TESIS DE DOCTORADO
PEDECIBA BIOLOGÍA

Genómica de los procesos de colonización en la región austral de Sudamérica

Mag. Facundo Giorello

Tutor: Dr. Enrique Lessa

Tribunal: Susana González, Arley Camargo, María Inés Fariello

Facultad de Ciencias, Universidad de la República
Montevideo, Uruguay

2019

ÍNDICE

RESUMEN

El ratón oliváceo, *A. olivacea*, es uno de los roedores cricétidos más ampliamente distribuido en la región austral de Sudamérica. Su distribución se extiende por casi todo Chile, la Patagonia Argentina y Tierra del Fuego, y habita una gran variedad de ambientes, desde desiertos costeros al norte, bosques Valdivianos y Magallánicos al sur y a lo largo de Chile, hasta la estepa Patagónica al este y sur de la Argentina, incluyendo Tierra del Fuego.

Su gran distribución y tolerancia a distintos ambientes vuelven a *A. olivacea* un excelente modelo para evaluar, por un lado, las distintas hipótesis sobre cómo los períodos glaciales afectaron las poblaciones patagónicas y por otro, para analizar su respuesta a los desafíos que le imponen los distintos ambientes, particularmente los más secos.

En base a datos mitocondriales, para *A. olivacea* así como para otras especies, los bosques valdivianos han sido señalados como los refugios "clásicos" durante los fases glaciales, sin embargo, para el ratón oliváceo también se ha propuesto un refugio sureño adicional en Tierra del Fuego. Se ha hallado que la especie presenta, además, una historia demográfica de expansión asociada al último período glacial.

En relación a su tolerancia a los ambientes xéricos, se ha descrito que posee una gran capacidad de concentración de orina comparable a varias especies adaptadas al desierto. Para las especies de desiertos diversos estudios moleculares ya han identificado genes interesantes, sin embargo los estudios que aborden la fisiología molecular para especies "generalistas" son escasos.

Para comprender la respuesta de las poblaciones patagónicas a los ciclos glaciales, analizar la hipótesis de un refugio adicional en Tierra del Fuego, y estudiar la respuesta molecular renal de la especie a los distintos ambientes, decidimos utilizar una aproximación a escala genómica para obtener inferencias más robustas que aquellas basadas en un solo locus. Para inferir el comportamiento de la especie en los períodos glaciales en la región Patagónica, genotipamos 172 individuos para más de 50.000 SNPs utilizando un protocolo de captura de exones. Para evaluar la respuesta molecular de la especie a los ambientes, analizamos el transcriptoma renal de 39 individuos procedentes de la estepa patagónica seca y de los bosques húmedos valdivianos y magallánicos.

Nuestros resultados indicaron que, en la región Patagónica-Fueguina, el ratón oliváceo se encuentra estructurada genéticamente en dos grandes grupos que corresponden, por un lado, a las poblaciones patagónicas al oeste de la cordillera, incluyendo Tierra del Fuego y, por otro, a las poblaciones del lado este de los Andes. Los resultados genético-poblacionales descartaron la posibilidad de un refugio en Tierra del Fuego mientras que confirman la existencia de un refugio en o cercano al refugio Valdiviano desde donde comenzó la colonización. Nuestros resultados sugirieron, además, un nuevo refugio glacial periférico hasta ahora solo reportado en plantas. Ambos grupos presentaron una expansión demográfica posterior al Último Máximo Glacial o asociada al fin de dicho ciclo.

Respecto a la fisiología molecular renal y su respuesta a los ambientes estepa y bosque, nuestros resultados indican que *A. olivacea* recurre a los genes clásicos relacionados con la conservación de agua para lidiar con el déficit hídrico. Nuestros análisis también indican un posible rol de la selección natural en estos sistemas, dado que genes claves, como los expresados diferencialmente, tienden a presentar una sutil pero mayor divergencia genética que los demás genes entre las poblaciones de estepa y bosque.

INTRODUCCIÓN

Factores determinantes para la Biogeografía de América del Sur.

América del Sur está compuesto por varias regiones biogeográficas, albergando la mayor biodiversidad del planeta y conteniendo 5 de los *hotspots* mundiales de biodiversidad (Myers 2003). Su extensión, entre las latitudes 12°N y 56°S, y sus características geomorfológicas (como la Cordillera de los Andes) hacen de América del Sur un continente con una gran variedad de climas que definen diferentes biomas y/o ecorregiones que contribuyen a albergar su biodiversidad (Morrone 2004; Morrone 2006; Aragón et al. 2011). Los procesos que han afectado la diversidad de especies y su distribución a lo largo del continente han sido varios y complejos (Rull 2011); sin embargo se pueden identificar 3 grandes factores que explican la biogeografía de la región: la formación del istmo de Panamá, hace 3 millones de años (MA), la formación de la Cordillera Andina, entre 22 MA y 4.5 MA antes del presente (AP), y las oscilaciones climáticas, principalmente del Cuaternario.

La formación del istmo de Panamá desencadenó el proceso conocido como “el Gran Intercambio -Biótico- Americano”, que resultó en un importante influjo de fauna desde Norteamérica. Por ejemplo, los carnívoros placentarios y herbívoros sudamericanos fueron reemplazados por los norteamericanos. Este intercambio terminó con aproximademente 40 MA de aislamiento relativo que habían experimentado, principalmente, los mamíferos en el continente. Previo a la formación del istmo de Panamá, América del Sur ya había sido receptora de roedores caviomorfos y primates desde África hace 42 a 30 MA (Leigh et al. 2014) y de roedores sigmondontinos hace 5 a 7 Ma (Marshall 1979) (aunque ver Leite et al. 2014 para tiempos más antiguos) desde Norteamérica. La formación del istmo de Panamá también afectó la biota marina de manera considerable. El istmo provocó la extinción de las surgencias marinas en el mar Caribe produciendo el declive de moluscos (que se alimentaban por filtración) y sus predadores asociados, que fueron reemplazados por arrecifes de coral (Leigh et al. 2014). Más en general, el impacto de istmo fue global, al dar lugar, por ejemplo, a la corriente del Golfo y volviendo más cálidas y húmedas las masas continentales asociadas al Atlántico Norte y cambiando la temperatura del mar, la salinidad y la sedimentación del carbón de manera diferencial entre el Caribe y el océano Pacífico (Leigh et al. 2014).

La formación de la cordillera de los Andes, por su lado, también tuvo un impacto significativo en el continente. Su gran extensión (alrededor de 7000km de longitud) y altitud (con puntos de casi 7000 metros de altura), en conjunto con su rápida formación (Hoorn et al. 2010; Folguera et al. 2011), propiciaron una gran cantidad de ambientes con condiciones climáticas contrastantes entre las tierras bajas y altas y de un lado y otro de la cordillera (Hoorn et al. 1995; Montgomery et al. 2001). Hacia el sur de la cordillera andina, por ejemplo, los contrastes de precipitación de un lado y otro de la cordillera constituyeron la estepa Patagónica al este (al menos desde el Cuaternario) y los bosques lluviosos Magallánicos y Valdivianos al oeste (Pizarro et al. 2012; Ruzzante & Rabassa 2011). Todo esto, sumado a la barrera geográfica que implican los Andes para la mayoría de las especies, ha propiciado a su vez una enorme biodiversidad mediante la especiación alopátrica (e.g Esquerré et al., 2019), así como por disminución del flujo génico transandino (e.g Trénel et al., 2008).

Por último, las oscilaciones climáticas del Cuaternario también tuvieron un gran impacto en la biogeografía al desencadenar los períodos glaciales e interglaciales. Típicamente, en los períodos glaciales grandes masas de hielo cubrían gran parte de la cordillera de los Andes desde los 37°S hasta Cabo de Hornos (56°S), mientras descubría gran parte de la plataforma continental del océano Atlántico en el lado este del continente (Ponce et al. 2011). Aunque las primeras glaciaciones Patagónicas se desarrollaron desde el final del Mioceno (6 MA), las cinco últimas glaciaciones ocurridas en el último millón de años fueron las que dieron lugar a los paisajes glaciales actuales de la región austral de América del Sur (Rabassa et al. 2011). De estos cinco glaciaciones las más importantes fueron la Gran Glaciación Patagónica (GGP), hace 1 MA, y el último máximo glacial (UMG). La GGP se caracterizó principalmente por su extensión (Ferrando 2002; Kaplan et al. 2009) mientras que el UMG por ser particularmente intensa hace aproximadamente 20 – 25 KA.

Las oscilaciones climáticas y sus ciclos glaciales asociados han servido de hipótesis para explicar la biodiversidad de América del Sur (Hooghiemstra & Van Der Hammen 1998). La “hipótesis de refugios” supone que la fragmentación ambiental producto de la extensión de los glaciares, al sur, y la fragmentación de selvas y bosques al norte de América del Sur, favoreció la divergencia y especiación de distintas especies al aislar en refugios a sus poblaciones. El papel de las oscilaciones climáticas y los cambios drásticos en la vegetación asociados a los ciclos glaciales e interglaciales son evidentes según reconstrucciones paleo-vegetales (Rull 1999; Iglesias et al. 2011; Quattrocchio et al. 2011).

Aunque esta hipótesis está hoy fuertemente discutida como explicación excluyente de la diversidad, y se considera una conjunción de procesos que no solo abarcan las oscilaciones climáticas sino también las tectónicas del Neógeno y la paleogeografía (Colinvaux et al. 2000; Rull 2011), es claro que tales oscilaciones han traído consigo cambios demográficos en varias especies (ver Turchetto-Zolet et al. 2013 y citas allí presentes). Típicamente la expansión de los glaciares desencadena una retracción poblacional, mientras que en los períodos interglaciales las poblaciones se expanden desde sus zonas de refugio.

De acuerdo a un meta-análisis en base a 214 artículos sobre la filogeografía de América del Sur, el 58 % de los estudios encontraron evidencia de cambios poblacionales en relación a los ciclos glaciares e interglaciares (Turchetto-Zolet et al. 2013). Más interesante aún, estos autores vieron que de aquellos que reportaron los tiempos de divergencia (un 38 %), la mitad precedía el UMG y la otra mitad lo procedía. También encontraron que la respuesta poblacional de las especies de bosque parecía ser más lineal que para las especies de ambientes xéricos y abiertos. Para las primeras, la asociación entre los períodos glaciales e interglaciares con los períodos de retracción y expansión poblacional, respectivamente, fue mucha más clara que para las segundas, donde el comportamiento parecía más errático o especie dependiente.

Por otro lado, dicho meta-análisis encontró que los tiempos de divergencia de los linajes de distintos filogrupos de Sudamérica se da en un 57% durante el Pleistoceno y el 43 % restante en el Mioceno/Plioceno aunque con diferencias según el grupo taxonómico. La formación de filogrupos intraespecíficos fue distinta para los mamíferos, aves, invertebrados y peces respecto a la herpetofauna (anfibios y reptiles); la divergencia de estos últimos parece haber comenzado mucho más temprano, durante el Plioceno/Mioceno.

En suma, la diversidad y distribución de las especies Sudamericanas fue moldeada tanto por eventos orogénicos y paleogeográficos (desde el Mioceno/Plioceno) como por las oscilaciones climáticas acontecidas en el cuaternario. Muchas especies habrían empezado a diversificarse en el Terciario (por eventos orogénicos) alcanzando la actual diversidad durante el Cuaternario cuando la estructuración poblacional y la formación de nuevas especies se originaron bajo los efectos de las oscilaciones climáticas (Rull 2011). Y aunque los estudios filogeográficos en América del Sur han aumentado considerablemente, todavía no está bien definido el rol de los ciclos glaciares y la generación de refugios en las regiones más australes de sudamérica, principalmente para las especies de ambientes abiertos (Turchetto-Zolet et al. 2013).

Métodos y análisis filogeográficos

Los datos moleculares, acoplados con la teoría de genética de poblaciones y abordajes filogenéticos, han demostrado su utilidad para entender cómo los procesos geológicos y climáticos han influido en la diversidad y distribución de una infinidad de especies (Avise 2001; Avise 2009). Particularmente, los estudios filogeográficos han provisto de información a la biogeografía sobre la biodiversidad, los modos de dispersión o colonización, tiempos de diversificación e identificación de zonas de refugio para múltiples especies. El marcador molecular por excelencia ha sido y es el ADN mitocondrial, seguido por el análisis de genes nucleares particulares (Morris & Shaw 2018; DeSalle et al. 2017). Sin embargo, con el advenimiento de las nuevas tecnologías de secuenciación masiva el análisis de miles de locus se ha vuelto posible (Shendure & Ji 2008). Para disponer o proveerse de múltiples loci es posible secuenciar el genoma completo de la especie de estudio, su transcriptoma (Wang et al. 2009), o utilizar distintas aproximaciones tales como *RADseq* (del inglés: *Restriction Site Associated DNA Sequecing*) (Hohenlohe et al. 2011) o captura de secuencias (Hodges et al. 2009), los cuales buscan reducir mediante un muestreo la representación del genoma para su posterior análisis, así como los costos asociados a la secuenciación. Brevemente, la familia de métodos entendidos como *RADseq* involucra la digestión de ADN genómico con una o más enzimas de restricción, la adición de *primers* específicos para la plataforma de secuenciación y la selección de fragmentos dentro de una determinada distribución de tamaño para su posterior secuenciación (Melville et al. 2017; Harvey et al. 2016) Por su lado, la captura de exones involucra la preparación de librerías a partir de fragmentos al azar de ADN genómico, la posterior hibridación de estas librerías con sondas oligonucleotídicas biotiniladas y la posterior captura, previo a la secuenciación, mediante *beads* (pequeñas bolitas) paramagnéticas cubiertas con estreptavidina que se utilizan para atraer las sondas. Estas sondas tienen en general entre 60 a 120 bases y las secuencias de éstas son complementarias a miles de regiones genómicas de interés seleccionadas de antemano (Harvey et al. 2016). Para el diseño de las sondas se pueden utilizar, además del propio genoma, secuencias transcriptómicas o genomas de otras especies (Bi et al. 2012).

Entre ambas técnicas, la de *RADseq* ha sido la más utilizado para aproximaciones filogeográficas, por su menor costo y principalmente porque no es necesario partir de ningún tipo de referencia (Harvey et al. 2016); luego de la secuenciación los *reads* se ensamblan *de novo* utilizando programas como *Stacks* (Catchen et al. 2013) y se realiza la detección de SNPs. Sin embargo, las principales limitaciones que adolecen los métodos *RADseq* son: (i) dificultad de establecer relaciones de ortología entre individuos y (ii) loci anónimos si no se cuenta con un genoma de referencia e incluso, en caso de contar con genomas

de especies relacionadas, no es trivial establecer la correspondencia. La captura de secuencias, por su naturaleza, no presenta estos problemas, lo que la vuelve una herramienta más interesante para análisis filogenéticos o poblaciones comparativos (Harvey et al. 2016). Como contrapartida, la captura está dirigida a regiones predeterminadas, y por tanto no produce un muestreo al azar de las regiones genómicas.

Con la aparición de estas nuevas aproximaciones moleculares han aparecido nuevas abordajes para su análisis. Para el caso de la inferencia demográfica, existen diversas aproximaciones que se pueden dividir en función del tipo de información genética que utilizan: regiones genómicas enteras o sitios independientes (SNPs, provenientes a su vez de loci independientes) (Schraiber & Akey 2015; Gao & Keinan 2016).

Las aproximaciones que trabajan con grandes regiones genómicas o directamente con el genoma entero capitalizan, principalmente, la información de ligamiento y el *phasing* (siempre que es posible). En particular, existen modelos basados en: (i) el Coalescente Markoviano Secuencial (SMC) siendo los programas *PSMC* (Li & Durbin 2011) y *MSCM2* (Schiffels & Durbin 2014) los más populares; (ii) el marco de distribución de muestreo condicional del Markoviano secuencial, como por ejemplo *diCAL* (Tataru et al. 2014) y *diCAL2* (Steinrücken et al. 2015); (iii) y aquellos que, utilizando la información de ligamiento, tratan de ajustar la distribución de ciertos estadísticos basados en el haplotipo (por ejemplo la distribución de los segmentos idénticos por descendencia o por estado), como *DoIRS* (Palamara & Pe'er 2013).

Por otro lado, los métodos que parten de sitios independientes hacen uso de la distribución de la frecuencia de las variantes o SFS (del inglés: *site frequency spectrum*; ver **Recuadro 1**), y buscan, en general, estimar las variables poblacionales (tamaño efectivo, momento de expansión, tiempo de bifurcación, tasa de crecimiento) que mejor ajusten a la SFS observada (Schraiber & Akey 2015). Estos modelos a su vez se pueden dividir en función de sí se basan en el coalescente (*e.g fastsimcoal2* (Excoffier & Foll 2011), *momi2* (Kamm et al. 2018), *stairway-plot* (Liu & Fu 2015), *fastneutrino* (Bhaskar et al. 2015) y *TNSFS* (Chen et al. 2015)) o en la teoría de difusión (*e.g dadi* (Gutenkunst et al. 2009) y *kimtree* (Gautier & Vitalis 2013)). Aunque estas aproximaciones basadas en el SFS se han utilizado en especies no modelos, su aplicación no es trivial para aquellas especies para las que no se cuenta con estimaciones de su tamaño poblacional, la tasa mutacional y la historia demográfica (en

particular los patrón de ramificación) (Schraiber & Akey 2015). Otra dificultad de estas aproximaciones basadas en el SFS es que son sensibles al tratamiento de los datos, en particular a filtrado de los SNPs (Harvey et al. 2016) y dependientes del sistema utilizado **(Recuadro 1).**

Recuadro 1

La distribución de la frecuencia de las variantes o SFS, es la proporción de alelos derivados que se encuentran en una frecuencia 1/2n, 2/2n,...(2n-1)/2n siendo n el número de individuos diploides analizados. En caso de no existir grupo externo se denomina *folded* SFS y se cuenta la proporción del alelo con menor frecuencia. Muchos estadísticos de genética de poblaciones (theta, *D* de Tajima, etc) pueden derivarse de esta distribución.

Las desviaciones que tiene el SFS, respecto a una población en equilibrio entre deriva y mutación neutral se utilizan para hacer inferencias demográficas y selectivas. Por ejemplo:

-*en un crecimiento demográfico reciente*: se evidenciará un exceso de alelos raros (a causa de una genealogía de ramas terminales largas donde "caerán" las mutaciones o influjo de mutaciones en los nuevos individuos nacidos). Desde el punto de vista de la prueba *D* de Tajima, van a aumentar proporcionalmente más los sitios segregantes (estrictamente theta K) que la heterocigosidad (considerada por theta pi) y por lo tanto *D* va a ser negativo (Holsinger 2012 p. 241).

-*en un cuello de botella reciente*: se evidenciará un pequeño deficit de *singletons* (variantes de frecuencia 1/2n) que desaparecen a medida que la población se recupera, un déficit más pronunciado de otros alelos raros y un exceso de alelos comunes (Gattepaille et al. 2013 pp 441). Generalmente la prueba *D* de Tajima será positiva (excepto si el cuello de botella barrió con todos los haplotipos, donde *D* sería negativo (Gattepaille et al. 2013)). La heterocigosidad (theta pi) depende más de los alelos comunes (o de alta frecuencia en la población) y media frecuencia que no se ven tan afectados. El número de sitios segregantes, en cambio, si ve verá mucho más afectado (Holsinger 2012 pp241).

-*barrido selectivo y selección purificadora*: también evidenciarán un exceso de alelos raros, aunque mucho más pronunciado en el caso de la selección purificadora, y para el caso del barrido selectivo habrá también un exceso de alelos/haplotipos comunes (los haplotipos con la variante ventajosa serán preferencialmente compartidos). El exceso de alelos raros bajo un régimen de selección purificadora se puede entender por la incapacidad de la selección de eliminar las nuevas variantes levemente deletéreas. Sin embargo, la selección si será lo suficientemente potente para evitar que estas variantes alcancen frecuencias comunes (Gibson 2012 pp 4).

Es importante tener en cuenta que, por ejemplo, bajo el marco de Campo aleatorio de Poisson las inferencias basadas en el SFS para algunos historias demográficas no son estadísticamente identificables (Schraiber & Akey 2015). Por otro lado, la captura de secuencias y el *RAD-seq* pueden tener un impacto diferente al momento de estimar la SFS. Se ha descrito que, al momento del filtrado los *reads*, establecer un umbral de cobertura menor aumenta proporcionalmente más los *singletons* en *RAD-seq* que en la captura (Harvey et al. 2016). Este efecto se puede deber a los sesgos en la cobertura (entre individuos para un mismo locus) en *RAD-seq*. En términos absolutos la captura presenta más *singletons* que los SNPs derivados de *RAD-seq* (y también valores *D* de Tajima negativos), principalmente si se tratan de regiones codificantes, al estar afectadas por la selección purificadora.

Las estimaciones de introgresión y tamaños poblacionales -ancestrales- también pueden variar si se utilizan SNPs derivados de una captura o de *RAD-seq*, y por lo tanto pueden sesgar las inferencias demográficas basadas en el SFS. Sesgos particulares de cada técnica pueden contribuir a estas diferencias: heterocigosidad sesgada por parálogos y secuencias conservadas insuficientemente dispersas en el genoma por el lado de captura, *allele dropout* en *RAD-seq*, más las diferencias ya descritas en relación a los *singletons* (Harvey et al. 2016).

Otra aproximación para estimar la historia demográfica, en conjunto con el árbol de especies y/o de poblaciones, a partir de SNPs independientes son aquellas basadas en el coalescente multiespecies, como el *SNAPP* (Bryant et al. 2012). Los modelos basados en el multicoalescente son aquellos donde se asume que el tamaño poblacional puede diferir entre las ramas del árbol de las especies (Rannala & Yang 2003). En particular, el *SNAPP* integra (calcula analíticamente) sobre todos los posibles árboles de genes (derivados de SNPs/AFLPs) evitando así su estimación. De esta manera puede estimar el árbol de las especies directamente a partir de los marcadores **(Recuadro 2)**. El *SNAPP* se ha utilizado tanto para estimar el árbol de especies como el de poblaciones (e.g Spalink et al. 2019) y no muestra gran variación si se utilizan SNPs derivados de una captura o de métodos RAD-seq (Harvey et al. 2016). Por otro lado, *SNAPP* es el único programa que, trabajando a partir de SNPs, también ha sido utilizado para calibrar los nodos del árbol utilizando información del registro fósil (Stange et al. 2018) y no a *posteriori* suponiendo una determinada tasa mutacional. Sin embargo, esta última aplicación tiene una fuerte limitante: asume, en general, un tamaño poblacional constante a lo largo de todo el árbol. De otro modo, el análisis de múltiples especies y/o poblaciones se vuelve inviable por los tiempos de corrida (Stange et al. 2018).

Recuadro 2

La aproximación filogenética basada en el multicoalescente busca principalmente resolver las incongruencias entre el árbol de especies y de genes causadas por "*lineage sorting*" (Mallo & Posada 2016). En general, los programas que construyen árboles de especies utilizan como *input*: topologías de los árboles de genes, los árboles con su longitud de rama (e.g BEST, Liu 2008) y últimamente directamente a partir de un alineamiento múltiple (e.g STAR-BEAST, Heled & Drummond 2010) (Yang 2014 pp340).

La probabilidad posterior del árbol de especies S y parametros Θ dado los datos X es:

$$f(S, \Theta|X) \propto f(S, \Theta) \times \prod_{i=1}^{L} \sum\nolimits_{G_i} \int_{\boldsymbol{t}_i} f(G_i, \boldsymbol{t}_i|S, \Theta) f(X_i|G_i, \boldsymbol{t}_i)\, \mathrm{d}\boldsymbol{t}_i,$$

Ecuación 9.49 pp 349 (Yang 2014) y/o ecuación 4 en (Bryant et al. 2012).

Donde Θ es un vector con los tamaños poblacionales de las especies y los tiempos de divergencia entre ellas; G_i árbol de genes $\boldsymbol{t}_i$ los tiempos de coalescencia de éste.

