UNIVERSIDAD DE GRANADA
DEPARTAMENTO DE GENÉTICA
DEPARTAMENTO DE ECOLOGÍA

Erysimum mediohispanicum at the evolutionary crossroad: phylogeography, phenotype, and pollinators

Tesis Doctoral

Antonio Jesús Muñoz Pajares

Granada 2013

Editor: Editorial de la Universidad de Granada
Autor: Antonio Jesús Muñoz Pajares
D.L.: GR 1888-2013
ISBN: 978-84-9028-613-5

Erysimum mediohispanicum at the evolutionary crossroad: phylogeography, phenotype, and pollinators

Memoria que el Licenciado Antonio Jesús Muñoz Pajares
presenta para aspirar al Grado de Doctor
por la Universidad de Granada

Esta memoria ha sido realizada bajo la dirección de:
Dr. Francisco Perfectti Álvarez y
Dr. José María Gómez Reyes

Ldo. Antonio Jesús Muñoz Pajares
Aspirante al Grado de Doctor

Granada, enero de 2013

Dr. Francisco Perfectti Álvarez,
Profesor Titular de Genética de la Universidad de Granada
y Dr. José María Gómez Reyes,
Catedrático de Ecología de la Universidad de Granada

CERTIFICAN

Que los trabajos de investigación desarrollados en la Memoria de Tesis Doctoral: "*Erysimum mediohispanicum* at the evolutionary crossroad: phylogeography, phenotype, and pollinators", son aptos para ser presentados por el Ldo. Antonio Jesús Muñoz Pajares ante el Tribunal que en su día se designe, para aspirar al Grado de Doctor por la Universidad de Granada.

Y para que así conste, en cumplimiento de las disposiciones vigentes, extendemos el presente certificado a 16 de enero de 2013

Dr. Francisco Perfectti Álvarez   Dr. José María Gómez Reyes

Durante el tiempo de realización de esta Tesis Doctoral he disfrutado de una Beca del Programa Nacional de Formación de Personal Universitario del Ministerio de Educación y Ciencia
Ref. (AP2006-00685).

Este trabajo estuvo financiado por los proyectos EVOFLOR (CGL2006–04883) y MONTES-CONSOLIDER Consolider-Ingenio (CSD2008-00040), del Ministerio español de Educación y Ciencias; (078/2007), del Ministerio español de Medio Ambiente y (P07-RNM-02869), de la Junta de Andalucía.

La investigación presentada en esta Tesis Doctoral se ha realizado en el Departamento de Genética y en el Departamento Ecología de la Universidad de Granada.

A los que me quieren.

¡Lo que ve el que vive!
(si sabe mirar...)
Isabel López (mi abuela)

Por muy larga que sea la tormenta,
el sol siempre vuelve a brillar entre las nubes.
Khalil Gibran

AGRADECIMIENTOS

Pensé que nunca llegaría este momento.

Delante del que será el último papel en blanco de esta Tesis Doctoral empiezo a recordar a todas las personas responsables de que este libro exista. Todos los que me llevaron a eligir el camino de la investigación. Todos los que me han enseñado a avanzar por él. Todos los que lo han recorrido a mi lado. Los que me han alumbrado el camino en las tinieblas. Los que me han ayudado a levantarme tras las caídas. Los que me animaron a ver la meta más cerca. Los que me curaron las rodillas ensangrentadas. De todo corazón, muchas gracias, porque sin vosotros sé que no lo habría conseguido.

Esta Tesis fue concebida por Francisco Perfectti y José María Gómez. A ellos agradezco que me brindaran la oportunidad de trabajar en un sistema complejo e interesante a partes iguales y que me enseñaran a mirarlo desde diferentes perspectivas, como sólo puede hacerse desde un grupo multidisciplinar. Con ellos empecé a entender los entresijos de la genética molecular, a disfrutar de los muestreos de campo y a comprender los detalles de la publicación de artículos en revistas. Gracias a ellos siento que puedo afrontar cualquier problema científico en el futuro. Gracias, Paco. Gracias, Rocka.

El tercer responsable de que haya dedicado estos años al estudio de *E. mediohispanicum* es Mohamed Abdelaziz, con quien he hecho dos Tesis a medias. Él me convenció de que me embarcara en esta línea cuando tenía ya pie y medio lejos de la Facultad de Ciencias. Con él he pasado la mayoría de horas de muestreo de campo, de campana en laboratorio, de pelea con el ordenador y de discusión de ideas en cualquier parte. Aunque le gusta jactarse de que en muchas ocasiones ha tenido que hacer funciones más propias de secretario que de compañero de trabajo, debo decir que realmente siempre ha actuado más como un amigo de los que se encuentran pocos, que como cualquier otra cosa. Gracias, Gorrión.

Sin duda que este camino no habría llegado a su destino sin Belén, quien, mientras la dejaron, sintió este trabajo más suyo que yo mismo. No es normal que haya que pedirle a una técnico que deje de dedicar horas extras a analizar tus muestras. No es normal que le quite el sueño acabar hoy en lugar de mañana. Pero es que Belén no es normal. No sólo por su forma de trabajar, si no por su forma de dar siempre más de lo que tiene a la gente que le rodea. Espero poder encontrarme muchos años dentro de esa gente privilegiada. Gracias, Belén.

Esta Tesis ha estado vinculada al Departamento de Genética, donde me han dado todas las facilidades para realizar tanto actividades investigadoras como docentes. Me gustaría agradecer a todos sus miembros por hacer del Departamento mi segundo hogar. Quiero destacar la labor de Esther Viseras, que ha estado siempre dispuesta a facilitar el trabajo y los trámites ya fuera con una firma, con una llamada o con un buen consejo. A ella debo, además, haber perdido parte del miedo que se siente al enfrentarse a una clase llena, gracias a su mentorización en el curso de profesores noveles. Siempre consideraré a Juan Pedro como mi padre científico, pues él fue capaz de hacer que un químico de segundo año pudiera aprender biología evolutiva en los libros, en sus clases, en sus tutorías y de una manera que yo no podía imaginar: participando en su grupo de investigación. Juan Pedro me abrió las puertas de este mundillo y por ello le deberé cualquier cosa que pueda conseguir en él. Agradezco a Pepi, además de sus lecciones de citogenética en el laboratorio, el cariño con el que siempre me ha tratado y su preocupación por mi bienestar en todo momento. A Lola, con quien disfruté horas de microscopio entre risas durante mis primeros años en el Grupo de Genética Evolutiva. A Ángel, al que basta "pitar" para que aparezca, como un superhéroe, a resolver cualquier problema informático. Gracias a él y Paquillo he podido manejarme con LaTex y Lyx para darle a esta Tesis el formato que tenía en mente. A Mohammed por no cejar en su empeño de hacerme ver la parte positiva de las cosas. A Eli, que desde que entré a formar parte del grupo dejó claro que se trabaja mucho mejor entre amigos que entre compañeros. Jamás podré olvidar su risa, sus ensayos de voz para el coro, las discusiones con rotulador y pizarra y su repulsión incontrolable por algunos chistes (y que tanto nos hizo reir a los que somos un poco quemasangres). A María Teruel por iniciarme en el mundo de las pipetas. A Tati, Bea, Mercy, Javi Valverde, Modesto, Rubén, María Lucena, Eva, Alejandra (la maestra): muchas gracias por hacer que las largas horas en la becaría fueran un poco más cortas.

Debido a que mis dos directores no forman un grupo en el sentido clásico, mi Tesis se ha desarrollado a medias en el Departamento de Ecologia. Agradezco a Rafael Morales y a José María Conde que me abrieran las puertas del

Departamento como a un ecólogo más y a Regino, José Antonio y Jorge las de su laboratorio, permitiendo que usáramos instrumentos del mismo. A Matías, ese McGiver de Purchena convertido en Braveheart del calentamiento global. Gracias por enseñarme lo que significa meticulosidad, por la ayuda con los papeleos, por las meriendas que se alargaban hasta después de la cena, por las palizas al mus. Pero sobre todo, gracias por ser un ejemplo a seguir en todos los sentidos. A Asier le agradezco su ayuda con el ArcGis y esas charlas por los pasillos, jarra de agua en mano. A Yedra y Leticia, que estuvieron dispuestas a echar una mano en lo que hiciera falta, siempre con una gran sonrisa. A Eulogio y Nacho, por su ayuda con las cajas de exclusión que, por desgracia, no llevaron a nada. A los que me acogieron en el Departamento como uno más: Susana, Ramón, Juande, Karim, Ana, Sara y Ayub. A los que estaban cuando llegué y dejaron un gran vacío al cambiar de aires: Javi Valdés, Ángela (¡¡Causty!!), Oti y José Luis.

A Paolo, por su ayuda con los análisis de fenotipo y su paciencia infinita en obtener los frutos en forma de publicación. A él, Mattia y Juan Lorite agradezco la ayuda que me dieron en los muestreos de campo. A Alfonso, Pepa y Juan Miguel por facilitar la odiosa burocracia.

A Xavier Picó, que se ofreció continuamente a resolver problemas, discutir ideas, analizar datos o escribir artículos de la manera más eficaz y concienzuda que he visto nunca. Además de agradecerle su disposición permanente tengo que pedirle disculpas por los múltiples asaltos a mano armada y sobre la bocina que ha sufrido por mi parte y a los que siempre respondió con gusto.

Jordi Bosch también sufrió estos asaltos de última hora a menudo, siempre accediendo a mis peticiones con su característica amabilidad extrema. En una de ellas me aceptó en su grupo durante unos meses y me enseñó lo poco que yo puedo aprender sobre clasificación de insectos. Con él hice los primeros muestreos lejos de Sierra Nevada y me quedé boquiabierto al ver cómo pueden identificarse pequeños puntos voladores hasta el nivel de género o especie (¡¡a veces hasta macho o hembra!!).

Estoy muy agradecido también al Profesor Michael Stumpf por permitirme trabajar en su laboratorio, dedicándome toda su atención durante los meses que pasé junto a él. A Christopher Barnes, que tutorizó el día a día de mi aprendizaje en métodos ABC durante esa estancia. Ellos, junto al resto del grupo de Teoría de Sistemas Biológicos (Suhail Islam, John Pinney, Rodrigo Liberal, Michal Komorowski, Maxime Huver, Heather Harrington, Georgia

Chan, Xia Sheng, Nathan Harmston) hicieron que mi paso por Londres fuera una de las mejores experiencias de mi vida.

A todos esos amigos para los que, muy a mi pesar, no he tenido todo el tiempo que merecían. A Sara (que casi me adelanta) y a Jose (siempre será El Canario), que me han dado fuerzas para seguir adelante cada poco tiempo con sus mails y sus llamadas. A Isa y David y sus visitas contra-reloj a Granada. A Maria José y Miguel, que me perdonaron un mal viaje a Mallorca. A Palma: por años que pasen sin vernos seguiremos hablando donde lo dejamos ayer. A Juan Diego, sus cursos y la comilona que nunca llegó. Gracias a todos por estar tan pendientes de mí estos años.

También me gustaría dedicar unas palabras a aquellos que, más que un ejemplo a seguir, han sido un molde del que alejarme. A ellos les estoy profundamente agradecido porque no hay mejor manera de calibrar el efecto de nuestras acciones que sintiendo en nuestras carnes las de los demás. Por eso, gracias de corazón a los que me han enseñado las actuaciones de las que tengo que alejarme.

Poco tiempo he tenido también para mis abuelos, mis tíos y mis primos, pero sé que me han tenido en mente, como han hecho desde que nací. Muchas gracias por ello a mis Encarnitas, Antonios y Pepes, a Enriqueta, Juanita, Verónica, Paco, Eli, Maria Eugenia, Esther y Blas. Recuerdo especialmente el día que mi tío Pepe usó su maña para arreglar mi coche, que creí que agonizaba justo antes de un muestreo y cómo mi tia Juanita me cargó de roscos para que "se me pasara la *enritación*". A los que me invitaron mil veces a desconectar junto a ellos durante "las próximas vacaciones": Roberto, Carmen, Amparo, Juanjo, Sandra, Antonio, David y Gema (cuando me cobre las invitaciones juntas me faltará verano). A Javi, que desde muy pequeñito me inculcó el vicio de la lectura, el interés por la ciencia y el gusto por las matemáticas. Esta Tesis, sin duda, es en parte culpa suya.

No puedo acabar los agradecimientos sin mencionar a las personas a las que debo todo, mucho más allá de las páginas que siguen a ésta. A mi abuela (mi mama Virgina), que me crió mientras mis padres se pasaban la mayor parte del día trabajando para poder salir adelante. A Laura, por los seis años y pico que separan Cuba de Antequera. Por sus intentos de calmar mis agobios, su aliento en los momentos de desánimo, por su ayuda aunque no se la pidiera, por su paciencia con mi desorden, por estar ahí cada día para lo bueno y para lo malo. A mis hermanas Leticia y Araceli. Por las canciones de piratas para atar zapatillas, por los partidos de fútbol en el patio y la pantalla, por las radios con micrófono, las pizarras, los vuelos de supermán, las estanterías que se caen en la noche, el lagrimeo al ver un ojo más cerca de la cuenta... Cuando llegasteis ya estaba yo aquí, pero os puedo asegurar que el mundo es mucho mejor desde que estáis vosotras en él. Es muy difícil plasmar con palabras la marca que en uno pueden dejar los mejores padres del mundo. No hay manera posible de expresar lo que un hijo siente al crecer, mirar atrás y darse cuenta de todo lo que han hecho por él. Al ser consciente de que todo lo bueno que pueda conseguir no será más que el reflejo de lo que le inculcaron desde la cuna. Gracias por darme la educación que no se aprende en ningún otro sitio. Gracias por pensar en mí antes que en vosotros. Gracias por infinidad de cosas. Gracias papá. Gracias mamá.

Finalmente llegó el momento y no puedo dejar de daros las gracias a todos.

RESUMEN

Las interacciones entre plantas y polinizadores, además de depender de una gran variedad de factores, pueden tener importantes consecuencias evolutivas. Debido a que los polinizadores muestran patrones de preferencia ante determinados rasgos fenotípicos, dichos caracteres influyen en el tipo y la frecuencia de polinizadores que visitan la planta. De esta manera, es esperable que esos rasgos sean seleccionados, dándose un proceso de coevolución que reducirá las varianzas fenotípicas para ajustar las morfologías florales y la de sus principales especies de polinizadores. Este proceso puede desembocar en una especialización de la interacción, haciendo que un número muy reducido de visitantes florales se encarguen de la polinización de la planta. Sin embargo, existen en la naturaleza plantas que son visitadas por un número muy elevado de polinizadores, cada uno con preferencias, patrones de forrajeo y morfologías diferentes, e incluso contrarias. Como consecuencia, es difícil predecir la influencia de los polinizadores en la evolución fenotípica de estas especies de plantas generalistas.

La interacción entre plantas y polinizadores no sólo influye en las características fenotípicas de ambos, sino que también puede repercutir en la composición genética de las poblaciones de plantas. Sin embargo, el conocimiento que se tiene sobre la importancia relativa de los polinizadores con respecto al resto de factores que determinan la evolución de las especies de plantas es muy limitado. En la presente Tesis Doctoral abordamos el estudio de esta cuestión, utilizando como sistema de estudio *Erysimum mediohispanicum*, una planta herbácea mediterránea, endémica de la Península Ibérica y extremada-

mente generalista en cuanto al número de especies de polinizadores que la visitan. *E. mediohispanicum* pertenece a un grupo especialmente conflictivo en cuanto a su taxonomía, debido a su origen reciente y a la existencia de flujo génico con otras especies cercanas.

En nuestro estudio hemos obtenido información de un total de 56 poblaciones de *E. mediohispanicum* para inferir los patrones filogeográficos y la estructuración genética de la especie. Además, para cada población hemos descrito la diversidad genética y el gremio de polinizadores, así como la media y la varianza de algunos rasgos fenotípicos involucrados en la interacción entre plantas y polinizadores. Nuestro diseño de muestreo comprende la distribución completa de la especie, que incluye poblaciones diploides (principalmente en Sierra Nevada) e hipotetraploides (en Cazorla y el noreste peninsular). Aunque hemos encontrado rasgos que difieren significativamente entre citotipos, ambos mostraron interacciones similares con los polinizadores.

Para obtener los patrones filogeográficos de la especie hemos analizado la secuencia de una región plastidial que presenta una frecuencia muy alta de inserciones y deleciones (indels) en crucíferas. Aunque este tipo de mutaciones hacen que la región sea altamente variable, el uso de la misma para las inferencias filogenética y filogeográfica es limitado. Esto se debe a los problemas asociados a la interpretación de la información evolutiva contenida en los indels. En esta Tesis Doctoral describimos un procedimiento general para extraer e interpretar dicha información y hemos implementado el método como un conjunto de funciones agrupadas en un paquete del lenguaje R. Dicho método consiste en la obtención de dos matrices de distancias, basadas respectivamente en el número de mutaciones observadas para indels y sustituciones. Ambas matrices pueden ser representadas por separado o combinarse para obtener una visión conjunta de la información proporcionada por ambos tipos de mutaciones. Debido a que nuestro objeto de estudio son poblaciones

de la misma especie, entre las que es probable que exista flujo génico, hemos preferido representar las distancias evolutivas obtenidas mediante redes de percolación.

De acuerdo al análisis filogeográfico, las poblaciones estudiadas se agrupan en cuatro linajes, cada uno con una ubicación geográfica bien delimitada que parece ser el resultado de múltiples eventos de dispersión. Las rutas de colonización compatibles con el conjunto de datos obtenidos sugieren que existió un primer flujo desde Sierra Nevada hacia el norte de la Península Ibérica, seguido de dos eventos norte-sur. El primero ha originado las poblaciones más próximas al litoral sureste, mientras que el segundo ha dejado su huella en las poblaciones del interior, desde la provincia de Soria hasta las Sierras de Cazorla, Segura y La Guillimona. Nuestro estudio remarca la complejidad de la historia filogeográfica de las plantas mediterráneas que habitan la Península Ibérica. Los múltiples eventos sucesivos de colonización, aislamiento y flujo génico hacen que las poblaciones de otras especies de *Erysimum* ibéricos estudiadas no constituyan unidades independientes, sino que se intercalen con las de *E. mediohispanicum* en función de su proximidad geográfica.

Las poblaciones estudiadas fueron extremadamente diversas tanto en sus rasgos fenotípicos como en su composición genética y en ambos casos encontramos patrones de variación geográficos latitudinales y altitudinales. Nuestros resultados sugieren la existencia de relaciones significativas entre la variación de la forma de la corola y la diversidad de polinizadores de las poblaciones, lo que, a su vez, se correlaciona con la diversidad genética de la planta. La forma de la corola parece haber evolucionado siguiendo los patrones filogeográficos de la especie y hemos encontrado, además, que las poblaciones más parecidas en morfologías florales son también visitadas por polinizadores más similares.

La presente Tesis Doctoral destaca la importancia del gremio de polinizadores en las propiedades genéticas y fenotípicas de las poblaciones de plantas, incluso en una especie tan generalista como *E. mediohispanicum*. Además, en ella hemos desarrollado marcadores microsatélites que pueden suponer un aporte importante para el estudio de las múltiples especies de *Erysimum* de la Península Ibérica. Por último, el método de obtención de información evolutiva a partir de indels supone una aportación generalizable no sólo a las crucíferas, sino a cualquier conjunto de organismos cuyas secuencias alineadas muestren una alta frecuencia de inserciones y deleciones.

ÍNDICE GENERAL

Part I

INTRODUCTION

1

INTRODUCCIÓN GENERAL

LA PENÍNSULA IBÉRICA COMO CRUCE DE CAMINOS EVOLUTIVOS

El área de distribución de las especies se ha modificado debido a las fluctuaciones climáticas ocurridas a escala global en los últimos miles de años. En el caso de Europa, a medida que el norte y el centro del continente se cubrían de hielo, los taxones que habitaban esas regiones se desplazaron hacia el sur buscando las condiciones ambientales más favorables que se daban en las Penínsulas Ibérica, Itálica y Balcánica (Taberlet et al., 1998; Hewitt, 2000, 2004; Nieto Feliner, 2011). El aumento global de temperatura característico de los periodos interglaciales llevó a las especies a colonizar de nuevo el centro y el norte de Europa a partir de sus refugios meridionales (Schmitt, 2007; Tzedakis, 2009; Jiménez-Moreno et al., 2010). El relieve montañoso de las tres penínsulas mencionadas permitió a las especies responder a los cambios climáticos, además de con desplazamientos latitudinales, mediante cambios altitudinales en una misma cordillera (Schmitt, 2007; Tzedakis, 2009; Jiménez-Moreno et al., 2010). Debido a esto, diferentes sistemas montañosos dentro de cada Península pudieron actuar como refugios relativamente independientes (Gómez and Lunt, 2007; Médail and Diadema, 2009; Nieto Feliner, 2011; Fuertes-Aguilar et al., 2011), haciendo que algunas especies (o linajes) permanecieran aisladas durante algún tiempo y luego volvieran a entrar en contacto (Johnsen et al., 2010; Zeng et al., 2011;

Singhal and Moritz, 2012). Por tanto, la región Mediterránea en general y las tres penínsulas en particular, han constituido al mismo tiempo un refugio para plantas relictas, al proporcionarles hábitats adecuados durante los periodo climáticos adversos, y un área de diversificación y especiación, como resultado de contactos secundarios y procesos de hibridación (Médail and Diadema, 2009).

Por todo, la región mediterránea puede calificarse como un cruce de caminos biogeográficos para especies de varios orígenes (Thompson, 2005; Blondel et al., 2010), donde es habitual encontrar patrones filogeográficos difícilmente generalizables ya que cada especie respondió de manera individual a las oscilaciones climáticas según sus requerimientos particulares, la geografía y el ambiente típico de cada localidad (Hewitt, 2000).

La Península Ibérica es, de los tres refugios mencionados, el que cuenta con mayor cantidad de estudios filogeográficos (Rodríguez-Sánchez et al., 2010), especialmente para especies de plantas leñosas, donde se han podido definir dos patrones filogeográficos principales, correspondientes a especies típicas del bosque mediterráneo y a plantas típicas de montaña (Arroyo et al., 2004). Las especies del primer grupo permanecieron aisladas en refugios costeros durante los periodos glaciales y se expandieron hacia el interior de la Península a medida que el clima se hacía más cálido. De acuerdo con el segundo patrón general, las especies de montaña respondieron a las fluctuaciones climáticas mediante cambios altitudinales en su rango de distribución. Aunque dichas especies también mostraron patrones migratorios latitudinales, éstos son mucho más difícilmente generalizables, ya que se han observado filogeografías discordantes entre especies pertenecientes al mismo género (Stebbins, 1984; Hillis et al., 1996; Barraclough and Nee, 2001) y entre especies con distribuciones y hábitats similares (Vargas, 2003).

En los últimos años se ha producido un incremento en el número de estudios filogeográficos tanto de plantas herbáceas como mediterráneas (e.g., Ortiz et al., 2008; Guzmán and Vargas, 2009; Médail and Diadema, 2009; Fernández-Mazuecos and Vargas, 2011; Fuertes-Aguilar et al., 2011; Alarcón et al., 2012; Valtueña et al., 2012; Beatty and Provan, 2012; Vrancken et al., 2012). Muchos de estos trabajos sugieren la existencia de distribuciones fragmentadas en múltiples refugios glaciales, pero son necesarios más estudios sobre linajes genuinamente mediterráneos para determinar si existen patrones filogeográficos generales en las especies herbáceas de la Península Ibérica.

filogeografía

Debido a que diferentes marcadores y técnicas permiten resolver distintas cuestiones evolutivas (Nybom, 2004), es fundamental elegir aquéllos que se ajusten mejor a los objetivos del trabajo y las características de la especie estudiada. La finalidad fundamental de la filogeografía es el estudio de las relaciones entre filogenia y distribución geográfica de variantes intraespecíficas y especies estrechamente emparentadas (Avise, 2000). A pesar de que la tasa de sustitución del ADN nuclear suele ser mayor que la de los dos orgánulos (Wolfe et al., 1987), los polimorfismos neutros del ADN plastidial (cpDNA) tienen más potencia que los polimorfismos nucleares para resolver las historias evolutivas de plantas (Hamilton et al., 2003), por lo que han sido ampliamente utilizados (Shaw et al., 2005).

El desarrollo de la tecnología de secuenciación permitió el estudio de regiones funcionales del cpDNA para inferir las relaciones filogenéticas entre familias, basadas en un solo gen al principio (por ejemplo, entre angiospermas con el gen rbcL; Chase et al., 1993) o mediante la concatenación de varios genes, poco después (como el caso de escrofulariáceas, basado en rbcL y ndhF; Olmstead and Sweere, 1994). Debido a que las regiones no codificantes suelen estar sometidas a menos constricciones funcionales que las regiones codificantes, el estudio de sus secuencias puede proporcionar mayores niveles de variación, lo que incrementa la capacidad resolutiva de los análisis filogenéticos (Gielly and Taberlet, 1994). Por esta razón, las regiones no codificantes del cpDNA han sido ampliamente estudiadas para explorar las relaciones evolutivas entre los niveles taxonómicas más bajos (Taberlet et al., 1991; Golenberg et al., 1993; Johnson and Soltis, 1994). Por último, si se quieren conocer características poblacionales como la diversidad o la estructuración genética, suele recurrirse a métodos basados en marcadores aún más variables, como son los basados en tamaño de banda (Petit et al., 2005). De entre estos métodos destacan los que implican corte del ADN mediante encimas de restricción (RFLP, AFLP), amplificación con PCR usando cebadores poco específicos (RAPD) y la amplificación de regiones con repeticiones en tándem de secuencias cortas (microsatélites o SSR).

Uno de los marcadores basados en secuencia más comúnmente usados en estudios filogenéticos de plantas es la región de cpDNA comprendida entre los genes transferentes de leucina (trnL_{UAA}) y fenilalanina (trnF_{GAA}), de aquí en adelante, trnL y trnF, respectivamente (Stuessy, 2009). Ambos genes están separados por una secuencia intergénica (trnL-trnF IGS) que, en algunas especies de plantas, está compuesta parcialmente por pseudogenes del trnF originados por duplicación del gen funcional. Los principales mecanismos involucrados en dicha duplicación parecen ser la recombinación desigual (tanto entre distintas regiones del mismo cromosoma como entre distintas copias idénticas presentes en un mismo orgánulo) y el desplazamiento de hebra durante la replicación del ADN (Dobeš et al., 2007).

La región trnL-trnF IGS ha sido descrita en profundidad en crucíferas (Koch et al., 2007), especialmente en *Arabidopsis* y *Boechera* (Dobeš et al., 2004, 2007), donde se ha encontrado polimorfismo en el número de pseudogenes entre especies, poblaciones e incluso entre individuos de una misma población (Koch et al., 2008; Tedder et al., 2010). Las diferencias en el número de pseudogenes presentes en cada organismo produce una alta variabilidad en las longitudes de sus secuencias. Para estudiar la historia evolutiva de secuencias homólogas con diferente longitud es necesario incorporar huecos (o gaps) en aquéllas posiciones de la secuencia que no presenten nucleótidos homólogos con las demás. Dichas posiciones han sido producidas por procesos que implican la inserción o deleción de secuencias, denominados de manera genérica "indels".

En los últimos años se ha incrementado el número de estudios que intentan extraer la información evolutiva de la región trnL-trnF IGS para dilucidar tanto divergencia evolutiva antigua (por ejemplo, las angiospermas basales, Borsch et al., 2003) como reciente (por ejemplo, entre poblaciones de la misma especie, Dobeš et al., 2004; Tedder et al., 2010), incluyendo la variación en indels de diferentes formas (Koch et al., 2005, 2007; Koch and Matschinger, 2007; Schmickl et al., 2008).

INDELS COMO MARCADOR FILOGEOGRÁFICO

La información proporcionada por los indels, combinada con la obtenida a partir de las sustituciones, puede mejorar la capacidad para resolver las relaciones evolutivas entre organismos (Simmons et al., 2001; Vogt, 2002; Young and Healy, 2003; Müller, 2006; Blair and Murphy, 2011). La ventaja derivada del uso de los indels puede ser importante si las secuencias muestran bajas tasas de sustitución (Redelings and Suchard, 2007). Sin embargo, obtener dicha información es una tarea compleja y laboriosa, debido principalmente a dos razones: la dificultad para obtener un alineamiento fiable y la ausencia de una metodología estándar para obtener la información contenida en los indels (Golenberg et al., 1993; Castresana, 2000; Loytynoja and Milinkovitch, 2001). Estas dificultades son las responsables de que diferentes autores hayan eliminado del alineamiento las posiciones con gaps (y en muchos casos también las regiones adyacentes) antes de realizar cualquier análisis filogenético (Talavera and Castresana, 2007).

Para realizar una inferencia precisa de las relaciones evolutivas es crucial partir de un alineamiento correcto, ya que la incorporación de gaps de manera correcta permitirá la comparación entre nucleótidos homólogos (Kumar and Filipski, 2007). Aunque existen procedimientos para estimar simultáneamente el alineamiento y la filogenia (e.g. Wheeler, 1996; Wheeler et al., 2003; Lunter et al., 2005; Redelings and Suchard, 2007; Liu et al., 2012), lo más habitual es que ambos procesos se realicen por separado.

El principal escollo a la hora de alinear secuencias con una alta proporción de indels radica en la incertidumbre asociada a la presencia de múltiples posiciones redundantes y/o no homólogas, especialmente si las inserciones se han producido por duplicación. Tanto es así que, aunque se han desarrollado un gran número de programas informáticos para alinear secuencias, todos ellos muestran poca precisión cuando las secuencias presentan un alto número de eventos indel (Liu et al., 2010). Varios autores recomiendan que los alineamientos sean refinados a mano después de usar alguno de estos programas, teniendo en cuenta la estructura secundaria y los mecanismos que subyacen a la evolución molecular de la región estudiada (Kjer, 1995; Kelchner, 2000; Müller, 2006; Kjer et al., 2007; Morrison, 2009; Blair and Murphy, 2011).

El segundo problema fundamental que subyace a la obtención de información evolutiva a partir de indels es la falta de procedimientos estándar para inferir las relaciones evolutivas de posiciones con gap, debido a la dificulad de obtener modelos evolutivos que expliquen satisfactoriamente el proceso mutacional de los mismos (Saakian, 2008). Para poder reconstruir la historia evolutiva de un conjunto de secuencias es necesario conocer (o estimar) el número de eventos mutacionales que se han producido entre cada pareja de secuencias. En el caso de las sustituciones, sabemos que cada evento de mutación hace que un nucleótido cambie a cualquiera de los otros tres, mientras que es difícil predecir el resultado de un evento indel, ya que puede afectar a un número variable de nucleótidos. Además, en cualquier proceso mutacional pueden darse múltiples cambios en una misma posición, haciendo que el número de mutaciones observadas sea menor que el de mutaciones producidas realmente. La magnitud de este error aumenta a medida que lo hace la distancia evolutiva entre las secuencias estudiadas. De nuevo esta limitación es más importante cuando se estudian indels que con sustituciones, donde se puede subsanar (al menos en parte) en base a las probabilidades de transición entre nucleótidos (e.g., Jukes and Cantor, 1969; Tajima and Nei, 1984; Tamura, 1992). En definitiva, inferir la historia evolutiva de un conjunto de secuencias a partir de los indels se ve dificultado también por la incertidumbre asociada tanto al recuento del número de eventos mutacionales observados como a la estima de los eventos realmente producidos.

A pesar de estas dificultades, cada vez más estudios están incluyendo la información contenida en los indels en la reconstrucción filogenética, la mayoría de ellos mediante métodos de máxima parsimonia (Ogden and Rosenberg, 2007), siendo el quinto estado y los métodos de codificación de indels las aproximaciones más comunes. El primero de dichos métodos considera que cada posición del alineamiento puede tener cinco estados distintos, los cuatro correspondientes a cada nucleótido más la ausencia del mismo (es decir, que la posición haya sido eliminada mediante un evento indel). La principal crítica a este procedimiento es que cada posición del alineamiento es considerada como un evento independiente, a pesar de que es habitual que un solo evento mutacional elimine múltiples posiciones adyacentes.

El segundo de los métodos de parsimonia más usados trata los indels como "dato nulo" ("missing data") en el alineamiento original de secuencias, pero su información se incluye mediante la codificación de los mismos como caracteres adicionales. Existen diferentes métodos de codificación, dependiendo del tratamiento que se le dé a las regiones con indels solapados (e.g., Baum et al., 1994; Freudenstein and Chase, 2001) y a la forma de cálculo de la matriz de costes de transformación. Esta matriz contiene el número de eventos indel que se asumen entre cada pareja de secuencias. De entre los métodos de codificación, los más ampliamente usados son los denominados Codificación Simple de Indels (SIC, Simmons and Ochoterena, 2000), Codificación Compleja de Indels (CIC, Simmons and Ochoterena, 2000) y Codificación Compleja de Indels Modificada (MCIC, Müller, 2006). Otros métodos de codificación no consideran los indels como dato nulo en ningún momento, si no que los codifican directamente en sus posiciones originales. Ése es el caso de la Codificación de Caracteres Multiestado (Lutzoni et al., 2000), el Método de Elisión (Wheeler et al., 1995), la Codificación Sensible a Mayúsculas (Swofford, 1989), la Codificación Estirada (Geiger, 2002) y la Codificación de Bloques (Geiger, 2002).

A pesar de que los métodos de máxima parsimonia son los más extendidos, también existen modelos evolutivos en inferencia probabilística (es decir, máxima verosimilitud e inferencia bayesiana) que utilizan la información contenida tanto en sustituciones como en indels (Rivas and Eddy, 2008, y métodos citados por estos autores). Sin embargo, estos modelos son complejos y es habitual que los gaps sean eliminados del alineamiento o tratados como dato nulo sin posterior recodificación cuando se realizan reconstrucciones evolutivas basadas en máxima verosimilitud o en inferencia bayesiana (Ogden and Rosenberg, 2007).

En contraste con la diversidad metodológica mencionada, el análisis de indels con métodos basados en distancias necesita ser explorado (Ogden and Rosenberg, 2007). Por el momento, el tratamiento que los métodos de distancias dan a los gaps se reduce a controlar la cantidad de información proporcionada por ellos que es descartada (Tamura et al., 2011). El método denominado "deleción completa" elimina del alineamiento aquéllas posiciones homólogas que presenten al menos un gap en al menos una de las secuencias alineadas. En el caso de la "deleción por parejas", la eliminación de la posición homóloga se realiza después de comparar cada pareja de secuencias, lo que permite obtener información de posiciones homólogas con gaps cuando las secuencias comparadas presentan un nucleótido en dicha posición.

Aunque la precisión filogenética de los métodos de distancias se ha considerado menor que la de otros métodos (Talavera and Castresana, 2007; Dwivedi and Gadagkar, 2009), recientemente se ha demostrado que son muy poderosos tanto para árboles evolutivos largos como para aquéllos con ramas cortas (independientemente de la profundidad del árbol; Roch, 2010). En este último caso, los métodos de distancias permiten resolver fácilmente el problema de la representación filogenética derivada de la poca divergencia entre los organismos estudiados, mediante la aplicación de cualquier método de redes a la matriz de distancias obtenida: La evolución suele representarse mediante árboles en los que cada taxón sólo puede estar vinculado directamente a un único ancestro. Sin embargo, esta premisa es violada si existe flujo génico entre los organismos estudiados (Makarenkov and Legendre, 2004), algo que es más frecuente cuanto más reciente sea la relación entre ellos. En el caso de la filogeografía, las unidades de estudio suelen ser poblaciones de la misma especie, entre las que es esperable que exista migración y flujo génico, por lo que la topología en forma de árbol seguramente no representa correctamente las relaciones evolutivas existentes (Morrison, 2005;

Mardulyn, 2012). Por el contrario, la visualización en forma de red permite que un taxón se conecte con más de un ancestro, lo que refleja mejor la incertidumbre asociada a la genealogía poblacional.

ESTRUCTURACIÓN, DIVERSIDAD Y FLUJO GÉNICO

Además de las relaciones de ancestría, para tener un conocimiento completo de la historia evolutiva de un conjunto de poblaciones es necesario conocer tanto la diversidad genética de cada una de ellas, como la diferenciación genética existente entre cada pareja de poblaciones. La primera nos informa de lo diferentes que son genéticamente los individuos que coexisten en una misma población, y de los valores obtenidos pueden inferirse, por ejemplo, patrones de flujo génico intrapoblacional o la existencia de pequeños tamaños poblacionales en el pasado. Por su parte, la diferenciación poblacional nos da idea de lo distintos genéticamente que son los individuos que habitan en dos poblaciones diferentes, lo que permite estimar, entre otras cosas, la intensidad de migración entre poblaciones (es decir, los patrones interpoblacionales de flujo génico).

Tanto la diversidad como la diferenciación genéticas dependen de una gran cantidad de factores característicos de cada especie, como el rango de distribución, los sistemas reproductivos, los tiempos de generación, los tamaños de población efectivos, el flujo génico entre poblaciones, los mecanismos de dispersión de semillas y la existencia de presiones selectivas (Gottlieb, 1977; Hamrick and Godt, 1990, 1996; Nybom, 2004; Aguilar et al., 2008; Pearson et al., 2009). De entre los factores mencionados, en la presente Tesis Doctoral prestaremos especial atención al flujo génico pasado y a algunas presiones selectivas actuales.

El flujo génico engloba cualquier mecanismo que implique el intercambio de genes entre poblaciones, ya sea mediante flujo gamético o mediante la migración de adultos o propágulos (Slarkin, 1985). Debido a que el flujo génico reduce la diferenciación genética entre las poblaciones que conecta (Luck et al., 2003; Thompson, 2005), la ausencia completa de dicho flujo mantenida durante largos periodos de tiempo puede hacer que la diferenciación entre poblaciones desemboque en especiación (Irwin et al., 2005). Sin embargo, es habitual encontrar linajes que han permanecido aislados durante un tiempo, pero no el suficiente para que los mecanismos de aislamiento pre y postcigótico estén completamente desarrollados (Wiens et al., 2006). Esto es especialmente habitual en grupos que presentan múltiples eventos de especiación reciente (Rieseberg and Willis, 2007). En ellos, la existencia de flujo génico puede modificar de manera considerable el tipo y la frecuencia de los alelos encontrados en las poblaciones, es decir, sus estructuras genéticas (Nei, 1975).

Como cualquier flujo génico implica movimiento de genes entre poblaciones, es esperable que exista una mayor probabilidad de intercambio genético entre poblaciones cercanas que entre poblaciones distantes. Como consecuencia, las poblaciones cercanas serán más similares entre sí que con las más distantes, dando lugar a un fenómeno denominado aislamiento por distancia (Wright, 1943). El aislamiento por distancia se utiliza habitualmente como hipótesis nula con la que contrastar los patrones geográficos observados, ya que existen multitud de factores que provocan que la variación geográfica en las propiedades genéticas de las poblaciones sea muy diferente a las esperadas bajo las asunciones de aislamiento por distancia. Tal es el caso, por ejemplo, de contracciones en tamaños poblacionales, expansiones del rango de distribución de las especies, eventos de mezcla genética y selección natural ejercida por cualquier factor que no muestre un patrón espacial (Epperson, 2003).

La existencia de presiones selectivas diferentes puede moldear tanto la diversidad como la estructuración genética de las poblaciones naturales a diferentes escalas geográficas. Cabe destacar el caso de aquellas poblaciones localizadas en gradientes altitudinales, donde drásticas variaciones ambientales ocurren a escalas espaciales especialmente cortas (Wen and Hsiao, 2001; Jump et al., 2006; Gonzalo-Turpin and Hazard, 2009). Aunque existen multitud de presiones selectivas actuando sobre las poblaciones naturales,, el efecto de los polinizadores puede ser especialmente importante para las especies de plantas autoincompatibles que presenten pequeñas distancias de dispersión de semilla, ya que la mayor parte del flujo genético, tanto entre poblaciones como dentro de población, se debe al movimiento de polen realizado por ellos. Por tanto, la abundancia y la identidad de visitantes florales pueden jugar también un papel fundamental en la diversidad genética poblacional, debido a que los polinizadores pueden incrementar el flujo génico y reducir los niveles de endogamia (Rozzi et al., 1997; Cascante et al., 2002). Variaciones en la abundancia, diversidad y composición de los polinizadores han sido frecuentemente encontradas también a diferentes escalas espaciales (Herrera, 1995; Fishbein and Venable, 1996; Thomson, 2001; Boyd, 2004; Price et al., 2005; Moeller, 2005, 2006; Kühn et al., 2006; Cosacov et al., 2008; Espíndola et al., 2011).

La polinización depende de una gran variedad de rasgos fenotípicos de la planta, tales como el color de la flor, la morfología de la corola o la longitud de los estambres. Es esperable que dichos rasgos evolucionen para que se ajusten a las preferencias y morfología de las principales especies de polinizadores, tal y como confirman varios estudios filogenéticos que encuentran caracteres florales divergentes en linajes que difieren en el tipo de polinizadores (Armbruster, 1993; Wilson et al., 2004; Muchhala, 2006; Alcantara and Lohmann, 2010; Knapp, 2010). Sin embargo, en plantas generalistas en cuanto a su polinización (es decir, aquéllas visitadas por un amplio número de especies de polinizadores), la existencia de múltiples polinizadores, cada uno con preferencias y morfologías características, dificultan la posibilidad de un ajuste perfecto entre planta y polinizador (Johnson and Steiner, 2000). En este caso es muy frecuente encontrar fluctuaciones espacio-temporales en el gremio de visitantes florales, convirtiendo a las interacciones planta-polinizador en uno de los factores que afectan a la variación geográfica en los fenotipos y genotipos poblacionales (Grant, 1949; Grant and Grant, 1965; Campbell et al., 1997; Anderson and Johnson, 2008). A pesar de su importancia potencial, las relaciones entre variación geográfica en diversidad genética, variación fenotípica y gremio de polinizadores han sido escasamente estudiadas.

Por último, además de los factores mencionados, poblaciones cercanas geográficamente pueden ser genéticamente más distantes de lo esperado debido a que tengan historias evolutivas independientes. Un ejemplo típico de este caso sería el aislamiento de linajes durante largos periodos de tiempo y la posterior expansión de las áreas de distribución hasta regiones colindantes (Durand et al., 1999; García-París et al., 2000; Cooke et al., 2012).

En resumen, existen multitud de factores que pueden moldear la distribución de diversidades y estructuras genéticas de las poblaciones naturales tanto latitudinal como altitudinalmente (Ohsawa and Ide, 2008). El estudio de los patrones geográficos de variación genética inter e intraespecífica permiten inferir efectos evolutivos de factores selectivos (ambientales y no ambientales) y movimientos históricos de especies y poblaciones (Epperson, 2003). Por tanto, la divergencia genética de las poblaciones refleja tanto los patrones actuales de flujo génico como los patrones históricos de aislamiento e intercambio genético (Olsen and Schaal, 1999).

Erysimum mediohispanicum

Erysimum L. (Brassicaceae) es un género que presenta multitud de aspectos conflictivos en su taxonomía actual, basada fundamentalmente en caracteres fenotípicos y cariotípicos (Clot, 1991; Nieto-Feliner, 1993; Blanca et al., 1991). Como consecuencia de dicha incertidumbre taxonómica, el número de especies que conforman el género varía según los autores entre 180 y 223 especies (Al-Shehbaz et al., 2006; Warwick and Al-Shehbaz, 2006; Koch and Al-Shehbaz, 2008). Aunque varias especies pueden encontrarse en África, Macaronesia y América del Norte, el género se distribuye principalmente en Eurasia (Koch and Al-Shehbaz, 2008), siendo el Mediterráneo occidental una importante región de diversificación (Greuter et al., 1986).

También existen discrepancias en el número de especies de *Erysimum* que habitan en la Península Ibérica, oscilando las estimas entre 13 y 22 (López González, 1998). Esta variación se debe a las diferencias entre autores en el concepto de especie taxonómica y a la importancia que dan a los distintos caracteres para delimitar un taxón como especie (López González, 1998). La dificultad del establecimiento de clados entre los *Erysimums* de la Península Ibérica queda reflejada también por la existencia de plantas bienales y perennes, así como monocárpicas y policárpicas dentro de grupos estrechamente emparentados (Blanca et al., 1991; Nieto-Feliner, 1993; López González, 1998). Además también existen diferencias en los números cromosómicos, lo que parece no ser una barrera para que exista flujo genético entre los grupos de *Erysimum* más conflictivos taxonómicamente (Correvon and Favarger, 1979).

De todo el género, *Erysimum mediohispanicum* es la especie que muestra una distribución más amplia en la Península Ibérica (de donde es endémica), extendiéndose en dos áreas principales, una en el noreste y otra en el sureste de la misma (Nieto-Feliner, 1993). Esto contrasta con el resto de especies de *Erysimum* descritas en la Península, que en su mayor parte son endemismos de distribución limitada (Nieto-Feliner, 1993). *E. mediohispanicum* es una herbácea principalmente bienal y monocárpica que se encuentra en regiones montañosas, predominando en territorios calizos entre matorrales subalpinos aclarados (especialmente tomillares), terrenos incultos y parameras (Nieto-Feliner, 1993; Fig.1.1A). Las poblaciones suelen ser pequeñas (entre decenas y varios cientos de individuos) y dispersas (con una densidad de plantas de 6.8 individuos en 100m^2; Gómez, 2005b), encontrándose entre los 600 y los 2000 m de altitud.

Los individuos de *E. mediohispanicum* viven durante dos o tres años como rosetas vegetativas (Fig.1.1F), muriendo tras haber desarrollado entre uno y ocho escapos que alcanzan una altura de unos 50cm (Fig.1.1G). Cada escapo está coronado por varias decenas (o incluso cientos) de botones florales que entre mayo y julio (dependiendo de la altitud y la latitud de la población) dan lugar a flores de color amarillo y morfología muy variable, incluso entre individuos de una misma población (Figs. 1.1B y 1.1C). Dichas flores son hermafroditas, ligeramente protándricas y presentan entre 30 y 40 óvulos cada una (Gómez et al., 2009). El polen se produce en un androecio tetradimo (es decir, que consta de cuatro estambres largos y dos cortos), mientras que el néctar es producido por cuatro glándulas localizadas en la base de los sépalos. Las inflorescencias de *E. mediohispanicum* son indeterminadas (es decir, capaces de producir flores continuamente) y acropétalas (con flores basales desarrollándose antes que las medias y apicales). Consecuentemente, hay un fuerte efecto de la posición en los patrones de fructificación, siendo mayor la probabilidad de fructificación en posiciones intermedias del escapo (cuyas flores están abiertas en el pico de floración) y menor en las posiciones basales y apicales (abiertas antes o después del pico de floración). *E. mediohispanicum* produce pequeñas semillas (menos de 4 mm de longitud y 0.36 mg de peso) que se dispersan por gravedad entre agosto y septiembre, cuando los frutos dehiscentes son abiertos debido al movimiento de la vegetación, el viento, la lluvia o el contacto físico. No se ha observado dispersión secundaria de semillas en esta especie, por lo que la distancia de dispersión es extremadamente corta (con una distancia máxima promedio de menos de 20cm, Gómez, 2007).

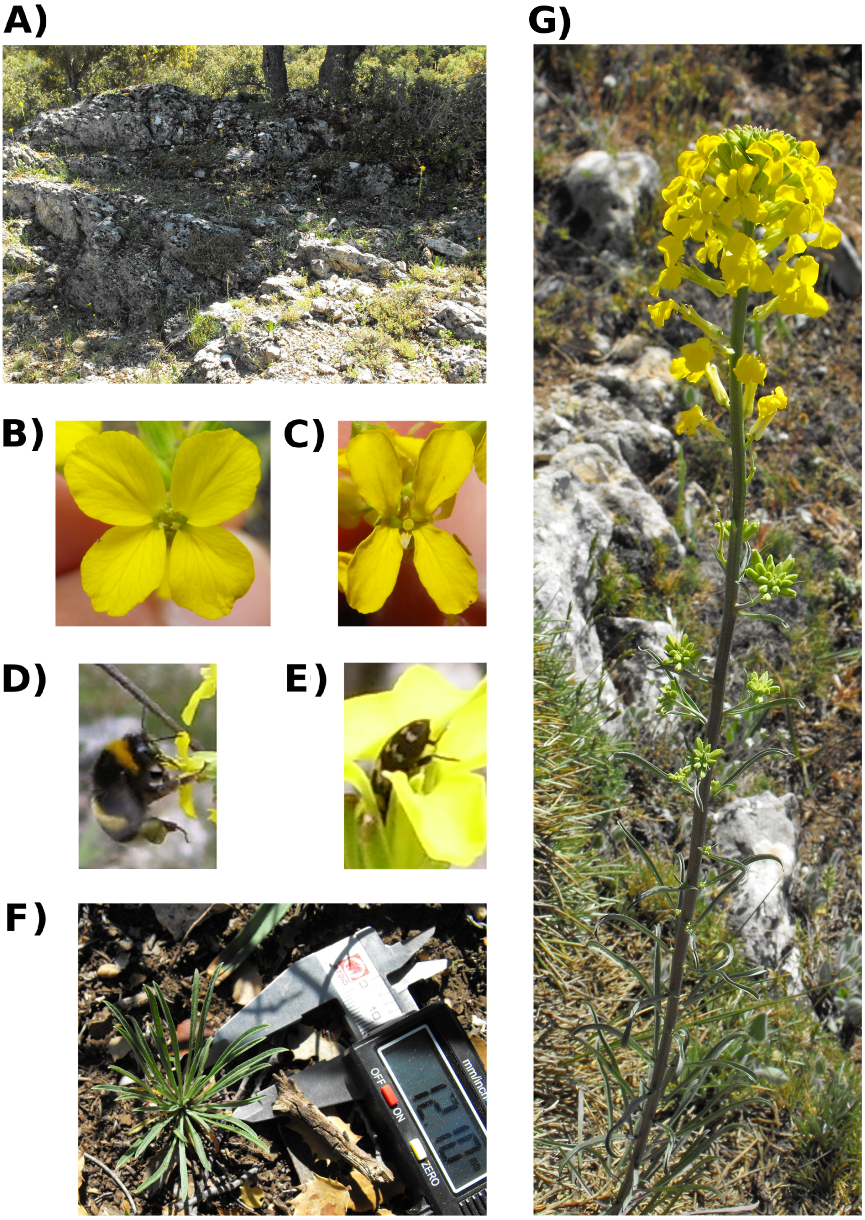


Figura 1.1: Historia natural de *E. mediohispanicum*. A) Hábitat típico de la especie. B) y C) Ejemplos de variación en morfología floral de dos individuos pertenecientes a la misma población. D) *Bombus terrestris* y E) *Attagenus trifasciatus*, ejemplos de dos visitantes florales de la especie con importantes diferencias en tamaño y patrón de forrajeo. D) Roseta vegetativa y E) Individuo reproductivo de *E. mediohispanicum*.

Las semillas son consumidas por ratones (*Apodemus sylvaticus*, Muridae), escarabajos (*Iberozabrus sp.*, Carabidae), hormigas (*Lasius niger*, Formicidae) e incluso pájaros (*Fringilla coelebs*, Fringillidae) (Gómez, 2005a).

E. mediohispanicum es parcialmente autocompatible, requiriendo la participación de vectores polínicos para conseguir el set completo de semillas (Gómez, 2005a). La mayor parte de estos visitantes florales actúan como polinizadores efectivos debido a la morfología abierta que presenta la flor, lo que permite que los órganos reproductivos sean extremadamente accesibles (Gómez et al., 2009). Por ello, incluso visitantes florales no especializados actúan como polinizadores (Gómez, 2005b) y pueden ejercer presiones selectivas significativas sobre varios rasgos florales de la especie (Gómez et al., 2006, 2008, 2009). En este sentido, se ha documentado que la variación espacial en el gremio de polinizadores lleva a adaptación local en esta especie, incluso a pequeñas escalas geográficas (Gómez et al., 2009).

El sistema de polinización de *E. mediohispanicum* es altamente generalista. Durante cinco años de estudio en ocho poblaciones muy próximas geográficamente se han encontrado más de 150 especies de visitantes florales pertenecientes a más de 25 familias y seis órdenes (Gómez et al., 2008, 2009), siendo los más abundantes las especies de himenopteros (más de 60 especies) y coleopteros (más de 40 especies). Por tanto, los visitantes florales de *E. mediohispanicum* muestran un amplio rango de variación en características tales como la longitud de probóscide, el comportamiento de forrajeo o el tamaño del cuerpo (Figs. 1.1D y 1.1E). Por ejemplo, mientras que un *Melighetes minutus* (Nitidulidae) tiene una masa de 0.3 mg, la de una *Anthophora aestivalis* (Anthophoridae) es más de 400 veces mayor, llegando a los 130 mg.

Además de la interacción con polinizadores, se ha documentado que los individuos reproductores de *E. mediohispanicum* son atacados por varias especies de herbívoros. Algunos botones florales no llegan a abrirse tras ser atacados por dípteros (e.g., *Dasineura sp.*, Cecidomidae), los escapos son perforados por algunas especies de gorgojos (e.g., *Lixus sp.*, Curculionidae) que consumen tejidos internos, mientras que otras especies (e.g., *Ceutorhynchus sp.*, Curculionidae) se desarrollan en el interior de los frutos y actuan como predadores predispersivos de semillas (Gómez, 2005a; Gómez and González-Megías, 2007). Los escapos son consumidos también por la cabra ibérica (*Capra pyrenaica*, Bovidae), que consume principalmente flores y frutos verdes (Gómez, 2003, 2005b). Por su parte, algunas especies de chinches (e.g., *Eurydema sp.*, Pentatomidae) se alimentan de la savia de los escapos reproductivos durante la floración y la fructificación (Gómez and González-Megías, 2007).

E. mediohispanicum está estrechamente emparentado con otros cinco taxones (*E.nevadense*, *E. merxmuelleri*, *E. rondae*, *E. ruscinonense* y *E. gomezcampoi*), a los que se les atribuye la categoría de microespecies por conformar el denominado complejo nevadense (Favarger, 1980; Clot, 1991; Nieto-Feliner, 1993). El tratamiento de microespecie es criticado por López González (1998), que parece más partidario de incluir la mayoría de los taxones ibéricos bajo una misma entidad biológica. Inferir relaciones taxonómicas entre las especies de este complejo se ve dificultado por la baja diferenciación y por la existencia de áreas de contacto entre ellas (Fig.1.2), donde la hibridación puede ocurrir (Clot, 1991). Consecuentemente, es usual encontrar individuos e incluso poblaciones enteras que no pueden asignarse de manera inequívoca por presentar características intermedias a varias especies (Clot, 1991; Nieto-Feliner, 1993).

Otro factor que dificulta el establecimiento de las relaciones entre poblaciones de estas especies es la existencia de diferencias en ploidía (Fig.1.2), lo que complica la aplicación de técnicas moleculares basadas en marcadores nucleares. De nuevo, el uso de marcadores genéticos localizados en los orgánulos puede ser especialmente útil,

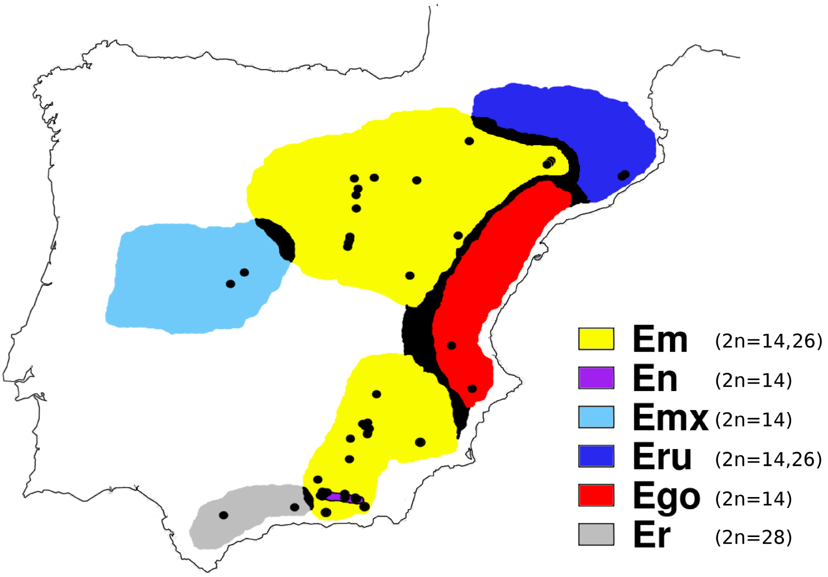


Figura 1.2: Distribución estimada de las cinco especies de *Erysimum* que conforman el complejo nevadense. Cada especie se representa con un color diferente y las regiones con posibilidad de contacto entre especies en negro. Los puntos negros representan las poblaciones estudiadas. (Em: *E. mediohispanicum*, En: *E.nevadense*, Emx: *E. merxmuelleri*, Eru: *E. ruscinonense*, Ego: *E. gomezcampoi*, Er: *E. rondae*).

al permitir la comparación directa de genotipos individuales independientemente de sus ploidías nucleares. Los números cromosómicos varían entre las especies del complejo (2n = 14 para *E.nevadense*, *E. merxmuelleri*, *E. ruscinonense* y *E. gomezcampoi*, 2n = 26 para *E. mediohispanicum* y 2n = 28 para *E. rondae*) e incluso dentro de especie (2n = 26 para las poblaciones de *E. ruscinonense* más occidentales y 2n = 14 para las poblaciones de *E. mediohispanicum* localizadas más al sur; Clot, 1991; Nieto-Feliner, 1993; Capítulo 5).

variación de ploidía en e. *mediohispanicum*

Los individuos de *E. mediohispanicum* con ploidías superiores a dos están descritos como autopoliploides (Clot, 1991), es decir, que los genomas duplicados pertenecían a la misma especie ancestral, en contraste con los denominados alopoliploides, que necesitan la participación de dos especies diferentes para la formación del poliploide (Müntzing, 1936; Darlington, 1937; Clausen, 1945). Durante mucho tiempo se pensó que los autopoliploides tenían desventajas evolutivas con respecto a los alopoliploides, por lo que se propuso que debían ser escasos y representar el fin de un camino evolutivo (Clausen, 1945; Stebbins, 1971). Sin embargo, Ramsey and Schemske (2002) estimaron que la tasa de formación de autopoliploides es mayor que la de alopoliploides, por lo que los primeros deben ser más frecuentes de lo que se pensaba hasta entonces. Así pues, los organismos autopoliploides han pasado de ser descritos como extremadamente raros (Stebbins, 1950; Clausen, 1945) a plantearse si son igual de frecuentes que los alopoliploides en las poblaciones naturales (Soltis et al., 2004). La expansión en el rango de distribución de varios autopoliploides naturales (Manton, 1937; Soltis et al., 2007) sugiere que este tipo de poliploides también puede presentar ventajas evolutivas, aunque éstas no sean obvias (Levin, 2002; Ramsey and Schemske, 2002).

Según el número de copias cromosómicas, los individuos poliploides de *E. mediohispanicum* son hipotetraploides, es decir, que presentan un número de cromosomas algo menor al doble de la dotación diploide (2n=14 para los diploides y 2n=26 para los poliploides; Clot, 1991; Nieto-Feliner, 1993). Este defecto en el número de cromosomas se debe a que los procesos de poliploidización son complejos e incluyen, además de duplicaciones genómicas, fenómenos posteriores de diploidización (Grant, 1981; Wolfe, 2001; Levy and Feldman, 2002; Doyle et al., 2008; Parisod et al., 2010) consistentes en cambios tanto del comportamiento citológico celular como de la constitución genética (es decir, pérdidas de material genético). El conocimiento tanto de la base molecular de dichos procesos como de las fuerzas que los impulsan es limitado (Wolfe, 2001; Soltis et al., 2004), aunque en los últimos años se han producido avances importantes en ambos aspectos (Doyle et al., 2008).

En *E. mediohispanicum*, ambos citotipos presentan distribuciones geográficas separadas. Las poblaciones poliploides se encuentran en la región norte, mientras que las diploides se restringen a las cordilleras orientales de Andalucía (Blanca et al., 1991; Nieto-Feliner, 1993). Estas diferencias geográficas y de ploidía llevaron inicialmente a (Polatschek, 1979) a considerar las poblaciones septentrionales bajo el nombre de *E. mediohispanicum* mientras que las diploides las refirió a *E. nevadense*. Autores posteriores (Blanca et al., 1991; Nieto-Feliner, 1993) han atribuido la última especie sólo a las formas de alta montaña, incluyendo la mayor parte de formas diploides del sur dentro de *E. mediohispanicum*. Dichos autores consideran que la ploidía no debe ser determinante en la taxonomía, ya que en la mayoría de los casos no hay diferencias morfológicas entre citotipos o las mismas se solapan ampliamente (Blanca et al., 1991).

Por el momento se desconocen los factores que motivan la distribución separada de las distintas ploidías en *E. mediohispanicum*. De hecho, no se han encontrado patrones generales de asociación clara entre condiciones ambientales y la existencia de poliploidía en ninguna especie (Manzaneda et al., 2012), aunque diferentes ventajas se han atribuido a los poliploides para justificar su carácter adaptativo. Por ejemplo, se ha postulado que tienen mayor flexibilidad genómica que los diploides, ya que la redundancia genética puede representar una ventaja ecológica en ambientes cambiantes como los mediterráneos (Balao et al., 2010). También se les han atribuido menores tasas de extinción, mayores tasas de diversificación (Soltis et al., 2010) y mayor resistencia a estrés abiótico que a sus progenitores diploides (Levy and Feldman, 2002), por lo que algunos autores aseguran que es más frecuente encontrar poliploides en lugares con climas fluctuantes o extremos (Stebbins, 1971; Brochmann et al., 2004; Parisod et al., 2010). Sin embargo son necesarios más estudios centrados en comparar directamente la resistencia de variantes con distinta ploidía a estrés, ya que los resultados obtenidos hasta la fecha no son concluyentes. Por ejemplo, Manzaneda et al. (2012) estudiaron correlaciones entre estrés y ploidía a lo largo del área de distribución de *Brachypodium distachyon* en la Península Ibérica. Aunque concluyeron que un origen adaptativo puede subyacer a la distribución ploídica de la especie, no pudieron descartar las explicaciones no adaptativas.

Existen también estudios que apuntan a que la ploidía puede tener efectos importantes en interacciones ecológicas tales como la intensidad de herbivoría o el gremio de polinizadores (Segraves et al., 1999; Thompson et al., 2004; Kennedy et al., 2006; Münzbergová, 2006; Arvanitis et al., 2007). Además, la historia evolutiva de la especie estudiada puede jugar un papel importante, como sucede en las especies búlgaras de *Erysimum*, donde la distribución de ploidías se explica por procesos de hibridación junto con cambios climáticos producidos durante la última glaciación (Ančev, 2006).

OBJETIVOS Y ESTRUCTURA DE LA TESIS

La presente Tesis Doctoral engloba dos objetivos principales. El primer objetivo se centra en explorar la historia evolutiva y filogeografía de *E. mediohispanicum* en la Peninsula Ibérica, prestando especial atención al posible papel que juegan los polinizadores como motor evolutvo. El segundo objetivo es puramente metodológico. En él hemos desarrollado las herramientas moleculares y analíticas que permitirán responder las cuestiones evolutivas relativas a *E. mediohispanicum*. Sin embargo, los métodos aquí presentados trascienden al sitema de estudio de esta Tesis y creemos que pueden ser de utilidad a otros investigadores que estudien problemas evolutivos similares.

Para estudiar la historia evolutiva de *E. mediohispanicum* hemos realizado un muestreo exhaustivo de la especie que nos ha permitido describir la ploidía, los principales visitantes florales, la media y la variacion fenotípica, la diversidad genética y las relaciones evolutivas de poblaciones naturales localizadas en todo su área de distribución. Dicho muestreo incluye un total de 56 poblaciones de *E. mediohispanicum*, a las que hay que añadir dos poblaciones por cada una de las otras cinco especies que componen el complejo nevadense (Fig.1.2; Tabla 1.1). De esta manera hemos podido establecer la filogeografía de *E. mediohispanicum* en el contexto de especies más próximas. Para determinar la diversidad y la estructuración genéticas nos hemos ceñido a las poblaciones diploides de la especie focal, lo que supone un total de 32 poblaciones ubicadas en su mayoría en el sur de la Península.

La Tesis está dividida en dos bloques que se corresponden con los marcos conceptuales descritos anteriormente. El primer bloque se centra en el desarrollo y la descripción de las herramientas metodológicas que se aplicarán, en el segundo bloque, a las poblaciones mostradas en las Tablas 1.1 y 1.2. Concretamente, las herramientas descritas serán los marcadores microsatélites empleados para estimar la diversidad genética (y que parecen poder extrapolarse a un gran número de especies del género; Capítulo 2) y un nuevo método de reconstrucción filogeográfica basado en distancias (Capítulo 4). Dicho método combina la información procedente de sustituciones e indels y representa las relaciones obtenidas en forma de red. En este bloque se incluye también la descripción de un nuevo paquete de R que permite la automatización del método para aplicarlo fácilmente a cualquier alineamiento (Capítulo 3).

Tabla 1.1: Descripción geográfica y tamaños de muestra de las 56 poblaciones de *Erysimum mediohispanicum* estudiadas en la presente Tesis Doctoral. Para cada población se muestra la latitud y la longitud (en grados decimales), la altitud (m sobre el nivel del mar) y el sistema montañoso en el que fueron muestreadas. Además se indican el número de visitantes florales registrados en cada población, así como el número de plantas usadas para estimar la estructura poblacional, la media y varianza de rasgos fenotípicos, y las relaciones filogeográficas.

Población	Longitud	Latitud	Altitud	Sistema Montañoso - (provincia)	Visitantes florales	Estructura genética	Fenotipo	Filogeografía
Em01	-3.428	37.133	1750	Sierra Nevada (GR)	222	15	196	5
Em02	-3.431	37.122	2099	Sierra Nevada (GR)	332	15	193	5
Em03	-3.472	37.085	1654	Sierra Nevada (GR)	263	15	60	5
Em04	-3.464	37.08	1826	Sierra Nevada (GR)	296	14	660	5
Em05	-2.555	38.072	1450	Sierra Cazorla (J)	213	15	60	5
Em06	-2.594	37.987	1448	Sierra Cazorla (J)	206	15	60	5
Em07	-3.505	37.083	1413	Sierra Nevada (GR)	274	15	60	5
Em08	-3.431	37.133	1690	Sierra Nevada (GR)	141	15	198	5
Em09	-3.371	37.13	1280	Sierra Nevada (GR)	158	15	51	5
Em10	-3.468	37.072	1811	Sierra Nevada (GR)	204	15	60	5
Em11	-3.473	37.071	1890	Sierra Nevada (GR)	108	14	60	2
Em12	-2.588	38.009	1732	Sierra Cazorla (J)	209	15	60	5
Em13	-2.641	38.112	1542	Sierra Cazorla (J)	22	15	60	5
Em14	-2.683	38.142	1469	Sierra Cazorla (J)	202	15	60	4
Em15	-2.415	38.581	1384	Sierra Cazorla (J)	218	15	60	5
Em16	-2.922	37.922	1063	Sierra Cazorla (J)	45	15	35	5
Em17	-3.424	37.116	2182	Sierra Nevada (GR)	202	15	230	5
Em18	-2.647	36.918	1215	Sierra Nevada (GR)	126	15	47	5
Em19	-3.476	37.092	1765	Sierra Nevada (GR)	42	15	61	5
Em20	-3.479	37.089	1718	Sierra Nevada (GR)	209	15	53	4
Em21	-3.428	37.134	1723	Sierra Nevada (GR)	89	15	164	5
Em22	-3.428	37.131	1802	Sierra Nevada (GR)	139	15	169	5
Em23	-3.426	37.128	1874	Sierra Nevada (GR)	125	15	166	5
Em24	-3.435	37.125	1943	Sierra Nevada (GR)	152	15	150	5
Em25	-3.434	37.121	2064	Sierra Nevada (GR)	142	15	332	5

Continúa en la página siguiente

Tabla 1.1 – Continuación de la página anterior

Población	Longitud	Latitud	Altitud	Sistema Montañoso - (provincia)	Visitantes florales	Estructura genética	Fenotipo	Filogeografía
Em26	-2.589	38.159	999	Sierra Cazorla (J)	145	14	46	5
Em27	-3.395	36.833	1752	Sierra Nevada (GR)	287	15	60	4
Em28	-3.029	37.051	1834	Sierra Nevada (GR)	214	15	60	5
Em29	-2.826	37.054	1918	Sierra Nevada (GR)	188	15	120	4
Em30	-2.804	37.036	1667	Sierra Nevada (GR)	-	-	15	4
Em31	1.024	42.032	1006	Montsec (Le)	212	15	30	5
Em32	0.916	41.99	732	Montsec (Le)	7	10	12	5
Em33	-2.933	40.909	983	Meseta Norte (Gu)	173	15	90	5
Em34	-2.81	41.524	890	Meseta Norte (So)	72	15	60	5
Em35	-2.845	41.766	1126	Meseta Norte (So)	-	15	60	5
Em36	-3.451	37.102	1471	Sierra Nevada (GR)	216	15	30	5
Em37	-3.467	37.138	1425	Sierra Nevada (GR)	163	15	30	4
Em38	-2.826	37.044	1783	Sierra Nevada (Al)	222	15	90	5
Em39	-3.552	37.319	1272	Sierra Nevada (Al)	64	18	30	5
Em40	-2.822	37.033	1664	Sierra Nevada (Al)	324	15	90	5
Em41	0.921	41.992	715	Montsec (Le)	197	15	30	5
Em42	0.94	41.998	882	Montsec (Le)	190	15	30	5
Em43	0.899	41.979	786	Montsec (Le)	248	15	30	5
Em44	1.028	42.017	790	Montsec (Le)	244	11	30	5
Em45	-2.938	40.862	963	Meseta Norte (Gu)	-	15	60	5
Em46	-2.96	40.817	1000	Meseta Norte (Gu)	234	15	60	5
Em47	-2.807	41.326	1063	Meseta Norte (So)	2	14	60	4
Em48	-2.456	41.781	1026	Meseta Norte (So)	20	15	60	5
Em49	-2.774	41.616	986	Meseta Norte (So)	223	15	60	5
Em50	-2.973	40.76	1000	Meseta Norte (Gu)	117	15	65	5
Em51	-1.567	37.869	1396	Sierra Espuña (Mu)	219	15	30	5
Em52	-2.941	37.619	1189	Sierra Cazorla (J)	69	15	30	5
Em53	-1.767	40.333	1633	S. de Albarracín (Te)	-	14	14	5
Em54	-0.616	42.321	932	Pre-pirineo (Hu)	-	-	60	5
Em55	-1.637	41.74	806	Sierra del Moncayo (Z)	-	15	60	5
Em56	-0.833	40.925	955	Sierra de Cucalón (Te)	-	-	60	5

Especie	Población	Longitud	Latitud	Altitud	Sistema montañoso (provincia)	Filogeografía
E. merxmuelleri	Emx01	-4.98	40.38	1550	Sierra de Gredos (Av)	4
E. merxmuelleri	Emx02	-5.25	40.21	1000	Sierra de Gredos (Av)	5
E. nevadense	En05	-3.03	37.11	2100	Sierra Nevada (Al)	5
E. nevadense	En10	-3.42	37.11	2200	Sierra Nevada (Gr)	5
E. rondae	Er02	-4.00	36.91	1250	Sierra de Tejeda (Gr)	3
E. rondae	Er03	-5.38	36.79	1050	S. de Grazalema (Ca)	5
E. ruscinonense	Eru01	2.35	41.80	750	Montseny (B)	5
E. ruscinonense	Eru02	2.40	41.83	1000	Montseny (B)	5
E. gomezcampoi	Eg002	-0.56	38.66	1100	Sierra de Font Roja (Al)	5
E. gomezcampoi	Eg003	-0.96	39.30	800	Sierra Martés (V)	5

Tabla 1.2: Descripción geográfica y tamaños de muestra de las 10 poblaciones no pertenecientes a *E. mediohispanicum* estudiadas en la presente Tesis Doctoral. Para cada población se muestra la latitud y la longitud (en grados decimales), la altitud (m sobre el nivel del mar) y el sistema montañoso en el que fueron muestreadas. Además se indican el número de plantas usadas para estimar las relaciones filogeográficas.

Por su parte, en el segundo bloque se estudian las relaciones existentes entre gremio de polinizadores y diversidad genética (mediadas por la diversidad fenotípica) para las poblaciones compuestas por individuos diploides (Capítulo 5), así como las relaciones filogeográficas entre todas las poblaciones mostradas en la Tabla 1.1 y 1.2 (Capítulo 6). Debido a que el volumen de datos manejados en los capítul°os de este segundo bloque es considerable, en cada uno se discuten únicamente los resultados obtenidos en ese capítulo. Por último, en la discusión general (Capítulo 7), además de resaltarse los aspectos más relevantes tratados en la Tesis, relacionamos resultados obtenidos en capítulos distintos.

Específicamente, los objetivos de la presente Tesis doctoral son los siguientes:

1) Caracterizar marcadores microsatélites para *E. mediohispanicum* que permitan el estudio de la estructura y diferenciación genéticas de la especie (Cap. 5).

2) Desarrollar un método para obtener información evolutiva contenida en los indels (Caps. 4 y 3). Para ello es necesario:

2.1) Describir un método de alineamiento específico del marcador plastidial trnF IGS, una región pseudogénica rica en indels (Cap. 4).

2.2) Describir un método general basado en distancias para investigar las relaciones entre secuencias alineadas combinando sustituciones e indels (Cap. 4).

3) Desarrollar herramientas (en este caso, un paquete de R) que permitan la aplicación del nuevo método de manera general (Cap. 3).

4) Analizar las relaciones existentes entre diversidad genética y gremio de polinizadores, mediado por los rasgos fenotípicos que determinan la interacción planta-polinizador en un sistema generalista (Cap. 5). Para completar este objetivo se han muestreado un alto número de poblaciones naturales de la especie y para cada una de ellas ha sido necesario:

4.1) Describir tanto los valores medios como las varianzas de caracteres fenotípicos involucrados en la interacción planta-polinizador (Cap. 5).

