UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE CIENCIAS

SISTEMÁTICA

DIFERENCIACIÓN GENÉTICA Y AMBIENTAL EN EL COMPLEJO *ARREMON BRUNNEINUCHA* (AVES: PASSERELLIDAE) EN MESOAMÉRICA

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

M. en C. Israel Moreno Contreras

TUTOR:
Dr. Adolfo Gerardo Navarro Sigüenza/Facultad de Ciencias, UNAM

COMITÉ TUTOR:
Dr. Luis Antonio Sánchez González/ Facultad de Ciencias, UNAM
Dra. María del Coro Arizmendi Arriaga/Facultad de Estudios Superiores Iztacala UNAM

Ciudad Universitaria, CD. MX. **Mayo, 2024**

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE CIENCIAS

SISTEMÁTICA

DIFERENCIACIÓN GENÉTICA Y AMBIENTAL EN EL COMPLEJO *ARREMON BRUNNEINUCHA* (AVES: PASSERELLIDAE) EN MESOAMÉRICA

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

M. en C. Israel Moreno Contreras

TUTOR:
Dr. Adolfo Gerardo Navarro Sigüenza/Facultad de Ciencias, UNAM

COMITÉ TUTOR:
Dr. Luis Antonio Sánchez González/ Facultad de Ciencias, UNAM
Dra. María del Coro Arizmendi Arriaga/Facultad de Estudios Superiores Iztacala UNAM

Ciudad Universitaria, CD. MX. **Mayo, 2024**

COORDINACIÓN GENERAL DE ESTUDIOS DE POSGRADO
COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE CIENCIAS
OFICIO: CGEP/CPCB/FC/0271/2024

ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Comité de Posgrado en Ciencias Biológicas, celebrada el día **27 de noviembre de 2023** se aprobó el siguiente jurado para el examen de grado de **DOCTOR EN CIENCIAS** del estudiante **MORENO CONTRERAS ISRAEL** con número de cuenta **517007590** con la tesis titulada: **"Diferenciación genética y ambiental en el complejo *Arremon brunneinucha* (Aves: Passerellidae) en Mesoamérica",** realizada bajo la dirección del **DR. ADOLFO GERARDO NAVARRO SIGÜENZA:**

Presidente:	**DR. JULIÁN ANDRÉS VELASCO VINASCO**
Vocal:	**DR. HERNANDO ALONSO RODRÍGUEZ CORREA**
Vocal:	**DRA. TANIA GARRIDO GARDUÑO**
Vocal:	**DRA. CAROLINA GRANADOS MENDOZA**
Secretario:	**DR. LUIS ANTONIO SÁNCHEZ GONZÁLEZ**

Sin otro particular, me es grato enviarle un cordial saludo.

A T E N T A M E N T E
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 22 de marzo de 2024

COORDINADOR DEL PROGRAMA

DR. ARTURO CARLOS II BECERRA BRACHO

c. c. p. Expediente del alumno

ACBB/AAC/GEMF/EARR/ipp

Agradecimientos Institucionales

Al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México (UNAM) por bridarme la oportunidad y el espacio de formarme como Doctor en Ciencias, lo cual enriqueció mi crecimiento no sólo en el ámbito académico sino también en lo personal.

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por el otorgamiento de una beca de Doctorado (No. 749969) la cual me permitió terminar mis estudios en tiempo y forma.

Al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (UNAM DGAPA-PAPIIT), que proporcionó el apoyo económico necesario para la elaboración de este proyecto de investigación a través de los Proyectos IN214621 e IN215818 (responsable A. G. Navarro).

Al Departamento de Supercómputo de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (UNAM) por apoyo a través del proyecto LANCAD-UNAM-DGTIC-402 (responsables A. G. Navarro y A. Llanes-Quevedo), y al Departamento de Cómputo de Alto Rendimiento de la Facultad de Ciencias (UNAM).

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) de la UNAM, quién brindó oportunamente el recurso económico para la realización de dos estancias de investigación en los laboratorios del Dr. Scott Vernon Edwards (Harvard University) y Michael G. Harvey (University of Texas at El Paso). Esto representó para mí un acercamiento a las distintas formas en la que se hace investigación, y de esa forma mejorar mi red de colaboración en futuros proyectos académicos.

Al Comité Tutor, constituido por los Dres. Adolfo G. Navarro Sigüenza, Luis Antonio Sánchez González y María del Coro Arizmendi Arriaga, por todo el apoyo, sus enseñanzas, sugerencias y críticas que enriquecieron el contenido del presente proyecto de investigación.

Agradecimientos a título personal

A mi familia (Moreno-Contreras), mis amados padres Ramiro y Antonia, y a mis hermanos Missael y Valeria, por todo el apoyo económico y moral, la motivación, empeño, entusiasmo y confianza que han impregnado en mí para poder emprender mi sueño de realizar un posgrado en la UNAM. Sin su ayuda jamás habría podido emprender esta gran aventura.

A mi hija, Itayetzi Moreno ("Ita-Ita"), por todas las sonrisas y hermosos momentos que compartimos juntos mientras emprendía el viaje de realizar estudios de doctorado. Su compañía siempre ha sido (y es) un bálsamo para la vida. Te amo hija.

A Adolfo ("Mi Apá"), primero por aceptarme en ser su estudiante (maestría y doctorado), y segundo, por su amistad e infinita paciencia para guiarme en ser un buen investigador. Los cientos (posiblemente miles) de horas en las que compartió conmigo sus anécdotas y enseñanzas representaron un punto importante en mi vida, no sólo en lo académico, sino también en muchos espacios ajenos a la ciencia. Estoy eternamente agradecido contigo por ayudarme a cumplir mis sueños.

A mi Comité Tutoral, Dres. Luis Antonio Sánchez González, María del Coro Arizmendi Arriaga, y Adolfo Navarro, por instruirme durante todos mis estudios de posgrado, y de esa forma, realizar con éxito cualquier actividad académica relacionada a mi proyecto doctoral. Especialmente a Luis ("Howell") por explicarme detalladamente el proceso de egreso en el doctorado, y por las muchas veces que compartió conmigo su experiencia en la ciencia y en la vida misma. Te lo agradezco mucho, Howell.

A la familia de la Vega López, en especial a la señora Socorro y a sus hijos Joel e Irvin, por acogerme dentro de su entorno familiar, por hacerme sentir uno de ellos, y darme siempre todo su apoyo desde mi llegada a su casa en la Ciudad de México, especialmente en los momentos de mayor incertidumbre. Sin su apoyo, este trabajo no hubiera terminado de la mejor forma posible. Aprovecho también este apartado para agradecer a los señores Alejandro, Rosalba, Juan, Guillermo, Isabel y a su

pareja Pablo, por las incontables horas de risas y diversión que, junto con Doña Soco y Joel, hicieron más amena mi estancia en la ciudad. ¡Jamás los olvidaré!

A la familia Mondaca-Neder, en especial a Fernando y a su esposa Ruth por aceptarme en su casa en Delicias para realizar mis proyectos de investigación ajenos a mi tesis doctoral. En especial a Fernando, por años de colaboración científica, sus consejos en lo profesional y personal, y una amistad de toda la vida.

A la familia Escalante-Termignoni, por recibirme en su hogar, su amabilidad, y su paciencia para resolver mis dudas durante mi estancia en USA. En especial a la Dra. Flavia por compartir su valioso conocimiento en Genómica, Evolución y Ornitología, pero sobre todo, por la calidez humana con la que me compartió sus consejos y experiencia.

A todos los amigos del Edwards Lab en Harvard (Dr. Scott V. Edwards) y sus estudiantes, Nicole Mejia, Inbar Mayaan, Sophie Orzechowski, Jonathan Schmitt, Kelsie Lopez, Bohao Fang, Subir Shakya y sobre todo Flavia, por bridarme una de mis mejores experiencias académicas.

A mis amigos del Harvey Lab en UTEP (Dr. Michael G. Harvey), incluyendo a sus alumnos y personal académico, Oscar Johnson, Manuelita Sotelo, Emily Johns, Ximena Larrieu, Pablo Solis y Matt Brady, por acogerme en su laboratorio, y el tiempo invertido en el que compartieron su conocimiento y compañía. En especial a Mike y Oscar, que hicieron mi estancia en su laboratorio extremadamente útil en mi vida académica.

A mis colaboradores y amigos que durante todos mis años en el Doctorado me enseñaron a hacer ciencia bajo distintas perspectivas, Jukka Jokimäki (University of Lapland, Finlandia), Jukka Suhonen (University of Turku, Finlandia), Marja-Liisa Kaisanlahti-Jokimäki (University of Lapland, Finlandia), Lucas Leveau (Universidad de Buenas Aires, Argentina), Piotr Tryjanowski (Poznań University of Life Sciences, Polonia), Joseph A. Tobias (Imperial College London, Reino Unido), Douglas Lewis (USA), Fernando Mondaca Fernández (Universidad Autónoma de Chihuahua, México), Dra. Ana Bertha Gatica Colima (Universidad Autónoma de Ciudad Juárez,

México), Dra. Cuauhcihuatl Vital García (Universidad Autónoma de Ciudad Juárez, México), Dr. Marco Fabio Ortiz Ramírez (Universidad Nacional Autónoma de México, México), Dr. Erick Rodríguez (Instituto Tecnológico de Ciudad Victoria, México) y Dr. Héctor Gómez de Silva (Ciudad de México, México). También a los observadores de aves por su colaboración en el ámbito científico: Cristina Pérez, Isaac Miramontes, Manuel Jurado, William Howe, Daniel Pérez y Marisa Tirado.

A todos los integrantes que conforman (y han conformado) el Laboratorio de Ornitología del Museo de Zoología “Alfonso L. Herrera”: Howell, Gordillo, Erick, Blanca, Alberto, Tania, Vicente, Enrique (Tapatío), Jorge Ramírez (Pollo), David Prieto (Maracucho), Lynna, Angela, Alejandra, Pepe, Mayte, Alexandra, Paulina, Genaro, Orlando, Manuel (Manolo), Fernando, Alexander, Rodolfo, Ronald y Reinhard; por todo el apoyo personal y académico en las que numerosas veces tuve que solicitar su valioso tiempo durante mi tiempo en el laboratorio (“La Pajarera”). En particular, sin la ayuda de Rein, Rodo y Manolo, muchos de los análisis jamás hubieran terminado a tiempo. A todos ellos, la mejor de la suerte en sus futuros proyectos.

A la Dra. Lynna M. Kiere por su infinita paciencia en revisar y editar el inglés de todo el contenido escrito en la presente obra.

Al Dr. Leopoldo Vázquez, Manuel Grosselet y Juan Carmona por sus bellas fotos que lograron capturar el misticismo de los Rascadores (*Arremon*).

A la Lic. Erendira Ramírez por su gran amistad y las incontables veces que me asesoró en el ámbito legal para enfocarme directamente (y únicamente) a mis estudios de posgrado. ¡Muchas gracias, Lic!

Al equipo de fútbol del laboratorio (“MZFC”), por todos esos gratos momentos de triunfos que compartimos cada semana durante los torneos de fútbol, y por su empeño en dar su máximo esfuerzo.

A todo el plantel docente conformado por investigadores de alta calidad que me dieron clases durante mis estudios de posgrado en la UNAM: Dres. Alicia Mastretta, Luis Eguiarte, Ella Vázquez, Susana Magallón, Ivalú Camacho, Ricardo García,

Mark Olson, Hibrahim Pérez, Tania Escalante, Juan J. Morrone, David Prieto, Jorge Ramírez, Carlos Yáñez y Pablo Jaramillo. Todo lo aprendido es un reflejo de lo expuesto en esta tesis doctoral, muchas gracias.

A los Dres. Luis Antonio Sánchez González, Luis Eguiarte, Tania Garrido Garduño, Flor Rodríguez y Roberto Munguía Steyer, por sus valiosos comentarios y sugerencias durante mi Examen de Candidatura los cuales mejoraron sustancialmente el contenido de los proyectos científicos derivados en esta tesis.

Al personal del Posgrado en Ciencias Biológicas: Lilia Espinosa, Dr. Ernesto Armando Rodríguez, Erica Flores y Patricia Oliva; así como a los integrantes de la División de Posgrado de la Facultad de Ciencias, por instruirme de forma rápida, eficiente y amable en los tramites de inscripción (semestre a semestre) y egreso del programa de Doctorado.

A la M. en C. Laura Margarita Márquez Valdelamar del Laboratorio de Secuenciación Genómica de la Biodiversidad y la Salud del Instituto de Biología (UNAM), y María Eugenia Muñiz Díaz de León del Departamento de Biología Comparada de la Facultad de Ciencias (UNAM), por su incondicional apoyo en el trabajo de laboratorio.

A los miembros del honorable Jurado evaluador del presente proyecto de investigación, Dres. Luis Antonio Sánchez González, Carolina Granados Mendoza, Hernando Alonso Rodríguez Correa, Tania Garrido Garduño y Julián Andrés Velasco Vinasco, por sus enriquecedores comentarios y críticas constructivas que elevaron la calidad del presente escrito.

ÍNDICE

RESUMEN 1

ABSTRACT 3

INTRODUCCIÓN GENERAL 5

OBJETIVO GENERAL 13

HIPÓTESIS 14

CAPÍTULO I. Aplicaciones de los métodos de secuenciación masiva de nueva generación en estudios evolutivos 16

CAPÍTULO II. Influencia de las variables paisajísticas sobre la variación genómica en aves asociadas a los Bosques Montanos Neotropicales 61

CAPÍTULO III. Relaciones evolutivas dentro del complejo *Arremon brunneinuchalvirenticeps* (Aves: Passerellidae) con base en datos de secuenciación de nueva generación 159

DISCUSIÓN Y CONCLUSIONES GENERALES 210

REFERENCIAS BIBLIOGRÁFICAS 219

RESUMEN

Uno de los temas centrales en la biología evolutiva y la sistemática es comprender los patrones que explican la variación genética en las especies y sus relaciones de parentesco. Los taxa de los Bosques Montanos Neotropicales (BMN) han estado influenciados bajo modelos de contracción-expansión siguiendo un esquema dinámico de ciclos glaciares. Esto les ha conferido de una alta diferenciación genética y baja diversidad genética, o una baja diferenciación genética pero alta diversidad genética. Nosotros esperamos una alta estructura genética y filogenética para taxa con una baja vagilidad. En el primer capítulo realicé una revisión de la literatura sobre estudios que han empleado métodos de secuenciación masiva (NGS) en taxa del BMN. Un modelo que consistió en variables espaciales y temporales fue el mejor en comparación a aquellos modelos en los que se incluyó la economía o la cantidad de investigadores involucrados. El grupo taxonómico mejor representado fue Plantae. La Faja Volcánica Transmexicana fue la provincia biogeográfica que más ha sido estudiada. Existió una mayor prevalencia de ddRAD respecto a otros métodos (e.g. UCEs o RADseq). Pocos estudios han usado nuevos métodos bioinformáticos para el filtrado de loci previamente sugeridos para el post-procesamiento de datos genómicos. Además, la aplicación de métodos espaciales, métodos del tipo correlativos y técnicas filogenéticas basadas en la teoría coalescente fueron poco empleados. En la presente tesis se utilizó como caso de estudio a un grupo de aves asociados a los bosques montanos de Mesoamérica, los rascadores del género *Arremon*. En el capítulo dos se evalúan las distintas hipótesis de aislamiento que explican la variación genómica en este grupo de aves. Hay un patrón significativo de aislamiento por distancia, y la resistencia paisajística de las condiciones actuales y la topografía han reforzado su diversidad genómica. El capítulo tres se enfocó en evaluar las relaciones evolutivas dentro de este complejo, lo cual determinó que hay al menos seis especies. El sorteo incompleto de linajes ha contribuido a la obtención de topologías contrastantes. Se evidencia que los procesos históricos y contemporáneos han dado forma a la diferenciación genética y ambiental de taxa de los BMN. Además, se manifiesta que la diversidad biológica de este ecosistema se encuentra subestimada, debido a la multiplicidad

de procesos evolutivos y ecológicos que oscurecen los límites de especies en este complejo montañoso.

ABSTRACT

One of the main goals of evolutionary biology and systematics is understanding the underlying patterns driving genetic variation in species. The taxa inhabiting Neotropical Montane Forests (NMF) have undergone cycles of expansion and contraction due to glacial cycles. This has resulted in high genetic differentiation but low genetic diversity, or low genetic differentiation but high genetic diversity. Thus, we expect high genetic and phylogenetic structure for low-vagility taxa. In the first chapter, we conducted a systematic literature review of genomic studies applied to NMF taxa. We found that a model that consisted of spatial and temporal variables was the best compared to those models that included economy-based variables or the number of researchers involved. The best represented taxonomic group was Plantae. The Transmexican Volcanic Belt was the biogeographic province that has received most attention in genomic studies. There was a higher prevalence of ddRAD than other NGS techniques. Only a handful of previous genomic studies have incorporated new filtering methods during genome-wide data postprocessing. Likewise, the application of spatial methods, regression-based methods, and phylogenetic techniques based on coalescent theory are still underrepresented in genomic studies. In the next two chapters, we employed a group of birds associated with the NMF—the Brushfinches of the genus *Arremon* (Aves: Passerellidae)—as a study case to further explore the following ideas. Chapter two is aimed at evaluating the different isolation hypotheses that explain the genomic variation in these low-vagility birds. We determined that there is a significant pattern of isolation by distance. Furthermore, landscape resistance of current conditions and topography have reinforced their genomic diversity. Chapter three focused on evaluating the evolutionary relationships within this species complex. Phylogenomics suggest the presence of six species. Incomplete lineage sorting could have contributed to conflicting topologies in earlier studies. The research presented in this thesis provides solid evidence of the historical and contemporary processes that have contributed to genomic and environmental differentiation in NMF taxa. It is shown that the biological diversity of this ecosystem is underestimated, due to multiple

evolutionary and ecological processes, largely obscuring the species limits inhabiting this archipelago-like mountain complex.

INTRODUCCIÓN GENERAL

El estudio de la variación biológica es el enfoque central en la biología evolutiva (Ghiselin 1997; Coyne y Orr 2004; Wiley y Lieberman 2011). Para comprender los patrones que subyacen la variación biológica (divergencia genética y ambiental), varios enfoques a distintas escalas espaciales y temporales han sido propuestos. Uno de esos enfoques es la genética del paisaje, el cual busca determinar las variables ambientales y geográficas que conducen a los distintos patrones de variación genómica y ambiental entre las poblaciones, usualmente (pero no exclusivamente) a escalas espaciales pequeñas (Manel y Holderegger 2013). Por su parte, la filogeografía infiere los patrones geográficos y demográficos que exhiben grupos genéticos cercanamente relacionados (Avise 2009), y su escala espacial de trabajo involucra áreas geográficas de mayor extensión en comparación al enfoque espacial de la genética del paisaje (Wang 2010). Una vez comprendidos los patrones observados de divergencia genética en las poblaciones naturales, uno de sus fines es su aplicación en la conservación biológica (Scoble y Lowe 2010; Perez *et al.* 2018). Particularmente, una de las zonas que amerita una necesidad urgente para el estudio y la comprensión de ambos factores (históricos y actuales) que gobiernan la variación genómica en las especies es el Neotrópico (Dantas-Queiroz *et al.* 2023).

Dicho esto, el enfoque de la presente tesis se inclina hacia un área geográfica que se extiende desde México hasta América del Sur (Morrone 2014; Morrone *et al.* 2022), los Bosques Montanos Neotropicales (BMN, o “NMF” por sus siglas en inglés; Churchill *et al.* 1995). Los NMF abarcan un gradiente latitudinal que se extiende desde 30°N hasta 29°S en América (Fig. 1 del **Capítulo 1**) y se componen de bosques templados de elevación moderada a alta (pino, pino-encino) en regiones secas, así como bosques mesófilos (Hernández-Baños *et al.* 1995; Morrone *et al.* 2022). El arreglo de la distribución espacial de los NMF se asemeja a la de un sistema insular, donde zonas de vegetación a cierta altitud (usualmente arriba de los 1500 m sobre el nivel del mar) se encuentran separadas por extensos valles o planicies en las tierras bajas (Gentry 1995), siendo que, en las latitudes más

meridionales, los NMF configuran una distribución menos discontinua (Flantua y Hooghiemstra 2018). Esto ha dado lugar a la presencia de una gran cantidad de taxa endémicos con distribuciones significativamente restringidas, y una alta riqueza específica, que resaltan su importancia biogeográfica y biológica (Churchill *et al.* 1995; Gentry 1995). Además, la biota del NMF se encuentra entre los mayores niveles de importancia para su conservación a nivel global (Myers *et al.* 2000), debido a que esta zona biogeográfica sufre altas tasas de extinción proporcionado por el constante cambio de uso de suelo y cambio climático, y consecuentemente afectando uns gran cantidad de especies endémicas (Foster 2001; Rojas-Soto *et al.* 2012; Palacio *et al.* 2020).

Este complejo biogeográfico (es decir, BMN) ha sido intensamente estudiado y se tiene propuesto que corresponde a una mezcla de elementos bióticos de distintos orígenes (Churchill e*t al.* 1995; Sánchez-González *et al.* 2008; Morrone 2015). Por ejemplo, hay evidencia que respalda la idea de que algunos ensamblajes de especies Neotropicales habitando los BMN están compuestas por grupos (por ejemplo, aves; Buainain *et al.* 2022) que llegaron después del surgimiento del Istmo de Panamá (Coates y Obando 1996; ver referencias en Mastretta-Yanes *et al.* 2015). Esto a su vez también permitió que otros grupos biológicos provenientes de Centroamérica durante el Oligoceno mostraran esa misma tendencia de llegar después del surgimiento del Istmo de Panamá (Morrone 2010; Mastretta-Yanes *et al.* 2015). Empero, es una hipótesis en la que no se ha llegado a un consenso, dado que algunos estudios ubican el cierre del Istmo de Panamá en una época anterior (desde el Eoceno tardío hasta el Mioceno tardío; Montes *et al.* 2012). Por su parte, algunas especies montanas (por ejemplo, carábidos del género *Platynus*) de México con afinidades Neárticas podrían ser el producto de migraciones hacia el sur desde Estados Unidos (Marshall y Liebherr 2000) debido a oscilaciones climáticas en el Mioceno medio (Graham 1999). A pesar de las distintas hipótesis, la idea principal es que muchas de esas especies Neárticas (por ejemplo, plantas de los géneros *Pinus*, *Abies* y *Quercus*) provinieron de eventos de migración significativamente antiguos (Graham 1999; Morrone 2010). Por el contrario, algunas coníferas de la familia Podocarpaceae (*Podocarpus matudae*; Ornelas *et al.* 2010), varios linajes

de reptiles (por ejemplo, *Bothrops*, *Crotalus*; Castoe *et al.* 2009; Dantas *et al*. 2010) y artrópodos del orden Orthoptera (género *Stenopelmatus*; Gutiérrez-Rodríguez *et al.* 2022) de orígenes Neotropicales parecen haber divergido durante eventos pre-Pleistocénicos (Mioceno: 10–20 Ma) o incluso mucho antes (entre el Oligoceno tardío y el Paleoceno tardío: 56–28 Ma), respectivamente, precediendo la formación del Istmo de Panamá en el Mioceno tardío (6–7 Ma; ver Coates y Obando 1996 y referencias en Montes *et al.* 2012). No obstante, estimaciones recientes colocan un cierre del Istmo de Panamá en el Mioceno (Stange *et al.* 2018).

Como sugiere lo arriba mencionado, movimientos tectónicos, oscilaciones climáticas y aspectos biológicos, como la demografía histórica de las especies, han dado origen a una gran diversidad (y disparidad) de trayectorias filogeográficas, algunas compartidas entre taxa, y otras siendo especie-específicas (Ramírez-Barahona y Eguiarte 2013; Ornelas y González 2014, Mastretta-Yanes *et al.* 2015; Mastretta-Yanes *et al.* 2018; Flantua *et al*. 2019). Bajo esa perspectiva, varias hipótesis filogeográficas han sido propuestas para explicar la variación genética de las poblaciones naturales de los NMF. Dos de esas hipótesis, llamadas los “refugios secos” (“*dry refugia*”) y “bosques húmedos” (“*moist forests*”), representan un continuo de dos estados opuestos (contracción y expansión) que describen las consecuencias genéticas en taxa distribuidos en los NMF (Ramírez-Barahona y Eguiarte 2013). La primera hipótesis explica que la distribución de los bosques montanos exhibió una contracción (una disminución de la distribución geográfica) debido a los efectos combinados del enfriamiento y aridez durante los ciclos glaciares (Ramírez-Barahona y Eguiarte 2013). Dada las condiciones climáticas adecuadas en términos de humedad y temperaturas cálidas, las especies realizarían movimientos de dispersión relativamente largos en las tierras bajas durante ciclos interglaciares. Todo esto conllevaría a una pérdida de diversidad genética (dada los tamaños poblacionales pequeños), pero incrementaría la estructura poblacional ocasionada por la fijación de alelos de diferentes alelos en distintos refugios, dando como resultado monofilia recíproca entre grupos genéticos cercanos (Ramírez-Barahona y Eguiarte 2013). Por el contrario, el segundo modelo evoca un efecto en donde las condiciones de precipitación persistieron en los ciclos

glaciares facilitando la migración entre poblaciones en las tierras bajas. Bajo ese modelo, se esperaría que las especies realizaron migraciones hacia las tierras altas en los ciclos interglaciares. Tales movimientos dan como consecuencia una alta diversidad genética debido a los tamaños poblacionales altos, pero una baja estructura poblacional debido a la presencia de flujo genético (Ramírez-Barahona y Eguiarte 2013).

Si bien, estas hipótesis filogeográficas explican la variación genómica de algunas especies montanas (véase Hernández-Langford *et al.* 2020), no todos los taxa exhiben las expectaciones genéticas ahí propuestas. Por ejemplo, plantas herbáceas (*Moussonia deppeana*, Gesneriaceae) distribuidas en el norte de Mesoamérica no se alinearon bajo las dos hipótesis filogeográficas ("refugios secos" *vs.* "bosques húmedos" de Ramírez-Barahona y Eguarte 2013). Por el contrario, Ornelas y González (2014) encontraron que las poblaciones de esta especie de planta permanecieron aisladas en el último máximo glaciar (LGM, por sus siglas en inglés), referente a su distribución geográfica conocida, sin evidencia de alguna conexión en las tierras bajas. Aunado a eso, se encontró una estructura genética significativa y baja diversidad genética, sin señal de expansión demográfica durante el LGM (Ornelas y González 2014). Entonces, todos estos hallazgos condujeron a la creación de una nueva hipótesis, la "divergencia alopátrica in *situ*" (Ornelas y González 2014). Dicha hipótesis parece ser reforzada con lo determinado para algunas plantas de alta longevidad del género *Juniperus* involucrando poblaciones sólo de una provincia biogeográfica (Mastretta-Yanes *et al.* 2018). Así mismo, se ha dictaminado que la hipótesis de los dos estados podría simplificar la historia filogeográfica mostrada en taxa montanos de Sur América (Flantua y Hooghiemstra 2018).

Independientemente de la hipótesis filogeográfica mejor respaldada en la literatura, todas ellas se encuentran contextualizadas bajo el uso de métodos de secuenciación y herramientas estadísticas tradicionales comúnmente usadas en filogeografía y genética del paisaje. Principalmente, la mayoría de los estudios han hecho uso de la secuenciación Sanger (secuencias mitocondriales, nucleares, y/o

de cloroplasto) para explicar la filogeografía de taxa montanos (ver Ramírez-Barahona y Eguiarte 2013; Flantua y Hooghiemstra 2018; Flantua *et al*. 2019), aunque algunas excepciones usando técnicas moleculares modernas empiezan a hacer presencia en la literatura (Mastretta-Yanes *et al.* 2019; Ortego *et al.* 2023). El uso de herramientas moleculares basadas en secuencias de un sólo gen, o concatenados de varios genes para la elaboración de hipótesis evolutivas ha sido criticado (Edwards y Bensch 2009; Edwards *et al.* 2015). Específicamente, el uso de un solo gen podría estar explicando la historia evolutiva de ese mismo marcador empleado, y no de un conjunto de genes que podrían representar mejor la historia evolutiva de un grupo de especies cercanamente relacionadas (Edwards *et al.* 2015). Algunos factores como la discordancia mitonuclear (y en plantas en los tres compartimentos genómicos [mitocondrial, nuclear, y de cloroplasto]), hibridación, sorteo incompleto de linajes, el aislamiento por distancia, y la estructura poblacional anidada, actúan de forma conjunta obscureciendo el proceso detrás de esa señal filogenética que da origen a un patrón filogeográfico compartido a través de escalas taxonómicas (Maddison 1997; Edwards 2009; Edwards y Bensch 2009; Toews y Brelsford 2012).

Con base en estos retos y lagunas de conocimiento que permanecen en la búsqueda de patrones evolutivos que tratan de dilucidar la variación genómica de los taxa asociados a los NMF, en la presente tesis se evalúan algunas hipótesis evolutivas usando técnicas de secuenciación de nueva generación (NextRAD; Russello *et al.* 2015). Esto nos permitió adentrarnos en distintas perspectivas espaciales y temporales en la búsqueda de patrones evolutivos tomando como marco de estudio las dos hipótesis basadas en regímenes de precipitación contrastantes (Ramírez-Barahona y Eguiarte 2013).

Para ello, la presente tesis doctoral se compone de los siguientes capítulos. En el **capítulo I** se hace una revisión crítica de la literatura sobre el uso de métodos modernos de secuenciación masiva basados en subrepresentación genómica, y herramientas estadísticas y analíticas que se suelen emplear para su estudio. Específicamente, la revisión de literatura hace énfasis sobre taxa montanos del

Neotrópico (acorde al mapa de Chuchill *et al.* 1995; ver también Fig. 1 del **capítulo 1**), y como varias áreas de incertidumbre y sesgos de publicación dificultan la reconstrucción de la historia evolutiva bajo distintos enfoques (por ejemplo, genómica de poblaciones, filogeografía). Bajo esta premisa, se emplea un marco de trabajo que involucró varias líneas de investigación, las cuales son principalmente técnicas bibliométricas, bioinformáticas, geográficas y estadística basada en la teoría robusta. Esto con el fin de determinar los sesgos de publicación referentes al uso de métodos modernos de secuenciación que han empleado los investigadores para estudiar determinados procesos geográficos y evolutivos en taxa de los NMF.

En los subsecuentes **capítulos** (**II** y **III**) empleamos como modelo de estudio un grupo de rascadores del género *Arremon brunneinucha*/*virenticeps* (Aves, Passerellidae). En un principio, se consideraban que todas las poblaciones de plumaje rojizo (grupo *A. brunneinucha*) eran recíprocamente monofiléticos respecto a las poblaciones de plumaje verde (grupo *A. virenticeps*) (Paynter 1978). De hecho, se pensaba que *A. virenticeps* (distribuido alopátricamente en la Faja Volcánica Transmexicana y Sierra Madre Occidental, ambos en México) estaba evolutivamente cercano a otra especie con plumaje verde (el Cerquero Vientre Blanco, *A. torquatus*) el cual tiene una distribución geográfica ubicada en Sur América (Paynter 1978). Aunque tradicionalmente descrito como dos especies diferentes (Hellmayr 1938; Parkes 1954; 1957; Paynter 1978), análisis moleculares basados en aloenzimas (Peterson *et al.* 1992) y en loci mitocondriales y/o nucleares parecen contradecir dicha premisa (Cadena *et al.* 2007; Navarro-Sigüenza *et al.* 2008; Flórez-Rodríguez *et al.* 2011; Navarro-Sigüenza *et al.* 2013). Todas las hipótesis evolutivas de parentesco han arrojado topologías contrastantes para este complejo de especies, siendo las causantes de este patrón la presencia significativa del sorteo incompleto de linajes y ramas internas cortas (ver Flórez-Rodríguez *et al.* 2011).

Uno de los patrones más interesantes que se han obtenido gracias a los análisis de Sistemática Molecular (Cadena *et al.* 2007; Navarro-Sigüenza *et al.* 2008; Flórez-Rodríguez *et al.* 2011; Navarro-Sigüenza *et al.* 2013) es que el grupo

virenticeps (Rascador Cejas Verdes) no es recíprocamente monofilético respecto a todos los linajes de *brunneinucha* (Rascador Gorra Castaña) en la Zona de Transición Mexicana (terminología siguiendo a Morrone *et al.* 2022). Estos hallazgos han sugerido que el establecimiento de hipótesis evolutivas basados en caracteres morfológicos (plumaje) pueden oscurecer las relaciones de parentesco incluso en grupos con patrones significativos de divergencia genética y tiempos de aislamiento relativamente largos. Consecuentemente, esto da indicios de la existencia de factores paisajísticos o geográficos complejos que han dado forma a los patrones de divergencia genética y evolutiva dentro de este grupo de aves y otras especies. Posiblemente, sus rasgos de historia de vida (poca vagilidad, hábitos relacionados al forrajeo en el suelo, poca capacidad de dispersión, distribuciones restringidas) también hayan contribuido a la diferenciación genética observada.

La distribución geográfica que presenta este grupo de aves también ofrece un interesante punto de partida para comprender la historia evolutiva de los BMN. Puntualmente, cada linaje se encuentra fuertemente asociado a una determinada provincia biogeográfica (Sierra Madre Oriental [*brunneinucha*], Sierra Madre Occidental [*virenticeps*], por mencionar algunos miembros), o al menos dos miembros dentro de una provincia (Sierra Madre del Sur: Guerrero [*kuehnerii*[y Oaxaca [*suttoni*]) de la Zona de Transición Mexicana (Howell y Webb 1995). A diferencia de las zonas montañosas de Sur América, estas provincias del área de estudio se encuentran más aisladas geográficamente.

Siguiendo el espíritu de cuantificar señales del acervo evolutivo compartido por taxa Neotropicales en regiones montañosas, el **capítulo II** funge como artículo de requisito para el programa de Doctorado en Ciencias Biológicas. En esa sección (publicado en ***Molecular Ecology***) se evaluaron distintas hipótesis de aislamiento genético (Wright 1943; McRae 2006; Wang y Bradburd 2014; Balkenhol *et al.* 2017) que pueden potencialmente explicar los patrones de variación genómica en el complejo *Arremon brunneinucha*/*virenticeps*. El **capítulo III** consistió en dilucidar las relaciones evolutivas de este complejo de especies de Mesoamérica, por medio de herramientas de genómica del paisaje, genómica de poblaciones y reconstrucción

filogenética. El aislamiento por distancia, la estructura genética y el sorteo incompleto de linajes han contribuido en la resolución de los resultados filogenómicos obtenidos en el presente estudio.

OBJETIVO GENERAL

Evaluar los patrones de divergencia genómica, filogenómica y ambiental que han dado forma e identidad a especies con rasgos de historia natural relacionados a la baja vagilidad y con distribuciones restringidas, usando como caso de estudio a los Rascadores del género *Arremon* (Aves: Passerellidae) de la Zona de Transición Mexicana.

Objetivos particulares:

1) Analizar los patrones espaciales y temporales que influyen en el número de estudios que han evaluado los procesos microevolutivos usando métodos de secuenciación de nueva generación sobre taxa asociados a los Bosques Montanos Neotropicales.

2) Evaluar las variables paisajísticas y biogeográficas que han dado forma a los patrones de variación y estructura genómica en el grupo de Rascadores (*Arremon brunneinucha*/*virenticeps*) asociados a sitios montañosos de Mesoamérica.

3) Determinar las relaciones evolutivas de los Rascadores que habitan la Zona de Transición Mexicana (*Arremon brunneinucha*/*virenticeps*) con base en datos provenientes de secuenciación de nueva generación (NextRAD).

HIPÓTESIS

Debido a las condiciones de aislamiento en los Bosques Montanos Neotropicales impuesta por una dinámica glaciar-interglaciar (contracción y expansión), se espera una divergencia genómica, filogenómica y ambiental significativamente alta dentro de un complejo de especies de aves (*Arremon brunneinucha*/*virenticeps*), relacionada al modelo de los refugios secos.

Lo anterior permite la formulación de las siguientes hipótesis principales, mismas que serán puestas a prueba en el presente proyecto de tesis doctoral en forma de **capítulos** (**I**, **II** y **III**):

1) Dado los sesgos históricos de publicación sobre los efectos de retardo en la implementación de métodos nuevos en la ciencia, esperamos que un modelo basado aspectos espaciales y temporales tenga mayor explicación estadística referente a la distancia que hay del sitio de origen del método de secuenciación (considerando todas las afiliaciones de los autores publicando esa técnica) y el área estudiada (los centroides de las provincias biogeográfica estudiadas). Es decir, transcurrido el tiempo y a mayor distancia del sitio de publicación de los métodos de secuenciación masiva, mayor será la distancia (el alcance de su implementación) de estas técnicas en los Bosques Montanos Neotropicales.

2) Debido a la limitada capacidad de dispersión de este grupo de aves y la divergencia de nicho hacia gradientes de elevación altos en poblaciones alopátricas, las condiciones paisajísticas relacionadas a la complejidad topográfica (aislamiento por resistencia) tendrán un mayor efecto en la variación genómica de este grupo de aves en comparación a otros modelos de aislamiento genético como el geográfico (aislamiento por distancia), el impuesto por la presencia (o ausencia) de barreras vicariantes (aislamiento por barrera) o por gradientes ecológicos específicos (aislamiento por ambiente basado en distancias euclidianas).

3) Si la generación de pocos marcadores genéticos (mitocondriales, nucleares) obtenidos con métodos de secuenciación tradicional (por ejemplo Sanger) presentan poco soporte estadístico para establecer las relaciones evolutivas de este complejo, entonces técnicas moleculares novedosas como las obtenidas por medio de subrepresentación genómica masiva asociadas a sitios de restricción (NextRAD) permitirán establecer relaciones evolutivas estadísticamente robustas e independientes de los métodos de reconstrucción empleados.

CAPÍTULO I. Aplicaciones de los métodos de secuenciación masiva de nueva generación en estudios evolutivos

CAPÍTULO I. Aplicaciones de los métodos de secuenciación masiva de nueva generación en estudios evolutivos

Este capítulo será enviado a revisión como artículo científico en **Biology & Philosophy** (ISI WEB Impact factor: **2.5**) bajo el título "**Advances in the study of genomic variation in taxa associated with Neotropical Montane Forests**"

Arremon b. suttoni en Oaxaca. Fotografía por: Manuel O. Grosselet.

Advances in the study of genomic variation in taxa associated with Neotropical Montane Forests

Israel Moreno-Contreras[1,2]

[1]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1º Piso, Circuito de Posgrados, Ciudad Universitaria, Universidad Nacional Autónoma de México, Coyoacán, Mexico City 04510, Mexico.

[2]Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, A.P. 70–399, Mexico City 04510, Mexico.

E-mail address: israel.moreno.contreras@gmail.com

ORCID: https://orcid.org/0000-0001-6000-7587

Declarations

Conflict of interest There are no relevant conflicts to disclose.

Acknowledgements

This paper partially fulfills the requirements to obtain a Doctoral degree in the Posgrado en Ciencias Biológicas, UNAM, for which the author received a Doctoral scholarship provided by Consejo Nacional de Ciencia y Tecnología (CONACyT, number 749969). I acknowledge the helpful comments and criticisms from C. Granados, H. Rodríguez-Correa, T. Garrido, L. A. Sánchez-González, and J. Velasco. L M. Kiere edited and reviewed the manuscript. Finally, I appreciate the support provided by the Moreno-Contreras family during the elaboration of this work.

Abstract

Several of the hypotheses that attempt to explain genomic variation and distribution of Neotropical montane taxa have so far been tested only with traditional sequencing methods (i.e. Sanger). However, the advent of next-generation sequencing (NGS) technologies has made it necessary to incorporate novel methods from different disciplines to understand speciation processes and phylogeography. NGS has been accompanied by parallel advances in the statistical approaches used to assess a wide variety of evolutionary hypotheses. Yet both isolation by distance and population genetic structure promote genetic-phenotypic divergence, and both must be analyzed during species delimitation studies or formulation of phylogeographical hypotheses. In this review, I present the importance of applying landscape genomics/population genomics tools to assess areas of uncertainty when studying phylogenomic reconstructions in species complexes or testing phylogeographical hypotheses. The objective of this review is to quantify the spatial and temporal constraints limiting the implementation of novel NGS methods and explore their utility, both to date and into the future, for formulating phylogeographic hypotheses about Neotropical taxa. Besides, I outline and summarize the theoretical basis of what I consider should become part of the basic toolbox for evolutionary biologists interested in elucidating rapid speciation processes in changing landscapes. This paper focuses mainly (but not exclusively) on taxa associated with Neotropical Montane Forests (NMF), since significant phylogenetic and genetic structure is expected for montane taxa.

Keywords

glacial-interglacial dynamics, landscape genomics, maximum likelihood population-effects, Neotropical Montane Forests, next-generation sequencing, phylogenomics

Introduction

Since the time of Charles Darwin's Origin of Species, there has been a growing interest in understanding patterns of biological variation (Ghiselin 1997). Although nature is by default continuous, biologists discretize nature in order to understand it (De Queiroz 2007). In biological systematics, the basis for the identification and delimitation of species is morphological, genomic, and ecological variation (Wägele 2005; Wiley and Lieberman 2011). However, closely related sympatric, parapatric (or nearly allopatric) lineages often exhibit subtle genomic differentiation, which leads to the formation of species complexes (Fišer et al. 2018; Giles-Pérez et al. 2022). Genomic variation (both differentiation and diversity) is the output of divergent processes that have promoted the historical isolation and/or secondary contact of the species and their populations (Coyne and Orr 2004). This presents an opportunity in the study of phylogeography, since the establishment of (phylogeographic) hypotheses (which involve one or several biogeographic provinces) is based on a comprehensive understanding of genomic variation and distribution of the analyzed species (see Ramírez-Barahona and Eguiarte 2013).

The Neotropical Montane Forests (NMF; Churchill et al. 1995) are a biogeographic complex that offers a comparative framework for the study of genomic variation and phylogeography. The NMF encompass a latitudinal gradient that extends from 30 °N to 29 °S in the Americas (Fig. 1). NMF are composed of moderate- to high-elevation temperate forests (pine, pine-oak) and cloud forests. The northern extreme of their distribution is in the Mexican Transition zone (Morrone et al. 2022), passing through isolated highlands in Mexico (e.g., Sierra de Los Tuxtlas, Veracruz) and Central America (e.g., Talamancan montane forests); as they progress southward, NMF are less patchily distributed until reaching their southern extreme in the Southern Andean Yungas (Bolivia and Argentina). NMF are characterized by their archipelago-like spatial distribution, the presence of restricted-habitat and endemic taxa, and their high taxonomic diversity, which highlight their biogeographic and biological importance (Churchill et al. 1995; Gentry 1995). Besides, they have been

extensively characterized ecologically and geographically (Sánchez-González et al. 2008; Sánchez-González and Navarro-Sigüenza 2009; Morrone 2014).

NMF have generally presented a glacial-interglacial dynamic through Quaternary oscillations and previous periods (see Ramírez-Barahona and Eguiarte 2013; Flantua and Hooghiemstra 2018). This has led to repeated contraction-expansion cycles, which have facilitated many microevolutionary processes (e.g., gene flow, demographic shifts, natural selection) that have permeated the genomic variation of natural populations in several fashions (Ramírez-Barahona and Eguiarte 2013; Ornelas and González 2014; Dantas-Queiroz et al. 2023). For instance, in some cases, the combined effect of warm and humid conditions has allowed gene flow among highland populations, leading to high genetic diversity (see some phylogeographical hypothesis in Ramírez-Barahona and Eguiarte 2013). At the same time, the accumulation of different historical (e.g., tectonic, climatological, and geographical) events have given rise to the island-like spatial arrangement of NMF patches (Ornelas et al. 2013). Other ecological aspects related to landscape barriers (Mastretta-Yanes et al. 2015; Mastretta-Yanes et al. 2018) and Grinnellian niche tracking toward to a specific subset of environmental conditions (Moreno-Contreras et al. 2020; Linck et al. 2021) seem to have determined distributional patterns of montane taxa. Thus, the joint effect of historical and contemporary factors has contributed to a complex pattern of speciation for many lineages associated with NMF (McCormack et al. 2008; Mastretta-Yanes et al. 2018; Hernández-Langford et al. 2020). This complexity is especially evident when evaluating different species-specific phylogeographic trajectories, since allopatric taxa usually exhibit marked genetic structure, high genomic differentiation, and reciprocal monophyly (e.g., León-Paniagua et al. 2007; Rocha-Méndez et al. 2019; but see Cortés-Rodríguez et al. 2008). It is important to note that much of that body of knowledge has been built using molecular data obtained through Sanger sequencing (Ramírez-Barahona and Eguiarte 2013; Ornelas et al. 2013; Ornelas and González 2014; Flantua et al. 2019; Dantas-Queiroz et al. 2023), with a few exceptions (Mastretta-Yanes et al. 2018; Ortego et al. 2023).

The use of traditional sequencing methods (e.g., Sanger) of different molecular-marker types (mitochondrial, nuclear, or chloroplast) have been fruitful, generating a large body of knowledge (Suárez-Díaz and Anaya-Muñoz 2008). However, interpretations based on these techniques and markers are strongly limited (Teske et al. 2018; Harvey et al. 2019; Mason et al. 2020) by genetic factors such as incomplete lineage sorting (hereafter ILS, Edwards and Bensch 2009), and the joint effects of isolation by distance (hereafter IBD; Wright 1943) and population genetic structure (hereafter PGS; Wright 1951; Meirmans et al. 2012; Rellstab et al. 2015). These factors are especially prevalent in species that have remained isolated or have come into secondary contact, as expected for taxa inhabiting NMF (Mastretta-Yanes et al. 2015; Mastretta-Yanes et al. 2018; Flantua et al. 2019). Therefore, conflicting topologies are the norm in these cases, especially in species with low vagility traits or a historical pattern of restricted isolation (e.g., *Arremon* brushfinches, Navarro-Sigüenza et al. 2008).

So-called "next-generation sequencing" (NGS) is an alternative to traditional sequencing methods (e.g., Baird et al. 2008; Elshire et al. 2011; Faircloth et al. 2012) that provides several advantages over previous methods. NGS allows the sequencing of numerous loci across the genome; this greatly increases statistical power, provides more detailed information on the evolutionary history of a species, and of the analysis of the generated genetic diversity data can be more robust than that obtained with traditional methods (Lemopoulos et al. 2019). The use of NGS data requires several molecular biology and bioinformatic precautions to ensure that they are correctly applied to studying organismal evolution (Joost et al. 2013; Puritz et al. 2014; Edwards et al. 2015; Linck and Battey 2019). For example, it is crucial that investigators are careful to avoid including paralogous loci or large amounts of missing data in the set of genomic data, as these can lead to biased inferences when assigning genetic groups (Chattopadhyay et al. 2014). With these caveats in mind, NGS methods are making major inroads into Neotropical evolutionary studies. However, in phylogeographic and landscape genomics applications (Mastretta-Yanes et al. 2018) they have continued to perpetuate several of the problems of traditional sequencing methods, such as issues surrounding IBD and PGS (Ramírez-

Barahona and Eguiarte 2013). In addition, the incorporation of species tree (hereafter ST) methods in evolutionary studies is still lacking (Liu et al. 2010; Bryant et al. 2012; Mirarab et al. 2016; Zhang et al. 2018), as most phylogeographic inferences are still based on concatenated mtDNA or cpDNA sequences. Ultimately, such challenges and knowledge gaps in the "-omics era" persist as areas of uncertainty when characterizing the phylogeographic patterns in NMF species in general.

The objective of this review is to quantify the spatial and temporal patterns of the implementation of novel NGS methods and explore their utility, both to date and into the future, for formulating phylogeographic hypotheses about Neotropical taxa. The aim of this paper is also to provide advices on "best methodological practices" in genomic studies aimed at biodiversity conservation of NMF. I will not evaluate whether the current body of literature supports the idea of a shared phylogeographic pattern for NMF taxa (which has been addressed repeatedly elsewhere and is almost exclusively based on pre-NGS techniques; Ramírez-Barahona and Eguiarte 2013). Rather, I will address this question whether the use of NGS to have been limited by the spatial and temporal constraints of the permeation of these techniques into the study of NMF taxa. In the literature there are proposals for the application of novel NGS methods (Baird et al. 2008; Faircloth et al. 2012; Russello et al. 2015) or statistical tools with greater explanatory power (e.g., Legendre et al. 1994; Clarke et al. 2002; Liu et al. 2011; Bradburd et al. 2013) aimed to formulate robust phylogeographic hypotheses. However, it is quantitatively unknown to what extent researchers have applied these novel tools in microevolutionary studies. Furthermore, these techniques may not necessarily be adopted homogeneously. In spatial terms, scientific revolutions and methods are more likely to be applied sooner in regions that are in close proximity to the institutions where they were first published (Kuhn 1962). In temporal terms, the gradual permeation of new techniques tends to lead to the trend that the longer ago the method or analytical tool was proposed, the more literature has been accumulated utilizing it (Kuhn 1962; Laudan 1977).

To explore these trends, I tested three hypotheses. First (*H1*), I tested whether a model of spatial (the distance between the site of origin of an NGS and the affiliation of the research group leading the studies focused on NMF taxa) and temporal predictors (difference between the year of publication of a genomics study and the year of publication of the implemented NGS in that study) influence on the distance between the origin site of each NGS method (considering all author affiliations for each study publishing a novel sequencing method) and the site (centroids of the studied biogeographic provinces) where this sequencing method is utilized. Given historic global biases in publication, I predicted that this would result in positive relationships between predictors (spatial and temporal) and explanatory variables. In the second hypothesis (*H2*), I tested whether earlier NGS methods (RADseq; Baird et al. 2008) proposed for the study of biological variation have significantly higher representation in the scientific literature than more recently proposed methods (UCEs; Faircloth et al. 2012). Finally, for the third hypothesis (*H3*), I tested whether relatively recent analytical methods (species trees [ST methods such as MP-EST, ASTRAL]) and landscape genomics/population genomics tools (e.g., spatial assignment groups, novel regression methods) have been adopted in studies focused on taxa associated with the NMF. I predicted that the implementation of all these analytic methods is underrepresented in the literature.

As a study model, I used the specialized literature focused on studying the genomic variation of the Neotropical highland species (Churchill et al. 1995; Morrone et al. 2022). I restricted the scope of this review to evolutionary studies using reduced representation genome-sequencing technologies (RADseq, ddRAD, NextRAD, 3RAD, GBS, TEC, and UCEs), whose affordability has led to relatively widespread feasibility and presence in the literature, and excluded techniques that are generally cost-prohibitive and have therefore been implemented only rarely (e.g., whole-genome sequencing, see Nater et al. 2015). To address these hypotheses, I used an integrative approach involving bibliometric techniques, bioinformatics pipelines, and robust modeling. Finally, I hope that the knowledge gaps found in this literature review could guide future phylogeographic studies in the Neotropics.

Methods

Bibliographic research and selection of studies

To test the three hypotheses stated above, I performed the following literature search and analysis protocol. I first did a systematic literature search considering scientific papers that were either published or under review in scientific journals or peer-reviewed books from 2008–2023 in two sources: Web of Science (WoS) and Google Scholar (GS). For the first source of information (WoS), I performed one search in March 2024 (in English language, reviewing both Spanish and English titles or abstracts) on literature published focused on genomic studies for NMF taxa using the following string of Boolean operators in an advanced search format: TS=(("next-generation sequencing" OR "high-throughput sequencing" OR "single nucleotide polymorphism" OR RADseq OR UCE OR "ultraconserved elements" OR *genomic* OR SNPs) AND ("neotropical montane forest" OR "cloud forest" OR *paramos* OR "temperate forest" OR "tropical mountains" OR "sky islands" OR *highlands* OR *forested* OR *Mesoamerica* OR *Andes*)). For the Scopus search engine (English language), we gathered published paper using the query string on 22 March 2024 as follows: TITLE-ABS-KEY (rad* OR genomic* OR phylogenomic* OR taxonomy* OR delimitation* AND montane* OR cloud* OR highlands* OR temperate* OR sky-islands* AND mesoamerica* OR andes*). Complementary to those sources, I used the following search string in GS, sorted by relevance, in March 2024 spanning 2008–2024: neotropical montane forest radseq* OR uce.

The searches in WoS, Scopus, and GS were restricted to the title, abstract, and keywords. Due to the possible duplication of records and the inclusion of records not suitable for the present trial, a bioinformatic analysis was performed for data cleaning using R v.4.2.3 (R Development Core Team 2023). The same keywords used in the literature searches were used as bioinformatic filters, first on the abstracts of the papers, and subsequently on their titles. Studies had to meet three specific criteria to be included in the systematic review. (1) The study is focused on the framework of landscape genomics/genetics (hereafter LG, Manel et al. 2003;

Balkenhol et al. 2016), population genomics, phylogeography, and phylogenomics (Philippe et al. 2005; Edwards et al. 2015). (2) The study was carried out within NMF (following to Churchill et al. 1995). (3) The genomic study was performed using NGS markers, specifically those related to reduced representation genome-sequencing technologies (RADseq, ddRAD, NextRAD, 3RAD, GBS, TEC, and UCEs).

Measurement of variables

To address *H1*, I measured seven variables (six predictors and one response variable). In the case of the predictor variables, I measured the straight line between the origin site of each NGS method (considering all author affiliations for each study publishing a novel NGS) and the affiliation of the research group leading the studies focused on NMF taxa, here called as $D_{NGS–AF}$. Also, I measured another distance-based predictor (hereafter $D_{AF–BP}$), calculating the distance between author affiliations of these genomic studies and centroids of the studied biogeographic provinces (Morrone et al. 2022). The age of the NGS method (hereafter Δ_{AGE}) was calculated as the difference between the year of publication of a genomics study and the year of publication of the implemented NGS in that study. The number of authors (hereafter NA_{GS}) and the number of nations (hereafter NN_{GS}) involved in each genomic study were also considered as predictor variables. Finally, a variable related to the economy in a place—gross domestic product (hereafter GDP)—was also implemented in the regression-based modeling. By doing so, I extracted the values of annual GIS layers proposed by Chen et al. (2022) considering the author affiliations who performed a genomic study focused on NMF taxa. For studies published from 2019 to up, I extracted the values using the GDP 2019 layer, while the GDP values for studies prior to 2019 correspond to their respective annual layers (2014, 2016, and 2018). Since most scholars and their respective institutes/universities give little transparency in accessing real budget data used in their research works, GDP seems a reasonable approximation of the economic aspect influencing scientific output. Finally, the response variable (hereafter $D_{NGS–BP}$) was calculated as the straight line (distance) between the origin site of each NGS

method (considering all author affiliations for each study publishing a novel sequencing method) and centroids of the studied biogeographic provinces (Morrone et al. 2022). As many of the genomic studies here analyzed have more than one author (including studies publishing a novel NGS method), all measurements were averaged. I performed all geospatial analyses (management of GIS layers and measurements) under a Behrmann projection.

Statistical analyses

To evaluate *H1*, I followed a robust regression-based approach. I contrasted potential spatial, temporal, and economical factors using nine nested linear models and an information-theoretic approach (Burnham and Anderson 2002). Before modeling, the response variable (distance between NGS origin site and biogeographic province) and three explanatory variables were log-transformed to improve distributional properties. Besides, I standardized the predictor variables to a mean of zero and standard deviation of one to allow comparison of the coefficients. To account for the inability of the measured data to meet all of the assumptions of ordinary least-squares (OLS) regression, I fitted robust regressions implemented in the *rlm* function in the library 'MASS' v.7.3-58.2 (Venables and Ripley 2002) and *p*-values were estimated using robust *F*-tests in the library 'sfsmisc' v.1.1-17 (Maechler 2024) in R v.4.2.3 (R Development Core Team 2023). The creation of robust regressions was done using the iterated reweighted least-squares with 10,000 maximum iterations and the Tukey bisquare weight function within each model. In brief, robust regressions use an M-estimator whose starting coefficients and fixed scale were determined by an S-estimator (Yohai et al. 1991). Each model was run 1,000 times, and the resulting Akaike Information Criterion with a correction for small sample size (AICc) values (Burnham and Anderson 2002) were averaged across runs. We ranked the nine nested models based on their AICc scores, with smaller values indicating models with greater relative statistical support. The nine models included a full model that contained all six predictors, one spatio-temporal model with two predictors, and one geoeconomic model with four predictors. The spatio-

temporal and geoeconomic models were further divided into six submodels. The spatio-temporal models considered $D_{NGS–AF}$ and Δ_{AGE} features related with lag-effects implementation of NGS (both spatial and temporal ones, respectively). The geoeconomic models considered the collaborative nature of genomic studies such number of authors (NA_{GS}), number of nations (NN_{GS}), and economy (GDP) features, as well as the distance between author affiliations of these genomic studies and centroids of the studied biogeographic provinces ($D_{AF–BP}$).

To assess *H2* and *H3*, I used a Chi-square analyses in R v.4.2.3 (R Development Core Team 2023) to determine whether there was a significant difference in the proportion of studies among the levels of the following categorical variables: biogeographic province, use of NGS methods, preferred genomic approach, application of minor allele count (MAC) filter, spatial clustering model-based methods, use of ST methods, use of novel regression methods, and quantification of genomic population summary metrics (differentiation and diversity). To determine if there is a connection between LG and phylogenomics, I screened the main phylogenetic textbooks (Wägele 2005; Williams and Ebach 2008; Wiley and Lieberman 2011) using a combination of words for two phrases ("IBD" and "PGS") commonly employed in the population or landscape genetics/genomics literature.

Results

Bibliographic research

The literature search phrases in WoS and Scopus yielded the following numbers of records: 1,093 and 737, respectively. On the other hand, GS searches had 790 records. After a detailed review of each article and excluding records that did not fit the objective of this paper, a final database of 33 records was obtained. Knowledge of the genomics in NMF began to be formalized in 2014 (Fig. 2); then, there was a gap in two years (2015 and 2017) without records, and stabilized during 2018 when standardized NGS methods were applied in Mexico and South America. The year with most publications was 2022. The 33 references, which addressed a total of 29

taxonomic genera, were published between 2014 and 2024 (mean = 3.0, SD = 2.82, range = 0–8 studies per year; Fig. 2). I found spatial clustering of the most studied biogeographic provinces (Fig. 3), with a clear lack of genomic approaches in South America.

Drivers of publication biases

The distance between the origin site of each NGS method (considering all affiliations of the leading research group publishing that NGS technique) and centroids of the studied biogeographic provinces was best explained by spatial and temporal factors (Table 1). Albeit, the spatial predictor variable ($D_{NGS–AF}$) of this model was the only statistically significant (Table 2).

