

UNIVERSIDAD DE CONCEPCIÓN

Facultad de Ciencias Naturales y Oceanográficas



**DIVERGENCIA DE *ERIOSYCE* SENSU STRICTO (CACTACEAE)
EN EL DESIERTO DE ATACAMA:
UNA PERSPECTIVA DESDE LA GENÓMICA ADAPTATIVA**

Tesis para optar al grado académico de Magíster en Ciencias con mención en Botánica

María Paz Melo Gaymer

Profesor Guía: Dr. Pablo Guerrero

Profesor Co-guía: Dr. Ricardo Segovia

Concepción | Marzo de 2024

© 2024. María Paz Melo Gaymer

Se autoriza la reproducción total o parcial con fines académicos, incluyendo la cita bibliográfica del documento

*« Inspiración. No hay prisa. Paciencia.
Al ritmo de la naturaleza se revela que el mejor calor es el del sol en invierno.
Adaptarse a inmiscuirse en el placer de un bosque sin sendero,
concede descubrir que las sonrisas de la tierra son las flores.
... y siempre habrá flores para aquellos que quieran verlas »*

A mi lindo amor Jesús A.

AGRADECIMIENTOS

Mi mayor agradecimiento es siempre a mi mamá Margarita Gaymer y papá Eduardo Melo, por estar presente, por su apoyo y por darme todas las herramientas posibles para mi crecimiento personal y profesional. A mis hermanos Catalina, Teresita y Gabriel, sobre todo por compartir conmigo la ternura de sus peques.

Quiero agradecer muy especialmente a mis amigas y amigos botánicos Magnolia, Javiera Paz, Francisca, Andrea, Javiera, Diego, Simona, Claudio, Paulo y Noemí. También a Perla y Camila. Por ser empatía y cariño tantas veces. A mis amigas de la vida, Daniela, Catalina, Romina y Fran, por tantas risas y pese a la distancia física, siempre estar. A mi Coni del mar, por la sincera amistad.

A mi pololo y compañero Jesús Morales, por la comprensión, paciencia, cariño, amor y dedicación. Y todas las mañanas/tardes/noches de largas conversaciones y trabajo. También a nuestra Petra y Grisú que sin duda, apoyaron para que esto se logre.

Agradezco a las grandes personas que me dejó ReGeneC: Gabriela Narváez por ser escucha e inspiración, y a Carlos Cocho por darme el impulso de entrada en mis análisis que tanto me ayudó. A Andrea por su gran ayuda en el proyecto ReGeneC.

Agradezco a la profesora Ángela Sierra por compartir la primera aventura de terreno y sus cálidos saludos para saber cómo estaba avanzando. También al profesor Alfredo Saldaña, por las largas reflexiones y apoyo en el ingreso del programa del Magíster.

Agradezco al laboratorio que me acogió con todos sus integrantes BIOMAS, que más de algún chascarro de terreno queda para recordar e incluso amigas y amigos. Especiales gracias a Antonio Varas por su ayuda, tiempo dedicado y conocimientos cuando estaba en una pendiente cuesta arriba.

Por último, quiero agradecer a mi tutor Pablo Guerrero por aceptarme en el laboratorio, por la paciencia, la enseñanza y por incentivar me siempre a participar en grandes actividades. Por su puesto, también a mi Co-tutor Ricardo Segovia, por entregar nuevas perspectivas a mi trabajo, las críticas constructivas y los desafíos.

Esta tesis fue financiada por IEB-BASAL: FB210006 y FONDECYT: 1211441 y 11200967

ÍNDICE

ÍNDICE FIGURAS	7
ÍNDICE TABLAS	9
RESUMEN	10
1. INTRODUCCIÓN	13
1.1 ESPECIACIÓN DE CACTÁCEAS EN EL DESIERTO DE ATACAMA	16
1.2 PREGUNTAS DE INVESTIGACIÓN	18
1.3 HIPÓTESIS	19
1.4 OBJETIVO GENERAL	19
1.5 OBJETIVOS ESPECÍFICOS	20
2. MATERIALES Y MÉTODOS	20
2.1 DISEÑO DE MUESTREO	20
2.2 EXTRACCIÓN DE ADN	24
2.3 SECUENCIACIÓN DARTSEQ	24
2.4 ANÁLISIS ESTADÍSTICOS	25
2.4.1 IDENTIFICACIÓN PRELIMINAR DE DIFERENCIAS CLIMÁTICAS ENTRE ESPECIES	25
2.4.2 SELECCIÓN DE PREDICTORES CLIMÁTICOS	26
2.4.3 BÚSQUEDA DE ADAPTACIÓN	28
2.4.4 DETERMINACIÓN DE GRUPOS GENÉTICOS SEGÚN MODELO MIXTO DE FACTORES LATENTES	31
2.4.5 DIFERENCIACIÓN GENÉTICA NEUTRALES V/S CANDIDATOS SEGÚN MODELO MIXTO DE FACTORES LATENTES	32
2.4.6 PREDICTOR AMBIENTAL MÁS CONTRIBUYENTE EN LA DIVERGENCIA GENÉTICA	33
3. RESULTADOS	36
3.1 IDENTIFICACIÓN PRELIMINAR DE DIFERENCIAS CLIMÁTICAS ENTRE ESPECIES	36
3.2 PREDICTORES CLIMÁTICOS	37
3.3 BÚSQUEDA DE ADAPTACIÓN	38
3.4 DETERMINACIÓN DE GRUPOS GENÉTICOS SEGÚN MODELO MIXTO DE FACTORES LATENTES	39
3.5 DIFERENCIACIÓN GENÉTICA NEUTRALES V/S CANDIDATOS SEGÚN MODELO MIXTO DE FACTORES LATENTES	45
3.6 PREDICTOR AMBIENTAL MÁS CONTRIBUYENTE EN LA DIVERGENCIA GENÉTICA	48
4. DISCUSIÓN	57
4.1 PREDICTORES CLIMÁTICOS Y BÚSQUEDA DE ADAPTACIÓN	57
4.2 ESTRUCTURA GENÉTICA NEUTRAL	58
4.3 ESTRUCTURA GENÉTICA ADAPTATIVA	59

4.4 DIVERGENCIA ASOCIADA A LA TEMPERATURA Y AL ÍNDICE DE ARIDEZ	60
5. <u>CONCLUSIÓN</u>	66
6. <u>REFERENCIAS</u>	67

ÍNDICE FIGURAS

- Figura 1:** Distribución de las poblaciones de *E. rodentiophila* (puntos azules), *E. aurata* var. *spinibarbis* (punto verde) y *E. aurata* (puntos rojos), en el gradiente latitudinal de aridez. Se indican los nombres de las localidades de muestreo.....21
- Figura 2:** Individuos y ambientes de zonas de muestreo. Paisajes evidencian el gradiente árido. (A) *E. rodentiophila* de Cifuncho. (B) Localidad de Chañaral. (C) *E. aurata* de Chollay. (D) Localidad de Pangué. (E) *E. aurata* de Los Patos. (F) Localidad de Batuco.22
- Figura 3:** Diagrama de flujo de metodología sobre análisis bioinformáticos con información molecular.....25
- Figura 4:** PCA representando el 77% de la diferenciación del nicho climático de *Eriosyce aurata* (en rojo) y *E. rodentiophila* (en azul). No se incluye variedad.....35
- Figura 5:** Tabla final de predictores luego de eliminar variables por colinealidad.....37
- Figura 6:** Gráfico indica el número de poblaciones preliminares inferidas mediante *snmf* y utilizadas únicamente para realizar la imputación de datos faltantes.38
- Figura 7:** Gráfico indica el número de clústeres genéticos inferidos mediante *snmf* con matriz Neutral. K=6.....39
- Figura 8:** Bar plot *snmf*. (A) Según matriz de SNPs Neutrales, K=6. Los colores indican cada clúster genético. (B) Según matriz de SNPs candidatos a adaptación (K=8). Cada color indica un grupo genético distinto.....41
- Figura 9:** Mapa de clústeres genéticos (matriz 1 neutral). Al igual que en barplot (Fig. 8A) los colores son: *Eriosyce rodentiophila*: Celeste= Cifuncho- Chañaral. *E. aurata* var. *spinibarbis*: Lila= Puquios. *E. aurata*: Coral= Carrizalillo, Mostaza= Chollay, Verde= Coquimbo, Azul= Putaendo.....42

Figura 10: Mapa de grupos genéticos (matriz 2 putativa a adaptación). Al igual que en bar plot (Fig. 8B) los colores son: *Eriosyce rodentiophila*: Amarillo= Cifuncho, Negro= Chañaral. *E. aurata* var. *spinibarbis*: Azul= Puquios. *E. aurata*: Marrón= Carrizalillo, Naranja= Vallenar, Gris= Chollay 2, Rosado= Chollay, Celeste= Coquimbo- Putaendo.....43

Figura 11: Análisis de coordenadas principales (PCoA) considerando K=6 según neutralidad. (A) Realizado con matriz 1 neutral. (B) Realizado con matriz 2 putativos a adaptación.....44

Figura 12: Heat map FST. (A) Similitud de los clústeres genéticos según datos neutrales. (B) Similitud de los clústeres genéticos según datos candidatos a adaptación.....46

Figura 13: Resultado GDM para cada predictor ambiental significativo. En orden descendente de la variable que más contribuye a la que menos contribuye a la diferenciación de los grupos. La altura máxima de cada curva indica la cantidad total de cambio en las frecuencias alélicas asociadas con ese predictor (importancia de la variable). La forma de cada curva indica cómo varía la tasa de cambio en las frecuencias alélicas a lo largo del gradiente predictivo.....49

Figura 14: Tabla correlación final de variables ambientales para análisis RDA.....50

Figura 15: Biplot que resume resultados del RDA ejes 1 y 2. La nube de puntos gris oscuro en el centro del gráfico representa los SNP, y los puntos coloreados representan los individuos codificados según población (K=6). Los colores son: celeste= Cifuncho- Chañaral, verde= Coquimbo, Azul= Putaendo, púrpura= Puquios, mostaza= Chollay y rojo= Carrizalillo. Los vectores azules representan predictores ambientales (consulte el texto para ver las abreviaturas), donde la longitud de la flecha refleja la cantidad de variación en los genotipos SNP explicada por esa variable y los ángulos de las flechas representan la correlación entre las variables.....52

Figura 16: Resultado RDA para resaltar las cargas de SNP. (A) los ejes 1 y 2 de RDA. (B) los ejes 4 y 5. Forma gráfica de visualizar la relación de SNPs con predictores. Los SNP candidatos se muestran como puntos coloreados con codificación según el predictor ambiental más altamente correlacionado. Los SNP no identificados como candidatos (SNP neutrales) se muestran en blanco. Los vectores azules representan predictores ambientales.....54

ÍNDICE TABLAS

Tabla 1

Tabla de número individuos por localidades de muestreo ordenados de Norte a Sur.....20

Tabla 2:

26 variables iniciales consideradas para los análisis moleculares.....26

Tabla 3

Valores F_{ST} para los grupos obtenidos con la matriz de SNPs neutrales (sobre la diagonal) y para datos candidatos a adaptación (bajo la diagonal). Valor 0 indica grupos idénticos, mientras que 1 indica diferenciación absoluta.....45

Tabla 4

Resumen AMOVA a distintos niveles de jerarquía.....47

Tabla 5

Importancia (Imp) y significancia de los predictores (Pred) según el modelo GDM.....49

Tabla 6

Puntuaciones biplot que indican la magnitud de la asociación de las variables climáticas con cada eje RDA. El aumento de la asociación entre el eje RDA y la variable climática se indica mediante puntajes biplot positivos más grandes.....51

Tabla 7

Cantidad de SNPs candidatos a adaptación, clasificados según la variable ambiental con la que muestran mayor asociación en el análisis RDA.....53

Tabla 8

Cantidad de SNPs candidatos a adaptación concatenados entre RDA y LFMM, clasificados según la variable ambiental con la que muestran mayor asociación. Resaltadas en color gris, las variables de temperatura.....55

Tabla 9

Tabla resumen AMOVA y Phi (Φ) a distintos niveles de jerarquía.....56

RESUMEN

La estructura genética de las especies en biomas desérticos es fuertemente influenciada por factores climáticos como la escasez hídrica y las altas temperaturas, las cuales pueden constituir una presión selectiva que promueve la divergencia. En el desierto de Atacama durante los ciclos glaciales del Pleistoceno, el enfriamiento y el aumento de la aridez global provocó la expansión de biomas áridos y diversificación de especies xerofíticas.

En Chile se distribuyen dos especies hermanas de cactáceas a lo largo de un amplio gradiente latitudinal de aridez. Las poblaciones más septentrionales pertenecen a la especie *Eriosyce rodentiophila*, habitando zonas costeras de la región de Antofagasta y Atacama. Mientras que las poblaciones de *E. aurata* habitan zonas costeras y valles de la cordillera de los Andes, desde la región de Atacama a Valparaíso.

En esta investigación, se evaluó si existe una relación entre el gradiente de aridez y la divergencia de grupos genéticos de estas especies y cuál es el principal predictor ambiental que influye en esta diferenciación. Hipotetizamos que el gradiente de aridez ha actuado como una fuerza selectiva conduciendo la diferenciación genómica adaptativa, promoviendo la especiación entre *E. rodentiophila* y *E. aurata*.

Para llevar a cabo el estudio, se extrajo ADN de 87 individuos de 14 localidades distribuidas latitudinalmente y fueron genotipificadas mediante la técnica DarTseq, obteniendo 121.209 SNPs. De predictores ambientales se utilizaron las 19 variables bioclimáticas básicas, elevación, radiación, índice de aridez, evapotranspiración y humedad. Para evaluar adaptación genómica, se utilizaron análisis bayesianos como el modelo mixto de factores latentes que arrojó 475 SNPs candidatos a adaptación y mediante algoritmos de factorización se determinó la presencia de 6 clústeres

genéticos según datos neutrales. Finalmente, mediante un análisis de disimilitud generalizada, se determinó que la temperatura media del trimestre más seco y el índice de aridez son las variables climáticas que mejor explican la variación adaptativa de los 6 clústeres genéticos. Complementariamente el análisis de redundancia apoya la importancia de la temperatura y la aridez en la divergencia adaptativa.

Se comprobó según el índice de fijación que hay mayor divergencia genética adaptativa y, además, mediante un análisis de varianza molecular se determinó que las diferencias entre los 6 clústeres genéticos son mayormente explicadas por datos adaptativos en comparación con datos neutrales. La temperatura reforzada por la sequía es determinante en la diferenciación de grupos genéticos. El efecto de estas variables sobre la diferenciación genética tiene un efecto significativo en el aislamiento de las poblaciones, promoviendo las diferencias locales adaptativas. De esta forma, las distintas poblaciones deben adaptarse a condiciones específicas generando patrones de aislamiento ecológico y geográfico.

Dado que el gradiente de aridez es un factor clave en la divergencia adaptativa de las poblaciones de estas especies, es muy importante de considerar en el escenario de cambio climático para dar un enfoque de priorización de poblaciones en planes conservación del desierto de Atacama.

Palabras Claves

Cactáceas, Aridez, Adaptación, Especiación ecológica, Selección, Evolución

1. INTRODUCCIÓN

La diversidad biológica se organiza en agrupaciones de organismos vivos que comparten características adaptativas, estas agrupaciones son llamadas especies (Darwin 1859; Mallet 2008). Las especies se van conformando como consecuencia de procesos de divergencia, que imponen restricciones al entrecruzamiento entre dos poblaciones de una especie ancestral (Futuyma & Kirkpatrick, 2017). La restricción al flujo genético se produce debido a barreras pre y/o postcigóticas. Un ejemplo de barrera precigótica es el aislamiento geográfico entre poblaciones a lo largo de un gradiente ambiental, causado por las diferencias locales en condiciones abióticas, como el clima (Mayr, 1947; Shen et al. 2022). Además del aislamiento al flujo genético, la heterogeneidad ambiental podría promover cambios evolutivos de las poblaciones originando diferencias en la selección natural a nivel local, la cual es uno de los principales motores de divergencia de las poblaciones e incluso de la especiación (Schneider 2000; Qian et al. 2021). Contribuyendo en gran medida al origen de la diversidad de las especies.

La especiación ecológica ocurre por las respuestas adaptativas específicas o locales al ambiente, que les permiten a los organismos tanto persistir como colonizar nuevos hábitats (Schluter 2000). El proceso adaptativo conduce a la generación de aislamiento reproductivo entre poblaciones, debido a que las poblaciones se adaptan a condiciones locales limitando el intercambio de propágulos o migrantes, facilitando la formación de nuevas especies. Estas adaptaciones no son azarosas y tienen una dirección que favorece la aptitud para sobrevivir y/o reproducirse en un determinado ambiente (Futuyma and Kirkpatrick 2017). Esta variabilidad en la adaptación de los organismos a su ambiente genera una ocupación diferencial de hábitats separados geográficamente llevando a un aislamiento reproductivo,

restringiendo el flujo genético, intensificando la divergencia y conduciendo a la especiación ecológica (Schluter 2000; Aguirre-Liguori et al. 2017, 2019).

Un tipo de selección natural que cumple un rol particularmente importante, es la selección disruptiva o también llamada diversificadora, la cual selecciona a favor de dos o más fenotipos extremos y en contra de los intermedios de ellos (Futuyma and Kirkpatrick 2017). De esta forma, actúa en contra de la media del rasgo, generando divergencia genómica adaptativa entre poblaciones, facilitando el aislamiento reproductivo y eventualmente la especiación (Thompson 2016; Qian et al. 2021). Por ejemplo, un estudio de (Aguirre-Liguori et al. 2019) se centró en averiguar los impulsores evolutivos de las diferencias entre dos subespecies de *Teosintes* anuales *Zea mays* subsp. *mexicana* y *Z. mays* subsp. *parviglumis*, identificando que la mayoría de las asociaciones significativas de SNPs fueron con la temperatura y la concentración de fósforo en el suelo. Este patrón de asociación ambiental de la estructura genómica en los *Teosintes* sugiere un proceso de divergencia adaptativa, presumiblemente porque uno de los grupos surgió en suelos volcánicos. En otro ejemplo reciente, se encontró que la precipitación del trimestre más cálido, la radiación solar en julio y la altitud podrían tener los mayores efectos sobre la divergencia interespecífica e intraespecífica en las especies de *Liriodendron* (Shen et al. 2022). En este caso, además, se encontró que estas variables ambientales tendrían asociación con genes adaptativos a regulación estomática, lo cual podría desempeñar un papel crucial en impulsar la adaptación local en los grupos. Esto ocurre, tal como indica (Jiang et al. 2019), porque las diferencias en las condiciones ambientales pueden disminuir el éxito de establecimiento de los inmigrantes en nuevas áreas. Esto, a su vez, acelera la tasa de fijación genética en las poblaciones locales debido a la reducción en las oportunidades de cruzamiento con individuos foráneos, fortaleciendo así el aislamiento genético entre poblaciones.

La integración de información geográfica, ambiental y genómica ayuda a comprender cómo las variaciones ambientales han influido en la divergencia genómica adaptativa de especies hermanas (Sork et al. 2013; Jia et al. 2020). Incluso mediante el análisis de las frecuencias alélicas de las poblaciones se puede obtener indicios sobre adaptación (Kawecki and Ebert 2004; Aguirre-Liguori et al. 2021), ya que se estudia cómo las frecuencias de ciertos alelos varían entre poblaciones y cómo estas variaciones se correlacionan con diferencias en el ambiente o rasgos adaptativos. Por lo tanto, los patrones genéticos proporcionan valiosa información acerca de procesos evolutivos y las bases moleculares de la adaptación (Jia et al. 2020).

Una aproximación que permite identificar adaptación genómica consiste en asociar polimorfismos de un solo nucleótido (SNP) con factores ambientales responsables de ejercer selección natural diferencial entre poblaciones (Aguirre-Liguori et al. 2019). Los SNP son mutaciones de un solo par de bases en un locus nuclear (Georges and Gruber 2019). Los SNPs candidatos a adaptación se pueden definir como porciones del genoma que están bajo presiones selectivas (Storz et al. 2005). De esta forma podemos incluso identificar las potenciales presiones ecológicas responsables de la adaptación local (Garot et al. 2021), revelando los factores ambientales más relevantes en la estructuración genética poblacional de las plantas (Shi et al. 2011; Fuentes et al. 2021). Han sido ampliamente documentados patrones de divergencia con las frecuencias alélicas de loci adaptativos, siendo la temperatura y las precipitaciones los dos factores climáticos más importantes que influyen la distribución, diferenciación y diversidad de las especies (Franks and Hoffmann 2012).