$f(S, \Theta)$: Es la densidad *a priori* (o *prior*) del árbol S con parámetros Θ.

$f(G_i, \boldsymbol{t}_i|S, \Theta)$: Es la densidad del coalescente multiespecie para el locus i (de un total L).

$f(X_i|G_i, \boldsymbol{t}_i)$: Es la función de verosimilitud de Felsentein.

Notar que para evaluar la verosimilitud se tiene que integrar sobre todo los posibles genes para cada uno de los loci, lo que en la práctica es imposible de resolver mediante una aproximación Bayesiana si se cuenta con cientos o miles de loci (como para el caso de los SNPs). Bryant et al. 2012 calculan esta integral analíticamente y evitando así las dificultades computacionales. De esta manera, estiman la verosimilitud del árbol de las especies dado la información genotípica proveniente de los SNPs. Por lo tanto, la MCMC muestrea sólo el árbol de especies, los tiempos de divergencia y los tamaños poblacionales.

Otro programa que resuelve el árbol de especies sorteando el aŕbol de genes es el *PoMo* (De Maio et al. 2015). No se basa en el multicoalescente, sino en las frecuencias alélicas, y la principal limitación es que requiere, ademas de los SNPs, todos los sitios invariantes (a diferencia del *SNAPP*) lo que dificulta su aplicación. Además de la realizar la detección de SNP tradicional (*SNP calling*) se requieren las *reference calling* (es decir, si el sitio analizado es efectivamente igual a la referencia o si no no hay suficiente información).

Otro método que puede acomodar historias de varias poblaciones más allá de los métodos basados en SFS o el coalescente es el Treemix (Pickrell & Pritchard 2012). Este programa utiliza la covarianza de las frecuencias alélicas entre poblaciones para capturar un árbol subyacente *ab initio* y para inferir eventos de introgresión.

Los análisis y aproximaciones filogeográficos que han buscado identificar el origen, y determinar las características y dirección de la colonización o expansión demográfica también han experimentado una evolución importante. Los primeros análisis, generalmente llevados a cabo en humanos, analizaban principalmente los patrones de heterocigosidad y de desequilibrio de ligamiento para los cuales se espera una reducción y un aumento, respectivamente, a medida que se analizan poblaciones más lejanas al origen de la expansión (Nielsen et al. 2017). Sin embargo, se ha descrito que ambos estadísticos, bajo ciertos escenarios, son incapaces de diferenciar entre patrones de colonización secuencial e instantáneo (DeGiorgio et al. 2009). Nuevos métodos han buscado aprovechar la clina de frecuencias de alelos derivados que se produce durante una expansión demográfica, fenómeno conocido como *allele surfing* (Excoffier & Ray 2008). Este fenómeno, descrito y observado en colonia de bacterias (Hallatschek et al. 2007), revela que el modelo típico de genética de poblaciones y sus predicciones son dependientes de la ubicación de los individuos cuando existe una expansión poblacional. En el borde migratorio o expansivo de la población los efectos de la deriva génica se amplifican (dado el número reducido de individuos) y por lo tanto también el destino de los alelos, que aceleran su marcha hacia la fijación o eliminación. Si consideramos el efecto de la mutación y la correspondiente formación de alelos derivados (es decir, nuevos) vamos a esperar que su frecuencia aumente en el borde migratorio y termine siendo más alto en las nuevas poblaciones que se hayan constituido si descontamos las variantes que han sido eliminadas. El mismo fenómeno se puede cuantificar haciendo uso de la frecuencia de alelos compartidos; las

poblaciones más lejanas al origen de la expansión tendrán una frecuencia de alelos compartidos mayor que las más cercanas (simplemente porque perdieron, se fijaron alelos). Esta forma de establecer direccionalidad entre pares de poblaciones está establecido en el índice de direccionalidad (Ψ) desarrollado por (Peter & Slatkin 2013). Para establecer el origen de la colonización a partir de la matriz Ψ conformada por cada par de poblaciones los autores utilizan un método de localización denominado Estimación Local de los Tiempos de Diferencia de Arribo (*TDOA*) (Aatique 1997).

RNA-seq y expresión diferencial

Las aproximaciones moleculares primero y la posterior llegada de las secuenciación masiva han permitido explorar y entender nuevas cascadas bioquímicas y génicas que le han aportado un nuevo nivel explicativo a la fisiología (Lemoine & Pradeu 2018). Unas de estas técnicas es el *RNA-seq* o secuenciación del ARN o del transcriptoma (Wang et al. 2009). Esta técnica ha permitido acceder a la secuencia de un gran número de genes, estudiar la estructura de los transcriptos, la información alélica, y la expresión génica (Wang et al. 2009). Últimamente, esta técnica se ha acoplado con la capacidad de secuenciar células particulares y descubrir así, por ejemplo, nuevas líneas celulares al interrogar la expresión génica de más de 50.000 células de un determinado tejido (e.g Park et al. 2018).

En el campo de ecología y evolución, conocer la fisiología molecular también es vital para entender, por ejemplo, qué procesos ecológicos pudieron haber desencadenado determinadas respuestas evolutivas (Pavey et al. 2010). La principal aproximación que ofrece la transcriptómica para entender cuál es el fenómeno biológico subyacente es la expresión diferencial entre dos condiciones biológicas, es decir, determinar qué genes presentan diferencias en su expresión a causa del tratamiento o condiciones (Todd et al. 2016).

Típicamente, el procedimiento bioinformático para llevar a cabo el análisis de expresión pretende "convertir" las millones de lecturas que se obtuvieron como resultado de la secuenciación masiva, en una medida de expresión. Este procedimiento se puede esquematizar en cuatro etapas (Oshlack et al. 2010). La primera consiste en mapear o alinear las lecturas a un transcriptoma o genoma. La idea de esta etapa es encontrar para cada lectura, su ubicación única (en caso de que corresponda) en la referencia. La segunda, consiste en agregar las lecturas en una unidad con sentido biológico, como exones, transcriptos o genes. Luego, se procede a estandarizar (o normalizar) la expresión de los genes para volverlos comparables.

Este procedimiento difiere según qué comparaciones se pretenden hacer; se puede comparar tanto los perfiles de expresión de genes entre muestras como dentro de cada muestra. Para la comparación entre muestras existen varias formas de normalizar los datos, entre ellas el *TMM* (Robinson et al. 2010) o el *DESeq size factors* (Anders et al. 2013). Estos métodos de normalización asumen que el nivel de expresión para la mayoría de los genes es igual entre las dos condiciones y tienen en cuenta también la distinta profundidad con la que fueron secuenciado las muestras.

Para modelar el conteo, es decir, el número de *reads* que se alinean a una referencia dada, se han utilizados distintas distribuciones (Poisson, Binomial negativa, Binomial Beta) y métodos no paramétricos. Los modelos paramétricos se utilizan, en general, dentro un modelo lineal generalizado (GLM) y se aplican distintos *tests* (exactos, razón de verosimilitud, etc) para definir si la expresión de gen difiere o no entre las dos condiciones o tratamientos (Qin et al. 2015). Estas distintas distribuciones tratan de capturar las principales fuentes de variación: el error de conteo de Poisson, la varianza técnica y la varianza biológica. El error de conteo de Poisson hace referencia a la incertidumbre relacionada a las medidas basadas en conteos. La incertidumbre es desproporcionalmente mayor para los conteos bajos y dominan la varianza para conteos menores que 10. Esto es consecuencia de que en la distribución de Poisson la media es igual a la varianza. En relación al rendimiento de las distintos métodos, varios trabajos comparativos han demostrado que aquellos basados en la binomial negativa, en general, son los que obtienen los mejores resultados (Rapaport et al. 2013; Soneson & Delorenzi 2013). Estos métodos, en conjunto con otros surgieron para acomodar la sobre-dispersión, es decir cuando la varianza entre réplicas biólogicas de un gen dado excede la media, que los métodos basados en la distribución de Poisson eran incapaces de explicar. Varios de las aproximaciones basadas en la binomial negativa (como el *Deseq2* y *EdgeR* (Robinson et al. 2010; Love et al. 2014)) pueden incorporar también covariables o corregir por efecto de lote mediante los modelos GLM.

La capacidad de la transcriptómica de investigar miles de loci y SNPs sin la necesidad de una referencia genómica, en conjunto con la capacidad de interrogar la expresión de los genes, ha hecho crecer considerablemente la presencia del *RNA-seq* en el campo de la ecología y evolución (Todd et al. 2016). Por ejemplo, haciendo uso de la información que ofrece el transcriptoma, varios trabajos han abordado el fenómeno de la especiación de varias especies no modelos analizando las secuencias codificantes, SNPs, así como la expresión génica (Wolf et al. 2010; Gagnaire et al. 2012; Pavey et al. 2010; Renaut et al. 2010). Otros trabajos han utilizado el transcriptoma para inferir la historia demográfica de especies no

modelos (McCoy et al. 2014) mediante la identificación cuidadosa de SNPs (Gayral et al. 2013) y haciendo uso de nuevos modelos y métodos poblacionales (Gutenkunst et al. 2009; Excoffier et al. 2013).

Antecedentes filogeográficos y ecológicos de *Abrothrix olivacea*

El ratón oliváceo, *Abrothrix olivacea* (Waterhouse 1837) es un roedor de la subfamilia Sigmodontinae (familia Cricetidae), una de las más grandes dentro de los mamíferos (86 géneros, y unas 400 especies) (D'Elía et al. 2007; D'Elía & Pardiñas 2015.). Se encuentra ampliamente distribuido en América del Sur, desde 18°S al 56°S de latitud, extendiéndose por casi todo Chile, la Patagonia Argentina y Tierra del Fuego, y habita una gran variedad de ambientes, desde desiertos costeros al norte, bosques Valdivianos y Magallánicos al sur y a lo largo de Chile, hasta la estepa Patagónica al este y sur de la Argentina, incluyendo Tierra del Fuego (Mann 1978). Al menos 6 subespecies se han identificado a lo largo de su distribución (Rodríguez-Serrano et al. 2006) y trabajos genético-poblacionales han señalado la existencia de al menos 3 filogrupos: en Chile central, Patagonia Continental y Tierra del Fuego (Lessa et al. 2010; Abud et al. 2011). Su gran distribución la vuelven una especie interesante para estudiar y caracterizar tanto el papel de las oscilaciones climáticas en las especies de ambientes abiertos como la fisiología molecular que permite a la especie habitar ambientes contrastantes.

Los primeros trabajos que abordaron la respuesta de *A. olivacea* a los ciclos glaciales propusieron refugios costeros en Valdivia durante el UMG, desde donde, posteriormente, ocurrirían expansiones hacia el oeste y al sur (Smith et al. 2001). Durante el UMG, cabe recordar que el lado oeste de la cordillera, lo que son hoy los bosques Valdivianos y Magallánicos, estaba prácticamente cubierto por hielo, mientras que el lado este se mantuvo libre de éste. Trabajos subsiguientes basados en citocromo b de cientos de individuos y con una cobertura espacial mayor, propusieron un refugio del lado este de la cordillera (latitud 42.47°S) dado sus niveles de diversidad nucleotídica y la falta de evidencia de expansión poblacional en las poblaciones de estepa de Patagonia Continental (Abud et al. 2011). La importante divergencia genética entre estas poblaciones y las correspondientes del lado chileno hizo descartar a los autores la hipótesis previa de un refugio costero, cercano a Valdivia, en Chile. Para las poblaciones de Tierra del Fuego, se hipotetizó un refugio propio, dada la divergencia entre ésta y el resto de las poblaciones continentales (3.2 %) y la evidencia de expansión poblacional que presentaban (Abud et al. 2011). Interesantemente, *A. olivacea* es la única especie de varias analizadas que presentaría dicho quiebre filogeográfico (Lessa et al. 2010) .

Las estimaciones de los tiempos de expansión poblacional para *A. olivacea* en general preceden el UMG (Abud et al. 2011; Lessa et al. 2010). Las estimaciones bayesianas para las poblaciones de Patagonia continental utilizando citocromo b son de 100.000 años, mientras la estimación utilizando modelos de expansión demográfica y la información de distribución de diferencias pareadas (mismatch distribution analysis) son más recientes (50.000 años) aunque también previos al UMG. La historia de expansión de las poblaciones de Tierra del Fuego es la más compatible con una expansión posterior al UMG (Abud et al. 2011; Lessa et al. 2010).

En relación a la respuesta y adaptaciones fisiológicas, tanto de *A. olivacea* así como de otros roedores sudamericanos a los distintos ambientes, nuevos estudios han descartado la idea tradicional de que los roedores sudamericanos, desérticos o de ambientes xéricos, carecen de adaptaciones para enfrentar la escasez de agua como las que presentan las especies norteamericanas (Bozinovic & Gallardo 2006). En particular, se han encontrado que varias especies presentan una gran capacidad de concentrar orina, una alta proporción relativa de médula corteza y bajas tasas de evaporación pulmocutánea que evitan la pérdida de agua por las vías respiratorias. Para el caso particular de *Abrothrix olivacea*, estudios fisiológicos han sugerido adaptaciones locales para las poblaciones que habitan ambientes xéricos (Bozinovic et al. 2011). Estudiando poblaciones de ambiente arbustivo semiárido (latitud 31° S) y poblaciones de bosque, Bozinovic et al. 2011 encontró que las primeras presentan una mayor capacidad de concentración de orina y una tasa evaporativa pulmo-cutánea menor. Además de estos sistemas fisiológicos, las poblaciones de *A. olivacea* de ambientes áridos o desérticos presenta una estructura renal comparable con varias especies de desierto (McNab & Brown 2002; Al-Kahtani et al. 2004; Cortés et al. 2000). A nivel molecular, estudios fisiológicos moleculares han encontrado variación en la expresión de los acuaporinas, proteínas de membrana relacionados al transporte de agua que se encuentran a lo largo de los túbulos de las nefrona, asociada a cambios estacionales, en varias especies de roedores sudamericanos (Gallardo et al. 2005; Bozinovic et al. 2003). La expresión de estas acuaporinas, principalmente acuaporina 2, fue mayor en los individuos durante las estaciones secas (Bozinovic et al. 2003). Las aproximaciones han sido principalmente inmunocitoquímicas y solo recientemente se ha caracterizado el transcriptoma renal de dos especies, entre ellas *A. olivacea* (Giorello et al. 2014) y *Abrothrix hirta* (Valdez et al. 2015).

OBJETIVOS

Objetivos generales:

1. Analizar, a nivel molecular, la respuesta fisiológica de *A. olivacea* para entender como ésta enfrenta ambientes con niveles de precipitación contrastantes, como la estepa y el bosque Patagónico.
2. Comprender cómo las poblaciones de la especie respondieron a los períodos glaciales del Cuaternario.

Objetivos específicos:

Para la especie *A. olivacea*:

1. Identificar SNPs a partir de transcriptomas y exones capturados.
2. Caracterizar la estructura genética de la especie en la región austral de Sudamérica a partir de SNPs.
3. Inferir magnitud, antigüedad y duración de las expansiones demográficas experimentadas por la especie a partir de SNPs.
4. Inferir origen, dirección de la colonización y posibles refugios glaciales.
5. Identificar genes expresados diferencialmente entre poblaciones de estepa y bosque.
6. Explorar si los genes expresados diferencialmente muestran evidencia de selección positiva con respecto a la distribución empírica del resto de los genes.

CAPÍTULOS

Capítulo 1.

Estudios moleculares recientes han encontrado diferencias sorprendentes entre las especies adaptadas al desierto y las especies que se han utilizado como modelo para estudiar los mecanismos para la conservación del agua, como el ratón doméstico. En particular, la acuaporina 4, un gen clásico involucrado en la regulación del agua en las especies modelo, está ausente o no se expresa en los riñones de las especies adaptadas al desierto. Para comprender mejor la respuesta molecular a la disponibilidad de agua, estudiamos el ratón oliváceo, *Abrothrix olivacea*, especie que presenta una tolerancia ecológica inusualmente amplia y ha mostrado una gran capacidad para concentrar orina. La especie habita la estepa patagónica como los bosques lluviosos valdivianos y magallánicos. Para investigar la respuesta de esta especie a las condiciones ambientales contrastantes, muestreamos 95 individuos de *A. olivacea* de cuatro localidades (dos en la estepa y dos en los bosques) y analizamos diversas variables fenotípicas en conjunto con el transcriptoma renal de 39 individuos. El tamaño relativo del riñón y el cociente concentración de orina y plasma se correlacionaron, como se esperaba, negativamente con la precipitación anual. Los análisis de expresión revelaron casi 3.000 genes que se expresaron de manera diferencial entre las muestras de estepa y bosque e indicaron que esta especie recurre a las cascadas génicas "clásicas" para la regulación del agua. La expresión diferencial entre los biomas también involucra genes que participan en las funciones inmunes y de desintoxicación. En general, los genes que se expresaron diferencialmente mostraron una ligera tendencia a ser más divergente y a mostrar un exceso de frecuencias de alelos intermedios, en relación a los loci restantes. Nuestros resultados indican que tanto la expresión diferencial como la variación alélica son importantes para la ocupación de ambientes contrastantes por parte del ratón oliváceo.

Received: 13 June 2017 | Revised: 30 May 2018 | Accepted: 31 May 2018

DOI: 10.1111/mec.14778

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*)

Facundo M. Giorello[1,2] | Matias Feijoo[1] | Guillermo D'Elía[3] | Daniel E. Naya[1] | Lourdes Valdez[3] | Juan C. Opazo[3] | Enrique P. Lessa[1]

[1]Departamento de Ecología y Evolución, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

[2]Espacio de Biología Vegetal del Noreste, Centro Universitario de Tacuarembó, Universidad de la República, Tacuarembó, Uruguay

[3]Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia, Chile

Correspondence
Facundo M. Giorello, Departamento de Ecología y Evolución, Facultad de Ciencias, Universidad de la República, Iguá 4225, Montevideo 11400, Uruguay.
Email: fagire@gmail.com

Funding information
Fondo Nacional de Desarrollo Científico y Tecnológico, Grant/Award Number: 1180366; Agencia Nacional de Investigación e Innovación, Grant/Award Number: ANII FCE 2014 103508

Abstract

Recent molecular studies have found striking differences between desert-adapted species and model mammals regarding water conservation. In particular, aquaporin 4, a classical gene involved in water regulation of model species, is absent or not expressed in the kidneys of desert-adapted species. To further understand the molecular response to water availability, we studied the Patagonian olive mouse *Abrothrix olivacea*, a species with an unusually broad ecological tolerance that exhibits a great urine concentration capability. The species is able to occupy both the arid Patagonian steppe and the Valdivian and Magellanic forests. We sampled 95 olive mouse specimens from four localities (two in the steppe and two in the forests) and analysed both phenotypic variables and transcriptomic data to investigate the response of this species to the contrasting environmental conditions. The relative size of the kidney and the ratio of urine to plasma concentrations were, as expected, negatively correlated with annual rainfall. Expression analyses uncovered nearly 3,000 genes that were differentially expressed between steppe and forest samples and indicated that this species resorts to the "classical" gene pathways for water regulation. Differential expression across biomes also involves genes that involved in immune and detoxification functions. Overall, genes that were differentially expressed showed a slight tendency to be more divergent and to display an excess of intermediate allele frequencies, relative to the remaining loci. Our results indicate that both differential expression in pathways involved in water conservation and geographical allelic variation are important in the occupation of contrasting habitats by the Patagonian olive mouse.

KEYWORDS
adaptation, detoxification, kidney trancriptomics, mammals, population genomics

1 | INTRODUCTION

To understand how adaptation occurs, evolutionary biologists often compare species at opposite ends of variation in a trait of interest. For instance, contrasts between desert animals and their mesic counterparts have uncovered many morphological and physiological characteristics that have evolved in response to arid environments. The most common adaptation involves the kidney and the enlarged loops of Henle of juxtamedullary nephrons, which consists of the descending and ascending limbs separated by the interstitial space of the renal medulla (Barrett, Kriz, Kaissling, & de Rouffignac, 1978; Mbassa, 1988). In brief, the production of concentrated urine is possible by the active transport of sodium chloride and other solutes, out of the tubular fluid in the thick ascending limb, making the interstitial fluid hyperosmotic. This creates an osmotic pressure, and water from the descending limb diffuses out to the interstitial fluid,

while the tubular fluid becomes hyperosmotic. The high water permeability and low solute permeability of the descending limb, opposite to those of the ascending limb, allow the dynamic of water and solutes between the loops of Henle and renal medulla. As new tubular fluid enters the descending tube, the countercurrent multiplication starts to work and an osmotic gradient starts to develop. The longer the loop of Henle, the greater the osmotic gradient and the size of the medulla. This, in turn, elicits more water reabsorption from the collecting ducts, especially in the presence of vasopressin that increases water permeability through aquaporins in collecting duct cells (reviewed in Fenton & Knepper, 2007; Sands & Layton, 2014). However, the correlation between an ability to concentrate urine and the length of the loops of Henle is not straightforward (Diaz, Ojeda, & Rezende, 2006; McNab & Brown, 2002). Other strategies to limit dehydration include reduction of water evaporative loss (MacMillen & Lee, 1967; Withers, Louw, & Henschel, 1980), increase of metabolic water production (Macmillen & Hinds, 1983) and behavioural strategies, such as nocturnal activity or microhabitat selection (Degen, Kam, Hazan, & Nagy, 1986; Macmillen & Hinds, 1983).

Molecular studies have already identified interesting genes putatively under selection (MacManes & Eisen, 2014; Marra, Romero, & Dewoody, 2014), highly expressed (Marra, Eo, Hale, Waser, & Dewoody, 2012) or differentially expressed in desert-adapted species (MacManes, 2017; Marra et al., 2014). Recent studies on desert species have found interesting differences in key genes relative to house mouse and rat. In particular, aquaporin 4, a member of the vasopressin-regulated water reabsorption pathway, is apparently absent in desert species such as kangaroo rats (Huang et al., 2001) and camelids (Wu et al., 2014). Aquaporins (*AQP*) are a gene family of integral membrane proteins, which facilitate transport of water between cells (Heymann & Engel, 1999). In terms of the effects on water reabsorption, knockout experiments in the house mouse have shown *AQP2* to be the most important member of the family, followed by *AQP1* and *AQP4*. Mice lacking *AQP2* suffer severe polyuria and are totally unable to concentrate urine, but those who lack *AQP1* and *AQP4* also presented defects in water reabsorption (Fenton & Knepper, 2007). AQP2 and AQP4 are located in the apical and basolateral membrane of the collecting duct cell (Fenton & Knepper, 2007), respectively. Another study carried out in the desert-adapted cactus mouse, *Peromyscus eremicus*, also obtained interesting results regarding *AQP4*. In this work, a control and a water deprivation group were analysed after 72 hr of treatment, and unexpectedly, this protein was significantly overexpressed in the control group instead in the water-deprived group (MacManes, 2017). Typically, water restriction leads to an overexpression of mRNA of *AQP4* in model species (Murillo-Carretero, Ilundáin, & Echevarria, 1999).

In sum, it appears that desert specialists may have dropped or significantly reduced expression of *AQP4*, a key gene for water reabsorption in the laboratory house mouse. Species that are capable of tolerating a wide range of water regimes (including arid conditions) provide significant opportunities to improve our understanding of the adaptation processes: Do they resort to the classical pathways uncovered by knockout and other laboratory experiments in model laboratory species or are there more like desert animals, which do not utilize key genes such as *AQP4*? Advances in transcriptomics allow a systems approach that includes, but is not restricted to particular genes of interest (e.g., MacManes, 2017; Marra et al., 2014).