4.2) Describir el gremio de polinizadores de la especie (Cap. 5).

4.3) Describir la diversidad y la estructuración genética (Cap. 5).

5) Determinar el papel que juegan los polinizadores en la evolución de *E. mediohispanicum* con respecto a la inercia filogeográfica (Cap. 7). Para ello ha sido necesario:

5.1) Inferir los patrones filogeográficos de *E. mediohispanicum* en la Península Ibérica (Cap. 6).

REFERENCES

Aguilar, R., M. Quesada, L. Ashworth, Y. Herrerias-Diego, and J. Lobo (2008). Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology 17*(24), 5177–5188.

Al-Shehbaz, I., M. Beilstein, and E. Kellogg (2006). Systematics and phylogeny of the brassicaceae (cruciferae): an overview. *Plant Systematics and Evolution 259*(2), 89–120.

Alarcón, M., P. Vargas, L. Sáez, J. Molero, and J. J. Aldasoro (2012). Genetic diversity of mountain plants: Two migration episodes of mediterranean erodium (geraniaceae). *Molecular Phylogenetics and Evolution 63*(3), 866–876.

Alcantara, S. and L. G. Lohmann (2010). Evolution of floral morphology and pollination system in bignonieae (bignoniaceae). *American Journal of Botany 97*(5), 782 –796.

Anderson, B. and S. D. Johnson (2008). The geographical mosaic of coevolution in a plant–pollinator mutualism. *Evolution 62*(1), 220–225.

Ančev, M. (2006). Polyploidy and hybridization in bulgarian brassicaceae: distribution and evolutionary role. *Phytol. Balcan 12*(3), 357–366.

Armbruster, S. W. (1993). Evolution of plant pollination systems: Hypotheses and tests with the neotropical vine dalechampia. *Evolution 47*(5), 1480–1505.

Arroyo, J. M., J. S. Carrión, A. Hampe, and P. Jordano (2004). La distribución de las especies a diferentes escalas espacio-temporales. http://digital.csic.es/handle/10261/40604. Peer reviewed.

Arvanitis, L., C. Wiklund, and J. Ehrlén (2007). Butterfly seed predation: effects of landscape characteristics, plant ploidy level and population structure. *Oecologia 152*(2), 275–285.

Avise, J. (2000). *Phylogeography : the history and formation of species*. Cambridge, Mass.: Harvard University Press.

Balao, F., L. M. Valente, P. Vargas, J. Herrera, and S. Talavera (2010). Radiative evolution of polyploid races of the iberian carnation dianthus broteri (caryophyllaceae). *New Phytologist 187*(2), 542–551.

Barraclough, T. G. and S. Nee (2001). Phylogenetics and speciation. *Trends in Ecology & Evolution 16*(7), 391–399.

Baum, D. A., K. Sytsma, and P. C. Hoch (1994). A phylogenetic analysis of epilobium (onagraceae) based on nuclear ribosomal DNA sequences. *Systematic Botany 19*, 363–388.

Beatty, G. E. and J. Provan (2012). Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the lusitanian plant species daboecia cantabrica (ericaceae). *Journal of Biogeography*, 335–344.

Blair, C. and R. W. Murphy (2011). Recent trends in molecular phylogenetic analysis: Where to next? *Journal of Heredity 102*(1), 130 –138.

Blanca, G., M. C. Morales Torres, and M. Ruiz Rejón (1991). El género.^{er}ysimum"l.(çruciferae") en andalucía (españa). In *Anales del Jardín Botánico de Madrid*, Volume 49, pp. 201–214.

Blondel, J., J. Aronson, J. Y. Bodiou, and G. Boeuf (2010). *The Mediterranean region: biological diversity through time and space*. Oxford University Press, Oxford.

Borsch, T., K. Hilu, D. Quandt, V. Wilde, C. Neinhuis, and W. Barthlott (2003). Noncoding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms. *J Evol Biol 16*(4), 558–76.

Boyd, A. E. (2004). Breeding system of macromeria viridiflora (boraginaceae) and geographic variation in pollinator assemblages. *American Journal of Botany 91*(11), 1809 –1813.

Brochmann, C., A. K. Brysting, I. G. Alsos, L. Borgen, H. H. Grundt, A.-C. Scheen, and R. Elven (2004). Polyploidy in arctic plants. *Biological Journal of the Linnean Society 82*(4), 521–536.

Campbell, D., N. Waser, and E. Melendez-Ackerman (1997). Analyzing pollinator-mediated selection in a plant hybrid zone: Hummingbird visitation patterns on three spatial scales. *The American Naturalist 149*(2), 295–315.

Cascante, A., M. Quesada, J. J. Lobo, and E. A. Fuchs (2002). Effects of dry tropical forest fragmentation on the reproductive success and genetic structure of the tree samanea saman. *Conservation Biology 16*(1), 137–147.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution 17*(4), 540 –552.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedren, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte,

E. Golenberg, G. H. Learn, S. W. Graham, S. C. H. Barrett, S. Dayanandan, and V. A. Albert (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden 80*(3), 528.

Clausen, R. T. (1945). Hybrids of the eastern north american subspecies of lycopodium complanatum and l. tristachyum. *American Fern Journal 35*(1), 9.

Clot, B. (1991). Caryosystématique de quelques erysimum l. dans le nord de la péninsule ibérique. *Anales del Jardín Botánico de Madrid 49*(2), 215–229.

Cooke, G., N. Chao, and L. Beheregaray (2012). Divergent natural selection with gene flow along major environmental gradients in amazonia: insights from genome scans, population genetics and phylogeography of the characin fish triportheus albus. *Molecular Ecology 21*(10), 2410–2427.

Correvon, P. and C. Favarger (1979). Croisements expérimentaux entre «Races chromosomiques» dans le genreErysimum. *Plant Systematics and Evolution 131*(1-2), 53–69.

Cosacov, A., J. Nattero, and A. A. Cocucci (2008). Variation of pollinator assemblages and pollen limitation in a locally specialized system: The oil-producing nierembergia linariifolia (solanaceae). *Annals of Botany 102*(5), 723 –734.

Darlington, C. D. (1937). Recent advances in cytology. *Recent advances in cytology.* (2nd Ed).

Dobeš, C., C. Kiefer, M. Kiefer, and M. A. Koch (2007). Plastidic trnFUUC pseudogenes in north american genus boechera (brassicaceae): Mechanistic aspects of evolution. *Plant biol (Stuttg) 9*(04), 502,515. 502.

Dobeš, C. H., T. Mitchell-Olds, and M. A. Koch (2004). Extensive chloroplast haplotype variation indicates pleistocene hybridization and radiation of north american arabis drummondii, a. divaricarpa, and a. holboellii (brassicaceae). *Molecular Ecology 13*(2), 349–370.

Doyle, J. J., L. E. Flagel, A. H. Paterson, R. A. Rapp, D. E. Soltis, P. S. Soltis, and J. F. Wendel (2008). Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics 42*(1), 443–461. PMID: 18983261.

Durand, J. D., H. Persat, and Y. Bouvet (1999). Phylogeography and postglacial dispersion of the chub (leuciscus cephalus) in europe. *Molecular Ecology 8*(6), 989–997.

Dwivedi, B. and S. Gadagkar (2009). Phylogenetic inference under varying proportions of indel-induced alignment gaps. *BMC Evolutionary Biology 9*(1), 211.

Epperson, B. (2003). *Geographical genetics*. Monographs in population biology;38. Princeton, NJ [etc.]: Princeton University Press.

Espíndola, A., L. Pellissier, and N. Alvarez (2011). Variation in the proportion of flower visitors of arum maculatum along its distributional range in relation with community-based climatic niche analyses. *Oikos 120*(5), 728–734.

Favarger, C. (1980). Le nombre chromosomique des populations alticoles d'Erysimum des picos de europa (espagne). *Bull. Soc. Neuchatel. Sci. Nat 103*, 85–90.

Fernández-Mazuecos, M. and P. Vargas (2011). Historical isolation versus recent long-distance connections between europe and africa in bifid toadflaxes (linaria sect. versicolores). *PLoS ONE 6*(7), e22234.

Fishbein, M. and D. L. Venable (1996). Diversity and temporal change in the effective pollinators of asclepias tuberosa. *Ecology 77*(4), 1061–1073.

Freudenstein, J. V. and M. W. Chase (2001). Analysis of mitochondrial nad1b-c intron sequences in orchidaceae: Utility and coding of length-change characters. *Systematic Botany 26*(3), 643–657.

Fuertes-Aguilar, J., B. G. Gutierrez-Larena, and G. N. Nieto-Feliner (2011). Genetic and morphological diversity in armeria (plumbaginaceae) is shaped by glacial cycles in mediterranean refugia. In *Anales del Jardín Botánico de Madrid*, Volume 68, pp. 175–197.

García-París, M., D. A. Good, G. Parra-Olea, and D. B. Wake (2000). Biodiversity of costa rican salamanders: Implications of high levels of genetic differentiation and phylogeographic structure for species formation. *Proceedings of the National Academy of Sciences 97*(4), 1640–1647.

Geiger, D. (2002). Stretch coding and block coding: two new strategies to represent questionably aligned DNA sequences. *J Mol Evol 54*(2), 191–9.

Gielly, L. and P. Taberlet (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Molecular Biology and Evolution 11*(5), 769–777.

Golenberg, E. M., M. T. Clegg, M. L. Durbin, J. Doebley, and D. P. Ma (1993). Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution 2*(1), 52–64.

Gómez, A. and D. Lunt (2007). Refugia within refugia: Patterns of phylogeographic concordance in the iberian peninsula. In *Phylogeography of Southern European Refugia*, pp. 155–188.

Gómez, J. M. (2003). Herbivory reduces the strength of pollinator-mediated selection in the mediterranean herb erysimum mediohispanicum: Consequences for plant specialization. *The American Naturalist 162*(2), 242–256. ArticleType: research-article / Full publication date: Aug., 2003 / Copyright Â© 2003 The University of Chicago Press.

Gómez, J. M. (2005a). Long-term effects of ungulates on performance, abundance, and spatial distribution of two montane herbs. *Ecological Monographs 75*(2), 231–258.

Gómez, J. M. (2005b). Non-additive effects of herbivores and pollinators on *Erysimum mediohispanicum* (cruciferae) fitness. *Oecologia 143*(3), 412–418.

Gómez, J. M. (2007). Dispersal-mediated selection on plant height in an autochorously dispersed herb. *Plant Systematics and Evolution 268*(1), 119–130–130.

Gómez, J. M., M. Abdelaziz, J. P. M. Camacho, A. J. Muñoz-Pajares, and F. Perfectti (2009). Local adaptation and maladaptation to pollinators in a generalist geographic mosaic. *Ecology Letters 12*(7), 672–682.

Gómez, J. M., J. Bosch, F. Perfectti, J. D. Fernández, M. Abdelaziz, and J. P. M. Camacho (2008). Association between floral traits and rewards in. *Annals of Botany 101*(9), 1413 –1420.

Gómez, J. M. and A. González-Megías (2007). Long-term effects of ungulates on phytophagous insects. *Ecological Entomology 32*(2), 229–234.

Gómez, J. M., F. Perfectti, J. Bosch, and J. P. M. Camacho (2009). A geographic selection mosaic in a generalized plant–pollinator–herbivore system. *Ecological Monographs 79*(2), 245–263.

Gómez, J. M., F. Perfectti, and J. P. M. Camacho (2006). Natural selection on erysimum mediohispanicum flower shape: Insights into the evolution of zygomorphy. *The American Naturalist 168*(4), 531–545.

Gonzalo-Turpin, H. and L. Hazard (2009). Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species festuca eskia. *Journal of Ecology 97*(4), 742–751.

Gottlieb, L. D. (1977). Electrophoretic evidence and plant systematics. *Annals of the Missouri Botanical Garden 64*(2), 161–180.

Grant, V. (1949). Pollination systems as isolating mechanisms in angiosperms. *Evolution; international journal of organic evolution 3*(1), 82–97.

Grant, V. (1981). *Plant speciation*. New York: Columbia University Press.

Grant, V. and K. Grant (1965). *Flower Pollination in the Phlox Family*. Columbia University Press.

Greuter, W., H. M. Burdet, , and G. Long. (1986). Med-checklist 3, dicotyledones (convolvulaceae-labiatae).

Guzmán, B. and P. Vargas (2009). Long-distance colonization of the western mediterranean by cistus ladanifer (cistaceae) despite the absence of special dispersal mechanisms. *Journal of Biogeography 36*(5), 954–968.

Hamilton, M. B., J. M. Braverman, and D. F. Soria-Hernanz (2003). Patterns and relative rates of nucleotide and Insertion/Deletion evolution at six chloroplast intergenic regions in new world species of the lecythidaceae. *Molecular Biology and Evolution 20*(10), 1710–1721.

Hamrick, J. L. and M. J. W. Godt (1990). Allozyme diversity in plant species. In A. Brown, M. Clegg, A. Kahler, and B. Weir (Eds.), *Plant population genetics, breeding, and genetic resources*, pp. 43–63.

Hamrick, J. L. and M. J. W. Godt (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351*(1345), 1291–1298.

Herrera, C. (1995). Microclimate and individual variation in pollinators: Flowering plants are more than their flowers. *Ecology 76*(5), 1516–1524.

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature 405*(6789), 907–913.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359*(1442), 183–195.

Hillis, D., C. Moritz, and B. Mable (1996). Applications of molecular systematics: the state of the field and look to the future. In D. Hillis, C. Moritz, and B. Mable (Eds.), *Molecular Systematics*, pp. 515–543. Sinauer Associates.

Irwin, D. E., S. Bensch, J. H. Irwin, and T. D. Price (2005). Speciation by distance in a ring species. *Science 307*(5708), 414–416.

Jiménez-Moreno, G., S. Fauquette, and J.-P. Suc (2010). Miocene to pliocene vegetation reconstruction and climate estimates in the iberian peninsula from pollen data. *Review of Palaeobotany and Palynology 162*(3), 403–415.

Johnsen, A., E. Rindal, P. G. P. Ericson, D. Zuccon, K. C. R. Kerr, M. Y. Stoeckle, and J. T. Lifjeld (2010). DNA barcoding of scandinavian birds reveals divergent lineages in trans-atlantic species. *Journal of Ornithology 151*(3), 565–578.

Johnson, L. A. and D. E. Soltis (1994). matk dna sequences and phylogenetic reconstruction in saxifragaceae s. str. *Systematic Botany 19*(1), 143.

Johnson, S. D. and K. E. Steiner (2000). Generalization versus specialization in plant pollination systems. *Trends in Ecology & Evolution 15*(4), 140–143.

Jukes, T. and C. Cantor (1969). Evolution of protein molecules. In M. Munro (Ed.), *Mammalian protein metabolism*, Volume III, pp. 21–132. Academic Press.

Jump, A., J. Hunt, J. Martínez-Izquierdo, and J. Peñuelas (2006). Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in fagus sylvatica. *Molecular Ecology 15*(11), 3469–3480.

Kelchner, S. A. (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annual Missouri Botanic Garden 87*(4), 482–498.

Kennedy, B. F., H. A. Sabara, D. Haydon, and B. C. Husband (2006). Pollinator-mediated assortative mating in mixed ploidy populations of chamerion angustifolium (onagraceae). *Oecologia 150*(3), 398–408.

Kjer, K. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Molecular Phylogenetics and Evolution 4*(3), 314–330.

Kjer, K. M., J. J. Gillespie, and K. A. Ober (2007). Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. *Systematic Biology 56*(1), 133 –146.

Knapp, S. (2010). On various contrivances: pollination, phylogeny and flower form in the solanaceae. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*(1539), 449 –460.

Koch, M. A. and I. A. Al-Shehbaz (2008). Molecular systematics and evolution of wild crucifers (brassicaceae or cruciferae). In *Biology and breeding of crucifers* (Gupta ed.). London, UK.: Taylor and Francis,. draft version, 1–19.

Koch, M. A., C. Dobeš, C. Kiefer, R. Schmickl, L. Klimeš, and M. A. Lysak (2007). Supernetwork identifies multiple events of plastid trnF(GAA) pseudogene evolution in the brassicaceae. *Molecular Biology and Evolution 24*(1), 63 –73.

Koch, M. A., C. Dobeš, M. Matschinger, W. Bleeker, J. Vogel, M. Kiefer, and T. Mitchell-Olds (2005). Evolution of the trnF(GAA) gene in arabidopsis relatives and the brassicaceae family: Monophyletic origin and subsequent diversification of a plastidic pseudogene. *Molecular Biology and Evolution 22*(4), 1032 –1043.

Koch, M. A. and M. Matschinger (2007). Evolution and genetic differentiation among relatives of arabidopsis thaliana. *Proceedings of the National Academy of Sciences 104*(15), 6272 –6277.

Koch, M. A., M. Wernisch, and R. Schmickl (2008). Arabidopsis thaliana's wild relatives: an updated overview on systematics, taxonomy and evolution. *Taxon 57*(3).

Kühn, I., S. M. Bierman, W. Durka, and S. Klotz (2006). Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. *New Phytologist 172*(1), 127–139.

Kumar, S. and A. Filipski (2007). Multiple sequence alignment: In pursuit of homologous DNA positions. *Genome Research 17*(2), 127 –135.

Levin, D. A. (2002). *The Role of Chromosomal Change in Plant Evolution*. Oxford University Press.

Levy, A. A. and M. Feldman (2002). The impact of polyploidy on grass genome evolution. *Plant Physiology 130*(4), 1587–1593.

Liu, K., C. Linder, and T. Warnow (2010). Multiple sequence alignment: a major challenge to large-scale phylogenetics. *PLoS Curr 2*, RRN1198.

Liu, K., T. J. Warnow, M. T. Holder, S. M. Nelesen, J. Yu, A. P. Stamatakis, and C. R. Linder (2012). SATé-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology 61*(1), 90 –106.

López González, G. (1998). Sobre algunos erysimun l.(cruciferae) madrileños. In *Anales del Jardín Botánico de Madrid*, Volume 56, pp. 370–378.

Loytynoja, A. and M. Milinkovitch (2001). SOAP, cleaning multiple alignments from unstable blocks. *Bioinformatics 17*(6), 573–4.

Luck, G. W., G. C. Daily, and P. R. Ehrlich (2003). Population diversity and ecosystem services. *Trends in Ecology & Evolution 18*(7), 331–336.

Lunter, G., I. Miklos, A. Drummond, J. Jensen, and J. Hein (2005). Bayesian coestimation of phylogeny and sequence alignment. *BMC Bioinformatics 6*(1), 83.

Lutzoni, F., P. Wagner, V. Reeb, and S. Zoller (2000). Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Systematic Biology 49*(4), 628 –651.

Makarenkov, V. and P. Legendre (2004). From a phylogenetic tree to a reticulated network. *Journal of Computational Biology 11*(1), 195–212.

Manton, I. (1937). The problem of biscutella laevigata l. II. the evidence from meosis. *Annals of Botany 1*(3), 439–462.

Manzaneda, A. J., P. J. Rey, J. M. Bastida, C. Weiss-Lehman, E. Raskin, and T. Mitchell-Olds (2012). Environmental aridity is associated with cytotype segregation and polyploidy occurrence in brachypodium distachyon (poaceae). *New Phytologist 193*(3), 797–805.

Mardulyn, P. (2012). Trees and/or networks to display intraspecific DNA sequence variation? *Molecular Ecology 21*(14), 3385–3390.

Médail, F. and K. Diadema (2009). Glacial refugia influence plant diversity patterns in the mediterranean basin. *Journal of Biogeography 36*(7), 1333–1345.

Moeller, D. (2005). Pollinator community structure and sources of spatial variation in plant–pollinator interactions in clarkia xantiana ssp. xantiana. *Oecologia 142*(1), 28–37.

Moeller, D. A. (2006). Geographic structure of pollinator communities, reproductive assurance, and the evolution of self-pollination. *Ecology 87*(6), 1510–1522.

Morrison, D. A. (2005). Networks in phylogenetic analysis: new tools for population biology. *International Journal for Parasitology 35*(5), 567–582.

Morrison, D. A. (2009). Why would phylogeneticists ignore computerized sequence alignment? *Systematic Biology 58*(1), 150 –158.

Muchhala, N. (2006). The pollination biology of burmeistera (campanulaceae): specialization and syndromes. *American Journal of Botany 93*(8), 1081 –1089.

Müller, K. (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution 38*(3), 667–676.

Müntzing, A. (1936). The evolutionary significance of autopolyploidy. *Hereditas 21*(2-3), 363–378.

Münzbergová, Z. (2006). Ploidy level interacts with population size and habitat conditions to determine the degree of herbivory damage in plant populations. *Oikos 115*(3), 443–452.

Nei, M. (1975). *Molecular population genetics and evolution*. Frontiers of biology;vol. 40. Amsterdam [etc.]: [s.n.].

Nieto-Feliner, G. (1993). Erysimum. In S. Castroviejo, C. Aedo, C. Gómez-Campo, M. Lainz, P. Monserrat, R. Morales, F. Muñoz-Garmendia, G. Nieto-Feliner, E. Rico, S. Talavera, and L. Villar (Eds.), *Flora Iberica*, Volume 4, Cruciferae-Monotropaceae., pp. 48–76. Madrid: Real Jardín Botánico CSIC.

Nieto Feliner, G. (2011). Southern european glacial refugia: A tale of tales. http://digital.csic.es/handle/10261/35607. Peer reviewed.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology 13*(5), 1143–1155.

Ogden, T. H. and M. S. Rosenberg (2007). How should gaps be treated in parsimony? a comparison of approaches using simulation. *Molecular Phylogenetics and Evolution 42*(3), 817–826.

Ohsawa, T. and Y. Ide (2008). Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography 17*(2), 152–163.

Olmstead, R. G. and J. A. Sweere (1994). Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the solanaceae. *Systematic Biology 43*(4), 467–481.

Olsen, K. M. and B. A. Schaal (1999). Evidence on the origin of cassava: Phylogeography of manihot esculenta. *Proceedings of the National Academy of Sciences 96*(10), 5586–5591.

Ortiz, M. Á., K. Tremetsberger, A. Terrab, T. F. Stuessy, J. L. García-Castaño, E. Urtubey, C. M. Baeza, C. F. Ruas, P. E. Gibbs, and S. Talavera (2008). Phylogeography of the invasive weed hypochaeris radicata (asteraceae): from moroccan origin to worldwide introduced populations. *Molecular Ecology 17*(16), 3654–3667.

Parisod, C., R. Holderegger, and C. Brochmann (2010). Evolutionary consequences of autopolyploidy. *New Phytologist 186*(1), 5–17.

Pearson, G. A., A. Lago-Leston, and C. Mota (2009). Frayed at the edges: selective pressure and adaptive response to abiotic stressors are mismatched in low diversity edge populations. *Journal of Ecology 97*(3), 450–462.

Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini, and G. G. Vendramin (2005). INVITED REVIEW: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology 14*(3), 689–701.

Polatschek, F. (1979). Die arten der gattung erysimum auf der iberischen halbinsel. *Ann. Naturhistor. Mus. Wien 82*, 325–362.

Price, M. V., N. M. Waser, R. E. Irwin, D. R. Campbell, and A. K. Brody (2005). Temporal and spatial variation in pollination of a montane herb: a seven-year study. *Ecology 86*, 2106–2116.

Ramsey, J. and D. W. Schemske (2002). Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, 589–639.

Redelings, B. and M. Suchard (2007). Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. *BMC Evolutionary Biology* 7(1), 40.

Rivas, E. and S. Eddy (2008). Probabilistic phylogenetic inference with insertions and deletions. *PLoS Comput Biol* 4(9), e1000172.

Roch, S. (2010). Toward extracting all phylogenetic information from matrices of evolutionary distances. *Science 327*(5971), 1376–1379.

Rodríguez-Sánchez, F., A. Hampe, P. Jordano, and J. Arroyo (2010). Past tree range dynamics in the iberian peninsula inferred through phylogeography and palaeodistribution modelling: A review. *Review of Palaeobotany and Palynology 162*(3), 507–521.

Rozzi, R., M. Arroyo, and J. Armesto (1997). Ecological factors affecting gene flow between populations of anarthrophyllum cumingii (papilionaceae) growing on equatorial- and polar-facing slopes in the andes of central chile. *Plant Ecology 132*(2), 171–179.

Saakian, D. B. (2008). Evolution models with base substitutions, insertions, deletions, and selection. *Physical Review E 78*(6), 061920.

Schmickl, R., C. Kiefer, C. Dobeš, and M. A. Koch (2008). Evolution of trnF(GAA) pseudogenes in cruciferous plants. *Plant Systematics and Evolution 282*, 229–240.

Schmitt, T. (2007). Molecular biogeography of europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology 4*(1), 11.

Segraves, K. A., J. N. Thompson, P. S. Soltis, and D. E. Soltis (1999). Multiple origins of polyploidy and the geographic structure of heuchera grossulariifolia. *Molecular Ecology 8*(2), 253–262.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small (2005, January). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany 92*(1), 142–166.

Simmons, M., H. Ochoterena, and T. Carr (2001, June). Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst Biol 50*(3), 454–62.

Simmons, M. P. and H. Ochoterena (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology 49*(2), 369–381. ArticleType: research-article / Full publication date: Jun., 2000 / Copyright Â© 2000 Society of Systematic Biologists.

Singhal, S. and C. Moritz (2012). Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution 66*(5), 1474–1489.

Slarkin, M. (1985, November). Gene flow in natural populations. *Annual Review of Ecology and Systematics 16*(1), 393–430.

Soltis, D. E., R. J. Buggs, J. J. Doyle, and P. S. Soltis (2010). What we still don't know about polyploidy. *Taxon 59*(5), 1387–1403.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd (2007). Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon 56*(1), 13–30.

Soltis, D. E., P. S. Soltis, and J. A. Tate (2004). Advances in the study of polyploidy since plant speciation. *New Phytologist 161*(1), 173–191.

Stebbins, G. L. (1950). Variation and evolution in plants.

Stebbins, G. L. (1971). Chromosomal evolution in higher plants. *Chromosomal evolution in higher plants.*.

Stebbins, G. L. (1984). Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. *Botanica helvetica 94*(1), 1–13.

Stuessy, T. F. (2009). *Plant taxonomy: The systematic evaluation of comparative data* (2 ed.). New York: Columbia University Press.

Swofford, D. L. (1989). Phylogenetic analysis using parsimony. *Illinois Natural History Survey, Champaign*.

Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy, and J.-F. Cosson (1998). Comparative phylogeography and postglacial colonization routes in europe. *Molecular Ecology 7*(4), 453–464.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology 17*(5), 1105–1109.

Tajima, F. and M. Nei (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution 1*(3), 269–285.

Talavera, G. and J. Castresana (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology 56*(4), 564 –577.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution 9*(4), 678–687.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution 28*(10), 2731–2739.

Tedder, A., P. Hoebe, S. Ansell, and B. Mable (2010). Using chloroplast trnF pseudogenes for phylogeography in arabidopsis lyrata. *Diversity 2*, 653–678.

Thompson, J. (2005). *The geographic mosaic of coevolution*. University Of Chicago Press.

Thompson, J. N., S. L. Nuismer, and K. Merg (2004). Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biological Journal of the Linnean Society 82*(4), 511–519.

Thomson, J. D. (2001). How do visitation patterns vary among pollinators in relation to floral display and floral design in a generalist pollination system? *Oecologia 126*(3), 386–394.

Tzedakis, P. C. (2009). Cenozoic climate and vegetation change. In J. Woodward (Ed.), *The physical geography of the Mediterranean*, pp. 89–137. Oxford University Press.

Valtueña, F. J., C. D. Preston, and J. W. Kadereit (2012). Phylogeography of a tertiary relict plant, meconopsis cambrica (papaveraceae), implies the existence of northern refugia for a temperate herb. *Molecular Ecology 21*(6), 1423–1437.

Vargas, P. (2003). Molecular evidence for multiple diversification patterns of alpine plants in mediterranean europe. *Taxon 52*(3), 463.

Vogt, L. (2002). Weighting indels as phylogenetic markers of 18S rDNA sequences in diptera and strepsiptera. *Organisms Diversity & Evolution 2*(4), 335–349.

Vrancken, J., C. Brochmann, and R. A. Wesselingh (2012). A european phylogeography of rhinanthus minor compared to rhinanthus angustifolius: unexpected splits and signs of hybridization. *Ecology and Evolution 2*(7), 1531–1548.

Warwick, S. and I. Al-Shehbaz (2006). Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution 259*(2), 237–248.

Wen, C. and J. Hsiao (2001). Altitudinal genetic differentiation and diversity of taiwan lily (lilium longiflorum var. formosanum; liliaceae) using RAPD markers and morphological characters. *International Journal of Plant Sciences 162*(2), 287–295. ArticleType: research-article / Full publication date: March 2001 / Copyright Â© 2001 The University of Chicago Press.

Wheeler, W. (1996). Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics 12*(1), 1–9.

Wheeler, W., D. Gladstein, and J. De Laet (2003). POY, phylogeny reconstruction via optimization of DNA and other data version 3.0. 11 (may 6 2003). *American Museum of Natural History*.

Wheeler, W. C., J. Gatesy, and R. DeSalle (1995). Elision: A method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. *Molecular Phylogenetics and Evolution 4*(1), 1–9.

Wiens, J. J., T. N. Engstrom, and P. T. Chippindale (2006). Rapid diversification, incomplete isolation, and the speciation clock in north american salamanders (genus plethodon): testing the hybrid swarm hypothesis of rapid radiation. *Evolution 60*(12), 2585–2603.

Wilson, P., M. C. Castellanos, J. N. Hogue, J. D. Thomson, and W. S. Armbruster (2004). A multivariate search for pollination syndromes among penstemons. *Oikos 104*(2), 345–361.

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics 2*(5), 333–341.

Wolfe, K. H., W. H. Li, and P. M. Sharp (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences 84*(24), 9054–9058.

Wright, S. (1943). Isolation by distance. *Genetics 28*(2), 114–138.

Young, N. and J. Healy (2003). GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics 4*(1), 6.

Zeng, Y., W.-J. Liao, R. J. Petit, and D. A. Y. Zhang (2011). Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Molecular ecology*.

Part II

DEVELOPMENT OF METHODS

2

CHARACTERIZATION OF MICROSATELLITE LOCI IN *ERYSIMUM MEDIOHISPANICUM* (BRASSICACEAE) AND CROSS-AMPLIFICATION IN RELATED SPECIES

ABSTRACT

We have developed and optimized microsatellite loci from a genomic library of *Erysimum mediohispanicum*. Microsatellites were also tested for cross-amplification in 32 other *Erysimum* species. A total of 10 microsatellite loci were successfully amplified. They were polymorphic for 81 *E. mediohispanicum* individuals from two locations in Sierra Nevada (south-east Spain), which showed similar patterns of genetic diversity. On average, microsatellites had 8.6 alleles per locus, and an expected heterozygosity of 0.69. Only one locus significantly departed from Hardy-Weinberg equilibrium in both locations. Most of the markers successfully amplified in other *Erysimum* species. The genetic attributes of microsatellite loci will allow their application for population genetic studies in *Erysimum*, such as genetic differentiation and structure, gene flow, pollinator-mediated speciation and hybridization studies.[1][2]

[1] Authors: A. Jesús Muñoz-Pajares, M. Belén Herrador, Mohamed Abdelaziz, F. Xavier Picó, Timothy F. Sharbel, José M. Gómez, and Francisco Perfectti

[2] Published: American Journal of Botany Primer Notes & Protocols (2011) e287-e289.

INTRODUCTION

Erysimum (Brassicaceae) is a genus comprising approximately 223 species (Al-Shehbaz et al., 2006) mainly distributed in the Northern hemisphere. *Erysimum mediohispanicum* Polatschek is a biennial to perennial monocarpic herb endemic of the Iberian Peninsula, where it is distributed in two geographically separated areas: one in north-east and the other in south-east. *E. mediohispanicum* typically exhibits high within- and among-population variation in floral traits, in particular corolla shape and size (Gómez et al., 2009). It has been shown that different pollinators discriminate among different flower shapes, suggesting that phenotypic evolution and diversification can occur in this species (Gómez et al., 2008, 2009). We have characterized ten new polymorphic microsatellite loci for *E. mediohispanicum* and tested their cross-amplification in 32 other *Erysimum* species from different locations across Europe and Africa. These microsatellite loci will be primarily used for population genetic and gene flow studies in *E. mediohispanicum* populations characterized by different patterns of variation in floral shape and size as well as in the species' pollinator community.

METHODS AND RESULTS

Microsatellite libraries were developed by Genetic Identification Services (Chatsworth, CA, USA, www.genetic-id-services.com) following Jones et al. (2002). Genomic DNA was extracted using the DNeasy Plant Extraction kit (Qiagen, Venlo, Netherlands) from one individual. Approximately 100 µg of DNA were digested with RsaI, HaeIII, BsrB1, PvuII, StuI, ScaI and EcoRV (New England Biolabs, Ipswich, MA, USA) and subsequent fragments were enriched in four motives: CA-, AAC-, ATG- and GA- using Biotin and Steptavidin magnetic beads for reversible capture (CPG Inc. East Bank Demerara, Guyana). Resulting fragments were ligated into pUC19 plasmid and cloned into an E. coli strain DH5alpha (Invitrogen Carlsbad, CA, USA). After incubation, a total of 100 randomly chosen recombinant clones were selected, purified and sequenced. Primer pairs were designed for 26 clones showing tandemly repeated motifs flanked by high quality sequence regions. Primer design was conducted using Designer PCR 1.03 (Research Genetics Inc., Huntsville, AL, USA). Such primers were tested on a total of 81 *E. mediohispanicum* individuals from two locations in south-east Spain: El Dornajo (37°7.67′N, 3°25.77′W; N = 30, UGR herbarium specimen: GDAC17407-1-2) and La Cortijuela (37°4.66′N, 3°28.29′W; N = 51, UGR herbarium specimen: GDAC17421-1-2).

DNA was isolated from silica-dried leaf samples using the GenElute™ Plant Genomic DNA Miniprep kit (Sigma-Aldrich, St. Louis, MO, USA). Polymerase chain reactions were performed in 15 µl of reaction mixture containing 0.17 ng of template genomic DNA, 1× Buffer (ref. M0273S, New England BioLabs, Ipswich, MA, USA), 0.16 mM each dNTP (Sigma-Aldrich, St. Louis, MO, USA), 0.33 µM each forward (fluorescently tagged, Applied Biosystems, Forster City, CA, USA) and reverse primer, and 0.02 U/µL Taq polymerase (ref. M0273S, New England Biolabs, Ipswich, MA, USA). Polymerase chain reactions were conducted in a Gradient Master Cycler Pro S (Eppendorf, Hamburg, Germany) with an initial 30 s of denaturation at 94 °C, 35 cycles at 94 °C for 15 s, annealing temperatures (Ta; Table 2.1) for 30 s, extension at 72 °C for 30 s, and a final extension at 72 °C for 3 min. PCR products were diluted 1:15 and analyzed by MACROGEN analyzers (Geumchun-gu, Seoul, Korea, www.macrogen.com) using 400HD ROX (Applied Biosystems, Forster City, CA, USA) as standard. Alleles were called using Peak Scanner ™ Software v.1.0 (Applied Biosystems, Forster City, CA, USA).

Seven out of the 26 tested primers failed to amplify, nine were monomorphic or showed complex patterns (primer sequences available upon request) and ten were polymorphic (Table 2.1). We es-

Locus (GenBank accession)	Repeat motif	Primer sequence (5′-3′)	Fluorescent label	Product size(bp)	Ta (ºC)
C5 (JF766210)	CCA$_8$	F: TCTTTCTTCTGCGGTTTATTC R: CGTTTTTGTTGTTGTTCTGG	6-FAM	164–182	56
D2 (JF766211)	CAT$_{23}$	F: ACGGAAGATGACGATGATCGACTG R: CAATGTCCCTAATTGGTCAATGG	HEX	117–189	54
D4 (JF766212)	ATC$_7$	F: TAAGGTGTTACCGGATTGTC R: GTGACGATTCGCTCCTTG	NED	200–215	57
E4 (JF766213)	CT$_{20}$	F: CCTTCCTCCGACTACTCTCC R: TGAGCGACTGATGATGATTC	HEX	145–178	57
E8 (JF766214)	CT$_{50}$	F: AGCTCACAGCCGTCGATGTTTGC R: GAGGTGAAATACACGTAGAACCT	6-FAM	157–229	50
D11 (JF766215)	TCA$_{14}$	F: TCCAGGGTCTGAGTCAATATG R: TTACCACTCCTTGCTTCTGAA	6-FAM	179–197	53
E6 (JF766216)	TC$_{14}$	F: CTTGTAACCGAGCCACTCA R: ATACGGAGAAGAAAGCGAATC	NED	131–159	53
D10 (JF766217)	TCA$_{12}$	F: ACTGCCATCAAACGACCTC R: TTGGTTGGAAAAGGGATTG	NED	166–185	53
E5 (JF766218)	GA$_{13}$	F: TCCATTTACACAATCCGTTCAT R: CCAACCTGACATCTTTGCTTC	6-FAM	167–195	50
E3 (JF766219)	GA$_{17}$	F: TTCCTCCAGATGAAACTACACAGG R: ACTTACATCGGATCGGTTGAG	HEX	215-253	56

Tabla 2.1: Characteristics of ten microsatellite markers in *E. mediohispanicum*. Loci names, GenBank accession number, repeat motifs, forward (F) and reverse (R) primers, F primer fluorescent tag, allele size ranges and optimal annealing temperatures (Ta) are given.

timated the number of observed alleles per locus (na), allelic richness (RS), observed heterozygosity (Ho), and expected heterozygosity (HS) using FSTAT v.2.9.3 (Goudet, 1995). Linkage disequilibrium was tested with FSTAT using Bonferroni correction and departures from Hardy-Weinberg equilibrium (HWE) were performed with GenAlEx v.6.3 (Peakall and Smouse, 2006).

For El Dornajo, the number of alleles per locus varied between two and 19, allelic richness between 2.0 and 18.5, observed heterozygosity

	El Dornajo						La Cortijuela					
Locus	N	Na	RS	HO	HS	HWE	N	Na	RS	HO	HS	HWE
C5	27	7	7	0.74	0.737	0.984 ns	50	6	5.94	0.76	0.778	0.766 ns
D2	29	19	18.5	0.72	0.925	0.057 ns	50	17	16.87	0.78	0.891	0.041 *
D4	30	7	6.6	0.5	0.629	0.277 ns	51	6	5.92	0.59	0.64	0.567 ns
E4	29	9	8.92	0.62	0.844	0.141 ns	51	13	12.83	0.71	0.854	0.001 **
E8	29	10	9.93	0.31	0.861	0.000 ***	47	14	14	0.45	0.851	0.000 ***
D11	27	3	3	0.52	0.562	0.600 ns	48	5	4.98	0.46	0.564	0.003 **
E6	29	7	6.93	0.72	0.664	0.920 ns	50	8	7.87	0.68	0.696	0.993 ns
D10	29	5	5	0.59	0.645	0.734 ns	51	5	5	0.47	0.481	0.221 ns
E5	28	12	11.89	0.71	0.87	0.798 ns	51	12	11.68	0.75	0.797	0.131 ns
E3	28	2	2	0.11	0.278	0.001 **	49	4	3.96	0.14	0.155	1.000 ns

Tabla 2.2: Genetic diversity of two populations of *E. mediohispanicum*. Sample size (N), number of observed alleles (Na), allelic richness (RS), observed heterozygosity (HO), expected heterozygosity (HS), and P-values for departure from Hardy-Weinberg equilibrium (HWE) test are given for each microsatellite marker and population. HWE significance: ns = not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

between 0.11 and 0.74, and expected heterozygosity between 0.28 and 0.93 (Table 2.2) . For La Cortijuela, the number of alleles per locus varied between four and 17, allelic richness between 4.0 and 16.9, observed heterozygosity between 0.14 and 0.78, and expected heterozygosity between 0.16 and 0.89 (Table 2.2). All loci showed no linkage disequilibrium ($P > 0.004$ in all cases; nominal level = 0.0011). For El Dornajo and La Cortijuela, two and four out of ten loci significantly departed from HWE, respectively (Table 2.2). However, only one out of ten loci significantly departed from HWE in both locations (Table 2.2).

We analyzed the cross-amplification success of the ten polymorphic microsatellite loci in a total of 32 *Erysimum* species (Table 2.3). Most of the loci amplified in several species. On average, each loci amplified in 79.1 % (range = 50.0 – 96.9 %) of the species (Table 2.3).

	Locus									
Species	C5	D2	D4	E4	E8	D11	E6	D10	E5	E3
E. baeticum	+	-	+	+	+	-	+	+	+	+
E. bonnanianum var. aetnense	+	+	+	+	+	+	+	+	-	+
E. bonnanianum	+	+	+	+	-	-	+	+	-	+
E. bicolor	+	-	+	+	+	+	+	+	+	+
E. corinthium	+	+	+	+	+	+	+	+	+	-
E. collisparsum	+	+	+	+	-	+	+	+	-	+
E. crepidifolium	+	+	+	+	-	+	+	+	-	+
E. duriaei	+	+	+	+	+	-	+	+	+	+
E. fitzii	+	+	+	+	+	+	+	+	+	+
*E. gomezcampo*i	+	+	+	+	-	-	+	+	+	+
E. gorbeanum	+	+	+	+	-	+	+	+	+	+
E. incanum subsp. mairei	+	-	-	+	+	+	+	+	-	+
E. jugicola	+	+	+	+	-	+	+	+	+	+
E. nevadense	+	+	+	+	+	+	+	+	+	+
E. nervosum	+	-	+	+	+	-	-	-	+	-
E. metlesicsii	+	+	+	+	-	+	+	-	-	-
E. mexmuellieri	+	-	+	+	+	+	+	+	+	+
E. myriophylum	+	-	+	-	+	-	-	-	+	+
E. myriophylum var. cazorlense	+	+	+	+	+	+	+	+	+	+
E. odoratum	+	+	+	+	-	+	+	+	-	+
E. popovii	+	-	+	+	+	-	+	+	+	+
E. penyalarense	+	+	+	+	+	+	+	+	+	+
E. pseudorhaeticum	+	+	+	+	-	-	+	+	-	+
E. rondae	+	+	+	+	-	+	+	+	+	+
E. rhaeticum	+	+	+	+	-	+	+	+	-	+
E. riphaeanum	+	+	+	+	-	+	+	+	+	-
E. ruscinonense	+	+	+	+	-	-	+	+	+	+
E. seipkae	+	+	+	+	-	+	+	+	+	+
E. scoparium	+	+	+	+	-	+	+	+	+	+
E. semperflorens	+	-	+	+	+	-	+	+	+	-
E. sylvestre	+	-	+	+	-	-	+	-	-	+
E. wilczekianum	-	-	-	-	+	-	+	-	-	-

Tabla 2.3: Characteristics of ten microsatellite markers of *E. mediohispanicum*. GenBank accession number (below loci names), repeat motifs, forward (F) and reverse (R) primers, allele size ranges and optimal annealing temperatures (Ta) are given.

CONCLUSIONS

Genetic diversity parameters indicate that these microsatellite loci can be a useful tool to study neutral genetic variation in *E. mediohispanicum* populations. Questions like the relationship between genetic diversity and phenotypic diversity, the effects of geographical variation in pollinator abundance and diversity on genetic diversity, differentiation and structure, or the extent of pollinator-mediated gene flow within and among-populations can now be addressed. Given that most of the primers successfully amplified a band of the expected size in several *Erysimum* species, these microsatellites also have the potential to become an efficient molecular tool to address similar questions in other *Erysimum* species, as well as to explore speciation processes and contact zones commonly found in this highly diverse genus across its distribution range.

ACKNOWLEDGEMENTS

The authors thank the staff of Genetic Identification Services (Chatsworth, CA, USA) for microsatellite development, Leticia Ayllón-Egea and Yedra García-García for technical support.

REFERENCES

Al-Shehbaz, I., M. Beilstein, and E. Kellogg (2006). Systematics and phylogeny of the brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution 259*(2), 89–120.

Gómez, J. M., M. Abdelaziz, J. P. M. Camacho, A. J. Muñoz-Pajares, and F. Perfectti (2009). Local adaptation and maladaptation to pollinators in a generalist geographic mosaic. *Ecology Letters 12*(7), 672–682.

Gómez, J. M., M. Abdelaziz, J. Muñoz-Pajares, and F. Perfectti (2009). Heritability and genetic correlation of corolla shape and size in erysimum mediohispanicum. *Evolution 63*(7), 1820–1831.

Gómez, J. M., J. Bosch, F. Perfectti, J. Fernández, M. Abdelaziz, and J. Camacho (2008). Spatial variation in selection on corolla shape in a generalist plant is promoted by the preference patterns of its local pollinators. *Proceedings of the Royal Society B: Biological Sciences 275*(1648), 2241 –2249.

Goudet, J. (1995). Fstat (version 1.2): A computer program to calculate f-statistics. *Journal of Heredity 86*(6), 485 –486.

Jones, K. C., K. F. Levine, and J. D. Banks (2002). Characterization of 11 polymorphic tetranucleotide microsatellites for forensic applications in california elk (Cervus elaphus canadensis). *Molecular Ecology Notes 2*(4), 425–427.

Peakall, R. and P. E. Smouse (2006). genalex 6: genetic analysis in excel. population genetic software for teaching and research. *Molecular Ecology Notes 6*(1), 288–295.

3

SIDIER: SUBSTITUTION AND INDEL DISTANCES TO INFER EVOLUTIONARY RELATIONSHIPS

ABSTRACT

SIDIER is a new software package in the R language for inferring evolutionary relationships from gapped alignments using information contained in both substitutions and insertions and deletions (indels). SIDIER estimates the number of indel events occurred during sequence evolution to obtain a distance matrix that may be combined with the substitutions matrix calculated separately from the same dataset. The inferred evolutionary process may be represented by means of percolated networks.[1]

[1] Author: A. Jesús Muñoz-Pajares

INTRODUCTION

Although insertions and deletions (indels) may provide valuable information for evolutionary reconstruction (Simmons et al., 2001; Young and Healy, 2003; Blair and Murphy, 2011), phylogenetic and phylogeographic inferences are usually based on nucleotide substitutions, dismissing gapped positions before conducting the analysis(Talavera and Castresana, 2007). Because combining the information provided by indels and substitutions will surely improve the accuracy of evolutionary inferences (Vogt, 2002; Müller, 2006), an increasing number of studies are exploring this combination under different theoretical frameworks such as maximum parsimony, maximum likelihood, and Bayesian inference (Baum et al., 1994; Freudenstein and Chase, 2001; Geiger, 2002; Rivas and Eddy, 2008). However, the question has been scarcely explored using distance-based methods (Ogden and Rosenberg, 2007) despite they allow intuitively combining data from different sources and require low computation times (Tamura et al., 2004).

Distance methods are particularly powerful to accurately reconstruct both large evolutionary trees and short-branch trees (Roch, 2010). In these two scenarios, information provided by nucleotide substitutions may be limited due to sequence saturation and low divergence, respectively. Consequently, using indels are especially advantageous in these two situations (e.g., Borsch et al., 2003; Redelings and Suchard, 2007). Whereas the evolutionary relationships among taxa showing large branches may surely be properly represented by a tree topology, the evolution of organisms showing low differentiation (such as close related species or populations belonging to the same species) may be better represented using network approaches (Morrison, 2005; Mardulyn, 2012).

The main objective of the SIDIER package is to disentangle the evolutionary relationships among gapped sequences (and the popu-

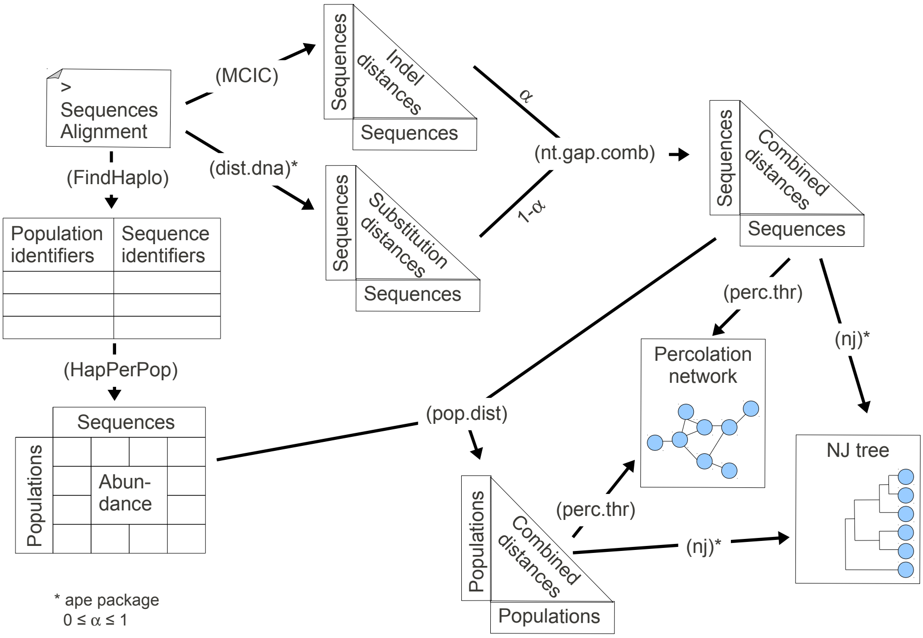


Figura 3.1: Representation of the complete process to obtain a percolation network combining the information provided by both, substitutions and indels. Functions used for each step are represented in brackets. The procedure uses an alignment in fasta format to estimate the distances between sequences based on indels and substitutions (the latter using the *ape* package; Paradis et al., 2004), which are combined giving different weights to each matrix (α and 1-α, respectively). The resulting combined distance matrix may be represented as a percolation network and used (together with the abundance of sequences) to estimate a population distance matrix, which may also be represented as a network. Alternatively, the inferred distance matrices may be represented as a tree (for example, using the *ape* package; Paradis et al., 2004).

lations they represent) combining the information provided by both, indels and substitutions, in a distance-based framework. The complete process is represented in Fig. 3.1 and described below, together with the main features of the software implementation. See package

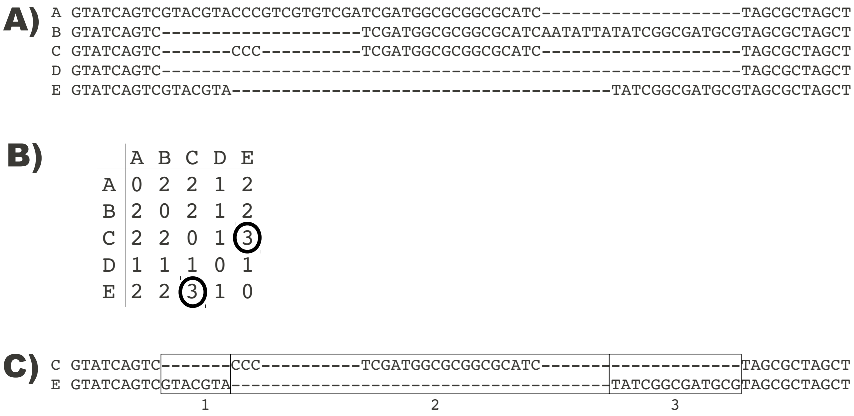

A)
A GTATCAGTCGTACGTACCCGTCGTGTCGATCGATGGCGCGGCGCATC--------------------TAGCGCTAGCT
B GTATCAGTC--------------------TCGATGGCGCGGCGCATCAATATTATATCGGCGATGCGTAGCGCTAGCT
C GTATCAGTC-------CCC----------TCGATGGCGCGGCGCATC--------------------TAGCGCTAGCT
D GTATCAGTC--TAGCGCTAGCT
E GTATCAGTCGTACGTA--------------------------------------TATCGGCGATGCGTAGCGCTAGCT
B)
A B C D E
A 0 2 2 1 2
B 2 0 2 1 2
C 2 2 0 1 3
D 1 1 1 0 1
E 2 2 3 1 0
C)
C GTATCAGTC-------CCC----------TCGATGGCGCGGCGCATC--------------------TAGCGCTAGCT
E GTATCAGTCGTACGTA--------------------------------------TATCGGCGATGCGTAGCGCTAGCT
1
2
3

Figura 3.2: Example of an indel distance matrix estimation using the *MCIC* function. A) Gapped alignment. B) Inferred indel distance matrix. C) Details on distance estimation, showing the three mutation events considered between sequences C and E. Because gapped regions shared by two sequences are not taken into account for the distance estimation, the two nucleotide motifs shown in the central region of the sequence C compute as a single mutation event (adapted from Müller, 2006).

manual (Supplementary Information) for further details on specific functions and data examples.

DESCRIPTION AND IMPLEMENTATION

Given an alignment, SIDIER estimates the number of indel events occurred between each pair of sequences following the Modified Complex Indel Coding rationale (MCIC; Müller, 2006) to estimate pairwise transformation costs. This method ignores gap positions shared by each pair of sequences being compared, yielding a biologically realistic distance matrix (Müller, 2006, Fig. 3.2).

To estimate the pairwise indel distance matrix (Fig. 3.3) SIDER considers the complete repetitive region as a single character (i.e. removing step I in Müller, 2006). Briefly, for each pair of haplotypes

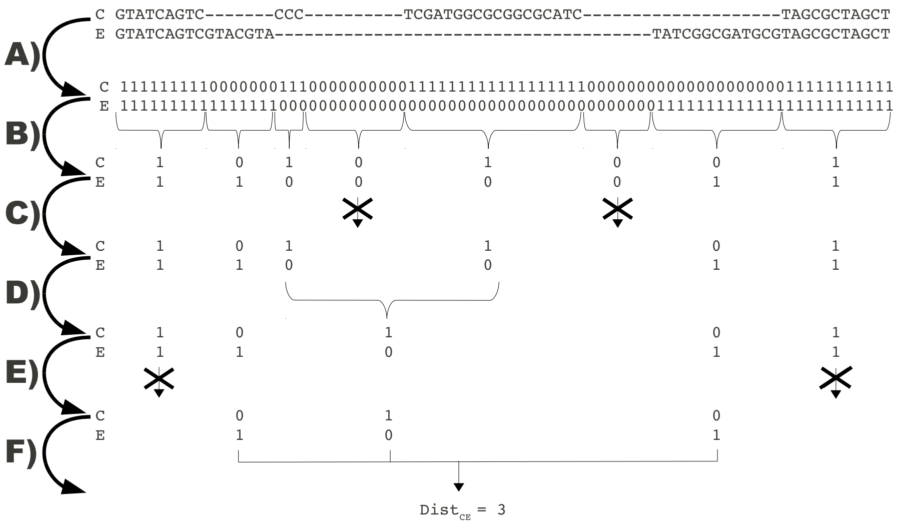

A)
B)
C)
D)
E)
F)
C GTATCAGTC-------CCC----------TCGATGGCGCGGCGCATC--------------------TAGCGCTAGCT
E GTATCAGTCGTACGTA--------------------------------------TATCGGCGATGCGTAGCGCTAGCT
C 111111111000000011100000000001111111111111111110000000000000000000011111111111
E 111111111111111100000000000000000000000000000000000000111111111111111111111111
C 1 0 1 0 1 0 0 1
E 1 1 0 0 0 0 1 1
C 1 0 1 1 0 1
E 1 1 0 0 1 1
C 1 0 1 0 1
E 1 1 0 1 1
C 0 1 0
E 1 0 1
Dist$_{CE}$ = 3

Figura 3.3: Schematic view of the pairwise indel distance estimation. A) Sequences are codified as binary (0 for gaps and 1 for A, T, C or G). B) Identical adjacent positions are merged. C) Positions showing 0 in both sequences are removed. D) Among the remaining positions, those showing either 0-1 or 1-0 are merged. E) Positions showing 1 in both sequences are removed. F) The number of indel events are estimated by counting the number of positions maintained after these steps. (Figure adapted from Müller, 2006).

the software codifies nucleotides as binary (0 for gaps and 1 for A, T, C or G; Fig. 3.3A) and identical adjacent positions are removed (Fig. 3.3B). Then, positions showing a value 0 in both sequences are removed (Fig. 3.3C), and those with either 0-1 or 1-0 are merged (Fig. 3.3D). Finally, 1-1 positions are ignored (Fig. 3.3E), and the number of remaining positions, equal to the absolute number of indel events, are used as the indel-based distances (Fig. 3.3F).

To represent a distance matrix as a tree or as a network it is needed to define which distance values should be depicted as a link between nodes and which values should not (that is, a connection threshold must be defined). For that, SIDIER determines the percolation threshold as described by Rozenfeld et al. (2008) and connects only populations showing genetic distances lower than this value. To determine the percolation threshold, the software calculates networks assuming different connection thresholds (provided by user). To estimate each network, SIDIER computes a new distance matrix by setting to zero all distances higher than the defined connection threshold and the new matrix is handled as an adjacency matrix by *igraph* (Csardi and Nepusz, 2006) and *network* (Butts, 2008) packages. For each network, SIDIER estimates the average size of clusters excluding the largest one (<s>; Rozenfeld et al., 2008) as follows:

$$< s >= \frac{1}{N} \sum_{s<s_{max}} s^2 n_s$$

where N is the number of nodes not included in the largest cluster and n_s is the number of nodes containing s nodes. The percolation threshold is calculated as an increase in <s> value as connection threshold decreases (Rozenfeld et al., 2008). By default, the software calculates 101 networks using 101 different connection thresholds, ranging from 100 % of the maximum distance found in the matrix (where all nodes are connected) to 0 % of this maximum distance (all nodes isolated). For each network, SIDIER may find modules (defined as subsets of nodes conforming densely connected subgraphs) by means of random walks as implemented in the *igraph* package (Csardi and Nepusz, 2006).

Using the mentioned methodology, SIDIER allows to visualize and compare results estimated from indel and substitution distances of the same sequence dataset (the latter distances can be estimated using, for instance, the *dist.dna* function in *ape* package). Additionally, it is also possible to visualize a single network from both datasets by obtaining the weighted combination of both matrices using the *nt.gap.comb* function (Fig. 3.1).

If the studied sequences represent different geographic locations, SIDIER may also estimate relationships among populations. For that, it is needed to obtain a genetic matrix representing the pairwise population distances (Fig. 3.1). In this matrix, each element is calculated as the arithmetic mean of the distances among all the sequences sampled in both populations:

$$\mathrm{dist}(i,j) = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n} \mathrm{dist}(H_{ki}, H_{lj})}{m * n}$$

Where *dist(i,j)* represents the distance between populations *i* and *j*, *m* and *n* are the number of sequences corresponding to populations *i* and *j*, respectively, and *dist(*H_{ki}*,*H_{lj}) is the distance between the *k-th* sequence found in population *i* and the *l-th* sequence found in population *j*. To obtain the population distance matrix, SIDIER uses as input the abundance of each sequence per population, which also may be estimated with the software (Fig. 3.1).

Besides extract evolutionary information from indels, as illustrated in Fig. 3.1, SIDIER provides functions that may be useful by themselves, for example, to identify unique sequences in a given alignment, to estimate the sequences abundance per population, or to obtain the weighted combination of two matrices yielded from different sources (e.g., genetic and phenotypic data). Additionally, the package provides one of the first available functions to represent the information contained in any distance matrix as a percolation network.

REFERENCES

Baum, D. A., K. Sytsma, and P. C. Hoch (1994). A phylogenetic analysis of epilobium (onagraceae) based on nuclear ribosomal DNA sequences. *Systematic Botany 19*, 363–388.

Blair, C. and R. W. Murphy (2011). Recent trends in molecular phylogenetic analysis: Where to next? *Journal of Heredity 102*(1), 130 –138.

Borsch, T., K. Hilu, D. Quandt, V. Wilde, C. Neinhuis, and W. Barthlott (2003). Noncoding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms. *J Evol Biol 16*(4), 558–76.

Butts, C. T. (2008). network: A package for managing relational data in r. *Journal of Statistical Software 24*(2), 1–36.

Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Freudenstein, J. V. and M. W. Chase (2001). Analysis of mitochondrial nad1b-c intron sequences in orchidaceae: Utility and coding of length-change characters. *Systematic Botany 26*(3), 643–657.

Geiger, D. (2002). Stretch coding and block coding: two new strategies to represent questionably aligned DNA sequences. *J Mol Evol 54*(2), 191–9.

Mardulyn, P. (2012). Trees and/or networks to display intraspecific DNA sequence variation? *Molecular Ecology 21*(14), 3385–3390.

Morrison, D. A. (2005). Networks in phylogenetic analysis: new tools for population biology. *International Journal for Parasitology 35*(5), 567–582.

Müller, K. (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution 38*(3), 667–676.

Ogden, T. H. and M. S. Rosenberg (2007). How should gaps be treated in parsimony? a comparison of approaches using simulation. *Molecular Phylogenetics and Evolution 42*(3), 817–826.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in r language. *Bioinformatics 20*, 289–290.

Redelings, B. and M. Suchard (2007). Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. *BMC Evolutionary Biology 7*(1), 40.

Rivas, E. and S. Eddy (2008). Probabilistic phylogenetic inference with insertions and deletions. *PLoS Comput Biol 4*(9), e1000172.

Roch, S. (2010). Toward extracting all phylogenetic information from matrices of evolutionary distances. *Science 327*(5971), 1376 –1379.

Rozenfeld, A. F., S. Arnaud-Haond, E. Hernández-García, V. M. Eguíluz, E. A. Serrão, and C. M. Duarte (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences 105*(48), 18824 –18829.

Simmons, M., H. Ochoterena, and T. Carr (2001, June). Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst Biol 50*(3), 454–62.

Talavera, G. and J. Castresana (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology 56*(4), 564 –577.

Tamura, K., M. Nei, and S. Kumar (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America 101*(30), 11030–11035.

Vogt, L. (2002). Weighting indels as phylogenetic markers of 18S rDNA sequences in diptera and strepsiptera. *Organisms Diversity & Evolution 2*(4), 335–349.

Young, N. and J. Healy (2003). GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics 4*(1), 6.

SUPPLEMENTARY INFORMATION

Package ‘sidier’

December, 2012

Version 0.1

Date 2012-12-01

Title Substitution and Indel Distances to Infer Evolutionary Relationships

Author A. Jesús Muñoz-Pajares

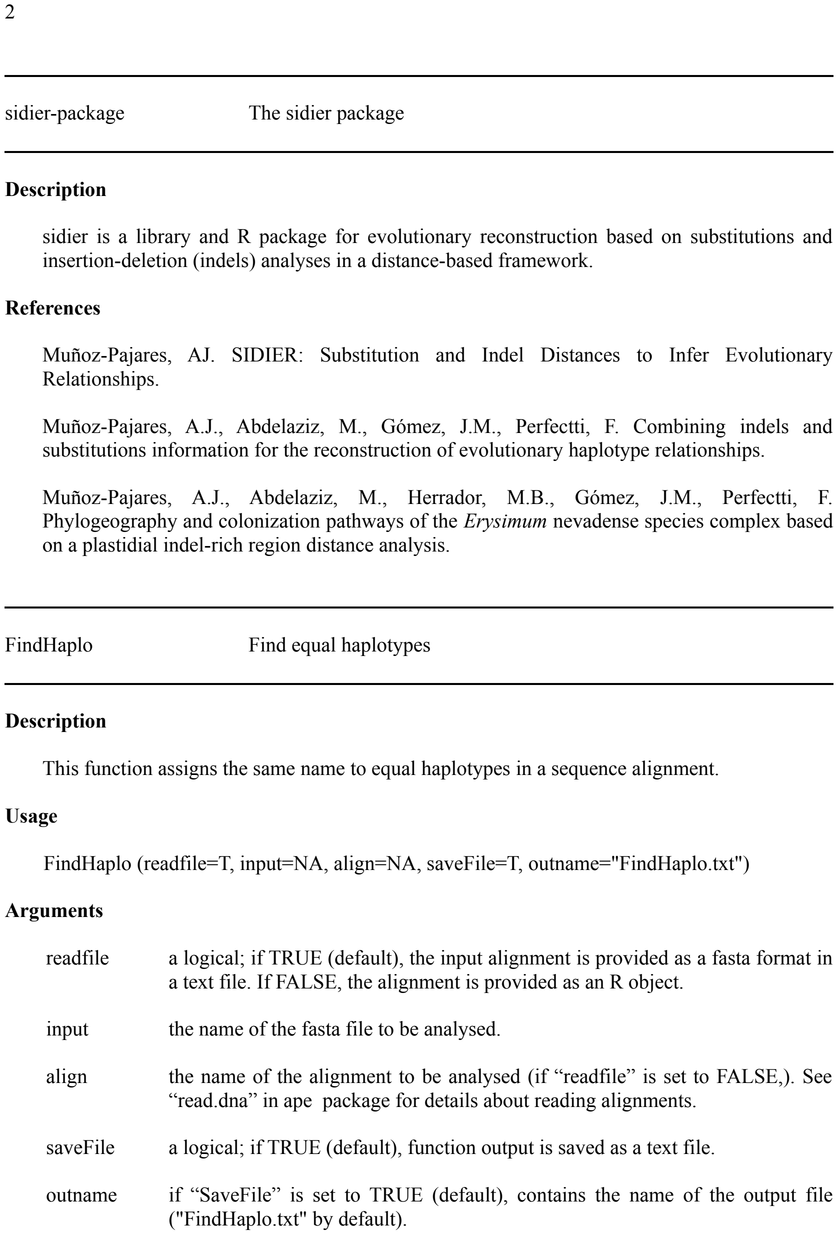
Maintainer A. Jesús Muñoz-Pajares <ajesusmp@ugr.es>

Imports ape, igraph, network

ZipData no

Description sidier provides functions for reading and writing fasta sequences, finding unique haplotypes, estimating genetic distances based on gap positions and lengths, combining distance matrices and estimating and plotting percolation networks.

R topics documented:



sidier-package The sidier package

Description

sidier is a library and R package for evolutionary reconstruction based on substitutions and insertion-deletion (indels) analyses in a distance-based framework.

References

Muñoz-Pajares, AJ. SIDIER: Substitution and Indel Distances to Infer Evolutionary Relationships.

Muñoz-Pajares, A.J., Abdelaziz, M., Gómez, J.M., Perfectti, F. Combining indels and substitutions information for the reconstruction of evolutionary haplotype relationships.

Muñoz-Pajares, A.J., Abdelaziz, M., Herrador, M.B., Gómez, J.M., Perfectti, F. Phylogeography and colonization pathways of the *Erysimum* nevadense species complex based on a plastidial indel-rich region distance analysis.

FindHaplo Find equal haplotypes

Description

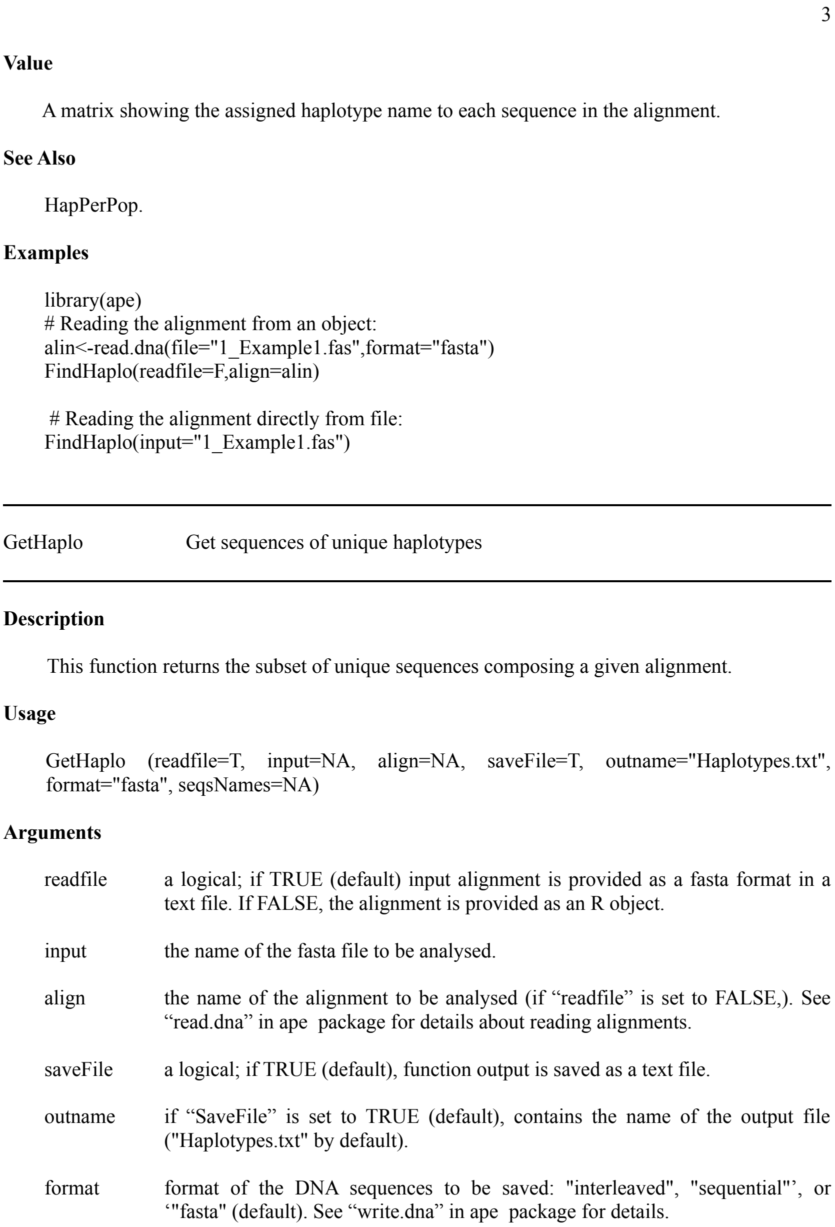
This function assigns the same name to equal haplotypes in a sequence alignment.

Usage

```
FindHaplo (readfile=T, input=NA, align=NA, saveFile=T, outname="FindHaplo.txt")
```

Arguments

readfile	a logical; if TRUE (default), the input alignment is provided as a fasta format in a text file. If FALSE, the alignment is provided as an R object.
input	the name of the fasta file to be analysed.
align	the name of the alignment to be analysed (if "readfile" is set to FALSE,). See "read.dna" in ape package for details about reading alignments.
saveFile	a logical; if TRUE (default), function output is saved as a text file.
outname	if "SaveFile" is set to TRUE (default), contains the name of the output file ("FindHaplo.txt" by default).



Value

A matrix showing the assigned haplotype name to each sequence in the alignment.

See Also

HapPerPop.

Examples

```
library(ape)
# Reading the alignment from an object:
alin<-read.dna(file="1_Example1.fas",format="fasta")
FindHaplo(readfile=F,align=alin)

 # Reading the alignment directly from file:
FindHaplo(input="1_Example1.fas")
```

GetHaplo Get sequences of unique haplotypes

Description

This function returns the subset of unique sequences composing a given alignment.

Usage

```
GetHaplo (readfile=T, input=NA, align=NA, saveFile=T, outname="Haplotypes.txt",
format="fasta", seqsNames=NA)
```

Arguments

readfile a logical; if TRUE (default) input alignment is provided as a fasta format in a text file. If FALSE, the alignment is provided as an R object.

input the name of the fasta file to be analysed.

align the name of the alignment to be analysed (if “readfile” is set to FALSE,). See “read.dna” in ape package for details about reading alignments.

saveFile a logical; if TRUE (default), function output is saved as a text file.

outname if “SaveFile” is set to TRUE (default), contains the name of the output file ("Haplotypes.txt" by default).

format format of the DNA sequences to be saved: "interleaved", "sequential"’, or ‘"fasta" (default). See “write.dna” in ape package for details.

4

seqsNames	names for each DNA sequence saved: Three choices are possible: if n unique sequences are found, "Inf.Hap" assign names from H1 to Hn (according to input order). The second option is to define a vector containing n names. By default, input sequence names are used.

Details

If two equal sequences are not identically aligned, they will be considered as different haplotypes. To avoid misleading results in uncertain alignments it is recommended to use as input the original unaligned sequences, including gaps after the last nucleotide of short sequences to make all sequence lengths equal.

Value

A file containing unique sequences from the input file.

Examples

```
library(ape)
 # Reading the alignment from an object and saving haplotypes names as sequential numbers:
alin<-read.dna(file="2_Example2.fas",format="fasta")
GetHaplo(readfile=F,align=alin,outname="Haplotypes_sequentialNames.txt",seqsNames="Inf.
Hap")

# Reading the alignment directly from file and saving using sequence input names:
GetHaplo(input="2_Example2.fas")
```

HapPerPop Returns the number of haplotypes per population.

