pyrenean Morales grasshopper *Chorthippus saulcyi moralesi* Uvarov 1954 as model system to investigate the relative role of genetic drift and local adaptation on phenotypic trait variation across range-wide distributed populations of the species (Spitze 1993; Brommer 2011; Leinonen *et al*. 2013). The Pyrenean Morales grasshopper is a brachypterous and small body-sized (males: 14.0-18.0 mm; females: 16.5-21.5 mm) gomphorecine (Orthoptera: Acrididae) belonging to the *Chorthippus* group *binotatus* species complex (Defaut 2011). This grasshopper is a narrow endemism whose patchy distribution is restricted to the central and eastern portion of the Pyrenees (Llucià-Pomares 2002; Defaut 2011). It inhabits different montane environments, from mesophilic shrubby habitats to subalpine grasslands (Llucià-Pomares 2002) at altitudes above 1 000 meters (from 1 100 up to 2 400 m.a.s.l.; Defaut 2011). The intricate orography of this area and the spatial configuration of the mountains that conform the Pyrenees are likely to be responsible for the strong spatial genetic structure observed in populations of this species (Noguerales *et al*. 2016). Thus, both the abiotic/geographic framework and the early stage of population genetic differentiation represented in this species provides a well-suited scenario to investigate the evolutionary forces shaping phenotypic variation across an entire species distribution range (Storz 2002; Hangartner *et al*. 2012). Specifically, we first tested the null hypothesis that phenotypic differentiation arises from the effect of genetic drift, the main evolutionary force underlying the deep genetic structure observed among populations of the study species (Noguerales *et al*. 2016). Second, we used distance-based redundancy analyses (dbRDA; Legendre & Anderson 1999) to analyze the association between phenotypic differentiation and environmental variation in order to test whether morphological trait variation is shaped by local adaptation in response to spatially varying selection pressures (Defaveri & Merilä 2013). Finally, we tested the hypothesis that the development of dispersal-related traits is associated with different parameters related with population persistence and turnover, including habitat suitability stability over the last 120 Kya, distance to the species distribution core, and population genetic variability.

MATERIAL AND METHODS

SAMPLING AND STUDY AREA

Between 2012 and 2014, we collected 202 individuals of *C. saulcyi moralesi* from 11 populations dispersed across the Pyrenees (Spain, France and Andorra). According to our own surveys and occurrence records available in the literature (Llucià-Pomares 2002; Defaut 2011), our sampling covered the entire distribution range of the taxon (~7 000 km^2; Fig. S1). Population codes and more information on sampling sites are presented in Table S1.

GENETIC DATA AND ANALYSIS

Genomic DNA from muscle tissue of the hind femur was extracted using a salt extraction protocol (Aljanabi & Martinez 1997). All individuals were genotyped at 18 polymorphic microsatellites markers whose characteristics and PCR cycling conditions are detailed in Basiita *et al*. (2016). We performed PCR amplifications and genotyping following the procedure described in Ortego *et al*. (2015a). We tested for deviations from Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and the presence of null alleles as described in Noguerales *et al*. (2016). Two loci (Cbin08 and Cbin36) were discarded from all downstream analyses because of HW disequilibrium in all populations and the presence of null alleles (Noguerales *et al*. 2016). We did not find evidence for linkage disequilibrium between any pair of loci in any sampling population after sequential Bonferroni corrections (Rice 1989).

We quantified neutral genetic differentiation among populations calculating pairwise F_{ST}-values and testing their significance with Fisher´s exact tests after 10 000 permutations using ARLEQUIN (Excoffier & Lischer 2010). We calculated global F_{ST}-values over all populations and 95% confidence intervals (95% CI) by 10 000 bootstrapping replicates over loci using FSTAT (Goudet 1995). Due to the high frequency of null alleles in Orthoptera, we applied the so-called ENA-method to calculate global and pairwise F_{ST}-values corrected for null alleles (F_{ST}NA) using the software FREENA (Chapuis & Estoup 2007; *e.g.* Ortego *et al*. 2015a). FREENA was also used to obtain 95% CI by 10 000 bootstrapping replicates over loci.

MORPHOLOGICAL DATA AND PHENOTYPIC DIVERGENCE

A total of 167 adult specimens (94 males and 73 females) were selected for linear and geometric morphometric analyses of body size and forewing length and shape, allowing us to obtain measurements for about 8 individuals (range: 5-9) of each sex per population. However, in one population (Creueta) no female could be captured and this locality was excluded from subsequent analyses for this sex. Body size strongly correlates with life-history and fitness-related traits and, thus, it constitutes a key character and target of selection (Blanckenhorn 2000; Whitman 2008; Kanuch *et al*. 2012). Forewings are strongly sclerotized structures involved in sound production and courtship rituals in Orthoptera (Petit *et al*. 2006; Routtu *et al*. 2007; Klingenberg *et al*. 2010). Forewing length is considered a good proxy of dispersal ability (Thomas *et al*. 2001; Simmons & Thomas 2004; Heidinger *et al*. 2010). Thus, this trait is expected to evolve under both natural and sexual selection (Routtu *et al*. 2007; Klingenberg *et al*. 2010).

We took linear measurements of left femur and left forewing length to the nearest 0.1 mm using a stereoscopic microscope LEICA S8 APO (Leica Microsystems GmbH, Wetzlar, Germany) and the Leica LAS image analysis software v.2.8.1. We used femur length as a proxy for body size since the total length of females varies substantially with the oviposition cycle (Hochkirch & Gröning 2008). Femur length was strongly correlated with structural body length excluding the abdomen (*i.e*. head + thorax) ($r^2 = 0.96$, $p < 0.001$) as it has been reported for a large number of Orthoptera species (*e.g*. Ortego *et al*. 2012; Laiolo *et al*. 2013; Eweleit & Reinhold 2014). We calculated an unbiased index of forewing size by expressing its length relative to femur length. We took digital images of forewings and digitized 10 homologous landmarks using TPSDIG (Rohlf 2015) in order to characterize shape variation for this trait (Fig. 1). Once all specimens were digitized, they were aligned and superimposed to a common coordinate system using a generalized Procrustes analyses (GPA; Rohlf & Slice 1990). Next, morphometric variation in forewing shape was assessed separately for each sex by a relative warps (RW) analysis on the adjusted landmark coordinates using the default weighting factor ($\alpha = 0$). Given that the components of variance are not differentially weighted by their bending energy, this analysis is equivalent to a principal component (PC) analysis (Zelditch *et al*. 2004). We retained the two first RW, which accounted for a high proportion of shape variation in both males (RW1: 57.51%; RW2: 19.60%) and females (RW1: 44.38%; RW2: 33.52%). RW scores (analogous to PC scores) were used in subsequent morphological analyses. GPA and RW analyses were conducted using TPSRELW (Rohlf 2015). Forewing shape variation was visually displayed using thin-plate spline diagrams as implemented

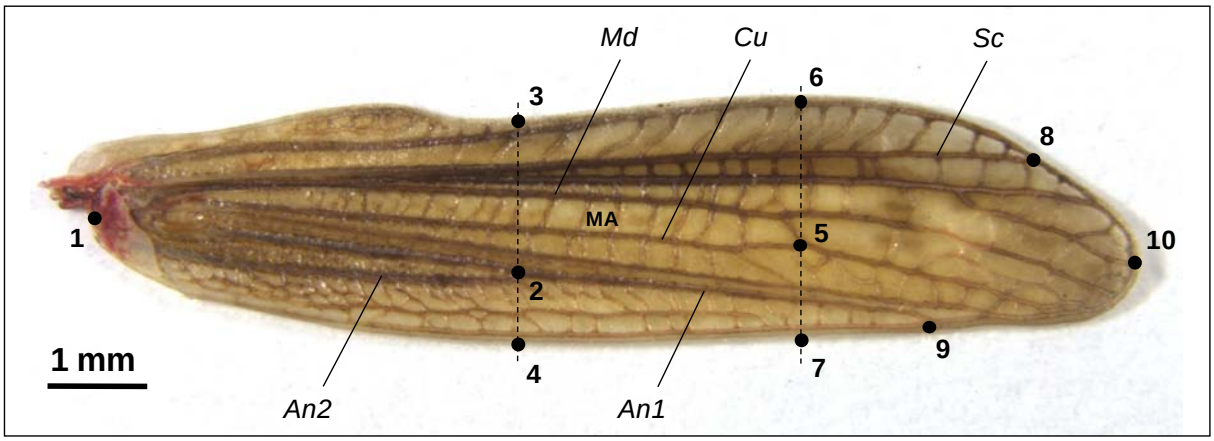


Figure 1 – Positions of the 10 landmarks used to characterize tegmina shape in the Pyrenean Morales grasshopper. The scale bar was used to standardize landmark distances to the same absolute scale across all images. We indicate the main traits (veins and areas) used to define the position of the landmarks: *Md*, median vein; *Cu*, cubitus vein; *Sc*, subcosta vein; *An1* and *An2*, anal 1 and anal 2 veins, respectively. MA, median area, delimited by media and cubitus veins (see Bethoux & Nel 2001; Petit *et al*. 2006).

in TPSSPLIN (Rohlf 2015). Finally, we conducted a Canonical Variates Analyses (CVA) to examine whether differences among populations in forewing shape were significant. We calculated Mahalanobis distances (D_2) between populations and tested their significance by permutation tests with 10 000 replicates. CVA of forewing shape was conducted using MORPHOJ v.1.05d (Klingenberg 2011). Differences among populations in femur length, forewing length, and wing length relative to femur length were analyzed using one-way ANOVAs.

We studied the processes underlying morphological differentiation by comparing phenotypic and neutral genetic differentiation. Morphological differentiation was determined using P_{ST} statistics (Spitze 1993; Brommer 2011; Leinonen *et al*. 2013). P_{ST} is used as a surrogate of Q_{ST} when estimating purely additive genetic variance and genotype-environment interactions is not straightforward. P_{ST} is statistically equivalent to F_{ST} for morphological characters (Brommer 2011). If the phenotypic differentiation is solely attributable to random genetic drift, P_{ST} and neutral F_{ST} should exhibit a positive correlation, whereas if they are uncoupled would indicate that the phenotype is plastic or controlled by selection. In the later scenario, P_{ST} could exhibit either smaller or higher values than neutral genetic differentiation that would suggest divergent and stabilizing selection, respectively (Merilä & Crnokrak 2001; McKay & Latta 2002; Leinonen *et al*. 2006; Whitlock 2008). Pairwise P_{ST}-values were calculated as:

$$P_{\mathrm{ST}} = \frac{\left(\frac{c}{h^2}\right) \sigma^2_{\mathrm{GB}}}{\left(\frac{c}{h^2}\right) \sigma^2_{\mathrm{GB}} + 2\,\sigma^2_{\mathrm{GW}}}$$

where the scalar *c* indicates the proportion of the total variance that is presumed to be due to additive genetic effects across populations, h^2 is the assumed additive genetic proportion of differences between individuals within populations (narrow sense "heritability"), and σ^2_{GB} and σ^2_{GW} are the between and within population variance components, respectively. The variation in *c* and h^2 parameters affects the estimate of P_{ST} and, consequently, the reliability of the comparisons between P_{ST} and F_{ST}. We considered that the proportion of phenotypic variance that is due to additive genetic effects is the same within and among populations (*i.e*. $c = h^2$), a biologically realistic assumption (Brommer 2011). The reported results were obtained using $h^2 = 0.5$, an estimate of heritability equal to the mean value previously calculated for several morphological traits in a close grasshopper species, *Chorthippus brunneus* (Butlin & Hewitt 1986; see also Klingenberg *et al*. 2010 for heritabilities of shape variables in Orthoptera forewings). However, we computed P_{ST}-values by varying the *c* and h^2 parameters (c/h^2 ratio: 0.1-2.0) to check if our conclusions hold even if the proportion of phenotypic variance across populations due to additive genetic effects (*c*) is much lower than heritability h^2 (sensitivity analysis; Wojcik *et al*. 2006). We calculated pairwise P_{ST}-values separately for females and males and for each morphological trait: femur length (FL), forewing length (FWL), forewing length relative to femur length (RFWL) and the two retained RW components summarizing information on variation in forewing shape. In addition, we calculated a weighted P_{ST} index of forewing shape variation (FWS) averaging P_{ST}-values obtained for the two RW components and using its respective explained variances as offset. We estimated the 95% CI for each P_{ST}-value by 10 000 bootstrap replicates using the BOOT package (Ripley 2016) as implemented in R 3.2.3 (R Core Team 2015).

We followed two complementary methods for comparing P_{ST} and F_{ST} according to the suggestions by Barley *et al*. (2015) (see also Lehtonen *et al*. 2009; Wojcieszek & Simmons 2012). At first, we tested the association between neutral genetic (F_{ST}) and phenotypic differentiation (P_{ST}) using one-tailed Mantel tests, in order to assess the role of genetic drift or local adaptation on morphological divergence (Lehtonen *et al*. 2009). A significant positive relationship would suggest that population phenotypic differentiation is mostly driven by neutral genetic drift. Conversely, the absence of association between genetic and phenotypic differentiation would suggest that morphological variation among populations is shaped by selective pressures favoring different phenotypes under certain

ecological conditions (Lehtonen *et al*. 2009; Ortego *et al*. 2015b). Secondarily, we compared 95% confidence intervals of global F_{ST}, F_{ST}NA and P_{ST} for each trait and sex to determine whether their mean values were significantly different. Additionally, we used one-tailed Mantel tests to analyze the potential association between phenotypic differentiation (P_{ST}) and pairwise Euclidean geographical distances in order to test if more isolated demes show a higher degree of morphological differentiation. All Mantel tests were conducted with 10 000 permutations using ZT software (Bonnet & Van de Peer 2002).

ENVIRONMENTAL EFFECTS ON PHENOTYPIC DIFFERENTIATION (P_{ST})

In order to investigate the potential effects of elevation and climate on phenotypic differentiation, we characterized the environmental space of each population using a principal component analysis (PCA) applied to the 19 present-day bioclimatic variables from the WorldClim dataset (Hijmans *et al*. 2005). Subsequently, we applied a 'Kaiser varimax' rotation to maximize differences in factor loadings on the principal components and make easier their interpretation (Norman & Streiner 2000). After applying rotation, PC scores for the three PC axes retained for subsequent analyses remained uncorrelated (Pearson correlations, PC1 and PC2: r = 0.04, P = 0.896; PC1 and PC3: r = -0.43, P = 0.188; PC2 and PC3: r = 0.40, P = 0.223). Bioclimatic variables were extracted from sampling sites, occurrence data from the extant literature (Llucià-Pomares 2002; Defaut 2011) and 1000 random points within the study area using ARCGIS 10.0 (ESRI, Redlands, CA, USA). This procedure allowed us to capture the environmental variation of the study area and avoid potential biases resulting from considering exclusively the environmental conditions of our sampling sites. Then, we obtained for each population the PC scores of the first three PC (eigenvalues > 1), which accounted for 63.43%, 20.57% and 8.48% of environmental variance, respectively. Additionally, we extracted the elevation (in meters) of each sampling site from a 90-m resolution digital elevation model obtained from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data, *http://srtm.csi.cgiar.org/*).

We tested the relationship between phenotypic differentiation and geography, environment and elevation using distance-based redundancy analyses (dbRDA) (Legendre & Anderson 1999). We performed dbRDA using the '*capscale*' function in the R package VEGAN (Oksanen *et al*. 2016). Phenotypic distance matrices (pairwise P_{ST}-values) were tested against the following predictors: (i) inter-population geographical distances (IBD) transformed by principal coordinates analyses (PCoA) using the '*pcnm*' function in the package VEGAN, (ii) elevation and (iii) population PC scores of the first

three axes from the PCA on bioclimatic data. We assessed the significance of the predictors using multivariate *F*-statistics with 9 999 permutations using the '*anova.cca*' function included in VEGAN. dbRDA analyses were conducted separately for males and females. Initially, we analyzed the relationship between the P_{ST} matrices and each variable separately (marginal test) and then we performed a partial dbRDA (conditional test) for each variable while controlling for the influence of geographic distance (fitted as covariate).

DISPERSAL-RELATED MORPHOLOGY IN RELATION TO CLIMATIC SUITABILITY STABILITY, DISTANCE TO THE DISTRIBUTION CORE AND GENETIC VARIABILITY

We modelled the distribution of the Pyrenean Morales grasshopper at different time periods to estimate the stability of climatically suitable habitats for this species during the last 120 Kya. First, we used MAXENT 3.3.3 (Phillips *et al*. 2006; Phillips & Dudik 2008) to model the present distribution of the species using the 19 bioclimatic variables available in WorldClim at 30 arc-sec resolution (Hijmans *et al*. 2005) and 47 occurrence points obtained from the extant literature and our own records (Llucià-Pomares 2002; Defaut 2011). We estimated the distribution of climatically suitable habitats for the study species in the Last Glacial Maximum (LGM, *c*. 21 Kya; CCSM3 model; Collins *et al*. 2006) and the Last Interglacial (LIG, *c*. 120 Kya; Otto-Bliesner *et al*. 2006) projecting contemporary species-climate relationships into these periods. Model and variable selection was performed as detailed in Noguerales *et al*. (2016). The final model was built with six bioclimatic variables: annual mean temperature (*Bio1*), mean temperature of the coldest quarter (*Bio11*), annual precipitation (*Bio12*), precipitation of the driest month (*Bio14*), precipitation seasonality (*Bio15*) and precipitation of the warmest quarter (*Bio18*). The final climate niche model showed an overall good performance (AUC = 0.919 ± 0.067 SD). We summed current, LGM and LIG suitability layers to generate a map of climate suitability stability, with pixel values ranging from 0 to 3 (minimum and maximum climate suitability in all periods, respectively). Average climate suitability stability ($CLIM_{STA}$) for each population was estimated in an area of 10 km^2 around sampling locations.