The best represented taxonomic group was Plantae (n = 10) while the least represented was Fungi (n = 1), and this difference was statistically significant (Chi-square test, $\chi^2 = 13.455$, df = 6, $P = 0.036$). The prevalence of different categories varied in each topic (Fig. 4). In summary, the Transmexican Volcanic Belt was the biogeographic province that has received most attention in genomic studies, followed by the Sierra Madre Oriental (Chi-square test, $\chi^2 = 73$, df = 11, $P < 0.05$; Fig. 4a). Evolutionary biologists had a particular interest in a single NGS method, specifically ddRAD was the most prevalent in genomic studies, and this difference was statistically significant (Chi-square test, $\chi^2 = 13.455$, df = 6, $P = 0.036$; Fig. 4b). Regarding the analytical approach, the phylogeographic framework had more presence in the NMF literature than other approaches, but this difference was not statistically significant (Chi-square test, $\chi^2 = 8.969$, df = 4, $P = 0.061$; Fig. 4c).

Notably, few researchers have incorporated the MAC filtering process into their bioinformatic pipelines for cleaning genomic data (Chi-square test, $\chi^2 = 22.091$, df = 1, $P < 0.05$; Fig. 4d). The application of ST methods is still little represented in the literature and its use is not significantly more common than not using it (Chi-square test, $\chi^2 = 0.030$, df = 1, $P = 0.861$; Fig. 4e). Only four studies (12%) have used clustering approaches accounting for IBD (GENELAND or conStruct); this was

significantly fewer than the studies that did not use spatial clustering methods (Chi-square test, $\chi^2 = 18.939$, df = 1, $P < 0.05$; Fig. 4f).

A minority of studies have incorporated new regression methods (i.e., other than Mantel/partial Mantel tests [hereafter MT/PMT]) to evaluate the influence of landscape or geographic factors on genomic variation (Chi-square test, $\chi^2 = 16.03$, df = 1, $P < 0.05$; Fig. 4g). The situation for population summary metrics is comparable, as both genomic differentiation (Chi-square test, $\chi^2 = 0.030$, df = 1, $P = 0.861$; Fig. 4h) and genome-wide diversity (Chi-square test, $\chi^2 = 0.757$, df = 1, $P = 0.384$; Fig. 4i) metrics are mostly quantified but, in both cases, the likelihood of using these metrics were not significantly higher than not using them.

Discussion

When testing *H1*, I found that a model consisting of spatial and temporal features outperformed other models including collaboration-based variables (e.g., economy or number of involved researchers). This suggests that new NGS methods were generally first applied to areas near the place where the method was first published. This also makes sense, since most genomic studies (as in the case of UCEs application) had as authors some researchers who were previously involved in collaborative projects publishing a given NGS method. In broad terms, the southernmost NGS method created (UCEs, Faircloth et al. 2012) was applied only to highland taxa (e.g., *Isthmura* salamanders [Bryson et al. 2018], *Dendrortyx* Wood-Partridges [Tsai et al. 2019]) from the Mexican Transition Zone subregion (e.g., Sierra Madre del Sur, Sierra Madre Occidental). On the other hand, the easternmost NGS method (ddRAD, Peterson et al. 2012) was unevenly applied to montane species from the Transmexican Volcanic Belt and adjacent provinces, until reaching southern latitudes in Central America and South America. A potential explanation for this trend is that many phylogeographic studies have been performed by collaborations between curators of North American museum specimens and scholars of institutes located in the tropics. In addition, it may be necessary to control for finer socio-economic factors (e.g., available infrastructure, trained human

resources) other than coarser geographic factors, which could in turn, be related to the implementation of a particular NGS method in regions from NMF. Nonetheless, many of them are difficult to quantify.

Contrary to *H2*, evolutionary biologists did not prefer RADseq (the earliest NGS method) over more recently proposed NGS technologies (NextRAD, UCEs). Rather, UCEs have been the most frequently employed in phylogenetic studies of NMF taxa. This may be because RADseq was designed for population-genomic level questions, while UCEs are used for deeper phylogenetics (Harvey et al. 2016; Manthey et al. 2016; Toews et al. 2016).

ST methods are beginning to abound in the specialized applied genomics literature (*H3*). Even though several of these phylogenetic reconstruction methods coincided with the dawn of NGS technologies (Liu et al. 2010; Bryant et al. 2012), they began to be used from 2016 onwards for the study of NMF taxa (Zarza et al. 2016). This is an important finding, as phylogeographic hypotheses have been proposed based solely on concatenated gene trees and Sanger sequenced-based genetic results (Ramírez-Barahona and Eguiarte 2013, Ornelas et al. 2013; Ornelas and González 2014), many of which could present conflicting topologies even for well-structured lineages (León-Paniagua et al. 2007; Navarro-Sigüenza et al. 2008), or even using NGS methods (DeRaad et al. 2022).

As expected, fewer studies have accounted for the IBD effect in their genomic inferences studying neotropical highland taxa (Perez et al. 2018). In at least a decade since the first paper on genomics of montane taxa (Mastretta-Yanes et al. 2014), few researchers have applied spatial assignments software (*H3*), despite the need to account for IBD in clustering based-models (Guillot et al. 2005; Meirmans 2012; Bradburd et al. 2016). Yet this represents a publication bias given that IBD is highly expected in the study of genomic variation in natural populations (Sexton et al. 2014). Furthermore, the presence IBD in montane taxa is often used to describe phylogeographic hypotheses (Ramírez-Barahona and Eguiarte 2013; Mastretta-Yanes et al. 2015). The pattern is similar for the use of new regression-based methods (*H3*; Legendre et al. 1994; Clarke et al. 2002) other than MT/PMT (Mantel

1967). Reviewing the literature, only three genomic studies have controlled for the non-independence of pairwise comparison (i.e., maximum likelihood population-effects [hereafter MLPE]) in regression-based methods (*H3*). This goes against what several studies (using real and simulated data) have suggested, as MLPE are one of the most robust methods to evaluate the importance of landscape factors on genetic (dis)similarity/distance (Balkenhol et al. 2009, Shirk et al. 2018; but see Peterman and Pope 2021).

Despite being created just over 20 years ago (Clarke et al. 2002), MLPEs were first introduced in a LG context just over a decade ago (van Strien et al. 2012). Since then, they have had an increasingly significant presence in the literature. This pattern is reinforced by the comparison of regression-based methods in LG (Balkenhol et al. 2009, Shirk et al. 2018), and then the advent of ResistanceGA to optimize resistance GIS layers in microevolutionary studies (Peterman 2018). The latter pattern coincides with the two genomic studies having implemented MLPE through ResistanceGA. Recently, an extension of the original MLPEs now considers the spatial component between sampling locations (Jaffé et al. 2019).

Notably, very few studies report summary metrics of genomic variation (*H3*). This is concerning, given that not having these useful metrics would lead to spurious genomic expectations studying different phylogeographic settings (as proposed in Ramírez-Barahona and Eguiarte 2013). Still, the few studies that calculate these genomic variation metrics took as reference subjectively delimited populations or sampling sites. Therefore, it becomes imperative to calculate these metrics based on groups previously delimited with spatial and not spatial assessment approaches.

Why the lag in the adoption of NGS and analytic approaches?

To date, it is clear that novel genomic or phylogenomic methods have not been fully adopted for the delimitation of species and subsequent generation of phylogeographic hypotheses. This begs the question: What maintains the persistent disconnection between these disciplines? Textbooks are one potential explanation

(Kuhn 1962). Importantly, I did not find a single reference to the phrase “population genetic structure” (or any combination of those words) in a single phylogenetics book (Wägele 2005; Williams and Ebach 2008; Wiley and Lieberman 2011). However, Wiley and Lieberman (2011, p. 60) were the only authors to mention “isolation by distance” precisely in a section dedicated to “*Empirical Methods for Determining Species Limits*.” Hence, the lack of integration in prominent textbooks may be the origin of the disconnect (Kuhn 1962), as many evolutionary biologists in their early training careers are not exposed to these two areas of uncertainty in species delimitation or phylogeographical issues. Although both concepts were formalized decades ago (Wright 1943; 1951), their implementation in the early training careers of systematic biologists could partially resolve this publication bias.

The trouble with the joint effect of IBD and PGS

Although the incorporation of MLPE in the literature means a step forward in the study of genomic variation, it is not without its pitfalls. In my opinion, one aspect that has been overlooked is non-independence due to close evolutionary relationships, and not just in statistical terms. This has been resolved, in part, with a novel modification of MLPE called “Hierarchical MLPE” (HMLPE) that includes PGS as a random variable (Carvalho et al. 2019). While this modification has been implemented for the study of Neotropical lowland taxa (Carvalho et al. 2019; Dalapicolla et al. 2021), it is expected to be even more useful in other ecosystems. For instance, many taxa inhabiting high-elevational habitats have a significant PGS (for plants with low to moderate genetic differentiation [$F_{ST} < 0.2$]; Mastretta-Yanes et al. 2018; Ortego et al. 2023) or marked phylogenetic distinctiveness, as in the case of birds (Navarro-Sigüenza et al. 2008; Rocha-Méndez et al. 2019). The inclusion of a variable representing PGS in evolutionary studies is not a new idea. For instance, LG studies identifying loci or genome-wide regions through outlier detection methods (Joost et al. 2013; Rellstab et al. 2015; Balkenhol et al. 2017), have included PGS as a random effect to account for confounding factor in linear mixed effects models (LMEM).

There are many reasons why populations may be structured in a hierarchical manner (Dionne et al. 2008; Meirmans et al. 2012). Glacial-interglacial dynamics (expansion and contraction) from multiple refugia can lead to heterogeneous landscapes resulting in hierarchical genetic structures (Ramírez-Barahona and Eguiarte 2013; Mastretta-Yanes et al. 2015; Flantua and Hooghiemstra 2018; Flantua et al. 2019). Consequently, demographic movements usually drive subpopulation-level segregation that confers a hierarchy at different spatial scales (Balkenhol et al. 2014). Historical demographic processes may even result in allele gradients that largely overlap with major environmental gradients. This latter assertion is particularly true for many Neotropical montane taxa that have followed a niche-tracking conservatism due to specific elevational or climatic constrains (Linck et al. 2021). Niche conservatism implies inherent proximity (tokogenetic or phylogenetic relationships) between entities (individuals or demes) that share something (usually a physical space), for a long time in evolutionary scales (generally through glacial-interglacial cycles). Thus, PGS (Wright 1953) arises from the mere historical relationships among individuals in places (biogeographic provinces) where they have shared a similar environment at evolutionary scales (Rellstab et al. 2015).

For its part, IBD arises as a result of dispersal limitation and genetic drift irrespective of environmental differences, which lead to genetic differentiation (Wright 1943). Under this premise, IBD should be more pronounced in small populations (Sexton et al. 2014). In the case of NMF, populations of some highland species have repeatedly invaded adjacent suitable sites (usually mountains or peaks of similar elevation) during different glacial and interglacial cycles, giving rise to an isolation by distance effect (Mastretta-Yanes et al. 2015). The further the mountain is from the source population, the greater the genetic differences will be between the two populations (Mastretta-Yanes et al. 2015).

Quantifying PGS: some state-of-the-art suggestions

Molecular ecologists and evolutionists interested in controlling for PGS in their evolutionary studies will be inclined to apply a set of good practices in the post-

processing steps of the genomic pipeline. One of the main issues that must be avoided is altering the distribution of allele frequencies across sites, also known as site frequency spectrum (SFS), as this affects PGS inference (Shafer et al. 2017). During the literature review, I did find that few NMF genomic studies have heeded previous recommendations by Linck and Battey (2019) of excluding singletons (rare variants with a genetic sequence that is present exactly once) from downstream genomic analysis by using a minor allele count (MAC) filter rather than using commonly applied thresholds (0.05 or 0.01) of the “thumb-rule” filter of minor allele frequency (MAF). Some genomic studies continue to use MAF filters and other traditional bioinformatic practices, which may lead to spurious or biased genetic estimates due to strong IBD throughout NMF complexes (Ramírez-Barahona and Eguiarte 2013; Mastretta-Yanes et al. 2015). However, this assertion will also depend on the particular interests of each researcher, the biology of the focal species, phylogeographical history of that taxon, and the genomic marker used (e.g., RADseq, GBS, UCE).

PGS can be modeled considering genetic assignment tests using distance-based methods (e.g., PCA, DAPC) or model-based methods such as STRUCTURE (Pritchard et al. 2000) or GENLAND (Guillot et al. 2005) (among others; see Corander et al. 2008; Caye et al. 2016). In cases where researchers have obtained contrasting results (e.g., different ***K*** values) across methods, an objective method to discriminate among them is to use SpaceMix (Bradburd et al. 2016). SpaceMix creates a bidimensional plot in which the geographic distances between individuals correspond to their expected GD under IBD (Bradburd et al. 2016). This geo-genetic map allows the researcher to determine whether the geo-genetic circles of the samples overlap; one can then infer whether a given set of clusters (obtained by assignment methods) should be merged into a single cluster or not (Bradburd et al. 2016). One alternative to the assignment of samples to genetic groups is the generation of phylogenetic trees in which each sample belongs to a clade (i.e., monophyletic group), which can be incorporated as a random variable (***K***) in MLPE. Whatever the desired focus point, the regression-based approach can then be applied, once the ***K*** value is objectively supported.

Accounting for PGS in evolutionary studies

If we decide to control for the underlying PGS of a given set of pairwise distance data (i.e., $\boldsymbol{K}_{ij}$), we must include it as a random effect (1 | $\boldsymbol{K}_{ij}$) in the MLPE analyses (Carvalho et al. 2019; Dalapicolla et al. 2021). Here, ***K*** is the selected value of population structure results, having employed a "total evidence" approach (combination of model-based clustering methods) of a given sample, where sample "*i"* corresponds to the sample of reference and "*j"* the sample for comparison (i.e., to generate the pairwise distance among all samples within and across genetic groups). Incorporating this random variable ($\boldsymbol{K}_{ij}$) in the MLPE equation may minimize possible time-lag effects in broad scale evolutionary studies (Epps and Keyghobadi 2015). In addition, this procedure increases the number of observations and statistical power of the study (Shirk et al. 2017), without the need for separate regressions for each population at a phylogeographic scale (individuals inhabiting a same biogeographic province) (Meirmans 2012; Balkenhol et al. 2016). This method simultaneously controls for the effect of statistical non-independence and the effect of non-independence due to evolutionary relationships, while favoring models that are best in terms of parsimony (*AIC* and its variants) and statistical performance (R^2). Accordingly, the following formula describes HMLPE: $\boldsymbol{D_Y \sim D_X + D_Z + (1 \mid K_{ij})}$. Where $\boldsymbol{D_Y}$ es the response variable (a vector of genetic distances), $\boldsymbol{D_X}$ is a vector of a given landscape predictor, $\boldsymbol{D_Z}$ is a vector of pairwise geographic distances (IBD), and the value **1** indicates the random intercept effect (with fixed random slope).

In philosophical terms, an MLPE that incorporates PGS as a random variable ('HMLPE') and IBD as a fixed variable approaches the study model according to the biological individuality thesis (Ghiselin 1997) in an ontological sense. If populations are biological individuals (Millstein 2009), then they have spatial and temporal boundaries. Herein, IBD represents the spatial boundary while PGS represents the temporal boundary. In accordance with other topics of biological individuality, this model takes into account the issue of degree (under the conception that individuality is a continuum rather than a discrete state) when determining that samples with

varying levels of admixture correspond to different genetic groups. This is highly expected in some biogeographic provinces, as several montane species present a relatively continuous distribution due to a clinal variation caused by IBD throughout the Transmexican Volcanic Belt (DeRaad et al. 2023). In any case, both MLPE and HLMPE act at broader temporal and spatial scales than other methods (Fig. 5).

The MT/PMT role in the genomics era

There is a rich and well-documented (e.g., Balkenhol et al. 2009) body of research pointing out the flaws in the MT/PMT. However, the growing literature suggests that molecular ecologists studying NMF taxa continue to use MT/PMT for hypothesis testing, and there is apparent reluctance to adopt alternative methods. Sufficient time has passed for MT/PMT to prevail in the literature (Mantel 1967), and it seems that decades must pass for a given analytical method to perpetuate (Kuhn 1962). MT/PMT in genomics are an analogue of the total evidence underlying gene concatenations in phylogenomics. Importantly, both violate several of the model assumptions in which they are framed, and represent a comfort zone for the evolutionary biologist (see Edwards 2009 for such a criticism in phylogenomics).

In my view, MT/PMT should only continue to be used to explore the relationships between predictors and response variables, but not to determine the statistical significance of IBD or other alternatives (e.g., isolation by resistance [McRae 2006] and isolation by environment [Wang and Bradburd 2014]). This will help the researcher determine if the preliminary results deviate from the implementation of other biologically and statistically more robust statistical methods (e.g., MLPE). On the other hand, controlling for IBD (Wright 1943) and/or PGS (Wright 1951; Slatkin and Voelm 1991) in regression-based approaches (e.g., HMLPE) as a tool to detect areas of uncertainty should become standard in any study focusing on microevolutionary processes. This will greatly benefit the study of recent speciation processes (like those expected in species complexes). Notably, several studies focused on delimitation of species are beginning to use LG concepts (e.g., Genty et al. 2022), such as “isolation by environment” (Wang and Bradburd 2014) and

“isolation by resistance” (McRae 2006). In this regard, other alternatives have been considered to take into account the effect of PGS in microevolutionary studies. For example, Prunier et al. (2017) considered that it is necessary to evaluate landscape effects on genomic variation of structured populations by including a response variable (a hierarchical genetic distance between individuals) obtained directly from a binary transformation of the z-transformed STRUCTURE ancestry-based outcomes. However, it is currently unknown whether the method proposed by Prunier et al. (2017) is better than HMLPE (Carvalho et al. 2019) for detecting landscape variables driving gene flow of montane taxa. Meanwhile, previous morphology- or sequence-based phylogenetic topologies can be evaluated using ASTRAL-III (Zhang et al. 2018) to measure their degree of uncertainty (quartet support, ILS) using gene trees (previously estimated with NGS data under frequentist and Bayesian philosophies).

Final considerations

The analyses covered in this review, which control for IBD, PGS, and/or ILS, add value to our evolutionary inferences, in a way that will also allow us to evaluate how much agreement exists between tree-based methods and model-based assignment techniques. The implementation of all of these analyses will yield more generalizable and comparable results, which will offer a more solid framework for the proposal of phylogeographical hypotheses with a better biological realism (e.g., Ramírez-Barahona and Eguiarte 2013; Flantua and Hooghiemstra 2018; Mastretta-Yanes et al. 2018). Although this paper emphasizes studies at broad scales (involving more than one biogeographic province), HMLPE can also be applied in fine-scale studies (a few populations within a single biogeographic province). For instance, PGS has been reported in some studies that considered only a few populations (Borja-Martínez et al. 2022). Therefore, HMLPE are indispensable tools for evolutionary studies on NMF taxa, regardless of the scale of the analysis.

I hope that this paper encourages researchers to use a truly integrative framework at different temporal and spatial scales. If the main goal of evolutionary biology is to

determine patterns of organismal variation (Clarke and Okasha 2013), then it makes no sense to omit such variation by following the traditional reductionist approach (e.g., MT/PMT, MRDM [multiple regression of distance matrices]). The use of both HMLPE and incorporation of spatial assignment methods in evolutionary hypotheses will allow the merging of three disciplines: phylogenomics, phylogeography, and LG. The persistent distinction among these disciplines makes little sense when evaluating species limits and/or phylogeographic hypotheses at a large scale (across one or more biogeographic provinces). Working with SNPs (single nucleotide polymorphism) bridges the rift across different evolutionary domains, confirming the previous assertion by Wang (2010) that the advent of NGS could help to fill in knowledge gaps between phylogeography and LG (see also Rissler 2016). Therefore, the implementation of an integrative framework will allow us to diagnose accurate levels of biodiversity in endangered ecosystems, as expected in the NMF.

References

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376. https://doi.org/10.1371/journal.pone.0003376

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818–830. https://doi.org/10.1111/j.1600-0587.2009.05807.x

Balkenhol N, Holbrook J.D, Onorato D, Zager P, White C, Waits LP (2014) A multi-method approach for analyzing hierarchical genetic structures: A case study with cougars *Puma concolor*. Ecography 37:552–563. https://doi.org/10.1111/j.1600-0587.2013.00462.x

Balkenhol N, Cushman SA, Storfer A, Waits LP (2016) Introduction to landscape genetics – concepts, methods, applications. In: Balkenhol N, Cushman SA,

Storfer A, Waits LP (eds) Landscape genetics: concepts, methods, applications. John Wiley & Sons Ltd, Chichester, United Kingdom, pp 1–7

Balkenhol N, Dudaniec RY, Krutovsky KV, Johnson JS, Cairns DM, Segelbacher G, Selkoe KA, von der Heyden S, Wang IJ, Selmoni O, Joost S (2017) Landscape genomics: Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora OP (ed) Population genomics concepts, strategies and approaches. Springer, Cham, Switzerland, pp 261–372

Barker FK, Burns KJ, Klicka J, Lanyon SM, Lovette IJ (2015) New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. Auk 132:333–348. https://doi.org/10.1642/AUK-14-110.1

Bird CE, Karl SA, Mouse PE, Toonen RJ (2011) Detecting and measuring genetic differentiation. In: Held C, Koenemann S, Schubart CD (eds) Phylogeography and population genetics in Crustacea. CRC Press, Florida, pp 31–55

Borja-Martínez G, Tapia-Flores D, Shafer AB, Vázquez-Domínguez E (2022) Highland forest's environmental complexity drives landscape genomics and connectivity of the rodent *Peromyscus melanotis*. Landsc Ecol 37:1653–1671. https://doi.org/10.1007/s10980-022-01428-6

Bradburd GS, Ralph PL, Coop GM (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67:3258–3273. https://doi.org/10.1111/evo.12193

Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. PLoS Genet 12:e1005703. https://doi.org/10.1371/journal.pgen.1005703

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol 29:1917–1932. https://doi.org/10.1093/molbev/mss086

Bryson RW Jr, Zarza E, Grummer JA, Parra-Olea G, Flores-Villela O, Klicka J, McCormack JE (2018) Phylogenomic insights into the diversification of salamanders in the *Isthmura bellii* group across the Mexican highlands. Mol Phylogenet Evol 125:78–84. https://doi.org/10.1016/j.ympev.2018.03.024

Burnham KP, Anderson RD (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York.

Cadena CD, Klicka J, Ricklefs RE (2007) Evolutionary differentiation in the Neotropical montane region: Molecular phylogenetics and phylogeography of *Buarremon* brush-finches (Aves, Emberizidae). Mol Phylogenet Evol 44:993–1016. https://doi.org/10.1016/j.ympev.2006.12.012

Carvalho CS, Lanes ÉCM, Silva AR, Caldeira CF, Carvalho-Filho N, Gastauer M, Imperatriz-Fonseca VL, Nascimento WJr, Oliveira G, Siqueira JO, Viana PL, Jaffé R (2019) Habitat loss does not always entail negative genetic consequences. Front Genet 10:1011. https://doi.org/10.3389/fgene.2019.01101

Caye K, Deist TM, Martins H, Michel O, François, O (2016) TESS3: fast inference of spatial population structure and genome scans for selection. Mol Ecol Resour 16:540–548. https://doi.org/10.1111/1755-0998.12471

Chattopadhyay B, Garg KM, Ramakrishnan U (2014) Effect of diversity and missing data on genetic assignment with RAD-Seq markers. BMC Res Notes 7:841. https://doi.org/10.1186/1756-0500-7-841

Chen J, Gao M, Cheng S, Hou W, Song M, Liu X, Liu Y (2022) Global 1 km × 1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data. Sci Data 9:202. https://doi.org/10.1038/s41597-022-01322-5

Churchill SP, Balslev H, Forero E, Luteyn JM (eds) (1995) Biodiversity and conservation of Neotropical Montane Forests. Proceeding of the Neotropical Montane Forest Biodiversity and Conservation Symposium, The New York

Botanical Garden, 21–26 June 1993. The New York Botanical Garden, Bronx, New York, pp 103–126

Clarke RT, Rothery P, Raybould AF (2002) Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. J Agric Biol Environ Stat 7:361–372. https://doi.org/10.1198/108571102320

Clarke E, Okasha S (2013) Defining the individual. In: Bouchard F, Huneman P (eds) From groups to individuals: evolution and emerging individuality. MIT Press, Massachusetts, pp 37–53

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9:539. https://doi.org/10.1186/1471-2105-9-539

Cortés-Rodríguez N, Hernández-Baños BE, Navarro-Sigüenza AG, Peterson AT, García-Moreno J (2008) Phylogeography and population genetics of the Amethyst-throated Hummingbird (*Lampornis amethystinus*). Mol Phylogenet Evol 48:1–11. https://doi.org/10.1016/j.ympev.2008.02.005

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Massachusetts

Dalapicolla J, do Prado JR, Percequillo AR, Knowles LL (2021) Functional connectivity in sympatric spiny rats reflects different dimensions of Amazonian forest-association. J Biogeogr 48:3196–3209. https://doi.org/10.1111/jbi.14281

Dantas-Queiroz MV, Hurbath F, de Russo Godoy FM. Lanna FM, Versieux LM, Palma-Silva C (2023) Comparative phylogeography reveals the demographic patterns of neotropical ancient mountain species. Mol Ecol 32:3165–3181. https://doi.org/10.1111/mec.16929

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886. https://doi.org/10.1080/10635150701701083

DeRaad DA, McCormack JE, Chen N, Peterson AT, Moyle RG (2022) Combining species delimitation, species trees, and tests for gene flow clarifies complex

speciation in scrub-jays. Syst Biol 71:1453–1470. https://doi.org/10.1093/sysbio/syac034

DeRaad DA, Applewhite EE, Tsai WLE, Terrill RS, Kingston SE, Braun MJ, McCormack JE (2023) Hybrid zone or hybrid lineage: a genomic reevaluation of Sibley's classic species conundrum in *Pipilo* towhees. Evolution 77:852–869. https://doi.org/10.1093/evolut/qpac068

Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. Mol Ecol 17:2382–2396. https://doi.org/10.1111/j.1365-294X.2008.03771.x

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Mol Ecol 18:2930–2933. https://doi.org/10.1111/j.1365-294X.2009.04270.x

Edwards SV, Shultz AJ, Campbell-Staton SC (2015) Next-generation sequencing and the expanding domain of phylogeography. Folia Zool 64:187–206. https://doi.org/10.25225/fozo.v64.i3.a2.2015

Edwards SV, Robin VV, Ferrand N, Moritz C (2022) The evolution of comparative phylogeography: putting the geography (and more) into comparative population genomics. Genome Biol Evol 14:evab176. https://doi.org/10.1093/gbe/evab176

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Mol Ecol 24:6021–6040. https://doi.org/10.1111/mec.13454

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol 61:717–726. https://doi.org/10.1093/sysbio/sys004

Fišer C, Robinson CT, Malard F (2018) Cryptic species as a window into the paradigm shift of the species concept. Mol Ecol 27:613–635. https://doi.org/10.1111/mec.14486

Flantua SGA, Hooghiemstra H (2018) Historical connectivity and mountain biodiversity (with special reference to the Northern Andes). In: Hoorn C, Perrigo A, Antonelli A (eds) Mountains, Climate and Biodiversity. John Wiley & Sons Ltd, Chichester, United Kingdom, pp 171–185

Flantua SGA, O'Dea A, Onstein RE, Giraldo C, Hooghiemstra H (2019) The flickering connectivity system of the north Andean páramos. J Biogeogr 46:1808–1825. https://doi.org/10.1111/jbi.13607

Gentry AH (1995) Patterns of diversity and floristic composition in Neotropical montane forests. In: Churchill SP, Balslev H, Forero E, Luteyn JM (eds) Biodiversity and conservation of Neotropical Montane Forests. Proceeding of the Neotropical Montane Forest Biodiversity and Conservation Symposium, The New York Botanical Garden, 21–26 June 1993. The New York Botanical Garden, Bronx, New York, pp 103–126

Genty G, Guarnizo CE, Ramírez JP, Barrientos LS, Crawford AJ (2022) Landscape genetics and species delimitation in the Andean palm rocket frog (Aromobatidae, *Rheobates*). J Zool Syst Evol Res 2022:6774225. https://doi.org/10.1155/2022/6774225

Ghiselin MT (1997) Metaphysics and the Origin of Species. State University of New York Press, New York

Giles-Pérez GI, Aguirre-Planter E, Eguiarte LE, Jaramillo-Correa JP (2022) Demographic modelling helps track the rapid and recent divergence of a

conifer species pair from Central Mexico. Mol Ecol 31:5074–5088. https://doi.org/10.1111/mec.16646

Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. Mol Ecol Notes 5:712–715. https://doi.org/10.1111/j.1471-8286.2005.01031.x

Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT (2016) Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst Biol 65:910–924. https://doi.org/10.1093/sysbio/syw036

Harvey MG, Singhal S, Rabosky DL (2019) Beyond reproductive isolation: Demographic controls on the speciation process. Annu Rev Ecol Evol Syst 50:75–95. https://doi.org/10.1146/annurev-ecolsys-110218-024701

Hernández-Langford DG, Siqueiros-Delgado ME, Ruíz-Sánchez E (2020) Nuclear phylogeography of the temperate tree species *Chiranthodendron pentadactylon* (Malvaceae): Quaternary relicts in Mesoamerican cloud forests. BMC Evol Biol 20:44. https://doi.org/10.1186/s12862-020-01605-8

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6:e1000862. https://doi.org/10.1371/journal.pgen.1000862

Jaffé R, Veiga JC, Pope NS, Lanes ÉCM, Carvalho CS, Alves R., Andrade SCS, Arias MC, Bonatti V, Carvalho AT, de Castro MS, Contrera FAL, Francoy TM, Freitas BM, Giannini TC, Hrncir M, Martins CF, Oliveira G, Saraiva AM, Souza BA, Imperatriz-Fonseca VL (2019) Landscape genomics to the rescue of a tropical bee threatened by habitat loss and climate change. Evol Appl 12:1164–1177. https://doi.org/10.1111/eva.12794

Joost S, Vuilleumier S, Jensen JD, Schoville S, Leempoel K, Stucki S, Widmer I, Melodelima C, Rolland J, Manel S (2013) Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical

population genetics. Mol Ecol 22:3659–3665. https://doi.org/10.1111/mec.12352

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Laudan L (1977) Progress and its problems. University of California Press, California

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487–1499. https://doi.org/10.1111/j.1558-5646.1994.tb02191.x

León-Paniagua L, Navarro-Sigüenza AG, Hernández-Baños BE, Morales JC (2007) Diversification of the arboreal mice of the genus *Habromys* (Rodentia: Cricetidae: Neotominae) in the Mesoamerican highlands. Mol Phylogenet Evol 42:653–664. https://doi.org/10.1016/j.ympev.2006.08.019

Lemopoulos A, Prokkola JM, Uusi-Heikkilä S, Vasemägi A, Huusko A, Hyvärinen P, Koljonen M-L, Koskiniemi J, Vainikka A (2019) Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. Ecol Evol 9:2106–2120. https://doi.org/10.1002/ece3.4905

Linck E, Battey CJ (2019) Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol Ecol Resour 19:639–647. https://doi.org/10.1111/1755-0998.12995

Linck EB, Freeman BG, Cadena CD, Ghalambor CK (2021) Evolutionary conservatism will limit responses to climate change in the tropics. Biol Lett 17:20210363. https://doi.org/10.1098/rsbl.2021.0363

Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol Biol 10:302. https://doi.org/10.1186/1471-2148-10-302

Maechler M (2024) sfsmisc: Utilities from 'Seminar fuer Statistik' ETH Zurich. R package version 1.1-17. https://CRAN.R-project.org/package=sfsmisc

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197. https://doi.org/10.1016/S0169-5347(03)00008-9

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220.

Manthey JD, Campillo LC, Burns KJ, Moyle RG (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). Syst Biol 65:640–650. https://doi.org/10.1093/sysbio/syw005

Mason NA, Fletcher NK, Gill BA, Funk WC, Zamudio KR (2020) Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. Syst Biodivers 18:269–280. https://doi.org/10.1080/14772000.2020.1730475

Mastretta-Yanes A, Zamudio S, Jorgensen TH, Arrigo N, Alvarez N, Piñero D, Emerson BC (2014) Gene duplication, population genomics, and species-level differentiation within a tropical mountain shrub. Genome Biol Evol 6:2611–2624. https://doi.org/10.1093/gbe/evu205

Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC (2015) Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J Biogeogr 42:1586–1600. https://doi.org/10.1111/jbi.12546

Mastretta-Yanes A, Xue AT, Moreno-Letelier A, Jorgensen TH, Alvarez N, Piñero D, Emerson BC (2018) Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. Mol Ecol 27:432–448. https://doi.org/10.1111/mec.14461

McCormack JE, Peterson AT, Bonaccorso E, Smith TB (2008) Speciation in the highlands of Mexico: genetic and phenotypic divergence in the Mexican jay

(*Aphelocoma ultramarina*). Mol Ecol 17:2505–2521. https://doi.org/10.1111/j.1365-294X.2008.03776.x

McRae BH (2006) Isolation by resistance. Evolution 60:1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

Meirmans PG (2012) The trouble with isolation by distance. Mol Ecol 21:2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Millstein RL (2009) Populations as individuals. Biol Theory 4:267–273. https://doi.org/10.1162/biot.2009.4.3.267

Mirarab S, Bayzid MS, Warnow T (2016) Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst Biol 65:366–380. https://doi.org/10.1093/sysbio/syu063

Moreno-Contreras I, Sánchez-González LA, Arizmendi MC, Prieto-Torres DA, Navarro-Sigüenza AG (2020) Climatic niche evolution in the *Arremon brunneinucha* complex (Aves: Passerellidae) in a Mesoamerican landscape. Evol Biol 47:123–132. https://doi.org/10.1007/s11692-020-09498-7

Morrone JJ (2014) Biogeographical regionalisation of the Neotropical region. Zootaxa 3782:1–110. https://doi.org/10.11646/zootaxa.3782.1.1

Morrone JJ, Escalante T, Rodríguez-Tapia G, Carmona A, Arana M, and Mercado-Gómez JD (2022) Biogeographic regionalization of the Neotropical region: New map and shapefile. An Acad Bras Cienc 94:e20211167. https://doi.org/10.1590/0001-3765202220211167

Navarro-Sigüenza AG, Peterson AT, Nyari A, García-Deras GM, García-Moreno J (2008) Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. Mol Phylogenet Evol 47:21–35. https://doi.org/10.1016/j.ympev.2007.11.030

Nater A, Burri R, Kawakami T, Smeds L, Ellegren H (2015) Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Syst Biol 64:1000–1017. https://doi.org/10.1093/sysbio/syv045

Ornelas JF, Sosa V, Soltis DE, Daza JM, González C, Soltis PS, Gutiérrez-Rodríguez C, Espinosa de los Monteros A, Castoe TA, Bell C, Ruiz-Sanchez E (2013) Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. PLoS ONE 8:e56283. https://doi.org/10.1371/journal.pone.0056283

Ornelas JF, González C (2014) Interglacial genetic diversification of *Moussonia deppeana* (Gesneriaceae), a hummingbird-pollinated, cloud forest shrub in northern Mesoamerica. Mol Ecol 23:4119–4136. https://doi.org/10.1111/mec.12841

Ortego J, Espelta JM, Armenteras D, Díez MC, Muñoz A, Bonal R (2023) Demographic and spatially explicit landscape genomic analyses in a tropical oak reveal the impacts of late Quaternary climate change on Andean montane forests. Mol Ecol 32:3182–3199. https://doi.org/10.1111/mec.16930

Perez MF, Franco FF, Bombonato JR, Bonatelli IA, Khan G, Romeiro-Brito M, Fegies AC, Ribeiro PM, Silva GAR, Moraes EM (2018) Assessing population structure in the face of isolation by distance: are we neglecting the problem? Divers Distrib 24:1883–1889. https://doi.org/10.1111/ddi.12816

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Peterman WE (2018) ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. Methods Ecol Evol 9:1638–1647. https://doi.org/10.1111/2041-210X.12984

Peterman WE, Pope NS (2021) The use and misuse of regression models in landscape genetic analyses. Mol Ecol 30:37–47. https://doi.org/10.1111/mec.15716

Philippe H, Delsuc F, Brinkmann H, Lartillot N (2005) Phylogenomics. Annu Rev Ecol Evol Syst 36:541–562. https://doi.org/10.1146/annurev.ecolsys.35.112202.130205

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

Prunier JG, Colyn M, Legendre X, Flamand MC (2017) Regression commonality analyses on hierarchical genetic distances. Ecography 40:1412–1425. https://doi.org/10.1111/ecog.02108

Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE (2014) Demystifying the RAD fad. Mol Ecol 23:5937–5942. https://doi.org/10.1111/mec.12965

Ramírez-Barahona S, Eguiarte LE (2013) The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. Ecol Evol 3:725–738. https://doi.org/10.1002/ece3.483

R Development Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. Mol Ecol 24:4348–4370. https://doi.org/10.1111/mec.13322

Rissler LJ (2016) Union of phylogeography and landscape genetics. Proc Natl Acad Sci USA 113:8079–8086. https://doi.org/10.1073/pnas.160107311

Rocha-Méndez A, Sánchez-González LA, González C, Navarro-Sigüenza AG (2019) The geography of evolutionary divergence in the highly endemic avifauna from the Sierra Madre del Sur, Mexico. BMC Evol Biol 19:237. https://doi.org/10.1186/s12862-019-1564-3

Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106. https://doi.org/10.7717/peerj.1106

Sánchez-González LA, Morrone JJ, Navarro-Sigüenza AG (2008) Distributional patterns of the Neotropical humid montane forest avifaunas. Biol J Linn Soc 94:175–194. https://doi.org/10.1111/j.1095-8312.2008.00979.x

Sánchez-González LA, Navarro-Sigüenza AG (2009) History meets ecology: a geographical analysis of ecological restriction in the Neotropical humid montane forests avifaunas. Divers Distrib 15:1–11. https://doi.org/10.1111/j.1472-4642.2008.00506.x

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68:1-15. https://doi.org/10.1111/evo.12258

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol 8:907–917. https://doi.org/10.1111/2041-210X.12700

Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI (2010) Inferring landscape effects on gene flow: a new model selection framework. Mol Ecol 19:3603–3619. https://doi.org/10.1111/j.1365-294X.2010.04745.x

Shirk AJ, Landguth EL, Cushman SA (2017) A comparison of individual-based genetic distance metrics for landscape genetics. Mol Ecol Resour 17:1308–1317. https://doi.org/10.1111/1755-0998.12684

Shirk AJ, Landguth EL, Cushman SA (2018) A comparison of regression methods for model selection in individual-based landscape genetic analysis. Mol Ecol Resour 18:55–67. https://doi.org/10.1111/1755-0998.12709

Slatkin M, Voelm L (1991) F_{ST} in a hierarchical island model. Genetics 127:627–629. https://doi.org/10.1093/genetics/127.3.627

Suárez-Díaz E, Anaya-Muñoz VH (2008) History, objectivity, and the construction of molecular phylogenies. Stud Hist Philos Biol Biomed Sci 39:451–468. https://doi.org/10.1016/j.shpsc.2008.09.002

Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, van der Lingen CD, von der Heyden S, Chiazzari B, van Vuuren BJ, Beheregaray LB (2018) Mitochondrial DNA is unsuitable to test for isolation by distance. Sci Rep 8:8448. https://doi.org/10.1038/s41598-018-25138-9

Toews DPL, Campagna L, Taylor SA, Balakrishnan CN, Baldassarre DT, Deane-Coe PE, Harvey MG, Hooper DM, Irwin DE, Judy CD, Mason NA, McCormack JE, McCracken KG, Oliveros CH, Safran RJ, Scordato ESC, Stryjewski KF, Tigano A, Uy JAC, Winger BM (2016) Genomic approaches to understanding population divergence and speciation in birds. Auk 133:13–30. https://doi.org/10.1642/AUK-15-51.1

Tsai WLE, Mota-Vargas C, Rojas-Soto O, Bhowmik R, Liang EY, Maley JM, Zarza E, McCormack JE (2019) Museum genomics reveals the speciation history of *Dendrortyx* wood-partridges in the Mesoamerican highlands. Mol Phylogenet Evol 136:29–34. https://doi.org/10.1016/j.ympev.2019.03.017

van Strien MJ, Keller D, Holderegger R (2012) A new analytical approach to landscape genetic modelling: Least-cost transect analysis and linear mixed models. Mol Ecol 21:4010–4023. https://doi.org/10.1111/j.1365-294X.2012.05687.x

Venables WN, Ripley BD (2002) Modern applied statistics with S. Springer, New York.

Wägele J-W (2005) Foundations of phylogenetic systematics. Verlag Dr Friedrich Pfeil, Germany

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. Mol Ecol 19:2605–2608. https://doi.org/10.1111/j.1365-294X.2010.04715.x

Wang IJ, Bradburd GS (2014) Isolation by environment. Mol Ecol 23:5649–5662. https://doi.org/10.1111/mec.12938

Wiley EO, Lieberman BS (2011) Phylogenetics: theory and practice of phylogenetic systematics, 2nd edn. John Wiley & Sons, Inc, New Jersey

Williams DM, Ebach MC (2008) Foundations of Systematics and Biogeography. Springer Science+Business Media, New York

Wright S (1943) Isolation by distance. Genetics 28:114–138. https://doi.org/10.1093/genetics/28.2.114

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354. https://10.1111/j.1469-1809.1949.tb02451.x

Yohai V, Stahel WA, Zamar RH (1991) A procedure for robust estimation and inference in linear regression. In: Stahel WA, Weisberg SD (eds) Directions in robust statistics and diagnostics, part II. Springer, New York, USA, pp 365–374

Zarza E, Faircloth BC, Tsai WLE, Bryson RW Jr, Klicka J, McCormack JE (2016) Hidden histories of gene flow in highland birds revealed with genomic markers. Mol Ecol 25:5144–5157. https://doi.org/10.1111/mec.13813

Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform 19:153. https://doi.org/10.1186/s12859-018-2129-y

FIGURES

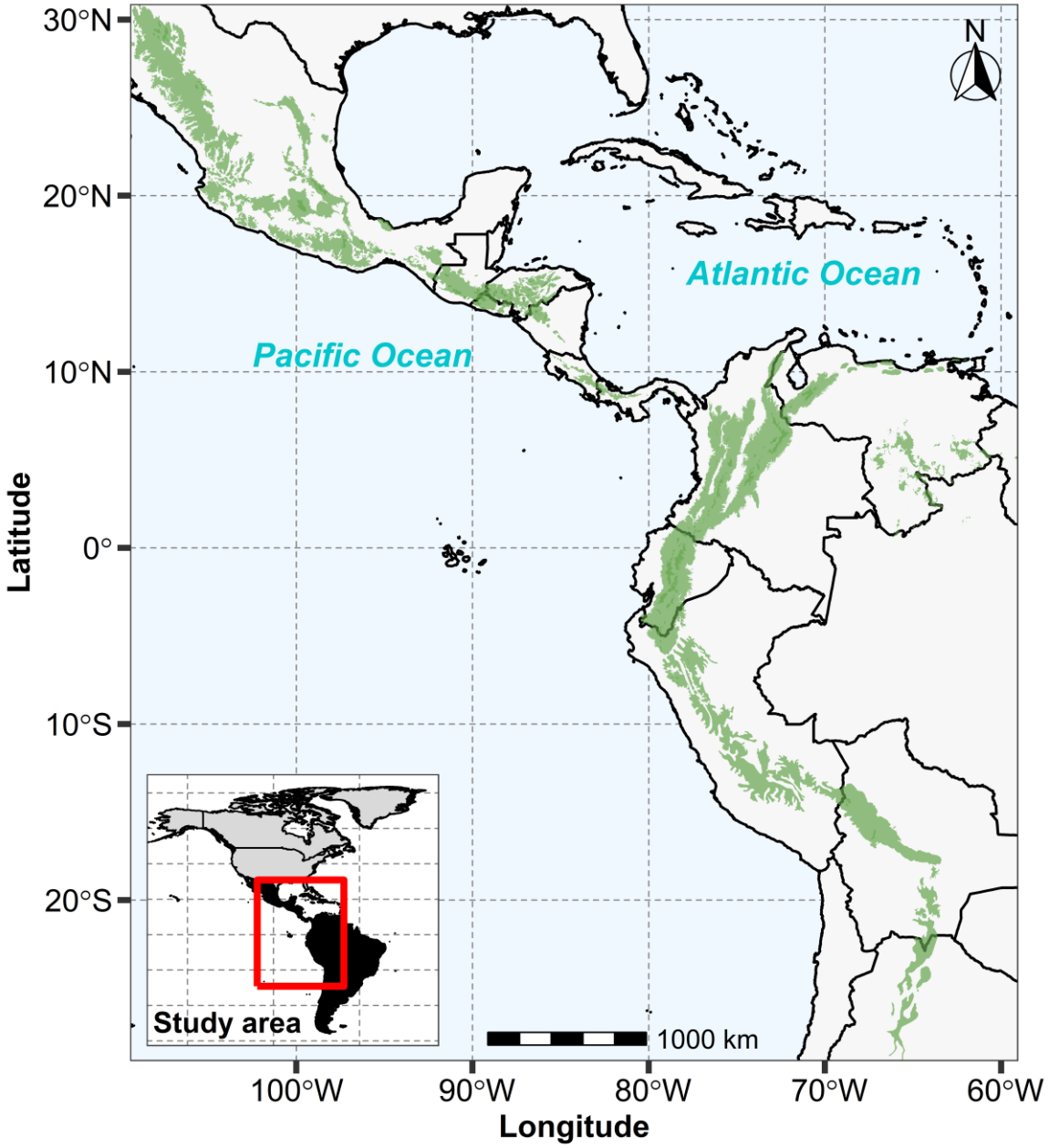


Fig. 1 Map showing the geographic distribution (green polygons) of the Neotropical Montane Forests (NMF). The inset map shows the study area highlighted with a red rectangle.

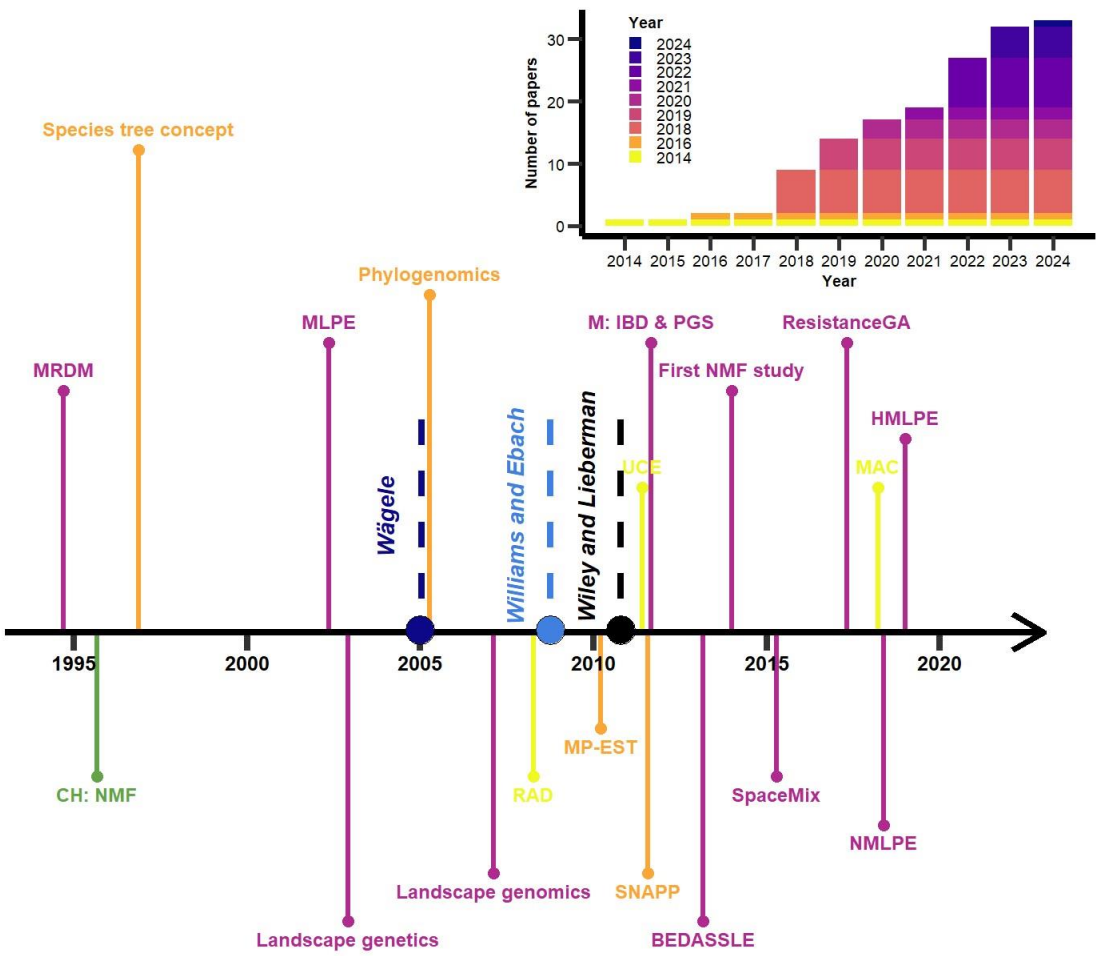


Fig. 2 Timeline indicating some scientific revolutions related to next-generation sequencing (NGS), and main events on the literature for Neotropical Montane Forests (NMF) taxa. Purple lollipops correspond to landscape genetics/genomics literature, orange lollipops are related to phylogenomics literature, yellow lollipops refer to publication of a new NGS method or important filtering process for genomic data (e.g., minor allele count [MAC]), and the green lollipop refers to publication date of a main scientific book (Churchill et al. 1995) on the biogeography or ecology of NMF. Dashed lines indicate main phylogenetic textbooks. Inverse lollipops refer to main textbooks on phylogenetics (Wägele 2005 [dark blue], Williams and Ebach 2008 [light blue], Wiley and Lieberman 2011 [black]). The embedded bar plot shows the accumulation of genomic papers studying NMF taxa since the first one was published.

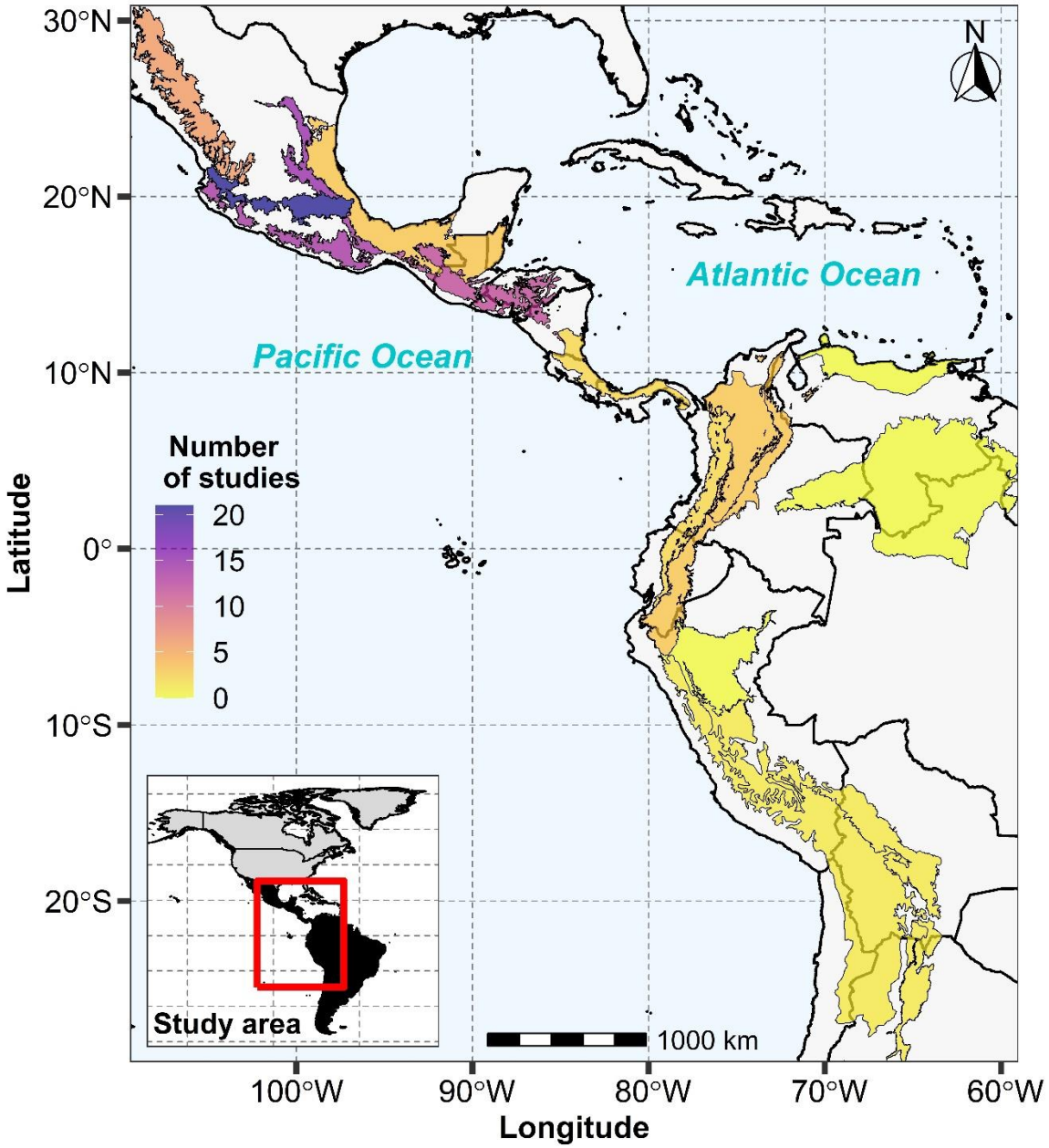


Fig. 3 Map showing the spatial distribution of the number of genomic studies having surveyed a determined biogeographic province (Morrone et al. 2022) from the Neotropical Montane Forests (NMF) complex. The inset map indicates the study area.

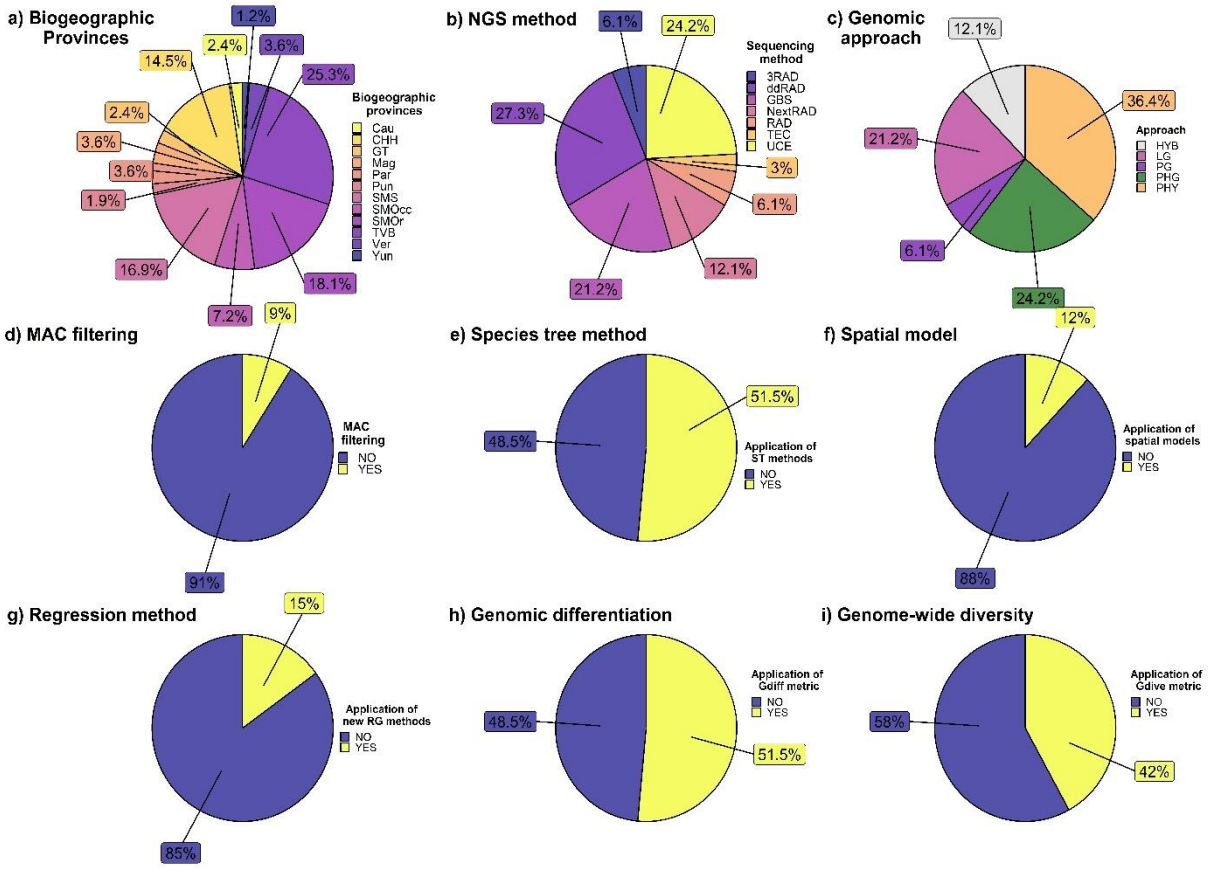


Fig. 4 Pie charts depicting percentages of studies for each level of the following categorical variables: a) Biogeographic provinces (Cau = Cauca, CHH = Chiapas Highlands, GT = Guatuso-Talamanca, Mag = Magdalena, Par = Paramo, Pun = Puna, SMS = Sierra Madre del Sur, SMOcc = Sierra Madre Occidental, SMOr = Sierra Madre Oriental, TVB = Transmexican Volcanic Belt, Ver = Veracruzan, Yun = Yungas); b) next-generation sequencing (NGS) method; c) Genomic approach (HYB = hybrid study, LG = landscape genomics, PG = population genomics, PHG = phylogeography, PHY = phylogenomics); d) Use minor allele count (MAC) filtering; e) Use of spatial clustering based-models; f) Use of species tree methods; g) Use of regression models other than Mantel test/partial Mantel test; h) Quantification of genomic differentiation metric; i) Quantification of genome-wide diversity.