1.1 Especiación de Cactáceas en el desierto de Atacama

El desierto de Atacama es el desierto no polar más seco del mundo (Clarke 2006) y al contrario de lo habitualmente pensado sobre que es un ambiente estéril, la aridez puede ser un motor evolutivo al actuar como presión selectiva y con ello promover adaptaciones y diferenciación entre poblaciones (Axelrod 1972). Por ejemplo, la aridificación global del Mioceno promovió la expansión de biomas terrestres modernos, incluidos pastizales y desiertos en grandes áreas de América, África, Australia y Eurasia (Potter and Szatmari 2009; Folk et al. 2020). El proceso de aridificación coincide con la radiación de muchas plantas xerofíticas, como la mayoría de las suculentas y plantas de cojín (Arakaki et al. 2011; Hernández-Hernández et al. 2014; Folk et al. 2020). Por lo tanto, no es coincidencia que el desierto más árido del mundo sea uno de los lugares con mayor diversidad de cactáceas (Hernández-Hernández et al. 2011; Guerrero et al. 2019b).

Los linajes de plantas perennes ricos en especies presentes en lugares áridos como el desierto de Atacama son principalmente suculentos como las cactáceas. Esta familia se habría originado y diversificado luego de la tendencia global de aridificación durante el Mioceno tardío y el Plioceno (Arakaki et al. 2011; Hernández-Hernández et al. 2014), o incluso posteriormente como ha sido reportado en algunos grupos por ejemplo *Consolea* (Opuntioideae) en el Pleistoceno (Majure et al. 2021). Las adaptaciones relacionadas a la aridez y al uso eficiente de agua permiten a las especies la colonización de nuevos hábitats, incluyendo hábitats extremos, teniendo importantes implicancias de modificaciones en sus nichos ecológicos (Matuszak et al. 2016). La evolución del nicho podría mantener las poblaciones de una especie separadas ocupando hábitats diferentes, favoreciendo la diferenciación genética y viéndose reflejado en la divergencia adaptativa ecológica entre ellas (Shen et al. 2022).

La aridez es un fenómeno multidimensional (Zhang et al. 2021; Ullah et al. 2022), dado que a diferencia de otros parámetros climáticos que pueden ser monitoreados directamente, es el resultado de la interacción entre varios factores climáticos, en los que la precipitación y la temperatura juegan conjuntamente un papel crítico (Quan et al. 2013; Wallén 2017). Los cambios en varias variables hidrometeorológicas, como el aumento de la temperatura, las fluctuaciones anormales de las precipitaciones, el aumento de la evapotranspiración, la baja humedad atmosférica entre otros, tienen influencia significativa en la aridez (Lickley and Solomon 2018; Ullah et al. 2022).

La familia de las Cactáceas es uno de los linajes con tasa de diversificación más elevada del reino vegetal (Hernández-Hernández et al. 2014; Magallón et al. 2015), mostrando grandes modificaciones evolutivas: morfológicas, anatómicas, celulares y metabólicas (Gibson and Nobel 1986). A pesar de que se han llevado a cabo estudios moleculares que manifiestan el efecto de acontecimientos biogeográficos y climáticos en los patrones genéticos, especialmente durante el Plio-Pleistoceno (Guerrero et al. 2013, 2019a), es interesante considerar en las evaluaciones de la variación genética intraespecífica, si la estructura observada ha sido causada por la aridificación. La evidencia más sólida de la especiación ecológica podría provenir de la combinación de inferencia de pares de especies con resultados intraespecíficos, como, por ejemplo, fuerte divergencia genética entre individuos en hábitats contrastantes (Hernández-Hernández et al. 2021).

Dentro de la familia Cactaceae, el género *Eriosyce* contiene dos especies hermanas *Eriosyce aurata* (Pfeiff.) Backeb. y *E. rodentiophila* F.Ritter, que se distribuyen en un gradiente latitudinal de aridez entre la región de Antofagasta (-25.001315, -70.455961) y la región de Valparaíso en Chile (-32.661564, -70.670113) (Kattermann 1994). *E. rodentiophila* es la especie cuyas poblaciones son más septentrionales

distribuyéndose a lo largo de la costa desde La Chimba (Antofagasta) hasta el sur de Chañaral (Fig. 1). Por su parte, las poblaciones de *E. aurata* comienzan su distribución desde Copiapó hasta Putaendo (Fig. 1). Además, entre la distribución de ambas especies en Puquios, al norte de Copiapó se encuentra un taxon descrito como *E. aurata* var. *spinibarbis* (Fig. 1).

La taxonomía de *Eriosyce aurata*, originalmente descrita como *Echinocactus auratus* por Pfeiffer en 1847 y posteriormente reasignada por Backeb. en 1936, ha experimentado numerosas reevaluaciones y sinonimias, incluyendo su sinonimia con *Eriosyce ceratistes* y otras variantes. Por otro lado, *Eriosyce rodentiophila*, descrita por F.Ritter en 1980, ha visto su nomenclatura igualmente debatida, con sinonimias como *Eriosyce megacarpa* (Ritter). 1980 o *Rodentiophila atacamensis*. La inclusión de estas sinonimias y evolución taxonómica (valga decir, sin análisis estadísticos asociados), pone en manifiesto la complejidad y las incertidumbres asociadas a la clasificación precisa de estas especies (Korotkova et al., 2021).

1.2 Preguntas de Investigación

¿Existe una relación entre el gradiente de aridez y la diferenciación de grupos genéticos en *E. aurata* y *E. rodentiophila*? Si es así, ¿cuál es el principal predictor ambiental que influye en esta diferenciación?

1.3 Hipótesis

Las especies hermanas de cactus *E. aurata* y *E. rodentiophila*, distribuidas a lo largo de un gradiente latitudinal de aridez, presentan historias adaptativas distintas reflejadas en este gradiente. Se propone que el gradiente de aridez es un factor clave en la divergencia de sus poblaciones, actuando como una presión selectiva que impulsa la adaptación local de sus nichos climáticos. Esta influencia resulta en la divergencia de grupos genéticos y promueve la especiación ecológica de ambas especies. Si esta hipótesis es verdadera debe cumplirse que:

- 1- Exista una menor diferenciación genética en SNPs neutrales comparado SNPs candidatos a adaptación.
- 2- Las especies tengan una mayor divergencia genética en SNPs relacionados con la adaptación a factores asociados al gradiente de aridez.

1.4 Objetivo general

El objetivo general de este estudio consistió en identificar señales de adaptación presentes en los genomas de especies hermanas de cactáceas. Estas adaptaciones estarán relacionadas con la especiación ecológica, específicamente promovida por la capacidad de adaptación a diferentes ambientes áridos.

1.5 Objetivos específicos

- i. Determinar predictores climáticos independientes, para comprender las influencias climáticas de manera más precisa.
- ii. Identificar SNPs neutrales y candidatos a adaptación.
- iii. Establecer la cantidad de clústeres genéticos dentro del gradiente latitudinal.
- iv. Evaluar si hay mayor diferenciación genética en SNPs candidatos a adaptación comparado a SNPs neutrales.
- v. Determinar cuáles son las variables ambientales que más contribuyen en la diferenciación genética adaptativa de los grupos.

2. MATERIALES Y MÉTODOS

2.1 Diseño de muestreo

Se muestrearon 87 individuos de las especies *Eriosyce rodentiophila* y *E. aurata* desde el Norte de Paposo (-24.4779, -70.5225) al Sur de Putaendo (-32.4392, -70.5707). Las catorce localidades de muestreo se señalan en el mapa de la Figura 1 y en la Tabla 1 se indica el número de individuos muestreados por cada localidad. Además, en la Figura 2 se muestran imágenes de algunos individuos y ambientes de las localidades muestreadas, donde se observa una disminución de vegetación hacia el norte. Cada individuo muestreado fue georreferenciado y se recolectó tejido de raíz el cual fue conservado en Buffer salino CTAB a -4°C.

Tabla 1

Tabla de número individuos por localidades de muestreo ordenados de Norte a Sur.

Localidad muestreo	Nro. individuos	Especie
Cifuncho	5	<i>E. rodentiophila</i>
Atacama	13	<i>E. rodentiophila</i>
Chañaral	4	<i>E. rodentiophila</i>
Puquios	7	<i>E. aurata</i> var. <i>spinibarbis</i>
Carrizalillo	7	<i>E. aurata</i>
Chollay	10	<i>E. aurata</i>
Los Choros	5	<i>E. aurata</i>
Pangue	5	<i>E. aurata</i>
Río Hurtado	6	<i>E. aurata</i>
Quebrada Seca	3	<i>E. aurata</i>
Fray Jorge	5	<i>E. aurata</i>
Cuesta El Espino	4	<i>E. aurata</i>
Batuco	4	<i>E. aurata</i>
Los Patos	9	<i>E. aurata</i>
Total muestreado:	87 individuos	

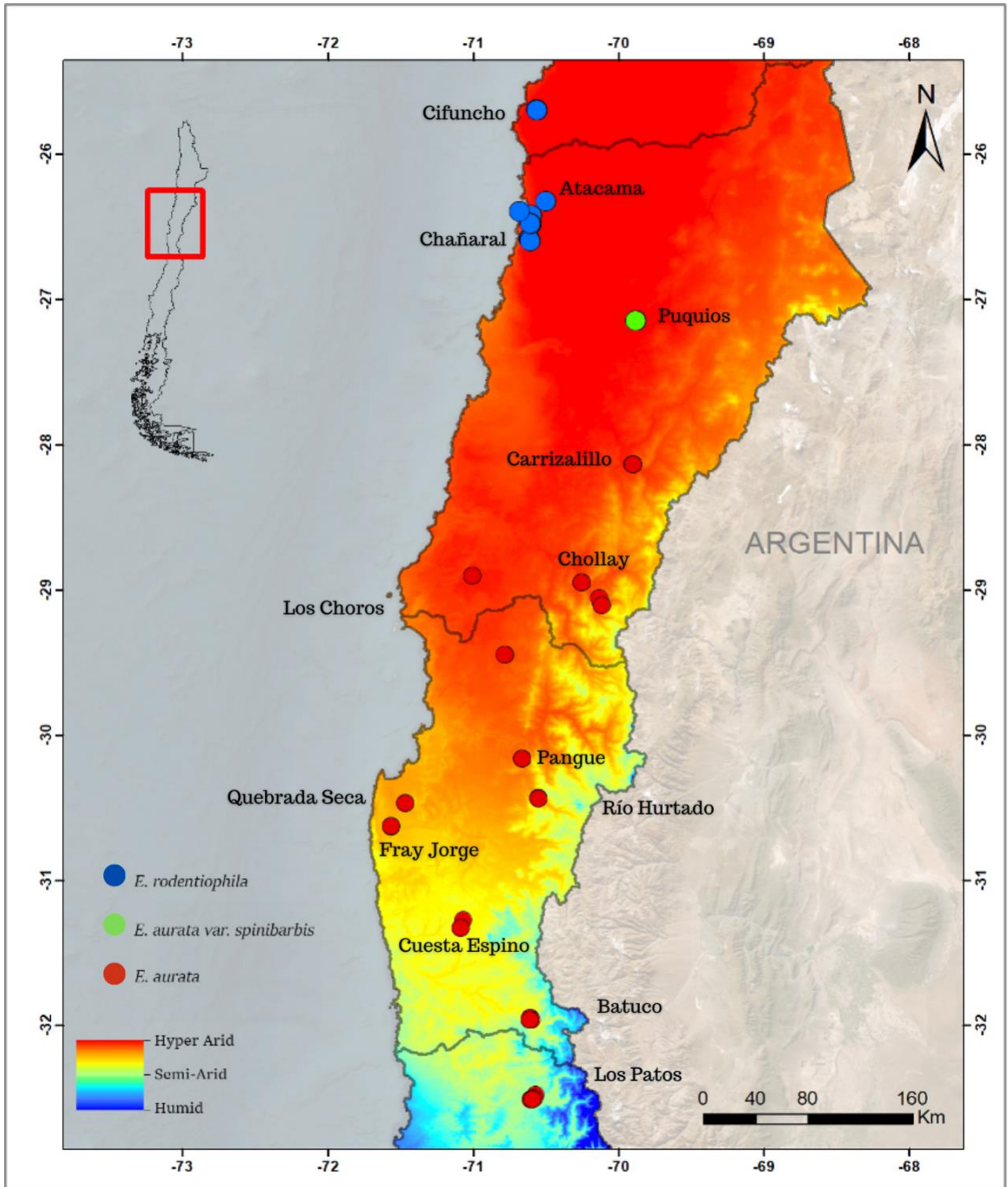


Figura 1. Distribución de las poblaciones de *E. rodentiophila* (puntos azules), *E. aurata* var. *spinibarbis* (punto verde) y *E. aurata* (puntos rojos), en el gradiente latitudinal de aridez. Se indican los nombres de las localidades de muestreo.

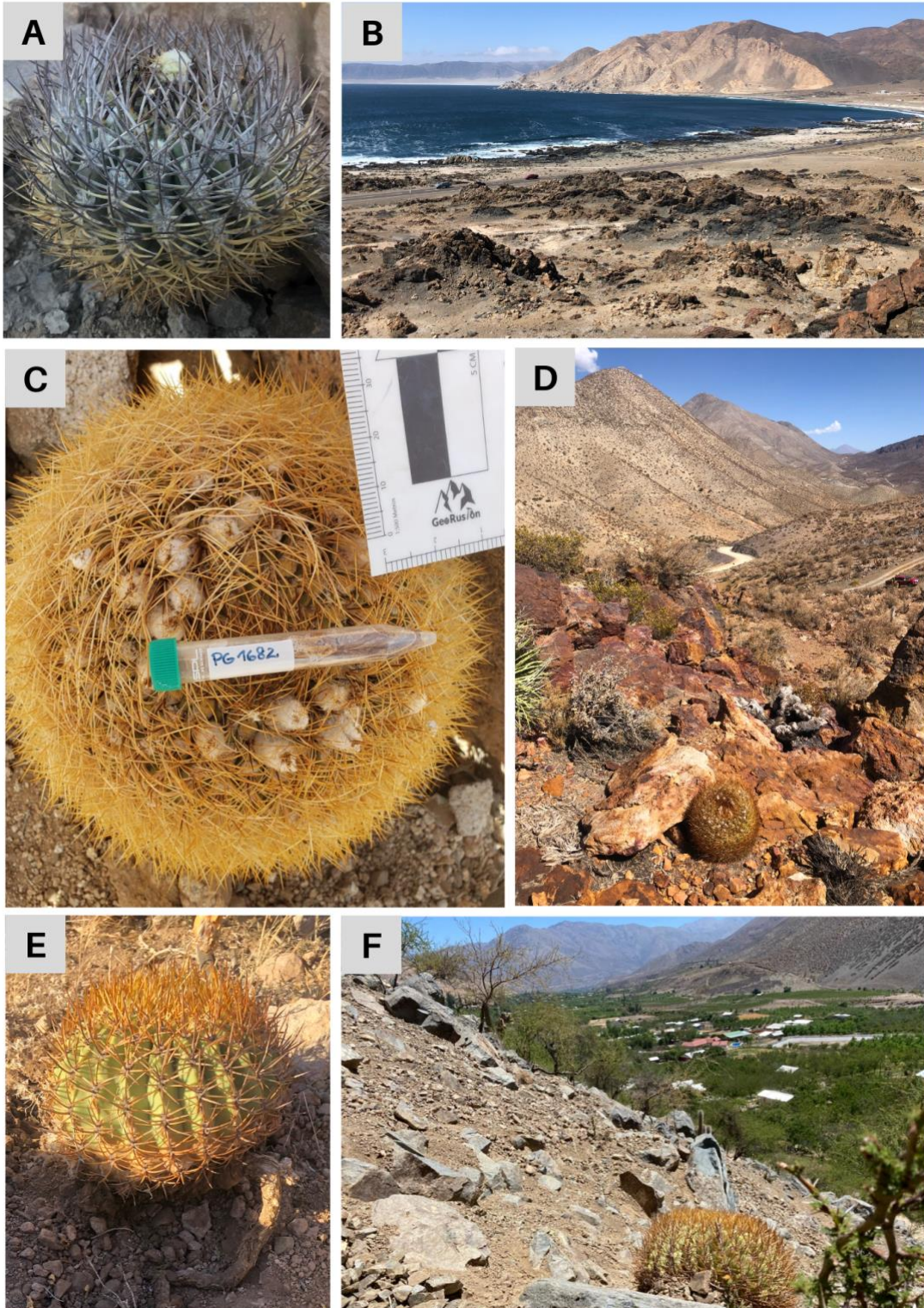


Figura 2. Individuos y ambientes de zonas de muestreo. Paisajes evidencian el gradiente árido. (A) *E. rodentiophila* de Cifuncho. (B) Localidad de Chañaral. (C) *E. aurata* de Chollay. (D) Localidad de Pangue. (E) *E. aurata* de Los Patos. (F) Localidad de Batuco.

2.2 Extracción de ADN

Se realizó extracción de ADN de las muestras de tejido de raíz mediante una metodología estandarizada (Edwards et al. 1991). En primer lugar, se procedió a moler 20mg de tejido de raíz de manera manual con un mortero y nitrógeno líquido hasta obtener un polvo. Luego, se añadieron 600µl de buffer de extracción y se aplicó vortex por 20 segundos. Después de centrifugar por 1 minuto (13.000 RPM), se transfirió el sobrenadante a un tubo nuevo. Se añadió y mezcló 400µl de isopropanol, se dejó actuar por 2 minutos a temperatura ambiente, para luego centrifugar por 5 minutos (13.000 RPM) y descartar el sobrenadante. Luego, enjuagamos con 500µl de EtOH 70%, se centrifugó nuevamente por 5 minutos y se descartó el sobrenadante. Finalmente se dejó secar el pellet de 10 a 15 minutos a temperatura ambiente y se agregó 50µl de TE. Se conserva el ADN a -80°C.

Se comprueba mediante electroforesis la presencia de bandas de ADN en el gel de agarosa tras su tinción y visualización bajo luz UV. Por último, el ADN se estandarizó a aproximadamente 50 ng/µl con NanoDrop.

2.3 Secuenciación DarTseq

El ADN fue procesado en Diversity Arrays Technology Pty. Ltd. (Canberra, Australia), utilizando para el genotipado la tecnología DArTseq (Sansaloni et al. 2011; Steane et al. 2014). DArTseq es un método de la familia Radseq, basado en la reducción del genoma que se diferencia de otros métodos por su capacidad para seleccionar de forma inteligente en un genoma las áreas predominantemente activas (baja secuencia de copias) y las secuencias repetitivas de menor valor.

El primer paso del proceso DArTseq implica la selección de enzimas de restricción para que proporcionen el mejor equilibrio entre: 1. Obtener una fracción adecuada del genoma representado, 2. Una profundidad de lectura adecuada para cada fragmento y 3. Niveles adecuados de polimorfismo. Esto es específico de la especie y, por lo tanto, requiere una optimización inicial. Una vez seleccionadas las enzimas, se corta el ADN en fragmentos aleatorios de la secuencia y se agregan primers y “códigos de barra”. Luego se apilan según la similitud de la secuencia y se aplican una serie de filtros patentados para seleccionar las etiquetas de secuencia que incluyen un marcador SNP fiable (Gruber et al. 2019). Mediante este genotipado se obtuvieron 121.209 SNPs con 33.76% de datos faltantes (NA).

2.4 Análisis estadísticos

2.4.1 Identificación preliminar de diferencias climáticas entre especies

Se realizó un Análisis de Componentes Principales (PCA) con las 19 variables bioclimáticas obtenidas desde WorldClim, extraídas para cada coordenada geográfica de los puntos muestreados de ambas especies con la función *PCA* {FactoMineR} (Lê et al. 2008). Esto con el objetivo de visualizar las ocurrencias de las especies en el espacio climático.

Posteriormente se procedió a realizar análisis bioinformáticos relacionados a la asociación genotipo- ambiente. Cada etapa y análisis fue en búsqueda de responder los objetivos específicos. El diagrama de flujo (Figura 3) resume los pasos que se fueron siguiendo.