We estimated the core of the current distribution range of the Pyrenean Morales grasshopper by calculating the central point of the minimum convex polygon that included all known occurrences of the species. Then, we calculated the Euclidean geographical distance between the species distribution core and each sampled population ($DIST_{COR}$). All GIS calculations were conducted in ARCGIS 10.0.

In order to visualize spatial patterns of genetic variation, we conducted a Genetic Landscape Shape Interpolation analysis using ALLELES IN SPACE 1.0 (AIS) software (Miller 2005). We employed the Delaunay triangulation-based connectivity network to link neighboring sampling sites and calculate genetic distances based on microsatellites. We used residual genetic distances to avoid possible effects of isolation by distance and their values were extracted from sampling sites (GEN_{VAR}) to be used in further analyses. By this approach, a genetic surface of inter-individual genetic distances (*i.e.* genetic variability between nearby demes) is expressed as 'surface heights' and visually displayed as a 3D graph.

We assessed the interdependence between population genetic variability (GEN_{VAR}), climate suitability stability ($CLIM_{STA}$) and the distance from each population to the species distribution core ($DIST_{COR}$) by means of Pearson Rank correlations. Finally, we used generalized linear mixed models (GLMMs) and an information-theoretic model selection approach (Burnham & Anderson 2002) to analyze the association between individual-based morphological traits related to dispersal capability (FL, FWL, and RFWL) and $CLIM_{STA,}$ $DIST_{COR}$ and GEN_{VAR}. We built GLMMs using a Gaussian error distribution and an identity link function. We fitted separated models for males and males and included locality as a random effect to control for the statistical dependence of individuals from the same population. GLMMs were built in the R package LME4 (Bates *et al*. 2015) and model selection and averaging were performed using the R package MUMIN (Barton 2015) as detailed in Noguerales *et al*. (2015) and Ortego *et al*. (2015a).

RESULTS

GENETIC AND PHENOTYPIC DIVERGENCE

Global F_{ST} and F_{ST}NA-values were 0.128 and 0.094, respectively (Table S2). Pairwise F_{ST} and F_{ST}NA-values ranged from 0.021 to 0.216 and 0.014 to 0.148, respectively (see Supplementary Material in Noguerales *et al*. 2016; see Chapter V in this Ph.D. Thesis) and both genetic matrices were highly correlated (Mantel $r = 0.941$; $P < 0.001$).

We found that FL, FWL and RFWL traits differed significantly among populations in both sexes (all *P*s < 0.05; see Fig. S2). RW analyses of shape variation showed a clustering pattern that roughly

grouped individuals from the same populations (Fig. S3). Accordingly, CVA revealed significant differences in forewing shape between populations for both sexes. After Bonferroni correction, Mahalanobis distances (D_2) exhibited statistical significance for 89% and 84% of the total pair-wise comparisons in males and females, respectively (Table S3). Accordingly, global and pairwise P_{ST}-values indicated a high degree of phenotypic divergence for both sexes and all morphological traits (Table S2 and Table S4). Pairwise P_{ST}-values were not correlated with pairwise F_{ST} or F_{ST}NA-values (Mantel tests: all $r < 0.119$, all Ps > 0.193; Table S5) or geographical distances (Mantel tests: all $r < 0.309$, all Ps > 0.055; Table S5) in any morphological trait, which suggests that morphological variation does not conform with that expected under a pattern of neutral genetic differentiation due to population isolation. Additionally, 95% confidence intervals of global P_{ST}-values for any morphological trait did not overlap with those obtained for global F_{ST} or F_{ST}NA (Fig. 2; Table S2), suggesting a predominant effect of divergent selection on the observed patterns of phenotypic differentiation among populations. Sensitivity analyses showed that our results remained similar even under very conservative scenarios considering small values of c/h^2 ratio (Fig. S4).

ENVIRONMENTAL EFFECTS ON PHENOTYPIC DIVERGENCE

After controlling for the influence of geographic distance, dbRDA tests for males showed a significant association between population divergence of forewing length (P_{ST} FWL) and elevation and climate PC2, and between population divergence of forewing length relative to femur length (P_{ST} RFWL) and climate PC3 (all Ps < 0.044), which explained 23.62%, 31.70% and 28.81% of variation, respectively (Table 1). Analyses focused on females indicated that climate PC3 significantly explained 31.76% of morphological variation of forewing length relative to femur length (P_{ST} RFWL) (Table 1). Climate PC2 was explained by a pool of bioclimatic variables related to annual temperature variation (*Bio2*, *Bio3* and *Bio7*) whereas climate PC3 is mainly associated with mean temperature during the driest period of the year (*Bio8*) (Table S6).

Table 1 – Results of distance-based redundancy analyses (dbRDA) testing the effects of geography, elevation and climate on phenotypic differentiation (pairwise P_{ST}-values) quantified for four morphological traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length; FWS: forewing shape) in eleven populations of the Pyrenean Morales grasshopper. In marginal tests, each predictor was tested separately, while in conditional tests geography was always included as covariate. The proportion of the multivariate phenotypic variation explained (% var) by a given predictor or set of predictors is indicated. Predictors with $P < 0.05$ after controlling for geographic influence are highlighted in bold.

Males								Females							
Marginal tests				Conditional tests (*vs*. geography)				Marginal tests				Conditional tests (*vs*. geography)			
Variable	*F*	*P*	% var	Variable	*F*	*P*	% var	Variable	*F*	*P*	% var	Variable	*F*	*P*	% var
P_{ST} FL								P_{ST} FL							
Geography	0.697	0.685	7.18					Geography	0.750	0.605	8.57				
Elevation	1.565	0.125	14.81	Elevation	1.815	0.100	17.16	Elevation	0.952	0.449	10.63	Elevation	1.012	0.411	11.55
Climate PC1	1.403	0.180	13.49	Climate PC1	1.162	0.305	11.77	Climate PC1	0.342	0.936	4.10	Climate PC1	0.785	0.551	9.21
Climate PC2	0.430	0.962	4.56	Climate PC2	1.500	0.153	14.65	Climate PC2	1.535	0.176	16.10	Climate PC2	1.058	0.376	12.04
Climate PC3	0.702	0.713	7.23	Climate PC3	0.747	0.619	7.93	Climate PC3	0.384	0.953	4.58	Climate PC3	1.783	0.173	18.56
P_{ST} FWL								P_{ST} FWL							
Geography	0.849	0.535	8.62					Geography	2.085	0.092	20.68				
Elevation	2.628	0.044	22.60	**Elevation**	2.789	0.044	23.62	Elevation	0.648	0.686	7.49	Elevation	0.708	0.577	7.29
Climate PC1	1.066	0.410	10.59	Climate PC1	1.327	0.266	13.00	Climate PC1	1.108	0.372	12.1	Climate PC1	0.556	0.759	5.83
Climate PC2	1.256	0.272	12.25	**Climate PC2**	4.251	0.007	31.70	Climate PC2	0.846	0.524	9.56	Climate PC2	0.346	0.934	3.74
Climate PC3	1.095	0.338	10.85	Climate PC3	1.577	0.193	15.04	Climate PC3	4.288	0.001	34.90	Climate PC3	2.57	0.055	32.32
P_{ST} RFWL								P_{ST} RFWL							
Geography	0.881	0.527	8.91					Geography	1.366	0.251	14.59				
Elevation	0.824	0.545	8.39	Elevation	0.811	0.601	8.38	Elevation	0.563	0.799	6.57	Elevation	0.542	0.849	6.16
Climate PC1	0.945	0.484	9.50	Climate PC1	1.187	0.309	11.77	Climate PC1	1.223	0.313	13.27	Climate PC1	0.607	0.778	6.82
Climate PC2	0.617	0.722	6.42	Climate PC2	0.831	0.574	8.57	Climate PC2	0.845	0.533	9.654	Climate PC2	0.725	0.658	8.02
Climate PC3	2.032	0.055	18.42	**Climate PC3**	3.702	0.001	28.81	Climate PC3	4.288	0.002	34.90	**Climate PC3**	4.143	0.004	31.76
P_{ST} FWS								P_{ST} FWS							
Geography	0.216	0.993	2.34					Geography	1.495	0.152	15.75				
Elevation	1.019	0.400	10.18	Elevation	0.839	0.562	9.27	Elevation	0.276	0.988	3.34	Elevation	0.285	0.975	3.30
Climate PC1	0.869	0.529	8.81	Climate PC1	1.095	0.388	11.76	Climate PC1	0.743	0.737	8.49	Climate PC1	0.435	0.934	4.93
Climate PC2	0.449	0.908	4.75	Climate PC2	0.455	0.907	5.25	Climate PC2	0.649	0.781	7.50	Climate PC2	0.491	0.871	5.52
Climate PC3	1.564	0.110	14.81	Climate PC3	1.855	0.094	18.38	Climate PC3	1.693	0.105	17.47	Climate PC3	0.762	0.660	8.27

DISPERSAL-RELATED MORPHOLOGY IN RELATION TO CLIMATIC SUITABILITY STABILITY, DISTANCE TO THE DISTRIBUTION CORE AND GENETIC VARIABILITY

Landscape interpolation of inter-individual genetic distances revealed the existence of an area with higher genetic diversity at the core of the species distribution range, whereas either western and eastern margins of the distribution were characterized by lower genetic diversity (*i.e.* higher genetic similarity between nearby demes) (Fig. 3). Accordingly, GEN_{VAR} was negatively correlated with $DIST_{COR}$ (r = -0.668; P = 0.025). However, we did not find a significant correlation between $CLIM_{STA}$ and GEN_{VAR} (r = 0.172; P = 0.613). Likewise, $CLIM_{STA}$ and $DIST_{COR}$ were not significantly correlated (r = 0.241; P = 0.474). Model selection results showed that $CLIM_{STA,}$ $DIST_{COR}$ and GEN_{VAR} were all included in the best ranked models ($\Delta AIC_c \leq 2$) for all the analyzed morphological traits (Table S7). However, exclusively $DIST_{COR}$ had a significant and positive effect on FWL and RFWL in both sexes (Table 2; Fig. 4). None of the predictors included in the averaged model for FL had a significant effect on this trait (*i.e.* all unconditional CIs crossed zero; Table 2).

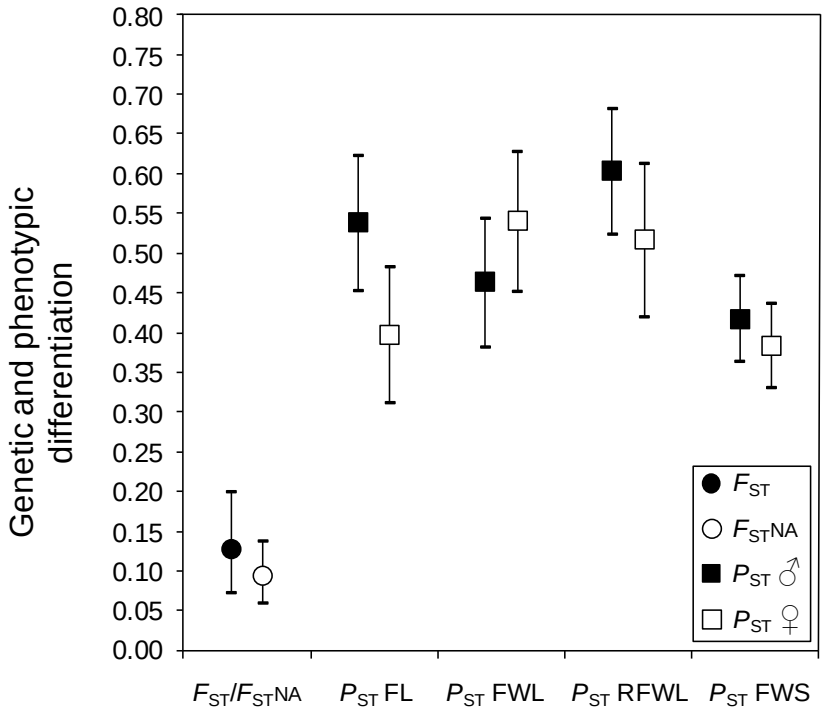


Figure 2 – Global F_{ST}, F_{ST}NA corrected for null alleles, and P_{ST} for four morphological traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length; FWS: forewing shape) and their respective 95% confidence intervals across all populations of the Pyrenean Morales grasshopper. Global P_{ST}-values are shown for each sex separately. The reported P_{ST}-values are those obtained assuming $c = h^2 = 0.5$.

Table 2 – Generalized linear mixed models (GLMM) for three dispersal-related traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length) in the Pyrenean Morales grasshopper testing the effects of climate stability during the last 120 Kya ($CLIM_{STA}$), distance from each population to the distribution core ($DIST_{COR}$) and population genetic variability (GEN_{VAR}). Model averaging was performed for the best ranked equivalent models ($\Delta AIC_c \leq 2$) in order to obtain parameter estimates and unconditional standard errors (S.E.) (for the model selection results, see Table S7). The relative importance of each predictor is indicated ($\sum \omega_i$, sum of Akaike weights of models with $\Delta AIC_c \leq 2$ in which the predictor was present). Bold type indicates predictors excluding the value 0 in their 95% CI and whose effect is considered significant. Models were fitted and averaged separately for each sex.

Males						Females					
Model	Predictor	Estimate ± SE	$\sum \omega_i$	Lower 95% CI	Upper 95% CI	Model	Predictor	Estimate ± SE	$\sum \omega_i$	Lower 95% CI	Upper 95% CI
FL	Intercept	9.1052 ± 0.5981				FL	Intercept	12.5133 ± 0.3449			
	$CLIM_{STA}$	0.5192 ± 0.4203	0.78	-0.3056	1.3442		$CLIM_{STA}$	-0.8124 ± 0.1913	0.23	-0.4562	0.2937
	$DIST_{COR}$	-0.0012 ± 0.0027	0.33	-0.0071	0.0044		$DIST_{COR}$	0.0004 ± 0.0011	0.21	-0.3746	0.3754
	GEN_{VAR}	2.4626 ± 3.7148	0.43	-4.8188	9.7439		GEN_{VAR}	-	-	-	-
FWL	Intercept	9.4348 ± 0.2029				FWL	Intercept	11.2895 ± 0.5737			
	$CLIM_{STA}$	-	-	-	-		$CLIM_{STA}$	-0.1196 ± 0.2618	0.24	-0.6328	0.9353
	$DIST_{COR}$	0.0071 ± 0.0033	1.00	0.0019	0.0165		$DIST_{COR}$	0.0122 ± 0.0048	0.77	0.0027	0.0216
	GEN_{VAR}	1.8227 ± 3.1316	0.37	-4.3146	7.9601		GEN_{VAR}	-3.0428 ± 5.104	0.23	-13.0475	6.9618
RFWL	Intercept	1.0096 ± 0.0488				RFWL	Intercept	0.9054 ± 0.0363			
	$CLIM_{STA}$	-0.0501 ± 0.0355	0.79	-0.1198	0.0185		$CLIM_{STA}$	-	-	-	-
	$DIST_{COR}$	0.0012 ± 0.0003	1.00	0.0007	0.0018		$DIST_{COR}$	0.0007 ± 0.0003	0.70	9.7175×10^{-6}	1.4808×10^{-3}
	GEN_{VAR}	0.2184 ± 0.3259	0.39	-0.4191	0.8553		GEN_{VAR}	0.0002 ± 0.0173	0.50	-0.0921	0.0927

DISCUSSION

Studies combining information on genetic and ecological data provide a powerful approach to assess the relative role of natural selection and neutral mechanisms in shaping phenotypic variation in natural populations and infer the proximate factors involved in such evolutionary processes (Storz 2002; Hangartner *et al*. 2012; Defaveri & Merilä 2013). Although populations of the Pyrenean Morales grasshopper exhibit a strong spatial genetic structure (Noguerales *et al*. 2016), our results indicate that phenotypic differentiation is not driven by genetic drift and point to a more important role of local adaptation processes across the environmentally heterogeneous landscape characterizing the distribution range of this species (Whitlock 2008). Accordingly, phenotypic differentiation (P_{ST}) for all the analyzed traits largely exceeded the background level of genetic neutral differentiation (F_{ST}), which points to divergent selection as the main evolutionary process explaining inter-population phenotypic variation (Leinonen *et al*. 2006; Lehtonen *et al*. 2009; see however Edelaar *et al*. 2011). Our results are in agreement with previous studies where it has been shown a predominant role of selection, rather than

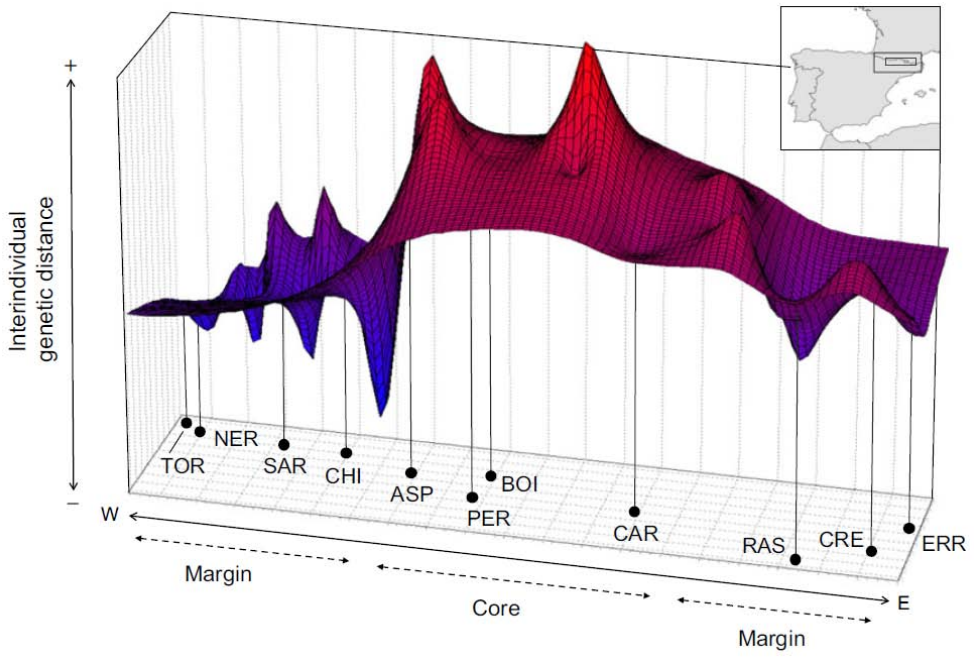


Figure 3 – Genetic landscape shape interpolation analysis based on 100 x 60 grids and a distance weighted value of 2. Surface plot heights are proportional to inter-individual genetic distances. Peaks (red) and troughs (blue) reflect areas with high or low genetic distances (*i.e.* regions with high or low levels of genetic variability) between neighboring demes, respectively. Population codes are described in Table S1.