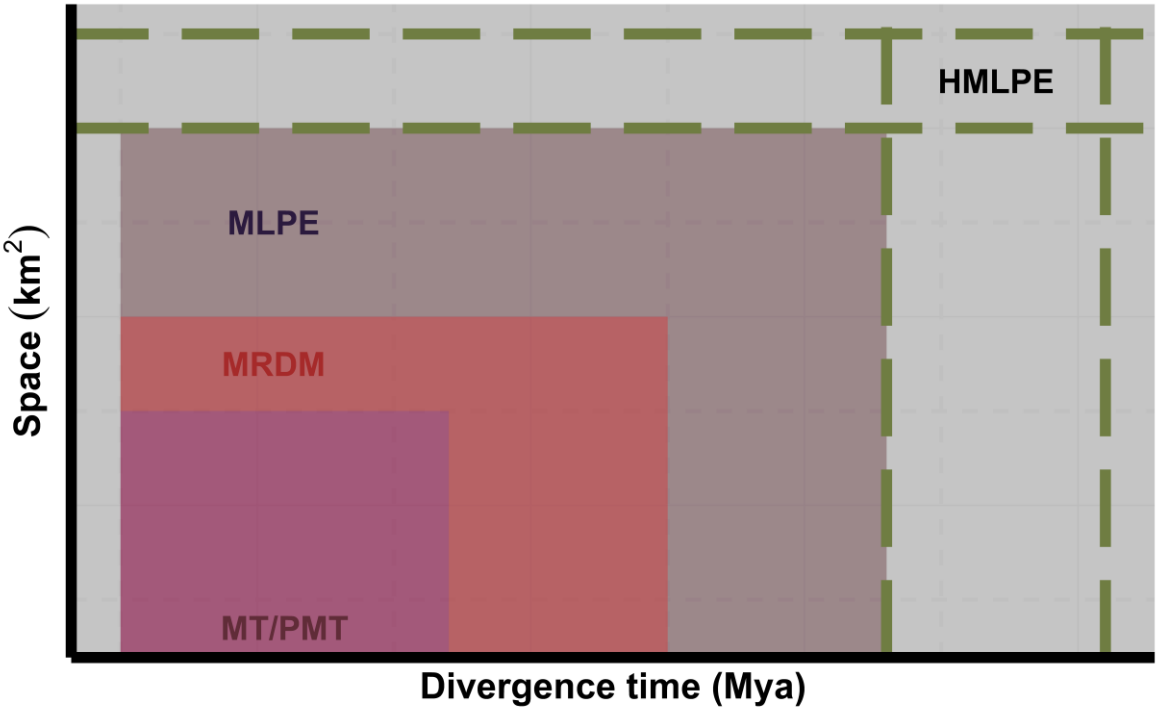


Fig 5. A fictitious bivariate plot showing the temporal and spatial dimensions at which different regression-based methods may be relevant in (micro-)evolutionary studies. It is important to note that in highland species complexes, statistically significant IBD only makes sense when controlling for PGS. The zone of statistical significance is the (possible) overlap between (vertical and horizontal) dashed lines. The different methods are depicted as follows: MT/PMT = Mantel tests/partial Mantel tests, MRDM = multiple regression of distance matrices, MLPE = maximum likelihood population-effects, and HMLPE = hierarchical maximum likelihood population-effects.

TABLES

Table 1 Robust regression models contrasting spatial, temporal, and economy-based correlates of distance between the origin site of each NGS and centroids of the studied biogeographic provinces ($D_{NGS–BP}$) in genomic studies focused on NMF taxa. Change in model AICc (Δ_i) represents the difference between model *i* and the model with the lowest AICc score; AICc weight (w_i) is the level of evidence for model *i* based on the entire set of models; the model with the minimum AICc is shown in italics. The response variable was log-transformed before analysis.

	$D_{NGS–BP}$		
Model	**AICc**	**Δ_i**	**w_i**
Full	29.087	9.794	0.003
Spatio-temporal	*19.293*	*0.00*	*0.430*
Spatial	19.525	0.232	0.383
Temporal	24.680	5.387	0.029
Geoeconomic	28.547	9.254	0.004
Near from home	22.131	2.838	0.104
Resources	28.420	9.127	0.004
Economy	24.620	5.327	0.030
Collaboration	26.680	7.387	0.010

Table 2 Robust regression coefficients, standard error, and test statistics for the best model (in terms of AICc) modeling distance between the origin site of each NGS (considering the averaged values of all author affiliations publishing that method, respectively) and centroids of the studied biogeographic provinces ($D_{NGS–BP}$) in genomic studies focused on NMF taxa. $D_{NGS–AF}$ is the geographic distance between the origin site of each NGS method (considering all author affiliations for each study publishing a novel NGS) and affiliation of the research group leading the studies focused on NMF taxa; Δ_{AGE} is the difference between the year of publication of a genomics study and the year of publication of the implemented NGS in that study. The predictor influencing statistically the response variable is shown in italics. SE represents the standard errors for the best model.

Model	**Predictors**	**Coefficients**	**SE**	***t*-value**	***F***	***P***
Spatio-temporal	Intercept	8.078	0.047	168.479		
	$D_{NGS–AF}$	*0.153*	*0.048*	*3.139*	*9.418*	*0.004*
	Δ_{AGE}	0.088	0.048	1.805	3.401	0.075

CAPÍTULO II. Influencia de las variables paisajísticas sobre la variación genómica en aves asociadas a los Bosques Montanos Neotropicales

CAPÍTULO II. Influencia de las variables paisajísticas sobre la variación genómica en aves asociadas a los Bosques Montanos Neotropicales

Este capítulo fue publicado como artículo científico en **MOLECULAR ECOLOGY** (ISI WEB Impact factor: **6.622**) bajo el título "**Isolation by resistance explains genetic diversity in the *Arremon* brushfinches of northern Mesoamerica**".

Dicho documento fue sometido como **artículo de requisito** para el Programa de **Doctorado en Ciencias Biológicas**.

Tradicionalmente llamado *Arremon b. apertus* (pero ver **Capítulo 3**) en Veracruz.
Fotografía por: Juan Manuel Carmona-Victoria.

Check for updates

Received: 4 February 2022 | Accepted: 27 March 2023

DOI: 10.1111/mec.16948

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Isolation by resistance explains genetic diversity in the *Arremon* brushfinches of northern Mesoamerica

Israel Moreno-Contreras[1,2] | Alexander Llanes-Quevedo[1,2] | Luis A. Sánchez-González[2] | María del Coro Arizmendi[3] | Adolfo G. Navarro-Sigüenza[2]

[1]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1° Piso, Circuito de Posgrados, Ciudad Universitaria, Universidad Nacional Autónoma de México, Mexico City, Mexico

[2]Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico

[3]UBIPRO, FES Iztacala Universidad Nacional Autónoma de México, Tlalnepantla, Mexico

Correspondence
Adolfo G. Navarro-Sigüenza, Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, A.P. 70-399, Mexico City 04510, Mexico.
Email: adolfon@ciencias.unam.mx

Funding information
DGTIC, UNAM, Grant/Award Number: LANCAD-UNAM-DGTIC-402; Dirección General de Asuntos del Personal Académico, UNAM, Grant/Award Number: IN214621

Handling Editor: David Coltman

Abstract

Genetic differentiation between and within natural populations is the result of the joint effects of neutral and adaptative processes. In addition, the spatial arrangement of the landscape promotes connectivity or creates barriers to gene flow, directly affecting speciation processes. In this study, we carried out a landscape genomics analysis using NextRAD data from a montane forest specialist bird complex, the Mesoamerican Chestnut-capped/Green-striped Brushfinch of the genus *Arremon*. Specifically, we examined population genomic structure using different assignment methods and genomic differentiation and diversity, and we tested alternative genetic isolation hypotheses at the individual level (e.g., isolation by barrier, IBB; isolation by environment, IBE; isolation by resistance, IBR). We found well-delimited genomic structuring ($K=5$) across Mesoamerican montane forests in the studied group. Individual-level genetic distances among major montane ranges were mainly explained by IBR hypotheses in this sedentary Neotropical taxon. Our results uncover genetic distances/differentiation and patterns of gene flow in allopatric species that support the role of tropical mountains as spatial landscape drivers of biodiversity. IBR clearly supports a pattern of conserved niche-tracking of suitable habitat conditions and topographic complexity throughout glacial-interglacial dynamics.

KEYWORDS
allopatry, Chestnut-capped Brushfinch, isolation by barrier, isolation by resistance, landscape genomics, NextRAD

1 | INTRODUCTION

Evolutionary theory suggests that genetic differentiation between and within natural populations is the result of neutral (e.g., gene flow) and adaptive (e.g., natural selection) processes (Mayr, 1963; Nosil et al., 2009; Slarkin, 1985). In addition, historical, geological, and contemporary environmental factors also contribute to spatial patterns of genetic variation (Bradburd & Ralph, 2019; Orsini et al., 2013; Turelli et al., 2001). Thus, the spatial arrangement of landscape features drives connectivity or creates barriers to gene flow, directly affecting divergence processes (Richardson et al., 2016; Sexton et al., 2014).

Since the emergence of landscape genetics (Manel et al., 2003), molecular ecologists have used this interdisciplinary framework to

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

Isolation by resistance explains genetic diversity in the *Arremon* brushfinches of northern Mesoamerica

Running title: Landscape genomics of Mesoamerican brushfinches

Authors and Affiliations

Israel Moreno-Contreras[1,2], Alexander Llanes-Quevedo[1,2], Luis A. Sánchez-González[2], María del Coro Arizmendi[3], Adolfo G. Navarro-Sigüenza[2,*]

[1]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1º Piso, Circuito de Posgrados, Ciudad Universitaria, Universidad Nacional Autónoma de México, Coyoacán, Mexico City 04510, Mexico.

[2]Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, A.P. 70–399, Mexico City 04510, Mexico.

[3]UBIPRO, FES Iztacala Universidad Nacional Autónoma de México, Av. de los Barrios, Los Reyes Iztacala, Tlalnepantla 54090, Estado de México, Mexico.

*Corresponding author: adolfon@ciencias.unam.mx

Abstract

Genetic differentiation between and within natural populations is the result of the joint effects of neutral and adaptative processes. In addition, the spatial arrangement of the landscape promotes connectivity or creates barriers to gene flow, directly affecting speciation processes. In this study, we carried out a landscape genomics analysis using NextRAD data from a montane forest specialist bird complex, the Mesoamerican Chestnut-capped/Green-striped Brushfinch of the genus *Arremon*. Specifically, we examined population genomic structure using different assignment methods and genomic differentiation and diversity, and we tested alternative genetic isolation hypotheses at the individual level (e.g., isolation by barrier, IBB; isolation by environment, IBE; isolation by resistance, IBR). We found well-delimited genomic structuring (K = 5) across Mesoamerican montane forests in the studied group. Individual-level genetic distances among major montane ranges were mainly explained by IBR hypotheses in this sedentary Neotropical taxon. Our results uncover genetic distances/differentiation and patterns of gene flow in allopatric species that support the role of tropical mountains as spatial landscape drivers of biodiversity. IBR clearly supports a pattern of conserved niche-tracking of suitable habitat conditions and topographic complexity throughout glacial-interglacial dynamics.

Keywords

allopatry, Chestnut-capped Brushfinch, isolation by barrier, isolation by resistance, landscape genomics, NextRAD

Introduction

Evolutionary theory suggests that genetic differentiation between and within natural populations is the result of neutral (e.g., gene flow) and adaptive (e.g., natural selection) processes (Mayr, 1963; Slatkin, 1985; Nosil et al., 2009). In addition, historical, geological, and contemporary environmental factors also contribute to spatial patterns of genetic variation (Turelli et al., 2001; Orsini et al., 2013; Bradburd & Ralph, 2019). Thus, the spatial arrangement of landscape features drives connectivity or creates barriers to gene flow, directly affecting divergence processes (Sexton et al., 2014; Richardson et al., 2016).

Since the emergence of landscape genetics (Manel et al., 2003), molecular ecologists have used this interdisciplinary framework to unravel how landscapes impede or promote gene flow within and among populations (Dyer, 2015; Balkenhol et al., 2016). The implementation of next-generation sequencing methods has overcome some of the most important limitations of traditional markers by expanding the number of sampled loci and covering the whole genome (or at least a representative part of it; Goodwin et al., 2016). For example, novel genomic reduced-representation methods (e.g., RAD-seq, GBS) have allowed the use of thousands of single nucleotide polymorphisms (SNPs) to address important topics in landscape genomics (hereafter LG) (Puritz et al., 2014; McKinney et al., 2017). Such applications include evaluating genetic isolation hypotheses or carrying out genome scans for loci under selection (Balkenhol et al., 2017; Termignoni-García et al., 2017; Robles-Bello et al., 2022). At the same time, the availability of high-resolution spatial data has made it possible to generate isolation hypotheses that reflect the neutral genomic structure and variation of natural populations at different biologically relevant spatial scales (Balkenhol et al., 2016; Balkenhol & Fortin, 2016).

Although much of population genetics theory is based on an idealized model of discretely bounded and panmictic populations, there are common scenarios that violate assumptions of random mating (Mason et al., 2020). For instance, inbreeding and limited panmixia driven by a higher probability of mating with nearby groups of individuals can leave populations spatially distant and functionally isolated

(Meirmans, 2012; Cushman et al., 2016). This has led to the development of several isolation hypotheses to explain the neutral genetic variation across a given landscape.

The ‘isolation by distance’ (IBD) hypothesis is typically considered the null hypothesis (H_0). IBD predicts a positive relationship between genetic distance (hereafter GD) and geographic distance in natural populations, in absence of natural selection (Wright, 1943; Sexton et al., 2014). The ‘isolation by barrier’ (IBB; Balkenhol et al., 2017) hypothesis considers genetic variation to be the result of fragmented or discontinuous landscapes (Shirk et al., 2010). It is frequently invoked in species with disjunctly distributed populations, where contemporary barriers reinforce genetic divergence that occurred due to an initial vicariant event that may have occurred long ago (Mayr, 1963). In this scenario, dispersal is affected by landscape heterogeneity, and populations may be sub-structured by barriers that limit gene flow (Cushman et al., 2006; Balkenhol et al., 2017). The ‘isolation by environment’ (IBE) hypothesis predicts increasing GD with increasing (Euclidean) environmental distance among sampling units (Wang & Bradburd, 2014). IBE may be the result of processes such as selection against immigrants, reduced hybrid fitness (selection against hybrids), or biased rates of dispersal modulated by climatic conditions (Sexton et al., 2014; Wang & Bradburd, 2014).

Finally, the ‘isolation by resistance’ (IBR) hypothesis proposes that the landscape matrix poses heterogeneous resistance to movement and gene flow (“functional connectivity”) in the absence of natural selection (McRae, 2006). IBR may be particularly evident if the reproductive strategy of an organism involves low vagility and it has a limited ability to invade semipermeable areas that lack suitable habitat. This is typically observed in montane forest species that are unable to establish in drier lowland valleys (Mastretta-Yanes et al., 2018; Thom et al., 2021). IBR represents features or effective distances in the landscape that may be acting as physically isolating barriers to dispersal. Unlike IBB, IBR predicts that GD may be the result of a gradient measured as continuous effective distances between sampling sites, rather than of a categorical patch-based entity, measured as the

number of barriers between samples (Cushman et al., 2006). IBR in conjunction with IBD creates clinal spatial differences in allele frequencies (Ruiz-González et al., 2015). These four hypotheses are not mutually exclusive and therefore, should be considered simultaneously to study microevolutionary processes such as those addressed in LG.

On the other hand, phylogeographic and ecological evidence suggest a significant pattern of historical genetic isolation in Neotropical highland taxa (e.g., Navarro-Sigüenza et al., 2008; Ornelas et al., 2013; Mastretta-Yanes et al., 2018; Rocha-Méndez et al., 2019; Thom et al., 2021). Montane taxa are particularly sensitive to global climate change, as most species that inhabit the Neotropics have tracked their climate tolerances over time (Ramírez-Barahona & Eguiarte, 2013; Flantua & Hooghiemstra, 2018; Rahbek et al., 2019; Linck et al., 2021). Furthermore, glacial cycles appear to have shaped biogeographic and genetic diversity patterns in many taxa (Pielou, 1991; Hewitt, 1996; Hewitt, 2004). Therefore, Neotropical montane species are an interesting study model to test genetic isolation hypotheses from a LG perspective (Mastretta-Yanes et al., 2015).

The Mesoamerican montane forests (Morrone, 2014; Morrone, 2020) are distributed in the major mountain ranges from Mexico (southern Tamaulipas and southern Sinaloa) to the Nicaragua highlands at elevations ranging between 500 and 3,500 m (Hamilton et al., 1995; Webster, 1995). Given that the distribution and spatial configuration of these forests resembles an archipelago, bird species are allopatrically distributed and typically show genetically and phenotypically differentiated populations among different forest patches (Sánchez-González et al., 2008). Thus, Mesoamerican montane forests constitute an attractive study area for testing hypotheses of genetic variation being influenced by landscape drivers (Balkenhol et al., 2016). Mesoamerican montane forests and other ecosystems in the Neotropics remain understudied at microevolutionary scales using LG methods (Monteiro et al., 2019). The understanding of microevolutionary processes is fundamental to develop conservation strategies for threatened montane taxa and other Neotropical range-restricted endemics (Stattersfield et al., 1998).

One classic example of allopatric taxa in Mesoamerican highlands is the Chestnut-capped Brushfinch/Green-striped Brushfinch complex (*Arremon brunneinucha*/*A. virenticeps*, Aves: Passerellidae; hereafter *Arremon* bird complex, ABC). It is distributed from central Mexico to Peru at elevations of 400–3,500 m (Paynter, 1978; Howell & Webb, 1995). The ABC in northern Mesoamerica consists of two currently recognized species: *A. brunneinucha*, from cloud forests and includes five or six independent evolutionary lineages in the region (e.g., Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013), and *A. virenticeps*, which inhabits comparatively drier montane forests (Paynter, 1978; Howell & Webb, 1995). Although there is little morphological variation within this group (Parkes, 1954), phylogenetic studies have shown that its biogeographic history and evolutionary relationships are complex (Cadena et al., 2007; Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013; Buainain et al., 2022). Phenotypic studies suggested that allopatric populations of the chestnut-capped and the green-striped brushfinches are reciprocally monophyletic (Paynter, 1978). However, molecular phylogenetic hypotheses have shown that the green-striped group is embedded within a large group of chestnut-capped populations and is sister to population in the Sierra Madre del Sur (Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013; Buainain et al., 2022). A recent study combining phylogenetics and an ecological perspective has further suggested that these populations have distinct climate preferences, leading to segregation of their Grinnellian niches throughout their allopatric distribution (Moreno-Contreras et al., 2020).

Geographic isolation due to the presence of lowland barriers (IBB; Mayr, 1963) may not fully explain GD in allopatric populations, leaving a role for other landscape drivers of neutral genomic variation in montane species. Herein, using Mesoamerican ABC we tested genetic isolation hypotheses, employing a link-level analysis (distance-based approach) within an individual-based LG framework (Wagner & Fortin, 2013; Kierepka & Latch, 2015; DiLeo & Wagner, 2016; Peterson et al., 2019). Birds are an underrepresented model group in LG studies; this may reflect the underlying assumption that spatial genetic structure will be difficult to detect in birds due to their high vagility (Kozakiewicz et al., 2018). However, bird

species that disperse short distances (< 15 km) are relatively common (Manthey & Moyle, 2015, Castaño et al., 2019). Furthermore, birds have been shown to be highly sensitive to the availability of forest patches to which they can move (Kozakiewicz et al., 2018) or changing landscapes through time (Buainain et al., 2020; Moreira et al., 2020), providing a chance to study birds under a LG approach.

We predict that ABC will show relatively deep genetic differentiation due to low vagility, as expected in restricted ground-foraging species with sedentary habits in montane habitats (Janzen, 1967; Thom et al., 2021). Due to its ecological (climate-based) preferences (Moreno-Contreras et al., 2020), phylogenetic distinctiveness and historical isolation (Cadena et al., 2007; Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013; Buainain et al., 2022), we expect the Mesoamerican ABC to follow “dry refugia expectations” (Ramírez-Barahona & Eguiarte, 2013). We expect the genomic variation (measured here as individual-based GD) to support an IBR (H_1) pattern based on habitat suitability conditions and topography (Linck et al., 2021). Such a pattern may have shaped its contemporary genomic variation as previously documented in other tropical montane species (Mastretta-Yanes et al., 2018; Linck et al., 2021; Thom et al., 2021). We therefore aimed to determine the causes of the genomic structure in the ABC under several assignment methods using next-generation sequencing data (Nextera-tagmented, Reductively Amplified DNA; hereafter NextRAD). Then, we evaluated both the genomic differentiation and summary statistics based on genome-wide diversity. We also aimed to test the four different isolation hypotheses to determine which may explain the neutral genomic variation in ABC (including all sequenced samples; “regional-scale analysis”). In addition, because the spatial scale may influence LG inferences (Lee-Yaw et al., 2009), we analyzed the best sampled group separately, which includes *brunneinucha* individuals from the Sierra Madre Oriental (hereafter “fine-spatial scale” analysis). Accordingly, we employed a multi-criteria approach for model selection avoiding method-dependent results (Balkenhol et al., 2009). This study will shed light on the relative influence of different landscape variables (considering both past and contemporary conditions) in testing the various genetic isolation hypotheses (e.g., IBD, IBB, IBE, and IBR) in allopatric species. This will provide a

broader view of the ecological factors driving speciation processes in montane species.

Methods

Study area and sampling locations

We collected frozen or ethanol-preserved tissue (muscle, heart or liver) samples from 41 individuals of the ABC in northern Mesoamerica stored at the Museo de Zoología, Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM). Individuals were collected between June 1990 and December 2013 and grouped into 35 nonrandom locations encompassing the Mesoamerican distributional range of *Arremon,* including *A. virenticeps* (with the exception of the Sinaloan *A. virenticeps* population) and *A. bruneinucha* occurring in the main mountain systems of Mexico (Figure 1, Table S1): *A. b. apertus* (*n* = 2; Sierra de Los Tuxtlas: Veracruz), *A. b. brunneinucha* (*n* = 10; Sierra Madre Oriental), *A. kuehnerii* (*n* = 8; Sierra Madre del Sur: Guerrero), *A. b. suttoni* (*n* = 7; Sierra Madre del Sur: Oaxaca), *A. b. macrourus* (*n* = 6; Chiapas Highlands: Chiapas), and *A. virenticeps* (*n* = 6; Transmexican Volcanic Belt: Mexico City, Jalisco, and Michoacán). We extracted genomic DNA from preserved tissue samples using Proteinase K digestion, followed by isolation with the DNeasy Blood & Tissue Kit (Qiagen Inc., Valencia, CA, USA) according to manufacturer's protocols. Once the DNA was extracted, samples were sent for NextRAD sequencing.

NextRAD sequencing

NextRAD genotyping-by-sequencing libraries were generated and sequenced by SNPsaurus, LLC (Oregon, USA) following the Russello et al. (2015) protocol. These libraries yielded short sequences (~ 150 base pairs) flanked by engineered transposon cut sites, which results in thousands of loci distributed throughout the genome (Russello et al., 2015). This method requires only small amounts of DNA (< 50 ng; Russello et al., 2015), allowing the recovery of sequence data even from

samples with low DNA concentration (e.g., museum specimens, Llanes-Quevedo et al., 2022). Briefly, genomic DNA was fragmented using a Nextera reagent (Illumina, Inc; California, USA), which also attaches short adapter sequences to the ends of the fragments. The Nextera reaction was scaled to fragments of ~ 10 ng of genomic DNA. Fragmented DNA was then amplified with one of the primers matching the adapter, and then extended 10 nucleotides into the genomic DNA with the selective sequence GTGTAGAGCC. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer were efficiently amplified by PCR. Purified libraries were size-selected to 350–800 bp and sequenced on a HiSeq 4000 with one lane of 150 bp reads (Genomics Core Facility, University of Oregon).

De novo assembly and post-processing

Prior to *de novo* assembly, raw NextRAD reads were submitted for quality control in 'FastQC' v.0.11.9 (Andrews, 2019) and 'Trimmomatic' v.0.38 (Bolger et al., 2014). We used the *SLIDINGWINDOW* function (5:33), which scans with a 5-base wide sliding window, cutting when the average quality per base drops below a threshold of 33. We then used the *ILLUMINACLIP* function (1:30:10) to prune the adapters of the NextRAD sequencing. Specifically, we allowed a maximum of 2 mismatches, and seeds were extended and clipped when a score of 30 was reached for paired-end reads (about 50 bases) or a score of 10 was reached for single-ended reads (about 17 bases). We processed clean Illumina reads using the software pipeline 'ipyrad' v.0.9.26 (Eaton & Overcast, 2020) to generate *de novo* assemblies in the 'Miztli' supercomputer of the Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM. We explored how the matrix output was affected by the level of sequence similarity (clust_threshold parameter) and the minimum number of samples (min_samples_locus parameter) required for a given locus to be retained in the final data set. To do this, we generated a total of 14 matrices with a combination of the following parameters: 0.85, 0.87, 0.91, and 0.93 values for the clust_threshold parameter, and 4, 8, 12, 16, 24, and 32 values for the min_samples_locus parameter. We kept the ipyrad default depth-related

parameters. The matrix produced by the combination of the parameters clust_threshold = 0.85 and min_samples_locus = 16 yielded the highest number of SNPs, lowest percentage of missing data, and one of the lowest Euclidean GD values among the generated matrices, so it was chosen for use in the subsequent analyses (see Figure S2 in Supplementary Material).

During post-processing, we pruned and filtered single nucleotide polymorphisms (SNPs) from the NextRAD dataset using 'VCFtools' v.0.1.17 (Danecek et al., 2011) and 'PLINK' v.1.90b6.4 (Chang et al., 2015), only retaining those loci with < 25% missing data. We followed the recommendation of Linck & Battey (2019) to remove singletons (SNPs where the minor allele only occurs in a single individual) in favor of setting a minor allele frequency (MAF) threshold. Here, we set the MAC (minor allele count) threshold to 2 because different studies in Neotropical montane forest taxa have found that birds have experienced different demographic scenarios in the past, which may have influenced their allele frequencies (Ramírez-Barahona & Eguiarte, 2013; Flantua & Hooghiemstra, 2018; Rahbek et al., 2019). We also retained SNPs in approximate linkage equilibrium (hereafter LD; sliding window of 50 SNPs, the step size of 5 SNPs, VIF = 2). Because background selection can bias population structure analyses (Nordborg et al., 1996; Manel et al., 2003), outlier loci were identified and removed. We identified outlier loci using a principal component analysis (PCA) following a two-step procedure in the R-package 'pcadapt' v.4.3.3 (Luu et al., 2017). First, a PCA captures the genetic structure of the dataset. Next, the Mahalanobis distance of the z-scores on the first *K*-components (PCs = 6) of each locus detects the SNPs that are most strongly related to population structure (Luu et al., 2017). Here, we used a $q < 0.05$ in the R-library 'qvalue' v.2.24.0 (Storey et al., 2021) to identify outlier SNPs, defined as SNPs with significantly larger Mahalanobis distances than the SNPs related to population structure. Outliers were removed from the SNP dataset to retain only approximate neutral markers and avoid spurious results (Balkenhol et al., 2017). The percentage of missing data of the final NextRAD data set for each sample are shown in Supplementary Material 1.1 (*Collected samples, assembly parameters, and quantification of missing data*; Table S1).

Spatial data

The study area included all of the sampled locations plus a 50-km buffer (14°32'56.0" N–21°35'53.3" N and 105°21'23.8" W–91°06'09.6" W, WGS84 geographic projection; Figure 1) to avoid possible computational problems when optimizing resistance surface layers for LG analyses. Because sequenced samples came from museum specimens that did not coincide temporally with conventional GIS climate data (WorldClim), we used other GIS climate alternatives to test the IBE and IBR hypotheses. Bioclimatic layers representing current conditions were obtained from the CHELSA v.1.2 project (Karger et al., 2017). To assess the relative influence of past climate conditions on genomic variation, we downloaded past climate conditions representing the Last Glacial Maximum (hereafter LGM; 21,000 years ago) based on PMIP3 (Paleoclimate Modeling Intercomparison Project), from the Model for Interdisciplinary Research on Climate (MIROC-ESM) from CHELSA-LGM project (https://chelsa-climate.org/last-glacial-maximum-climate/). We also obtained a digital elevation model (DEM) from the HYDRO1k project (USGS-United States Geological Survey, 2021). All environmental layers were downloaded at high resolution (30 arc-second, approximately 1 × 1 km) with a WGS84 projection, then converted to projected coordinates using the Mexico ITRF2008 Lambert Conformal Conical projection (900 × 900 m resolution). We chose a resolution of 1 km due to the known limited dispersal in the *Arremon* species complex (Castaño et al., 2019), which may reflect the importance of landscape features impeding gene flow. Geoprocessing, manipulation, and visualization of spatial data were performed using ArcGIS v.10.4.1 (ESRI, Redlands, California, USA), QGIS v.3.16.9 (QGIS Development Team), and the following geospatial packages in R: 'dismo' v.1.3-3 (Hijmans et al., 2020), 'ggmap' v.3.0.0 (Kahle & Wickham, 2013), 'ggplot2' v.3.3.5 (Wickham, 2016), 'raster' v3.4-13 (Hijmans, 2021), 'rgdal' v.1.-23 (Bivand et al., 2021), 'rgeos' v. 0.5-5 (Bivand & Rundel, 2020), and 'sf' v.1.0-2 (Pebesma, 2018).

Population structure

We followed a “total evidence” approach to assign individuals to genetic groups (“*K*”) using different clustering methods (Cullingham et al., 2020). Based on the allopatric distribution of ABC in northern Mesoamerica (Howell & Webb, 1995), we explored genetic structure, testing $K \in \{2, 3, 4, 5, 6, 7\}$ in different model-based methods (GENELAND, TESS3, and STRUCTURE; see details in Supplementary Material 1.2, *Population structure analyses*). Using ‘GENELAND’ v.2.1.4 (Guillot et al., 2005), we inferred the number of genetic clusters in the complex using a Bayesian model under an MCMC scheme to locate genetic discontinuities using individual geo-referenced multilocus genotypes (François & Durand, 2010). We then used the ‘TESS3’ algorithm as implemented in the R-package ‘tess3r’ v.1.1.0 (Caye et al., 2016). TESS3 attempts to minimize the Walhund effect by incorporating local dependencies in their model. Unlike GENELAND, Voronoi cells in TESS3 are not associated with territories, but with ‘individuals’ (Caye et al., 2016). As a complementary approach to spatial assignment tests, we explored the genome-wide SNP data using the individual-based Bayesian clustering approach in ‘STRUCTURE’ v.2.3.4 (Pritchard et al., 2000).

We employed distance-based methods as a visualization tool for investigating and summarizing population structure identified by model-based methods. PCA (Menozzi et al., 1978) was performed using the *dudi.pca* function of ‘ade4’ v.1.7-17 (Dray & Dufour, 2007), centering but not scale-transforming the PCA to preserve differences in the magnitude of the genetic importance among environmental variables. We also carried out a discriminant analysis of principal components (DAPC), which is a multivariate method free of Hardy-Weinberg and linkage disequilibrium assumptions (Jombart et al., 2010). We considered the clusters identified by the “total-evidence” approach as *a priori* within the *dapc* function, while the optimal number of PCs to retain was optimized using the function *xvalDapc* in ‘adegenet’ v.2.1.4 (Jombart, 2008).

We followed a spatial population genetics approach to understand whether the population assignments mentioned above are true genetic clusters or the result of a geo-genetic map. To visualize the spatial arrangement of the sequenced samples,

we generated a geo-genetic map of the Mesoamerican ABC using 'SpaceMix' v.2.3.1 (Bradburd et al., 2016). We further investigated the spatial autocorrelation of the allele frequencies in our study area by performing a spatial principal component analysis (sPCA, Jombart et al., 2008) in the R-package 'adegenet' v.2.1.4 (Jombart, 2008), which does not require genomic data to be in Hardy-Weinberg or linkage equilibrium.

Landscape genomics analyses

To quantify individual-based GD, we used two response variables. First, as a measure of similarity between the multiple locus genotype of two individuals, we calculated the proportion of alleles (*Ps*; Bowcock et al., 1994). The formula denoting the following GD is: $Ps = \frac{\sum_u S}{2u}$, where the number of shared alleles (*S*) is summed over loci (*u*). A distance measure between pairs of individuals was calculated as (1 – *Ps* = D_{PS}), estimated using the R package 'adegenet' (Jombart, 2008). Second, we used Euclidean distance from a PCA (see *Population structure*), retaining the first 10 axes ($Euclidean_{PCA\text{-}10\ axes}$) for accounting 56.97% of variation, which contemplates the inherent complexity of codominant data, performed with the *dist* function in 'stats' v.3.6.2 (R Development Core Team, 2021). Due to the sample size, the "fine-spatial scale" analysis was based on the first 9 axes ($Euclidean_{PCA\text{-}9\ axes}$).

We quantified the basic population summary statistics, such as genetic differentiation (F_{ST}; Weir & Cockerham, 1984) between genetic clusters using the *genet.dist* function, and genomic diversity measures such as observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{IS}) with the *basic.stats* function, and allelic richness (*Ar*) using *allelic.richness* function, all available in 'hierfstat' v.0.5-7 (Goudet & Jombart, 2020). To visualize the spatial arrangement of the sPCA and inter-individual pairwise response variables (D_{PS} and $Euclidean_{PCA\text{-}10\ axes}$), interpolation maps were generated with the R-package 'akima' v.0.6-2.2 (Akima & Gebhardt, 2021).

Landscape distance matrices

We compared the ability of nine competing isolation models to predict neutral GD in the ABC: one null model included only isolation by distance (IBD), and eight candidate models representing all of the alternative hypothesis while controlling for IBD. Specifically, the eight alternative models consisted of one model for isolation by barrier (IBB), two models for isolation by environment (IBE), and five models for isolation by resistance (IBR) considering a least-cost path (hereafter LCP), and random-walk commute times or circuit-theory. In addition, for the fine-scale analysis (SMOr samples only), eight candidate models were tested (we excluded the IBB model).

The IBD matrix was built from the Euclidean distance estimated from the geographic coordinates of sampled locations using 'stats' v.3.6.2 (R Development Core Team, 2021). Regarding the IBB hypothesis, we used the number of times a straight line between two samples crosses potential barriers to estimate a pairwise discretized distance (Balkenhol et al., 2017). We defined barriers as areas representing lowland ecoregions. In the case of the IBE hypothesis, we generated two candidate models based on a global climatic model including non-correlated bioclimatic variables (r < 0.7) used in the construction of ENMs (see below), and a single climatic model including only BIO12 (annual precipitation), since it has been shown that BIO12 influences the segregation of climatic niches within this complex (Moreno-Contreras et al., 2020), therefore affecting the breeding biology throughout its distribution (Rice et al., 2020). This abiotic driver has also influenced distributional movements of Neotropical highland taxa in past and current times (Ramírez-Barahona & Eguiarte, 2013; Flantua & Hooghiemstra, 2018). Therefore, we expected that precipitation may explain the genetic variation of this complex. Both climatic datasets for the IBE were converted into a Euclidean distance following the same procedure as for the IBD model.

IBR was tested using least-cost path algorithms (LCP; seeking single optimal pathways), and random-walk commute time or using circuit-based resistance distances (tracking multiple pathways). To test IBR models, we generated ecological

niche models (ENM) representing current and past conditions during LGM (MIROC-ESM), retaining only uncorrelated bioclimatic variables ($r < 0.7$). Further details on the elaboration of the ENM, fine-tuning process, as well as past projections are provided in the Supplementary Material 1.3 (*Elaboration of ecological niche models for IBR tests*). Once each ENM was obtained, we constructed five candidate models representing the different IBR hypotheses (see details in *Elaboration of resistance distance matrices* of Supplementary Material 1.4).

Model selection

Prior to analyses, all predictor variables were standardized to a mean equal to zero and a standard deviation equal to one. Several studies with real or simulated data have emphasized the importance of using an optimal combination of statistics to test hypotheses of genetic isolation (Balkenhol et al., 2009; Kierepka et al., 2015; Shirk et al., 2018; Peterman & Pope, 2021). Most LG studies have used information theory metrics (e.g., *AIC* and its variants) for model selection (Balkenhol et al., 2016). However, because candidate models can often be equally parsimonious and therefore indistinguishable, we followed a multi-criteria approach for model selection using metrics derived from maximum-likelihood population effects models (MLPE, AIC_c and cR^2, Clarke et al., 2002) and multiple regression of distance matrices (MRDM, R^2, Legendre et al., 1994; see details in *Model Selection* of Supplementary Material 1.5). To determine if our results were consistent across spatial scales (Lee-Yaw et al., 2009), we also analyzed separately the best sampled group (*brunneinucha* of the SMOr) using only MLPEs.

Results

Pre- and post-processing of NextRAD

The 41 samples sequenced had an average of 3,231,536 raw reads; two samples were excluded because the FastQC review showed poor quality reads. After removing adapters and poor-quality samples, we obtained an average of 3,092,344

± 1,681,393 (SD) raw reads. An average of 15,463 ± 6,992 (SD) loci (*n* = 39 samples) were assembled after demultiplexing, with a minimum of 1,123 and a maximum of 23,258 loci. Sequence processing through the 'ipyrad' pipeline resulted in 315,693 high-quality SNPs for the 39 remaining samples when using the selected parameters for our genomic data set. This assembly matrix had one of the lowest Euclidean GD values ("run 14," see Figure S2-S3) as expected, as having a high percentage of missing data may increasing Euclidean GD between samples. In addition, the genomic data had a high number of SNPs and a relatively low percentage of missing data (Table S2). The mean genotype depth for the selected unfiltered ipyrad matrix was 19.65 (SD = 19.41), with some samples having extreme depth values of 64.54 (e.g., YAGILA_08) and 0.25 (e.g., BMM_804) (Figure S1A). The mean percent missing data for the unfiltered data set was 51.31 (Figure S1B). The average depth of each sample was 10.39 (Fig. 1). Regarding bioinformatic post-processing, we retained 5,953 SNPs after applying the 25% missing data filter, which was reduced to 3,276 SNPs after excluding genomic data with MAC < 2. This was further reduced to 2,663 SNPs after filtering for LD. Multivariate tests using 'pcadapt' identified 29 outlier loci, which we excluded before further analyses, leaving a total of 2,634 neutral and unlinked SNPs for use in the LG framework. The amount of missing data for the final NextRAD set was 20.70% ± 19.59% (SD) per individual and 20.70% ± 3.23% per site (SD). See Table S1for further details.

Population structure

We selected the sixth independent run in GENELAND since it had the highest posterior probability values (mean log PP = 25,899.45; Figure S4A), and it indicated that the Mesoamerican ABC could be split into $K = 5$ populations (Figure 2A–B). The frequency of visits of the chain for the sixth run when $K = 6–7$ was very low (Figure S4B–S4C). On the other hand, the fine-tuned TESS3 algorithm indicated an optimal $K = 6$ (Figure 2C–D), with a cross-validation entropy value of 0.28 (Figure S4D).

We found the highest values of ln Pr($X|K$) and ΔK (Figure S5A, Figure S5D) when using the CAF model (independently of the ancestry prior) in genome-wide

individual-based assignment tests in STRUCTURE. The $\widehat{K}_{\text{Pritchard}}$ and $\widehat{K}_{\text{Evanno}}$ estimated *K* = 6 (CAF model) and *K* = 2 (UAF model) using a default prior in both cases. The alternative ancestry prior was *K* = 5 (CAF model), however this prior worked well and yielded accurate individual assignments only when sampling was unbalanced, thus corroborating the same optimal *K* as was obtained in GENELAND. This was important since the *apertus* population was poorly represented (*n* = 2), while most of the collected samples belonged to the *brunneinucha* population (*n* = 10). Summary statistics for STRUCTURE are shown in Supplementary Material (Figure S5A–D). Bar plots of STRUCTURE results (CAF model with alternative ancestry prior) displaying admixture values are shown in Supplementary Material (Figure S6A–D).

According to GENELAND (Figure 2A–B) and STRUCTURE, using *K* = 5 (Figure S6C) in the "total evidence" approach, grouped the samples in the following clusters: (1) *apertus* (the chestnut-capped population without pectoral band) from Sierra de Los Tuxtlas (SLT); (2) *brunneinucha* from the Sierra Madre Oriental (SMOr), (3) *kuehnerii* and *suttoni* from the Sierra Madre del Sur (SMS); (4) *macrourus* from the Chiapas Highlands (CHH); and (5) the phenotypically different group *virenticeps* from the Transmexican Volcanic Belt (TVB; Figure 2A–B, Figure S6C). Unlike GENELAND and STRUCTURE, TESS3 separated individuals of *macrourus* into two clusters: one included one individual from the northernmost portion of the Chiapas Highlands (NoCH), while the other included the rest of the individuals distributed in the southernmost Chiapas Highlands (SoCH; Figure 2C–D).

Although they are not directly comparable, the PCA method (Figure 3A) was outperformed by the DAPC method (Figure 3B) in that it achieved a clearer variance difference among the genetic clusters identified by the model-based methods. According to the PCA, 14.8% of the genetic structure variance was explained by axis 1 and 9.4% by axis 2 (Figure 3A), while the first two axes of the DAPC explained 46.4% and 32.5% of the variation (Figure 3B). The discriminant function axis 1 showed the separation of SLT and TVB from the rest of the clusters. Discriminant function axis 2 showed separation of SMS, although there was considerable overlap

between the SMOr and CHH clusters. SpaceMix provided evidence that *macrourus* individuals are close to each other on the geo-genetic map, and none of the ellipses from any group overlapped with individuals from another group (Figure 3C–D). The estimated long-distance admixture proportions were minimal for all samples, and the model with these proportions fixed at 0 was not supported over one where they were allowed to vary (Bayes Factor < 0.0001), which suggests that patterns of relatedness in these genomic data were well-described by an IBD model.

Bivariate plots showing normal PCA and sPCA outputs for the three first axes are found in Supplementary Material (Figure S7A–F). For the sPCA, we found significant global genetic structure within the study area ($p < 0.05$, $n = 39$) as noted by the positive eigenvalues that explained the observed structure. The first global (positive) sPCA axis identified a genetic break (Figure S8A) between *brunneinucha* (SMOr) and *macrourus* (CHH). For the second global sPCA axis (Figure S8B), most of the clusters showed negative spatial autocorrelation, except for individuals of *macrourus*, which had positive values of spatial genetic variation. The third global sPCA axis (Figure S8C) suggested a genetic break for individuals belonging to *virenticeps* (TVB) located in the westernmost part of our study area, which had negative autocorrelation. The third global sPCA axis also showed SMS individuals having positive autocorrelation (Figure S8C).

Landscape genomic analysis

The different genetic clustering methods (GENELAND, STRUCTURE, and TESS3) applied in this study, in conjunction with spatial patterns found by SpaceMix and sPCA, were largely congruent with previous phylogeographical hypotheses and distributional evidence. Given this finding, we decided to base our LG analysis on $K = 5$. Regarding inter-individual GD for D_{PS} (Figure S8D), *macrourus* (CHH) individuals presented the highest pairwise GD, while the least genetically distant individuals correspond to *virenticeps* (TVB), *apertus* (SLT), and *kuehnerii* (SMS). In the case of $Euclidean_{PCA\text{-}10\ axes}$ (Figure S8E), we found a higher GD for individuals of *macrourus* (CHH), *virenticeps* (TVB) and *apertus* (SLT), which closely resembled

the spatial genetic population structure obtained in GENELAND (Figure 2A). Genetic differentiation based on genetic groups determined that the comparison between SMS and TVB had the lowest F_{ST} values, while the comparison between SLT and TVB had the highest F_{ST} values (Figure 4). CHH had the highest and SLT the lowest values of H_O and H_E (Table 1). The cluster with the highest values of F_{IS} was *virenticeps* from the TVB (Table 1). The genetic population with the highest values of *Ar* was CHH (Table 1).

Model selection

Exploratory analyses based on Mantel tests for the null hypotheses (IBD) in both GD metrics were statistically significant ($p < 0.05$; Figure S9A–B). IBD was better explained by D_{PS} (Mantel r = 0.622, $p < 0.05$) than $Euclidean_{PCA\text{-}10\ axes}$ (Mantel r = 0.476, $p < 0.05$). Maps showing raster pixel values of some alternative hypothesis such as IBB, IBE_{bio12}, IBR_{tlcp}, and ENM during LGM-MIROC conditions are found in Supplementary Material (Figure S10A–D). Some of the resistance surfaces optimized in ResistanceGA and radish are available in Supplementary Material (Figure S11A–D). Partial Mantel tests and their respective values for each alternative hypothesis were plotted in both D_{PS} and $Euclidean_{PCA\text{-}10\ axes}$ (Figures 5–6). As shown in the partial Mantel tests, the LGM-MIROC conditions model (IBR_{LGMrad} | IBD) optimized using the radish algorithm had the highest Mantel r values when the GD response was D_{PS} (Figure 5H). On the contrary, the IBR model based on the topographic least-cost paths (IBR_{tlcp} | IBD) showed the highest Mantel r values when the GD response was $Euclidean_{PCA\text{-}10\ axes}$ (Figure 6D).

We rejected the null hypothesis of IBD under a multi-criteria approach for model selection in the LG analyses that considered the Mesoamerican ABC as a whole ("regional-scale analysis"). This pattern was consistent regardless of which specific GD response variable was used in the regression analysis and regardless of whether we controlled for population structure (Tables 2–5, Table S3, Table S4). However, we only rejected the null hypothesis in the case of $Euclidean_{PCA\text{-}9\ axes}$ (Table S5 and Table S6) in the fine-scale analysis (only SMOr samples). In the regional-scale

analysis (including all samples), the models that controlled for population structure showed the lowest AIC_c values (Table 2 and Table 4). The candidate model that best balanced parsimony and statistical accuracy while explaining the neutral genetic variation for D_{PS} (Table 2–3) was the IBR model (optimized in ResistanceGA) corresponding to ENM current conditions (IBR_{PRErad} | IBD; AIC_c = -4647.971, cR^2 = 0.700, R^2 = 0.391). Regarding the $Euclidean_{PCA\text{-}10\ axes}$ (Table 4–5), the IBR model based on topographic least-cost paths conditioned for IBD (IBR_{tlcp} | IBD) had the best balance between parsimony and the greatest explained variation (AIC_c = 3304.2, cR^2 = 0.444, R^2 = 0.327). MLPEs that did not control for population structure also found that IBR models had the lowest AIC_c values (see detailed accounts in Table S3 and Table S4).

Discussion

The Mesoamerican montane forests are considered biodiversity hotspots harboring an extremely diverse and heterogeneous mix of tropical and temperate species with different phylogeographic trajectories (Ornelas et al., 2013). Most of these phylogeographic patterns have emerged from demographic scenarios during glacial climate fluctuations (Ramírez-Barahona & Eguiarte, 2013; Flantua & Hooghiemstra, 2018; Rahbek et al., 2019).

In the present study, we used reduced-representation genome sequencing methods to identify patterns of neutral genomic structure and variation related to different isolation hypothesis (IBD, IBB, IBE, and IBR) in a Neotropical bird species complex. Although the evolutionary relationships in the Mesoamerican ABC are challenging to characterize, our study contributes to a better understanding of this issue through the lens of LG. Amplifying more (and more variable) loci is likely to increase the power of LG inferences (Landguth et al., 2012; Teske et al., 2018). As such, our LG estimates using NextRAD loci should offer benefits for describing the genomic variation in this study group. In accordance with our expectations, the ABC exhibited well-delimited genomic structuring across Mesoamerican montane forests (K = 5, Figure 2A–D; Figure 3A–D; Figure S6C), coinciding with previous phylogeographic

studies based on nuclear and mitochondrial DNA (Cadena et al., 2007; Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013), and Ultra Conserved Elements (UCE; Buainain et al., 2022). Unlike previous evolutionary hypotheses, our population structure analyses suggest a close genetic relationship between *kuehnerii* (western group) and *suttoni* (eastern group) from the Sierra Madre del Sur (see Figure 2A–D; Figure 3A–B; Figure S6C), with little GD at the individual level (Figure S8D–E). However, only STRUCTURE results displayed low admixture values for a sixth cluster containing all *suttoni* samples (see Figure S6D), and TESS3 suggested one cluster (NoCH) containing only one sample (see Figure 2C–D). For both cases, these findings may reflect either phylogenetic distances rather than admixture fractions (Lawson et al., 2018), or a "ghost" cluster due to IBD (Frantz et al., 2009). The latter explanation is supported by SpaceMix analyses, in which these two groups did not overlap (Figure 3C–D).

We predicted that the ground-foraging and sedentary habits of the ABC would show low genome-wide genomic diversity but high genomic differentiation at individual- (Figure S8D–E) and population-levels (Figure 4), as suggested by Janzen's hypothesis (Janzen, 1967). This finding suggests that genomic structure of the group leads directly to a model describing the "dry refugia expectations" (Ramírez-Barahona & Eguiarte, 2013). According to genomic diversity values, the genome-wide estimates obtained herein (H_E = 0.107, NextRAD = 2,634 SNPs) for Mesoamerican ABC were lower than those of the congeneric Pectoral Sparrow (*Arremon taciturnus*; H_E = 0.335, ddRAD = 647 SNPs; Buainain et al., 2020), but higher than other passerines, such as the Yucatan Jay (*Cyanocorax yucatanicus*), which is a cooperative breeder and canopy forager (H_E = 0.0075, ddRAD = 1,649 SNPs; Termignoni-García et al., 2017). These comparisons support the results of Burney & Brumfield (2009), who found that understory birds showed significantly higher genetic diversity than canopy birds. Alternatively, it has been suggested that other ecological and evolutionary differences such as dispersal propensity and rates of diversification are associated with genetic divergence between understory and canopy birds (Burney & Brumfield, 2009). Our genome-wide diversity estimates were unsurprising, since ABC usually have low effective population sizes with little support

for hypotheses of recent population expansion (Navarro-Sigüenza et al., 2008) and high levels of allele fixation due to restricted distribution (e.g., *apertus* of the SLT cluster). It is though that the populations with the highest genomic diversity and effective population sizes (e.g., *macrourus* of the CHH cluster) may be better able to tolerate a wide range of environmental pressures, including climate extremes, parasites, and novel pathogens (Eo et al., 2011).

Our study provides partial support for the “moist forest” rather than the “dry forest” expectation (Ramírez-Barahona & Eguiarte, 2013) for this Neotropical taxon. Admixture values were very low, and the Bayes Factor support were also seemingly very low; this shows no significant admixture across all Mesoamerica *Arremon* samples, suggesting that patterns of relatedness in genomic data of ABC may be well-described by a model of IBD alone, according to SpaceMix. This is also supported by the presence of spatial autocorrelation in the allele frequencies obtained on the three axes of the sPCA (Figure S8A–C), as well as the statistical significance and positive relationships found in the Mantel tests for the IBD models (Figure S9A–B). Furthermore, our IBD analysis (regression-based approach) indicates that this Neotropical bird displays a case IV pattern at the regional scale (Hutchinson & Templeton, 1999), in which both GD and its confidence intervals increased with the geographic distance between individuals (Hutchinson & Templeton, 1999). In general terms, there is a significant lack of regional equilibrium of IBD in this group. Consequently, given that the ABC showed significant genomic structure, we would not expect a significant IBD pattern under the “dry refugia hypothesis” (Ramírez-Barahona & Eguiarte, 2013). However, it is also possible that genetic distance/differentiation reaches the equilibrium state of IBD (case I pattern; Hutchinson & Templeton, 1999) relatively quickly (Slatkin, 1993) under colonization dynamics out of a single refugia (Ramírez-Barahona & Eguiarte, 2013). According to the literature, a pattern consistent with regional equilibrium could form across a region if conditions for dispersal remain sufficiently stable (e.g., refugial zones) across the region for enough time (Hutchinson & Templeton, 1999).

To ensure that our LG inferences would not be method- or scale-dependent, we used a wide range of statistical approaches to reject the null hypothesis (IBD). Our MLPEs-based regression approach supported IBR models (either LCP and random-walk commute times) as the main predictors shaping GD in this montane and sedentary species complex. Specifically, contemporary suitable habitat conditions ("IBR_{PREga}") and topographic complexity ("IBR_{tlcp}") appear to be important landscape drivers for Mesoamerican ABC, when controlling for IBD. The finding was mostly consistent across different spatial scales, considering both the "regional-scale analysis" of the ABC (*n* = 39 samples, 741 pairwise comparisons) or only *brunneinucha* samples of SMOr (*n* = 10 samples, 45 pairwise comparisons). That pattern was also congruent either regardless of whether population structure was controlled for (with: Table 2, Table 4; without: Table S3–S4). Therefore, the interconnection of *Arremon* populations is due to the reduced effect of environmental resistance in the Mesoamerican highlands. Our study is in line with the findings of Thom et al. (2021), who found that the effective distance ("landscape resistance") of suitable contemporary conditions was one of the most important predictors of genomic variation in montane taxa in the Atlantic Forest.

Unexpectedly, neither discrete-based entity lowland vicariant barriers (IBB) nor the null model (IBD) or IBE predicted GD in this species complex. It has long been though that geographic isolation (Mayr, 1963) represented by vicariant barriers (IBB) promoted GD of allopatric species. However, an important environmental factor conditioning the breeding in this taxon is the pronounced seasonal rainfall across Mesoamerica (Rice et al., 2020). Nonetheless, according to our results, contemporary local climatic conditions (IBE, measured as Euclidean distances) do not explain the overall neutral genomic variation in the studied group, although IBE has been reported as significant for non-montane taxa like the Yucatan Jay (*Cyanocorax yucatanicus*; ddRAD = 1,649 SNPs; Termignoni-García et al., 2017). In comparison to IBB, IBR surfaces displayed high resistance values (or low conductance values), emerging as a continuous barrier across lowland dry regions such as the Isthmus of Tehuantepec, the Balsas Depression, and the Chiapas Central Depression (Figure S11A). These environmentally different but contiguous

areas (areas presenting high resistance values) probably disfavor movement of taxa with low vagility, ground-foraging habits, and specific habitat preferences, like the *Arremon* species. Thus, isolation of populations would be a consequence of the expansion of highly unsuitable habitats, such as tropical deciduous forest on hillsides, semideciduous forest in lowland valleys and along stream beds, and open vegetation. Therefore, the composition and structure of the intervening lowland landscape mosaics may determine the permeability to movements in this species group. According to ENM predictions in the present, BIO1 (annual mean temperature) was the most important bioclimatic variable, and BIO12 was the second most important. This finding contrasts with natural history observations at smaller spatial scales, where it is asserted that precipitation is a main abiotic driver governing southern chestnut-capped populations (e.g., *suttoni*) distribution (Rice et al., 2020). However, it partially coincided with the "moist forest" expectation that there will be little influence of precipitation on Mesoamerican highland taxa. Similarly, the ENMs projected onto LGM-MIROC conditions (Figure S10D) predicted areas with high environmental suitability in the lowlands, suggesting in part, past range expansions and population connectivity in accordance with the "moist forest" hypothesis (Ramírez-Barahona & Eguiarte, 2013). The only exception to this trend were areas in the westernmost portion of the TVB ecoregion, where the green-striped population is located. This distribution is more similar to expectation of the "dry refugia" hypothesis, which predicts montane taxa with mid-elevation distribution and more stable climate areas, like the documented elevational range for *virenticeps* (Paynter, 1978; Howell & Webb, 1995).

Partial Mantel tests (Figure 4F and Figure 4H) and MRDMs (Table 3) indicated that past environmental conditions may be a strong promoter (though not the best model) for GD in the Mesoamerican ABC. Although it has been suggested that these methods are not entirely reliable for model selection (Balkenhol et al., 2009; Franckowiak et al., 2017; Shirk et al., 2018), prior LG studies in Neotropical lowland birds such as the Altamira Oriole (*Icterus gularis*; ddRAD = 11,873 SNPs; Moreira et al., 2020) and Pectoral Sparrow (*Arremon taciturnus*; ddRAD = 647 SNPs; Buainain et al., 2020), or long-lived montane plants (*Juniperus monticola*; ddRAD = 2,925

SNPs; Mastretta-Yanes et al., 2018) have suggested that IBR models corresponding to past environmental conditions are the primary predictor of GD. A previous study using partial Mantel tests simulations suggested that the effect of past landscapes on GD may rapidly disappear due to high rates of long-distance dispersal (Landguth et al., 2010), which are thought to be common in birds (Kozakiewicz et al., 2018). However, in species with short-range dispersal, IBD will be more pronounced at the regional scale (Landguth et al., 2010; Blair et al., 2012). While the effective resistance distances during the LGM-MIROC conditions could explain GD, the accumulated effect due to contemporary habitat suitability conditions and topographic complexity on genomic variation likely promoted more significant responses for the GD of the Mesoamerican brushfinches. These historical events include demographic movements, possibly during global climate fluctuations in the past (Ramírez-Barahona & Eguiarte, 2013; Flantua & Hooghiemstra, 2018), climatic niche tracking or segregation (Moreno-Contreras et al., 2020), geological and tectonic complexity (Vázquez-Selem & Heine, 2011; Mastretta-Yanes et al., 2018), and phylogenetic distinctiveness (Navarro-Sigüenza et al., 2008; Sánchez-González et al., 2008; Ornelas et al., 2013).

In conclusion, our findings do not refute either of the two hypotheses explaining genetic consequences for Neotropical taxa during LGM. Rather, our study provides insights into an alternative model for explaining genomic diversity for montane taxa in Mesoamerica (including both cloud forest and temperate forest populations). The idea of two opposing hypotheses—the “dry refugia” vs. “moist forests” hypotheses (Ramírez-Barahona & Eguiarte, 2013)—may oversimplify the divergence processes in Mesoamerican highland taxa with low vagility and ground-foraging habits. Indeed, previous studies have highlighted the need for alternative models to describe the complex scenarios of genetic variation in montane taxa (Ornelas & González, 2014; Flantua & Hooghiemstra, 2018). We found that relatively old montane taxa with Central America origins (Buainain et al., 2022) that diverged over past glacial cycles (even before recent glacial maxima), might have expanded their geographic distributions into the Mesoamerican lowlands during the LGM. Both historical isolation and divergence are imprinted in a strong genomic structure between

populations associated with different mountains ranges. Furthermore, this allopatric bird complex shows a significant IBD (geographical distance) following a case IV pattern. In this pattern, gene flow is more effective at shorter distances of geographic separation, and genetic drift is more influential at greater distances of geographic separation (Hutchinson & Templeton, 1999). When controlling for IBD, an emergent pattern of IBR has been reinforcing the historical divergence of the Mesoamerican ABC in the present. The latter genetic isolation hypothesis clearly supports a pattern of conserved niche-tracking of suitable habitat conditions and topographic complexity through glacial-interglacial dynamics (Mastretta-Yanes et al., 2018; Linck et al., 2021; Thom et al., 2021). Finally, our study provides a wider picture of microevolutionary processes shaping genomic structure and variation in allopatric species of the Neotropics. This information will not only be relevant to molecular ecologists and evolutionary biologists, but also to conservationists seeking to preserve protected natural areas from a functional connectivity perspective (Balkenhol et al., 2017).