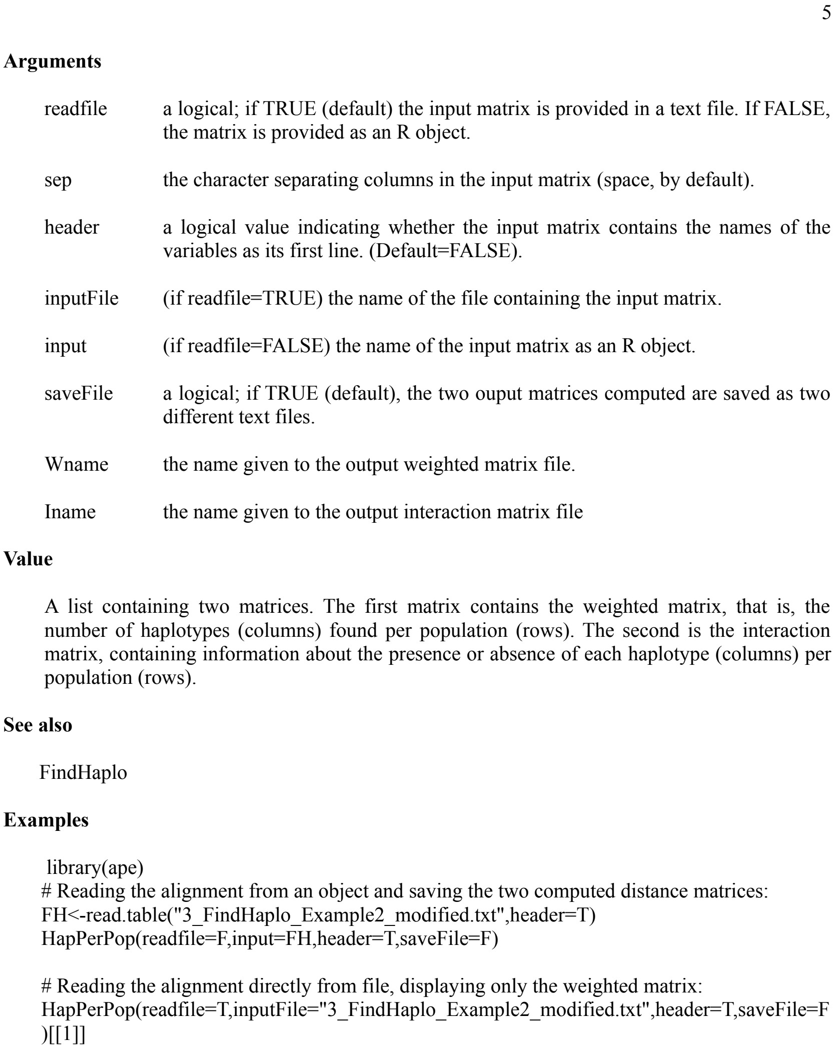
Description

Given a two column matrix, this function returns the number of haplotypes per population. The input matrix must contain one row per individual. The first column must contain the population name, while the second must contain the name of the haplotype. The desired matrix can be obtained using "FindHaplo".

Two output matrices are estimated, one giving the abundance of each haplotype per population (named weighted matrix) and the other representing presence/absence of each haplotype per population by 1/0 (named interaction matrix).

Usage

```
HapPerPop (readfile=T, sep=" ", header=F, inputFile=NA, input=NA, saveFile=T,
Wname=NA, Iname=NA)
```



5

Arguments

readfile a logical; if TRUE (default) the input matrix is provided in a text file. If FALSE, the matrix is provided as an R object.

sep the character separating columns in the input matrix (space, by default).

header a logical value indicating whether the input matrix contains the names of the variables as its first line. (Default=FALSE).

inputFile (if readfile=TRUE) the name of the file containing the input matrix.

input (if readfile=FALSE) the name of the input matrix as an R object.

saveFile a logical; if TRUE (default), the two ouput matrices computed are saved as two different text files.

Wname the name given to the output weighted matrix file.

Iname the name given to the output interaction matrix file

Value

A list containing two matrices. The first matrix contains the weighted matrix, that is, the number of haplotypes (columns) found per population (rows). The second is the interaction matrix, containing information about the presence or absence of each haplotype (columns) per population (rows).

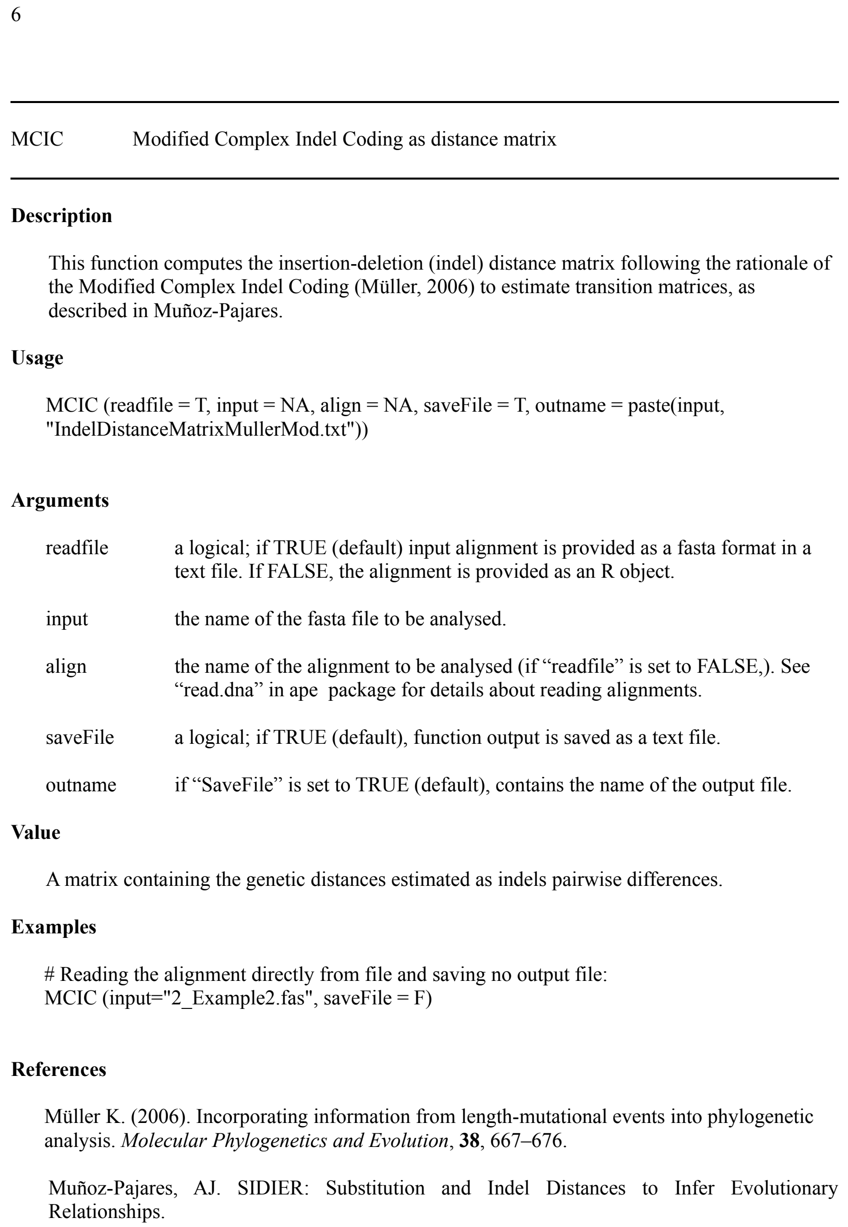
See also

FindHaplo

Examples

```
 library(ape)
# Reading the alignment from an object and saving the two computed distance matrices:
FH<-read.table("3_FindHaplo_Example2_modified.txt",header=T)
HapPerPop(readfile=F,input=FH,header=T,saveFile=F)

# Reading the alignment directly from file, displaying only the weighted matrix:
HapPerPop(readfile=T,inputFile="3_FindHaplo_Example2_modified.txt",header=T,saveFile=F
)[[1]]
```



6

MCIC Modified Complex Indel Coding as distance matrix

Description

This function computes the insertion-deletion (indel) distance matrix following the rationale of the Modified Complex Indel Coding (Müller, 2006) to estimate transition matrices, as described in Muñoz-Pajares.

Usage

```
MCIC (readfile = T, input = NA, align = NA, saveFile = T, outname = paste(input,
"IndelDistanceMatrixMullerMod.txt"))
```

Arguments

readfile	a logical; if TRUE (default) input alignment is provided as a fasta format in a text file. If FALSE, the alignment is provided as an R object.
input	the name of the fasta file to be analysed.
align	the name of the alignment to be analysed (if "readfile" is set to FALSE,). See "read.dna" in ape package for details about reading alignments.
saveFile	a logical; if TRUE (default), function output is saved as a text file.
outname	if "SaveFile" is set to TRUE (default), contains the name of the output file.

Value

A matrix containing the genetic distances estimated as indels pairwise differences.

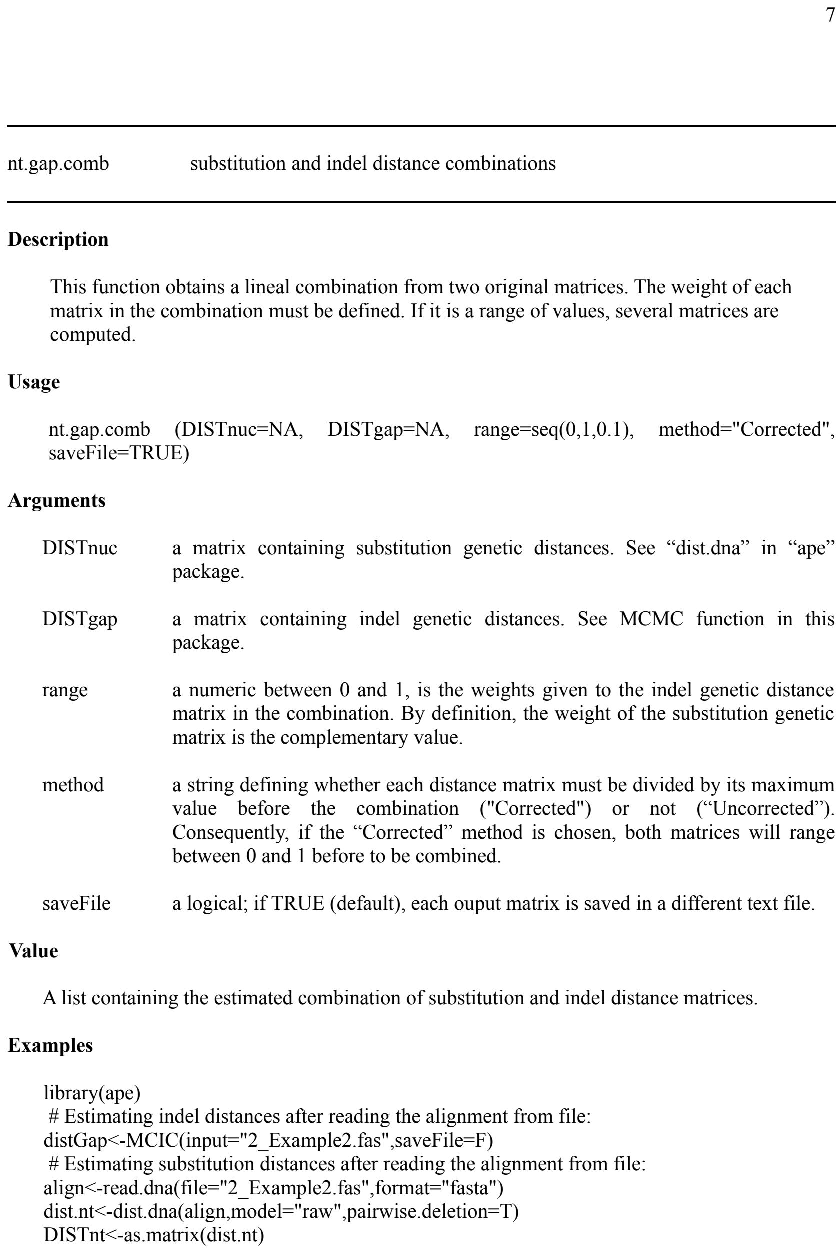
Examples

```
# Reading the alignment directly from file and saving no output file:
MCIC (input="2_Example2.fas", saveFile = F)
```

References

Müller K. (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution*, **38**, 667–676.

Muñoz-Pajares, AJ. SIDIER: Substitution and Indel Distances to Infer Evolutionary Relationships.



nt.gap.comb substitution and indel distance combinations

Description

This function obtains a lineal combination from two original matrices. The weight of each matrix in the combination must be defined. If it is a range of values, several matrices are computed.

Usage

```
nt.gap.comb (DISTnuc=NA, DISTgap=NA, range=seq(0,1,0.1), method="Corrected", saveFile=TRUE)
```

Arguments

DISTnuc a matrix containing substitution genetic distances. See “dist.dna” in “ape” package.

DISTgap a matrix containing indel genetic distances. See MCMC function in this package.

range a numeric between 0 and 1, is the weights given to the indel genetic distance matrix in the combination. By definition, the weight of the substitution genetic matrix is the complementary value.

method a string defining whether each distance matrix must be divided by its maximum value before the combination ("Corrected") or not (“Uncorrected”). Consequently, if the “Corrected” method is chosen, both matrices will range between 0 and 1 before to be combined.

saveFile a logical; if TRUE (default), each ouput matrix is saved in a different text file.

Value

A list containing the estimated combination of substitution and indel distance matrices.

Examples

```
library(ape)
 # Estimating indel distances after reading the alignment from file:
distGap<-MCIC(input="2_Example2.fas",saveFile=F)
 # Estimating substitution distances after reading the alignment from file:
align<-read.dna(file="2_Example2.fas",format="fasta")
dist.nt<-dist.dna(align,model="raw",pairwise.deletion=T)
DISTnt<-as.matrix(dist.nt)
```

```
# Obtaining 11 corrected combined matrices using a range of alpha values:
nt.gap.comb(DISTgap=distGap, range=seq(0,1,0.1), method="Corrected", saveFile=FALSE,
DISTnuc=DISTnt)
 # Obtaining the arithmetic mean of both matrices using both the corrected and the uncorrected
methods.
nt.gap.comb(DISTgap=distGap, range=0.5, method="Both", saveFile=FALSE,
DISTnuc=DISTnt)
```

See also

MCIC

---

perc.thr Percolation threshold network

Description

This function computes the percolation network following Rozenfeld et al. (2008), as described in Muñoz-Pajares.

Usage

```
perc.thr (dis, threshold = seq (0,1,0.01), ptPDF = TRUE, ptPDFname =
"PercolatedNetwork.pdf", estimPDF = TRUE, estimPDFname = "PercThr Estimation.pdf",
estimOutfile = TRUE, estimOutName = "PercThresholdEstimation.txt", appendOutfile =
TRUE, plotALL = FALSE, bgcol = "white", label.col = "black", label = colnames(dis),
modules = FALSE)
```

Arguments

dis	the distance matrix to be represented
threshold	a numeric vector between 0 and 1, is the range of thresholds (referred to the maximum distance in a matrix) to be screened (by default, 101 values from 0 to 1).
ptPDF	a logical, must the percolated network be saved as a pdf file?
ptPDFname	if ptPDF=TRUE, the name of the pdf file containing the percolation network to be saved ("PercolatedNetwork.pdf", by default)
estimPDF	a logical, must the percolation threshold estimation be saved as a pdf file? If estimPDF=TRUE (default) the value of <s> for each threshold is also saved
estimPDFname	if estimPDF=TRUE (default), defines the name of the pdf file containing the percolation threshold estimation ("PercThr Estimation.pdf" by default).

estimOutfile a logical, must the matrix containing percolation threshold estimation variables be saved as a pdf file?

estimOutName if estimOutfile=TRUE (default), contains the name of the text file containing the percolation threshold estimation ("PercThr Estimation.txt" by default).

appendOutfile a logical, if estimOutfile=TRUE, it defines whether results must be appended to an existing file with the same name (TRUE) or the existing file must be replaced (FALSE).

plotALL a logical, must all the networks calculated during the percolation threshold estimation be saved as different pdf files? (FALSE, by default). If TRUE, for each value defined in threshold, one file is generated.

bgcol string, defining the colour of the background for each node in the network. Can be equal for all nodes (if only one colour is defined), customized (if several colours are defined), or can represent different modules (see modules option).

label.col string, defining the colour of labels for each node in the network. Can be equal for all nodes (if only one colour is defined) or customized (if several colours are defined),

label string, labels for each node. By default are the column names of the distance matrix (dis). (See substr function in base package to automatically reduce name lengths).

modules a logical, must nodes belonging to different modules be represented as different colours?

Details

By default, percolation threshold is estimated with an accuracy of 0.01, but it may be increased by setting the decimal places in threshold function (e.g., seq(0,1,0.0001)). However, it may strongly increase computation times (in this example, it is required to estimate 100 001 instead of 101 networks). It is also possible to increase accuracy with a low increase in computation time by repeating the process and increasing decimal places only in a range close to a previously estimated percolation threshold.

Examples

```
library(ape)
 # Estimating indel distances after reading the alignment from file:
distGap<-MCIC(input="2_Example2.fas",saveFile=F)
 # Estimating substitution distances after reading the alignment from file:
align<-read.dna(file="2_Example2.fas",format="fasta")
dist.nt <-dist.dna(align,model="raw",pairwise.deletion=T)
DISTnt<-as.matrix(dist.nt)
```

10

```
 # Obtaining the arithmetic mean of both matrices using the corrected method:
CombinedDistance<-nt.gap.comb(DISTgap=distGap, range=0.5, method="Corrected",
saveFile=FALSE, DISTnuc=DISTnt)
 # Estimating the percolation threshold of the combined distance, modifying labels:
perc.thr(dis=as.data.frame(CombinedDistance$Corrected),label=paste(substr(row.names(as.dat
a.frame(CombinedDistance$Corrected)),11,11),substr(row.names(as.data.frame(CombinedDista
nce$Corrected)),21,21),sep="-"))
 # The same network showing different modules as different colours (randomly selected):
perc.thr(dis=as.data.frame(CombinedDistance$Corrected),label=paste(substr(row.names(as.dat
a.frame(CombinedDistance$Corrected)),11,11),substr(row.names(as.data.frame(CombinedDista
nce$Corrected)),21,21),sep="-"), module=T)
```

References

Rozenfeld AF, Arnaud-Haond S, Hernández-García E, Eguíluz VM, Serrão EA, Duarte CM. (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences,***105**, 18824 –18829.

Muñoz-Pajares, AJ. SIDIER: Substitution and Indel Distances to Infer Evolutionary Relationships.

plot.thr.network   Plot a network given a threshold or a range of thresholds.

Description

Given a distance matrix, this function plots a network taking into account only distances shorter than a defined value (threshold). Multiple networks are estimated if a list of thresholds are provided.

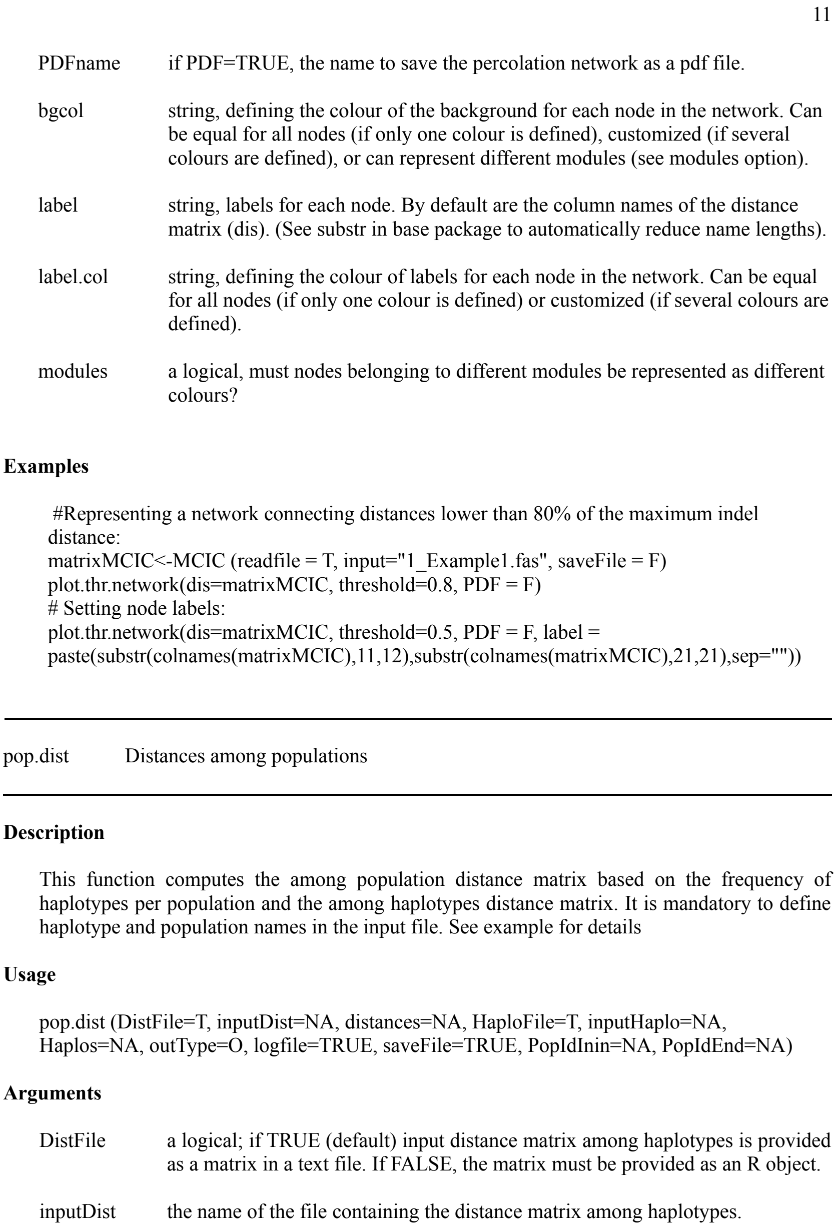
The defined threshold must range between 0 and 1 and represents the percentage of the maximum value of the distance matrix. Consequently, if the threshold value is set to 0.5, only distances lower than half the maximum will be represented.

Usage

```
plot.network(dis, threshold, bgcol = "white", label.col = "black", label = colnames(dis),
        modules = FALSE, PDF = TRUE, PDFname = paste("Network_thr = ",threshold,
        sep = ""))
```

Arguments

dis	the distance matrix to be represented
threshold	a numeric or a vector between 0 and 1, is the threshold or range of thresholds to be plotted.
PDF	a logical, must the network be saved as a pdf file?



11

PDFname — if PDF=TRUE, the name to save the percolation network as a pdf file.

bgcol — string, defining the colour of the background for each node in the network. Can be equal for all nodes (if only one colour is defined), customized (if several colours are defined), or can represent different modules (see modules option).

label — string, labels for each node. By default are the column names of the distance matrix (dis). (See substr in base package to automatically reduce name lengths).

label.col — string, defining the colour of labels for each node in the network. Can be equal for all nodes (if only one colour is defined) or customized (if several colours are defined).

modules — a logical, must nodes belonging to different modules be represented as different colours?

Examples

```
 #Representing a network connecting distances lower than 80% of the maximum indel
distance:
matrixMCIC<-MCIC (readfile = T, input="1_Example1.fas", saveFile = F)
plot.thr.network(dis=matrixMCIC, threshold=0.8, PDF = F)
# Setting node labels:
plot.thr.network(dis=matrixMCIC, threshold=0.5, PDF = F, label =
paste(substr(colnames(matrixMCIC),11,12),substr(colnames(matrixMCIC),21,21),sep=""))
```

pop.dist   Distances among populations

Description

This function computes the among population distance matrix based on the frequency of haplotypes per population and the among haplotypes distance matrix. It is mandatory to define haplotype and population names in the input file. See example for details

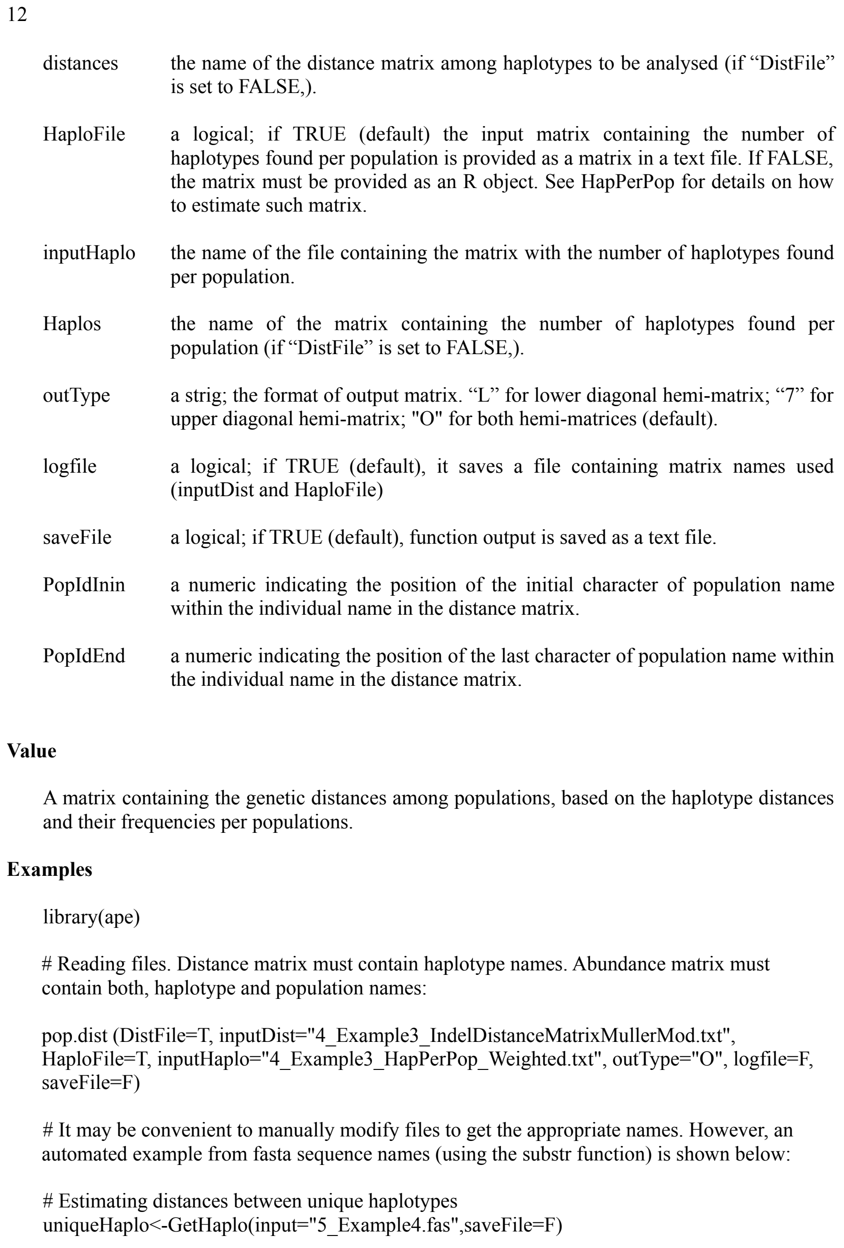
Usage

```
pop.dist (DistFile=T, inputDist=NA, distances=NA, HaploFile=T, inputHaplo=NA,
Haplos=NA, outType=O, logfile=TRUE, saveFile=TRUE, PopIdInin=NA, PopIdEnd=NA)
```

Arguments

DistFile — a logical; if TRUE (default) input distance matrix among haplotypes is provided as a matrix in a text file. If FALSE, the matrix must be provided as an R object.

inputDist — the name of the file containing the distance matrix among haplotypes.



12

distances — the name of the distance matrix among haplotypes to be analysed (if “DistFile” is set to FALSE,).

HaploFile — a logical; if TRUE (default) the input matrix containing the number of haplotypes found per population is provided as a matrix in a text file. If FALSE, the matrix must be provided as an R object. See HapPerPop for details on how to estimate such matrix.

inputHaplo — the name of the file containing the matrix with the number of haplotypes found per population.

Haplos — the name of the matrix containing the number of haplotypes found per population (if “DistFile” is set to FALSE,).

outType — a strig; the format of output matrix. “L” for lower diagonal hemi-matrix; “7” for upper diagonal hemi-matrix; "O" for both hemi-matrices (default).

logfile — a logical; if TRUE (default), it saves a file containing matrix names used (inputDist and HaploFile)

saveFile — a logical; if TRUE (default), function output is saved as a text file.

PopIdInin — a numeric indicating the position of the initial character of population name within the individual name in the distance matrix.

PopIdEnd — a numeric indicating the position of the last character of population name within the individual name in the distance matrix.

Value

A matrix containing the genetic distances among populations, based on the haplotype distances and their frequencies per populations.

Examples

```
library(ape)

# Reading files. Distance matrix must contain haplotype names. Abundance matrix must
contain both, haplotype and population names:

pop.dist (DistFile=T, inputDist="4_Example3_IndelDistanceMatrixMullerMod.txt",
HaploFile=T, inputHaplo="4_Example3_HapPerPop_Weighted.txt", outType="O", logfile=F,
saveFile=F)

# It may be convenient to manually modify files to get the appropriate names. However, an
automated example from fasta sequence names (using the substr function) is shown below:

# Estimating distances between unique haplotypes
uniqueHaplo<-GetHaplo(input="5_Example4.fas",saveFile=F)
```

```
distGap<-MCIC(readfile=F,align=uniqueHaplo,saveFile=F)
dist.nt <-dist.dna(uniqueHaplo,model="raw",pairwise.deletion=T)
DISTnt<-as.matrix(dist.nt)

# Replacing sequence names by haplotype names in both distance matrices
for (Hi in 1:length(colnames(distGap)))
colnames(distGap)[Hi]<-FindHaplo(input="5_Example4.fas",saveFile=F)
[which(colnames(distGap)[Hi]==FindHaplo(input="5_Example4.fas",saveFile=F)[,1]),2]
row.names(distGap)<-colnames(distGap)
for (Hi in 1:length(colnames(DISTnt)))
colnames(DISTnt)[Hi]<-FindHaplo(input="5_Example4.fas",saveFile=F)
[which(colnames(DISTnt)[Hi]==FindHaplo(input="5_Example4.fas",saveFile=F)[,1]),2]
row.names(DISTnt)<-colnames(DISTnt)

#Combining distance matrices and setting haplotype names
CombinedDistance<-as.data.frame(nt.gap.comb(DISTgap=distGap, range=0.5,
method="Corrected", saveFile=FALSE, DISTnuc=DISTnt)[[2]])
colnames(CombinedDistance)<-row.names(CombinedDistance)

# Estimating haplotype abundance per population and setting population names:
Haplotypes<-FindHaplo(input="5_Example4.fas",saveFile=F)
Haplotypes[,1]<-substr(Haplotypes[,1],1,11)
Weighted<-as.data.frame(HapPerPop(readfile=F,header=T,input=Haplotypes)[1])
colnames(Weighted)<-substr(colnames(Weighted),10,11)

# Estimating population distances
pop.dist (DistFile=F, distances=CombinedDistance, HaploFile=F, Haplos=Weighted,
outType="O", logfile=F, saveFile=F)
```

Index

4

COMBINING INDELS AND SUBSTITUTIONS INFORMATION FOR THE RECONSTRUCTION OF EVOLUTIONARY HAPLOTYPE RELATIONSHIPS

ABSTRACT

The use of indel-rich regions in phylogenetic and phylogeographic inference is hindered by the difficulty in obtaining a correct sequence alignment, as well as by the problem of interpreting the evolutionary information contained in indels. In this chapter, we propose a procedure to accurately extract phylogenetic information from indel-rich regions in three multi-step stages. First, we align indel-rich sequences according to the main mutational processes affecting the studied DNA region to infer indel and nucleotide homology. Second, we obtain both substitution and indel distance matrices, to subsequently combine them in a single distance matrix. Finally, using the information from this inclusive matrix, we build a haplotype network using percolation theory. We have applied this methodology to a plastid indel-rich region widely used in plant phylogeny: the trnL-trnF intergenic spacer (IGS). Using this marker, we have studied 317 individuals from 66 populations belonging to the six species conforming the *Erysimum nevadense* complex. The yielded percolation network will allow us to infer the evolutionary history of the *E. mediohispanicum* in the context of closely related species.[1]

[1] A. Jesús Muñoz-Pajares, Mohamed Abdelaziz, M. Belén Herrador, José M. Gómez, and Francisco Perfectti

INTRODUCTION

Insertion and deletion events (indels) may provide valuable information for phylogenetic and phylogeographic inferences (Simmons et al., 2001; Vogt, 2002; Young and Healy, 2003; Müller, 2006; Blair and Murphy, 2011). The advantage of using indels may be even greater when sequences show low substitution rates (Redelings and Suchard, 2007). However, obtaining this information is a complex, time-consuming task, for two main reasons: the difficulty in confidently aligning these indel-rich sequences and the lack of a standard methodology to deal with the evolutionary information contained in gaps (Golenberg et al., 1993; Castresana, 2000; Loytynoja and Milinkovitch, 2001). Although some procedures exist to simultaneously estimate alignments and phylogenies (e.g. Wheeler, 1996; Wheeler et al., 2003; Lunter et al., 2005; Redelings and Suchard, 2005; Liu et al., 2012), in most studies they are performed separately.

A correct alignment is crucial to infer the evolutionary relationships between species or populations (Kumar and Filipski, 2007). Indel events produce sequences with different lengths, and the introduction of gaps into the alignment is required to establish nucleotide homology. The achievement of an alignment from sequences showing a high proportion of indels (especially when duplication exits) is hindered by the uncertainty associated with the presence of multiple redundant and/or non-homologous positions. For this reason, although a great deal of software has been developed to align sequences, all of them show poor accuracy in the presence of high indel rates (Liu et al., 2010). These difficulties have traditionally led to the proposal to remove gap positions before conducting phylogenetic analyses (Talavera and Castresana, 2007). It has been suggested that, in order to avoid losing the information from indels, the alignment should be manually refined after the use of any software (Kjer et al., 2007; Morrison, 2009; Blair and Murphy, 2011). Manual alignment may be further improved when secondary structures and molecular evolutionary processes are considered (Kjer, 1995; Kelchner, 2000; Müller, 2006).

There are no standard procedures for inferring evolutionary relationships from gapped positions. This is because several adjacent positions could be involved in a single evolutionary event (indels involving more than one nucleotide), and because different indel events could affect the same nucleotide position in different taxa (producing overlapping gap sites). However, an increasing number of studies are including the phylogenetic information provided by indels, most of them within a maximum parsimony framework. The fifth state and indel coding methods are the most common approaches. The former method considers each gap position as a new state in addition to the four nucleotides. The latter method treats indels as missing data in the original sequence but their information is included by codifying them as additional characters. There are several different coding methods, depending on the treatment of the overlapping gap positions (e.g., Baum et al., 1994; Freudenstein and Chase, 2001) and the calculation of the pairwise transformation cost matrix. The Simple Indel Coding (SIC, Simmons and Ochoterena, 2000), the Complex Indel Coding (CIC, Simmons and Ochoterena, 2000), and the Modified Complex Indel Coding (MCIC, Müller, 2006) are the most widely used of these coding methods. Other coding methods codify gaps in their original positions, as in the case of multi-state character coding (Lutzoni et al., 2000), the elision method (Wheeler et al., 1995), case-sensitive coding (Swofford, 1989), stretch coding (Geiger, 2002), and block coding (Geiger, 2002). Several evolutionary models in probabilistic inference methods (i.e., maximum likelihood and Bayesian inference) have accounted for both substitution and indel events (e.g., Rivas and Eddy, 2008, and references therein). However, such models are complex and in most cases gaps are either removed from the alignment or handled as missing data when either maximum likelihood or Bayesian inference are used (Ogden and Rosenberg, 2007).

In contrast with the diverse methodology mentioned above, dealing with indels in distance-based algorithms needs to be explored (Ogden and Rosenberg, 2007). Distance methods discard the information provided by a given homologous position if that position is a gap in at least one of the aligned sequences (complete deletion) or in at least one of the two compared sequences (pairwise deletion). In both instances the information provided by indels is lost. In this study, we propose a method to include indel evolutionary information within a distance-based framework. Although the phylogenetic accuracy of distance methods is assumed to be lower than that of other methods (Talavera and Castresana, 2007; Dwivedi and Gadagkar, 2009), they are very powerful for large evolutionary trees and short-branch trees (regardless of tree depth) (Roch, 2010). In fact, these methods are widely used in phylogeography to infer neighbor joining (NJ; Saitou and Nei, 1987) trees and minimum spanning networks (MSN; Excoffier et al., 1992).

One of the most commonly used markers in plant phylogenetic studies (Stuessy, 2009) was analyzed using this method: the plastidial DNA region, encompassing the transfer RNA genes for leucine (trnL$_{\mathrm{UAA}}$) and phenylalanine (trnF$_{\mathrm{GAA}}$), hereafter trnL and trnF, respectively. The trnL gene is composed of two exons separated by one group I intron, while an intergenic spacer (IGS) separates the trnL 3′ exon and the trnF (Fig.4.1) The IGS contains an indel-rich region in several plant genera due to the existence of multiple tandemly repeated trnF pseudogene copies (Fig.4.1). These pseudogenes were originated by duplication and are subjected to divergence due to the accumulation of independent substitutions. The functional trnF is composed of five domains: the D-domain, the anticodon-domain, the T-domain and two flanking acceptor stems (Koch et al., 2005). Consequently the indel-rich IGS region is composed of different mo-

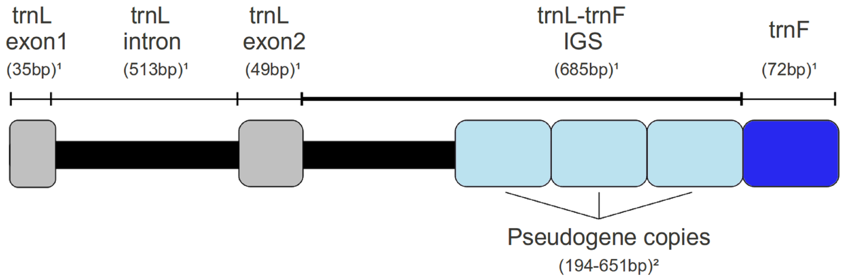


Figura 4.1: Schematic representation of the trnL-trnF cpDNA region. Both transfer genes are separated by an intergene spacer (IGS) showing a variable number of trnF pseudogene copies. (1) Sizes according to the *A. thaliana* complete genome (accession number NC_000932.1). (2) Size range found in *E. mediohispanicum*.

tifs showing similarity with each of these functional trnF domains, usually conserving the same arrangement (acceptor - D-domain - anticodon - T-domain - acceptor). The IGS region has been described in depth in the Brassicaceae (Koch et al., 2007), especially in *Arabidopsis* and *Boechera*, where the presence of a variable number of pseudogene copies produces high variability in sequence length (Dobeš et al., 2004, 2007). To determine the number of pseudogene copies per individual, most of these studies count the number of motifs showing similarity with the anticodon region (Koch et al., 2007; Schmickl et al., 2008). An increasing number of studies are trying to extract evolutionary information from this region, including indel variation in different ways (Koch et al., 2005; Koch and Matschinger, 2007; Koch et al., 2007; Schmickl et al., 2008). The indel-rich trnL-trnF IGS region has been used to elucidate both ancient (e.g., basal angiosperms, Borsch et al., 2003) and recent (e.g., among populations of the same species, Dobeš et al., 2004; Tedder et al., 2010) divergence.

The specific goals of the present study are: 1) To describe a method to align the indel-rich trnF IGS pseudogene region; and 2) To propose a distance-based method to investigate haplotype relationships combining information from both substitutions and indels. We illustrate this methodology analyzing the trnL-trnF IGS pseudogene variation in 56 populations of *Erysimum mediohispanicum* (Brassicaceae) and the five closely related *Erysimum* species that conform the *E. nevadense* species complex.

MATERIAL AND METHODS

Plant species

Erysimum comprises more than 200 species recently grouped in the unigeneric tribe *Erysimeae* (Couvreur et al., 2010). *Erysimum* is one of the few crucifer polybasic genera (i.e. characterized by multiple base chromosome numbers; Warwick and Al-Shehbaz, 2006). The evolution of the genera is complex, with events of inter-specific hybridization, polyploidization, incomplete sorting, and reticulate evolution (Clot, 1991; Ančev, 2006; Marhold and Lihová, 2006), frequently resulting in species complexes and cryptic species (Ančev, 2006; Turner, 2006).

E. mediohispanicum is a monocarpic herb endemic to the Iberian Peninsula, distributed into two regions, one in the north-east and the other in the south-east of the Peninsula (Nieto-Feliner, 1993). More than twenty *Erysimum* species inhabit the Iberian Peninsula (Nieto-Feliner, 1993), of which five species (*E. nevadense*, *E. merxmuelleri*, *E. rondae*, *E. ruscinonense*, and *E. gomezcampoi*) are closely related to *E. mediohispanicum*, conforming the *E. nevadense* species complex (Nieto-Feliner, 1993).

Molecular methods

We collected fresh leaf tissue from five individuals belonging to 56 populations of *E. mediohispanicum* covering the complete species distribution area (Table S4). We also collected samples from five individuals belonging to two populations of the remaining five species from the *E. nevadense* complex (Table S4). Leaves were dried up and stored in silica gel until DNA extractions were performed using the GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich). We amplified the trnL-trnF spacer using the TabC and TabF primers (Taberlet et al., 1991) in 25 µL PCR reactions containing 0.2 µM each primer (MWG), 0.1 µM each dNTP (Sigma-Aldrich), 0.02 U/µL Taq polymerase (New England BioLabs), 1X Taq buffer (New England BioLabs), and 5 ng of template genomic DNA. PCR mix reactions were conducted in a Gradient Master Cycler Pro S (Eppendorf) with an initial denaturation step of 94 ºC for 180 s followed by 35 cycles of 94 ºC for 15 s, 58 ºC for 30 s, and 72 ºC for 90 s, and a single final extension step at 72 ºC for 180 s. PCR products were visualized in a 1.5 % agarose gel, showing sizes ranging from ~900 to ~1500 bp, and precipitated using sodium acetate 3 M pH= 4.6 and absolute ethanol at 4 ºC. The resulting pellets were re-suspended in water and sent to MACROGEN Inc. for sequencing. Chromatograms were checked using FinchTV v.1.3 (http: //www.geospiza.com/Products/finchtv.shtml) and sequences were edited with Bioedit v7.0.9.0 (Hall, 1999).

As we focused our analysis on the repetitive IGS region, we only conserved 37 bp of the non-repetitive sequence at 5′. As few sequences showed unclear chromatograms for the last repetitive motif, we decided to remove this region from the complete set of studied sequences. Afterwards, identical sequences were defined as the same haplotype.

Haplotype alignment

The complicated trnL-trnF IGS evolutionary dynamic has led to the existence of homologous and paralogous pseudogene regions, making it highly difficult to obtain a confident alignment using conventional softwares. For this reason, we manually aligned haplotypes according to the following steps (Fig. 4.2):

1) Motifs conforming haplotypes were identified and classified into families according to their similarity with functional trnF motifs (Fig. 4.2B). To find all the motif variants per family, we searched for motifs allowing local dissimilarities using FastPCR v4.0 (http: //primerdigital.com/fastpcr/version-history). Motifs were named using one letter (family) and one number (motif variant) (Fig. 4.2B).

2) We codified haplotypes substituting motif sequences for their names (Fig. 4.2C).

3) Haplotypes were classified into different groups according to their similarity in motif arrangements (Fig. 4.2D).

4) For each haplotype group, we performed independent alignments (Fig. 4.2E) following Kelchner (2000) and Borsch et al. (2003), and using the longest haplotype in a group as a reference. At least we have contrary evidence, we have assumed that indels occurs in the trnL-trnF IGS region involving complete pseudogenes and we consequently introduced gaps to try to maintain pseudogene sequence integrity. To avoid an alignment compatible with multiple origins of the same motif, identical motifs were considered homologous (i.e., they were aligned in the same column, if possible). We assumed motif divergence (i.e., different motif sequences in the same column) when flanking motifs were identical.

5) To facilitate the alignment procedure, we obtained one consensus sequence per group (Fig. 4.2F). For each homologous position (column), this consensus sequence is composed of the most abundant motif of the column (Fig. 4.2F).

6) We aligned consensus sequences following the same rules described in step 4 (Fig. 4.2G).

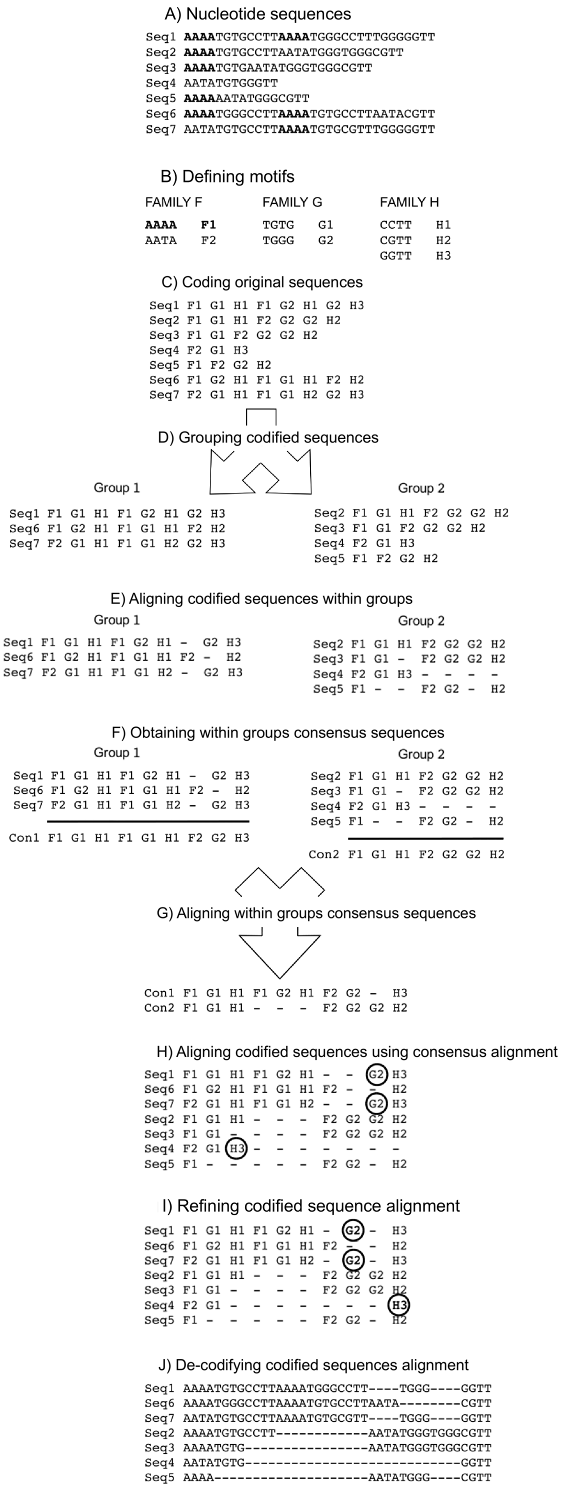
7) We used the consensus sequence alignment as a guide to perform the overall alignment with the entire set of sequences (Fig. 4.2H).

8) We refined the yielded alignment trying to maintain complete pseudogenes and similar motifs in homologous positions (Fig. 4.2I), following rules described in step 4. For example, in Fig. 4.2I the position of H3 in Seq4 is modified to be homologous with all the remaining H3, and the positions of G2 in Seq1 and Seq7 are modified to maintain all the non-duplicated G2 motifs in the same column.

9) Finally, to produce the final alignment, we re-codified haplotypes by changing motif names for their real sequences (Fig. 4.2J).

Figura 4.2: Alignment workflow exemplified by using hypothetical sequences. (A) Sequences showing different lengths due to the presence of different concatenated motifs. (B) These motifs where classified into families according to their sequence similarity. (C) Sequences were recodified using motif names and (D) classified into groups according to their similar motif structure. (E) Sequences belonging to different groups were separately aligned. (F) The most common motif per site was used to yield a consensus sequence per group. (G) We used group consensuses to perform a guide alignment (H) to align all the studied sequences. (I) This overall alignment was refined to place low frequency motifs as homologous (see H3 motif) and considering all single motifs as homologous in the same position within a duplicated region (see G2 motif). (J) Finally, we obtained the final alignment by decoding motif names of the nucleotide sequences. This alignment was used to obtain both substitution and indel distance matrices (see Fig. 4.3).

(Next page) ►



A) Nucleotide sequences

```
Seq1 AAAATGTGCCTTAAAATGGGCCTTTGGGGGTT
Seq2 AAAATGTGCCTTAATATGGGTGGGCGTT
Seq3 AAAATGTGAATATGGGTGGGCGTT
Seq4 AATATGTGGGTT
Seq5 AAAAAATATGGGCGTT
Seq6 AAAATGGGCCTTAAAATGTGCCTTAATACGTT
Seq7 AATATGTGCCTTAAAATGTGCGTTTGGGGGTT
```

B) Defining motifs

FAMILY F		FAMILY G		FAMILY H	
AAAA	**F1**	TGTG	G1	CCTT	H1
AATA	F2	TGGG	G2	CGTT	H2
				GGTT	H3

C) Coding original sequences

```
Seq1 F1 G1 H1 F1 G2 H1 G2 H3
Seq2 F1 G1 H1 F2 G2 G2 H2
Seq3 F1 G1 F2 G2 G2 H2
Seq4 F2 G1 H3
Seq5 F1 F2 G2 H2
Seq6 F1 G2 H1 F1 G1 H1 F2 H2
Seq7 F2 G1 H1 F1 G1 H2 G2 H3
```

D) Grouping codified sequences

Group 1

```
Seq1 F1 G1 H1 F1 G2 H1 G2 H3
Seq6 F1 G2 H1 F1 G1 H1 F2 H2
Seq7 F2 G1 H1 F1 G1 H2 G2 H3
```

Group 2

```
Seq2 F1 G1 H1 F2 G2 G2 H2
Seq3 F1 G1 F2 G2 G2 H2
Seq4 F2 G1 H3
Seq5 F1 F2 G2 H2
```

E) Aligning codified sequences within groups

Group 1

```
Seq1 F1 G1 H1 F1 G2 H1 -  G2 H3
Seq6 F1 G2 H1 F1 G1 H1 F2 -  H2
Seq7 F2 G1 H1 F1 G1 H2 -  G2 H3
```

Group 2

```
Seq2 F1 G1 H1 F2 G2 G2 H2
Seq3 F1 G1 -  F2 G2 G2 H2
Seq4 F2 G1 H3 -  -  -  -
Seq5 F1 -  -  F2 G2 -  H2
```

F) Obtaining within groups consensus sequences

Group 1

```
Seq1 F1 G1 H1 F1 G2 H1 -  G2 H3
Seq6 F1 G2 H1 F1 G1 H1 F2 -  H2
Seq7 F2 G1 H1 F1 G1 H2 -  G2 H3
_______________________________
Con1 F1 G1 H1 F1 G1 H1 F2 G2 H3
```

Group 2

```
Seq2 F1 G1 H1 F2 G2 G2 H2
Seq3 F1 G1 -  F2 G2 G2 H2
Seq4 F2 G1 H3 -  -  -  -
Seq5 F1 -  -  F2 G2 -  H2
_________________________
Con2 F1 G1 H1 F2 G2 G2 H2
```

G) Aligning within groups consensus sequences

```
Con1 F1 G1 H1 F1 G2 H1 F2 G2 -  H3
Con2 F1 G1 H1 -  -  -  F2 G2 G2 H2
```

H) Aligning codified sequences using consensus alignment

```
Seq1 F1 G1 H1 F1 G2 H1 -  -  G2 H3
Seq6 F1 G2 H1 F1 G1 H1 F2 -  -  H2
Seq7 F2 G1 H1 F1 G1 H2 -  -  G2 H3
Seq2 F1 G1 H1 -  -  -  F2 G2 G2 H2
Seq3 F1 G1 -  -  -  -  F2 G2 G2 H2
Seq4 F2 G1 H3 -  -  -  -  -  -  -
Seq5 F1 -  -  -  -  -  F2 G2 -  H2
```

I) Refining codified sequence alignment

```
Seq1 F1 G1 H1 F1 G2 H1 -  G2 -  H3
Seq6 F1 G2 H1 F1 G1 H1 F2 -  -  H2
Seq7 F2 G1 H1 F1 G1 H2 -  G2 -  H3
Seq2 F1 G1 H1 -  -  -  F2 G2 G2 H2
Seq3 F1 G1 -  -  -  -  F2 G2 G2 H2
Seq4 F2 G1 -  -  -  -  -  -  -  H3
Seq5 F1 -  -  -  -  -  F2 G2 -  H2
```

J) De-codifying codified sequences alignment

```
Seq1 AAAATGTGCCTTAAAATGGGCCTT----TGGG----GGTT
Seq6 AAAATGGGCCTTAAAATGTGCCTTAATA--------CGTT
Seq7 AATATGTGCCTTAAAATGTGCGTT----TGGG----GGTT
Seq2 AAAATGTGCCTT------------AATATGGGTGGGCGTT
Seq3 AAAATGTG----------------AATATGGGTGGGCGTT
Seq4 AATATGTG----------------------------GGTT
Seq5 AAAA--------------------AATATGGG----CGTT
```

Haplotype relationships

To study haplotype relationships, we combined information provided by substitutions and indels within a distance-based method framework (Figs. 4.2J and 4.3). Using the aligned sequences (Fig.4.3A), we estimated the pairwise substitution distance matrix as p-distance (proportion of sites that differ between each pair of sequences) using *ape* package in R (Paradis et al., 2004), and treating gaps as pairwise deletion (Fig.4.3B). We estimated p-distances rather than more complex substitution models because it has been demonstrated that distance methods of phylogenetic inference perform better when simpler distance measures are used (Takahashi and Nei, 2000; Rosenberg and Kumar, 2001, 2003). Additionally, the divergence found among trnL-trnF IGS sequences (mean number of substitutions per site: 0.016) was lower than the recommended to apply any correction (0.05 according to Nei and Kumar, 1998). Saturation plots showed a close lineal relationship between corrected and uncorrected substi-

Figura 4.3: Workflow from alignment to network illustrated with hypothetical sequences. (A) The aligned sequences (see Fig. 4.2 for methods) are used to compute (B) the substitution distance matrix (as the proportion of polymorphic sites, considering gaps as missing data) (C) the indel distance matrix (according to method described by Müller, 2006) and. (D-E) Both distance matrices were transformed (dividing by the maximum distance value of each matrix) to enable a comparison of processes with different mutation rates. We tested for genetic, phylogenetic, and topological congruence by performing different CADM tests (although only genetic congruence tests are represented in this figure). To study the effect of this transformation, we compared the congruence values obtained by using both transformed and untransformed distance matrices. (F) If significant congruence was found, we obtained the combined genetic distance matrix as a lineal combination of both previous distances. (G) We estimated the percolation threshold for this distance matrix (black arrow) and (H) built a network, linking haplotypes showing the distance values below this threshold.

(Next page) ▶

A) Fasta sequences

```
>Seq1
AAAATGTGCCTTAAAATGGGCCTT----TGGG----GGTT
>Seq6
AAAATGGGCCTTAAAATGTGCCTTAATA--------CGTT
>Seq7
AATATGTGCCTTAAAATGTGCGTT----TGGG----GGTT
>Seq2
AAAATGTGCCTT------------AATATGGGTGGGCGTT
>Seq3
AAAATGTG----------------AATATGGGTGGGCGTT
>Seq4
AATATGTG----------------------------GGTT
>Seq5
AAAA--------------------AATATGGG----CGTT
```

B) Substitution distance matrix.
(Proportion of variable sites. Gaps as missing data)

	Seq1	Seq6	Seq7	Seq2	Seq3	Seq4	Seq5
Seq1							
Seq6	0.11						
Seq7	0.09	0.14					
Seq2	0.05	0.05	0.10				
Seq3	0.06	0.06	0.13	0.00			
Seq4	0.08	0.25	0.00	0.17	0.17		
Seq5	0.08	0.00	0.17	0.00	0.00	0.25	

CADM
tests
(see Campbell
et al. 2011)

C) Indel distance matrix (see Müller 2006)

	Seq1	Seq6	Seq7	Seq2	Seq3	Seq4	Seq5
Seq1							
Seq6	2						
Seq7	0	2					
Seq2	3	2	3				
Seq3	3	2	3	1			
Seq4	1	1	1	1	1		
Seq5	2	2	2	2	2	2	

D) Corrected substitution distance matrix.
(Dividing by the maximum distance: 0.25)

	Seq1	Seq6	Seq7	Seq2	Seq3	Seq4	Seq5
Seq1							
Seq6	0.43						
Seq7	0.38	0.57					
Seq2	0.20	0.20	0.40				
Seq3	0.25	0.25	0.50	0.00			
Seq4	0.33	1.00	0.00	0.67	0.67		
Seq5	0.33	0.00	0.67	0.00	0.00	1.00	

CADM
tests
(see Campbell
et al. 2011)

E) Corrected indel distance matrix.
(Dividing by the maximum distance: 3)

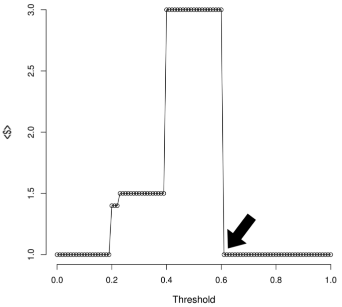
	Seq1	Seq6	Seq7	Seq2	Seq3	Seq4	Seq5
Seq1							
Seq6	0.67						
Seq7	0.00	0.67					
Seq2	1.00	0.67	1.00				
Seq3	1.00	0.67	1.00	0.33			
Seq4	0.33	0.33	0.33	0.33	0.33		
Seq5	0.67	0.67	0.67	0.67	0.67	0.67	

F) Combined corrected genetic distance matrix.
(Lineal combination of matrices D and E. We gave equal weights to both matrices)
(For other weights see Appendix1)

	Seq1	Seq6	Seq7	Seq2	Seq3	Seq4	Seq5
Seq1							
Seq6	0.55						
Seq7	0.19	0.62					
Seq2	0.60	0.43	0.70				
Seq3	0.63	0.46	0.75	0.17			
Seq4	0.33	0.67	0.17	0.50	0.50		
Seq5	0.50	0.33	0.67	0.33	0.33	0.83	

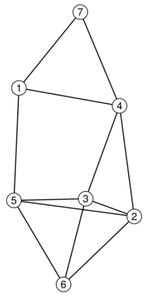
G) Percolation threshold
(see Rozenfeld et al., 2008)

$$\langle S\rangle^* = \frac{1}{N}\sum_{s<S_{max}} s^2 n_s$$



H) Network
(see Rozenfeld et al., 2008)

threshold = 0.6



tution distances (Fig. 4.S1), confirming the low divergence between sequences.

We used the number of indel events as pairwise indel distance matrix, which was estimated as the transition matrix calculated following the MCIC method (Müller, 2006) as implemented in sidier R package (Chapter 3). We used this distance, rather than a p-distance, because indel events involve different numbers of sites and affect variable sequence lengths, making it difficult to estimate the potential number of indel sites (i.e., the denominator of the p-distance).

Despite the low divergence, different evolutionary rates could be expected for substitutions and indels. To reduce the effect of such differences, we transformed each matrix by dividing it by its maximum value (Figs. 4.4D and 4.4E). To establish whether indel and substitution distance matrices provided consistent information about haplotype evolutionary history, we performed Congruence Among Distance Matrices tests (CADM, Legendre and Lapointe, 2004), as implemented in *ape* (Paradis et al., 2004). The CADM test is an extension of the Mantel test of matrix correspondence which null hypothesis assumes complete incongruence of the analyzed distance matrices. As well as a p-value (estimated using 10,000 permutations), the CADM test provide the W statistic as an estimate of the degree of congruence among the studied matrices, ranging between 0 (absence of congruence) and 1 (complete congruence) (Campbell et al., 2011). We performed the three different incongruence tests described in Campbell et al. (2011) to test for genetic, phylogenetic, and topological congruence. For the first CADM test, we compared the two distance matrices described above (substitution and indel matrices). For the second CADM test, we firstly calculated two NJ trees: one using the substitution distance matrix and the other using the indel distance matrices. Then, we tested the congruence between patristic distances estimated using both trees. For the last CADM test, all NJ tree branch lengths were set to one before patristic distance calculations. All patristic distances were estimated between all nodes (internal and terminal), using *ape* (Paradis et al., 2004). In addition, to test the effect of matrix transformation (through division by the maximum value in the matrix), we performed the three described CADM tests using both transformed and untransformed distance matrices.

We combined both (substitution and indel) corrected distance matrices, giving the same weight to both matrices to obtain the final genetic distance matrix (i.e., estimating each element in the final matrix as the arithmetic mean of the same elements in the other two matrices, Fig. 4.4F; see Fig. 4.S3 for results obtained using different weights of both matrices). To visualize haplotype relationships based on distance matrices, we have applied the network analysis method described by Rozenfeld et al. (2008) as implemented in sidier package (Chapter 3). This procedure estimates the maximum genetic distance to be represented as a link between nodes (named percolation threshold; black arrow in Fig. 4.4G. See Rozenfeld et al. (2008) and Chapter 3 for details). We estimated the percolation threshold with an accuracy of 10^{-6} to build the network (Fig. 4.4H), which allocates nodes spatially according to their genetic distances (Chapter 3). According to Rozenfeld et al. (2008), the resulting links show the relevant genetic relationships.

RESULTS

Haplotype characterization and alignment

We successfully sequenced the trnL-trnF IGS region of 317 individuals belonging to the 66 *Erysimum* sp. populations described in Table 4. Only four different haplotypes were found according to the short non-repetitive region studied (37 bp; NR1-NR4, first column in Fig. 4.5), but 69 different haplotypes were found when the complete IGS pseudogenic region was considered (Table 4). Because did not find any strong differences in haplotype arrangement between

Erysimum species. Thus, we analyzed the haplotypes of all species together. Haplotypes were named according to their frequencies, being the most abundant haplotype (H01) found in 48 individuals (15.1 % of the studied individuals, Table 4). Only the first eight haplotypes were found in more than 3 % of the total individuals and 24 haplotypes (H46 to H69) appeared in only one individual (Table 4). Sequence lengths ranged from 194 bp (H39) to 651 bp (H46 and H40) and no significant relationship was found between haplotype length and frequency (r^2=0.053, p=0.056, df=67).

The full length of the indel-rich trnL-trnF IGS region can be described as the concatenation of several motifs related to the functional trnF. Although each haplotype showed a different number and type of motifs, no other kind of sequences were found in this region (Table 4). According to their similarity with the functional trnF domains, we classified motifs conforming haplotypes into three families (Fig. 4.4): the T-domain-like family (T), the anticodon-domain-like family (A), and X family (X). T motifs showed similarity with the T-domain and the flanking acceptor stem. We found 12 different T motifs, 9 of them (T21-T29) lacking part of the T-domain, while three motifs (T01, T03, and T11) showed the complete length (Fig. 4.4). Eleven different motifs were classified as A motifs (A01-A11) due to their similarity with the anticodon domain. Finally, the 17 different X motifs showed the same length as the region comprising the D-domain and the flanking acceptor but, due to low similarity (Fig. 4.4), homology was unsure.

The shortest haplotype found (H39) showed seven motifs conforming two trnF pseudogenes according to the number of A motifs (Table 4). The longest haplotypes (H40 and H46) comprised a total of 26 motifs corresponding to eight trnF pseudogene copies, according to the number of A motifs (Table 4). Most of the 69 haplotypes showed between 17 and 23 motifs and the number of motifs among haplotypes differed by three (or multiples of three, Fig. 4), as expected when indel events involve complete trnF pseudogenes.

Motif	Acceptor stem	D domain	Anticodon domain	T domain	Acceptor stem
	gccgggata	gctcagttggtagagca	gaggact**gaa**aatcctcgtgtc	accagttcaaatctggt----	tcctggca
A01			GAGGACTGAAAATCCTTGTGTC		
A02			GAGGACTTTAAATCCTTGTGTC		
A03			GGGGACTTTAAATCCTTGTGTC		
A04			AAGGACTGAAAATCCTTGTGTC		
A05			GGGGACTGAAAATCCTTGTGTC		
A06			GAGGACTTAAAATCCTTGTGTC		
A07			GGGGACTTTAAATCCCTGTGTC		
A08			GGTGACTTTAAATCCTCGTGTC		
A09			GGGGACTTTAAATCCTCGTGTC		
A10			GGGGACTTTAAATCCTC-----		
A11			GGGGGCTTTAAATCCTC-----		
T01				ACCATTAGGAAAACGCGGATG	ATACTTCA
T03				ACCATTAGGAAAACGAGGATG	ATACTTCA
T11				ACCATTAGGAAAAGCAAGATG	ATCCTTCG
T21				----------ACATGATAATG	ATCCTTCG
T22				------------ATGATACTG	ATACTTCA
T23				----------ACATGATAATG	ATCCCTCG
T24				----------ACATGATAATG	AT------
T25				----------ACATGATACTG	ATCCTTCG
T26				------------CTGATACTG	ATACTTCA
T27				------------CTGATAATG	ATACTTCA
T28				--------------GTAGATG	ATCCTTCG
T29				--------------GTAGATG	ATACCTCA
X01	GTAATGGTC	GACATAGCTTAATTGCG			
X02	GTAATGGTC	GACATAGCTTAACTGCG			
X03	GTAATGGTA	GACATAGCTTAATTGCG			
X05	GTAATGGTG	GACATAGCTTAATTGCG			
X06	GTAATGGTC	GGCATAGCTTAATTGCG			
X08	GTAATGGTG	GACATCGCTTAATTGCG			
X09	GCAATGGTC	GGCATCGCTTTTTTGCG			
X10	TCAACGGTC	GGCATCGCTTTTTTGCG			
X11	GTAATGGTG	GCCATAGCTTAATTGCG			
X12	TCAATGGTC	GGCATCGCTTTTTTGCG			
X13	GTAATGGTA	GACATAGCTTAACTGCG			
X14	TCAATGGTC	GGCATAGCTTAATTGCG			
X15	GTAATGGTA	GACATAGGTTAATTGCG			
X16	CTAATGGTC	GACATAGCTTAACTGCG			
X17	GTAATGGTG	GACATCGCTTTTTTGCG			
X18	GTAATGGTA	GACATAGATTAAGTGCG			
X19	GTAATGGTC	GGCATAGCTTAACTGCG			

Figura 4.4: Motifs composing the repetitive trnL-trnF IGS region of the 317 individuals analyzed in this study. Motifs were clasified into three families according to their similarities with the *A. thaliana* functional trnF$_{GAA}$ (complete sequence shown in grey).

Different motif arrangements were found, yielding 57 different pseudogene variants (Table 4.S3). Although in several cases motif order was expected according to the functional trnF (i.e. X-A-T; Fig. 4.4), other different arrangements were also found. For instance, ten pseudogenes showed four motifs due to the presence of an additional T29 motif (Table 4.S3). As the last motif of each haplotype was removed, nine out of the 57 pseudogenes were incomplete and showed less than three motifs (Table 4.S3).

We classified the 69 different haplotypes found into five haplogroups according to their motif arrangement (Fig. 4) and additional features such as the presence of exclusive motifs (e.g. X01 in Group I, T23 in Group II, X10 in Group III, or A04 in Group IV). Haplogroups were composed of an unequal number of haplotypes (Tables 4): Groups I and II consisted of 21 haplotypes each (104 and 90 individuals, respectively); Groups III and V consisted of eight haplotypes (found in 63 and 17 individuals, respectively); and Group IV presented 11 haplotypes (43 individuals). Although the 5′ non-repetitive region showed no resolution to completely discriminate between these groups, they were not randomly distributed: NR3 appeared in Group II, NR4 in Group I, NR1 was found in Groups III to V, and NR2 in Groups I and II (Fig.4.5).

Figura 4.5: Classification and within-group alignment of the 69 studied haplotypes. Haplotype sequences are represented by motif names. Haplotype groups are depicted in different colors (red=Group I, Grey=Group II, Blue=Group III, brown=Group IV, purple= Group V). Empty cells represent gaps and numbers represent motif variants. For each group, the first row represents the family aligned per column (NR=Non-repetitive region, T=T domain-like family, A=A domain-like family, X=X family), whereas the last row represents the consensus sequence (built using the most frequent motif per column). The duplicated pseudogenes defining each group are delimited by black boxes in the consensus sequences.

(Next page) ▶

	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	T	X	A
H58	2	22	28	1	1	11	1																1	1	29	5	9
H43	2	26	28	1	1	11	1																1	1	29	5	9
H18	2	22	28	1	1	11	1							1	11	2	1	21	3	3	1	3	1	1	29	5	9
H53	2	22	28	1	1	11	2							1	11	2	1	21	3	3	1	3	1	1	29	5	
H35	2	22	28	1	1	11	2	1	21	3	3	1	3	1	11	2	1	21	3	3	1	3	1	1	29	5	
H46	2	22	28	1	1	11	2	6	21	3	3	1	3	1	11	2	6	21	3	3	1	3	1	1	29	5	9
H05	2	22	28	1	1	11	2	6	21	3	3	1	3										1	1	29	5	9
H01	2	22	28	1	1	11	2	1	21	3	3	1	3										1	1	29	5	9
H48	2	22	28	1	1	11	2																1	1	29	5	10
H28	2	22	28	1	1	11	2	1	21	3	3	1	3										1	1	29	5	10
H54	2	22	28	1	1	11	2	1	21	3	3	1													29	5	9
H25	2	22	28	1	1	11	2	1	21	3	3	1	3										2	1	29	5	9
H08	2	22	28	1	1	11	2	1	21	3	5	1	3										1	1	29	5	9
H38	2	26	28	1	1	11	2	1	21	3	3	1	3										1	1	29	5	9
H61	4	26	28	1	1	11	2	1	21	3	3	1	3										1	1	29	5	9
H65	2	26	28	1	1	11	2	1	21	15	3	1	3										1	1	29	5	9
H44	2	26	28	2	1	11	2	1	25	15	3	1	3										1	1	29	5	9
H47	2	26	28	1	1	11	2	6	21	15	3	1	3										1	1	29	5	9
H60	2	26	28	1	1	11	13	1	21	15	3	1	3										1	1	29	5	9
H62	2	27	28	1	1	11	13	1	25	15	3	1	3										1	1	29	5	9
H63	4	27	28	2	1	11	18	1	25	15	3	1	3										1	1	29	5	9
GroupI	2	22	28	1	1	11	2	1	21	3	3	1	3	1	11	2	1	21	3	3	1	3	1	1	29	5	9

	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	T	X	A	T	X	A
H11	3	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3				1	3	29	5				9
H23	3	22	28	6										1	1	3				1	3	29	5				9
H45	3	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3	1	1	3	1	3	29	5				9
H67	3	22	28	6	1	11	2	1	21	15	3	23	3	1	1	3	1	1	3	1	3	29	5				9
H68	3	26	28	6	1	11	2	1	25	3	3	23	3	1	1	3	1	1	3	1	3	29	5				9
H17	3	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3				1	3	29	5	3	3	9	9
H64	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3				1	3	29	5	7	3	9	9
H66	2	22	28	6	1	11	2	1	21	3	3	23	3							1	3	29	5	7	3	9	9
H24	2	22	28	6	1	11	2	1	21	3	5	23	3	1	1	3				1	3	29	5	3	3	9	9
H42	2	22	28	6	1	11	2							1	1	3				1	3	29	5	3	3	9	9
H59	2	22	28	6	6	11	2							1	1	3				1	3	29	5	3	3	9	9
H03	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3				1	1	29	5	3	3	9	9
H20	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3				1	1	29	5	3	1	9	9
H14	2	22	28	6	1	11	2	1				23	3	1	1	3				1	1	29	5	3	3	9	9
H39	2	22	28	6	1																				3	9	9
H29	2	22	28	6	1	11	2	1	21	3	3	23	3	1	23	3				1	1	29	5	3	3	9	9
H04	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1							29	5	3	3	9	9
H33	2	22	28	6										1	1							29	5	3	3	9	9
H40	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3	1	1	3	1	1	29	5	3	3	9	9
H27	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3	1	1	3	1	1	29				17	9
H69	2	22	28	19	1	11	2	1	21	15	3	23	3	1	1	3	1	1	3	1	1	29				17	9
GroupII	2	22	28	6	1	11	2	1	21	3	3	23	3	1	1	3	1	1	3	1	3	29	5	3	3	9	9

	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	T	X	A	T	T	X	A	T	X	A
H32	1	22	28	6	1	21	2	1	1	3	1	1	3	1	1	29	5					10	3	10	8
H56	1	22	28	6										1	1	29	5					10	3	10	8
H30	1	22	28	6	1	21	2							1	1	29	5					10	3	10	8
H09	1	22	28	6	1	21	2							1	1	29	5	1	1	29	5	10	3	10	8
H02	1	22	28	6	1	21	2	1	1	3				1	1	29	5					10	3	10	8
H15	1	22	28				2	1	1	3				1	1	29	5					10	3	10	8
H19	1	22	28	6	1	21	2	1	1	3				1	1	29	5					11	3	10	8
H06	1	22	28	6	1	21	2	1	1	3				1	1	29	5	1	1	29	5	10	3	10	8
GroupIII	1	22	28	6	1	21	2	1	1	3	1	1	3	1	1	29	5	1	1	29	5	10	3	10	8

	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	A	T	X	A	T	T	X							
H16	1	22	28	6	1	21	2				1	1	3						1	1	3	1	1	29	5	3	3				12	8
H55	1	22	28	6	1	24	16															1	1	29	5	3	3				12	8
H51	1	22	28	6	1	21	2				1	1									3	1	1	29	5	3	3				12	8
H50	1	22	28	6	1	21	2	1	21	2	1	1									3	1	1	29	5	3	3				12	8
H13	1	22	28	6	6	21	2				1	1									3	1	1	29	5	3	3				12	8
H07	1	22	28	6	1	21	2				1	1									3	1	1	29	5	3	3	14	3	3	12	8
H26	1	22	28	6	1	21	2				1	1									3	1	1	29	5	3	3	14	3			
H10	1	22	28	6	1	21	2				1	1	3	4	1						3	1	1	29	5	3	3				12	8
H12	1	22	28	6	1	21	2				1	1	3	4	1									29	11	3	3				12	8
H57	1	22	28	6	1	21	2	1	21	2	1	1	3	4	1									29	11	3	3				12	8
H21	1	22	28	6	1	21	2				1	1	3	4	1	3	4	1						29	11	3	3				12	8
GroupIV	1	22	28	6	1	21	2	1	21	2	1	1	3	4	1	3	4	1	1	1	3	1	1	29	5	3	3	14	3	3	12	8

	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	T	X	A	T	X	A
H49	1	22	28	2													1	1	29	8				9
H22	1	22	28	2													1	1	29	8	3	3	9	9
H34	1	22	28	2										1	1	3	1	1	29	8	3	3	9	9
H37	1	22	28	2	1	21	13							1	1	3	1	1	29	8	3	3	9	9
H31	1	22	28	2	1	21	13	1	21	13							1	1	29	8	3	3	9	9
H41	1	22	28	2	1	21	13	1	21	13				1	1	3	1	1	29	8	3	3	9	9
H36	1	22	28	2	1	21	13	1	21	13				1	1	3	1	1	29	8	3	1	9	9
H52	1	22	28				13	1	21	13	1	1	3	1	1	3	1	1	29	8	3	3	9	9
GroupV	1	22	28	2	1	21	13	1	21	13	1	1	3	1	1	3	1	1	29	8	3	3	9	9

We used each motif as a single site to perform the sequence alignment, trying to maintain complete pseudogene arrangements and the same motifs in homologous positions. As sequences belonging to different groups showed great differences in motif arrangement, we aligned sequences independently within each group (Fig. 4.5). We obtained a consensus sequence per group showing the most representative features of each group (Fig. 4.5): Group I showed a long duplication involving three pseudogene copies (A01-T11-X02, A01-T21-X03, and A03-T01-X03; black boxes in consensus sequences, Fig. 4.5). Group II showed a maximum of three tandem repeats of one complete pseudogene (X03-A01-T01). Group III haplotypes showed two tandemly repeated copies of two different pseudogenes (A01-T01-X03 and A01-T01-T29-X05). Group IV showed the most complex pattern, with different duplication events involving structures A01-T21-X02, T01-X03-A04, and probably A03-T03-X14 and A01-T01-X03, with subsequent mutations (Fig. 4.5). Group V showed two copies of two tandemly repeated pseudogenes (A01-T21-X13 and A01-T01-X03). We used the alignment of the five resulting consensus sequences per group as a guide to perform the complete alignment (Fig.4.6).