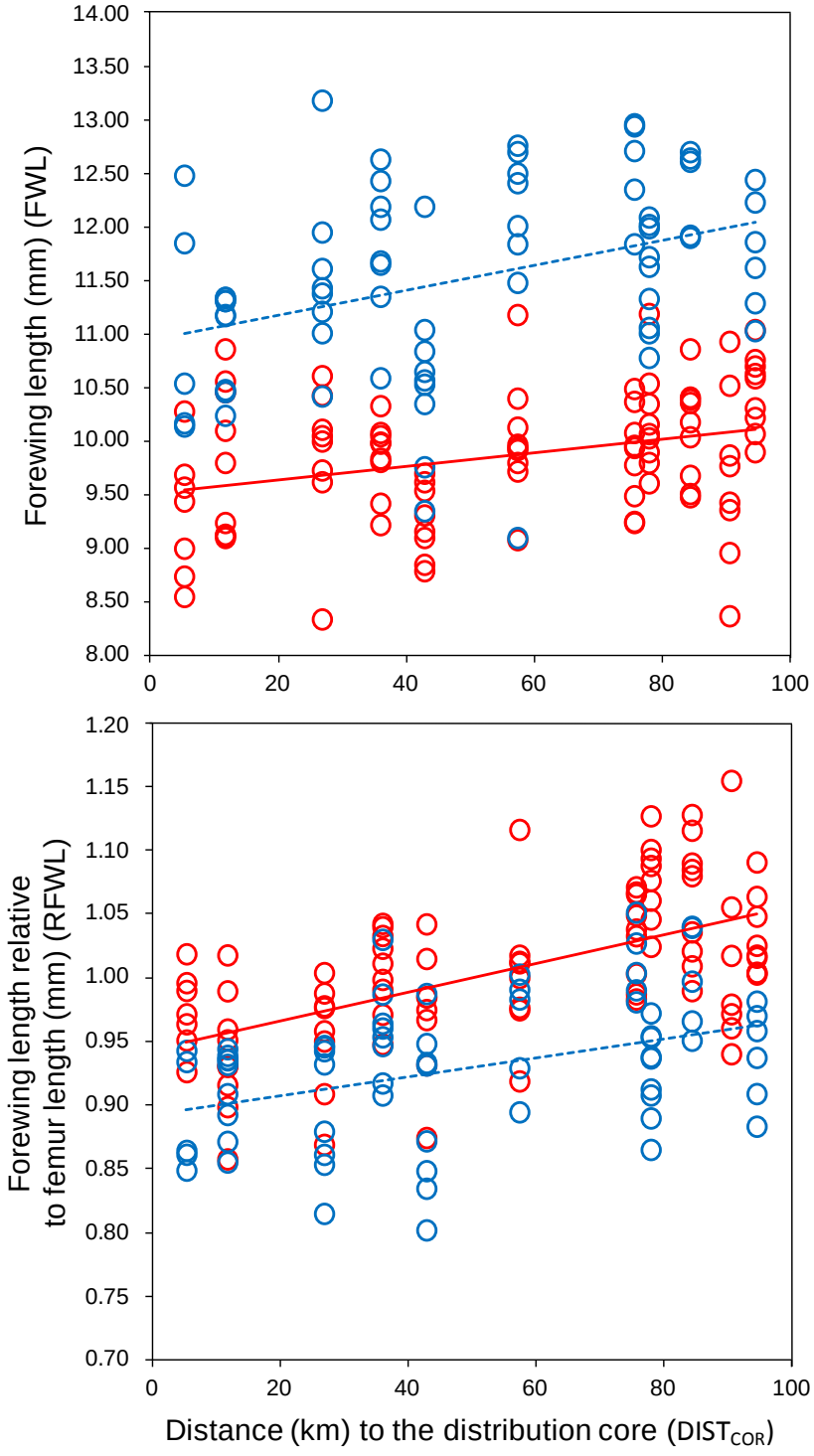


Figure 4 – Relationships between forewing length (FWL) and forewing length relative to femur length (RFWL) and the distance to the distribution range core ($DIST_{COR}$) in eleven populations of the Pyrenean Morales grasshopper. Regression lines for each sex (males: red dots and solid red line; males: blue dots and dashed blue line) are shown.

drift, on morphological differentiation across many taxonomic groups (Merilä & Crnokrak 2001; McKay & Latta 2002; Leinonen *et al*. 2008). At this point we must note that our results should be taken cautiously due to limitations and potential biases underlying P_{ST}-F_{ST} comparisons (Brommer 2011). On the one hand, it must be emphasized that because we do not know the exact value of the c/h^2 coefficient, any inferences made on the basis of P_{ST} estimates can be only considered indicative. Regarding this, we consider unlikely that variation among populations is strongly affected by differences in the environmental conditions experienced across populations (*i.e.* non-additive genetic effects), since the studied traits are heritable ($h^2 \sim 0.5$; Butlin & Hewitt 1986; Roff 1986; Mousseau & Roff 1987; see also Berggren *et al*. 2012). Even so, our conclusions remained valid when considering other more conservative scenarios ($c < h^2$) (Fig. S4). On the other hand, it has been suggested that the employment of hyper-variable markers such as microsatellite loci can diminish estimates of neutral genetic differentiation (F_{ST}) and thus induce a bias towards inferring divergent selection (Whitlock 2008; Edelaar & Björklund 2011; Edelaar *et al*. 2011). However, our analyses based on pair-wise comparisons lack the caveats associated to overall F_{ST}-P_{ST} comparisons and suggest that genetic drift cannot solely explain morphological trait variation (Lehtonen *et al*. 2009).

Our analyses aimed to identify the proximate factors shaping phenotypic variation revealed that differentiation in dispersal-related morphology is associated with temperature gradients. Many environmental factors, such as climate and elevation, have been found to be involved in local adaptation processes and shape morphological and life-history trait variation in insects (*e.g.* Berner *et al*. 2004; Wojcieszek & Simmons 2012; Keller *et al*. 2013; Laiolo & Obeso 2015). With regard to temperature influence, our distance-based redundancy analyses indicate that temperature gradients significantly explained variation of population differentiation in forewing length and forewing length relative to body size (Pitchers *et al*. 2013). Different hypothesis have been proposed to explain the influence of temperature on the evolution of more developed wings in insects. The most accepted hypothesis argues that larger wings (*i.e.* lower wing loadings) would facilitate flight in ectotherms organisms under cooler conditions because low temperatures impair locomotor performance (Gilchrist & Huey 2004; Dillon *et al*. 2006; Pitchers *et al*. 2013). Alternatively, the thermoregulatory hypothesis posits that the evolution of wings in insects is associated to its thermoregulatory function as heat absorption capability increases with wing size (Kingsolver & Koehl 1985; Lewin 1985).

Several studies on insect taxa including orthopterans, have examined evolutionary responses to climate along latitudinal or altitudinal gradients (Shelomi 2012). The most frequently reported pattern is

the existence of a negative association between body size and temperature and/or elevation (converse Bergmann's rule; *e.g.* Blanckenhorn & Demont 2004; Laiolo *et al*. 2013; Levy & Nufio 2015). Body size is a key trait associated with thermoregulation and fecundity in insects (Gilchrist 1990; Honek 1993; Blanckenhorn 2000; Whitman 2008) and this pervasive trend has been traditionally explained as an adaptation to short growing seasons in cooler climates and/or a consequence of resource shortage under harsh environmental conditions (Hodkinson 2005; Roff & Mousseau 2005; Dillon *et al*. 2006). Here, we failed to find an altitudinal or environmental cline for body size in this species. The absence of environmental effects on body size variation among populations of the Pyrenean Morales grasshopper may be a consequence of the faster development rates that orthopterans inhabiting montane habitats exhibit (Telfer & Hassall 1999; Berner *et al*. 2004; Berner & Blanckenhorn 2006). Alternatively, the studied environmental variables become more relevant at a micro-geographic scale (*e.g.* microhabitat sun exposure, slope, etc.) or body size evolution may be driven by other agents not considered in our study (*e.g.* sexual selection, habitat structure, predation risk; Basolo & Wagner 2004; Berner *et al*. 2004; Grace *et al*. 2010; Heidinger *et al*. 2010).

Regarding forewing shape, there was neither a significant association between this variable and elevation or climate. This is in contrast to previous studies in flying insects where it has been shown the existence of environmental clines for wing morphology (Pitchers *et al*. 2013; Perrard *et al*. 2014). As far as we know no other study has investigated forewing shape variation in a grasshopper species, which prevent us from determining whether this lack of variability is widespread in this group or not. In many grasshoppers, including our study species, males produce conspicuous songs by rubbing the inner part of hind legs against forewings whereas females emit short syllables if the specific song of a male matches with their preferences (Harz 1975; Lampe *et al*. 2012). Thus, it would be reasonable to expect that stabilizing (sexual) selection against immigrants exhibiting different forewing shapes impacts phenotypic divergence on this trait regardless of the studied environmental factors (Wojcieszek & Simmons 2012; Oneal & Knowles 2013).

Dispersal capability plays a key role in determining gene flow, ultimately shaping range-wide patterns of genetic variability and spatial genetic structure (Kanuch *et al*. 2012; García-Navas *et al*. 2014; Dussex *et al*. 2016). We found support to our hypothesis predicting a higher development of dispersal-related traits in populations located in the periphery of the species distribution range. Thus, our findings are consistent with previous studies reporting that individuals from populations located at range boundaries, fragmented habitats, and expanding edges possess a higher dispersal capacity than

those inhabiting core areas or stable/continuous habitats (Thomas *et al*. 2001; Therry *et al*. 2014a,b; Hughes *et al*. 2003; Kanuch *et al*. 2012; Fountain *et al*. 2016). According to theory on evolutionary stable dispersal strategies, selection for increasing dispersal capability should be stronger under conditions of high population turnover (Comins *et al*. 1980; Berggren *et al*. 2012). Under this scenario, individuals with more "willing to disperse" phenotypes would be able to settle at the dynamic range boundaries (Shine *et al*. 2011), where habitat suitability may experience stronger temporal changes than in more stable populations at the core of the distribution range (Hardie & Hutchings 2010). The notion that peripheral populations of the Pyrenean Morales grasshopper are less stable than those located in the core area is also reinforced by the lower levels of genetic diversity observed in these populations, which indicates that they sustain smaller effective population sizes according with the central-marginal hypothesis (Eckert *et al*. 2008; Lira-Noriega & Manthey 2014). The fact that we did not find a direct relationship between habitat suitability stability and our estimates of dispersal-related morphology or genetic variability may be consequence of the uncertainties associated with the projection of the current species' distribution models into past climates (*e.g.* Diniz-Filho *et al*. 2015). Alternatively, it's also plausible that past demographic dynamics are not reflected in current evolutionary processes, which could be happening at much shorter temporal scales than those reflected by our long-term habitat stability estimates (Thomas *et al*. 2001; Simmons & Thomas 2004; Fountain *et al*. 2016).

CONCLUSIONS

Studying ongoing differentiation processes in small-scale situations is essential to gain insight about the mechanisms driving adaptive divergence. Here, we sustain local adaptation, rather than genetic drift, as the main factor shaping phenotypic divergence in our study system. Analyses of the proximate environmental factors potentially determining range-wide patterns of phenotypic variation indicate that temperature gradients seem to be involved in morphological divergence. Our data also provide evidence for a higher development of dispersal-related traits in populations located at range boundaries characterized by higher population turnover, unpredictable habitat dynamics and lower levels of genetic diversity in comparison with those located at the core of the species distribution. Overall, our study highlights the importance of integrating genetic and environmental data to get robust inferences about the different evolutionary processes and selective agents shaping phenotypic variation across a species` distribution range.

Acknowledgements

We wish to thank Bernard Defaut for providing valuable information about sampling locations. Chris Klingenberg, Nicola Nadeau and one anonymous referee provided useful and valuable comments on an earlier draft of this manuscript. VN was supported by a FPI pre-doctoral fellowship (BES-2012-053741) (Ministerio de Economía y Competitividad). VGN was supported by a Juan de la Cierva postdoctoral fellowship (Ministerio de Economía y Competitividad). JO was supported by Severo Ochoa (SEV-2012-0262) and Ramón y Cajal (RYC-2013-12501) research fellowships. This work received financial support from research grants CGL2011-25053 (Ministerio de Ciencia e Innovación and European Social Fund), POII10-0197-0167, PEII-2014-023-P (Junta de Comunidades de Castilla-La Mancha and European Social Fund) and UNCM08-1E-018 (European Regional Development Fund).

Availability of data and materials

Nuclear microsatellite, morphology and population geographic and climatic data are available in Dryad repository (*http://datadryad.org/resource/doi:10.5061/dryad.qv56d*). All other data supporting the results of this article are included within the article and its additional files.

REFERENCES

Aljanabi SM, Martínez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692-4693.

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM (2015) Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology*, 24, 1696-1712.

Barton K (2015) *MuMIn: multimodel Inference*. R Package, version 1.15.6. *http://r-forge.r-project.org/projects/mumin/*

Basiita RK, Bruggemann JH, Cai N, *et al*. (2016) Erratum to: Microsatellite records for volume 7, issue 4. *Conservation Genetic Resources,* 8, 85-87.

Basolo AL, Wagner WE (2004) Covariation between predation risk, body size and fin elaboration in the green swordtail, *Xiphophorus helleri*. *Biological Journal of Linnean Society,* 83, 87-100.

Bates D, Maechler M, Bolker BM, Walker SC (2015) Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software,* 67, 1-48.

Berggren H, Tinnert J, Forsman A (2012) Spatial sorting may explain evolutionary dynamics of wing polymorphism in pygmy grasshoppers. *Journal of Evolutionary Biology,* 25, 2126-2138.

Berner D, Blanckenhorn WU (2006) Grasshopper ontogeny in relation to time constraints: adaptive divergence and stasis. *Journal of Animal Ecology*, 75, 130-139.

Berner D, Korner C, Blanckenhorn WU (2004) Grasshopper populations across 2000 m of altitude: is there life history adaptation? *Ecography*, 27, 733-740.

Bertrand JAM, Delahaie B, Bourgeois YXC, *et al*. (2016) The role of selection and historical factors in driving population differentiation along an elevational gradient in an island bird. *Journal of Evolutionary Biology,* 29, 824–836.

Bethoux O, Nel A (2001) Venation pattern of Orthoptera. *Journal of Orthoptera Research*, 10, 195-198.

Blanckenhorn WU (2000) The evolution of body size: what keeps organisms small? *Quarterly Review of Biology*, 75, 385-407.

Blanckenhorn WU, Demont M (2004) Bergmann and converse Bergmann latitudinal clines in arthropods: two ends of a continuum? *Integrative and Comparative Biology*, 44, 413-424.

Bonnet E, Van de Peer Y (2002) ZT: a software tool for simple and partial Mantel tests. *Journal of Statistical Software,* 7, 1-12.

Brommer JE (2011) Whither P_{ST}? The approximation of Q_{ST} by P_{ST} in evolutionary and conservation biology. *Journal of Evolutionary Biology*, 24, 1160-1168.

Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, USA.

Butlin RK, Hewitt GM (1986) Heritability estimates for characters under sexual selection in the grasshopper *Chorthippus brunneus*. *Animal Behaviour*, 34, 1256-1261.

Calabuig G, Ortego J, Cordero PJ, Aparicio JM (2010) Colony foundation in the lesser kestrel: patterns and consequences of the occupation of empty habitat patches. *Animal Behaviour*, 80, 975-982.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24, 621-631.

Collins WD, Bitz CM, Blackmon ML*, et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, 19, 2122-2143.

Comins HN, Hamilton WD, May RM (1980) Evolutionarily stable dispersal strategies. *Journal of Theoretical Biology*, 82, 205-230.

Debeffe L, Morellet N, Cargnelutti B*, et al.* (2012) Condition-dependent natal dispersal in a large herbivore: heavier animals show a greater propensity to disperse and travel further. *Journal of Animal Ecology*, 81, 1327-1337.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (*Caelifera*, *Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Defaveri J, Merilä J (2013) Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *Journal of Evolutionary Biology*, 26, 1700-1715.

Denno RF (1994) The evolution of dispersal polymorphisms in insects: The influence of habitats, host plants and mates. *Researches on Population Ecology*, 36, 127-135.

Denno RF, Roderick GK, Olmstead KL, Dobel HG (1991) Density-related migration in planthoppers (Homoptera: Delphacidae): the role of habitat persistence. *American Naturalist*, 138, 1513-1541.

Dillon ME, Frazier MR, Dudley R (2006) Into thin air: physiology and evolution of alpine insects. *Integrative and Comparative Biology*, 46, 49-61.

Diniz-Filho JAF, Rodrigues H, Telles MPdC*, et al.* (2015) Correlation between genetic diversity and environmental suitability: taking uncertainty from ecological niche models into account. *Molecular Ecology Resources*, 15, 1059-1066.