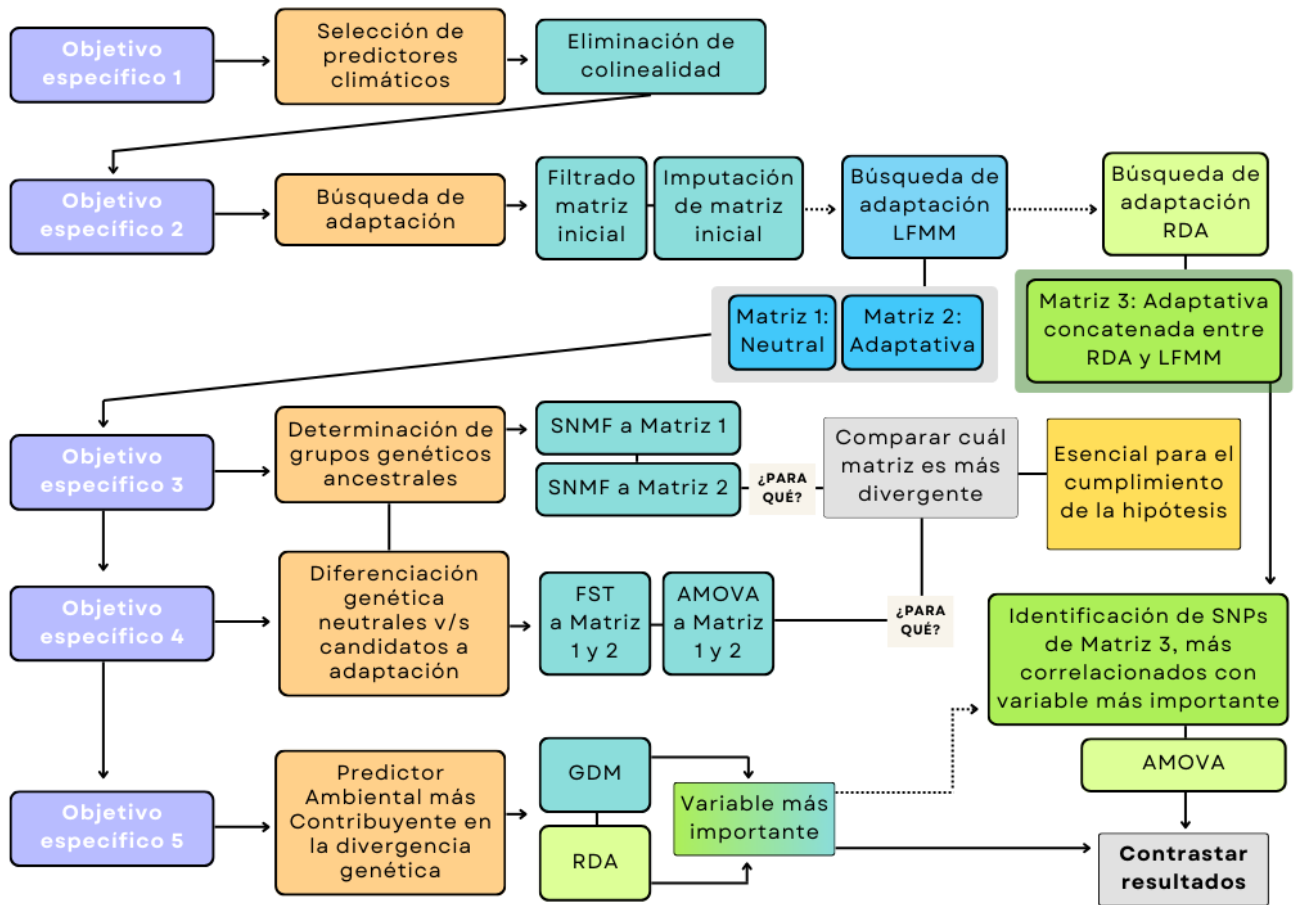


Figura 3. Diagrama de flujo de metodología sobre análisis bioinformáticos con información molecular.

2.4.2 Selección de predictores climáticos

En relación con el objetivo específico número 1, para los análisis preliminares se descargaron 26 variables (Tabla 2). En primer lugar, se obtuvo de la plataforma Chelsa (Brun et al. 2022) las 19 bioclimáticas clásicas, además de CMI-mean, HURS-mean y PET-mean (resolución 30 arc-sec, 1981-2010), considerando que los valores deben multiplicarse por 0.1 y en algunos casos restar -273.15 para recuperar valores en unidades correctas. Desde WorldClim (Fick and Hijmans 2017) se obtuvo Radiación anual y Elevación (resolución 30 arc-sec, 1970-2000). Por último, desde un trabajo de Zomer y colaboradores en 2022 (resolución 30 arc-sec, 1970-2000), se

obtuvo el índice de aridez (los valores deben multiplicarse por 0,0001 para recuperar los valores en las unidades correctas) y de evapotranspiración. Para eliminar la colinealidad de las variables se utilizó complementariamente *vifcor* {usdm} (Naimi et al. 2014) y *chart.Correlation* {PerformanceAnalytics} (Peterson and Carl 2020).

Tabla 2

26 variables iniciales consideradas para los análisis moleculares.

N°	Código	Variable
1	BIO1	Annual Mean Temperature
2	BIO2	Mean Diurnal Range
3	BIO3	Isothermality
4	BIO4	Temperature Seasonality
5	BIO5	Max Temperature of Warmest Month
6	BIO6	Min Temperature of Coldest Month
7	BIO7	Temperature Annual Range
8	BIO8	Mean Temperature of Wettest Quarter
9	BIO9	Mean Temperature of Driest Quarter
10	BIO10	Mean Temperature of Warmest Quarter
11	BIO11	Mean Temperature of Coldest Quarter
12	BIO12	Annual Precipitation
13	BIO13	Precipitation of Wettest Month
14	BIO14	Precipitation of Driest Month
15	BIO15	Precipitation Seasonality
16	BIO16	Precipitation of Wettest Quarter
17	BIO17	Precipitation of Driest Quarter
18	BIO18	Precipitation of Warmest Quarter
19	BIO19	Precipitation of Coldest Quarter
20	CMI- mean	Average Monthly Climatic Humidity Index
21	HURS-mean	Monthly Average Relative Humidity near the Surface
22	PET-mean	Average Monthly Potential Evaporation
23	Rad	Radiation
24	E	Elevation
25	AI	Global Aridity Index
26	ET0	Global Reference Potential Evapotranspiration

2.4.3 Búsqueda de Adaptación

En relación con el objetivo específico número 2 se realizaron los pasos siguientes (3.4.3.a-e):

a) Filtrado de matriz inicial

De los SNPs que se obtuvieron con el método DarTseq sin genoma de referencia, sino que apilando las secuencias idénticas, se realizó un filtrado con las recomendaciones del manual Dart (Gruber et al. 2017) y el paquete de R {dartR} (Mijangos et al. 2022) según: loci monomórficos, profundidad de lectura con $th=$ default, repetibilidad promedio de los alelos en un locus con $th=0.9$, eliminación de datos faltantes según el método de locus con $th=0.6$ y por último frecuencia de alelos menores.

b) Imputación de matriz inicial

Para poder cumplir con el segundo objetivo específico y sus respectivos análisis, se generó una matriz sin datos faltantes. Para poder completarla fue necesario realizar un análisis preliminar de grupos genéticos, dado que la imputación de los genotipos faltantes se basa en los coeficientes de ascendencia estimados y en las frecuencias de los genotipos ancestrales. Probamos valores de K que van de 1 a 10 utilizando un algoritmo de factorización de matriz no negativa dispersa (Frichot et al. 2014; Forester et al. 2018) implementando la función *snmf* {LEA} (Frichot and François 2015) con $K = 1:10$, 1000 iteraciones y $entropy = TRUE$. El análisis estima correlaciones entre variables geográficas, ecológicas y frecuencias alélicas, simultáneamente infiere los niveles de fondo de la estructura de la población (Frichot et al. 2013; Frichot and François 2015), similar a lo que realiza STRUCTURE (Pritchard et al. 2000). Graficamos los valores de entropía cruzada y seleccionamos K en función del punto de inflexión en estas gráficas; cuando el punto de inflexión

no estaba claro, usamos el valor donde la pérdida de entropía cruzada adicional era mínima (Forester et al. 2018). De esta forma, tenemos un valor de grupos genéticos preliminares, que nos permite obtener la matriz genética imputada sin NA.

c) Búsqueda de adaptación según Modelo Mixto de Factores Latentes

Para identificar SNPs candidatos a adaptación y SNPs neutrales, se utilizó la matriz genética imputada y la matriz ambiental para realizar un Modelo Mixto de Factores Latentes (LFMM), el cual es un método bayesiano que se utiliza para probar asociaciones entre loci y gradientes ambientales (Malekmohammadi et al. 2023). La variable de respuesta es una matriz genotípica y las variable explicativa es la matriz ambiental factores ambientales. Para evitar falsos negativos por efecto de la estructura o aislamiento por distancia y la distribución nula subyacente (Frichot et al. 2013), *lfmm* utiliza como factor latente la estructura de la población, que procede del análisis *snmf*. El programa estima correlaciones entre las variables ambientales y frecuencias alélicas, contrastando con una hipótesis nula. El objeto resultante puede ser utilizado para identificar SNPs que muestren asociación con gradientes ecológicos o fenotipos (Forester et al. 2018). Todo esto realizado mediante la función *lfmm2* {LEA} (Frichot and François 2015). El resultado del LFMM son dos matrices, la primera es una matriz de datos neutrales (matriz 1) y la segunda, una matriz de datos candidatos a adaptación (matriz 2).

d) Búsqueda de adaptación según Análisis de Redundancia

Adicionalmente, se realizó un Análisis de Redundancia (RDA) para encontrar SNPs candidatos adaptación mediante un método complementario. Este método identifica efectivamente los loci asociados con las variables ambientales (Malekmohammadi et al. 2023). Se utilizó la matriz imputada total inicial de SNPs y

de variables ambientales. Se eliminaron los altos factores de inflación de varianza (índice que cuantifica la multicolinealidad) que podrían sobreestimar los resultados para este análisis, por lo que fueron eliminadas con la función *vif.cca* {vegan} (Oksanen et al. 2022). Se calcula la proporción de la varianza explicada por los predictores ambientales, la cual se ajusta según el número de predictores y se procede a realizar el análisis RDA mediante la función *cca* {vegan} (Oksanen et al. 2022).

Mediante un Análisis de Varianza (ANOVA) se comprueba si el modelo RDA es significativo y si hay relación entre los datos SNPs y los predictores ambientales con la función *anova.cca* {vegan} (Oksanen et al. 2022). Además, se realizó un análisis para verificar la importancia de cada eje del RDA *anova.cca* {vegan} by="axis" $p < 0,05$ (Oksanen et al. 2022) y determinar en cuáles ejes se debe buscar candidatos a adaptación local (Legendre et al. 2011). Los histogramas de las cargas en cada eje RDA generalmente muestran distribuciones normales. Es más probable que los SNP que se cargan en las colas estén bajo selección en función de los predictores. Por lo tanto, se identificaron valores atípicos (candidatos a adaptación) como SNP que se cargan en las colas de estas distribuciones, límite de desviación estándar 2.5 (p de dos colas = 0,012).

Identificamos SNP para la carga de adaptación local para los primeros 5 ejes significativamente restringidos. Se eliminan detecciones duplicadas, es decir, SNPs que se identifican como candidatos a la adaptación en más de un eje RDA (Forester et al. 2018).

e) Vínculo Análisis de Redundancia y Modelo Mixto de Factores Latentes

Con esta información se realizó un vínculo con metodología *lfmm*, donde se concatenaron los SNPs candidatos a adaptación que se encontraron en común según

ambas metodologías de búsqueda de adaptación (RDA y LFMM). Por lo tanto, se obtiene un tercer conjunto de datos (matriz 3).

Esta tercera matriz contiene específicamente los SNPs candidatos a adaptación concatenados entre ambas metodologías y además la información de qué predictor ambiental está más correlacionado con cada uno de estos SNPs, lo cual fue obtenido según información entregada por RDA (Forester et al. 2018). Esta tercera matriz de datos SNPs fue utilizada sólo para el último análisis, no para determinación de clústeres genéticos, ni para modelo de disimilitud generalizado (GDM).

2.4.4 Determinación de grupos genéticos según Modelo Mixto de Factores Latentes

En relación con el objetivo específico número 3, para determinar los grupos o clústeres genéticos, es decir, los grupos que comparten un conjunto de características genéticas que los distinguen de los demás, se realizó nuevamente el análisis *snmf* con la matriz de datos neutrales (matriz 1), *snmf* {LEA} con $K = 1:10$, 1000 iteraciones y *entropy = TRUE* (Frichot and François 2015).

Este mismo análisis también se realizó para la matriz de datos candidatos a adaptación (matriz 2) a modo de comparar en cuál matriz de datos hay mayor divergencia genética de SNPs. En caso de haber mayor divergencia genética en la matriz candidata a adaptación, se sugiere que están reflejados los procesos adaptativos. Esto sería esencial para el cumplimiento de la hipótesis.

Además, se hizo un Análisis de Coordenadas Principales (PCoA), tanto con el set de SNPs neutrales como con aquellos candidatos a adaptación, para visualizar la similitud genética de individuos y poblaciones mediante la ordenación del Análisis de Coordenadas Principales (PCoA) (Gower 1966; Gruber et al. 2019). Esto mediante la función *gl.pcoa* {dartR} (Mijangos et al. 2022).

2.4.5 Diferenciación genética neutrales v/s candidatos según Modelo Mixto de Factores Latentes

En relación con el objetivo específico número 4, se comparan matrices 1 y 2 (neutral y candidata a adaptación), para determinar qué datos explican mejor la cantidad de clústeres genéticos y su divergencia genética. Primero, evaluando el índice de fijación (F_{ST}) para poder observar qué tan distantes o diferenciadas genéticamente están los clústeres genéticos, según cada matriz de datos.

Entonces, tanto como con los clústeres del set de datos neutrales, como con el set de SNPs candidatos a adaptación, se realizaron análisis para determinar los valores del índice de fijación (F_{ST}), *gl.fst.pop* {dartR}, bootstraps= 10000, intervalo de confianza 95% (Mijangos et al. 2022).

Posteriormente y siguiendo la misma lógica, se realizó un Análisis de Varianza Molecular (AMOVA) con cada uno de estos conjuntos de datos por separado (matriz de datos neutrales y matriz de datos candidatos a adaptación), considerando los clústeres genéticos obtenidos de la neutralidad como Poblaciones y las localidades de muestreo como Subpoblaciones. Esto con el objetivo de medir la estructura genética de las poblaciones en términos de diferencias a distintos niveles y comparar ambos conjuntos de datos para encontrar cuál de ellos explica de mayor forma la diferenciación entre los clústeres genéticos. Se calcularon los valores de los coeficientes de correlación genética Phi (Φ) que miden la correlación de la información genética alelos/haplotipos entre diferentes niveles jerárquicos. El valor de Phi es entre -1 y 1. El valor cercano a 1 indica fuerte diferenciación genética. Esto mediante la función *poppr.amova* {poppr} (Kamvar et al. 2015). Por último, se realizó test Monte-Carlo con la función *randtest* {ade4} (Thioulouse et al. 2018) con 99 pruebas de permutación, considerando un umbral de significancia de 0.05.

2.4.6 Predictor ambiental más contribuyente en la divergencia genética

En relación con el objetivo específico número 5 se realizaron los pasos siguientes (3.4.6.a-c):

a) Modelo de Disimilitud Generalizada según Modelo Mixto de Factores Latentes

Para determinar cuáles variables ambientales explican mayormente la variación adaptativa de los grupos, se realizó un Modelo de Disimilitud Generalizada (GDM) (Ferrier et al. 2007). GDM es una técnica estadística con enfoque no lineal basado en distancias, que modela asociaciones genotipo-ambiente (Fitzpatrick et al. 2022) y que se adapta a una tasa variable de cambio en las frecuencias alélicas de las poblaciones presentes en gradientes ambientales (Varas-Myrik et al. 2022).

El enfoque se adapta a dos tipos de no linealidad que se encuentran comúnmente en conjuntos de datos ecológicos: 1. La relación no lineal entre el aumento de la distancia ecológica y la disimilitud de composición genética observada entre sitios; y 2. Las diferencias en la tasa de variación de la composición genética en diferentes poblaciones a lo largo de los gradientes ambientales (Ferrier et al. 2007).

Ajustar un GDM requiere primero preparar los datos de entrada en un formato específico. Se necesita una tabla de predictores por sitio (que sería la matriz ambiental) y otra de variables respuesta, es decir, una matriz de distancia que contiene los índices de disimilitud genética entre poblaciones, la cual se obtuvo al aplicar `vegdist {vegan} method = "Jaccard"` (Oksanen et al. 2022) sobre la matriz bialélica que contiene los loci candidatos a adaptación. Luego, la función `formatsitepair {gdm}, dist = "bray"` (Fitzpatrick et al. 2022), toma los datos predictivos (que incluyen datos ambientales y geográficos), más los datos biológicos (la matriz de distancia de los SNPs candidatos a adaptación) y crea una tabla de pares de sitios.

El formato de la tabla de entrada para GDM se ha desarrollado para permitir el modelado en una amplia variedad de circunstancias posibles, permitiendo la opción de ajustar modelos sin utilizar todos los pares de sitios posibles (Mokany et al. 2022). Se eliminaron de la matriz todos los pares que sean de individuos de un mismo clúster genético para analizar exclusivamente las diferencias entre los clústeres. Por último, con esta tabla final de pares, se procede al último paso que entregará el resultado que determina cuál predictor explica mayormente la divergencia de los grupos genéticos, esto con la función *gdm* {gdm}, geo=T (Fitzpatrick et al. 2022).

Se realizó el test de validación cruzada *gdm.crossvalidation* {gdm}, train.proportion=0.9, n.crossvalid.tests=default (Fitzpatrick et al. 2022), para ver si el modelo GDM se ajusta bien a los datos, utilizando todos los predictores de la tabla de entrada y la función *gdm.varImp* {gdm}, predSelect=FALSE, nPerm=50, pValue=0.05 (Fitzpatrick et al. 2022), para realizar pruebas que estimen la importancia de los predictores y del modelo de disimilitud generalizada. Esto mediante permutaciones que aleatorizan los valores de un predictor y luego ajusta nuevamente el modelo para ver cómo afecta al rendimiento de este. La importancia del predictor se cuantifica como el cambio porcentual en la desviación explicada entre un modelo ajustado con y sin ese predictor permutado (Ferrier et al. 2007; Mokany et al. 2022). Cualquier predictor en el que los coeficientes I-spline sume cero, se elimina de forma preventiva.

b) Análisis de Redundancia (RDA)

Adicionalmente, se realizó un análisis de redundancia (RDA, ver sección 3.4.3.d) mediante la función *cca* {vegan} (Oksanen et al. 2022). Este es un método que

combina regresión multivariable y PCA. El modelo encuentra combinaciones lineales de múltiples predictores (aquí variables ambientales) que explican las combinaciones lineales de múltiples variables de respuesta (aquí los SNPs), mientras maximiza la varianza explicada en las variables respuesta. Las combinaciones lineales de variables respuesta se llaman ejes canónicos del RDA, y cada eje está correlacionado con múltiples predictoras (Harrisson et al. 2017). De esta forma, también podemos identificar los predictores ambientales que están contribuyendo en la diferenciación de los clústeres genéticos. La disposición relativa de estos elementos en el espacio de ordenación refleja su relación con los ejes de ordenación, que son combinaciones lineales de las variables predictoras.

- c) Contraste con resultado Modelo Generalizado de Disimilitud según matriz concatenada Modelo Mixto de Factores Latentes y Análisis de Redundancia

Con la tercera matriz de datos SNPs candidatos a adaptación concatenada entre ambas metodologías (RDA y LFMM) (obtenida en la sección 3.4.3.e) y la información del predictor ambiental más correlacionado con cada uno de estos SNPs se contrastan los resultados del análisis GDM. Primero se eligen únicamente los SNPs correlacionados a las variables que fueron más importantes según resultado GDM. Con este set de datos entonces, se realizó AMOVA *poppr.amova* {*poppr*} (Kamvar et al. 2015), con el objetivo de determinar cuánto explican estos SNPs asociados al predictor climático, la variación de los clústeres genéticos. De esta forma se contrasta el resultado sobre la importancia del predictor ambiental que resultó del análisis GDM.

Todos los análisis bioinformáticos se realizaron en el software R, versión 4.2.2 (R Core Team 2023).

3. RESULTADOS

3.1 Identificación preliminar de diferencias climáticas entre especies

Los resultados del Análisis de componentes principales (PCA) con datos geográficos y climáticos básicos muestra que ambas especies estrechamente emparentadas, presentan poblaciones en una amplia distribución, pero sin solaparse. No hay barreras vicariantes evidentes, lo que nos hace considerar al clima para dilucidar la diferenciación de las poblaciones. Para este caso no se incluyó *E. aurata* var. *Spinibarbis* para mantener la homogeneidad de los datos a comparar, ya que su variabilidad podría sesgar el resultado hacia otras interpretaciones. Se obtuvo como resultado que ambas especies tienen rangos climáticos distintos, por lo tanto, están separadas en el espacio climático (Fig. 4). De esta forma, es posible que algún predictor climático en específico no solo segregue a las especies geográficamente, sino también de manera genética.