Acknowledgments

This paper is a partial fulfillment for obtaining a Doctoral degree in the Posgrado en Ciencias Biológicas, UNAM of IM-C, who received a Doctoral scholarship provided by CONACyT (number 749969). M. E. Muñiz (Facultad de Ciencias, UNAM) and L. Márquez-Valdelamar (Instituto de Biología, UNAM) provided access and logistical support for the laboratory work. L. Vázquez-Reyes, M. Grosselet, and J. Carmona-Victoria kindly provided photos for Figure 1. We acknowledge the helpful comments from L. Kiere, E. Vázquez, J. P. Jaramillo, L. Eguiarte, F. Rodríguez, T. Garrido, R. Munguía-Steyer, and three anonymous reviewers. G. Bradburd (Michigan State University) provided help for SpaceMix. A. Gordillo-Martínez provided useful information on *Arremon* specimens hosted at Museo de Zoología, UNAM. Financial support was obtained from PAPIIT (IN 214621 to AGN-S). Bioinformatic analyses were performed using the supercomputer ‘Miztli’ through DGTIC-UNAM (LANCAD-UNAM-DGTIC-402 to AGN-S and AL-Q). Finally, IM-C appreciates the support provided by the Moreno-Contreras family during the elaboration of this work.

References

Akima, H., & Gebhardt, A. (2021). *akima: Interpolation of irregularly and regularly spaced data*. R package version 0.6-2.2. https://CRAN.R-project.org/package=akima

Andrews, S. (2019). FastQC: a quality control application for FastQ files (Version 0.11.9). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Balkenhol, N., Waits, L. P., & Dezzani, R. J. (2009). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, *32*(5), 818–830. https://doi.org/10.1111/j.1600-0587.2009.05807.x

Balkenhol, N., & Fortin, M.-J. (2016). Basics of study design: heterogeneity and genetic variation for landscape genetic studies. In N. Balkenhol, S. A. Cushman, A. T. Storfer, & L. P. Waits (Eds.), *Landscape genetics: concepts, methods, applications* (1st ed, pp. 58–76). John Wiley & Sons.

Balkenhol, N., Cushman, S. A., Storfer, A. T., & Waits, L. P. (Eds). (2016). *Landscape genetics: concepts, methods, applications* (1st ed). John Wiley & Sons.

Balkenhol, N., Dudaniec, R. Y., Krutovsky, K. V., Johnson, J. S., Cairns, D. M., Segelbacher, G., Selkoe, K. A., von der Heyden, S., Wang, I. J., Selmoni, O., & Joost, S. (2017). Landscape genomics: Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In O. P. Rajora (Ed.), *Population genomics concepts, strategies and approaches* (pp. 261–322). Springer.

Bivand, R., & Rundel, C. (2020). *rgeos: Interface to geometry engine - open source ('GEOS')*. R package version 0.5-5. https://CRAN.R-project.org/package=rgeos

Bivand, R., Keitt, T., & Rowlingson, B. (2021). *rgdal: Bindings for the 'geospatial' data abstraction library*. R package version 1.5-23. https://CRAN.R-project.org/package=rgdal

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T., Walker, F. M., Landguth, E., Cushman, S., Murphy, M., Waits, L., & Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, *12*(5), 822–833. https://doi.org/10.1111/j.1755-0998.2012.03151.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, *368*(6470), 455–457. https://doi.org/10.1038/368455a0

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2016). A spatial framework for understanding population structure and admixture. *PLoS Genetics*, *12*, e1005703. https://doi.org/10.1371/journal.pgen.1005703

Bradburd, G. S., & Ralph, P. L. (2019). Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics*, *50*, 427–449. https://doi.org/10.1146/annurev-ecolsys-110316-022659

Buainain, N., Canton, R., Zuquim, G., Tuomisto, H., Hrbek, T., Sato, H., & Ribas, C. C. (2020). Paleoclimatic evolution as the main driver of current genomic diversity in the widespread and polymorphic Neotropical songbird *Arremon taciturnus*. *Molecular Ecology*, *29*(15), 2922–2939. https://doi.org/10.1111/mec.15534

Buainain, N., Ferreira, M., Avendaño, J. E., Cadena, C. D., Faircloth, B. C., Brumfield, R. T., Cracarft, J., & Ribas, C. C. (2022). Biogeography of a neotropical songbird radiation reveals similar diversification dynamics

between montane and lowland clades. *Journal of Biogeography*, *49*(7), 1260–1273. https://doi.org/10.1111/jbi.14379

Burney, C. W., & Brumfield, R. T. (2009). Ecology predicts levels of genetic differentiation in Neotropical birds. *The American Naturalist*, *174*(3), 358–368. https://doi.org/10.1086/603613

Cadena, C. D., Klicka, J., & Ricklefs, R. E. (2007). Evolutionary differentiation in the Neotropical montane region: Molecular phylogenetics and phylogeography of Buarremon brush-finches (Aves, Emberizidae). *Molecular Phylogenetics and Evolution*, *44*(3), 993–1016. https://doi.org/10.1016/j.ympev.2006.12.012

Castaño, M. I., Cadena, C. D., & Avendaño, J. E. (2019). Home-range size of an Andean bird: Assessing the role of physical condition. *Biotropica*, *51*(4), 591–599. https://doi.org/10.1111/btp.12673

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. https://doi.org/10.1111/1755-0998.12471

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, *7*(3), 361–372. https://doi.org/10.1198/108571102320

Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct *K* value using the Δ*K* method: When does *K* = 2? *Molecular Ecology*, *29*(5), 862–869. https://doi.org/10.1111/mec.15374

Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, *168*(4), 486–499. https://doi.org/10.1086/506976

Cushman, S. A., McRae, B. H., & McGarigal, K. (2016). Basics of landscape ecology: An introduction to landscapes and population processes for landscape geneticists. In N. Balkenhol, S. A. Cushman, A. T. Storfer, & L. P. Waits (Eds.), *Landscape genetics: concepts, methods, applications* (1st ed, pp. 11–34). John Wiley & Sons.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DiLeo, M. F., & Wagner, H. H. (2016). A landscape ecologist's agenda for landscape genetics. *Current Landscape Ecology Reports*, *1*(3), 115–126. https://doi.org/10.1007/s40823-016-0013-x

Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04

Dyer, R. J. (2015). Is there such a thing as landscape genetics? *Molecular Ecology*, *24*(14), 3518–3528. https://doi.org/10.1111/mec.13249

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Eo, S. H., Doyle, J. M., & DeWoody, J. A. (2011). Genetic diversity in birds is associated with body mass and habitat type. *Journal of Zoology*, *283*(3), 220–226. https://doi.org/10.1111/j.1469-7998.2010.00773.x

Flantua, S., & Hooghiemstra, H. (2018). Historical connectivity and mountain biodiversity (with special reference to the Northern Andes). In C. Hoorn, A. Perrigo, & A. Antonelli (Eds.), *Mountains, Climate and Biodiversity* (pp. 171–185). John Wiley & Sons.

Franckowiak, R. P., Panasci, M., Jarvis, K. J., Acuña-Rodriguez, I. S., Landguth, E. L., Fortin, M.-J., & Wagner, H. H. (2017). Model selection with multiple regression on distance matrices leads to incorrect inferences. *PLoS ONE*, *12*(4), e0175194. https://doi.org/10.1371/journal.pone.0175194

François, O., & Durand, E. (2010). Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, *10*(5), 773–784. https://doi.org/10.1111/j.1755-0998.2010.02868.x

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, *46*(2), 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*, 333–351. https://doi.org/10.1038/nrg.2016.49

Goudet, J., & Jombart, T. (2020). *hierfstat: Estimation and tests of hierarchical F-statistics*. R package version 0.5-7. https://CRAN.R-project.org/package=hierfstat

Guillot, G., Mortier, F., & Estoup, A. (2005). Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, 5(3), 712–715. https://doi.org/10.1111/j.1471-8286.2005.01031.x

Hamilton, L. S., Juvik, J. O., & Scatena, F. N. (Eds). (1995). *Tropical montane cloud forests*. Springer-Verlag.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247–276. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2020). *dismo: Species distribution modeling*. R package version 1.3-3. https://CRAN.R-project.org/package=dismo

Hijmans, R. J. (2021). *raster: geographic data analysis and modeling*. R package version 3.4-13. https://CRAN.R-project.org/package=raster

Howell, S. N. G., & Webb, S. (1995). *A guide to the birds of Mexico and northern Central America*. Oxford University Press.

Hutchison, D. W., & Templeton, A. R. (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, *53*(6), 1898–1914. https://doi.org/10.1111/j.1558-5646.1999.tb04571.x

Janzen, D. H. (1967). Why mountain passes are higher in the tropics. *The American Naturalist*, *101*(919), 233–249. https://doi.org/10.1086/282487

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, *101*(1), 92–103. https://doi.org/10.1038/hdy.2008.34

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. https://doi.org/10.1186/1471-2156-11-94

Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, *5*(1), 144–161. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmerman, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kierepka, E. M., & Latch, E. K. (2015). Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources*, *15*(3), 512–525. https://doi.org/10.1111/1755-0998.12332

Kozakiewicz, C. P., Carver, S., & Burridge, C. P. (2018). Under-representation of avian studies in landscape genetics. *Ibis*, *160*(1), 1–12. https://doi.org/10.1111/ibi.12532

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179–4191. https://doi.org/10.1111/j.1365-294X.2010.04808.x

Landguth, E. L., Fedy, B. C., Oyler-McCance, S. J., Garey, A. L., Emel, S. L., Mumma, M., Wagner, H. H., Fortin, M.-J., & Cushman, S. A. (2012). Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology Resources*, 12(2), 276–284. https://doi.org/10.1111/j.1755-0998.2011.03077.x

Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*, 3258. https://doi.org/10.1038/s41467-018-05257-7

Lee-Yaw, J. A., Davidson, A., McRae, B. H., & Green, D. M. (2009). Do landscape processes predict phylogeographic patterns in the wood frog? *Molecular Ecology*, *18*(9), 1863–1874.

Legendre, P., Lapointe, F.-J., & Casgrain, P. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, *48*(5), 1487–1499. https://doi.org/10.1111/j.1558-5646.1994.tb02191.x

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. https://doi.org/10.1111/1755-0998.12995

Linck, E. B., Freeman, B. G., Cadena, C. D., & Ghalambor, C. K. (2021). Evolutionary conservatism will limit responses to climate change in the tropics. *Biology Letters*, *17*(10), 20210363. https://doi.org/10.1098/rsbl.2021.0363

Llanes-Quevedo, A., Mastretta-Yanes, A., Sánchez-González, L. A., Castillo-Chora, V. J., & Navarro-Sigüenza, A. G. (2022). The tangled evolutionary history of a long-debated Mesoamerican taxon: The Velazquez Woodpecker (*Melanerpes santacruzi*, Aves: Picidae). *Molecular Phylogenetics and Evolution*, *170*, 107445. https://doi.org/10.1016/j.ympev.2022.107445

Luu, K., Bazin, E., & Blum, M. G. (2017). *pcadapt*: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, *18*(4), 189–197. https://doi.org/10.1016/S0169-5347(03)00008-9

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*, *24*(14), 3628–3638. https://doi.org/10.1111/mec.13258

Mason, N. A., Fletcher, N. K., Gill, B. A., Funk, W. C., & Zamudio, K. R. (2020). Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*, *18*(3), 269–280. https://doi.org/10.1080/14772000.2020.1730475

Mastretta-Yanes, A., Moreno-Letelier, A., Piñero, D., Jorgensen, T. H., & Emerson, B. C. (2015). Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, *42*(9), 1586–1600. https://doi.org/10.1111/jbi.12546

Mastretta-Yanes, A., Xue, A. T., Moreno-Letelier, A., Jorgensen, T. H., Alvarez, N., Piñero, D., & Emerson, B. C. (2018). Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. *Molecular Ecology*, *27*(2), 432–448. https://doi.org/10.1111/mec.14461

Mayr, E. (1963). *Animal species and evolution*. Harvard University Press.

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RAD seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). *Molecular Ecology Resources*, *17*(3), 356–361. https://doi.org/10.1111/1755-0998.12649

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science*, *201*(4358), 786–792. https://doi.org/10.1126/science.356262

Monteiro, W. P., Veiga, J. C., Silva, A. R., Carvalho, C. S., Lanes, É. C. M., Rico, Y., & Jaffé, R. (2019). Everything you always wanted to know about gene flow in tropical landscapes (but were afraid to ask). *PeerJ*, *7*, e6446. https://doi.org/10.7717/peerj.6446

Moreira, L. R., Hernández-Baños, B. E., & Smith, B. T. (2020). Spatial predictors of genomic and phenotypic variation differ in a lowland Middle American bird (*Icterus gularis*). *Molecular Ecology*, *29*(16), 3084–3101. https://doi.org/10.1111/mec.15536

Moreno-Contreras, I., Sánchez-González, L. A., Arizmendi M. C., Prieto-Torres, D. A., & Navarro-Sigüenza, A. G. (2020). Climatic niche evolution in the *Arremon brunneinucha* complex (Aves: Passerellidae) in a Mesoamerican landscape. *Evolutionary Biology*, *47*(2), 123–132. https://doi.org/10.1007/s11692-020-09498-7

Morrone, J. J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa*, *3782*(1), 1–110. https://doi.org/10.11646/zootaxa.3782.1.1

Morrone, J. J. (2020). *The Mexican Transition Zone: a natural biogeographic laboratory to study biotic assembly*. Springer.

Navarro-Sigüenza, A. G., Peterson, A. T., Nyari, A., García-Deras, G. M., & García-Moreno, J. (2008). Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. *Molecular Phylogenetics and Evolution*, *47*(1), 21–35. https://doi.org/10.1016/j.ympev.2007.11.030

Navarro-Sigüenza, A. G., García-Hernández, M. A., & Peterson, A. T. (2013). A new species of brush-finch (*Arremon*; Emberizidae) from western Mexico. *Wilson Journal of Ornithology*, *125*(3), 443–453. https://doi.org/10.1676/12-136.1

Nordborg, M., Charlesworth, B., & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetics Research*, *67*(2), 159–174. https://doi.org/10.1017/S0016672300033619

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Ornelas, J. F., Sosa, V., Soltis, D. E., Daza, J. M., González, C., Soltis, P. S., Gutiérrez-Rodríguez, C., Espinosa de los Monteros, A., Castoe, T. A., Bell,

C., & Ruiz-Sanchez, E. (2013). Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PloS ONE*, *8*(2), e56283. https://doi.org/10.1371/journal.pone.0056283

Ornelas, J. F., & González, C. (2014). Interglacial genetic diversification of *Moussonia deppeana* (Gesneriaceae), a hummingbird-pollinated, cloud forest shrub in northern Mesoamerica. *Molecular Ecology*, *23*(16), 4119–4136. https://doi.org/10.1111/mec.12841

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, *22*(24), 5983–5999. https://doi.org/10.1111/mec.12561

Parkes, K. C. (1954). A revision of the Neotropical finch *Atlapetes brunneinucha*. *Condor*, *56*(3), 129–138.

Paynter, R. A., Jr. (1978). Biology and evolution of the avian genus *Atlapetes* (Emberizinae). *Bulletin of the Museum of Comparative Zoology*, *148*(7), 323–369.

Pebesma, E. J. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Peterson, E. E., Hanks, E. M., Hooten, M. B., Ver Hoef, J. M., & Fortin, M.-J. (2019). Spatially structured statistical network models for landscape genetics. *Ecological Monographs*, *89*(2), e01355. https://doi.org/10.1002/ecm.1355

Pielou, E. C. (1991). *After the ice age: the return of life to glaciated North America*. University of Chicago Press.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology*, *23*(24), 5937–5942. https://onlinelibrary.wiley.com/doi/full/10.1111/mec.12965

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., Richardson, K., Rosing, M. T., Whittaker, R. J., & Fjeldså, J. (2019). Building mountain biodiversity: geological and evolutionary processes. *Science*, *365*(6458), 1114–1119. https://doi.org/10.1126/science.aax0151

Ramírez-Barahona, S., & Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. *Ecology and Evolution*, *3*(3), 725–738. https://doi.org/10.1002/ece3.483

R Development Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/ (accessed May 23, 2021).

Rice, A. E., Vinciguerra, N. T., McCullough, J. M. (2020). Early nest record and additional notes on the breeding biology of the Chestnut-capped Brushfinch (*Arremon brunneinucha suttoni*) in southern Mexico. *Ornitología Neotropical*, *31*(1), 76–78. https://journals.sfu.ca/ornneo/index.php/ornneo/article/view/557

Richardson, J. L., Brady, S. P., Wang, I. J., & Spear, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, *25*(4), 849–863. https://doi.org/10.1111/mec.13527

Robles-Bello, S. M., Vázquez-López, M., Ramírez-Barrera, S. M., Terrones-Ramírez, A. K., & Hernández-Baños, B. E. (2022). Drivers of phenotypic divergence in a Mesoamerican highland bird. *PeerJ*, *10*, e12901. https://doi.org/10.7717/peerj.12901

Rocha-Méndez, A., Sánchez-González, L. A., González, C., & Navarro-Sigüenza, A. G. (2019). The geography of evolutionary divergence in the highly endemic

avifauna from the Sierra Madre del Sur, Mexico. *BMC Evolutionary Biology*, *19*, 237. https://doi.org/10.1186/s12862-019-1564-3

Ruiz-Gonzalez, A., Cushman, S. A., Madeira, M. J., Randi, E., & Gómez-Moliner, B. J. (2015). Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular Ecology*, *24*(20), 5110–5129. https://doi.org/10.1111/mec.13392

Russello, M. A., Waterhouse, M. D., Etter, P. D., & Johnson, E. A. (2015). From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ*, *3*, e1106. https://doi.org/10.7717/peerj.1106

Sánchez-González, L. A., Morrone, J. J., & Navarro-Sigüenza, A. G. (2008). Distributional patterns of the Neotropical humid montane forest avifaunas. *Biological Journal of the Linnean Society*, *94*(1), 175–194. https://doi.org/10.1111/j.1095-8312.2008.00979.x

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15. https://doi.org/10.1111/evo.12258

Shirk, A. J., Wallin, D. O., Cushman, S. A., Rice, C. G., & Warheit, K. I. (2010). Inferring landscape effects on gene flow: a new model selection framework. *Molecular Ecology*, *19*(17), 3603–3619. https://doi.org/10.1111/j.1365-294X.2010.04745.x

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources*, *17*(6), 1308–1317. https://doi.org/10.1111/1755-0998.12684

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2018). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, *18*(1), 55–67. https://doi.org/10.1111/1755-0998.12709

Slarkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics*, *16*(1), 393–430. https://doi.org/10.1146/annurev.es.16.110185.002141

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, *47*(1), 264–279. https://doi.org/10.1111/j.1558-5646.1993.tb01215.x

Stattersfield, A. J., Crosby, M. J., Long, A. J., & Wege, D. C. (1998). *Endemic bird areas of the world*. Birdlife International.

Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2021). *qvalue: Q-value estimation for false discovery rate control*. R package version 2.24.0. http://github.com/jdstorey/qvalue

Termignoni-García, F., Jaramillo-Correa, J. P., Chablé-Santos, J., Liu, M., Shultz, A. J., Edwards, S. V., & Escalante-Pliego, P. (2017). Genomic footprints of adaptation in a cooperatively breeding tropical bird across a vegetation gradient. *Molecular Ecology*, *26*(17), 4483–4496. https://doi.org/10.1111/mec.14224

Teske, P. R., Golla, T. R., Sandoval-Castillo, J., Emami-Khoyi, A., van der Lingen, C. D., von der Heyden, S., Chiazzari, B., van Vuuren, B. J., & Beheregaray, L. B. (2018). Mitochondrial DNA is unsuitable to test for isolation by distance. *Scientific Reports*, *8*, 8448. https://doi.org/10.1038/s41598-018-25138-9

Thom, G., Gehara, M., Smith, B. T., Miyaki, C. Y., & do Amaral, F. R. (2021). Microevolutionary dynamics show tropical valleys are deeper for montane birds of the Atlantic Forest. *Nature Communications*, *12*, 6269. https://doi.org/10.1038/s41467-021-26537-9

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology & Evolution*, *16*(7), 330–343. https://doi.org/10.1016/S0169-5347(01)02177-2

USGS-United States Geological Survey. (2021). *HYDRO1k*. https://doi.org/10.5066/F77P8WN0

Vázquez-Selem, L., & Heine, K. (2011). Late Quaternary glaciation in Mexico. In J. Ehlers, P. L. Gibbard, & P. D. Hughes (Eds.), *Developments in Quaternary Sciences* (pp. 849–861). Elsevier.

Wagner, H. H., & Fortin, M.-J. (2013). A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, *14*(2), 253–261. https://doi.org/10.1007/s10592-012-0391-5

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Webster, G. L. (1995). The panorama of neotropical cloud forests. In S. P. Churchill, H. Balslev, E. Forero, & J. L. Luteyn (Eds.), *Biodiversity and conservation of Neotropical montane forest* (pp. 53–77). The New York Botanical Garden.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

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

Author contributions

IM-C, LAS-G, MCA, and AGN-S designed the study; IM-C extracted genomic DNA for the NextRAD sequencing libraries; IM-C, AL-Q, LAS-G, and AGN-S conducted the analyses; all authors contributed to paper writing that was led by IM-C. AGN-S obtained funding.

Data availability statement

Demultiplexed NextRAD short-read data with sample metadata have been deposited in NCBI's SRA archives under BioProject ID PRJNA950112 (accession numbers SAMN33975154–SAMN33975192). Associated files (VCF and metadata) have been deposited at GitHub repository of the Museo de Zoología "Alfonso L. Herrera" (https://github.com/Ornitologia-MZFC/).

TABLES

Table 1. Genomic diversity summary statistics for the different populations ('clusters') of the Mesoamerican *Arremon* bird complex delimited by the "total evidence" population structure approach (K = 5). Summary statistics were calculated using 2,634 NextRAD SNPs: observed heterozygosity = H_O; expected heterozygosity = H_E; inbreeding coefficient = F_{IS}; allelic richness = *Ar*.

Genetic cluster (taxonomic group / ecoregion name)	H_O	H_E	F_{IS}	*Ar*
SLT (*apertus* / Sierra de Los Tuxtlas)	0.027	0.055	0.246	1.039
SMOr (*brunneinucha* / Sierra Madre Oriental)	0.071	0.111	0.307	1.109
SMS (*kuehnerii* and *suttoni* / Sierra Madre del Sur)	0.070	0.109	0.307	1.107
CHH (*macrourus* / Chiapas Highlands)	0.075	0.173	0.464	1.155
TVB (*virenticeps* / Transmexican Volcanic Belt)	0.029	0.089	0.597	1.082

Table 2. Summary statistics of MLPE models controlling for population structure (*lme* function) using the proportion of shared alleles (D_{PS}) as the response variable (GD), and the predictors representing different genetic isolation hypotheses. For each model we provide estimates, standard errors (SE), degrees of freedom (df), *t*-values, *p*-values, confidence intervals (CI), and conditional R^2 (cR^2). Abbreviations for the fixed and random variables are: PS = population structure as a random effect, GeoD = geographic distance, NB = number of barriers, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish.

Hypothesis	**Predictor**	**Estimate**	**SE**	**df**	***t***	***p***	**CI (min/max)**	***AIC$_c$***	**c*R*^{2}**
IBD	Intercept	0.162	0.008	729	19.964	0	0.146/0.178	-4639.297	0.668
	GeoD	0.011	0.001	729	10.813	0	0.009/0.013		
	PS	0.025					0.015/0.041		
IBB | IBD	Intercept	0.161	0.008	728	20.014	0.000	0.145/0.177	-4629.582	0.667
	NB	0.001	0.0009	728	2.106	0.035	0.0001/0.003		
	GeoD	0.010	0.001	728	8.447	0.000	0.007/0.012		

	PS	0.025					0.015/0.041		
IBE_{global} | IBD	Intercept	0.161	0.008	728	19.906	0.000	0.145/0.177	-4626.236	0.669
	ED_{global}	0.0008	0.0006	728	1.306	0.191	-0.0004/0.002		
	GeoD	0.011	0.001	728	10.584	0.000	0.009/0.013		
	PS	0.025					0.015/0.041		
IBE_{bio12} | IBD	Intercept	0.162	0.008	728	19.932	0.000	0.146/0.178	-4625.32	0.668
	ED_{bio12}	0.0006	0.0007	728	0.839	0.401	-0.0008/0.002		
	GeoD	0.011	0.001	728	10.722	0.000	0.009/0.013		
	PS	0.025					0.015/0.041		
IBR_{tlcp} | IBD	Intercept	0.159	0.008	728	19.767	0.000	0.143/0.175	-4638.737	0.680
	RD_{tlcp}	0.010	0.003	728	3.362	0.0008	0.004/0.016		
	GeoD	0.003	0.002	728	1.490	0.136	-0.001/0.008		
	PS	0.025					0.015/0.041		
IBR_{PREga} | IBD	Intercept	0.163	0.008	728	19.188	0.000	0.147/0.180	-4647.971	0.700
	RD_{PREga}	0.006	0.001	728	4.827	0.000	0.003/0.008		
	GeoD	0.003	0.001	728	1.835	0.066	-0.0002/0.007		
	PS	0.027					0.017/0.043		

IBR_{LGMga} | IBD	Intercept	0.162	0.008	728	19.750	0.000	0.146/0.178	-4636.054	0.676
	RD_{LGMga}	0.003	0.001	728	3.279	0.001	0.001/0.005		
	GeoD	0.006	0.001	728	3.597	0.0003	0.002/0.010		
	PS	0.026					0.016/0.042		
IBR_{PRErad} | IBD	Intercept	0.161	0.007	728	20.824	0.000	0.145/0.176	-4642.178	0.670
	RD_{PRErad}	0.009	0.002	728	3.902	0.0001	0.004/0.014		
	GeoD	0.003	0.002	728	1.788	0.074	-0.0003/0.008		
	PS	0.024					0.015/0.039		
IBR_{LGMrad} | IBD	Intercept	0.161	0.007	728	20.572	0.000	0.146/0.177	-4637.34	0.664
	RD_{LGMrad}	0.006	0.001	728	3.321	0.0009	0.002/0.009		
	GeoD	0.005	0.002	728	2.601	0.009	0.001/0.009		
	PS	0.024					0.015/0.040		

Table 3. Summary statistics of the MRDM models using the proportion of shared alleles (D_{PS}) as the response variable (GD) and the predictors representing different genetic isolation hypotheses. For each model we provide estimates (β), p-values for each variable and for each model, and the coefficient of variation (R^2). GeoD = geographic distance, NB = number of barriers, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish

Hypothesis	**Predictor**	***β***	***p (var)***	***R^2***	***p (model)***
IBD	Intercept	0.152	0.001	0.387	0.001
	GeoD	0.023	0.001		
IBB | IBD	Intercept	0.152	0.001	0.397	0.001
	NB	0.006	0.082		
	GeoD	0.018	0.001		
IBE_{global} | IBD	Intercept	0.152	0.001	0.393	0.001
	ED_{global}	0.003	0.273		
	GeoD	0.022	0.001		
IBE_{bio12} | IBD	Intercept	0.152	0.001	0.392	0.001
	ED_{bio12}	0.002	0.294		
	GeoD	0.022	0.001		
IBR_{tlcp} | IBD	Intercept	0.152	0.001	0.404	0.001
	RD_{tlcp}	0.012	0.064		
	GeoD	0.012	0.066		

IBR_{PREga} | IBD	Intercept	0.152	0.001	0.391	0.001
	RD_{PREga}	0.004	0.446		
	GeoD	0.019	0.001		
IBR_{LGMga} | IBD	Intercept	0.152	0.001	0.407	0.001
	RD_{LGMga}	0.015	0.001		
	GeoD	0.008	0.052		
IBR_{PRErad} | IBD	Intercept	0.152	0.001	0.405	0.001
	RD_{PRErad}	0.012	0.087		
	GeoD	0.012	0.051		
IBR_{LGMrad} | IBD	Intercept	0.152	0.001	0.463	0.001
	RD_{LGMrad}	0.034	0.001		
	GeoD	-0.009	0.176		

31

Table 4. Summary statistics of the MLPE models controlling for population structure (*lme* function) using Euclidean distance of a PCA retaining the first 10 axes ($Euclidean_{PCA\text{-}10\ axes}$) as the response variable (GD), and the predictors representing different genetic isolation hypotheses. For each model we provide estimates, standard errors (SE), degrees of freedom (df), *t*-values, *p*-values, confidence intervals (CI), and conditional R^2 (cR^2). Abbreviations for the fixed and random variables are: PS = population structure as a random effect, GeoD = geographic distance, NB = number of barriers, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish.

Hypothesis	**Predictor**	**Estimate**	**SE**	**df**	***t***	***p***	**CI (min/max)**	***AIC*$_c$**	**c*R*^{2}**
IBD	Intercept	16.900	0.895	729	18.880	0	15.143/18.658	3329.041	0.399
	GeoD	3.420	0.229	729	14.901	0	2.969/3.870		
	PS	2.232					1.009/4.938		
IBB | IBD	Intercept	16.842	0.880	728	19.135	0.000	15.114/18.570	3331.282	0.398

	NB	0.217	0.203	728	1.070	0.284	-0.181/0.617		
	GeoD	3.283	0.263	728	12.461	0.000	2.766/3.801		
	PS	2.163					0.960/4.873		
IBE_{global} | IBD	Intercept	16.872	0.888	728	18.987	0.000	15.128/18.617	3332.817	0.398
	ED_{global}	0.073	0.147	728	0.501	0.616	-0.215/0.363		
	GeoD	3.404	0.231	728	14.714	0.000	2.950/3.859		
	PS	2.201					0.978/4.953		
IBE_{bio12} | IBD	Intercept	16.909	0.899	728	18.800	0.000	15.143/18.675	3332.969	0.400
	ED_{bio12}	-0.018	0.153	728	-0.121	0.902	-0.320/0.282		
	GeoD	3.422	0.230	728	14.850	0.000	2.969/3.874		
	PS	2.245					1.024/4.918		
IBR_{tlcp} | IBD	Intercept	16.217	0.731	728	22.174	0.000	14.781/17.653	3304.2	0.444
	RD_{tlcp}	3.405	0.654	728	5.203	0.000	2.120/4.690		
	GeoD	0.829	0.547	728	1.515	0.130	-0.245/1.903		
	PS	1.500					0.480/4.683		
IBR_{PREga} | IBD	Intercept	17.465	0.963	728	18.130	0	15.574/19.356	3291.892	0.376
	RD_{PREga}	1.216	0.187	728	6.496	0	0.849/1.584		

	GeoD	1.793	0.333	728	5.382	0	1.139/2.448		
	PS	2.570					1.309/5.045		
IBR_{LGMga} | IBD	Intercept	17.153	0.888	728	19.298	0	15.408/18.898	3305.094	0.352
	RD_{LGMga}	0.697	0.129	728	5.373	0	0.442/0.952		
	GeoD	2.183	0.319	728	6.836	0	1.556/2.810		
	PS	2.255					1.087/4.679		
IBR_{PRErad} | IBD	Intercept	16.782	0.842	728	19.915	0.000	15.127/18.436	3317.671	0.378
	RD_{PRErad}	1.072	0.283	728	3.776	0.000	0.514/1.629		
	GeoD	2.341	0.365	728	6.399	0.000	1.623/3.060		
	PS	2.014					0.901/4.505		
IBR_{LGMrad} | IBD	Intercept	16.796	0.851	728	19.729	0.000	15.124/18.467	3323.978	0.383
	RD_{LGMrad}	1.066	0.396	728	2.691	0.007	0.288/1.844		
	GeoD	2.464	0.426	728	5.774	0.000	1.626/3.302		
	PS	2.032					0.882/4.680		

Table 5. MRDM models using Euclidean distance of a PCA retaining the first 10 axes (*Euclidean*$_{PCA\text{-}10\ axes}$) as the response variable (GD) and the predictors representing different genetic isolation hypotheses. For each model we provide estimates (***β***), *p*-values for each variable and for each model, and coefficient of variation (R^2). Abbreviations for the predictor variables are: GeoD = geographic distance, NB = number of barriers, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish

Hypothesis	**Predictor**	***β***	***p (var)***	***R²***	***p (model)***
IBD	Intercept	15.695	0.001	0.226	0.001
	GeoD	3.042	0.001		
IBB | IBD	Intercept	15.695	0.001	0.231	0.001
	NB	0.800	0.254		
	GeoD	2.390	0.003		
IBE_{global} | IBD	Intercept	15.695	0.001	0.234	0.001
	ED_{global}	0.611	0.279		
	GeoD	2.848	0.001		
IBE_{bio12} | IBD	Intercept	15.695	0.001	0.234	0.001
	ED_{bio12}	0.590	0.348		
	GeoD	2.855	0.001		
IBR_{tlcp} | IBD	Intercept	15.695	0.001	0.327	0.001
	RD_{tlcp}	5.048	0.003		

	GeoD	-1.582	0.234		
IBR_{PREga} | IBD	Intercept	15.695	0.001	0.238	0.001
	RD_{PREga}	0.828	0.314		
	GeoD	2.588	0.002		
IBR_{LGMga} | IBD	Intercept	15.695	0.001	0.242	0.001
	RD_{LGMga}	1.300	0.010		
	GeoD	2.029	0.007		
IBR_{PRErad} | IBD	Intercept	15.695	0.001	0.228	0.001
	RD_{PRErad}	0.665	0.621		
	GeoD	2.448	0.048		
IBR_{LGMrad} | IBD	Intercept	15.695	0.001	0.227	0.001
	RD_{LGMrad}	0.450	0.723		
	GeoD	2.656	0.021		

FIGURES

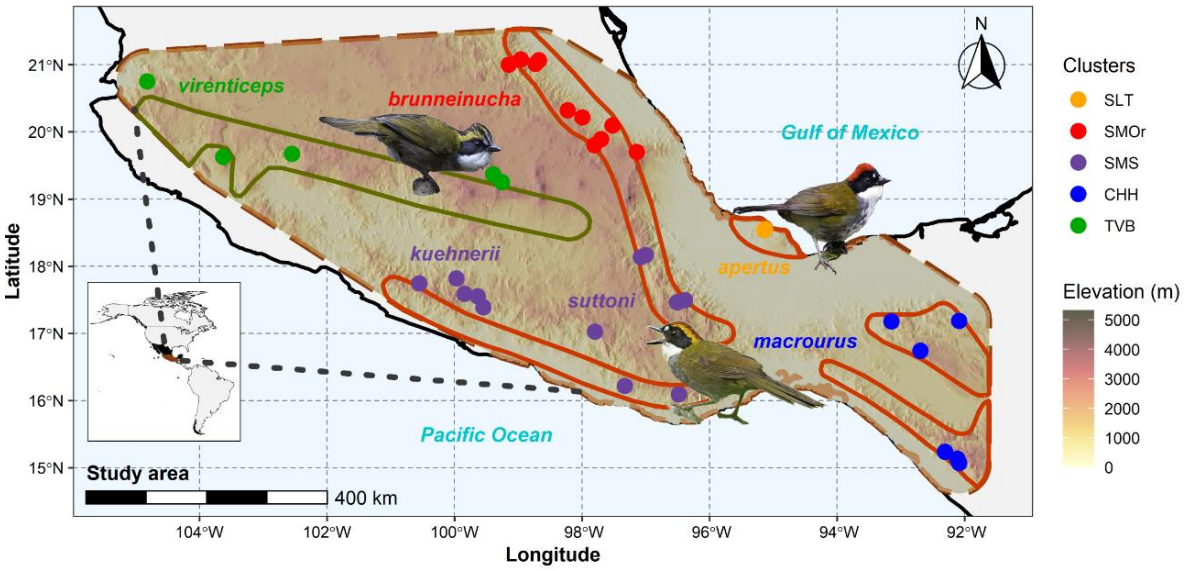


Figure 1. Map of the 35 sampling localities (representing 39 samples) overlaid on a digital elevation model from the Hydro1K project in northern Mesoamerica following the taxonomy of both the International Ornithological Congress (IOC) and Navarro-Sigüenza et al. (2013). The dotted line encompasses our study area. Red and green polygons represent distributional ranges of the Chestnut-capped Brushfinch/Green-striped Brushfinch complex. Colored dots represent identified genetic clusters according to the "total evidence" approach (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS= Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt. Photographs were kindly provided by Leopoldo Vázquez-Reyes (*A. virenticeps*), Manuel O. Grosselet (*A. b. suttoni*), and Juan M. Carmona-Victoria (*A. b. apertus*).

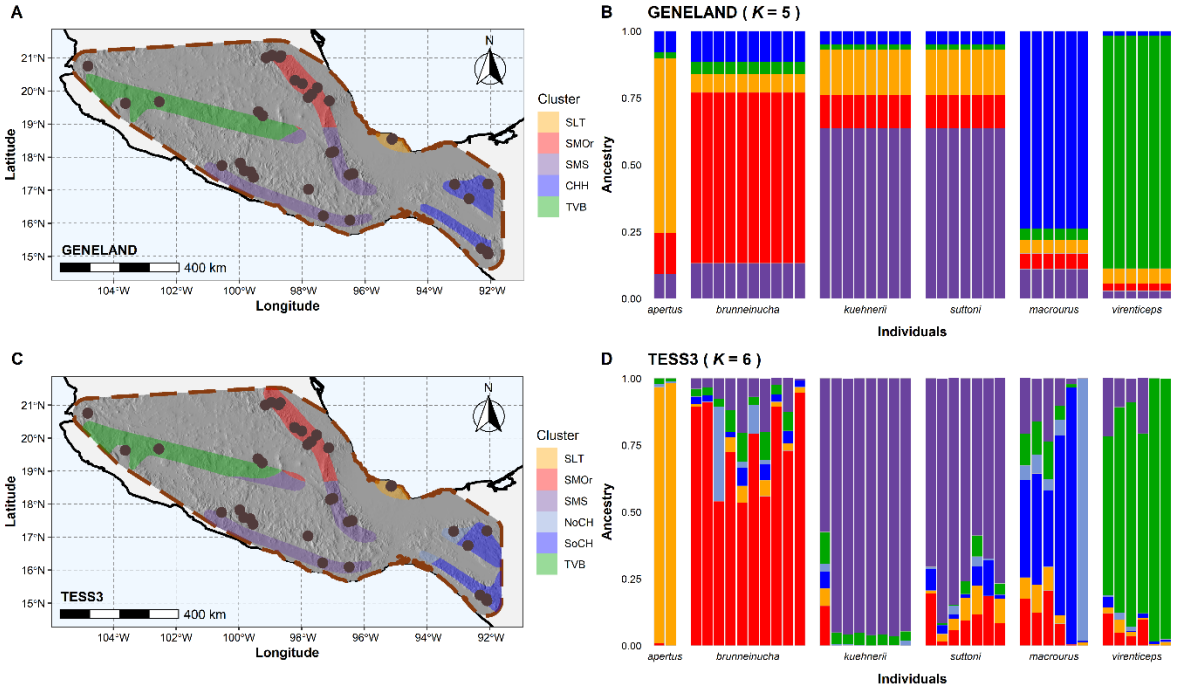


Figure 2. Results from GENELAND (A-B) and TESS3R (C-D) performed on the set of 2,634 filtered neutral SNP data (excluding outliers) from all individual samples (*n* = 39). The dotted line encompasses our study area. Colored dots represent identified genetic clusters (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS= Sierra Madre del Sur, CHH = Chiapas Highlands, NoCH = Northen Chiapas Highlands, SoCH = Southern Chiapas Highlands, TVB = Transmexican Volcanic Belt.

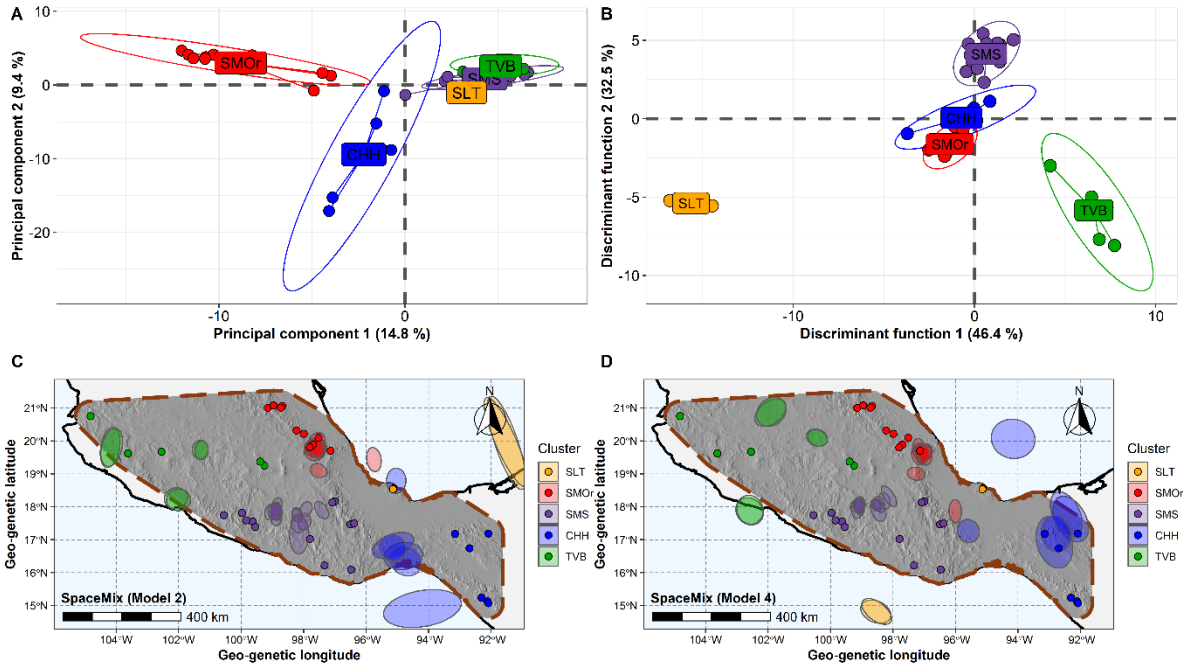


Figure 3. Principal component analysis (PCA) explaining genetic variation of the two first axes (A). Discriminant analysis of principal components (DAPC) depicting the two first discriminant axes (B). SpaceMix results showing the geo-genetic map of Model 2 ("estimating geo-genetic locations but not admixture"; C) and Model 4 ("estimating geo-genetic locations and admixture source locations"; D). Ellipses in SpaceMix models represent the 95% Bayesian credible intervals around each sample's location on the geo-genetic map, and colored dots represent actual sampling locations. Colored dots represent identified genetic clusters according to the "total evidence" approach (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS= Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt.

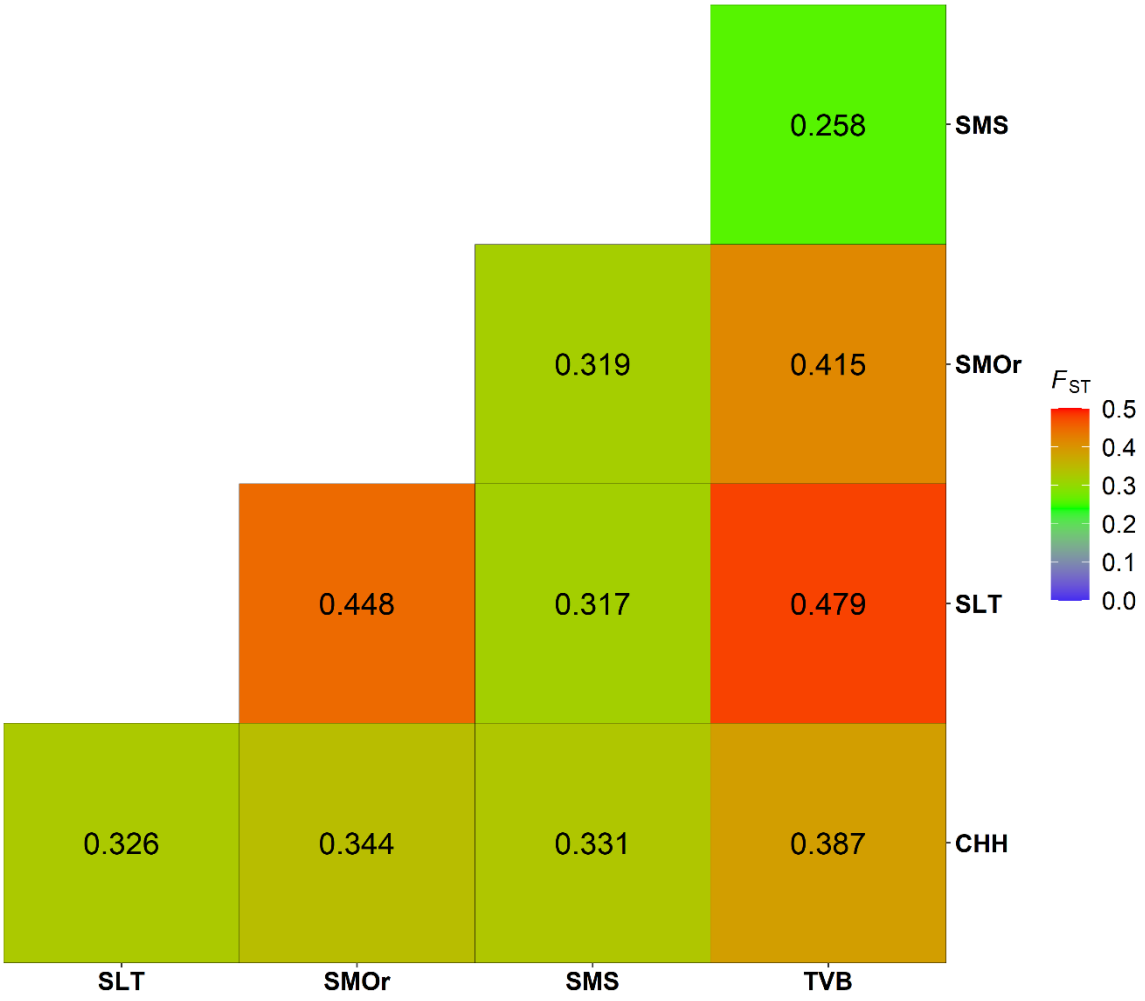


Figure 4. Pairwise genetic differentiation (F_{ST}) between genetic clusters. Blue and green colors indicate weaker genetic differentiation. SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS= Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt.

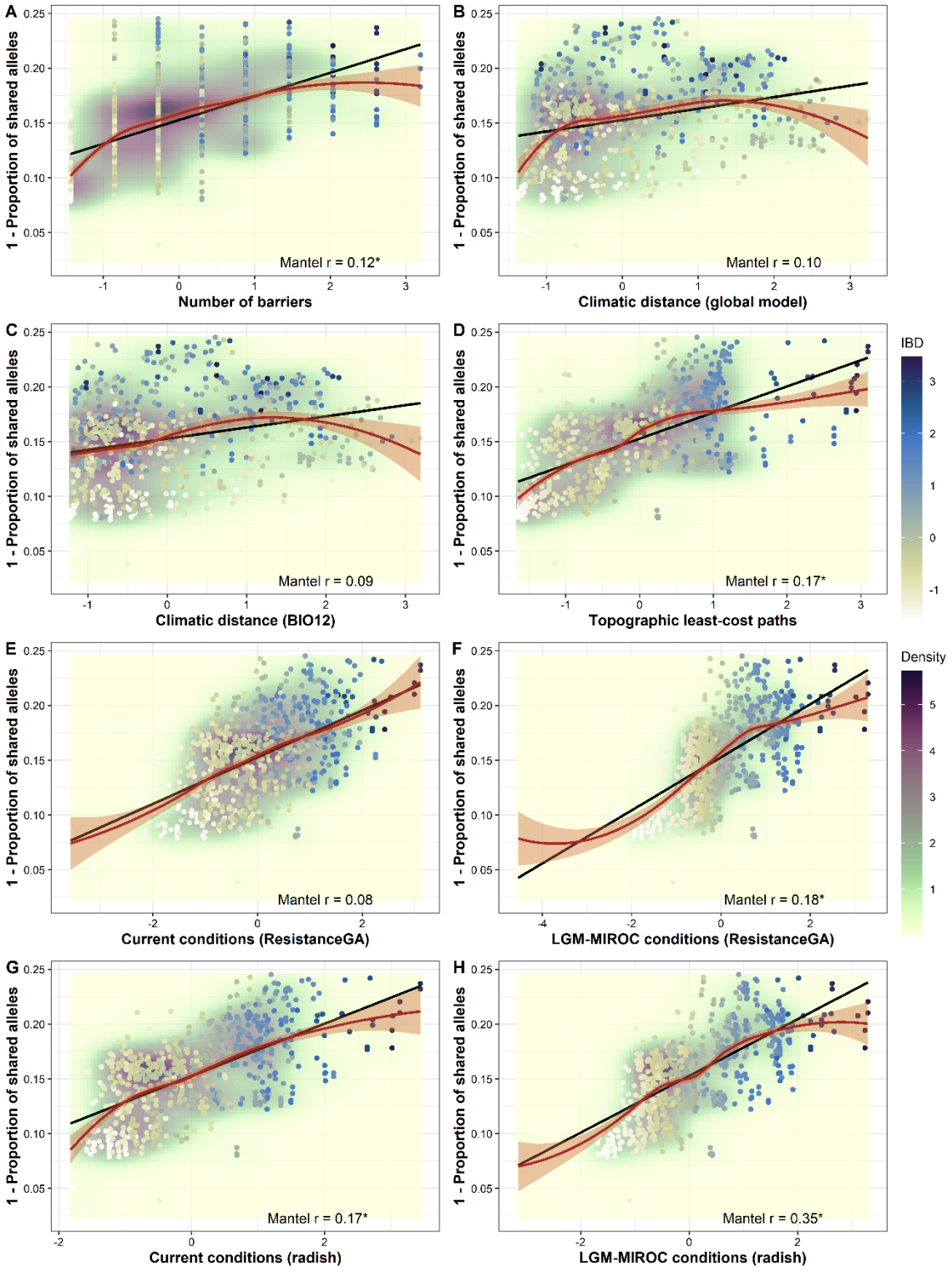


Figure 5. Plots of linear relationships representing each genetic isolation hypothesis using the proportion of shared alleles (D_{PS}) while controlling for IBD. (A) IBB model; (B)

IBE_{global} model including all non-correlated bioclimatic layers; (C) IBE_{bio12} model which includes only BIO12; (D) IBR_{tlcp} model representing IBR based on topographic least-cost paths; (E) IBR_{PREga} model representing IBR of a surface of ENM (current conditions) and optimized in ResistanceGA; (F) IBR_{LGMga} model representing IBR of a surface of ENM (LGM-MIROC conditions) and optimized in ResistanceGA; (G) IBR_{PRErad} model representing IBR of a surface of ENM (current conditions) and optimized in radish; (H) IBR_{LGMrad} model representing IBR of a surface of ENM (LGM-MIROC conditions) and optimized in radish. The red line indicates the observed trend (displaying 95% confidence intervals), while the black line represents the expected trend. Dot colors indicate the pairwise IBD values and the Density bar values illustrates the point conglomerations for its model respectively. * = statistical significance ($p < 0.05$).

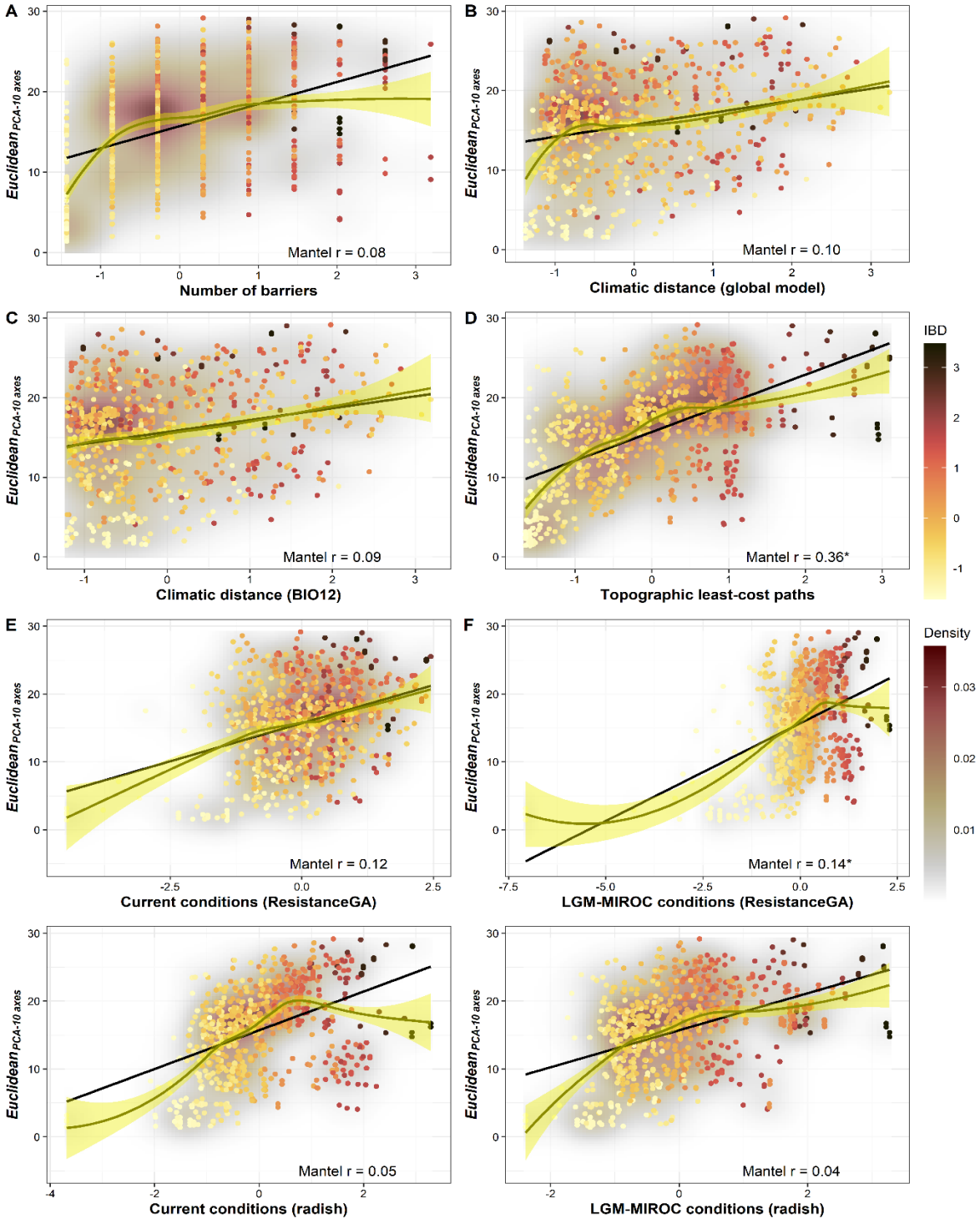


Figure 6. Plots of linear relationships representing each genetic isolation hypothesis using Euclidean distance of a PCA retaining the first 10 axes (*Euclidean*$_{PCA\text{-}10\ axes}$) while

controlling for IBD. (A) IBB model; (B) IBE_{global} model including all non-correlated bioclimatic layers; (C) IBE_{bio12} model which includes only BIO12; (D) IBR_{tlcp} model representing IBR based on topographic least-cost paths; (E) IBR_{PREga} model representing IBR of a surface of ENM (current conditions) and optimized in ResistanceGA; (F) IBR_{LGMga} model representing IBR of a surface of ENM (LGM-MIROC conditions) and optimized in ResistanceGA; (G) IBR_{PRErad} model representing IBR of a surface of ENM (current conditions) and optimized in radish; (H) IBR_{LGMrad} model representing IBR of a surface of ENM (LGM-MIROC conditions) and optimized in radish. The green line indicates the observed trend (displaying 95% confidence intervals), while the black line represents the expected trend. Dot colors indicate the pairwise IBD values and the Density bar values illustrate the point conglomerations for its model respectively. * = statistical significance ($p < 0.05$).

Supplementary Material

Isolation by resistance explains genetic diversity in the *Arremon* brushfinches of northern Mesoamerica

Running title: Landscape genomics of Mesoamerican brushfinches

Authors and Affiliations

Israel Moreno-Contreras[1,2], Alexander Llanes-Quevedo[1,2], Luis A. Sánchez-González[2], María del Coro Arizmendi[3], Adolfo G. Navarro-Sigüenza[2]

[1]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1º Piso, Circuito de Posgrados, Ciudad Universitaria, Universidad Nacional Autónoma de México, Coyoacán, Mexico City 04510, Mexico.

[2]Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, A.P. 70–399, Mexico City 04510, Mexico.

[3]UBIPRO, FES Iztacala Universidad Nacional Autónoma de México, Av. de los Barrios, Los Reyes Iztacala, Tlalnepantla 54090, Estado de México, Mexico.

*Corresponding author: adolfon@ciencias.unam.mx

Supplementary Material

1. Methods

1.1 Collected samples, assembly parameters, and quantification of missing data

We collected frozen or ethanol-preserved tissue (muscle, heart or liver) samples from 41 individuals of the *Arremon* bird complex in northern Mesoamerica stored at the Museo de Zoología, Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM). Individuals were collected between June 1990 and December 2013 and grouped into 35 nonrandom locations encompassing the Mexican part of the range of the complex (Table S1). As mentioned in the Method section of the main paper, samples were sent for NextRAD sequencing in SNPsaurus, LLC (Oregon, USA).

To determine the performance of each *de novo* assembly of the ipyrad pipeline, we computed the Euclidean genetic distance between each individual (n = 741 paired comparisons). Euclidean distances were estimated with the *dist* function in 'stats' v.3.6.2 (R Development Core Team, 2021). The assembly matrices analyzed were the unfiltered ipyrad data set and the filtered genomic data containing those loci with < 25% missing data using 'VCFtools' v.0.1.17 (Danecek et al., 2011). We selected ipyrad parameter settings to be those that both yielded the highest number of resulting SNPs and a lowest amount of missing data, while minimized the Euclidean distances. Once the ipyrad parameters were selected, we quantified the genotype depth and percentage of missing data of the ipyrad output (the unfiltered data set). We also calculated the percentage of missing data of final NextRAD data set for each sample and for each locus in 'VCFtools' v.0.1.17 (Danecek et al., 2011). This final NextRAD data set was obtained after cleaning of genomic data, standardize *de novo* assembly parameters, and post-processing bioinformatic steps.

Table S1. Geographical information of the sequenced samples. * = indicate samples removed prior to *de novo* assembly in ipyrad. Sex: F = female, M = male, U = unidentified sex.