Dussex N, Chuah A, Waters JM (2016) Genome-wide SNPs reveal fine-scale differentiation among wingless alpine stonefly populations and introgression between winged and wingless forms. *Evolution*, 70, 38-47.

Dytham C (2009) Evolved dispersal strategies at range margins. *Proceedings of the Royal Society B-Biological Sciences*, 276, 1407-1413.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170-1188.

Edelaar P, Alonso D, Lagerveld S, Senar JC, Bjorklund M (2012) Population differentiation and restricted gene flow in Spanish crossbills: not isolation-by-distance but isolation-by-ecology. *Journal of Evolutionary Biology*, 25, 417-430.

Edelaar P, Björklund M (2011) If F_{ST} does not measure neutral genetic differentiation, then comparing it with Q_{ST} is misleading. Or is it? *Molecular Ecology*, 20, 1805-1812.

Edelaar P, Bolnick DI (2012) Non-random gene flow: an underappreciated force in evolution and ecology. *Trends in Ecology & Evolution*, 27, 659-665.

Edelaar P, Burraco P, Gómez-Mestre I (2011) Comparisons between Q_{ST} and F_{ST} - how wrong have we been? *Molecular Ecology*, 20, 4830-4839.

Egan SP, Ragland GJ, Assour L*, et al.* (2015) Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecology Letters*, 18, 817-825.

Eweleit L, Reinhold K (2014) Body size and elevation: do Bergmann's and Rensch's rule apply in the polytypic bushcricket *Poecilimon veluchianus*? *Ecological Entomology*, 39, 133-136.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567.

Fornel R, Cordeiro-Estrela P, Freitas TRO (2010) Skull shape and size variation in *Ctenomys minutus* (Rodentia: Ctenomyidae) in geographical, chromosomal polymorphism, and environmental contexts. *Biological Journal of the Linnean Society*, 101, 705-720.

Fountain T, Nieminen M, Sirén J, Wong SC, Hanski I (2016) Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences of the United States of America,* 113, 2678-2683.

García-Navas V, Ferrer ES, Sanz JJ, Ortego J (2014) The role of immigration and local adaptation on fine-scale genotypic and phenotypic population divergence in a less mobile passerine. *Journal of Evolutionary Biology*, 27, 1590-1603.

Gilchrist GW (1990) The consequences of sexual dimorphism in body size for butterfly flight and thermoregulation. *Functional Ecology*, 4, 475-487.

Gilchrist GW, Huey RB (2004) Plastic and genetic variation in wing loading as a function of temperature within and among parallel clines in *Drosophila subobscura*. *Integrative and Comparative Biology*, 44, 461-470.

Goudet J (1995). F_{STAT} (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, 86, 485-486.

Grace T, Wisely SM, Brown SJ, Dowell FE, 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, 866-878.

Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, 8, 461-467.

Hangartner S, Laurila A, Räsänen K (2012) Adaptive divergence in moor frog (*Rana arvalis*) populations along an acidification gradient: inferences from Q_{ST}-F_{ST} correlations. *Evolution*, 66, 867-881.

Hanski I, Eralahti C, Kankare M, Ovaskainen O, Siren H (2004) Variation in migration propensity among individuals maintained by landscape structure. *Ecology Letters*, 7, 958-966.

Hardie DC, Hutchings JA (2010) Evolutionary ecology at the extremes of species' ranges. *Environmental Reviews*, 18, 1-20.

Harz K (1975) *The Orthoptera of Europe II*. W. Junk Publishers, The Hague.

Heidinger IMM, Hein S, Bonte D (2010) Patch connectivity and sand dynamics affect dispersal-related morphology of the blue-winged grasshopper *Oedipoda caerulescens* in coastal grey dunes. *Insect Conservation and Diversity*, 3, 205-212.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Hill JK, Griffiths HM, Thomas CD (2011) Climate change and evolutionary adaptations at species' range margins. *Annual Review of Entomology*, 56, 143-159.

Hochkirch A, Gröning J (2008) Sexual size dimorphism in Orthoptera (sens. str.) – a review. *Journal of Orthoptera Research*, 17, 189-196.

Hodkinson ID (2005) Terrestrial insects along elevation gradients: species and community responses to altitude. *Biological Reviews*, 80, 489-513.

Honek A (1993) Intraspecific variation in body size and fecundity in insects: a general relationship. *Oikos*, 66, 483-492.

Hughes CL, Dytham C, Hill JK (2007) Modelling and analysing evolution of dispersal in populations at expanding range boundaries. *Ecological Entomology*, 32, 437-445.

Hughes CL, Hill JK, Dytham C (2003) Evolutionary trade-offs between reproduction and dispersal in populations at expanding range boundaries. *Proceedings of the Royal Society B-Biological Sciences*, 270, S147-S150.

Kanuch P, Jarcuska B, Schlosserova D*, et al.* (2012) Landscape configuration determines gene flow and phenotype in a flightless forest-edge ground-dwelling bush-cricket, *Pholidoptera griseoaptera*. *Evolutionary Ecology*, 26, 1331-1343.

Kekkonen J, Jensen H, Brommer JE (2012) Morphometric differentiation across House Sparrow *Passer domesticus* populations in Finland in comparison with the neutral expectation for divergence. *Ibis*, 154, 846-857.

Keller I, Alexander JM, Holderegger R, Edwards PJ (2013) Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *Journal of Evolutionary Biology*, 26, 2527-2543.

Kingsolver JG, Koehl MAR (1985) Aerodynamics, thermoregulation, and the evolution of insect wings: differential scaling and evolutionary change. *Evolution*, 39, 488-504.

Klingenberg CP (2011) MORPHOJ: an integrated software packagefor geometric morphometrics. *Molecular Ecology Resources,* 11: 353-357.

Klingenberg CP, Debat V, Roff DA (2010) Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. *Evolution*, 64, 2935-2951.

Laiolo P, Illera JC, Obeso JR (2013) Local climate determines intra- and interspecific variation in sexual size dimorphism in mountain grasshopper communities. *Journal of Evolutionary Biology*, 26, 2171-2183.

Laiolo P, Obeso JR (2015) Plastic responses to temperature versus local adaptation at the cold extreme of the climate gradient. *Evolutionary Biology*, 42, 473-482.

Lampe U, Schmoll T, Franzke A, Reinhold K (2012) Staying tuned: grasshoppers from noisy roadside habitats produce courtship signals with elevated frequency components. *Functional Ecology*, 26, 1348-1354.

Lande R (1976) Natural-selection and random genetic drift in phenotypic evolution. *Evolution*, 30, 314-334.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69, 1-24.

Lehtonen PK, Laaksonen T, Artemyev AV, *et al.* (2009) Geographic patterns of genetic differentiation and plumage colour variation are different in the pied flycatcher (*Ficedula hypoleuca*). *Molecular Ecology*, 18, 4463-4476.

Leinonen T, Cano JM, Makinen H, Merilä J (2006) Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*, 19, 1803-1812.

Leinonen T, McCairns RJS, O'Hara RB, Merilä J (2013) Q_{ST}-F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, 14, 179-190.

Leinonen T, O'Hara RB, Cano JM, Merilä J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, 21, 1-17.

Levy RA, Nufio CR (2015) Dispersal potential impacts size clines of grasshoppers across an elevation gradient. *Oikos*, 124, 610-619.

Lewin R (1985) On the origin of insect wings. *Science*, 230, 428-429.

Lira-Noriega A, Manthey JD (2014) Relationship of genetic diversity and niche centrality: a survey and analysis. *Evolution*, 68, 1082-1093.

Llucià-Pomares D (2002) *Revision of the Orthoptera (Insecta) of Catalonia (Spain).* Monografías S.E.A., vol. 7, Zaragoza, Spain.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, Q_{TL} and traits. *Trends in Ecology & Evolution*, 17, 285-291.

Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, 14, 892-903.

Miller MP (2005) ALLELES IN SPACE (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, 96, 722-724.

Mousseau TA, Roff DA (1987) Natural-selection and the heritability of fitness components. *Heredity*, 59, 181-197.

Noguerales V, Cordero PJ, Ortego J (2016) Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology*, 16, 96.

Noguerales V, Traba J, Mata C, Morales M (2015) Winter habitat selection and partitioning in two sympatric farmland small mammals: *Apodemus sylvaticus* and *Mus spretus*. *Revue d'écologie (La Terre et la Vie)*, 70, 70-82.

Norman GR, Streiner DL (2000) *Biostatistics*: *The Bare Essentials, 2nd edn*. BC Decker, London, UK.

Nosil P, Crespi BJ (2004) Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution*, 58, 102-112.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al*. (2016) *VEGAN: community ecology package*. R Package, version 2.3-4. *http://vegan.r-forge.r-project.org/*

Oneal E, Knowles LL (2013) Ecological selection as the cause and sexual differentiation as the consequence of species divergence? *Proceedings of the Royal Society B-Biological Sciences*, 280, 20122236.

O'Riain MJ, Jarvis JUM, Faulkes CG (1996) A dispersive morph in the naked mole-rat. *Nature*, 380, 619-621.

Ortego J, Aguirre MP, Cordero PJ (2012) Genetic and morphological divergence at different spatiotemporal scales in the grasshopper *Mioscirtus wagneri* (Orthoptera: Acrididae). *Journal of Insect Conservation*, 16, 103-110.

Ortego J, Aguirre MP, Noguerales V, Cordero PJ (2015a) Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications*, 8, 621-632.

Ortego J, García-Navas V, Noguerales V, Cordero PJ (2015b) Discordant patterns of genetic and phenotypic differentiation in five grasshopper species codistributed across a microreserve network. *Molecular Ecology*, 24, 5796-5812.

Otto-Bliesner BL, Marsha SJ, Overpeck JT*, et al.* (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science*, 311, 1751-1753.

Perrard A, Baylac M, Carpenter JM, Villemant C (2014) Evolution of wing shape in hornets: why is the wing venation efficient for species identification? *Journal of Evolutionary Biology*, 27, 2665-2675.

Petit D, Picaud F, Elghadraoui L (2006) Geometric morphometrics of Acrididae wings (Orthoptera, Caelifera): sex, stridulation and character. *Annales De La Societe Entomologique De France*, 42, 63-73.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

Phillips SJ, Dudik M (2008) Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Pitchers W, Pool JE, Dworkin I (2013) Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? *Evolution*, 67, 438-452.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org/.*

Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters,* 11, 624-636.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223-225.

Ripley B (2016) *BOOT: bootstrap functions*. R Package, version 1.3-18. *https://cran.r-project.org/web/packages/bootstrap/index.html*

Roff DA (1986) The genetic-basis of wing dimorphism in the sand cricket, *Gryllus firmus* and its relevance to the evolution of wing dimorphisms in insects. *Heredity*, 57, 221-231.

Roff DA, Mousseau T (2005) The evolution of the phenotypic covariance matrix: evidence for selection and drift in *Melanoplus*. *Journal of Evolutionary Biology*, 18, 1104-1114.

Rohlf FJ (2015) TPS SERIES. Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA. *http://life.bio.sunysb.edu/morph.*

Rohlf FJ, Slice D (1990) Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39, 40-59.

Routtu J, Mazzi D, Van Der Linde K*, et al.* (2007) The extent of variation in male song, wing and genital characters among allopatric *Drosophila montana* populations. *Journal of Evolutionary Biology*, 20, 1591-1601.

San Martín y Gómez G, Van Dyck H (2012) Ecotypic differentiation between urban and rural populations of the grasshopper *Chorthippus brunneus* relative to climate and habitat fragmentation. *Oecologia*, 169, 125-133.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford, UK.

Shelomi M (2012) Where are we now? Bergmann's Rule sensu lato in insects. *American Naturalist*, 180, 511-519.

Shine R, Brown GP, Phillips BL (2011) An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 5708-5711.

Simmons AD, Thomas CD (2004) Changes in dispersal during species' range expansions. *American Naturalist*, 164, 378-395.

Spitze K (1993) Population structure in *Daphnia obtusa*: quantitive genetic and allozymic variation. *Genetics*, 135, 367-374.

Storz JF (2002) Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Molecular Ecology*, 11, 2537-2551.

Telfer MG, Hassall M (1999) Ecotypic differentiation in the grasshopper *Chorthippus brunneus:* life history varies in relation to climate. *Oecologia*, 121, 245-254.

Therry L, Gyulavari HA, Schillewaert S, Bonte D, Stoks R (2014a) Integrating large-scale geographic patterns in flight morphology, flight characteristics and sexual selection in a range-expanding damselfly. *Ecography*, 37, 1012-1021.

Therry L, Zawal A, Bonte D, Stoks R (2014b) What factors shape female phenotypes of a poleward-moving damselfly at the edge of its range? *Biological Journal of the Linnean Society*, 112, 556-568.

Thomas CD, Bodsworth EJ, Wilson RJ*, et al.* (2001) Ecological and evolutionary processes at expanding range margins. *Nature*, 411, 577-581.

Tregenza T, Pritchard VL, Butlin RK (2000) Patterns of trait divergence between populations of the meadow grasshopper, *Chorthippus parallelus*. *Evolution*, 54, 574-585.

Whitlock MC (2008) Evolutionary inference from Qst. *Molecular Ecology*, 17, 1885-1896.

Whitman DW (2008) The significance of body size in the Orthoptera: a review. *Journal of Orthoptera Research*, 17, 117-134.

Wojcieszek JM, Simmons LW (2012) Evidence for stabilizing selection and slow divergent evolution of male genitalia in a millipede (*Antichiropus variabilis*). *Evolution*, 66, 1138-1153.

Wojcik AM, Polly PD, Sikorski MD, Wojcik JM (2006) Selection in a cycling population: differential response among skeletal traits. *Evolution*, 60, 1925-1935.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric morphometrics for biologists: a primer*. Elsevier Academic Press, San Diego, CA, USA.

Zera AJ, Denno RF (1997) Physiology and ecology of dispersal polymorphism in insects. *Annual Review of Entomology*, 42, 207-230.

Zhan J, Linde CC, Jurgens T*, et al.* (2005) Variation for neutral markers is correlated with variation for quantitative traits in the plant pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology*, 14, 2683-2693.

SUPPLEMENTARY MATERIAL

Table S1 – Geographical location, elevation and number of genotyped individuals (*n*) for the studied populations of the Pyrenean Morales grasshopper.

Locality	Code	*n*	Latitude	Longitude	Elevation (m.a.s.l)
Torla	TOR	20	42.63860	-0.08196	1822
Nerín	NER	20	42.59047	0.01062	1623
Saravillo	SAR	20	42.56089	0.23046	1313
Chía	CHI	20	42.56833	0.41708	1900
Aspes	ASP	20	42.44307	0.58389	1361
Boi	BOI	12	42.47945	0.86732	2046
Perves	PER	20	42.35250	0.83709	1370
Carmeniu	CAR	20	42.37243	1.33766	1140
Err	ERR	19	42.42762	2.05956	1800
Creueta	CRE	11	42.30131	1.99350	1928
Rasos	RAS	20	42.14165	1.76461	1840

Table S2 – Global genetic (F_{ST} and F_{ST}NA) and phenotypic differentiation (measured as global P_{ST}) for four morphological traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length; FWS: forewing shape) across eleven populations of the Pyrenean Morales grasshopper. The reported pairwise P_{ST}-values are those obtained assuming $c = h^2 = 0.5$.

	Males	Females
Global F_{ST}	0.128 [0.073 - 0.200]	
Global F_{ST}NA	0.094 [0.060 - 0.138]	
Global P_{ST} FL	0.538 [0.453 - 0.623]	0.398 [0.312 - 0.483]
Global P_{ST} FWL	0.463 [0.382 - 0.544]	0.540 [0.452 - 0.628]
Global P_{ST} RFWL	0.603 [0.524 - 0.682]	0.517 [0.420 - 0.613]
Global P_{ST} FWS	0.417 [0.364 - 0.472]	0.384 [0.331 - 0.437]

Table S3 – Results of Canonical Variates Analyses (CVA) testing for forewing shape differentiation between populations of Pyrenean Morales grasshopper. Mahalanobis distances (D_2) for males are presented below the diagonal and above the diagonal for females. No data was available for females in one population (CRE). Population codes as in Table S1. Significant Mahalanobis distances after sequential Bonferroni correction are shown in bold.

	TOR	NER	SAR	CHI	ASP	BOI	PER	CAR	ERR	CRE	RAS
TOR	-	3.066	3.450	4.998	4.232	5.260	4.764	3.359	3.550	-	3.384
NER	1.995	-	1.938	4.196	2.716	4.010	3.148	2.400	3.164	-	2.247
SAR	2.046	2.111	-	4.001	2.617	3.969	3.252	2.562	3.207	-	2.650
CHI	3.651	3.513	2.771	-	2.817	3.412	2.668	3.826	4.366	-	4.247
ASP	3.325	2.968	2.419	2.795	-	2.394	2.108	2.766	3.060	-	2.913
BOI	5.199	5.433	4.881	3.831	3.713	-	2.637	3.028	3.603	-	4.470
PER	3.692	4.054	3.333	3.033	2.511	2.778	-	3.014	3.170	-	3.195
CAR	2.944	2.767	3.057	2.939	2.460	4.189	2.896	-	3.140	-	3.530
ERR	3.618	3.931	3.781	3.392	2.930	2.744	2.195	2.737	-		2.498
CRE	3.137	3.461	2.880	3.363	3.034	3.870	2.644	3.798	3.147	-	-
RAS	2.870	3.496	3.291	3.009	3.655	3.869	3.058	3.376	2.615	3.126	-

Table S4 – Phenotypic differentiation (measured as P_{ST}) between eleven populations of the Pyrenean Morales grasshopper for four morphological traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length; FWS: forewing shape). Pairwise P_{ST}-values for males are presented below the diagonal and above the diagonal for females. No data was available for females in one population (CRE). Population codes as in Table S1. The reported pairwise P_{ST}-values are those obtained assuming $c = h^2 = 0.5$.