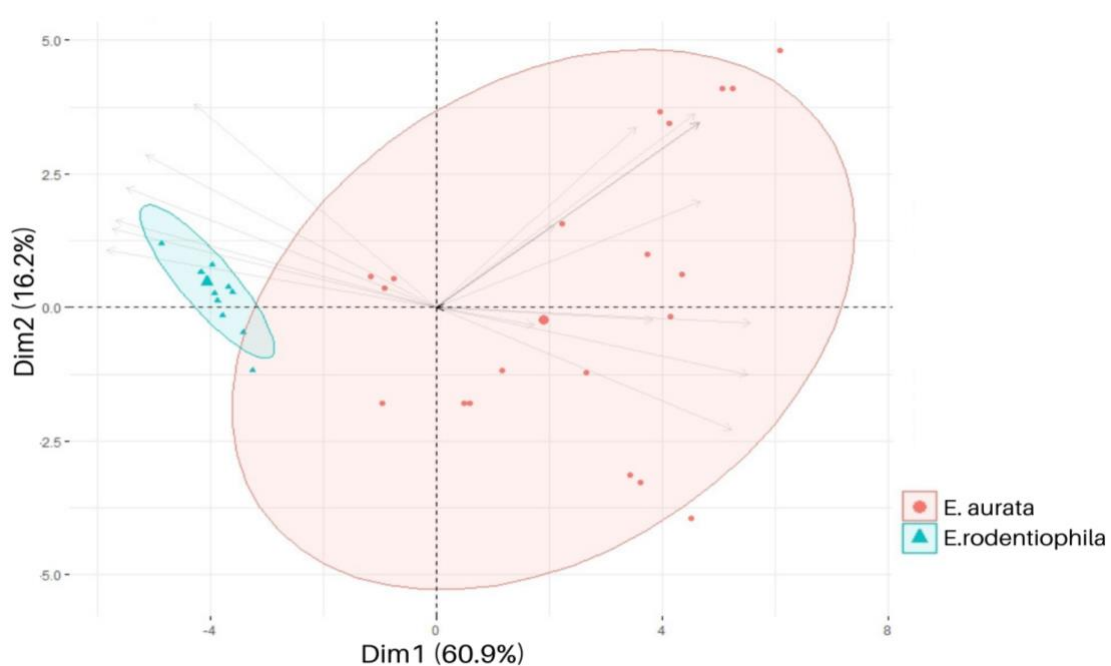


Figura 4. PCA representando el 77% de la diferenciación del nicho climático de *Eriosyce aurata* (en rojo) y *E. rodentiophila* (en azul). No se incluye variedad.

3.2 Predictores Climáticos

Para los análisis genómicos, se realizó la selección de variables climáticas. Se utilizaron 13 variables seleccionadas luego de la eliminación de colinealidad (Figura 5): Elevación, Radiación anual (RAD), Rango diurno medio de temperaturas promediadas durante 1 año (BIO2), Isotermalidad (BIO3), Temperatura media del Trimestre más seco del año (BIO9), Temperatura de trimestre más cálido del año (BIO10), Precipitación del mes más seco (BIO14), Coeficiente de Variación de Estacionalidad de las Precipitaciones (BIO15), Índice de humedad climática mensual promedio durante 1 año (CMI-mean), Humedad relativa media mensual cerca de la superficie durante 1 año (HURS-mean), Evaporación potencial mensual promedio durante 1 año; calculado con la ecuación de Penman-Monteith (PET-mean, Monteith 1965), índice de aridez global (Global-AI) y la evapotranspiración potencial de referencia global (Global-ET0, similar a PET).

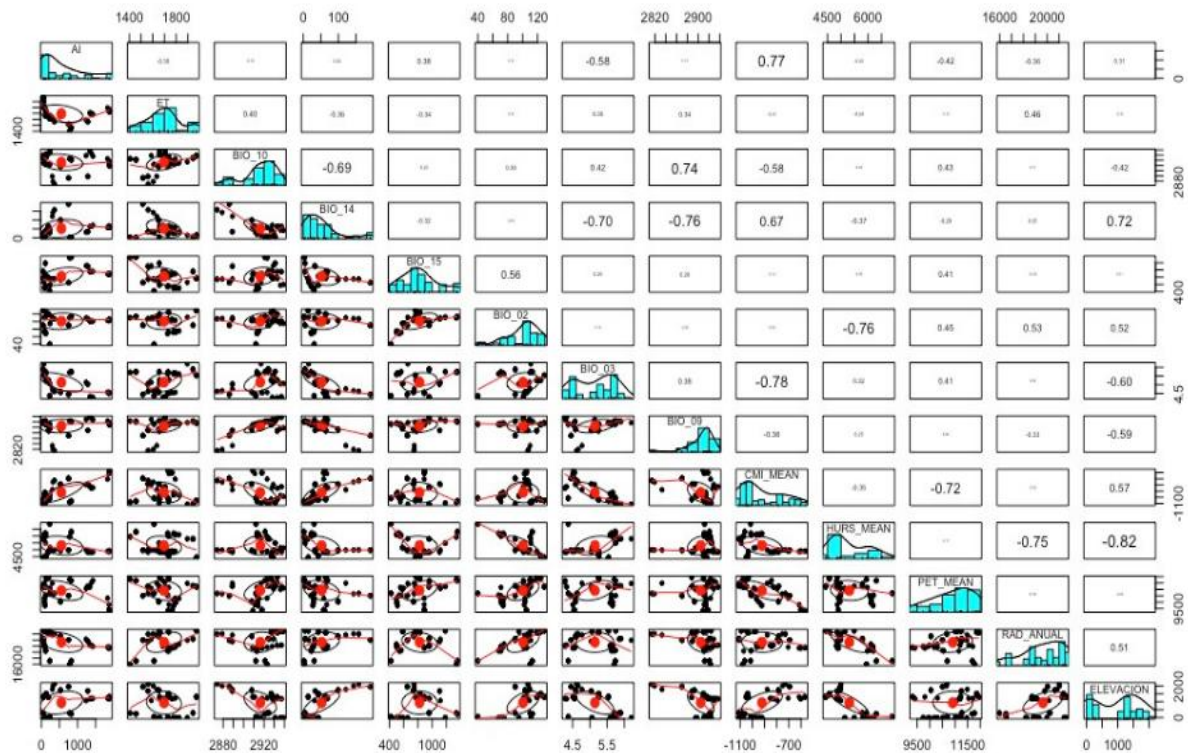


Figura 5. Tabla final de predictores luego de eliminar variables por colinealidad.

3.3 Búsqueda de adaptación

Para los 87 individuos de 14 localidades de muestreo, los datos SNPs originales recibidos fueron 121.209 SNPs con el 33.76% de datos faltantes (NA). Posterior al filtrado, la matriz de SNPs quedó con 26.965 SNPs con un 22.08% de datos faltantes. Para imputación de datos faltantes en la matriz se utilizó primer resultado de poblaciones según *snmf*. El valor fue $K=5$ (Figura 6). Este valor solo es utilizado para realizar la imputación de datos faltantes considerando el resultado preliminar de que hubiera 5 grupos genéticos.

Con esta matriz total imputada, se procedió a separar los datos SNPs para buscar adaptación según método LFMM, donde de los 26.965 SNPs totales se obtuvo un matriz de 26.490 SNPs neutrales y otra matriz de 475 SNPs candidatos a adaptación.

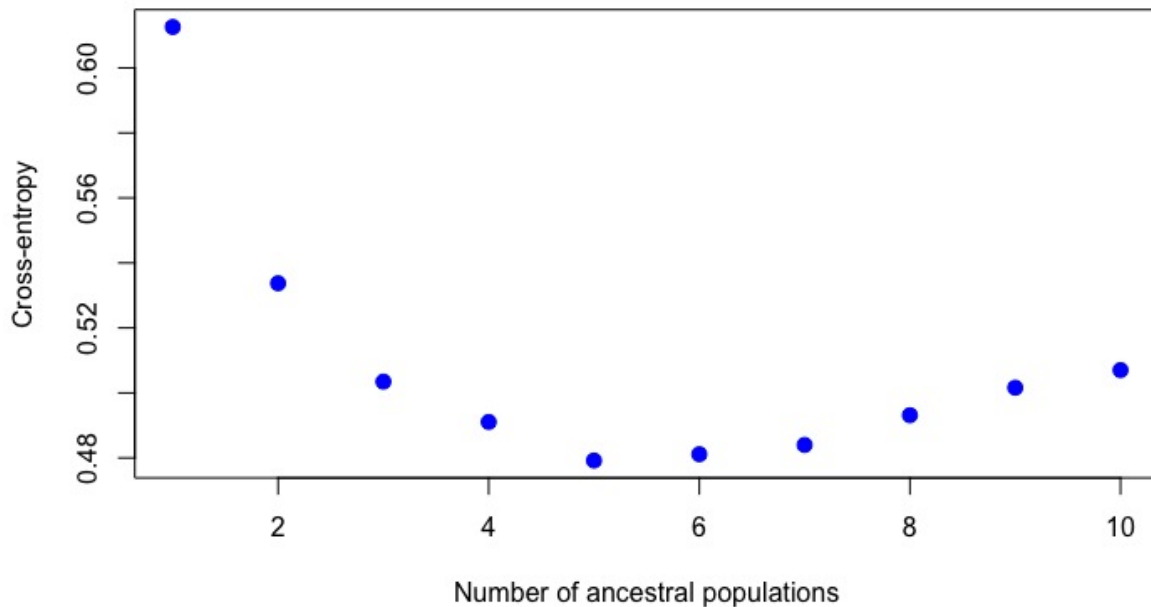


Figura 6. Gráfico indica el número de poblaciones preliminares inferidas mediante *snmf* y utilizadas únicamente para realizar la imputación de datos faltantes.

3.4 Determinación de grupos genéticos según Modelo Mixto de Factores Latentes

Con la matriz 1 de SNPs de datos neutrales (26.490 SNPs) se determinaron los clústeres genéticos según el valor mínimo de entropía cruzada *snmf*, resultando K=6 (Figura 7).

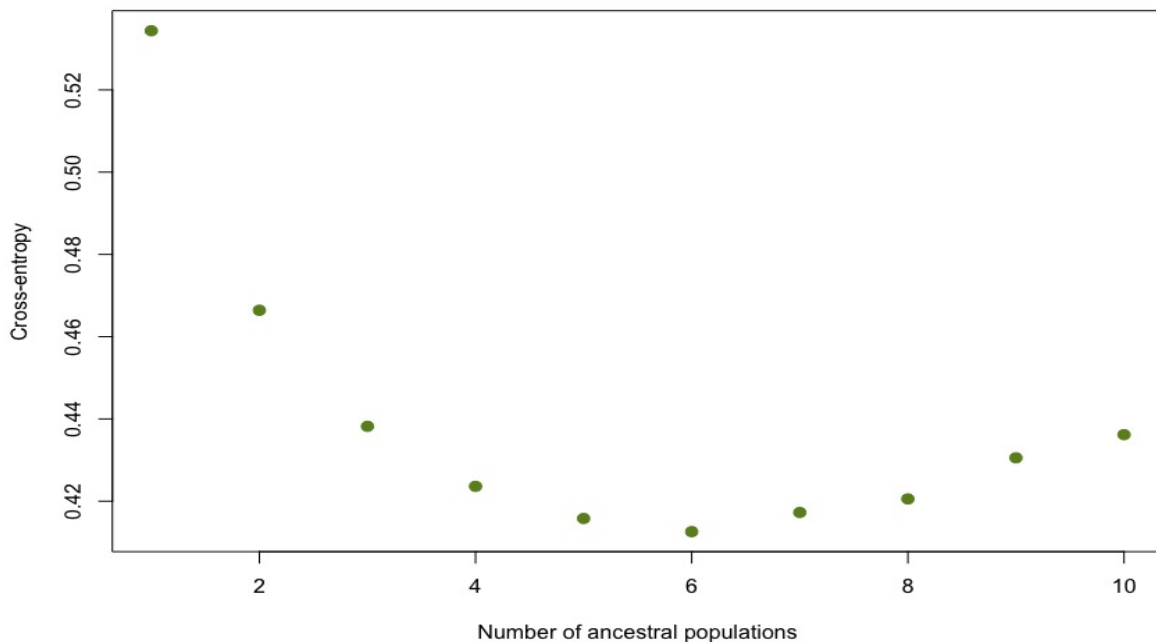


Figura 7. Gráfico indica el número de clústeres genéticos inferidos mediante *snmf* con matriz Neutral. K=6.

La especie *E. rodentiophila* se mantiene como único grupo genético nombrado como Cifuncho- Chañaral conteniendo las 3 localidades de muestreo: Atacama, Cifuncho y Chañaral (celeste), mientras que *E. aurata* se separa en 5 grupos diferenciados (Figura 8A y Figura 9), siendo una novedad dentro de la taxonomía de la especie. El primer grupo nombrado Puquios está formado por individuos de la localidad de Puquios (morado). El segundo grupo nombrado Carrizalillo contiene individuos de la localidad de Carrizalillo (coral). El tercer grupo, nombrado Chollay contiene individuos de la localidad de Chollay (mostaza). El cuarto grupo nombrado

Coquimbo contiene individuos de las localidades de Los Choros, Río Hurtado, Fray Jorge, Pangué, Bатуco, Cuesta el Espino, Quebrada Seca y Los Patos (verde). El quinto grupo de *E. aurata* es nombrado Putaendo y está formado por individuos de las localidades de Bатуco y Los Patos (azul). Estos son los 6 clústeres (K=6) utilizados en los análisis posteriores.

De estos 6 clústeres genéticos según la neutralidad, es interesante destacar a Carrizalillo (color coral, Figura 8A y Figura 9), por ser el grupo con menor interacción genética con otros grupos. Por otro lado, las poblaciones más septentrionales que correspondían a *E. rodentiophila*, resultaron ser un mismo clúster genético Cifuncho- Chañaral (color celeste, Figura 8A y Figura 9), pero compartiendo genes con casi todos los grupos, incluso el grupo lejano Coquimbo (color verde), exceptuando a grupo más al sur Putaendo (color azul).

Por otro lado, con la matriz de datos candidatos a adaptación (matriz 2) se obtuvieron 8 poblaciones (Figura 8B y Figura 10). Por lo tanto, la matriz de candidatos a adaptación muestra mayor diferenciación en comparación con datos neutrales.

El resultado del análisis de coordenadas principales representa una visualización de las similitudes entre poblaciones. Esto indica que para SNPs neutrales el eje 1 explica 22.9% y el eje 2 explica 9.4%. Esto quiere decir que en dos dimensiones la variación total es explicada un 32.3% por la neutralidad (Figura 11A). En el caso de candidatos a adaptación el eje 1 explica un 33.4% y el eje 2 explica 21.4%, por lo que la variación total de los datos candidatos a adaptación es explicada en un 54.8% en dos dimensiones (Figura 11B). En ambos casos se observa cómo los individuos de cada grupo están cercanos en el espacio indicando su similitud genética. En la matriz de datos candidatos a adaptación, la población Chollay se muestra como un grupo genéticamente más distante a los demás (Figura 11B).

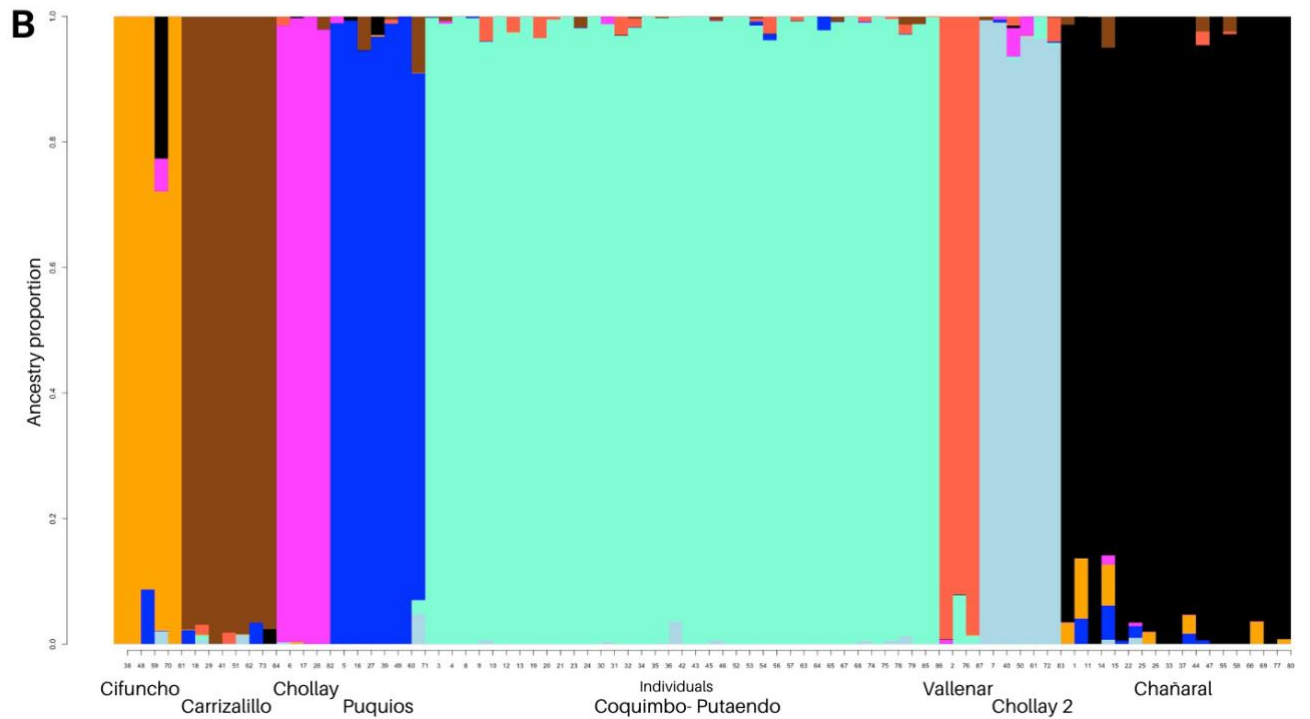
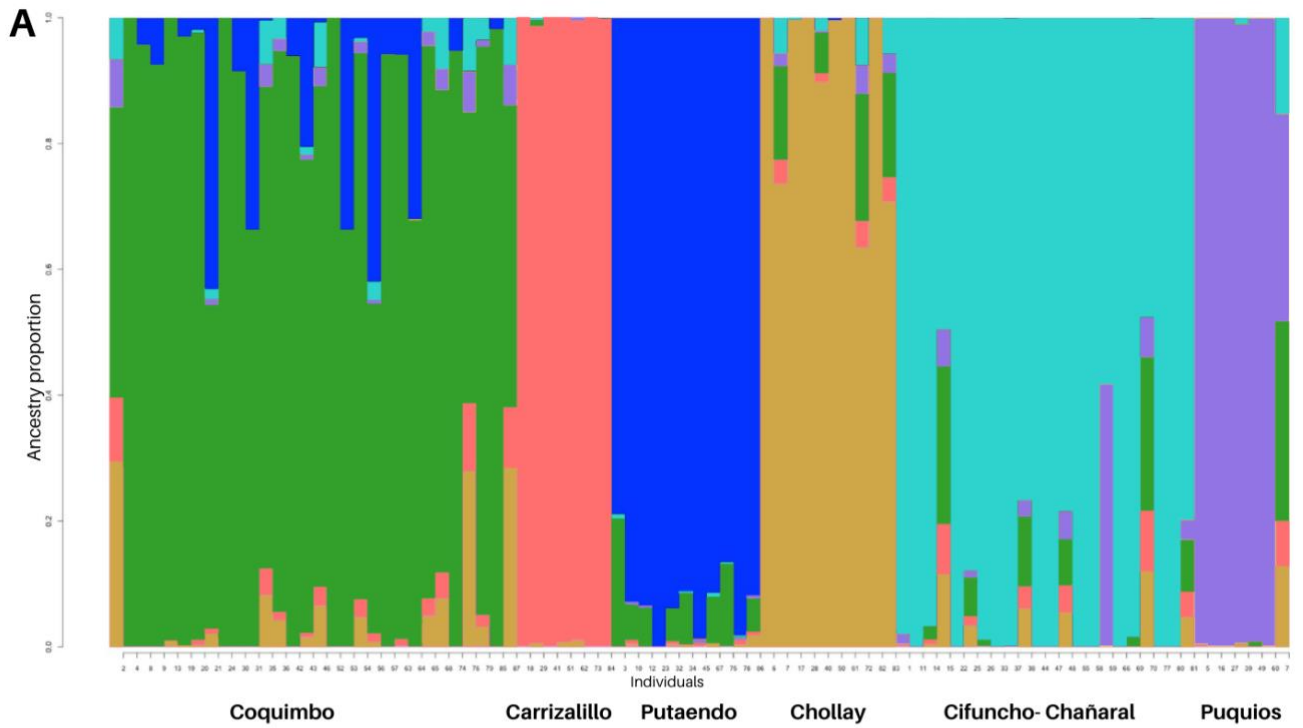


Figura 8. Bar plot *snmf*. (A) Según matriz de SNPs Neutrales, $K=6$. Los colores indican cada clúster genético. (B) Según matriz de SNPs candidatos a adaptación ($K=8$). Cada color indica un grupo genético distinto.