In this work, we use phenotypic and transcriptomic data to study the Patagonian olive mouse, *Abrothrix olivacea*, a species with an unusually broad ecological tolerance, in selected localities that include both temperate rainforests and arid Patagonian steppe. The olive mouse is commonly found in Chile from the northern dry lands to the Valdivian and Magellanic temperate rainforests and is also common in the Patagonian steppe and Fuegian grasslands (Patton, Pardiñas, & D'Elía, 2015). The species possesses a great urine concentration capability and a relative medullary thickness comparable to several desert-adapted species (Al-Kahtani, Zuleta, Caviedes-Vidal, & Garland, 2004; McNab & Brown, 2002). We examine variation in both gene expression and allele frequencies in two pairs of forest and steppe populations of the Patagonian olive mouse at two different latitudes (Figure 1). In southern South America, the dry Patagonian steppe and Valdivian and Magellanic humid forests are found side by side, from 40° to 50°S. Mean annual precipitation ranges from 160 to 2,000 mm among our four sampling sites, but mean annual temperatures only range from 6 to 10°C. Water availability, thus, is a major component of ecological variation in this system. We sampled 95 Patagonian olive mouse specimens from four localities (Figure 1) to obtain urine plasma concentrations and 26 kidney samples (which were combined with 13 samples reported by Giorello et al. (2014) for deep coverage individual transcriptome sequencing and associated sequences). Using measures of differential gene expression, population structure (F_{ST}) and neutrality tests (Tajima's D), we found that (a) overall, differential expression distinguishes samples from the forest and the steppe; (b) differentially expression analyses point out that this species resorts to the "classical" gene pathways for water regulation and the expression of some key genes showed a different expression pattern of those found in desert mammals; and (c) genes that were differentially expressed showed a slight tendency to be more divergent and an excess of intermediate allele frequencies, relative to the remaining loci. We also found that at least two additional systems, namely those of immune response and detoxification, also varied dramatically in gene expression across environments, indicating that multiple environmental factors vary across biomes and affect gene expression in the kidney.

2 | MATERIALS AND METHODS

2.1 | Data collection and RNA extraction

Adult individuals of *A. olivacea* were captured with Rodent Trap Special-S traps (Forma Ltda., Santiago, Chile) in the following localities: Río Oro, Provincia de Santa Cruz: n = 33; Gan Gan, Provincia de Chubut: n = 18, Argentina; Sector Barrancoso, Región de Aysén: n = 23; Fundo San Martín, Región de Los Ríos: n = 20, Chile. After

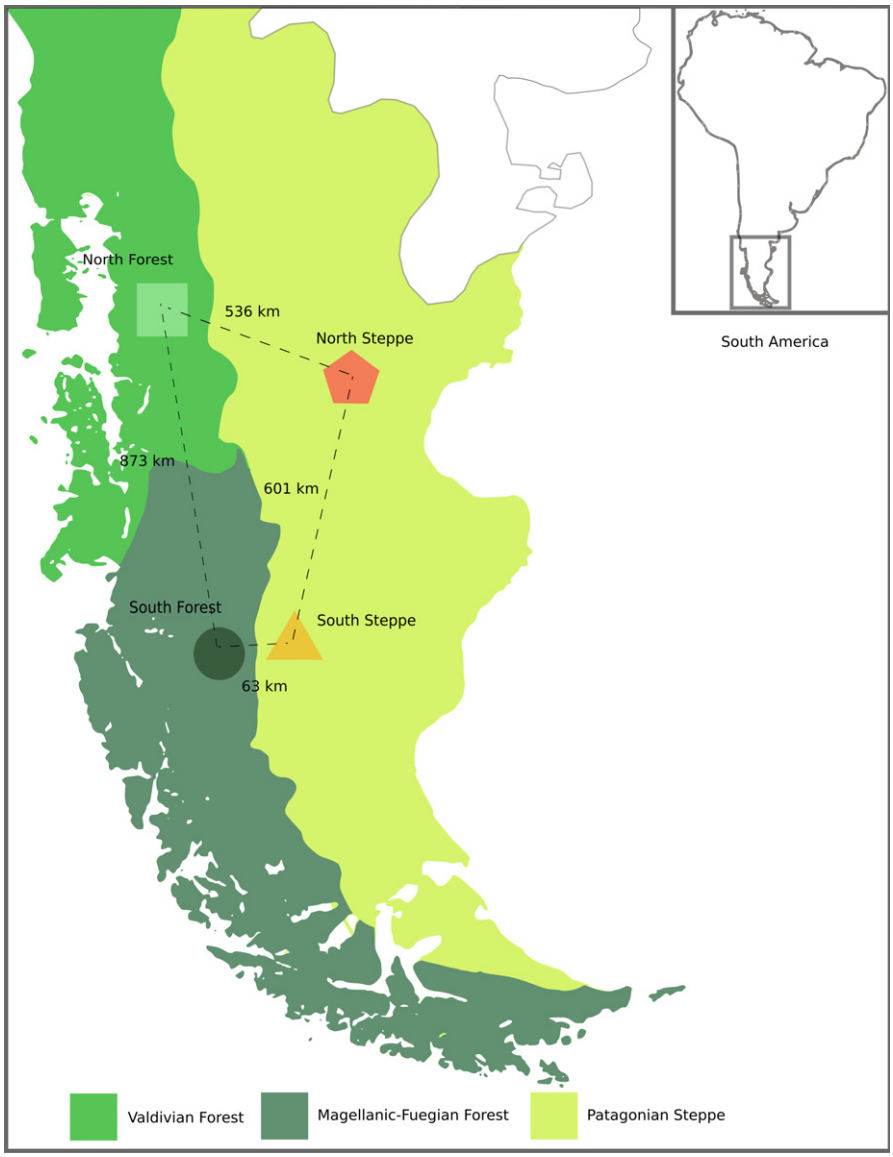

FIGURE 1 Map of southern South America and sampling localities. Map of southern South America showing the four localities where specimens of the Patagonian olive mouse *Abrothrix olivacea* were sampled; linear distances among sites are included

collection, animals were weighed (±0.1 g) and euthanized following guidelines of the American Society of Mammalogists, by properly trained personnel and outside the perceptive range of the other captive individuals (Sikes & Gannon, 2011). Then, animals were ventrally dissected, and samples of urine (from the urinary bladder) and blood (from the retro-orbital sinus) were obtained with a syringe. The right kidney of each individual was carefully dissected, weighted (±0.001) and stored in liquid nitrogen. We used a digital refractometer (Reichert r^2) to obtain temperature-corrected values of total solute concentration in urine (U) and blood (P) samples. Then, we calculated the [U/P] coefficient for each individual by dividing these two concentration values. Given that not all rodents had urine in the urinary bladder, we obtained [U/P] values only for a subsample of the individuals collected at each locality (Río Oro: n = 10, Gan Gan: n = 10, Sector Barrancoso: n = 15, Fundo San Martín: n = 13). RNA extractions were conducted in one half of a kidney (longitudinally sectioned) of each of 26 specimens as follows: Fundo San Martín: n = 6, Sector Barrancos: n = 7, Gan Gan: n = 7, Río Oro: n = 6 (further details in Supporting Information Table S1).

2.2 | Library construction, sequencing and trimming

The RNeasy mini kit (Qiagen) was employed for RNA extraction following recommendations of the manufacturer. RNA quantity and purity were assessed with NanoDrop 1000 Technologies spectrophotometer. RNA integrity was checked through electrophoresis in formaldehyde/agarose 1.2% denaturing gels. Libraries were constructed and sequenced commercially by Macrogen (Korea). Poly-A-based mRNA enrichment and paired-end library preparation were performed following the Illumina TruSeqTM RNA sample preparation kit, according to the instructions of the manufacturer. Library sequencing was performed on an Illumina HiSeq 2000 platform (2 × 100 bp paired-end mode). After adapter and primer filtering, 3′ end low-quality bases ($Q < 24$) were removed using Trimmomatic (Bolger, Lohse, & Usadel, 2014) before the SNP analysis. For

transcriptome assembly, reads were untrimmed or softy trimmed following (Q < 2) MacManes (2014). Quality of the reads was checked with FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The final number of reads prior to transcriptome assembly and SNP analysis is shown in Supporting Information Table S1.

2.3 | Transcriptome assembly

Assembly of the 26 new libraries was performed using Trinity with default settings (Haas et al., 2013). A reference transcriptome was constructed using the transcriptome assembly of each of the 26 samples and 13 previously published transcriptomic samples and multiread assembly (an assembly constructed joining the reads of 13 libraries) of Giorello et al. (2014), using custom Python scripts. For the construction of the reference transcriptome, we select the best hit between the contigs, among the 39 assembled transcriptomes and among the OMA browser mouse proteins (September 2014 database) (Altenhoff, Schneider, Gonnet, & Dessimoz, 2011). The alignment was carried out using BLASTX (*e*-value cut-offs < 1*e*−10).

2.4 | Differential gene expression analysis

One of the 26 individuals (PPA450) was conservatively removed from subsequent analyses because its locality of origin was mislabelled; the remaining 25 were combined with 13 samples from Giorello et al. (2014) for a net total of 38 individuals. Reads were aligned against the reference transcriptome using Bowtie2 in paired-end mode, implemented in RSEM 1.2.19 (Li & Dewey, 2011). RSEM handles reads that map to multiple transcripts and estimates the read counts for each gene, as well as for each transcript. Differential gene expression was examined with DESEQ2 (Love, Huber, & Anders, 2014) using the expected counts of each gene. We tested for differences in transcript abundance using a generalized linear model (GLM). A single main effect (biome, contrasting forest and steppe) was included in the model. Genes with adjusted (Benjamini & Hochberg, 1995) *p*-value < 0.05 were considered to be differentially expressed. Body mass has a direct effect on water regulation; thus, we analysed the correlation between gene expression and body mass within each of four localities. We did not treat body mass as a covariate in the GLM model because body mass itself varies with the treatment (see Section 3). Those genes which proved to be significantly correlated with body mass in any of the four analyses were disregarded (Supporting Information Table S2).

To determine gene ontology categories and pathways of the differentially expressed genes, we employed the bioinformatics database DAVID (Dennis et al., 2003), using Fisher's exact test ($p < 0.05$) as a significance test to evaluate enrichment. We further checked whether the differentially expressed genes were enriched in genes related to "classical" water conservation pathways or, rather, in genes related to water and sodium balance found in desert mammals. To this end, we constructed an a priori gene list of 133 "classical" genes (Supporting Information Table S3) including (i) those reported in the work of Fenton and Knepper (2007); this study lists mouse putative genes related to water conservation according, mainly, to gene knockout experiments; and (ii) all the genes associated with the "classical" pathways related to water reabsorption: the renin–angiotensin system (Sparks, Crowley, Gurley, Mirotsou, & Coffman, 2014), aldosterone-regulated sodium reabsorption (Rozansky, 2006) and vasopressin-regulated water reabsorption (Boone & Deen, 2008). On the other hand, we considered five solute carriers (*SLC2a9*, *SLC12a2*, *SLC12a3* and *SLC8a1*) to be related to water conservation in desert animals (MacManes, 2017; MacManes & Eisen, 2014; Marra et al., 2012). Finally, we checked the potential differential expression of inositol polyphosphate 5-phosphatase K (*INPP5K*), arginine vasopressin receptor 2 (*AVPR2*), sodium channel epithelial 1 alpha subunit (*SCNN1A*), angiotensinogen (*AGT*), aquaporin 4 (*AQP4*) and renin (*REN*), as they belong to the "classical" pathways and were also reported to be differentially expressed in desert animals in response to water deprivation (MacManes, 2017). The probability of observing an equal or higher given number of a priori genes was calculated assuming a binominal distribution.

2.5 | Gene coexpression network analysis and module detection

Clustering genes whose expression is highly correlated can reveal groups of interacting genes belonging to a particular gene pathway. WGCNA (version 1.61; Langfelder & Horvath, 2008) has been commonly used to detect such clusters of genes, identify modules (clusters of densely interconnected and highly correlated genes) and relate them with particular phenotypic variables. To analyse our transcriptomic data through WGCNA, we previously transformed our raw counts with *varianceStabilizingTransformation* DESEQ2's function to produce approximately homoscedastic expression values. For the module identification, we choose a soft thresholding power of $\beta = 16$, to ensure an approximately scale-free topology network. We then generated a clustering tree requiring a minimum 30 genes per module, and a threshold of cluster dissimilarity of 0.20 was chosen for merging the modules found. The module eigengenes [as defined by Langfelder and Horvath (2008)] was used to summarize the expression level of each module, and these values were then used to obtain their correlation (Pearson) with the [U/P] variables.

2.6 | Population genetics analysis

SNP calling and filtering were carried out using GATK version 3.5-0-g36282e4 (McKenna et al., 2010). First, reads were aligned against the reference transcriptome using BWA 0.7.12 (Li & Durbin, 2009). Reads were sorted and converted to bam, and duplicates removed using PICARD TOOLS 2.1.0 (https://broadinstitute.github.io/picard/). SNP calling for each sample was implemented through GATK's HaplotypeCaller using the 38 samples (-stand_call_conf 20.0 -stand_emit_conf 20.0). The resulting vcf file was then filtered using Variant Filtration (-window 35 -cluster 3 -filterName FS -filter "FS > 30.0" -filterName QD -filter "QD < 2.0") following GATK recommendations for RNA-seq data.

Population genetic analysis was performed using VCFTOOLS (Danecek et al., 2011). We evaluated genetic differentiation through the F_{ST} fixation index and different selection regimens using Tajima's D statistic. Only SNP sites without missing data were considered (–max-missing 1). For both statistics, the chosen bin size was larger than the max contig length; this way, we were able to obtain the statistics for each contig (and not for a given window size). We used the Kolmogorov–Smirnov (K-S) nonparametric test to evaluate whether the distributions of F_{ST} or D were different in differentially expressed genes relative to the remaining expressed genes. To this end, we only considered the differentially expressed genes with an average expression (baseMean column of DESEQ2 output) higher than 50. Imposing this threshold effectively eliminated a slight, but significant correlation between F_{ST} and expression values ($\tau = 0.05$, p-value < 0.01). To evaluate the false-positive rate of the Kolmogorov–Smirnov (K-S) test when analysing Tajima's D/F_{ST} distributions, we ran 1,000 tests taking random genes and evaluating them with the respective background gene resampled distribution. We found that the rate of positive results was equal to the p-value chosen for the K-S test, suggesting that this test can be safely applied to our data.

To estimate individual ancestries, the vcf file was first converted to a ped file, using VCFTOOLS and PLINK 1.9 (Chang et al., 2015), and then, the analysis was carried out using ADMIXTURE (Alexander, Novembre, & Lange, 2009) with $K = 4$ (four populations). Individual ancestry was calculated using one randomly selected SNP per gene to avoid skewing the analysis towards the longer genes.

3 | RESULTS

3.1 | Reference kidney transcriptome

The Illumina sequencing of the new 26 samples yielded a total of ~136 Gb of data. After assembly, the number of contigs, median contig length and transcriptome size ranged from 56,162 to 170,091, 300 to 527 bp and ~40 to 131 Mb, respectively (Supporting Information Table S4). The total number of genes annotated per individual ranged from 12,150 to 13,786 (Supporting Information Table S5). Using the newly 26 sequenced samples plus the 13 transcriptomes from Giorello et al. (2014), a reference kidney transcriptome for *A. olivacea* that included 17,397 genes with a total size of ~49 Mb was constructed (Table 1). This new reference transcriptome was improved (in terms of number of genes, percentage of CDS reconstructed and median contig length) to the one reported by Giorello et al. (2014).

TABLE 1 Main assembly metrics of the reference kidney transcriptome of the Patagonian olive mouse *Abrothrix olivacea*

Total contigs	19,587
Max contig length	24,023
Average length	2,491
Median length	1,787
Total base pairs	48,799,735
Genes identified	17,397
CDS reconstructed at >50% (with reference to the house mouse orthologs)	16,448

Note. CDS: coding sequences.

3.2 | Phenotypic variables

Body mass differed among local samples, with animals from the steppe being smaller than animals from the forests; in particular, animals from the northern steppe were markedly smaller (65% to 75%) than animals from the other three populations (Figure 2a). Relative kidney size (i.e., the residual values with respect to body mass) and urine-to-plasma concentration ratio [U/P] were, as expected, negatively correlated with annual rainfall (Figure 2b,c).

3.3 | Differential gene expression and admixture

Principal component analysis (PCA) based on expression profiles separated steppe and forest samples; the first principal component explains 25% of the variance. Within each biome, there is not a clear-cut distinction between northern and southern samples (Figure 3).

Estimated individual ancestries (Supporting Information Figure S1) were almost in complete agreement with the biome of origin of each individual and with the PCA results. The samples from the steppe were assigned to the same population, whereas the analyses suggested subdivision within the northern forest sample. Finally, the same southern forest samples that showed a mixed ancestry (50% red and green) are those that in the PCA appear closer to the steppe individuals.

Following the PCA results (Figure 3), we emphasize comparisons between steppe and forest for the differential expression analysis.

After discarding 86 genes whose differential expression could be explained by body size (Supporting Information Table S2), comparison between forest and steppe samples shows 2,881 significantly differentially expressed genes; of these, 1,383 and 1,498 gene transcripts were more abundant in the steppe and forest, respectively (Supporting Information Table S6). Gene transcripts overabundant in the forest were enriched in terms related to immune system processes (biological process), MHC class II protein complex (cellular component) and drug metabolism, in conjunction with other pathways associated with different diseases (KEGG pathways; Supporting Information Table S7).

Several immune systems and detoxification genes were found to be differentially expressed between biomes, including glutathione S-transferase, alpha 4 (*Gsta4*), which involved in the conjugation phase of drug metabolism (Sherratt & Hayes, 2002), ATP-binding cassette, subfamily C, member 2 (*Abcc2*), which involved in the excretion of xenobiotics (Glavinas, Krajcsi, Cserepes, & Sarkadi, 2004), and interleukin 18 (*IL-18*), a proinflammatory cytokine.

On the other hand, the functional analysis of the gene transcripts overabundant in the steppe pointed to, among other terms, the

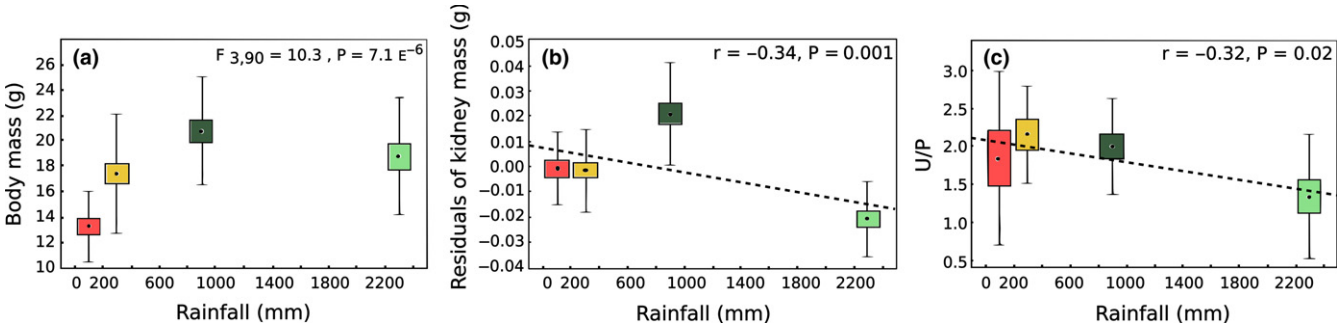

FIGURE 2 Morphological and physiological analyses of *Abrothrix olivacea*. (a) *F*-test for body mass variation between forest and steppe ($F_{3,90}$ = 10.3; *p*-value: 7.1 e^{-6}). (b) Correlation between residual of kidney mass and rainfall. (c) Correlation between urine-to-plasma concentration ratio [U/P] and rainfall. Samples: first box: north steppe; second box: south steppe; third box: south forest; and fourth box: north forest

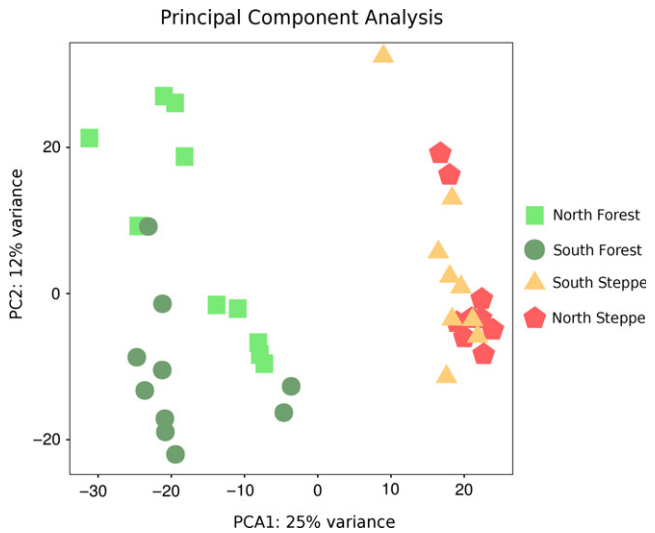

FIGURE 3 Principal component analysis of expression values for 38 individuals

sodium ion transmembrane transport (biological process), which is associated with water absorption processes (Supporting Information Table S8). Interestingly, the renin–angiotensin system pathway, which is involved in the regulation of plasma sodium concentration, was associated with gene transcripts significantly overabundant in the forest cluster (Supporting Information Table S7).

Several important "classical" genes involved in water conservation identified in laboratory studies were differentially expressed. Our transcriptome recovered 123 of 133 genes presumably related to water conservation (see Section 2). Of these, 18 are overabundant in the steppe (more than what is expected by chance, $p < 0.01$; Table 2; Figure 4). Among these genes, we found urea transporter 2, key to urine concentration, aquaporin 4, which expresses in the basolateral membrane of the collecting duct, angiotensinogen, the precursor of angiotensin I that ultimately triggers sodium and water reabsorption, and *SCNN1A*, which encodes for the α-subunit of the epithelial sodium channel ENaC. Meanwhile, 11 gene transcripts were overabundant in the forest, but they did not represent an enrichment of the a priori genes (Supporting Information Table S9). On the other hand, of those genes reported to be overexpressed in desert species (see Section 2), we found angiotensinogen (*AGT*) and solute carrier family 8 member A1 (*SLC8A1*) to be significantly overabundant in the steppe specimens, whereas renin (*REN*) was significantly overabundant in the forest individuals. The remaining genes reported to be relevant for water regulation in desert species were not significant.

Our differential expression results also uncovered genes presumably important to water conservation, which were not present in the a priori list. For example, we found 90 solute carriers (Supporting Information Table S6), in addition to the urea transporter, which are candidates for future studies. We also found that the lysine-deficient protein kinase 4, which is involved in water retention and potassium wasting through the regulation of sodium chloride symporter (San-Cristobal, De Los, Ponce-Coria, Moreno, & Gamba, 2008), is more abundant in transcriptomes of individuals from the steppe.

3.4 | WGCNA and correlation with [U/P]

The WGCNA analysis of our transcriptome resulted in 36 modules, each including between 31 and 2,951 genes (Supporting Information Table S10). The most interesting modules were 10, 19 and 36 as they presented many significant GO terms (biological process and KEGG pathways) previously found in the differential expression analysis: Module 10 presented GO terms strongly associated with immune response; 19 was enriched in xenobiotic metabolism related to GO terms; and 36 grouped two of the main pathways related to water and sodium balance: aldosterone-regulated sodium reabsorption and vasopressin water reabsorption. Another interesting module was 21, which included genes related to "renin secretion," a hormone that regulates blood pressure and fluid balance.