		TOR	NER	SAR	CHI	ASP	BOI	PER	CAR	ERR	CRE	RAS
P_{ST} FL	TOR	-	0.135	0.821	0.267	0.578	0.050	0.356	0.046	0.240	-	0.129
	NER	0.004	-	0.773	0.437	0.426	0.014	0.666	0.031	0.031	-	0.003
	SAR	0.822	0.748	-	0.818	0.233	0.742	0.945	0.795	0.727	-	0.671
	CHI	0.006	0.000	0.811	-	0.743	0.366	0.114	0.467	0.531	-	0.523
	ASP	0.896	0.873	0.640	0.886	-	0.452	0.842	0.575	0.362	-	0.410
	BOI	0.037	0.015	0.667	0.016	0.832	-	0.509	0.002	0.071	-	0.024
	PER	0.903	0.886	0.575	0.892	0.054	0.850	-	0.534	0.726	-	0.579
	CAR	0.508	0.441	0.165	0.462	0.742	0.319	0.758	-	0.118	-	0.053
	ERR	0.848	0.817	0.413	0.833	0.095	0.760	0.385	0.535	-	-	0.010
	CRE	0.617	0.517	0.310	0.585	0.757	0.618	0.803	0.003	0.620	-	-
	RAS	0.000	0.004	0.830	0.064	0.892	0.675	0.897	0.493	0.841	0.620	-

		TOR	NER	SAR	CHI	ASP	BOI	PER	CAR	ERR	CRE	RAS
P_{ST} FWL	TOR	-	0.215	0.287	0.912	0.692	0.769	0.941	0.564	0.682	-	0.841
	NER	0.414	-	0.439	0.926	0.770	0.810	0.950	0.702	0.776	-	0.883
	SAR	0.058	0.202	-	0.770	0.170	0.435	0.673	0.001	0.019	-	0.231
	CHI	0.898	0.817	0.844	-	0.743	0.287	0.454	0.857	0.827	-	0.815
	ASP	0.262	0.002	0.113	0.707	-	0.307	0.603	0.250	0.144	-	0.000
	BOI	0.810	0.663	0.737	0.034	0.557	-	0.009	0.582	0.506	-	0.406
	PER	0.335	0.004	0.177	0.676	0.008	0.525	-	0.830	0.028	-	0.382
	CAR	0.442	0.002	0.205	0.865	0.000	0.716	0.011	-	0.028	-	0.382
	ERR	0.633	0.839	0.672	0.961	0.724	0.918	0.752	0.870	-	-	0.270
	CRE	0.499	0.158	0.368	0.438	0.132	0.274	0.094	0.193	0.786	-	-
	RAS	0.065	0.550	0.191	0.913	0.390	0.838	0.456	0.585	0.509	0.582	-

		TOR	NER	SAR	CHI	ASP	BOI	PER	CAR	ERR	CRE	RAS
P_{ST} RFWL	TOR	-	0.000	0.538	0.846	0.889	0.879	0.923	0.556	0.692	-	0.895
	NER	0.500	-	0.600	0.872	0.909	0.947	0.939	0.800	0.875	-	0.920
	SAR	0.700	0.500	-	0.053	0.059	0.048	0.000	0.333	0.100	-	0.067
	CHI	0.895	0.800	0.400	-	0.000	0.000	0.143	0.727	0.538	-	0.400
	ASP	0.925	0.867	0.625	0.200	-	0.000	0.200	0.789	0.600	-	0.429
	BOI	0.882	0.875	0.429	0.000	0.200	-	0.500	0.778	0.636	-	0.667
	PER	0.944	0.909	0.750	0.556	0.200	0.556	-	0.818	0.600	-	0.333
	CAR	0.778	0.600	0.000	0.556	0.857	0.667	0.833	-	0.333	-	0.714
	ERR	0.500	0.000	0.429	0.778	0.926	0.857	0.902	0.600	-	-	0.333
	CRE	0.556	0.250	0.000	0.429	0.600	0.455	0.724	0.000	0.143	-	-
	RAS	0.200	0.800	0.846	0.917	0.938	0.952	0.953	0.913	0.818	0.727	-

		TOR	NER	SAR	CHI	ASP	BOI	PER	CAR	ERR	CRE	RAS
P_{ST} FWS	TOR	-	0.348	0.533	0.378	0.456	0.424	0.506	0.348	0.681	-	0.498
	NER	0.129	-	0.198	0.425	0.241	0.374	0.641	0.507	0.482	-	0.158
	SAR	0.485	0.312	-	0.613	0.614	0.454	0.793	0.671	0.324	-	0.284
	CHI	0.398	0.226	0.085	-	0.287	0.063	0.017	0.238	0.383	-	0.465
	ASP	0.753	0.730	0.575	0.518	-	0.098	0.464	0.317	0.178	-	0.107
	BOI	0.610	0.524	0.374	0.293	0.157	-	0.135	0.267	0.259	-	0.272
	PER	0.290	0.164	0.208	0.049	0.571	0.377	-	0.389	0.569	-	0.664
	CAR	0.646	0.553	0.589	0.608	0.431	0.207	0.689	-	0.505	-	0.380
	ERR	0.474	0.338	0.056	0.179	0.634	0.367	0.343	0.585	-	-	0.301
	CRE	0.159	0.388	0.594	0.374	0.640	0.531	0.279	0.796	0.638	-	-
	RAS	0.197	0.240	0.445	0.228	0.633	0.532	0.101	0.812	0.620	0.252	-

Table S5 – Results of Mantel tests analyzing the association between pairwise P_{ST}-values for four morphological traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length; FWS: forewing shape) and Euclidean geographic distances, F_{ST} and F_{ST}NA (F_{ST} corrected for null alleles) matrices among eleven populations of the Pyrenean Morales grasshopper. *P* values were obtained by 10 000 permutations.

	Geography		F_{ST}		F_{ST}NA	
	Mantel *r*	*P*	Mantel *r*	*P*	Mantel *r*	*P*
Males						
P_{ST} FL	-0.063	0.361	0.006	0.437	0.031	0.370
P_{ST} FWL	0.042	0.349	0.038	0.388	0.071	0.330
P_{ST} RFWL	-0.034	0.441	-0.016	0.479	0.045	0.354
P_{ST} FWS	-0.308	0.308	-0.008	0.490	-0.069	0.333
Females						
P_{ST} FL	-0.237	0.088	0.051	0.398	0.029	0.449
P_{ST} FWL	0.154	0.187	-0.119	0.244	-0.146	0.212
P_{ST} RFWL	0.309	0.055	-0.069	0.363	-0.136	0.242
P_{ST} FWS	0.141	0.202	-0.146	0.193	-0.120	0.246

Table S6 – Factor loadings after a 'Kaiser' varimax rotation for the first three principal components (PC) axes and the 19 bioclimatic variables obtained from the WorldClim dataset. Bold type indicates bioclimatic variables with factor loadings higher than 0.7.

	PC1	PC2	PC3
Bio1	**0.920**	0.207	0.268
Bio2	0.271	**0.952**	0.059
Bio3	-0.058	**0.885**	0.328
Bio4	0.656	0.398	-0.522
Bio5	**0.909**	0.358	0.146
Bio6	**0.920**	0.026	0.323
Bio7	0.538	**0.788**	-0.213
Bio8	**0.901**	0.091	-0.157
Bio9	**0.246**	0.213	**0.800**
Bio10	**0.939**	0.218	0.196
Bio11	**0.911**	0.136	0.333
Bio12	**-0.899**	-0.412	0.066
Bio13	**-0.814**	-0.512	0.145
Bio14	**-0.922**	-0.336	-0.109
Bio15	**0.874**	-0.020	-0.005
Bio16	**-0.860**	-0.449	0.102
Bio17	**-0.924**	-0.349	0.013
Bio18	**-0.816**	-0.475	-0.153
Bio19	**-0.898**	-0.232	0.318

Table S7 – Model selection testing the association between three dispersal-related traits (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length) in the Pyrenean Morales grasshopper and climate stability during the last 120 Kya ($CLIM_{STA}$), distance from each population to the distribution core ($DIST_{COR}$) and population genetic variability (GEN_{VAR}). Details are given for the best ranked equivalent models ($\Delta AIC_c \leq 2$). For each model we indicate *K*, number of parameters in the model; AIC_c, corrected Akaike's information criterion (AIC) value; ΔAIC_c, difference in AIC_c value from that of the strongest model; ωi, AIC_c weight.

Males						Females					
Model	Model parameters	*K*	AIC_c	ΔAIC_c	ω_i	Model	Model parameters	*K*	AIC_c	ΔAIC_c	ω_i
FL						FL					
1	$CLIM_{STA} + DIST_{COR}$	2	123.50	0.00	0.33	1	Null model	0	142.80	0.00	0.56
2	$CLIM_{STA}$	1	124.19	0.69	0.24	2	$CLIM_{STA}$	1	144.60	1.80	0.23
3	GEN_{VAR}	1	124.36	0.86	0.22	3	$DIST_{COR}$	1	144.74	1.94	0.21
4	$CLIM_{STA} + GEN_{VAR}$	2	124.39	0.89	0.21						
FWL						FWL					
1	$DIST_{COR}$	1	169.50	0.00	0.63	1	$DIST_{COR}$	1	178.60	0.00	0.53
2	$DIST_{COR} + GEN_{VAR}$	2	170.55	1.05	0.37	2	GEN_{VAR}	1	180.18	1.58	0.23
						3	$CLIM_{STA} + DIST_{COR}$	2	180.24	1.64	0.24
RFWL						RFWL					
1	$CLIM_{STA} + DIST_{COR}$	2	-289.30	0.00	0.40	1	$DIST_{COR}$	1	-188.00	0.00	0.50
2	$CLIM_{STA} + DIST_{COR} + GEN_{VAR}$	3	-289.35	0.05	0.39	2	GEN_{VAR}	1	-187.01	0.99	0.30
3	$DIST_{COR}$	1	-288.08	1.27	0.21	3	$DIST_{COR} + GEN_{VAR}$	2	-186.19	1.81	0.20

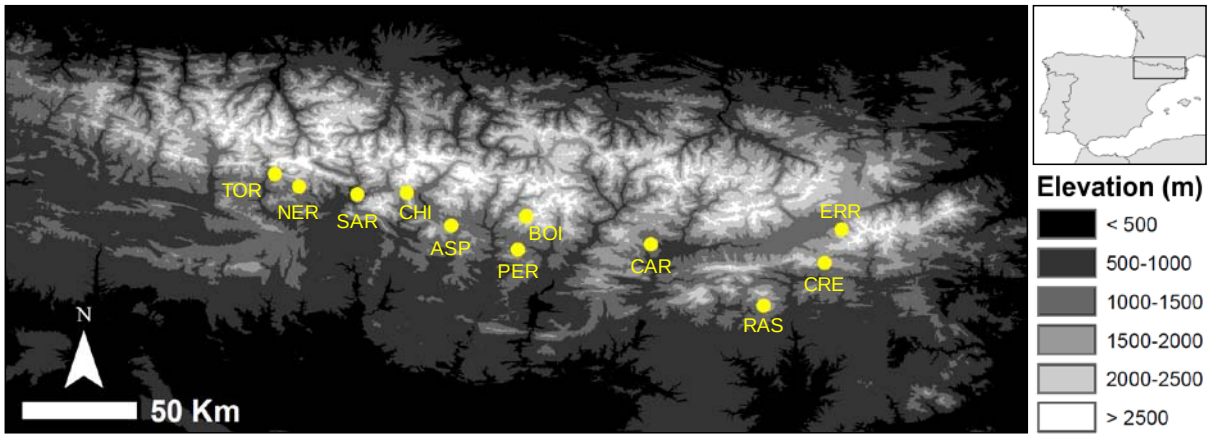


Figure S1 – Sampling sites of the Pyrenean Morales grasshopper plotted on an elevation map of the study area, the Pyrenees mountains. Topographic data was obtained from NASA Shuttle Radar Topographic Mission (SRTM Digital Elevation Data, *http://srtm.csi.cgiar.org/*). Detailed information for each sampling site is given in Table S1.

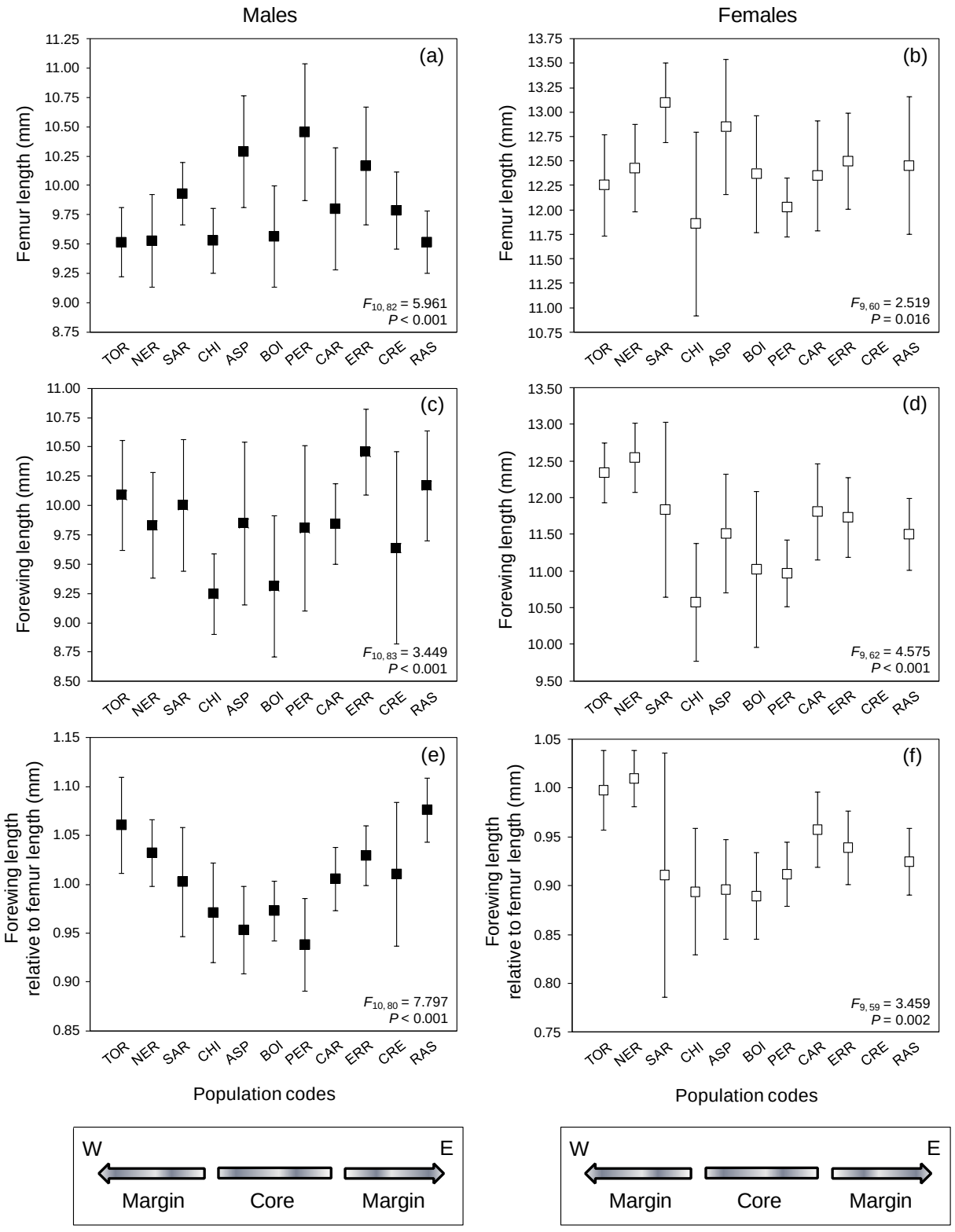


Figure S2 – Population mean (± SD) values for linear morphological traits in eleven populations of the Pyrenean Morales grasshopper (males: left panels; females: right panels). Results of one-way ANOVAs testing for differences in morphological traits among populations are shown. Population codes as in Table S1. No data was available for females in one population (CRE). Populations are sorted from west to east, with those in the middle being located at the species distribution core and those at the right and left extremes representing marginal populations.