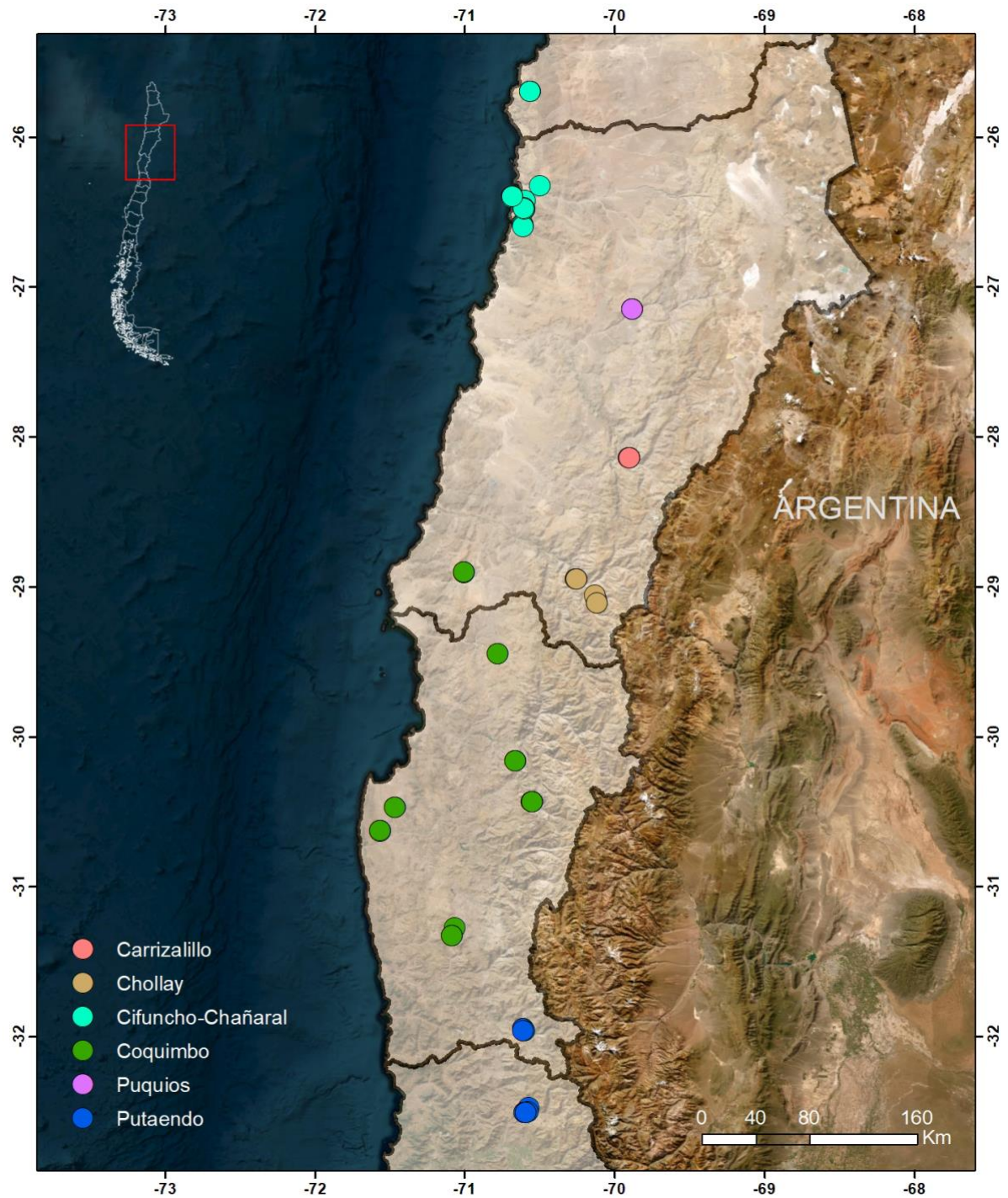


Figura 9. Mapa de clústeres genéticos (matriz 1 neutral). Al igual que en barplot (Fig. 8A) los colores son: *Eriosyce rodentiophila*: Celeste= Cifuncho- Chañaral. *E. aurata* var. *spinibarbis*: Lila= Puquios. *E. aurata*: Coral= Carrizalillo, Mostaza= Chollay, Verde= Coquimbo, Azul= Putaendo.

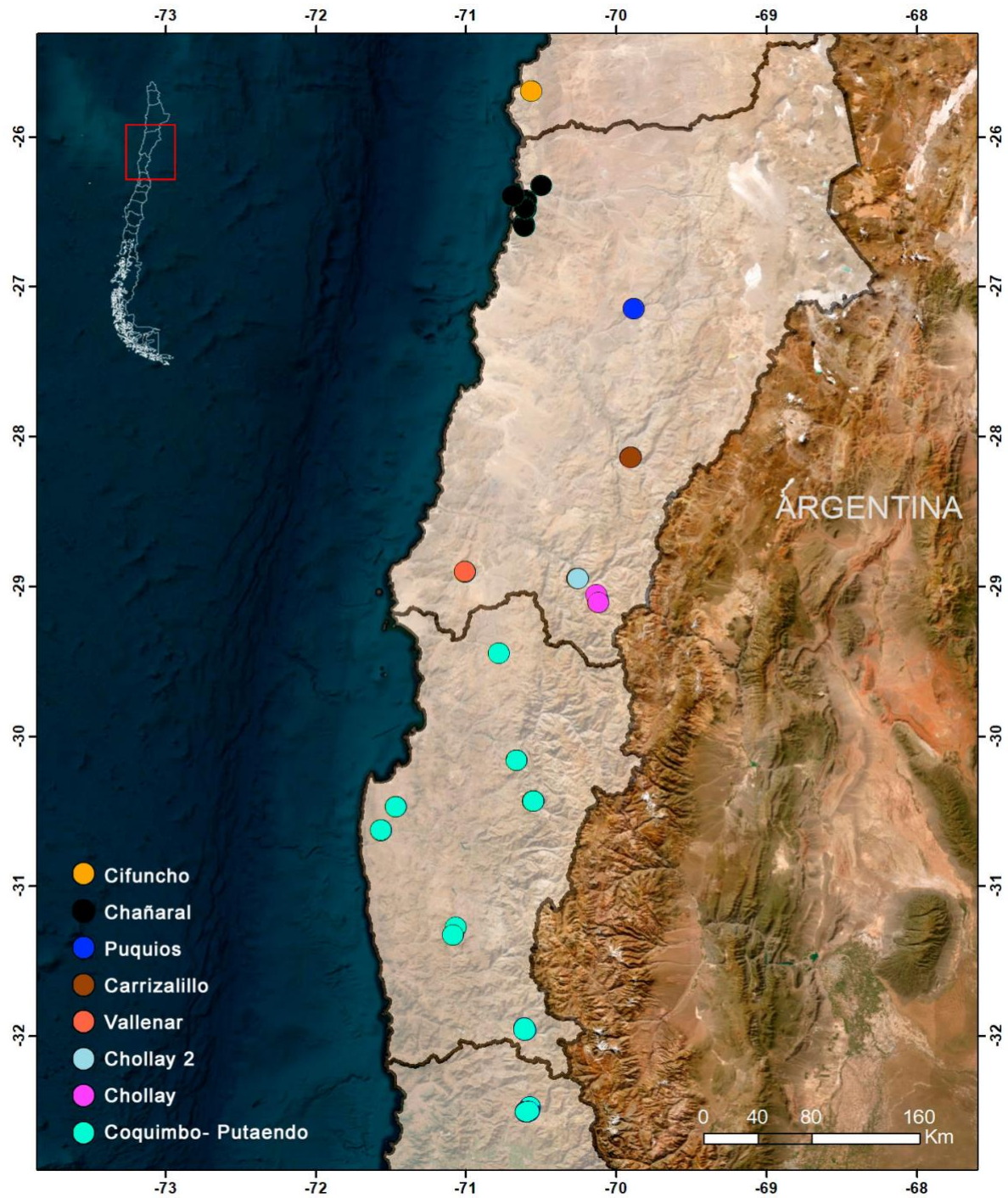


Figura 10. Mapa de grupos genéticos (matriz 2 putativa a adaptación). Al igual que en bar plot (Fig. 8B) los colores son: *Eriosyce rodentiophila*: Amarillo= Cifuncho, Negro= Chañaral. *E. aurata* var. *spinibarbis*: Azul= Puquios. *E. aurata*: Marrón= Carrizalillo, Naranja= Vallenar, Gris= Chollay 2, Rosado= Chollay, Celeste= Coquimbo- Putaendo.

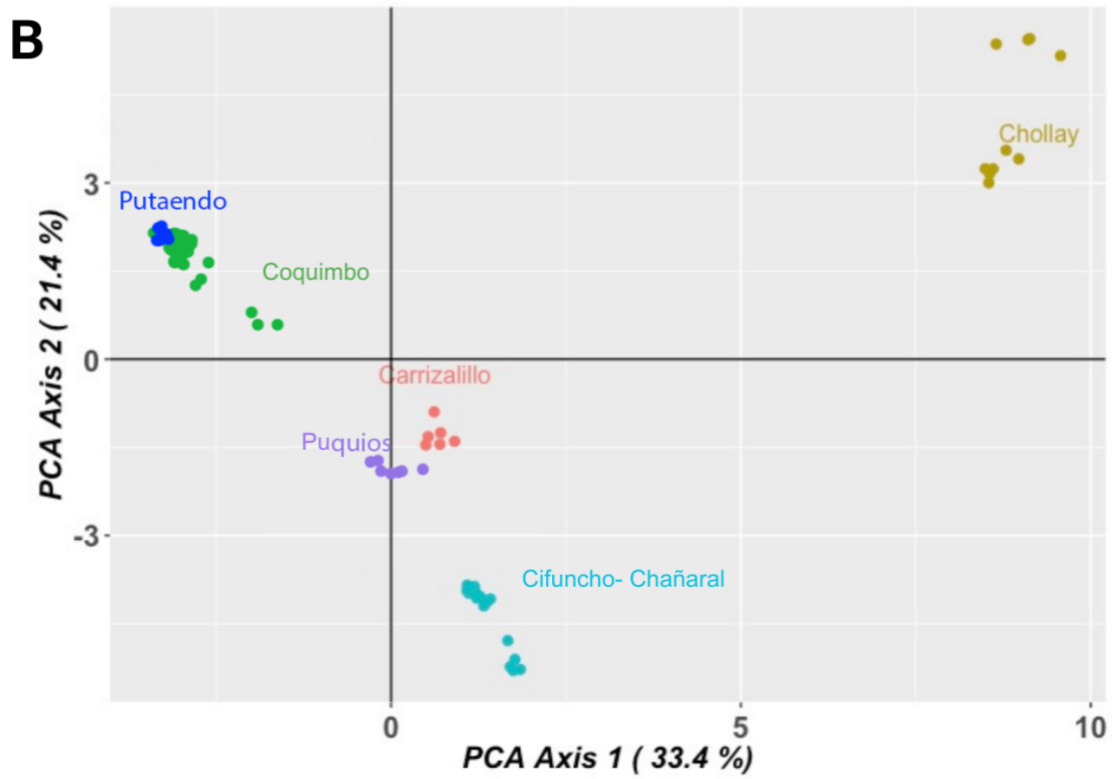
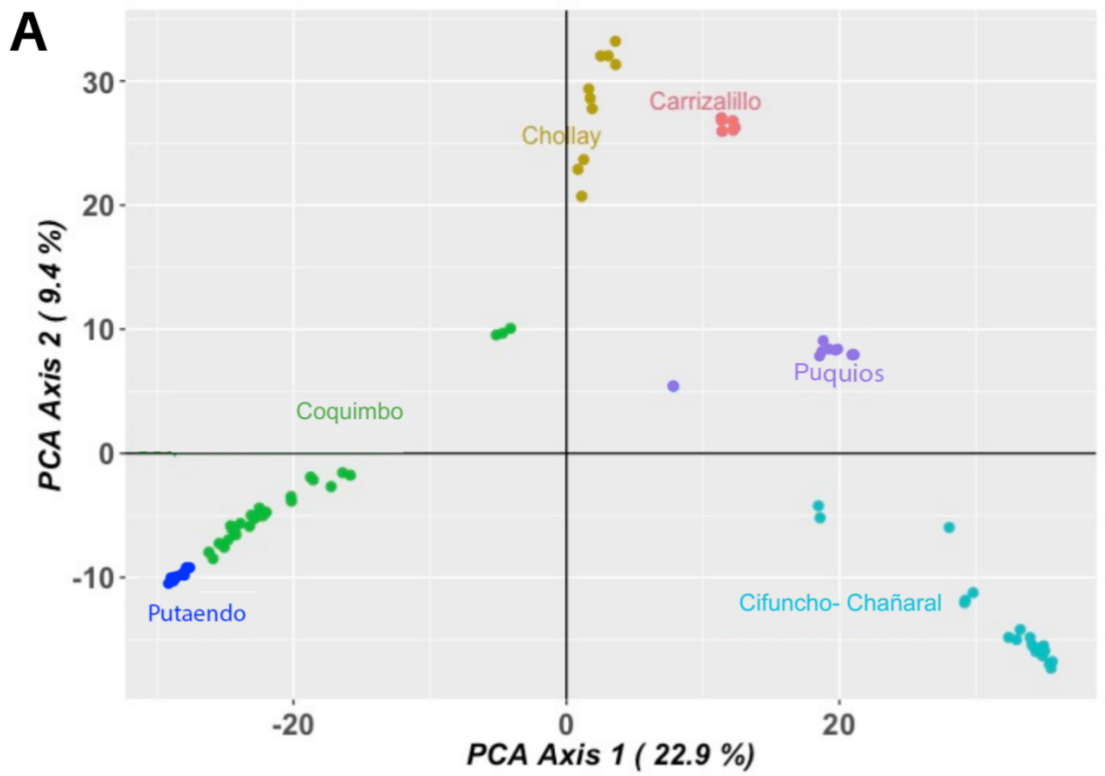


Figura 11. Análisis de coordenadas principales (PCoA) considerando K=6 según neutralidad. (A) Realizado con matriz 1 neutral. (B) Realizado con matriz 2 putativos a adaptación.

3.5 Diferenciación genética neutrales v/s candidatos según Modelo Mixto de Factores Latentes

Para los grupos genéticos obtenidos con ambas matrices, el índice de fijación (F_{ST}) dio significativo ($p < 0.05$). Para los grupos obtenidos con la matriz neutral (Tabla 3 y Figura 12A) se obtuvieron valores más cercanos a 0 en comparación con los valores de F_{ST} en la matriz de datos candidatos a adaptación (Tabla 3 y Figura 12B), la cual tiene valores más cercanos a 1.

El resultado para el set neutral (Tabla 3 y Figura 12A), es concordante con el análisis de estructura, indicando que el grupo más diferenciado es Carrizalillo, dado que tiene valores más cercanos a 1, lo cual es coincidente con el set de datos putativamente adaptativos que también indica ser el grupo más diferenciado (Tabla 3 y Figura 12B). De todas formas, hay pocos valores muy cercanos a 0, lo que indica que hay bastante estructura para ambos conjuntos de datos, pero en general hay mayor estructura para datos candidatos a adaptación.

Tabla 3

Valores F_{ST} para los grupos obtenidos con la matriz de SNPs neutrales (sobre la diagonal) y para datos candidatos a adaptación (bajo la diagonal). Valor 0 indica grupos idénticos, mientras que 1 indica diferenciación absoluta.

	Cifuncho- Ch	Coquimbo	Putando	Puquios	Chollay	Carrizalillo
Cifuncho- Ch	0	0.415	0.482	0.356	0.431	0.449
Coquimbo	0.633	0	0.12	0.427	0.363	0.458
Putando	0.683	0.066	0	0.514	0.475	0.547
Puquios	0.518	0.596	0.749	0	0.491	0.524
Chollay	0.663	0.727	0.681	0.599	0	0.533
Carrizalillo	0.556	0.625	0.761	0.649	0.579	0

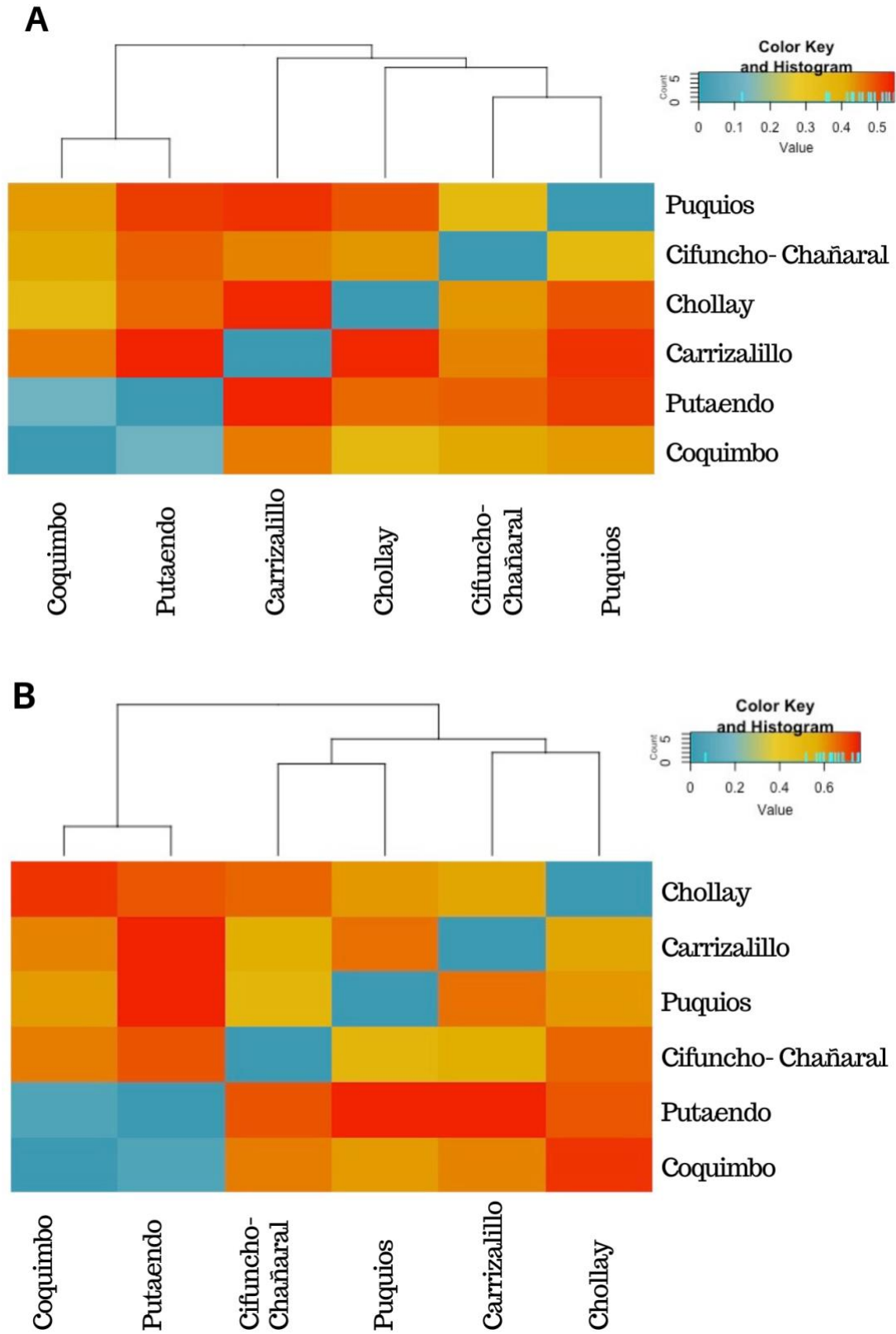


Figura 12. Heat map FST. (A) Similitud de los clústeres genéticos según datos neutrales. (B) Similitud de los clústeres genéticos según datos candidatos a adaptación.

Tabla 4

Tabla resumen AMOVA a distintos niveles de jerarquía.