This alignment showed that the repetitive region was composed of a total of 14 pseudogenes (Ψ1 to Ψ14; Fig.4.6). Between the non-repetitive region and Ψ1 all haplotypes showed a highly conserved

Figura 4.6: A) Alignment of the five group consensus sequences used as a guide to perform B) the alignment of the 69 haplotypes found in this study. Haplotype sequences are represented by motif names. Empty cells represent gaps and numbers represent motif variants. The second row represents the motif family aligned per column (NR= Non-repetitive region, T= T domain-like, A= A domain-like, X= X family). The overall alignment is composed of a total of 14 pseudogenes (first row: $\Uppsi$1-$\Uppsi$14), showing between three and four motifs each (e.g. $\Uppsi$1 and $\Uppsi$11, respectively). Colors represent Haplogroups (Red= Group I, Grey= Group II, Blue= Group III, Brown= Group IV, Purple= Group V).

(Next page) ▶

				Ψ1			Ψ2			Ψ3			Ψ4			Ψ5			Ψ6			Ψ7			Ψ8			Ψ9			Ψ10				Ψ11			Ψ12			Ψ13			Ψ14	
	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	A	T	X	A	T	T	X	A	T	T	X	A	T	X	A	T	X	A
GROUP I	2	22	28	1	1	11	2	1	21	3	3	1	3	1	11	2	1	21	3	3	1	3									1	1	29	5											9
GROUP II	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3						1	1	3	1	3	29	5					3	3				9	9
GROUP III	1	22	28	6				1	21	2										1	1	3						1	1	3	1	1	29	5	1	1	29	5	10	3				10	8
GROUP IV	1	22	28	6				1	21	2	1	21	2							1	1	3	4	1	3	4	1	1	1	3	1	1	29	5					3	3	14	3	3	12	8
GROUP V	1	22	28				2	1	21	13	1	21	13							1	1	3						1	1	3	1	1	29	8					3	3				9	9

				Ψ1			Ψ2			Ψ3			Ψ4			Ψ5			Ψ6			Ψ7			Ψ8			Ψ9			Ψ10				Ψ11			Ψ12			Ψ13			Ψ14	
	NR	T	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	X	A	T	A	T	X	A	T	T	X	A	T	T	X	A	T	X	A	T	X	A
H58	2	22	28	1	1	11	1																								1	1	29	5											9
H43	2	26	28	1	1	11	1																								1	1	29	5											9
H18	2	22	28	1	1	11	1							1	11	2	1	21	3	3	1	3									1	1	29	5											9
H53	2	22	28	1	1	11	2							1	11	2	1	21	3	3	1	3									1	1	29	5											
H35	2	22	28	1	1	11	2	1	21	3	3	1	3	1	11	2	1	21	3	3	1	3									1	1	29	5											
H46	2	22	28	1	1	11	2	6	21	3	3	1	3	1	11	2	6	21	3	3	1	3									1	1	29	5											9
H05	2	22	28	1	1	11	2	6	21	3	3	1	3																		1	1	29	5											9
H01	2	22	28	1	1	11	2	1	21	3	3	1	3																		1	1	29	5											9
H48	2	22	28	1	1	11	2																								1	1	29	5											10
H28	2	22	28	1	1	11	2	1	21	3	3	1	3																		1	1	29	5											10
H54	2	22	28	1	1	11	2	1	21	3	3	1																					29	5											9
H25	2	22	28	1	1	11	2	1	21	3	3	1	3																		2	1	29	5											9
H08	2	22	28	1	1	11	2	1	21	3	5	1	3																		1	1	29	5											9
H38	2	26	28	1	1	11	2	1	21	3	3	1	3																		1	1	29	5											9
H61	4	26	28	1	1	11	2	1	21	3	3	1	3																		1	1	29	5											9
H65	2	26	28	1	1	11	2	1	21	15	3	1	3																		1	1	29	5											9
H44	2	26	28	2	1	11	2	1	25	15	3	1	3																		1	1	29	5											9
H47	2	26	28	1	1	11	2	6	21	15	3	1	3																		1	1	29	5											9
H60	2	26	28	1	1	11	13	1	21	15	3	1	3																		1	1	29	5											9
H62	2	27	28	1	1	11	13	1	25	15	3	1	3																		1	1	29	5											9
H63	4	27	28	2	1	11	18	1	25	15	3	1	3																		1	1	29	5											9
H11	3	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3									1	3	29	5											9
H23	3	22	28	6																1	1	3									1	3	29	5											9
H45	3	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3						1	1	3	1	3	29	5											9
H67	3	22	28	6	1	11	2	1	21	15	3	23	3							1	1	3						1	1	3	1	3	29	5											9
H68	3	26	28	6	1	11	2	1	25	3	3	23	3							1	1	3						1	1	3	1	3	29	5											9
H17	3	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3									1	3	29	5					3	3				9	9
H64	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3									1	3	29	5					7	3				9	9
H66	2	22	28	6	1	11	2	1	21	3	3	23	3																		1	3	29	5					7	3				9	9
H24	2	22	28	6	1	11	2	1	21	3	5	23	3							1	1	3									1	3	29	5					3	3				9	9
H42	2	22	28	6	1	11	2													1	1	3									1	3	29	5					3	3				9	9
H59	2	22	28	6	6	11	2													1	1	3									1	3	29	5					3	3				9	9
H03	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3									1	1	29	5					3	3				9	9
H20	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3									1	1	29	5					3	1				9	9
H14	2	22	28	6	1	11	2	1				23	3							1	1	3									1	1	29	5					3	3				9	9
H39	2	22	28	6	1																																			3				9	9
H29	2	22	28	6	1	11	2	1	21	3	3	23	3							1	23	3									1	1	29	5					3	3				9	9
H04	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1												29	5					3	3				9	9
H33	2	22	28	6																1	1												29	5					3	3				9	9
H40	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3						1	1	3	1	1	29	5					3	3				9	9
H27	2	22	28	6	1	11	2	1	21	3	3	23	3							1	1	3						1	1	3	1	1	29											17	9
H69	2	22	28	19	1	11	2	1	21	15	3	23	3							1	1	3						1	1	3	1	1	29											17	9
H32	1	22	28	6				1	21	2										1	1	3						1	1	3	1	1	29	5					10	3				10	8
H56	1	22	28	6																											1	1	29	5					10	3				10	8
H30	1	22	28	6				1	21	2																					1	1	29	5					10	3				10	8
H09	1	22	28	6				1	21	2																					1	1	29	5	1	1	29	5	10	3				10	8
H02	1	22	28	6				1	21	2										1	1	3									1	1	29	5					10	3				10	8
H15	1	22	28							2										1	1	3									1	1	29	5					10	3				10	8
H19	1	22	28	6				1	21	2										1	1	3									1	1	29	5					11	3				10	8
H06	1	22	28	6				1	21	2										1	1	3									1	1	29	5	1	1	29	5	10	3				10	8
H16	1	22	28	6				1	21	2										1	1	3						1	1	3	1	1	29	5					3	3				12	8
H55	1	22	28	6				1	24	16																					1	1	29	5					3	3				12	8
H51	1	22	28	6				1	21	2										1	1									3	1	1	29	5					3	3				12	8
H50	1	22	28	6				1	21	2	1	21	2							1	1									3	1	1	29	5					3	3				12	8
H13	1	22	28	6				6	21	2										1	1									3	1	1	29	5					3	3				12	8
H07	1	22	28	6				1	21	2										1	1									3	1	1	29	5					3	3	14	3	3	12	8
H26	1	22	28	6				1	21	2										1	1									3	1	1	29	5					3	3	14	3			
H10	1	22	28	6				1	21	2										1	1	3	4	1						3	1	1	29	5					3	3				12	8
H12	1	22	28	6				1	21	2										1	1	3	4	1									29	11					3	3				12	8
H57	1	22	28	6				1	21	2	1	21	2							1	1	3	4	1									29	11					3	3				12	8
H21	1	22	28	6				1	21	2										1	1	3	4	1	3	4	1						29	11					3	3				12	8
H49	1	22	28				2																								1	1	29	8											9
H22	1	22	28				2																								1	1	29	8					3	3				9	9
H34	1	22	28				2																					1	1	3	1	1	29	8					3	3				9	9
H37	1	22	28				2	1	21	13																		1	1	3	1	1	29	8					3	3				9	9
H31	1	22	28				2	1	21	13	1	21	13																		1	1	29	8					3	3				9	9
H41	1	22	28				2	1	21	13	1	21	13															1	1	3	1	1	29	8					3	3				9	9
H36	1	22	28				2	1	21	13	1	21	13															1	1	3	1	1	29	8					3	1				9	9
H52	1	22	28							13	1	21	13							1	1	3						1	1	3	1	1	29	8					3	3				9	9

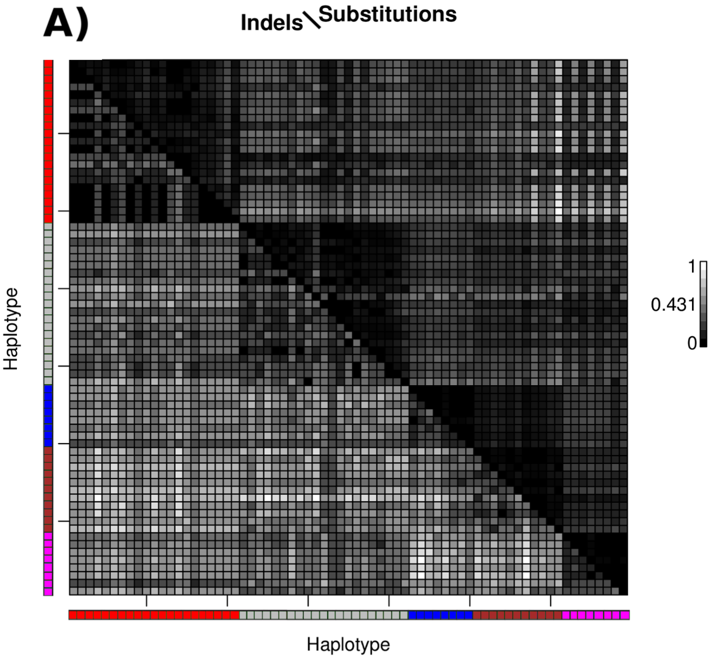
region containing two T family motifs: T22-T28. Only Haplogroups I and II showed variation in this region (nine and one haplotypes, respectively).

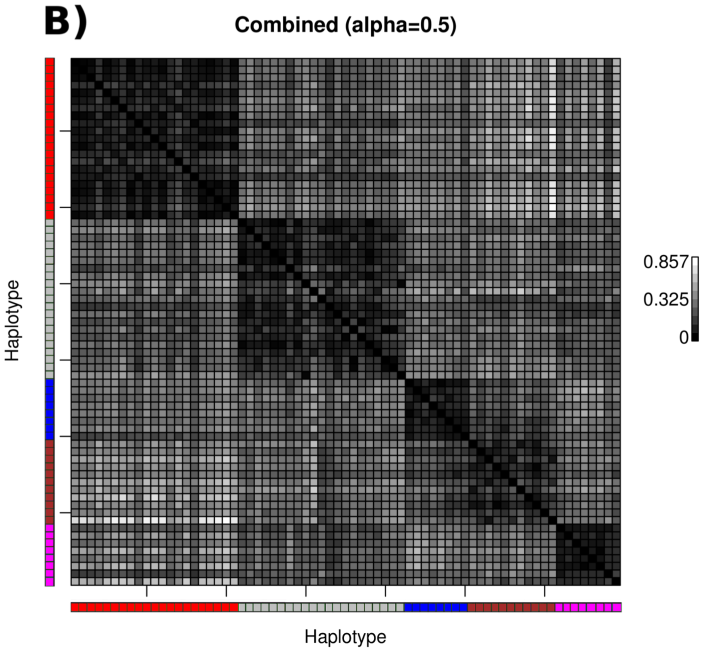
Motif names were replaced by the sequences they represent to yield the final nucleotide alignment, which was used to calculate the two distance matrices, based on substitutions and indels (Fig 4.7). Both datasets showed significant congruence according to the three CADM tests performed. The direct comparison between both distance matrices yielded the most congruent result (W=0.727, p<0.001). Lower congruence values were obtained using NJ patristic distances performed with both untransformed (W=0.535, p=0.020) and transformed (W=0.562, p<0.001) distance matrices. Even lower congruence values were obtained when the length of all the branches was collapsed into one. In this case, NJ patristic distances estimated without transformation were not congruent (Untransformed distances: W=0.511, p=0.200; Transformed distances: W=0.543, p=0.004).

Networks built using the combined genetic distance matrix kept all the haplotypes connected in a single graph, even when thresholds were as low as the 24 % of the maximum distance value (Fig.4.8A). The network built by connecting haplotypes with distances lower than the estimated percolation threshold (0.231188; Fig. 4.8A) showed four modules (Fig. 4.8B). Three of them correspond to Haplogroups I, II and V described above, while the fourth module encom-

Figura 4.7: A) Representation of the haplotypic transformed distance matrices based on substitutions (p-distances, above diagonal) and indels (number of mutation events, below diagonal). Note that distances ranged between 0 and 1 in both cases. B) Representation of the combined distance matrix, giving equal weight to both matrices. Colors on axis represent Haplogroups (Red= Group I, Grey= Group II, Blue= Group III, Brown= Group IV, Purple= Group V).

(Next page) ▶


A)
Indels\Substitutions
Haplotype
Haplotype
1
0.431
0


B)
Combined (alpha=0.5)
Haplotype
Haplotype
0.857
0.325
0

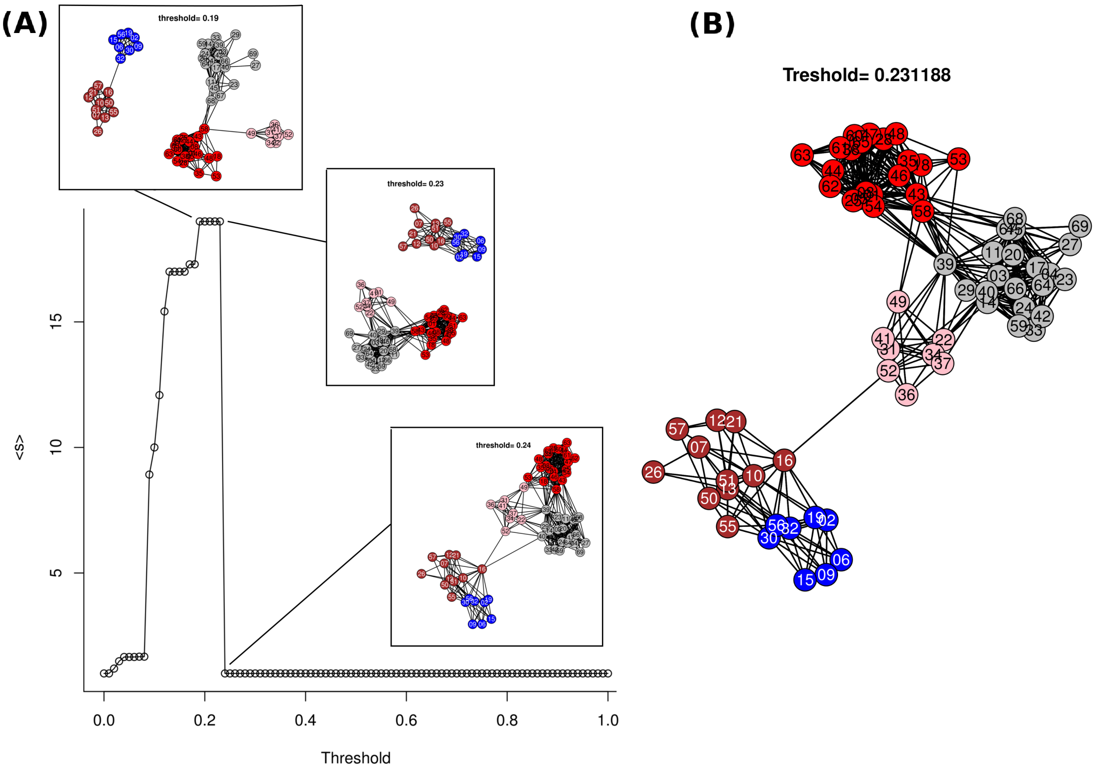


Figura 4.8: Depiction of the percolation threshold estimate and the network produced (A). The threshold determines the maximum distance (as the percentage of the maximum distance found) permitting the link between haplotypes, ranging from 100 % (saturated network) to 0 % (no links between haplotypes). We built a network (depicted for thresholds 0.19, 0.23, and 0.24) for each threshold value and estimated the average cluster size, excluding the largest one (<s>). The percolation threshold is determined by the strong increase in <s> between 0.23 and 0.24. (B) Network yielded with the estimated percolation threshold (0.231188). Node numbers and colors represent haplotype numbers and groups, respectively (Red = Group I; Grey = Group II; Blue = Group III; Brown = Group IV; Purple = Group V).

passes haplotypes belonging to Haplogroups III and IV, separately placed according to the Fruchterman-Reingold Algorithm (Fig. 4.8B).

Whereas haplotypes belonging to Haplogroups I and II were connected to 18.8 and 16.8 nodes in average, haplotypes belonging to the remaining Haplogroups showed an average of 9.1, 8.5, and 9.4 connections per node in the yielded percolation network (Table 4). H39 (the shortest haplotype) was the most connected haplotype (linked to 32 haplotypes) and presented a central position between Haplogroups I, II, and V. Although less connected (14 links), H16 was essential to link Haplogroups II and IV to the remaining network through H52 (Fig. 4.8B). H26 and H57 (both belonging to Haplogroup IV) were the least connected haplotypes (only three and four links, respectively; Table 4). No correlation was found between the number of links per haplotype and haplotype frequency (r^2=0.006, p=0.514, df=67) or length (r^2<0.001, p=0.961, df=67). Using lower thresholds, Haplogroups III and IV were disconnected from the rest of the network (threshold= 0.23) and Group V was only linked to Group I (threshold= 0.19, Fig.4.8A).

DISCUSSION

In this study we propose a new method for finding out the evolutionary relationship between indel-rich sequences, using information from both substitutions and indels. This approach has been followed via three multi-step stages. The first stage involved building an alignment based on motif identification, grouping sequences according to motif codification, obtaining within-group consensus sequences, aligning these consensus sequences, and, finally, decoding motifs to obtain the final alignment (Fig. 4.4). The second stage involved obtaining distance matrices for indels and substitutions, and combining both matrices. The third stage involved applying a percolation threshold to obtain an inclusive haplotype network (Fig. 4.2). While the first stage is highly-system dependent, the second and third stages can be applied to a wide range of evolutionary situations.

Indel-rich regions may offer valuable information about the recent evolution of many organisms (Redelings and Suchard, 2007). However, the use of these regions has been restricted in phylogeographic or phylogenetic inference due to the difficulty in obtaining an accurate alignment, among other reasons (Castresana, 2000). These regions have therefore been traditionally removed from the alignment (e.g., Fischer et al., 2007; Schmickl et al., 2010), used as a size polymorphism (Provan et al., 2001; García-Verdugo et al., 2010), or used to group haplotypes (Wittzell, 1999). To reduce the possibility of obtaining misleading alignments, a common strategy is to use the information coming from sequence secondary structure (e.g., Subbotin et al., 2007; Letsch et al., 2010; Goertzen et al., 2003). As the functionality of the secondary structure of the trnL-trnF IGS region has been questioned (Bakker et al., 2000), we have not implemented it in our alignment strategy. However, we have implicitly maintained secondary structure, since we have maintained pseudogene integrity during the alignment procedure.

Some authors have recently aligned the trnL-trnF IGS region in Brassicaceae using different criteria to determine nucleotide homology. Alignments have been performed manually (eg, Koch et al., 2005, 2007; Dobeš et al., 2007; Koch and Matschinger, 2007; Hoebe et al., 2009; Ansell et al., 2010; Poczai and Hyvönen, 2011), manually but software-assisted (Bakker et al., 2000; Bailey et al., 2002), or purely software-assisted (Drábková et al., 2004; Müller et al., 2006). Borsch et al. (2003) have detailed the rules followed to align the trnT-trnL region, improving the repeatability of their manual procedure. We have classified similar haplotypes into Haplogroups and, separately, manually aligned the sequences of each group, following the rationale of Kelchner (2000) and Borsch et al. (2003), to obtain one consensus sequence per group. We used these consensus sequences to obtain the final alignment, not only because the number of sequences to align was lower, but also because it was easier to find common pseudogenes among consensus sequences. A conceptually similar alignment procedure ("divide-and-conquer strategy") has been recently implemented (Liu et al., 2010) to simultaneously estimate alignment and phylogeny. However, this method uses an ML approach and treats gaps as missing data.

According to the abundant information available for the trnL-trnF IGS region in several plant species (e.g., Vijverberg and Bachmann, 1999; Wittzell, 1999; Dobeš et al., 2004, 2007; Drábková et al., 2004; Stoneberg Holt et al., 2004; Fischer et al., 2007; Koch et al., 2007; Pirie et al., 2007; Schmickl et al., 2008; Poczai and Hyvönen, 2011), we have assumed that most indel events in this region involved one or several whole pseudogenes, composed of three motifs each. This assumption was confirmed by the fact that haplotypes mostly differed in 3x number of motifs (Fig. 4.S2). In contrast with this constancy, haplotypes may differ in their motifs arrangement (e.g. ATX instead of the canonical XAT), suggesting that the number of motifs involved in an indel event is more frequently preserved than its arrangement. Consequently, we have used motifs (instead of nucleotides) as sequence units to perform the alignment, trying to maintain the pseudogene sequences as complete as possible.

Indels and substitutions have been used together in different approaches, such as parsimony (Ogden and Rosenberg, 2007) and Bayesian Inference (Rivas and Eddy, 2008). However, the treatment of gaps in distance-based methods has not been explored (Ogden and Rosenberg, 2007), although these methods adequately reconstruct the phylogeny in short-branch scenarios (Roch, 2010), as we found for the trnL-trnF IGS region. We have implemented a conditional combination method strategy (sensu Huelsenbeck et al., 1996) combining congruent distance matrices obtained separately from each kind of mutation. Although they may evolve at different rates, the

CADM test has shown that the phylogenetic information they provide is congruent. From this perspective, considering substitutions and indels separately will make phylogenetic inferences more prone to sampling error than combining them (Huelsenbeck et al., 1996). In fact, Löhne and Borsch (2005), exploring the phylogenetic relationship of basal angiosperms using petD intron, found that the information provided by indels was congruent with the signal inferred from substitutions, improving the phylogenetic resolution.

The trnL-trnF IGS region probably shows complex relationships between haplotypes (Tedder et al., 2010), due to its high mutation rate. These haplotype relationships are better visualized as networks than as trees, because networks can depict the complex relationships usually appearing between haplotypes without assuming a bifurcation process (Morrison, 2005; Mardulyn, 2012). However, one major challenge of using networks in this context is finding an objective criterion to establish links between haplotypes. There are several methods for constructing haplotype networks, including minimum spanning network, split decomposition, statistical parsimony, Median-joining, etc (Cassens et al., 2003; Morrison, 2005; Huson and Bryant, 2006). In our study we used the percolation threshold analysis (Rozenfeld et al., 2008). This approach maintains a higher number of between-haplotype connections than other network techniques, thereby capturing the uncertainty associated with the low divergence usually found in haplotype evolution. The low value of the percolation threshold (23.1 % of the maximum distance) points to the existence of multiple low distances between node pairs, thus confirming the many possible connections between haplotypes. For these reasons, we think that the method we have used in this study is particularly useful for sequences showing low divergence. This is especially true in phylogeographic studies, where a small deviation between the number of observed and actual mutations is expected (Fig. 4.S1). In this scenario, substitutions and indels will provide low information separately, and combining both matrices will surely increase the degree of phylogenetic signal. However, more empirical and theoretical studies are necessary to examine the suitability of this method for other evolutionary situations.

TrnF pseudogenes in Erysimum

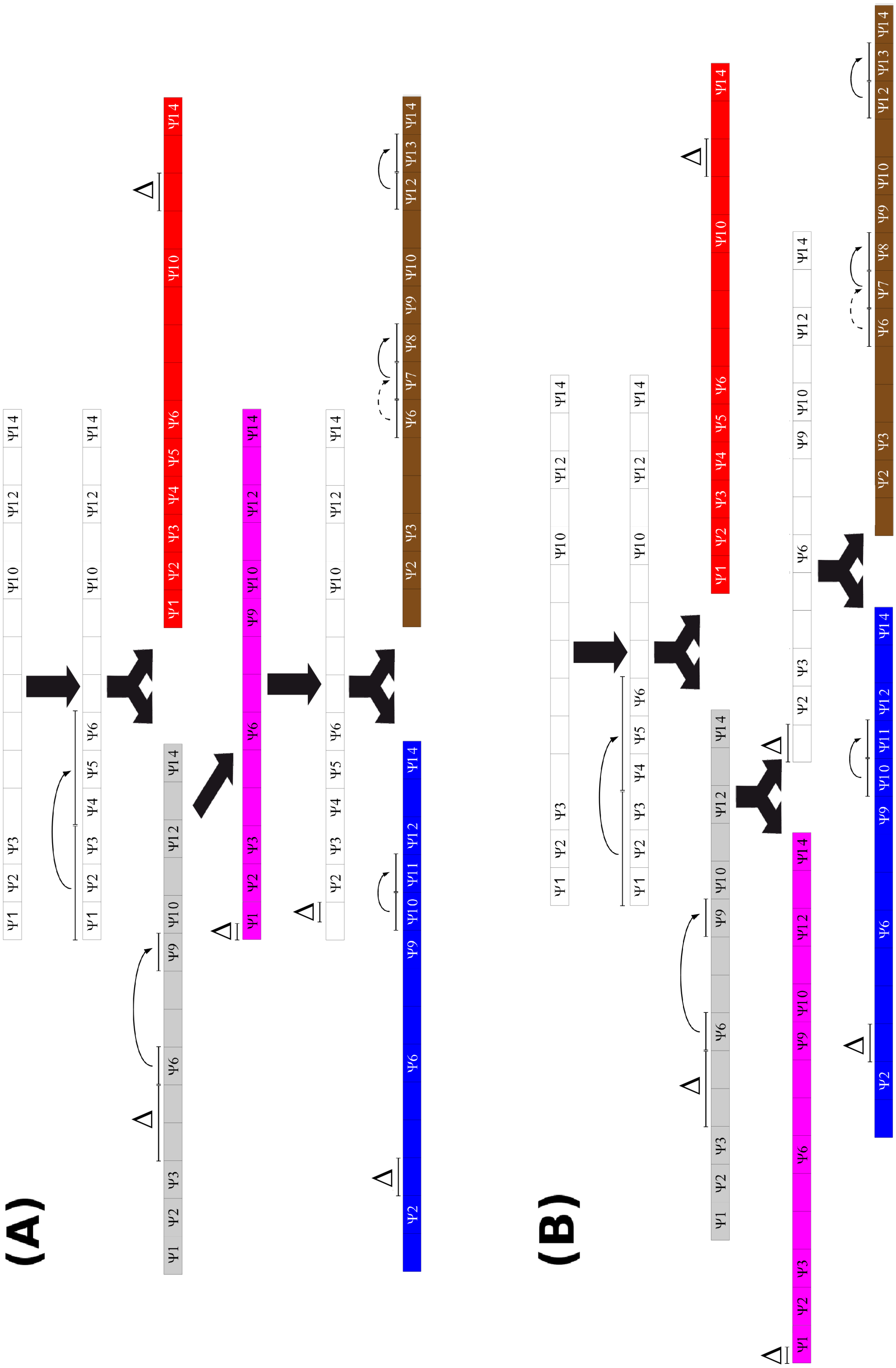
The variation in the number of pseudogenes of the trnF IGS region in *Erysimum* spp. is high even compared to other Brassicaceae species (e.g., Dobeš et al., 2004, 2007; Koch et al., 2005). The alignment of the studied region were composed of fourteen different tandemly repeated pseudogenes. Most of them showed a structure similar to that of the functional trnF (i.e. three domains with different motifs of X, A, and T families) (Table 4.S3). While A motifs can be considered homologous to the functional trnF anticodon region and T motifs showed a partial similarity with functional trnF T-domain, X motifs showed no similarity with functional trnF (Fig. 4.4). This X family of motifs are also present in the trnF pseudogenes of other species (e.g., Koch et al., 2005; Poczai and Hyvönen, 2011) and could be interpreted as a modification of the D-domain, allowing a secondary arrangement to improve folding of the spacer transcript. The same explanation could be applied to the evolution of T-like motifs. Nevertheless, Koch et al. (2005) concluded that the X motif region is the result of the rearrangement of the trnF gene and its original flanking regions.

We have found an average of one haplotype per population (represented by five individuals each), and 24 out of the 69 haplotypes were found in only one individual (Table 4). This abundance of rare haplotypes could be attributable to PCR polymerase errors if haplotype differences were caused by single nucleotide mutations. However, most between-haplotype differences were due to indels (Table 4). Furthermore, we were conservative and removed any doubtful substitutions (e.g. the last motif in all haplotypes) and, consequently, haplotype diversity has probably been underestimated. In any case, haplogroup definitions and evolutionary relationships were clear, even with a large variation in motif families. The high number of haplotypes at a low frequency could also indicate a high mutation rate and/or non-random processes affecting this region. In fact, the region encompassing both transfer genes and the IGS is co-transcribed (Kanno and Hirai, 1993), and it has been proposed that there are constraints on length variation in the IGS (Koch et al., 2005). The marginally significant correlation between haplotype frequency and length (r^2=0.053, p=0.056, df=67) suggests that *E. mediohispanicum* IGS length variation may also be constrained.

The observed haplotype lengths suggest that the main mechanism increasing trnL-trnF IGS variability is the insertion or deletion of DNA fragments with the same length as the functional trnF (i.e., 3x motifs). Such indels occur even with different motif arrangements. For example, H14 showed deletion of pseudogene arrangements such as XAT in Ψ3 and Ψ13, or ATX in Ψ4 and Ψ5 (Fig. 4.7B). Three out of the 14 pseudogenes composing the final alignment also showed arrangements different to AXT, namely Ψ10 and Ψ11 (with an extra T motif) and Ψ9 (lacking the X motif; Fig. 4.7B). Conservation of the structure of trnF pseudogenes has been also observed in other Brassicaceae, such as *Arabidopsis* (Dobeš et al., 2004) and *Boechera* (Dobeš et al., 2007), as well as in other plant families, such as Asteraceae (Vijverberg and Bachmann, 1999; Wittzell, 1999), Juncaceae (Drábková et al., 2004), Poaceae (Stoneberg Holt et al., 2004), Annonaceae (Pirie et al., 2007), Orchidaceae (Fischer et al., 2007), and Solanaceae (Poczai and Hyvönen, 2011). The presence of tandemly arranged pseudogenes are the outcome of several molecular mechanisms, including inter- and intra-molecular recombination, impairing during DNA duplication, and slippage (Ansell et al., 2007; Dobeš et al., 2007).

Haplotypes from different *Erysimum* species did not group in independent clusters (Fig. 4.S4). This may be a consequence of several non-exclusive processes, such as homoplasy, incomplete lineage sorting, recent origin, or interspecific hybridization. Because haplotypes from different species were grouped together according to their geographic proximity (Chapter 6), we believe that homoplasy was not a major causative factor behind the observed pattern. A differentiation between the other three potential explanations would require the use of additional molecular markers and methods (Morando et al., 2004; Chapters 6 and 7).

Using the information provided by the methodological approach described in this study, it is possible to explore the evolutionary relationship between haplogroups (Fig. 4.9). On the basis of their high frequency (Fig. 4.7B), we hypothesize that pseudogenes Ψ1, Ψ2, Ψ3, Ψ10, Ψ12, and Ψ14 were present in the ancestral haplotype. Sequence similarity suggests that pseudogenes Ψ4 to Ψ6 were originated via duplication from pseudogenes Ψ1 to Ψ3, also in the ancestral haplotype (Figs. 4.6 and 4.9). We propose that Haplogroups I and II originated from this hypothetical haplotype. Whereas Haplogroup II appeared as a consequence of Ψ4-Ψ5 deletion and the duplication of Ψ6 producing Ψ9, Haplogroup I appeared as a consequence of Ψ12 deletion (Fig. 4.9). From this point, two alternative routes are possible (Fig. 4.9). According to the route depicted in Fig 4.10A, Haplogroup V originated from Haplogroup II by a partial deletion of Ψ1. The other two remaining Haplogroups III and IV originated from Haplogroup V through an intermediate step involving a complete Ψ1 deletion. Haplogroup III appeared after deletion of Ψ3 and duplication of Ψ10 to produce Ψ11, whereas Haplogroup IV originated after Ψ6 duplication to produce Ψ7, with Ψ7 duplication producing Ψ8, and Ψ12 duplication producing Ψ13. According to the route depicted in Fig 4.10B, Haplogroup II also originated Haplogroup V, but this group was not the ancestor of the remaining haplogroups. In


(A)
(B)
Δ
Ψ1
Ψ2
Ψ3
Ψ4
Ψ5
Ψ6
Ψ7
Ψ8
Ψ9
Ψ10
Ψ11
Ψ12
Ψ13
Ψ14

◀ (Previous page)

Figura 4.9: Main indel events explaining the arrangement of the 14 pseudogenes found in the *E. mediohispanicum* trnL-trnF IGS region. We assumed duplication events between identical pseudogene copies (solid thin arrows), but no identical pseudogene was found for Ψ7 and a putative origin is indicated (dashed thin arrows). Deletion events are represented with Δ. Two alternative hypotheses are described: Haplogroup V (A) as ancestor and (B) as paraphyletic of Haplogroups III and IV. Colors represent group consensus sequences (Red = Group I; Grey = Group II; Blue = Group III; Brown = Group IV; Purple = Group V).

contrast, Haplogroup II produced an intermediate haplotype involving a complete Ψ1 deletion. From this intermediate, Haplogroup III originated through the deletion of Ψ13 and duplication of Ψ10 to produce Ψ11, whereas Haplogroup IV originated from Ψ6 duplication to produce Ψ7, with duplication of Ψ7 producing Ψ8, and duplication of Ψ12 producing Ψ13.

Due to the restricted geographical location of the populations showing Haplogroup V haplotypes, the two described alternatives imply two highly different colonization paths (Chapter 6). Overall, our findings suggest the occurrence of two independent evolutionary pathways, one generating Haplogroup I haplotypes, the other producing the remaining haplogroups found in this study.

REFERENCES

Ansell, S. W., H. Schneider, N. Pedersen, M. Grundmann, S. J. Russell, and J. C. Vogel (2007). Recombination diversifies chloroplast trnF pseudogenes in arabidopsis lyrata. *Journal of Evolutionary Biology 20*(6), 2400–2411.

Ansell, S. W., H. K. Stenoien, M. Grundmann, H. Schneider, A. Hemp, N. Bauer, S. J. Russell, and J. C. Vogel (2010). Population structure and historical biogeography of european arabidopsis lyrata. *Heredity 105*(6), 543–553.

Ančev, M. (2006). Polyploidy and hybridization in bulgarian brassicaceae: distribution and evolutionary role. *Phytologia Balcanica 12*(3), 357–366.

Bailey, C. D., R. A. Price, and J. J. Doyle (2002). Systematics of the halimolobine brassicaceae: Evidence from three loci and morphology. *Systematic Botany 27*(2), 318–332.

Bakker, F. T., A. Culham, R. Gomez-Martinez, J. Carvalho, J. Compton, R. Dawtrey, and M. Gibby (2000). Patterns of nucleotide substitution in angiosperm cpDNA trnL (UAA)–trnF (GAA) regions. *Molecular Biology and Evolution 17*(8), 1146 –1155.

Baum, D. A., K. Sytsma, and P. C. Hoch (1994). A phylogenetic analysis of epilobium (onagraceae) based on nuclear ribosomal DNA sequences. *Systematic Botany 19*, 363–388.

Blair, C. and R. W. Murphy (2011). Recent trends in molecular phylogenetic analysis: Where to next? *Journal of Heredity 102*(1), 130 –138.

Borsch, T., K. Hilu, D. Quandt, V. Wilde, C. Neinhuis, and W. Barthlott (2003). Noncoding plastid trnt-trnf sequences reveal a well resolved phylogeny of basal angiosperms. *J Evol Biol 16*(4), 558–76.

Campbell, V., P. Legendre, and F.-J. Lapointe (2011). The performance of the congruence among distance matrices (CADM) test in phylogenetic analysis. *BMC Evolutionary Biology 11*(1), 64.

Cassens, I., K. Van Waerebeek, P. Best, E. Crespo, J. Reyes, and M. Milinkovitch (2003). The phylogeography of dusky dolphins (lagenorhynchus obscurus): a critical examination of network methods and rooting procedures. *Mol Ecol 12*(7), 1781–92.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution 17*(4), 540 –552.

Clot, B. (1991). Caryosystématique de quelques erysimum l. dans le nord de la péninsule ibérique. *Anales del Jardín Botánico de Madrid 49*(2), 215–229.

Couvreur, T. L. P., A. Franzke, I. A. Al-Shehbaz, F. T. Bakker, M. A. Koch, and K. Mummenhoff (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (brassicaceae). *Molecular Biology and Evolution 27*(1), 55 –71.

Dobeš, C., C. Kiefer, M. Kiefer, and M. A. Koch (2007). Plastidic trnFUUC pseudogenes in north american genus boechera (brassicaceae): Mechanistic aspects of evolution. *Plant biol (Stuttg) 9*(04), 502,515. 502.

Dobeš, C. H., T. Mitchell-Olds, and M. A. Koch (2004). Extensive chloroplast haplotype variation indicates pleistocene hybridization and radiation of north american arabis drummondii, a. divaricarpa, and a. holboellii (brassicaceae). *Molecular Ecology 13*(2), 349–370.

Drábková, L., J. Kirschner, Čestmír Vlček, and V. Pačes (2004). TrnL-TrnF intergenic spacer and TrnL intron define major clades within luzula and juncus (juncaceae): Importance of structural mutations. *Journal of Molecular Evolution 59*(1), 1–10.

Dwivedi, B. and S. Gadagkar (2009). Phylogenetic inference under varying proportions of indel-induced alignment gaps. *BMC Evolutionary Biology 9*(1), 211.

Excoffier, L., P. E. Smouse, and J. M. Quattro (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics 131*(2), 479–491.

Fischer, G. A., B. Gravendeel, A. Sieder, J. Andriantiana, P. Heiselmayer, P. J. Cribb, E. de Camargo Smidt, R. Samuel, and M. Kiehn (2007). Evolution of resupination in malagasy species of bulbophyllum (orchidaceae). *Molecular Phylogenetics and Evolution 45*(1), 358–376.

Freudenstein, J. V. and M. W. Chase (2001). Analysis of mitochondrial nad1b-c intron sequences in orchidaceae: Utility and coding of length-change characters. *Systematic Botany 26*(3), 643–657.

García-Verdugo, C., A. D. Forrest, M. F. Fay, and P. Vargas (2010). The relevance of gene flow in metapopulation dynamics of an oceanic island endemic, olea europaea subsp. guanchica. *Evolution 64*(12), 3525–3536.

Geiger, D. (2002). Stretch coding and block coding: two new strategies to represent questionably aligned DNA sequences. *J Mol Evol 54*(2), 191–9.

Goertzen, L. R., J. J. Cannone, R. R. Gutell, and R. K. Jansen (2003). ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the asteraceae. *Molecular Phylogenetics and Evolution 29*(2), 216–234.

Golenberg, E. M., M. T. Clegg, M. L. Durbin, J. Doebley, and D. P. Ma (1993). Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution 2*(1), 52–64.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series 41*, 95–98.

Hoebe, P. N., M. Stift, A. Tedder, and B. K. Mable (2009). Multiple losses of self-incompatibility in north-american arabidopsis lyrata?: Phylogeographic context and population genetic consequences. *Molecular Ecology 18*(23), 4924–4939.

Huelsenbeck, J. P., J. Bull, and C. W. Cunningham (1996). Combining data in phylogenetic analysis. *Trends in Ecology & Evolution 11*(4), 152–158.

Huson, D. H. and D. Bryant (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution 23*(2), 254 –267.

Kanno, A. and A. Hirai (1993). A transcription map of the chloroplast genome from rice (oryza sativa). *Curr Genet 23*(2), 166–74.

Kelchner, S. A. (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annual Missouri Botanic Garden 87*(4), 482–498.

Kjer, K. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Molecular Phylogenetics and Evolution 4*(3), 314–330.

Kjer, K. M., J. J. Gillespie, and K. A. Ober (2007). Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. *Systematic Biology 56*(1), 133 –146.

Koch, M. A., C. Dobeš, C. Kiefer, R. Schmickl, L. Klimeš, and M. A. Lysak (2007). Supernetwork identifies multiple events of plastid trnF(GAA) pseudogene evolution in the brassicaceae. *Molecular Biology and Evolution 24*(1), 63 –73.

Koch, M. A., C. Dobeš, M. Matschinger, W. Bleeker, J. Vogel, M. Kiefer, and T. Mitchell-Olds (2005). Evolution of the trnF(GAA) gene in arabidopsis relatives and the brassicaceae family: Monophyletic origin and subsequent diversification of a plastidic pseudogene. *Molecular Biology and Evolution 22*(4), 1032 –1043.

Koch, M. A. and M. Matschinger (2007). Evolution and genetic differentiation among relatives of arabidopsis thaliana. *Proceedings of the National Academy of Sciences 104*(15), 6272 –6277.

Kumar, S. and A. Filipski (2007). Multiple sequence alignment: In pursuit of homologous DNA positions. *Genome Research 17*(2), 127 –135.

Legendre, P. and F. J. Lapointe (2004). Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Australian & New Zealand Journal of Statistics 46*(4), 615–629.

Letsch, H. O., P. Kück, R. R. Stocsits, and B. Misof (2010). The impact of rRNA secondary structure consideration in alignment and tree reconstruction: Simulated data and a case study on the phylogeny of hexapods. *Molecular Biology and Evolution 27*(11), 2507 –2521.

Liu, K., C. Linder, and T. Warnow (2010). Multiple sequence alignment: a major challenge to large-scale phylogenetics. *PLoS Curr 2*, RRN1198.

Liu, K., T. J. Warnow, M. T. Holder, S. M. Nelesen, J. Yu, A. P. Stamatakis, and C. R. Linder (2012). SATé-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology 61*(1), 90 –106.

Löhne, C. and T. Borsch (2005). Molecular evolution and phylogenetic utility of the petD group II intron: A case study in basal angiosperms. *Molecular Biology and Evolution 22*(2), 317 –332.

Loytynoja, A. and M. Milinkovitch (2001). SOAP, cleaning multiple alignments from unstable blocks. *Bioinformatics 17*(6), 573–4.

Lunter, G., I. Miklos, A. Drummond, J. Jensen, and J. Hein (2005). Bayesian coestimation of phylogeny and sequence alignment. *BMC Bioinformatics 6*(1), 83.

Lutzoni, F., P. Wagner, V. Reeb, and S. Zoller (2000). Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Systematic Biology 49*(4), 628 –651.

Mardulyn, P. (2012). Trees and/or networks to display intraspecific DNA sequence variation? *Molecular Ecology 21*(14), 3385–3390.

Marhold, K. and J. Lihová (2006). Polyploidy, hybridization and reticulate evolution: lessons from the brassicaceae. *Plant Systematics and Evolution 259*(2), 143–174.

Morando, M., L. J. Avila, J. Baker, and J. W. Sites (2004). Phylogeny and phylogeography of the liolaemus darwinii complex (squamata: Liolaemidae): Evidence for introgression and incomplete lineage sorting. *Evolution 58*(4), 842–859.

Morrison, D. A. (2005). Networks in phylogenetic analysis: new tools for population biology. *International Journal for Parasitology 35*(5), 567–582.

Morrison, D. A. (2009). Why would phylogeneticists ignore computerized sequence alignment? *Systematic Biology 58*(1), 150 –158.

Müller, K. (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution 38*(3), 667–676.

Müller, K. F., T. Borsch, and K. W. Hilu (2006). Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: Contrasting matK, trnT-F, and rbcL in basal angiosperms. *Molecular Phylogenetics and Evolution 41*(1), 99–117.

Nei, M. and S. Kumar (1998). *Molecular Evolution and Phylogenetics*. Oxford University Press.

Nieto-Feliner, G. (1993). Erysimum. In S. Castroviejo, C. Aedo, C. Gómez-Campo, M. Lainz, P. Monserrat, R. Morales, F. Muñoz-Garmendia, G. Nieto-Feliner, E. Rico, S. Talavera, and L. Villar (Eds.), *Flora Iberica*, Volume 4, Cruciferae-Monotropaceae., pp. 48–76. Madrid: Real Jardín Botánico CSIC.

Ogden, T. H. and M. S. Rosenberg (2007). How should gaps be treated in parsimony? a comparison of approaches using simulation. *Molecular Phylogenetics and Evolution 42*(3), 817–826.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in r language. *Bioinformatics 20*, 289–290.

Pirie, M. D., M. P. B. Vargas, M. Botermans, F. T. Bakker, and L. W. Chatrou (2007). Ancient paralogy in the cpDNA trnL-F region in annonaceae: implications for plant molecular systematics. *American Journal of Botany 94*(6), 1003–1016.

Poczai, P. and J. Hyvönen (2011). Identification and characterization of plastid trnF(GAA) pseudogenes in four species of solanum (solanaceae). *Biotechnology Letters 33*(11), 2317–2323.

Provan, J., W. Powell, and P. M. Hollingsworth (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution 16*(3), 142–147.

Redelings, B. and M. Suchard (2007). Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. *BMC Evolutionary Biology 7*(1), 40.

Redelings, B. D. and M. A. Suchard (2005). Joint bayesian estimation of alignment and phylogeny. *Systematic Biology 54*(3), 401 –418.

Rivas, E. and S. Eddy (2008). Probabilistic phylogenetic inference with insertions and deletions. *PLoS Comput Biol 4*(9), e1000172.

Roch, S. (2010). Toward extracting all phylogenetic information from matrices of evolutionary distances. *Science 327*(5971), 1376 –1379.

Rosenberg, M. S. and S. Kumar (2001). Traditional phylogenetic reconstruction methods reconstruct shallow and deep evolutionary relationships equally well. *Molecular Biology and Evolution 18*(9), 1823–1827.

Rosenberg, M. S. and S. Kumar (2003). Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. *Molecular Biology and Evolution 20*(4), 610–621.

Rozenfeld, A. F., S. Arnaud-Haond, E. Hernández-García, V. M. Eguíluz, E. A. Serrão, and C. M. Duarte (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences 105*(48), 18824 –18829.

Saitou, N. and M. Nei (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution 4*(4), 406–425.

Schmickl, R., M. Jorgensen, A. Brysting, and M. Koch (2010). The evolutionary history of the arabidopsis lyrata complex: a hybrid in the amphi-beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evolutionary Biology 10*(1), 98.

Schmickl, R., C. Kiefer, C. Dobeš, and M. A. Koch (2008). Evolution of trnF(GAA) pseudogenes in cruciferous plants. *Plant Systematics and Evolution 282*, 229–240.

Simmons, M., H. Ochoterena, and T. Carr (2001). Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst Biol 50*(3), 454–62.

Simmons, M. P. and H. Ochoterena (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology 49*(2), 369–381.

Stoneberg Holt, S., L. Horova, and P. Bures (2004). Indel patterns of the plastid DNA trnL- trnF region within the genus poa (poaceae). *J Plant Res 117*(5), 393–407.

Stuessy, T. F. (2009). *Plant taxonomy: The systematic evaluation of comparative data* (2 ed.). New York: Columbia University Press.

Subbotin, S. A., D. Sturhan, N. Vovlas, P. Castillo, J. T. Tambe, M. Moens, and J. G. Baldwin (2007). Application of the secondary structure model of rRNA for phylogeny: D2–D3 expansion segments of the LSU gene of plant-parasitic nematodes from the family hoplolaimidae filipjev, 1934. *Molecular Phylogenetics and Evolution 43*(3), 881–890.

Swofford, D. L. (1989). Phylogenetic analysis using parsimony. *Illinois Natural History Survey, Champaign*.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology 17*(5), 1105–1109.

Takahashi, K. and M. Nei (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution 17*(8), 1251–1258.

Talavera, G. and J. Castresana (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology 56*(4), 564 –577.

Tedder, A., P. Hoebe, S. Ansell, and B. Mable (2010). Using chloroplast trnF pseudogenes for phylogeography in arabidopsis lyrata. *Diversity 2*, 653–678.

Turner, B. L. (2006). Taxonomy and nomenclature of the erysimum asperum - e. capitatum complex (brassicaceae). *Phytologia 88*, 279–287.

Vijverberg, K. and K. Bachmann (1999). Molecular evolution of a tandemly repeated trnF(GAA) gene in the chloroplast genomes of microseris (asteraceae) and the use of structural mutations in phylogenetic analyses. *Molecular Biology and Evolution 16*(10), 1329 –1340.

Vogt, L. (2002). Weighting indels as phylogenetic markers of 18S rDNA sequences in diptera and strepsiptera. *Organisms Diversity & Evolution 2*(4), 335–349.

Warwick, S. and I. Al-Shehbaz (2006). Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution 259*(2), 237–248.

Wheeler, W. (1996). Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics 12*(1), 1–9.

Wheeler, W., D. Gladstein, and J. De Laet (2003). POY, phylogeny reconstruction via optimization of DNA and other data version 3.0. 11 (may 6 2003). *American Museum of Natural History*.

Wheeler, W. C., J. Gatesy, and R. DeSalle (1995). Elision: A method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. *Molecular Phylogenetics and Evolution 4*(1), 1–9.

Wittzell, H. (1999). Chloroplast DNA variation and reticulate evolution in sexual and apomictic sections of dandelions. *Molecular Ecology 8*, 2023–2035.

Young, N. and J. Healy (2003). GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 4(1), 6.

SUPPLEMENTARY INFORMATION

Tabla 4.S1: Geographical location and altitude of the 66 *Erysimum* populations studied.

Species	Population code	Longitude	Latitude	Altitude (m a.s.l.)
E. mediohispanicum	Em01	-3.428	37.133	1750
E. mediohispanicum	Em02	-3.431	37.122	2099
E. mediohispanicum	Em03	-3.472	37.085	1654
E. mediohispanicum	Em04	-3.464	37.080	1826
E. mediohispanicum	Em05	-2.555	38.072	1450
E. mediohispanicum	Em06	-2.594	37.987	1448
E. mediohispanicum	Em07	-3.505	37.083	1413
E. mediohispanicum	Em08	-3.431	37.133	1690
E. mediohispanicum	Em09	-3.371	37.130	1280
E. mediohispanicum	Em10	-3.468	37.072	1811
E. mediohispanicum	Em11	-3.473	37.071	1890
E. mediohispanicum	Em12	-2.588	38.009	1732
E. mediohispanicum	Em13	-2.641	38.112	1542
E. mediohispanicum	Em14	-2.683	38.142	1469
E. mediohispanicum	Em15	-2.415	38.581	1384
E. mediohispanicum	Em16	-2.922	37.922	1063
E. mediohispanicum	Em17	-3.424	37.116	2182
E. mediohispanicum	Em18	-2.647	36.918	1215
E. mediohispanicum	Em19	-3.476	37.092	1765
E. mediohispanicum	Em20	-3.479	37.089	1718
E. mediohispanicum	Em21	-3.428	37.134	1723
E. mediohispanicum	Em22	-3.428	37.131	1802
E. mediohispanicum	Em23	-3.426	37.128	1874

Continued on next page

Tabla 4.S1 – continued from previous page

Species	Population code	Longitude	Latitude	Altitude (m a.s.l.)
E. mediohispanicum	Em24	-3.435	37.125	1943
E. mediohispanicum	Em25	-3.434	37.121	2064
E. mediohispanicum	Em26	-2.589	38.159	999
E. mediohispanicum	Em27	-3.395	36.833	1752
E. mediohispanicum	Em28	-3.029	37.051	1834
E. mediohispanicum	Em29	-2.826	37.054	1918
E. mediohispanicum	Em30	-2.804	37.036	1667
E. mediohispanicum	Em31	1.024	42.032	1006
E. mediohispanicum	Em32	0.916	41.990	732
E. mediohispanicum	Em33	-2.933	40.909	983
E. mediohispanicum	Em34	-2.810	41.524	890
E. mediohispanicum	Em35	-2.845	41.766	1126
E. mediohispanicum	Em36	-3.451	37.102	1471
E. mediohispanicum	Em37	-3.467	37.138	1425
E. mediohispanicum	Em38	-2.826	37.044	1783
E. mediohispanicum	Em39	-3.552	37.319	1272
E. mediohispanicum	Em40	-2.822	37.033	1664
E. mediohispanicum	Em41	0.921	41.992	715
E. mediohispanicum	Em42	0.940	41.998	882
E. mediohispanicum	Em43	0.899	41.979	786
E. mediohispanicum	Em44	1.028	42.017	790
E. mediohispanicum	Em45	-2.938	40.862	963
E. mediohispanicum	Em46	-2.960	40.817	1000
E. mediohispanicum	Em47	-2.807	41.326	1063

Continued on next page

Tabla 4.S1 – continued from previous page

Species	Population code	Longitude	Latitude	Altitude (m a.s.l.)
E. mediohispanicum	Em48	-2.456	41.781	1026
E. mediohispanicum	Em49	-2.774	41.616	986
E. mediohispanicum	Em50	-2.973	40.760	1000
E. mediohispanicum	Em51	-1.567	37.869	1396
E. mediohispanicum	Em52	-2.941	37.619	1189
E. mediohispanicum	Em53	-1.767	40.333	1633
E. mediohispanicum	Em54	-0.616	42.321	932
E. mediohispanicum	Em55	-1.637	41.740	806
E. mediohispanicum	Em56	-0.833	40.925	955
E. merxmuelleri	Emx01	-4.980	40.380	1550
E. merxmuelleri	Emx02	-5.245	40.210	1000
E. nevadense	En05	-3.030	37.110	2100
E. nevadense	En11	-3.420	37.110	2200
E. rondae	Er02	-4.000	36.910	1250
E. rondae	Er03	-5.380	36.790	1050
E. ruscinonense	Eru01	2.347	41.800	750
E. ruscinonense	Eru02	2.400	41.830	1000
E. gomezcampoi	Eg002	-0.560	38.660	1100
E. gomezcampoi	Eg003	-0.960	39.300	800

Tabla 4.S2: Description of the 69 trnL-trnF IGS spacer haplotypes found in the 66 studied *Erysimum* populations. For each haplotype, the name, haplogroup, number of motifs, number of pseugenes, sequence length (in bp), number of individual and frequency (in brackets), and number of connections in the percolated network (Fig. 4.8) are shown.

Haplotype code	Haplo-group	Motifs	Pseudo-genes	Length (pb)	Individuals (frequency)	Network connections
H01	I	17	5	430	48 (15.14)	21
H02	III	17	5	425	28 (8.83)	9
H03	II	23	7	574	22 (6.94)	22
H04	II	20	6	497	16 (5.05)	17
H05	I	17	5	430	15 (4.73)	21
H06	III	21	6	517	13 (4.1)	7
H07	IV	20	6	507	10 (3.15)	6
H08	I	17	5	430	10 (3.15)	21
H09	III	18	5	440	6 (1.89)	6
H10	IV	20	6	507	6 (1.89)	13
H11	II	20	6	497	6 (1.89)	18
H12	IV	17	5	430	5 (1.58)	6
H13	IV	17	5	430	5 (1.58)	12
H14	II	20	6	507	5 (1.58)	16
H15	III	14	4	358	5 (1.58)	6
H16	IV	20	6	507	5 (1.58)	15
H17	II	23	7	574	5 (1.58)	19
H18	I	20	6	507	5 (1.58)	15
H19	III	17	5	425	5 (1.58)	9

Continued on next page

Tabla 4.S2 – continued from previous page

Haplotype code	Haplo-group	Motifs	Pseudo-genes	Length (pb)	Individuals (frequency)	Network connections
H20	II	23	7	574	5 (1.58)	21
H21	IV	20	6	507	4 (1.26)	6
H22	V	11	3	286	4 (1.26)	12
H23	II	11	3	286	4 (1.26)	15
H24	II	23	7	574	4 (1.26)	18
H25	I	17	5	430	4 (1.26)	19
H26	IV	17	5	430	4 (1.26)	3
H27	II	23	7	574	4 (1.26)	11
H28	I	17	5	425	3 (0.95)	18
H29	II	23	7	564	3 (0.95)	9
H30	III	14	4	348	3 (0.95)	13
H31	V	17	5	420	3 (0.95)	8
H32	III	20	6	502	2 (0.63)	10
H33	II	11	3	286	2 (0.63)	12
H34	V	14	4	363	2 (0.63)	9
H35	I	25	7	629	2 (0.63)	19
H36	V	20	6	497	2 (0.63)	7
H37	V	17	5	430	2 (0.63)	9
H38	I	17	5	430	2 (0.63)	19
H39	II	7	2	194	2 (0.63)	32
H40	II	26	8	651	2 (0.63)	24
H41	V	20	6	497	2 (0.63)	8
H42	II	17	5	440	2 (0.63)	14
H43	I	11	3	286	2 (0.63)	25

Continued on next page

Tabla 4.S2 – continued from previous page

Haplotype code	Haplo-group	Motifs	Pseudo-genes	Length (pb)	Individuals (frequency)	Network connections
H44	I	17	5	430	2 (0.63)	18
H45	II	23	7	574	2 (0.63)	17
H46	I	26	8	651	1 (0.32)	22
H47	I	17	5	430	1 (0.32)	19
H48	I	11	3	281	1 (0.32)	18
H49	V	8	2	209	1 (0.32)	13
H50	IV	20	6	497	1 (0.32)	9
H51	IV	17	5	430	1 (0.32)	12
H52	V	20	6	507	1 (0.32)	9
H53	I	19	5	485	1 (0.32)	7
H54	I	14	4	353	1 (0.32)	18
H55	IV	14	4	347	1 (0.32)	7
H56	III	11	3	281	1 (0.32)	13
H57	IV	20	6	497	1 (0.32)	4
H58	I	11	3	286	1 (0.32)	30
H59	II	17	5	440	1 (0.32)	14
H60	I	17	5	430	1 (0.32)	18
H61	I	17	5	430	1 (0.32)	18
H62	I	17	5	430	1 (0.32)	16
H63	I	17	5	430	1 (0.32)	13
H64	II	23	7	574	1 (0.32)	19
H65	I	17	5	430	1 (0.32)	19
H66	II	20	6	497	1 (0.32)	17
H67	II	23	7	574	1 (0.32)	17

Continued on next page

Tabla 4.S2 – continued from previous page

Haplotype code	Haplo-group	Motifs	Pseudo-genes	Length (pb)	Individuals (frequency)	Network connections
H68	II	23	7	574	1 (0.32)	15
H69	II	23	7	574	1 (0.32)	6

Tabla 4.S3: Pseudogenes found in the 69 haplotypes.

Pseudogene	Sequence
X03A03T01T29	GTAATGGTAGACATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X03A01T01T29	GTAATGGTAGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X01A01T01T29	GTAATGGTCGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X03A02T01T29	GTAATGGTAGACATAGCTTAATTGCGGAGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X03A01T03T29	GTAATGGTAGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGAGGATGATACTTCAGTAGATGATACCTCA
X06A01T01T29	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X03A04T01T29	GTAATGGTAGACATAGCTTAATTGCGAAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X16A01T01T29	CTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X13A01T01T29	GTAATGGTAGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X02A01T01T29	GTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCAGTAGATGATACCTCA
X01A01T11	GTAATGGTCGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAAGCAAGATGATCCTTCG
X02A01T01	GTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X02A01T21	GTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATAATGATCCTTCG
X03A03T01	GTAATGGTAGACATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X03A01T01	GTAATGGTAGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X02A01T11	GTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAAGCAAGATGATCCTTCG
X03A05T01	GTAATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X03A01T11	GTAATGGTAGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAAGCAAGATGATCCTTCG
X02A06T21	GTAATGGTCGACATAGCTTAACTGCGGAGGACTTAAAATCCTTGTGTC----------ACATGATAATGATCCTTCG
X15A03T01	GTAATGGTAGACATAGGTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X13A01T21	GTAATGGTAGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATAATGATCCTTCG
X02A01T25	GTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATACTGATCCTTCG
X13A01T25	GTAATGGTAGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATACTGATCCTTCG
X18A01T25	GTAATGGTAGACATAGATTAAGTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATACTGATCCTTCG
X03A03T23	GTAATGGTAGACATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTC----------ACATGATAATGATCCCTCG
X03A05T23	GTAATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTC----------ACATGATAATGATCCCTCG
X06A01T11	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAAGCAAGATGATCCTTCG
X06A06T11	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTTAAAATCCTTGTGTCACCATTAGGAAAAGCAAGATGATCCTTCG
X06A01T03	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGAGGATGATACTTCA
X06A01T01	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X02A01T23	GTAATGGTCGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATAATGATCCCTCG
X03A01T23	GTAATGGTAGACATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATAATGATCCCTCG
X15A03T23	GTAATGGTAGACATAGGTTAATTGCGGGGGACTTTAAATCCTTGTGTC----------ACATGATAATGATCCCTCG
X19A01T11	GTAATGGTCGGCATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAAGCAAGATGATCCTTCG
X03A04T01	GTAATGGTAGACATAGCTTAATTGCGAAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X06A01T21	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATAATGATCCTTCG
X06A01T24	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTGAAAATCCTTGTGTC----------ACATGATAATGAT------
X06A06T21	GTAATGGTCGGCATAGCTTAATTGCGGAGGACTTAAAATCCTTGTGTC----------ACATGATAATGATCCTTCG
X13A01T01	GTAATGGTAGACATAGCTTAACTGCGGAGGACTGAAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X05A03T03	GTAATGGTGGACATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGAGGATGATACTTCA
X05A07T03	GTAATGGTGGACATAGCTTAATTGCGGGGGACTTTAAATCCCTGTGTCACCATTAGGAAAACGAGGATGATACTTCA
X05A03T01	GTAATGGTGGACATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X05A10T03	GTAATGGTGGACATAGCTTAATTGCGGGGGACTTTAAATCCTC-----ACCATTAGGAAAACGAGGATGATACTTCA
X05A11T03	GTAATGGTGGACATAGCTTAATTGCGGGGGGCTTTAAATCCTC-----ACCATTAGGAAAACGAGGATGATACTTCA
X11A03T03	GTAATGGTGGCCATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGAGGATGATACTTCA
X14A03T03	TCAATGGTCGGCATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGAGGATGATACTTCA
X08A03T03	GTAATGGTGGACATCGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGAGGATGATACTTCA
X08A03T01	GTAATGGTGGACATCGCTTAATTGCGGGGGACTTTAAATCCTTGTGTCACCATTAGGAAAACGCGGATGATACTTCA
X05A10	GTAATGGTGGACATAGCTTAATTGCGGGGGACTTTAAATCCTC-----
X05A09	GTAATGGTGGACATAGCTTAATTGCGGGGGACTTTAAATCCTCGTGTC
X09A09	GCAATGGTCGGCATCGCTTTTTTGCGGGGGACTTTAAATCCTCGTGTC
X17A09	GTAATGGTGGACATCGCTTTTTTGCGGGGGACTTTAAATCCTCGTGTC
X10A08	TCAACGGTCGGCATCGCTTTTTTGCGGGTGACTTTAAATCCTCGTGTC
X12A08	TCAATGGTCGGCATCGCTTTTTTGCGGGTGACTTTAAATCCTCGTGTC
X14A03	TCAATGGTCGGCATAGCTTAATTGCGGGGGACTTTAAATCCTTGTGTC
X08A09	GTAATGGTGGACATCGCTTAATTGCGGGGGACTTTAAATCCTCGTGTC
X05	GTAATGGTGGACATAGCTTAATTGCG

Figura 4.S1: Saturation plots comparing uncorrected and corrected distances using A) Galtier-Gouy and B) Tamura-Nei models.

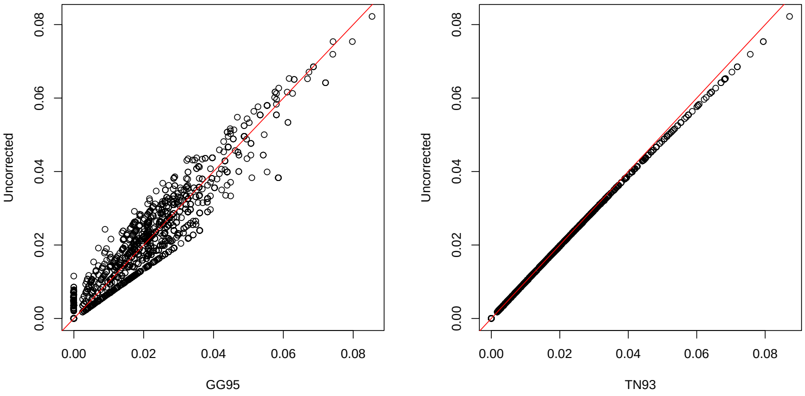
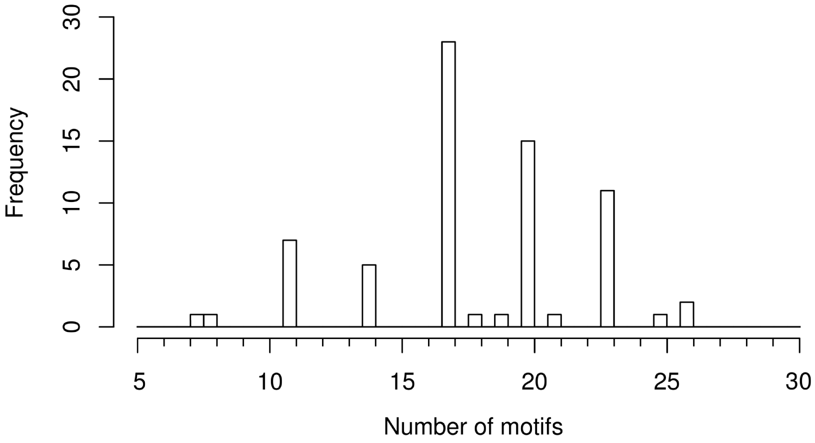
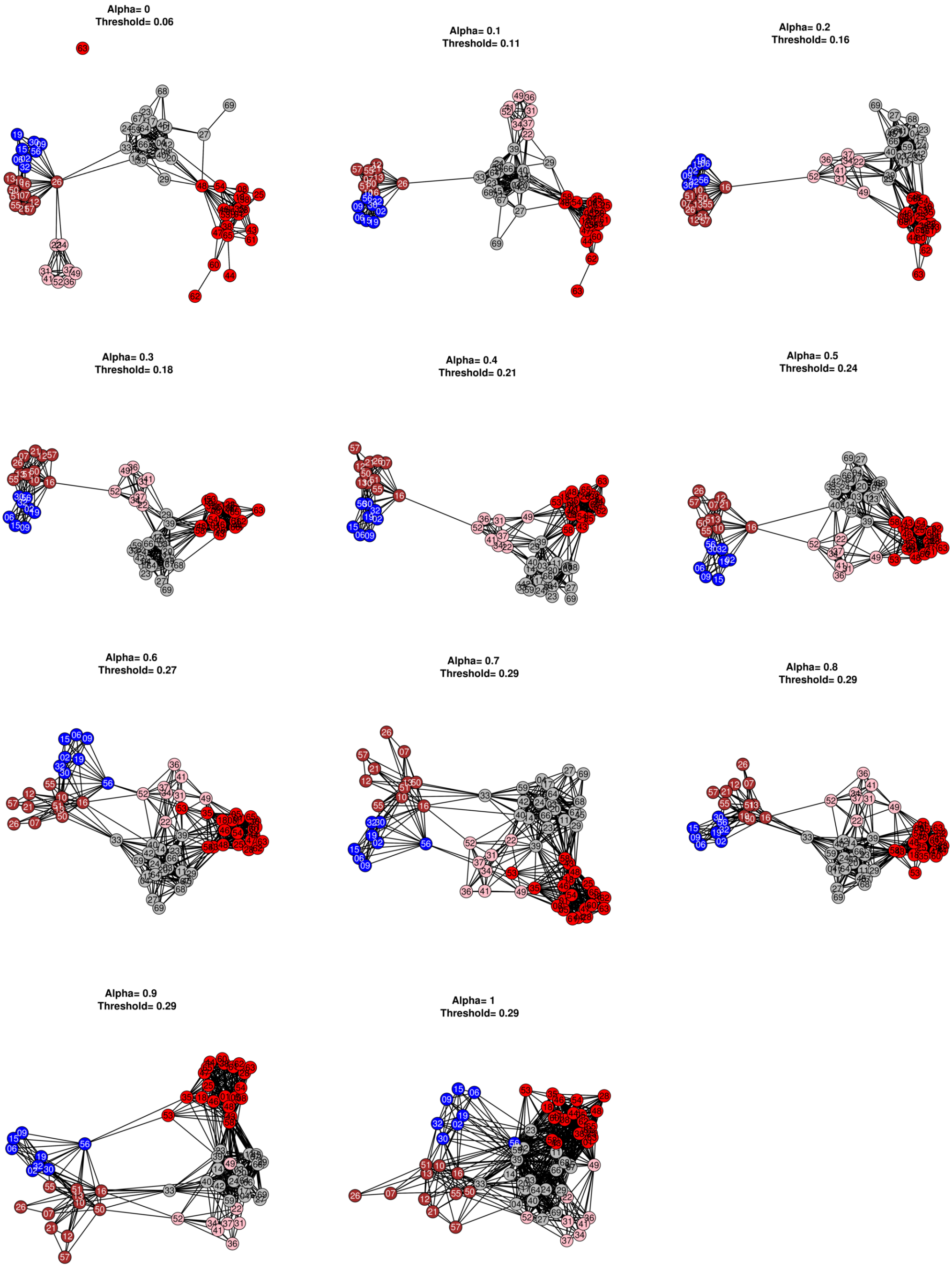


Figura 4.S2: Number of repetitive motifs found in all the studied haplotypes.

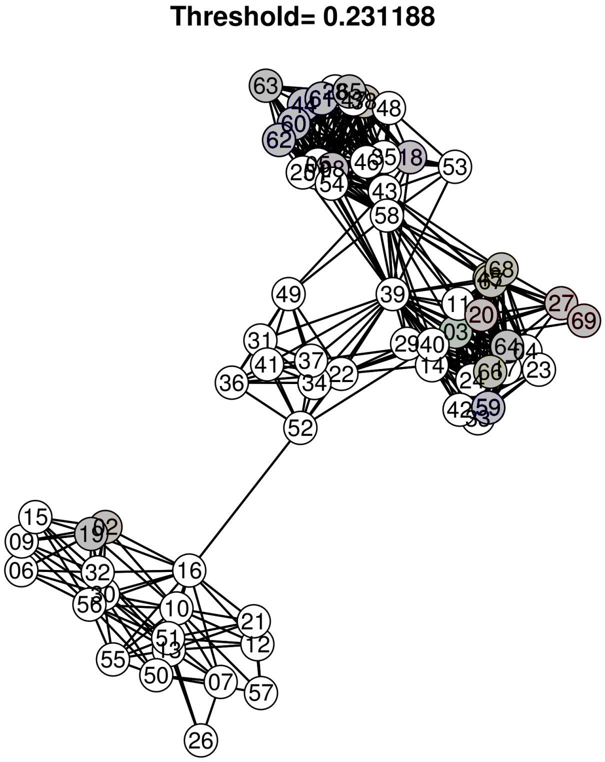



Alpha= 0
Threshold= 0.06
Alpha= 0.1
Threshold= 0.11
Alpha= 0.2
Threshold= 0.16
Alpha= 0.3
Threshold= 0.18
Alpha= 0.4
Threshold= 0.21
Alpha= 0.5
Threshold= 0.24
Alpha= 0.6
Threshold= 0.27
Alpha= 0.7
Threshold= 0.29
Alpha= 0.8
Threshold= 0.29
Alpha= 0.9
Threshold= 0.29
Alpha= 1
Threshold= 0.29

◀ (Previous page)

Figura 4.S3: Percolation networks obtained using different weights for substitution and indel distance matrices. "Alpha" represents indel weights, as they are substitution Alpha-1 weights. The percolation threshold is shown for each alpha value. Colors represent haplogroups (Red = Group I; Grey = Group II; Blue = Group III; Brown = Group IV; Purple = Group V).

Figura 4.S4: Percolation network obtained combining indel and substitution distances (giving equal weight to both matrices; Fig. 4.8), highlighting haplotypes found in species other than *E. mediohispanicum*.