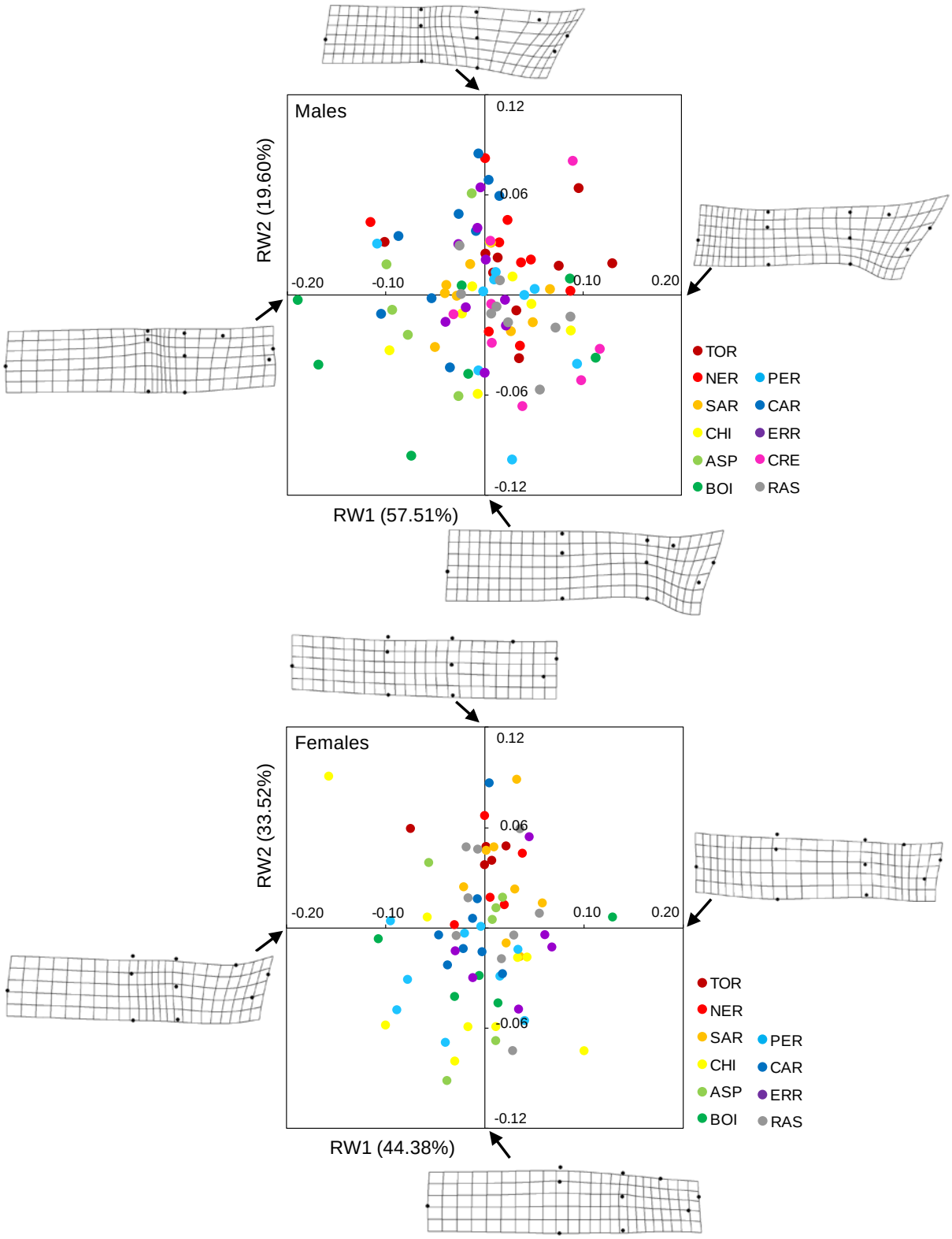


Figure S3 – Forewing shape variation in males (upper panel) and females (lower panel) of Pyrenean Morales grasshopper along the two first relative warps. No data was available for females in one population (CRE). Thin-plate spline transformation grids show extreme shapes for each relative warp. Population codes as in Table S1.

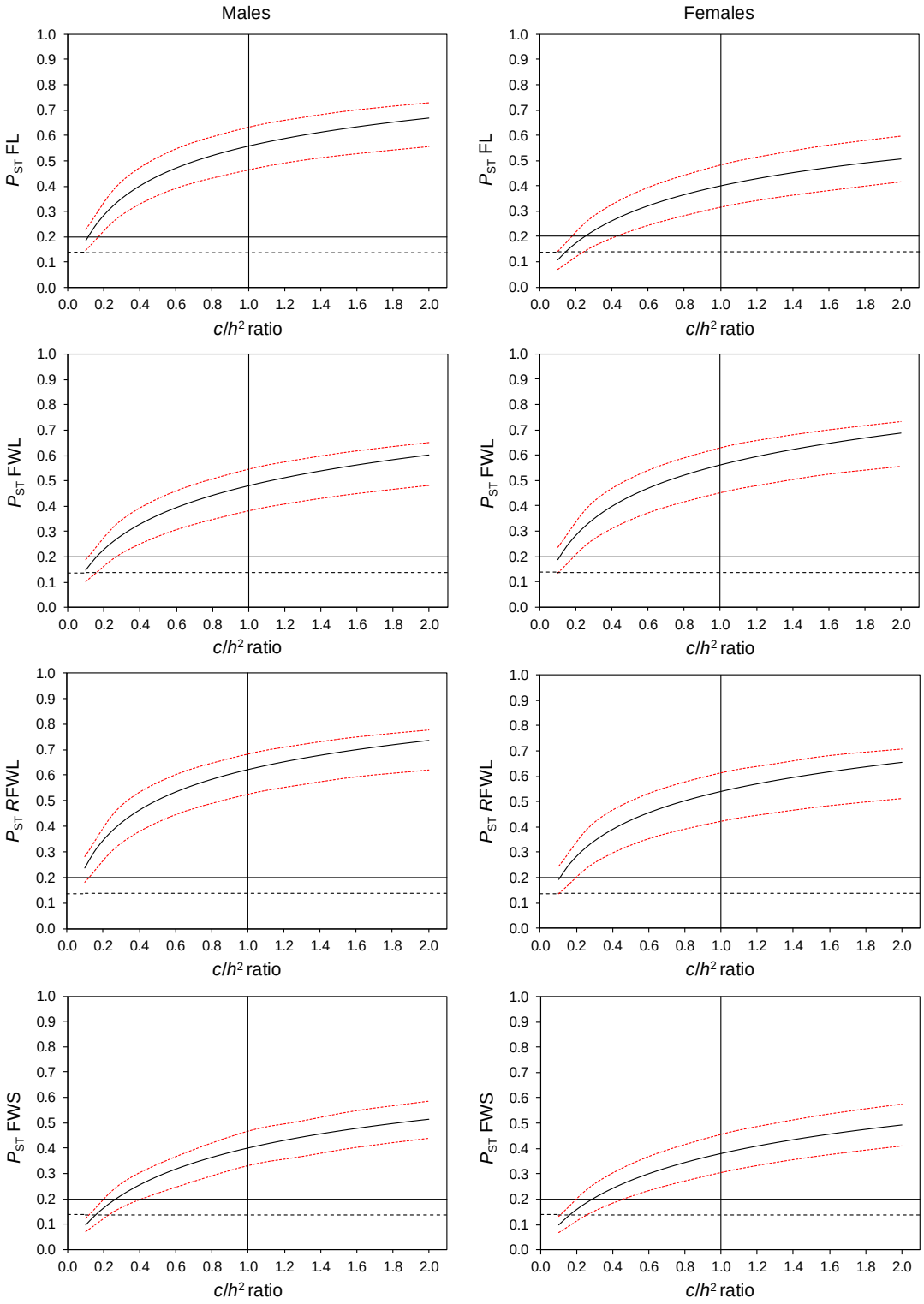


Figure S4 – P_{ST}-values and their 95% confidence intervals (black solid and red dashed lines, respectively) for varying values of c/h^2 ratio and different morphological traits of the Pyrenean Morales grasshopper (FL: femur length; FWL: forewing length; RFWL: forewing length relative to femur length; FWS: forewing shape). For each trait and sex, upper 95% confidence intervals for F_{ST} and F_{ST}NA (F_{ST} corrected for null alleles) are shown as black solid and dashed horizontal lines, respectively. The vertical solid line represents the *x*-axis value where $c/h^2 = 1$ (*i.e.* the proportion of phenotypic variance that is due to additive genetic effects is assumed to be the same within and among populations). Left panels for males and right for females.

DISCUSIÓN GENERAL

Paisaje de la Sierra de Gredos. En primer plano, piornales de Cytisus oromediterraneus, hábitat característico de Chorthippus binotatus binotatus (C. binotatus). Fotografía tomada por Cipriano Noguerales Esteban.

La combinación de datos genómicos, morfológicos y ecológicos bajo un marco metodológico multi-disciplinar ha permitido determinar el papel de los diferentes mecanismos evolutivos (neutrales y adaptativos) que generan y mantienen la diversidad biológica a diferentes escalas espacio-temporales (Leaché *at al*. 2009). En particular, esta aproximación proporciona información relevante acerca de la contribución relativa de la geografía, el ambiente y las fluctuaciones climáticas sobre los patrones de diferenciación genética y fenotípica a lo largo del continuo de divergencia (Crispo *et al*. 2006; Voje *et al*. 2009; Barley *et al*. 2015; Ortego *et al*. 2015). La selección como modelo de estudio de una radiación evolutiva reciente compuesta por entidades taxonómicas que presentan distintos grados de diferenciación genética/fenotípica ha hecho factible abordar el estudio de los procesos evolutivos a distintos niveles biológicos: especies, linajes y poblaciones (Prates *et al*. 2016). Este sistema de estudio ha permitido además examinar comparativamente las trayectorias demográficas y evolutivas en taxones que presentan requerimientos ecológicos marcadamente distintos (Papadopoulou y Knowles 2016; Massatti y Knowles 2016). En esta misma línea de estudio, el conocimiento generado en esta Tesis Doctoral proporciona evidencia empírica de los distintos procesos evolutivos estocásticos y deterministas que actúan a lo largo de todo el espectro de diferenciación genómica y fenotípica, contribuyendo a una mejor comprensión de los mecanismos que subyacen al fenómeno de la especiación (Butlin *et al*. 2012; Shafer y Wolf 2013).

Se han determinado los límites taxonómicos y las relaciones filogenéticas entre las distintas entidades biológicas que conforman el complejo de especies objeto de estudio mediante la generación de datos genómicos por ddRADSeq (Peterson *et al*. 2012) y la utilización de un amplio conjunto de métodos basados en coalescencia (Yang y Rannala 2010, 2014; Leaché *et al*. 2014; Solís-Lemus *et al*. 2015). Estos análisis indican de modo consistente que los ocho taxones, considerados hasta ahora infra-específicos, serían elevados a la categoría de especies. A pesar de que nuestros análisis de sensibilidad han mostrado la eficacia del método BPP (Yang y Rannala 2010, 2014, 2017; Rannala y Yang 2013) para inferir de forma consistente el mismo modelo de delimitación de especies bajo diferentes escenarios, dicha aproximación no está exenta de limitaciones (Sukumaran y Knowles 2017) (Capítulo I). Recientemente se ha sugerido que el modelo de coalescencia multi-específica en el que se basa BPP y otros métodos similares pudiera no ser eficaz a la hora de discriminar la estructura genética promovida por procesos a nivel poblacional de aquella resultante de mecanismos de divergencia inter-específica, lo que podría conllevar una sobreestimación del número de especies, particularmente en escenarios evolutivos caracterizados por una marcada estructura genética

poblacional (Sukumaran y Knowles 2017). Mientras se desarrollan nuevos modelos que solventen esta limitación, dichos autores recomiendan la integración de datos genéticos y fenotípicos bajo un mismo marco conceptual y analítico de delimitación de especies (Sukumaran y Knowles 2017), una demanda que ha sido puesta de manifiesto en la literatura científica de los últimos años (Wiens 2007; Yeates *et al*. 2011). En este sentido, el esquema taxonómico de ocho especies queda también fuertemente apoyado por los resultados procedentes de iBPP, un método recientemente desarrollado que integra datos genómicos y morfológicos para analizar cuantitativa y estadísticamente diferentes hipótesis de delimitación de especies (Solís-Lemus *et al*. 2015). El empleo de esta aproximación integradora permitió evidenciar que todos los taxones infra-específicos y específicos que conforman el *Chorthippus* grupo *binotatus* son estadísticamente equivalentes y, por tanto, merecen ser considerados efectivamente como especies distintas (*i.e.* linajes evolutivamente independientes, *sensu* de Queiroz 2007; ver también Huang y Knowles 2016). Por todo lo anterior, el complejo de especies quedaría compuesto por las siguientes especies: *Chorthippus algoaldensis*, *C. armoricanus*, *C. atlasi*, *C. binotatus*, *C. daimei*, *C. moralesi*, *C. saulcyi* y *C. vicdessossi*, una conclusión que implicaría la revisión del estatus de conservación de algunos de estos taxones (Hochkirch *et al*. 2016). El resultado aquí obtenido ofrece pistas para la comprensión de la verdadera diversidad específica de este grupo y a su vez tiene importantes implicaciones taxonómicas de cara al desarrollo de la actual estrategia de gestión y conservación de los ortópteros a escala europea (Hochkirch *et al*. 2016) (Capítulo I). Aunque nuestros resultados implican un cambio en el estatus taxonómico de las diferentes especies y subespecies del complejo *Chorthippus* grupo *binotatus* elevándolas al rango de especies, sería interesante realizar más estudios que integren información bioacústica, ecológica y corológica procedente de toda el área de distribución potencial de cada taxón para confirmar la validez de los resultados obtenidos en nuestros análisis. No obstante, a lo largo de esta discusión se emplea la nomenclatura más aceptada hasta la fecha (Defaut 2011, 2015) con el objetivo de facilitar su lectura y evitar confusiones entre los diferentes capítulos de esta memoria.

Los análisis dirigidos a examinar el impacto de la información fenotípica procedente de distintas fuentes sobre la delimitación de especies indicaron que el mayor grado de apoyo estadístico a la divergencia entre especies fue obtenido cuando se emplearon matrices que incluían datos de rasgos morfológicos relativos a la longitud y forma de la tegmina (Capítulo I). Estos rasgos presentaron una señal filogenética más fuerte que aquellos que describían la variación morfológica del pronoto, un hecho que podría estar asociado a que la tegmina es una estructura implicada en múltiples funciones biológicas como el vuelo, la termorregulación o la producción de sonidos y, por tanto, posiblemente

sometida a una selección más intensa que la forma del pronoto (Thomas *et al*. 2001; Petit *et al*. 2006) (Capítulo VI). Este resultado, no obstante, varió en función del sexo debido probablemente al modo en el que las presiones evolutivas afectan diferencialmente los caracteres en machos y hembras. Independientemente de las diferencias entre rasgos y sexos, nuestros análisis señalaron la importancia de incorporar información procedente de múltiples rasgos con el propósito de capturar la mayor proporción de información del espectro completo de variación fenotípica. Esta aproximación adquiere especial importancia cuando se abordan escenarios de especiación reciente donde los patrones de divergencia fenotípica pueden ser muy sutiles. El nuevo método iBPP parece configurarse como una potente herramienta que solventa las deficiencias que hasta la fecha existían a la hora de establecer los límites inter-específicos empleando diferentes tipos de información (Andujar *et al*. 2014; Watcher *et al*. 2015) y el posible sesgo hacia la sobreestimación del número de especies cuando se emplean métodos coalescentes basados exclusivamente en datos genéticos (Sukumaran y Knowles 2017) (Capítulo I).

Las relaciones filogenómicas dentro de la especie nominativa *C. binotatus* fueron resueltas de modo robusto y sugieren que la divergencia entre los taxones que la conforman (*C. b. atlasi*, *C. b armoricanus* y *C. b. binotatus*) probablemente se pudo producir en alopatría como consecuencia del aislamiento generado por barreras geográficas (*i.e.* estrecho de Gibraltar y Pirineos) (Capítulo I y III). Por el contrario, las relaciones filogenéticas dentro del grupo *C. saulcyi* no fueron tan claras y los resultados variaron en función de las características de las distintas matrices de SNPs analizadas (número de loci y proporción de *missing data*, *i.e.* SNPs homólogos sin información para determinados individuos). Este resultado pone de manifiesto el impacto que puede ocasionar la variación de algunos parámetros durante el tratamiento bioinformático de los datos de ddRADSeq sobre la inferencia de relaciones filogenéticas, un hecho que ya ha sido documentado previamente por otros autores (Takahashi *et al*. 2014; Leaché *et al*. 2015; Herrera y Shank 2016). La topología más apoyada estadísticamente situó a *C. s. algoaldensis* como un taxón independiente pero incluido en el clado *C. saulcyi*, lo que indicaría que las especies nominativas *C. binotatus* y *C. saulcyi* son monofiléticas, en línea con lo establecido por Defaut (2011) en base a caracteres morfológicos. Los restantes taxones (*C. s. moralesi, C. s. vicdessossi, C. s. saulcyi* y *C. s. daimei*) incluidos en el clado *C. saulcyi* conformarían una politomía que probablemente refleja un evento de divergencia simultaneo y posterior aislamiento en diferentes macizos montañosos. En este sentido, las condiciones de aislamiento geográfico acompañados por mecanismos evolutivos neutrales (*ej.* deriva genética) parecen haber tenido una importancia evidente en la diversificación de ambos grupos (*C. binotatus* y *C. saulcyi*;

Knowles 2000) (Capítulo I). Sin embargo, futuros estudios que examinen la diferenciación de nicho ecológico en aquellos taxones que actualmente exhiben áreas de distribución parapátricas o ligeramente solapantes (*ej*. *C. s. moralesi*, *C. s. vicdessossi* y *C. s. saulcyi* en Pirineos) podrían proporcionar información más precisa acerca de la contribución del ambiente en los mecanismos de divergencia inter-especifica (Huang y Knowles 2016) (Capítulo I).