Of the 36 modules, two were shown to be negatively correlated with the [U/P] data, 33 and 31 (Supporting Information Figure S2). None of these modules presented clear GO terms that could suggest a link to water reabsorption, except for very punctual terms, such as "bradykinin catabolic process" and "regulation of systemic arterial blood pressure by renin–angiotensin" in the case of 33 (Supporting Information Table S11). Bradykinin is a vasodilator, part of the kinin–

TABLE 2 Population genetics and expression data for the 18 genes from the a priori gene list that are overexpressed in the transcriptomes of the steppe cluster of *Abrothrix olivacea*

Ensembl gene code	WikiGene name	WikiGene description	Tajima's D steppe	Tajima's D forest	F_{ST}	logFC	BaseMean
ENSMUSG00000022994	*Adcy6*	Adenylate cyclase 6	−1.5	−1.08	0.21	−0.44	1,361
ENSMUSG00000004929	*Thop1*	Thimet oligopeptidase 1	−2	−0.66	0.14	−0.53	173
ENSMUSG00000059336	*Slc14a1*	Solute carrier family 14 (urea transporter), member 1	−0.91	−0.1	0.15	−1	701
ENSMUSG00000040016	*Ptger3*	Prostaglandin E receptor 3 (subtype EP3)	−0.43	−0.93	0.1	−0.54	325
ENSMUSG00000019970	*Sgk1*	Serum/glucocorticoid-regulated kinase 1	0.88	−0.37	0.14	−1.55	5,199
ENSMUSG00000035451	*Foxa1*	Forkhead box A1	NA	NA	NA	−1.03	19
ENSMUSG00000021367	*Edn1*	Endothelin-1	−1.13	−1.49	−0.01	−0.74	33
ENSMUSG00000031980	*Agt*	Angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	−1.78	−0.47	0.14	−1.12	4,338
ENSMUSG00000002504	*Slc9a3r2*	Solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 2	NA	−0.95	0.04	−0.42	1,039
ENSMUSG00000035770	*Dync1li2*	Dynein, cytoplasmic 1 light intermediate chain 2	−1.28	−1.2	0.13	−0.77	2,907
ENSMUSG00000030340	*Scnn1a*	Sodium channel, nonvoltage-gated 1 alpha	−0.55	−1.04	0.23	−0.51	5,086
ENSMUSG00000024411	*Aqp4*	Aquaporin 4	−1.32	−0.29	0.15	−0.78	746
ENSMUSG00000041248	*Kcnj1*	Potassium inwardly rectifying channel, subfamily J, member 1	NA	−1.12	0	−0.44	4,223
ENSMUSG00000046971	*Pla2g4f*	Phospholipase A2, group IVF	NA	NA	NA	−1.06	19
ENSMUSG00000050211	*Pla2g4e*	Phospholipase A2, group IVE	NA	NA	NA	−1.13	29
ENSMUSG00000007097	*Atp1a2*	ATPase, Na^+/K^+ transporting, alpha 2 polypeptide	0.7	−0.75	0.09	−1.07	359
ENSMUSG00000000711	*Rab5b*	RAB5B, member RAS oncogene family	NA	NA	NA	−0.68	2,563
ENSMUSG00000031891	*Hsd11b2*	Hydroxysteroid 11-beta dehydrogenase 2	NA	NA	NA	−1.44	2,261

Note. logFC indicates the fold change in base 2; negative values indicate overexpression at the steppe cluster. baseMean is the mean of normalized counts of all samples, normalizing for sequencing depth. All *p*-adjusted values (Benjamini–Hochberg) were ≤0.02. Not available (NA). Tajima's D corresponds to loci that are monomorphic in at least one biome.

kallikrein system, which is involved in the regulation of blood pressure (Golias, Charalabopoulos, Stagikas, Charalabopoulos, & Batistatou, 2007).

3.5 | Genetic divergence and evidence of selection

Fixation index (F_{ST}) values were estimated between forest and steppe samples, as well as among pairs of populations, using 182,832 SNPs (one SNP every 267 bp) identified and unequivocally genotyped from the transcriptome in all individuals of *A. olivacea*. Average pairwise estimates of F_{ST} varied from very minimal divergence between the northern steppe and southern steppe samples (601 km, F_{ST} = 0.04) to moderate divergence between the northern steppe and southern forest samples (F_{ST} = 0.27, Figure 5a). The southern forest and steppe samples, which are separated by only 63 km, also show moderate levels of divergence (F_{ST} = 0.28, Figure 5a).

As expected, F_{ST} values of individual genes vary broadly in comparisons between forest and steppe for both differentially expressed and baseline genes (Figure 5b). However, the distribution is right-shifted in differentially expressed genes relative to the remaining genes (Figure 5b). This shift is also evident, although not always statistically significant, considering separately the southern and northern pairs (Supporting Information Figure S3). The average F_{ST} value between the steppe and forest clusters was 0.119, whereas the average F_{ST} value of differentially expressed genes was 0.125. Genes that are differentially expressed appear, on average, to be slightly more likely to differentiate across environments than those that not differ in expression levels.

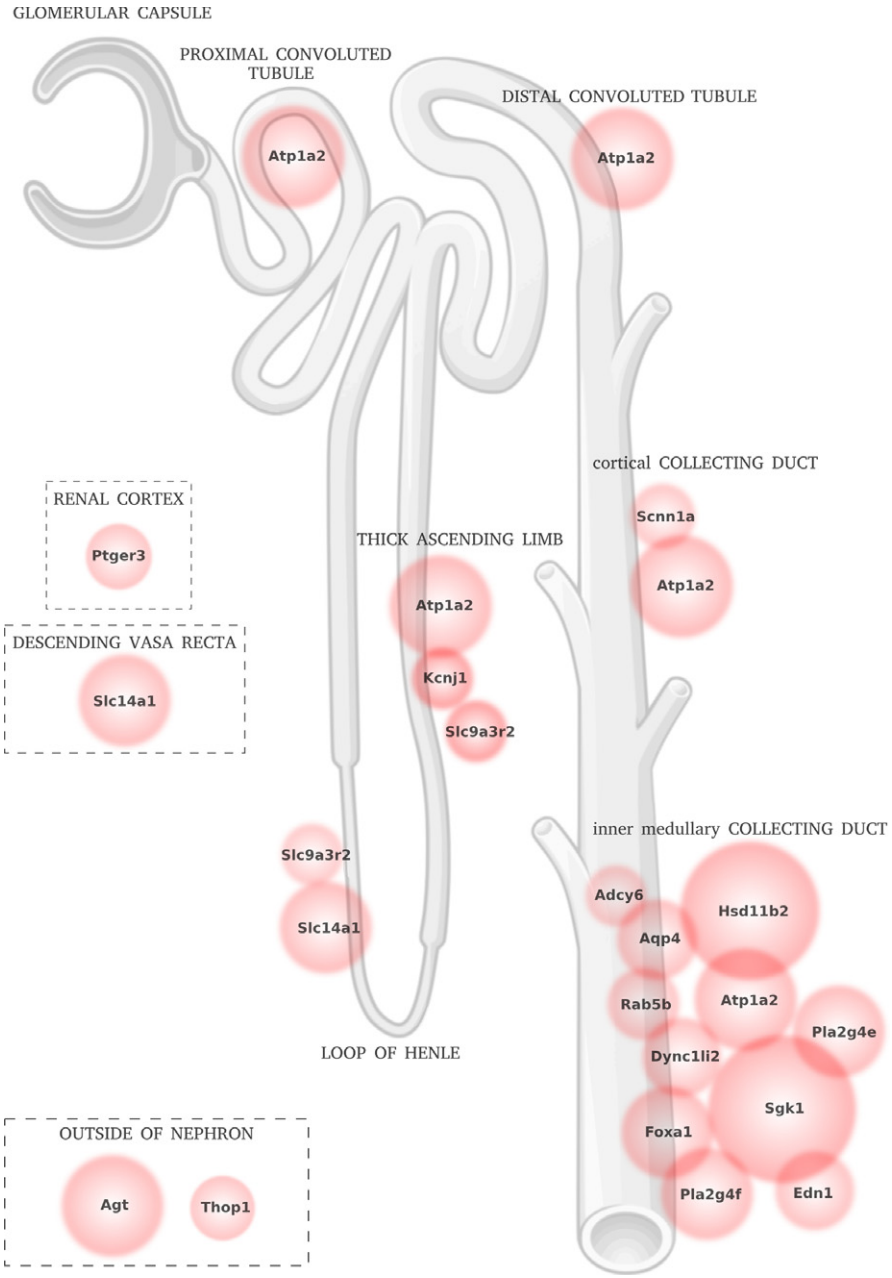

FIGURE 4 Nephron localization of some of the genes significantly overexpressed in transcriptomes of steppe specimens of the Patagonian olive mouse *Abrothrix olivacea*. Mapped genes are those present in the a priori list of 133 genes (Supporting information Table S3), which were significantly overexpressed in transcriptomes of the steppe samples. Circle diameters are proportional to the fold-change magnitude of gene expression. As reference, the expressions of Adcy6 and Sgk1 are 1.34 and 2.9 higher in the steppe than in the forest, respectively. Localization of gene expression is based on the review of house mouse studies by Fenton and Knepper (2007)

Genes with most extreme F_{ST} values (top 1%) that were not differentially expressed across environments include flavin-containing monooxygenase 3 (F_{ST} = 0.52) and glutathione S-transferase (F_{ST} = 0.58), which are involved in xenobiotic and drug metabolism (Krueger & Williams, 2005; Sherratt & Hayes, 2002). This further illustrates the importance of these processes across contrasting environments. No other genes associated with water absorption or immune response pathways were found among these F_{ST} outliers (Supporting Information Table S12).

Tajima's D is a statistic sensitive to both demographic history and departures from neutrality. Nondifferentially expressed genes cover a broad range of *D* values, with means of −0.74 and −0.85 in the forest and steppe, respectively (Figure 5c,d). Differentially expressed genes also show broad, but right-shifted distributions of *D* values (means were −0.69 and −0.79 in the forest and steppe, respectively). Overall, all distributions of *D* are centred on negative values that imply an excess of low-frequency alleles; meanwhile, right-shifted distributions of *D* in differentially expressed genes suggest that intermediate frequency alleles are more common in these genes. The same contrasts were observed when comparing *D* separately by locality (Supporting Information Figure S3).

4 | DISCUSSION

4.1 | Phenotypic variables and gene expression analysis

Body mass was smaller in the steppe—particularly in the northern steppe site (i.e., the driest sampling point)—than in forests, reinforcing the idea that more productive habitats allow for larger body sizes

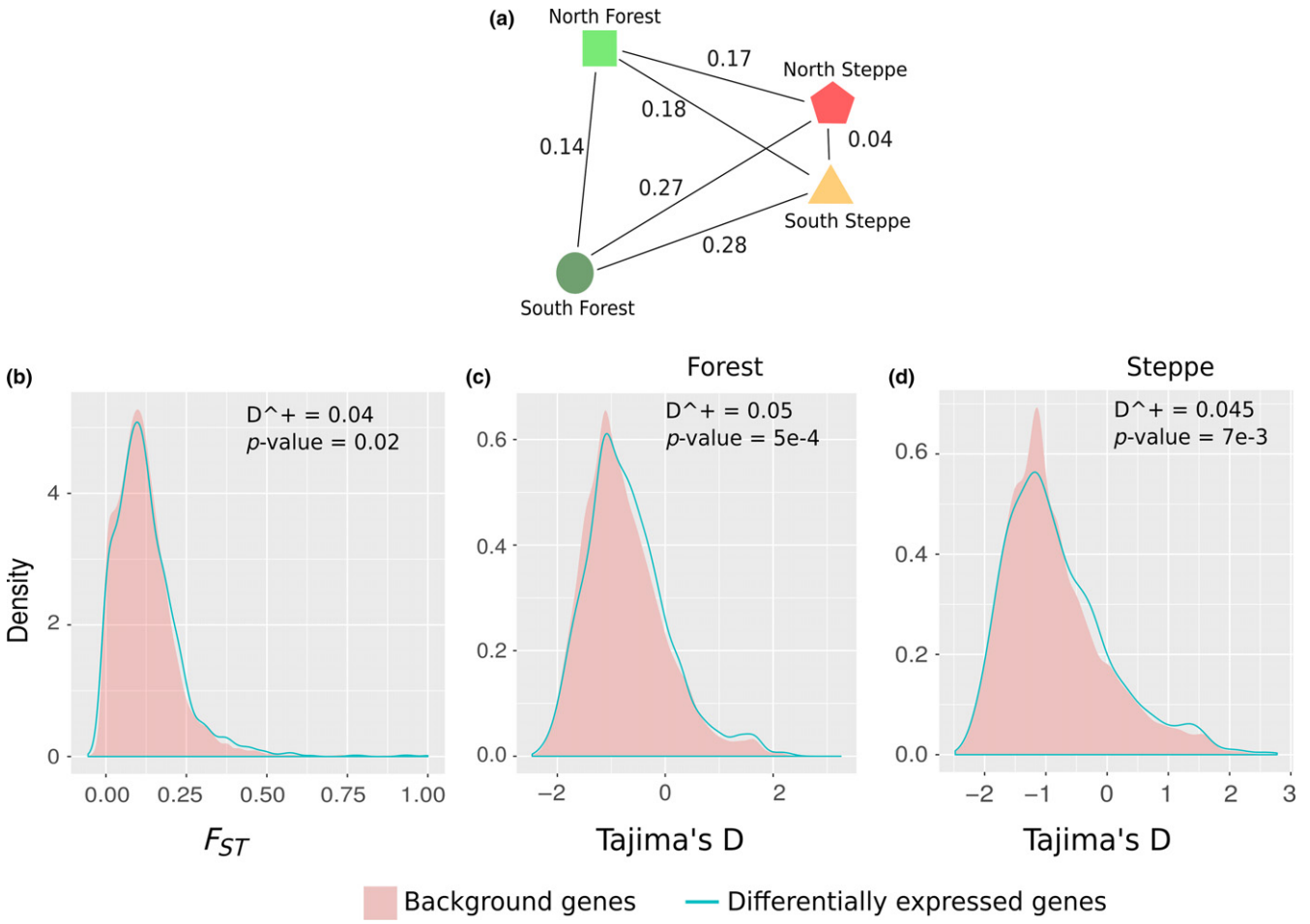

FIGURE 5 Population genetics analysis. (a) Pairwise F_{ST} estimation for the four localities. (b) Density plot of F_{ST} values between steppe and forest samples. (c) Density plot of Tajima's D for forest samples. (d) Density plot of Tajima's D for steppe samples. For b–d, distributions for differentially expressed genes are shown in blue, while those for remaining genes are shown in red. Kolmogorov–Smirnov test values and their significance are indicated. Positive D^+ values indicate that the cumulative distribution function is higher in background genes than in differentially expressed genes; thus, the density distribution of the differentially expressed genes is it moved to the right (towards higher values)

(Al-Kahtani et al., 2004; Gür & Kart Gür, 2012). We found that both kidney relative mass and [U/P] values were moderately but significantly correlated with annual rainfall; animals from more xeric habitats showed larger kidneys as well as a greater urine concentration. This is in agreement with previous studies analysing morphological and physiological variability in renal function across geographical gradients of water availability (Al-Kahtani et al., 2004; Sabat, Maldonado, Canals, & Del Rio, 2006). However, the specimens of southern forest showed an erratic pattern, as they unexpectedly presented [U/P] and kidney relative size values comparable or higher than the steppe populations (Figure 2b,c). Osmolarity of urine and plasma varies widely in response to water intake and thus can be strongly affected by short-term conditions (Sands & Layton, 2014). In fact, although we find a significant trend, [U/P] values presented a large variation and all the populations presented discrete maximum [U/P] values (Figure 2c); the maximum [U/P] value registered by *A. olivacea* is 11.4, while the higher value found in our populations was 3 (McNab & Brown, 2002). This is in part a contrast between measurements taken in the field and those using a water deprivation protocol. Variation in the relative size of the kidneys of the southern forest population is harder to interpret and will require further scrutiny.

We found marked differences in expression levels between forest and steppe samples, rather than between localities within each of these biomes. Interestingly, we found a significant enrichment of key genes related to water conservation among those that were overabundant in the steppe. Our analyses clearly indicate that this species resorts to the "classical" gene pathways of water conservation and solute balance to face water scarcity in the steppe. Among these genes, the differential expression of *AQP4* is interesting because this gene is apparently not expressed in the kidneys of desert mammals, such as the kangaroo rat *Dipodomys merriami* (Huang et al., 2001) and camelids (Wu et al., 2014), for which alternative mechanisms of water conservation have been proposed. Likewise, a laboratory study carried out in the desert-adapted cactus mouse, *Peromyscus eremicus* (MacManes, 2017), found *AQP4* and *SCNN1A* significantly overexpressed in the control group, instead of the treatment group which was under water deprivation for 72 hr.

According to knockout experiments in the house mouse, both of these genes are important for urine concentration (by means of water or sodium balance). Thus, although the Patagonian olive mouse has a capacity to concentrate urine comparable to several desert-adapted species (McNab & Brown, 2002), it apparently differs in the mechanism to achieve water conservation. The comparison of our results with genes exclusively reported to be important for water conservation in the cactus mouse also denoted striking differences. We found the sodium/calcium exchanger *SLC8A1* to be significantly overabundant in the transcriptome of steppe specimens, whereas the study carried out in cactus mouse found it to be overexpressed in the control group (MacManes, 2017).

We also found other genes that belong to the classical pathways of water and sodium balance to be differentially expressed between biomes. These included both angiotensinogen and the urea transporter, which are overabundant in the steppe. Liver angiotensinogen is expected to be responsible for the production of angiotensin II in the kidney (Matsusaka et al., 2012), but our results would suggest that angiotensinogen locally expressed in the kidney could be also important for water conservation. It is also noteworthy that 11 of the 18 a priori genes overexpressed in the steppe biome are expressed in the collecting tube, a key nephron region to concentrate urine (Figure 4). These include the urea transporter, one of the most important genes to achieve concentrated urine. Only one of the 18 genes, endothelin-1, shows an inverted expression direction relative to what is known about its function (Fenton & Knepper, 2007) but its receptor, endothelin receptor type A, is significantly overabundant in the forest cluster and also a member of the a priori gene list. It is presumed that the interaction between these two genes limits water reabsorption (Fenton & Knepper, 2007).

Differential expression was not restricted only to genes involved in water conservation. The pathway analysis also highlighted the well-known detoxification function of kidneys (Lohr, Willsky, & Acara, 1998) and a strong divergence in immune response genes, suggesting that these two functions also vary across biomes. The detoxification and drug metabolism functions of the kidney could be activated by the secondary compounds of the plants consumed in each of the environments (see Haley, Lamb, Franklin, Constance, & Dearing, 2007a; Haley, Lamb, Franklin, Constance, & Denise Dearing, 2007b). It has been reported that the diet of *A. olivacea* differs considerably between the forest and the steppe (Silva (2005) and references therein); in the forest, the diet is predominantly herbivorous and animals have larger intestines, whereas in the more arid environments, the diet becomes more omnivorous–granivorous and animals have smaller intestines (Naya et al., 2014). Therefore, over-represented pathways related to drug metabolism found almost exclusively on the forest samples may be due to the load of secondary plant compounds in an herbivorous diet.

As for immune responses, steppe and forest individuals may differ in levels of exposure to micro-organisms and parasites. For instance, *A. olivacea* is a reservoir of leptospirosis (Zamora & Riedemann, 1999), a disease strongly associated with rainy environments like our northern forest site. As expected, the immune response pathways were mainly associated with increased gene expression in the forest. Finally, the genes related to renin–angiotensin system pathway were unexpectedly over-represented in the forest; besides its link to water reabsorption pathways (Sparks et al., 2014), this system is also related to various inflammatory responses (Capettini et al., 2012; Husain, Hernandez, Ansari, & Ferder, 2015; Stegbauer et al., 2009), such as those expected under an immune response. In this regard, we found an overexpression of renin in the forest, a direction that was, again, in contraposition of what has been described in *Peromyscus eremicus*. Renin catalyses the first step in the activation pathway of angiotensinogen (Fenton & Knepper, 2007), and it is noteworthy because we found angiotensinogen and several key genes that the renin–angiotensin pathways ultimately activate to be significantly overexpressed in transcriptomes of individuals from the steppe. Further studies are needed to disentangle these interactions, in particular those involving immune response and water conservation.

In summary, our results show that (a) relevant phenotypic changes are associated—although not perfectly—with the two biomes and more generally with ecogeographical variation; (b) gene expression differences between steppe and forests involve, at least in part, "classical" genes related to water conservation, rather than the particular modification found in desert-adapted rodents and camelids; and (c) genes related to other pathways (including detoxification and immune responses) also show differences in gene expression between steppe and forest. Differential expression in the kidney could be achieved by several means, such as differential activity in specialized tissues or changes in their relative development (e.g., variation in the proportions of kidney cortex and medulla), but these are not distinguished by our analyses.

4.2 | Genomic and expression divergence

According to our PCA results, the sharpest expression differences are between the steppe and forest biomes, although a difference between the two forest sites was also noticeable. This is in line with the pairwise F_{ST} results, where the largest differences were found between the biomes (F_{ST} from 0.17 to 0.28) followed by the divergence between the two forest sites (F_{ST} ~0.14), while the divergence between the two steppe sites was only of 0.04. Following the PCA partition, we found almost 3,000 genes differentially expressed between the biomes, which also presented, on average, slightly larger values of F_{ST}. Based on our results, it appears that differentially expressed genes would be important for coping with different environment conditions in two ways: first, through gene expression modulation, and second, through local genetic adaptation, as these genes may have different alleles or combinations of alleles in different biomes. However, the association between differential expression and genetic divergence is complex. For instance, the distributions of F_{ST} values of both differentially expressed and baseline genes are broad and largely overlapping. Also, gene expression modulation may be obtained plastically, as well as via genetic differences in regulatory factors. Additional detailed studies focused on the interplay

between these factors in particular sets of genes are needed. For example, geographical variation in SNPs found in untranslated gene regulatory regions should be examined in relation to differential expression.

In analysing Tajima's D distributions, we found them biased towards negative values, which is expected if, as has been suggested (Lessa, D'Elía, & Pardiñas, 2010), the species has a history of recent demographic expansion/colonization; in addition, purifying selection against slightly deleterious variants may be common in protein-coding genes (e.g., Carneiro et al., 2012; Jackson, Campos, & Zeng, 2015), also resulting in negative values of *D*. Relative to that overall pattern, the right-shifted distribution of *D* values for the differentially expressed genes suggests that balancing selection may be more common in differentially expressed loci. Establishing whether and, if so, how balancing selection is acting within environments is challenging. Among the "classical" mechanism of balancing selection, heterozygote advantage may be the most likely mechanism behind the genes related to immune system. The association between the MHC-related genes and balancing selection through heterozygote advantage is well known (Rikowski & Grammer, 1999; Thornhill et al., 2003). The joint interpretation of both statistics, namely F_{ST} and Tajima's D, is even more complicated. In any case, studies of specific sets of genes are needed to further establish the role of selection in their patterns of variation.

Altogether, our results strongly suggest that the broadly adapted olive mouse recurs to the "classical" gene pathways of water conservation and solute balance to face water scarcity in the arid steppe. At least for some key genes, our species seems to respond in a manner different from desert-adapted species, especially the cactus mouse. Also, differentially expressed genes showed a slight tendency to be more divergent than the remaining, baseline genes. In short, despite having a remarkable capacity to concentrate urine, our work points out that olive mouse does not present the modifications in the "classical" gene pathways of water conservation proposed for some desert rodents. Future experiments in controlled conditions are needed to understand (a) how much of our gene expression results could be driven by other environmental factors and (b) whether under severe water deprivation conditions, our observed pattern of gene expression is maintained. On the other hand, genotyping at the DNA level could provide a better assessment of genetic variation by avoiding the problems related to the RNA-seq approaches (e.g., differential allele expression (Castel, Levy-Moonshine, Mohammadi, Banks, & Lappalainen, 2015)).

ACKNOWLEDGMENTS

We are grateful to Pablo Teta, Daniel Udrízar-Sauthier and Ulyses Pardiñas for assistance in the field and to José Boggia and Liliana Gadola for their helpful comments on the function of genes located in the nephron. We thank Federico Hoffmann and three anonymous reviewers for most helpful comments on an earlier version of the manuscript. This work was supported by grants from CSIC-Universidad de la República, Agencia Nacional de Investigación e Innovación (ANII FCE 2014 103508), Uruguay, and FONDECYT 1141055 and 1180366, Chile. Facundo Giorello is supported by graduate fellowships from ANII.

DATA ACCESSIBILITY

Illumina sequence reads: NCBI SRA: Reads can be found with the entries for NCBI BioProjects: PRJNA471316. Final transcriptome assembly: Dryad https://doi.org/10.5061/dryad.7nh50k7.