Threshold= 0.231188

Part III

THE EVOLUTIONARY HISTORY OF *ERYSIMUM MEDIOHISPANICUM*

5

GEOGRAPHIC PATTERNS IN GENETIC DIVERSITY AND ITS EFFECTS IN THE DIVERSITY OF POLLINATORS

ABSTRACT

In this chapter, we described phenotypic mean and variation, the genetic diversity, and the pollinator assemblage in a large set of populations of the pollination-generalist plant *E. mediohispanicum*. Our sampling design encompass the entire distribution of the species, which showed differences in ploidy between the southernmost distribution range and the remaining populations. Our results suggest the existence of significant relationship between corolla shape variation and pollinator functional groups diversity, which also were correlated to plant genetic diversity. The studied populations were extremely diverse phenotypically and genotypically and showed significant altitudinal patterns regardless the spatial scale considered. We have also found significant effect of population distances on their phenotypes and genotypes. Population structure suggests the existence of two main linages, one in Sierra Nevada (where probably recent within range gene flow existed) and the other in the rest of the Iberian Peninsula (which showed incipient differentiation between ranges). Our results highlight the important relationship between pollination assemblage, and both genotypic and phenotypic properties of plant populations, even in pollination-generalist plant species.[1]

[1] Authors: A. Jesús Muñoz-Pajares, Mohamed Abdelaziz, Jordi Bosh, José M. Gómez, and Francisco Perfectti

INTRODUCTION

Genetic diversity is a key factor in the evolution and ecology of natural populations (Hughes et al., 2008). In the case of plant species, genetic diversity may be influenced by a variety of factors such as the range of distribution, seed dispersal mechanisms, breeding system, generation time, population effective size, gene flow between populations, and biotic and abiotic selective pressures (Gottlieb, 1977; Hamrick and Godt, 1990, 1996; Nybom, 2004; Aguilar et al., 2008; Pearson et al., 2009). Some of these factors usually vary among populations producing spatial patterns on genetic diversity and differentiation. Genetic differentiation can be geographically structured due not only to spatial variation in the aforementioned factors but also to limited dispersal between populations (Heywood, 1991). In this latter case, a scenario known as isolation by distance arises (Wright, 1943; Slatkin, 1993).

In mountain regions, population genetic differentiation is usually more important among than within contiguous mountain ranges, although long-distance dispersal may also exits (Shafer et al., 2011). Moreover, mountain populations are located in altitudinal gradients, where drastic environmental variation occur at particularly short spatial scales (Wen and Hsiao, 2001; Jump et al., 2006; Gonzalo-Turpin and Hazard, 2009). Therefore, horizontal and vertical environmental gradients may affect plant populations by shaping their genetic diversity distributions (Ohsawa and Ide, 2008).

In animal pollinated plant species, abundance and identity of flower visitors may also play a key role in their population genetic diversity by increasing gene flow and reducing endogamy levels (Rozzi et al., 1997; Cascante et al., 2002). Variation in pollinator abundance, diversity, and composition is commonly found at different spatial scales (Herrera, 1995; Fishbein and Venable, 1996; Boyd, 2004; Irwin et al., 2005; Moeller, 2005, 2006; Kühn et al., 2006; Cosacov et al., 2008;

Espíndola et al., 2011). Consequently, geographic fluctuation in plant-pollinator interactions may be one of the several factors affecting the geographic variation in plant population genotypes and phenotypes (Grant, 1949; Grant and Grant, 1965; Campbell et al., 1997; Anderson and Johnson, 2008).

Pollination depends on a wide variety of plant phenotypic traits (e.g., flower color, corolla morphology, stamen length), which are expected to evolve matching the preferences and morphology of their pollinator species. In pollination specialist plants, this idea is supported by phylogenetic studies showing divergent floral characters in lineages differing in pollinator type (Armbruster, 1993; Wilson et al., 2004; Muchhala, 2006; Alcantara and Lohmann, 2010; Knapp, 2010). However, in pollinator generalist plants, the existence of multiple pollinator preferences and morphologies together with the frequent spatio-temporal fluctuations in the pollinator fauna hinder the possibility of perfect fitting between plant and pollinator (Johnson and Steiner, 2000). In this case, relationships between geographic variation in genetic diversity, phenotype variation and pollinator fauna, although potentially important, has been scarcely explored.

In this chapter we assessed the genetic variation, pollinator assemblage, and phenotypic traits mean and variation in the pollination-generalist species *Erysimum mediohispanicum*. This plant species is a biennial herb endemic to the Iberian Peninsula, visited by more than 150 different insect species, including 25 families and six insect orders (Gómez and Perfectti, 2010). Most of these flower visitors act as effective pollinators due to the open morphology of the flower and the extremely accessibility of the reproductive organs (Gómez et al., 2009). *E. mediohispanicum* is autocompatible, but needs pollinators to obtain the complete fruit set (Gómez, 2003). Pollinators exert significant selection on several *E. mediohispanicum* phenotypic traits (Gómez et al., 2009) and spatial variation in pollinator assembling lead to local adaptation at short spatial scales (Gómez et al., 2009). Contrasting with this accumulated knowledge on phenotype and pollinator variations at short spatial scales, there is no available information about the effect of pollinator variation on *E. mediohispanicum* phenotype through its entire distribution range, encompassing two main regions, one in the north-east and other in the south-east of the Iberian Peninsula (Nieto-Feliner, 1993). *E. mediohispanicum* plants are diploids (2n = 14) in the south region and hypotetraploids (2n = 26) in the north (Nieto-Feliner, 1993), but nothing is known about the genetic structure of *E. mediohispanicum* populations at any spatial scale. The main objective of this chapter is to determine the existence of relationships among genetic diversity, pollinator assemblage, and phenotypic trait diversity and mean values. Prior to analyze these relationships, we have characterized the genetic structure, flower visitor fauna, phenotypic traits, and spatial patterns of variation in a representative number of *E. mediohispanicum* populations located along its whole distribution area.

MATERIAL AND METHODS

Sampling design

We sampled a total of 56 *E. mediohispanicum* populations, encompassing the entire distribution range of the species, including 37 populations from the southern region and 19 populations from the northern region (Fig.5.1; Table 5.1; Table 5.S1). In the northern region, we sampled populations in Aragón, Sierra del Montsec (hereafter, Lérida), and Northern Moorlands (hereafter, Guadalajara-Soria; Fig.5.1). Southern populations were sampled in Sierra Espuña, Sierras de Cazorla, Segura and Guillimona (hereafter, Sierra de Cazorla), and Sierra Nevada (Fig.5.1). Within Sierra Nevada we sampled, among others, the area surrounding the El Dornajo mountain, where several *E. mediohispanicum* populations have been widely studied (e.g., Gómez et al., 2007, 2009,a).

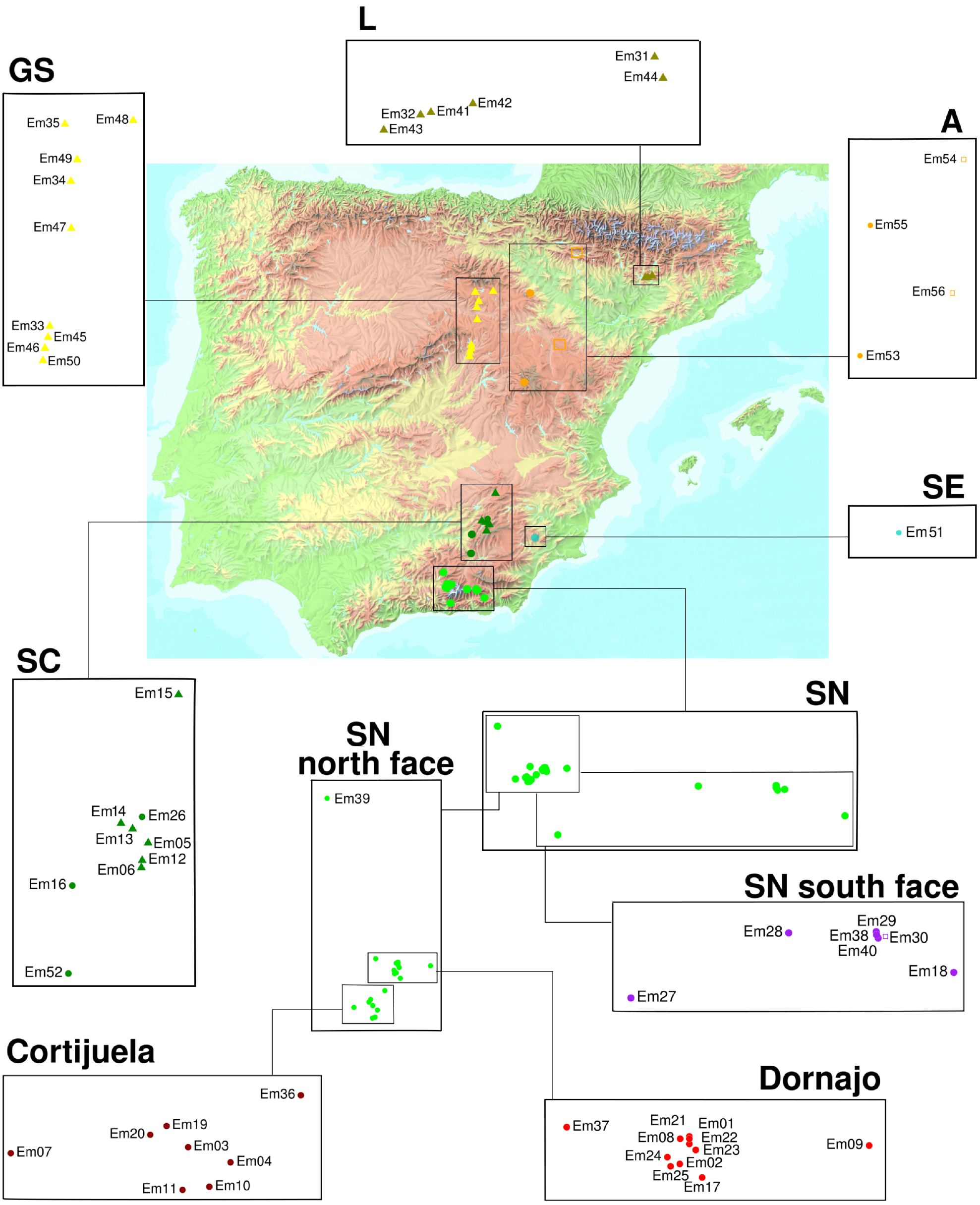


Figura 5.1: Location of the 56 studied Erysimum mediohispanicum populations. Sampling areas are represented with different colors. Circles and triangles represent diploid and polyploid populations, respectively. Empty squares represent populations which ploidy is unknown but were sampled for pollinator visitors and/or phenotypic traits. A= Aragón; L= Lérida; GS= Guadalajara-Soria; SC= Sierra de Cazorla; SE= Sierra Espuña; SN =Sierra Nevada.

Tabla 5.1: Summary of the number of populations composing the phenotype, genotype, and pollinators datasets, grouped by their geographic locations. Number of genotyped diploid populations are shown in brackets.

	Phenotype	Pollinators	Genotype
Aragón	4	0	2 (2)
Lérida	6	5	6 (0)
Guadalajara-Soria	9	4	9 (0)
Northern Total	19	9	17 (2)
Sierra Nevada	27	23	26 (26)
Sierra Cazorla	9	6	9 (3)
Sierra Espuña	1	1	1 (1)
Southern Total	37	30	36 (30)
TOTAL	56	39	53 (32)

This sampling design allowed us to study population differences manifested at large spatial scale (Iberian Peninsula, 56 populations; maximum distance: 758 Km), medium scale (Sierra Nevada, 27 populations; maximum distance: 92 Km), and short scale (area surrounding the Dornajo mountain in the Sierra Nevada north face, 11 populations; maximum distance: 11 Km). We obtained information about phenotypic traits, pollinator fauna, and genetic diversity for each population. Sample sizes for each kind of data are shown in Table 5.1.

Molecular methods

We collected fresh leaf tissue from 15 individuals per population. Leaves were dried and stored in silica gel before DNA extraction, performed using the GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich). For each individual sample, we amplified ten microsatellite loci following the procedure described in Chapter 2. PCR products were diluted to 10 ng/µL and send to MACROGEN (Geumchun-gu, Seoul, Korea; http://www.macrogen.com) for microsatellite fragment separation. We analyzed the electropherograms and called alleles with Peak Scanner Software version 1.0 (Applied Biosystems).

Determination of population ploidy

Depending on the frequency of individuals showing more than two peak alleles in several loci, we classified populations as diploids or polyploids. Because partial genome duplications may be commonly found in plants (Camacho, 2005; and references therein), it is possible to find individuals with multiple allele copies for only a few markers. We considered a population as diploid if we found no individuals with more than two peaks in more than one locus, and as polyploid in any other case. The establishment of these two groups was clear and we found no doubtful populations.

Population genetic structure

The high uncertainty associated to the assessment of allele dosage in polyploid individuals (Nybom, 2004) precludes an accurate estimation of gene frequencies for polyploid populations. For this reason, we have performed all genetic analyses only in diploid populations. We excluded from analyses individuals showing more than 30 % of missing data. Consequently, we used for subsequent analyses 465 individuals belonging to 32 populations. We tested whether each loci and population were in Hardy-Weinberg equilibrium with Arlequin (Evanno et al., 2005), using 1,000,000 steps in a Markov chain after 100,000 burn-in steps. For each population, we estimated the number of non-redundant multilocus genotypes as the number of genotypes showing at least one different allele (excluding missing data). We also estimated the following population genetic parameters, averaging among loci: 1) Mean number of alleles per locus (N_A). 2) Mean allelic richness per locus (R_S), which is an estimate of the number of alleles independent of sample size. Allelic richness was calculated as the probability of sampling the allele *i* at least once among the *2n* genes of a sample. 3) Expected heterozygosity (H_S), which is the proportion of expected heterozygous individuals assuming Hardy-Weinberg equilibrium. Expected heterozygosity was calculated using Nei (1987) estimator. 4) Observed heterozygosity (H_O). We estimated individual heterozygosities as the ratio between the number of heterozygote loci and the number of successfully genotyped loci (excluding missing data) and then, we estimated the mean of the individual heterozygosities per population. 5) The inbreeding coefficient (F_{IS}), which provide information about Hardy-Weinberg equilibrium departures due to either excess or defect of homozygotes (being F_{IS} positive or negative, respectively). It was estimated as the ratio of the difference between the expected and observed heterozygosities to the expected heterozygosity (Nei, 1987). 6) Population differentiation, measured based on pairwise D_{ST} matrices. We used the harmonic mean of D_{ST} values across loci as an estimator of actual differentiation (Jost, 2008) because D_{ST} reflects genetic differentiation better than F_{ST} when markers show more than two alleles per locus (Jost, 2008; Gerlach et al., 2010). To estimate N_A, R_S, H_S, and F_{IS} we used FSTAT v.2.9.3 (Goudet, 1995), while D_{ST}-based genetic distances were computed using SMOGD version 1.2.5 (Crawford, 2010). 7) Selfing rate (S) was estimated using RMES software (David et al., 2007), which is based on the computation on $\hat{g}2$ (an estimator of the two-locus heterozygosity disequilibrium) instead of on F_{IS}. The latter method has been traditionally used, but is more sensitive to null alleles than the former (Jarne and David, 2008).

We inferred the number of clusters of individuals (K) composing the 32 studied diploid populations, and assigned individuals to these genetic groups using STRUCTURE software (Falush et al., 2007), which showed consistent results regardless the model selected. We performed simulations with ten replicates for each K value ranging from 1 to 15. Each run consisted on 50,000 MCMC steps after 20,000 burn-in steps. To detect the optimum value of K, we used the STRUCTURE HARVESTER website (Earl and vonHoldt, 2012), that implement the Evanno method (Evanno et al., 2005).

We visually represented the similitude among all diploid genotyped individuals by performing Principal Components Analysis (PCA) in R using the *stats* package (R Development Core Team, 2009). For that, we estimated genetic distances among individuals using Bruvo distances (Bruvo, 2004) obtained with the R *polysat* package (Clark and Jasieniuk, 2011). We used the first two PCA components as individual coordinates and plotted them in a two-dimensional space. With this method, distances among plotted points are proportional to the original Bruvo genetic distances.

Plant phenotypic traits

We characterized the phenotype of at least 30 individual plants per population or all the reproductive individuals present if the population were composed by less than 30 individuals (see Table 5.S1 for population sample sizes). For each individual we quantified the following phenotypic traits: 1) Stalk height, measured as the height of the tallest flowering stalk. 2) Stalk diameter, quantified as the diameter at the base of the tallest flowering stalk. 3) Flower number. We counted all flowers and flower buds, thus obtaining the total flower production. 4) Corolla diameter, quantified as the distance between the apical edges of two opposite petals. 5) Corolla tube length, measured as the distance between the corolla tube aperture and the base of the sepals. 6) Corolla tube width, defined as the internal space between petals at the top of the corolla tube aperture, and estimated as the difference between corolla diameter minus the length of two opposite petals. Trait 1 was measured using a measuring tape (precision = 0.5 cm). Traits 2 and 4 to 6 were measured using a digital calliper with 0.1 mm resolution. 7) Corolla shape, measured as a multidimensional trait by means of geometric morphometric tools using a landmark-based methodology (Zelditch et al., 2004). For each individual we selected a flower at anthesis and took a digital photo in front view and planar position. We defined 32 co-planar landmarks covering the corolla shape and using midrib, primary and secondary veins and petal extremes and connections as references (see Gómez et al., 2006 for details on landmark positions, and Abdelaziz et al., 2011 for a full justification of landmark homology). Landmarks were captured using tpsDig v1.4 (available at http://life.bio.sunysb.edu/morph/index.html). Using the two-dimensional coordinates of landmarks we computed the generalized orthogonal least-squares Procrustes average using the Generalized Procrustes Analysis (GPA) superposition method. Then, to visualize shape differences, we computed the relative warps (RWs), which are the principal components of the covariance matrix of the partial warp scores and uniform components (Adams et al., 2004), using tpsRelw v1.11 (available at http://life.bio.sunysb.edu/morph/index.html). Our set of 32 landmarks generates 60 RWs, sorted in decreasing order of the proportion of total variance they explain (i.e., RW1 explains a higher proportion variance than any other RW). Shape differences among individuals can be described and analyzed using RW scores (Zelditch et al., 2004; Gómez and Perfectti, 2010). In this study, only the first four RWs explained more than 5 % of the total variation (Table 5.S2). Together, these four RWs explained more than 71 % of the total variance (Table 5.S2). RW1 defines corolla shapes ranging from petals being almost parallel and partially overlapping along a vertical axis, to petals completely separated and diverging along a vertical axis (Fig.5.S1). RW2 defines zygomorphy with either upper or lower petals overlapping. RW3 defines left and right petal asymmetry. RW4 captures the range of petal slimness to roundness (Fig.5.S1). These four RWs were selected to perform subsequent corolla shape analyses.

We log-transformed traits showing non-normal distributions (i.e. number of flowers, stalk diameter, and stalk height; Fig.5.S2) and characterized the phenotype of each population by estimating mean values of the described traits. We also calculated trait standard deviation values as an estimate of the phenotypic variation found within populations.

Flower visitors surveys

E. mediohispanicum flowering starts between mid May and early June depending on population altitude and latitude, and spans until late July in northern and highest populations. To perform pollinator surveys, we visited each population at least twice during the flowering period. Each survey consisted on walking through the population and recording the insects contacting the reproductive organs of flowers. We focused in plant-pollinator interactions and discarded multiple visits of the same insect to different flowers of the same plant. We also avoided following the same flower visitor for more than 30 seconds. During our first surveys, we captured insect specimens to be sent to specialists for determination. Thus, we assembled a reference pollinator collection that allowed us to identify insect species in field during subsequent surveys.

We studied the *E. mediohispanicum* flower visitors by classifying them into functional groups (Fenster et al., 2004; Wilson et al., 2004). Each functional group included insects of similar body and proboscis sizes, foraging behavior, and feeding habits. We established nine functional groups: 1) ants: including nectar-collecting Formicidae; 2) beetles: species collecting nectar and/or pollen Coleoptera; 3) beeflies: nectar-collecting large-tongued Bombyliidae; 4) butterflies: nectar collectors lepidoptera; 5) hoverflies: nectar- and pollen-collecting Syrphidae and short-tongued Bombyliidae; 6) large bees: pollen- and nectar-collecting female bees over 10 mm in body length; 7) small bees: pollen- and nectar-collecting female bees with body lengths lower than 10 mm; 8) wasps: wasps and cleptoparasitic bees collecting only nectar; 9) others: nectar-collecting flies, grasshoppers, and bugs.

We characterized pollinator diversity per population by calculating Hurlbert's PIE (Colwell, 2005), which is the probability that two randomly sampled pollinator individuals pertain to two different species; Species Richness, which is the number of pollinator species found; and the Functional Group Dominance, which is the frequency of the most abundant functional group. We used EstimateS v8.2.0 (http://viceroy.eeb.uconn.edu/estimates/; Colwell, 2005) to compute Hurlbert's PIE and Species Richness values.

Geographic partition of variance

To evaluate the amount of genetic variance maintained among and within- populations, we studied these variances at three different spatial scales: 1) Large spatial scale: Including all diploid populations (maximum distance: 579 Km; altitudinal range: 806 - 2182 m); 2) Medium spatial scale: Including only Sierra Nevada populations (maximum distance: 92 Km; altitudinal range: 1215 - 2182 m); 3) Short spatial scale: Including only populations sampled in the area surrounding El Dornajo Mountain in Sierra Nevada (maximum distance: 11 Km; altitudinal range: 1280 - 2182 m). We partitioned the genetic variance by means of locus-by-locus AMOVAs as implemented in Arlequin (Excoffier et al., 2005), using the Rst-like distances (sum of squared size differences; (Slatkin, 1995)) as distance matrix and 1,000 permutations to test significance. Due to the unbalanced number of diploid populations found among different geographic areas, we did not include "group of populations" as a level in the variation analyses.

To determine the partition of the phenotypic variance at these different spatial scales, we also performed three independent ANOVA analyses with the same setup used for genetic variance for medium and short spacial scales. However, for large spatial scale, we included both diploid and polyploid populations (maximum distance: 758 Km; altitudinal range: 715 - 2182 m). For these ANOVA we used population as random factor, and the phenotypic traits as dependent variables as implemented in JMP v7.0 (SAS institute Inc., Cary, NC).

Spatial analyses

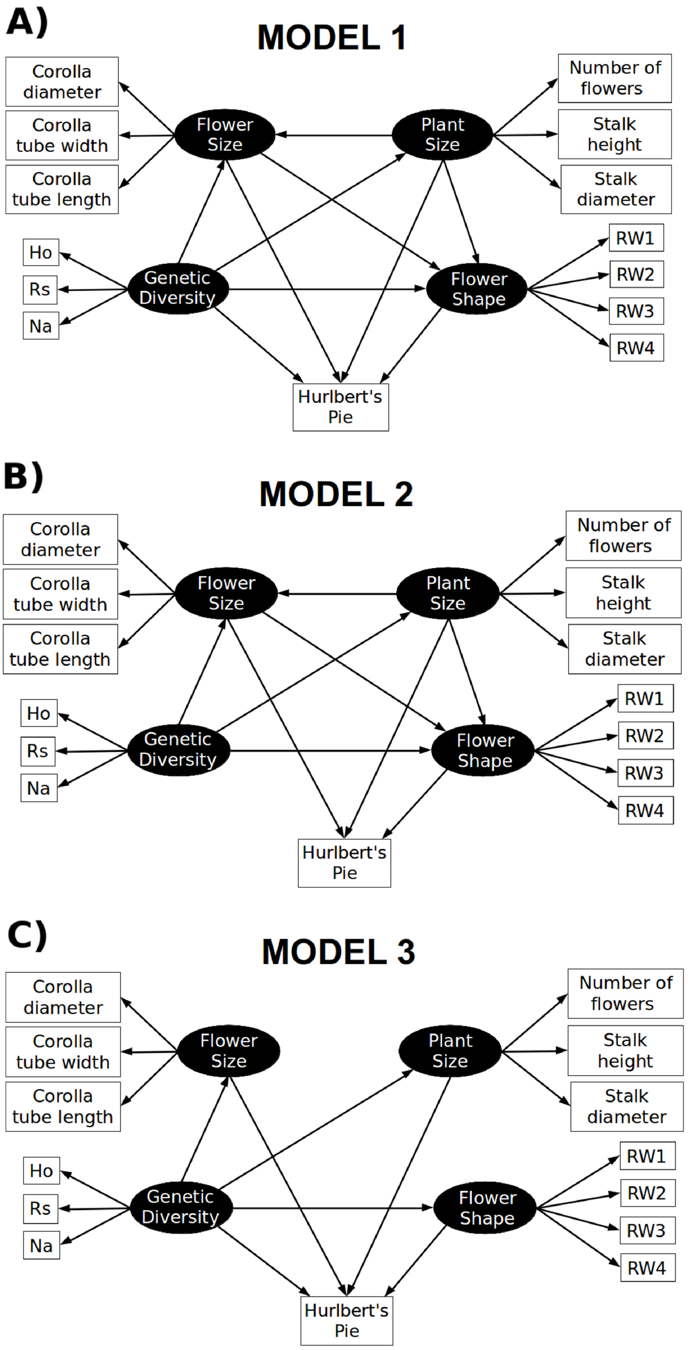
We studied the spatial patterns of variation of the three datasets (i.e., flower visitors, phenotypic traits, and genotypic information) independently, with respect to both population altitude and geographic distances. For the phenotypic dataset, we used population mean values of the traits described above. The flower visitor dataset was represented by Hurlbert's PIE, Species Richness, and Functional Group Dominance, but only populations with more than 100 plant-pollinator recorded interactions were included in the analyses (Table 5.S1). We also included the number of visits per functional group to get a better estimation of the specific pollinator assemblage differences between populations. Finally, the genotypic dataset was tested using R_S, N_A, H_S, H_O, F_{IS}, and D_{ST}.

To test for the existence of significant correlations between population altitude and each of the mentioned population variables we used spatial autoregressive models, specifically Simultaneous Autoregression (SAR) as implemented in SAM v4.0 (Rangel et al., 2010). To test for correlation between geographic and genetic distances (measured as pairwise D_{ST} values), we used Mantel tests (Mantel, 1967) as implemented in the *vegan* package in R (Oksanen et al., 2012).

We also tested correlation between geographic distances and the euclidean pairwise distances for the above mentioned population variables. For corolla-shape, we used the 60 RWs in the computation of the population euclidean distance. For all the described statistical analyses, we applied the sequential Bonferroni correction.

Relationships among pollinator, genetic, and phenotypic diversities

To test for the existence of significant correlations between genetic diversity, pollinator diversity, and plant phenotype, we used the 25 diploid populations with more than 100 plant-pollinator interactions (Table 5.S1). We performed structural equation modelling (SEM) as implemented in the *sem* R package (Fox et al., 2012). We created four latent variables (Fig.5.2), one denoting genetic diversity, and the other representing three different phenotypic group of traits (plant size, flower size, and flower shape). We connected genetic diversity to the genetic parameters R_S, H_O, and N_A. Plant size were connected to log-transformed values for stalk diameter, stalk height, and number of flowers. Flower size was connected to corolla diameter, corolla tube width, and corolla tube length. Flower shape was connected to the first four relative warps (RW1 to RW4; Fig.5.2). We connected plant size, flower size, and flower shape among them and to Hurlbert's Pie values (this last variable acting as a surrogate for pollinator diversity). Finally, we connected genetic diversity to the other three latent variables and to Hurlbert's PIE to obtain the "full model" (Model1, Fig.5.2A). We compared this full model and those models obtained by removing the path between genetic diversity and Hurlbert's PIE (model 2, Fig.5.2B) and all paths connecting directly the three phenotypic latent variables (model 3, Fig.5.2C). Model 2 allowed us to test the existence of relationship between pollinator


A)
MODEL 1
Corolla diameter
Corolla tube width
Corolla tube length
Flower Size
Plant Size
Number of flowers
Stalk height
Stalk diameter
Ho
Rs
Na
Genetic Diversity
Flower Shape
RW1
RW2
RW3
RW4
Hurlbert's Pie
B)
MODEL 2
Corolla diameter
Corolla tube width
Corolla tube length
Flower Size
Plant Size
Number of flowers
Stalk height
Stalk diameter
Ho
Rs
Na
Genetic Diversity
Flower Shape
RW1
RW2
RW3
RW4
Hurlbert's Pie
C)
MODEL 3
Corolla diameter
Corolla tube width
Corolla tube length
Flower Size
Plant Size
Number of flowers
Stalk height
Stalk diameter
Ho
Rs
Na
Genetic Diversity
Flower Shape
RW1
RW2
RW3
RW4
Hurlbert's Pie

◀ (Previous page)

Figura 5.2: Depiction of the three structural equation models tested. A) Full model, B) Model assuming that the only relationship between genetic diversity and Hurlbert's PIE is mediated by the measured phenotypic traits. C) Model assuming no interaction between traits related to plant size, flower size, and corolla shape. For each population, we included in the model the mean values of R_S , H_O , and N_A, and Hurlbert's PIE. For the remaining traits we tested separately the effect of mean and standard values.

diversity and genetic diversity not mediated by the measured phenotypic traits. Model 3 allowed us to explore the existence of integration among phenotypic trait groups. We contrasted the models using a likelihood ratio test (Burnham and Anderson, 2002) and the most appropriate model was selected according to the smaller Bayesian Information Criteria (BIC) value. Because we were interested on studying the relationships among *E. mediohispanicum* genetic diversity, pollinators diversity, and both phenotypic mean and variance values, we performed two independent SEM analyses, one connecting the three phenotypic latent variables with mean trait values and the other with trait standard deviations.

RESULTS

Population ploidy

Of the 53 populations analyzed genetically, 32 populations (479 individuals) were diploids and 21 populations were composed of polyploid individuals (305 individuals) (Table 5.S1). There was a clear difference in ploidy between regions. Whereas 30 out of the 36 southern populations were diploid (all the Sierra Nevada populations, three out of the nine Cazorla populations, and the Sierra Espuña population), only two out of the 17 northern populations were diploid (Fig.5.1).

	All diploid	Sierra Nevada	Dornajo
No-HWE	2.2 ±1.0	2.1 ±1.0	2.3 ±0.6
Monomorphic	0.2 ±0.5	0.2 ±0.5	0.3 ±0.6
R_S	5.341 ±1.248	5.443 ±1.359	5.507 ±1.645
H_S	0.658 ±0.045	0.665 ±0.044	0.683 ±0.025
H_O	0.550 ±0.054	0.556 ±0.054	0.544 ±0.047
N_A	5.867 ±0.834	6.062 ±0.796	6.246 ±0.818
F_{IS}	0.163 ±0.065	0.164 ±0.065	0.201 ±0.057
S	0.042 ±0.068	0.055 ±0.053	0.044 ±0.09

Tabla 5.2: Summary of population genetic diversities for large (all diploid populations), medium (Sierra Nevada), and short (Dornajo) spatial scales. Mean and standard deviation values for allelic richness (R_S), expected heterozygosity (H_S), observed heterozygosity (H_O), number of alleles (N_A), inbreeding coefficient (F_{IS}), and selfing coefficient (S). Number of loci in Hardy-Weinberg disequilibrium and frequency of monomorphic loci are also included.

Population genetic structure and differentiation

Allelic richness and number of alleles per locus showed similar values in the studied diploid populations (R_S = 5.34 ± 1.25; range = 1.57 - 6.84; N_A = 5.87 ± 0.83; range = 4.00 - 7.10; Table 5.2). These populations showed an average of two microsatellite loci deviating from Hardy-Weinberg Equilibrium (HWE) (Table 5.2; Table 5.S3), ranging between a minimum of zero loci (Em03) and a maximum of five loci (Em55). Several loci were found in HWE in most of the populations. This is the case of D10 and D11 that appeared in HWE in 30 out of the 32 diploid populations (Table 5.S3). Contrary, E8 were in HWE in only eight populations (Table 5.S3). Inbreeding coefficients were positive in all populations (F_{IS} = 0.16 ± 0.06; range: 0.06 - 0.27; Table 5.2), indicating that populations showed less heterozygotes than expected under HWE. In fact, only two of the observed HWE departures were due to heterozygote excess (loci C5 in Em51 and Em53, Table 5.S3). Significantly higher values were obtained for expected heterozygosity (H_S = 0.66 ± 0.04; range = 0.53 - 0.72) than for observed heterozygosity (H_O = 0.55 ± 0.05; range = 0.41 - 0.64; Table 5.2), (t = 8.7, df = 60, p <0.001). Contrastingly, selfing rate values were low (S = 0.042 ± 0.068; range 0.00 - 0.25; Table 5.2). Among the 465 studied individuals, we only found two identical multilocus genotypes, corresponding to two individuals belonging to the same population (Em36, which also showed the lowest H_O value).

We assessed population differentiation through D_{ST} distances (range 0.0001 - 0.57). While the lowest pairwise D_{ST} values were found between Sierra Nevada populations, the highest values were observed between populations sampled in Sierra Nevada and Sierra de Cazorla. In fact, all the D_{ST} values higher than 0.4 involved one population from Sierra de Cazorla (Table 5.3).

Genotyped individuals were assigned to two clusters according to the Bayesian estimation (Fig.5.5A). One of them included populations from Aragón, Sierra Espuña, and Sierra de Cazorla, while the remaining 26 populations (all of them belonging to Sierra Nevada) encompassed the second group (Fig.5.5A). PCA based on Bruvo distances between individuals also grouped all Sierra Nevada populations together, but separated the remaining populations into three groups related to Sierra de Cazorla, Aragón, and Sierra Espuña (Fig.5.5B). These three groups of individuals showed peripheral, partially overlapped positions regarding Sierra Nevada individuals, and those belonging to Sierra Espuña showed an intermediate position between Sierra de Cazorla and Aragón individuals (Fig.5.5B). At medium and short spatial scales, PCA showed that individuals belonging to different populations presented overlapped positions (Figs.5.5C and 5.5D).

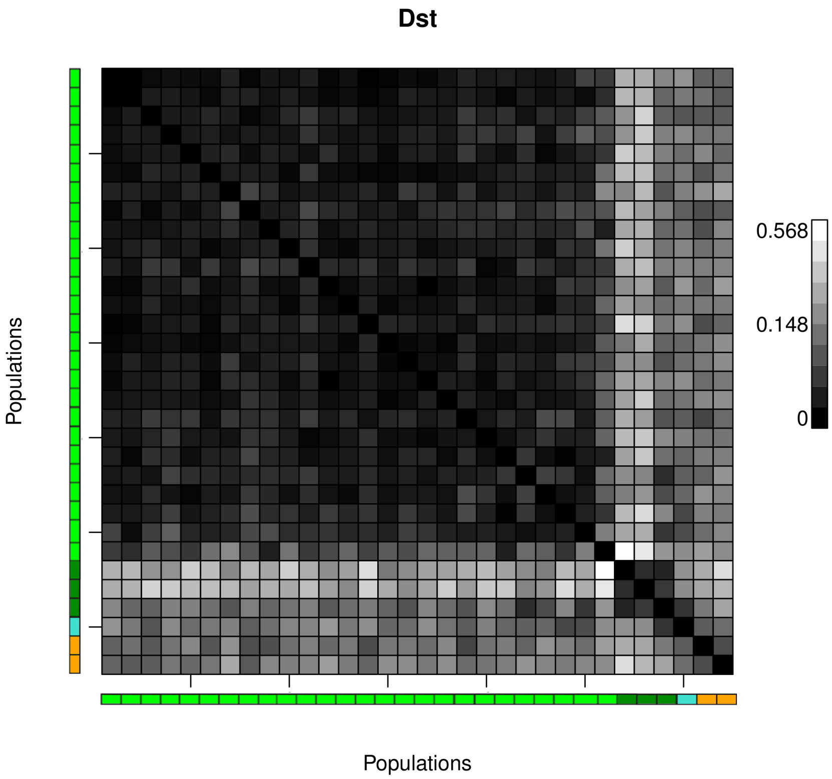
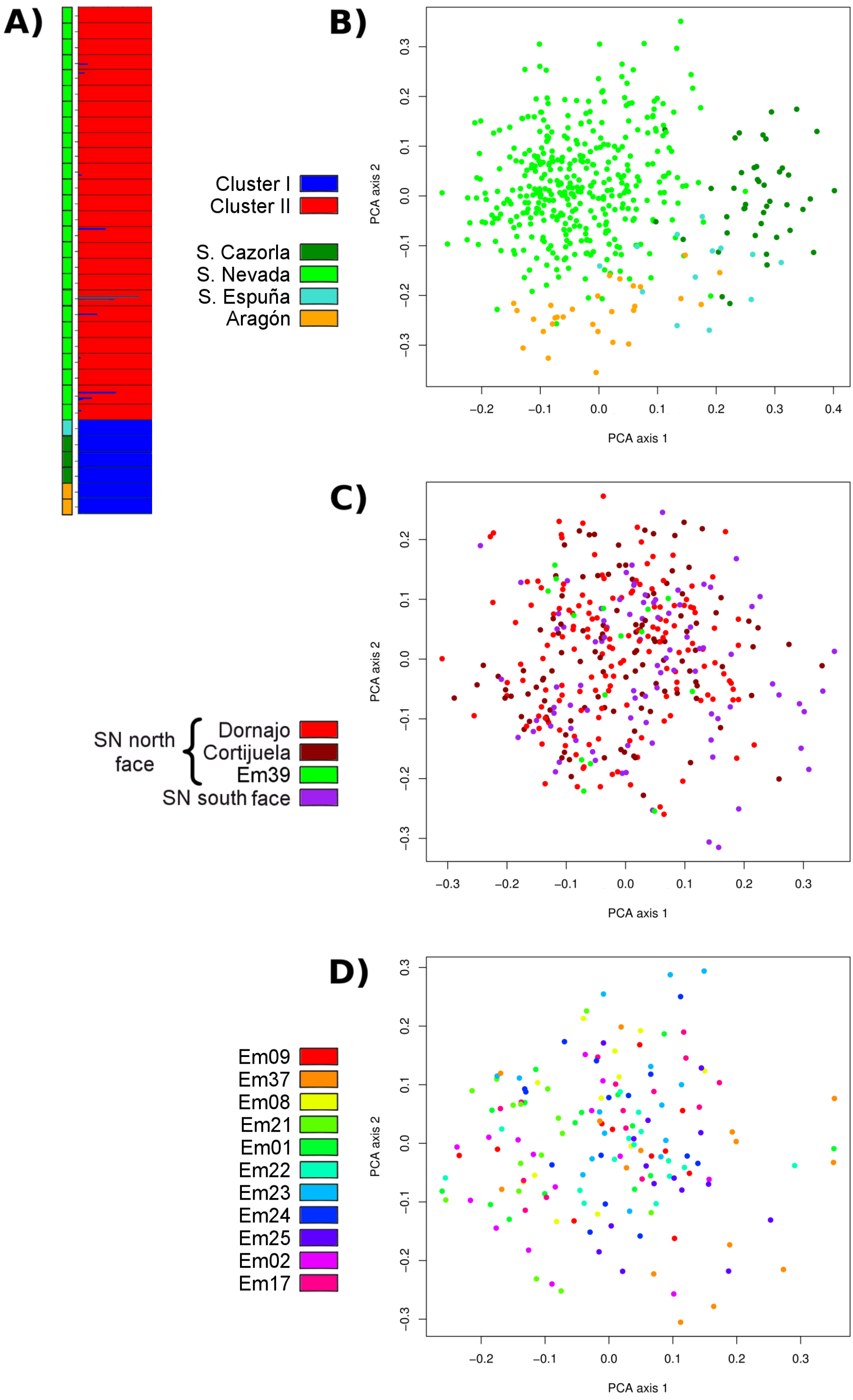


Figura 5.3: Representation of the pairwise D_{ST} matrix. Colors on axis represent population origin (Light green= Sierra Nevada, dark green= Sierra de Cazorla, turquoise= Sierra Espuña, orange= Aragón).

Figura 5.4: Genetic differentiation of diploid *E. mediohispanicum* populations. A) Assignment of individuals to the optimal number of clusters (two) according to Bayesian models. B) Representation of the two principal components of individual Bruvo genetic distances at large scale (including all diploid populations). C) Representation of the two principal components of individual Bruvo genetic distances at medium scale (including only Sierra Nevada populations). D) Representation of the two principal components of individual Bruvo genetic distances at short scale (including only El Dornajo populations).

(Next page) ▶


A)
Cluster I
Cluster II
S. Cazorla
S. Nevada
S. Espuña
Aragón
B)
PCA axis 1
PCA axis 2
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
C)
SN north face
Dornajo
Cortijuela
Em39
SN south face
PCA axis 1
PCA axis 2
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
D)
Em09
Em37
Em08
Em21
Em01
Em22
Em23
Em24
Em25
Em02
Em17
PCA axis 1
PCA axis 2
−0.2
−0.1
0.0
0.1
0.2
0.3

	All	S. Nevada	Dornajo	North	South
Stalk diameter	0.751 ±0.156	0.805 ±0.140	0.762 ±0.126	0.675 ±0.172	0.790 ±0.134
Stalk height	3.616 ±0.315	3.475 ±0.346	3.422 ±0.301	3.739 ±0.242	3.553 ±0.331
Number of flowers	3.798 ±0.402	3.941 ±0.336	3.891 ±0.350	3.688 ±0.497	3.854 ±0.337
Corolla diameter	12.742 ±1.195	12.676 ±1.206	11.813 ±0.676	13.296 ±1.043	12.457 ±1.179
Corolla width	1.281 ±0.453	1.050 ±0.374	0.890 ±0.356	**1.583 ±0.470**	**1.126 ±0.362**
Corolla length	11.431 ±0.893	11.396 ±0.877	10.830 ±0.545	11.725 ±0.907	11.281 ±0.860
RW1	-0.003 ±0.044	-0.010 ±0.034	0.002 ±0.019	-0.018 ±0.043	0.004 ±0.043
RW2	-0.002 ±0.023	-0.004 ±0.021	0.001 ±0.012	0.001 ±0.027	-0.003 ±0.020
RW3	-0.007 ±0.020	0.003 ±0.019	0.012 ±0.015	**-0.022 ±0.015**	**0.001 ±0.018**
RW4	-0.012 ±0.028	0.006 ±0.022	0.017 ±0.020	**-0.034 ±0.021**	**-0.001 ±0.024**

Tabla 5.3: Summary of population phenotypes. For each trait we estimated mean and standard deviation values for large (all sampled populations), medium (Sierra Nevada), and short (El Dornajo) spatial scales, as well as for northern and southern populations separately. Bold values represent significant differences (according to t-test) between northern and southern mean trait values. Logarithmic values are represented for stalk diameter, stalk height, and number of flowers.

Plant phenotype description

Table 5.3 summarizes mean values and standard deviations for each phenotypic trait at different spatial scales and geographic regions. Overall, plants showed in average 70 flowers with a diameter of 12 mm each and corolla tubes width and length of 1.1 and 11 mm, respectively. Stalk diameter and height mean values were 2.3 and 36 cm, respectively. As expected, RW mean values were close to zero (Table5.3).

Whereas plant size traits showed no significant differences between regions, corolla width was significantly smaller in southern than in northern populations (t = 3.71; p < 0.0001). We have also found significant differences in both RW3 and RW4 between northern and southern populations (RW3 p = -4.91; p < 0.0001; RW4 t = -5.18; p < 0.0001).

Composition and diversity of pollinator assemblage

The flower visitor assemblage was extremely diverse. Indeed, more than 250 species from seven orders visited *E. mediohispanicum* flowers when pooling all populations (49 populations and 8389 flower visits). Beetles and ants were the most abundant functional groups while wasps and hoverflies were the least abundant (Table 5.4A; Fig.5.5). Specifically, beetles represented more than 32 % of the total visits and was the main functional group in both northern (29 %) and southern (34 %) populations (Table 5.4A). Ants was the second functional group in abundance, but while in northern populations showed a frequency similar to beetles (26 %), in southern populations represented less than a half of beetle visits (14 %). Wasps represented 1 % of the total visits and was the least abundant functional group in both northern (1.4 %) and southern (0.9 %) populations (Table 5.4A).

Pollinator diversity was high according to the estimated Hurlbert's PIE values (0.8471 ± 0.1275; Table 5.4B; range = 0.3882-0.9654). Most of the studied populations showed Hurlbert's PIE values higher than 0.8 and only one population showed a value lower than 0.5. In average, more than 30 insect species visited each *E. mediohispanicum* population (Species richness: 31 ± 7; Table 5.4B; range = 5-53) and the most abundant species per population represented less than 30 % of the total visits (Functional group dominance: 0.442 ± 0.168; Table 5.4B; range = 0.187-0.900). Despite the existence of variation among populations (Fig.5.5), the pollinator diversity indices were almost identical for northern and southern populations (Table 5.4B).

To assess whether pollinator diversity were independent of pollinator identity, we performed linear regressions between Hurlbert's PIE and the frequency of each functional group. We found a significant negative correlation between Hurlbert's PIE and abundance of beetles (slope = -0.35; $p<0.0001$) and marginal positive tendencies for large bees, wasps, and "others" (5.4C).

A)

	All	Sierra Nevada	Dornajo	North	South
Ants	0.169 ± 0.126	0.149 ± 0.079	0.136 ± 0.080	0.263 ± 0.198	0.141 ± 0.081
Beetles	0.327 ± 0.229	0.365 ± 0.241	0.330 ± 0.240	0.286 ± 0.164	0.339 ± 0.246
Beeflies	0.078 ± 0.116	0.105 ± 0.137	0.041 ± 0.048	0.049 ± 0.078	0.086 ± 0.125
Butterflies	0.072 ± 0.099	0.078 ± 0.080	0.120 ± 0.104	0.034 ± 0.036	0.083 ± 0.109
Hoverflies	0.031 ± 0.062	0.014 ± 0.021	0.031 ± 0.022	**0.074 ± 0.105**	**0.017 ± 0.034**
LargeBees	0.091 ± 0.108	0.094 ± 0.104	0.116 ± 0.086	0.067 ± 0.042	0.098 ± 0.120
SmallBees	0.113 ± 0.133	0.063 ± 0.085	0.047 ± 0.042	0.139 ± 0.109	0.105 ± 0.140
Wasps	0.010 ± 0.020	0.011 ± 0.022	0.022 ± 0.031	0.014 ± 0.020	0.009 ± 0.020
Others	0.111 ± 0.104	0.122 ± 0.097	0.158 ± 0.108	0.074 ± 0.067	0.122 ± 0.112

B)

	All	Sierra Nevada	Dornajo	North	South
Species Richness	30 ± 9	28 ± 9	30 ± 11	35 ± 8	28 ± 9
Hulbert's PIE	0.847 ± 0.128	0.841 ± 0.139	0.851 ± 0.174	0.891 ± 0.050	0.834 ± 0.141
Dominance GF	0.442 ± 0.168	0.440 ± 0.183	0.368 ± 0.212	0.424 ± 0.106	0.447 ± 0.184

C)

	HP regression	SR regression
Ants	-0.028 (0.868)	-0.012 (0.918)
Beetles	**-0.349 ($1.9*10^{-5}$)**	-0.144 (0.02)
Beeflies	0.203 (0.262)	0.002 (0.985)
Butterflies	0.33 (0.116)	0.176 (0.234)
Hoverflies	0.493 (0.143)	0.370 (0.115)
LargeBees	0.346 (0.071)	0.216 (0.109)
SmallBees	0.094 (0.551)	-0.004 (0.973)
Wasps	1.869 (0.077)	1.582 (0.03)
Others	0.421 (0.032)	0.143 (0.308)

Tabla 5.4: Summary of pollinator assemblages. A) Mean and standard deviation of visit frequencies of each functional group at large (all sampled populations), medium (Sierra Nevada), and short (El Dornajo) spatial scales, as well as for northern and southern populations. Significant differences between northern and southern mean values (according to Wilcoxon test) are shown in bold. B) Mean and standard deviation of Species Richness, Hurlbert's PIE, and Functional Group Dominance estimated in the same groups of populations. C) Slope and p-value (in brackets) of lineal regressions performed between either Hurlbert's PIE or Species Richness and the frequency of each functional group. Significant correlations after Bonferroni correction are shown in bold.

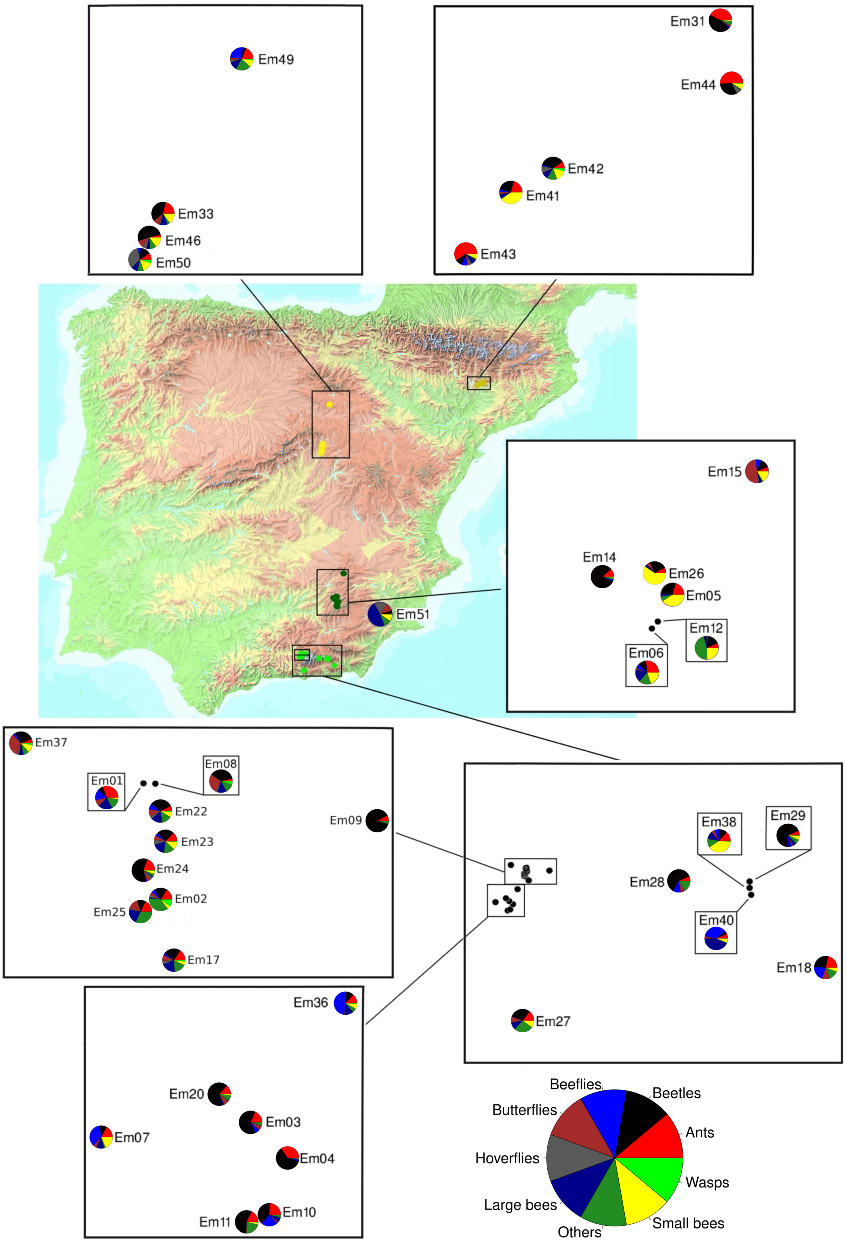


Figura 5.5: Pollinator functional group frequencies per population. Only populations with more than 100 plant-pollinator interactions are represented.

A)

	All diploid	Sierra Nevada	Dornajo
Among populations	10.46	6.34	4.53
Whithin populations	89.54	93.66	95.47
F_{ST}(p-val)	0.105 (<0.00001)	0.063 (<0.00001)	0.045 (<0.00001)

B)

	ALL	S. Nevada	Dornajo
Stalk diameter	14.41	11.45	8.31
Stalk height	40.51	42.51	32.98
Corolla diameter	27.77	29.31	10.29
Cororolla width	17.81	12.60	10.94
Corolla length	25.23	25.82	9.36
Flower number	20.04	15.55	15.13
RW1	10.94	6.51	1.44
RW2	4.95	3.62	1.51
RW3	9.80	8.24	1.71
RW4	20.49	12.40	10.64
MEAN	19.20	16.80	10.23

Tabla 5.5: Variance partition analyses. A) Within and among populations genetic variation components for the three spatial scales: large (all the diploid populations studied), medium (Sierra Nevada), and short (El Dornajo). For each scale, fixation index (FST) and its p-value are also shown. B) Percentage of variance explained among populations for each phenotypic trait at large (all the studied populations), medium (Sierra Nevada), and short (El Dornajo) spatial scales. For phenotypic traits, the residual component of each trait may be estimated as the difference to 100.

Geographic partition of variance

Most of the genetic variance (between 90 and 96 %) was present at within population level independently of the spatial scale assessed, and the among population genetic variance was consistently low (4 to 10 %; Table 5.5). Contrasting with the low variation found among populations, AMOVA results showed significant levels of genetic structure regardless the spatial scale considered (F_{ST} values range from 0.045 to 0.105, p<0.0001 in all cases; Table 5.5A).

At the large spatial scale, an average of 19 % of the total phenotypic variance was found among populations (mean = 19 %; range 5-41 %; Table 5.5B), being lower at the shortest spatial scale (Dornajo populations, mean = 10 %; range = 1-33 %; Table 5.5B). Stalk height was the trait showing the highest among populations variance, while corolla shape components showed the lowest values (Table 5.5B).

Spatial analyses

Allelic richness showed significant positive correlation with population altitude after Bonferroni correction regardless the spatial scale considered ($p < 0.003$; Table 5.6A). The number of alleles per locus also showed a significant correlation with altitude, maintained after Bonferroni correction at large and medium scales ($p = 0.002$ in both cases) and marginal for short spatial scales ($p = 0.033$; Table 5.6A). Stalk height significantly decreased with population altitude at all large, medium, and short spatial scales ($p<0.002$, Table 5.6C) even after Bonferroni correction. Although Hurlbert's PIE marginal positive correlations with altitude were maintained at the three spatial scales ($p<0.085$, Table 5.6B), altitude showed no significant effect on either diversity, abundance, and dominance of insect visiting *E. mediohispanicum* (Table 5.6B).

Genetic distances between populations (measured as D_{ST} harmonic mean values) positively co-varied with geographic distances. Significant results were maintained after Bonferroni correction at large ($r = 0.57$; $p = 0.0026$) and medium scale ($r = 0.37$; $p = 0.0078$, Table 5.7A), and marginally significant for short spatial scale ($r=0.42$, $p=0.029$). R_S also showed a significant spatial autocorrelation pattern at short spatial scale ($r = 0.77$; $p = 0.03$).

Among the phenotypic traits (Table 5.7C), corolla shape showed the most consistent geographic pattern, with significant autocorrela-

A)

	Diploid	Sierra Nevada	Dornajo
R_S	**0.002 (0.003)**	**0.004 (<0.001)**	**0.005 (<0.001)**
N_a	**0.002 (0.002)**	**0.002 (0.002)**	0.002 (0.032)
H_S	<0.001 (0.305)	<0.001 (0.159)	<0.001 (0.673)
H_O	<0.001 (0.118)	<0.001 (0.144)	<0.001 (0.123)
F_{IS}	<0.001 (0.230)	<0.001 (0.491)	<0.001 (0.217)

B)

	All	Sierra Nevada	Dornajo
Species Richness	0.004 (0.170)	0.011 (0.128)	0.012 (0.264)
Hulbert's PIE	<0.001 (0.062)	<0.001 (0.085)	<0.001 (0.069)
Dominance GF	<0.001 (0.310)	<0.001 (0.425)	<0.001 (0.063)

C)

	All	Sierra Nevada	Dornajo
Stalk diameter	<0.001 (0.233)	<0.001 (0.469)	<0.001 (0.759)
Stalk height	**-0.001 (<0.001)**	**-0.001 (<0.0001)**	**-0.001 (0.002)**
Number of flowers	<0.001 (0.528)	<0.001 (0.628)	<0.001 (0.557)
Corolla diameter	<0.001 (0.381)	-0.002 (0.296)	<0.001 (0.997)
Corolla width	<0.001 (0.751)	<0.001 (0.675)	<0.001 (0.864)
Corolla length	<0.001 (0.772)	-0.002 (0.091)	<0.001 (0.524)
RW1	<0.001 (0.476)	<0.001 (0.821)	<0.001 (0.455)
RW2	<0.001 (0.244)	<0.001 (0.020)	<0.001 (0.140)
RW3	<0.001 (0.440)	<0.001 (0.454)	<0.001 (0.405)
RW4	<0.001 (0.055)	<0.001 (0.213)	<0.001 (0.032)

Tabla 5.6: Spatial autoregressive model results for altitude and the different datasets described in this study. For each geographic scale, regression slope and p-value (in brackets) are provided for A) genetic indices (allelic richness, R_S; number of alleles per locus, N_A; expected heterozygosity H_S; observed heterozygosity, H_O; and inbreeding coefficient F_{IS}), B) pollinator assemblage (Species richness, Hurlbert's PIE, and functional group dominance), and C) phenotypic traits (stalk diameter, stalk height, corolla diameter, corolla tube width, corolla tube length, and the first four RW). Bold values represent significant correlations after Bonferroni correction.

tion at large and medium scales (r = 0.24; p = 0.0002 and r = 0.74; p = 0.0001, respectively). Corolla tube width showed significant correlations at large scales (r = 0.26; p = 0.0005), whereas corolla diameter were significantly correlated to geographic distances at medium spatial scales (r = 0.36; p = 0.0004).

Regarding pollinator assemblage, closer populations were more similar in number of visits for each functional group at large scale (r = 0.15; p = 0.04), but not at medium (r = -0.01; p = 0.5) and short (r = 0.10; p = 0.3) spatial scales. Mantel tests showed no significant correlation between other pollinator variables (diversity, abundance, and dominance) and spatial distances at any geographic scale (p>0.01; Table 5.7B).

Relationships among genetic diversity, pollinator diversity and plant phenotype.

For mean phenotypic traits, the full model (model 1; Fig.5.2A) was the best structural model tested (BIC = 250; Fig.5.6A) despite it did not fit the observed variance-covariance matrix (p <0.0001). According to this model, pollinator diversity (measured as Hurlbert's PIE) was significantly affected by corolla shape (see Tables 5.S4 and 5.S5 for details). Specifically, the correlation was negative for RW3 and positive for RW4, meaning that flowers represented by left petal asymmetry (negative RW3) and rounder corollas (positive RW4; Fig.5.S1) were visited by a more diverse assemblage of insect species. The model also showed that larger plants produced larger flowers, and genetic diversity had not significant effect on mean values of any of the three phenotypic components (corolla size, plant size, and corolla shape). Finally, a significant positive relationship exists between pollinator diversity and genetic diversity (Fig.5.6A; Tables 5.S4 and 5.S5).

A)

	Diploid	Sierra Nevada	Dornajo
R_S	-0.09 (0.8202)	0.05 (0.3059)	**0.77 (0.0024)**
N_A	0.09 (0.1818)	0.15 (0.1255)	0.56 (0.0650)
H_S	0.15 (0.1483)	0.08 (0.2428)	0.42 (0.0328)
H_O	0.20 (0.0852)	0.10 (0.2116)	0.49 (0.0663)
F_{IS}	0.07 (0.2500)	0.04 (0.3341)	0.30 (0.0992)
D_{ST}	**0.57 (0.0026)**	**0.37 (0.0078)**	0.42 (0.0289)

B)

	All	Sierra Nevada	Dornajo
Species Reachness	0.01 (0.3048)	-0.13 (0.8365)	0.69 (0.0188)
Hulbert's PIE	-0.08 (0.8646)	-0.15 (0.9102)	0.73 (0.1002)
Dominance	-0.09 (0.8791)	-0.08 (0.7299)	0.63 (0.0929)

C)

	All	Sierra Nevada	Dornajo
Stalk diameter	0.14 (0.0184)	0.05 (0.3109)	0.17 (0.1776)
Plant height	-0.02 (0.5884)	0.25 (0.0369)	0.64 (0.0102)
Number of flowers	0.09 (0.0866)	0.00 (0.4703)	0.13 (0.2237)
Corolla diameter	0.10 (0.0330)	**0.36 (0.0004)**	0.44 (0.0032)
Corolla width	**0.26 (0.0005)**	0.20 (0.0451)	0.12 (0.2984)
Corolla length	0.09 (0.0653)	0.19 (0.0355)	0.35 (0.1297)
Corolla shape	**0.24 (0.0002)**	**0.74 (0.0001)**	0.50 (0.0583)

Tabla 5.7: Mantel test results. For each geographic scale, correlation and p values (in brackets) are provided for A) genetic indices (allelic richness, R_S; number of alleles per locus, N_A; expected heterozygosity H_S; observed heterozygosity, H_O; and inbreeding coefficient F_{IS}; and genetic distances, D_{ST}) and pollinator diversity (Hurlbert's PIE), B) pollinator assemblage (Species richness, Hurlbert's PIE, and functional group dominance), C) phenotypic traits (stalk diameter, stalk height, corolla diameter, corolla tube width, corolla tube length, and the first four RW). Bold values represent significant correlations after Bonferroni correction.

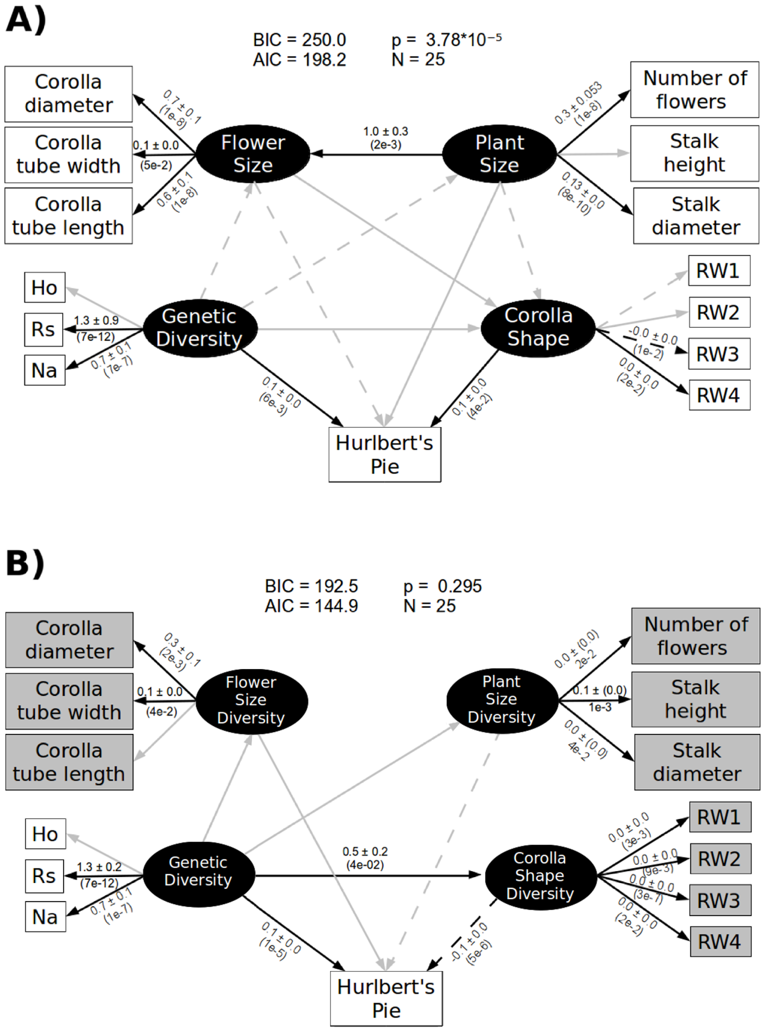


Figura 5.6: A) Best structural model obtained for mean phenotypic trait values. B) Best structural model obtained for standard deviations of phenotypic traits. For each model we also provide coefficients, standard errors, and p-values (in brackets) for significant paths (for full results see Tables 5.S4 and 5.S5). Bayesian and Akaike information criteria, significance of the model, and sample size (N) are also shown. White boxes represent mean values and grey boxes, standard deviations. Black circles represent latent variables. Black, grey, solid and dashed arrows represent significant, no significant, positive and negative relationships, respectively.

For the phenotypic variability, the model showing no relationships among the phenotypic latent variables (model 3) was the best one (BIC = 192.5; Fig.5.6B). This model showed no significant differences with the observed variance-covariance matrix (p = 0.295) and confirmed the significant positive relationship between pollinator and genetic diversities (Fig.5.6B; Tables 5.S4 and 5.S5). Genetic diversity was also positively related to variation in corolla shape, which also was negatively related to pollinator diversity. Consequently, genetic diversity showed two effects on pollinator diversity, one direct positive effect and another, negative, mediated by corolla shape variation (Fig.5.6B).

DISCUSSION

Phenotypic variation

Despite the wide geographic range covered by the studied populations, most of the variation in phenotypic traits was found within population (Table 5.5B). Although high intra-population variation is typically found for neutral characters (Orr, 1998), previous studies on *E. mediohispanicum* have shown that most of the traits measured in this study are not selectively neutral. For example, corolla size and shape are related to pollen and nectar production, and pollinators are able to discriminate between them and show preference patterns (Gómez et al., 2008b,a). Therefore, we have found high intra-population variation in traits under selection. This is somewhat counter-intuitive, because directional selection is expected to deplete variation (Walsh and Blows, 2009). A combination of three possible reasons may explain these results: low heritability of some traits, the generalist pollinator system of *E. mediohispanicum*, and the existence of correlational selection.

Irrespective of the strength of the selection acting on floral traits, low heritability could contribute to maintain high intra-population variability (Conner et al., 2003). To properly discuss the relationship between trait variation and heritability it would be necessary to quantify it on every single population (Falconer, 1981) and this was not possible in our study. Despite these limitations, we have previously demonstrated that several traits showed significant heritability in a subset of the population studied in this Thesis (Gómez et al., 2009b).

High intra-population variation could also be maintained by the existence of a diverse assemblage of flower visitors differing in floral preferences, foraging behavior, and pollinating effectiveness (Aigner, 2001; Gómez, J. M. and Zamora, R., 2006; Siepielski and Benkman, 2010). Due to this pollinator diversity, selective pressures acting on *E. mediohispanicum* are also diverse Gómez et al. (2009) and, consequently, high phenotypic variability can be maintained in natural populations.

Finally, intra-population variation in single traits despite the occurrence of strong selection may also be a consequence of selection acting on a combination of traits via correlational selection and the occurrence of genetic constraints (Hunt et al., 2007; Walsh and Blows, 2009). Under this perspective, we have found in *E. mediohispanicum* that genetic correlation is strong amongst most phenotypic traits (Gómez et al., 2009b), and correlational selection between some traits may also occur (Gómez et al., 2006).

Genetic diversity

The amount of genetic variation within and among populations in a plant species strongly depend on intrinsic factors (such as migration rates or mating system) and extrinsic factors (such as environment or evolutionary history; Gómez-Gómez et al., 2012). In

E. mediohispanicum, almost all the genetic variation was found within populations (Table 5.5A), as is characteristic of insect pollinated plant species showing wide distribution and low self-fertilization levels (Hamrick and Loveless, 1986; Schoen and Brown, 1991; Celka et al., 2010).

All the studied populations showed less heterozygotic individuals than expected and almost all the observed HWE departures per loci and population were due to heterozygote deficit (positive F_{IS} values; Table 5.S3). Inbreeding is the usual cause for heterozygote deficit, but the estimated selfing rate value was also low (averaged $S = 0.042$), suggesting that contribution of autogamous seeds to *E. mediohispanicum* natural populations is negligible. Additionally, we have found high levels of genetic variation in *E. mediohispanicum* populations according to the extremely large number of multilocus genotypes found and the high value of expected heterozygosities.

Pollinator preferences for specific plant phenotypes also may produce an increase in mating frequency among related individual plants (Abdelaziz, 2013 and references therein), although generalist pollination interactions will contribute to reduce this structured network of mating (Gómez et al., 2011). Finally, allelic dropout (that is, the PCR amplification failure of one out of the two alleles present in heterozygous individuals) yield a spurious increase of the homozygote frequency (Wang et al., 2012). Although autogamy seems unlikely, particular studies focused on allelic drop out and population subdivision by assortative mating are needed to elucidate their relative contributions to the observed pattern in *E. mediohispanicum*.

Genetic differentiation and geographic patterns

E. mediohispanicum showed well-differentiated populations according to AMOVA results (significant F_{ST} values, Tabla 5.5A). Nevertheless, the magnitude of F_{ST} found for *E. mediohispanicum* (0.045-0.105) was lower than for other plant species with similar life form (0.31), geographic range (0.28), breading system (0.22), and seed dispersal mode (0.34) (Nybom, 2004).

We found significant correlation between geographic distance and genetic distance (measured as D_{ST}; Table 5.7A), indicating the occurrence of isolation by distance. Using other statistical techniques, such as Bayesian clustering and principal component analyses, we found strong genetic differentiation mostly among areas (i.e., Aragón, Sierra de Cazorla, Sierra Espuña, and Sierra Nevada; Fig.5.4). These results suggest the existence of differentiation associated to separated mountain ranges. It also suggest that within mountain range gene flow is frequent. Altogether, this outcome hinders the ability of discerning different lineages at medium and short spatial scales. Similar patterns have been found in other Iberian genera, for example the genus Armeria (Fuertes-Aguilar et al., 2011).

Genetic distance between populations was not monotonically related with geographic distances. Higher genetic distances were found among populations sampled in Sierra de Cazorla and Sierra Nevada than among populations from Sierra Nevada and Aragón, as expected geographically. This result may be consequence of historical colonization during glacial and interglacial periods (e.g. Kropf et al., 2006, 2008; Kramp et al., 2009), suggesting a late colonization of Sierra de Cazorla from northern populations (Chapters 6 and 7). The fact that, contrary to previously described (Nieto-Feliner, 1993), most of Sierra de Cazorla populations were composed by polyploid individuals (Fig.5.1 and Table 5.S1) may also reinforce this idea, supporting the colonization of these mountains from northern polyploid populations instead of from closer Sierra Nevada diploid populations (Chapters 6 and 7). These results are compatible with the inferred *E. mediohispanicum* colonization paths, suggesting that the species occurred into two isolated focuses during certain geological times, one in the north and other in the south of the Peninsula. The existence of multiple refuges have been described for other plant species within the Iberian Peninsula (Olalde et al., 2002; Gómez and Lunt, 2007; Feliner, 2011; Beatty and Provan, 2012) (Chapters 6 and 7).

We have found significant effect of population geographic distance at large scale not only on genetic distances (Table 5.7A), but also on flower size, and corolla shape traits (tables 5.7C and 5.3). More specific studies are required to establish the main factors influencing these patterns in *E. mediohispanicum*, but limited dispersion, the evolutionary history of populations, and selective pressures (e.g., temperature) have been previously postulated to influence these associations (Hu et al., 2006; Kunin et al., 2009). The lack of spatial patterns on plant size traits at any spatial scale (Tables 5.7C and 5.3) may be due to their higher dependence to micro-environmental conditions which lead to rapid divergence of related populations living in different locations (Pélabon et al., 2011). The existence of spatial patterns is doubtful for the number of pollinator functional group visits ($r = 0.15$; $p = 0.04$) and negligible for pollinator diversity (Tables 5.6B and 5.7B).

The influence of space on *E. mediohispanicum* phenotype and genotype is not limited to geographic distances. Population altitude had a strong effect on plant height in the sense that higher plants were found in populations located at lower altitudes. We have also found a positive correlation between population altitude and genetic diversity (measured as R_S and N_A). In both cases, significant correlations found at large geographic scale were maintained at medium and short spatial scales (Table 5.7). Consequently, population altitude seems to affect both genetic diversity and mean phenotypic traits more consistently (i.e. regardless the spatial scale considered) than geographic distances.

The effect of population altitude on both phenotypic traits and genetic diversity has been widely studied. Whereas a negative correlation between plant height and population altitude has been frequently found in different species and environments (Moles et al., 2009), the effect of population altitude on genetic diversity is highly system dependent. Four almost equally represented patterns have been reported (Ohsawa and Ide, 2008), including the lack of any significant correlation, the existence of either positive or negative correlations, and the existence of higher diversities in medium than in extreme altitudes. According to Ohsawa and Ide (2008), an increase in genetic diversity with elevation is more frequent in herbaceous species (such as *E. mediohispanicum*). These authors also reviewed possible factors accounting for altitudinal gradients, such as high mutation rate at high altitude due to higher levels of ultraviolet-B radiation (Li et al., 1997), the decrease of human activities with altitude (Wen and Hsiao, 2001), the existence of interspecific hybridization (Maghuly et al., 2006), the particular history of colonization (Ohsawa et al., 2007) or adaptation to different environments (Gämperle and Schneller, 2002). Specific studies are also needed to account for the importance of these factors in the observed altitudinal patterns at different spatial scales in *E. mediohispanicum*.

Is genetic diversity mediating the diversity of interactions?

Our resuts suggest that plant phenotypic trait variation is more important than phenotypic trait mean values to explain the pollinator diversity in a given population (see best models p-value, Fig.5.6). We have also found that populations with larger phenotypic variation were more diverse in floral visitors. Additionally, populations showing more genetic diversity (measured using neutral markers) were also more variable in corolla shape and visited by a more diverse pollinator assemblage (Fig.5.6).