Species	Sample ID	State	Locality	Sex	Date	Longitude	Latitude	Missing data (%)
A. kuehnerii	AGM_141	Guerrero	Cerro El Balcón	F	02/2008	-100.5483	17.7416	49.81
A. brunneinucha	AHC_060	Puebla	Cerro Miqueco	U	06/2013	-97.6938	19.8868	13.02
A. brunneinucha	AHC 107	Puebla	Chopilco Alto	M	08/2013	-97.8034	19.7994	4.67
A. kuehnerii	AMT_017	Guerrero	Carrizal de Bravo	M	04/2003	-99.9675	17.8167	18.64
A. virenticeps	BIODF_015	Mexico City	San Nicolás Totolapan	M	10/2008	-99.2516	19.2465	48.94
A. brunneinucha	BMM_430	Chiapas	Volcán Tacaná, Río Malá, x km x Unión Juárez	M	06/1990	-92.11	15.1316	61.58
A. brunneinucha	BMM_568	Chiapas	6 km NE de Pueblo Nuevo, Camino Aurora-Ermita	F	04/1991	-92.0833	17.1833	50.72
A. brunneinucha	BMM_804	Chiapas	Volcán Tacaná ladera, vereda	M	04/1992	-92.0833	15.0666	86.07

			a Tapalapa, Rancho Chiquihuite					
A. brunneinucha	BMM_915	Hidalgo	El Potrero, 5 km Tenango de Doria	M	08/1992	-98.2216	20.3172	25.55
A. brunneinucha	CONACYT_0773	Oaxaca	Puerto de la Soledad	M	07/2001	-96.9966	18.165	3.87
A. brunneinucha	CRGA_41	Chiapas	Cerro Boquerón en la cima	U	06/2012	-92.3046	15.2354	7.56
A. virenticeps	DEUT_047	Michoacán	La Verdura	M	12/2000	-102.545	19.6683	15.49
A. virenticeps	ENT_078*	Jalisco	El Floripondio (Nevado de Colima) aprox. 2 km S	M	06/2008	-103.6233	19.6216	NA
A. virenticeps	ENT_098	Jalisco	El Floripondio (Nevado de Colima) aprox. 2 km S	M	06/2008	-103.6233	19.6216	12.38
A. kuehnerii	GMS_894*	Guerrero	Carrizal de Bravo	F	01/2004	-99.9675	17.8167	NA

A. brunneinucha	HGO-SLP 025	Hidalgo	Cerro Jarros, 1 Km E El Sotano	F	12/1998	-99.1449	20.9983	25.32
A. brunneinucha	HUAS_108	Hidalgo	Hueyapa	F	01/2005	-98.6727	21.0642	52.51
A. brunneinucha	IALL_133	Puebla	Salto Chico, camino a Presa Necaxa	M	06/2009	-97.9920	20.2143	12.87
A. kuehnerii	LAB_011	Guerrero	Xocomanatlán	M	05/2006	-99.6333	17.55	19.48
A. brunneinucha	LAM_01	Oaxaca	El Portillo, 4km SW de Juquila	M	10/2009	-97.3266	16.2153	5.92
A. virenticeps	MAG_23	State of Mexico	Río Magdalena Contreras	U	05/2007	-99.3897	19.3674	17.01
A. brunneinucha	MFOR_004	Hidalgo	Tlanchinol 5 km E	F	05/2006	-98.7238	21.0072	56.91
A. kuehnerii	MM_101	Guerrero	Carrizal de Bravo	M	01/2004	-99.9675	17.8167	6.00
A. brunneinucha	MOL13_020	Chiapas	San Nicolás Buenavista, Cerro Huitepec	M	06/2013	-92.6880	16.7380	12.45
A. brunneinucha	MOL13_183	Oaxaca	Carretera San Miguel Suchitepec a 6	U	06/2013	-96.4797	16.0891	14.58

			km hacia Puerto Angel					
A. kuehnerii	MOLGRO_153	Guerrero	Ejido Carrizal de Bravo, El Asoleadero	M	05/2013	99.8370	17.5866	9.04
A. kuehnerii	MOLGRO_175	Guerrero	Ejido Carrizal de Bravo, El Asoleadero	F	05/2013	99.8370	17.5866	9.19
A. kuehnerii	MOLGRO_302	Guerrero	Ejido Carrizal de Bravo, El Asoleadero	M	05/2013	99.8370	17.5866	11.28
A. brunneinucha	MT_362	Oaxaca	Vistahermosa, km 77 carretera Tuxtepec-Oaxaca	F	06/1991	-96.5033	17.4616	22.63
A. brunneinucha	OMVP_0211	Oaxaca	Reyes Llano Grande	F	03/1993	-97.795	17.025	33.49
A. brunneinucha	OR_05	Oaxaca	Puerto de la Soledad 20 km E Teotitlán del Camino	M	09/2003	-97.0766	18.1319	9.64

A. kuehnerii	PBSMS_07	Guerrero	Las Antenas, Ejido Palo Blanco	F	10/2003	-99.5440	17.3853	8.43
A. brunneinucha	PUE_096	Puebla	Cuitchat, 8 km NE Cuetzalan	F	04/1993	-97.5166	20.0916	7.90
A. virenticeps	SIT_024	Jalisco	Carretera San Sebastián hacia La Bufa a 20 minutos en auto	M	11/2007	-104.8169	20.7490	6.87
A. virenticeps	SIT_025	Jalisco	Carretera San Sebastián hacia La Bufa a 20 minutos en auto	M	11/2007	-104.8169	20.7490	6.42
A. brunneinucha	SIT_143	Chiapas	Coapilla a 6.5 km N	F	12/2007	-93.1463	17.1741	7.59
A. brunneinucha	SLA_031	Hidalgo	El Coyol, 1 Km E	M	08/2000	-98.955	21.0749	15.57
A. brunneinucha	TXT_29	Veracruz	Ejido La Perla de San Martín, Campamento B	M	07/2013	-95.1236	18.5512	21.15

A. brunneinucha	TXT_45	Veracruz	Ejido Adolfo Ruiz Cortínez	F	12/2013	-95.1379	18.5353	7.71
A. brunneinucha	XCOA_50	Veracruz	Rancho de la Virgen	F	11/2012	-97.1349	19.6973	4.78
A. brunneinucha	YAGILA_08	Oaxaca	San Juan Yagila	M	11/2007	-96.3814	17.4914	4.56

1.2 Population structure analyses

We followed a "total evidence" approach to assign individuals to genetic groups (i.e., "*K*") using different clustering methods (Cullingham et al., 2020). Based on the allopatric distribution of the Arremon complex in northern Mesoamerica (Howell & Webb, 1995), we explored genetic structure, testing $K \in \{2, 3, 4, 5, 6, 7\}$ in different model-based methods (i.e., GENELAND, TESS3, and STRUCTURE). Using 'GENELAND' v.2.1.4 (Guillot et al., 2005), we inferred the number of genetic clusters in the complex using a Bayesian model under an MCMC scheme to locate genetic discontinuities using individual geo-referenced multilocus genotypes (François & Durand, 2010). We then performed ten independent Bayesian runs with 200,000 iterations, saving every two hundredth run, treating the number of genetic clusters as unknown with correlated allele frequencies, and using the spatial D-model as a prior for allele frequencies. We applied a burn-in of 400 iterations (40%), and MCMC convergence was assessed by comparing the number of clusters across replicated runs, with the mean posterior density used as a criterion to choose the best run (Guillot et al., 2005).

We then used the 'TESS3' algorithm as implemented in the R-package 'tess3r' v.1.1.0 (Caye et al., 2016). TESS3 attempts to minimize the Walhund effect by incorporating local dependencies in their model. Unlike GENELAND, Voronoi cells in TESS3 are not associated with territories, but with 'individuals' (Caye et al., 2016). This alternative clustering estimation was computed with 10 repetitions (tolerance = 1×10^{-5}, max. iterations = 100,000, mask = 0.5), and different lambda values (0–1.5, varying each 0.05). Fine-tuning these parameters resulted in a total of 31 analyses; to choose among these analyses, we selected the most likely lambda value based on the minimum cross-validation score and the best K value following the first trendline break (Caye et al., 2016).

As a complementary approach to spatial assignment tests, we explored the genome-wide SNP data using the individual-based Bayesian clustering approach in 'STRUCTURE' v.2.3.4 (Pritchard et al., 2000). We ran STRUCTURE for K with ten replicates per estimation. Each run was performed over 200,000 MCMC iterations with a burn-in of 100,000 iterations, excluding the LOCPRIOR option. We performed a fine-tune

exploration of parameter settings to determine which STRUCTURE results best fit those obtained using spatial methods (GENELAND and TESS). First, we used the admixture model with the correlated allele frequencies (hereafter CAF) model (with all options set at their default settings). Second, we tested the admixture model with the uncorrelated allele frequencies (hereafter UAF) model (with all options kept at their default parameters). Third, we implemented the alternative model (gamma prior) setting recommended by Wang (2017) for heterogeneous sample sizes among actual genetic clusters (admixture model with α varying for each population), with the initial α set at 1/*K* (i.e., 1/6 = 0.16666667), and a CAF model. We then determined the optimal *K* using both the highest ln Pr(*X*|*K*) values ($\widehat{K}_{\text{Pritchard}}$; Pritchard et al., 2000) and Δ*K* scores ($\widehat{K}_{\text{Evanno}}$; Evanno et al., 2005) in the 'POPHELPER' v.2.3.1 package for R (Francis, 2017). The results from 10 replicates of the selected *K* values were summarized into a single result, which were then aligned and analyzed in CLUMPP v.1.1.2 (Jakobsson & Rosenberg, 2007) using the greedy algorithm with 10,000 repeats.

We followed a spatial population genetics approach to understand whether the population assignments mentioned above are true genetic clusters or the result of a geo-genetic map. To visualize the spatial arrangement of the sequenced samples, we generated a geo-genetic map of the Mesoamerican Arremon species complex using 'SpaceMix' v.2.3.1 (Bradburd et al., 2016). The map consists of a bidimensional plot in which the geographic distances between individuals correspond to their expected GD under IBD. Differences between geographic and geo-genetic locations may reflect historical rates of gene flow across the landscape. For example, if these geo-genetic circles overlap, one can infer that populations (or clusters obtained by assignment methods) should probably be collapsed into one. As input for SpaceMix, we used allele frequency covariance data generated from the filtered NextRAD genomic dataset and set real geographic coordinates as location priors. Two SpaceMix analyses were implemented, incorporating the model options "estimating geo-genetic locations but not admixture" (Model 2) and "estimating geo-genetic locations and admixture source locations" (Model 4; Bradburd et al., 2016), respectively.

We executed both SpaceMix analyses using 10 initial fast runs, each with 1,000,000 iterations followed by a long MCMC run of 10,000,000 iterations sampled every 1,000 iterations. The first 4,000 samples were discarded as burn-in, leaving 6,000 samples per run. We plotted geo-genetic maps showing 95% credible ellipses for both models. Since populations of Neotropical montane forest taxa have probably been in contact with disjunct populations by extending or contracting their distributions during the LGM (Ramírez-Barahona & Eguiarte, 2013; Flantua & Hooghiemstra, 2018; Rahbek et al., 2019), we evaluated the support of our data for models with and without long-distance gene flow between populations (i.e., all w parameters set to zero vs. w allowed to vary). For that, we calculated the Bayes Factor (Kass & Raftery, 1995) between the two models using the Savage-Dickey density ratio (Dickey & Lientz, 1970). These analyses were performed using an R-script published by Márquez et al. (2020).

We further investigated the spatial autocorrelation of the allele frequencies in our study area by performing a spatial principal component analysis (sPCA, Jombart et al., 2008), which does not require genomic data to be in Hardy-Weinberg or linkage equilibrium. sPCA also adds a spatial element by the isolation of the global structure, representing disconnected groups or clines. A global pattern would differentiate between two spatial groups or find a cline (or any intermediate state), helping to resolve genetic clustering suggested by model-based methods. Using the R-package 'adegenet' v.2.1.4 (Jombart, 2008), we estimated sPCA given neutral SNPs and location of each genotyped sample using three positive eigenvalues (global structures). To connect individual locations, we built a connection network using the neighborhood by distance option. We ran Monte-Carlo tests with 999 permutations using the *spca_randtest* function to determine the significance in global spatial structures (Montano & Jombart, 2017).

1.3 Elaboration of ecological niche models for IBR hypothesis tests

Distributional data

We compiled distributional records for the allopatric Chestnut-capped/Green-striped Brushfinches complex in Mesoamerica (genus *Arremon*) from Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) using the *occ_search* function of 'rgbif' v3.6.0 R-library (Chamberlain et al., 2022) for two representative taxa of the complex (*A. brunneinucha* and *A. virenticeps*). Our search retrieved 51,786 records for *A. brunneinucha* and 4,392 records for *A. virenticeps* (see DOI for each dataset in 'References for distributional data'), which included the entire distribution range of the *Arremon* bird complex in northern Mesoamerica (Howell & Webb, 1995). To reduce sampling biases for the ecological niche modeling, the occurrences were cleaned setting filters of the *clean_coordinates* function in 'CoordinateCleaner' v2.0-18 (Zizka et al., 2019).

After applying these filters, we kept those records that fell within the temporal window between 1990 and 2013 (*n* = 3,689 records). Most citizen-science data and museum records are collected opportunistically and often from convenient locations (i.e., near roads or towns and cities), leading to geographic clusters of unique locations (Boria et al., 2014). Therefore, we spatially thinned the entire dataset at 10 km for the simulation and configuration of the M area (calibration area), and for the elaboration of ENMs. The spatial thin process was determined based on the number of available records and the spatial extent of the allopatric populations.

After this filtering, both datasets comprised 219 occurrences for the creation of calibration area and 213 occurrences for the creation of ENMs for IBR tests considering all *Arremon* sequenced samples. The species occurrences used for ENMs were equally split in two datasets, one for building the ENMs (*n* = 80 records) and the second for data testing (i.e., evaluation; *n* = 133 records). For the fine-spatial analysis, 39 records were employed for the creation of the calibration area of *brunneinucha* (Sierra Madre Oriental), which 24 records were applied for calibration and 15 records were used for evaluation model.

Environmental data

Because sequenced samples came from museum specimens temporarily not coinciding with conventional GIS climate data (e.g., WorldClim), we used other GIS climate alternatives for IBE and IBR hypothesis testing. Bioclimatic layers representing current conditions were obtained from the CHELSA v.1.2 project (Karger et al., 2017). Past climate conditions representing the last glacial maximum (hereafter LGM; 21,000 years ago) based on PMIP3 (Paleoclimate Modeling Intercomparison Project), were gathered of the Model for Interdisciplinary Research on Climate (MIROC-ESM, (Hasumi & Emori, 2004) from CHELSA-LGM project (https://chelsa-climate.org/last-glacial-maximum-climate/). All environmental layers were downloaded at high resolution (30 arc-second, approximately 1 x 1 km) with a WGS84 projection. One environmental dataset (all bioclimatic variables) was clipped using one polygon representing the G-area for the generation of the calibration area (see *Elaboration of calibration area*).

We selected a relevant set of environmental factors ("scenopoetic variables") potentially influencing the geographical distribution of the *Arremon* bird complex in northern Mesoamerica (Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013; Moreno-Contreras et al., 2020; Rice et al., 2020). The geographic coordinates (WGS84) were converted into projected coordinates under a Mexico ITRF2008 Lambert Conformal Conical projection (900 x 900 m of resolution). We choose such a resolution due to known limited dispersal capabilities in the *Arremon* species (Castaño et al. 2019), which may reflect the importance of landscape features impeding gene flow. For that, we extracted the environmental values of bioclimatic variables using the occurrence dataset for the ENM building (n = 213 records) with the *extract* function of 'raster' v3.4-13 (Hijmans, 2021).

We discarded highly correlated variables ($r \geq 0.7$) using Spearman's rank correlation coefficients, and we kept the following environmental data set: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), annual precipitation (BIO12), precipitation of driest month (BIO14), precipitation seasonality (BIO15), and precipitation of warmest quarter (BIO18). That final

environmental dataset was included for the creation of ENMs (see Constructing ecological niche models). Geoprocessing, manipulation, and visualization of spatial data were performed using ArcGIS v.10.4.1 (ESRI, Redlands, California, USA), QGIS v.3.16.9 (QGIS Development Team), and the following geospatial packages in R: 'dismo' v.1.3-3 (Hijmans et al., 2020), 'ggmap' v.3.0.0 (Kahle & Wickham, 2013), 'ggplot2' v.3.3.5 (Wickham, 2016), 'raster' v3.4-13 (Hijmans, 2021), 'rgdal' v.1.-23 (Bivand et al., 2021), 'rgeos' v. 0.5-5 (Bivand & Rundel, 2020), and 'sf' v.1.0-2 (Pebesma, 2018).

Elaboration of calibration area

Traditionally, calibration areas (also called "M area"; Barve et al., 2011) employed in the building of ENMs were delimited using ecoregions or biogeographic provinces (Machado-Stredel et al., 2021a). Nonetheless, the accessible area should include all sites to which the species has had access through dispersal, a subset of which would be the sites that are appropriate for the species in terms of abiotic and biotic environment (Barve et al., 2011). This finding emerges by the well-known phenomenon of "sky islands," in which species have presently isolated populations at sites well south of the main distributional area that were colonized by the species during glacial periods (Rahbek et al., 2019).

Instead, we followed a novel approach for the elaboration of the M area based on simulations (Machado-Stredel et al., 2021a). We generated simulations with the *M_simulationR* function in 'grinnell' v0.0.21 (Machado-Stredel et al., 2021b) and 216 species occurrences including all representatives of the *Arremon* bird complex in northern Mesoamerica. In the case of brunneinucha samples, 39 species occurrences were used for simulations. The simulations for both scales were parameterized with the following: 2 as maximum number of dispersers that depart from each colonized pixel, 350 dispersal events, and 10 replicates. The resultant polygons were considered as the hypotheses representing the calibration areas for both cases.

Constructing ecological niche models

To obtain resistance layers of the *Arremon* bird complex in our landscape genomics approach, we developed ecological niche models (ENMs) under current and past (LGM-

MIROC) climate conditions using the uncorrelated environmental variables implementing a maximum entropy algorithm, as implemented in Maxent v3.4.1 (Phillips et al., 2017), through the 'dismo' v1.3-3 (Hijmans et al., 2020), and 'kuenm' v1.1.7 (Cobos et al., 2017) for R (R Development Core Team, 2021). Maxent is a machine-learning technique that predicts species' distributions using detailed environmental GIS data sets together with occurrence data and a previously delimited background accessible area (i.e., M, see above; Phillips et al., 2006). Maxent is widely considered a robust modeling algorithm compared to other techniques (Elith et al., 2006) and it is mathematically equivalent to a Poisson regression, a particular type of generalized linear model (Renner and Warton, 2013).

Considering that our study is including almost all allopatric populations of *Arremon* in northern Mesoamerica (Howell & Webb, 1995), and because their life histories (Paynter, 1978) and morphological traits (Navarro-Sigüenza et al., 2008; Navarro-Sigüenza et al., 2013) are similar, we built ENMs using all spatially thinned occurrences (*n* = 213 records) as implemented in Castaño-Quintero et al. (2020). We included 39 occurrences for the elaboration of ENMs of *brunneinucha*. Levels of model complexity were evaluated, such as overfitting by varying the regularization multiplier (RM) values (0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 8) and feature classes linear (L), quadratic (Q), product (P), threshold (T), and hinge (H), in 31 combinations (e.g., L, LQ, LQP, LQPT, LQPTH, among many other combinations), and 10,000 background points. This resulted in 589 candidate models.

The candidate ENMs were evaluated based on the statistical significance of the partial Receiver Operating Characteristic (*pROC*; Peterson et al., 2008), 10,000 random points, 500 iterations, and 0.5% omission rates. We generated several metrics to rank the best models: AUC, omission rate, AIC_c, *delta* AIC_c, parameter number, and $wAIC_c$. The ENMs selected with the best parameters were constructed in the same polygon (i.e., M area) and then projected to current and past (LGM-MIROC) conditions in a polygon extending to Mexico and some areas of northern Central America with ten bootstrap iterations.

1.4 Elaboration of resistance distance matrices

For the first model, we calculated the LCP distance corrected for topography by using the DEM and the ENM of the current conditions applying the *topoLCP* function of 'topoDistance' v.1.0.1 (Wang, 2019). For the second (present) and third (LGM) models, we optimized both resistance surfaces (ENMs) in 'ResistanceGA' v.4.2 (Peterman, 2018) following random-walk commute time. Then, we calculated the effective distance between sample points on the true resistance surface as the random walk commute time between points using the 'gdistance' v.1.3-6 (van Etten, 2017) R-package. This measure is analogous to resistance distance calculated using 'CIRCUITSCAPE' (McRae, 2006), and represents the effective distance between points averaged over multiple pathways. The optimization procedure was performed twice using the *GA.prep* (probability of crossover = 0.95, probability of mutation = 0.500) and *SS_optim* functions (Peterman, 2018).

For the fourth (present) and fifth (LGM) models, we used the alternative optimization procedure in 'radish' v.0.0.2 (Pope & Peterman, 2020). This R package (Peterman & Pope, 2021) optimizes IBR models using maximum-likelihood estimation, where conductance is a function of spatial covariates, the observed data are GD, and the likelihood of the "measurement process" is cheap to compute (Pope & Peterman, 2020). All of the processes in ResistanceGA and radish were performed using an eight-neighbor R-connection scheme and optimized for each response variable (see *Landscape genomics analyses*) to assess functional connectivity.

1.5 Model selection

Because candidate models can often be equally parsimonious and therefore indistinguishable, we followed a multi-criteria approach for model selection using metrics

derived from maximum-likelihood population effects models (MLPE; AIC_c and cR^2) and multiple regression of distance matrices (MRDM; R^2).

First, we searched for parsimonious models selecting five candidate models out of a total of 9 models for each response variable with the lowest AIC_c values. Of these, we selected the final model with the highest cR^2 values. This approach allowed us to obtain the candidate model that best explained the neutral genetic variation, balancing between parsimony and statistical accuracy.

The alternative hypotheses (IBB, IBE, and IBR) were conditioned by including them as covariates to IBD in the MLPEs and MRDMs, as suggested in exploratory analysis of Mantel tests and partial Mantel tests using 'vegan' v.2.5-7 (Oksanen et al., 2020). To perform the multi-criteria approach, we fit maximum-likelihood population effects models (MLPE) to account for the non-independence of pair-wise distance data (Clarke et al., 2002). The "total evidence" assignment approach revealed several genetic clusters within the *Arremon* bird species complex (see Results). Thus, we also controlled for the underlying population structure of pairwise distance data (K_{ij}) as a random effect (1 | K_{ij}) in the MLPE analyses (Carvalho et al., 2019; Dalapicolla et al., 2021). Here, K is the optimal value of population structure employing the "total evidence" approach of a given sample, where sample "*i*" corresponds to the sample of reference and "*j*" the sample being compared in each pairwise distance comparison. Incorporating this random variable (K_{ij}) in the MLPE equations may minimize possible time-lag effects in LG studies (Epps & Keyghobadi, 2015). In addition, this procedure increased the number of observations and statistical power of our study. The correlation structure of the random effects for the MLPEs was implemented using the R-code 'corMLPE' v.0.0.3 (Pope, 2021). We applied the MLPEs with the "REML = TRUE" option by fitting linear mixed-effects models (LME) in the *lme* function available for the R-library 'nlme' v.3.1-152 (Pinheiro et al., 2021). We calculated AIC_c (Burnham & Anderson, 2002) and the conditional R^2 (cR^2; Nakagawa et al., 2017) to estimate the probability and variance explained by each candidate model using the R-package 'MuMIn' v.1.43.17 (Barton, 2020). For the MLPEs that did not control for population structure, we fitted generalized least squares (GLS) in the *gls* function of the R-library 'nlme' v.3.1-152 (Pinheiro et al., 2021). Finally, we used multiple regression

of distance matrices (MRDM; Legendre et al., 1994) to determine the relative importance (R^2) of predictors in explaining GD using the *MRM* function of 'ecodist' v.2.0.7 (Goslee & Urban, 2007). This method has been shown to have a good balance between type 1 error and power, and it has high levels of accuracy compared with other methods (Balkenhol et al., 2009).

2. Results

2.1 Standardization of assembly parameters in ipyrad

Table S2. Different runs performed in the ipyrad pipeline to call SNPs on brushfinches from northern Mesoamerica. In bold is the name of ipyrad run employed in subsequent landscape genomic analyses.

ipyrad runs	Number of SNPs	Missing data (%)
run 1	896,407	69.02
run 2	914,699	69.13
run 3	930,751	69.19
run 4	938,140	69.46
run 5	911,789	70.34
run 6	864,720	71.39
run 7	678,549	62.70
run 8	343,423	50.21
run 9	643,589	63.34
run 10	310,726	50.37
run 11	1,261,944	35.07
run 12	7,585	25.28
run 13	631,142	63.40
run 14	**315,693**	**51.31**

Figure S1. Genotype depth and percentage of missing data for each sample of the selected ipyrad output ("run14").

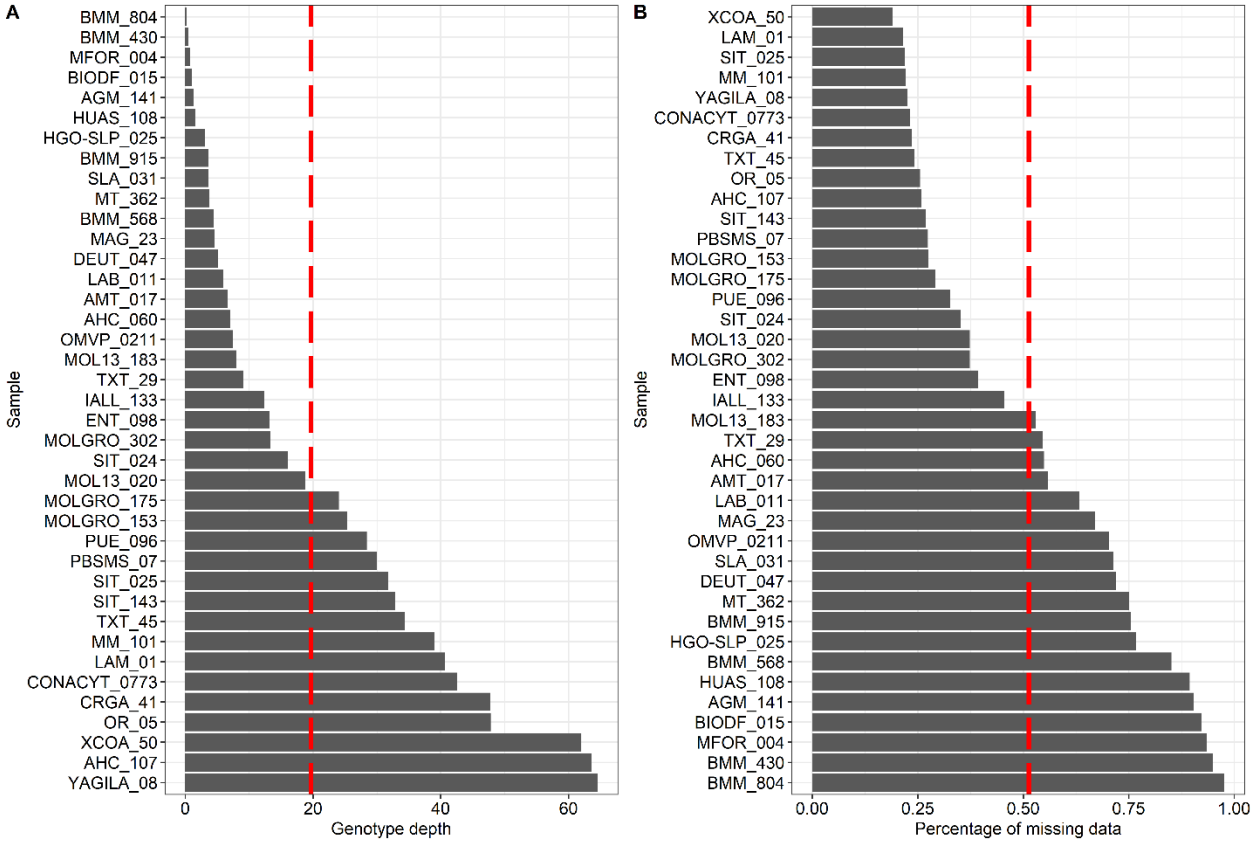


Figure S2. Boxplots plots of Euclidean genetic distances between samples of the unfiltered genomic data sets as a function of different values of two ipyrad parameters set to call SNPs on brushfinches from northern Mesoamerica. Red dots indicate mean Euclidean genetic distance values for each run.

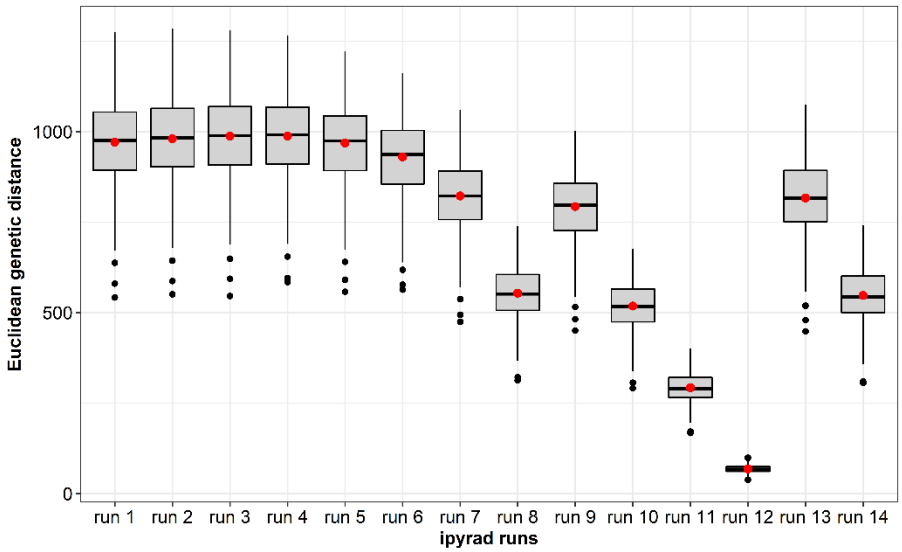
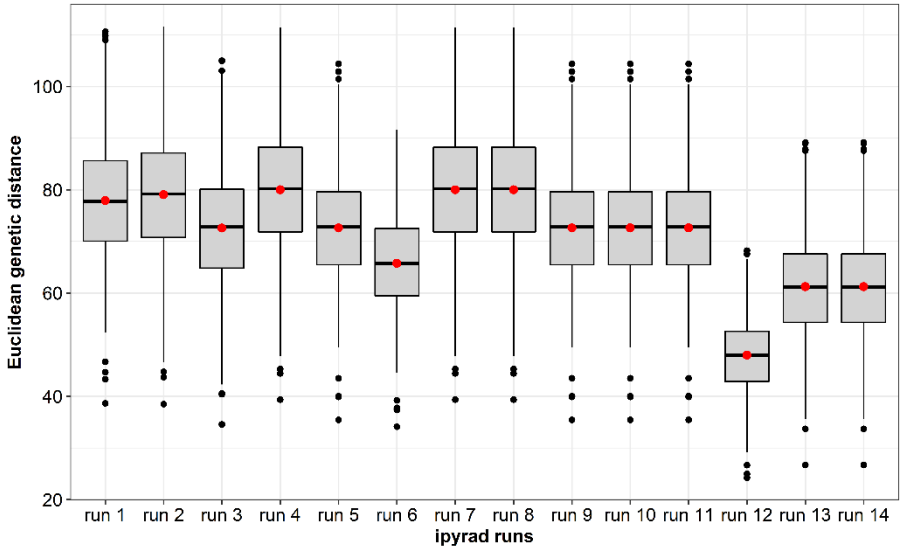


Figure S3. Boxplots plots of Euclidean genetic distances between samples of the genomic data set after applying a threshold of 25% of missing data (i.e., SNPs) allowed, and considering different values of two ipyrad parameters set to call SNPs on brushfinches from northern Mesoamerica. Red dots indicate mean Euclidean genetic distance values for each run.



2.2 Population structure using model-based assignment methods and landscape genomic analyses

Figure S4. Diagnostic plots of GENELAND results showing log probability posterior after burn-in across independent runs (A), number of classes (i.e., clusters) where MCMC chain visited (B), and the density of number of clusters along the chain after burn-in (C). Diagnostic plot of the fine-tuning process implemented in TESS3R depicting the optimal *K* according to cross-validation scores (D).

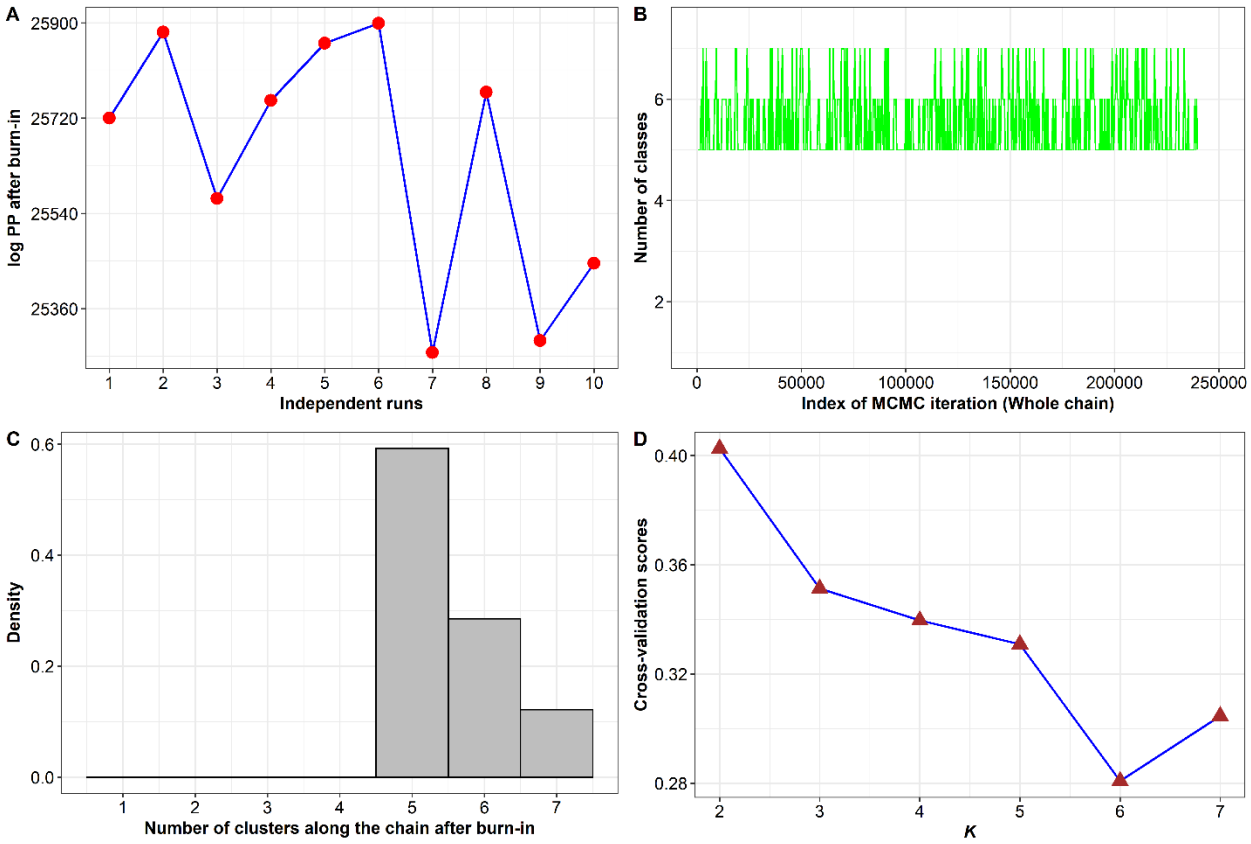


Figure S5. Diagnostic plots of the fine-tuning process in STRUCTURE according to ln Pr(*X*|*K*) values (A), L' [*K*] ± SD values (B), |L'' [*K*]| ± SD values (C), and Δ*K* scores (D). Yellow color lines and dots represent the admixture model with the correlated allele frequencies (CAF) model with default prior. Purple color lines and dots indicate the admixture model with the uncorrelated allele frequencies (UAF) model with default prior, and blue color lines and dots depict the admixture model with the CAF model with alternative prior.

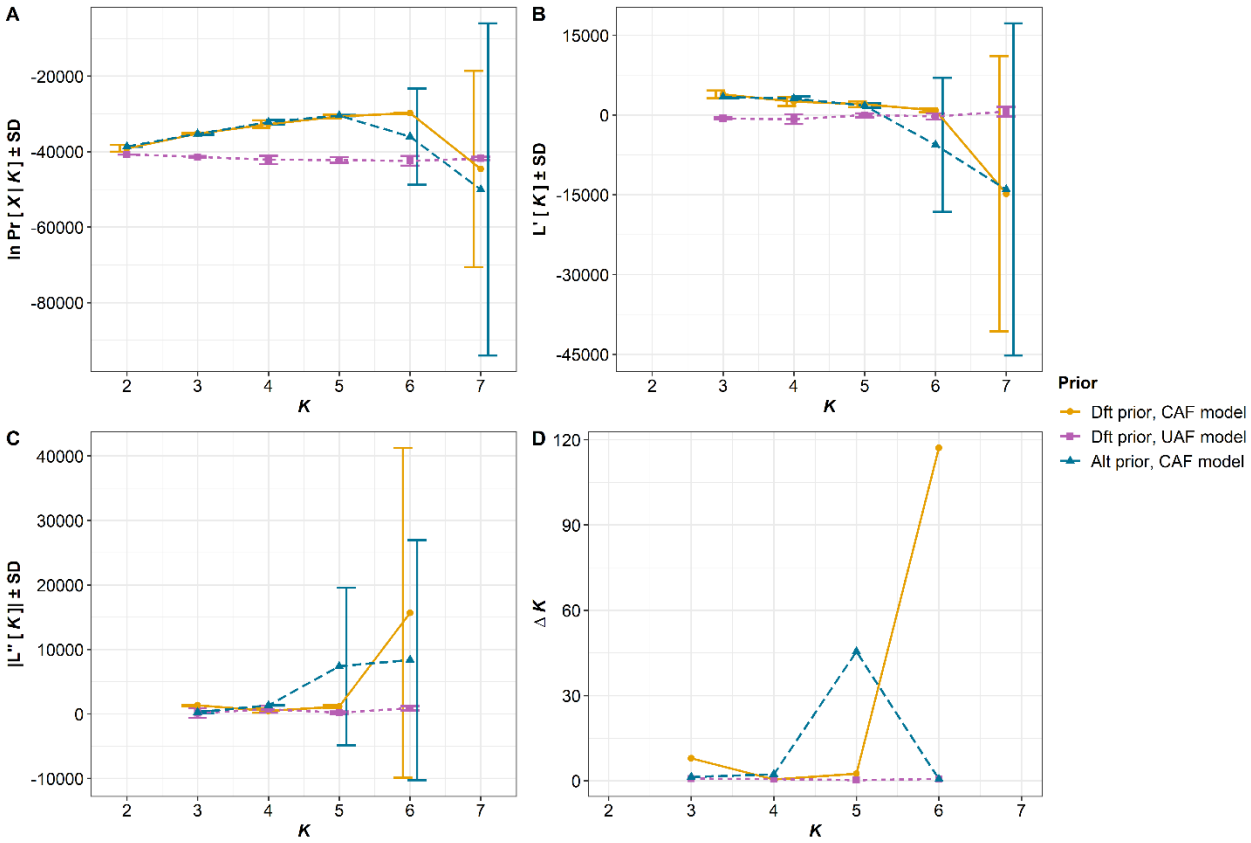


Figure S6. Bar plots of the STRUCTURE analysis using correlated allele frequencies and alternative prior. STRUCTURE results across independent runs were aligned in CLUMPP, where (A) shows *K* = 3, (B) *K* = 4, (C) *K* = 5, and (D) *K* = 6. All figures display admixture values for each individual.

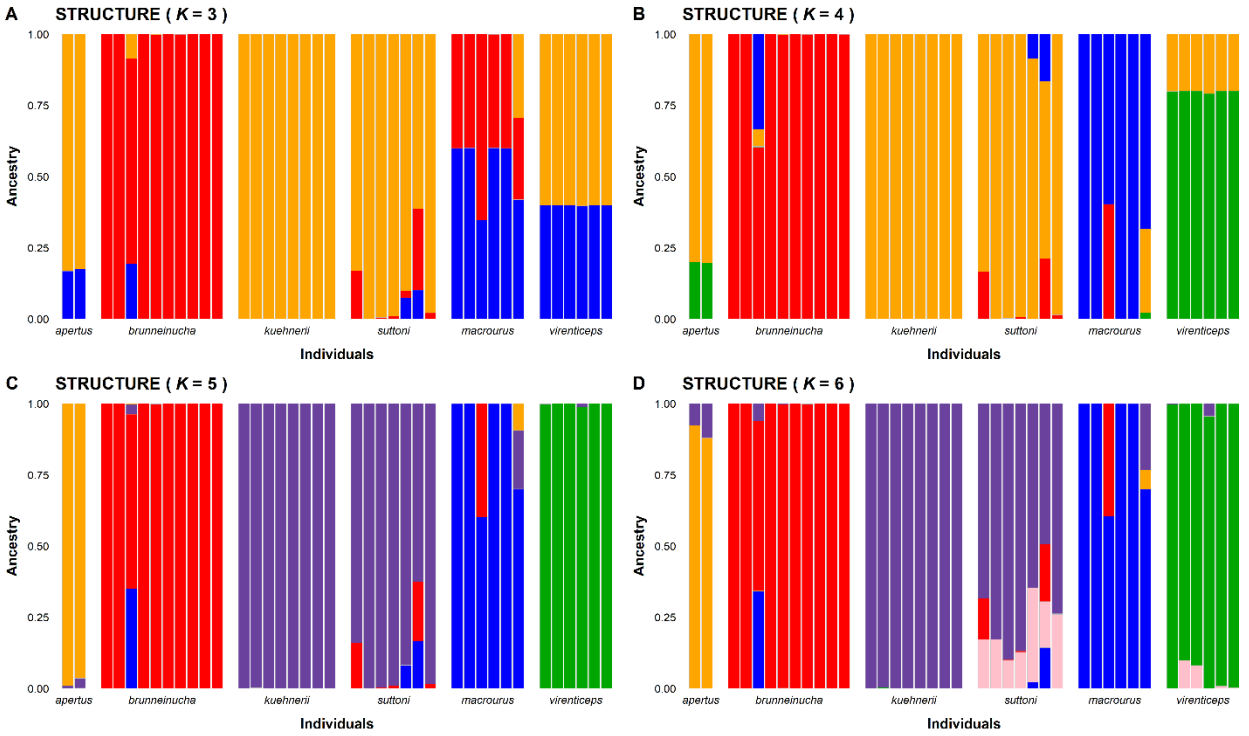


Figure S7. Bivariate plots showing normal PCA and sPCA outputs (considering the first three axes) side-by-side. Colored dots represent identified genetic clusters according to the "total evidence" approach (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS = Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt

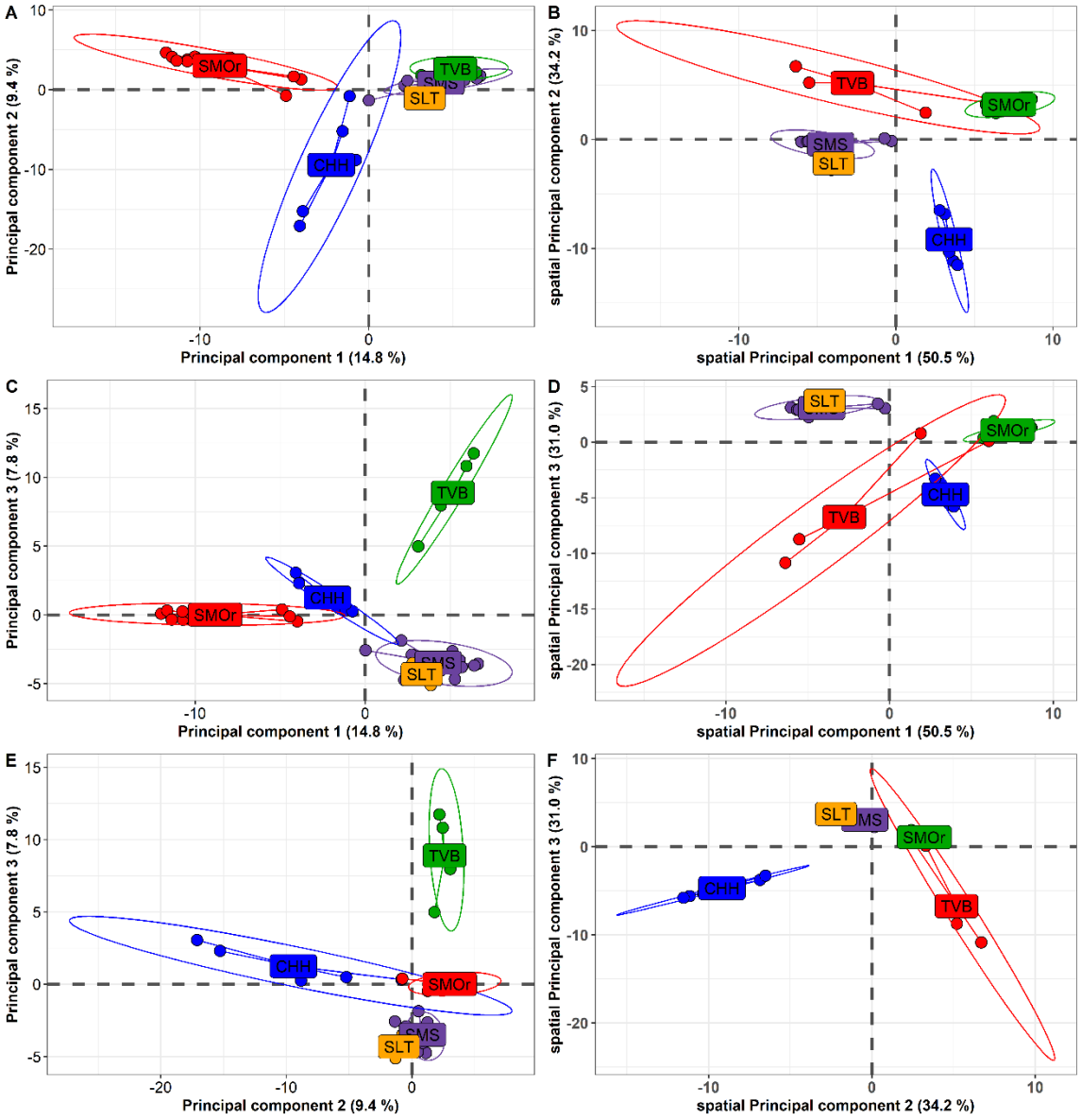


Figure S8. Interpolation maps of the sPCA components of the first component (A), second component (B), third component (C), and inter-individual pair-wise genetic distance representing proportion of shared alleles (D_{PS}, D), and Euclidean distance of a PCA retaining the first 10 axes ($Euclidean_{PCA\text{-}10\ axes}$, E). sPCA and genetic distances were calculated using 2,634 filtered neutral SNP data (excluding outliers) from all samples (n = 39). The dotted line encompasses our study area. Colored dots represent identified genetic clusters according to the "total evidence" approach (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS = Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt.

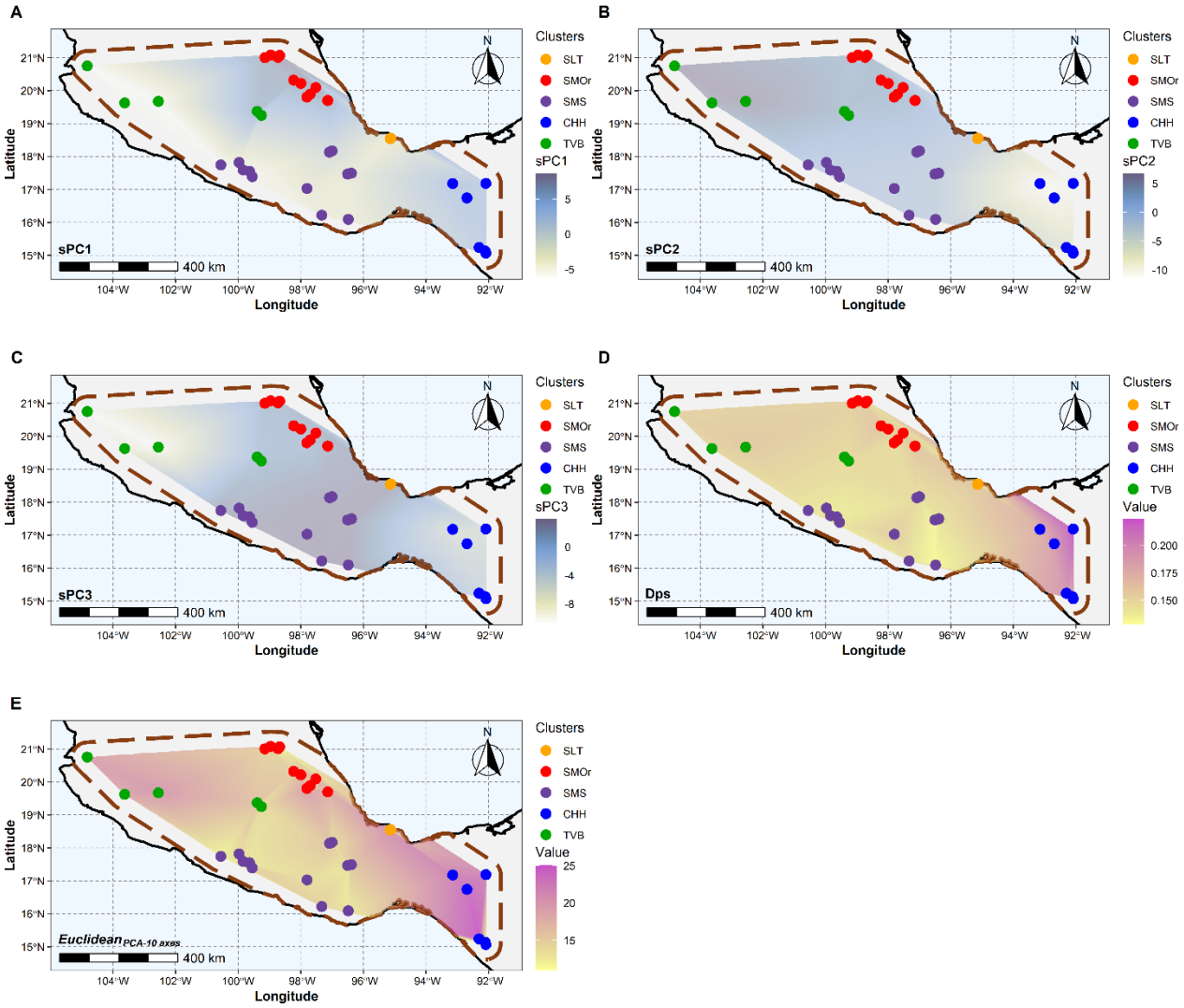


Figure S9. Plots of linear relationships representing the null hypotheses (IBD) by using the genetic response variable of (A) the proportion of shared alleles (D_{PS}), and (B) Euclidean distance of a PCA retaining the first 10 axes ($Euclidean_{PCA\text{-}10\ axes}$). The red or green lines indicate the observed trend (displaying 95% confidence intervals), while the black line represents the expected trend. Dot colors indicate the pairwise IBD values and the Density bar values illustrate the point conglomerations for its model respectively. * = statistical significance ($p < 0.05$).

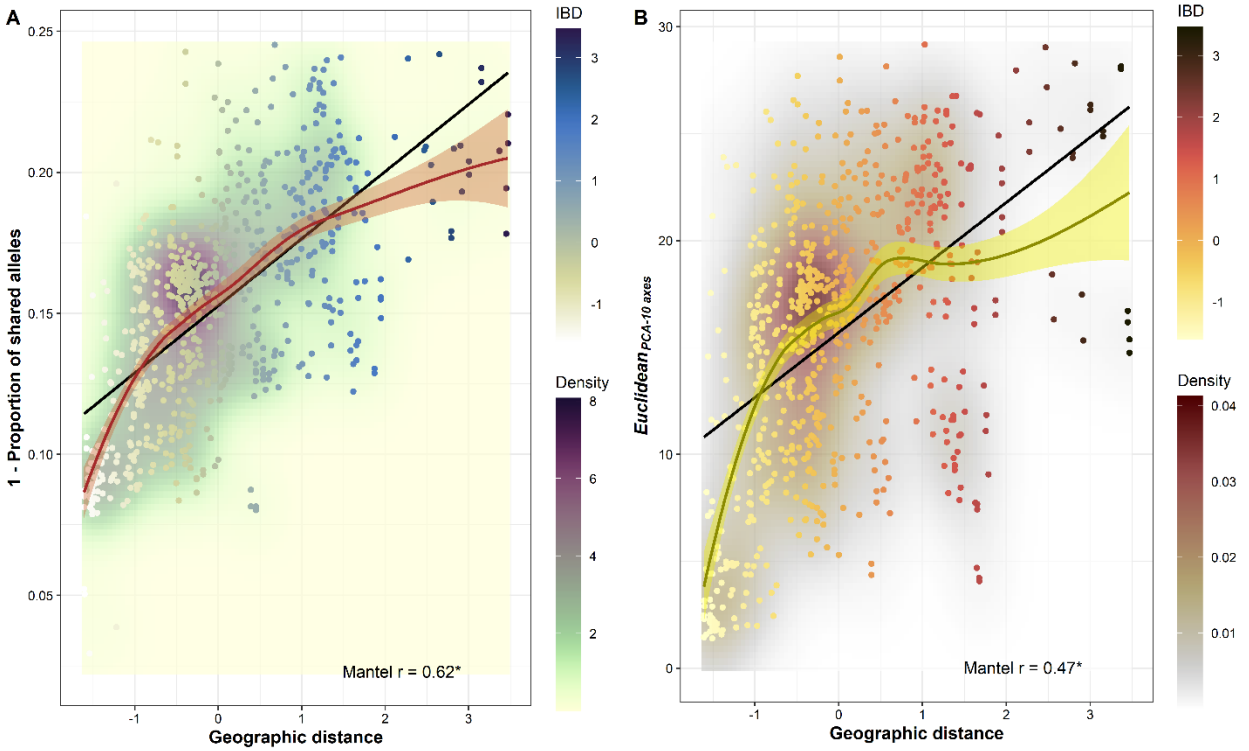


Figure S10. (A) Map detailing the presence (= 1) and absence (= 0) of lowland barriers limiting gene flow; (B) Map describing the spatial arraignment of the BIO12 (annual precipitation) values from the CHELSA project; (C) Clipped raster layer of the fine-tuned ecological niche model (ENM) during current conditions for the *Arremon* bird complex based on maximum entropy. Red lines represent the topographic least-cost paths (tLCP); (D) Clipped raster layer of the fine-tuned ecological niche model (ENM) during LGM-MIROC conditions for the *Arremon* bird complex based on maximum entropy. The dotted line encompasses our study area. Colored dots represent identified genetic clusters according to the "total evidence" approach (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS = Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt.

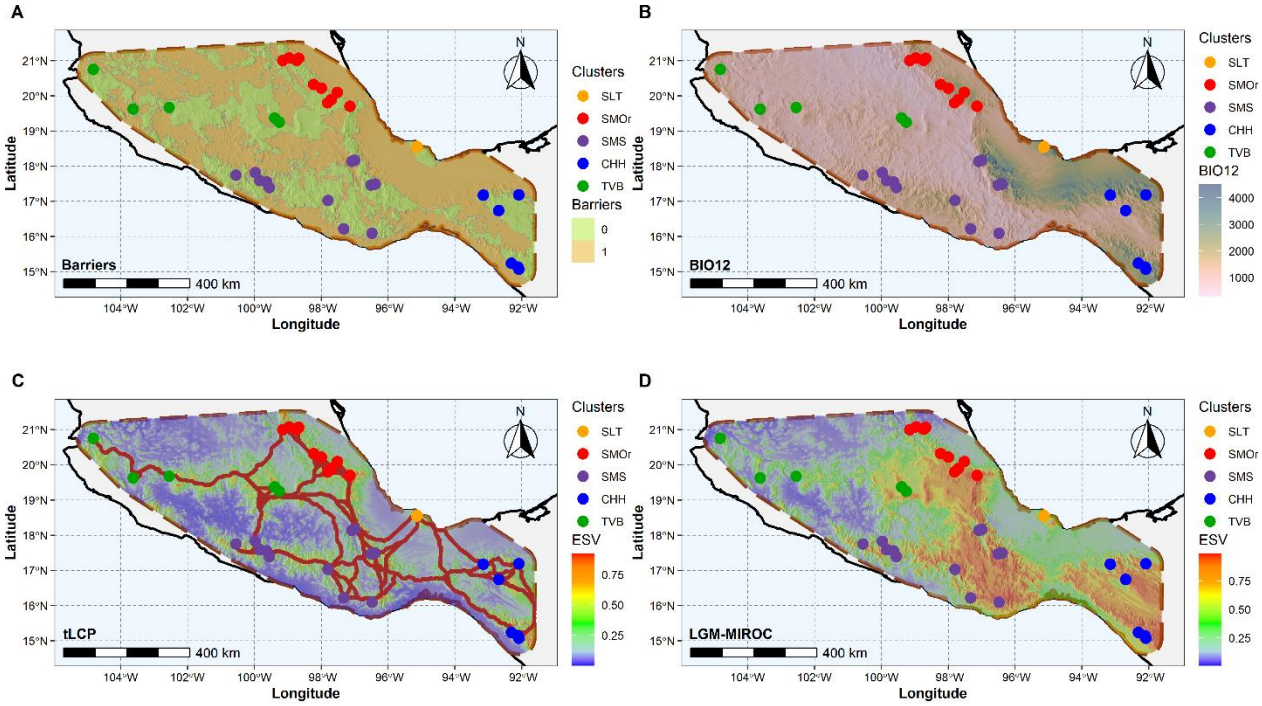


Figure S11. (A) Surface resistance layer of the ecological niche model (ENM) during current conditions, and optimized in ResistanceGA using proportion of shared alleles (D_{PS}) as response variable; (B) Surface resistance layer of the ecological niche model (ENM) during LGM-MIROC conditions, and optimized in ResistanceGA using proportion of shared alleles (D_{PS}) as response variable; (C) Surface resistance layer of the ecological niche model (ENM) during current conditions, and optimized in radish using proportion of shared alleles (D_{PS}) as response variable; (D) Surface resistance layer of the ecological niche model (ENM) during current conditions, and optimized in radish using Euclidean distance of a PCA retaining the first 10 axes ($Euclidean_{PCA\text{-}10\ axes}$) as response variable. The dotted line encompasses our study area. Colored dots represent identified genetic clusters according to the "total evidence" approach (see Results section): SLT = Sierra de Los Tuxtlas, SMOr = Sierra Madre Oriental, SMS = Sierra Madre del Sur, CHH = Chiapas Highlands, TVB = Transmexican Volcanic Belt.