Variación	(c) Adaptativos	Neutrales
Entre poblaciones	57% (p < 0.05)	37.64% (p < 0.05)
Entre subpoblaciones dentro de las poblaciones	12.44% (p < 0.05)	8.42% (p < 0.05)
Entre muestras dentro de las subpoblaciones	16.93% (p < 0.05)	27.27% (p < 0.05)
Dentro de las muestras	13.62% (p < 0.05)	26.66% (p < 0.05)

Los resultados del análisis de varianza molecular (AMOVA) indican tanto en la matriz neutral como candidata a adaptación que, de todos los niveles jerárquicos, la mayor variación es entre poblaciones (Tabla 4), es decir, entre los 6 clústeres genéticos. Además, las pruebas de significancia muestran que las variaciones en todas las categorías son significativas ($p < 0.05$) con respecto a lo que se esperaría por azar. Valores Phi (Φ) indican mismo resultado, pero en decimal, por lo que no se indican en una tabla, para evitar redundancia.

3.6 Predictor ambiental más contribuyente en la divergencia genética

a) Modelo de Disimilitud Generalizada según Modelo Mixto de Factores Latentes

Las métricas de la prueba de validación cruzada indican que el modelo GDM se ajusta bien a los datos y está haciendo predicciones precisas. El modelo está explicando el 86.2% de la variación de los datos, con un error promedio de 0.01. Además, la medida de correlación entre los valores observados y predichos es de 0.95, muy cercano a 1, lo que indica un excelente ajuste.

Los resultados del GDM indican que hay una fuerte relación entre la distancia ecológica y la disimilitud en la composición (Figura 13).

Solo 4 predictores son significativos del modelo ($p\text{-value} \leq 0.05$). Entre estos predictores significativos la variable BIO09 (Temperatura media del trimestre más seco) es la variable ambiental más importante seguido por el índice de aridez y la elevación. Por otro lado, PET (Evaporación) tiene valor NA, dado que no converge en ninguna de las permutaciones (Tabla 5). Por lo tanto, según estos resultados, la variable que explica mayormente las diferencias entre los grupos es la Temperatura media del trimestre más seco, la cual exhibió gran importancia entre los 8 y 15 °C. Sobre los 15 °C la pendiente disminuye mostrando un comportamiento constante, lo cual representa que sobre esta temperatura no influye en la diferenciación de los grupos. En el caso del índice de Aridez, que es la segunda variable más importante, los valores más cercanos a 0 indican hiper aridez mientras que los valores >0.65 indican humedad (United Nations Environment Programme et al. 1997). El GDM demuestra la importancia en valores hiper áridos muy cercanos a 0, e inmediatamente la pendiente se mantiene constante.

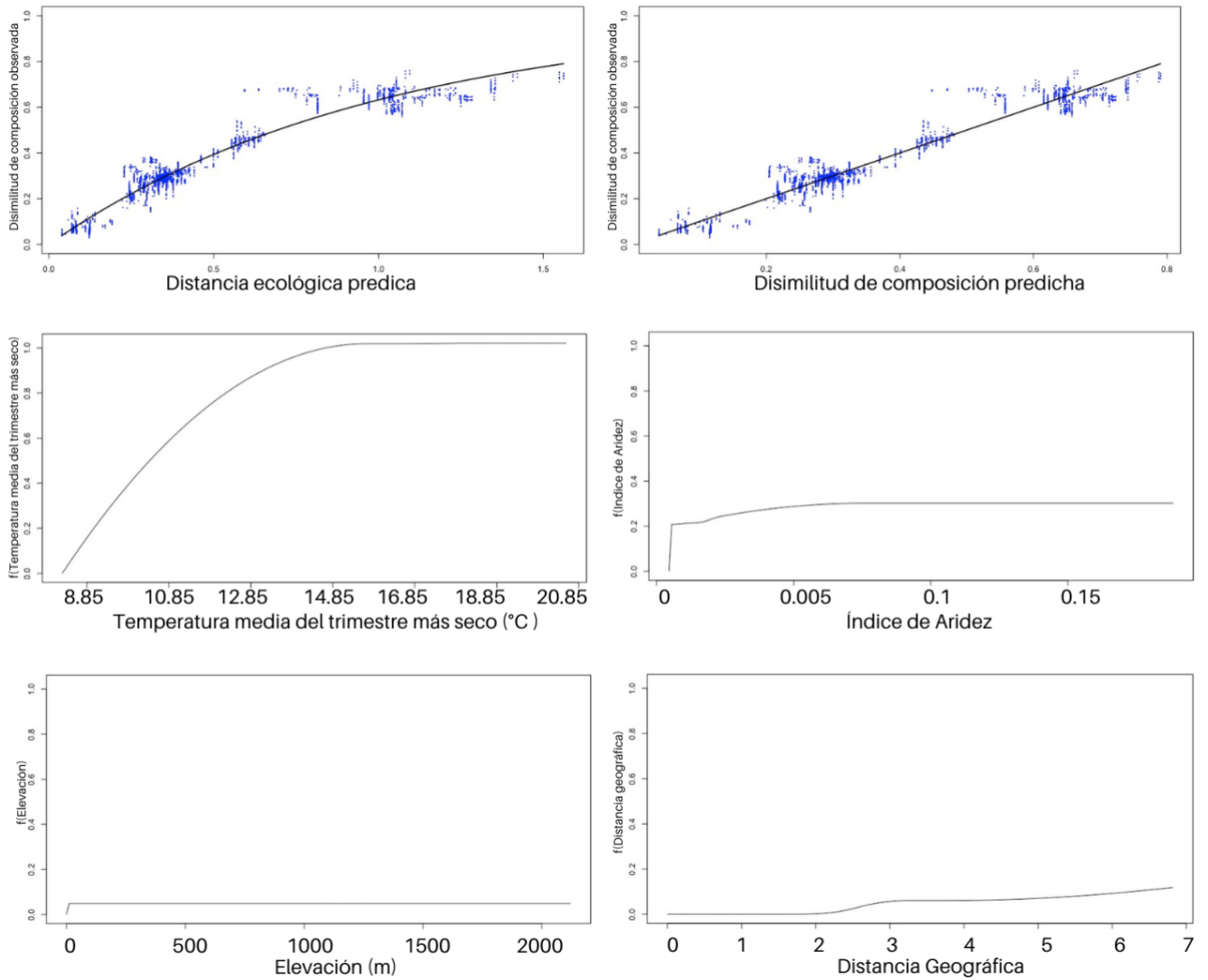


Figura 13. Resultado GDM para cada predictor ambiental significativo. En orden descendente de la variable que más contribuye a la que menos contribuye a la diferenciación de los grupos. La altura máxima de cada curva indica la cantidad total de cambio en las frecuencias alélicas asociadas con ese predictor (importancia de la variable). La forma de cada curva indica cómo varía la tasa de cambio en las frecuencias alélicas a lo largo del gradiente predictivo.

Tabla 5

Importancia (Imp) y significancia de los predictores (Pred) según el modelo GDM.

Pred	Geo	AI	BIO10	BIO15	BIO02	BIO03	BIO09	CMI	HURS	PET	RAD	E
Imp	0.262	0.965	0.349	0.185	0.083	0.04	16.93	0.162	0.019	NA	0.245	0.325
P-val	0	0	0.14	0.2	0.36	0.96	0.02	0.24	0.8	NA	0.1	0.05

b) Análisis de Redundancia (RDA)

Para el análisis RDA se intentó utilizar las mismas variables anteriores, pero al ser otro método, mostraba fuertes correlaciones que afectan a este tipo de análisis. Por lo que, siguiendo las recomendaciones del manual, se volvió a eliminar la colinealidad de variables, quedando solo 10 variables (Figura 14): BIO_02, BIO_03, BIO_09, BIO_10, BIO_14, BIO_15, PET_MEAN, RAD_ANUAL, AI y ET. El resultado del análisis de redundancia (Figura 15) apoya significativamente el papel de los predictores ambientales en la configuración de la distribución de los genotipos ($p = 0,001$; $R^2 = 0,39173$; R^2 ajustado = $0,31170$). Además, se encontró que 5 de sus ejes son significativos para buscar adaptación (Tabla 6). Explicando de manera significativa un 84% de la varianza representada en 5 biplots.

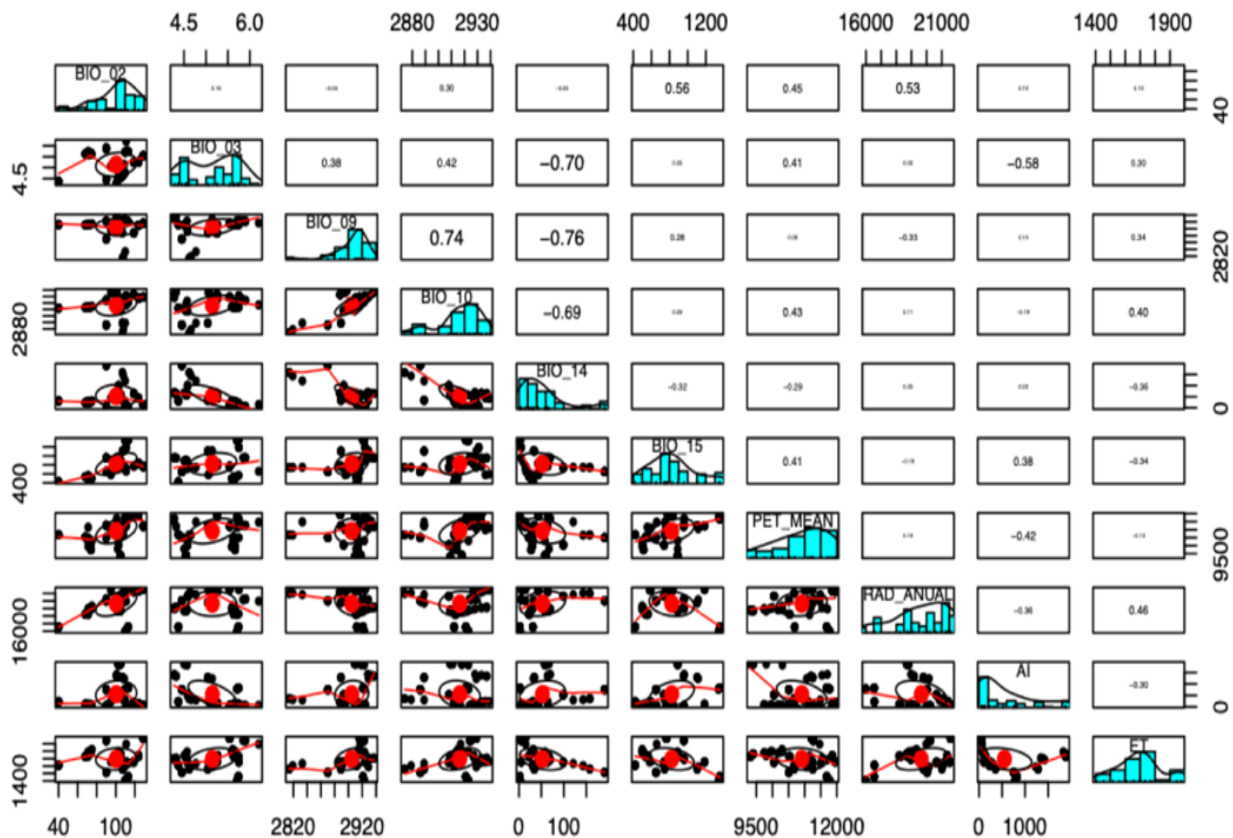


Figura 14. Tabla correlación final de variables ambientales para análisis RDA.

Tabla 6

Puntuaciones biplot que indican la magnitud de la asociación de las variables climáticas con cada eje RDA. El aumento de la asociación entre el eje RDA y la variable climática se indica mediante puntajes biplot positivos más grandes.

Código variable	Nombre Variable	RDA1	RDA2	RDA3	RDA4	RDA5
BIO_02	Rango diurno medio de temperaturas promediadas durante 1 año	0.44	-0.62	0.47	0.19	0.11
BIO_03	Isotermalidad	-0.55	-0.09	0.12	-0.04	0.39
BIO_09	Temperatura media del Trimestre más seco del año	-0.15	0.56	0.52	0.13	0.19
BIO_10	Temperatura de trimestre más cálido del año	-0.13	-0.05	0.41	0.14	0.13
BIO_14	Precipitación del mes más seco	0.4	-0.2	-0.29	-0.16	-0.14
BIO_15	Coefficiente de Variación de Estacionalidad de las Precipitaciones	0.58	-0.03	0.15	0.09	0.5
PET_MEAN	Evaporación potencial mensual promedio durante 1 año	-0.06	-0.57	0.18	-0.23	0.35
RAD_ANUAL	Radiación	-0.05	-0.67	0.25	0.13	0.13
AI	Índice de aridez global	0.79	0.46	-0.01	0.29	-0.24
ET	Evapotranspiración potencial de referencia global	-0.56	-0.04	0.38	0.64	-0.01

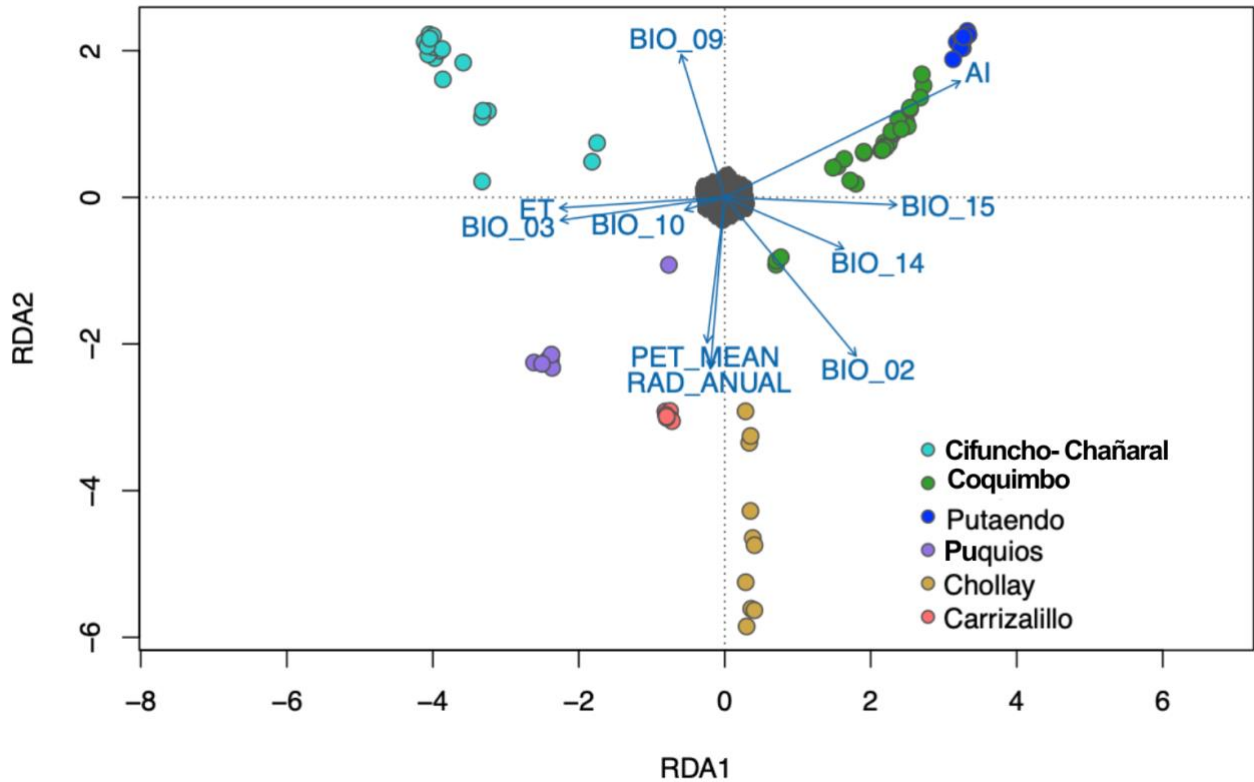


Figura 15: Biplot que resume resultados del RDA ejes 1 y 2. La nube de puntos gris oscuro en el centro del gráfico representa los SNP, y los puntos coloreados representan los individuos codificados según población (K=6). Los colores son: celeste= Cifuncho-Chañaral, verde= Coquimbo, Azul= Putaendo, púrpura= Puquios, mostaza= Chollay y rojo= Carrizalillo. Los vectores azules representan predictores ambientales (consulte el texto para ver las abreviaturas), donde la longitud de la flecha refleja la cantidad de variación en los genotipos SNP explicada por esa variable y los ángulos de las flechas representan la correlación entre las variables.

El biplot de la figura 15, muestra dos de los cinco ejes RDA significativos para nuestros datos. Los predictores climáticos más influyentes en la variación de los datos según RDA son el índice de aridez (AI) para el eje RDA 1 y la temperatura media del trimestre más seco (BIO_09) para el eje RDA 2 (Tabla 6).

Por otro lado, se encontraron 1179 SNPs candidatos a adaptación, pero luego al eliminar duplicados en los ejes quedó un total de 1064 SNPs candidatos a adaptación. En el eje 1 se encontraron 24 SNPs, en el eje 2 hay 187 SNPs, en el eje 3 hay 90 SNPs, en el eje 4 hay 225 SNPs y en el eje 5 se encontraron 653 SNPs candidatos a adaptación. Cada uno de estos SNPs tienen mayor correlación con determinadas variables (Tabla 7).

Tabla 7

Cantidad de SNPs candidatos a adaptación, clasificados según la variable ambiental con la que muestran mayor asociación en el análisis RDA.

Predictor	AI	BIO02	BIO03	BIO09	BIO10	BIO14	BIO15	ET	PET	RAD
Cantidad SNPs	70	94	79	72	3	4	47	542	9	144

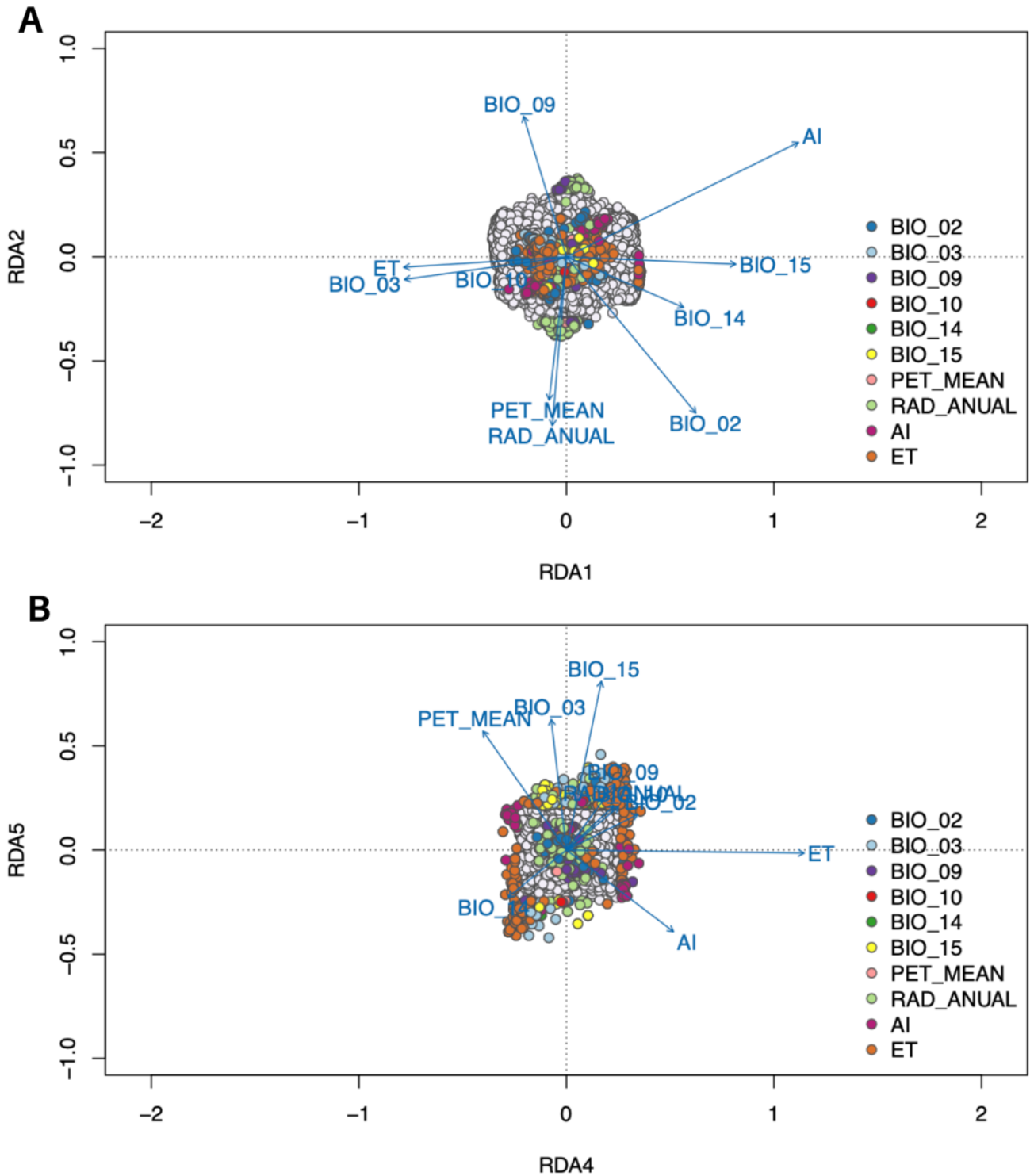


Figura 16. Resultado RDA para resaltar las cargas de SNP. (A) los ejes 1 y 2 de RDA. (B) los ejes 4 y 5. Forma gráfica de visualizar la relación de SNPs con predictores. Los SNP candidatos se muestran como puntos coloreados con codificación según el predictor ambiental más altamente correlacionado. Los SNP no identificados como candidatos (SNP neutrales) se muestran en blanco. Los vectores azules representan predictores ambientales.