Remarkably, we have also found that populations with lower corolla shape variation are more diverse in pollinators (measured as Hurlbert's PIE) than populations showing more corolla shape variation (Fig.5.6B). This is a surprising result because it could be expected that increasing the variation of corolla shape will also increase the probability of interaction with pollinators showing different corolla shape preferences (Aigner, 2001; Gómez, J. M. and Zamora, R., 2006; Gómez et al., 2008b,a). However, the observed negative correlation between beetles and Hurlbert's PIE (Table 5.4C) may produce a decrease in pollinator diversity as shape diversity increases. This negative correlation is consequence of the particular foraging pattern of beetles, which usually spend long time in the same flower, dissuading other pollinator species to visit these flowers (authors' personal observation). These findings, allow us to hypothesize that populations with more variation in corolla shape probably attract more pollinators, including more beetles which reduce the visitation frequency of other functional groups. According to that, an increase on corolla diversity may result in the observed increase of beetles (which is the most abundant functional group with more than 30 % of total visits) and the consequent decrease of Hurlbert's PIE values. Our SEM analyses did not allow us to test this hypothesis because of the high number of variables needed given the actual sample size (N=25). Poor models are yielded when the number of visits per functional group is connected to a latent variable (called Pollinator Identity) for both mean (BIC=429; $p < 0.0001$) and standard deviation (BIC=400; $p < 0.0001$) values of plant phenotypic traits (Fig.5.S4A and 5.S4B, respectively). Over-parametrization is evident when the number of variables is reduced (Fig 5.S4C), by decreasing BIC and increasing p-values (compare Fig.5.S4B and 5.S4C). For this reason, we performed SEM analyses using the minimum number of studied variables yielding the maximum amount of information.

Taken together, our data suggest the influence of genotypic and phenotypic variation on the diversity of insects visiting *E. mediohispanicum* flowers. Corolla shape seems to be more important than flower size and plant size traits for this interaction, which is complex due to the large number of pollinators visiting this species and the putative negative interactions between some functional groups. Our findings highlight the importance of phenotypic diversity on pollinator variability, even in plant species with generalist pollination systems.

ACKNOWLEDGMENTS

We thank specialists who kindly identified pollinator specimens: M. A. Alonso Zarazaga, R. Constantin, S. Fernandez Gayubo, M. Goula, F. Gusenleitner, J. Hava, P. Leblanc, M. A. Marcos, J. C. Otero, A. Sanchez Ruiz, A. Sanchez Terron, M. Schwarz, A. Tinaut, F. Vallhonrat, and D. Ventura. We also thank permissions granted by Junta de Andalucía, Junta de Castilla la Mancha, Junta de Castilla-León, Gobierno de Aragón, and Generalitat de Catalunya to collect samples from different National and Natural Parks. We are grateful to M. B. Herrador for her help in microsatellite amplifications and M. Castiglione, P. Biella, H. Barril-Graells, M. Berbel and J. Lorite for their help in field works.

REFERENCES

Abdelaziz, M. (2013). *How species are evolutionary maintained? Pollinator-mediated divergence and hybridization in Erysimum mediohispanicum and Erysimum nevadense*. PhD. Thesis.

Abdelaziz, M., J. Lorite, A. J. Muñoz-Pajares, M. B. Herrador, F. Perfectti, and J. M. Gómez (2011, June). Using complementary techniques to distinguish cryptic species: A new erysimum (brassicaceae) species from north africa. *American Journal of Botany 98*(6), 1049 – 1060.

Adams, D. C., F. J. Rohlf, and D. E. Slice (2004). Geometric morphometrics: Ten years of progress following the revolution. *Italian Journal of Zoology 71*(1), 5–16.

Aguilar, R., M. Quesada, L. Ashworth, Y. Herrerias-Diego, and J. Lobo (2008). Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology 17*(24), 5177–5188.

Aigner, P. A. (2001). Optimality modeling and fitness trade-offs: when should plants become pollinator specialists? *Oikos 95*(1), 177–184.

Alcantara, S. and L. G. Lohmann (2010). Evolution of floral morphology and pollination system in bignonieae (bignoniaceae). *American Journal of Botany 97*(5), 782 –796.

Anderson, B. and S. D. Johnson (2008). The geographical mosaic of coevolution in a plant–pollinator mutualism. *Evolution 62*(1), 220–225.

Armbruster, S. W. (1993). Evolution of plant pollination systems: Hypotheses and tests with the neotropical vine dalechampia. *Evolution 47*(5), 1480–1505.

Beatty, G. E. and J. Provan (2012). Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the lusitanian plant species daboecia cantabrica (ericaceae). *Journal of Biogeography*, 335–344.

Boyd, A. E. (2004). Breeding system of macromeria viridiflora (boraginaceae) and geographic variation in pollinator assemblages. *American Journal of Botany 91*(11), 1809 –1813.

Bruvo, R. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology 13*(7), 2101–2106.

Burnham, K. and D. Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

Camacho, J. (2005). B chromosomes. In T. Gregory (Ed.), *The evolution of the genome edited by TR Gregory*.

Campbell, D., N. Waser, and E. Melendez-Ackerman (1997). Analyzing pollinator-mediated selection in a plant hybrid zone: Hummingbird visitation patterns on three spatial scales. *The American Naturalist 149*(2), 295–315.

Cascante, A., M. Quesada, J. J. Lobo, and E. A. Fuchs (2002). Effects of dry tropical forest fragmentation on the reproductive success and genetic structure of the tree samanea saman. *Conservation Biology 16*(1), 137–147.

Celka, Z., K. Buczkowska, A. Baczkiewicz, and M. Drapikowska (2010). Genetic differentiation among geographically close populations of malva alcea. *Acta Biologica Cracoviensia Series Botanica 52*(2), 32–41.

Clark, L. and M. Jasieniuk (2011). polysat: an r package for polyploid microsatellite analysis. *Molecular Ecology Resources 11*(3), 562–566.

Colwell, R. (2005). EstimateS: statistical estimation of species richness and shared species from samples, version 7.5.

Conner, J. K., R. Franks, and C. Stewart (2003). Expression of additive genetic variances and covariances for wild radish floral traits: comparison between field and greenhouse environments. *Evolution 57*(3), 487–495.

Cosacov, A., J. Nattero, and A. A. Cocucci (2008). Variation of pollinator assemblages and pollen limitation in a locally specialized system: The oil-producing nierembergia linariifolia (solanaceae). *Annals of Botany 102*(5), 723 –734.

Crawford, N. (2010). smogd: software for the measurement of genetic diversity. *Molecular Ecology Resources 10*(3), 556–557.

David, P., B. Pujol, F. Viard, V. Castella, and J. Goudet (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology 16*(12), 2474–2487.

Earl, D. and B. vonHoldt (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conservation Genetics Resources 4*(2), 359–361.

Espíndola, A., L. Pellissier, and N. Alvarez (2011). Variation in the proportion of flower visitors of arum maculatum along its distributional range in relation with community-based climatic niche analyses. *Oikos 120*(5), 728–734.

Evanno, G., S. Regnaut, and J. Goudet (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology 14*(8), 2611–2620.

Excoffier, L., G. Laval, and S. Schneider (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online 1*, 47–50.

Falconer, D. S. (1981). *Introduction to quantitative genetics.*

Falush, D., M. Stephens, and J. Pritchard (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes 7*(4), 574–578.

Feliner, G. N. (2011). Southern european glacial refugia: A tale of tales. *Taxon 60*(2), 365–372.

Fenster, C., Âw Armbruster, Âpaul Wilson, Âmichele Dudash, and J. Thomson (2004). Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution, and Systematics 35*(1), 375–403.

Fishbein, M. and D. L. Venable (1996). Diversity and temporal change in the effective pollinators of asclepias tuberosa. *Ecology 77*(4), 1061–1073.

Fox, J., Z. Nie, and J. Byrnes (2012). sem: Structural equation models. r package version 3.0-0.

Fuertes-Aguilar, J., B. G. Gutierrez-Larena, and G. N. Nieto-Feliner (2011). Genetic and morphological diversity in armeria (plumbaginaceae) is shaped by glacial cycles in mediterranean refugia. In *Anales del Jardín Botánico de Madrid*, Volume 68, pp. 175–197.

Gämperle, E. and J. J. Schneller (2002). Phenotypic and isozyme variation in cystopteris fragilis (pteridophyta) along an altitudinal gradient in switzerland. *Flora - Morphology, Distribution, Functional Ecology of Plants 197*(3), 203–213.

Gerlach, G., A. Jueterbock, P. Kraemer, J. Deppermann, and P. Harmand (2010). Calculations of population differentiation based on GST and d: forget GST but not all of statistics! *Molecular Ecology 19*(18), 3845–3852.

Gómez, A. and D. Lunt (2007). Refugia within refugia: Patterns of phylogeographic concordance in the iberian peninsula. In *Phylogeography of Southern European Refugia*, pp. 155–188.

Gómez, J., J. Bosch, F. Perfectti, J. Fernández, and M. Abdelaziz (2007). Pollinator diversity affects plant reproduction and recruitment: the tradeoffs of generalization. *Oecologia 153*(3), 597–605.

Gómez, J. M. (2003). Herbivory reduces the strength of pollinator-mediated selection in the mediterranean herb erysimum mediohispanicum: Consequences for plant specialization. *The American Naturalist 162*(2), 242–256. ArticleType: research-article / Full publication date: Aug., 2003 / Copyright Â© 2003 The University of Chicago Press.

Gómez, J. M., M. Abdelaziz, J. P. M. Camacho, A. J. Muñoz-Pajares, and F. Perfectti (2009). Local adaptation and maladaptation to pollinators in a generalist geographic mosaic. *Ecology Letters 12*(7), 672–682.

Gómez, J. M., M. Abdelaziz, J. Muñoz-Pajares, and F. Perfectti (2009a). Heritability and genetic correlation of corolla shape and size in erysimum mediohispanicum. *Evolution 63*(7), 1820–1831.

Gómez, J. M., M. Abdelaziz, J. Muñoz-Pajares, and F. Perfectti (2009b). Heritability and genetic correlation of corolla shape and size in erysimum mediohispanicum. *Evolution 63*(7), 1820–1831.

Gómez, J. M., J. Bosch, F. Perfectti, J. Fernández, M. Abdelaziz, and J. Camacho (2008a). Spatial variation in selection on corolla shape in a generalist plant is promoted by the preference patterns of its local pollinators. *Proceedings of the Royal Society B: Biological Sciences 275*(1648), 2241 –2249.

Gómez, J. M., J. Bosch, F. Perfectti, J. D. Fernández, M. Abdelaziz, and J. P. M. Camacho (2008b). Association between floral traits and rewards in. *Annals of Botany 101*(9), 1413 –1420.

Gómez, J. M. and F. Perfectti (2010). Evolution of complex traits: the case of erysimum corolla shape. *International Journal of Plant Siences 171*, 987–998.

Gómez, J. M., F. Perfectti, J. Bosch, and J. P. M. Camacho (2009). A geographic selection mosaic in a generalized plant–pollinator–herbivore system. *Ecological Monographs 79*(2), 245–263.

Gómez, J. M., F. Perfectti, and J. P. M. Camacho (2006). Natural selection on erysimum mediohispanicum flower shape: Insights into the evolution of zygomorphy. *The American Naturalist 168*(4), 531–545.

Gómez, J. M., F. Perfectti, and P. Jordano (2011). The functional consequences of mutualistic network architecture. *PLoS ONE 6*(1), e16143.

Gómez-Gómez, L., O. Ahrazem, J. M. Herranz, and P. Ferrandis (2012). Genetic characterization and variation within and among populations of anthyllis rupestris coss., and endangered endemism of southern spain. *Biochemical Systematics and Ecology 45*(0), 138–147.

Gómez, J. M. and Zamora, R. (2006). Ecological factors that promote the evolution of generalization in pollination systems. In *Plant-pollinator interactions: from specialization to generalization*, pp. 145–166. University of Chicago Press.

Gonzalo-Turpin, H. and L. Hazard (2009). Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species festuca eskia. *Journal of Ecology 97*(4), 742–751.

Gottlieb, L. D. (1977). Electrophoretic evidence and plant systematics. *Annals of the Missouri Botanical Garden 64*(2), 161–180.

Goudet, J. (1995). FSTAT (version 1.2): A computer program to calculate f-statistics. *Journal of Heredity 86*(6), 485 –486.

Grant, V. (1949). Pollination systems as isolating mechanisms in angiosperms. *Evolution; international journal of organic evolution 3*(1), 82–97.

Grant, V. and K. Grant (1965). *Flower Pollination in the Phlox Family*. Columbia University Press.

Hamrick, J. and M. Loveless (1986). Isozyme variation in tropical trees: procedures and preliminary results. *Biotropica. 18*(3), 201–207.

Hamrick, J. L. and M. J. W. Godt (1990). Allozyme diversity in plant species. In A. Brown, M. Clegg, A. Kahler, and B. Weir (Eds.), *Plant population genetics, breeding, and genetic resources*, pp. 43–63.

Hamrick, J. L. and M. J. W. Godt (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351*(1345), 1291–1298.

Herrera, C. (1995). Microclimate and individual variation in pollinators: Flowering plants are more than their flowers. *Ecology 76*(5), 1516–1524.

Heywood, J. S. (1991). Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics 22*(1), 335–355.

Hu, X.-S., F. He, and S. P. Hubbell (2006). Neutral theory in macroecology and population genetics. *Oikos 113*(3), 548–556.

Hughes, A. R., B. D. Inouye, M. T. J. Johnson, N. Underwood, and M. Vellend (2008). Ecological consequences of genetic diversity. *Ecology Letters 11*(6), 609–623.

Hunt, J., M. W. Blows, F. Zajitschek, M. D. Jennions, and R. Brooks (2007). Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (teleogryllus commodus). *Genetics 177*(2), 875 –880.

Irwin, D. E., S. Bensch, J. H. Irwin, and T. D. Price (2005). Speciation by distance in a ring species. *Science 307*(5708), 414–416.

Jarne, P. and P. David (2008). Quantifying inbreeding in natural populations of hermaphroditic organisms. *Heredity 100*(4), 431–439.

Johnson, S. D. and K. E. Steiner (2000). Generalization versus specialization in plant pollination systems. *Trends in Ecology & Evolution 15*(4), 140–143.

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology 17*(18), 4015–4026.

Jump, A., J. Hunt, J. Martínez-Izquierdo, and J. Peñuelas (2006). Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in fagus sylvatica. *Molecular Ecology 15*(11), 3469–3480.

Knapp, S. (2010). On various contrivances: pollination, phylogeny and flower form in the solanaceae. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*(1539), 449 –460.

Kramp, K., S. Huck, M. Niketić, G. Tomović, and T. Schmitt (2009). Multiple glacial refugia and complex postglacial range shifts of the obligatory woodland plant polygonatum verticillatum (convallariaceae). *Plant Biology 11*(3), 392–404.

Kropf, M., H. P. Comes, and J. W. Kadereit (2006). Long-distance dispersal vs vicariance: the origin and genetic diversity of alpine plants in the spanish sierra nevada. *New Phytologist 172*(1), 169–184.

Kropf, M., H. P. Comes, and J. W. Kadereit (2008). Causes of the genetic architecture of south-west european high mountain disjuncts. *Plant Ecology & Diversity 1*(2), 217–228.

Kühn, I., S. M. Bierman, W. Durka, and S. Klotz (2006). Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. *New Phytologist 172*(1), 127–139.

Kunin, W. E., P. Vergeer, T. Kenta, M. P. Davey, T. Burke, F. I. Woodward, P. Quick, M.-E. Mannarelli, N. S. Watson-Haigh, and R. Butlin (2009). Variation at range margins across multiple spatial scales: environmental temperature, population genetics and metabolomic phenotype. *Proceedings of the Royal Society B: Biological Sciences 276*(1661), 1495–1506.

Li, J., P. Wang, D. Han, F. Chen, L. Deng, and Y. Guo (1997). Mutation effect of high altitude balloon flight on rice and green pepper seeds. *Hang tian yi xue yu yi xue gong cheng = Space medicine & medical engineering 10*(2), 79–83.

Maghuly, F., W. Pinsker, W. Praznik, and S. Fluch (2006). Genetic diversity in managed subpopulations of norway spruce [picea abies (l.) karst.]. *Forest Ecology and Management 222*(1–3), 266–271.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research 27*(2 Part 1), 209–220.

Moeller, D. (2005). Pollinator community structure and sources of spatial variation in plant–pollinator interactions in clarkia xantiana ssp. xantiana. *Oecologia 142*(1), 28–37.

Moeller, D. A. (2006). Geographic structure of pollinator communities, reproductive assurance, and the evolution of self-pollination. *Ecology 87*(6), 1510–1522.

Moles, A. T., D. I. Warton, L. Warman, N. G. Swenson, S. W. Laffan, A. E. Zanne, A. Pitman, F. A. Hemmings, and M. R. Leishman (2009). Global patterns in plant height. *Journal of Ecology 97*(5), 923–932.

Muchhala, N. (2006). The pollination biology of burmeistera (campanulaceae): specialization and syndromes. *American Journal of Botany 93*(8), 1081 –1089.

Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.

Nieto-Feliner, G. (1993). Erysimum. In S. Castroviejo, C. Aedo, C. Gómez-Campo, M. Lainz, P. Monserrat, R. Morales, F. Muñoz-Garmendia, G. Nieto-Feliner, E. Rico, S. Talavera, and L. Villar (Eds.), *Flora Iberica*, Volume 4, Cruciferae-Monotropaceae., pp. 48–76. Madrid: Real Jardín Botánico CSIC.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology 13*(5), 1143–1155.

Ohsawa, T. and Y. Ide (2008). Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography 17*(2), 152–163.

Ohsawa, T., Y. Tsuda, Y. Saito, H. Sawada, and Y. Ide (2007). Altitudinal genetic diversity and differentiation of quercus crispula in the chichibu mountains, central japan. *International journal of plant sciences 168*(3), 333–340.

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. O'Hara, G. Simpson, P. Solymos, M. Stevens, and H. Wagner (2012). vegan: Community ecology package. r package version 2.0-4.

Olalde, M., A. Herran, S. Espinel, and P. Goicoechea (2002). White oaks phylogeography in the iberian peninsula. *Forest ecology and management 156*(1/3), 89–102.

Orr, H. A. (1998). Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics 149*(4), 2099–2104.

Pearson, G. A., A. Lago-Leston, and C. Mota (2009). Frayed at the edges: selective pressure and adaptive response to abiotic stressors are mismatched in low diversity edge populations. *Journal of Ecology 97*(3), 450–462.

Pélabon, C., W. S. Armbruster, and T. F. Hansen (2011). Experimental evidence for the berg hypothesis: vegetative traits are more sensitive than pollination traits to environmental variation. *Functional Ecology 25*(1), 247–257.

R Development Core Team (2009). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rangel, T. F., J. A. F. Diniz-Filho, and L. M. Bini (2010). SAM: a comprehensive application for spatial analysis in macroecology. *Ecography 33*(1), 46–50.

Rozzi, R., M. Arroyo, and J. Armesto (1997). Ecological factors affecting gene flow between populations of anarthrophyllum cumingii (papilionaceae) growing on equatorial- and polar-facing slopes in the andes of central chile. *Plant Ecology 132*(2), 171–179.

Schoen, D. J. and A. H. Brown (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences 88*(10), 4494–4497.

Shafer, A. B. A., S. D. Côté, and D. W. Coltman (2011). Hot spots of genetic diversity descended from multiple pleistocene refugia in an alpine ungulate. *Evolution 65*(1), 125–138.

Siepielski, A. and C. Benkman (2010). Conflicting selection from an antagonist and a mutualist enhances phenotypic variation in a plant. *Evolution; international journal of organic evolution 64*(4), 1120–1128.

Slatkin, M. (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution 47*, 264–279.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics 139*(1), 457–462.

Walsh, B. and M. W. Blows (2009). Abundant genetic variation + strong selection = multivariate genetic constraints: A geometric view of adaptation. *Annu. Rev. Ecol. Evol. Syst. 40*(1), 41–59.

Wang, C., K. B. Schroeder, and N. A. Rosenberg (2012). A maximum likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics*.

Wen, C. and J. Hsiao (2001). Altitudinal genetic differentiation and diversity of taiwan lily (lilium longiflorum var. formosanum; liliaceae) using RAPD markers and morphological characters. *International Journal of Plant Sciences 162*(2), 287–295. ArticleType: research-article / Full publication date: March 2001 / Copyright Â© 2001 The University of Chicago Press.

Wilson, P., M. C. Castellanos, J. N. Hogue, J. D. Thomson, and W. S. Armbruster (2004). A multivariate search for pollination syndromes among penstemons. *Oikos 104*(2), 345–361.

Wright, S. (1943). Isolation by distance. *Genetics 28*(2), 114–138.

Zelditch, M., D. Swiderski, D. Sheets, and W. Fink (2004). *Geometric morphometrics for biologist:a primer*. San Diego: Elsevier Academic Press.

SUPPLEMENTARY INFORMATION

Tabla 5.S1: Description of all populations sampled for this study, providing information about location (sampling area, and region), ploidy (estimated according to the number of alleles per individuals), number of genotyped individuals, number of studied genotyped individuals (diploid individuals showing 30 % or less missing data), number of flower visitor contacts, and number of phenotype measured individuals.

Population code	Sampling area	Locality	Ploidy	# Genotyped individuals	# Analysed individuals	# Pollinator visits	# Individuals Phenotype
Em01	Dornajo	South	Diploid	15	15	222	196
Em02	Dornajo	South	Diploid	15	15	332	193
Em03	Cortijuela	South	Diploid	15	15	263	60
Em04	Cortijuela	South	Diploid	14	14	296	660
Em05	S.Cazorla	South	Polyploid	15	-	213	60
Em06	S.Cazorla	South	Polyploid	15	-	206	60
Em07	Cortijuela	South	Diploid	15	15	274	60
Em08	Dornajo	South	Diploid	15	12	141	198
Em09	Dornajo	South	Diploid	15	11	158	51
Em10	Cortijuela	South	Diploid	15	15	204	60
Em11	Cortijuela	South	Diploid	14	13	108	60
Em12	S.Cazorla	South	Polyploid	15	-	209	60
Em13	S.Cazorla	South	Polyploid	15	-	22	60
Em14	S.Cazorla	South	Polyploid	15	-	202	60
Em15	S.Cazorla	South	Polyploid	15	-	218	60
Em16	S.Cazorla	South	Diploid	15	15	45	35
Em17	Dornajo	South	Diploid	15	15	202	230
Em18	South S.Nevada	South	Diploid	15	15	126	47
Em19	Cortijuela	South	Diploid	15	15	42	61
Em20	Cortijuela	South	Diploid	15	15	209	53
Em21	Dornajo	South	Diploid	15	15	89	164
Em22	Dornajo	South	Diploid	15	15	139	169
Em23	Dornajo	South	Diploid	15	15	125	166
Em24	Dornajo	South	Diploid	15	15	152	150
Em25	Dornajo	South	Diploid	15	15	142	332
Em26	S.Cazorla	South	Diploid	14	14	145	46
Em27	South S.Nevada	South	Diploid	15	14	287	60
Em28	South S.Nevada	South	Diploid	15	15	214	60
Em29	South S.Nevada	South	Diploid	15	15	188	120
Em30	South S.Nevada	South	Unknown	-	-	-	15
Em31	Lérida	North	Polyploid	15	-	212	30
Em32	Lérida	North	Polyploid	10	-	7	12
Em33	Guadalajara-Soria	North	Polyploid	15	-	173	90
Em34	Guadalajara-Soria	North	Polyploid	15	-	72	60
Em35	Guadalajara-Soria	North	Polyploid	15	-	-	60
Em36	Cortijuela	South	Diploid	15	15	216	30
Em37	Dornajo	South	Diploid	15	15	163	30
Em38	South S.Nevada	South	Diploid	15	15	222	90
Em39	S.Nevada	South	Diploid	18	15	64	30
Em40	South S.Nevada	South	Diploid	15	15	324	90
Em41	Lérida	North	Polyploid	15	-	197	30
Em42	Lérida	North	Polyploid	15	-	190	30
Em43	Lérida	North	Polyploid	15	-	248	30
Em44	Lérida	North	Polyploid	11	-	244	30
Em45	Guadalajara-Soria	North	Polyploid	15	-	-	60
Em46	Guadalajara-Soria	North	Polyploid	15	-	234	60
Em47	Guadalajara-Soria	North	Polyploid	14	-	2	60
Em48	Guadalajara-Soria	North	Polyploid	15	-	20	60
Em49	Guadalajara-Soria	North	Polyploid	15	-	223	60
Em50	Guadalajara-Soria	North	Polyploid	15	-	117	65
Em51	S.Espuña	South	Diploid	15	14	219	30
Em52	S.Cazorla	South	Diploid	15	14	69	30
Em53	Aragón	North	Diploid	14	14	-	14
Em54	Aragón	North	Unknown	-	-	-	60
Em55	Aragón	North	Diploid	15	15	-	60
Em56	Aragón	North	Unknown	-	-	-	60

Tabla 5.S2: Percentage of variance explained by relative warps (RWs) defining *E. mediohispanicum* flower shape. Values for individual variance (variance explained by each RW: Ind Var) and cumulative variance (variance explained by summing individual RW variances: Cum Var) are shown.

RW	Ind Var	Cum Var	RW	Ind Var	Cum Var
1	36.13	36.13	31	0.21	96.77
2	17.00	53.13	32	0.21	96.98
3	9.32	62.45	33	0.21	97.19
4	8.73	71.18	34	0.20	97.39
5	3.33	74.51	35	0.17	97.56
6	2.78	77.29	36	0.17	97.73
7	2.60	79.89	37	0.16	97.89
8	1.97	81.86	38	0.16	98.05
9	1.80	83.66	39	0.15	98.2
10	1.60	85.26	40	0.14	98.34
11	1.47	86.73	41	0.13	98.47
12	1.25	87.98	42	0.13	98.6
13	1.00	88.98	43	0.13	98.73
14	0.78	89.76	44	0.11	98.84
15	0.77	90.53	45	0.11	98.95
16	0.72	91.25	46	0.10	99.05
17	0.64	91.89	47	0.10	99.15
18	0.58	92.47	48	0.09	99.24
19	0.45	92.92	49	0.08	99.32
20	0.43	93.35	50	0.09	99.41
21	0.39	93.74	51	0.08	99.49
22	0.38	94.12	52	0.08	99.57
23	0.37	94.49	53	0.07	99.64
24	0.34	94.83	54	0.07	99.71
25	0.33	95.16	55	0.07	99.78
26	0.31	95.47	56	0.06	99.84
27	0.29	95.76	57	0.05	99.89
28	0.29	96.05	58	0.04	99.93
29	0.26	96.31	59	0.04	99.97
30	0.25	96.56	60	0.03	100.00

Tabla 5.S3: Hardy-Weinberg equilibrium departure for each loci and population. Lower heterozygosity than expected is represented by minus symbol (-) while the excess of homozygotes is represented by plus symbol (+). "M" represents monomorphic loci in a given population. The number of loci significantly departured per population and the number of populations significantly departured per locus are also shown.

	C5	D2	D4	D10	D11	E3	E4	E5	E6	E8	Total
Em01		-								-	2
Em02			-					-		-	3
Em03											0
Em04										-	1
Em07			-				-		-	-	4
Em08							-			-	2
Em09		-			M	M				-	2
Em10						M				-	1
Em11		-								-	2
Em16					-					-	2
Em17		-								-	2
Em18							-		-	-	3
Em19							-			-	2
Em20										-	1
Em21		-								-	2
Em22		-					-			-	3
Em23						M				-	1
Em24		-					-			-	3
Em25		-					-	-			3
Em26						-				-	2
Em27	-					-				-	3
Em28		-				-	-				3
Em29			-					-		-	3
Em36	-	-					-	-			4
Em37	-						-				2
Em38		-									1
Em39				-							1
Em40								-			1
Em51	+						-			M	2
Em52						-			-	M	2
Em53					-					-	2
Em55	+			-		M		-		-	4
Total	5	11	3	2	2	4	11	6	3	22	

Tabla 5.S4: Best structural models obtained for phenotypic mean, corresponding to Model1 in Fig.5.S3. For each coefficient we provide its description, the estimated value, the standard error, the z value, and the p value.

Best model for standard deviation trait values (MODEL 1)

Coefficient	Estimate	Std Error	z value	Pr(>|z|)
RsMean <— Div.Gen	1.2836e+00	0.18738956	6.849649	7.4031e-12
HoMean <— Div.Gen	1.5652e-02	0.00949531	1.648346	9.9282e-02
NaMean <— Div.Gen	6.7430e-01	0.13552389	4.975508	6.5077e-07
cor.length <— Fl.Size	5.6723e-01	0.0989424	5.732931	9.8710e-09
coro.width <— Fl.Size	8.9836e-02	0.04581784	1.960720	4.9912e-02
cor.diam <— Fl.Size	7.2734e-01	0.12767288	5.696867	1.2203e-08
logFlowers <— Pl.Size	3.0025e-01	0.05294413	5.670978	1.4198e-08
logHeight <— Pl.Size	2.3896e-02	0.06459794	0.369918	7.1144e-01
logDiamStalk <— Pl.Size	1.2544e-01	0.02038286	6.154189	7.5462e-10
RW1.30 <— Fl.Shape	-7.7694e-03	0.00467414	-1.662206	9.6471e-02
RW2.30 <— Fl.Shape	2.0077e-03	0.0024046	0.834946	4.0375e-01
RW3.30 <— Fl.Shape	-5.3444e-03	0.00214685	-2.489414	1.2795e-02
RW4.30 <— Fl.Shape	1.0492e-02	0.00435031	2.411818	1.5873e-02
Fl.Size <— Div.Gen	-4.6383e-01	0.23726792	-1.954871	5.0598e-02
Pl.Size <— Div.Gen	-2.4536e-01	0.20354233	-1.205459	2.2803e-01
Fl.Shape <— Div.Gen	1.8789e-01	0.35309396	0.532132	5.9463e-01
HP <— Div.Gen	8.5186e-02	0.03114898	2.734793	6.2419e-03
HP <— Fl.Size	-6.2258e-02	0.06373901	-0.976758	3.2869e-01
HP <— Fl.Shape	6.6624e-02	0.03261657	2.042637	4.1088e-02
HP <— Pl.Size	1.0517e-01	0.05651111	1.861080	6.2733e-02
Fl.Size <— Pl.Size	9.8373e-01	0.31369764	3.135931	1.7131e-03
Fl.Shape <— Pl.Size	-7.5038e-01	0.59885673	-1.253013	2.1020e-01
Fl.Shape <— Fl.Size	1.1624e+00	0.59624913	1.949548	5.1230e-02
logDiamStalk <–> logDiamStalk	2.8538e-03	0.00193607	1.474029	1.4047e-01
logHeight <–> logHeight	1.1712e-01	0.03077186	3.806061	1.4120e-04
cor.diam <–> cor.diam	1.6182e-01	0.08630545	1.875003	6.0792e-02
coro.width <–> coro.width	1.2897e-01	0.03412296	3.779467	1.5716e-04
cor.length <–> cor.length	7.9264e-02	0.05017265	1.579815	1.1415e-01
logFlowers <–> logFlowers	2.9816e-02	0.01288192	2.314577	2.0636e-02
RW1.30 <–> RW1.30	9.8497e-04	0.00026918	3.659185	2.5302e-04
RW2.30 <–> RW2.30	3.4181e-04	0.00009043	3.779764	1.5698e-04
RW3.30 <–> RW3.30	6.4185e-05	0.00002546	2.521050	1.1701e-02
RW4.30 <–> RW4.30	4.1190e-04	0.00013273	3.103393	1.9132e-03
RsMean <–> RsMean	1.1570e-02	0.2039514	0.056728	9.5476e-01
HoMean <–> HoMean	2.4502e-03	0.00064456	3.801384	1.4389e-04
NaMean <–> NaMean	2.5470e-01	0.08744653	2.912665	3.5836e-03
HP <–> HP	7.2110e-03	0.00424622	1.698222	8.9466e-02

Tabla 5.S5: Best structural models obtained for phenotypic standard deviation, corresponding to Model3 in Fig.5.S3. For each coefficient we provide its description, the estimated value, the standard error, the z value, and the p value.

Best model for standard deviation trait values (MODEL 3)				
Coefficient	Estimate	Std Error	z value	Pr(>|z|)
RsMean <— Div.Gen	1.2543e+00	1.8307e-01	6.85154	7.3060e-12
HoMean <— Div.Gen	1.6050e-02	9.6161e-03	1.66908	9.5101e-02
NaMean <— Div.Gen	6.9148e-01	1.3145e-01	5.26032	1.4381e-07
cor.length <— Fl.Size	6.9704e-02	3.6325e-02	1.91891	5.4996e-02
coro.width <— Fl.Size	7.4285e-02	3.6605e-02	2.02936	4.2422e-02
cor.diam <— Fl.Size	2.9918e-01	9.5764e-02	3.12412	1.7834e-03
logFlowers <— Pl.Size	4.4807e-02	1.8987e-02	2.35993	1.8279e-02
logHeight <— Pl.Size	7.1399e-02	2.2402e-02	3.18723	1.4364e-03
logDiamStalk <— Pl.Size	1.6953e-02	8.0625e-03	2.10275	3.5488e-02
RW1.30 <— Fl.Shape	6.2034e-03	2.1142e-03	2.93413	3.3448e-03
RW2.30 <— Fl.Shape	3.6622e-03	1.4040e-03	2.60836	9.0978e-03
RW3.30 <— Fl.Shape	4.8765e-03	9.5296e-04	5.11724	3.1003e-07
RW4.30 <— Fl.Shape	3.1721e-03	1.3446e-03	2.35915	1.8317e-02
Fl.Size <— Div.Gen	-3.3221e-02	1.5969e-01	-0.20804	8.3520e-01
Pl.Size <— Div.Gen	3.8493e-01	2.2459e-01	1.71388	8.6550e-02
Fl.Shape <— Div.Gen	4.8753e-01	2.4259e-01	2.00967	4.4466e-02
HP <— Div.Gen	1.2837e-01	2.9650e-02	4.32951	1.4944e-05
HP <— Fl.Size	2.2317e-02	1.7307e-02	1.28945	1.9724e-01
HP <— Fl.Shape	-1.0594e-01	2.3242e-02	-4.55816	5.1602e-06
HP <— Pl.Size	-6.0459e-03	1.5793e-02	-0.38283	7.0185e-01
logDiamStalk <–> logDiamStalk	1.7324e-03	4.7939e-04	3.61372	3.0183e-04
logHeight <–> logHeight	-1.0238e-03	2.9649e-03	-0.34532	7.2985e-01
cor.diam <–> cor.diam	-2.3638e-02	5.5311e-02	-0.42737	6.6911e-01
coro.width <–> coro.width	2.2100e-02	6.6647e-03	3.31589	9.1350e-04
cor.length <–> cor.length	2.3962e-02	6.9008e-03	3.47231	5.1600e-04
logFlowers <–> logFlowers	8.5637e-03	2.5082e-03	3.41432	6.3942e-04
RW1.30 <–> RW1.30	1.1485e-04	3.2125e-05	3.57514	3.5003e-04
RW2.30 <–> RW2.30	5.3055e-05	1.4599e-05	3.63421	2.7883e-04
RW3.30 <–> RW3.30	7.6921e-06	5.3161e-06	1.44695	1.4791e-01
RW4.30 <–> RW4.30	5.0162e-05	1.3664e-05	3.67101	2.4159e-04
RsMean <–> RsMean	8.5737e-02	1.4869e-01	0.57660	5.6421e-01
HoMean <–> HoMean	2.4376e-03	6.4333e-04	3.78901	1.5125e-04
NaMean <–> NaMean	2.3124e-01	7.5625e-02	3.05768	2.2305e-03
HP <–> HP	5.8455e-03	3.2029e-03	1.82505	6.7993e-02

Figura 5.S1: Representation of positive and negative extreme values of the first four relative warps (RWs) defining flower shape for the studied diploid populations.

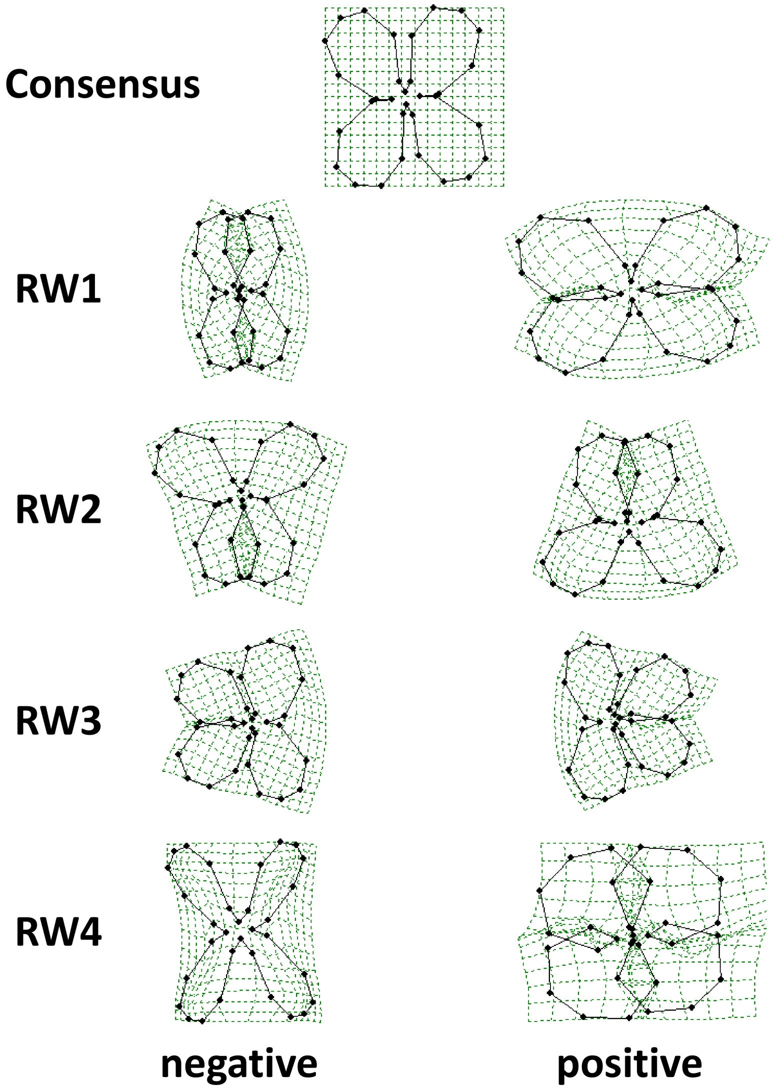


Figura 5.S2: Distribution of values obtained for each trait among all phenotype individuals. According to these results, stalk diameter, stalk height, and number of flowers were log-transformed.

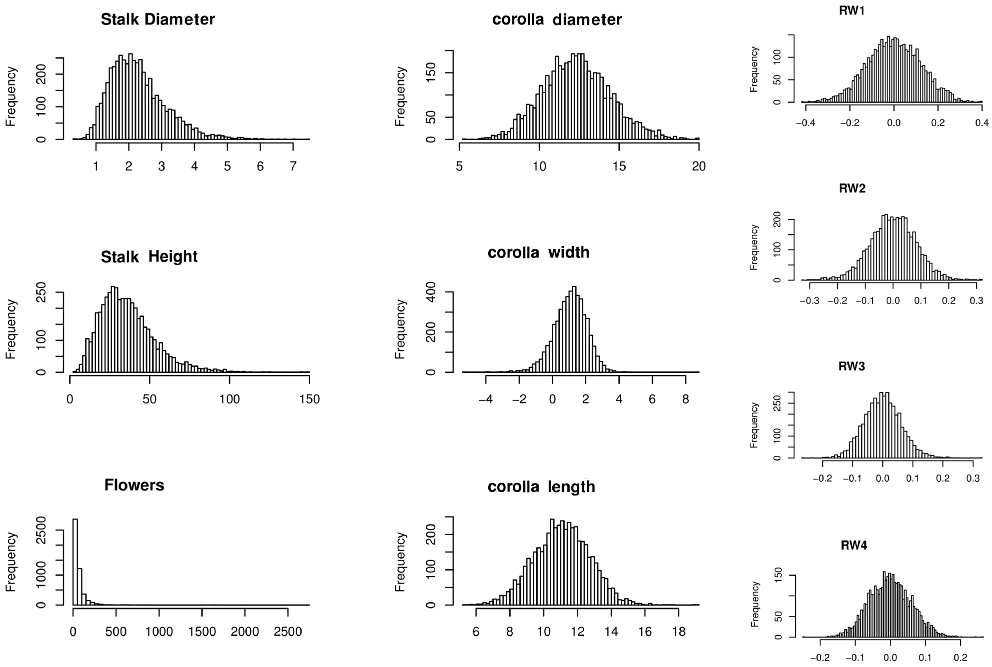
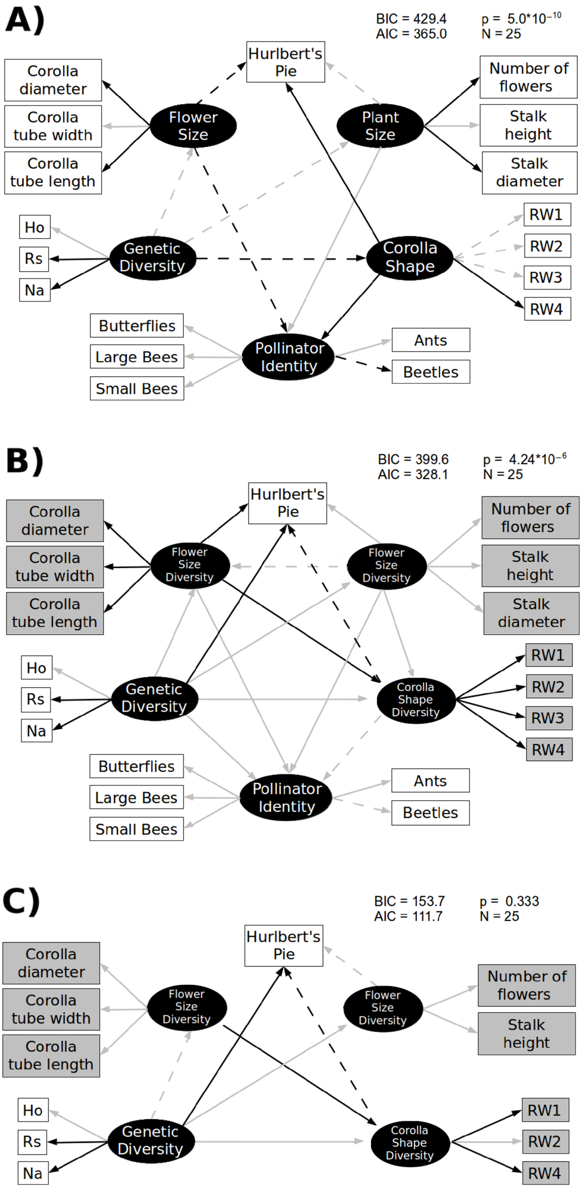


Figura 5.S3: Best structural models obtained for A) mean phenotypic trait values after including pollinator identity as latent variable, B) standard deviation of phenotypic traits after including pollinator identity as latent variable, C) standard deviation of phenotypic traits after reducing the number of variables tested. The comparison of BIC and p-values among them and with results shown in Fig.5.6 confirm the overparametrization of the model as more variables are included.



6

DISPERSAL, ISOLATION, AND REUNION: THE PHYLOGEOGRAPHY OF *ERYSIMUM MEDIOHISPANICUM* IN THE IBERIAN PENINSULA

ABSTRACT

In the present chapter, we explore the phylogeographic pattern of a Mediterranean herb from the Iberian Peninsula, *Erysimum mediohispanicum* (Brassicaceae), using the highly variable trnL-trnF IGS cpDNA region. We also include in our analysis five closely related *Erysimum* species. The 66 studied populations are grouped into four genetic lineages, each spanning different geographic areas. Dispersal explain better than vicariance the current distribution of linages, but two alternative colonization pathways are compatible with the obtained results. The first potential pathway encompasses two colonization events, one from north to south followed by a second one from south to north. The second pathway encompasses two independent north-to-south colonization events. The relationship between *E. mediohispanicum* and the other studied *Erysimum* species can be explained by two alternative processes. The fist one implies the origin of the other species from *E. mediohispanicum* meanwhile this species was colonizing the Iberian Peninsula. The other scenario involves the hybridization between *E. mediohispanicum* and the other already existing *Erysimum* species when they contacted. Our study remarks the complexity of the phylogeographic history of Mediterranean plants.[1]

[1] A. Jesús Muñoz-Pajares, Mohamed Abdelaziz, M. Belén Herrador, José M. Gómez, and Francisco Perfectti

INTRODUCTION

During glacial periods, as the north of Europe was covered with ice, the distribution of species inhabiting this area shifted southward, spreading into the three major southern Mediterranean peninsulas (Iberia, Italy, and the Balkans), and even reaching Africa (Hewitt, 2004; Nieto Feliner, 2011). During inter-glacial periods, warmer temperatures lead species to re-colonize central and north Europe from its southern refugia before re-starting the cycle during the next glacial and inter-glacial periods (Schmitt, 2007; Tzedakis, 2009; Jiménez-Moreno et al., 2010). Apart from these latitudinal shifts, the mountainous relief of the three Mediterranean peninsulas allowed the existence of altitudinal shifts on species distributions as a response to climatic changes (Schmitt, 2007; Tzedakis, 2009; Jiménez-Moreno et al., 2010). As a result, different ranges may have acted as different refugia within each peninsula (Gómez and Lunt, 2007; Médail and Diadema, 2009; Nieto Feliner, 2011; Fuertes-Aguilar et al., 2011), where species may remain isolated during some periods before coming into contact during others (Johnsen et al., 2010; Zeng et al., 2011; Singhal and Moritz, 2012). Consequently, phylogeographic patterns of these peninsulas are usually complex due to the admixture of previously isolated lineages but also to the multiple latitudinal and altitudinal distribution movements, and because species responded individually to these severe climatic oscillations depending on their particular requirements, geography, and environment (Hewitt, 2000).

Within the Iberian Peninsula, most of the plant phylogeographic knowledge is focused on woody species, where two main phylogeographic patterns have been defined (Arroyo et al., 2004), one corresponding to the Mediterranean forest species and the other typically found in mountain species. According to the former pattern, species were isolated in coastal refugia during glacial periods and expanded to the interior of the Peninsula as climate became warmer. The latter showed altitudinal migration patterns, responding to climate fluctuation ascending or descending its distribution range. Although mountain species also showed migration between areas differing in climate conditions, discordant phylogeographic patterns have been found among species belonging to the same genus (Stebbins, 1984; Hillis et al., 1996; Barraclough and Nee, 2001) and among species showing similar habitats and distributions (e.g., Taberlet et al., 1998; Vargas, 2003).

In recent years, an increasing number of studies are focused on Iberian herbs (e.g., Ortiz et al., 2008; Guzmán and Vargas, 2009; Fernández-Mazuecos and Vargas, 2011; Fuertes-Aguilar et al., 2011; Alarcón et al., 2012; Valtueña et al., 2012; Beatty and Provan, 2012; Vrancken et al., 2012). Many of these studies suggest the existence of past fragmented distributions into multiple glacial refugia. However, most of these studies have been conducted with plant taxa inhabiting high-mountain ecosystems and related to temperate or boreo-alpine lineages. Nevertheless, more studies on genuine Mediterranean lineages are required to establish whether a general phylogeographic pattern exists for Iberian plant species (Fernández-Mazuecos and Vargas, 2010; Migliore et al., 2012; Valtueña et al., 2012).

Erysimum (Brassicaceae) is a recent genus, showing several problematic issues in traditional taxonomy due to morphological similarities, probably reflecting rapid speciation processes (Favarger, 1978; Nieto-Feliner, 1993; Abdelaziz et al., 2011). As a consequence of such taxonomic disagreement, the number of species conforming the genus is unknown and ranges from 180 to 223 species, depending on authors (Al-Shehbaz et al., 2006; Warwick and Al-Shehbaz, 2006; Koch and Al-Shehbaz, 2008). Although several species can be found in north Africa, Macaronesia, and North America, the genus is mainly distributed in Eurasia (Koch and Al-Shehbaz, 2008) being the western Mediterranean region an important diversification area (Greuter et al., 1986). Twenty-two *Erysimum* species have been described in the Iberian Peninsula, being most of them narrow endemisms

(Nieto-Feliner, 1993). Six out of these twenty-two species (namely, *E. mediohispanicum*, *E. nevadense*, *E. merxmuelleri*, *E. rondae*, *E. ruscinonense*, and *E. gomezcampoi*) are considered as microspecies conforming the *E. nevadense* complex (Nieto-Feliner, 1993). *E. mediohispanicum* is the most widespread *Erysimum* species in the Iberian Peninsula, being distributed in two main areas, one in the north-east and other in the south-east (Nieto-Feliner, 1993) (Fig.6.1). Inferring relationships among populations of such *E. nevadense* species is hindered by their recent origin and consequent low differentiation, and by the existence of among species contact areas (Fig. 1.2 en la página 47), where hybridization can occur (Clot, 1991). Consequently, it is usual to find individuals (and even entire populations) showing problematic taxonomic diagnostic traits, particularly in sympatry areas (Nieto-Feliner, 1993; Clot, 1991). An additional factor hindering the establishment of relationships among these species using nuclear markers is the different ploidy level. Chromosome number varies among species (2n = 14 for *E. nevadense*, *E. merxmuelleri*, *E. gomezcampoi*, and *E. ruscinonense*, 2n = 26-28 for *E. mediohispanicum* and 2n = 28 for *E. rondae*) and within species (2n = 26 for western *E. ruscinonense* populations, and 2n = 14 for the southernmost *E. mediohispanicum* populations; Nieto-Feliner, 1993; Clot, 1991; Chapter 5). Using organelle markers would allow direct comparison among all these species, regardless ploidy levels.

Neutral chloroplast DNA (cpDNA) polymorphisms are more powerful than nuclear polymorphisms to resolve plant population evolutionary histories (Hamilton et al., 2003). One of the most commonly used cpDNA markers is the region encompassing the transfer genes for leucine (trnL) and phenylalanine (trnF) (Stuessy, 2009). These two genes are separated by an intergenic sequence (IGS), showing pseudogenes of the functional trnF in some plant species (Fig. 4.1 en la página 126). Moreover, in several clades (particularly in Brassicaceae) this IGS shows polymorphism in the number of pseudogenes not

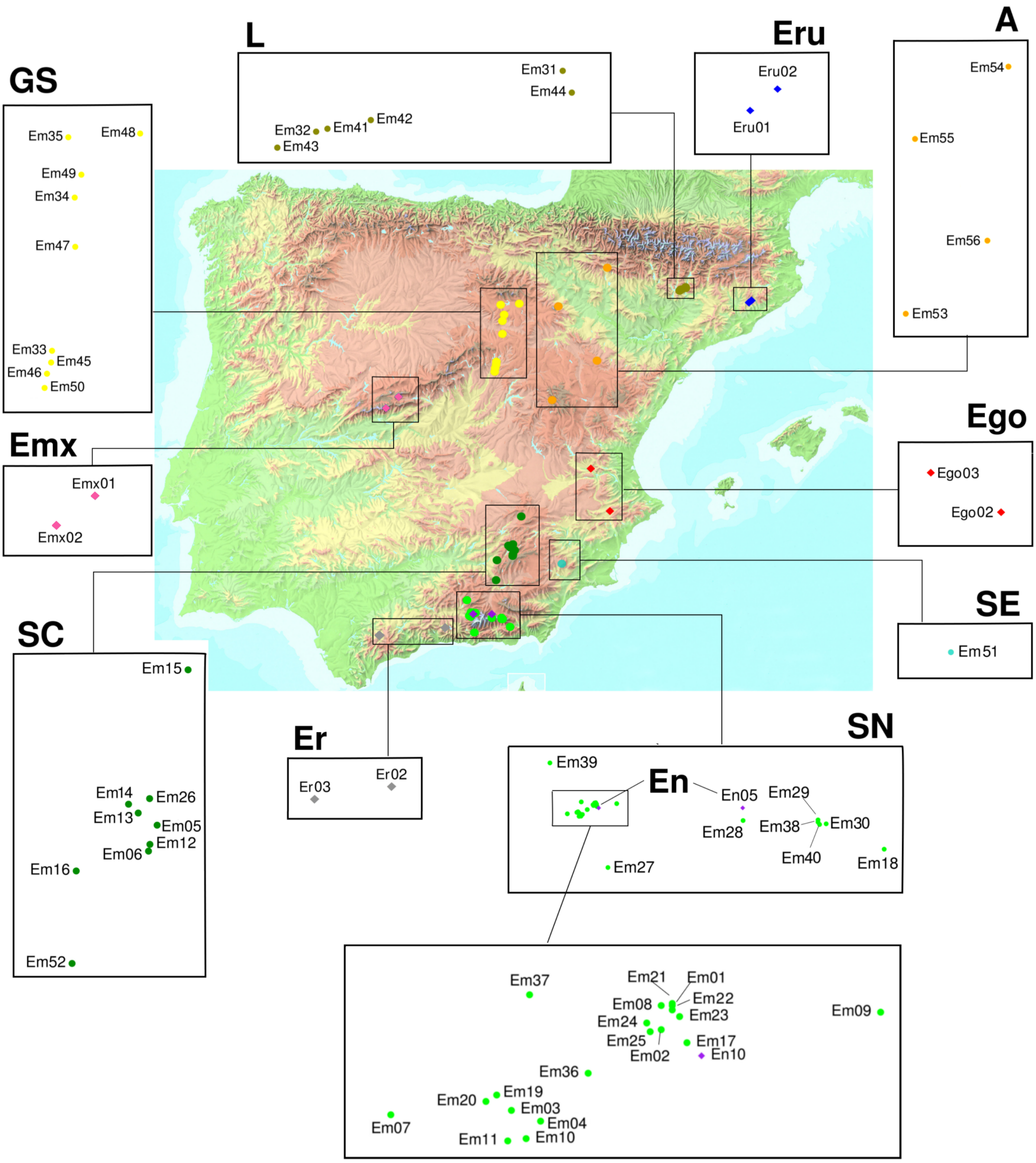


Figura 6.1: Location of the 66 studied populations and sampling areas definitions (L: Lérida; A: Aragón; GS: Guadalajara-Soria; SE: Sierra Espuña; SC: Sierra de Cazorla, Segura and Guillimona; SN: Sierra Nevada).

only among species, but also within and among populations of the same species. For these reason, the region is being used to infer phylogeographic relationships (Koch et al., 2008; Tedder et al., 2010). We used the sequence length polymorphism to establish the genetic relationships between haplotypes found in different species of *Erysimum* in the Iberian Peninsula, using a distance-based method integrating substitution and indel information (Chapter 4). In addition, we have used a percolation network approach (Rozenfeld et al., 2008) because this methodology better represent the uncertainty associated to the population genealogy (Chapter 4).

In the present study, we infer the phylogeographic pattern of *E. mediohispanicum* in the Iberian Peninsula and the evolutionary relationship with the other species from the *E. nevadense* complex.

MATERIAL AND METHODS

Sampling design and population genetic relationships

We sampled a total of 56 *E. mediohispanicum* populations in both northern and southern regions (Fig.S1, Table S1). Southern populations were sampled in three main areas: Sierra Nevada (27 populations), Sierras de Cazorla, Segura and Guillimona (hereafter Sierra de Cazorla; 9 populations), and Sierra Espuña (1 population). Northern populations were sampled in Sierra del Montsec (hereafter, Lérida; 6 populations), Northern Moorlands (hereafter, Guadalajara-Soria; 9 populations) and Aragón (4 populations). We sampled two populations per species of the remaining five *E. nevadense* complex species (Fig.S1, Table S1: *E. ruscinonense*, *E. gomezcampoi*, *E. merxmuelleri*, *E. nevadense*, and *E. rondae*). In all cases populations were represented by five individuals (Table S1).

For each individual, we extracted DNA and amplified the trnL-trnF region as described in Chapter 4. The trnL-trnF intergenic spacer (IGS) region (composed by a variable number of different trnF pseudogene copies) was successfully sequenced in 317 individuals, and 69 different haplotypes were found (Chapter 4). Seventeen out of these 69 haplotypes were specific to *Erysimum* species different from *E. mediohispanicum*, while four haplotypes were shared by two species (H02 and H38: *E. mediohispanicum* and *E. merxmuelleri*; H08: *E. mediohispanicum* and *E. nevadense*; H03: *E. mediohispanicum* and *E. ruscinonense*) (Table S1). The 69 haplotypes were classified into five haplogroups according to their different pseudogene arrangements, and genetic relationships were estimated by combining (giving equal weights) both substitutions and indel distance matrices (see Chapter 4 for details). We estimated genetic distances between populations taking into account haplotypic distances and haplotype frequency per population as implemented in *sidier* R package (Chapter 3). We used the population distance matrix to infer genetic relationships among populations into a network framework (Dyer and Nason, 2004) using percolation networks as described by Rozenfeld et al. (2008). Networks and modules (i.e. subsets of nodes conforming densely connected subgraphsCsardi and Nepusz, 2006) were estimated using *sidier* R package (Chapter 3).

Genetic variance partition

We performed analyses of molecular variance (AMOVA; Excoffier et al., 1992) as implemented in ARLEQUIN version 3.0 (Excoffier et al., 2005) to analyze genetic variance partition at three levels: among groups of populations, among populations within groups, and within populations. We assessed whether grouping populations by network modules increased the similarity among populations within groups compared to grouping populations by geographical sampling areas (Table S1). For that, we performed two separate analyses classifying populations using either network modules or population geographic origins (i.e., mountain ranges) and compared the amount of variance explained by group (module or range) and population within group. We performed each AMOVA using two different approaches, one based on haplotype frequencies and the other based on haplogroup frequencies.

Genetic and geographic distances

To analyze the relationship between genetic differences and geographical distances, we performed three different spatial analyses using the population genetic distance matrix and the geographic distance matrix among populations (estimated from population coordinates using gmt library in R, Magnusson, 2011): First, we tested for correlation between geographical and genetic distances using Mantel tests (Mantel, 1967) as implemented in ape (Paradis et al., 2004). Second, we tested for isolation by distance using the software SGS (Degen et al., 2001), which estimates the Gregorious genetic distances (Gregorius, 1978) from haplotype frequencies per population. Third, to search for genetic discontinuity among regions, we used the software Barrier v2.2 (Gerlach et al., 2010) that implement the maximum difference algorithm of Monmonier (1973) to find edges associated with the highest rate of change in a genetic distance measure (Manni et al., 2004). As a result, Barrier estimates the geographic area where populations show the highest genetic distances.

Phylogenetic relationships

We have also explored the phylogenetic relationships between populations of the *E. nevadense* complex species by performing Bayesian Inference (BI), Maximum Likelihood (ML), Neighbor-Joining

(NJ) and Maximum Parsimony (MP) phylogenies, using nuclear and plastidial sequences in a subset of the studied populations (Table S1). We amplified two chloroplast markers (trnT-trnL, ~1300bp and ndhF, ~2000bp) and two nuclear regions (ITS1 and ITS2, ~350bp each) in one individual per population. We used primers tabA, tabD (Taberlet et al., 1991), ndhF5 and ndhF2100 (Olmstead and Sweere, 1994) to amplify the plastidial markers and ITS1, ITS2, ITS3, and ITS4 (White, 1990) for the two nuclear regions. PCR reactions contained 1.5 mM MgCl2 (New England BioLabs), 0.1 mM each dNTP, 0.2 mM each primer, and 0.02 U Taq DNA polymerase (New England Biolabs) in a total volume of 50 µL. Reactions were conducted in a Gradient Master Cycler Pro S (Eppendorf) with a initial denaturing step of 3 minutes at 94°C and a final extension step of 3 minutes at 72°C in all cases. To amplify the mentioned markers we included 35 cycles at the following parameters: ndhF: 94°C for 15 s, 47°C for 30 s, and 72°C for 90 s; trnT-trnL: 94°C for 15 s, 53°C for 30 s, and 72°C for 90 s; ITS1: 94°C 15 s, 64°C 30 s, and 72°C 45 s; ITS2: 94°C 15 s, 53°C 30 s, and 72°C 45 s. To purify PCR products we centrifuged at 4°C the mix performed with such products, 0.15 volume of 3 M sodium acetate (pH 4.6) and 3 volumes 95 % (v/v) ethanol. Purified PCR products were sequenced by Macrogen Inc. Chromatograms were checked using Finch TV software v1.4.0 (Geospiza) and sequences were edited using BioEdit v7.0.5.3 (Hall, 1999). We included in the analyses the sequences of three species (as outgroups). We used one species of the same genus but distant geographically: *E. passgalense* from Iran. Additionally, we used *Moricandia moricandioides* and *Arabidopsis thaliana*, which are relatives of the *Erysimum* genus (Couvreur et al., 2010). We sequenced from tissue *E. passgalense* and *M. moricandioides* but for *A. thaliana* we used sequences from GenBank (ITS1, X52322; ITS2, X52322; ndhF, AP000423; tabAD, AP000423).

The significant congruence among markers (CADM test: W=0.722; p=0.0018) allowed us to concatenate sequences and perform combined phylogenetic analyses. BI tree was estimated using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) with two independent runs of two million MCMC generations of four chains each, sampling trees every 100 generations. We used Tracer v1.4.1 (Rambaut and Drummond, 2007) to check convergence of runs (average standard deviation of split frequencies <0.008) and to establish in 5000 the number of burn-in generations. To obtain the ML tree, we used PhyML v3.0 (Guindon and Gascuel, 2003) with a BioNJ starting tree and the Nearest Neighbor Interchange (NNI) as the heuristic algorithm for tree topology search. Branch supports were estimated both by approximate likelihood ratio test (with SH-like supports) and by 500 non parametric bootstrap replicates. As PhyML has not yet incorporated the analysis of different DNA partitions, we considered the concatenated sequences as a unique partition. The best evolutionary model for our dataset was the GTR+I+G, according to the Akaike information criterion (AIC) as implemented in MrModeltest (Nylander, 2004). The NJ phylogeny was estimated with MEGA5 (Tamura et al., 2011) using Tamura 3-parameter as distance method (Tamura, 1992) and homogeneous substitutions pattern among lineages. Rate variation among sites was modeled with a gamma distribution (shape parameter = 4) and branch supports were estimated using 500 bootstrap replicates. Finally, MP trees were inferred using Close-Neighbor-Interchange (CNI) on Random Trees option in MEGA5 (Tamura et al., 2011), using 10 initial trees and producing 61 equally most parsimonious trees, from which the consensus tree was obtained. We performed 500 bootstrap replicates to estimate branch supports. Gaps were computed as missing data for BI and ML analyses, and treated as "partial gap deletion" (with 95 % site coverage cut-off) for NJ and MP trees.

We compared the congruence among distance matrices obtained for the trnL-trnF IGS region and for the phylogenetic markers mentioned above for the 22 populations shared in both datasets. For that, we estimated separately for nuclear and plastidial markers the matrix of pairwise distances using Kimura 2-parameters substitution model (Kimura, 1980) and gap pairwise deletion. We assessed the congruence among these two distance matrices and the substitutions and indels combined distance matrix by means of global CADM test, and evaluated the contribution of individual matrices to the overall congruence using a posteriori CADM tests. We performed these analyses using the *ape* package (Paradis et al., 2004).

RESULTS

We have found an average of 1.91 haplotypes and 1.33 haplogroups per population among the 66 studied populations (Table 6.2A). The most diverse species was *E. merxmuelleri* with 3.5 haplotypes and 2.0 haplogroups per population, while *E. nevadense* showed the lowest diversity with 1.0 haplotypes and 1.0 haplogroups per population (Table 6.2B). In *E. mediohispanicum*, the mean number of haplotypes found per population was 1.88 and the mean number of haplogroups per population was 1.34 (Table 6.2B). Southern *E. mediohispanicum* populations showed more haplotypes per population than northern populations (2.00 and 1.63, respectively) although in both regions the number of haplogroups found were highly similar (1.35 and 1.32, respectively), (Table 6.2C). Regarding sampling areas, Lérida populations were the less diverse (1.33 haplotypes in average, contrasting with the 2.00 haplotypes in average found in Sierra Nevada, Sierra de Cazorla, and Sierra Espuña) (Table 6.2D).

Genetic and geographic distances

The distribution and frequency of haplotypes found in the studied populations showed that close populations share haplotypes (Fig.6.2). Considering haplogroups instead of haplotypes enhanced this geo-

A)

	Populations	Haplotype/population	Group/population
All	66	1.91	1.33

B)

	Populations	Haplotype/population	Group/population
E. mediohispanicum	56	1.88	1.34
E. merxmuelleri	2	3.50	2.00
E. nevadense	2	1.00	1.00
E. rondae	2	2.00	1.50
E. ruscinonense	2	2.50	1.00
E. gomezcampoi	2	1.50	1.00

C)

	Populations	Haplotype/population	Group/population
Southern Em	37	2.00	1.35
Northern Em	19	1.63	1.32

D)

	Populations	Haplotype/population	Group/population
Sierra Nevada	27	2.00	1.37
Sierra de Cazorla	9	2.00	1.33
Lérida	6	1.33	1.17
Guadalajara-Soria	9	1.78	1.44
Aragón	4	1.75	1.25
Sierra de Espuña	1	2.00	1.00

E)

	Populations	Haplotype/population	Group/population
Module 1	22	2.00	1.27
Module 2	17	1.65	1.12
Module 3	23	1.91	1.48
Module 4	4	2.50	1.75

Tabla 6.1: Number of populations studied, average number of haplotypes per population, average number of haplogroups per population found in A) overall populations; B) *E. nevadense* complex species; C) *E. mediohispanicum* distribution regions; and D) *E. mediohispanicum* mountain ranges. E) Percolation network modules.

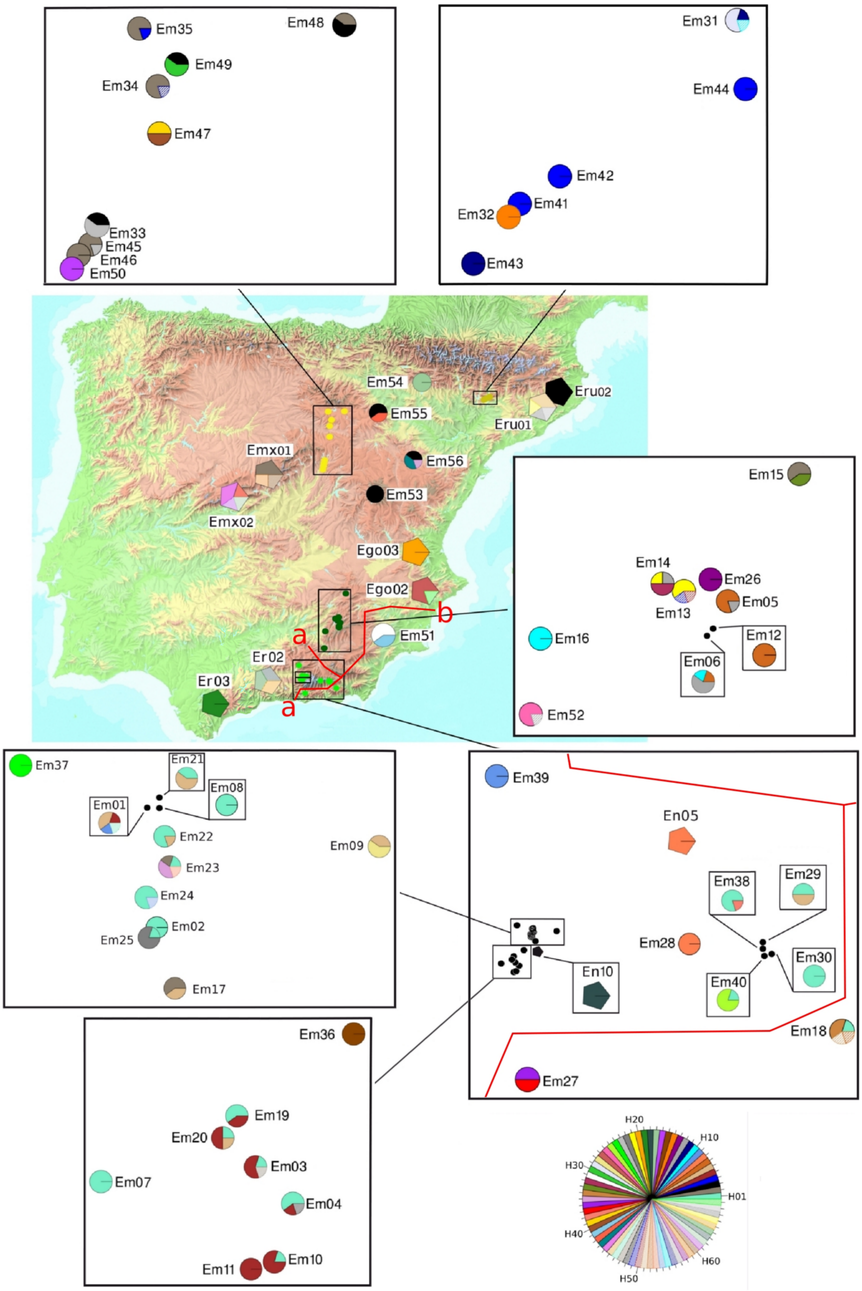


Figura 6.2: Haplotype frequencies per population. Each haplotype is represented by a different color. *E. mediohispanicum* populations are shown as circles and other *Erysimum* species are shown as pentagons. Red lines represent the two most important genetic discontinuities ("a" represents the most important one). (Em: *E. mediohispanicum*, En: *E. nevadense*, Emx: *E. merxmuelleri*, Eru: *E. ruscinonense*, Ego: *E. gomezcampoi*, Er: *E. rondae*).

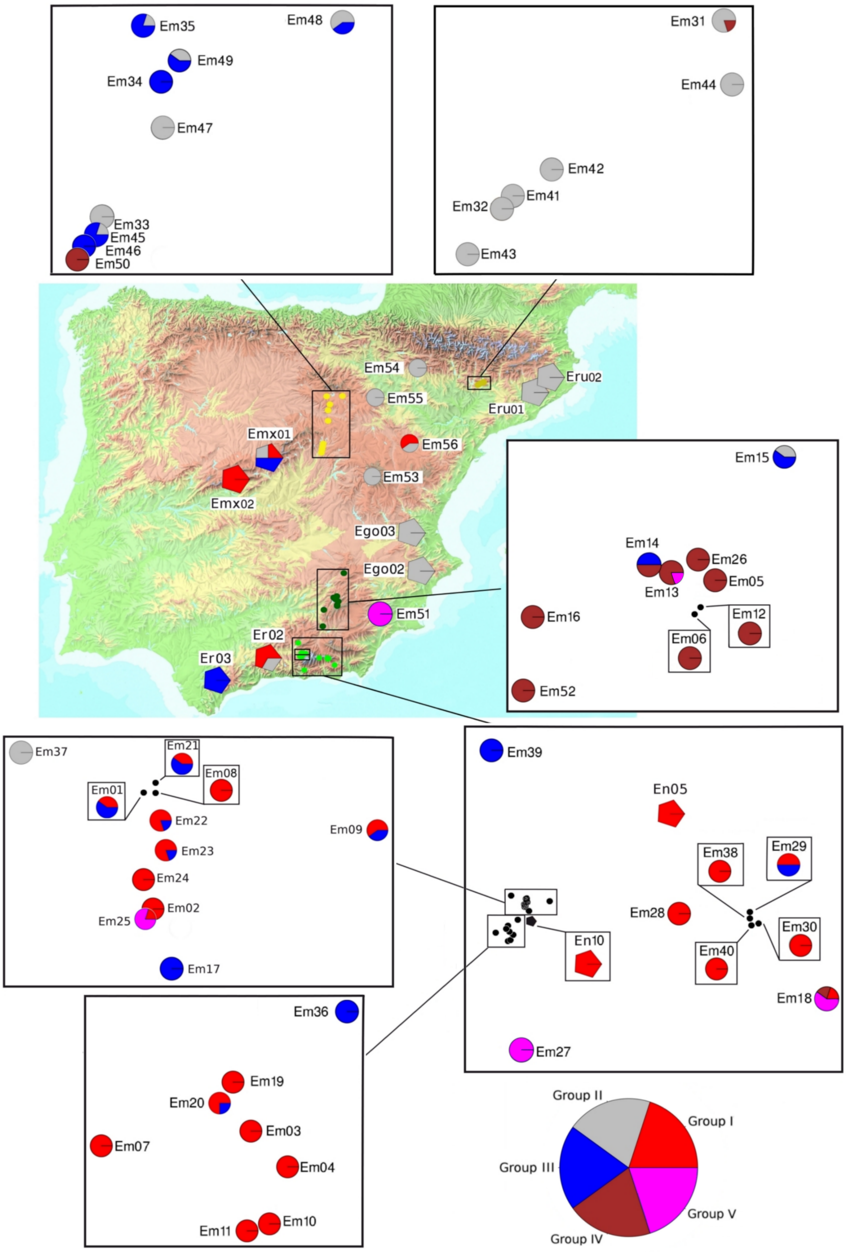


Figura 6.3: Haplogroup frequencies per population. Each haplogroup is represented by a different color. *E. mediohispanicum* populations are shown as circles and other *Erysimum* species are shown as pentagons. (Em: *E. mediohispanicum*, En: *E. nevadense*, Emx: *E. merxmuelleri*, Eru: *E. ruscinonense*, Ego: *E. gomezcampoi*, Er: *E. rondae*).

graphic pattern (Fig.6.3 and see Chapter 4 for haplogroups definition). Haplogroup I appeared mainly distributed in Sierra Nevada. Haplogroup II was found in northern and eastern populations (Lérida, Aragón, Guadalajara-Soria) and in *E. merxmuelleri* and *E. gomezcampoi* (Fig.6.3). Haplogroup III was more frequent in Guadalajara-Soria populations although it was also present in Sierra Nevada, Sierra de Cazorla, and in *E. merxmuelleri* and *E. rondae*. Haplogroup IV was almost exclusively found in Sierra de Cazorla populations (Fig.6.3). Finally, Haplogroup V only occurred in five populations belonging to Sierra Espuña (Em51), Sierra de Cazorla (Em13), and Sierra Nevada (Em18, Em25, Em27).