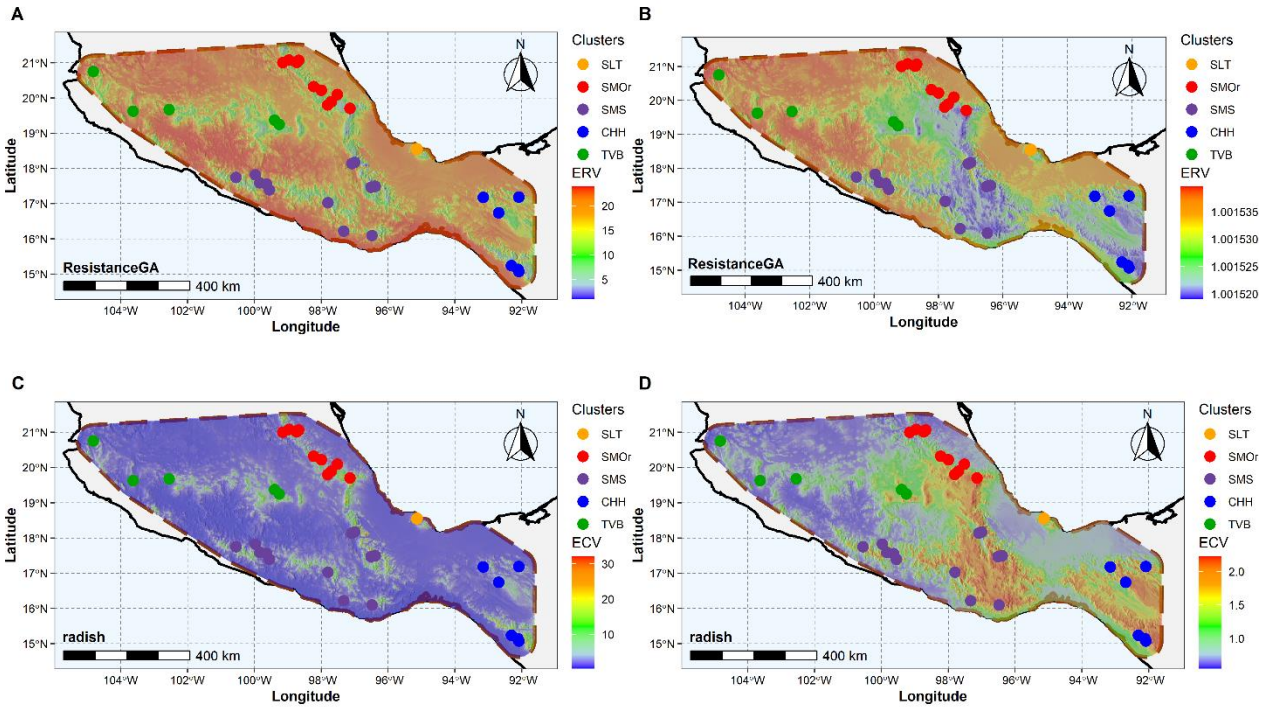


Table S3. Summary statistics of MLPE models without controlling for population structure (*gls* function) using the proportion of shared alleles (D_{PS}) as the response variable (GD), and the predictors representing different genetic isolation hypotheses. For each model we provide estimates, standard errors (SE), degrees of freedom (df), *t*-values, *p*-values, and confidence intervals (CI). Predictor variable abbreviations are the same as Table 1. Abbreviations for the predictor variables are: CS = correlation structure, GeoD = geographic distance, NB = number of barriers, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish.

Hypothesis	**Predictor**	**Estimate**	**SE**	**df**	***t***	***p***	**CI (min/max)**	***AIC*$_c$**
IBD	Intercept	0.152	0.005	739	30.137	0	0.142/0.162	-3482.891
	GeoD	0.022	0.001	739	21.068	0	0.020/0.024	
	CS	0.264					0.204/0.322	
IBB | IBD	Intercept	0.152	0.005	738	30.015	0	0.142/0.162	-3499.191
	NB	0.008	0.001	738	5.480	0	0.005/0.011	
	GeoD	0.015	0.001	738	9.826	0	0.012/0.018	
	CS	0.270					0.211/0.326	

IBE_{global} | IBD	Intercept	0.152	0.005	738	29.858	0.000	0.142/0.162	-3470.111
	ED_{global}	-0.001	0.001	738	-0.969	0.332	-0.003/0.001	
	GeoD	0.022	0.001	738	19.600	0.000	0.020/0.025	
	CS	0.266					0.207/0.324	
IBE_{bio12} | IBD	Intercept	0.152	0.005	738	29.392	0.000	0.142/0.162	-3474.076
	ED_{bio12}	-0.002	0.001	738	-2.219	0.026	-0.004/-0.0003	
	GeoD	0.023	0.001	738	20.287	0.000	0.021/0.025	
	CS	0.271					0.211/0.329	
IBR_{tlcp} | IBD	Intercept	0.152	0.005	738	27.041	0.000	0.141/0.163	-3611.872
	RD_{tlcp}	0.032	0.002	738	12.596	0.000	0.027/0.037	
	GeoD	-0.006	0.002	738	-2.578	0.010	-0.011/-0.001	
	CS	0.317					0.258/0.368	
IBR_{PREga} | IBD	Intercept	0.152	0.005	738	26.664	0.000	0.141/0.163	-3541.55
	RD_{PREga}	0.022	0.002	738	8.740	0.000	0.017/0.027	
	GeoD	0.004	0.002	738	2.035	0.042	0.0001/0.009	
	CS	0.308					0.249/0.361	
IBR_{LGMga} | IBD	Intercept	0.152	0.005	738	30.186	0.000	0.142/0.162	-3519.848

	RD_{LGMga}	0.016	0.002	738	7.135	0.000	0.012/0.021	
	GeoD	0.006	0.002	738	2.872	0.004	0.002/0.011	
	CS	0.272					0.213/0.329	
IBR_{PRErad} | IBD	Intercept	0.152	0.005	738	29.977	0.000	0.142/0.162	-3505.24
	RD_{PRErad}	0.020	0.003	738	5.891	0.000	0.013/0.026	
	GeoD	0.006	0.002	738	2.491	0.012	0.001/0.012	
	CS	0.271					0.213/0.327	
IBR_{LGMrad} | IBD	Intercept	0.152	0.004	738	32.82	0.000	0.143/0.161	-3494.814
	RD_{LGMrad}	0.016	0.003	738	4.917	0.000	0.010/0.023	
	GeoD	0.007	0.003	738	2.328	0.020	0.001/0.013	
	CS	0.245					0.202/0.288	

Table S4. Summary statistics of the MLPE models without controlling for population structure (*gls* function) using Euclidean distance of a PCA retaining the first 10 axes ($Euclidean_{PCA\text{-}10\ axes}$) as the response variable (GD), and the predictors representing different genetic isolation hypotheses. For each model we provide estimates, standard errors (SE), degrees of freedom (df), *t*-values, *p*-values, confidence intervals (CI), and conditional R^2. Predictor variable abbreviations are the same as Table 1. Abbreviations for the predictor variables are: CS = correlation structure, GeoD = geographic distance, NB = number of barriers, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish.

Hypothesis	**Predictor**	**Estimate**	**SE**	**df**	***t***	***p***	**CI (min/max)**	***AIC*$_c$**
IBD	Intercept	15.695	1.069	739	14.672	0	13.595/17.795	4016.4
	GeoD	2.660	0.168	739	15.828	0	2.330/2.991	
	CS	0.335					0.279/0.382	
IBB | IBD	Intercept	15.695	1.074	738	14.611	0	13.586/17.804	3995.437
	NB	1.210	0.245	738	4.934	0	0.728/1.691	
	GeoD	1.714	0.253	738	6.774	0	1.217/2.211	
	CS	0.339					0.284/0.386	

IBE_{global} | IBD	Intercept	15.695	1.080	738	14.531	0.000	13.575/17.815	4017.743
	ED_{global}	-0.275	0.184	738	-1.496	0.134	-0.636/0.085	
	GeoD	2.775	0.184	738	15.035	0.000	2.412/3.137	
	CS	0.337					0.282/0.384	
IBE_{bio12} | IBD	Intercept	15.695	1.095	738	14.330	0.000	13.545/17.845	4011.143
	ED_{bio12}	-0.559	0.187	738	-2.983	0.002	-0.927/-0.191	
	GeoD	2.875	0.182	738	15.793	0.000	2.518/3.233	
	CS	0.342					0.286/0.388	
IBR_{tlcp} | IBD	Intercept	15.695	1.019	738	15.393	0	13.693/17.697	3849.5
	RD_{tlcp}	5.512	0.400	738	13.781	0	4.727/6.298	
	GeoD	-2.230	0.385	738	-5.791	0	-2.986/-1.474	
	CS	0.350					0.296/0.394	
IBR_{PREga} | IBD	Intercept	15.695	1.260	738	12.451	0.000	13.220/18.170	3837.769
	RD_{PREga}	3.712	0.254	738	14.575	0.000	3.212/4.212	
	GeoD	0.334	0.216	738	1.540	0.123	-0.091/0.759	
	CS	0.394					0.349/0.428	
IBR_{LGMga} | IBD	Intercept	15.695	1.120	738	14.011	0.000	13.496/17.894	3880.291

	RD_{LGMga}	2.393	0.192	738	12.455	0.000	2.016/2.770	
	GeoD	0.511	0.229	738	2.224	0.026	0.060/0.962	
	CS	0.366					0.315/0.407	
IBR_{PRErad} | IBD	Intercept	15.695	1.080	738	14.528	0.000	13.574/17.816	4002.375
	RD_{PRErad}	1.885	0.473	738	3.986	0.0001	0.957/2.814	
	GeoD	1.076	0.430	738	2.498	0.012	0.230/1.921	
	CS	0.339					0.284/0.386	
IBR_{LGMrad} | IBD	Intercept	15.695	1.118	738	14.129	0.000	13.514/17.876	3993.909
	RD_{LGMrad}	2.737	0.553	738	4.943	0.000	1.650/3.825	
	GeoD	0.706	0.428	738	1.650	0.099	-0.133/1.547	
	CS	0.347					0.293/0.392	

Table S5. Summary statistics of MLPE models (*gls* function) including only *brunneinucha* samples (*n* = 10) of the Sierra Madre Oriental. The MLPEs were fitted using the proportion of shared alleles (D_{PS}) as the response variable (GD), and the predictors representing different genetic isolation hypotheses. For each model we provide estimates, standard errors (SE), degrees of freedom (df), *t*-values, *p*-values, and confidence intervals (CI). Predictor variable abbreviations are the same as Table 1. Abbreviations for the predictor variables are: CS = correlation structure, GeoD = geographic distance, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish.

Hypothesis	**Predictor**	**Estimate**	**SE**	**df**	***t***	***p***	**CI (min/max)**	***AIC*$_c$**
IBD	Intercept	0.103	0.007	43	13.215	0.000	0.087/0.119	-280.883
	GeoD	0.0005	0.0008	43	0.710	0.4812	-0.001/0.002	
	CS	0.460					0.401/0.485	
IBE_{global} | IBD	Intercept	0.103	0.007	42	13.301	0.000	0.087/0.119	-266.979
	ED_{global}	-0.0008	0.0009	42	-0.884	0.381	-0.002/0.001	
	GeoD	0.0005	0.0008	42	0.588	0.559	-0.001/0.002	
	CS	0.460					0.399/0.485	
IBE_{bio12} | IBD	Intercept	0.103	0.007	42	13.303	0.000	0.087/0.119	-267.003
	ED_{bio12}	-0.0008	0.0009	42	-0.890	0.378	-0.002/0.001	

	GeoD	0.0004	0.0008	42	0.578	0.566	-0.001/0.002	
	CS	0.460					0.399/0.485	
IBR_{tlcp} | IBD	Intercept	0.103	0.007	42	13.264	0.000	0.087/0.119	-272.765
	RD_{tlcp}	0.014	0.014	42	1.049	0.299	-0.013/0.043	
	GeoD	-0.014	0.014	42	-1.005	0.320	-0.042/-0.014	
	CS	0.460					0.400/0.485	
IBR_{PREga} | IBD	Intercept	0.103	0.007	42	13.188	0.000	0.087/0.119	-271.272
	RD_{PREga}	0.004	0.010	42	0.420	0.676	-0.017/0.026	
	GeoD	-0.003	0.010	42	-0.361	0.719	-0.024/0.017	
	CS	0.460					0.399/0.485	
IBR_{LGMga} | IBD	Intercept	0.103	0.008	42	12.885	0.000	0.087/0.119	-270.130
	RD_{LGMga}	0.003	0.005	42	0.638	0.526	-0.007/0.014	
	GeoD	-0.001	0.005	42	-0.385	0.702	-0.006/0.004	
	CS	0.462					0.403/0.486	
IBR_{PRErad} | IBD	Intercept	0.103	0.007	42	13.173	0.000	0.087/0.119	-271.297
	RD_{PRErad}	0.004	0.010	42	0.425	0.672	-0.017/0.026	
	GeoD	-0.003	0.010	42	-0.366	0.715	-0.024/0.017	
	CS	0.460					0.400/0.485	
IBR_{LGMrad} | IBD	Intercept	0.103	0.008	42	12.649	0.000	0.087/0.120	-270.143
	RD_{LGMrad}	0.002	0.005	42	0.486	0.629	-0.009/0.014	
	GeoD	-0.0006	0.002	42	-0.248	0.805	-0.006/0.004	
	CS	0.463					0.406/0.487	

Table S6. Summary statistics of MLPE models (*gls* function) including only *brunneinucha* samples (*n* = 10) of the Sierra Madre Oriental. The MLPEs were fitted using Euclidean distance of a PCA retaining the first 9 axes ($Euclidean_{PCA\text{-}9\ axes}$) as the response variable (GD), and the predictors representing different genetic isolation hypotheses. For each model we provide estimates, standard errors (SE), degrees of freedom (df), *t*-values, *p*-values, confidence intervals (CI), and conditional R^2. Predictor variable abbreviations are the same as Table 1. Abbreviations for the predictor variables are: CS = correlation structure, GeoD = geographic distance, RD_{tlcp} = resistance distance based on topographic least-cost path, ED_{global} = environmental distance including all non-correlated bioclimatic variables, ED_{bio12} = environmental distance including only bioclimatic 12 variable, RD_{PREga} = resistance distance of ENM during current conditions and optimized on ResistanceGA, RD_{LGMga} = resistance distance of ENM during LGM-MIROC conditions and optimized on ResistanceGA, RD_{PRErad} = resistance distance of ENM during current conditions and optimized on radish, RD_{LGMrad} = resistance distance of ENM during LGM-MIROC conditions and optimized on radish.

Hypothesis	**Predictor**	**Estimate**	**SE**	**df**	***t***	***p***	**CI (min/max)**	***AIC*$_c$**
IBD	Intercept	17.299	0.901	43	19.196	0.000	15.482/19.117	86.342
	GeoD	0.093	0.053	43	1.761	0.085	-0.013/0.200	
	CS	0.487					0.466/0.495	
IBE_{global} | IBD	Intercept	17.299	0.901	42	19.199	0.000	15.481/19.118	92.324
	ED_{global}	-0.037	0.058	42	-0.633	0.529	-0.155/0.081	
	GeoD	0.089	0.053	42	1.651	0.106	-0.019/0.198	
	CS	0.487					0.465/0.495	

IBE_{bio12} | IBD	Intercept	17.299	0.901	42	19.196	0.000	15.481/19.118	92.304
	ED_{bio12}	-0.037	0.058	42	-0.637	0.527	-0.156/0.081	
	GeoD	0.088	0.054	42	1.642	0.107	-0.020/0.197	
	CS	0.487					0.465/0.495	
IBR_{tlcp} | IBD	Intercept	17.299	0.887	42	19.482	0.000	15.507/19.091	85.279
	RD_{tlcp}	1.271	0.904	42	1.405	0.167	-0.554/3.096	
	GeoD	-1.166	0.897	42	-1.298	0.201	-2.978/0.645	
	CS	0.487					0.465/0.495	
IBR_{PREga} | IBD	Intercept	17.299	0.866	42	19.955	0.000	15.550/19.049	88.432
	RD_{PREga}	0.298	0.366	42	0.812	0.420	-0.442/1.038	
	GeoD	-0.072	0.210	42	-0.342	0.733	-0.497/0.353	
	CS	0.486					0.461/0.495	
IBR_{LGMga} | IBD	Intercept	17.299	0.863	42	20.025	0.000	15.556/19.043	88.156
	RD_{LGMga}	0.325	0.458	42	0.708	0.482	-0.600/1.251	
	GeoD	-0.060	0.223	42	-0.269	0.788	-0.511/0.390	
	CS	0.485					0.458/0.495	
IBR_{PRErad} | IBD	Intercept	17.299	0.863	42	20.026	0.000	15.556/19.042	86.642

	RD_{PRErad}	0.702	0.551	42	1.274	0.209	-0.409/1.814	
	GeoD	0.009	0.084	42	0.112	0.910	-0.161/0.180	
	CS	0.486					0.461/0.495	
IBR_{LGMrad} | IBD	Intercept	17.299	0.888	42	19.463	0.000	15.505/19.093	87.668
	RD_{LGMrad}	0.412	0.560	42	0.736	0.465	-0.718/1.544	
	GeoD	0.009	0.126	42	0.073	0.941	-0.245/1.544	
	CS	0.486					0.462/0.495	

References

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541–545. https://doi.org/10.1111/ecog.01132

Balkenhol, N., Waits, L. P., & Dezzani, R. J. (2009). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, *32*(5), 818–830. https://doi.org/10.1111/j.1600-0587.2009.05807.x

Barton, K. (2020). *MuMIn: Multi-model inference*. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*(11), 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Bivand, R., & Rundel, C. (2020). *rgeos: Interface to geometry engine - open source ('GEOS')*. R package version 0.5-5. https://CRAN.R-project.org/package=rgeos

Bivand, R., Keitt, T., & Rowlingson, B. (2021). *rgdal: Bindings for the 'geospatial' data abstraction library*. R package version 1.5-23. https://CRAN.R-project.org/package=rgdal

Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, *275*, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag.

Carvalho, C. S., Lanes, É. C. M., Silva, A. R., Caldeira, C. F., Carvalho-Filho, N., Gastauer, M., Imperatriz-Fonseca, V. L., Nascimento, W. Jr., Oliveira, G., Siqueira, J. O., Viana, P. L., & Jaffé, R. (2019). Habitat loss does not always entail negative

genetic consequences. *Frontiers in Genetics*, *10*, 1011. https://doi.org/10.3389/fgene.2019.01101

Castaño, M. I., Cadena, C. D., & Avendaño, J. E. (2019). Home-range size of an Andean bird: Assessing the role of physical condition. *Biotropica*, *51*(4), 591–599. https://doi.org/10.1111/btp.12673

Castaño-Quintero, S., Escobar-Luján, J., Osorio-Olvera, L., Peterson, A. T., Chiappa-Carrara, X., Martínez-Meyer, E., & Yañez-Arenas, C. (2020). Supraspecific units in correlative niche modeling improves the prediction of geographic potential of biological invasions. *PeerJ*, *8*, e10454. https://doi.org/10.7717/peerj.10454

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. https://doi.org/10.1111/1755-0998.12471

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2022). *rgbif: Interface to the Global Biodiversity Information Facility API*. R package version 3.6.0. https://CRAN.R-project.org/package=rgbif

Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, *7*(3), 361–372. https://doi.org/10.1198/108571102320

Cobos, M. E., Peterson, A. T., Barve, N., & Osorio-Olvera, L. (2019). kuenm: an R package for detailed development of ecological niche models using Maxent. *PeerJ*, *7*, e6281. https://doi.org/10.7717/peerj.6281

Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct *K* value using the Δ*K* method: When does *K* = 2? *Molecular Ecology*, *29*(5), 862–869. https://doi.org/10.1111/mec.15374

Dalapicolla, J., do Prado, J. R., Percequillo, A. R., & Knowles, L. L. (2021). Functional connectivity in sympatric spiny rats reflects different dimensions of Amazonian

forest-association. *Journal of Biogeography*, 48(12), 3196–3209. https://doi.org/10.1111/jbi.14281

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, *41*(1), 214–226. https://doi.org/10.1214/aoms/1177697203

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, *24*(24), 6021–6040. https://doi.org/10.1111/mec.13454

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Flantua, S., & Hooghiemstra, H. (2018). Historical connectivity and mountain biodiversity (with special reference to the Northern Andes). In C. Hoorn, A. Perrigo, & A. Antonelli (Eds.), *Mountains, Climate and Biodiversity* (pp. 171–185). John Wiley & Sons.

Francis, R. M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

François, O., & Durand, E. (2010). Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, *10*(5), 773–784. https://doi.org/10.1111/j.1755-0998.2010.02868.x

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19. https://doi.org/10.18637/jss.v022.i07

Guillot, G., Mortier, F., & Estoup, A. (2005). Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, *5*(3), 712–715. https://doi.org/10.1111/j.1471-8286.2005.01031.x

Hasumi, H., Emori, S. (2004). *K-1 coupled GCM (MIROC) description*. Center for Climate System Research.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2020). *dismo: Species distribution modeling*. R package version 1.3-3. https://CRAN.R-project.org/package=dismo

Hijmans, R. J. (2021). *raster: geographic data analysis and modeling*. R package version 3.4-13. https://CRAN.R-project.org/package=raster

Howell, S. N. G., & Webb, S. (1995). *A guide to the birds of Mexico and northern Central America*. Oxford University Press.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, *101*(1), 92–103. https://doi.org/10.1038/hdy.2008.34

Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *The R Journal*, *5*(1), 144–161. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmerman, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Legendre, P., Lapointe, F.-J., & Casgrain, P. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, *48*(5), 1487–1499. https://doi.org/10.1111/j.1558-5646.1994.tb02191.x

Machado-Stredel, F., Cobos, M. E., & Peterson, A. T. (2021a). A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. *Frontiers of Biogeography*, *13*(4), e48814. https://doi.org/10.21425/F5FBG48814

Machado-Stredel, F., Cobos, M. E., & Peterson, A. T. (2021b). *grinnell: Dispersal simulations based on ecological niches*. R package version 0.0.21. https://github.com/fmachados/grinnell

Márquez, R., Linderoth, T. P., Mejía-Vargas, D., Nielsen, R., Amézquita, A., & Kronforst, M. R. (2020). Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of colour pattern variation in *Phyllobates* poison-dart frogs. *Molecular Ecology*, *29*(19), 3702–3719. https://doi.org/10.1111/mec.15598

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

Montano, V., & Jombart, T. (2017). An eigenvalue test for spatial principal component analysis. *BMC Bioinformatics*, *18*(1), 562. https://doi.org/10.1186/s12859-017-1988-y

Moreno-Contreras, I., Sánchez-González, L. A., Arizmendi M. C., Prieto-Torres, D. A., & Navarro-Sigüenza, A. G. (2020). Climatic niche evolution in the *Arremon brunneinucha* complex (Aves: Passerellidae) in a Mesoamerican landscape. *Evolutionary Biology*, *47*(2), 123–132. https://doi.org/10.1007/s11692-020-09498-7

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Navarro-Sigüenza, A. G., Peterson, A. T., Nyari, A., García-Deras, G. M., & García-Moreno, J. (2008). Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. *Molecular Phylogenetics and Evolution*, *47*(1), 21–35. https://doi.org/10.1016/j.ympev.2007.11.030

Navarro-Sigüenza, A. G., García-Hernández, M. A., & Peterson, A. T. (2013). A new species of brush-finch (*Arremon*; Emberizidae) from western Mexico. *Wilson Journal of Ornithology*, *125*(3), 443–453. https://doi.org/10.1676/12-136.1

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *vegan: Community ecology package*. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Paynter, R. A., Jr. (1978). Biology and evolution of the avian genus *Atlapetes* (Emberizinae). *Bulletin of the Museum of Comparative Zoology*, *148*(7), 323–369.

Pebesma, E. J. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*(6), 1638–1647. https://doi.org/10.1111/2041-210X.12984

Peterman, W. E., & Pope, N. S. (2021). The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology*, *30*(1), 37–47. https://doi.org/10.1111/mec.15716

Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, *213*(1), 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3-4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893. https://doi.org/10.1111/ecog.03049

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2021). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-152. https://CRAN.R-project.org/package=nlme

Pope, N. S., & Peterman, B. (2020). *radish: Fast gradient-based optimization of resistance surfaces*. R package version 0.0.2. https://github.com/nspope/radish

Pope, N. (2021). *corMLPE: A correlation structure for symmetric relational data*. R package version 0.0.3. https://github.com/nspope/corMLPE

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., Richardson, K., Rosing, M. T., Whittaker, R. J., & Fjeldså,

J. (2019). Building mountain biodiversity: geological and evolutionary processes. *Science*, *365*(6458), 1114–1119. https://doi.org/10.1126/science.aax0151

Ramírez-Barahona, S., & Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. *Ecology and Evolution*, *3*(3), 725–738. https://doi.org/10.1002/ece3.483

R Development Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/ (accessed May 23, 2021).

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, *69*(1), 274–281. https://doi.org/10.1111/j.1541-0420.2012.01824.x

Rice, A. E., Vinciguerra, N. T., McCullough, J. M. (2020). Early nest record and additional notes on the breeding biology of the Chestnut-capped Brushfinch (*Arremon brunneinucha suttoni*) in southern Mexico. *Ornitología Neotropical*, *31*(1), 76–78. https://doi.org/10.58843/ornneo.v31i1.557

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(13), 1–21. https://doi.org/10.18637/jss.v076.i13

Wang, I. (2019). *topoDistance: Calculating topographic paths and distances*. R package version 1.0.1. https://CRAN.R-project.org/package=topoDistance

Wang, J. (2017). The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources*, *17*(5), 981–990. https://doi.org/10.1111/1755-0998.12650

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5), 744–751. https://doi.org/10.1111/2041-210X.13152

CAPÍTULO III. Relaciones evolutivas dentro del complejo *Arremon brunneinuchalvirenticeps* (Aves: Passerellidae) con base en datos de secuenciación de nueva generación

Capítulo III. Relaciones evolutivas dentro del complejo Arremon brunneinucha/virenticeps (Aves: Passerellidae) con base en datos de secuenciación de nueva generación

Título: "**Evolutionary History and Species Delimitation Within Mesoamerican *Arremon* Brushfinches (Aves: Passerellidae)**".

Arremon virenticeps en Ciudad de México, México. Fotografía por: Leopoldo Vázquez-Reyes.

Evolutionary History and Species Delimitation Within Mesoamerican *Arremon* Brushfinches (Aves: Passerellidae)

Authors and Affiliations

Israel Moreno-Contreras[1,2], Scott V. Edwards[3], Flavia Termignoni-García[3], Michael G. Harvey[4], Oscar Johnson[4], Alexander Llanes-Quevedo [1,2], Luis A. Sánchez-González[2], María del Coro Arizmendi[5], and Adolfo G. Navarro-Sigüenza[2,*]

[1]*Posgrado en Ciencias Biológicas, Unidad de Posgrado, Universidad Nacional Autónoma de México, Coyoacán, Mexico City 04510, Mexico.*

[2]*Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, A.P. 70–399, Mexico City 04510, Mexico.*

[3]*Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA.*

[4]*Department of Biological Sciences, University of Texas at El Paso, El Paso, Texas, USA.*

[5]*UBIPRO, FES Iztacala Universidad Nacional Autónoma de México, Av. de los Barrios, Los Reyes Iztacala, Tlalnepantla 54090, Estado de México, Mexico.*

*Corresponding author: adolfon@ciencias.unam.mx

Abstract

A species complex is an assemblage of closely related species in which there are blurry boundaries, and from which species could arise from different speciation processes and/or a speciation continuum. Species complexes provide an opportunity to investigate the evolutionary mechanisms that act on speciation. However, the presence of horizontal gene transfer, incomplete lineage sorting, isolation by distance, and hierarchical population genetic structure can yield conflicting phylogenetic inferences. The need for an integrative approach to species delimitation that simultaneously evaluates molecular, ecological, and geographic lines of evidence is therefore becoming increasingly apparent. The *Arremon brunneinucha*/*virenticeps* complex (Aves: Passerellidae) is a monophyletic group of small highland birds in northern Mesoamerica. It consists of two low-vagility species that are distinguishable by subtle morphological variation, have relatively segregated Grinnellian niches, and have allopatric distribution patterns. In this study, we used NextRAD data to construct a well-supported phylogeny of this species complex, which we then supplemented with population- and landscape genomics approaches to elucidate the evolutionary relationships and delimit species within this complex in Mesoamerican montane forests. We concluded that the Mesoamerican Brushfinches consists of six or seven species. Geographical and ecological factors each contributed to the current diversity and distribution patterns of this species complex, which is composed of both well-diverged species and diverging lineages that are on the path toward speciation.

Keywords allopatric speciation; genomics; incomplete lineage sorting; isolation by distance; phylogenomics; systematics.

Introduction

Systematics mainly seeks to elucidate the evolutionary relationships among species. One of the challenges of phylogenetic reconstruction is the expected heterogeneity of topological evolutionary histories displayed by gene trees and species trees (Maddison 1997; Degnan and Rosenberg 2009; Edwards 2009a; Edwards 2009b). This variation among the different evolutionary histories of gene trees ("gene discordance") can be illustrated as a cloudogram (Maddison 1997). In broad terms, gene discordance (*sensu* Maddison 1997) may arise due to multiple processes, including gene duplication and extinction ("hidden paralogy"), horizontal gene transfer, and incomplete lineage sorting (ILS). At the same time, a handful of continuous spatial processes such as isolation by distance (IBD), gene flow, and hierarchical population structure can together contribute to spurious inferences in the quantitative delimitation of biodiversity (Leaché et al. 2014a; Chambers and Hillis 2020). Furthermore, ILS can cause short phylogenetic branches due to geographic closeness between allopatric populations (Tomasello et al. 2020); it is therefore urgent that these factors be accounted for in phylogenomic studies.

Advances in next-generation sequencing (NGS) technologies and access to supercomputing clusters have greatly facilitated the construction of phylogenomic data matrices across different evolutionary timescales in diverse systematic groups (Jarvis et al. 2014; Manthey et al. 2016; Eaton et al. 2017; Scornavacca and Galtier 2017). Systematic biologists have incorporated, in parallel, the advances for the delimitation of species under an integrative approach, with a better representation of biological realism (Smith and Carstens 2020; Douglas and Bouckaert 2022). This emerging synthesis proposes that validating species in a fully integrative taxonomy should focus on testing delimitation hypotheses under novel explicit evolutionary models (Leaché et al. 2014b; Carstens et al. 2022). However, the estimation of species limits under the multispecies coalescent model (MSC) poses significant theoretical and practical challenges, with several simplifying assumptions (Carstens et al. 2013).

Integrative species delimitation studies are of particular importance for understanding patterns of biodiversity (Pante et al. 2015). Particularly, recognition of the incredible species diversity that resides in the Neotropics has accrued over the last decade (Moura

and Jetz 2021), and continued exploration suggests that Neotropical species richness has been dramatically underestimated (Moura and Jetz 2021). Specifically, Neotropical montane regions host many species that are threatened or at risk of extinction due to historical and contemporary modification of land use and ongoing global climate change (Bellard et al. 2014). At the same time, these natural systems present interesting evolutionary divergence mechanisms (e.g., IBD) caused by the combination of landscape attributes and glacial-interglacial dynamics (Mastretta-Yanes et al. 2015; Mastretta-Yanes et al. 2018; Moreno-Contreras et al. 2023). This has been reflected in high indices of genomic structure and differentiation for montane taxa over time, offering an excellent opportunity to assess species limit hypotheses and their causal mechanisms.

The Mesoamerican Chestnut-capped/Green-striped Brushfinches of the genus *Arremon* present an excellent case in which to apply a comprehensive framework for studying complex speciation using genomic data. The systematics of *Arremon* Brushfinches was initially thought to be well-characterized (Hellmayr 1938; Parkes 1954; Parkes 1957). This led to a long history of taxonomic revision, from early lumping of many allopatric *brunneinucha* populations, to considering them phylogenetically distant from the Green-striped group (Paynter 1978, see Figure 1a). Previous (allozyme-based) systematics suggested significant genetic differentiation within Chestnut-capped populations (Peterson et al. 1992). Subsequently, one of the most striking findings in avian systematics was that the Green-striped group is embedded within a large group of Chestnut-capped populations. First, *virenticeps* was treated as sister species to *brunneinucha* from the Sierra Madre Oriental (Cadena et al. 2007, Figure 1b). Shortly after, it was also aligned as closest relative to a Chestnut-capped population in the Sierra Madre del Sur (Navarro-Sigüenza et al. 2008; Navarro-Sigüenza et al. 2013; Figure 1c). Ultimately, phylogenetic analysis using Ultra Conserved Elements (UCEs) obtained a relatively contrasting topology, but that similarly placed *virenticeps* within *brunneinucha* (Buainain et al. 2022; Figure 1d).

The rich body of prior research in this complex has shown that the evolutionary relationships within this interesting group have yet to be unraveled. For instance, the conflicting topologies are likely the result of stochastic sorting of ancestral polymorphisms

following rapid divergence (Flores-Rodríguez et al. 2011), incomplete reproductive isolation due to niche conservatism (Moreno-Conteras et al. 2020) and/or low resistance to landscape barriers (Moreno-Contreras et al. 2023) in geographically proximate lineages. Additionally, despite extensive geographic sampling throughout the range of the complex, previous phylogeographic studies of the group have been limited to a handful of loci (Cadena et al. 2007; Navarro-Sigüenza et al. 2008; Flores-Rodríguez et al. 2011; Navarro-Sigüenza et al. 2013), or lineage-based systematics (Buainain et al. 2022). A new and promising genotyping technology called NextRAD may be able to overcome these constraints by fragmenting and ligating adaptor sequences to genomic DNA through engineered transposons. It requires very small amounts of DNA of non-model organisms, so it is particularly useful in scenarios where no reference genome is available (Russello et al. 2015). This approach has been successfully utilized to study fine-scale population genomics and diversity in a whitefly species complex (Wosula et al. 2017) and phylogeographic patterns in birds (Conklin et al. 2022).

Here, we combined multiple statistical approaches for inferring species discovery, tests for hierarchical population structure, and species tree reconstruction (Figure 2), to elucidate the complex evolutionary history of the entire *Arremon brunneinucha*/*virenticeps* group in Mesoamerica (Figure 3), while also identifying sources of statistical uncertainty. As input, we sequenced thousands of genome-wide NextRAD loci from the entire geographic distribution of *Arremon* Brushfinches from Mesoamerica. We implemented a clustering-based species discovery process (Derkarabetian et al. 2019), which compares species delimitation schemes across a handful of analytically unique dimensionality reduction techniques (Figure 2). We then tested for a hierarchical population structure using non-spatial and spatial analysis of Mesoamerican Brushfinches to resolve areas of persistent uncertainty (Figure 2). We also employed multiple phylogenomic methods to account for heterogeneity of evolutionary histories among gene trees generated by both ILS and gene flow during the speciation process (Maddison 1997; Edwards 2009a), a problem exacerbated by recent and rapid speciation (Kubatko and Degnan 2007; Gatesy and Springer 2014; Giarla and Esselstyn 2015). In addition, we evaluated the usefulness of the NextRAD loci in supporting alternative phylogenetic topologies (based on published literature and the backbones obtained in this study) using two sets of estimated gene

trees. Since mtDNA sequenced with Sanger methods usually provides discordant topologies with respect to NGS-based topologies for species complexes of the Mesoamerican highlands (Zarza et al. 2018; Tsai et al. 2019), we hypothesize that species trees estimated with NextRAD loci will have higher quartet support values than the topologies based on mtDNA or morphology. Despite ongoing inter-species gene flow and rapid diversification, we demonstrate the potential of a comprehensive genomic approach to reveal a new and nuanced understanding of evolutionary history, even in non-model groups without reference genome.

Materials and Methods

Taxon Sampling and NextRAD Sequencing

We extracted genomic DNA from 36 tissue samples collected throughout the Mesoamerican distribution range (Moreno-Contreras et al. 2023) of the Chestnut-capped/Green-striped Brushfinch species complex (see Figure 3), as well as two samples of *Arremon aurantiirostris* (Orange-billed Sparrow) as outgroups. The latter was chosen as outgroup due to closely relationship with the Mesoamerican Brushfinches (Buainain et al. 2022). Whole genomic DNA was extracted from tissues using the Qiagen DNeasy Kit (Qiagen Inc., Valencia, CA, USA) following the manufacturer's protocol. Prior to library preparation, DNA concentration was quantified using a Qubit fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA), as well as by visual examination using a 1% agarose gel.

NextRAD genotyping-by-sequencing libraries were generated and sequenced for 38 samples by SNPsaurus, LLC following the protocol of Russello et al. (2015). These libraries yielded short sequences (~150 base pairs) flanked by engineered transposon cut sites, which results in thousands of loci distributed throughout the nuclear genome (Russello et al. 2015). This method requires only small amounts of DNA (<50 ng; Russello et al. 2015), allowing the recovery of sequence data even from samples with low DNA concentration. Briefly, genomic DNA was fragmented using a Nextera reagent (Illumina Inc., San Diego, CA, USA), which also attaches short adapter sequences to the ends of

the fragments. The Nextera reaction was scaled to fragments of ~10 ng of genomic DNA. Fragmented DNA was then amplified with one of the primers matching the adapter, and then extending 10 nucleotides into the genomic DNA with the selective sequence GTGTAGAGCC. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer were efficiently amplified by PCR. Purified libraries were size-selected to 350–800 bp and sequenced on a HiSeq 4000 with one lane of 150 bp reads (Genomics Core Facility, University of Oregon, Eugene, OR, USA).

Raw genomic data were trimmed for adapters and quality filtered before calling SNPs. The quality of sequencing data was checked with FastQC v.0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The cleaning process for Illumina reads was performed using Trimmomatic v.0.39 (Bolger et al. 2014). We used the *SLIDINGWINDOW* function (5:33), which scans with a 5-base wide sliding window, cutting when the average quality per base drops below a threshold of 33. We then used the *ILLUMINACLIP* function (1:30:10) to prune the NextRAD sequencing adapters. Specifically, we allowed a maximum of 2 mismatches, and seeds were extended and clipped when a score of 30 was reached for paired-end reads (about 50 bases) or a score of 10 was reached for single-ended reads (about 17 bases).

We used the ipyrad v.0.9.26 pipeline (Eaton and Overcast 2020) to identify homology among clean reads, make SNP calls, and format output files in a *de novo* framework. We followed a *de novo* approach, as few reference genomes exist for sister species within the family Passerellidae. The closest relative species was the Chipping Sparrow (*Spizella passerina*), which diverged around 9 Mya (Barker et al. 2015). Since the Mesoamerican *Arremon* bird complex diverged 3.36 Mya (Buainain et al. 2022), we would have recovered very few loci during SNP calling by using reference genomes from more distant taxa due to deep divergences (Harvey et al. 2016). Preliminary results employing available genome references supported this conclusion, as some assembly matrices had low numbers of loci after SNP calls and post-processing filters. In order, the following ipyrad parameter settings were explored to obtain *de novo* assembly matrices for downstream analyses: max_low_qual_bases (maximum low quality base calls) = 0, 5; phred_Qscore_offset (phred Q score offset) = 20, 33; mindepth_statistical (minimum depth for statistical base

calling) = 5, 6, 7, 8, 9, 10; mindepth_majrule (minimum depth for majority-rule base calling) = 3, 6, 9; maxdepth (maximum cluster depth within samples) = 1000, 5000, 10000; clust_threshold (clustering threshold for de novo assembly) = 0.7, 0.75, 0.8, 0.85, 0.9, 0.95; min_samples_locus (minimum number of samples per locus for output): 2, 4, 8, 9, 19, 29; max_SNPs_locus (maximum number of SNPs per locus) = 0.05, 0.15, 0.2, 0.25, 0.3, 0.4; max_Indels_locus (maximum number of indels per locus) = 5, 6, 8; and max_shared_Hs_locus (maximum number of heterozygous sites per locus) = 0.05, 0.15, 0.25, 0.3, 0.4, 0.5. The exploration of the parameter space tallied 72 assembly matrices (49 matrices using a *de novo* approach). We chose the matrix based on the recovery of low amount of missing data, high number of SNPs, and phylogenetic resolution.

Quality Filtering

We used VCFTools v.0.1.16 (Danecek et al. 2011) as part of an interactive protocol of good practices for SNP quality filtering (O'Leary et al. 2018; Linck and Battey 2019). Briefly, we filtered for maximum depth, keeping only sites less than or equal to the mean depth values with a threshold of 50 (--max-meanDP 50). This threshold was a good indicator to eliminate possible paralogous loci resulting from sequencing errors (O'Leary et al. 2018), which could obscure phylogenetically informative loci. To avoid violation of the assumptions of maximum likelihood (CA-ML) and coalescent theory-based phylogenetic methods, only biallelic SNPs (--min-alleles 2 --max-alleles 2) were retained.

RAD-seq and its variants generate genomic data sets with large numbers of loci and high amount of missing data. Previously, it has been demonstrated in quantitative studies that setting the thresholds of the amount of missing data too strict may produce data sets that exclude loci with the highest mutation rates. As a consequence, data sets with lower proportions of missing data may also include fewer variable sites, which could have an impact on the patterns of genome-wide variation inferred from phylogenetic, genotypic clustering, and species delimitation analyses (Leaché et al. 2015; Huang and Knowles 2016; Eaton et al. 2017; Molloy and Warnow 2018). In order to test the effect of missing data on downstream analyses, we constructed three distinct scenarios that respectively reflect permissiveness (--max-missing 0.65, 0.70, 0.75) thresholds of 35%, 30%, and

25%, of missing data. Exploratory analysis suggested that a permissiveness threshold of 35% missing data recovered well-delineated clades and genetic clusters.

Subsequently, SNPs with a minor allele count (MAC) of at least 2 were kept (--mac 2) to eliminate singletons, whose inclusion may affect population genomic inferences (Linck and Battey 2019). In order to minimize the effects of linkage disequilibrium, downstream analyses were performed retaining only one SNP per RAD locus, discarding all but the first SNP at each locus. The first SNP is likely to have lower sequencing error, since sequencing errors rates tend to increase towards the end of a read. This "unlinked SNP dataset" was obtained using the Python script *vcf_parser.py* (https://github.com/CoBiG2/RAD_Tools/blob/master/vcf_parser.py).

Species Discovery

We followed the clustering-based species discovery protocol outlined by Derkarabetian et al. (2019), which involves dimensionality reduction methods. It is well-documented that a high prevalence of missing data in genome-wide sets may lead to artifactual clustering results. So, we further filtered the "unlinked SNP dataset" using a 25% missing data threshold with VCFTools v.0.1.16 (Danecek et al. 2011). We also excluded monomorphic loci in BCFtools v.1.10.2 (Danecek et al. 2021). That process resulted in a genomic set containing 1,733 SNPs, which was subsequently used for the clustering-based species discovery protocol. In many of our exploratory analyses, we found that these methods were still yielding nonsensical clustering results, so we will focus only on discriminant analysis of principal components (DAPC) and principal component analysis (PCA) using a hierarchical agglomerative clustering (Scrucca et al. 2016). This is expected as multivariate methods tend to have poor resolution in differentiating groups when there are few SNPs or a large amount of missing data (Yi and Latch 2022). Both methods were implemented using the R package adegenet v.2.1.10 (Jombart and Ahmed 2011), and for the hierarchical agglomerative clustering we used the R-library mclust v.6.0.0 (Scrucca et al. 2016). This entire workflow is depicted graphically in Figure 2.

Sample Clustering

We then sought to reconstruct and visualize the relationships among individual samples using clustering based-model approaches. Specifically, we used STRUCTURE v.2.3.4 (Pritchard et al. 2000) to assign samples to population clusters, which can aid in identifying both population structure and admixed individuals. We used our filtered, unlinked set of 4,490 SNPs as input, and performed 10 independent runs, testing $K \in \{2, 3, 4, 5, 6, 7, 8\}$. Each independent run was performed over 300,000 MCMC iterations with a burn-in of 200,000 iterations, excluding the LOCPRIOR option. We set the admixture model to the correlated allele frequencies (with all remaining options kept at their default settings). Optimal *K* values were determined using three indices: ln Pr(*X*|*K*) values ($\widehat{K}_{\text{Pritchard}}$; Pritchard et al. 2000) and Δ*K* scores ($\widehat{K}_{\text{Evanno}}$; Evanno et al. 2005) in the 'POPHELPER' v.2.3.1 package for R (Francis 2017); while Puechmaille metrics (e.g., MedMeaK, MaxMeaK; Puechmaille 2016) with different threshold values (0.5, 0.6, 0.7, and 0.8) were computed in StructureSelector (Li and Liu 2018).

In our analysis that included all samples (*n* = 36 samples, excluding outgroups), a *K* = 6 was suggested. However, the Puechmaille method indicated a *K* = 5. This makes sense, as Δ*K* identifies the uppermost hierarchical level of population structure (Evanno et al. 2005) and Puechmaille metrics does not recognize clusters with membership < 0.5 ("ghost clusters," according to Puechmaille 2016). This outcome may indicate the presence of a hierarchical structure phenomenon in Mesoamerican brushfinches. To address this issue, we followed a "hierarchical STRUCTURE analysis" (Vähä et al. 2007), which included only the *kuehnerii* and *suttoni* samples (both SM and SMS). We further filtered the genomic data by excluding monomorphic loci using BCFtools v.1.10.2 (Danecek et al. 2021), resulting in a data set of 2,136 SNPs. In this reduced analysis (*n* = 15 samples), we explored $K \in \{1, 2, 3, 4\}$, keeping the same default parameters as mentioned above. Finally, the results from 10 replicates of the selected *K* values for both analyses were summarized into a single result, which were then aligned and analyzed in CLUMPP v.1.1.2 (Jakobsson and Rosenberg 2007) using the greedy algorithm with 10,000 repetitions.

We then sought to reconstruct and visualize the spatial arrangement among individual samples onto a geo-genetic map using SpaceMix (Bradburd et al. 2016). This landscape

genetic method enhances the species delimitation scheme by controlling for the effect of IBD. Our working hypothesis was that if the ellipses of a given (problematic) sample overlapped with the samples of another cluster, then the two should be merged into a single genetic cluster. This technique is particularly useful in situations where STRUCTURE results are not robust enough to assign samples to specific genetic clusters at fine scales (e.g., samples from the Guerrero and Oaxaca montane ranges). In this way, we used SpaceMix v.0.13 (Bradburd 2023) with an allele frequency covariance data (generated from the filtered NextRAD genomic dataset) and setting real geographic coordinates as location priors. The implemented model was “estimating geo-genetic locations and admixture source locations” (Bradburd et al. 2016), which was evaluated using the Bayes Factor (Kass and Raftery 1995) through custom R-scripts (Márquez et al. 2020). We executed the SpaceMix analysis with 10 initial fast runs, each with 1,000,000 iterations followed by a long MCMC run of 100,000,000 iterations sampled every 10,000 iterations. The first 2,500 samples were discarded as burn-in, leaving 7,500 samples per run. We plotted geo-genetic maps showing 95% credible ellipses.

Concatenation-based Methods

We inferred concatenation-based phylogenies (“supergene trees”) using maximum likelihood (CA-ML) in two ML programs. The first was IQ-TREE v.1.6.12 (Nguyen et al. 2015), whose heuristic search algorithm uses nearest-neighbor-interchange (NNI). The second was RAxML v.8.4.2 (Stakamatis 2014), whose algorithm uses subtree pruning and regrafting (SPR). IQ-TREE analyses automatically select (only if is requested) the best-fitting model. The node support of each internal branch was evaluated using 1,000 ultrafast bootstraps (-bb, Minh et al. 2013). On the other hand, we used the JC69 model of nucleotide substitution with rate heterogeneity for the filtered data set (-m GTRGAMMA --JC69) in RAxML. This model of sequence evolution was the best fit for the concatenated NextRAD data (using matrix miss35), which was determined using jModelTest v.2.1.10 (Darriba et al. 2012). For each RAxML analysis, branch support was estimated using the rapid bootstrap function (-f a) with 1,000 bootstraps. CA-ML trees were rooted with *Arremon aurantiirostris* (Orange-billed Sparrow).

We also concatenated the SNP-based matrix as input for SplitsTree4 v.4.17.1 (Huson and Bryant 2006) to generate a phylogenetic network implementing a Neighbor-Net, which can effectively present conflicting underlying data patterns like those that can occur with next-generation sequencing (Bryant and Moulton 2004).

Coalescent-based Methods

To account for gene-tree discordance among NextRAD loci, we analyzed the genomic data set using several coalescent-based species tree approaches. In all analyses we designated all identified lineages (including the outgroup) as individual tips (K = 8), except for SNAPP, in which we did not include the outgroup samples. The first analysis was SVDquartets v.1.0 (Chifman and Kubatko 2014) in PAUP* v.4.0a169 (Swofford 2002), which infers the relationships among four species at a time (= quartets) under a coalescent model directly from SNP sites. We evaluated all possible quartets with 1,000 bootstrap replicates, and then the quartet assembly method Quartet Fiduccia-Mattheyses (QFM) was used to produce a summary tree.

As a complement to the site-based coalescent approach, we constructed species trees based on two summary coalescent methods: MP-EST v.1.5 (Liu et al. 2010) and ASTRAL-III v.5.7.8 (hereafter ASTRAL, Zhang et al. 2018). As input for both summary methods, we used the LOCI file output (containing 18,946 locus) generated by the ipyrad pipeline. Later, we created a separate FASTA alignment for each RAD locus using the Python script *loci2fasta.py* (https://github.com/btmartin721/file_converters/blob/master/loci2fasta.py). Then, we built custom R scripts to elaborate an "allowlist" containing the NextRAD locus names from the "unlinked SNP dataset" VCF file (inferred from the 65% occupancy matrix). This list allowed us to select only the FASTA files containing the unlinked 4,490 loci. We concatenated these selected FASTA files with the Perl script *FASconCAT-G_v1.05.1.pl* (https://github.com/PatrickKueck/FASconCAT-G/blob/master/FASconCAT-G_v1.05.1.pl). Subsequently, we built a partition of the concatenated FASTA file to obtain separate FASTA files for each locus in the R package ape v.5.7-1 (Paradis and Schliep 2019). Next, we converted these FASTA locus alignments into either NEXUS or PHYLIP

formats using the publicly available Python script *convert.py* (https://github.com/mmatschiner/tutorials/blob/master/ml_species_tree_inference/src/convert.py). Finally, we generated a separate gene tree for each NEXUS file using IQ-TREE v.1.6.12 (Nguyen et al. 2015) and performed 1,000 bootstrap replicates for each estimated gene tree (hereafter $GT_{IQ\text{-}TREE}$). Likewise, gene trees were estimated (hereafter GT_{RAxML}) in RAxML v.8.4.2 (Stakamatis 2014) with the PHYLIP files using a search for the best-scoring tree in the same run (-f a option), in conjunction with 1,000 bootstrap replicates, and the GTRGAMMA approximation of rate heterogeneity. We estimate two sets of gene trees to compare node support using different previously published topologies (see below).

Although our NextRAD loci contain few base pairs (~150 bp) and could poorly estimate gene trees, MP-EST and ASTRAL are capable of handling polytomies and missing data in input gene trees. Poorly or incorrectly resolved gene trees can mislead species trees (Gatesy and Springer 2014). We therefore analyzed the data after collapsing branches in estimated gene trees with less than 10%, 20%, and 30% bootstrap support using Newick utilities v.1.6 (Junier and Zdobnov 2010). This can increase the normalized quartet score of our overall ASTRAL species tree reconstruction. However, exploratory analyses suggested that collapsing the branches did not improve the performance of phylogenetic reconstructions relative to the raw estimated gene trees.

We first used MP-EST, whose algorithm uses the frequencies of rooted gene trees of triplets of taxa to estimate the topology and branch lengths (in coalescent units) of the overall species tree (Liu et al. 2010). This algorithm has good accuracy and is fast enough to run on hundreds to thousands of estimated gene trees, as is necessary in phylogenomic studies (Jarvis et al. 2014; Mirarab et al. 2016). We applied MP-EST to the $GT_{IQ\text{-}TREE}$ set computed on bootstrap replicates of each gene sequence alignment. With bootstrapped ML trees for each gene, MP-EST was performed with the site-only multilocus bootstrapping (MLBS) procedure (Seo 2008). In order, bootstrapped gene trees estimated from entire locus DNA sequence data were rooted with the Orange-billed Sparrow samples. In the case of bootstrapped gene trees where samples of this outgroup were not available, we used the samples that had the lowest amount of missing data with the aid

of LINUX custom scripts and the *root.multiPhylo* function in ape v.5.7-1 (Paradis and Schliep 2019). Next, 200 replicates ($GT_{IQ\text{-}TREE}1$, $GT_{IQ\text{-}TREE}2$, …, $GT_{IQ\text{-}TREE}200$) of the input datasets for the MP-EST were created such that $GT_{IQ\text{-}TREE}i$ contained the *ith* bootstrap tree across all genes. MP-EST was run 10 times, selecting the species tree with the best pseudo-likelihood score for each one of these 200 MLBS input replicates. Finally, the best scoring MP-EST species trees inferred from the bootstrap replicates were summarized with a 50% majority-rule consensus tree using SumTrees v.4.6.0, part of the DendroPy v.4.6.0 package (Sukumaran and Holder 2010).

We used the unrooted $GT_{IQ\text{-}TREE}$ set as the input for ASTRAL and generated a full annotated species tree along with quartet frequencies (for the main topology, the first alternative, and the second alternative), and local posterior probabilities for each of the internal branches by specifying the "-t 2" option. The two node supports were visualized using the R-scripts of Crameri et al. (2022). Finally, we also used the "-t 10" command in ASTRAL to perform polytomy tests (Sayyari and Mirarab 2018). For a given internal branch in a species tree, a polytomy test evaluate the statistical power to reject the null hypothesis that the given internal branch has a true length of zero (i.e., a hard polytomy; see Sayyari and Mirarab 2018). We annotated each internal branch in our species tree reconstruction with the posterior probability calculated by ASTRAL and the *P*-value for the polytomy test for the given branch.

We used SNAPP v.1.6.1 (Bryant et al. 2012) implemented in BEAST v.2.7.4 (Bouckaert et al. 2019) to infer a species tree directly and visualize potential uncertainty in species tree reconstruction from SNP data in a full-coalescent analysis without an outgroup. This Bayesian approach has the advantage of bypassing gene tree estimation and minimizing error due to the low information content of individual NextRAD loci. Because of computational constraints, we subsampled from each clade the three individuals (where available) that had the greatest number of NextRAD loci and the lowest number of missing data recovered in order to maximize the number of loci for the SNAPP analyses. In this analysis we treated as tips each of the clades identified in the IQ-TREE 30% phylogeny, STRUCTURE, and SpaceMix analyses. To alleviate potential stochasticity associated with sample inclusion, we repeated this procedure twice. Based on the tree height of the

IQ-TREE 30% phylogeny of 0.1755 substitutions/site, we applied a gamma distributed prior on the speciation rate (λ = 10, α = 11.75, β = 109.73), and a gamma-distributed prior on the expected divergence (θ: α = 2, β = 23). The tree height was measured using the *cophenetic.phylo* function of ape v.5.7-1 (Paradis and Schliep 2019), previously excluding the outgroup samples and diagonal cells containing ceros. We ran all analyses for 20 million generations, storing every 10,000 generations, a burn-in of 10%, and checking that runs converged in Tracer v.1.7.2 (Rambaut et al. 2018) based on ESS values over 200. Two independent runs were combined in LogCombiner implemented in BEAST v.2.7.4 (Bouckaert et al. 2019), and from the posterior of the species trees we calculated an MCC tree in TreeAnnotator v.2.7.4 implemented in BEAST v.2.7.4 (Bouckaert et al. 2019).

Quartet Support for Alternative Topologies

We are interested in how closely NextRAD loci support each of the other topologies, which are relatively contrasting to each other. To address this, we additionally evaluated the quartet support and local probabilities for six different species tree topologies (i.e., all published phylogenetic trees with *b. macrourus*, *b. brunneinucha*, *b. apertus*, *virenticeps*, *b. suttoni* [SM], *b. suttoni* [SMS], and *kuehnerii* as ingroups). The six topologies were gathered according through a literature review (Paynter 1978; Cadena et al. 2007; Navarro-Sigüenza et al. 2008) and those obtained in this study (SVDquartets, SNAPP). Some of these topologies (e.g., Paynter 1978) lacked specific details about the phylogenetic relationships within this species complex (see Fig. 1). Accordingly, we readjusted these topologies based on the geographic proximity for each mountain containing a specific allopatric species (Howell and Webb 1995; Navarro-Sigüenza et al. 2008). This is supported by the hypothesis that *Arremon* Brushfinches in Mesoamerica present a significant IBD pattern (Moreno-Contreras et al. 2023). Therefore, we used the tree scoring option (-t 2) in ASTRAL under two different schemes of estimated gene tree sets ($GT_{IQ\text{-}TREE}$ and GT_{RAxML}), and a file containing the assignment of taxa to species or the outgroup.

Node support was considered statistically significant under the following criteria: BS ≥ 95% for CA-ML analyses; MLBS or BS ≥ 95% for coalescent-based methods (MP-EST

and SVDquartets, respectively); and PP ≥ 95% for ASTRAL and SNAPP. All phylogenetic outcomes (trees and networks) were edited and displayed using R-packages such as ape v.5.7-1 (Paradis and Schliep 2019), panghorn v.2.11.1 (Schliep et al. 2017), cowplot v.1.1.1 (Wilke 2020), ggplot2 v.3.4.2 (Wickham 2016), ggimage v.0.3.3 (Yu 2023), ggpubr v.0.6.0 (Kassambara 2023), ggtree v.3.6.2 (Yu et al. 2017), tanggle v.1.4.0 (Schliep et al. 2022), treeio v.1.22.0 (Wang et al. 2020).

Genomic Differentiation

We quantified the basic population summary statistics, such as genetic differentiation (F_{ST}; Weir and Cockerham 1984) between identified genetic clusters and clades (K = 6–7) using the *genet.dist* function available in hierfstat v.0.5-11 (Goudet and Jombart 2022).