A una escala evolutiva menor (*i.e.* a nivel intra-específico), la geografía también parece ser el factor responsable de la diversificación dentro de la subespecie *C. b. binotatus.* A lo largo de su rango de distribución, este taxón exhibe una estructuración genética (linajes) resultante del aislamiento de sus poblaciones en diferentes refugios climáticos durante los ciclos glaciales e inter-glaciales del Pleistoceno (Hewitt 1996), un patrón de diversificación acorde con el concepto denominado "refugios dentro de refugios" (*refugia-within-refugia concept*; Gómez y Lunt 2007; Abellán y Svenning 2014) (Capítulo III y IV). Contrariamente a lo esperado, la presencia de un linaje genuinamente francés se explicaría por la persistencia de la especie en micro-refugios situados en el límite septentrional de su rango de distribución actual (Provan y Bennet 2008; Stewart *et al*. 2010) y no por la colonización post-glacial desde la península Ibérica como ha sido documentado para una buena parte de organismos europeos (Schmitt 2007). La persistencia de poblaciones de *C. b. binotatus* en el límite norte de su distribución durante periodos climáticos desfavorables estaría estrechamente vinculada a la presencia de comunidades vegetales conformadas por algunas de sus plantas nutricias, una posibilidad que fue apoyada por los modelos de nicho climático potencial durante los últimos 21000 años obtenidos para las distintas especies de leguminosas arbustivas de las que se alimenta. A pesar de que la estabilidad de estas comunidades vegetales en áreas septentrionales posiblemente permitió el mantenimiento de poblaciones viables de *C. b. binotatus*, nuestros resultados sugieren que la mayor intensidad de las fluctuaciones climáticas en esa región generaron cuellos de botella demográficos y redujeron los niveles de diversidad genética a escala local y regional (Bidegaray-Batista *et al*. 2016). En esta línea, nuestros análisis mostraron una disminución de la variabilidad genética con la latitud acorde con el patrón denominado "riqueza en el sur *vs*. pureza en el norte" (*southern richness vs. northern purity*; Hewitt 2004) (Capítulo III). Sin embargo, cuando se redujo la escala espacial y se puso el foco únicamente en el sureste de la península ibérica, la diversidad genética poblacional no fue explicada por la estabilidad climática durante los últimos 21000 años. Los modelos de nicho climático específicos para esta región indicaron que la mayoría de las poblaciones de *C. b. binotatus* en esta zona se asentaban sobre áreas que han presentado una estabilidad climática muy alta. Este resultado sugiere que el sureste de la península Ibérica actuó para *C. b. binotatus* como un gran refugio climático

independiente evolutiva y demográficamente de otras áreas climáticamente estables situadas a lo largo del rango de distribución de la especie (Capítulo IV). En su conjunto, estos resultados ponen de manifiesto la importancia de integrar factores abióticos (*i.e.* clima y geografía) y bióticos (interacciones inter-específicas) de cara a comprender mejor los factores que generan los patrones de variación genética en las poblaciones naturales (Tsai y Manos 2010) (Capítulo III y IV).

Mientras a escala de todo el rango de distribución la formación de los distintos linajes de *C. b. binotatus* es consecuencia de procesos de divergencia en alopatría, a una menor escala espacio-temporal (sureste de la península Ibérica) son las características fisionómicas del paisaje las responsables de modelar el patrón espacial de estructura y flujo genético (Capítulo IV). Este hallazgo concuerda con los resultados obtenidos para el taxón pirenaico *C. s. moralesi*, cuya diferenciación genética inter-poblacional fue explicada igualmente por la complejidad topográfica del paisaje (Capítulo V). En ambos casos, las regiones de estudio (sureste de la península Ibérica y Pirineos) están caracterizadas por un relieve muy abrupto, lo que unido a la baja capacidad dispersiva de estos taxones explicaría el papel relevante de la complejidad topográfica como predictor de la diferenciación genética neutral de sus poblaciones. Resultados similares han sido obtenidos para otros organismos terrestres, un hecho que ha sido relacionado con el mayor gasto energético y la dificultad que supone desplazarse a través de paisajes muy irregulares topográficamente (Castillo *et al*. 2014; Velo-Antón *et al*. 2013; Benham y Witt 2016; Hemp *et al*. 2016).

Todos los resultados anteriormente expuestos apuntan a que los mecanismos evolutivos neutrales (*i.e.* deriva genética), en contraposición a los adaptativos, parecen haber contribuido primordialmente a la divergencia genética a lo largo del continuo de especiación del complejo de especies *Chorthippus* grupo *binotatus* (Capítulos I, III, IV y V). Cuando se examinó explícitamente la importancia del aislamiento por ambiente (Wang y Bradburd 2014), no se encontró un efecto importante de los procesos de adaptación local (*i.e.* divergencia ecológica/ambiental) sobre el patrón de flujo genético entre las poblaciones de *C. s. moralesi* a lo largo de su rango de distribución. Este hallazgo fue corroborado por la consistencia en las inferencias obtenidas a partir de diferentes aproximaciones estadísticas, las cuales permitieron discriminar entre el efecto de la geografía y el ambiente y concluir que éste último presenta un papel muy limitado sobre la divergencia genética inter-poblacional (Capítulo V). Este resultado contrasta con otros estudios previos que han encontrado una importante contribución del ambiente a la diferenciación genética en numerosos taxones (Shafer y Wolf 2013) e incluso su mayor contribución relativa en comparación con la geografía en determinados grupos de

organismos, como los insectos (Sexton *et al*. 2014). La ausencia de efecto del ambiente sobre el flujo genético podría deberse a que nuestra caracterización ambiental de las poblaciones no incluyó la variación de determinados factores ecológicos que pudieran estar actuando como agentes de selección (*ej.* diferentes plantas nutricias, micro-estructura del hábitat; Grace *et al*. 2010) (Capítulo V). No obstante, aunque no se encontró un efecto significativo del ambiente como limitador del flujo genético neutral (*i.e.* basado en microsatélites), nuestros análisis apuntan a que la selección divergente es el principal mecanismo evolutivo responsable de la variación en distintos rasgos morfológicos como el tamaño corporal y la longitud y forma de la tegmina (Leinonen *et al*. 2006; Lehtonen *et al*. 2009) (Capítulo VI). El hecho de que las diferencias inter-poblacionales en el grado de desarrollo de la tegmina estuvieran explicadas por la temperatura sugiere una respuesta adaptativa a lo largo de un gradiente climático y altitudinal (Pitchers *et al*. 2013) (Capítulo VI). Adicionalmente, nuestros resultados mostraron que las poblaciones localizadas en los márgenes del rango de distribución presentan un mayor desarrollo de rasgos involucrados en la capacidad dispersiva (*i.e.* tegminas de mayor longitud) y una menor variabilidad genética en comparación con aquellas situadas en las zonas centrales del rango. Este hallazgo concuerda con la hipótesis evolutiva sobre estrategias estables de dispersión, la cual predice un incremento en la fuerza de selección hacia fenotipos más dispersivos bajo condiciones de mayor tasa de recambio poblacional, como pueden ser las imperantes en los bordes de la distribución de muchos organismos (Shine *et al*. 2011; Berggren *et al*. 2012; Fountain *et al*. 2016). La menor variabilidad genética en las poblaciones periféricas en comparación con las centrales apunta a que las primeras sustentan menores tamaños efectivos poblaciones y recurrentes dinámicas de extinción-recolonización, un patrón concordante con la hipótesis central-marginal de diversidad genética (*marginal-core hypothesis*; Eckert *et al*. 2008; Lira-Noriega y Manthey 2014) (Capítulo VI).

REFERENCIAS

Abellán P, Svenning J-C (2014) Refugia within refugia - patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, 113, 13-28.

Andujar C, Arribas P, Ruiz C, Serrano J, Gómez-Zurita J (2014) Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation of *Mesocarabus* (Coleoptera: Carabidae). *Molecular Ecology*, 23, 4344-4361.

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM (2015) Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology*, 24, 1696-1712.

Benham PM, Witt CC (2016) The dual role of Andean topography in primary divergence: functional and neutral variation among populations of the hummingbird, *Metallura tyrianthina*. *BMC Evolutionary Biology*, 16, 22.

Berggren H, Tinnert J, Forsman A (2012) Spatial sorting may explain evolutionary dynamics of wing polymorphism in pygmy grasshoppers. *Journal of Evolutionary Biology*, 25, 2126-2138.

Bidegaray-Batista L, Sánchez-Gracia A, Santulli G*, et al.* (2016) Imprints of multiple glacial refugia in the Pyrenees revealed by phylogeography and palaeodistribution modelling of an endemic spider. *Molecular Ecology*, 25, 2046-2064.

Butlin RK, Debelle A, Kerth C*, et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, 27, 27-39.

Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology*, 23, 843-856.

Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, 15, 49-62.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, 56, 879-886.

Defaut B (2011) Preliminary revision of *Chorthippus* of the *binotatus* group (Charpentier, 1825) (*Caelifera*, *Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 16, 17-54.

Defaut B (2015) *Gomphocerippus binotatus armoricanus*, sous-espèce nouvelle de France (*Acrididae*, *Gomphocerinae*). *Materiaux Orthopteriques et Entomocenotiques*, 20, 25-26.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170-1188.

Fountain T, Nieminen M, Siren J, Wong SC, Hanski I (2016) Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 2678-2683.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N (Eds.) *Phylogeography in southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity*. Springer Verlag, Dordrecht, The Netherlands.

Grace T, Wisely SM, Brown SJ, Dowell FE, 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, 866-878.

Hemp C, Grzywacz B, Warchalowska-Sliwa E, Hemp A (2016) Topography and climatic fluctuations boosting speciation: biogeography and a molecular phylogeny of the East African genera *Afroanthracites* Hemp & Ingrisch and *Afroagraecia* Ingrisch & Hemp (Orthoptera, Tettigoniidae, Conocephalinae, Agraeciini). *Organisms Diversity & Evolution*, 16, 211-223.

Herrera S, Shank TM (2016) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, 100, 70-79.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247-276.

Hewitt GM (2004) The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology*, 1, 4.

Hochkirch A, Nieto A, García Criado M, *et al*. (2016) *European Red List of Grasshoppers, Crickets and Bush-crickets*. Publications Office of the European Union , Luxembourg.

Huang J-P, Knowles LL (2016) The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single bayesian approach in Hercules beetles. *Systematic Biology*, 65, 685-699.

Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, 54, 1337-1348.

Leaché AD, Chavez AS, Jones LN*, et al.* (2015) Phylogenomics of Phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, 7, 706-719.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, 63, 534-542.

Leaché AD, Koo MS, Spencer CL*, et al.* (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12418-12423.

Lehtonen PK, Laaksonen T, Artemyev AV*, et al.* (2009) Geographic patterns of genetic differentiation and plumage colour variation are different in the pied flycatcher (*Ficedula hypoleuca*). *Molecular Ecology*, 18, 4463-4476.

Leinonen T, Cano JM, Makinen H, Merilä J (2006) Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*, 19, 1803-1812.

Lira-Noriega A, Manthey JD (2014) Relationship of genetic diversity and niche centrality: a survey and analysis. *Evolution*, 68, 1082-1093.

Massatti R, Knowles LL (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, 3974-3986.

Ortego J, Noguerales V, Gugger PF, Sork VL (2015) Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology*, 24, 6188-6208.

Papadopoulou A, Knowles LL (2016) Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 8018-8024.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One*, 7, e37135.

Petit D, Picaud F, Elghadraoui L (2006) Geometric morphometrics of Acrididae wings (Orthoptera, Caelifera): sex, stridulation and character. *Annales De La Societe Entomologique De France*, 42, 63-73.

Pitchers W, Pool JE, Dworkin I (2013) Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? *Evolution*, 67, 438-452.

Prates I, Xue AT, Brown JL*, et al.* (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7978-7985.

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, 23, 564-571.

Rannala B, Yang Z (2013) Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, 194, 245-253.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4, 11.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, 68, 1-15.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, 16, 940-950.

Shine R, Brown GP, Phillips BL (2011) An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 5708-5711.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, 69, 492-507.

Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences*, 277, 661-671.

Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 1607-16120

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.

Thomas CD, Bodsworth EJ, Wilson RJ*, et al.* (2001) Ecological and evolutionary processes at expanding range margins. *Nature*, 411, 577-581.

Tsai Y-HE, Manos PS (2010) Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana*. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17035-17040.

Velo-Antón G, Parra JL, Parra-Olea G, Zamudio KR (2013) Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, 22, 3261-3278.

Voje KL, Hemp C, Flagstad O, Saetre G-P, Stenseth NC (2009) Climatic change as an engine for speciation in flightless Orthoptera species inhabiting African mountains. *Molecular Ecology*, 18, 93-108.

Wachter GA, Muster C, Arthofer W*, et al.* (2015) Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: Megabunus). *Molecular Ecology*, 24, 863-889.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, 23, 5649-5662.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, 56, 875-878.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9264-9269.

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, 31, 3125-3135.

Yang Z, Rannala B (2017) Bayesian species identificacion under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology*, in press. DOI: 10.1111/mec.14093

Yeates DK, Seago A, Nelson L*, et al.* (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, 36, 209-217.

CONCLUSIONES

Chorthippus binotatus binotatus (C. binotatus) ♂, Puerto de La Morcuera (Madrid).
Fotografía tomada por Víctor Noguerales Rodríguez.

1. Los datos genómicos y fenotípicos apoyan que todos los taxones del complejo *Chorthippus* grupo *binotatus*, tanto a nivel de especie como de subespecie, son estadísticamente equivalentes. En su conjunto, estos resultados indican que la taxonomía del complejo queda definida por ocho entidades biológicas independientes que merecen ser elevadas a la categoría de especies (*Chorthippus algoaldensis*, *C. armoricanus*, *C. atlasi*, *C. binotatus*, *C. daimei*, *C. moralesi*, *C. saulcyi* y *C. vicdessossi*), cuyo estatus de conservación debería ser revisado (Capítulo I).

2. Los taxones que conforman el complejo de especies se agruparon en dos clados monofiléticos (*Chorthippus binotatus* y *Chorthippus saulcyi*). El clado *Chorthippus binotatus* incluye los taxones *atlasi*, *armoricanus* y *binotatus* y el clado *Chorthippus saulcyi* los taxones *algoaldensis*, *daimei*, *moralesi*, *saulcyi* y *vicdessossi*, lo cual concuerda con la taxonomía propuesta con anterioridad en base a caracteres morfológicos y comportamentales. Sin embargo, las relaciones filogenéticas entre la mayoría de los taxones que conforman el clado *C. saulcyi* no pudieron ser totalmente resueltas y los resultados muestran una politomía que incluye a *C. daimei*, *C. moralesi*, *C. saulcyi* y *C. vicdessossi* (Capítulo I).

3. El método Bayesiano empleado para delimitar especies integrando datos genómicos y fenotípicos bajo un marco estadístico unificado (iBPP) mostró una alta eficacia para discriminar los límites taxonómicos bajo diferentes escenarios demográficos. Los resultados obtenidos indican que la utilización de diferentes fuentes de información fenotípica (morfometría clásica y geometría morfométrica) procedentes de ambos sexos puede tener un impacto importante en la delimitación de especies, mientras que el número de loci empleados tiene una escasa repercusión en las inferencias obtenidas (Capítulo I).

4. Se aislaron dieciocho marcadores microsatélites para el taxón *Chorthippus binotatus binotatus* cuya funcionalidad y grado de polimorfismo ha sido caracterizado para todas las especies del complejo (Capítulo II). El empleo de estos marcadores permitió el estudio de la demografía y estructura y diversidad genética de los taxones *Chorthippus binotatus binotatus* y *Chorthippus saulcyi moralesi* a escalas tanto filogeográficas como de paisaje (Capítulos III y VI).

5. El taxón *Chorthippus binotatus binotatus* presentó una marcada subdivisión genética a lo largo de toda su área de distribución (península Ibérica y Francia) cuyo origen probablemente radica en procesos de divergencia en alopatría resultantes tanto de la presencia de barreras geográficas (Pirineos) como del aislamiento de sus poblaciones en refugios climáticos durante del Pleistoceno. La existencia de un linaje propio del centro y oeste de Francia indica la persistencia a largo plazo de la especie en áreas ecológicamente estables situadas en el actual límite septentrional de su rango de distribución, lo que descarta que el origen de estas poblaciones esté vinculado a una colonización post-glacial desde refugios climáticos ibéricos (Capítulo III).

6. El linaje de *Chorthippus binotatus binotatus* distribuido actualmente por el centro y oeste de Francia presenta señales genéticas indicativas de un cuello de botella demográfico, probablemente ocasionado por la severidad de las fluctuaciones climáticas de los últimos 21000 años a las que se han visto sometidas sus poblaciones a escala regional. La mayor inestabilidad de los hábitats en las zonas más norteñas del rango de distribución de la especie ha dado lugar a una reducción de la diversidad genética con la latitud, un patrón semejante al observado en otros organismos europeos (Capítulo III).

7. A una escala espacial menor, las poblaciones de *Chorthippus binotatus binotatus* en el sureste de la península Ibérica presentaron una sub-estructuración genética coherente geográficamente con la distribución de los diferentes sistemas montañosos de la región. Dicha estructuración genética sería consecuencia de procesos neutrales de diferenciación ocasionados por el aislamiento geográfico de sus poblaciones. Esta conclusión fue apoyada por el hecho de que el escenario de aislamiento por resistencia basado en la complejidad topográfica del paisaje fue el que mejor explicó la diferenciación genética entre las poblaciones del área de estudio (Capítulo IV).

8. El patrón de flujo genético entre las poblacio*nes* del taxón pirenaico *Chorthippus saulcyi moralesi* no estuvo determinado por procesos de adaptación local a gradientes ambientales, si no por mecanismos neutrales resultantes del aislamiento poblacional generado por la fisionomía del territorio. A pesar de su reducido rango de distribución, este taxón exhibió una marcada estructura genética espacial que probablemente es debida a su limitada capacidad dispersiva y a la escasa conectividad inter-poblacional como consecuencia del abrupto relieve que caracteriza los Pirineos (Capítulo V).

9. La selección divergente mediada por gradientes climáticos parece ser responsable de la diferenciación fenotípica inter-poblacional para todos los rasgos morfológicos estudiados en *Chorthippus saulcyi moralesi*. En particular, el grado de desarrollo de la tegmina se asoció a un gradiente temperatura, un hecho que sugiere que la variación genética de loci potencialmente implicados en la expresión de dicho rasgo responde a los diferentes regímenes climáticos locales (Capítulo VI).

10. Las poblaciones de *Chorthippus saulcyi moralesi* localizadas en los límites del rango de distribución de la especie presentaron una menor diversidad genética y tegminas proporcionalmente más largas en comparación con aquellas poblaciones situadas en el centro de la distribución. Este hallazgo sugiere que las poblaciones periféricas experimentan una menor estabilidad demográfica que reduce sus tamaños efectivos poblacionales y favorece la selección de fenotipos con mayor capacidad dispersiva (Capítulo VI).