According to the population distance matrix, the most important genetic discontinuity separates the core of the Sierra Nevada populations from both Sierra de Cazorla and the two southernmost populations of Sierra Nevada (Em27 and Em18; Fig.6.2). A second genetic discontinuity separates these two last populations, together with Em51 (Sierra Espuña), from the remaining *Erysimum* populations (Fig.6.2). Mantel tests showed significant correlation between geographic and genetic distances ($Z = 237076$; $p < 0.00001$) and the distogram (Fig.6.4) is compatible with isolation by distance, because genetic distances were lower than expected for short spatial distances and higher than expected for large distances.

Phylogeographic relationships among populations

We used the population distance matrix to study population relationships based in a network representation. The estimated percolation threshold was 0.33 (Fig.6.5A) and the percolation network showed four modules (Fig.6.5B). Each module was distributed in a different geographic area (Figs. 6.5C and 6.5D), and each haplogroup was mainly found in one module (Fig.6.5E):

Module 1 was almost exclusive from Sierra Nevada (Fig.6.5B and Fig.6.5C), encompassing 18 out of the 27 populations belonging to such mountains, the two populations of *E. nevadense*, and one out of the two populations of both *E. rondae* and *E. merxmuelleri* (Fig.6.5D). Among the 102 haplotypes classified in Module 1, 94 of them belonged to Haplogroup I (Fig.6.5E).

Module 2 was mainly composed of eastern and north-eastern populations (Fig.6.5). Sixteen out of the 17 populations included in this module were sampled in the northern region: all the populations belonging to Lérida and Aragón, two out of the nine Guadalajara-Soria populations, and the two populations of both *E. ruscinonense* and *E. gomezcampoi* (Fig.6.5D). Additionally, this module included one population from Sierra Nevada and 79 out of the 83 haplotypes present in populations of this module belong to Haplogroup II (Fig.6.5E).

Module 3 included interior populations from both southern (mainly Sierra de Cazorla) and northern (Guadalajara-Soria) regions. This module encompassed the nine Sierra de Cazorla populations, five from Sierra Nevada, seven from Guadalajara-Soria, and one out of the two populations of both *E. rondae* and *E. merxmuelleri* (Fig.6.5). Populations belonging to this module were mainly composed by haplotypes belonging to Haplogroups III (56 out of 113) and IV (41 out of 113) (Fig.6.5E).

Finally, Module 4 encompassed the remaining southern populations, including three populations from Sierra Nevada and the Sierra Espuña population (Fig.6.5C). Although these four populations belonged to the southern region, they were connected to the module 2, almost exclusively composed by northern populations (Fig.6.5D). Sixteen out of the 19 haplotypes in this module belonged to Haplogroup V (Fig.6.5E).

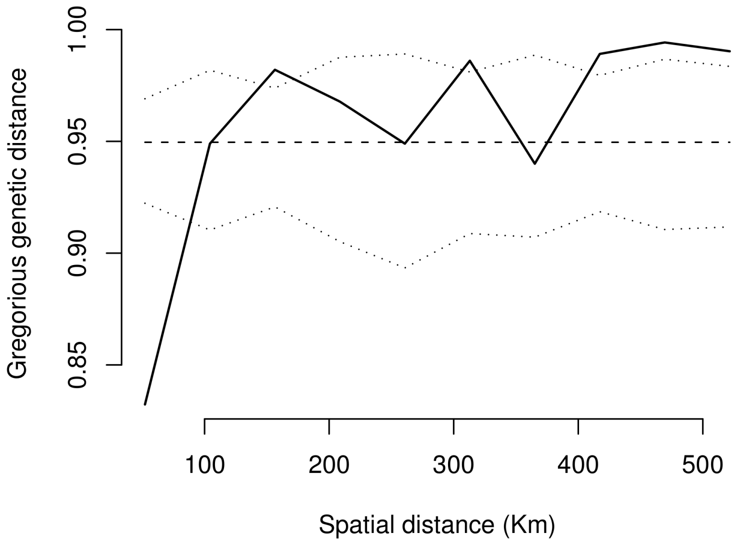
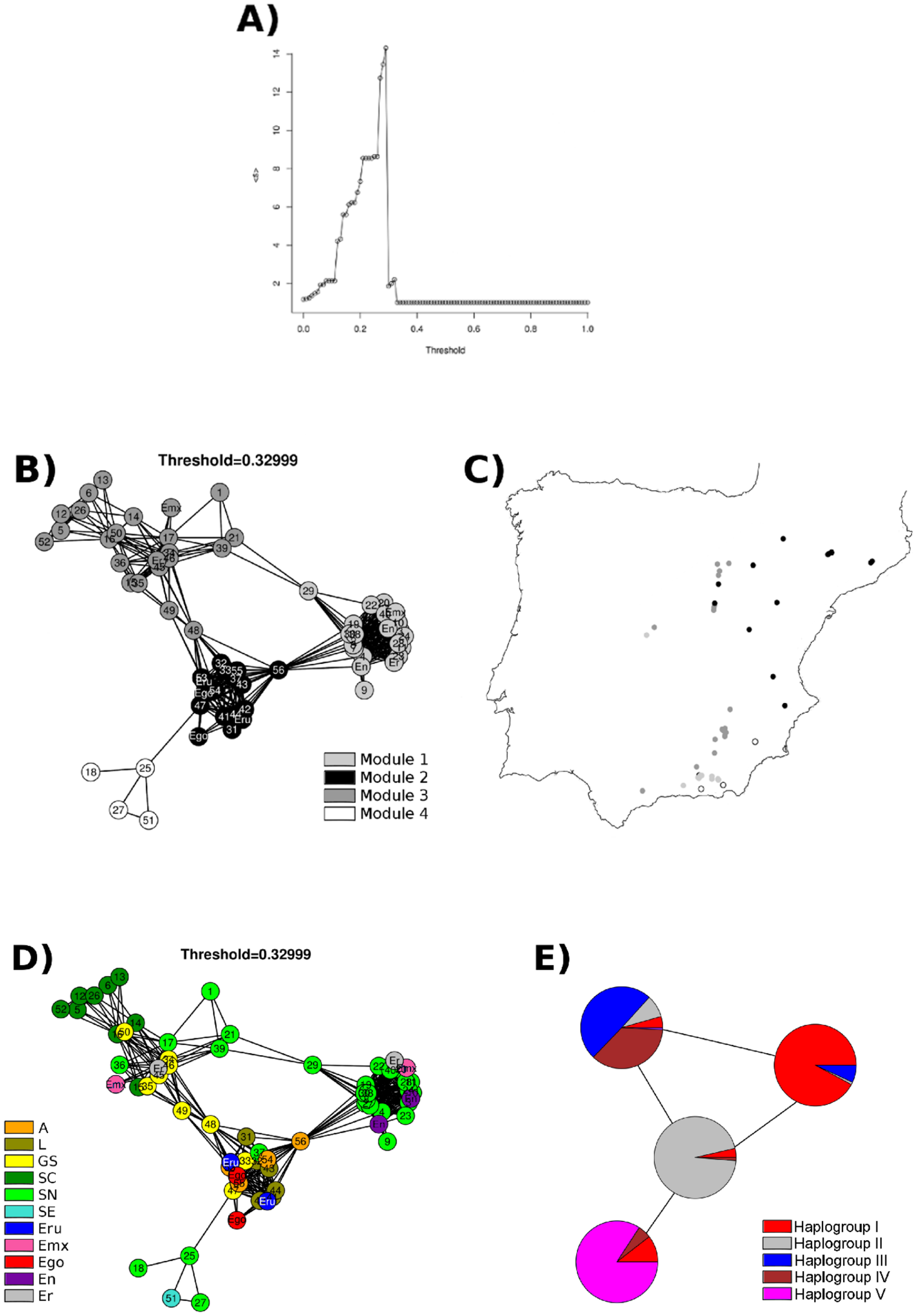


Figura 6.4: Gregorious genetic distance (solid line) versus spatial distance classes. Dashed lines indicate absence of spatial autocorrelation. Dotted lines represent 95 % Confidence Intervals.

Figura 6.5: A) Percolation threshold estimation. B) Network built connecting distances lower than the estimated percolation threshold. Modules are represented in different grey tones. Numbers represent *E. mediohispanicum* populations. Other species are represented by acronyms. C) Geographic location of the four modules found in B. D) Network built connecting distances lower than the estimated percolation threshold. Sampling areas are shown in different colors. Numbers represent *E. mediohispanicum* populations. Other species are represented by acronyms. E) Haplogroup composition of each network module (network depicted in the same orientation than in B and D). Acronyms: L: Lérida; A: Aragón; GS: Guadalajara-Soria; SE: Sierra Espuña; SC: Sierra de Cazorla, Segura and Guillimona; SN: Sierra Nevada; En: *E. nevadense*; Emx: *E. merxmuelleri*; Eru: *E. ruscinonense*; Ego: *E. gomezcampoi*; Er: *E. rondae*.

(Next page) ►


A)
<S>
2
4
6
8
10
12
14
0.0
0.2
0.4
0.6
0.8
1.0
Threshold
B)
Threshold=0.32999
Module 1
Module 2
Module 3
Module 4
C)
D)
Threshold=0.32999
A
L
GS
SC
SN
SE
Eru
Emx
Ego
En
Er
E)
Haplogroup I
Haplogroup II
Haplogroup III
Haplogroup IV
Haplogroup V

	Mountain ranges		Modules	
	Haplotype	Groups	Haplotype	Groups
Among ranges/modules	40.76	49.79	55.47	66.29
Among pops within range/module	36.48	31.54	23.71	16.76
Within populations	22.76	18.66	20.81	16.95
F_{SC}	0.6158	0.6283	0.5327	0.4971
F_{ST}	0.7724	0.8134	0.7919	0.8305
F_{CT}	0.4076	0.4979	0.5547	0.6629

Tabla 6.2: Analysis of molecular variance (AMOVA) and fixation indices estimated grouping populations according to geographic distances (mountain ranges) and network modules. For each case, we based the analyses in both haplotype and haplogroup frequencies per population. F_{SC} is the proportion of the group variance due to differences among populations. F_{ST} is the proportion of the total variance due to differences among populations. F_{CT} is the proportion of the total variance due to differences among groups.

Population genetic structure

Although the number of haplotypes and the number of haplogroups were clearly different (69 and five, respectively), AMOVA results based on haplotypes and haplogroups were similar, with most of the variance found among groups and with the lowest variation in the within population component (Table 6.2). When populations were grouped according to geographic distribution (Table 6.2), sampling areas explained between 41 and 50 % of the variance and among populations within mountain range explained between 32 and 36 %. When populations were classified according to network modules (Table 6.2), differences among groups were higher (among modules variation: 55-66 %) and each group was more homogeneous (among populations within module variation: 17-24 %). A lower F_{SC} value (the proportion of the variance found in a group due to differences among populations) was obtained after grouping according to modules (0.50-0.53) than according to geography (0.62-0.63). F_{ST} values were high and almost identical regardless populations were grouped according to sampling areas (0.77-0.81) or modules (0.79-0.83) (Table 6.2).

Phylogenetic relationships among populations

Phylogenies inferred using the concatenated dataset including four DNA markers (ITS1, ITS2, trnT-trnL, and ndhF) showed low resolution. Only three branches were well-supported (⩾75 %) with all BI, ML, NJ, and MP methods (Fig.6.6). Such branches grouped the studied populations of *E. nevadense*, *E. ruscinonense*, and *E. gomezcampoi* in three different clades. Populations belonging to the remaining species did not constitute independent clades according to all BI, ML, NJ, and MP trees. Sierra Nevada populations were classified into two different clades (Fig.6.6).

The three distance matrices (i.e., the nuclear distance matrix, the plastidial distance matrix, and the substitutions and indels combined distance matrix) were congruent (N=22, W=0.452, p= 0.0004; Table 6.3A). The relative contribution of each matrix to this congruence was different, being the trnL-trnF IGS region determinant for the congruence found (W=0.310, P=0.0003) whereas the contribution of the two markers based on substitutions were marginally significant (W<0.129, p>0.065; Table 6.3B). The trnL-trnF IGS combined distance matrix was also congruent with both plastidial (W=0.673, p= 0.0004) and nuclear (W=0.638, p=0.0015) markers when pairwise comparisons were performed (Table 6.3A), showing values similar to the congruence between nuclear and plastidial substitution distance matrices (W=0.722, p=0.015; Table 6.3A).

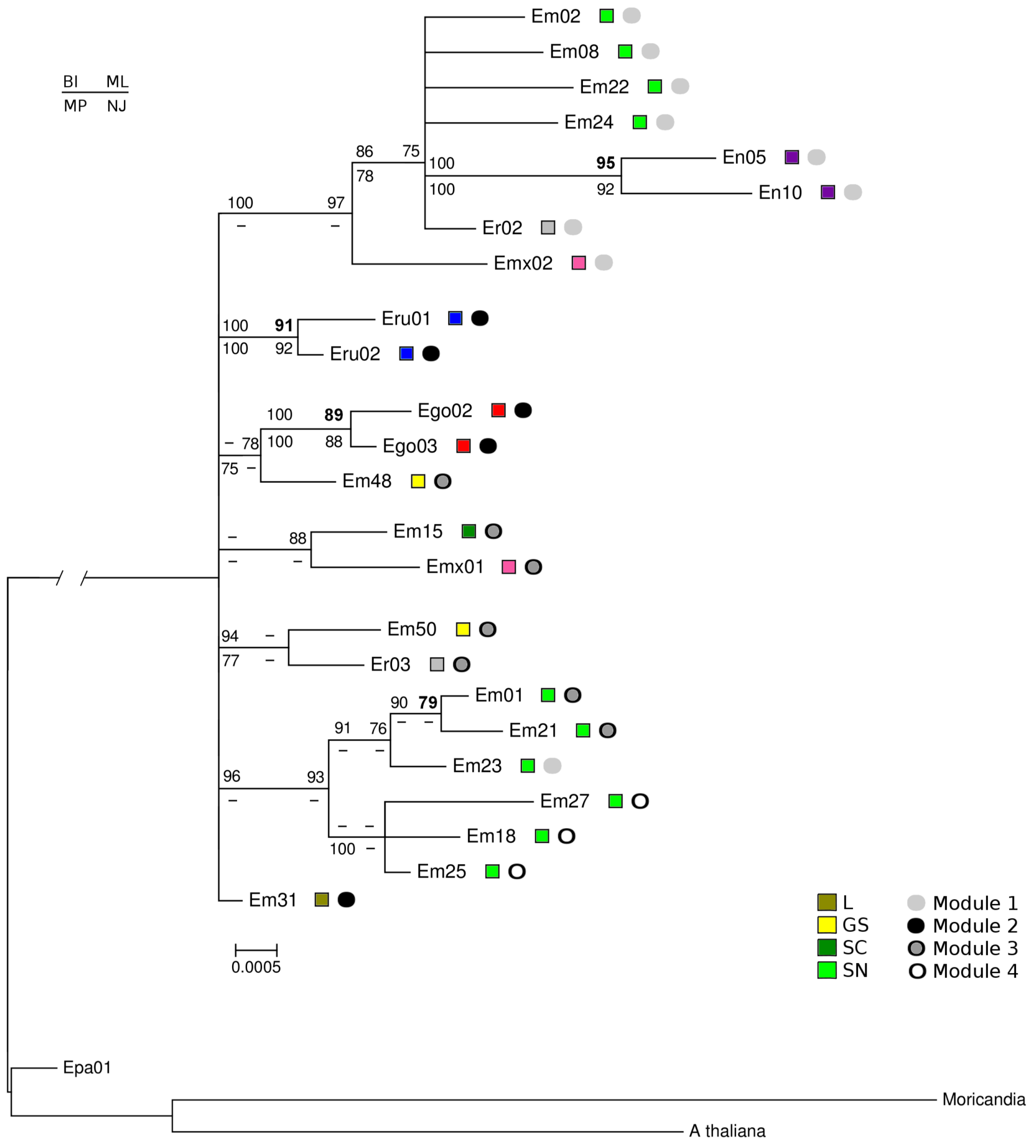
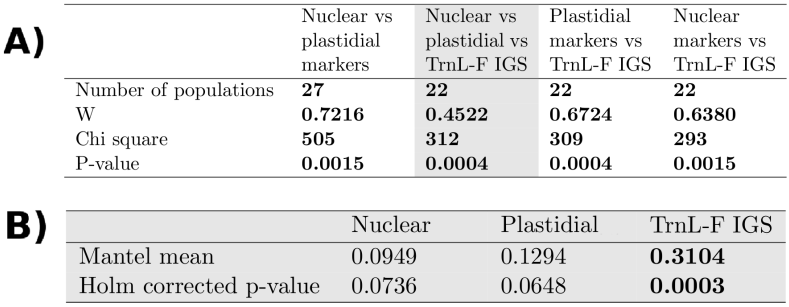


Figura 6.6: Bayesian inference (BI) phylogenetic relationships in a subset of the studied populations (represented by one individual each) according to concatenated sequences from both nuclear (ITS1 and ITS2) and chloroplast (trnT-trnL and ndhF) markers. Branch supports higher than 0.75 are shown for BI, MP, NJ, and ML (values for aLRT. In bold, branch supports showing bootstrap values > 0.75). Filled square colors represent sampling areas and species. Filled circle grey tones represent modules in trnL-trnF IGS network analysis.



A)

	Nuclear vs plastidial markers	Nuclear vs plastidial vs TrnL-F IGS	Plastidial markers vs TrnL-F IGS	Nuclear markers vs TrnL-F IGS
Number of populations	**27**	**22**	**22**	**22**
W	**0.7216**	**0.4522**	**0.6724**	**0.6380**
Chi square	**505**	**312**	**309**	**293**
P-value	**0.0015**	**0.0004**	**0.0004**	**0.0015**

B)

	Nuclear	Plastidial	TrnL-F IGS
Mantel mean	0.0949	0.1294	**0.3104**
Holm corrected p-value	0.0736	0.0648	**0.0003**

Tabla 6.3: Congruence among distance matrices (CADM) test results of comparing nuclear and plastidial sequences (represented by distance matrices based on substitutions) and trnL-trnF IGS region (which distance matrix is based on both indels and substitutions). (A) Global CADM comparing matrices pairwise and the three matrices together (shadowed column). (B) Partial contributions of each matrix to the global congruence.

DISCUSSION

Phylogeographic relationships

The disjunct distribution of *E. mediohispanicum*, showing two separate regions may be originated by two mechanisms (Ronquist, 1997): the colonization of new areas by individual from a given population (dispersal) or the fragmentation of a large original species distribution (vicariance). Several methods are available to discriminate between both processes based on dating the divergence, the topology of phylogenetic relationships, the amount of genetic diversity, and the geographic distribution of alleles (Ronquist, 1997; Kropf et al., 2008).

The method based on dating the divergence consists on estimating and comparing both the lineage split time and the age of the barrier separating the different distribution areas. However, applying this method to lineages showing recent divergence (as is the case for *E. mediohispanicum*) may produce misleading results (Kropf et al., 2006). To apply this method we also need dating the phylogeography obtained by combining indels and substitutions. The different natures (e.g., mutation rates, mechanisms) of these two kind of mutations also hinder the application of this method to our dataset.

One of the methods used in cases of recent divergence compares the geographically separated lineages genetic diversities. Whereas it is expected that separated regions showed similar diversities under vicariance, areas colonized by dispersal must be less diverse than source areas. However, if colonization is produced from different areas, unclear patterns may be obtained (Kropf et al., 2006). Accepting the number of haplotypes found per population as a population genetic diversity estimator (despite based only on five individuals per population), we have found higher genetic diversity in southern populations (with an average of 2.00 haplotypes per population) than in northern populations (with an average of 1.63 haplotypes per population), supporting the existence of dispersal mechanisms. Within the northern region, it is especially low the haplotype diversity found in Lérida (1.33), whereas the intermediate values obtained in Aragón and Guadalajara-Soria (1.75 and 1.78, respectively) may be consequence of the multiple origins of colonization for these areas (Kropf et al., 2006). Unfortunately, this method has an important drawback: the genetic diversity of a given population not only depends on its origin but also on several other factors such as the population size, breeding system or divergence rate (Nybom and Bartish, 2000; Charlesworth and Pannell, 2001; Landergott et al., 2001).

It is also possible to distinguish between vicariance and dispersal events depending on the phylogenetic branching order (Brunsfeld et al., 2001; Cartens et al., 2005). Lineages of the different geographic areas must constitute monophyletic clades in the case of vicariance, but a paraphyletic pattern of geographically distant populations composing a phylogenetic clade is expected according to the dispersal hypothesis (Kropf et al., 2006). According to the percolation network, modules and geographic areas showed a clear correspondence (Figs. 6.5B and 6.5C) but existing regions with paraphyletic origins (e.g. Sierra Nevada, Guadalajara-Soria) and modules were composed by geographically distant populations (e.g., Module 3) (Fig.6.5). These findings also support the hypothesis of dispersal.

Finally, due to the wide original distribution underlying vicariance, it is expected the existence of particular alleles (or haplotypes) common to all lineages regardless their geographic distribution (Ronquist, 1997). We have found no haplotypes present in the whole *E. mediohispanicum* distribution range, supporting dispersal as the mechanism better explaining the current distribution of the species in the Iberian Peninsula.

AMOVA results also support the existence of dispersal rather than vicariance because the genetic structure is stronger when populations are grouped according to the estimated network modules instead to their geographic location (compare variance partition among ranges and among modules, Table 6.2). This result points to the existence of geographic areas where different genetic lineages are overlapped (e.g., Sierra Nevada, Guadalajara-Soria; Figs. 6.5C and 6.5D), probably due to different colonization events that occurred at different time periods and allowed genetic differentiation between lineages before reunion. Glacial and inter glacial periods are the most invoked causes to explain that kind of events (Cooper and Hewitt, 1993; Hewitt, 2000).

According to our network analyses, we propose below two alternative colonization pathways, based on the two alternative topologies describing haplotype relationships. These topologies are the results of different interpretations of mutation events (Fig.6.7 and Chapter 4) and each of these topologies is supported by different

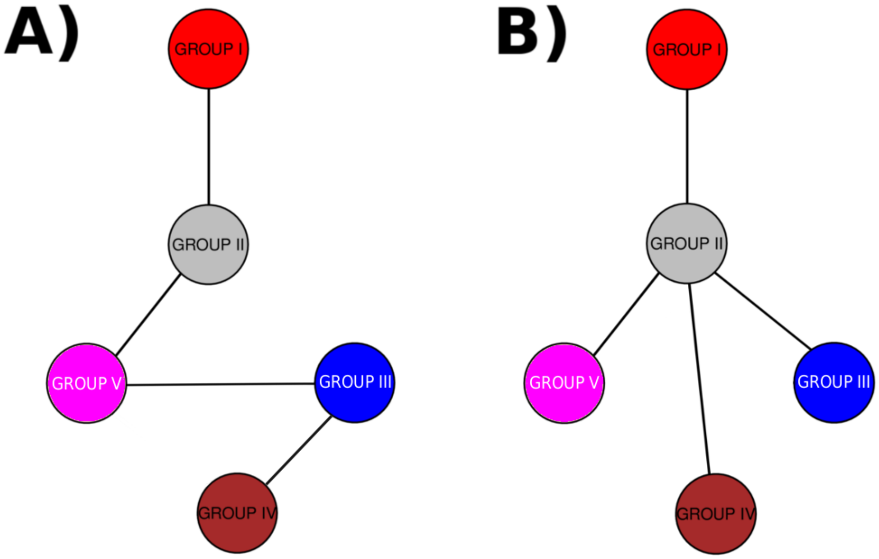


Figura 6.7: Schematic view of the haplogroups relationships inferred in Chapter 4 combining interpretations of duplication and deletion events and (A) among haplotype distance matrices and (B) among haplogroup distance matrices.

network representations. Percolation network estimated using the among haplotype distance matrix (Chapter 4) supports the origin of Haplogroups III and IV from Haplogroup V (Fig.6.7A). However, percolation networks estimated using the among haplogroup distance matrix (not shown) and the among populations distance matrix (Fig.6.5B) supports the origin of Haplogroups III and IV from Haplogroup II (Fig.6.7B).

Despite these mentioned discrepancies, all the analyses performed showed coincident results in several issues: Mutation events, the large number of haplotypes per haplogroup, the higher variation, and the higher number of connections per node points to Haplogroups I and II as the most ancient (Table S1 and Chapter 4). Mutation events and all network analyses also coincide to suggest a possible origin for the southern Haplogroup V (Module 4 populations) from the northern Haplogroup II (Module 2 populations) (Fig.6.5 and Fig.6.7). The two strongest genetic barriers (Fig.6.2) and clades obtained in phylogenetic analyses(Fig.6.6) also support this idea by isolating the Module 4 from other Sierra Nevada and Sierra de Cazorla populations (which are much more closer geographically than Module 2 populations). Module 3 is the only showing two majority haplogroups (III and IV; Fig.6.5), which showed the lowest haplotype distances (Chapter 4). However, these haplogroups showed different geographic distribution, being Haplogroup IV almost exclusive from Sierra de Cazorla populations (Fig.6.3). Although such distribution difference were not included as an additional network module, a tendency is found when populations were allocated according to the Fruchterman-Reingold algorithm (Fruchterman and Reingold, 1991) and most Sierra de Cazorla populations are grouped together (dark green in Fig.6.5D). This result, together with the mutation events (Chapter 4), suggest that haplotypes belonging to Haplogroups III and IV are the most recent and they had have no time to accumulate differences to be classified as two different modules.

Evolution of the E. nevadense *complex*

Despite the phylogeny resolution is low, the yielded clades were consistent with trnL-trnF IGS network modules (Fig.6.6). The only difference between these two analyses was the position of the Sierra Nevada Em23 population, which was one of the most diverse studied populations (four different haplotypes in the five studied individuals). Sampling effects could explain the mentioned difference in highly diverse populations, especially if the within population diversity have different origins, as may occur in Sierra Nevada populations (Hillis, 1998; Poe, 1998).

The *E. nevadense* complex species other than *E. mediohispanicum* did not form a monophyletic clade (Fig.6.5 and Fig.6.6), suggesting that each of these species has an independent origin. This result together with the large distribution range of *E. mediohispanicum* and the fact that haplotypes shared for different species were in all cases found in *E. mediohispanicum* (Table S1) point to *E. mediohispanicum* as the origin of the remaining *E. nevadense* complex species. Because most of the other species were clustered with *E. mediohispanicum* populations according to their geographical proximity, these speciation processes were probably associated to mountain ranges, due to the isolation of low size populations during certain geological periods, a process several times reported for the Iberian Peninsula (Arroyo et al., 2004; Gómez and Lunt, 2007 and references therein). However, this phylogeographic pattern could be also produced by recurrent hybridization between *E. mediohispanicum* and the other *Erysimum* species in the geographic areas where they contact. In this latter scenario the *E. mediohispanicum* is not the ancestor species of the remaining species. Although our results better support the former situation, further studies focused in hybrid zones are required to properly evaluate which of both scenarios is more likely.

The two studied populations per species of *E. ruscinonense*, *E. gomezcampoi*, and *E. nevadense* were grouped in the same module according to the network analysis, and constituted three well-supported clades in phylogenetic analyses. However, for *E. merxmuelleri* and *E. rondae* each population appeared in a different module (Fig.6.5), and were polyphyletic according to phylogenetic analyses (Fig.6.6). First, these two species also showed the highest intrapopulation variability, measured as mean number of haplogroups per population (2.0 and 1.5, respectively. Table 6.2D), which may also influence the result as described above for population Em23. Several processes may cause this pattern. Second, it could be caused by hybridization with nearby populations of *E. mediohispanicum* or other *Erysimum* species when co-occurring spatially, as it has been observed in other organisms (e.g. Pelser et al., 2010; Ramdhani et al., 2011). In fact, we have found that some *Erysimum* species from the *E. nevadense* complex, such as *E. mediohispanicum* and *E. nevadense* may hybridize and hybrid zones exist in nature (Abdelaziz, 2013, Muñoz-Pajares, unpublished data). The effects of hybridization in one but not in the other studied populations of *E. rondae* and *E. merxmuelleri* (due to their respective distances to hybrid zones) may explain that these species do not constitute monophyletic clades. This issue may be specially determinant in the case of Er02, located in the overlap of the two species distribution (Sierra de Alhama), in a range where populations has not been clearly attributed to one of the two species according to phenotypic characters (Nieto-Feliner, 1993). In fact, several authors have classified *E. rondae* as a subspecies of *E. mediohispanicum* (Blanca et al., 1991).

Other possible mechanism producing these patterns is incomplete lineage sorting of ancestral polymorphisms. Because different alleles have particular histories (even within a single species), they may coalesce more recently or more anciently simply by random (Pamilo and Nei, 1988). Thus, the incomplete lineage sorting may be a potential source of polyphyly in gene trees based on few loci (Funk, 2003). It is also more frequently found in species showing short evolutionary times and large effective population sizes (Pamilo and Nei, 1988). Phylogenetic patterns generated by hybridization and incomplete lineage sorting are similar (Avise and Ball, 1990; Morando et al., 2004), and both may lead to discordance between gene and species trees (Avise et al., 1983; Pamilo and Nei, 1988; Takahata, 1989; Doyle, 1992; Harrison, 1991; Maddison, 1997; Knowles, 2001; Sullivan et al., 2002; Rosenberg, 2003; Maddison and Knowles, 2006).

It is difficult to discriminate between these two processes, especially when closely related species are studied (Avise and Ball, 1990). One way to differentiate between incomplete lineage sorting and hybridization is based on the geographic distribution of the alleles shared by the studied species (Goodman et al., 1999; McGuire et al., 2007; Morando et al., 2004). While incomplete lineage sorting must lead to a random distribution of shared alleles across the range of the species descending from a common ancestor, hybridization must lead to a pattern where shared alleles are more frequently found in nearby populations (Barbujani et al., 1994; Hare and Avise, 1998; Masta et al., 2002). Despite it is not possible to properly evaluate the distribution of shared haplotypes based only in two populations per species, we have found that *E. merxmuelleri* shares two haplotypes with *E. mediohispanicum* (H02 and H38). The population Emx01 (closer to *E. mediohispanicum* populations) showed two individuals with H02, whereas only one individual with H38 was found in population Emx02. Contrastingly, *E. rondae* shares no haplotypes with other species, despite one population of *E. rondae* (Er02) is much more closer to *E. mediohispanicum* populations than the other (Fig.6.2). Contrasting with the lack of shared alleles in *E. rondae*, we have found a decrease in haplotype richness as distance to *E. mediohispanicum* increases. Whereas all the individuals sampled in population Er02 showed different haplotypes, all the individuals sampled in the population Er03 (more distant to *E. mediohispanicum*) showed the same haplotype (H19). Although ambiguous, these results support the pattern expected under the existence of hybridization processes.

We have followed the two approaches usually employed to reduce the impact of incomplete lineage sorting: the concatenation of several markers and the inference of phylogenetic relationships using different methods (but see Mossel and Roch, 2010). The observed polyphyletic relationship of populations belonging to *E. rondae* and *E. merxmuelleri* was based in multiple markers of both nuclear and plastidial DNA. It also could be expected a low effect of lineage sorting in the studied dataset because plastidial markers are more represented than nuclear markers. The reason is that the haploid and maternally inherited nature of plastidial genome lead to the occurrence of lineage sorting more unlikely in chloroplast than in nuclear

DNA (Hudson and Turelli, 2003). Despite all these arguments, we are not able to reject the influence of incomplete lineage sorting in the observed results and more populations of the mentioned species must be studied to clarify the mechanisms leading to the observed pattern.

E. nevadense *complex evolutionary history: routes and speciation events*

According to the information provided by the trnL-trnF IGS region, Haplogroups I and II are the oldest (Chapter 4), but nothing could be said about the relative origin of these two haplogroups. However, the orientation of fruits (parallel to stalk) allowed Nieto-Feliner (1993) to propose that northern populations were originated from southern populations. The reason is that this character is found in several southern *Erysimum* species, but only in *E. mediohispanicum* in the northern region. According to this idea, the southern Haplogroup I could be the origin of the Haplogroup II haplotypes, mainly found in the north of the Iberian Peninsula, implying the existence of a first south to north colonization (Fig.6.8A). Because Haplogroup V is derived from Haplogroup II (Fig.6.7), probably occurred a genetic flow from north to south, spreading *E. mediohispanicum* distribution to the south-eastern Mediterranean basin (Fig.6.8B).

Due to the correspondence between haplogroups, network modules, and geography (Fig.6.5), the information provided by the indel-rich region of the trnL-trnF IGS in *E. nevadense* species complex is compatible with two alternative pathways, both sharing the first stage after the initial colonization of northern and southern regions: The first alternative pathway assumes that Haplogroup V is the origin of the most recent Haplogroups III and IV (based on percolation network using haplotype distances; Fig.6.7A) and points to the existence of a south to north flow reaching Sierra Nevada and Guadalajara-Soria populations (Fig.6.8C). Alternatively, Haplogroup II could be

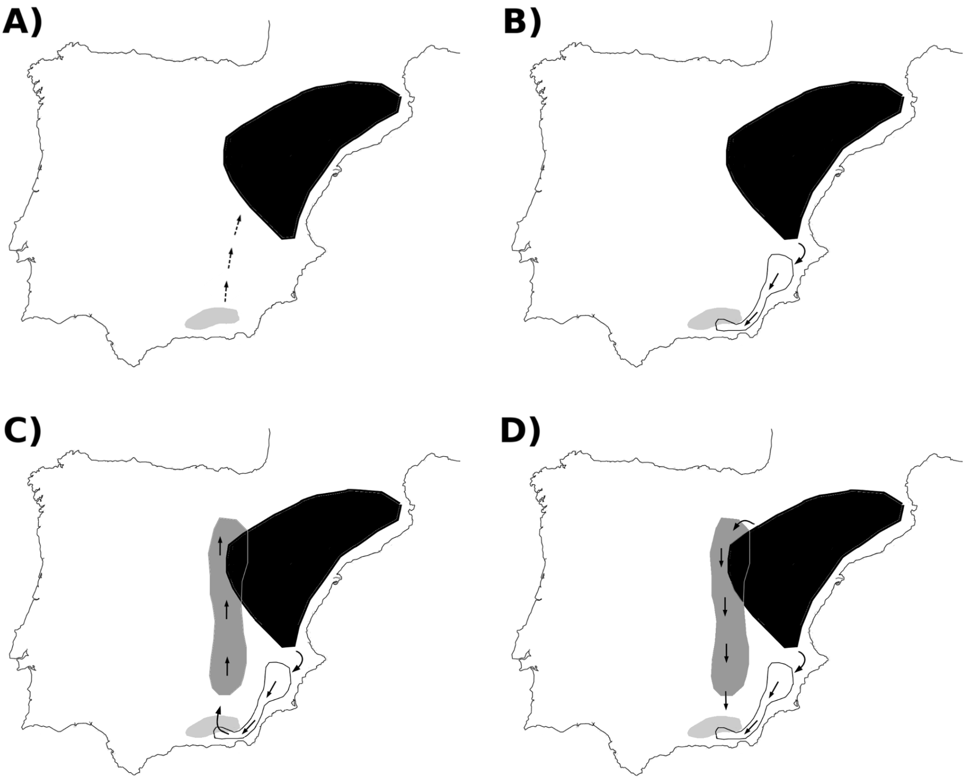


Figura 6.8: Depiction of the hypothetical colonization pathways inferred from trnL-trnF IGS region (solid arrows). A) Haplogroups I and II were the most ancient. Dashed arrows represent putative pathways inferred by Nieto-Feliner (1993). B) Haplogroup V was originated from Haplogroup I, implying a north to south migration event. From this point two alternative pathways are compatible with the obtained information: C) Haplogroups III and IV may be derived from Haplogroup V (as represented in Fig.6.7A), implying a south to north migration event. D) Alternatively, Haplogroups III and IV may be derived from Haplogroup II (as represented in Fig.6.7B), implying a second north to south migration event.

the direct ancestor Haplogroups III and IV (based on percolation network using haplogroup distances; Fig.6.7B), and the origin of Sierra de Cazorla and Guadalajara-Soria populations may be a second north to south migration flow (Fig.6.8D). In both scenarios, all sampling areas but Lérida were colonized in several occasions, which could explain differences in haplotype diversity found between this range and the remaining populations (Haplotypes per population: Lérida = 1.3, Others = 1.75-2.00; Table 6.1).

Both proposed pathways imply the existence of two separate distributions for *E. mediohispanicum* within the Iberia Peninsula during given geological moments. This result is in accordance with a crescent number of evidences suggesting that several species have survived during glacial periods isolated in multiple refugia within the Iberia, several of then located in the north of the Peninsula (Olalde et al., 2002; Gómez and Lunt, 2007; Nieto Feliner, 2011).

Conclusions

Summarizing, the information provided by the trnL-trnF IGS using a distance-based method and network representation was congruent with the information obtained using conventional phylogenetic tools and results of both approaches were compatible. Consequently, the mentioned method seems to successfully extract evolutionary information from an indel-rich region, allowing us to infer the evolutionary history of the *E. nevadense* species complex. The six species conforming this complex showed low divergence and no important genetic differences in the studied cpDNA region was found among species. Haplogroups were distributed according to a clear geographic pattern, showing areas where different haplogroups are overlapped due to the isolation and posterior dispersal events during glacial and interglacial periods. Two different refugia existed in the Iberian Peninsula for *E. mediohispanicum*, one in the north-east and the other in the south-east. The current distribution of the species may be explained after either one north-to-south migration followed by a south-to-north pathway or two independent north-to-south pathways. These dispersal events determined the current taxonomic complexity of the studied *Erysimum* species due to both isolation associated to different mountain ranges and hybridization processes.

REFERENCES

Abdelaziz, M. (2013). *How species are evolutionary maintained? Pollinator-mediated divergence and hybridization in Erysimum mediohispanicum and Erysimum nevadense*. PhD. Thesis.

Abdelaziz, M., J. Lorite, A. J. Muñoz-Pajares, M. B. Herrador, F. Perfectti, and J. M. Gómez (2011). Using complementary techniques to distinguish cryptic species: A new erysimum (brassicaceae) species from north africa. *American Journal of Botany 98*(6), 1049 –1060.

Al-Shehbaz, I., M. Beilstein, and E. Kellogg (2006). Systematics and phylogeny of the brassicaceae (cruciferae): an overview. *Plant Systematics and Evolution 259*(2), 89–120.

Alarcón, M., P. Vargas, L. Sáez, J. Molero, and J. J. Aldasoro (2012). Genetic diversity of mountain plants: Two migration episodes of mediterranean erodium (geraniaceae). *Molecular Phylogenetics and Evolution 63*(3), 866–876.

Arroyo, J. M., J. S. Carrión, A. Hampe, and P. Jordano (2004). La distribución de las especies a diferentes escalas espacio-temporales. http://digital.csic.es/handle/10261/40604. Peer reviewed.

Avise, J. and R. Ball (1990). Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford surveys in evolutionary biology 7*, 45–67.

Avise, J. C., J. F. Shapira, S. W. Daniel, C. F. Aquadro, and R. A. Lansman (1983). Mitochondrial DNA differentiation during the speciation process in peromyscus. *Molecular Biology and Evolution 1*(1), 38–56.

Barbujani, G., A. Pilastro, S. De Domenico, and C. Renfrew (1994). Genetic variation in north africa and eurasia: Neolithic demic diffusion vs. paleolithic colonisation. *American Journal of Physical Anthropology 95*(2), 137–154.

Barraclough, T. G. and S. Nee (2001). Phylogenetics and speciation. *Trends in Ecology & Evolution 16*(7), 391–399.

Beatty, G. E. and J. Provan (2012). Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the lusitanian plant species daboecia cantabrica (ericaceae). *Journal of Biogeography*, 335–344.

Blanca, G., C. Morales, and M. Ruiz Rejon (1991). El género erysimum l. (cruciferae) en andalucía (españa). *Anales del Jardín Botánico de Madrid 49*(2), 201–214.

Brunsfeld, S. J., J. Sullivan, D. E. Soltis, and P. S. Soltis (2001). Comparative phylogeography of northwestern north america: a synthesis. In J. Silvertown and J. Antonovics (Eds.), *Integrating ecology and evolution in a spatial context*, Volume 14, pp. 319–340. Oxford, UK: Blackwell Science.

Cartens, B. C., S. J. Brunsfeld, J. R. Demboski, J. M. Good, and J. Sullivan (2005). Investigating the evolutionary history of the pacific northwest mesic forest ecosystem: Hpotehesis testing within a comparative phylogeographic framework. *Evolution 59*(8), 1639–1652.

Charlesworth, D. and J. Pannell (2001). Mating systems and population genetic structure in the light of coalescent theory. In J. Silvertown and J. Antonovics (Eds.), *Integrating Ecology and Evolution in a Spatial Context: 14th Special Symposium of the British Ecological Society*. Cambridge University Press.

Clot, B. (1991). Caryosystématique de quelques erysimum l. dans le nord de la péninsule ibérique. *Anales del Jardín Botánico de Madrid 49*(2), 215–229.

Cooper, S. J. B. and G. M. Hewitt (1993). Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper chorthippus parallelus. *Insect Molecular Biology 2*(3), 185–194.

Couvreur, T. L. P., A. Franzke, I. A. Al-Shehbaz, F. T. Bakker, M. A. Koch, and K. Mummenhoff (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (brassicaceae). *Molecular Biology and Evolution 27*(1), 55 –71.

Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Degen, B., R. Petit, and A. Kremer (2001). Sgs spatial genetic software: A computer program for analysis of spatial genetic and phenotypic structures of individuals and populations. *Journal of Heredity 92*(5), 447 –448.

Doyle, J. J. (1992). Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany 17*(1), 144.

Dyer, R. J. and J. D. Nason (2004). Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology 13*(7), 1713–1727.

Excoffier, L., G. Laval, and S. Schneider (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online 1*, 47–50.

Excoffier, L., P. E. Smouse, and J. M. Quattro (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics 131*(2), 479–491.

Favarger, C. (1978). Un exemple de variation cytogeographique: le complexe de l'Erysimum grandiflorum sylvestre. *Anales del Instituto Botánico AJ Cavanilles 35*, 361–393.

Fernández-Mazuecos, M. and P. Vargas (2010). Ecological rather than geographical isolation dominates quaternary formation of mediterranean cistus species. *Molecular Ecology 19*(7), 1381–1395.

Fernández-Mazuecos, M. and P. Vargas (2011). Historical isolation versus recent long-distance connections between europe and africa in bifid toadflaxes (linaria sect. versicolores). *PLoS ONE 6*(7), e22234.

Fruchterman, T. and E. Reingold (1991). Graph drawing by force-directed placement. *Software - Practice and Experience 21*(11), 1129–1164.

Fuertes-Aguilar, J., B. G. Gutierrez-Larena, and G. N. Nieto-Feliner (2011). Genetic and morphological diversity in armeria (plumbaginaceae) is shaped by glacial cycles in mediterranean refugia. In *Anales del Jardín Botánico de Madrid*, Volume 68, pp. 175–197.

Funk, D. (2003). Species-level paraphyly and polyphyly : frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst 34*, 397–423.

Gerlach, G., A. Jueterbock, P. Kraemer, J. Deppermann, and P. Harmand (2010). Calculations of population differentiation based on GST and d: forget GST but not all of statistics! *Molecular Ecology 19*(18), 3845–3852.

Gómez, A. and D. Lunt (2007). Refugia within refugia: Patterns of phylogeographic concordance in the iberian peninsula. In *Phylogeography of Southern European Refugia*, pp. 155–188.

Goodman, S. J., N. H. Barton, G. Swanson, K. Abernethy, and J. M. Pemberton (1999). Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus cervus) in argyll, scotland. *Genetics 152*(1), 355–371.

Gregorius, H.-R. (1978). The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Mathematical Biosciences 41*(3–4), 253–271.

Greuter, W., H. M. Burdet, , and G. Long. (1986). Med-checklist 3, dicotyledones (convolvulaceae-labiatae).

Guindon, S. and O. Gascuel (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology 52*(5), 696–704.

Guzmán, B. and P. Vargas (2009). Long-distance colonization of the western mediterranean by cistus ladanifer (cistaceae) despite the absence of special dispersal mechanisms. *Journal of Biogeography 36*(5), 954–968.

Hall, T. (1999). {BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT}. In *Nucleic acids symposium series*, Volume 41, pp. 95–98.

Hamilton, M. B., J. M. Braverman, and D. F. Soria-Hernanz (2003). Patterns and relative rates of nucleotide and Insertion/Deletion evolution at six chloroplast intergenic regions in new world species of the lecythidaceae. *Molecular Biology and Evolution 20*(10), 1710–1721.

Hare, M. P. and J. C. Avise (1998). Population structure in the american oyster as inferred by nuclear gene genealogies. *Molecular Biology and Evolution 15*(2), 119–128.

Harrison, R. G. (1991). Molecular changes at speciation. *Annual Review of Ecology and Systematics 22*(1), 281–308.

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature 405*(6789), 907–913.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359*(1442), 183–195.

Hillis, D., C. Moritz, and B. Mable (1996). Applications of molecular systematics: the state of the field and look to the future. In D. Hillis, C. Moritz, and B. Mable (Eds.), *Molecular Systematics*, pp. 515–543. Sinauer Associates.

Hillis, D. M. (1998). Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology 47*(1), 3–8.

Hudson, R. R. and M. Turelli (2003). Stochasticity overrules the three-times rule: Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial dna. *Evolution 57*(1), 182–190.

Jiménez-Moreno, G., S. Fauquette, and J.-P. Suc (2010). Miocene to pliocene vegetation reconstruction and climate estimates in the iberian peninsula from pollen data. *Review of Palaeobotany and Palynology 162*(3), 403–415.

Johnsen, A., E. Rindal, P. G. P. Ericson, D. Zuccon, K. C. R. Kerr, M. Y. Stoeckle, and J. T. Lifjeld (2010). DNA barcoding of scandinavian birds reveals divergent lineages in trans-atlantic species. *Journal of Ornithology 151*(3), 565–578.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution 16*(2), 111–120.

Knowles, L. L. (2001). Did the pleistocene glaciations promote divergence? tests of explicit refugial models in montane grasshopprers. *Molecular Ecology 10*(3), 691–701.

Koch, M. A. and I. A. Al-Shehbaz (2008). Molecular systematics and evolution of wild crucifers (brassicaceae or cruciferae). In *Biology* *and breeding of crucifers* (Gupta ed.). London, UK.: Taylor and Francis,. draft version, 1–19.

Koch, M. A., M. Wernisch, and R. Schmickl (2008). Arabidopsis thaliana's wild relatives: an updated overview on systematics, taxonomy and evolution. *Taxon 57*(3).

Kropf, M., H. P. Comes, and J. W. Kadereit (2006). Long-distance dispersal vs vicariance: the origin and genetic diversity of alpine plants in the spanish sierra nevada. *New Phytologist 172*(1), 169–184.

Kropf, M., H. P. Comes, and J. W. Kadereit (2008). Causes of the genetic architecture of south-west european high mountain disjuncts. *Plant Ecology & Diversity 1*(2), 217–228.

Landergott, U., R. Holderegger, G. Kozlowski, and J. J. Schneller (2001). Historical bottlenecks decrease genetic diversity in natural populations of dryopteris cristata. *Heredity 87*(3), 344–355.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology 46*(3), 523–536.

Maddison, W. P. and L. L. Knowles (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology 55*(1), 21–30.

Magnusson, A. (2011). gmt: Interface between GMT map-making software and r. R package version 1.1-9.

Manni, F., E. Guérard, and E. Heyer (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using monmonier's algorithm. *Human biology 76*(2), 173–190.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research 27*(2 Part 1), 209–220.

Masta, S. E., B. K. Sullivan, T. Lamb, and E. J. Routman (2002). Molecular systematics, hybridization, and phylogeography of the bufo americanus complex in eastern north america. *Molecular Phylogenetics and Evolution 24*(2), 302–314.

McGuire, J. A., C. W. Linkem, M. S. Koo, D. W. Hutchison, A. K. Lappin, D. I. Orange, J. Lemos-Espinal, B. R. Riddle, and J. R. Jaeger (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid lizards. *Evolution 61*(12), 2879–2897.

Médail, F. and K. Diadema (2009). Glacial refugia influence plant diversity patterns in the mediterranean basin. *Journal of Biogeography 36*(7), 1333–1345.

Migliore, J., A. Baumel, M. Juin, and F. Médail (2012). From mediterranean shores to central saharan mountains: key phylogeographical insights from the genus myrtus. *Journal of Biogeography 39*(5), 942–956.

Monmonier, M. S. (1973). Maximum-difference barriers: An alternative numerical regionalization method*. *Geographical Analysis 5*(3), 245–261.

Morando, M., L. J. Avila, J. Baker, and J. W. Sites (2004). Phylogeny and phylogeography of the liolaemus darwinii complex (squamata: Liolaemidae): Evidence for introgression and incomplete lineage sorting. *Evolution 58*(4), 842–859.

Mossel, E. and S. Roch (2010). Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM 7*(1), 166–171.

Nieto-Feliner, G. (1993). Erysimum. In S. Castroviejo, C. Aedo, C. Gómez-Campo, M. Lainz, P. Monserrat, R. Morales, F. Muñoz-Garmendia, G. Nieto-Feliner, E. Rico, S. Talavera, and L. Villar

(Eds.), *Flora Iberica*, Volume 4, Cruciferae-Monotropaceae., pp. 48–76. Madrid: Real Jardín Botánico CSIC.

Nieto Feliner, G. (2011). Southern european glacial refugia: A tale of tales. http://digital.csic.es/handle/10261/35607. Peer reviewed.

Nybom, H. and I. V. Bartish (2000). Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology, Evolution and Systematics 3*(2), 93–114.

Nylander, J. (2004). MrModeltest v2.

Olalde, M., A. Herran, S. Espinel, and P. Goicoechea (2002, February). White oaks phylogeography in the iberian peninsula. *Forest ecology and management 156*(1/3), 89–102.

Olmstead, R. G. and J. A. Sweere (1994, January). Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the solanaceae. *Systematic Biology 43*(4), 467–481.

Ortiz, M. Á., K. Tremetsberger, A. Terrab, T. F. Stuessy, J. L. García-Castaño, E. Urtubey, C. M. Baeza, C. F. Ruas, P. E. Gibbs, and S. Talavera (2008). Phylogeography of the invasive weed hypochaeris radicata (asteraceae): from moroccan origin to worldwide introduced populations. *Molecular Ecology 17*(16), 3654–3667.

Pamilo, P. and M. Nei (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution 5*(5), 568–583.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in r language. *Bioinformatics 20*, 289–290.

Pelser, P. B., A. H. Kennedy, E. J. Tepe, J. B. Shidler, B. Nordenstam, J. W. Kadereit, and L. E. Watson (2010). Patterns and causes of incongruence between plastid and nuclear senecioneae (asteraceae) phylogenies. *American Journal of Botany 97*(5), 856–873.

Poe, S. (1998). The effect of taxonomic sampling on accuracy of phylogeny estimation: Test case of a known phylogeny. *Molecular Biology and Evolution 15*(8), 1086.

Rambaut, A. and A. Drummond (2007). Tracer v1.4.

Ramdhani, S., N. P. Barker, and R. M. Cowling (2011). Revisiting monophyly in haworthia duval (asphodelaceae): Incongruence, hybridization and contemporary speciation. *Taxon 60*(4), 1001–1014.

Ronquist, F. (1997). Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology 46*(1), 195–203.

Ronquist, F. and J. P. Huelsenbeck (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics 19*(12), 1572–1574.

Rosenberg, N. A. (2003). The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution 57*(7), 1465–1477.

Rozenfeld, A. F., S. Arnaud-Haond, E. Hernández-García, V. M. Eguíluz, E. A. Serrão, and C. M. Duarte (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences 105*(48), 18824 –18829.

Schmitt, T. (2007). Molecular biogeography of europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology 4*(1), 11.

Singhal, S. and C. Moritz (2012). Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution 66*(5), 1474–1489.

Stebbins, G. L. (1984). Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. *Botanica helvetica 94*(1), 1–13.

Stuessy, T. F. (2009). *Plant taxonomy: The systematic evaluation of comparative data* (2 ed.). New York: Columbia University Press.

Sullivan, J. P., S. Lavoué, and C. D. Hopkins (2002). Discovery and phylogenetic analysis of a riverine species flock of african electric fishes (mormyridae: teleostei). *Evolution 56*(3), 597–616.

Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy, and J.-F. Cosson (1998). Comparative phylogeography and postglacial colonization routes in europe. *Molecular Ecology 7*(4), 453–464.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology 17*(5), 1105–1109.

Takahata, N. (1989). Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics 122*(4), 957–966.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution 9*(4), 678–687.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution 28*(10), 2731–2739.

Tedder, A., P. Hoebe, S. Ansell, and B. Mable (2010). Using chloroplast trnF pseudogenes for phylogeography in arabidopsis lyrata. *Diversity 2*, 653–678.

Tzedakis, P. C. (2009). Cenozoic climate and vegetation change. In J. Woodward (Ed.), *The physical geography of the Mediterranean*, pp. 89–137. Oxford University Press.

Valtueña, F. J., C. D. Preston, and J. W. Kadereit (2012). Phylogeography of a tertiary relict plant, meconopsis cambrica (papaveraceae), implies the existence of northern refugia for a temperate herb. *Molecular Ecology 21*(6), 1423–1437.

Vargas, P. (2003). Molecular evidence for multiple diversification patterns of alpine plants in mediterranean europe. *Taxon 52*(3), 463.

Vrancken, J., C. Brochmann, and R. A. Wesselingh (2012). A european phylogeography of rhinanthus minor compared to rhinanthus angustifolius: unexpected splits and signs of hybridization. *Ecology and Evolution 2*(7), 1531–1548.

Warwick, S. and I. Al-Shehbaz (2006). Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution 259*(2), 237–248.

White, T. J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols, A Guide to Methods and Applications*.

Zeng, Y., W. Liao, R. J. Petit, and D. Y. Zhang (2011). Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Molecular ecology*.

SUPPLEMENTARY INFORMATION

Tabla S1: Basic information of the 66 *Erysimum* sp. populations studied including location (longitude, latitude, altitude, region and locality), number of sampled individuals (for trnF and phylogeny), and trnF IGS haplotype information (haplotypes, haplogroups)

Species	Population code	Region	Locality	Ids trnF	Ids phylo	Haplotypes	Groups
E. mediohispanicum	Em01	South	Sierra Nevada	5	1	H05, H06, H09, H46	I, III
E. mediohispanicum	Em02	South	Sierra Nevada	5	1	H01	I
E. mediohispanicum	Em03	South	Sierra Nevada	5		H01, H05, H47	I
E. mediohispanicum	Em04	South	Sierra Nevada	5		H01, H05, H48	I
E. mediohispanicum	Em05	South	Sierra Cazorla	5		H07, H12	IV
E. mediohispanicum	Em06	South	Sierra Cazorla	5		H07, H10, H12	IV
E. mediohispanicum	Em07	South	Sierra Nevada	5		H01	I
E. mediohispanicum	Em08	South	Sierra Nevada	5	1	H01	I
E. mediohispanicum	Em09	South	Sierra Nevada	5		H06, H28	I, III
E. mediohispanicum	Em10	South	Sierra Nevada	5		H01, H05	I
E. mediohispanicum	Em11	South	Sierra Nevada	2		H05	I
E. mediohispanicum	Em12	South	Sierra Cazorla	5		H07	IV
E. mediohispanicum	Em13	South	Sierra Cazorla	5		H21, H49, H50	IV, V
E. mediohispanicum	Em14	South	Sierra Cazorla	4		H12, H21, H32	III, IV
E. mediohispanicum	Em15	South	Sierra Cazorla	5	1	H02, H33	II, III
E. mediohispanicum	Em16	South	Sierra Cazorla	5		H10	IV
E. mediohispanicum	Em17	South	Sierra Nevada	5		H02, H06	III
E. mediohispanicum	Em18	South	Sierra Nevada	5	1	H01, H34, H51, H52	I, IV, V
E. mediohispanicum	Em19	South	Sierra Nevada	5		H01, H05	I
E. mediohispanicum	Em20	South	Sierra Nevada	4		H01, H05, H06	I, III
E. mediohispanicum	Em21	South	Sierra Nevada	5	1	H01, H06	I, III
E. mediohispanicum	Em22	South	Sierra Nevada	5	1	H01, H06	I, III
E. mediohispanicum	Em23	South	Sierra Nevada	5	1	H01, H02, H35, H53	I, III
E. mediohispanicum	Em24	South	Sierra Nevada	5	1	H01, H54	I
E. mediohispanicum	Em25	South	Sierra Nevada	5	1	H01, H22	I, V
E. mediohispanicum	Em26	South	Sierra Cazorla	5		H13	IV
E. mediohispanicum	Em27	South	Sierra Nevada	4	1	H36, H37	V
E. mediohispanicum	Em28	South	Sierra Nevada	5		H08	I
E. mediohispanicum	Em29	South	Sierra Nevada	4		H01, H06	I, III
E. mediohispanicum	Em30	South	Sierra Nevada	4		H01	I
E. mediohispanicum	Em31	North	Lérida	5	1	H11, H29, H55	II, IV
E. mediohispanicum	Em32	North	Lérida	5		H14	II
E. mediohispanicum	Em33	North	Guadalajara-Soria	5		H03, H23	II
E. mediohispanicum	Em34	North	Guadalajara-Soria	5		H02, H56	III
E. mediohispanicum	Em35	North	Guadalajara-Soria	5		H02, H04	II, III
E. mediohispanicum	Em36	South	Sierra Nevada	5		H15	III
E. mediohispanicum	Em37	South	Sierra Nevada	4		H24	II
E. mediohispanicum	Em38	South	Sierra Nevada	5		H01, H38	I
E. mediohispanicum	Em39	South	Sierra Nevada	5		H09	III
E. mediohispanicum	Em40	South	Sierra Nevada	5		H01, H25	I
E. mediohispanicum	Em41	North	Lérida	5		H04	II
E. mediohispanicum	Em42	North	Lérida	5		H04	II
E. mediohispanicum	Em43	North	Lérida	5		H11	II
E. mediohispanicum	Em44	North	Lérida	5		H04	II
E. mediohispanicum	Em45	North	Guadalajara-Soria	5		H02, H23	II, III
E. mediohispanicum	Em46	North	Guadalajara-Soria	5		H02	III
E. mediohispanicum	Em47	North	Guadalajara-Soria	4		H39, H40	II
E. mediohispanicum	Em48	North	Guadalajara-Soria	5	1	H02, H03	II, III
E. mediohispanicum	Em49	North	Guadalajara-Soria	5		H03, H30	II, III
E. mediohispanicum	Em50	North	Guadalajara-Soria	5	1	H16	IV
E. mediohispanicum	Em51	South	Sierra Espuña	5		H31, H41	V
E. mediohispanicum	Em52	South	Sierra Cazorla	5		H26, H57	IV
E. mediohispanicum	Em53	North	Aragón	5		H03	II
E. mediohispanicum	Em54	North	Aragón	5		H17	II
E. mediohispanicum	Em55	North	Aragón	5		H03, H42	II
E. mediohispanicum	Em56	North	Aragón	5		H03, H43, H58	I, II
E. merxmuelleri	Emx01	North	Sierra de Gredos	4	1	H02, H59, H60	I, II, III
E. merxmuelleri	Emx02	North	Sierra de Gredos	5	1	H38, H44, H61, H62	I
E. nevadense	En05	South	Sierra Nevada	5	1	H08	I
E. nevadense	En10	South	Sierra Nevada	5	1	H18	I
E. rondae	Er02	South	Sierra de Tejeda	3	1	H63, H64, H65	I, II
E. rondae	Er03	South	S. de Grazalema	5	1	H19	III
E. ruscinonense	Eru01	North	Montseny	5	1	H45, H66, H67, H68	II
E. ruscinonense	Eru02	North	Montseny	5	1	H03	II
E. gomezcampoi	Eg002	North	Sierra de Font Roja	5	1	H27, H69	II
E. gomezcampoi	Eg003	North	Sierra Martés	5	1	H20	II

Part IV

DISCUSSION

7

DISCUSIÓN GENERAL

El objetivo principal de esta Tesis Doctoral ha sido desentrañar la historia evolutiva de *Erysimum mediohispanicum* atendiendo a la interacción entre morfología floral y polinizadores como motor evolutivo. Para ello hemos utilizado la población como unidad de estudio y hemos recogido la variación geográfica al incluir poblaciones localizadas en todo el área de distribución de la especie. Por lo tanto, cuatro son los pilares en los que se sustenta nuestro estudio. En primer lugar el análisis genético, con dos vertientes: las relaciones de parentesco entre poblaciones, por un lado, y la estructuración y diversidad poblacionales, por otro. En segundo lugar, nuestro estudio se basa en la descripción fenotípica, tanto de la media como de la variación, de rasgos relacionados con el tamaño de la planta, el tamaño de la flor y la forma de la corola. El tercero comprende el estudio de la identidad, la diversidad, la riqueza y la abundancia de los visitantes florales de cada población. En los tres casos hemos estudiado el efecto de la geografía (tales como distancia entre poblaciones o la altitud de las mismas), por lo que podemos considerar la localización geográfica como el cuarto pilar de esta Tesis Doctoral. Los objetivos fundamentales de esta discusión serán: 1) discutir aquellos resultados que, por su extensión, no fueron discutidos en los capítulos en los que se obtuvieron; 2) incluir la información relativa a ploidía y composición genética poblacional para mejorar la inferencia filogeográfica; 3) testar la existencia de relaciones entre los datos obtenidos en los distintos capítulos de esta Tesis, esto es, la filogeografía y las diversidades genética, fenotípica y de polinizadores.

Diferencias de ploidía entre poblaciones y sus efectos

Aunque un estudio riguroso de la ploidía debe basarse en la cuantificación explícita de la cantidad de ADN (por ejemplo, mediante citometría de flujo; Williams and Waller, 2012), el número de loci microsatélites por individuo en *E. mediohispanicum* nos ha permitido distinguir entre individuos diploides y poliploides. En plantas que muestran variación en ploidía, los linajes parentales (generalmente de ploidías menores) y los derivados poliploides pueden coexistir en la misma población (Weiss et al., 2002; Halverson et al., 2008; Sudová et al., 2010) o estar geográficamente separados (Lihová et al., 2003; Yeung et al., 2005; Buggs and Pannell, 2006; Balao et al., 2010). Algunos autores han interpretado la coexistencia de ploidías en una misma población como una característica exclusiva de algunas especies (Borrill and Lindner, 1971; Husband and Schemske, 1998; Hardy et al., 2000; Stuessy et al., 2004; Keeler, 2004) mientras que para otros es un fenómeno infrecuente dentro de un género, pero común a varias especies (Rothera and Davy, 1986; Lumaret et al., 1987; Schranz et al., 2005). Parece, por tanto, que todos los estudios coinciden en apuntar que es más frecuente encontrar poblaciones compuestas principalmente por individuos de la misma ploidía, aunque la especie muestre variación interpoblacional. Este patrón se ha encontrado incluso en especies con largas series poliploides, como *Dianthus broteri*, que presenta poblaciones diploides, triploides, tetraploides, hexaploides y dodecaploides (Balao et al., 2009, 2010). Éste es el caso también de *E. mediohispanicum*, donde ningún individuo poliploide fue encontrado en poblaciones catalogadas como diploides (Capítulo 5). En esta especie, además, poblaciones de diferente ploidía se distribuyen en distintas áreas geográficas (Capítulo 5). Más concre-

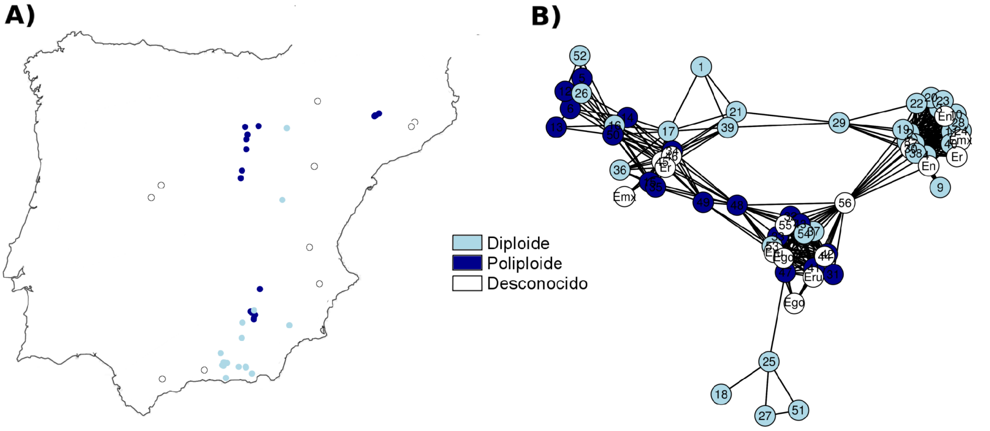


Figura 7.1: Distribución de las poblaciones diploides y poliploides de *E. mediohispanicum* estudiadas de acuerdo a: A) la localización geografía y B) la posición dentro de la red de percolación obtenida en el Capítulo 6.

tamente, las poblaciones del norte son mayoritariamente poliploides, mientras que en el sur abundan las poblaciones diploides (Fig.7.1A). Las sierras de Cazorla, Segura y la Guillimona, cuyas poblaciones fueron catalogadas como diploides (Nieto-Feliner, 1993), componen el único territorio geográfico estudiado en el que pueden encontrarse poblaciones diploides y poliploides. Es llamativo también que las únicas poblaciones diploides encontradas en el norte de la Península se ubiquen en la zona central de dicha región (las poblaciones Em53 y Em55, en Áragón; Fig.7.1A).

Tal y como sucede en *E. mediohispanicum*, otras especies de la Península Ibérica muestran un incremento en el número ploídico al desplazarse de sur a norte (por ejemplo, *Arenaria tetraquetra*; Vargas, 2003). Sin embargo, también existe el patrón contrario, con mayores ploidías en el sur que en el norte (como sucede en *Brachypodium distachyon*, Manzaneda et al., 2012; *Mercurialis annua*, Buggs and Pannell, 2006), así como otros patrones más complejos (por ejemplo, *Carda-*

mine pratensis, Lihová et al., 2003; *Dianthus broteri*, Balao et al., 2009). Esta falta de patrones generales ha contribuido a mantener la incertidumbre acerca de si el mecanismo que subyace a la diferente distribución geográfica de citotipos es adaptativo o si es consecuencia de otros procesos ecológicos (Lumaret et al., 1987; Baack and Stanton, 2005; Buggs and Pannell, 2007; Raabová et al., 2008; Duchoslav et al., 2010; Soltis et al., 2010). Algunas de las ventajas evolutivas que las hipótesis adaptativas atribuyen a los distintos citotipos se fundamentan en los efectos que la ploidía tiene sobre rasgos fenotípicos, especialmente sobre aquéllos que condicionan interacciones ecológicas tales como la herbivoría o la polinización (Segraves et al., 1999; Thompson et al., 2004; Kennedy et al., 2006; Münzbergová, 2006; Arvanitis et al., 2007). Para cuantificar el efecto de la ploidía sobre el gremio de visitantes florales de manera adecuada, sería necesario que ambos citotipos se encontraran en simpatría, ya sea de forma natural o experimental. Sin embargo, los datos obtenidos sobre la distribución natural de ploidías en *E. mediohispanicum*, con poliploides en el norte de la península y diploides del sur, puede hacer que los polinizadores que visitan cada citotipo sean distintos debido a diferencias en la composición de insectos específica de cada región geográfica. Sin embargo, a pesar de la distribución disjunta de citotipos, hemos explorado la variación en fenotipo y polinizadores entre poblaciones con distinta ploidía.

La Tabla 7.1 resume (para cada citotipo) los valores medios y las varianzas de los rasgos fenotípicos medidos en el Capítulo 5, así como la frecuencia de visitas de cada grupo funcional. En primer lugar, hemos encontrado que las poblaciones poliploides tienen escapos significativamente más largos (t=-3.23; p=0.002), y tubos de corola de mayor diámetro (t=-3.25; p=0.002) que las poblaciones diploides. También hemos encontrado diferencias significativas para los componentes de forma RW3 (t=3.86; p=3.3e-04) y RW4 (t=6.64; p=2.6e-08). Estas diferencias fenotípicas podrían jugar un papel esencial en la

A)

	Todas	Poliploides	Diploides
Diametro de escapo	0.751 ± 0.156	0.717 ± 0.161	0.791 ± 0.137
Altura de escapo	3.616 ± 0.315	**3.771 ± 0.234**	**3.517 ± 0.335**
Número de flores	3.798 ± 0.402	3.710 ± 0.475	3.871 ± 0.352
Diámetro de corola	12.74 ± 1.195	13.19 ± 1.069	12.402 ± 1.182
Anchura del tubo de corola	1.281 ± 0.453	**1.519 ± 0.455**	**1.123 ± 0.404**
longitud del tubo de corola	11.43 ± 0.893	11.71 ± 0.859	11.26 ± 0.905
RW1	-0.003 ± 0.044	0.004 ± 0.052	-0.005 ± 0.034
RW2	-0.002 ± 0.023	0.005 ± 0.025	-0.004 ± 0.020
RW3	-0.007 ± 0.020	**-0.019 ± 0.016**	**0.001 ± 0.020**
RW4	-0.012 ± 0.028	**-0.034 ± 0.018**	**0.003 ± 0.022**

B)

	Todas	Poliploides	Diploides
Hormigas	0.169 ± 0.126	0.220 ± 0.173	0.141 ± 0.081
Escarabajos	0.327 ± 0.229	0.291 ± 0.210	0.347 ± 0.241
Bombílidos	0.078 ± 0.116	0.044 ± 0.064	0.097 ± 0.134
Mariposas	0.072 ± 0.099	0.061 ± 0.131	0.078 ± 0.077
Sírfidos	0.031 ± 0.062	0.050 ± 0.089	0.020 ± 0.037
Abejas grandes	0.091 ± 0.108	0.062 ± 0.051	0.107 ± 0.127
Abejas pequeñas	0.113 ± 0.133	0.163 ± 0.121	0.085 ± 0.134
Avispas	0.010 ± 0.020	0.009 ± 0.017	0.010 ± 0.021
Otros	0.111 ± 0.104	0.101 ± 0.121	0.116 ± 0.096

C)

	Todas	Poliploides	Diploides
Riqueza específica	30 ± 9	33 ± 8	28 ± 9
Hulbert’s PIE	0.847 ± 0.128	0.874 ± 0.096	0.832 ± 0.142
Dominancia	0.442 ± 0.168	0.433 ± 0.155	0.447 ± 0.178

Tabla 7.1: Comparación de poblaciones diploides y poliploides mediante valores poblacionales medios y desviaciones estándar, correspondientes a A) rasgos fenotípicos y gremio de polinizadores representados por B) frecuencia de visitas de cada grupo funcional y C) riqueza, diversidad y dominancia. Los valores en negrita indican la existencia de diferencias significativas entre poblaciones diploides y poliploides.

atracción diferencial de polinizadores. Sin embargo, la comparación mediante MANOVA de las visitas realizadas por cada grupo funcional de polinizadores en poblaciones diploides y poliploides no mostró diferencias significativas (aunque los valores fueron cercanos a la significación: r^2=0.04; p=0.07). Más clara es la carencia de efecto de la ploidía (tras comparar los valores medios de cada población mediante tests de Wilcoxon) en los valores de diversidad de polinizadores (Hurlbert's PIE: W=196; p = 0.60), en la riqueza específica (W=223; p = 0.16) y en la dominancia (W=170; p = 0.90), a pesar de que dichos análisis no separan los efectos espaciales y de ploidía. Aunque son necesarios diseños experimentales enfocados específicamente a este objetivo, podemos concluir que, a pasar de las diferencias en fenotipo entre los dos citotipos y sus diferentes distribuciones geográficas, no hay grandes efectos de la ploidía sobre los grupos funcionales de insectos que visitan las poblaciones de *E. mediohispanicum*.

RELACIONES FILOGENÉTICAS Y DIVERSIDAD GENÉTICA POBLACIONAL

Para obtener las relaciones evolutivas entre poblaciones de una misma especie no es recomendable recurrir a la secuenciación de marcadores utilizados comúnmente en reconstrucciones filogenéticas, ya que, debido a la poca divergencia entre los taxones estudiados, es necesario secuenciar un gran número de nucleótidos para obtener resultados concluyentes. En el caso de *Erysimum*, el análisis filogenético de varias especies usando marcadores nucleares y plastidiales (4 Kb en total, Capítulo 6) no sólo mostró baja resolución a la hora de dilucidar las relaciones entre poblaciones de *E. mediohispanicum*, si no que en algunos casos tampoco proporcionó soportes aceptables entre distintas especies (Fig. 6.6 en la página 268). Esto corrobora la información botánica previa, que daba a esos taxones el estatus de complejo de especies, debido a la existencia de individuos que difícilmente pueden ser identificados como pertenecientes a una u otra especie (Clot, 1991; Nieto-Feliner, 1993). A pesar de la poca resolución obtenida, la topología del árbol construido nos ha permitido apoyar resultados encontrados con otros métodos (Capítulo 6).

Para obtener mayor resolución, los estudios filogeográficos utilizan generalmente marcadores basados en tamaño de banda (alozimas, microsatélites, RFLP , RAPD, AFLP, . . .), por ser más variables que los basados en secuencias. Sin embargo, todos los marcadores basados en tamaño de banda comparten una notable desventaja: la ambigüedad que presentan a la hora de reconstruir los estados ancestrales, por lo que deducir los patrones genealógicos con certeza a partir de ellos puede resultar problemático (Zhang and Hewitt, 2003). A pesar de ello, este tipo de marcadores han sido muy ampliamente utilizados para estimar la diversidad genética poblacional, especialmente los microsatélites, debido a sus mejores cualidades de reproducibilidad, polimorfismo, requerimientos técnicos y calidad necesaria del ADN inicial (Agarwal et al., 2008). Para *E. mediohispanicum* hemos desarrollado una batería de microsatélites que han demostrado ser polimórficos y estar en equilibrio Hardy-Weinberg en la mayoría de las poblaciones estudiadas (Capítulo 2, Table 5.2 en la página 200, and Table 5.S3 en la página 239). Sin embargo, el uso de los marcadores microsatélites en *E. mediohispanicum* se ha visto limitado por la existencia de variación en el número ploídico, debido a la incertidumbre asociada a la asignación genotípica de los individuos poliploides (Capítulo 5). De acuerdo al análisis de microsatélites en las poblaciones diploides, la mayor parte de la variación genética de la especie se encuentra dentro de población (Tabla 5.5 en la página 208). Hemos atribuido el mantenimiento de dicha variación intrapoblacional a la relación encontrada entre diversidad genética y fenotípica (Fig. 5.S3 en la página 244) y al amplio gremio de polinizadores, debido a que la heterogeneidad en visitantes florales supone que también sean seleccionados una gran variedad de rasgos morfológicos (y por lo tanto de genotipos; Capítulo 5).