En los ejes RDA 1 y 2 (Figura 16A) la mayoría de los candidatos están altamente correlacionados con RAD_ANUAL y BIO_09. Se puede observar cómo los SNP candidatos que se correlacionan con ET, AI y BIO_03 se ubican en el centro de la gráfica y no se detectarán como valores atípicos en estos ejes (Figura 16A). Sin embargo, estos loci se detectan como valores atípicos en los ejes 4 y 5 (Figura 16B).

c) Contraste con resultado Modelo Generalizado de Disimilitud según matriz concatenada Modelo Mixto de Factores Latentes y Análisis de Redundancia

Entre los 1064 SNPs candidatos a adaptación encontrados mediante RDA y los 475 mediante lfmm, 101 SNPs fueron detectados en común entre ambas metodologías. Según la información que se obtuvo del RDA, cada uno de estos SNPs está mayormente correlacionado a una variable ambiental (Tabla 8), las variables detectadas entre ambas metodologías fueron BIO02, BIO03, BIO09, BIO15, ET y Radiación (Tabla 8).

Tabla 8

Cantidad de SNPs candidatos a adaptación concatenados entre RDA y LFMM, clasificados según la variable ambiental con la que muestran mayor asociación. Resaltadas en color gris, las variables de temperatura. BIO02= Rango de temperatura medio diario, BIO03= Isotermalidad, BIO09= Temperatura media del cuarto más seco.

Variable	BIO02	BIO03	BIO09	BIO15	ET	RAD
SNPs	3	1	27	10	52	8

Para verificar la influencia de la temperatura en la variación de los datos y contrastar con resultado GDM, se realizó otra matriz según SNPs concatenados entre ambas metodologías, pero solo los SNPs más correlacionados con variables de temperatura. De este modo, se consideraron los predictores Rango medio diurno de temperatura (BIO02), Isotermalidad (BIO03) y Temperatura media del cuarto más seco (BIO09), resultando ser una matriz de 31 SNPs (tabla 8). Se han resaltado las variables de temperatura (tabla 8) para contrastar la importancia de variables de temperatura en general.

El análisis de varianza molecular (AMOVA) resuelve que los 31 SNPs concatenados entre ambas metodologías que estaban más correlacionados con variables de temperatura (BIO02, BIO03 y BIO09), explican un 88,6% de la variación genética entre los 6 clústeres genéticos (tabla 9). Los valores Phi soportan el mismo resultado, indicando alta diferenciación en los 6 clúster genéticos por SNPs asociados a temperatura.

Tabla 9

Tabla resumen AMOVA y Phi (Φ) a distintos niveles de jerarquía.

Variación	Temperatura	Phi (Φ)
Entre poblaciones	88.6% ($p < 0.05$)	0.887
Entre subpoblaciones dentro de las poblaciones	2.4% ($p < 0.05$)	0.216
Entre muestras dentro de las subpoblaciones	1.7% ($p < 0.05$)	0.2
Dentro de las muestras	7.1% ($p < 0.05$)	0.929

4. DISCUSIÓN

Mediante este estudio se confirma la existencia de variación genética adaptativa entre las poblaciones de las especies de *E. aurata* y *E. rodentiophila*. De tal manera que tal como se predijo, existe una menor diferenciación genética en SNPs neutrales comparado con SNPs candidatos a adaptación. La variación adaptativa se asocia significativamente con la temperatura media del trimestre más seco y también con el índice de aridez. Estos resultados sugieren que la temperatura de los meses en que menos llueve (verano para esta zona, es decir diciembre, enero y febrero) y la aridez, serían las variables que más influyen sobre la diferenciación adaptativa de los grupos genéticos y, de esta forma, promueven la especiación ecológica, tal como fue planteado en la hipótesis.

4.1 Predictores climáticos y búsqueda de adaptación

Luego de una reducción de variables ambientales, se detectaron 475 SNPs candidatos a adaptación (Matriz 2), lo que corresponde al 1,8% de los SNPs rescatados. Se definieron variables apropiadas para evitar colinealidad, lo cual se produce cuando hay una correlación fuerte entre dos o más variables predictivas (Graham 2003). De esta forma, evitamos que esto pueda causar inestabilidad en la estimación de parámetros en el modelo (Dormann et al. 2013). Se trabajó con distintas variables climáticas asociadas a temperatura, precipitaciones, radiación, entre otras. Con estos predictores se realizó la detección de SNPs con potencial a ser adaptativos según el método LFMM, encontrando un total de 475 SNPs candidatos a adaptación (Matriz 2) y a su vez obtuvimos la información de los datos SNPs neutrales (Matriz 1). De esta forma, los datos genéticos más influenciados por efectos

de la deriva génica y la historia demográfica quedan separados de los putativos a adaptación local (Novembre and Di Rienzo 2009; Fritchot et al. 2013), y así se pudo evaluar la evidencia de la adaptación local probando si las variables ambientales explicaban más la varianza que los datos nulos.

4.2 Estructura genética neutral

El análisis de estructura poblacional mediante datos neutrales (matriz 1) reveló 6 clústeres genéticos ($K=6$) entre ambas especies. La mayoría de la variación genética es entre grupos, en lugar de dentro de ellos (Tabla 4). Todos los clústeres genéticos tienen similares niveles de divergencia genética (Tabla 3). Sin embargo, entre los grupos Putaendo y Coquimbo (Fig. 12A) han acumulado menos diferencias genéticas en comparación que los otros grupos, pero sí suficientes para que sean grupos distintos según la neutralidad (Fig. 8A). Los 6 clústeres genéticos sugieren historias evolutivas independientes entre sí, y dado el flujo genético que hay, podría ser reciente y estar en vías de especiación. La divergencia en la especiación la comienza en una pequeña parte del genoma que participa en la generación del aislamiento reproductivo entre formas incipientes (Wu 2001). En un futuro, incluso podría haber repercusiones taxonómicas para estos grupos, porque a pesar de que la especie *E. rodentiophila* fue detectado como un único grupo genético, en la especie *E. aurata* fueron detectados cinco grupos genéticos, lo cual no había sido reportado con anterioridad.

4.3 Estructura genética adaptativa

Según los datos que resultaron ser putativamente adaptativos (Matriz 2) de 475 SNPs, los individuos de estas especies se agrupan en 8 poblaciones. Esto indica una respuesta a cambios micro ambientales que pueden ser a lo largo del gradiente continuo ecológico y geográfico, lo cual permite acumulación de diferencias genéticas, incluso formando estructuras genéticas (Duncan et al. 2015; Jiang et al. 2019; Yang et al. 2020). Es interesante reconocer que, según los datos putativamente adaptativos, *E. rodentiphila* es separado en 2 grupos genéticos. Por otro lado, *E. aurata* estaría formado por 6 grupos genéticos.

La diferenciación genética en SNPS candidatos a adaptación es mayor a la encontrada para SNPs neutrales (Figura 8). Al comparar los resultados entre datos neutrales (Matriz 1) y candidatos a adaptación (Matriz 2), para determinar en qué tipo de datos hay mayor diferenciación genética, logramos identificar que hay mayor divergencia adaptativa v/s la divergencia neutral (Tabla 3, 4 y Figura 12). El análisis de estructura mostró que, con el set de datos candidatos a adaptación hay mayor cantidad de grupos ($K=8$), en comparación con los datos neutrales, donde resultaron ser menos grupos genéticos ($K=6$). Los valores del índice de fijación (F_{ST}) que considera los clústeres neutrales ($K=6$) que estima los grados de diferenciación genética entre grupos (Weir and Cockerham 1984), también son más altos para los datos putativamente adaptativos, sugiriendo mayor diferenciación. El AMOVA igualmente indica que datos candidatos a adaptación explican mayormente la variación de los 6 clústeres genéticos en comparación con datos neutrales.

Estos resultados reflejan que los grupos genéticos están mayormente diferenciados por variaciones adaptativas, validando la primera predicción de este trabajos. De

hecho, el análisis RDA, determinó que las variables explicativas del modelo (variables climáticas) justifican el 31% de la varianza, lo cual es una cantidad relativamente alta de variación en los genotipos SNP explicada por el clima si tenemos en consideración que, solo proporciones relativamente pequeñas del genoma controlan rasgos adaptativos específicos (Harrisson et al. 2017).

Debido a todo esto, los resultados indican que las presiones de selección han llevado a una mayor diferenciación entre las poblaciones en términos de rasgos adaptativos en comparación con los rasgos neutrales. Resultados similares han sido reportados sobre divergencia genética significativa entre dos grupos, que fue detectada para loci selectivos en lugar de neutrales, sugiriendo que los rasgos fueron seleccionados por factores ambientales (Yang et al. 2020). Esto puede ser porque en entornos cambiantes, la selección determina la diversidad genética de los genes adaptados, en cambio la diversidad genética neutral se ve reducida por asociaciones aleatorias independiente de las aptitudes (Neher 2013; Jiang et al. 2019) que puedan tener los individuos.

4.4 Divergencia asociada a la temperatura y al índice de aridez

En esta investigación se reporta adaptación local en las distintas poblaciones de *E. aurata* y *E. rodentiophila*, impulsada principalmente por la temperatura media del trimestre más seco (BIO-09) y el índice de aridez (AI) validando la segunda predicción de este trabajo. Esto se determinó gracias a distintos análisis que se apoyan entre sí. Por ejemplo, mediante el RDA se pudo interpretar que el eje 1 es influenciado mayormente por el índice de aridez y el eje 2 por variables relacionadas a la temperatura como BIO-09 y BIO-02. Según esto, la población Cifuncho-Chañaral que corresponde a la especie *E. rodentiophila* destaca por estar en las condiciones más áridas y cálidas. Por otro lado, la población Chollay, que pertenece a la especie *E.*

aurata, tiene el menor valor en el índice de aridez y está en menores temperaturas. Por lo demás, Chollay tiene la particularidad que los individuos que lo componen difieren bastante entre sí dentro del rango de temperatura (Fig. 15, eje 2). Gracias al GDM se determinó el rango específico donde la temperatura media del trimestre más seco (BIO-09) contribuye más a las diferencias de los 6 clústeres genéticos, la cual se ve reflejada en la pronunciada pendiente dentro entre los 8 y 15 °C (Figura 13). En este rango de temperatura, se encuentran los individuos del grupo Chollay donde en promedio la temperatura media del trimestre más seco es de 13°C, sin embargo, los demás grupos oscilan entre los 17 y los 19°C. Esto sugiere que la población Chollay podría estar experimentando una diferenciación significativa de las otras poblaciones impulsadas por la temperatura de verano. Estas pequeñas diferencias están asociadas a grandes cambios en la disimilitud biológica (Ferrier et al. 2007), y que contribuyen a la adaptación local de los grupos. Estos resultados, podrían estar sugiriendo que la mayoría de las poblaciones se han expandido hacia los extremos gracias a las adaptaciones tanto a mayores temperaturas, como también a mayor aridez. Chollay podría estar en condiciones diferentes en comparación a las otras poblaciones, adaptada a menores temperaturas.

Es posible distinguir patrones que manifiesten el efecto de la temperatura en rasgos fenotípicos y su importancia sobre todo para ambientes desérticos. La temperatura media del trimestre más seco puede afectar a distintas áreas biológicas en el desarrollo de estas especies y dada la escasez del recurso hídrico, la temperatura podría estar tomando un papel protagónico. Se ha reportado que las actividades de las plantas del desierto están altamente reguladas por la disponibilidad de agua y el período de severa limitación hídrica (Gurera and Bhushan 2020; Mohanta et al. 2023). Durante el verano, cuando la humedad y el agua son limitados, la transferencia de calor se vuelve especialmente desafiante (Sandquist 2014). El estrés

por calor, solo o combinado con otro factor como la sequía, puede afectar gravemente la integridad de las membranas tilacoides y la función del fotosistema II (Zhu et al. 2021; Ostria-Gallardo et al. 2023), restringiendo la fotosíntesis. Al ser una época con reducidas precipitaciones, las plantas de desierto aseguran su supervivencia cuidando al máximo el agua de sus tejidos, lo que implica mantener cerrados sus estomas para disminuir la transpiración (Taiz and Zeiger 2007). Por lo tanto, la temperatura externa influiría en cuánto aumenta la temperatura interna de la planta, y podría poner en riesgo de desecación el tallo y el sistema radicular del cactus. Para una plántula nueva, la restricción de la fotosíntesis y procesos metabólicos podría ser determinante en su supervivencia. Se ha reportado que una adaptación importante para plantas de desiertos extremos es la tolerancia de las plántulas jóvenes a los períodos de desecación (Gutterman 2002; Huang and Gutterman 2007), por lo que no tener suficientes reservas hídricas o una gran longitud de su raíz para abastecerse de agua, podría ser crítico. De esta forma, la temperatura media del trimestre más seco podría afectar al reclutamiento de nuevas plántulas, debido a la desecación.

En las regiones frías del desierto, como, por ejemplo, donde se encuentra la población Chollay, las precipitaciones pueden ser en forma de nieve, lo cual podría tener un papel único respecto al crecimiento y desarrollo de las plantas (Fan et al. 2014), sobre todo si pensamos en las plántulas. También podría verse diferenciada de los otros grupos por no poder realizar con normalidad su apertura estomática nocturna debido a bajas temperaturas, afectando su captura de CO₂. Sería interesante cuantificar la importancia relativa en la divergencia de rasgos genéticos y funcionales, para revelar procesos subyacentes de las estrategias adaptativas (Thomas 2011), y así obtener respuestas certeras.

El efecto general de la temperatura en la divergencia adaptativa ha sido documentado, por ejemplo, en un trabajo se estudió la vinculación de los procesos ecológicos y evolutivos con la adaptación al ambiente, en un pariente del jengibre llamado *Zingiber kawagoii*, que habita en Taiwán en ambientes de bosques, laderas y valles (Li et al. 2022). El estudio, deja en evidencia que la divergencia fue impulsada por la selección, actuando en un estrecho rango latitudinal, y que esta divergencia adaptativa se vio afectada principalmente por el rango de temperatura anual (Li et al. 2022). Para el caso de los desiertos, una investigación sobre suculentas del clado *Euphorbia* identificó un efecto diferencial entre condiciones secas y frías en comparación con condiciones moderadamente secas y cálidas, evidenciando que las variaciones de temperatura inducen una respuesta adaptativa en los desiertos (Evans et al. 2014). Además, ese mismo trabajo demuestra la asociación de la suculencia como adaptación a la aridez.

Las adaptaciones a la aridez y su relación con la especiación han sido de interés para muchos autores. Por ejemplo, para el género biespecífico *Pugionium*, endémico del desierto de Kubuqi y de Mu Us del noroeste de China, se examinó la divergencia genómica en relación con el ambiente desértico de dichas especies a nivel de población (Hu et al. 2021). La evidencia indicó que estas especies comenzaron a divergir cuando hubo un aumento global de la aridez, promoviendo así la divergencia inicial de las dos especies a través de la adaptación del microhábitat desértico (Hu et al. 2021). Además, los genes con señales de selección estaban involucrados en desarrollo de las raíces, la morfogénesis de las hojas y la adaptación al microhábitat (germinación de semillas y tolerancia a la sequía/sal), lo que se corresponde bien con las divergencias interespecíficas en estos aspectos (Hu et al. 2021). También se ha documentado la relación de la aridez con la especiación en otros grupos de estudio, los reptiles, en que la colonización de ambientes áridos está

repetidamente asociada con cambios en la pérdida de agua por evaporación, un rasgo genéticamente adaptativo (Cox and Cox 2015). Este rasgo contribuye a procesos ecológicos y evolutivos exitosos para la colonización de nuevos hábitats, teniendo concordancia con la relación ancestral de los grupos y especies de reptiles. En el desierto de Atacama también se hizo un estudio en relación con lo mismo sobre poblaciones de lagartijas de la especie *Liolaemus fuscus* donde los resultados sugieren que hay especiación adaptativa incipiente en relación con el efecto de la aridez, proporcionando evidencia sobre mecanismos que permiten especiación y adaptación al desierto (Araya-Donoso et al. 2022).

Dada la hiper aridez del desierto de Atacama, las estrategias selectivas en los distintos grupos que habitan ahí, tienen que ver en la resistencia y tolerancia a la aridez y son de interés evolutivo porque la aridez actúa como una presión selectiva que promueve cambios evolutivos y especiación (Whitford and Duval 2019). En el desierto de Atacama, las oscilaciones de los períodos de glaciaciones del Pleistoceno crearon ciclos húmedo-secos que permitieron expansiones y contracciones de rangos, lo que sumado a la incipiente aridez durante el Holoceno (Maldonado et al. 2005) pudo haber promovido el aislamiento de los grupos como ya ha sido reportado anteriormente para taxones de Neoporteria (Guerrero et al. 2011). Esto pudo generar que los nichos se fueran segregando y conservando, promoviendo la divergencia genética (Wiens and Graham 2005; Finger et al. 2022). Como se puede ver, la hipótesis sobre la relación de la aridez y la diversificación no es algo nuevo, pero, lo complejo para plantas suculentas es que dada la ausencia de registros fósiles y reconstrucciones paleo climáticas confiables, es una hipótesis difícil de probar (Thompson et al. 2023).

Para este caso de estudio, las variables climáticas protagónicas son la temperatura y aridez, que generan la suficiente heterogeneidad ambiental que mantiene a los 6 grupos aislados genéticamente. Esto gracias a alguna adaptación específica que no fue objeto de estudio, que les fue confiriendo una ventaja en distintos rangos dentro del gradiente. No hay que dejar de lado que los rangos en las variables predictoras que separan a los 6 clústeres genéticos son muy estrechos, incluso para la variable que más contribuye a las diferencias genéticas (BIO-09). Por ejemplo, para el índice de aridez (AI), que es la segunda variable significativa más importante, en 5 poblaciones los rangos caen dentro de la categoría hiper árido (<0.03) (Zomer et al. 2022), y solo una población (Puquios) está fuera de este rango con un promedio que cae dentro de valores áridos ($0.03 - 0.2$). Para estos estrechos rangos hay que tener en consideración las limitaciones del modelo al extrapolar relaciones basándose en datos disponibles, o también la posibilidad del efecto de más factores que incluso pueden no haber sido incluidos en el modelo. Tampoco hay que olvidar la interacción entre variables que, de manera complementaria están influyendo en la disimilitud biológica. No es extraño esperar la complementariedad de las variables BIO_09 y AI, dado que la temperatura tiene un directo efecto en la aridez.

Por otro lado, se destaca en este trabajo que se utilizaron distintos métodos y perspectivas para obtener resultados más certeros y fueron capaces de complementarse entre sí, brindando robustez a nuestros resultados. De hecho, con los distintos métodos utilizados, podemos afirmar la importancia de variables de temperatura y la aridez en la diferenciación genética de los grupos. Además, los métodos utilizados para identificar adaptación local mediante la exhibición de SNPs con una alta correlación de variables ambientales, es una buena forma cuando hay efectos fenotípicos débiles (Joost et al. 2007; Frichot et al. 2013).