Results

The fine-tuned ipyrad pipeline contained 225,607 SNPs for 38 samples (including outgroups). During bioinformatic post-processing filtering, we retained 205,886 SNPs after applying the mean depth filter with a threshold of 50, which was reduced to 189,289 SNPs after excluding non-biallelic genomic data. This was further reduced to 28,259 SNPs after filtering for a stringent missing data threshold of 35%, then to 16,643 SNP after excluding singletons (MAC < 2 filter). Finally, filtering the genomic data by retaining only one SNP per RAD locus left a total of 4,490 biallelic and unlinked SNPs for downstream analyses. The mean amount of missing data for the filtered NextRAD set was 30% with a SD = 25.44% considering 38 samples, and 30% (mean) with 4.45% (SD) for 4490 SNPs.

Our species discovery pipeline resulted in two species delimitation iterations (DAPC and PCA), which can be categorized into one distinct scheme differentiated by the number of species supported (K = 6, Figure 4). Within each of these schemes, individual assignment to clusters was non-concordant between methodological approaches (Figure 4).

Sample assignments from STRUCTURE at K = 5 (Puechmaille metrics) and K = 6 ($\widehat{K}_{\text{Pritchard}}$ and $\widehat{K}_{\text{Evanno}}$) were nearly identical to sample assignments at K = 6 for all implemented species delimitation schemes (Figure 5a–b). At K = 6, *kuehnerii* and *suttoni* from Oaxaca displayed signs of nested structure when analyzed separately from the rest of the distribution (Figure 5c–d). According to hierarchical STRUCTURE analysis, ln Pr(X|K) values, ΔK scores, and Puechmaille methods favored K = 4, K = 2, and K = 3, respectively. SpaceMix showed evidence that *suttoni* individuals are close to each other on the geo-genetic map, and none of the ellipses from any group overlapped with individuals from another group (Figure 6). The estimated long-distance admixture proportions were minimal for all samples. The model with these proportions fixed at 0 was not supported over one where they were allowed to vary (Bayes Factor = 0.307), which suggests that patterns of relatedness in these genomic data were well-described by an IBD model.

Phylogenetic Reconstruction

A maximum-likelihood tree (obtained in IQ-TREE) from the analysis of 4,490 unlinked, concatenated loci was well supported (through ultrafast bootstrap values), with seven reciprocally monophyletic groups under the species delimitation schemes (Figure 7a). The best supported clade (BS > 100%) contained samples of the *apertus* group and the green-striped *virenticeps* (Figure 7a). The least supported clade (BS = 77%) was *suttoni* from the Sierra Mazateca. The clade containing *suttoni* from Sierra Madre del Sur was moderately well supported (BS = 83%), and it included a *suttoni* sample (MT_362) from northern Oaxaca. The RAxML topology recovered the same clades as those obtained in IQ-TREE, but with slightly lower bootstrap support (Figure S1).

Coalescent-based species tree reconstruction used gene trees estimated in IQ-TREE as input for summary-methods (MP-EST and ASTRAL-III), with each of the seven putatively identified species designated as separate tips (Figure 7b–c). In the case of MP-EST, all branches had excellent support (MLBS > 95%). ASTRAL-III also recovered the K = 6 species delimitation scheme with low posterior probability for the internal branch determining the branching order among *suttoni* from the Sierra Mazateca, *suttoni* (SMS),

and *kuehnerii* (Figure 7c). Additionally, the internal branch connecting *virenticeps* with the clade containing *kuehnerii* and *suttoni* received low support (Figure 7c). Visualizations of quartet frequencies revealed that alternate gene tree topologies (i.e., topologies that disagreed with the recovered species tree) were common at almost all of these internal branches, even in well-supported branches, whereas the most frequently recovered gene trees represented only a narrow plurality in both cases (Figure 7c). Polytomy tests confirmed that hard polytomies cannot be statistically rejected at three of these nodes, among which the *kuhnerii* and *suttoni* (SMS) relationship had the highest *P*-values (Figure 7c).

SVDquartets obtained a topology similar to that of the two-step (summary) methods, but placing *suttoni* (SM) with high support (BS = 97.2%) as sister group to the clade containing *virenticeps*, *suttoni* (SMS), and *kuehnerii* (Figure S2). However, the node indicating *virenticeps* as sister group to a clade of Sierra Madre del Sur populations (both Guerrero and Oaxaca) received low support (BS = 69.8%).

Coalescent-based species tree reconstruction under a full Bayesian framework, using unlinked biallelic SNPs as input for SNAPP (Figure 7d), recovered a topology similar to the summary (MP-EST and ASTRAL-III) and site-based (SVDquartets) reconstructions, except that the *kuehnerii* population from the Guerrero highlands was sister to the clade containing both populations from the Oaxacan *suttoni*, with high support (PP = 1). In addition, the internal branch of *brunneinucha* from the Sierra Madre Oriental had low support (PP < 0.5; Figure 7d).

The unrooted phylogenetic network constructed using the distance-based Neighbor-Net approach resulted in seven visually apparent clusters that correspond perfectly to sample assignment under a *K* = 7 species delimitation scheme (Figure 7e). Despite discrete clustering of *suttoni* individuals from the Sierra Madre del Sur and Sierra Mazateca in the phylogenetic network, the internal branches separating these groups are extremely short (Figure 7e). Clusters containing individuals from the Sierra de Los Tuxtlas (*apertus*) and *virenticeps* from the Transmexican Volcanic Belt had long branches (Figure 7e),

We evaluated quartet scores (i.e., the fraction of induced quartet trees) of six topologies for relationships among sampled species (Figure 8) using the tree scoring option in

ASTRAL-III in combination with a file that mapped taxa to species or to the outgroup. The ASTRAL topology (which is similar to MP-EST), presented in Figure 7c, showed the highest normalized quartet score (40.57%) using IQ-TREE gene trees. Alternative topologies (i.e., SVDquartets or SNAPP) had similar normalized quartet scores (40.56% and 40.51%, respectively; Figure 8). We mostly obtained higher quartet support values when using estimated gene trees of RAxML (Figure 8).

We assessed relative differentiation between all seven lineages identified via species discovery by calculating pairwise F_{ST}. The level of relative genetic differentiation and the number of fixed SNPs between lineages are both highly variable (F_{ST} range: 0.047–0.476) but with moderate average genomic differentiation (F_{ST} mean: 0.321). However, nearly all pairwise F_{ST} comparisons are above than 0.1, reflecting substantial differentiation between most identified lineages (Figure 9). The same pattern was found with $K = 6$ (Figure S3).

Discussion

Our results demonstrate that sequencing thousands of genomic loci and integrating multiple unique analytical approaches for species delimitation, genomic clustering methods, and species tree reconstruction, can reveal complex and reticulate evolutionary histories while helping to clarify species limits and the speciation process. A main takeaway from this comparative approach is that current taxonomy (two recognized species: *A. brunneinucha* and *A. virenticeps*) significantly underestimates the species-level diversity of the Brushfinches.

In this study, we inferred high statistical support for concatenated and coalescent-based trees of the *Arremon* Brushfiches from the Mesoamerican clade that combined thousands of NextRAD loci (Figure 1). Our concatenated empirical trees were well supported at most internal nodes, regardless of the maximum likelihood software used. Nonetheless, the bootstrap scores are likely biased, and the topology may be somewhat unreliable due to well-documented artifactual issues associated with concatenation approaches (Kubatko and Degnan 2007; Cranston et al. 2009; Gatesy and Springer 2014; Leaché et al. 2015).

At first glance, short branches relative to effective population sizes are expected in species complexes, probably reflecting recent divergences and contributing to biases (low) bootstrap values using ML concatenation approaches (Kubatko and Degnan 2007). Furthermore, none of our coalescent-based analyses produced a phylogeny that is well supported at all nodes, though some interspecific relationships do emerge fairly consistently and with moderate to strong support under different schemes of phylogenetic uncertainty.

In short, we confirmed previously reported species-level phylogenetic relationships of the *Arremon brunneinucha*/*virenticeps* complex (Cadena et al. 2007; Navarro-Sigüenza et al. 2008; Flores-Rodríguez et al. 2011; Navarro-Sigüenza et al. 2013), but the obtained topologies differ slightly from a species tree that has been used previously for comparative studies within the group (Buainain et al. 2022). The main pattern of discrepancy involves four lineages: *virenticeps* from the Transmexican Volcanic Belt, *kuehnerii* associated with the highlands of Guerrero, *suttoni* from northern Oaxaca (Sierra Mazateca), and *suttoni* from the Oaxacan portion of the Sierra Madre del Sur. In the first place, a monophyletic *virenticeps* was recovered with high nodal support, and it was placed as sister group to a clade composed of *kuehnerii* (Sierra Madre del Sur of Guerrero) and *suttoni* from Oaxaca (both Sierra Mazateca and Sierra Madre del Sur). This insight contrasts with earlier thoughts where *virenticeps* is evolutionarily distant from all Chestnut-capped populations across the Chestnut-capped's distributional range (Hellmayr 1938; Paynter 1978). At the same time, it is in good agreement with Sanger-based systematics using mitochondrial and nuclear data (Cadena et al. 2007; Navarro-Sigüenza et al. 2008; Navarro-Sigüenza et al. 2013), and recent (UCEs-based) species tree reconstructions (Buainain et al. 2022). The only previous study indicating a non-monophyletic *virenticeps* was based on mitochondrial DNA sequences (Navarro-Sigüenza et al. 2008). There, it was evidenced that a single sample (DEUT47; La Verdura, Michoacán) remains aligned as a sister lineage to all remaining Green-striped Brushfinches from the Transmexican Volcanic Belt and Chestnut-caped Brushfinches from Guerrero. One potential explanation of that previous paraphyletic arrangement is ILS in early-diverged groups. It is expected that lineages that consist of a single sample in mtDNA-based trees (using concatenated frameworks) will generate non-monophyletic topologies (McKay and Zink 2010), due to

the retention of ancestral polymorphisms and the low amount of phylogenetic signal of the mtDNA sequences prohibiting reciprocal monophyly (Flores-Rodríguez et al. 2011). Biogeographic discordance, which explicitly assumes that populations have remained isolated or are in secondary contact during glaciation cycles (Toews and Brelsford 2012), would be a plausible alternative hypothesis. Under this premise, it is expected that during geographic isolation in climate refugia, there has been a fixation of alleles that have provided identity in the structure of each isolated population (Ramírez-Barahona and Eguiarte 2013). The sister relationship of *virenticeps* with Chestnut-capped Brushfinches of the highlands of southern Mexico highlights the biogeographic coherence between the Transmexican Volcanic Belt and Sierra Madre del Sur. Other biological groups such as *Isthmura* salamanders (Bryson et al. 2018) and *Sarcohyla* frogs (Zarza et al. 2018) support this close biogeographical affinity. These two biogeographic provinces are known to be separated by the Balsas depression, an important geographic barrier for montane taxa (Marshall and Liebherr 2000; Sánchez-González et al. 2008).

Regarding the Chestnut-capped Brushfinches from the southern cloud-forests (Guerrero and Oaxaca), when we used a hierarchical genetic structure analysis focusing only on these three groups, we found that the significant structure is best explained by two lineages: *kuehnerii* and all members of *suttoni*. This is supported by a genetic structure analysis (SpaceMix) while controlling for the effect of IBD involving the 36 ingroup samples. Here, it is clearly observed that the ellipses corresponding to the *kuehnerii* samples did not overlap with any *suttoni* samples. By contrast, *suttoni* samples from the SMS overlap with *suttoni* samples from the SM. While the topology proposed by SNAPP is robust and is corroborated by the results of the hierarchical STRUCTURE analysis and the SpaceMix outcomes (i.e., no overlap between *kuehnerii* and *suttoni* ellipses), it was the third best topology in terms of the quartet support for ASTRAL. In fact, very few (ML) gene trees corroborated the sister relationship between the two *suttoni* populations (SM and SMS), which was supported with low posterior probabilities, as can be seen in Figure 6. Notably, we only found high posterior probabilities between the two *suttoni* lineages when they form a sister clade relative to the other Mesoamerican Brushfinches (as obtained in Navarro-Sigüenza et al. 2008). This finding suggests that throughout the Mesoamerican distribution of the *Arremon brunneinucha*/*virenticeps* complex, genetic

population-related factors act jointly (e.g., nested genetic structure and isolation by distance), obscuring phylogenetic resolution at fine scales. Another possible explanation could be that the NextRAD loci used in the summary-methods (MP-EST and ASTRAL) were too short (i.e., had either too few parsimony informative sites or too few total base pairs; see Harvey et al. 2016) to detect the close relationship of *kuehnerii* as a sister lineage of the clade containing all of the *suttoni* samples. However, the topologies obtained in both approaches (concatenation and coalescent) are largely similar to recent topologies (e.g., Buainain et al. 2022), so our study coincides with early ideas that genomic data (e.g. RAD loci) can yield concordant results regardless of the phylogenetic method used (Manthey et al. 2016). Furthermore, there is a biogeographic correspondence that reconciles the results obtained in SNAPP and the genetic structure analyses. First, the uplift of the Sierra Madre del Sur (Guerrero) and highlands of Oaxaca occurred in a west-to-east gradient (Nieto-Samaniego et al. 2006). This tectonic movement likely generated a phylogeographic pattern where there is genetic differentiation between allopatric (or nearly parapatric) sister populations. Second, contemporary landscape resistance (and quite possibly glaciation-interglaciation dynamics) has limited gene flow within and between populations of Mesoamerican Brushfinches (Moreno-Contreras et al. 2023).

The striking fact that both portions (east and west) of the Sierra Madre del Sur present different species that are evolutionarily proximate but with a high degree of genomic divergence is not new. This same phylogeographic pattern of long or mid-term period of significant isolation has been reported for *Eupherusa* Hummingbirds (Rocha-Méndez et al. 2019), *Dendrortyx* Wood-Partridges (Tsai et al. 2019), and *Aphelocoma* Jays (Venkatraman et al. 2019). The co-divergence in those taxa might relate to their suspected low dispersal behavior and their close association with high-elevation temperate forests (Paynter 1978; Howell and Webb 1995; Watson 2003). Our study adds information to already accumulating evidence that mtDNA topologies with short internodes (e.g., Navarro-Sigüenza et al. 2008; see also Barrera-Guzmán et al. 2012 for a similar case) show distant separation of lineages from geographically close areas (for example, between and/or within Guerrero and Oaxaca), while NGS-based topologies usually recovered them as closely sister lineages (see also Zarza et al. 2018; Tsai et al. 2019).

We demonstrate that this can be evidenced only with the implementation of novel species tree methods (e.g., by detecting hard polytomies and low posterior probabilities), and not with concatenated mtDNA datasets alone, as previously done. Furthermore, this apparently general finding is independent of the type of NGS marker (RADseq or UCE-based), and may be expected in Central or South American taxa that diverged just over two million years ago, with (allopatric) year-round distributions throughout the TVB, SMS (Guerrero and Oaxaca), and patchy and discontinuous minor mountain ranges in Oaxaca (e.g., SM). All of these taxa show contrasting life-history traits, however, which warrants further investigation as to whether this phylogeographic pattern is also repeated in other biological groups with similar criteria (as above explained) or is unique to avian species complexes. This premise contrasts with the hypothesis that the Oaxacan Highlands do not represent a natural biogeographic unit (León-Paniagua and Morrone 2009), or with what has been found with some highland taxa such as mammals (León-Paniagua et al. 2007) and birds without distributional evidence within the TVB (Maldonado-Sánchez et al. 2016). Therefore, we propose that this hypothesis be tested in subsequent phylogeographic studies. Notably, the TVB has served as an east-to-west colonization path (Mastretta-Yanes et al. 2015) for Neotropical taxa from the Sierra Madre Oriental (SMO) or regions further east (e.g., Sierra de Los Tuxtlas).

Regarding the eastern and southern lineages in this complex, the remaining Chestnut-capped Brushfinches have not been problematic. Particularly, the morphologically distinct Chestnut-capped Brusfinch, which lacks the black chest band (*apertus* from Sierra de Los Tuxtlas), has been well described in the literature as a diagnosable entity (Hellmayr 1938; Parkes 1957; Parkes 1957; Paynter 1978). In addition, the relatively long-term isolation of the Sierra de Los Tuxtlas uplift (Ornelas et al. 2013; Rodríguez-Elizarrarás et al. 2023) and narrow distribution of this group (Howell and Webb 1995) have contributed to high levels of genomic differentiation and long branches for *apertus* Brushfinches, as depicted in our network analysis. In any case, this contrasts with a previous finding where *apertus* was placed as a more basal lineage within this complex (Cadena et al. 2007). On the other hand, two morphologically similar Chestnut-capped groups remain as two well-delineated and basal lineages—*brunneinucha* from the SMO and *macrourus* from CHH—

with high node support. This outcome supports the hypothesis that this species complex had a South American origin (Buainain et al. 2022).

Despite the subtle morphological differences among Brushfinch lineages (Hellmayr 1938; Parkes 1954; Parkes 1957; Paynter 1978; Howell and Webb 1995; Navarro-Sigüenza et al. 2013), significant phylogenetic, genetic, and ecological evidence (Peterson et al. 1992; Cadena et al. 2007; Navarro-Sigüenza et al. 2008; Navarro-Sigüenza et al. 2013; Moreno-Contreras et al. 2023) support most of them as good species. Our phylogenomic approaches are also in line with previous findings. Specifically, regarding the *kuehnerii* lineage that is geographically isolated in the highlands of the westernmost portion of Guerrero (Howell and Webb 1995; Navarro-Sigüenza et al. 2008; Navarro-Sigüenza et al. 2013), we found repeated evidence from species delimitation, hierarchical structure analyses, tests of IBD, and phylogenomic approaches that supports the species status of this group. The overall phylogenetic uniqueness of this Guerrero endemic population is further supported by previous work indicating that *kuehnerii* is distinct in an mtDNA concatenated tree (Navarro-Sigüenza et al. 2008), nuDNA haplotype (Navarro-Sigüenza et al. 2013), ecology (at least in terms of segregation of Grinnellian niches; Moreno-Contreras et al. 2020), allopatric distribution (Howell and Webb 1995; Navarro-Sigüenza et al. 2013), and low-vagility habits (Paynter 1978; Moreno-Contreras et al. 2023). Although no individual criterion can universally delimit species (Pigliucci 2003), published lines of evidence now support the species status of *kuehnerii* under the phylogenetic (Donoghue 1985), genotypic cluster (Mallet 1995), and general lineage (De Queiroz 2007) species concepts. Field studies investigating the functional connectivity between individuals of *suttoni* inhabiting the Oaxacan Highlands, where putative hybrid individuals could be reported given the lowest landscape resistance to migration in that distributional area (Moreno-Contreras et al. 2023), will be essential for confirming the existence of reproductive barriers and the species status of both *suttoni* lineages. Notably, their narrowly restricted distributions throughout the Mesoamerican highlands, in conjunction with accelerated land use change (Watson 2003) and climate change (Sierra-Morales et al. 2021) in many of their distributional areas only exacerbate the problem as an urgent conservation issue.

Concluding remarks

Our study highlights the value of integrating multiple analytical methods for genomic analysis, including species tree methods, to reconcile discordant gene histories into a single bifurcating species history, as well as landscape genomic analyses, which can incorporate variation in both branching order and gene flow events (Bradburd et al. 2016). Ultimately, we have greatly clarified our understanding of the systematics of the *Arremon* Brushfinches, a group for which the complex glacial-interglacial dynamics of Mesoamerica (Mastretta-Yanes et al. 2015; Mastretta-Yanes et al. 2018) has reinforced a complex speciation history (Buainain et al. 2022; Moreno-Contreras et al. 2023). Phylogenomics is not only about evaluating evolutionary relationships using lots of genome-wide loci, but also about assessing the support that this genomic data gives to previous (alternative) phylogenetic hypotheses. Future studies formally testing the role of different evolutionary processes (e.g., ILS, hybridization) may elucidate the origin of conflicting topologies. In this fashion, we can better understand the nature of conflicting topologies that have remained as areas of phylogenetic uncertainty.

Acknowledgments

This paper partially fulfills the requirements for obtaining a Doctoral degree in the Posgrado en Ciencias Biológicas, UNAM of IM-C, who received a Doctoral scholarship provided by CONACyT (number 749969). PAEP-CGEP-UNAM provided financing for IM-C's research visit to Harvard University and University of Texas at El Paso. M. E. Muñiz (Facultad de Ciencias, UNAM) and L. Márquez-Valdelamar (Instituto de Biología, UNAM) provided access and logistical support for the laboratory work. We acknowledge the helpful comments from L. Kiere, C. Granados, H. Rodríguez-Correa, J. Velasco, E. Vázquez, J. P. Jaramillo, L. Eguiarte, F. Rodríguez, T. Garrido, and R. Munguía-Steyer. A. Gordillo-Martínez provided useful information on *Arremon* specimens hosted at Museo de Zoología, UNAM. Financial support was obtained from PAPIIT (IN 214621 to AGN-S). Bioinformatic analyses were performed using the supercomputer 'Miztli' through DGTIC-

UNAM (LANCAD-UNAM-DGTIC-402 to AGN-S and AL-Q). Finally, IM-C appreciates the support provided by the Escalante-Termignoni and Moreno-Contreras families during the elaboration of this work.

References

Barker F.K., Burns K.J., Klicka J., Lanyon S.M., Lovette I.J. 2015. New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. Auk. 132: 333–348.

Barrera-Guzmán A.O., Milá B., Sánchez-González L.A., Navarro-Sigüenza A.G. 2012. Speciation in an avian complex endemic to the mountains of Middle America (*Ergaticus*, Aves: Parulidae). Mol. Phylogenet. Evol. 62: 907–920.

Bellard C., Leclerc C., Leroy B., Bakkenes M., Veloz S., Thuiller W., Courchamp F. 2014. Vulnerability of biodiversity hotspots to global change. Glob. Ecol. Biogeogr. 23: 1376–1386.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30: 2114–2120.

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15: e1006650.

Bradburd G.S., Ralph P.L., Coop G.M. 2016. A spatial framework for understanding population structure and admixture. PLoS Genet. 12: e1005703.

Bradburd G. 2023. SpaceMix: Runs SpaceMix: a method for inferring and visualizing patterns of population genetic structure. R package version 0.13. https://github.com/gbradburd/SpaceMix

Bryant D., Moulton V. 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21: 255–265.

Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29: 1917–1932.

Bryson R.W. Jr., Zarza E., Grummer J.A., Parra-Olea G., Flores-Villela O., Klicka J., McCormack J.E. 2018. Phylogenomic insights into the diversification of salamanders in the *Isthmura bellii* group across the Mexican highlands. Mol. Phylogenet. Evol. 125: 78–84.

Buainain N., Ferreira M., Avendaño J.E., Cadena C.D., Faircloth B.C., Brumfield R.T., Cracraft J., Ribas C.C. 2022. Biogeography of a neotropical songbird radiation reveals similar diversification dynamics between montane and lowland clades. J. Biogeogr. 49: 1260–1273.

Cadena C.D., Klicka J., Ricklefs R.E. 2007. Evolutionary differentiation in the Neotropical montane region: Molecular phylogenetics and phylogeography of *Buarremon* brush-finches (Aves, Emberizidae). Mol. Phylogenet. Evol. 44: 993–1016.

Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. Mol. Ecol. 22: 4369–4383.

Carstens B.C., Smith M.L., Duckett D.J., Fonseca E.M., Thomé M.T.C. 2022. Assessing model adequacy leads to more robust phylogeographic inference. Trends Ecol. Evol. 37:402–410.

Chambers E.A., Hillis D.M. 2020. The multispecies coalescent over-splits species in the case of geographically widespread taxa. Syst. Biol. 69: 184–193.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30: 3317–3324.

Conklin J.R., Verkuil Y.I., Battley P.F., Hassell C.J., ten Horn J., Johnson J.A., Tomkovich P.S., Baker A.J., Piersma T., Fontaine M.C. 2022. Global flyway evolution in red

knots *Calidris canutus* and genetic evidence for a Nearctic refugium. Mol. Ecol. 31: 2124–2139.

Crameri S., Fior S., Zoller S., Widmer A. 2022. A target capture approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae). Mol. Ecol. Resour. 22: 3087–3105.

Cranston K.A., Hurwitz B., Ware D., Stein L., Wing R.A. 2009. Species trees from highly incongruent gene trees in rice. Syst. Biol. 58: 489–500.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R., 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27: 2156–2158.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10: giab008.

Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9: 772.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24: 332–340.

De Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56: 879–886.

Derkarabetian S., Castillo S., Koo P.K., Ovchinnikov S., Hedin M. 2019. A demonstration of unsupervised machine learning in species delimitation. Mol. Phylogenet. Evol. 139: 106562.

Donoghue M.J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist. 88: 172–181.

Douglas J., Bouckaert R. 2022. Quantitatively defining species boundaries with more efficiency and more biological realism. Commun. Biol. 5: 755.

Eaton D.A.R., Spriggs E.L., Park B., Donoghue M.J. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. 66: 399–412.

Eaton D.A.R., Overcast I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics. 36: 2592–2594.

Edwards S.V. 2009a. Is a new and general theory of molecular systematics emerging? Evolution. 63: 1–19.

Edwards S.V. 2009b. Natural selection and phylogenetic analysis. Proc. Natl. Acad. Sci. U.S.A. 106: 8799–8800.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611–2620.

Flórez-Rodríguez A., Carling M.D., Cadena C.D. 2011. Reconstructing the phylogeny of "*Buarremon*" brush-finches and near relatives (Aves, Emberizidae) from individual gene trees. Mol. Phylogenet. Evol. 58: 297–303.

Francis R.M. 2017. pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. 17: 27–32.

Gatesy J., Springer M.S. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80: 231–266.

Giarla T.C., Esselstyn J.A. 2015. The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. Syst. Biol. 64: 727–740.

Goudet J., Jombart T. 2020. hierfstat: Estimation and tests of hierarchical *F*-statistics. R package version 0.5-11. https://CRAN.R-project.org/package=hierfstat

Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst. Biol. 65: 910–924.

Hellmayr C.E. 1938. *Catalogue of birds of the Americas. Part XI: Ploceidae - Catamblyrhynchidae - Fringillidae*. Field Museum of Natural History Zoological Series, Volume XIII, Part XI. Chicago: Field Museum of Natural History.

Howell S.N.G., Webb S. 1995. *A guide to the birds of Mexico and northern Central America*. Oxford: Oxford University Press.

Huang H., Knowles L.L. 2016. Unforeseen consequences of excluding missing data from next generation sequences: simulation study of RAD sequences. Syst. Biol. 65: 357–365.

Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23: 254–267.

Jakobsson M., Rosenberg N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23: 1801–1806.

Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Núñez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Xiong Y., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldså J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G. 2014.Whole-genome

analyses resolve early branches in the tree of life of modern birds. Science. 346: 1320–1331.

Jombart T., Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 27: 3070–3071.

Junier T., Zdobnov E.M. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. Bioinformatics. 26: 1669–1670.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90: 773–795.

Kassambara A. 2023. *ggpubr: 'ggplot2' based publication ready plots*. R package version 0.6.0. https://CRAN.R-project.org/package=ggpubr

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56: 17–24.

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014a. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63: 17–30.

Leaché A.D., Fujita M.K., Minin V.N., Bouckaert R.R. 2014b. Species delimitation using genome-wide SNP data. Syst. Biol. 63: 534–542.

Leaché A.D., Banbury B.L., Felsenstein J., Nieto-Montes de Oca A., Stamatakis A. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64: 1032–1047.

León-Paniagua L., Navarro-Sigüenza A.G., Hernández-Baños B.E., Morales J.C. 2007. Diversification of the arboreal mice of the genus *Habromys* (Rodentia: Cricetidae: Neotominae) in the Mesoamerican highlands. Mol. Phylogenet. Evol. 42: 653–664.

León-Paniagua L., Morrone J.J. 2009. Do the Oaxacan Highlands represent a natural biotic unit? A cladistic biogeographical test based on vertebrate taxa. J. Biogeogr. 36: 1939–1944.

Li Y.-L., Liu J.-X. 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol. Ecol. Resour. 18: 176–177.

Linck E., Battey C.J. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol. Ecol. Resour. 19: 639–647.

Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10: 302.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46: 523–536.

Maldonado-Sánchez D., Gutiérrez-Rodríguez C., Ornelas J.F. 2016. Genetic divergence in the common bush-tanager *Chlorospingus ophthalmicus* (Aves: Emberizidae) throughout Mexican cloud forests: the role of geography, ecology and Pleistocene climatic fluctuations. Mol. Phylogenet. Evol. 99: 76–88.

Mallet J. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10: 294–299.

Manthey J.D., Campillo L.C., Burns K.J., Moyle R.G. 2016. Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). Syst. Biol. 65: 640–650.

Márquez R., Linderoth T.P., Mejía-Vargas D., Nielsen R., Amézquita A., Kronforst M.R. 2020. Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of colour pattern variation in *Phyllobates* poison-dart frogs. Mol. Ecol. 29: 3702–3719.

Marshall C.J., Liebherr J.K. 2000. Cladistic biogeography of the Mexican transition zone. J. Biogeogr. 27: 203–216.

Mastretta-Yanes A., Moreno-Letelier A., Piñero D., Jorgensen T.H., Emerson B.C. 2015. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J. Biogeogr. 42: 1586–1600.

Mastretta-Yanes A., Xue A.T., Moreno-Letelier A., Jorgensen T.H., Alvarez N., Pinero D., Emerson B.C. 2018. Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. Mol. Ecol. 27: 432–448.

McKay B.D., Zink R.M. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. Mol. Phylogenet. Evol. 54: 647–650.

Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30: 1188–1195.

Mirarab S., Bayzid M.S., Warnow T. 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65: 366–380.

Molloy E.K., Warnow T. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst. Biol. 67: 285–303.

Moreno-Contreras I., Sánchez-González L.A., Arizmendi M.C., Prieto-Torres D.A., Navarro-Sigüenza A.G. 2020. Climatic niche evolution in the *Arremon brunneinucha* complex (Aves: Passerellidae) in a Mesoamerican landscape. Evol. Biol. 47: 123–132.

Moreno-Contreras I., Llanes-Quevedo A., Sánchez-González L.A., Arizmendi M.C., Navarro-Sigüenza A.G. 2023. Isolation by resistance explains genetic diversity in the *Arremon brushfinches* of northern Mesoamerica. Mol. Ecol. 32: 3450–3470.

Moura M.R., Jetz W. 2021. Shortfalls and opportunities in terrestrial vertebrate species discovery. Nat. Ecol. Evol. 5: 631–639.

Navarro-Sigüenza A.G., Peterson A.T., Nyari A., García-Deras G.M., García-Moreno J. 2008. Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. Mol. Phylogenet. Evol. 47: 21–35.

Navarro-Sigüenza A.G., García-Hernández M.A., Peterson A.T. 2013. A new species of brush-finch (*Arremon*; Emberizidae) from western Mexico. Wilson J. Ornithol. 125: 443–453.

Nieto-Samaniego A.F., Alaniz-Álvarez S.A., Silva-Romo G., Eguiza-Castro M.H., Mendoza-Rosales C.C. 2006. Latest Cretaceous to Miocene deformation events in the eastern Sierra Madre del Sur, Mexico, inferred from the geometry and age of major structures. Geol. Soc. Am. Bull. 118: 238–252.

Nguyen L.T., Schmidt H.A., von Haeseler A., Bui Q.M. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32: 268–274.

O'Leary S.J., Puritz J.B., Willis S.C., Hollenbeck C.M., Portnoy D.S. 2018. These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. Mol. Ecol. 27: 3193–3206.

Ornelas J.F., Sosa V., Soltis D.E., Daza J.M., González C., Soltis P.S., Gutiérrez-Rodríguez C., Espinosa de los Monteros A. Castoe T.A., Bell C., Ruiz-Sanchez E. 2013. Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. PLoS ONE. 8: e56283.

Pante E., Schoelinck C., Puillandre N. 2015. From integrative taxonomy to species description: one step beyond. Syst. Biol. 64: 152–160.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35: 526–528.

Parkes K.C. 1954. A revision of the Neotropical finch *Atlapetes brunneinucha*. Condor. 56: 129–138.

Parkes K.C. 1957. The juvenal plumages of the finch genera *Atlapetes* and *Pipilo*. Auk. 74: 499–502.

Paynter R.A. Jr. 1978. Biology and evolution of the avian genus *Atlapetes* (Emberizinae). Bull. Mus. Comp. Zool. Harv. Coll. 148: 323–369.

Peterson A.T., Escalante P., Navarro, A. 1992. Genetic variation and differentiation in Mexican populations of Common Bush-Tanagers and Chestnut-capped Brush-Finches. Condor. 94: 244–253.

Pigliucci M. 2003. Species as family resemblance concepts: the (dis-) solution of the species problem? BioEssays 25: 596–602.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155: 945–959.

Puechmaille S.J. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol. Ecol. Resour. 16: 608–627.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard, M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67: 901–904.

Ramírez-Barahona S., Eguiarte L.E. 2013. The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. Ecol. Evol. 3: 725-738.

Rocha-Méndez A., Sánchez-González L.A., González C., Navarro-Sigüenza A.G. 2019. The geography of evolutionary divergence in the highly endemic avifauna from the Sierra Madre del Sur, Mexico. BMC Evol. Biol. 19: 237.

Rodríguez-Elizarrarás S R., Morales-Barrera W.V., Pompa-Mera V., Siebe C., Benowitz J., Layer, P.W., Lozano-Santacruz R., Girón P. 2023. Geochemistry and 40Ar/39Ar dating of the Sierra de Santa Marta in the tectonically controversial Los Tuxtlas Volcanic Complex (Veracruz, Mexico). J. South Am. Earth Sci. 124: 104250.

Russello M.A., Waterhouse M.D., Etter P.D., Johnson E.A. 2015. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ. 3: e1106.

Sánchez-González L.A., Morrone J.J., Navarro-Sigüenza A.G. 2008. Distributional patterns of the Neotropical humid montane forest avifaunas. Biol. J. Linn. Soc. 94: 175–194.

Sayyari E., Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes. 9: 132.

Schliep K., Potts A.J., Morrison D.A., Grimm G.W. 2017. Intertwining phylogenetic trees and networks. Methods Ecol. Evol. 8: 1212–1220.

Schliep K, Vidal-García M, Solís-Lemus C, Biancani L, Ada E, Henao Diaz L (2022). *tanggle: visualization of phylogenetic networks*. https://klausvigo.github.io/tanggle

Scornavacca C., Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. Syst. Biol. 66: 112–120.

Scrucca L., Fop M., Murphy T.B., Raftery A.E. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8: 289–317

Seo T.K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25: 960–971.

Sierra-Morales P., Rojas-Soto O., Ríos-Muñoz C.A., Ochoa-Ochoa L.M., Flores-Rodríguez P., Almazán-Núñez R.C. 2021. Climate change projections suggest severe decreases in the geographic ranges of bird species restricted to Mexican humid mountain forests. Glob. Ecol. Conserv. 30: e01794.

Smith M.L., Carstens B.C. 2020. Process-based species delimitation leads to identification of more biologically relevant species. Evolution. 74: 216–229.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30: 1312–1313.

Sukumaran J., Holder M.T. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics. 26: 1569–1571.

Swofford D.L. 2019. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*. Version 4. Available from: https://paup.phylosolutions.com.

Toews D.P.L., Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21: 3907–3930.

Tomasello S., Karbstein K., Hodač L., Paetzold C., Hörandl E. 2020. Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane plant species: a case study on the *Ranunculus auricomus* species complex. Mol Ecol 29: 2031–2049.

Tsai W.L.E., Mota-Vargas C., Rojas-Soto O., Bhowmik R., Liang E.Y., Maley J.M., Zarza E., McCormack J.E. 2019. Museum genomics reveals the speciation history of *Dendrortyx* wood-partridges in the Mesoamerican highlands. Mol. Phylogenet. Evol. 136: 29–34.

Vähä J.P., Erkinaro J., Niemelä E., Primmer C.R. 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Mol. Ecol. 16: 2638–2654.

Venkatraman M.X., DeRaad D.A., Tsai W.L.E., Zarza E., Zellmer A.J., Maley J.M., McCormack J.E. 2019. Cloudy with a chance of speciation: integrative taxonomy reveals extraordinary divergence within a Mesoamerican cloud forest bird. Biol. J. Linn. Soc. 126: 1–15.

Wang L.G., Lam T.T.-Y., Xu S., Dai Z., Zhou L., Feng T., Guo P., Dunn C.W., Jones B.R., Bradley T., Zhu H., Guan Y., Jiang Y., Yu G. 2020. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Mol. Biol. Evol. 37: 599–603.

Watson D.M. 2003. Long-term consequences of habitat fragmentation—highland birds in Oaxaca, Mexico. Biol. Conserv. 111: 283–303.

Weir B.S., Cockerham C.C. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution. 38: 1358–1370.

Wickham H. 2016. *ggplot2: Elegant graphics for data analysis*. 2nd edn. Switzerland: Springer.

Wilke C. 2020. *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. R package version 1.1.1. https://CRAN.R-project.org/package=cowplot

Wosula E.N., Chen W., Fei Z., Legg J.P. 2017. Unravelling the genetic diversity among cassava *Bemisia tabaci* whiteflies using NextRAD sequencing. Genome Biol. Evol. 9: 2958–2973.

Yi X., Latch, E.K. 2022. Nonrandom missing data can bias principal component analysis inference of population genetic structure. Mol. Ecol. Resour. 22: 602–611. https://doi.org/10.1111/1755-0998.13498

Yu G., Smith D.K., Zhu H., Guan Y., Lam T.T.-Y. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8: 28–36.

Yu G. 2023. *ggimage: use image in 'ggplot2'*. R package version 0.3.3. https://CRAN.R-project.org/package=ggimage

Zarza E., Connors E.M., Maley J.M., Tsai W.L.E., Heimes P., Kaplan M., McCormack J.E. 2018. Combining ultraconserved elements and mtDNA data to uncover lineage diversity in a Mexican highland frog (*Sarcohyla*; Hylidae). PeerJ. 6: e6045.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 19: 153.

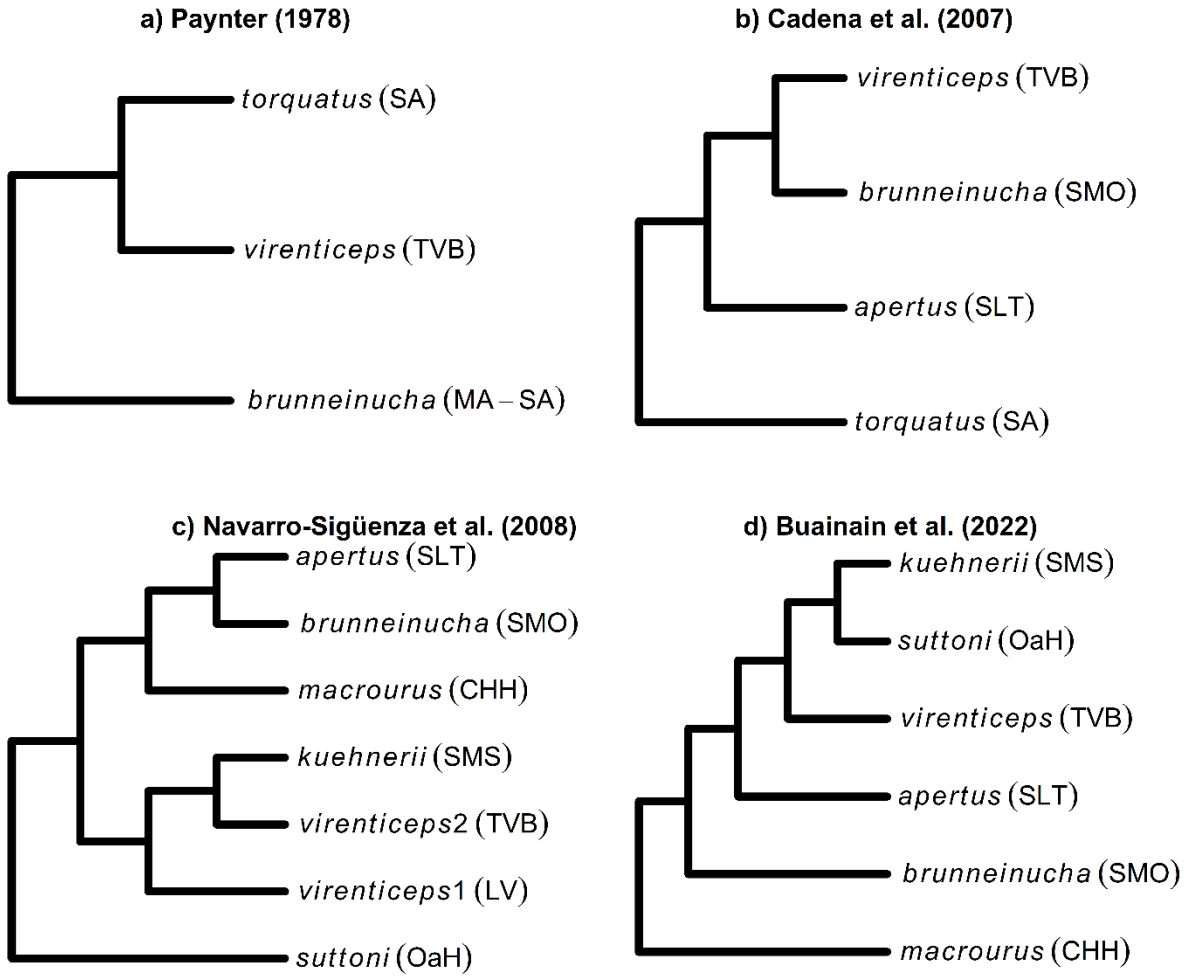


FIGURE 1. Example species trees based on the prior literature drawn according to the lineages analyzed in this study. a) Morphology-inferred species tree of Paynter (1978); b) Concatenated mitochondrial (four genes) and nuclear (two genes) species tree (4,208 bp) of Cadena et al. (2007); c) Mitochondrial (four genes) species tree (801 bp) of Navarro-Sigüenza et al. (2008); d) Ultra Conserved Elements species tree (2,201 loci) of Buainain et al. (2022).

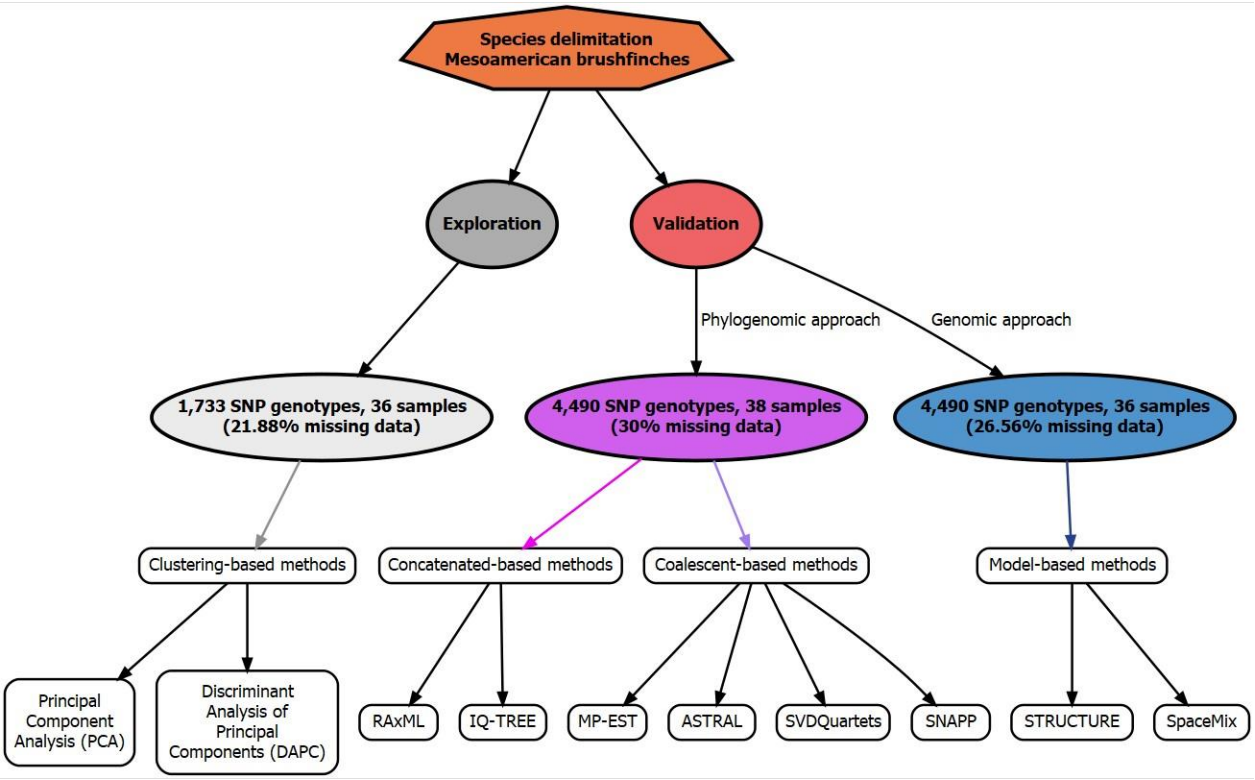


FIGURE 2. Diagram detailing the species discovery approaches (exploration), and the two quantitative species delimitation schemes (validation). Circles depicting the dimensionality-reduction process are shaded in gray. Circles depicting phylogenomic and genomic approaches are highlighted in purple and blue, respectively.

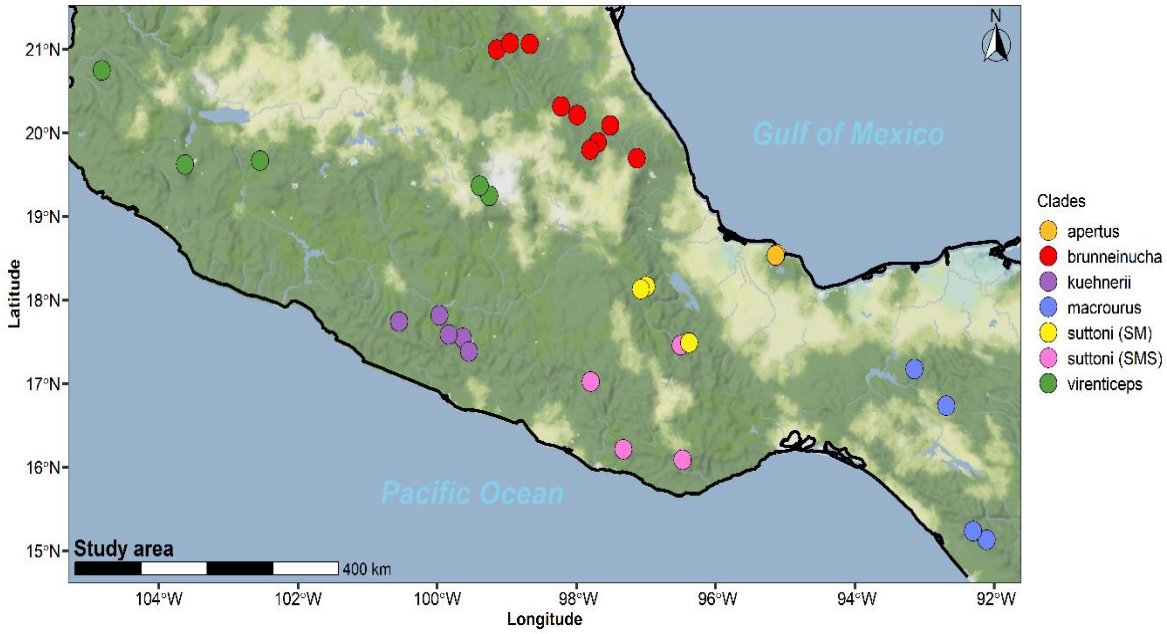


FIGURE 3. Sampling localities, with color corresponding to the *K* = 7 quantitative species delimitation scheme for the Mesoamerican *Arremon* Brushfinches.

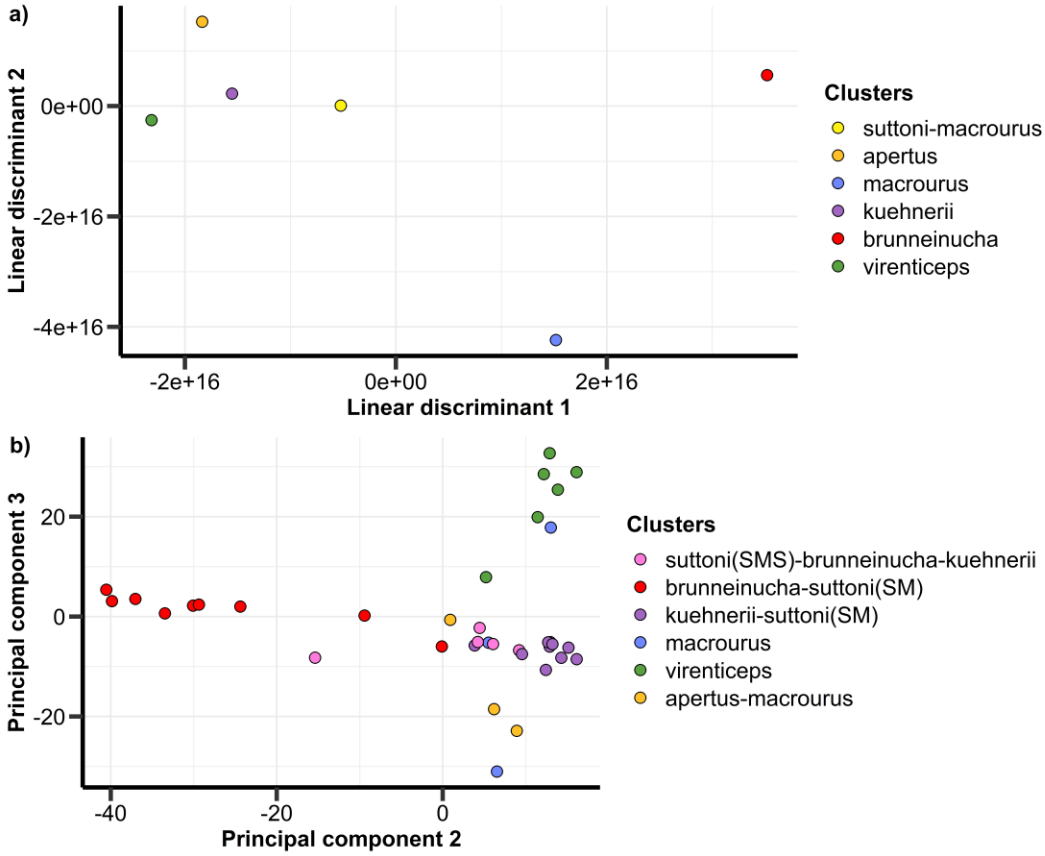


FIGURE 4. Species discovery and species clustering using range-wide genomic sampling of the Chestnut-capped/Green-striped Brushfinch species complex. Two-dimensional plots visualizing sample relatedness using the complete SNP data set (1,733 SNPs, 21.88% missing data) for a) DAPC and b) PCA. Each scheme shows the two dimensions where population clusters are most visually differentiated. If a species delimitation scheme identified a single cluster containing more than one of the color-coded localities on the map, the color of the cluster containing the most individual samples is used for depiction in the given plot

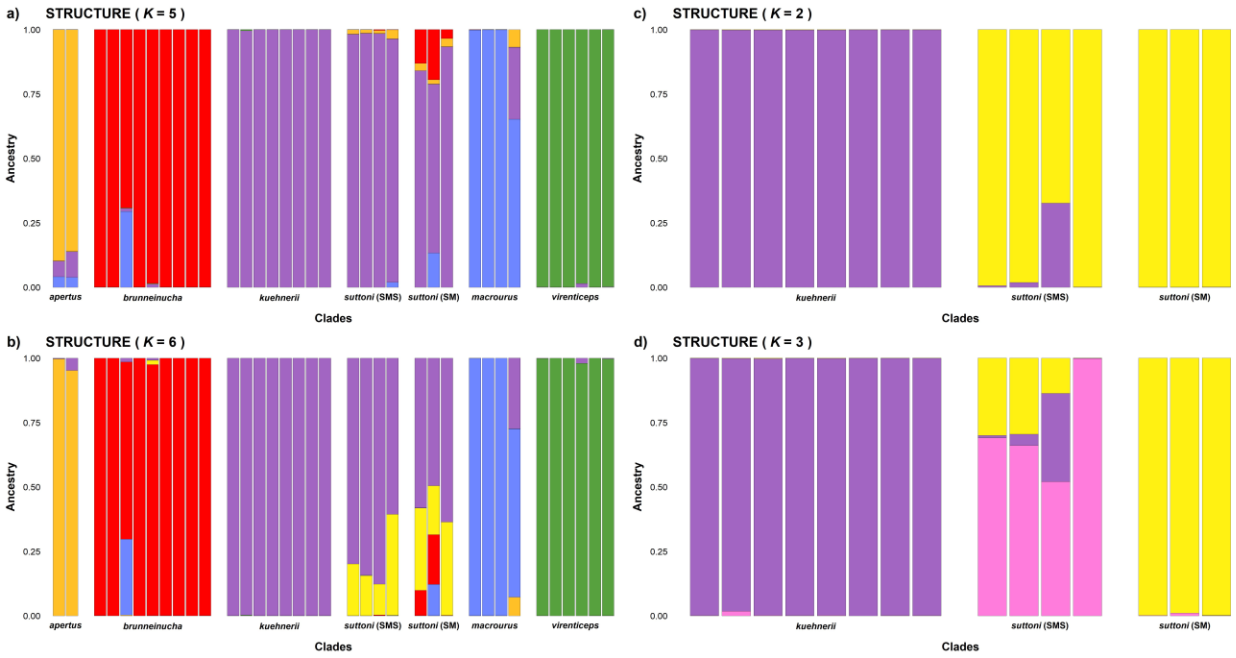


FIGURE 5. Ancestry assignments from STRUCTURE shown as bar plots for all 36 samples, for the two most optimal clustering schemes (a) *K* = 5 and b) *K* = 6) according to ln Pr(*X*|*K*) values, Δ*K* scores, and Puechmaille metrics. Right, additional STRUCTURE iterations (c and d) with only Chestnut-capped Brushfinch from Guerrero (*kuehnerii*) and northern (*suttoni* [SM]) and southern (*suttoni* [SMS]) Oaxaca, revealing hierarchical geographic structure. The colors of the sample assignments of relevant species delimitation schemes (*K* = 5–7) correspond to the colors used in the sampling map (Figure 2).

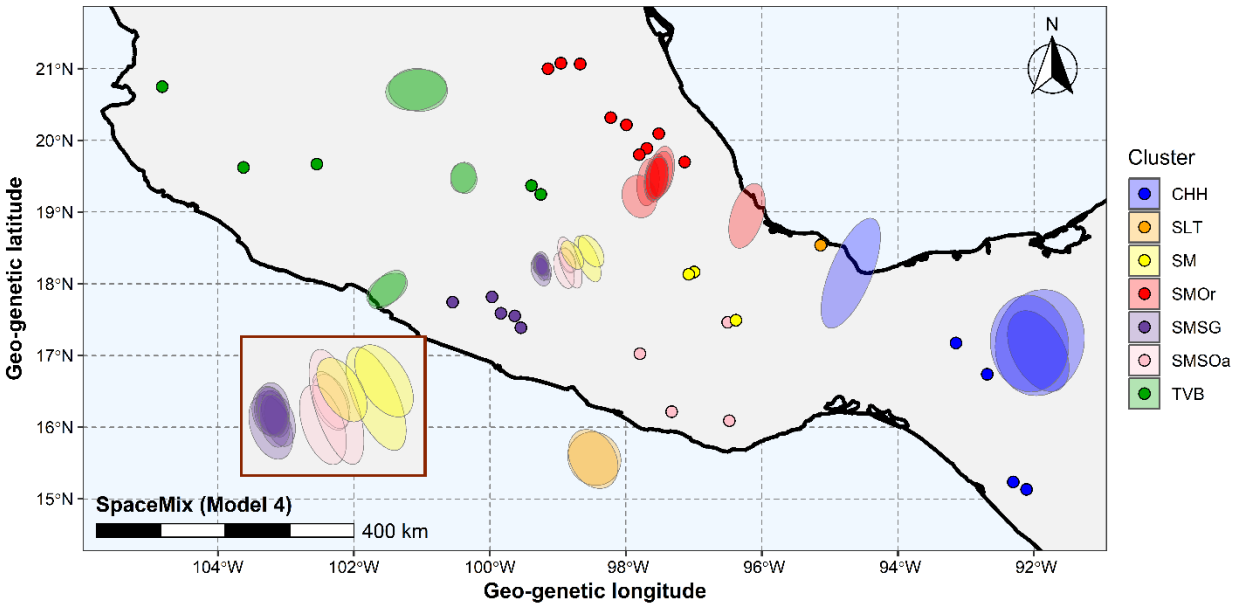


FIGURE 6. SpaceMix results showing the geo-genetic map of Model 4 ("estimating geo-genetic locations and admixture source locations"). Ellipses in SpaceMix models represent the 95% Bayesian credible intervals around each sample's location on the geo-genetic map. The top panel shows the geo-genetic map for all groups, while the inset panel shows a magnified view of the ellipses corresponding to *kuehnerii* (purple [SMS from Guerrero]) and *suttoni* (pink [SMS from Oaxaca] and yellow [Sierra Mazateca]).