El análisis de los marcadores microsatélites tampoco nos permitió establecer relaciones de parentesco detalladas entre poblaciones, quizá como consecuencia de la alta diversidad genética intrapoblacional y la baja divergencia interpoblacional mencionada (Tabla 5.5 en la página 208). Tanto el análisis de componentes principales como la inferencia bayesiana del número de unidades genéticas muestran el mismo patrón, con cierta separación entre las poblaciones de Aragón, Sierra de Cazorla y Sierra Espuña, que aparecen relativamente alejadas de las de Sierra Nevada (Fig. 5.4 en la página 202). Hemos atribuido la poca estructuración genética encontrada entre poblaciones de Sierra Nevada a la existencia de flujo reciente entre estas poblaciones, al menos a nivel de polen (Capitulo 5).

Hemos representado las distancias genéticas calculadas a partir de los marcadores microsatélites (matriz de distancias de Bruvo, Capítulo 5) como una red de percolación (Fig.3.2) para permitir una comparación más directa con los resultados obtenidos del estudio de la región plastidial (Capítulo 6). Esta nueva representación coincide con las anteriores (PCA) al mostrar sólo el patrón genético a gran escala, con un módulo muy interconectado donde se sitúan todas las poblaciones de Sierra Nevada. Unido a él, las tres poblaciones diploides de la Sierra de Cazorla conforman un pequeño módulo poco interconectado, mientras que las poblaciones de Sierra Espuña y Aragón aparecen como nodos independientes (Fig.3.2).

En resumen, el análisis de microsatélites muestra que las poblaciones diploides son muy diversas genéticamente y se agrupan de acuerdo a su procedencia geográfica.

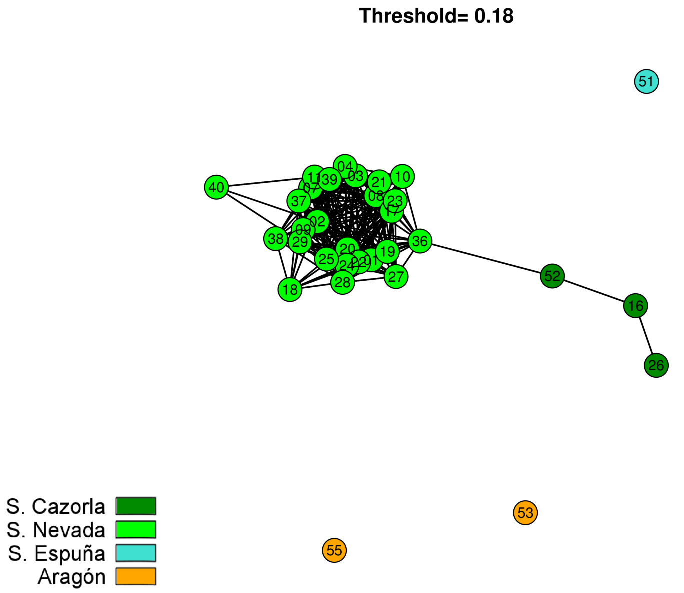


Figura 7.2: Red de percolación (umbral=0.18) construida a partir de las distancias de Bruvo calculadas a partir de los genotipos individuales de poblaciones diploides (Capítulo 5). Los números identifican las poblaciones (Tabla 5.S1 en la página 237) mientras que los colores indican el origen geográfico de las mismas.

MÉTODO DE RECONSTRUCCIÓN FILOGEOGRÁFICA BASADO EN INDELS

Para intentar completar la información proporcionada por los marcadores filogenéticos tradicionales y los microsatélites hemos recurrido a una región plastidial, lo que, además, permite la comparación de individuos independientemente de sus ploidías. Debido a las limitaciones de las sustituciones para obtener información filogeográfica, hemos añadido a este tipo de mutación la información procedente de inserciones y deleciones (indels), para lo que hemos desarrollado un nuevo método basado en distancias.

Para realizar este análisis hemos secuenciado una región muy bien caracterizada en plantas: la localizada entre los transferentes de leucina y fenilalanina, o trnL-trnF IGS (Taberlet et al., 1991; Koch et al., 2005, 2007). Dicha región no sólo es abundante en indels, si no que éstos se producen principalmente involucrando un número de nucleótidos bastante constante, similar al del trnF funcional (Dobeš et al., 2004, 2007; Koch et al., 2007; Schmickl et al., 2008). La poca distancia evolutiva existente entre las poblaciones estudiadas reduce la probabilidad de homoplasia, es decir, de que procesos mutacionales diferentes den lugar a secuencias idénticas. Para poder combinar la información procedente de sustituciones e indels, nuestro método se basa en la obtención de matrices de distancias para ambos tipos de mutaciones, cuya congruencia puede ser testada. En caso de que dicha congruencia exista, ambas matrices pueden ser fácilmente combinadas para obtener una única matriz de distancia genética total, que puede usarse para construir árboles o redes que representen las relaciones evolutivas entre las secuencias (Capitulo 4) o las poblaciones (Capitulo 6) estudiadas.

Este método tiene una primera parte muy dependiente de sistema (el alineamiento) y una segunda parte mucho más general (obtención y representación de la matriz de distancias). Para esta segunda etapa estamos implementando las herramientas necesarias en un paquete de R (Capítulo 3), lo que permitirá repetir el proceso de manera sencilla para cualquier alineamiento. La precisión del método no ha sido testada utilizando secuencias de historia conocida debido a que no existen programas para simular secuencias de acuerdo con los procesos evolutivos que suceden en la región estudiada. Sin embargo, creemos que nuestro método ha sido validado por la obtención de resultados coherentes al analizar las mismas poblaciones mediante métodos tradicionales como el análisis de sustituciones tanto de marcadores nucleares como plastidiales (Capítulo 6), tal como se describe en el siguiente apartado.

comparación entre marcadores genéticos en e. *mediohispanicum*

Esta Tesis Doctoral ha generado múltiples fuentes de información genética que nos han permitido inferir las relaciones evolutivas entre poblaciones de *E. mediohispanicum*, así como testar a grandes rasgos nuestro método. Para comparar los resultados obtenidos con cada pareja de marcadores, hemos utilizado las secuencias comunes a los mismos para calcular matrices de distancias bajo el modelo evolutivo Kimura 2-parámetros (Kimura, 1980) usando el paquete de R *ape* (Paradis et al., 2004). A continuación hemos usado dichas matrices de distancia para testar la congruencia mediante CADM (Legendre and Lapointe, 2004) también implementado en *ape* (Paradis et al., 2004). El CADM es una extensión del test de Mantel de correspondencia de matrices (Mantel, 1967) y asume como hipótesis nula la incongruencia completa entre matrices de distancias (Campbell et al., 2011). Para comparar las topologías visualmente, hemos construido árboles NJ de cada matriz de datos usando MEGA5 (Tamura et al., 2011).

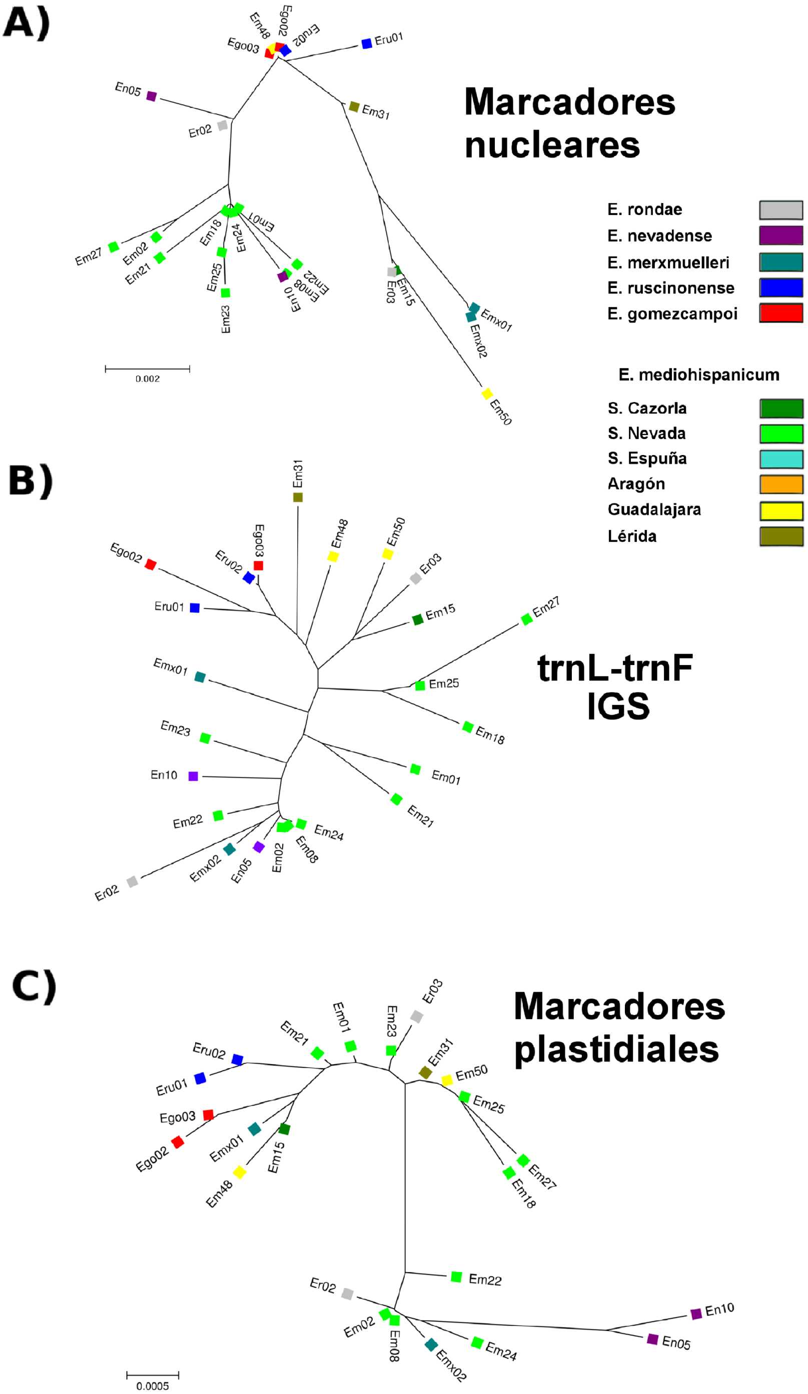
En primer lugar, hemos encontrado una alta congruencia entre marcadores nucleares y plastidiales (W=0.72; p=0.002; Tabla 7.2), aunque la representación en forma de árbol permite ver algunas discrepancias (Fig.7.3). En general podemos ver que el árbol basado en marcadores nucleares tiene mayor correspondencia con la distribución geográfica que el plastidial (Fig.7.3). De acuerdo a los marcadores nucleares, todas las poblaciones de *E. mediohispanicum* procedentes de Sierra Nevada son monofiléticas y aparecen separadas de poblaciones del norte de la Península (Fig.5.5A). Sin embargo, de acuerdo a los marcadores plastidiales, una rama larga separa las poblaciones de Sierra Nevada en dos grupos (Fig.7.4C). En uno de esos dos clados podemos encontrar poblaciones del norte (Em31 y Em50) junto con poblaciones que conforman el Módulo 4 en el análisis de redes del Capítulo 6 (Em18, Em25 y Em27; Fig.5.5C; Fig.6.5). Por lo tanto, los marcadores plastidiales apoyan la relación directa del Módulo 4

	N	W	χ^2	p
nDNA vs cpDNA	**27**	**0.7216**	**505**	**0.0018**
trnF vs cpDNA	**22**	**0.6724**	**309**	**0.0001**
trnF vs nDNA	**22**	**0.6380**	**293**	**0.0018**
trnF vs nDNA vs cpDNA	**22**	**0.4522**	**312**	**0.0005**
trnF vs D_{ST}	**32**	**0.7191**	**712**	**0.0002**
D_{ST} vs cpDNA	10	0.3678	32	0.9713
D_{ST} vs nDNA	10	0.6780	60	0.1008
D_{ST} vs nDNA vs cpDNA	10	0.3730	49	0.2825
D_{ST} vs nDNA vs cpDNA vs trnF	10	0.3695	65	0.0306

Tabla 7.2: Resultados de la congruencia entre marcadores de acuerdo a tests CADM. Para cada conjunto de matrices de distancia comparadas se muestra el número de poblaciones incluidas por matriz (N), el valor de la congruencia (coeficiente de concordancia de Kendall, W), el valor de ji cuadrado (χ^2) y la significación (p). Los valores en negrita indican la existencia de diferencias significativas tras corrección de Bonferroni secuencial. Las matrices de distancia fueron representadas por los siguientes abreviaturas: nDNA = distancia basada en marcadores nucleares (ITS); cpDNA = distancia basada en marcadores plastidiales (ndhF, trnT-trnL); trnF = distancia combinada de sustituciones e indels de la región trnL-trnF IGS; D_{ST} = distancia basada en microsatélites, estimada como D_{ST}.

Figura 7.3: Representaciones NJ de las matrices de distancias basadas en: A) marcadores nucleares; B) región trnL-trnF IGS; C) marcadores plastidiales. Los colores indican la especie o el origen geográfico de las poblaciones de *E. mediohispanicum*.

(Next page) ▶


A)
Marcadores nucleares
Em48
Ego02
Eru02
Ego03
Eru01
En05
Er02
Em31
Em01
Em27
Em02
Em21
Em18
Em25
Em24
Em23
Em22
Em08
En10
Er03
Em15
Emx01
Emx02
Em50
0.002
E. rondae
E. nevadense
E. merxmuelleri
E. ruscinonense
E. gomezcampoi
E. mediohispanicum
S. Cazorla
S. Nevada
S. Espuña
Aragón
Guadalajara
Lérida
B)
trnL-trnF IGS
Em31
Ego03
Em48
Em50
Ego02
Eru02
Er03
Eru01
Em15
Em27
Emx01
Em25
Em18
Em23
Em01
En10
Em21
Em22
Em24
Em08
Emx02
En05
Em02
Er02
C)
Marcadores plastidiales
Em01
Em23
Er03
Em21
Eru02
Em31
Em50
Eru01
Em25
Ego03
Emx01
Ego02
Em15
Em48
Em27
Em18
Em22
Er02
En10
Em02
Em08
Emx02
Em24
En05
0.0005

con poblaciones del norte, lo que hemos atribuido a la existencia de una vía de colonización norte-sur (Capitulo 6). Otra diferencia entre los árboles nuclear y plastidial (a pesar de la alta congruencia encontrada) viene dada por la agrupación de las poblaciones de otras especies en uno o varios clados, como ocurre con *E. nevadense* y *E. merxmuelleri* (Fig.7.3).

Las diferencias entre resultados proporcionados por marcadores plastidiales y nucleares pueden deberse a hibridación o a separación incompleta de linajes, tal y como ya discutimos en el Capítulo 6. Además de estos fenómenos, las señales evolutivas proporcionadas por ambos marcadores pueden no ser idénticas debido a las diferencias en flujo dispersivo existente entre polen y semillas. En *E. mediohispanicum*, la distancia de dispersión de semilla (unos 20 centímetros según Gómez, 2007) es mucho más pequeña que la distancia a la que pueden transportar polen los insectos (más de un kilómetro según Chifflet et al., 2011). Además, en el caso de angiospermas, la herencia del cloroplasto se produce exclusivamente por vía materna, por lo que generalmente carece de recombinación (Avise et al., 1987; Comes and Kadereit, 1998). Por ello es esperable que los marcadores plastidiales reflejen mejor los procesos ocurridos durante la colonización, mientras que la existencia de flujo genético vía polen entre poblaciones cercanas pero con diferentes historias de colonización, borrará dicha huella (Avise, 2000; Kropf et al., 2006). Como consecuencia los marcadores plastidiales tienden a mostrar patrones históricos, mientras los nucleares pueden mostrar una agrupación más coherente con la disposición geográfica.

Pero los marcadores nucleares y de orgánulos revelan diferentes aspectos de la historia evolutiva de los organismos estudiados no sólo por sus diferencias en tipo de herencia, frecuencia de recombinación y tasa de dispersión, si no que también presentan diferentes tasas de mutación y número poblacional efectivo (Zhang and Hewitt, 2003). Debido al menor tamaño poblacional efectivo de los orgánulos, aún cuando el resto de circunstancias permanezcan constantes, la acción de la deriva genética hace que la divergencia entre poblaciones sea menor cuando se mide con marcadores nucleares (Brito and Edwards, 2009). Esto hace que la estructura observada en filogeografías construidas con marcadores nucleares sea menor que cuando se han empleado marcadores de los orgánulos (Moore, 1995), tal y como observamos en nuestro estudio (Capítulos 5 y 6). Además, se ha encontrado que los marcadores plastidiales están más estructurados espacialmente por presentar la mayor parte de su varianza total entre poblaciones en lugar de dentro de población, como sucede con marcadores como las isoenzimas y los microsatélites (Comes and Abbott, 1998; Oddou-Muratorio et al., 2001; Nybom, 2004). La tasa de mutación del cloroplasto es aproximadamente la mitad de la observada en el núcleo, que a su vez es tres veces mayor que la de la mitocondria (Wolfe et al., 1987). Por todo, los orgánulos proporcionan una visión excepcionalmente clara de la historia evolutiva de las poblaciones (Brito and Edwards, 2009).

La información obtenida con el análisis de la región trnL-trnF IGS es congruente tanto con la obtenida con marcadores nucleares como con plastidiales (tabla 7.2), presentando valores similares en ambos casos (W=0.64; p=0.0018 y W=0.67; p=0.0001, respectivamente). En el árbol NJ de dicha región (Fig.7.4B) aparecen características ya comentadas en la filogenia basada en marcadores plastidiales (Fig.7.4C) y que no aparecen en el árbol nuclear (Fig.7.4A). Tal es el caso de las poblaciones del Módulo 4 (Em18, Em25 y Em27) separadas del resto de Sierra Nevada o del origen no monofilético de las poblaciones de *E. merxmuelleri*. Además de dichas correspondencias, el árbol obtenido con la región trnL-trnF IGS presenta características del árbol nuclear que no aparecen en el plastidial, como sucede con la rama que relaciona las poblaciones Er03, Em15 y Em50 o como el hecho de que las dos poblaciones de *E. nevadense* no formen un clado pro-

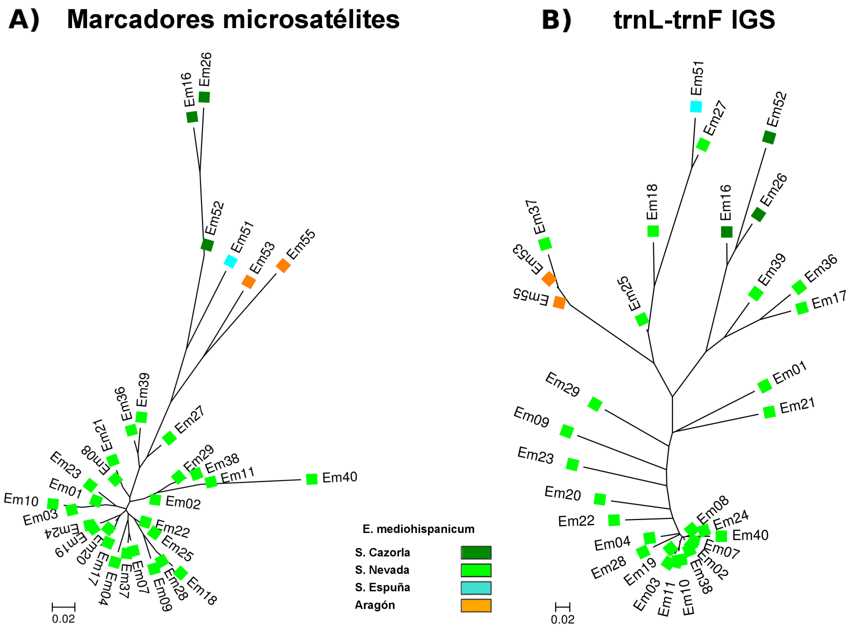


Figura 7.4: Representaciones NJ de las matrices de distancias basadas en: A) región trnL-trnF IGS de poblaciones diploides; B) distancias de Bruvo inferidas de los genotipos individuales de las mismas poblaciones. Los colores indican la especie o el origen geográfico de las poblaciones de *E. mediohispanicum*.

pio (Fig.7.3). La congruencia global de los tres marcadores, aunque menor, sigue siendo significativa (W=0.45; p=0.0005; Tabla 7.2).

La comparación de los árboles obtenidos a partir de la región trnL-trnF IGS y los microsatélites (Fig.7.4) deja patente la diferencia de señal comentada anteriormente entre marcadores nucleares y plastidiales. En el caso de los microsatélites sólo se aprecia una señal de gran escala, con ramas por lo general muy cortas entre poblaciones de Sierra Nevada, que aparecen como monofiléticas (Fig.7.4A). El análisis de las mismas poblaciones con el trnL-trnF IGS coincide en asignar las ramas más largas a las poblaciones de Cazorla, Aragón y Sierra Espuña, aunque en esta ocasión algunas poblaciones de Sierra Nevada se agrupan con ellas (Fig.7.4B). A pesar de la diferente resolución, la información proporcionada por ambos marcadores es congruente (tanto como lo son los marcadores nucleares y los plastidiales entre sí: W=0.72; p=0.0002; Tabla 7.2). Por el contrario, las distancias entre poblaciones obtenidas usando los microsatélites son incongruentes con las calculadas para los marcadores filogenéticos tanto nucleares como plastidiales (Tabla 7.2). Como era de esperar, la congruencia de los marcadores microsatélites, aunque no significativa, es mayor con los marcadores nucleares (W=0.68; p=0.1) que con los plastidiales (W=0.37; p=1.0) (Tabla 7.2). Debido a que para cada análisis de congruencia sólo hemos comparado las poblaciones comunes a los marcadores estudiados, los tamaños de muestra de cada test de congruencia son diferentes (Tabla 7.2). En concreto, la comparación de micros con los marcadores filogenéticos se ha realizado entre sólo 10 poblaciones, lo que puede influir en que dichos resultados sean no significativos.

En resumen, la información proporcionada por el análisis de la región trnL-trnF IGS siguiendo el método propuesto en esta Tesis Doctoral es congruente con todos los demás marcadores, aunque entre ellos presenten incongruencias. La mayor parte de las diferencias encontradas entre los marcadores pueden atribuirse a las diferentes naturalezas de los marcadores nucleares y plastidiales. En base a todos los resultados obtenidos hemos podido extraer la información filogeográfica recogida a continuación.

FILOGEOGRAFÍA

De acuerdo a la región trnL-trnF IGS, las poblaciones pertenecientes a las especies del complejo nevadense de *Erysimum* pueden agruparse en cuatro linajes (módulos según la representación de redes, Fig. 6.5 en la página 264), cada uno con una composición característica de haplogrupos y con diferente distribución geográfica. El módulo 1 (compuesto sobre todo por secuencias clasificadas dentro del Haplogrupo I) se encuentra mayoritariamente en Sierra Nevada, mientras que el módulo 2 (Haplogrupo II) se distribuye por la costa levantina y el noreste de la Península (Capítulo 6). El módulo 3 (compuesto mayoritariamente por los Haplogrupos III y IV) aparece en áreas del interior de la Península Ibérica (entre Jaén y Soria) y presenta indicios de una diferenciación incipiente entre las poblaciones de las sierras del sur y del norte (como indica la existencia de diferentes haplogrupos muy emparentados en ambas regiones; Capítulo 6). Finalmente, el módulo 4 (Haplogrupo V) se restringe a la costa sureste (Provincias de Murcia, Almería y Granada; Capítulo 6).

La distribución de esos cuatro grupos coincide con las principales unidades biogeográficas descritas por Rivas-Martínez (1973) (Fig.7.5). Dicho estudio se basa en las características geologías, climatológicas, edáficas y botánicas del territorio para clasificarlo en unidades jerárquicas denominadas reinos, regiones, superprovincias, provincias, sectores y subsectores. Toda la Península Ibérica pertenece al reino Holártico y está dividida en dos regiones, la Euro-siberiana (que se extiende por las regiones costeras del norte peninsular) y la Mediterránea (que ocupa el resto del territorio; Fig.7.5). Cada región, a su vez, está dividida en dos superprovincias (Rivas-Martínez, 1973). Las poblaciones de *Erysimum* estudiadas se encuentran todas en la región Mediterránea, perteneciendo las poblaciones de los módulos 1 y 3 a la Ibero-atlántica y los módulos 2 y 4 a la superprovincia Ibero-levantina (Fig.7.5). De hecho el módulo 4 se localiza en una única provincia (denominada murciano-almeriense por Rivas-Martínez, 1973). Por lo tanto, los resultados obtenidos con nuestro método parecen coherentes no sólo con otros resultados genéticos (métodos filogenéticos tradicionales aplicados a otros marcadores), si no también con los estudios corológicos.

También hemos podido determinar que las especies que, junto a *E. mediohispanicum*, conforman el denominado complejo nevadense

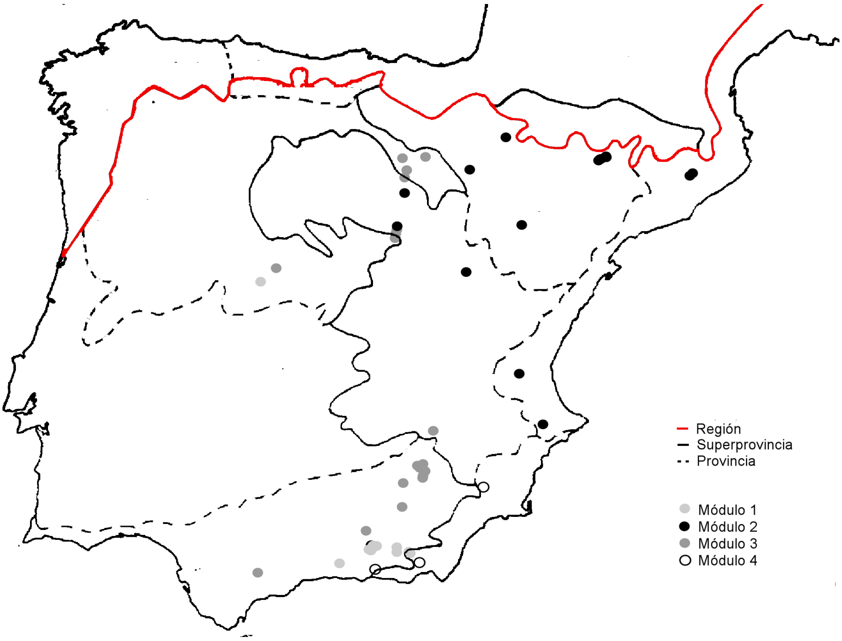


Figura 7.5: Superposición de las principales unidades biogeográficas descritas por Rivas-Martínez (1973) y la distribución geográfica de los cuatro linajes (módulos de la red de percolación, Capitulo 6) encontrados en *E. mediohispanicum*. Los tipos de línea representan la jerarquía de las unidades biogeográficas y los puntos en escala de grises representan los linajes.

no aparecen como linajes independientes, si no que se agrupan con otras poblaciones de acuerdo a su proximidad geográfica (Fig. 6.5 en la página 264). Esto nos ha llevado a proponer que dichas especies pueden derivar de *E. mediohispanicum* (con un área de distribución mucho mayor que el resto de especies del complejo), vinculadas a distintos sistemas montañosos, en las que algunas poblaciones permanecieron aisladas durante determinadas etapas geológicas. Alternativamente, el patrón observado puede haber sido producido por la hibridación de *E. mediohispanicum* con especies de *Erysimum* previamente establecidas en dichos sistemas montañosos (Capítulo 6).

Todas las interpretaciones de los distintos datos genéticos obtenidos apuntan a mecanismos de dispersión como los que más probablemente subyacen a la distribución actual de linajes en las especies del complejo nevadense (Capítulo 6). Mayor incertidumbre tenemos a la hora de asignar las rutas de colonización. El carácter adpreso de los frutos sugiere que las plantas poliploides del norte provienen de las diploides del sur, ya que dicho rasgo es compartido por distintas especies de las cordilleras béticas pero es exclusivo de *E. mediohispanicum* en el norte (Nieto-Feliner, 1993). Además, si utilizamos el número de haplotipos por haplogrupo (Capítulo 4) como indicador de la antigüedad de los mismos, obtenemos que el Haplogrupo I (distribuido principalmente en Sierra Nevada 6.5 en la página 264) es el más antiguo, seguido del Haplogrupo II (Levante y noreste). Las mismas conclusiones se ven apoyadas por el hecho de que ambos haplogrupos sean los únicos que mostraron variación en la región más conservada del alineamiento (motivos T22-T28 previos a $\Psi 1$; Fig. 4.7B en la página 141). Además, esos mismos haplogrupos mostraron un mayor número de conexiones por nodo (Capítulo 4).

En base a los resultados obtenidos en el Capítulo 6, hemos propuesto dos rutas alternativas de colonización. En ambas (y de acuerdo con los marcadores plastidiales, Fig. 7.4C en la página 311) el linaje del litoral murciano-almeriense (módulo 4) deriva de las poblaciones de Levante (módulo 2), tras una migración norte-sur a la que siguió una segunda ruta por el centro de la península que pudo ser sur-norte (derivando el módulo 3 del módulo 4) o norte-sur (en cuyo caso el módulo 3 se originó a partir del módulo 2; Fig. 6.8 en la página 278). Ambas rutas de colonización derivan de la interpretación en forma de red de las matrices de distancias (Fig. 6.5 en la página 264). Hemos preferido dicha representación a la de árbol porque en ella se recogen mejor las relaciones complejas existentes entre poblaciones, que no siempre están conectadas por procesos de bifurcación, como suponen los árboles (Morrison, 2005; Mardulyn, 2012).

Además, hemos preferido la representación de red de la matriz de distancias genéticas basada en la región trnL-trnF IGS ya que el árbol (Fig.7.6) muestra algunos aspectos difíciles de explicar a la luz de otros datos obtenidos, como la asignación bayesiana del número de poblaciones (Fig. 5.4 en la página 202) o las regiones geográficas donde se localizan las principales barreras entre poblaciones (Fig. 6.2 en la página 260).

Dicho árbol coincide en mostrar una cierta separación entre las poblaciones que pertenecen a distintos módulos en la red, apareciendo las poblaciones de Sierra Nevada en ramas centrales del árbol y los módulos del interior y Levante en los extremos (Fig.7.6). Si volvemos a asumir que las poblaciones de Sierra Nevada son las más antiguas, esta topología implicaría dos colonizaciones independientes sur-norte, lo que probablemente llevaría al establecimiento de la barrera genética más importante entre dos sistemas montañosos del norte de la Península Ibérica (Fig.7.7A), que deberían representar dos linajes genéticos diferenciados. Este escenario contrasta con el patrón observado, en el que la barrera genética más importante se encuentra entre las poblaciones de Sierra Nevada y las de Cazorla y los linajes genéticos separan Sierra Nevada del resto de la Península (Fig. 6.2 en la página 260).

Además, la topología del árbol requiere la existencia de múltiples eventos independientes de poliploidización para explicar que las poblaciones poliploides ocupen posiciones tan distantes en el árbol y que aparezcan poblaciones diploides y poliploides en las mismas ramas (Fig.7.6). A pesar de que la aparición recurrente de taxones poliploides es algo habitual en las representaciones filogenéticas en forma de árbol (por ejemplo, Van Dijk and Bakx-Schotman, 1997; Segraves et al., 1999; Halverson et al., 2008; Soltis et al., 2010), hemos preferido mantener la representación como red por ser más parsimoniosa, ya que poblaciones de ploidía similar aparecen conectadas de manera directa (Fig.7.1B) y son necesarios menos eventos de poliploi-

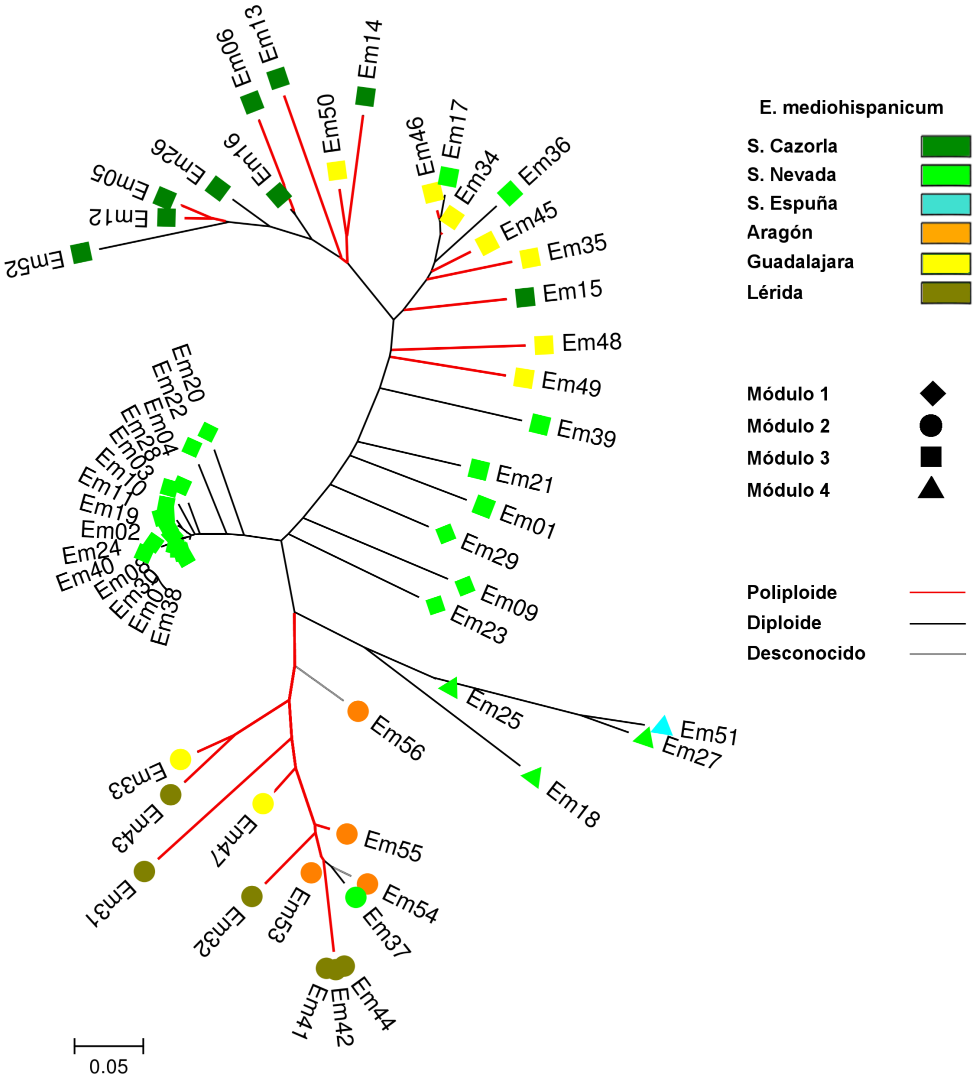


Figura 7.6: Representación como árbol NJ de la matriz de distancia combinada de sustituciones e indels de la región trnL-trnF IGS (Capítulo 6). La procedencia geográfica de cada población se representa en diferentes colores y la pertenencia a módulos de la red (Fig. 6.5 en la página 264) mediante formas geométricas. Las líneas rojas representan ramas que originan poblaciones poliploides.

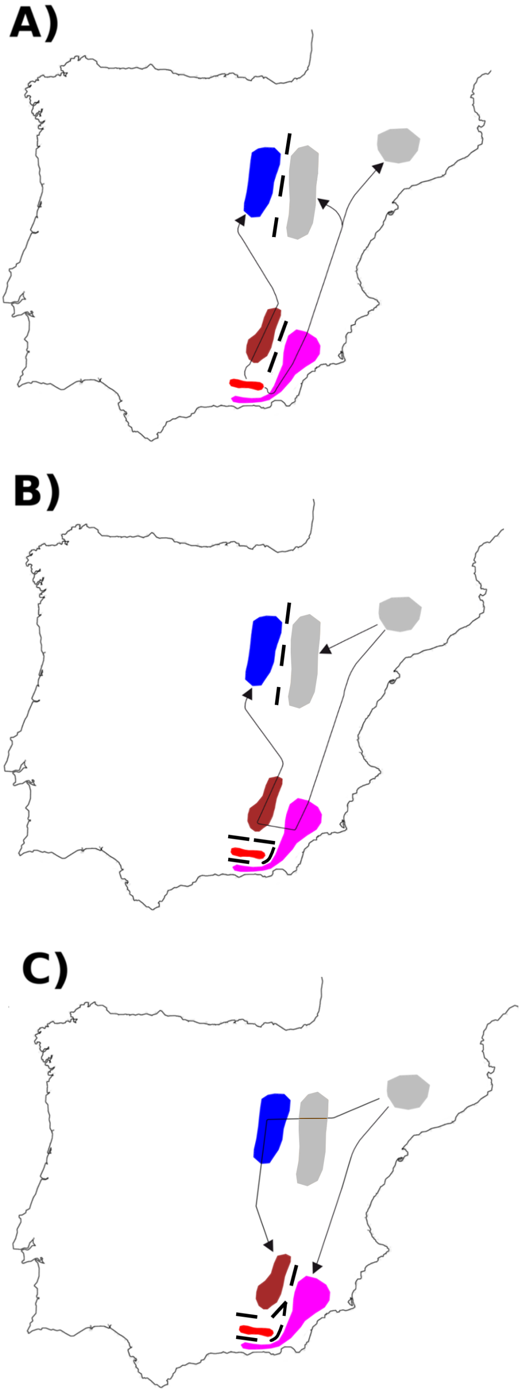


Figura 7.7: Representación de las rutas dispersivas que se deducen de la matriz de distancia combinada de sustituciones e indels de la región trnL-trnF IGS A) de acuerdo al árbol NJ; B) según la red de percolación, asumiendo dispersión sur-norte; B) según la red de percolación, asumiendo dispersión norte-sur. Las líneas discontinuas representan las principales barreras genéticas esperables bajo cada escenario.

dización. Por todo, hemos obviado cualquier referencia anterior a la representación como árbol y hemos utilizado únicamente la de red, más conservadora y coherente con otros resultados obtenidos.

Una de las dos rutas compatibles con la representación de red (Fig. 6.8 en la página 278) asume una migración sur-norte, lo que vuelve a ubicar una de las principales barreras genéticas entre poblaciones del norte, aunque también es previsible la aparición de una segunda barrera en el sur, separando las poblaciones de Sierra Nevada de las de Sierra de Cazorla y las pertenecientes del Módulo 4 (Fig.7.7B). La segunda ruta alternativa propuesta sería compatible con la aparición de las principales barreras genéticas únicamente en la región sur, separando Sierra Nevada de Cazorla y las poblaciones del Módulo 4 (Fig.7.7C). Este último patrón se corresponde exactamente con la localización de barreras genéticas estimadas a partir de la matriz de distancias de la región trnL-trnF IGS (Fig. 6.2 en la página 260). El mismo flujo de dispersión se ve apoyado también por las distancias de microsatélites y por la asignación bayesiana del número de poblaciones genéticamente diferentes (Capítulo 5). En todos los casos, las mayores diferencias aparecen entre poblaciones de Sierra Nevada y de Cazorla, indicando que, antes de ese flujo norte-sur, el área de distribución de la especie debió quedar reducida a dos núcleos, uno en el noreste y otro en Sierra Nevada, con el consiguiente aislamiento de los linajes genéticos de estas dos regiones.

Dicho aislamiento entre linajes genéticos suele vincularse a la supervivencia en distintos refugios glaciales durante el Pleistoceno (Cooper and Hewitt, 1993). Cada vez son más numerosas las evidencias de que en la Península Ibérica existieron múltiples refugios, algunos en regiones del norte (Olalde et al., 2002; Gómez and Lunt, 2007; Feliner, 2011), tal y como parece haber sucedido en *E. mediohispanicum*. En concreto podríamos identificar el área del refugio norte con el foco aparentemente aislado de poblaciones diploides en Aragón (Capítulo 5; Fig.7.8), si aceptamos la hipótesis que sugiere que los organismos diploides se encuentran típicamente en áreas que formaron parte de refugios, mientras que los poliploides son más abundantes en áreas que estuvieron bajo condiciones de glaciación (Ehrendorfer and Lewis, 1980; Parisod et al., 2010). Por lo tanto, se nos muestra un escenario en el que las plantas diploides de Sierra Nevada se dispersaron hacia el norte hasta Aragón (Fig.7.8). La duplicación del material genético que originó los poliploides del norte debió producirse después de la primera migración norte-sur (que originó el módulo 4, diploide), o durante la dispersión de la especie desde Aragón hacia Lérida (por el este) y Guadalajara (por el oeste; Fig.7.8). De esta manera, la distribución de ploidías parece apoyar también la ruta de colonización según la cual las poblaciones del interior de la península se originaron a partir de un flujo norte-sur desde las regiones poliploides del norte (Fig.7.8).

Uno de los aspectos más difíciles de interpretar en la filogeografía de *E. mediohispanicum* es el encontrado en las Sierras de Cazorla, Segura y La Guillimona, donde poblaciones diploides y poliploides están separadas por pequeñas distancias (Fig.5.1). Parece improbable que las poblaciones diploides de esta región se relacionen con la existencia de un refugio glacial y las poliploides con posteriores eventos de duplicación, ya que todas las representaciones de distancias (tanto los árboles, Fig.7.6, como las redes, Fig.6.5) apuntan a que dichas poblaciones son los nodos más terminales, estrechamente emparentados con las poblaciones poliploides de Soria y Guadalajara. El patrón observado podría explicarse también mediante una mezcla de linajes poliploides procedentes del norte y diploides emparentados con poblaciones del módulo 4 (lo que mantendría las grandes distancias con Sierra Nevada). Sin embargo, tanto las poblaciones diploides como las poliploides de los sistemas montañosos adyacentes a Cazorla presentan el mismo Haplogrupo mayoritario (el IV), lo que indica que comparten ancestría por vía materna a pesar de las diferencias en ploidía. Otra hipótesis que explicaría la distribución de ploidías sin

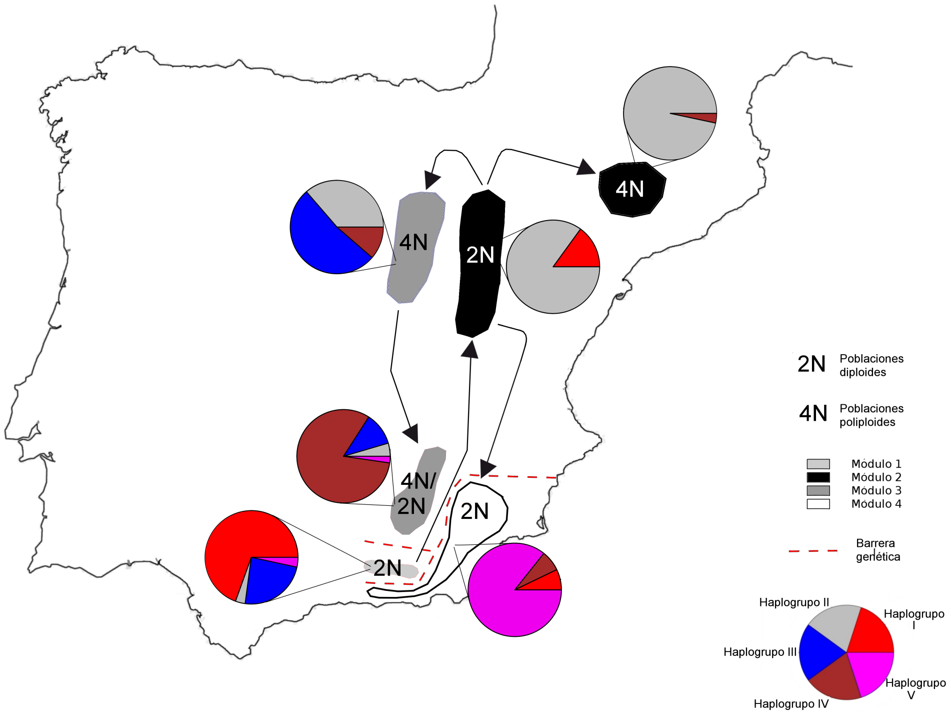


Figura 7.8: Representación de las rutas dispersivas mejor soportada por los datos obtenidos en los distintos capítulos de esta Tesis Doctoral. Dichos datos incluyen la distribución geográfica de los módulos de la red, la ploidía y la composición haplotípica de cada territorio geográfico muestreado, así como las principales barreras genéticas encontradas, todos ellos recogidos en esta figura.

los inconvenientes anteriores sería la diploidización, uno de los procesos más desconocidos de la evolución genómica (Wolfe, 2001). Este mecanismo lleva a una especie poliploide a tener un comportamiento meiotico similar a las especies diploides, produciéndose además pérdida de material genético y reagrupaciones cromosómicas (Grant, 1981; Wolfe, 2001; Levy and Feldman, 2002; Chen, 2007; Doyle et al., 2008; Parisod et al., 2010; Mandáková et al., 2010). Debido a que es un proceso paulatino, sería de esperar en este caso que las poblaciones catalogadas como poliploides en las sierras adyacentes a Cazorla presentaran cantidades de ADN menores a las poliploides de Soria y Guadalajara.

En resumen, la historia evolutiva de las especies estudiadas muestra un flujo dispersivo desde Sierra Nevada hasta el norte de la Península, llegando hasta Aragón. Tras un periodo de aislamiento en dos focos, tuvieron lugar dos flujos dispersivos norte-sur, uno llegando hasta las costas almeriense y granadina y otro, por el centro de la Península Ibérica, hasta Cazorla (Fig.7.8). Estos movimientos dispersivos parecen explicar la estrecha relación existente entre las especies del complejo nevadense, como consecuencia de peroiodos de aislamiento geográfico e la hibridación. Para dilucidar la importancia relativa que cada uno de estos procesos han jugado en la evolución del complejo son necesarios estudios específicos de las zonas de contacto.

RELACIONES ENTRE DATOS POBLACIONALES: FENOTIPO, PARENTESCO GENÉTICO, POLINIZADORES Y GEOGRAFÍA

Mediante modelos de ecuaciones estructurales hemos encontrado que la diversidad genética de las poblaciones diploides de *E. mediohispanicum* se relaciona tanto con la diversidad fenotípica como con la diversidad en el gremio de polinizadores, siendo la morfología floral el rasgo que juega un papel más destacado en la interacción planta-polinizador (Capitulo 5). En este apartado exploraremos la existencia de relaciones entre fenotipo, polinizadores y parentesco de las poblaciones estudiadas, minimizando el efecto de la distancia geográfica en dicha relación. Para ello hemos obtenido matrices de distancia para cada uno de los conjuntos de variables mencionados usando el paquete de R *vegan* (Oksanen et al., 2012), y las hemos comparado mediante tests de Mantel (Mantel, 1967). Como distancia genética hemos utilizado la matriz de distancias poblacionales (combinando indels y sustituciones con igual peso) calculada en el Capítulo 6. La diferencia en el gremio de polinizadores fue calculada como la distancia de Bray-Curtis (Bray and Curtis, 1957) aplicada al número de visitas de cada grupo funcional, para lo que hemos considerado sólo poblaciones con más de 100 visitas. Por último, obtuvimos una matriz de distancias euclídeas que incluye todas las variables fenotípicas medidas (Capítulo 5), así como otras tres matrices fenotípicas en las que incluimos, respectivamente, las variables relacionadas con el tamaño de la planta (diámetro y altura del escapo y número de flores), con el tamaño de la flor (diámetro de la corola y anchura y longitud del tubo de corola) y con la forma de la corola (los cuatro primeros componentes en forma: RW1-RW4).

La primera cuestión que nos planteamos es la existencia de patrones espaciales en cada una de las matrices mencionadas. Para ello testamos las correlaciones entre cada matriz de distancia descrita arriba y la matriz de distancia geográfica. La distancia geográfica entre cada pareja de poblaciones fue calculada como la distancia euclídea de sus coordenadas UTM. De acuerdo a los resultados de los tests de Mantel (Tabla 7.3) podemos concluir que no existe un patrón de variación geográfica en la identidad de los visitantes florales en las 39 poblaciones estudiadas con censos de más de 100 visitas (r=-0.08; p=0.80; Tabla 5.7A).

Al contrario que con los polinizadores, hemos encontrado autocorrelación espacial positiva tanto en las distancias genéticas (r=0.28; p=0.001) como en las fenotípicas (r=0.16; p=0.005; Tabla 5.7A). Al desglosar la variación fenotípica en las tres componentes descritas anteriormente, encontramos también una autocorrelación espacial significativa y positiva para la forma de la corola (r=0.24; p=0.001), pero no significativa para el tamaño de la planta (r=0.07; p=0.15). En el caso del tamaño floral encontramos una significación marginal (r=0.14; p=0.018), ya que desaparece después de aplicar la corrección de Bonferroni secuencial (Tabla 5.7A).

A)

	p	Mantel	N
Polinizadores	0.801	-0.080	39
Genética	**0.001**	**0.282**	**56**
Fenotipo	**0.005**	**0.159**	**56**
Tamaño de planta	0.153	0.067	56
Tamaño de flor	**0.018**	**0.142**	**56**
Forma de corola	**0.001**	**0.244**	**56**

B)

	p	Mantel	N
Genetica-Fenotipo	0.951	-0.090	56
Genética-Forma corola	**0.006**	**0.164**	**56**
Genética-Tamaño planta	0.475	0.001	56
Genética-Tamaño flor	0.921	-0.082	56
Genética-Polinizadores	0.0592	0.089	39
Fenotipo-Polinizadores	0.5694	-0.015	39
Forma corola-Polinizadores	**0.0004**	**0.232**	**39**
Tamaño planta-Polinizadores	0.309	0.035	39
Tamaño flor-Polinizadores	0.6425	-0.028	39

Tabla 7.3: A) Resultados de los tests de Mantel comparando las distancias geográficas con las obtenidas para cada matriz de datos. B) Resultados de los tests de Mantel parciales realizados comparando las parejas de matrices indicadas y controlando por las distancias geográficas. Para cada análisis se muestra la significación (p), el valor de la correlación (r) y el número de poblaciones incluidas en cada análisis (N)

En base a estos resultados podemos concluir que poblaciones más distantes no presentan mayores diferencias en gremio de polinizadores, pero que poblaciones más alejadas geográficamente están significativamente más alejadas genética y fenotípicamente, especialmente en cuanto a forma floral. Este patrón geográfico es coherente con el obtenido en el Capítulo 5, donde además demostramos la influencia de la altitud poblacional en la altura media de las plantas, así como en la diversidad genética de las poblaciones (Tabla 5.6).

A continuación, para cada matriz que mostró correlación espacial significativa, hemos testado la existencia de correlación con una tercera matriz mediante tests de Mantel parciales, condicionados a la matriz de distancias geográficas (Tabla 5.7B). De esta manera conseguimos información sobre la correlación entre dos matrices substrayendo el efecto que tenga la localización geográfica de las poblaciones sobre dicha correlación.

Existen dos correlaciones significativas tras la corrección de Bonferroni secuencial: entre genética y forma de la corola (r=0.16; p=0.006) y entre forma de la corola y gremio de polinizadores (r=0.23; p=0.0004), aunque también es marginalmente significativa la encontrada entre genética y polinizadores (p=0.038) (Tabla 5.7B). A pesar de lo rudimentario que puede resultar este análisis para trazar sobre la filogeografía de la especie otro tipo de información, nos permite observar que la morfología floral media de las poblaciones de *E. mediohispanicum* varía de acuerdo a un patrón coherente tanto con la filogeografía de la especie como con la abundancia de grupos funcionales que encontramos en cada población.

LÍNEAS FUTURAS

Los resultados de esta Tesis doctoral resuelven algunos aspectos de la historia evolutiva y las interacciones ecológicas de *E. mediohis-*

panicum, pero abren también nuevos interrogantes que merecería la pena estudiar con más detalle en futuras líneas de investigación.

Las Sierras de Cazorla, Segura y la Guillimona se postulan como una región interesante para hacer un muestreo más exhaustivo, ya que es único territorio estudiado en el que se encuentran poblaciones diploides y poliploides a muy poca distancia. El objetivo de dicho estudio sería desentrañar el origen de cada citotipo. Concretamente, las hipótesis que se contrastarían contemplan las posibilidades de que las poblaciones diploides deriven de las poliploides mediante procesos como la diploidización, sean el resultado de procesos de hibridación entre individuos de diferente ploidía, o sean el citotipo ancestral a partir del que aparecieron las poliploides de esa región mediante eventos de poliploidización independientes a los ocurridos en el norte de la Península. Para distinguir entre estos escenarios podrían compararse las cantidades de ADN de individuos de las distintas poblaciones (estimadas mediante citometría de flujo) con sus relaciones de parentesco (estimadas mediante la región trnL-trnF, el uso de microsatélites localizados en los orgánulos o el análisis de genomas plastidiales completos).

Una metodología similar podría emplearse en la comunidad de Aragón, donde se encuentran las únicas poblaciones aparentemente diploides del norte de la Península. El objetivo de este estudio sería testar las rutas de dispersión propuestas, que apuntan a Aragón como el primer destino de *E. mediohispanicum* en el norte de la Península, a partir del que se originaron el resto de linajes, excepto el más abundante en Sierra Nevada. La descripción de la ploidía en un mayor número de poblaciones de dicha región podría aportar información acerca de si las plantas diploides de Lérida y Guadalajara derivan de un mismo evento de diploidización en Aragón o si son consecuencia de dos eventos independientes.

Otro aspecto que también debe ser estudiado con detalle es el efecto de la ploidía sobre el gremio de visitantes florales. Para ello podría diseñarse un experimento en el que plantas de diferente ploidía crecidas en macetas sean dispuestas aleatoriamente en un lugar accesible a los polinizadores habituales de la especie. Tanto el recuento de las visitas a cada citotipo como los patrones de forrajeo pueden proporcionar una idea más precisa de cuanto influye la ploidía en las interacciones planta-polinizador en *E. mediohispanicum*.

Debido a que la especie focal de este estudio ha sido *E. mediohispanicum*, las relaciones evolutivas de otras especies emparentadas fueron obtenidas con un número de poblaciones mucho menor, lo que puede justificarse también por la diferencia en las superficies que ocupan las distribuciones de cada especie. Sin embargo, los resultados basados en el trnL-trnF IGS apenas mostraron diferencias entre las poblaciones de las distintas especies del complejo nevadense, por lo que sería deseable un estudio más detallado en el que se incluyan más poblaciones de dichas especies, poniendo especial atención a las zonas de contacto. Además, resultados recientes (Abdelaziz, 2013) apuntan a que especies que se suponían relativamente alejadas del complejo nevadense (por ejemplo, *E. baeticum*, *E. penyalarense*, *E. popovi*, *E. cazorlense*,...) no lo están tanto, por lo que sería interesante abordar el estudio de las relaciones entre ellas desde una perspectiva filogeográfica.

También merecería la pena explorar mediante métodos más potentes que los tests de Mantel la existencia de relaciones entre la filogeografía y las variables fenotípicas y de polinizadores. Una posibilidad sería utilizar análisis comparados, que no se basan en matrices de distancias si no en los valores medidos de las diferentes variables por población. Siguiendo esta metodología hemos explorado la existencia de asociaciones entre el fenotipo, el gremio de polinizadores y las relaciones evolutivas de las poblaciones de *E. mediohispanicum* descritas en esta Tesis Doctoral (Gómez et al., 2013). Dichas relaciones evolutivas han sido representadas mediante topología de árboles en ese análisis debido a que los análisis comparados para reded no han sido desarrollados aún.

En un ámbito más teórico, el desarrollo de un programa que simule la evolución de secuencias de acuerdo a los mecanismos que operan en la región trnL-trnF IGS, permitiría testar la precisión del método de reconstrucción evolutiva descrito en los Capítuos 3, 4 y 6. Dichos mecanismos involucran (además de sustituciones) deleciones y duplicaciones de fragmentos relativamente constantes en cuanto a longitud. También sería deseable que dicho programa simule individuos dentro de población y permita eventos de migración entre ellas, con lo que se podría testar también el método bajo un enfoque más filogeográfico.

Una aportación muy importante al paquete de R descrito sería una herramienta que permita alinear secuencias con un alto número de indels siguiendo la metodología propuesta en el Capítulo4. SIDIER también podría hacerse más completo incluyendo métodos adicionales de obtención de distancias de secuencias basadas en indels, tales como el CIC y el SIC (Simmons and Ochoterena, 2000).

Los datos poblacionales obtenidos pueden arrojar más información sobre la historia evolutiva de *E. mediohispanicum*, por ejemplo, la determinación de las tasas de migración entre poblaciones, las fluctuaciones en tamaños poblacionales o los tiempo de coalescencia puede llevarse acabo usando métodos de computación bayesiana aproximada (ABC; Wegmann et al., 2009). Esta metodología fue ya aplicada a un subconjunto de las poblaciones presentadas pero no obtuvimos resultados satisfactorios debido a que únicos datos genéticos disponibles en ese momento (polimorfismos de amplificación aleatoria del ADN o RAPD) no permitían distinguir la dosis de los alelos por individuo, problema que queda subsanado con la aplicación de los marcadores microsatélites.

REFERENCES

Abdelaziz, M. (2013). *How species are evolutionary maintained? Pollinator-mediated divergence and hybridization in Erysimum mediohispanicum and Erysimum nevadense*. PhD. Thesis.

Agarwal, M., N. Shrivastava, and H. Padh (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports 27*(4), 617–631.

Arvanitis, L., C. Wiklund, and J. Ehrlén (2007). Butterfly seed predation: effects of landscape characteristics, plant ploidy level and population structure. *Oecologia 152*(2), 275–285.

Avise, J. (2000). *Phylogeography : the history and formation of species*. Cambridge, Mass.: Harvard University Press.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics 18*, 489–522.

Baack, E. J. and M. L. Stanton (2005). Ecological factors influencing tetraploid speciation in snow buttercups (ranunculus adoneus): Niche differentiation and tetraploid establishment. *Evolution 59*(9), 1936–1944.

Balao, F., R. Casimiro-Soriguer, M. Talavera, J. Herrera, and S. Talavera (2009). Distribution and diversity of cytotypes in dianthus broteri as evidenced by genome size variations. *Annals of Botany 104*(5), 965–973.

Balao, F., L. M. Valente, P. Vargas, J. Herrera, and S. Talavera (2010). Radiative evolution of polyploid races of the iberian carnation dianthus broteri (caryophyllaceae). *New Phytologist 187*(2), 542–551.

Borrill, M. and R. Lindner (1971). Diploid-tetraploid sympatry in dactylis(gramineae). *New Phytologist 70*(6), 1111–1124.

Bray, J. R. and J. T. Curtis (1957). An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs 27*(4), 325–349.

Brito, P. H. and S. V. Edwards (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica 135*(3), 439–455.

Buggs, R. J. A. and J. R. Pannell (2006). Rapid displacement of a monoecious plant lineage is due to pollen swamping by a dioecious relative. *Current Biology 16*(10), 996–1000.

Buggs, R. J. A. and J. R. Pannell (2007). Ecological differentiation and diploid superiority across a moving ploidy contact zone. *Evolution 61*(1), 125–140.

Campbell, V., P. Legendre, and F.-J. Lapointe (2011). The performance of the congruence among distance matrices (CADM) test in phylogenetic analysis. *BMC Evolutionary Biology 11*(1), 64.

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual review of plant biology 58*, 377–406. PMID: 17280525 PMCID: PMC1949485.

Chifflet, R., E. K. Klein, C. Lavigne, V. Le Féon, A. E. Ricroch, J. Lecomte, and B. E. Vaissière (2011). Spatial scale of insect-mediated pollen dispersal in oilseed rape in an open agricultural landscape. *Journal of Applied Ecology 48*(3), 689–696.

Clot, B. (1991). Caryosystématique de quelques erysimum l. dans le nord de la péninsule ibérique. *Anales del Jardín Botánico de Madrid 49*(2), 215–229.

Comes, H. P. and R. J. Abbott (1998). The relative importance of historical events and gene flow on the population structure of a mediterranean ragwort, senecio gallicus (asteraceae). *Evolution 52*(2), 355.

Comes, H. P. and J. W. Kadereit (1998). The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science 3*(11), 432–438.

Cooper, S. J. B. and G. M. Hewitt (1993). Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper chorthippus parallelus. *Insect Molecular Biology 2*(3), 185–194.

Dobeš, C., C. Kiefer, M. Kiefer, and M. A. Koch (2007). Plastidic trnFUUC pseudogenes in north american genus boechera (brassicaceae): Mechanistic aspects of evolution. *Plant biol (Stuttg) 9*(04), 502,515. 502.

Dobeš, C. H., T. Mitchell-Olds, and M. A. Koch (2004). Extensive chloroplast haplotype variation indicates pleistocene hybridization and radiation of north american arabis drummondii, a. divaricarpa, and a. holboellii (brassicaceae). *Molecular Ecology 13*(2), 349–370.

Doyle, J. J., L. E. Flagel, A. H. Paterson, R. A. Rapp, D. E. Soltis, P. S. Soltis, and J. F. Wendel (2008). Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics 42*(1), 443–461. PMID: 18983261.

Duchoslav, M., L. Safarova, and F. Krahulec (2010). Complex distribution patterns, ecology and coexistence of ploidy levels of allium oleraceum (alliaceae) in the czech republic. *Annals of Botany 105*(5), 719–735.

Ehrendorfer, F. and W. H. Lewis (1980). Polyploidy and distribution. pp. 45–60.

Feliner, G. N. (2011). Southern european glacial refugia: A tale of tales. *Taxon 60*(2), 365–372.

Gómez, A. and D. Lunt (2007). Refugia within refugia: Patterns of phylogeographic concordance in the iberian peninsula. In *Phylogeography of Southern European Refugia*, pp. 155–188.

Gómez, J. M. (2007). Dispersal-mediated selection on plant height in an autochorously dispersed herb. *Plant Systematics and Evolution 268*(1), 119–130–130.

Gómez, J. M., A. J. Muñoz-Pajares, M. Abdelaziz, J. Lorite, and F. Perfectti (2013). Evolution of the pollination niches in the generalist plant erysimum mediohispanicum. *American Journal of Botany in press*.

Grant, V. (1981). *Plant speciation*. New York: Columbia University Press.

Halverson, K., S. B. Heard, J. D. Nason, and J. O. Stireman (2008). Origins, distribution, and local co-occurrence of polyploid cytotypes in solidago altissima (asteraceae). *American Journal of Botany 95*(1), 50–58.

Hardy, O. J., S. Vanderhoeven, M. De Loose, and P. Meerts (2000). Ecological, morphological and allozymic differentiation between diploid and tetraploid knapweeds (centaurea jacea) from a contact zone in the belgian ardennes. *New Phytologist 146*(2), 281–290.

Husband, B. C. and D. W. Schemske (1998). Cytotype distribution at a diploid–tetraploid contact zone in chamerion (epilobium) angustifolium (onagraceae). *American Journal of Botany 85*(12), 1688–1694.

Keeler, K. H. (2004). Impact of intraspecific polyploidy in andropogon gerardii (poaceae) populations. *The American Midland Naturalist 152*(1), 63–74.

Kennedy, B. F., H. A. Sabara, D. Haydon, and B. C. Husband (2006). Pollinator-mediated assortative mating in mixed ploidy populations of chamerion angustifolium (onagraceae). *Oecologia 150*(3), 398–408.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution 16*(2), 111–120.

Koch, M. A., C. Dobeš, C. Kiefer, R. Schmickl, L. Klimeš, and M. A. Lysak (2007). Supernetwork identifies multiple events of plastid trnF(GAA) pseudogene evolution in the brassicaceae. *Molecular Biology and Evolution 24*(1), 63 –73.

Koch, M. A., C. Dobeš, M. Matschinger, W. Bleeker, J. Vogel, M. Kiefer, and T. Mitchell-Olds (2005). Evolution of the trnF(GAA) gene in arabidopsis relatives and the brassicaceae family: Monophyletic origin and subsequent diversification of a plastidic pseudogene. *Molecular Biology and Evolution 22*(4), 1032 –1043.

Kropf, M., H. P. Comes, and J. W. Kadereit (2006). Long-distance dispersal vs vicariance: the origin and genetic diversity of alpine plants in the spanish sierra nevada. *New Phytologist 172*(1), 169–184.

Legendre, P. and F. J. Lapointe (2004). Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Australian & New Zealand Journal of Statistics 46*(4), 615–629.

Levy, A. A. and M. Feldman (2002). The impact of polyploidy on grass genome evolution. *Plant Physiology 130*(4), 1587–1593.

Lihová, J., A. Tribsch, and K. Marhold (2003). The cardamine pratensis (brassicaceae) group in the iberian peninsula: taxonomy, polyploidy and distribution. *Taxon 52*(4), 783–802.

Lumaret, R., J.-L. Guillerm, J. Delay, A. A. L. Loutfi, J. Izco, and M. Jay (1987). Polyploidy and habitat differentiation in dactylis glomerata l. from galicia (spain). *Oecologia 73*(3), 436–446.

Mandáková, T., S. Joly, M. Krzywinski, K. Mummenhoff, and M. A. Lysak (2010). Fast diploidization in close mesopolyploid relatives of arabidopsis. *The Plant Cell Online 22*(7), 2277–2290.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research 27*(2 Part 1), 209–220.

Manzaneda, A. J., P. J. Rey, J. M. Bastida, C. Weiss-Lehman, E. Raskin, and T. Mitchell-Olds (2012). Environmental aridity is associated with cytotype segregation and polyploidy occurrence in brachypodium distachyon (poaceae). *New Phytologist 193*(3), 797–805.

Mardulyn, P. (2012). Trees and/or networks to display intraspecific DNA sequence variation? *Molecular Ecology 21*(14), 3385–3390.

Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution 49*(4), 718.

Morrison, D. A. (2005). Networks in phylogenetic analysis: new tools for population biology. *International Journal for Parasitology 35*(5), 567–582.

Münzbergová, Z. (2006). Ploidy level interacts with population size and habitat conditions to determine the degree of herbivory damage in plant populations. *Oikos 115*(3), 443–452.

Nieto-Feliner, G. (1993). Erysimum. In S. Castroviejo, C. Aedo, C. Gómez-Campo, M. Lainz, P. Monserrat, R. Morales, F. Muñoz-Garmendia, G. Nieto-Feliner, E. Rico, S. Talavera, and L. Villar

(Eds.), *Flora Iberica*, Volume 4, Cruciferae-Monotropaceae., pp. 48–76. Madrid: Real Jardín Botánico CSIC.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology 13*(5), 1143–1155.

Oddou-Muratorio, S., R. J. Petit, B. Le Guerroue, D. Guesnet, and B. Demesure (2001). Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution 55*(6), 1123–1135.

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. O'Hara, G. Simpson, P. Solymos, M. Stevens, and H. Wagner (2012). vegan: Community ecology package. r package version 2.0-4.

Olalde, M., A. Herran, S. Espinel, and P. Goicoechea (2002). White oaks phylogeography in the iberian peninsula. *Forest ecology and management 156*(1/3), 89–102.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in r language. *Bioinformatics 20*, 289–290.

Parisod, C., R. Holderegger, and C. Brochmann (2010). Evolutionary consequences of autopolyploidy. *New Phytologist 186*(1), 5–17.

Raabová, J., M. Fischer, and Z. Münzbergová (2008). Niche differentiation between diploid and hexaploid aster amellus. *Oecologia 158*(3), 463–472.

Rivas-Martínez, S. (1973). Avance sobre una síntesis corológica de la península ibérica, baleares y canarias. *Anal. Inst. Bot. Cavanilles 30*, 69–87.

Rothera, S. L. and A. J. Davy (1986). Polyploidy and habitat differentiation in deschampsia cespitosa. *New Phytologist 102*(3), 449–467.

Schmickl, R., C. Kiefer, C. Dobeš, and M. A. Koch (2008). Evolution of trnF(GAA) pseudogenes in cruciferous plants. *Plant Systematics and Evolution 282*, 229–240.

Schranz, M. E., C. Dobeš, M. A. Koch, and T. Mitchell-Olds (2005, January). Sexual reproduction, hybridization, apomixis, and polyploidization in the genus boechera (brassicaceae). *American Journal of Botany 92*(11), 1797–1810.

Segraves, K. A., J. N. Thompson, P. S. Soltis, and D. E. Soltis (1999). Multiple origins of polyploidy and the geographic structure of heuchera grossulariifolia. *Molecular Ecology 8*(2), 253–262.

Simmons, M. P. and H. Ochoterena (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology 49*(2), 369–381. ArticleType: research-article / Full publication date: Jun., 2000 / Copyright Â© 2000 Society of Systematic Biologists.

Soltis, D. E., R. J. Buggs, J. J. Doyle, and P. S. Soltis (2010). What we still don't know about polyploidy. *Taxon 59*(5), 1387–1403.

Stuessy, T. F., H. Weiss-Schneeweiss, and D. J. Keil (2004). Diploid and polyploid cytotype distribution in melampodium cinereum and m. leucanthum (asteraceae, heliantheae). *American Journal of Botany 91*(6), 889–898.

Sudová, R., J. Rydlová, Z. Münzbergová, and J. Suda (2010). Ploidy-specific interactions of three host plants with arbuscular mycorrhizal fungi: Does genome copy number matter? *American Journal of Botany 97*(11), 1798–1807.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology 17*(5), 1105–1109.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution 28*(10), 2731–2739.

Thompson, J. N., S. L. Nuismer, and K. Merg (2004). Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biological Journal of the Linnean Society 82*(4), 511–519.

Van Dijk, P. and T. Bakx-Schotman (1997). Chloroplast DNA phylogeography and cytotype geography in autopolyploid plantago media. *Molecular Ecology 6*(4), 345–352.

Vargas, P. (2003). Molecular evidence for multiple diversification patterns of alpine plants in mediterranean europe. *Taxon 52*(3), 463.

Wegmann, D., C. Leuenberger, and L. Excoffier (2009). Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. *Genetics*, genetics.109.102509.

Weiss, H., C. Dobeš, G. M. Schneeweiss, and J. Greimler (2002). Occurrence of tetraploid and hexaploid cytotypes between and within populations in dianthus sect. plumaria (caryophyllaceae). *New Phytologist 156*(1), 85–94.

Williams, E. W. and D. M. Waller (2012). Phylogenetic placement of species within the genus botrychium s.s. (ophioglossaceae) on the basis of plastid sequences, amplified fragment length polymorphisms, and flow cytometry. *International Journal of Plant Sciences 173*(5), 516–531.

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics 2*(5), 333–341.

Wolfe, K. H., W. H. Li, and P. M. Sharp (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences 84*(24), 9054–9058.

Yeung, K., J. S. Miller, A. E. Savage, B. C. Husband, B. Igic, and J. R. Kohn (2005). Association of ploidy and sexual system in lycium californicum (solanaceae). *Evolution 59*(9), 2048–2055.

Zhang, D.-X. and G. M. Hewitt (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology 12*(3), 563–584.

CONCLUSSIONS

8

CONCLUSIONES

1. Los marcadores microsatélites que hemos desarrollado y optimizado en la presente Tesis Doctoral han permitido la caracterización genética de las poblaciones de *Erysimum mediohispanicum*. Estos marcadores están en equilibrio Hardy-Weinberg en la mayoría de las poblaciones estudiadas y amplifican con éxito en otras especies de *Erysimum*.

2. Hemos desarrollado un nuevo procedimiento asociado a un paquete del lenguaje de programación R para extraer información filogeográfica a partir de regiones con alta frecuencia de inserciones y deleciones (indels). La historia evolutiva inferida para *Erysimum mediohispanicum* con esta metodología fue consistente con los resultados obtenidos usando otros marcadores moleculares con metodologías convencionales.

3. Las poblaciones de *Erysimum mediohispanicum* estudiadas fueron extremadamente diversas, tanto fenotípica como genéticamente, y mostraron patrones de variación espacial latitudinales y altitudinales.

4. A pesar de que *Erysimum mediohispanicum* es una planta generalista en términos de polinización, hemos demostrado que existe una relación significativa entre la variación fenotípica y genética de sus poblaciones.

5. Al contrario de lo que sucede para los rasgos relacionados con el tamaño de la planta, la forma y el tamaño de la flor mostraron efectos significativos en las interacciones entre planta y polinizador, siendo la influencia de la forma de la corola especialmente influyente en dicha interacción.

6. Genéticamente, *Erysimum mediohispanicum* está constituido por cuatro linajes principales, que muestran distribuciones disjuntas localizadas en el interior, el este, el sur y el extremo sureste de la Península Ibérica.

7. La distribución actual de *Erysimum mediohispanicum* es el resultado de múltiples eventos de dispersión y de la existencia de múltiples refugios aislados durante ciertos periodos de tiempo. Las rutas de colonización que mejor explican las propiedades genéticas de la especie incluyen un evento dispersivo desde el sur hacia el norte de la Península Ibérica seguido de dos eventos de dispersión independientes desde el norte hacia el sur.

8. Durante los movimientos de colonización mencionados se sucedieron varios periodos de aislamiento y contacto de linajes, lo que explica la complejidad taxonómica observada entre las especies de *Erysimum* de la Península Ibérica.

LIST OF FIGURES AND TABLES

ÍNDICE DE FIGURAS

ÍNDICE DE CUADROS