AUTHOR CONTRIBUTIONS

M.F., G.D., L.V., D.E.N. and E.P.L. participated in the fieldwork and sample preparations. F.G., M.F. and L.V. carried out the analyses. G.D., J.C.O., D.E.N. and E.P.L. designed and supervised the project. F.G. wrote the manuscript. All authors approved the final manuscript.

ORCID

Facundo M. Giorello http://orcid.org/0000-0001-5158-7225

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Al-Kahtani, M. A., Zuleta, C., Caviedes-Vidal, E., & Garland, T. (2004). Kidney mass and relative medullary thickness of rodents in relation to habitat, body size, and phylogeny. *Physiological and Biochemical Zoology*, *77*, 346–365. https://doi.org/10.1086/420941

Altenhoff, A. M., Schneider, A., Gonnet, G. H., & Dessimoz, C. (2011). OMA 2011: Orthology inference among 1000 complete genomes. *Nucleic Acids Research*, *39*, D289–D294. https://doi.org/10.1093/nar/gkq1238

Barrett, J. M., Kriz, W., Kaissling, B., & de Rouffignac, C. (1978). The ultrastructure of the nephrons of the desert rodent (*Psammomys obesus*) kidney II. Thin limbs of Henle of long-looped nephrons. *American Journal of Anatomy*, *151*, 499–514. https://doi.org/10.1002/(ISSN)1553-0795

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*, 289–300.

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

Boone, M., & Deen, P. M. T. (2008). Physiology and pathophysiology of the vasopressin-regulated renal water reabsorption. *Pflugers Archiv European Journal of Physiology*, *456*, 1005–1024. https://doi.org/10.1007/s00424-008-0498-1

Capettini, L., Montecucco, F., Mach, F., Stergiopulos, N., Santos, R. A., & da Silva, R. F. (2012). Role of renin-angiotensin system in inflammation, immunity and aging. *Current Pharmaceutical Design*, *18*, 963–970. https://doi.org/10.2174/138161212799436593

Carneiro, M., Albert, F. W., Melo-Ferreira, J., Galtier, N., Gayral, P., Blanco-Aguiar, J. A., … Ferrand, N. (2012). Evidence for widespread positive and purifying selection across the European rabbit (*Oryctolagus cuniculus*) genome. *Molecular Biology and Evolution*, *29*, 1837–1849. https://doi.org/10.1093/molbev/mss025

Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., & Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biology*, *16*, 195. https://doi.org/10.1186/s13059-015-0762-6

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*, 7. https://doi.org/10.1186/s13742-015-0047-8

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Degen, A. A., Kam, M., Hazan, A., & Nagy, K. A. (1986). Energy expenditure and water flux in three sympatric desert rodents. *Journal of Animal Ecology*, *55*, 421–429. https://doi.org/10.2307/4728

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology*, *4*, P3. https://doi.org/10.1186/gb-2003-4-5-p3

Diaz, G. B., Ojeda, R. A., & Rezende, E. L. (2006). Renal morphology, phylogenetic history and desert adaptation of South American hystricognath rodents. *Functional Ecology*, *20*, 609–620. https://doi.org/10.1111/j.1365-2435.2006.01144.x

Fenton, R. A., & Knepper, M. A. (2007). Mouse models and the urinary concentrating mechanism in the new millennium. *Physiological Reviews*, *87*, 1083–1112. https://doi.org/10.1152/physrev.00053.2006

Giorello, F. M., Feijoo, M., D'Elía, G., Valdez, L., Opazo, J. C., Varas, V., … Lessa, E. P. (2014). Characterization of the kidney transcriptome of the South American olive mouse Abrothrix olivacea. *BMC Genomics*, *15*, 446.

Glavinas, H., Krajcsi, P., Cserepes, J., & Sarkadi, B. (2004). The role of ABC transporters in drug resistance, metabolism and toxicity. *Current Drug Delivery*, *1*, 27–42. https://doi.org/10.2174/1567201043480036

Golias, C., Charalabopoulos, A., Stagikas, D., Charalabopoulos, K., & Batistatou, A. (2007). The kinin system–bradykinin: Biological effects and clinical implications. Multiple role of the kinin system–bradykinin. *Hippokratia*, *11*, 124–128.

Gür, H., & Kart Gür, M. (2012). Is spatial variation in food availability an explanation for a Bergmannian size pattern in a North American hibernating, burrowing mammal? An information-theoretic approach. *Journal of Zoology*, *287*, 104–114. https://doi.org/10.1111/j.1469-7998.2011.00893.x

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haley, S. L., Lamb, J. G., Franklin, M. R., Constance, J. E., & Dearing, M. D. (2007a). Xenobiotic metabolism of plant secondary compounds in oak (Quercus Agrifolia) by specialist and generalist woodrat herbivores, genus Neotoma. *Journal of Chemical Ecology*, *33*, 2111–2122. https://doi.org/10.1007/s10886-007-9371-5

Haley, S. L., Lamb, J. G., Franklin, M. R., Constance, J. E., & Denise Dearing, M. (2007b). Xenobiotic metabolism of plant secondary compounds in juniper (*Juniperus monosperma*) by specialist and generalist woodrat herbivores, genus Neotoma. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, *146*, 552–560. https://doi.org/10.1016/j.cbpc.2007.06.007

Heymann, J., & Engel, A. (1999). Aquaporins: Phylogeny, structure, and physiology of water channels. *News in Physiological Sciences*, *14*, 187–193.

Huang, Y., Tracy, R., Walsberg, G. E., Makkinje, A., Fang, P., Brown, D., & Van Hoek, A. N. (2001). Absence of aquaporin-4 water channels from kidneys of the desert rodent Dipodomys merriami merriami. *American Journal of Physiology-Renal Physiology*, *280*, F794–F802. https://doi.org/10.1152/ajprenal.2001.280.5.F794

Husain, K., Hernandez, W., Ansari, R. A., & Ferder, L. (2015). Inflammation, oxidative stress and renin angiotensin system in atherosclerosis. *World Journal of Biological Chemistry*, *6*, 209–217. https://doi.org/10.4331/wjbc.v6.i3.209

Jackson, B. C., Campos, J. L., & Zeng, K. (2015). The effects of purifying selection on patterns of genetic differentiation between Drosophila melanogaster populations. *Heredity*, *114*, 163–174. https://doi.org/10.1038/hdy.2014.80

Krueger, S. K., & Williams, D. E. (2005). Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacology and Therapeutics*, *106*, 357–387. https://doi.org/10.1016/j.pharmthera.2005.01.001

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.

Lessa, E. P., D'Elía, G., & Pardiñas, U. F. J. (2010). Genetic footprints of late Quaternary climate change in the diversity of Patagonian-Fuegian rodents. *Molecular Ecology*, *19*, 3031–3037. https://doi.org/10.1111/j.1365-294X.2010.04734.x

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323. https://doi.org/10.1186/1471-2105-12-323

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lohr, J. W., Willsky, G. R., & Acara, M. A. (1998). Renal drug metabolism. *Pharmacological Reviews*, *50*, 107–141.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8

MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics*, *5*, 13.

MacManes, M. D. (2017). Severe acute dehydration in a desert rodent elicits a transcriptional response that effectively prevents kidney injury. *American Journal of Physiology - Renal Physiology*, *313*, F262–F272. https://doi.org/10.1152/ajprenal.00067.2017

MacManes, M. D., & Eisen, M. B. (2014). Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse *Peromyscus eremicus*. *PeerJ*, *2*, e642. https://doi.org/10.7717/peerj.642

Macmillen, R. E., & Hinds, D. S. (1983). Water regulatory efficiency in heteromyid rodents: A model and its application (Dipodomys, Perognathus). *Ecology*, *64*, 152–164. https://doi.org/10.2307/1937337

MacMillen, R. E., & Lee, A. K. (1967). Australian desert mice: Independence of exogenous water. *Science*, *158*, 383–385. https://doi.org/10.1126/science.158.3799.383

Marra, N. J., Eo, S. H., Hale, M. C., Waser, P. M., & Dewoody, J. A. (2012). A priori and a posteriori approaches for finding genes of evolutionary interest in non-model species: Osmoregulatory genes in the kidney transcriptome of the desert rodent *Dipodomys spectabilis* (banner-tailed kangaroo rat). *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *7*, 328–339.

Marra, N. J., Romero, A., & Dewoody, J. A. (2014). Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. *Molecular Ecology*, *23*, 2699–2711. https://doi.org/10.1111/mec.12764

Matsusaka, T., Niimura, F., Shimizu, A., Pastan, I., Saito, A., Kobori, H., … Ichikawa, I. (2012). Liver angiotensinogen is the primary source of renal angiotensin II. *Journal of the American Society of Nephrology*, *23*, 1181–1189. https://doi.org/10.1681/ASN.2011121159

Mbassa, G. K. (1988). Mammalian renal modifications in dry environments. *Veterinary Research Communications*, *12*, 1–18. https://doi.org/10.1007/BF00396399

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

McNab, B., & Brown, J. (2002). *Physiological Ecology of Vertebrates: A View from Energetics*. Ithaca: Cornell University Press.

Murillo-Carretero, M. I., Ilundáin, A. A., & Echevarria, M. (1999). Regulation of aquaporin mRNA expression in rat kidney by water intake. *Journal of the American Society of Nephrology*, *10*, 696–703.

Naya, D. E., Feijoo, M., Lessa, E. P., Pardiñas, U. F. J., Teta, P., Tomasco, I. H., … D'Elía, G. (2014). Digestive morphology of two species of *Abrothrix* (Rodentia, Cricetidae): Comparison of populations from contrasting environments. *Journal of Mammalogy*, *95*, 1222–1229. https://doi.org/10.1644/13-MAMM-A-261

Patton, J. L., Pardiñas, U. F. J., & D'Elía, G. (2015). *Mammals of South America, Volume 2: Rodents*.

Rikowski, A., & Grammer, K. (1999). Human body odour, symmetry and attractiveness. *Proceedings. Biological Sciences/The Royal Society*, *266*, 869–874. https://doi.org/10.1098/rspb.1999.0717

Rozansky, D. J. (2006). The role of aldosterone in renal sodium transport. *Seminars in Nephrology*, *26*, 173–181. https://doi.org/10.1016/j.semnephrol.2005.09.008

Sabat, P., Maldonado, K., Canals, M., & Del Rio, C. M. (2006). Osmoregulation and adaptive radiation in the ovenbird genus Cinclodes (Passeriformes: Furnariidae). *Functional Ecology*, *20*, 799–805. https://doi.org/10.1111/j.1365-2435.2006.01176.x

San-Cristobal, P., De Los, Heros. P., Ponce-Coria, J., Moreno, E., & Gamba, G. (2008). WNK kinases, renal ion transport and hypertension. *American Journal of Nephrology*, *28*, 860–870. https://doi.org/10.1159/000139639

Sands, J. M., & Layton, H. E. (2014). Advances in understanding the urine-concentrating mechanism. *Annual Review of Physiology*, *76*, 387–409. https://doi.org/10.1146/annurev-physiol-021113-170350

Sherratt, P. J., & Hayes, J. D. (2002). Glutathione S-transferases. In C. Ioannides (Ed.), *Enzyme systems that metabolise drugs and other xenobiotics* (pp. 319–352). Chichester, UK: Wiley-Blackwell. https://doi.org/10.1002/0470846305

Sikes, R. S., & Gannon, W. L. (2011). Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *Journal of Mammalogy*, *92*, 235–253. https://doi.org/10.1644/10-MAMM-F-355.1

Silva, S. I. (2005). Posiciones tróficas de pequeños mamíferos en Chile: Una revisión. *Revista Chilena de Historia Natural*, *78*, 589–599.

Sparks, M. A., Crowley, S. D., Gurley, S. B., Mirotsou, M., & Coffman, T. M. (2014). Classical Renin-Angiotensin system in kidney physiology. *Comprehensive Physiology*, *4*, 1201–1228. https://doi.org/10.1002/cphy.c130040

Stegbauer, J., Lee, D. H., Seubert, S., Ellrichmann, G., Manzel, A., Kvakan, H., … Linker, R. A. (2009). Role of the renin-angiotensin system in autoimmune inflammation of the central nervous system. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 14942–14947. https://doi.org/10.1073/pnas.0903602106

Thornhill, R., Gangestad, S. W., Miller, R., Scheyd, G., McCollough, J. K., & Franklin, M. (2003). Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology*, *14*, 668–678. https://doi.org/10.1093/beheco/arg043

Withers, P. C., Louw, G. N., & Henschel, J. (1980). Energetics and water relations of Namib Desert rodents. *South African Journal of Zoology*, *15*, 131–137. https://doi.org/10.1080/02541858.1980.11447700

Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., … Wang, J. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nature Communications*, *5*, 5188. https://doi.org/10.1038/ncomms6188

Zamora, J., & Riedemann, S. (1999). Animales silvestres como reservorios de leptospirosis en Chile. Una revisión de los estudios efectuados en el país. *Archivos de Medicina Veterinaria*, *31*, 151–156.

SUPPORTING INFORMATION

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

How to cite this article: Giorello FM, Feijoo M, D'Elía G, et al. An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). *Mol Ecol*. 2018;00:1–13. https://doi.org/10.1111/mec.14778

Capítulo 2.

Durante los períodos glaciares, las capas de hielo que se encontraban a lo largo de las laderas andinas y australes de América del Sur forzaron cambios en la distribución de muchas especies. La distribución poblacional de muchas especies se vio restringida a refugios glaciares, incluido posiblemente uno en lo que actualmente es el bosque Valdiviano. Estos cambios en la distribución trajeron consigo también cambios demográficos distintivos entre las especie de ambientes abiertos (e.g estepa) y ambientes cerrados como el bosque. Según un meta-análisis, las especies de bosque muestran, típicamente, una contracción poblacional asociada al avance de la capa de hielo mientras que la respuesta de las especies de ambientes abiertos parece ser mucho más errática. En lo que respecta particularmente a los roedores patagónicos, éstos han mostrado unidades filogeográficas dispuestas latitudinalmente, y no asociadas a los ambientes que se disponen preferentemente de manera longitudinal, y han mostrado también diferentes respuestas asociado a los períodos glaciales. Un patrón de colonización postglacial de norte a sur se ha propuesto para algunos roedores patagónicos, pero también se han descrito procesos más complejos, como la diferenciación dentro una región determinada y la persistencia de la población en posibles refugios. Para *Abrothrix olivacea*, una especie de particular interés al encontrarse ampliamente distribuida en la región Patagónica, se ha encontrado un clado mitocondrial distinto en Tierra del Fuego por lo cual se ha hipotetizado un refugio adicional en esta región. Con el propósito de caracterizar las respuestas demográficas de poblaciones de estepa y bosque, así como para evaluar las hipótesis particulares relacionadas a *A. olivacea*, genotipamos más de 172 ejemplares del ratón oliváceo con un protocolo de captura de exones obteniendo más 50.000 SNPs. Nuestros análisis muestran que el ratón oliváceo se encuentra estructurado genéticamente en dos grandes grupos que corresponden básicamente a la estepa y pastizales pataǵonicos del lado este de la cordillera (y que incluye a Tierra del Fuego, independientemente de sus diferentes biomas) y a los bosques valdivianos y magallánicos asociados a los Andes. Nuestras estimaciones del momento de inicio de la expansión demográfica para ambos grupos de población son compatibles con una expansión posterior al Último Máximo Glacial. También identificamos un proceso de colonización significativo y una clina en los valores de heterocigosidad desde el noroeste hacia la Patagonia oriental, y hasta Tierra del Fuego, lo que descarta la posibilidad de un refugio sureño en Tierra del Fuego y brinda evidencia adicional a favor de los refugios “clásicos”. Nuestros resultados sugieren, además, un nuevo refugio glacial periférico hasta ahora solo reportado en plantas. En conclusión, nuestros resultados apoyan la idea de un refugio costero en o cercano al refugio Valdiviano anteriormente propuesto, desde donde a su vez comenzó la colonización, y un posible refugio adicional.

Genomic footprints of Quaternary colonization and population expansion in the Patagonian olive mouse (*Abrothrix olivacea*).

Facundo M. Giorello [1,2], Enrique P. Lessa [1].

1 Departamento de Ecología y Evolución, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.

2 Espacio de Biología Vegetal del Noreste, Centro Universitario de Tacuarembó, Universidad de la República, Tacuarembó, Uruguay.

INTRODUCTION

Glacial cycles during the Quaternary significantly impacted the distribution of species all around the world. Typically, during glacial phases populations of high-latitude species are forced into ice-free refugia, experimenting population contraction, while in the interglacial phases populations expand and colonize newly available terrain. However, numerous phylogeographic studies in both northwestern North America and South America have already revealed much more complex scenarios. For instance, in northwestern North America, in addition to the two classical refugia, namely Beringia and the Pacific Northwest refugium, additional refugia have been suggested, including cryptic coastal refugia, as well as refugia enclosed by ice sheets; in addition, some refugia appear to have been subdivided, increasing the complexity of the emergent, postglacial patterns (Shafer et al. 2010). In South America, several studies have identified contrasting responses to glacial cycles in species from open and forested biomes (Turchetto-Zolet et al. 2013). According to Turchetto-Zolet et al. (2013), forest species showed the typical population contraction associated with the advance of ice-sheets, while the response of open environment species seems to have been much more erratic. Additionally, different demographic processes and trajectories (e.g., demographic or range expansion or stability, fragmentation, secondary contact) have been recorded in co-distributed species of plants and vertebrates (Sérsic et al. 2011).

In regards to Patagonian rodent species, they have been differentially impacted by past climatic dynamics. A pattern of north-to-south postglacial colonization has been proposed in several species of Patagonian rodents (e.g., *Eligmodontia morgani* and *Reithrodon auritus*), although more complex processes, such as within-region differentiation, and population persistence at putative, southern refugia, have also been suggested (Valdez & D'Elía 2018; Cañón et al. 2010; Lessa et al. 2010).

In Southern South America, five major glaciations were identified that dramatically altered Patagonian landscapes (Hulton et al. 2002; Rabassa et al. 2011; McCulloch et al. 2000). The “Great Patagonian Glaciation” (GPG) that occurred approximately 1 MYA was the most extensive, covering completely Tierra del Fuego and exposing significant segments of the southern Atlantic continental platform. After three other major glaciations (known as post GPG), the most recent of these events peaked at the Last Glacial Maximum (LGM), which began ∼20,000–19,000 YA and finished 14,000–10,000 YA (Glasser et al. 2008; Rabassa et al. 2011). During at least the five major glaciations an 1,800 km-long ice sheet built up along the axis of the southern Andes from 37°S to Cape Horn (56°S) (Rabassa et al. 2011; McCulloch et al. 2000). At Los Lagos Regions in Chile, pollen records indicate a vegetation change from Subantarctic Parkland to open woodland and later to North Patagonian Evergreen Forest after 14,000 YA (Heusser et al. 1996). In Tierra del Fuego and southernmost Patagonia, after 12,500 YA, steppe replaced the heathlands suggesting decrease in wind intensity, increase in effective moisture, and increase in temperatures (Markgraf 1993). Also during the LGM, sea level was at approximately 120 and 140 m below present sea level, exposing the Argentinian Continental shelf and generating an enormous coastal plain along the Pampa and Patagonia (Ponce et al. 2011). The separation of the Isla Grande de Tierra del Fuego from the rest of the continent and the formation of the present Magellan Straits would have taken place approximately 10,200 YA, when the sea level ascended above -35 m. Such changes are likely to have influenced patterns of faunal and plant abundance and distribution, and may have affected dispersal, migration and displacement processes since the LGM to Middle Holocene times (Ponce et al. 2011).

In this work, in order to understand the impact of climatic oscillation in Patagonian species we focus in the Patagonian olive mouse *Abrothrix olivacea,* one of the most widely distributed cricetid rodents of South America. The olive mouse is distributed along Chile and Argentinian Patagonia, from 18ºS to 55ºS latitude (D’Elía & Pardiñas 2015; Mann 1978) extending for over 1000 km latitudinally, and encompassing a great variety of open and closed environments: coastal deserts in the north, Mediterranean scrubs in central Chile, Valdivian and Magallanic forests through the south of Chile and Argentina and Patagonian steppe towards the Atlantic coast. For *A. olivacea*, as well for other species, the Valdivian region have been proposed as the classical refugium during glacial times (Smith et al. 2001; Palma et al. 2012). However, *Abrothrix olivacea*, presented also a distinct mtDNA clade in Tierra del Fuego, suggesting an additional, southern refugium. Moreover, time estimation of population expansion for this species leave open the possibility of a post LGM

expansion inside Tierra del Fuego (Lessa et al. 2010; Abud 2011). These results, however, are based solely on mtDNA sequences; selection processes specific to, and/or the particulars of the realization of demographic processes of the mitochondrial genome, make it difficult to generalize these demographic reconstructions to the species level.

In addition to the Valdivian and Fueguian, several other refugia have been proposed in Patagonia based on phylogeographic analyses of plants and vertebrates (Sérsic et al. 2011). In vertebrates, phylogeographic breaks are mainly present in eastern Patagonia, whereas in the Chilean forest the breaks are always in or north of the Valdivian Forest. In the forest, only one peripheral glacial refugium, at the same latitude of the Valdivian refugium, has been proposed along the Andean slopes. For plants, in contrast, many peripheral glacial refugia have been proposed and two additional phylogeograhical breaks have been described in the southern of Valdivian and Magellanic forests (Sérsic et al 2011). Some species of plants (e.g *Oreobolus obtusangulus*) may even have persisted in the heavily glaciated Patagonian Channels during the LGM according to a recent multilocus study (Pfanzelt et al. 2017).

To confirm the classical Valdivian refugium and test the hypothesis of additional refugia, we sampled 172 individuals of the ubiquitous *A. olivacea* species covering much of southern South America. *A. olivacea* individual, as well as 4 individuals of *Abrothrix lanosa* and 3 of *Abrothrix hirta*, were genotyped for more than 50,000 SNPs using an exon-capture protocol. Our main objectives were: (a) to evaluate possible refugia for *A. olivacea*, including the proposed Valdivian and Tierra del Fuego refugia; (b) to study the phylogeographic pattern of the species and estimate divergence times between major geographical units; (c) to evaluate the demographic dynamics of steppe and forest population in response to climatic oscillations during the Pleistocene; and (d) to establish the sources and direction of colonization events. Our analyses identified two genetic clusters, principally arranged longitudinally and broadly corresponding to open and forest biomes. These clusters of populations show contrasting geographical patterns, reflecting contrasting biogeographical histories, including levels of geographical differentiation and genetic diversity. However, our estimates of the timing of the onset of demographic expansion for both population clusters are consistent with post-glacial expansions. We also identified a significant range expansion and a heterozygosity cline from the Valdivian forest towards the Eastern Patagonia reaching Tierra del Fuego that discard the possibility of a separate, southern refugium in Tierra del Fuego. However, our results suggested new periglacial refugia close to the Magellanic forest so far only reported in plants. In general, our results support the idea of a coastal refugium in or near the

previously proposed Valdivian refugium from where, in turn, colonization of all open continental biomes started. Tierra del Fuego was colonized as a result of this southward, stepwise process, from neighboring continental areas.

MATERIALS AND METHODS

Samples

Adult individuals of 21 populations of *A. olivacea* (n = 172) and two other species, *Abrothrix hirta* (n = 3) and *Abrothrix lanosa* (n = 4) (Figure 1; Table S1), were captured with Rodent Trap Special-S traps (Forma Ltda, Santiago, Chile) and then euthanized following the guidelines of the American Society of Mammalogists (Sikes 2016), by properly trained personnel and outside the perceptive range of the other captive individuals. Liver and muscle tissue were preserved in the field using alcohol 90 %.