CONCLUSIONS

Paisaje de Pirineos, hábitat potencial de Chorthippus saulcyi moralesi (C. moralesi).
Fotografía tomada por Joaquín Ortego Lozano.

1. Genomic and morphological data support that all taxa within the *Chorthippus* group *binotatus* species complex, at both the specific and infra-specific level, are statistically equivalent. These results indicate that the taxonomy of the group would be defined by eight independent biological entities that merit full species status (*Chorthippus algoaldensis*, *C. armoricanus*, *C. atlasi*, *C. binotatus*, *C. daimei*, *C. moralesi*, *C. saulcyi* and *C. vicdessossi)* and whose conservation status should be thoroughly revised (Chapter I).

2. Taxa from the species complex were grouped into two monophyletic clades (*Chorthippus binotatus* and *Chorthippus saulcyi*). The *Chorthippus binotatus* clade includes *atlasi*, *armoricanus* and *binotatus*, while *Chorthippus saulcyi* clade comprises *algoaldensis*, *daimei*, *moralesi*, *saulcyi* and *vicdessossi*. This scheme is in agreement with the previously proposed taxonomy based on morphological and behavioral traits. Nevertheless, phylogenetic relationships among taxa from the *C. saulcyi* clade were not fully resolved and our results suggest a polytomy including *C. daimei*, *C. moralesi*, *C. saulcyi* and *C. vicdessossi* (Chapter I).

3. The Bayesian species delimitation method integrating genomic and morphological data in a unified statistical framework (iBPP) showed a high performance to distinguish species boundaries under different demographic scenarios. The results indicate that using different sources of morphological information (traditional *vs*. geometric morphometrics) from both sexes can have an important impact on species delimitation, while the number of loci employed had a little influence on the obtained inferences (Chapter I).

4. We isolated and characterized eighteen microsatellite loci for *Chorthippus binotatus binotatus* and described their functionality and degree of polymorphism for all taxa within the species complex (Chapter II). These molecular markers allowed us to study the genetic structure, genetic variation and demographic dynamics of *Chorthippus binotatus binotatus* and *Chorthippus saulcyi moralesi* at both phylogeographic- and landscape-level scales (Chapters III and VI).

5. The taxon *Chorthippus binotatus binotatus* presented a strong genetic subdivision throughout its entire distribution range (Iberian Peninsula and France) that was likely originated by divergence processes in allopatry as a consequence of population isolation by geographic barriers (Pyrenees) and climatic refugia during the Pleistocene. The existence of a lineage from western and central France indicated that the species long-term persisted in ecologically stable areas located at the present northern limit of its distribution range, rejecting the hypothesis that the origin of such populations has resulted from post-glacial colonization from Iberian climatic refugia (Chapter III).

6. The western and central French lineage of *Chorthippus binotatus binotatus* presented genetic signatures of demographic bottlenecks likely promoted by the severity of climatic fluctuations experienced by its populations at both local and regional scales during the last 21 000 years. The lower habitat stability in northern areas of the species distribution range has led to a negative association between genetic diversity and latitude, a pattern similar to that found in other European organisms (Chapter III).

7. At a lower spatial scale, the patterns of genetic structure of *Chorthippus binotatus binotatus* in southeastern Iberia matched the spatial distribution of the mountain ranges of the region. Such genetic structure was a consequence of neutral differentiation processes promoted by population geographic isolation. Accordingly, isolation by resistance defined by topographic complexity was the best-fitting scenario explaining genetic differentiation among populations in the study area (Chapter IV).

8. Gene flow among populations of the Pyrenean taxon *Chorthippus saulcyi moralesi* was not driven by local adaptation along environmental gradients but shaped by neutral processes resulting from population isolation linked to the rough topography of the region. Despite its small distribution range, this taxon exhibited a strong genetic structure that has probably resulted from its limited dispersal capability and scarce population connectivity across the abrupt relief characterizing the Pyrenees (Chapter V).

9. Divergent selection mediated by climatic gradients seems to be responsible of inter-population phenotypic differentiation for all studied traits in *Chorthippus saulcyi moralesi.* Particularly, the development of forewings was associated to a temperature gradient, a fact suggesting that genetic variation in loci involved in the expression of this trait responds to local climatic regimes (Chapter VI).

10. The populations of *Chorthippus saulcyi moralesi* located at the margins of its distribution range presented lower genetic diversity and longer forewings in comparison to those located at the species distribution core. This finding suggests that peripheral populations experience low demographic stability that reduces local effective population sizes and leads to selection toward more dispersive phenotypes (Chapter VI).

AGRADECIMIENTOS

Ciudad Real, 24 de mayo de 2017

Llegado este momento, me gustaría dedicar unas palabras de agradecimiento a las personas que han colaborado activamente en el desarrollo de esta Tesis Doctoral. Aprovechando la oportunidad que me brinda este escenario, también voy a dar las gracias a otras tantas personas que, aunque ajenas a los detalles de esta Tesis, han permanecido a mi lado durante este tiempo haciéndome disfrutar de la vida. Intentaré ser breve…

Esta Tesis Doctoral tiene otros dos grandes protagonistas: Joaquín y Pedro, sin cuya dirección científica e implicación personal el desarrollo de este trabajo no hubiera sido posible. Vosotros me habéis proporcionado los elementos con los que todo investigador pre-doctoral debe contar para llevar a cabo el inicio de su (larga) formación científica. Me habéis otorgado el apoyo, respeto, confianza, libertad e independencia necesarias para que yo me aventurará autónomamente en la creación y ejecución de los trabajos que conforman esta Tesis, siempre bajo vuestra acertada y cercana supervisión pero respetando mi modus operandi y mi manera de sentir e interpretar el fenómeno. Más allá de la amistad que nos une, también quiero agradeceros que me hayáis alentado y animado en esta carrera cuando en ocasiones hizo falta. Pero esto no lo hicisteis actuando como "jefes", sino como psicólogos que interpretasteis exquisitamente mis puntos fuertes y debilidades, y tras ello, me planteasteis nuevos retos con el objetivo de ir siempre un paso más allá, en dirección a la cresta de la ola. Espero que hayáis disfrutado de esta placentera aventura tanto como yo.

José Miguel, gracias por el soporte computacional que me proporcionaste con tus ordenadores, en los cuales corrieron durante semanas y semanas gran parte de los análisis que se muestran en esta Tesis. A Raúl Bonal, por las interesantes discusiones mantenidas y su continuo apoyo.

I also would like to thank to Bernard Defaut and David Morichon for allowing me to use their beautiful images of some specimens from Chorthippus group binotatus. I am very grateful for their exhaustive studies on such species complex. The field work and other achievements of this PhD Thesis would have been impossible without the accurate and valuable information included in their previous studies.

Si hablamos de PCRs, ADN, autoclaves y centrifugas y etc, es momento de hablar de Conchi y María Santoro. Yo sin vosotras, a fecha de hoy, aún estaría intentando abrocharme la bata mientras intento sacar la cabeza por una manga. Vuestra labor como formadoras fue sublime. Que grandes maestras, que andaluzas, que arte, que alegría, y que grandes amigas con las que he compartido momentos tan divertidos. Dávila, gracias a ti también por resolver numerosas dudas técnicas en el laboratorio y por acompañarlas de tus hilarantes e interminables anécdotas vitales. Y a ti, Calero, por tu inestimable compañía y amigables conversaciones durante las largas jornadas de pipeteo. Otra andaluza importante es Ana Benítez, que cordobesa tan buena gente y divertida. A ti también te debo muchas cosas, amiga. Gracias por involucrarte tan desinteresadamente en la etapa final de esta Tesis, y gracias por tu apoyo incondicional tanto fuera como dentro del IREC, por los transectos y por hacer de Omega nuestro himno vital. Andaluces, felices y sensacionales también son Marisa y Jesús. Gracias a vosotros también por abrirme vuestros corazones, hacerme sentir como en casa y ofrecerme vuestro cariño.

¡Vicente! Gracias por tu apoyo y colaboración en esta Tesis. Sin tu ayuda, las PST aún estarían saliéndome negativas. Gracias por enseñarme que es el proceso Browniano, Early-Burst, Ornstein-Uhlenbeck, y etc, y sobre todo por los boquerones en vinagre, los paseos en La Alameda, y esos meses tan geniales en Sevilla.

Si se menciona Sevilla, cabe nombrar a Carlangas. La experiencia en la EBD no hubiera sido lo mismo sin ti. Gracias por alojarme en tu piso durante mis dos primeros meses en la capital hispalense. Pero Carlos, sobre todo, gracias por sacarme numerosas veces del despacho, de las hipótesis y de los análisis, por dar la vuelta a mis esquemas, por zarandear mis ideales y por llevarme muy lejos con tus pensamientos, tu música y tu poesía, que tras muchos años juntos finalmente he comprendido que es lo que eres tú.

Tras dos años de soledad en mi despacho, llegó nueva sangre a la rebelión. Entonces, el P-37 se tiñó lenta y silenciosamente de una calmada esencia manchega. MJ, gracias por dotar al despacho de equilibrio y armonía, lo cual me ha permitido encontrar tan fácilmente profundos momentos de concentración. Gracias por tu amistad, respeto, confianza y especialmente por querer involucrarte tanto en las mejores como en las peores confidencias que han escuchado las paredes de ese despacho. Gracias también por tu paciencia cuando mi humor no era el mejor o cuando te respondía a tus preguntas horas después de que las formularas. No sería justo si no hablará aquí otra vez de María Santoro; que es, sencillamente, la alegría del IREC. Santoro, siente como tuyas también las anteriores palabras cuando las leas. Las dos habéis sido responsables de que yo alcanzará esta meta sin una disminución notable de mi salud mental. Gracias por vuestros rescates de emergencia que tomaron forma de fiestas y bailes, cuando al marcharos por las tardes me descubríais en pleno viaje astral. Gracias de corazón, también de parte de Gabriel y los enanos.

A Jorgito Yepes, por recibirme cada mediodía con una nueva reflexión vital. Durante años, has sido la primera persona (humana) con la que he hablado cada mañana. Gracias por tus conversaciones sobre la Teoría de Cuerdas, la Teoría de la Relatividad, el Principio de Incertidumbre de Heisenberg, y etc (aquí es mejor no enumerar otros temas…), por ayudarme en innumerables detalles logísticos y sobre todo por valorar mi trabajo e interesarte sobre especiación, relaciones filogenéticas y la vida y obra de nuestros queridos saltamontes. Eres una persona genial, aunque no seas del Atleti. Esta Tesis también va a dedicada a ti. Con ello deseo poner de relieve la importancia de los trabajadores que sustentan la estructura funcional de los centros de investigación desde sus labores de administración, gestión, limpieza o mantenimiento. Que sirvan estas breves palabras para reconocer lo indispensable que resultan vuestras misiones. Este agradecimiento va dirigido a Almudena, Carolina, Lucía, Pilar, Emilia, Juan Carlos y Arturo. Gracias a todas y todos.

Un elemento fundamental que atraviesa transversalmente y define esta Tesis Doctoral han sido mis estancias de investigación en la Universidad de Michigan. At this point, I would like to thank to Lacey Knowles for giving me the opportunity to join to her lab during my internships. There, from her hands, I learnt the importance of thinking big in science. Also, I would like to thank to Anna, Carlos, Melisa, Marcelo, JP, Qixin, Lucy, Andrea, Joyce, Rob y Jesusito for their help in the lab, for many interesting talks about science and also for making easier and happier my internships in Ann Arbor. Mis estancias en Ann Arbor no podrían ser entendidas sin la aparición de Anna, mi amiga Bayesiana. Tu modo de pensar y proceder en esto que hacemos llamar "investigación" es un ejemplo para mí. Gracias por todo, gracias no solo por enseñarme tanto sobre el empleo de metodologías y softwares, sino también por compartir juntos tan buenos momentos en la vida real (aquella que no sabe de índices de impacto ni de revisores). La gélida primavera norteamericana del 2014 ha dado paso a una agradable y permanente verano chipriota. Enhorabuena!

Ένα μεγάλο μέρος αυτής της διδακτορικής διατριβής πραγματοποιήθηκε στο Ann Arbor, το San Francisco, το Berkeley, το Seattle, τη Νέα Υόρκη, τη Ζυρίχη, τη Μαδρίτη, το Cádiz, την Κρήτη και την Αθήνα. Μαρία, εσύ ήσουν μαζί μου σε όλους αυτούς τους τόπους. Ανέχτηκες στωικά τις απουσίες μου και νοιάστηκες για να προχωρήσει καλά αυτό το διδακτορικό. Τις φτωχές μου γνώσεις της αγγλικής γλώσσας και το περιορισμένο λεξιλόγιό μου στα ελληνικά, τα οφείλω σε σένα και στην υπομονή σου. Εσύ μου έμαθες πόσο σημαντικές είναι οι προσομοιώσεις, όπως επίσης να σκέφτομαι σε βάθος κάθε αποτέλεσμα των αναλύσεών μου. Απο εσένα έμαθα ότι ο Ατλαντικός είναι μόνο ένα στενό ποτάμι, και έμαθα να απολαμβάνω αυτό το κομμάτι της ζωής όπου η δουλειά δε χωράει, ούτε είναι αναγκαία. Εσύ με έκανες καλύτερο άνθρωπο και αυτή η διδακτορική διατριβή για μεσογειακές ακρίδες είναι αφιερωμένη επίσης σε σένα.

A mediados de 2016, tras tanta estancia y tanto tiempo fuera, volví a Ciudad Real. Y descubrí que durante mis ausencias, el sector hispano y latino habían confluido, y posteriormente decretado una rigurosa estrategia de ofrendas y dádivas a la amistad y al ocio (especialmente nocturno); y a ella me abracé. Lo admito. Como habréis intuido, es momento de hablar de vosotros y vosotras: Felipe (procedente de las Tierras de Calatrava más intelectuales y amistosas), Vladi (del Mexico de Durito), Jhon (del Pacífico colombiano), Jordi (de la Lérida más manchega), Chai (del Almagro más nepalí) y Sandra (del Toledo más *country*); y por supuesto, nuevamente Dávila, Santoro y MJ. Quizá, hayáis retrasado ligeramente la consecución de esta Tesis (es broma), pero como lo hemos gozado… Gracias por formar parte y construir esta historia, que sin vosotros y vosotras hubiera sido tan solo una Tesis Doctoral más.

Un capítulo aparte merece Emma, que aunque ajena al mundo de la investigación has sido muy consciente del significado de realizar una Tesis Doctoral y en consecuencia, has padecido los efectos indirectos de la etapa final de su desarrollo. A pesar de ello, no has dudado en mostrarme lo mejor de ti: tu jovialidad y tu cariño. Gracias por cuidarme tanto y proveerme de alimentos, y especialmente por atreverte a mirar a través del cascarón de mi apariencia, e incluso ir más allá y explorar mis rincones más profundos. Has adornado nuestros paseos y las copiosas cenas de cazón en adobo con tu fabuloso y sincero humor con el cual me has dado el ánimo más sincero y puro para poder terminar felizmente está Tesis Doctoral. De repente, sin buscarlo, te has convertido en el mejor regalo de este final de fiesta. Ahora te debo un baile, un buen bocata de jamón y una canción.

Llegados a este punto, no puedo olvidarme de los amigos que me han dado su cariño y su apoyo en estos años. Amigos y amigas, que a pesar de la distancia y del poco tiempo que les he podido dedicar, han permanecido a mi lado. Con especial énfasis quiero agradecer a aquellos que me han visitado en Ciudad Real y que sufrieron las consecuencias físicas y mentales de dichas incursiones en tierras culipardas: Lean, Topo, Carlangas, Loquillo, Txupa, Irene R., Cristian... todos grandes amigos y amigas desde hace mucho tiempo, desde antes de que supiéramos lo que era una tesis, el trabajo, el paro o las oposiciones. A todos, y a otros tantos que se me olvidan, gracias. Robleño, gracias por funcionar como mi columna y contrafuerte desde el año 2004, y por despertarme los domingos por la mañana con tus llamadas. A este grupo de amigos también hay que añadir a Miguelito, Alfredito, Rosita, Morente, Raul, Eva, Bea… entusiastas biólogos y biólogas de profesión y vocación, y protagonistas de nuestros primeros pasos en la observación y estudio de diferentes grupos de seres vivos. En este sentido, agradecimiento especial requiere la persona del Dr. Pérez-Granados, quien me inculcó de primera mano la pasión por investigar. Todo comenzó con la ilusión de descubrir a ese fantasma de los páramos, la voluntad de llevar a cabo un estudio científico *amateur* de un modo autónomo y sin recursos económicos, con tan solo las ideas y el esfuerzo de ver amanecer en la Soria esteparia.

Y para terminar, mi familia. Para ellos, no tengo palabras; cualquier frase de agradecimiento que les dedicará, por profunda y elaborada que fuera, sería incapaz de describir fielmente mis emociones. A ellos va dedicada realmente esta Tesis Doctoral; por la ilusión, cariño, humildad y carácter con los que han impregnado el ambiente en el que he crecido y forjado como persona. Si soy como soy, si es que soy algo, es gracias a mi padre, a mi madre, a mi hermana, a mis tíos, a mis tías, así como a mis abuelas y mis abuelos… Estos últimos, mis antepasados, son ejemplos de personas humildes procedentes de pequeñas aldeas castellanas arraigadas a la tierra, y que como otras tantas familias emigraron a grandes ciudades con el único propósito de proporcionar nuevas oportunidades a sus hijos y a sus nietos, y de este modo intentar verles prosperar y desarrollarse a luz de la educación, el estudio, el conocimiento… Ellos merecen todo mi respeto, cariño y eterno agradecimiento.

Gracias a todos y todas.

"Todos somos genios.
Pero si juzgamos a un pez por su capacidad de subir a un árbol,
pasará toda la vida creyendo que es un estúpido"

Albert Einsten (1879-1955)