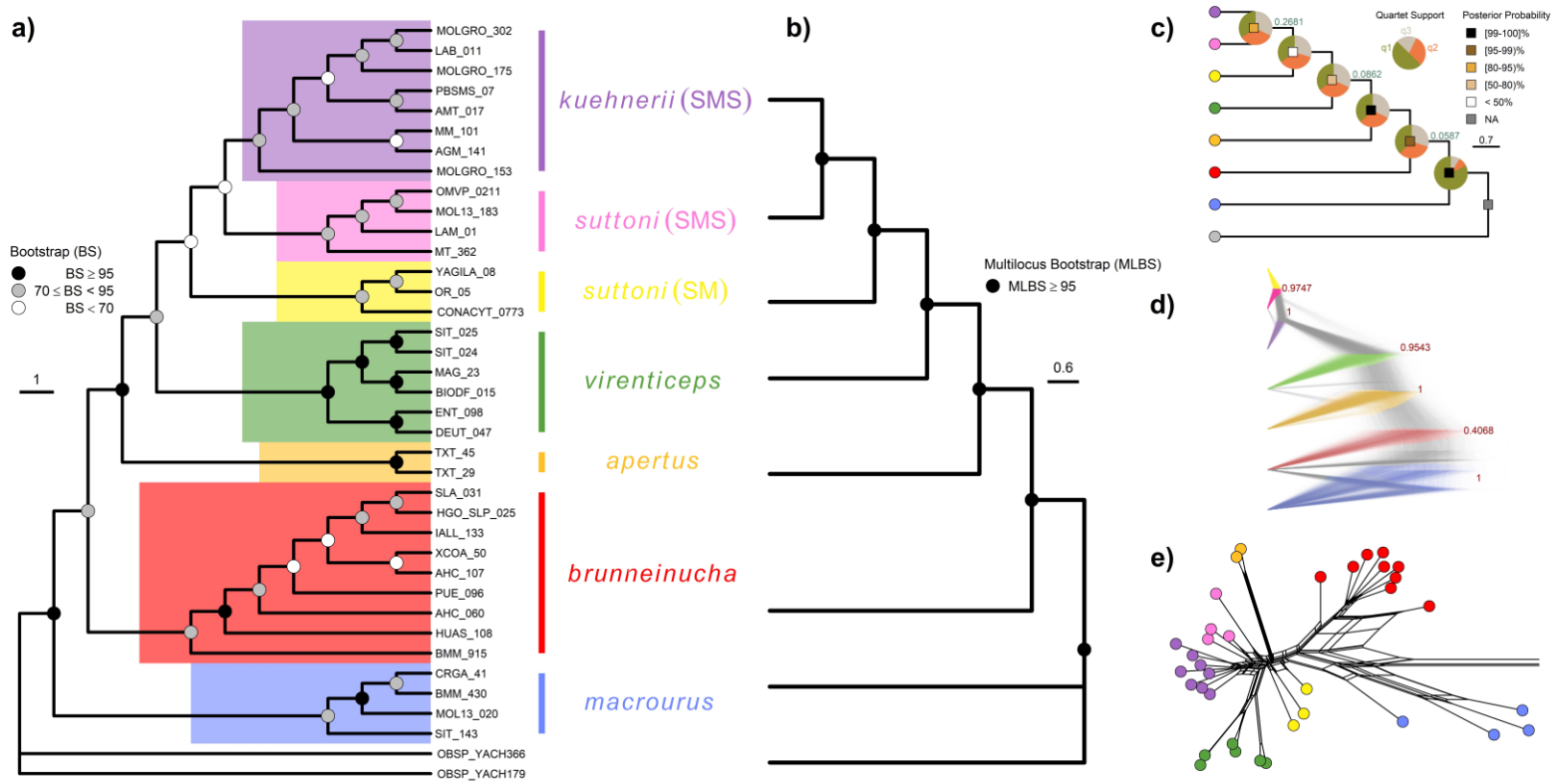


FIGURE 7. a) Concatenated IQ-TREE 30% phylogeny based on 4,490 SNPs depicting ultrafast bootstrap values at each internal node. Species trees (b, c, and d) inferred by assigning samples into seven lineages according to the most lenient (*K* = 7) species delimitation scheme. b) Summary species tree inferred from individual gene trees under the multi-species coalescent model in MP-EST with the multilocus bootstrap values for each internal node. c) Coalescent-based phylogeny inferred using ASTRAL-III, where pie charts on each node denote the fraction of gene trees that are consistent with the topology shown (q1; green), and with the first (q2; orange) and second (q3; grey) alternative topologies. Local posterior probabilities are shown as small color-coded dots in the center of each pie chart. *P*-values colored in green from a polytomy

test are shown for each internal branch (to the right of the corresponding node). A significant ($P < 0.05$) result from the polytomy test indicates a statistical rejection of the null hypothesis that the given branch represents a hard polytomy. d) Simultaneous visualization of 40,000,000 trees sampled via MCMC from the posterior distribution of species trees inferred under the multispecies coalescent model using SNAPP, depicting the posterior probability values (in red) based on the maximum clade credibility (MCC) tree. e) SplitsTree analysis showing the relatedness of all Brushfinches samples as a rooted phylogenetic network (the outgroup is not shown to improve clarity). Each tip is color-coded according to the $K = 7$ species delimitation scheme as in Fig 3. Branch lengths for summary-methods are not to scale.

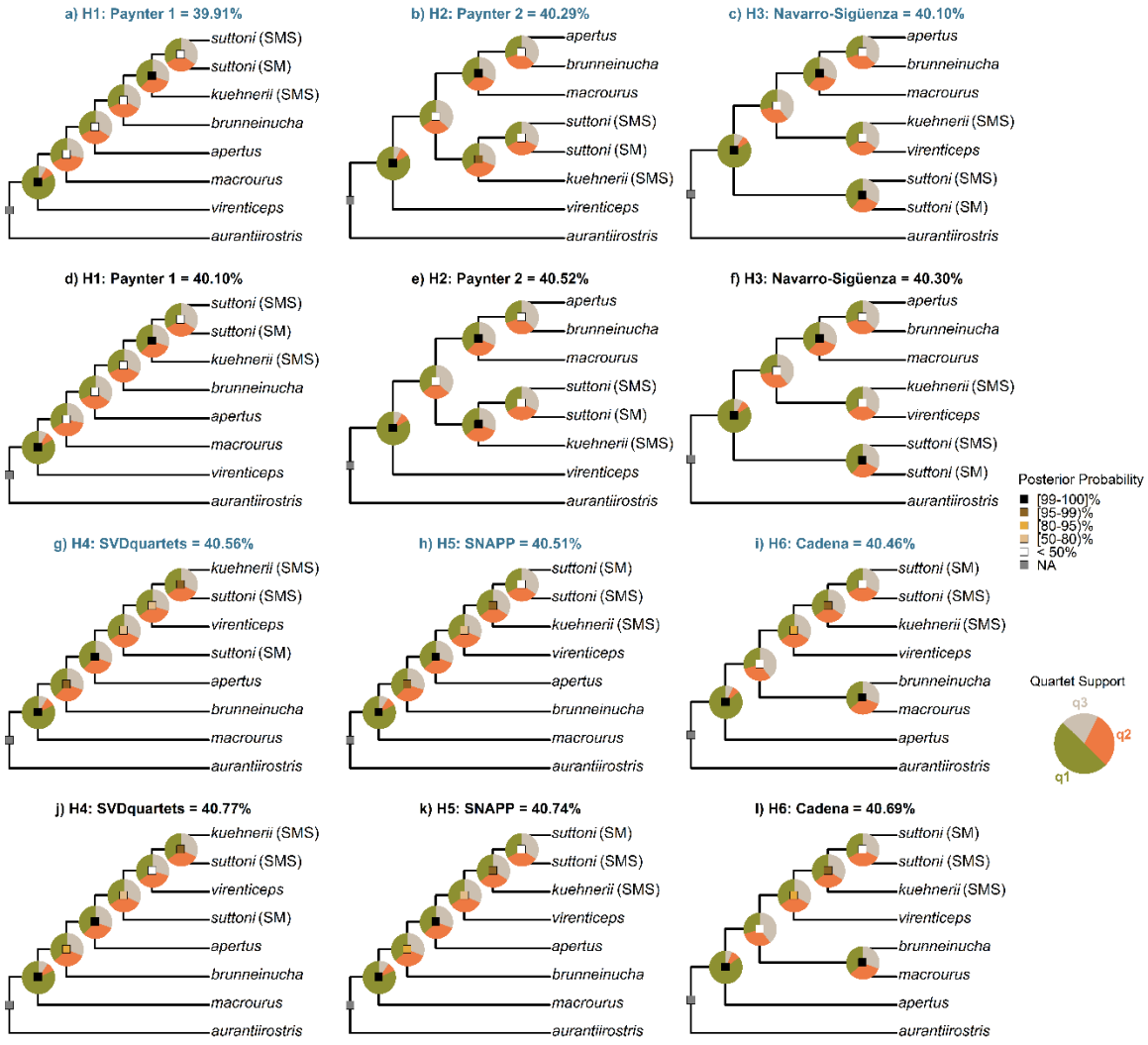


FIGURE 8. Quartet support for six hypotheses (H1 to H6) of relationships among Mesoamerican Brushfinches species. Panels a) to l) each represent one hypothesis given by one topology depending on estimated gene trees used as input: IQ-TREE (blue labels) and RAxML (black labels). We show normalized quartet scores computed by ASTRAL-III (i.e., the percentage of input gene tree quartet trees that are consistent with a topology). Pie charts on each node denote the fraction of estimated gene trees that are consistent with the shown topology (q1; green), and with the first (q2; orange) and second (q3; gray) alternative topologies. Local posterior probabilities are shown as color-coded circles on each node (see legend).

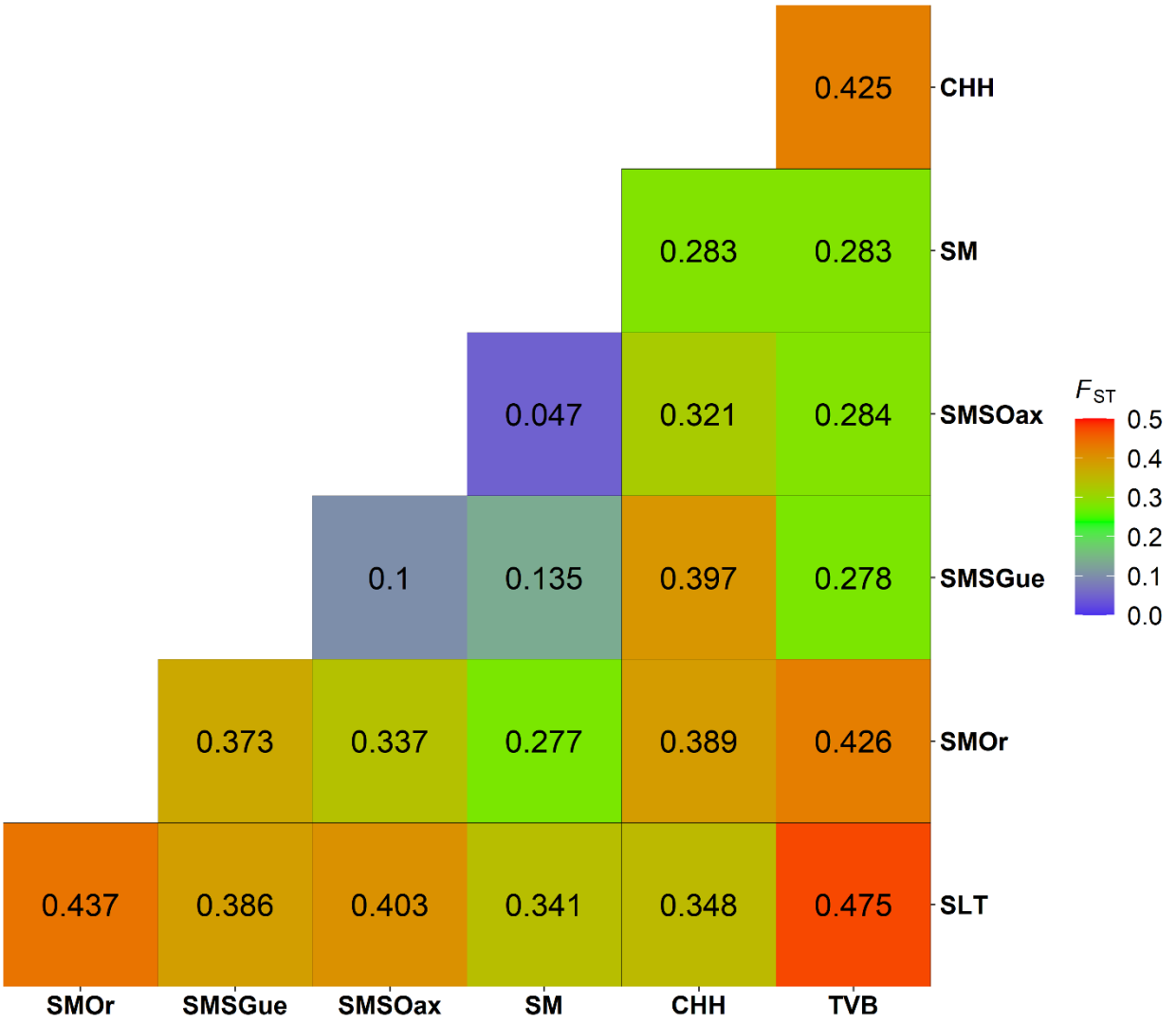


FIGURE 9. Genomic differentiation of the identified Mesoamerican Brushfinches lineages when *K* = 7. The heatmap color-coded according to pairwise F_{ST} between each of the seven putative species.

Supplementary Material

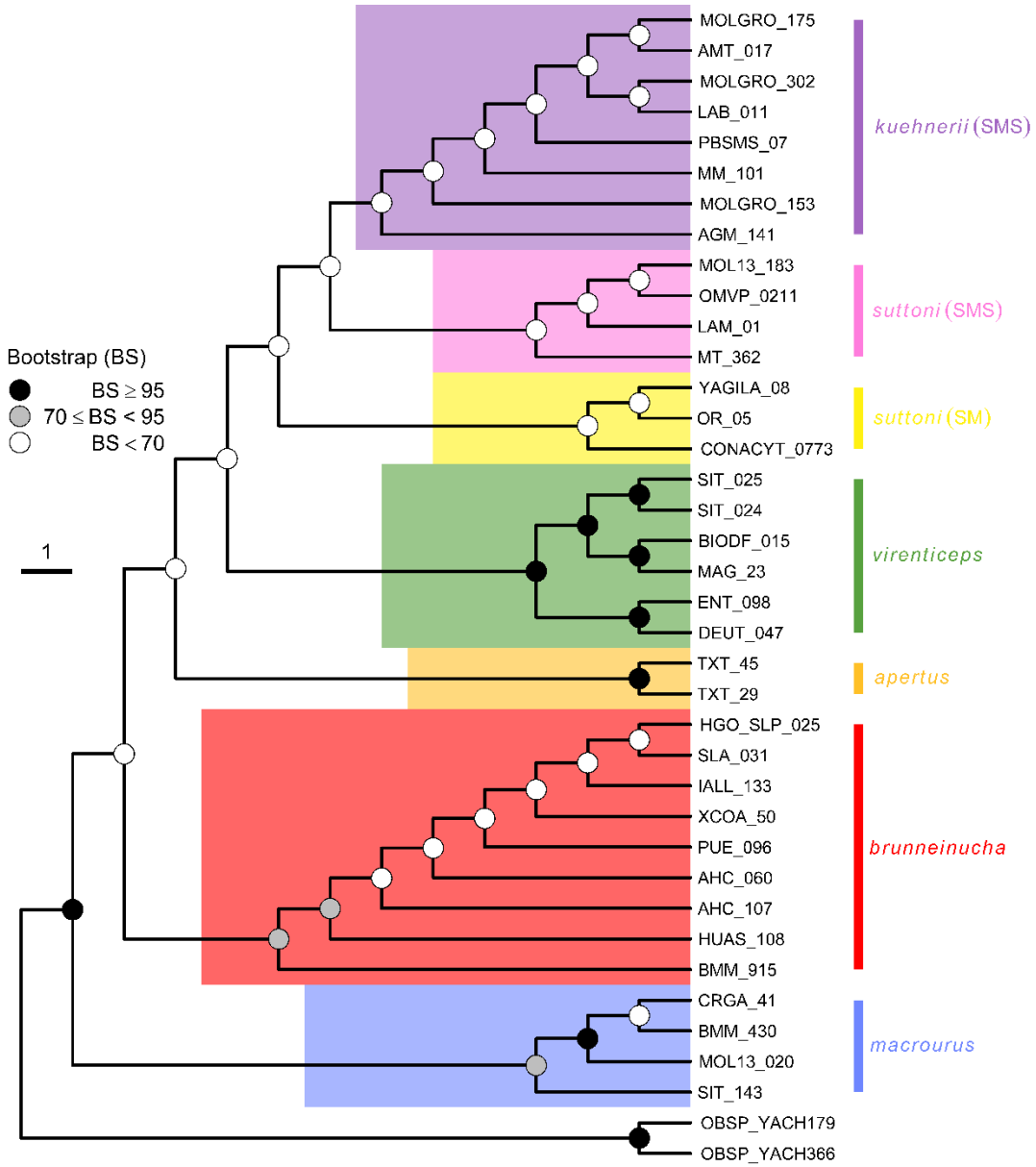


FIGURE S1. Concatenated RAxML 30% gene tree based on 4,490 SNPs depicting rapid bootstrap values on each internal node.

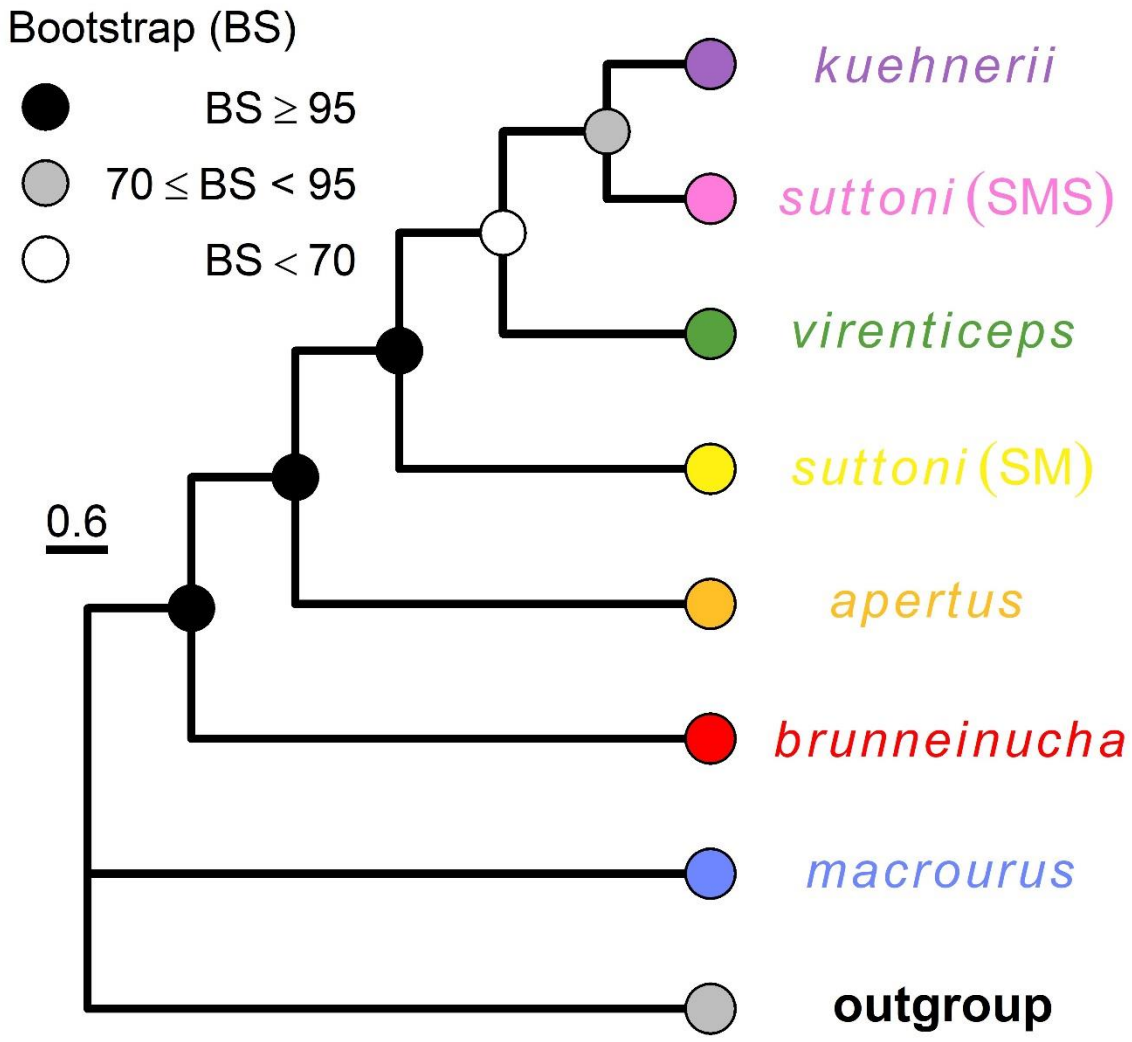


FIGURE S2. Species tree inferred using SVDquartets and 4,490 SNPs depicting bootstrap values on each internal node.

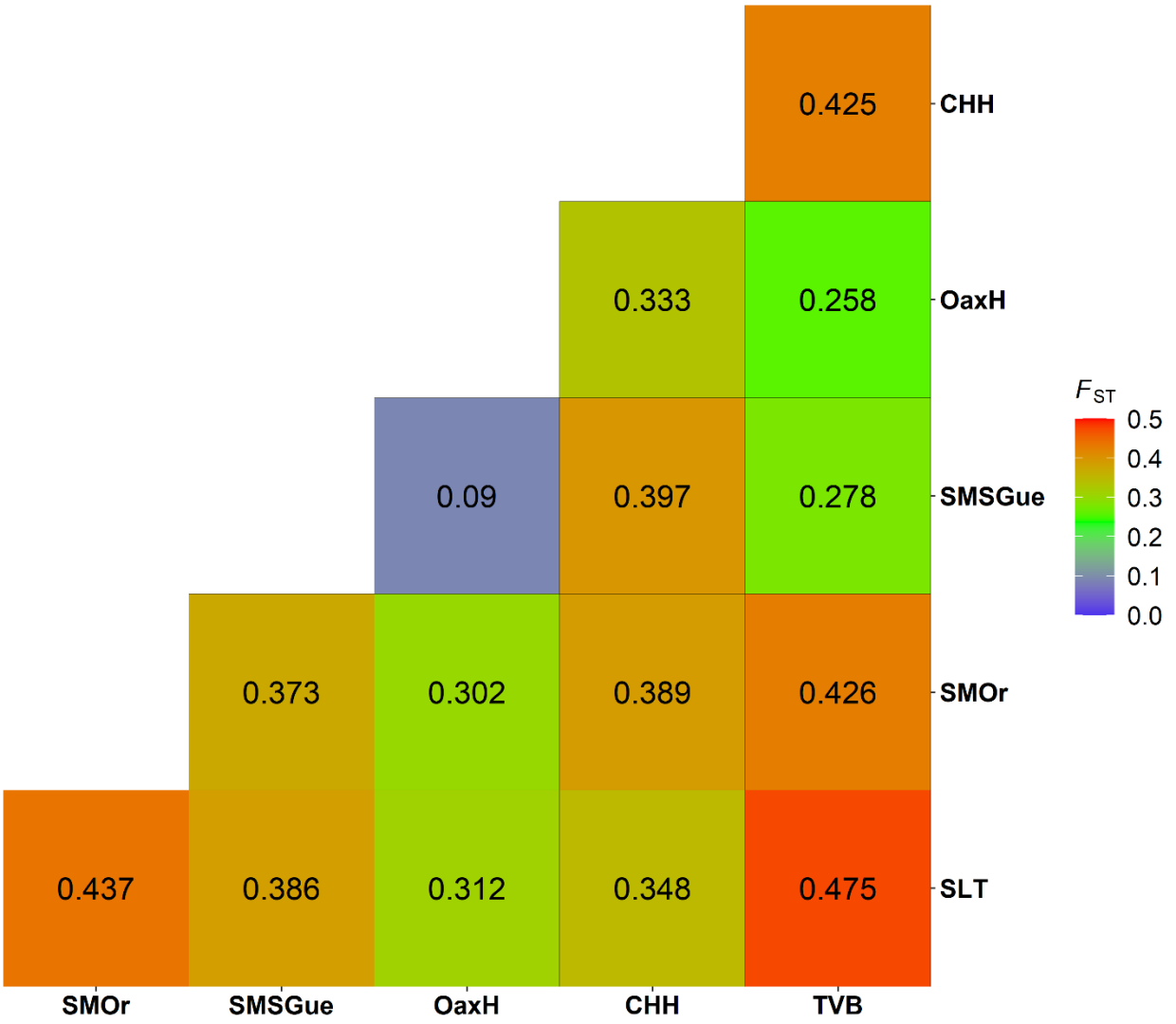


FIGURE S3. Genomic differentiation of the identified Mesoamerican Brushfinches lineages when K = 6. The heatmap is color-coded according to pairwise F_{ST} between each of the six putative species.

DISCUSIÓN Y CONCLUSIONES GENERALES

DISCUSIÓN Y CONCLUSIONES GENERALES

En la presente investigación doctoral se realizaron varios análisis los cuales abordaron temas relacionados a la historia de la ciencia, bibliometría, estadística espacial, biogeografía, ecología, genómica del paisaje y filogenómica. Sólo un marco integral de distintas áreas del conocimiento nos permitirá dilucidar y comprender los patrones históricos y actuales que han proporcionado identidad a la variación genómica observada de las especies. Bajo esta premisa, se vuelve necesario evaluar si los resultados obtenidos se encuentran condicionados por el tipo de escala espacial o temporal empleada, o en efecto, son el resultado genuino de un patrón que se mantiene constante a través de escalas.

Una vez que hemos determinado que cierto patrón se mantiene constante, podemos dar lugar a la formulación de hipótesis evolutivas (usualmente bajo un enfoque filogeográfico comparativo) que necesitarán ser refutadas bajo otros escenarios previamente no incluidos. En este marco general de investigación, se referenció como zona de estudio un área geográfica que se extiende desde México hasta América del Sur (Morrone 2014; Morrone *et al.* 2022), los Bosques Montanos Neotropicales (BMN, o "NMF" por sus siglas en inglés; Churchill *et al.* 1995). Este complejo biogeográfico ha sido intensamente estudiado bajo un enfoque filogeográfico (Ornelas *et al.* 2013; Ramírez-Barahona y Eguiarte 2013), biogeográfico (Sánchez-González *et al.* 2008; Morrone 2014; Morrone *et al.* 2022), y ecológico (Sánchez-González *et al.* 2009). Todas estas aportaciones han significado un paso adelante hacia la comprensión holística de estos ecosistemas, dado que permiten edificar estrategias de conservación basadas en evidencia (Toledo-Aceves *et al.* 2011). No obstante, los NMF aún se encuentran en constantes presiones antropogénicas. Tales presiones selectivas y estocásticas relacionadas al hombre han ocasionado un cambio significativo sobre las comunidades vegetales (modificación del uso de suelo), y un cambio climático que tiende a constreñir las distribuciones geográficas de una gran cantidad de especies endémicas (e.g., Sierra-Morales *et al.* 2021; Matadamas *et al.* 2022).

A pesar de este umbral de incertidumbre, varios investigadores han empleado herramientas moleculares para evaluar hipótesis evolutivas cada vez más sólidas sobre

los procesos subyacentes que delinean una historia evolutiva en común en los NMF (Ornelas *et al.* 2013; Flantua y Hooghiemstra 2018; Flantua 2019). Sin embargo, con la avenida de críticas relacionadas a los métodos tradicionales de secuenciación (e.g., Sanger) y los grandes obstáculos que permanecen al usar los métodos estadísticos tradicionales (Philippe *et al.* 2005; Edwards *et al.* 2015; Teske *et al.* 2018) es imperativo abordar estos temas usando datos de secuenciación masiva.

Específicamente, en esta tesis abordé algunos desafíos metodológicos y teóricos que identifiqué en el campo de la biología evolutiva al emplear datos de secuenciación masiva de nueva generación del tipo subrepresentación genómica (Puritz *et al.* 2014; Toews *et al.* 2016). Actualmente, hay mucha discusión sobre los métodos analíticos y estadísticos en la biología evolutiva (ver Balkenhol *et al.* 2009; Mirarab *et al.* 2016), y existe una necesidad urgente de desarrollar y perfeccionar una teoría centrada en genómica del paisaje, filogenómica y filogeografía (Edwards *et al.* 2009; Wang 2010; Rissler 2016).

En el **capítulo I** se demostró que las herramientas genómicas en el estudio de la variación biológica (genómica de poblaciones, genómica del paisaje, filogenómica, y filogeografía) de los NMF ha experimentado un auge notable a partir del primer artículo publicado en 2014 (Mastretta-Yanes *et al.* 2014). Si bien esto representó un avance en el estudio de los NMF, hubo una carencia de investigaciones publicadas hasta el 2018. A partir de ese momento, se comenzaron a visualizar patrones de acumulación de datos genómicos, y los cuales se ven reflejados en el espacio geográfico.

Notablemente, se encontró que variables espaciales y temporales son las causantes de un posible sesgo de publicación que influyen en que algunas provincias biogeográficas sean más estudiadas usando técnicas de secuenciación de nueva generación. Ambas variables involucradas en el mejor modelo que explica esta variación están relacionadas a un efecto de retardo en su implementación, es decir, a mayor distancia y tiempo del sitio de origen donde se publicaron estas técnicas de secuenciación, mayor será el alcance en que estos métodos de secuenciación serán implementados. Este patrón se mantiene incluso cuando comparamos modelos usando distintas variables relacionadas a la economía y a la naturaleza del tipo de colaboración

(número de investigadores y número de naciones involucradas en un determinado estudio genómico). Queda evidenciado entonces que las herramientas y métodos que se emplean para el estudio de la variación genómica para la formulación de hipótesis filogeográficas (consideradas aquí como “revoluciones científicas”, Kuhn 1962) suelen predominar en sitios cercanos a donde estos surgen. Caso contrario con las provincias ubicadas en Sur América, donde la escasez de tales estudios persiste a pesar de casi 10 años desde el primer estudio sobre taxa habitando los NMF (ver Ortego *et al.* 2023). La Faja Volcánica Transmexicana (Morrone *et al.* 2022) es ciertamente el área con mayor número de estudios genómicos.

Paralelo a esos hallazgos, también queda evidenciado que muchas de las NGS suelen presentar un patrón de retardo para su aplicación. Los NGS pueden aportar a una compresión más amplia del genoma, y la metodología surge a partir de esa indagación puede esclarecer las hipótesis filogeográficas analizadas (“revoluciones científicas”; Kuhn 1962; Laudan 1977). Es decir, posiblemente deban de pasar una o muchas décadas para que estos métodos sean mejor representados en la literatura científica. Nuestros resultados parecen apoyar esa hipótesis. Aunado a eso, no existe una predisposición de que algunos métodos NGS sean más utilizados que otros. Si bien es cierto que algunas NGS como RADseq (Baird *et al.* 2008) fueron creados para ser empleados a escalas espaciales y temporales finas (genómica de poblaciones y genómica del paisaje; Manthey *et al.* 2016; Toews *et al.* 2016), en realidad también son empleados en el estudio de patrones filogenéticos (ver **capítulo I**). Los UCEs que fueron concebidos para ser empleados en escalas temporales más profundas (filogenómica) han permeado mayormente en la literatura de este enfoque (**capítulo I**). Empero, a pesar de su aplicación indistinta en diversas áreas del conocimiento evolutivo, parece ser que se ha hecho caso omiso de algunas sugerencias sobre el tratamiento bioinformático de datos genómicos (Linck y Battey 2019).

Uno de los resultados más sobresaliente fue que muchas de las críticas teóricas y metodológicas de datos tradicionales de secuenciación, también se encuentran fuertemente arraigadas a los datos NGS. Por ejemplo, la persistencia de métodos como las pruebas de Mantel (Mantel 1967) o las regresiones múltiples de matrices de distancia

(“MRDM” por sus siglas en inglés; Legendre *et al.* 1994) aunque útiles en sus inicios, siguen presentando algunos por menores estadísticos (generalmente presentan errores del tipo I) que podrían sesgar las inferencias que tienen las variables paisajísticas sobre la variación genómica (Balkenhol *et al.* 2009; Shirk *et al.* 2018; Peterman y Pope 2021). Si bien, algunos autores han expresado la idea de definir bien las distintas escalas a las que un investigador opera en el estudio de la biología evolutiva (Wang 2010; Edwards *et al.* 2022), también hay otros que sugieren que tal distinción suele ser prematura bajo ciertas circunstancias (Rissler 2016). Posiblemente, esta distinción tradicional (y de alguna forma conveniente) entre disciplinas se deba a que en libros de texto especializados sobre filogenómica no abordan temas relacionados a cuestiones paisajísticas o al estudio de la variación genómica *per se* (pero ver Wiley y Lieberman 2011). Otra posible explicación podría ser la idea del tipo del marcador empleado (Wang 2010): microsatélites para genética del paisaje, mientras que para filogenética y filogeografía, secuencias (mitocondriales, nucleares y de cloroplasto) obtenidas por el método Sanger.

En la presente tesis se identificó que distinguir genómica del paisaje, filogenómica, y filogeografía carece de sentido bajo dos rubros principales: la delimitación de especies y la formulación de hipótesis evolutivas en un contexto comparativo. Esto se debe mayormente a dos áreas de incertidumbre que persisten en la reconstrucción filogenética dentro de los complejos de especies: aislamiento por distancia (“IBD”; Wright 1943) y estructura genética poblacional (“PGS”; Wright 1951). Ambos factores actúan separada o conjuntamente obscureciendo la delimitación de especies (Meirmans 2012; Mason *et al.* 2020), ya sea subestimando o sobreestimando la diversidad de especies que componen un determinado complejo. Por fortuna, muchas de las herramientas que se emplean en genómica del paisaje, cada vez toman mayor auge en el quehacer del filogenetista, típicamente para delimitar complejos de especies. Incluso cuando controlamos por el efecto del IBD o PGS (esta última como variable aleatoria) en nuestras inferencias evolutivas (usando efectos poblacionales de máxima verosimilitud, “MLPE” Clarke *et al.* 2002), necesariamente estamos evocando el factor histórico típico de las escalas filogeográficas.

Con los resultados obtenidos en el **capítulo I**, vimos la necesidad de incorporar la mayoría de esas áreas de incertidumbre (IBD y PGS) adentrándonos a dos escalas: fina (genómica del paisaje; **capítulo II**) y amplia (filogeográfica; **capítulo III**). Como modelo de estudio, se utilizó un complejo de especies que han permanecido desde su descripción taxonómica como dos especies (Hellmayr 1938; Parkes 1954; 1957; Paynter 1978), el complejo *Arremon brunneinucha*/*virenticeps* (Aves, Passerellidae) del norte de los bosques montanos de Mesoamérica. Si bien en este complejo se pensó que sus relaciones evolutivas estaban bien caracterizadas (Hellmayr 1938; Parkes 1954; 1957; Paynter 1978), análisis filogenéticos basados en secuencias mitocondriales y nucleares demuestran lo contrario (Cadena *et al.* 2007; Navarro-Sigüenza *et al.* 2008; Flórez-Rodríguez *et al.* 2011; Navarro-Sigüenza *et al.* 2013). Lo que complica la situación es que un linaje de plumaje verde (*Arremon virenticeps*) se encuentra anidado dentro de un grupo de linajes con plumaje rojo correspondientes a *Arremon brunneinucha* (Navarro-Sigüenza et al. 2008; Navarro-Sigüenza et al. 2013). Esto pone a la luz las limitantes del uso de secuencias Sanger para desentrañar la historia evolutiva de un complejo de especies asociadas a sitios montañosos, incluso para especies donde hay una estructura genética significativa y bien delimitada.

De esta forma, la información genómica analizada (NextRAD, Russello *et al.* 2015) en este trabajo permitió comprender y evaluar la estructura poblacional, así como la influencia que tiene el paisaje (actual e histórico) sobre la variación genómica de este grupo de especies con una baja vagilidad y fuertemente asociado a bosques montanos del Neotrópico. Específicamente, pudimos determinar que las variables relacionadas a la resistencia paisajística actual y la complejidad topográfica han estado reforzando los distintos niveles de diversidad genética en este grupo de aves (**capítulo II**). Esto coincide con las hipótesis evolutivas evaluadas (Janzen 1967; Thom *et al.* 2021) donde se sugiere que varias especies montanas del Neotrópico se encuentran fuertemente constreñidas por el arreglo topográfico del paisaje (Linck *et al.* 2021).

A pesar de que existe un fuerte componente de IBD dentro de este grupo (Moreno-Contreras *et al.* 2023), mientras condicionemos por esta variable, nuestros resultados permanecen constantes, incluso independientemente del método de

regresión empleado en la mayoría de los casos. Es de esta forma que un modelo de aislamiento por resistencia ("IBR"; McRae 2006) y no efectos de aislamiento por ambiente (Wang y Bradburd 2014) o de barrera (Balkenhol *et al.* 2017) da forma a la variación genómica dentro de este grupo de aves. Además, encontramos que este complejo de especies presenta atributos de dos hipótesis filogeográficas (ver Barahona y Eguiarte 2013), lo cual sugiere que existe un continuo que se pasa por alto al sobresimplificar los patrones de variación genómica en taxa de los bosques montanos en el Neotrópico (ver hipótesis alternativas en Ornelas y González 2014; Mastretta-Yanes *et al.* 2018; Flantua y Hooghiemstra 2018; Flantua *et al.* 2019).

Considerando lo arriba expuesto, subsecuentemente se dio paso a evaluar las relaciones evolutivas dentro del complejo *Arremon brunneinucha*/*virenticeps* (**capítulo III**). Dado que es un complejo de especies con tiempos de divergencias relativamente antiguos (Buainain *et al.* 2022), se esperan niveles significativos de sorteo incompleto de linajes (Maddison 1997; Edwards 2009). Lo que hace necesario el uso de herramientas que tomen en cuenta esta incertidumbre al momento de evaluar las relaciones evolutivas (Liu *et al.* 2010; Mirarab *et al.* 2016). Cual sea el método de reconstrucción filogenética empleado, una matriz concatenada de SNPs o árboles de especies basada en la teoría coalescente, se encontró una fuerte estructura filogenética. Todos los grupos fueron monofiléticos con altos valores de soporte en los nodos. No obstante, hubo algunas discrepancias en las topologías obtenidas.

La principal fue que el grupo *kuehnerii* (de Guerrero; Navarro-Sigüenza *et al.* 2008; Navarro-Sigüenza *et al.* 2013) es un grupo hermano a un clado que contiene a dos grupos de *suttoni* que son de la Sierra Mazateca en Oaxaca y la Sierra Madre del Sur de Oaxaca (Parkes 1954; 1957; Paynter 1978). Por otro lado, *suttoni* de la Sierra Mazateca es basal a un grupo que consiste en *kuehnerii* y de *suttoni*, ambos de la Sierra Madre del Sur. Ante estas discrepancias, métodos de estructura genética (espaciales y no espaciales) parecen aclarar la situación. Puntualmente, cuando evaluamos estructura genética el algoritmo sugiere que hay cinco o seis especies. Un clúster formado de *kuehnerii* y los *suttoni* parecen indicar que hay presencia de estructura jerárquica. Una vez que ahondamos dentro de este clúster, se sugiere una subdivisión de dos o tres

clústers. De hecho, el primer clúster que se formó es uno que consiste en puras muestras de *kuehnerii*, coincidiendo con una de las topologías obtenidas. Estos análisis se ven favorecidos con lo obtenido en SpaceMix (Bradburd *et al.* 2016), donde bajo un modelo de IBD, admixia y migración, ninguna muestra de *kuehnerii* se traslapa con las muestras de *suttoni*. Tales resultados coinciden geográficamente con lo expuesto en Nieto-Samaniego *et al.* (2006), dados los patrones de formación de la Sierra Madre del Sur que fueron de oeste (Guerrero) a este (Oaxaca).

Estos hallazgos relacionados a aspectos de la estructura genética y filogenética (controlando por IBD, PGS, y sorteo incompleto de linajes) en conjunto con factores tectónicos (Nieto-Samaniego *et al.* 2006) sugieren que *kuehnerii* amerita tener rango de especie (Navarro-Sigüenza et al. 2008; Navarro-Sigüenza et al. 2013). El hecho de que ambas porciones (este y oeste) de la Sierra Madre del Sur presenten especies diferentes que son evolutivamente cercanas, pero con un alto grado de divergencia genómica no es nuevo. Se tiene documentado en la literatura sobre este patrón filogeográfico de un periodo de aislamiento largo o mediano significativo para colibríes del género *Eupherusa* (Rocha-Méndez *et al.* 2019), codornices *Dendrortyx* (Tsai *et al.* 2019) y charas del género *Aphelocoma* (Venkatraman *et al.* 2019). La co-divergencia en esos taxa podría estar relacionada con sus rasgos de historia de vida como el presentar una baja vagilidad y su asociación estrecha con los bosques montanos (Paynter 1978; Howell y Webb 1995; Watson 2003). Nuestro estudio agrega información a la evidencia ya acumulada de que los análisis filogenéticos de ADNmt suelen presentan ramas cortas y topologías conflictivas (p. ej., Navarro-Sigüenza *et al.* 2008; ver también Barrera-Guzmán *et al.* 2012 para un caso similar). Específicamente, topologías de ADNmt separan evolutivamente los linajes que se encuentran geográficamente cercanos (p. ej., entre y/o dentro de Guerrero y Oaxaca), mientras que las topologías basadas en NGS generalmente recuperan ambos linajes como grupos hermanos (ver también Zarza *et al.* 2018; Tsai *et al.* 2019). Independientemente de si el análisis completo sugiere que son cinco, seis, o hasta siete especies (dependiendo del concepto de especie empleado; Coyne y Orr 2004; De Queiroz 2007), ciertamente encontramos que lo previamente observado sobreestima los altos niveles de diversidad y variación que hay en el grupo de los rascadores *Arremon* de Mesoamérica.

En conclusión, los tres capítulos de la presente tesis ofrecen un marco integrador al estudio evolutivo de los Bosques Montanos Neotropicales. El uso de técnicas moleculares modernas, así como la implementación de diversos análisis espaciales y ecológicos permiten la generación y refutación de hipótesis evolutivas. Además, comprobamos cuantitativamente que los datos de secuenciación masiva proveen de una mayor resolución al responder preguntas evolutivas que previamente se encontraban condicionadas por los métodos usados. Existe una idea errónea de que la filogenómica se trata únicamente de evaluar hipótesis evolutivas novedosas, al contar con una mayor cantidad de datos genómicos. No obstante, con el desarrollo en paralelo de métodos estadísticos cada vez más robustos, también nos permite cuantificar los niveles de incertidumbre de topologías filogenéticas previas. Con lo arriba expuesto, el presente estudio también hace hincapié en la necesidad de invertir recursos y fondos dirigidos a aquellas provincias biogeográficas poco estudiadas, particularmente en regiones que se encuentran en Centroamérica y América del Sur. Sólo con un panorama más amplio bajo distintos escenarios, obtendremos una pieza fundamental en el conocimiento de una zona que se caracteriza por presentar altos valores de diversidad taxonómica y altos grados de endemismos.

REFERENCIAS BIBLIOGRÁFICAS

Avise JC (2009) Phylogeography: retrospect and prospect. J Biogeogr 36:3–15.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376.

Balkenhol N, Dudaniec RY, Krutovsky KV, Johnson JS, Cairns DM, Segelbacher G, Selkoe KA, von der Heyden S, Wang IJ, Selmoni O, Joost S (2017) Landscape genomics: Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora OP (ed) Population genomics concepts, strategies and approaches. Springer, Cham, Switzerland, pp 261–372

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

Barrera-Guzmán AO, Milá B, Sánchez-González LA, Navarro-Sigüenza AG (2012) Speciation in an avian complex endemic to the mountains of Middle America (*Ergaticus*, Aves: Parulidae). Mol Phylogenet Evol 62:907–920.

Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. PLoS Genet 12:e1005703.

Buainain N, Ferreira M, Avendaño JE, Cadena CD, Faircloth BC, Brumfield RT, Cracraft J, Ribas CC (2022) Biogeography of a neotropical songbird radiation reveals similar diversification dynamics between montane and lowland clades. J Biogeogr 49:1260–1273.

Cadena CD, Klicka J, Ricklefs RE (2007) Evolutionary differentiation in the Neotropical montane region: Molecular phylogenetics and phylogeography of *Buarremon* brush-finches (Aves, Emberizidae). Mol Phylogenet Evol 44:993–1016.

Castoe TA., Daza JM, Smith EN, Sasa MM, Kuch U, Campbell JA, Chippindale PT, Parkinson CL (2009) Comparative phylogeography of pitvipers suggests a

consensus of ancient Middle American highland biogeography. J Biogeogr 36:88–103.

Churchill SP, Balslev H, Forero E, Luteyn JM (eds) (1995) Biodiversity and conservation of Neotropical Montane Forests. Proceeding of the Neotropical Montane Forest Biodiversity and Conservation Symposium, The New York Botanical Garden, 21–26 June 1993. The New York Botanical Garden, Bronx, New York, pp 103–126.

Clarke RT, Rothery P, Raybould AF (2002) Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. J Agric Biol Environ Stat 7:361–372.

Coates AG, Obando JA (1996) The geologic evolution of the Central American isthmus. In: Jackson JBC, Budd A, Coates AG (eds), Evolution and environment in tropical America. University of Chicago Press, Illinois, pp 21–56.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Massachusetts.

Dantas-Queiroz MV, Hurbath F, de Russo Godoy FM, Lanna FM, Versieux LM, Palma-Silva C (2023) Comparative phylogeography reveals the demographic patterns of neotropical ancient mountain species. Mol Ecol 32:3165–3181.

Daza JM, Castoe TA, Parkinson CL (2010) Using regional comparative phylogeographic data from snake lineages to infer historical processes in Middle America. Ecography 33.343–354.

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886.

Edwards SV (2009) Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Mol Ecol 18:2930–2933.

Edwards SV, Shultz AJ, Campbell-Staton SC (2015) Next-generation sequencing and the expanding domain of phylogeography. Folia Zool 64:187–206.

Edwards SV, Robin VV, Ferrand N, Moritz C (2022) The evolution of comparative phylogeography: putting the geography (and more) into comparative population genomics. Genome Biol Evol 14:evab176.

Flantua SGA, Hooghiemstra H (2018) Historical connectivity and mountain biodiversity (with special reference to the Northern Andes). In: Hoorn C, Perrigo A, Antonelli A (eds) Mountains, Climate and Biodiversity. John Wiley & Sons Ltd, Chichester, United Kingdom, pp 171–185.

Flantua SGA, O'Dea A, Onstein RE, Giraldo C, Hooghiemstra H (2019) The flickering connectivity system of the north Andean páramos. J Biogeogr 46:1808–1825.

Flórez-Rodríguez A, Carling MD, Cadena CD (2011) Reconstructing the phylogeny of "*Buarremon*" brush-finches and near relatives (Aves, Emberizidae) from individual gene trees. Mol Phylogenet Evol 58:297–303.

Foster P (2001) The potential negative impacts of global climate change on tropical montane cloud forests. Earth-Sci Rev 55:73–106.

Gentry AH (1995) Patterns of diversity and floristic composition in Neotropical montane forests. In: Churchill SP, Balslev H, Forero E, Luteyn JM (eds) Biodiversity and conservation of Neotropical Montane Forests. Proceeding of the Neotropical Montane Forest Biodiversity and Conservation Symposium, The New York Botanical Garden, 21–26 June 1993. The New York Botanical Garden, Bronx, New York, pp 103–126.

Ghiselin MT (1997) Metaphysics and the Origin of Species. State University of New York Press, New York.

Graham A (1999) The Tertiary history of the northern temperate element in the northern Latin American biota. Am J Bot 86:32–38.

Gutiérrez-Rodríguez J, Zaldívar-Riverón A, Weissman DB, Vandergast AG (2022) Extensive species diversification and marked geographic phylogenetic structure in

the Mesoamerican genus *Stenopelmatus* (Orthoptera: Stenopelmatidae: Stenopelmatinae) revealed by mitochondrial and nuclear 3RAD data. Invertebr Syst 36:1–21.

Hellmayr CE (1938) Catalogue of birds of the Americas. Part XI: Ploceidae - Catamblyrhynchidae - Fringillidae. Field Museum of Natural History Zoological Series, Volume XIII, Part XI. Field Museum of Natural History, Chicago.

Hernández-Baños BE, Peterson AT, Navarro-Sigüenza AG, Escalante-Pliego BP (1995) Bird faunas of the humid montane forests of Mesoamerica: biogeographic patterns and priorities for conservation. Bird Conserv Int 5:251–277.

Hernández-Langford DG, Siqueiros-Delgado ME, Ruíz-Sánchez E (2020) Nuclear phylogeography of the temperate tree species *Chiranthodendron pentadactylon* (Malvaceae): Quaternary relicts in Mesoamerican cloud forests. BMC Evol Biol 20:44.

Howell SNG, Webb S (1995) A guide to the birds of Mexico and northern Central America. Oxford University Press, United Kingdom.

Janzen DH (1967) Why mountain passes are higher in the tropics. Am Nat 101:233–249.

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago.

Laudan L (1977) Progress and its problems. University of California Press, California.

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487–1499.

Linck E, Battey CJ (2019) Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol Ecol Resour 19:639–647.

Linck EB, Freeman BG, Cadena CD, Ghalambor CK (2021) Evolutionary conservatism will limit responses to climate change in the tropics. Biol Lett 17:20210363.

Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol Biol 10:302.

Maddison WP (1997) Gene trees in species trees. Syst Biol 46:523–536.

Manel S, Holderegger R (2013) Ten years of landscape genetics. Trends Ecol Evol 28:614–621.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220.

Manthey JD, Campillo LC, Burns KJ, Moyle RG (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: Piranga). Syst Biol 65:640–650.

Marshall CJ, Liebherr JK (2000) Cladistic biogeography of the Mexican transition zone. J Biogeogr 27:203–216.

Mason NA, Fletcher NK, Gill BA, Funk WC, & Zamudio KR (2020) Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. Syst Biodivers 18:269–280.

Mastretta-Yanes A, Zamudio S, Jorgensen TH, Arrigo N, Alvarez N, Piñero D, Emerson BC (2014) Gene duplication, population genomics, and species-level differentiation within a tropical mountain shrub. Genome Biol Evol 6:2611–2624.

Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC (2015) Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J Biogeogr 42:1586–1600.

Mastretta-Yanes A, Xue AT, Moreno-Letelier A, Jorgensen TH, Alvarez N, Piñero D, Emerson BC (2018) Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. Mol Ecol 27:432–448.

Matadamas RE, Enríquez PL, Guevara L, Navarro-Sigüenza AG (2022) Stairway to extinction? Influence of anthropogenic climate change on distribution patterns of montane Strigiformes in Mesoamerica. Avian Conserv Ecol 17:37.

McRae BH (2006) Isolation by resistance. Evolution 60:1551–1561.

Meirmans PG (2012) The trouble with isolation by distance. Mol Ecol 21:2839–2846.

Mirarab S, Bayzid MS, Warnow T (2016) Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst Biol 65:366–380.

Montes C, Cardona A, McFadden R, Morón SE, Silva CA, Restrepo-Moreno S, Ramírez DA, Hoyos N, Wilson J, Farris D, Bayona GA, Jaramillo CA, Valencia V, Bryan J, Flores JA (2012) Evidence for middle Eocene and younger land emergence in central Panama: Implications for Isthmus closure. Geol Soc Am Bull 124:780–799.

Moreno-Contreras I, Sánchez-González LA, Arizmendi MC, Prieto-Torres DA, Navarro-Sigüenza AG (2020) Climatic niche evolution in the *Arremon brunneinucha* complex (Aves: Passerellidae) in a Mesoamerican landscape. Evol Biol 47:123–132.

Morrone JJ (2010) Fundamental biogeographic patterns across the Mexican Transition Zone: an evolutionary approach. Ecography 33:355–361.

Morrone JJ (2014) Biogeographical regionalisation of the Neotropical region. Zootaxa, 3782:1–110.

Morrone JJ (2015) Halffter's Mexican transition zone (1962–2014), cenocrons and evolutionary biogeography. J Zool Syst Evol Res, 53(3):249–257.

Morrone JJ, Escalante T, Rodríguez-Tapia G, Carmona A, Arana M, and Mercado-Gómez JD (2022) Biogeographic regionalization of the Neotropical region: New map and shapefile. An Acad Bras Cienc 94:e20211167.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858.

Navarro-Sigüenza AG, Peterson AT, Nyari A, García-Deras GM, García-Moreno J (2008) Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. Mol Phylogenet Evol 47:21–35.

Navarro-Sigüenza AG, García-Hernández MA, Peterson AT (2013) A new species of brush-finch (*Arremon*; Emberizidae) from western Mexico. Wilson J Ornithol 125:443–453.

Nieto-Samaniego AF, Alaniz-Álvarez SA, Silva-Romo G, Eguiza-Castro MH, Mendoza-Rosales CC (2006) Latest Cretaceous to Miocene deformation events in the eastern Sierra Madre del Sur, Mexico, inferred from the geometry and age of major structures. Geol Soc Am Bull 118:238–252.

Ornelas JF, Ruiz-Sánchez E, Sosa V (2010) Phylogeography of *Podocarpus matudae* (Podocarpaceae): pre-Quaternary relicts in northern Mesoamerican cloud forests. J Biogeogr 37:2384–2396.

Ornelas JF, Sosa V, Soltis DE, Daza JM, González C, Soltis PS, Gutiérrez-Rodríguez C, Espinosa de los Monteros A, Castoe TA, Bell C, Ruiz-Sanchez E (2013) Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. PLoS ONE 8:e56283.

Ornelas JF, González C (2014) Interglacial genetic diversification of *Moussonia deppeana* (Gesneriaceae), a hummingbird-pollinated, cloud forest shrub in northern Mesoamerica. Mol Ecol 23:4119–4136.

Ortego J, Espelta JM, Armenteras D, Díez MC, Muñoz A, Bonal R (2023) Demographic and spatially explicit landscape genomic analyses in a tropical oak reveal the impacts of late Quaternary climate change on Andean montane forests. Mol Ecol 32:3182–3199.

Palacio RD, Kattan GH, Pimm SL (2020) Bird extirpations and community dynamics in an Andean cloud forest over 100 years of land-use change. Conserv Biol 34:677–687.

Parkes KC (1954) A revision of the Neotropical finch *Atlapetes brunneinucha*. Condor. 56:129–138.

Parkes KC (1957) The juvenal plumages of the finch genera *Atlapetes* and *Pipilo*. Auk. 74:499–502.

Paynter RA Jr (1978) Biology and evolution of the avian genus *Atlapetes* (Emberizinae). Bull Mus Comp Zool Harv Coll 148:32–369.

Perez MF, Franco FF, Bombonato JR, Bonatelli IA, Khan G, Romeiro-Brito M, Fegies AC, Ribeiro PM, Silva GAR, Moraes EM (2018) Assessing population structure in the face of isolation by distance: are we neglecting the problem? Divers Distrib 24:1883–1889.

Peterman WE, Pope NS (2021) The use and misuse of regression models in landscape genetic analyses. Mol Ecol 30:37–47.

Peterson AT, Escalante-Pliego P, Navarro-Sigüenza A (1992) Genetic variation and differentiation in Mexican populations of Common Bush-Tanagers and Chestnut-capped Brush-Finches. Condor 94:244–253.

Philippe H, Delsuc F, Brinkmann H, Lartillot N (2005) Phylogenomics. Annu Rev Ecol Evol Syst 36:541–562.

Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE (2014) Demystifying the RAD fad. Mol Ecol 23:5937–5942.

Ramírez-Barahona S, Eguiarte LE (2013) The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. Ecol Evol 3:725–738.

Rissler LJ (2016) Union of phylogeography and landscape genetics. Proc Natl Acad Sci USA 113:8079–8086.

Rojas-Soto OR, Sosa V, Ornelas JF (2012) Forecasting cloud forest in eastern and southern Mexico: conservation insights under future climate change scenarios. Biodivers Conserv 21, 2671–2690.

Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106.

Sánchez-González LA, Morrone JJ, Navarro-Sigüenza AG (2008) Distributional patterns of the Neotropical humid montane forest avifaunas. Biol J Linn Soc 94:175–194.

Sánchez-González LA, Navarro-Sigüenza AG (2009) History meets ecology: a geographical analysis of ecological restriction in the Neotropical humid montane forests avifaunas. Divers Distrib 15:1–11.

Scoble J, Lowe AJ (2010) A case for incorporating phylogeography and landscape genetics into species distribution modelling approaches to improve climate adaptation and conservation planning. Divers Distrib 16:343–353.

Shirk AJ, Landguth EL, Cushman SA (2018) A comparison of regression methods for model selection in individual-based landscape genetic analysis. Mol Ecol Resour 18:55–67.

Sierra-Morales P, Rojas-Soto O, Ríos-Muñoz CA, Ochoa-Ochoa LM, Flores-Rodríguez P, Almazán-Núñez RC (2021) Climate change projections suggest severe decreases in the geographic ranges of bird species restricted to Mexican humid mountain forests. Glob Ecol Conserv 30:e01794.

Stange M, Sánchez-Villagra MR, Salzburger W, Matschiner M (2018) Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. Syst Biol 67:681–699.

Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, van der Lingen CD, von der Heyden S, Chiazzari B, van Vuuren BJ, Beheregaray LB (2018) Mitochondrial DNA is unsuitable to test for isolation by distance. Sci Rep 8:8448.

Thom G, Gehara M, Smith BT, Miyaki CY, do Amaral FR (2021) Microevolutionary dynamics show tropical valleys are deeper for montane birds of the Atlantic Forest. Nat Commun 12:6269.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21:3907–3930.

Toews DPL, Campagna L, Taylor SA, Balakrishnan CN, Baldassarre DT, Deane-Coe PE, Harvey MG, Hooper DM, Irwin DE, Judy CD, Mason NA, McCormack JE, McCracken KG, Oliveros CH, Safran RJ, Scordato ESC, Stryjewski KF, Tigano A, Uy JAC, Winger BM (2016) Genomic approaches to understanding population divergence and speciation in birds. Auk 133:13–30.

Toledo-Aceves T, Meave JA, González-Espinosa M, Ramírez-Marcial N (2011) Tropical montane cloud forests: current threats and opportunities for their conservation and sustainable management in Mexico. J Environ Manage 92:974–981.

Tsai WLE, Mota-Vargas C, Rojas-Soto O, Bhowmik R, Liang EY, Maley JM, Zarza E, McCormack JE (2019) Museum genomics reveals the speciation history of *Dendrortyx* wood-partridges in the Mesoamerican highlands. Mol Phylogenet Evol 136:29–34.

Venkatraman MX, DeRaad DA, Tsai WLE, Zarza E, Zellmer AJ, Maley JM, McCormack JE (2019) Cloudy with a chance of speciation: integrative taxonomy reveals extraordinary divergence within a Mesoamerican cloud forest bird. Biol J Linn Soc 126:1–15.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. Mol Ecol 19:2605–2608.

Wang IJ, Bradburd GS (2014) Isolation by environment. Mol Ecol 23:5649–5662.

Watson DM (2003) Long-term consequences of habitat fragmentation—highland birds in Oaxaca, Mexico. Biol Conserv 111:283–303.

Webb SD (2006) The Great American biotic interchange: patterns and processes. Ann Missouri Bot Gard 93:245–257.

Wiley EO, Lieberman BS (2011) Phylogenetics: theory and practice of phylogenetic systematics, 2nd edn. John Wiley & Sons, Inc, New Jersey.

Wright S (1943) Isolation by distance. Genetics 28:114–138.

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354.

Zarza E, Connors EM, Maley JM, Tsai WLE, Heimes P, Kaplan M, McCormack JE (2018) Combining ultraconserved elements and mtDNA data to uncover lineage diversity in a Mexican highland frog (*Sarcohyla*; Hylidae). PeerJ 6:e6045.