Targeted exons

Using a previously constructed transcriptome reference (Giorello et al. 2018) we targeted for 8499 (> 8Mb) exons to be captured based on: (i) a minimum length of 300 bp, (ii) a minimum of length alignment with mouse exons of 300pb, (iii) a GC content between 30 and 70 %, (iv) excluding those with repetitive sequences and (v) excluding pseudogenes and mitochondrial genes. Exons from multi-copy genes were not excluded since the selection of unique copy genes could bias the analysis to more conserved genes, thus impacting the type (coding or non-coding) and the total number SNPs. The alignments between transcriptomic contigs and mouse ensembl coding sequence (Yates et al. 2016) were carried out using reciprocal blast (Altschul et al. 1990); custom python scripts were used to finally select the exons based on the criteria described above. The repetitive sequences were evaluating using *RepeatMasker* (Nishimura 2002) whereas the IDs of pseudogenes and mitochondrial genes (ultimately excluded from analyses) were downloaded using Biomart from Ensembl. After exon selection, their coordinates at transcriptomes contigs were supplied to Nimblegen for probe design for use with NimbleGen SeqCap EZ Enrichment System.

Genomic DNA library preparation and multiplexing

DNA was extracted from muscle and liver tissue using a standard protocol (modified from Miller et al. 1988). DNA was sheared using a Bioruptor through 1-4 rounds of sonication of 15 minutes each on low mode (30s on, 30s off). The ideal DNA fragment distribution should range from 100 to 500 bp. Following the protocol outlined by Meyer & Kircher 2010, individual genomic libraries were indexed via PCR using 7-nt barcoded primers after the ligation of universal adapters. From each PCR-indexed library, 100ng of DNA was pooled for the hybridization experiment.

Exon-capture

First, pooled libraries were denatured in a solution containing excess blocking oligos and Mouse COT DNA. Then, the pool was hybridized with the designed probes at 47°C (lid at 57°C) for 64 to 72 hours. DNA was eluted and PCR amplified for 11 cycles. We verified target enrichment with qPCR analysis of amplified elutes. Three target and non-target region were evaluated independently. Finally, the capture was diluted and sequencing was performed using Illumina HiSeq 4000, generating 150bp of paired-end reads.

SNP calling and filtering

Previous to SNP calling we used *Trimmomatic* (Bolger et al. 2014) to remove any adapter sequence and read 3' end bases were trimmed if their phred-value dropped below 24 (Q < 24). Read alignment was carried out with *Bowtie2* (Langmead & Salzberg 2012) using as reference only the targeted sequences from the transcriptomic reference. Reads were sorted, converted to bam, and deduped using *Picard Tools* 2.1.0 (https://broadinstitute.github.io/picard/). *GATK*'s HaplotypeCaller (McKenna et al. 2010) was used for SNP calling in three different runs: (i) a first joint genotyping using the 172 *A. olivacea* samples, (ii) a second joint genotyping with 16 specimens (n = 9 *A.olivacea* ; n= 4 *A. lanosa*; n=3 *A. hirta*) and (iii) a third joint genotyping using the 6 outgroup specimens (n= 4 *A. lanosa*; n=3 *A. hirta*). The first run was used for all the analyses with the exception for *SNAPP*, for which the second ran was used. The third run was used in conjunction with the first to determine ancestral alleles when using the directionality index (see below). The SNPs were filtered using Variant Filtration (-window 35 -cluster 3 --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"). For subsequent analyses, all gaps were removed and only biallelic sites without missing data were considered. For all the analyses except for Tajima's D values, one random SNP per exon was used.

Population Genetic Analyses

Population genetic analyses were performed using *VCFtools* (Danecek et al. 2011). We evaluated genetic differentiation through the F_{ST} fixation index and demographic history using Tajima's D statistic. For Tajima's D, the chosen bin size was larger than the max contig length; this way, we were able to obtain the statistics for each contig (and not for a given window size). For Tajima's D value all SNPs present in the contig were used, however for the calculation of F_{ST} and all other following analyses one random SNP was evaluated for each contig. To further explore the population structure without making any assumptions about populations we ran a genotype-based PCA using *SNPrelate* (Zheng et al. 2012).

Admixture analysis were ran using the *LEA* R package (Frichot & François 2015). To estimate individual ancestries with *LEA*, vcf files were first converted to plink using VCFtools (--plink option) and to ped format using *PLINK* 1.9 (Chang et al. 2015).

Range expansion

To test the hypothesis of a range expansion in contrast with a simple equilibrium in an isolation-by-distance model, we calculated the directionality index (ψ), a measure based on the difference in derived allele frequencies for biallelic SNPs between two populations (Peter & Slatkin 2013). During an expansion, in particular through a series of founder events, derived alleles have higher probabilities of increasing its frequency in the new population as a consequence of more genetic drift, once those driven to extinction are excluded. This same index, coupled with time difference of arrival (*TDOA*) method (Gunnarsson & Gustafsson 2003), can be used to infer the most likely location of the origin of an expansion for a set of populations. The approach has been used in several other studies (e.g Anderson et al. 2019; Pedersen et al. 2018; Potter et al. 2016; Marques et al. 2017). We applied the directionality index to various subsets of data, namely: (1) including all populations; (2) only Eastern Patagonia populations with and without Tierra del Fuego; (3) only Western Patagonia population without population 23 (Tongoy, Coquimbo); (4) only Tierra del Fuego populations; and (5) Tierra del Fuego populations with neighboring continental population 19 and 20.

We estimated ψ using the script available in *X-Origin* (He et al. 2017) and ran *TDOA* method using the range expansion package available from github.com/BenjaminPeter/rangeexpansion. The sample of *A. lanosus* was used as an outgroup for determining the derived state in each SNP.

Population Divergence Times

Population divergence times were estimated using *SNAPP* (Bryant et al. 2012), a multi-coalescent phylogenetic method for biallelic markers, such as SNPs. This software is able to analyze thousands of SNPs, since it integrates over all the gene tree and thus bypassing the estimation of individual gene trees for each SNPs during the *MCMC* process. This software has been used extensively for the estimation of species trees, as well for population trees (e.g Spalink et al. 2019). Recently, *SNAPP* was used with calibrated nodes to estimate divergence times assuming constant mutation rate and population sizes (Stange et al. 2018). We ran *SNAPP*, implemented in *BEAST* (Bouckaert et al. 2014), using the approach of Stange et al. 2018 to estimate node ages. In this analysis, we used at least one non-admixed individual per inferred population (see Results), 7 outgroup individuals from 2 species, *Abrothrix hirta* and *Abrothrix lanosa*, and one random SNP per each of 1,000 loci. We used one sample per population since the accuracy of the tree reconstruction and divergence times estimations are virtually the same using one or more samples per species, considerably reducing running times (Stange et al. 2018). For the basal node of the tree we assumed a node age of 3.24 MY (node age estimated in Cañón et al. 2010) with a lognormal distribution with a variance of 0.1. We ran 1 million iterations and checked for convergence using Tracer (ESS > 200) (Rambaut et al. 2018), after a burn-in of 100.000 iterations. Tree output was produced using *FigTree* (Hancock et al. 2014).

Population demographic inference

To investigate the demographic history of *A. olivacea* Patagonian populations, we used *Stairway plot* (Liu & Fu 2015), a SFS-based analysis. We analyzed the Patagonian population clusters as defined by the admixture analyses, which clustered all Eastern Patagonia populations (Continental and Tierra del Fuego) and retained each individual forest population separately (pop. 1, pop. 4, pop. 9, pop. 11 and pop. 5). For Eastern Patagonia population only individuals inferred to be of >90% local ancestry were used, and two significantly admixed (see Results) individuals from population 5 were discarded. Demographic inference was based on a folded SFS and including only SNPs per locus with at least 6x coverage in all the individuals of each putative cluster. Absolute times and effective population sizes were estimated using a mutation rate of 5.5e-9 per site per year (from *Mus musculus* (Uchimura et al. 2015)) and a generation time of 1.61 years (Pacifici et al. 2013)

RESULTS

Target Enrichment for *A. olivacea* and outgroups species.

Across the 8499 targeted exons, our samples presented an average of 15X coverage and 7339 (86%) of the exons presented at least one SNP which could be positively genotyped in each of 172 samples (not allowing missing genotypes), indicating a high effectiveness of our exon capture protocol. Most samples presented also high specificity, with more than 80% reads mapping to targeted sequences. Thirty nine samples (n= 39) presented a lower performance during the capture (42% of aligned read); however, they presented an acceptable average coverage of 9X. This is a batch effect, as these samples corresponded to the first set of samples, after which the protocol was adjusted. Samples from different populations were evenly distributed across batches to minimize the impact of batch effects.

Before our conservative filtering, the number of raw SNPs was 222,884 for the *A. olivacea* samples (n=172), 155,403 for the outgroup individuals (n total = 7) and 167,818 for *SNAPP* analysis (n =16). After filtering, not allowing missing genotypes, we identified 59,359, 130,878 and 64,457 high quality SNPs for the 172 individuals, outgroup individuals and *SNAPP* analysis, respectively.

Population Genetic Structure

PCA clustering of individual genotypes for all samples showed three major groups: 1) Northwest Patagonian populations; 2) population 23 (Tongoy, Coquimbo), in central Chile; and 3) Eastern Patagonian populations (Figure 1A). Pairwise values of the fixation index F_{ST} ranged from 0.16 to 0.3 (Figure 1C, 1D and TableS2). When only Patagonian samples were analyzed, thus excluding Tongoy, individuals from population 5 formed a distinct cluster (except for two individuals that showed signals of admixture), whereas the remaining populations were split into Northwest and Eastern Patagonian clusters. Northwestern populations conformed a less compact cluster compared to the Eastern Patagonian population, suggesting greater genetic divergence and isolation within that cluster. However, F_{ST} values between population 8 and 11 (most differentiated population within the Northwest Patagonia) exhibit a F_{ST} of 0.086 (Table S2) lower than, for example, Fueguian population and Eastern Patagonia population (F_{ST} ranging from 0.09 to 0.17 Table S2). PCA of Eastern Patagonia individuals show roughly a linear organization from north to south, where individual from Tierra del Fuego are preferentially grouped (with two exceptions) (Figure 1E).

Using *LEA* analysis, the most likely number of ancestral populations across our all samples was estimated to be 7. The 7 ancestral populations were those associated with: 1) Eastern Patagonia (green cluster); 2) Tierra del Fuego populations plus individuals from neighbouring continental populations 19 and 20 (yellow cluster); 3) Population 5 (red cluster); 4) Forest populations 9 and 11 (black cluster); 5) Tongoy (blue cluster); 6) Forest population 8 (skyblue cluster); and 7) Forest population 1 and 4 (orange cluster) (Figure 1B). From the admixture analyses, it is clear that virtually all the populations present medium to high levels of admixture. For example, two individuals from population 5 were inferred to present high amount of admixture from Eastern Patagonia (in fact, one of these clustered in this group), whereas the remaining formed a distinct cluster in the PCA analyses. More generally, putatively admixed individuals showed indications various degrees of shared ancestry with geographically adjacent populations and clusters.

Divergence times between putative phylogroups /ancestral populations.

To estimate the genetic relationship and divergence times between putative populations (as defined by the admixture analysis) we ran *SNAPP* using one a non-admixed individual per population (Figure 2). Divergence time between the Eastern and Northwestern group was estimated to have occurred between 0.81 - 1.36 MYA, whereas the split of Tierra del Fuego from continental populations was timed between 55 KYA – 230 KYA. On the other hand, within the Northwestern group the populations, started to divergence 410 - 779 KYA and the last split occurred 200 - 510 KYA. In agreement with previous studies, the divergence estimation times between the two outgroups species, *A. lanosa* and *A. hirta*, was between 1.6 and 2.69 MYA (see "*longipilis* group" in Cañón et al. 2014).

Range Expansion

Population genetic analyses revealed significant relationship between heterozygosity levels and latitude, which broadly decrease southwards from the northwest (Figure 3A). Tajima's D values averaged across loci are always negative and also show a largely concordant gradient, with northern populations exhibiting higher values than southern ones (Figure 3B, Table S3).

Analysis of the directionality index favored a range expansion over an isolation by distance process in three subsets of data (p values < 0.05), namely: 1) when all the populations are considered

(Figure 3) and 2) when the analysis was restricted to the Eastern Patagonia population (with or without Tierra del Fuego) (Table S4). In the remaining tested scenarios, the analysis of the directionality index failed to reject the null hypothesis (Table S4).

When all the populations are used, the directionality index support an origin located in Chiloé. When only Argentina Patagonia population are considered, the directionality index points towards an origin near 44.1 S latitude in the Eastern Andean slopes, just between population 11 and 5 (Figure 3D). In general, a putative origin in or near the Valdivian rainforest is broadly supported by these analyses.

Population demographic history of populations.

Population size inference revealed a history of population expansion in all the putative populations (Figure 4). The timing of expansion, however, varies across geography, with populations to the north and west inferred to have expanded earlier, and the onset of expansion gradually becoming more recent as latitude increases. Coastal northwestern populations, 1 and 4, exhibited an expansion around 40 KYA and a recent population decline. Northwestern population 9 and 11, presented a more recent population expansion (22 KYA). Magellanic forest population 5, exhibited a rapid population growth close to 22 KYA and a recent population decline after a steady population size. Continental Eastern Patagonia and Tierra del Fuego populations presented the most recent population expansion, around 16 KYA and 9 KYA respectively. Tierra del Fuego populations also showed a population explosive growth. The effective population size varied significantly between populations: population 5 (Cochrane, Chile) presented the lower population size, whereas Continental Easterna Patagonia population presented the higher size. The final effective population size estimated for Tierra del Fuego was higher than all Patagonian forest populations with the exception of population 4.

Discussion

General phylogeographic patterns of *A. olivacea* in Patagonia

One of the main goals of this project was to evaluate the hypothesis, based on mtDNA sequences (Lessa et al. 2010; Abud 2011), that the Patagonian olive mouse *A. olivacea* occupied two refugia (one in the Valdivian forest and one in Tierra del Fuego) during the last glacial maximum and

subsequently colonized the rest of its distribution in the Patagonian region from these two sources. More generally, we sought to contribute to the understating of the historical responses of an ecologically flexible species to the glacial cycles in relation to changes in both open (i.e., the Patagonian steppe and grasslands) and forested (Valdivian and Magellanic forests) environments and evaluate additional refugia. In order to understand the historical demography, direction of colonization and origin of the expansion we applied an exon capture protocol, circumventing the need for a high-quality reference genome. We designed the capture of more than 8000 exons from protein coding sequences (8 Mb) and applied it to 172 individuals of *A. olivacea* and four and three individuals of *A. lanosa* and *A. hirta*, respectively. Our sampling consisted of 21 populations distributed across all steppe, grasslands and forests in the Patagonian-Fuegian region, with 17 reference samples external to the region in central Chile.

Our analyses show that *A. olivacea* is clearly geographically structured into three genetics major units in Patagonia (Fig. 1D; Fig. 2): a northwestern group, largely associated with the Valdivian and Magellanic forests, an eastern group, including Tierra del Fuego, associated with the steppe and southern grasslands, and most individuals of population 5 in the southwestern Magellanic forests. At a broad geographical scale, it is clear that the phylogeography of the species was highly influenced by the Andean Mountain range, separating western populations of the forested Andean foothills from those that occupy open steppe and grasslands to the east. In Tierra del Fuego, at least to the extent our sampling is concerned, all populations, including those in forested environments, are connected to the eastern group in the continent. Interestingly, previous studies based on mitochondrial sequences indicated all continental populations belonged to one phylogroup, whereas those in Tierra del Fuego constituted a separate phylogeographic unit (Lessa et al. 2010; Abud 2011).

Generally, the historical divergence of *A. olivacea* across the region took place after the retreat of the Great Patagonian glaciation (~1 MYA), though some of the most recent events took place in the late Pleistocene. Estimations of the time of divergence between the two major Patagonian population clusters uncovered by our SNP data suggest that the divergence took place about 0.81 - 1.36 MYA. Meanwhile, the split between Tierra del Fuego and and eastern continental Patagonia occurred much more recently, 55 - 230 KYA, possibly during the last glacial period.

Examination of the population tree obtained with *SNAPP* (Fig. 2), also indicates that there is far greater divergence and population structure in the cluster corresponding to populations in the forest. Meanwhile, populations in the continental steppe and grasslands, as well as those in Tierra del

Fuego, which belong to the same cluster, are far less divergent. Among populations linked to the forests, the split between population 5 and the remaining was estimated between 0.6 - 1.1 MYA, explaining why this population, though more related to the remaining forest population, appears as a distinct genetic unit in the PCA analysis when only the Patagonian population are considered (Figure 1D). In general, all the Patagonian population showed low to moderate genetic divergence with important levels of introgression level that are largely consistent with their geographical location.

Our data support previous results in that rodent species in the Patagonian region are away from drift-mutation equilibrium, reflecting recent demographic expansions (Lessa et al. 2010). As a general trend, we found that Tajima's D value and heterozygosity decreased almost linearly from north to south into Tierra del Fuego, where the populations exhibit the lowest heterozygosity and Tajima's D values. As expected for a recent demographic expansion Tajima's D values were always biased towards negative value. Although negative Tajima's D values also could be a consequence of purifying selection (e.g., Carneiro et al. 2012; Jackson et al. 2015) the latitudinal trend of Tajima's D values across Patagonia are much more likely reflecting a distinct history of population expansions from Northwestern Patagonia to Tierra del Fuego than different selection regimens. The gradient of southward decline of heterozygosity results also reinforces this fact: more recently funded populations are expected to have lower genetic diversity.

Western population cluster

As noted, one of the major population clusters uncovered both by PCA and the population tree corresponds to populations located in the continental Valdivian and Magellanic forests. These populations stand in contrast to the remaining ones as follows: 1) they presented higher heterozygosity levels and higher (i.e., negative, but closer to zero Tajima's D values than populations in the steppe; 2) they are more structured geographically (i.e., deeper divergences and greater geographical subdivision); and 3) they include or are more closely related to the source of colonization, identified by an analysis of directionality in or around Chiloé, near the hypothesized Valdivian Forest refugium (Smith et al. 2001).

Although all populations in the forest cluster show signs of expansion, those signs are more dramatic in populations 9 and 11, which are part of this cluster but are geographically transitional to the Patagonian steppe. Indeed, these populations show a relatively recent onset of expansion

(around 20 KYA) and a marked subsequent increase (Fig. 4A). Other forest populations are inferred to have expanded beginning some 40 KYA. These estimates are more recent than those obtained from mtDNA data for this (Lessa et al. 2010) and other rodent species in the region (e.g. *Abrothrix manni* (Valdez & D'Elía 2018)).

Current estimates place the last glacial maximum in Patagonia at about 20 KYA (Glasser et al. 2008). Thus, our estimates of the timing of the onset of demographic expansion in forest populations are somewhat before or right at the peak of the last glacial phase. Given the uncertainties associated with these estimates, they are consistent with largely post-glacial expansions.

Stairway plots also show modest decreases of population sizes closer to the present (e.g., in populations 1, 4 and 5, Fig. 4 A,B), a pattern that has also been uncovered in other species (Valdez & D'Elía 2018). This observation must be considered provisional, as the stairway plot may not be reliable for very recent (or very old) events.

It is also noteworthy that estimates of divergence between populations in the forest predate these estimates of expansion, strongly suggesting that geographical structure in that biome preceded the last glacial maximum. The divergence time between Western and Eastern group is estimated approx, between 0.81 - 1.36 MYA. Subsequent branching events are much older in the forest than the steppe cluster, beginning 0.61 - 1.1 MYA. These divergence times between the forest populations suggest that the population splitting already started well before the last quaternary glacial maximum, possibly during the GPG (1 MYA) or in the following glacial period (Post-GPG 1 : >710 KYA (Coronato et al. 2004; Kaplan et al. 2005)). Since forests were heavily glaciated during glacial periods, these results suggest the possibility of an additional peripheral glaciar refugium close to Cochrane, where population 5 was sampled. This population was the most divergent within the western cluster and is placed very close to the refugia proposed for several plant species (Sérsic et al. 2011). More generally, the deeper population relationships between forest populations in comparison to those from the Eastern Patagonia, may be the result of the glacier dynamics on the region, promoting not only structured refugia, as already noted, but cryptic refugia as well.

Eastern Continental Patagonia

The second major cluster of populations, uncovered by both PCA and the population tree, corresponds to the Patagonian steppe and grasslands further south, and also includes all populations from Tierra del Fuego, regardless of their habitat (which includes both forest and grasslands). Heterozygosity of continental eastern Patagonia populations is intermediate between higher values found in the forest cluster and the lower levels in Tierra del Fuego.

The directionality index further supports a southward range expansion across continental eastern Patagonia, rejecting an isolation by distance process in this region (Table S4). This result holds with or without incorporating the forest cluster or the populations from Tierra del Fuego, although, as is to be expected, the signal of colonization is stronger if all populations are included.

Estimation of the time of divergence between Eastern Continental Patagonia and Tierra del Fuego (55 - 230 KYA) are more recent than those estimated using mitochondrial data (384 -678 KYA (Abud 2011)). This is not surprising, as the mtDNA data uncovered a distinct phylogeographic clade, whereas our nuclear data clearly place populations in Tierra del Fuego in the eastern clade. SNP data result in a low to moderate fixation index between the Fuegian population and the remaining ones (F_{ST} values between 0.06 and 0.17) supporting the idea of a recent divergence.

The times of divergence of these populations in the continental steppe and grasslands are more recent than those in the forest cluster and are estimated to have begun about 0.8 KYA (compared to 0.61-1.1 KYA in the forest cluster). Along the same lines, population demographic inference using *stairway plot* points towards a possible post LGM expansion of the eastern continental populations (around 20 KYA). As noted above, values this close to the present are only found in some populations in the forest cluster that are located at or near transitional, ecotonal areas between forest and steppe.

Earlier estimates of expansion were obtained using mtDNA data, and separately for each of the major phylogeographic groups uncovered by those studies, i.e., continental areas and Tierra del Fuego. The resulting estimates were 90 - 150 KYA (Lessa et al. 2010) and 75 – 150 KYA (Abud 2011). As noted, our estimates are much more recent (~ 20 KYA), in line with onsets of expansion at or near the LGM. The multilocus nature of SNP data makes them more reliable and allows

reasonable estimates for individual populations, and the contrasts with mtDNA estimates likely reflect both of these factors.

Tierra del Fuego

Populations in Tierra del Fuego presented the lowest heterozygosity and Tajima's D values and the most recent estimates of divergence and population expansion times. These contrasts with continental populations are broadly in agreement with the previous mitochondrial studies (Abud 2011; Lessa et al. 2010). However, our estimations of population expansion are clearly posterior to the LGM, close to 10 KYA, and more abrupt compared to those found with mtDNA data. For comparison, the previous two studies based on mtDNA estimated the expansion in Tierra del Fuego around 50 KYA (Abud 2011) and between 21 - 100 KYA (Lessa et al. 2010).

In contrast with the previous mitochondrial studies, our results do not support the presence of a distinct phylogroup in Tierra del Fuego. Our fixation index values suggest a low to moderate genetic divergence and *SNAPP* divergence times point out the possibility of recent divergent times (55 KYA) between Tierra del Fuego and neighboring continental populations. Moreover, as have been previously noted, analyses based on the directionality index clearly indicated a colonization from the continental populations into the archipelago. During the last glacial phase, and up to 10,000 YA, the continental shelf exposure linked Tierra del Fuego to the continent, allowing southward colonization.

Taking together, our results do not support a Tierra del Fuego refugium for many reasons. First, the directionality index supports a colonization from the continent, rather than from a separate source, explaining the gradient of heterozygosity as part of a colonization process from north to south. Second, the population tree clearly includes Tierra del Fuego within a larger, eastern continental cluster. Third, several continental individuals from populations 19 and all individuals from population 20 are clustered within Tierra del Fuego individuals, suggesting that the genetic identity of populations of Tierra del Fuego is shared with neighboring continental populations. In regard to this last point, many rodents, as well as other species, present a phylogeographic break further north in continental Patagonia and a genetic variation between the continent and Tierra del Fuego (Sérsic et al. 2011).

In sum, our results discard an *in situ* or otherwise distinct differentiation in Tierra del Fuego but confirm a post glacial population expansion in these populations.