5. CONCLUSIÓN

Nuestros análisis indican 6 clústeres genéticos, sin grandes distancias geográficas de separación o diferencias morfológicas evidentes detectadas, lo que sugiere que las barreras climáticas que más contribuyen a estas diferencias (según la evidencia de estos resultados la temperatura y aridez), además del aislamiento reproductivo resultante, puede ser reciente. Por lo tanto, los 6 grupos se encuentran aún en etapas tempranas de divergencia. Gracias a esta investigación, encontramos la evidencia esperada de una alta divergencia en genes putativamente adaptativos por efecto de la aridez y la temperatura. Este trabajo logró detectar evidencia adaptativa y sus impulsores ambientales para la especiación ecológica, que apoyan la hipótesis de aridificación para las cactáceas. No obstante, quedaron muchas preguntas planteadas que en un futuro se podrían investigar. Este enfoque de adaptación climática local es muy útil para poder considerar según la resiliencia de las poblaciones, planes de conservación y priorización en climas futuros en el desierto de Atacama.

Por último, se confirma que el gradiente de aridez es un factor clave en la divergencia de las poblaciones de estas especies, actuando como una presión selectiva que impulsa la evolución de sus nichos.

6. REFERENCIAS

- Aguirre-Liguori, J. A., B. S. Gaut, J. P. Jaramillo-Correa, M. I. Tenaillon, S. Montes-Hernández, F. García-Oliva, S. J. Hearne, and L. E. Eguiarte. 2019. Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). *Mol Ecol* 28:2814–2830. Blackwell Publishing Ltd.
- Aguirre-Liguori, J. A., S. Ramírez-Barahona, and B. S. Gaut. 2021. The evolutionary genomics of species' responses to climate change. *Nature Research*.
- Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. E. Eguiarte. 2017. Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Mol Ecol* 26:4226–4240. Blackwell Publishing Ltd.
- Arakaki, M., P. A. Christin, R. Nyffeler, A. Lendel, U. Eggli, R. M. Ogburn, E. Spriggs, M. J. Moore, and E. J. Edwards. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci U S A* 108:8379–8384. National Academy of Sciences.
- Araya-Donoso, R., E. San Juan, Í. Tamburrino, M. Lamborot, C. Veloso, and D. Véliz. 2022. Integrating genetics, physiology and morphology to study desert adaptation in a lizard species. *Journal of Animal Ecology* 91:1148–1162.
- Axelrod, D. I. (1972). Edaphic Aridity as a Factor in Angiosperm Evolution. *The American Naturalist*, 106(949), 311–320. <https://doi.org/10.1086/282773>
- Clarke, J. D. A. 2006. Antiquity of aridity in the Chilean Atacama Desert. *Geomorphology* 73:101–114. Elsevier.
- Cox, C. L., and R. M. Cox. 2015. Evolutionary shifts in habitat aridity predict evaporative water loss across squamate reptiles. *Evolution (N Y)* 69:2507–2516. Blackwell Publishing Inc Malden, USA.
- Darwin, C. 1859. *On The Origin of Species*. London: Murray.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B.

- Reineking, B. Schröder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46. John Wiley & Sons, Ltd.
- Duncan, C. J., J. R. P. Worth, G. J. Jordan, R. C. Jones, and R. E. Vaillancourt. 2015. Genetic differentiation in spite of high gene flow in the dominant rainforest tree of southeastern Australia, *Nothofagus cunninghamii*. *Heredity* 2016 116:1 116:99–106. Nature Publishing Group.
- Mayr, E. 1947. Ecological factors in speciation. *JSTOR* 1.
- Edwards, K., C. Johnstone, and C. Thompson. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* 19:1349. Oxford University Press.
- Evans, M., X. Aubriot, D. Hearn, M. Lanciaux, S. Lavergne, C. Cruaud, P. P. Lowry, and T. Haevermans. 2014. Insights on the evolution of plant succulence from a remarkable radiation in Madagascar (*Euphorbia*). *Syst Biol* 63.
- Fan, L.-L., L.-S. Tang, L.-F. Wu, J. Ma, and Y. Li. 2014. The limited role of snow water in the growth and development of ephemeral plants in a cold desert. *Journal of Vegetation Science* 25:681–690.
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12), 4302-4315.
- Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib* 13:252–264. John Wiley & Sons, Ltd.
- Finger, N., K. Farleigh, J. T. Bracken, A. D. Leach, O. François, Z. Yang, T. Flouri, T. Charran, T. Jezkova, D. A. Williams, and C. Blair. 2022. Genome-Scale Data Reveal Deep Lineage Divergence and a Complex Demographic History in the Texas Horned Lizard (*Phrynosoma cornutum*) throughout the Southwestern and Central United States. *Genome Biol Evol* 14. Oxford Academic.
- Fitzpatrick M, Mokany K, Manion G, Nieto-Lugilde D, Ferrier S, Lisk M, and ... 2022. Package “gdm” Title Generalized Dissimilarity Modeling., doi: 10.1111/geb.13459.

- Folk, R. A., Siniscalchi, C. M., & Soltis, D. E. 2020. Angiosperms at the edge: Extremity, diversity, and phylogeny. *Plant, Cell & Environment*, 43(12), 2871-2893.
- Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol Ecol* 27.
- Franks, S. J., and A. A. Hoffmann. 2012. Genetics of Climate Change Adaptation., doi: 10.1146/annurev-genet-110711-155511.
- Frichot, E., and O. François. 2015. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol* 6:925–929. John Wiley & Sons, Ltd.
- Frichot, E., F. Mathieu, T. Trouillon, G. Bouchard, and O. François. 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196.
- Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Mol Biol Evol* 30:1687–1699. Oxford Academic.
- Fuentes, G., F. González, J. Saavedra, P. López-Sepúlveda, P. F. Victoriano, T. F. Stuessy, and E. Ruiz-Ponce. 2021. Assessing signals of selection and historical demography to develop conservation strategies in the Chilean emblematic *Araucaria araucana*. *Sci Rep* 11. Nature Research.
- Futuyma DJ, and Kirkpatrick M. 2017. *Evolution* 4th edition.
- Garot, E., S. Dussert, F. Domergue, T. Joët, I. Fock-Bastide, M. C. Combes, and P. Lashermes. 2021. Multi-Approach Analysis Reveals Local Adaptation in a Widespread Forest Tree of Reunion Island. *Plant Cell Physiol* 62.
- Georges A, and Gruber B. 2019. SNP Analysis using dartR Guide to Preparatory Analysis IA E Institute for Applied Ecology.
- Gibson, A., and P. Nobel. 1986. *The cactus primer*.
- Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325–338. Oxford Academic.

- Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Statistical Reports. Ecology* 84:2809–2815.
- Gruber, B., P. Unmack, ... O. B.-M. ecology, and undefined 2017. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Wiley Online Library
- Gruber, B., P. Unmack, O. Berry, A. G.-U. Manual, and undefined 2019. Introduction to dartR. cran.r-hub.io Gruber, P Unmack, O Berry, A GeorgesUser Manual, 2019•cran.r-hub.io.
- Guerrero, P. C., M. T. K. Arroyo, R. O. Bustamante, M. Duarte, T. K. Hagemann, and H. E. Walter. 2011. Phylogenetics and predictive distribution modeling provide insights into the geographic divergence of *Eriosyce* subgen. *Neoporteria* (Cactaceae). *Plant Systematics and Evolution* 297:113–128. Springer-Verlag Wien.
- Guerrero, P. C., L. C. Majure, A. Cornejo-Romero, and T. Hernández-Hernández. 2019a. Phylogenetic relationships and evolutionary trends in the cactus family. *Journal of Heredity*.
- Guerrero, P. C., M. Rosas, M. T. K. Arroyo, and J. J. Wiens. 2013. Evolutionary lag times and recent origin of the biota of an ancient desert (Atacama-Sechura). *Proc Natl Acad Sci U S A* 110:11469–11474.
- Guerrero P. C., H. E. Walter, M. T. K. Arroyo, C. M. Peña, I. Tamburrino, De Benedictis M., and Larridon I. (2019) Molecular phylogeny of the large South American genus *Eriosyce* (Notocactaceae, Cactaceae): Generic delimitation and proposed changes in infrageneric and species ranks, *TAXON*, 68, 557–573.
- Gurera, D., and B. Bhushan. 2020. Passive water harvesting by desert plants and animals: lessons from nature. *Philosophical Transactions of the Royal Society A* 378. The Royal Society Publishing.
- Gutterman, Y. 2002. *Survival strategies of annual desert plants*. Springer Science & Business Media.
- Harrisson, K. A., S. J. Amish, A. Pavlova, S. R. Narum, M. Telonis-Scott, M. L. Rourke, J. Lyon, Z. Tonkin, D. M. Gilligan, B. A. Ingram, M. Lintermans, H. M. Gan, C. M. Austin, G. Luikart, and P. Sunnucks. 2017. Signatures of polygenic adaptation

associated with climate across the range of a threatened fish species with high genetic connectivity. *Mol Ecol* 26.

Hernández-Hernández, T., J. W. Brown, B. O. Schlumpberger, L. E. Eguiarte, and S. Magall. 2014. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New phytologist* 202:1382–1397. Blackwell Publishing Ltd.

Hernández-Hernández, T., H. M. Hernández, J. Arturo De-Nova, R. Puente, L. E. Eguiarte, and S. Magallón. 2011. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot* 98.

Hernández-Hernández, T., E. C. Miller, C. Román-Palacios, and J. J. Wiens. 2021. Speciation across the Tree of Life. *Biological Reviews* 96:1205–1242. John Wiley & Sons, Ltd.

Hu, Q., Y. Ma, T. Mandáková, S. Shi, C. Chen, P. Sun, L. Zhang, L. Feng, Y. Zheng, X. Feng, W. Yang, J. Jiang, T. Li, P. Zhou, Q. Yu, D. Wan, M. A. Lysak, Z. Xi, E. Nevo, and J. Liu. 2021. Genome evolution of the psammophyte *Pugionium* for desert adaptation and further speciation. *Proceedings of the National Academy of Sciences* 118:e2025711118.

Huang, Z., and Y. Gutterman. 2007. Seedling desiccation tolerance of *Leymus racemosus* (Poaceae) (wild rye), a perennial sand-dune grass inhabiting the Junggar Basin of Xinjiang, China. *Seed Sci Res* 14:233–239. Cambridge University Press.

Jia, K.-H., W. Zhao, Paul, A. Maier, Xian-Ge Hu, Y. Jin, S.-S. Zhou, S.-Q. Jiao, Yousry, A. El-Kassaby, Tongli Wang, X.-R. Wang, and J.-F. Mao. 2020. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Wiley Online Library*.

Jiang, S., M. X. Luo, R. H. Gao, W. Zhang, Y. Z. Yang, Y. J. Li, and P. C. Liao. 2019. Isolation-by-environment as a driver of genetic differentiation among populations of the only broad-leaved evergreen shrub *Ammopiptanthus mongolicus* in Asian temperate deserts. *Scientific Reports* 2019 9:1 9:1–14. Nature Publishing Group.

Joost, S., A. Bonin, M. W. Bruford, L. Després, C. Conord, G. Erhardt, and P. Taberlet. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Mol Ecol* 16.

- Kamvar, Z. N., J. C. Brooks, and N. J. Grünwald. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.*
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4, 170122.
- Kattermann, F. 1994. *Eriosyce* (Cactaceae): the genus revised and amplified. *Succul Plant Res* 176. David Hunt.
- Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation.
- Korotkova, N., Aquino, D., Arias, S., Eggli, U., Franck, A., Gómez-Hinostrosa, C., Guerrero, P. C., Hernández, H. M., Kohlbecker, A., Köhler, M., Luther, K., Majure, L. C., Müller, A., Metzring, D., Nyffeler, R., Sánchez, D., Schlumpberger, B., & Berendsohn, W. G. (2021). Cactaceae at Caryophyllales. org—a dynamic online species-level taxonomic backbone for the family. *Willdenowia*, 51(2), 251-270.
- Lê, S., J. Josse, and François Husson. 2008. FactoMineR: A Package for Multivariate Analysis. *J Stat Softw* 25:1–18.
- Legendre, P., J. Oksanen, and C. J. F. ter Braak. 2011. Testing the significance of canonical axes in redundancy analysis. *Methods Ecol Evol* 2.
- Li, Y. S., P. C. Liao, C. Te Chang, and S. Y. Hwang. 2022. Pattern of Adaptive Divergence in *Zingiber kawagooii* Hayata (Zingiberaceae) along a Narrow Latitudinal Range. *Plants* 11:2490. MDPI.
- Lickley, M., and S. Solomon. 2018. Drivers, timing and some impacts of global aridity change. *Environmental Research Letters* 13:104010. IOP Publishing.
- Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207:437–453. Blackwell Publishing Ltd.
- Majure, L. C., D. Barrios, E. Díaz, B. A. Zumwalde, W. Testo, and V. Negrón-Ortíz. 2021. Pleistocene aridification underlies the evolutionary history of the Caribbean endemic, insular, giant *Consolea*(Opuntioideae). *American Journal of Botany* 108(2): 200–215.

- Maldonado, A., J. L. Betancourt, C. Latorre, and C. Villagran. 2005. Pollen analyses from a 50 000-yr rodent midden series in the southern Atacama Desert (25° 30' S). *J Quat Sci* 20.
- Malekmohammadi, L., M. Sheidai, F. Ghahremaninejad, A. Danekar, and F. Koohdar. 2023. Putative Local Adaptive SNPs in the Genus *Avicennia*. *Biochem Genet*, doi: 10.1007/s10528-023-10362-4. Springer.
- Mallet, J. 2008. Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:2971–2986. Royal Society.
- Matuszak, A Favre, J Schnitzler, AN Muellner-Riehl. 2016. Key innovations and climatic niche divergence as drivers of diversification in subtropical Gentianinae in southeastern and eastern Asia. *Journal of Botany, Wiley Online Library* 103:899–911. Botanical Society of America Inc.
- Mijangos, J. L., B. Gruber, O. Berry, C. Pacioni, and A. Georges. 2022. dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods Ecol Evol* 13:2150–2158.
- Mohanta, T. K., Y. K. Mohanta, P. Kaushik, and J. Kumar. 2023. Physiology, genomics, and evolutionary aspects of desert plants. *J Adv Res*, doi: 10.1016/J.JARE.2023.04.019. Elsevier.
- Mokany, K., Ware, C., Woolley, S. N., Ferrier, S., & Fitzpatrick, M. C. (2022). A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. *Global Ecology and Biogeography*, 31(4), 802-821.
- Naimi, B., N. a.s. Hamm, T. A. Groen, A. K. Skidmore, and A. G. Toxopeus. 2014. Where is positional uncertainty a problem for species distribution modelling. *Ecography* 37:191–203.
- Neher, R. A. 2013. Genetic Draft, Selective Interference, and Population Genetics of Rapid Adaptation. <https://doi.org/10.1146/annurev-ecolsys-110512-135920> 44:195–215. Annual Reviews.
- Novembre, J., and A. Di Rienzo. 2009. Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics* 2009 10:11 10:745–755. Nature Publishing Group.

- Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C. J. F. Ter Braak, and J. Weedon. 2022. *vegan: Community Ecology Package*.
- Ostria-Gallardo, E., E. Zúñiga-Contreras, D. E. Carvajal, T. C. de La Peña, E. Gianoli, and L. Bascuñán-Godoy. 2023. Two Congeneric Shrubs from the Atacama Desert Show Different Physiological Strategies That Improve Water Use Efficiency under a Simulated Heat Wave. *Plants* 12:2464. Multidisciplinary Digital Publishing Institute (MDPI).
- Peterson, B. G., and P. Carl. 2020. *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*.
- Potter, P. E., and P. Szatmari. 2009. Global Miocene tectonics and the modern world. *Earth Sci Rev* 96:279–295. Elsevier.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155:945–959. Oxford Academic.
- Qian, C., X. Yan, T. Fang, X. Yin, S. Zhou, X. Fan, Y. Chang, and X. F. Ma. 2021. Genomic Adaptive Evolution of Sand Rice (*Agriophyllum squarrosum*) and Its Implications for Desert Ecosystem Restoration. *Front Genet* 12. Frontiers Media S.A.
- Quan, C., S. Han, T. Utescher, C. Zhang, and Y. S. C. Liu. 2013. Validation of temperature–precipitation-based aridity index: Paleoclimatic implications. *Palaeogeogr Palaeoclimatol Palaeoecol* 386:86–95. Elsevier.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- Sandquist, D. R. 2014. Plants in deserts. *Ecology and the Environment* 297–326. Springer New York.
- Sansaloni, C., C. Petrolì, D. Jaccoud, J. Carling, F. Detering, D. Grattapaglia, and A. Kilian. 2011. Diversity Arrays Technology (DArT) and next-generation sequencing

combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings* 2011 5:7 5:1–2. BioMed Central.

Schluter, D. 2000. *The ecology of adaptive radiation*. OUP Oxford.

Schneider, C. J. 2000. Natural selection and speciation. *Proc Natl Acad Sci USA* 97:12398–12399. National Academy of Sciences.

Shen, Y., H. Xia, Z. Tu, Y. Zong, L. Yang, and H. Li. 2022. Genetic divergence and local adaptation of *Liriodendron* driven by heterogeneous environments. *Mol Ecol* 31:916–933. John Wiley and Sons Inc.

Shi, M. M., S. G. Michalski, X. Y. Chen, and W. Durka. 2011. Isolation by elevation: Genetic structure at neutral and putatively non-neutral loci in a dominant tree of subtropical forests, *castanopsis eyrei*. *PLoS One* 6.

Sork, V. L., S. N. Aitken, R. J. Dyer, A. J. Eckert, P. Legendre, and D. B. Neale. 2013. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genet Genomes* 9:901–911.

Steane, D. A., B. M. Potts, E. McLean, S. M. Prober, W. D. Stock, R. E. Vaillancourt, and M. Byrne. 2014. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Mol Ecol* 23:2500–2513. John Wiley & Sons, Ltd.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671–688. Blackwell Publishing Ltd.

Taiz, L., and E. Zeiger. 2007. *Fisiologia vegetal*. Universitat Jaume I.

Thioulouse, J., S. Dray, Anne-Béatrice Dufour, A. Siberchicot, T. Jombart, and Sandrine Pavoine. 2018. *Multivariate Analysis of Ecological Data with ade4*. Springer.

Thomas, S. C. 2011. Genetic vs. phenotypic responses of trees to altitude.

Thompson, J. N. 2016. Coevolution, local adaptation and ecological speciation. *Mol Ecol* 25:5608–5610. Blackwell Publishing Ltd.

Thompson, J., S. Ramírez-Barahona, N. Priest, and T. Hernández-Hernández. 2023. Did succulents diversify in response to aridity? Evolutionary analyses of major succulent

lineages around the world. bioRxiv 2023.05.23.541957. Cold Spring Harbor Laboratory.

Ullah, S., Q. You, D. A. Sachindra, M. Nowosad, W. Ullah, A. S. Bhatti, Z. Jin, and A. Ali. 2022. Spatiotemporal changes in global aridity in terms of multiple aridity indices: An assessment based on the CRU data. *Atmos Res* 268:105998. Elsevier.

United Nations Environment Programme, Nick. Middleton, and D. S. G. Thomas. 1997. *World Atlas of Desertification: Second Edition*. 182. Arnold.

Varas-Myrik, A., F. Sepúlveda-Espinoza, A. Fajardo, D. Alarcón, Ó. Toro-Núñez, E. Castro-Nallar, and R. Hasbún. 2022. Predicting climate change-related genetic offset for the endangered southern South American conifer *Araucaria araucana*. *Ecol Manage* 504:119856. Elsevier.

Wallén, C. C. 2017. Aridity Definitions and their Applicability. <http://dx.doi.org/10.1080/04353676.1967.11879765> 49:367–384. Taylor & Francis.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N Y)* 38:1358. Oxford University Press (OUP).

Whitford, W. G., and B. D. Duval. 2019. *Ecology of desert systems*. Academic Press.

Wiens, J. J., and C. H. Graham. 2005. Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. <https://doi.org/10.1146/annurev.ecolsys.36.102803.095431> 36:519–539. *Annual Reviews*.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of evolutionary biology*, 14(6), 851-865.

Yang, L., Y. He, S. Wang, Y. Zhang, Y. Wang, and Q. Liang. 2020. Population genetic divergence in *Lycium chinense* Mill. (Solanaceae) is driven by both hybridization and geo-environmental factors. *Genet Resour Crop Evol* 67:1505–1520. Springer.

Zhang, C., Y. Yang, D. Yang, and X. Wu. 2021. Multidimensional assessment of global dryland changes under future warming in climate projections. *J Hydrol (Amst)* 592:125618. Elsevier.

Zhu, L., W. Wen, M. R. Thorpe, C. H. Hocart, and X. Song. 2021. Combining heat stress with pre-existing drought exacerbated the effects on chlorophyll fluorescence rise kinetics in four contrasting plant species. *Int J Mol Sci* 22:10682. MDPI.

Zomer, R. J., J. Xu, and A. Trabucco. 2022. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data* 2022 9:1 9:1–15. Nature Publishing Group.