Acknowledgements

We thank Ke Bi for helping with the exon-capture design. We thank Eileen Lacey for sponsoring a stay of FG in the Museum of Vertebrate Zoology, University of California Berkeley, to carry out the exome capture experiments, and Lydia Smith for key assistance with laboratory work. We thank Guillermo D' Elía for providing samples of *Abrothrix olivacea*. We are grateful to Matias Feijoo, Guillermo D'Elía, Lourdes Valdez and Daniel Naya for participating in the field work. This work was supported by grants from Agencia Nacional de Investigación e Innovación (ANII FCE 2014 103508). Facundo Giorello is supported by graduate fellowships ANII.

Supplementary information

Table S1. Samples information.
Table S2. Pairwise Fst values between populations.
Table S3. Tajima's D value and heterozygosity per population.
Table S4. Range expansion analyses in different geographical region.
Table S5. Pairwise directionality index between populations.

Author contributions

E.P.L. designed and supervised the project and participated in the fieldwork. F.M.G carried out the exon-capture experiments and bioinformatics analyses. F.M.G. and E.P.L wrote the manuscript. Both authors approved the final manuscript.

References

Abud, C., 2011. *Variación genética y estructura filogeográfica de Abrothrix olivaceus en la Patagonia argentina y el sur chileno (Master thesis)*. Universidad de la República.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215, pp.403–410.

Anderson, B.M. et al., 2019. Recent range expansion in Australian hummock grasses (Triodia) inferred using genotyping-by-sequencing. *AoB PLANTS*.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120.

Bouckaert, R. et al., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*.

Bryant, D. et al., 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), pp.1917–1932.

Cañón, C. et al., 2014. A multilocus perspective on the phylogenetic relationships and diversification of rodents of the tribe Abrotrichini (Cricetidae: Sigmodontinae). *Zoologica Scripta*.

Cañón, C. et al., 2010. Phylogeography of Loxodontomys micropus with comments on the alpha taxonomy of Loxodontomys (Cricetidae: Sigmodontinae). *Journal of Mammalogy*.

Carneiro, M. et al., 2012. Evidence for widespread positive and purifying selection across the european rabbit (oryctolagus cuniculus) genome. *Molecular Biology and Evolution*, 29(7), pp.1837–1849.

Chang, C.C. et al., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), p.7.

Coronato, A., Martínez, O. & Rabassa, J., 2004. Glaciations in Argentine Patagonia, southern South America. *Developments in Quaternary Science*.

D'Elía, G. & Pardiñas, U.F.J., 2015. Subfamily Sigmodontinae Wagner, 1843. In J. L. Patton, U. F. J. Pardiñas, & G. D'Elía, eds. *Mammals of South America. Volume 2. Rodents.* Chicago: University of Chicago Press, p. 1384.

Danecek, P. et al., 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15), pp.2156–2158.

Frichot, E. & François, O., 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*.

Giorello, F.M. et al., 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (Abrothrix olivacea). *Molecular Ecology*.

Glasser, N.F. et al., 2008. The glacial geomorphology and Pleistocene history of South America between 38°S and 56°S. *Quaternary Science Reviews*.

Gunnarsson, F. & Gustafsson, F., 2003. Control theory aspects of power control in UMTS. *Control Engineering Practice*.

Hancock, J.M., Zvelebil, M.J. & Cummings, M.P., 2014. FigTree. In *Dictionary of Bioinformatics and Computational Biology*.

He, Q., Prado, J.R. & Knowles, L.L., 2017. Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program x-origin. *Molecular Ecology*.

Heusser, C.J. et al., 1996. Full glacial — late glacial palaeoclimate of the Southern Andes: evidence from pollen, beetle and glacial records. *Journal of Quaternary Science*.

Hulton, N.R.J. et al., 2002. The Last Glacial Maximum and deglaciation in southern South America. *Quaternary Science Reviews*.

Jackson, B.C., Campos, J.L. & Zeng, K., 2015. The effects of purifying selection on patterns of genetic differentiation between Drosophila melanogaster populations. *Heredity*, 114(2), pp.163–74.

Kaplan, M.R. et al., 2005. Cosmogenic nuclide chronology of pre-last glacial maximum moraines at Lago Buenos Aires, 46°S, Argentina. *Quaternary Research*.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, pp.357–359.

Lessa, E.P., D'Elía, G. & Pardiñas, U.F.J., 2010. Genetic footprints of late Quaternary climate change in the diversity of Patagonian-Fueguian rodents. *Molecular Ecology*, 19(15), pp.3031–3037.

Liu, X. & Fu, Y.X., 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics*.

Mann, G., 1978. Los pequeños mamíferos de Chile. *Gayana, Zoologia*, 40, pp.1–342.

Markgraf, V., 1993. Paleoenvironments and paleoclimates in Tierra del Fuego and southernmost Patagonia, South America. *Palaeogeography, Palaeoclimatology, Palaeoecology*.

Marques, J.P. et al., 2017. Range expansion underlies historical introgressive hybridization in the Iberian hare. *Scientific Reports*.

McCulloch, R.D. et al., 2000. Climatic inferences from glacial and palaeoecological evidence at the last glacial termination, southern South America. *Journal of Quaternary Science*.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20, pp.1297–303.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*.

Miller, S.A., Dykes, D.D. & Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*.

Nishimura, D., 2002. RepeatMasker. *Biotech Software & Internet Report*.

Pacifici, M. et al., 2013. Generation length for mammals. *Nature Conservation*.

Palma, R.E. et al., 2012. Glaciation effects on the phylogeographic structure of Oligoryzomys longicaudatus (rodentia: Sigmodontinae) in the southern andes. *PLoS ONE*.

Pedersen, C.E.T. et al., 2018. A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology and Evolution*.

Pfanzelt, S., Albach, D.C. & von Hagen, K.B., 2017. Tabula rasa in the Patagonian Channels? The phylogeography of Oreobolus obtusangulus (Cyperaceae). *Molecular Ecology*.

Ponce, J.F. et al., 2011. Palaeogeographical evolution of the Atlantic coast of Pampa and Patagonia from the last glacial maximum to the Middle Holocene. *Biological Journal of the Linnean Society*.

Potter, S. et al., 2016. Phylogenomics at the tips: Inferring lineages and their demographic history in a tropical lizard, Carlia amax. *Molecular Ecology*.

Rabassa, J., Coronato, A. & Martínez, O., 2011. Late Cenozoic glaciations in Patagonia and Tierra del Fuego: An updated review. *Biological Journal of the Linnean Society*.

Rambaut, A. et al., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*.

Sérsic, A.N. et al., 2011. Emerging phylogeographical patterns of plants and terrestrial vertebrates from Patagonia. *Biological Journal of the Linnean Society*.

Shafer, A.B.A. et al., 2010. Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, 19(21), pp.4589–4621.

Sikes, R.S., 2016. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy*.

Smith, M.F., Kelt, D.A. & Patton, J.L., 2001. Testing models of diversification in mice in the Abrothrix olivaceus/xanthorhinus complex in Chile and Argentina. *Molecular Ecology*, 10(2), pp.397–405.

Spalink, D., MacKay, R. & Sytsma, K.J., 2019. Phylogeography, population genetics and distribution modelling reveal vulnerability of Scirpus longii (Cyperaceae) and the Atlantic Coastal Plain Flora to climate change . *Molecular Ecology*, pp.0–3.

Stange, M. et al., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports miocene closure of the Panamanian Isthmus. *Systematic Biology*.

Turchetto-Zolet, A.C. et al., 2013. Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, 22(5), pp.1193–1213.

Uchimura, A. et al., 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Research*.

Valdez, L. & D'Elía, G., 2018. Local persistence of Mann's soft-haired mouse Abrothrix manni (Rodentia, Sigmodontinae) during Quaternary glaciations in southern Chile . *PeerJ*.

Yates, A. et al., 2016. Ensembl 2016. *Nucleic Acids Research*.

Zheng, X. et al., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*.

Figures

Figure 1.

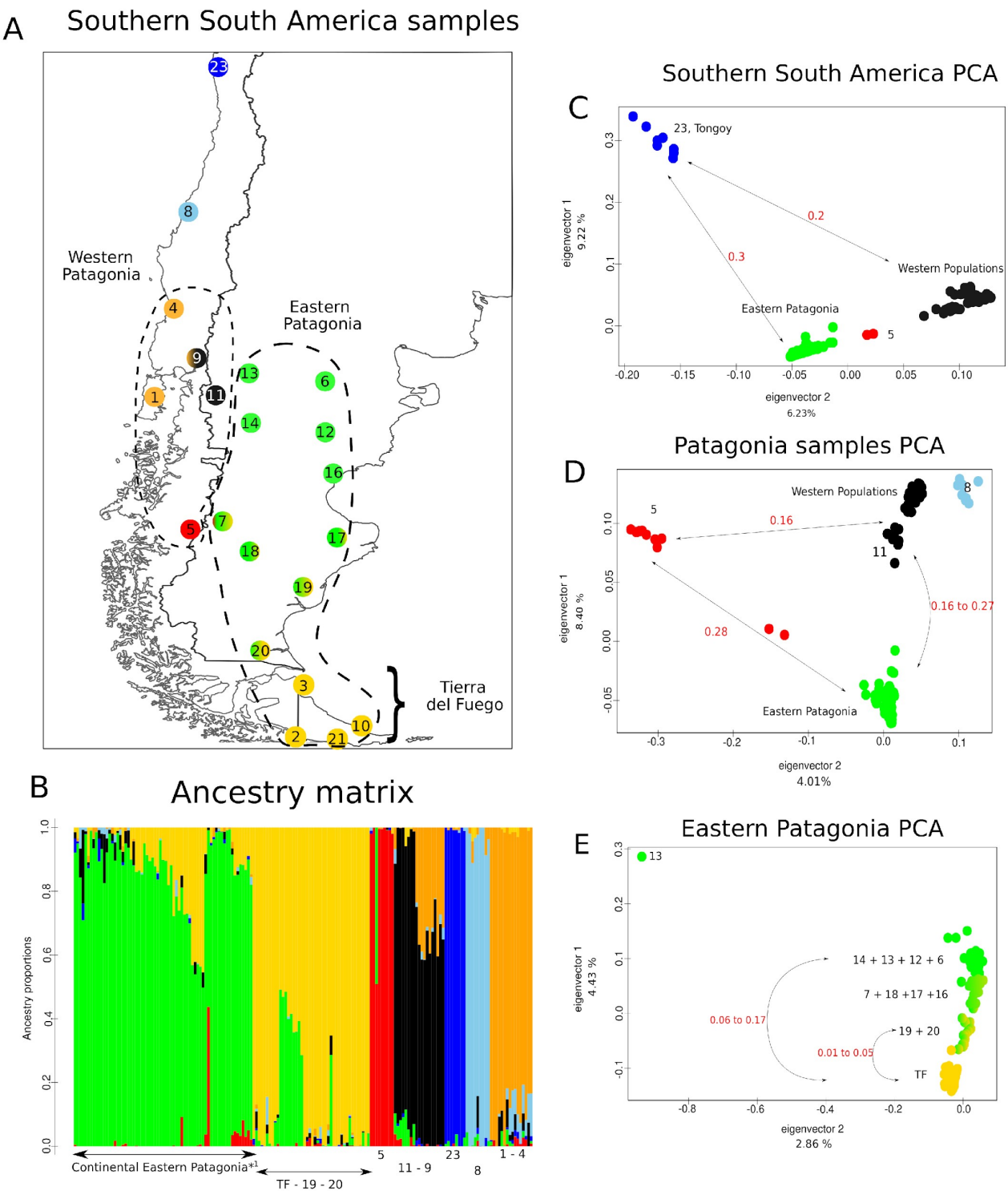
A
Southern South America samples
Western Patagonia
Eastern Patagonia
Tierra del Fuego
23
8
4
9
13
6
1
11
14
12
16
7
5
18
17
19
20
3
10
2
21
B
Ancestry matrix
Ancestry proportions
Continental Eastern Patagonia*1
TF - 19 - 20
5
23
1 - 4
11 - 9
8
C
Southern South America PCA
23, Tongoy
0.2
0.3
Western Populations
Eastern Patagonia
5
eigenvector 1
9.22 %
eigenvector 2
6.23%
D
Patagonia samples PCA
Western Populations
8
5
0.16
11
0.16 to 0.27
0.28
Eastern Patagonia
eigenvector 1
8.40 %
eigenvector 2
4.01%
E
Eastern Patagonia PCA
13
14 + 13 + 12 + 6
7 + 18 +17 +16
0.06 to 0.17
19 + 20
0.01 to 0.05
TF
eigenvector 1
4.43 %
eigenvector 2
2.86 %

Figure 2.

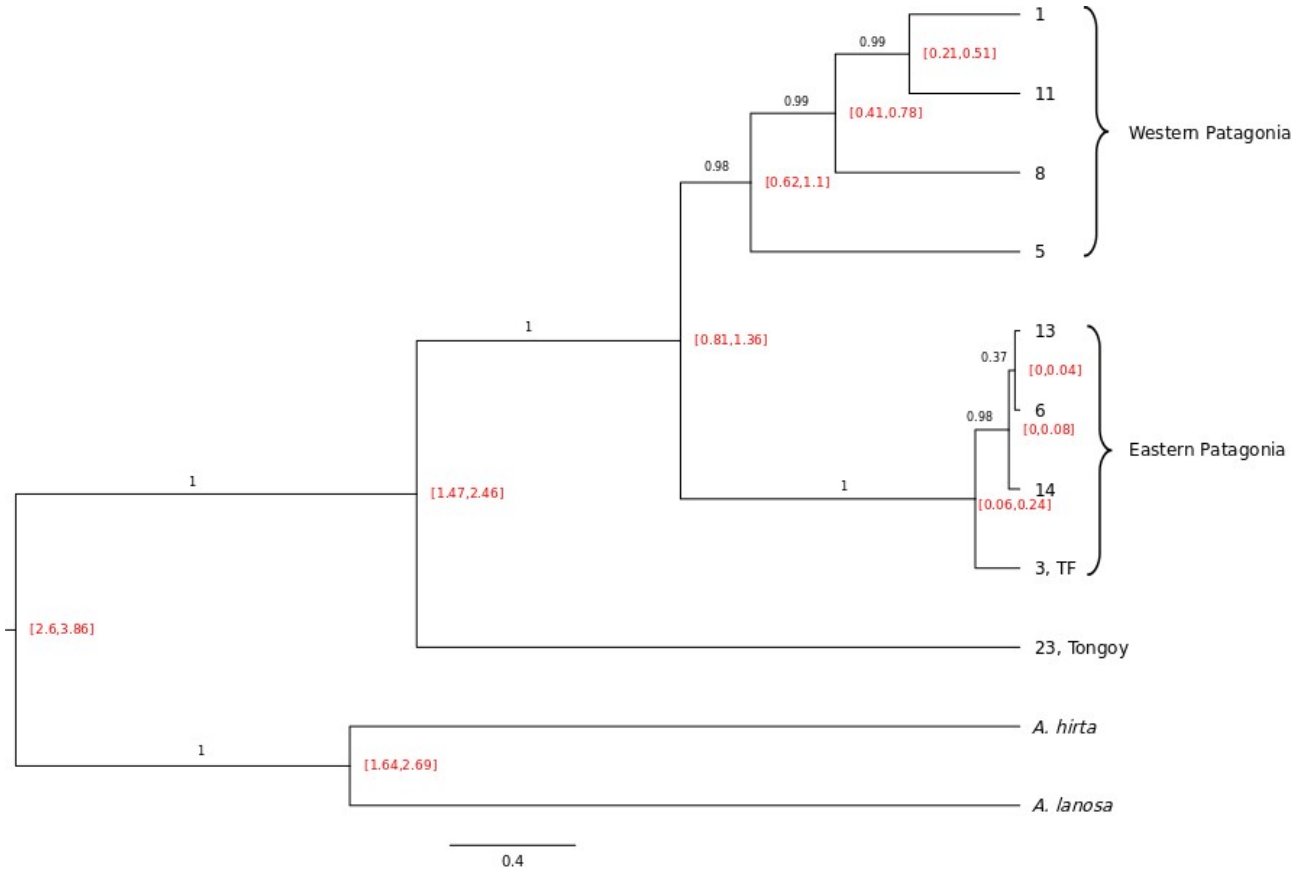
1
11
8
5
Western Patagonia
13
6
14
3, TF
Eastern Patagonia
23, Tongoy
A. hirta
A. lanosa
0.99
[0.21,0.51]
0.99
[0.41,0.78]
0.98
[0.62,1.1]
1
[0.81,1.36]
0.37
[0,0.04]
0.98
[0,0.08]
1
[0.06,0.24]
1
[1.47,2.46]
1
[2.6,3.86]
1
[1.64,2.69]
0.4

Figure 3.

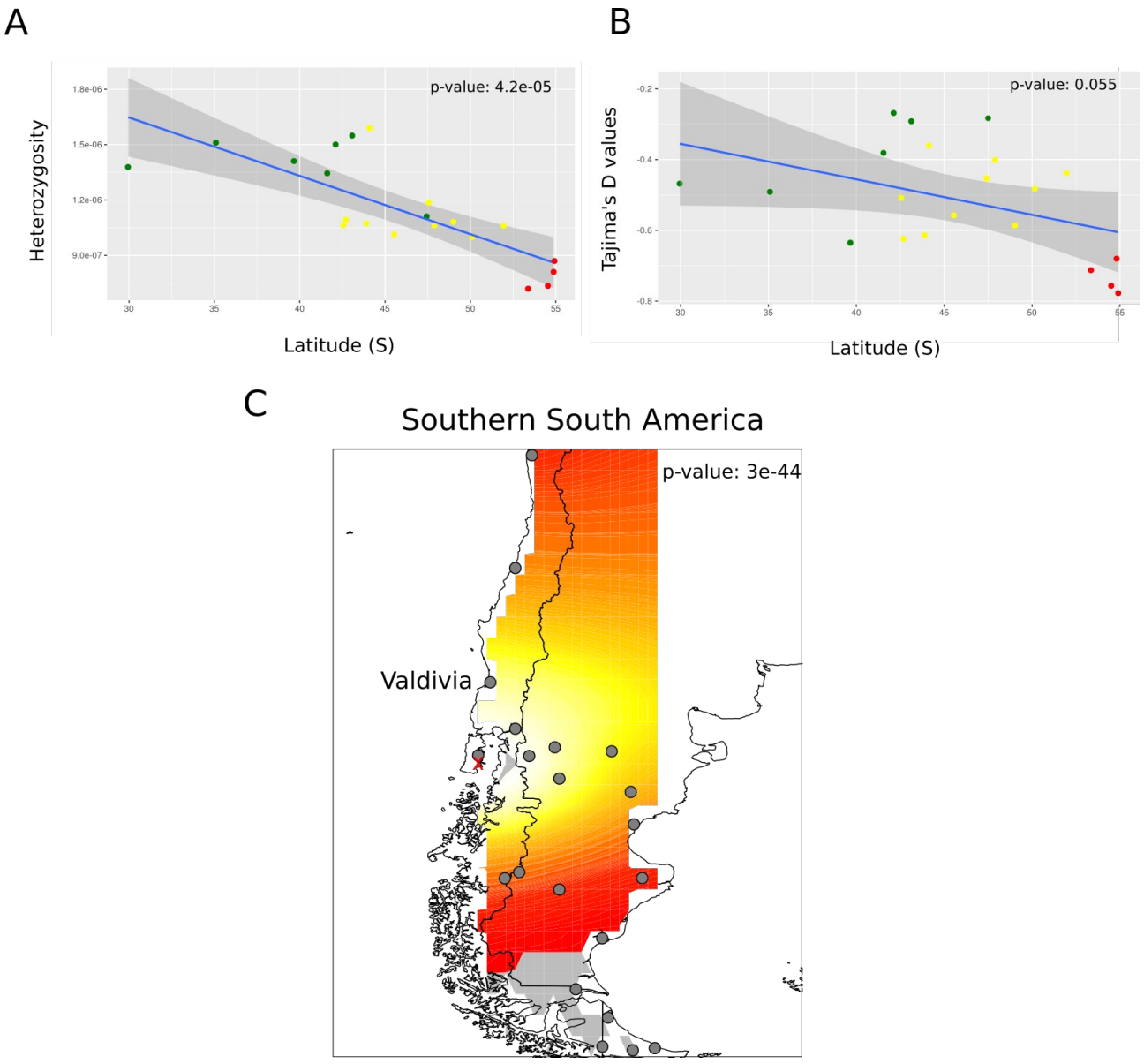
A
B
C
p-value: 4.2e-05
Heterozygosity
Latitude (S)
p-value: 0.055
Tajima's D values
Latitude (S)
Southern South America
p-value: 3e-44
Valdivia

Figure 4.

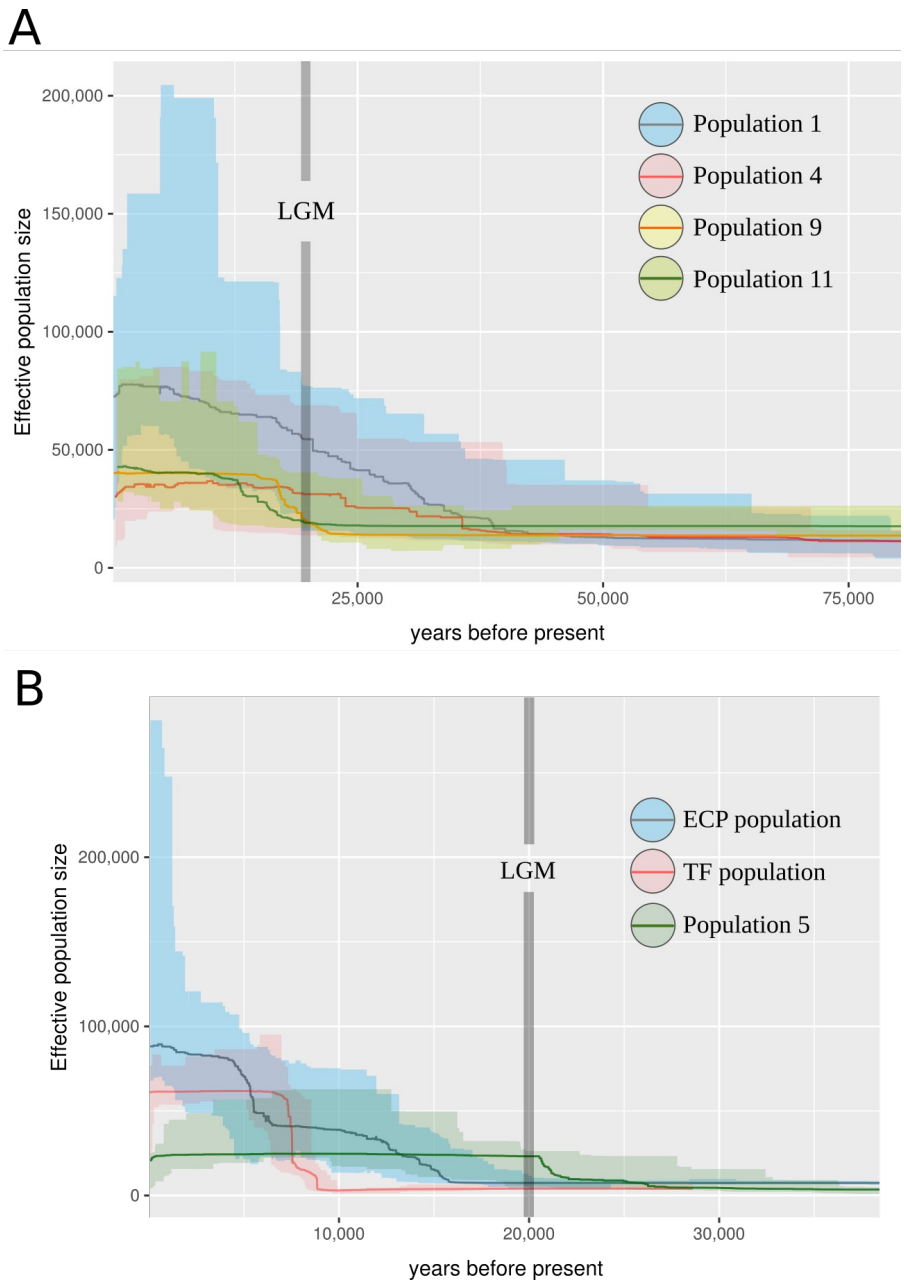

Figure 1. Population genetic analyses of *Abrothrix olivacea* in Southern South America.

A. Geographical sampling across Southern South America (n=172). **B**. Admixture analyses based on 7 ancestral populations. The seven ancestral populations correspond roughly to: (i) Continental Eastern Patagonia; (ii) Tierra del Fuego combined with populations 19 and 20; (iii) population 5; (iv) populations 9 and 11; (v) population 8; (vi) populations 1 and 4; and (vii) population 23. The

Continental Eastern Patagonia cluster also clustered individuals from population 5 (n=1), Tierra del Fuego (n=1) and population 19 (n=5). Levels of admixture between populations are consistent to their geographical proximity. **C**. Principal Component Analyses based on individual genotypes using all samples of *A. olivacea*. **D**. Principal Component Analyses based on individual genotypes excluding population 23 (Coquimbo, Tongoy). **E**. Principal Component Analyses based on individual genotypes of Eastern Patagonia individuals. Numbers in red corresponds to F_{ST} values between identified clusters. In **A**, population regions were coloured following admixture analyses results. *[1] Continental Eastern Patagonia cluster also grouped 2 individuals from TF and one from population five. TF: Tierra del Fuego

Figures 2.

Divergence estimation between the seven population clusters.

Divergences times (95% confidence intervals in red) estimated using SNAPP with a subset of 1000 SNPs. For each node posterior probabilities are shown in black.

Figure 3.

Evaluation of population expansion and origin of colonization.

A. Regression analysis between Heterozygosity and Latitude (Intercept: 6.742e+01***; Slope = -1.884e+07 ***)

B. Regression analysis between Tajima's D value and Latitude (Intercept: 36.523 ***; Slope = -18.022). Green: Western populations. Yellow: Eastern Continental Populations. Red: Tierra del Fuego populations.

C. Range expansion analyses based on the directionality index. The most likely source of colonization is shaded in yellow, with red and grey indicating successively least likely sources. The X points to the inferred origin by TDOA analysis.The sampling locations are represented by circles.

Figure 4. Stairway plots, representing demographic history of each of the seven *A. olivacea* clusters. as defined by Admixture, based on a mutation rate of 5.5e-9 per site per year and a generation time of 1.61 years.

A. Cluster combining populations 1, 4, 9 and 11. **B.** Populations 9, 11 and 5. Shadows represent 95% confidence intervals.

Table 1. Samples information

Population	Individual	Coverage	Latitude	Longitude	% read alignment
1	GD1148	8	-43.08	-73.91	44.69
1	GD1149	19	-43.08	-73.91	91.75
1	GD1153	22	-43.08	-73.91	91.21
1	GD1156	21	-43.08	-73.91	91.12
1	UWBM79699	20	-41.53	-73.41	92.56
1	UWBM79701	14	-41.53	-73.41	90.94
2	PNG0422	15	-54.82	-68.33	91.61
2	PNG0423	20	-54.82	-68.33	91.39
2	PNG0425	18	-54.82	-68.33	90.17
2	PNG0436	10	-54.82	-68.33	91.92
2	PNG0467	13	-54.82	-68.33	93.25
2	PNG0472	21	-54.82	-68.33	90.24
2	PNG0525	10	-54.82	-68.33	42.56
2	PNG0527	17	-54.82	-68.33	91.37
3	363	7	-53.32	-68.41	36.61
3	382	21	-53.32	-68.41	91.52
3	383	17	-53.32	-68.41	91.94
3	386	22	-53.32	-68.41	93.89
3	388	24	-53.32	-68.41	89.95
3	390	15	-53.32	-68.41	93.44
3	392	18	-53.32	-68.41	90.40
3	PNG0375	20	-53.32	-68.41	91.59
3	PNG0377	14	-53.32	-68.41	93.87
3	PNG0395	9	-53.32	-68.41	40.63
4	GD1403	8	-39.65	-73.19	42.29
4	GD1405	19	-39.65	-73.19	89.27
4	GD1407	22	-39.65	-73.19	88.95
4	GD1411	24	-39.65	-73.19	93.27
4	GD1412	16	-39.65	-73.19	93.77
4	GD1417	8	-39.65	-73.19	41.14
4	GD1426	21	-39.65	-73.19	91.52
4	GD1427	23	-39.65	-73.19	89.76
4	GD1516	22	-39.65	-73.19	91.76
4	GD1517	18	-39.65	-73.19	92.80
5	GD1464	7	-47.50	-72.82	40.47
5	GD1465	22	-47.50	-72.82	91.04
5	GD1469	11	-47.50	-72.82	91.78
5	GD1470	31	-47.50	-72.82	93.02
5	GD1481	15	-47.50	-72.82	92.01
5	GD1482	14	-47.50	-72.82	91.82
5	GD1484	21	-47.50	-72.82	92.11
5	GD1487	22	-47.50	-72.82	91.59
5	GD1489	23	-47.50	-72.82	93.66
5	GD1490	9	-47.50	-72.82	42.29
6	PPA278	20	-42.52	-68.28	92.89
6	PPA280	16	-42.52	-68.28	91.58
6	PPA281	22	-42.52	-68.28	91.38
6	PPA284	26	-42.52	-68.28	93.14
6	PPA313	19	-42.52	-68.28	90.94
6	PPA315	7	-42.52	-68.28	41.23
6	PPA328	24	-42.52	-68.28	91.05
6	PPA329	23	-42.52	-68.28	94.01
7	PPA350	9	-47.42	-71.95	40.29

7	PPA443	21	-47.42	-71.95	91.88
7	PPA444	24	-47.42	-71.95	91.04
7	PPA446	24	-47.42	-71.95	95.05
7	PPA448	18	-47.42	-71.95	93.08
7	PPA452	19	-47.42	-71.95	90.42
7	PPA529	21	-47.42	-71.95	93.50
7	PPA532	8	-47.42	-71.95	41.83
8	GD1725	8	-35.08	-72.17	43.86
8	GD1726	22	-35.08	-72.17	90.00
8	GD1727	17	-35.08	-72.17	90.87
8	GD1728	19	-35.08	-72.17	92.89
8	GD1729	7	-35.08	-72.17	42.03
8	GD1731	15	-35.08	-72.17	93.38
8	GD1753	20	-35.08	-72.17	93.05
8	GD1755	24	-35.08	-72.17	90.44
8	GD1756	20	-35.08	-72.17	91.71
9	1176	7	-41.56	-72.17	42.80
9	1177	20	-41.56	-72.17	89.54
9	1180	23	-41.56	-72.17	92.86
9	1186	22	-41.64	-72.21	92.18
9	1187	16	-41.64	-72.21	93.31
9	1191	21	-41.56	-72.17	90.23
9	GD1172	28	-41.56	-72.17	90.90
9	GD1174	19	-41.56	-72.17	92.12
9	GD1175	22	-41.56	-72.17	91.64
9	GD1184	9	-41.64	-72.21	42.65
9	GD1185	7	-41.64	-72.21	40.06
10	401	17	-54.49	-66.41	92.51
10	402	20	-54.49	-66.41	91.39
10	408	21	-54.49	-66.41	94.45
10	412	13	-54.61	-67.54	91.87
10	413	8	-54.61	-67.54	36.69
10	PNG0403	14	-54.49	-66.41	91.56
10	PNG0404	14	-54.49	-66.41	91.84
10	PNG0405	20	-54.49	-66.41	90.01
10	PNG0406	6	-54.49	-66.41	35.03
10	PNG0407	5	-54.49	-66.41	49.58
11	PNG0268	7	-42.09	-71.62	37.39
11	PNG0278	8	-42.89	-71.57	38.90
11	PNG0282	20	-42.89	-71.57	90.93
11	PNG0284	22	-42.89	-71.57	93.14
11	PNG0287	4	-42.89	-71.57	82.30
11	PNG0288	15	-42.89	-71.57	93.72
11	PNG0289	18	-42.89	-71.57	92.23
11	PNG0304	22	-42.89	-71.57	92.52
12	353	19	-44.10	-67.99	92.36
12	PNG0343	21	-44.10	-67.99	92.44
12	PNG0351	9	-44.10	-67.99	36.88
13	239	7	-42.67	-70.09	40.72
13	241	18	-42.67	-70.09	90.69
13	1390	10	-42.06	-71.16	91.84
13	PNG0238	19	-42.67	-70.09	92.62
13	PNG0244	8	-42.67	-70.09	40.06
13	PNG0247	18	-42.67	-70.09	92.31
13	PNG0248	6	-42.67	-70.09	41.10
13	PNG1376	19	-42.33	-70.55	90.21

13	PNG1389	15	-42.06	-71.16	93.34
14	PPA26	9	-43.87	-70.73	42.15
14	PPA34	22	-43.87	-70.73	91.86
14	PPA4	24	-43.87	-70.73	91.90
14	PPA41	23	-43.87	-70.73	94.18
14	PPA44	17	-43.87	-70.73	91.43
14	PPA62	18	-43.87	-70.73	92.01
14	PPA63	7	-43.87	-70.73	41.30
14	PPA8	21	-43.87	-70.73	93.20
14	PPA88	15	-43.87	-70.73	92.20
16	PNG1620	7	-45.51	-67.49	42.96
16	PNG1626	8	-45.51	-67.49	42.03
16	PNG1630	19	-45.51	-67.49	89.60
16	PNG1631	21	-45.51	-67.49	92.78
16	PNG1633	13	-45.51	-67.49	95.00
16	PNG1639	16	-45.51	-67.49	93.87
16	PNG1640	25	-45.51	-67.49	90.79
16	PNG1641	19	-45.51	-67.49	92.37
17	PNG0980	9	-47.86	-66.44	40.21
17	PNG0998	19	-47.86	-66.44	91.48
17	PNG1515	25	-47.86	-66.44	92.25
17	PNG1532	24	-47.86	-66.44	91.89
17	PNG1555	16	-47.86	-66.44	91.77
17	PNG1604	21	-47.86	-66.44	90.55
17	PNG1605	22	-47.86	-66.44	92.75
17	PNG1606	9	-47.86	-66.44	42.68
18	PNG0805	16	-48.99	-70.25	93.37
18	PNG0806	8	-48.99	-70.25	39.75
18	PNG0807	20	-48.99	-70.25	91.51
18	PNG0810	13	-48.99	-70.25	95.64
18	PNG0813	17	-48.99	-70.25	92.88
18	PNG0840	22	-48.99	-70.25	91.06
18	PNG0844	21	-48.99	-70.25	93.33
19	LTU638	16	-50.10	-68.47	90.80
19	LTU639	11	-50.10	-68.47	92.14
19	LTU640	20	-50.10	-68.47	93.01
19	LTU642	7	-50.10	-68.47	41.12
19	LTU643	19	-50.10	-68.47	91.86
19	LTU644	21	-50.10	-68.47	92.45
19	LTU645	22	-50.10	-68.47	92.24
19	LTU646	8	-50.10	-68.47	45.05
20	LTU671	10	-51.94	-69.57	41.54
20	LTU672	17	-51.94	-69.57	93.86
20	LTU673	29	-51.94	-69.57	90.37
20	LTU676	24	-51.94	-69.57	93.32
20	LTU677	18	-51.94	-69.57	94.27
20	LTU678	18	-51.98	-69.75	92.67
21	PNG0486	20	-54.87	-67.37	92.94
21	PNG0535	6	-54.87	-67.37	45.37
21	PNG0536	7	-54.87	-67.37	94.99
21	PNG0539	18	-54.87	-67.37	91.82
21	PNG0542	14	-54.87	-67.37	93.75
21	PNG0554	21	-54.87	-67.37	93.97
21	PNG0557	18	-54.87	-67.37	91.96
21	PNG0558	8	-54.87	-67.37	40.39
23	GD1782	8	-29.95	-71.36	43.17

23	GD1802	17	-28.95	-70.36	91.05
23	GD1809	22	-27.95	-69.36	91.27
23	GD1811	23	-26.95	-68.36	89.49
23	GD1827	24	-24.95	-66.36	91.77
23	GD1829	20	-23.95	-65.36	91.63
23	GD1831	23	-22.95	-64.36	90.15
23	GD1832	8	-21.95	-63.36	39.20

Table S2. Pairwise Fst values between populations.

pop	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	18	19	20	21	23
1	-																				
2	0.270	-																			
3	0.278	0.013	-																		
4	0.022	0.178	0.184	-																	
5	0.185	0.299	0.297	0.129	-																
6	0.222	0.104	0.109	0.153	0.245	-															
7	0.216	0.085	0.091	0.149	0.241	0.035	-														
8	0.074	0.196	0.204	0.051	0.150	0.168	0.164	-													
9	0.054	0.225	0.228	0.033	0.161	0.193	0.189	0.080	-												
10	0.268	0.009	0.010	0.179	0.283	0.099	0.083	0.198	0.222	-											
11	0.075	0.232	0.239	0.050	0.166	0.186	0.181	0.086	0.038	0.229	-										
12	0.198	0.158	0.176	0.123	0.270	0.023	0.030	0.137	0.183	0.147	0.160	-									
13	0.191	0.085	0.091	0.136	0.206	0.010	0.029	0.150	0.170	0.085	0.157	-0.003	-								
14	0.201	0.086	0.091	0.141	0.216	0.006	0.024	0.155	0.177	0.085	0.166	-0.001	0.005	-							
16	0.221	0.086	0.091	0.152	0.245	0.017	0.020	0.167	0.193	0.083	0.186	0.019	0.016	0.009	-						
17	0.232	0.090	0.094	0.158	0.258	0.029	0.026	0.173	0.200	0.086	0.195	0.042	0.026	0.020	0.012	-					
18	0.209	0.069	0.076	0.145	0.238	0.025	0.015	0.158	0.186	0.068	0.175	0.015	0.020	0.015	0.008	0.013	-				
19	0.235	0.051	0.055	0.159	0.262	0.049	0.035	0.175	0.202	0.049	0.199	0.065	0.042	0.039	0.031	0.033	0.018	-			
20	0.240	0.044	0.046	0.159	0.283	0.061	0.043	0.175	0.208	0.040	0.205	0.090	0.048	0.046	0.041	0.046	0.023	0.012	-		
21	0.237	0.006	0.012	0.161	0.262	0.077	0.058	0.176	0.203	0.005	0.203	0.095	0.064	0.064	0.061	0.063	0.043	0.031	0.019	-	
23	0.235	0.358	0.360	0.183	0.293	0.312	0.309	0.188	0.224	0.350	0.235	0.319	0.284	0.288	0.310	0.323	0.302	0.325	0.338	0.325	-

Table S3. Tajima's D value and heterozygosity (H) per population.

populations	D	H
10	-0.76	7.31E-07
11	-0.27	1.50E-06
12	-0.36	1.59E-06
13	-0.62	1.09E-06
14	-0.61	1.07E-06
16	-0.56	1.01E-06
17	-0.40	1.06E-06
18	-0.59	1.08E-06
19	-0.49	9.98E-07
1	-0.29	1.55E-06
20	-0.44	1.06E-06
21	-0.78	8.65E-07
23	-0.47	1.38E-06
2	-0.68	8.06E-07
3	-0.71	7.14E-07
4	-0.64	1.41E-06
5	-0.28	1.18E-06
6	-0.51	1.06E-06
7	-0.45	1.11E-06
8	-0.49	1.51E-06
9	-0.38	1.34E-06

Table S4. Range expansion analyses in different geographical regions

	Southern South America	Eastern Patagonia with TF	TF + 19 + 20	Western Patagonia without 23	Eastern Patagonia without TF	TF
longitude	−73.7	-71.95	-68.63	-71.60	-71.95	-67.67
latitude	−42.98	-46.24	-50.05	-47.50	-46.00	-54.12
q	8.18E−05	8.98E-05	3.46E-05	1.39E-05	1.36E-04	2.73E-03
r1	1.00	1.00	1.00	1.00	1.00	0.99
r10	1.00	1.00	1.00	1.00	1.00	0.95
r100	0.98	0.98	0.99	1.00	0.97	0.65
d1	61.73	56.23	145.58	361.42	36.97	1.85
rsq	0.64	0.68	0.06	0.04	0.76	0.96
pval	1,76E−44	7.30E-20	*not significant*	*not significant*	3.06E-11	*not significant*

p-value: probability of observing the pattern of pairwise psi under a null hypothesis of isolation by distance
longitude/latitude: are the coordinates of the most likely inferred origin of the expansion
q: is the strenght of the founder effect
r1/r10/r100: are the decrease in diversity over 1, 10 and 100 km, respectively
rsq/pval: are the correlation coefficient and correlation p-value for the most likely origin
TF: Tierra del Fuego.

Table S5. Pairwise directionality index between populations.

pop	10	11	12	13	14	16	17	18	19	1	20	21	23	2	3	4	5	6	7	8	9
10	0.000																				
11	-0.157	0.000																			
12	-0.060	0.065	0.000																		
13	-0.061	0.055	0.005	0.000																	
14	-0.053	0.078	0.014	0.011	0.000																
16	-0.039	0.104	0.028	0.023	0.013	0.000															
17	-0.046	0.113	0.019	0.017	0.005	-0.007	0.000														
18	-0.074	0.072	-0.011	-0.014	-0.024	-0.038	-0.032	0.000													
19	-0.041	0.121	0.025	0.022	0.014	0.000	0.007	0.039	0.000												
1	-0.080	0.058	-0.001	0.030	0.000	-0.025	-0.041	0.005	-0.051	0.000											
20	-0.006	0.162	0.058	0.053	0.047	0.031	0.036	0.068	0.034	0.084	0.000										
21	-0.026	0.134	0.029	0.032	0.025	0.010	0.016	0.044	0.013	0.054	-0.021	0.000									
23	0.026	0.175	0.079	0.115	0.098	0.071	0.065	0.096	0.046	0.113	0.012	0.046	0.000								
2	0.000	0.158	0.051	0.053	0.046	0.031	0.039	0.067	0.036	0.076	0.001	0.026	-0.024	0.000							
3	0.022	0.195	0.093	0.095	0.085	0.073	0.075	0.100	0.066	0.118	0.027	0.049	0.019	0.024	0.000						
4	-0.105	0.033	-0.020	0.003	-0.023	-0.050	-0.064	-0.026	-0.071	-0.023	-0.109	-0.079	-0.141	-0.100	-0.141	0.000					
5	-0.111	0.040	-0.040	-0.025	-0.049	-0.067	-0.075	-0.028	-0.072	-0.014	-0.109	-0.089	-0.125	-0.112	-0.141	-0.007	0.000				
6	-0.044	0.108	0.021	0.018	0.007	-0.005	0.002	0.035	-0.003	0.028	-0.037	-0.014	-0.069	-0.034	-0.073	0.057	0.076	0.000			
7	-0.090	0.057	-0.026	-0.028	-0.039	-0.051	-0.044	-0.013	-0.051	-0.024	-0.080	-0.058	-0.123	-0.082	-0.116	0.005	0.014	-0.050	0.000		
8	-0.070	0.050	-0.006	0.016	-0.004	-0.030	-0.038	-0.007	-0.043	0.003	-0.079	-0.051	-0.127	-0.078	-0.116	0.025	0.032	-0.033	0.016	0.000	
9	-0.081	0.055	0.003	0.022	-0.003	-0.028	-0.043	-0.003	-0.047	-0.004	-0.090	-0.064	-0.116	-0.083	-0.121	0.019	0.026	-0.034	0.010	-0.002	0.000

Cell (ij), respresent the directionality index between population of row i and population of column j.

If directionality index is negative, population of row i is closer to the origin and population of column j present more derived alleles.

If directionality index is positive, population of column j is closer to the origin and population of row i present more derived alleles.

Absolute values of directionality index are simmetrical between pair of populations.

CONCLUSIONES

En la región Patagónica, el ratón oliváceo, *A. olivacea*, se encuentra estructurada genéticamente en dos grandes grupos que corresponden básicamante a la estepa y pastizales patagónicos del lado oeste de la cordillera y a los bosques valdivianos y magallánicos asociados a los Andes. Estos dos grandes grupos mostraron niveles de diferenciación geográfica y diversidad genética contrastantes que evidencian historias biogeográficas dispares. En particular, el índice de direccionalidad sumado a la clina de heterocigosidad indican que Tierra del Fuego fue colonizado desde el continente y éste a su vez desde los bosques valdivianos. Nuestros resultados sugirieron, además, un nuevo refugio glacial periférico hasta ahora solo reportado en plantas. Nuestras estimaciones del momento de inicio de la expansión demográfica para ambos grupos de población fueron consistentes con una expansión posterior al último máximo glacial. En general, nuestros resultados apoyan la idea de un refugio costero cercano a los refugios valdivianos propuestos anteriormente, desde donde a su vez comenzó la colonización, y un posible refugio adicional más al sur y cercano a los bosques Magallánicos.

Respecto a la fisiología molecular de la especie y su respuesta a los ambientes estepa y bosque, nuestros resultados indican que *A. olivacea* recurre a los genes clásicos relacionados con la conservación de agua para lidiar con el déficit hídrico. Nuestros análisis también indican un posible rol de la selección dado que genes claves, como los expresados diferencialmente, presentaron una mayor divergencia genética, aunque sutil, que los demás genes al comparar las poblaciones de estepa y bosque.

PERSPECTIVAS

Dado que nuestros resultados ya se basan en miles de loci, los futuros esfuerzos para entender la filogeografía de la región deberían focalizarse en: (i) analizar otras especies de roedores Patagónicos para evaluar si existen patrones biogeográficos en común que, además de interesantes por sí mismos, apuntalarían los hallazgos individuales; (ii) incorporar análisis que tengan en cuenta la heterogeneidad espacial y temporal del ambiente para obtener resultados más precisos; (iii) incorporar datos que exploten otros estadísticos genético-poblacionales. Respecto a este punto sería conveniente incorporar, por ejemplo, regiones genómicas más amplias para complementar nuestra aproximación con análisis que exploten la información de ligamiento, y no solo las frecuencias alélicas.

Por otro lado, y en relación a la fisiología molecular de la especie frente a las condiciones que le imponen los ambientes, sería valioso incorporar análisis en condiciones de laboratorio. De esta manera, podremos librarnos de los efectos de la dieta y del sistema imnune para evaluar más precisamente la respuesta molecular a la restricción hidríca. En condiciones de laboratorio también podríamos discernir si la respuesta que encontramos es fundamentalmente plástica o si es producto de la divergencia genética en regiones reguladores estudiando la norma de reacción de las poblaciones de bosque y estepa. Finalmente, para evaluar genes directamente relacionados a la conservación del agua y que pueden estar sujetos a selección, se pueden llevar a cabo experimentos como XP-GWAS (del inglés: *Extreme Phenotype Genome Wide Association*). Aunque este método adolece de falta de sensibilidad puede ser muy útil como primera aproximación por su fácil aplicación y bajos costos.

BIBLIOGRAFÍA

Aatique, M., 1997. *Evaluation of TDOA Techniques for Position Location in CDMA Systems*.

Abud, C., 2011. *Variación genética y estructura filogeográfica de Abrothrix olivaceus en la Patagonia argentina y el sur chileno (Master thesis)*. Universidad de la República.

Al-Kahtani, M. a et al., 2004. Kidney mass and relative medullary thickness of rodents in relation to habitat, body size, and phylogeny. *Physiological and biochemical zoology : PBZ*, 77(3), pp.346–365.

Anders, S. et al., 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols*, 8(9), pp.1765–86.

Aragón, E. et al., 2011. Palaeogeography and palaeoenvironments of northern Patagonia from the Late Cretaceous to the Miocene: The Palaeogene Andean gap and the rise of the North Patagonian High Plateau. *Biological Journal of the Linnean Society*.

Avise, J.C., 2009. Phylogeography: Retrospect and prospect. *Journal of Biogeography*.

Avise, J.C., 2001. *Phylogeography: The History and Formation of Species*,

Bhaskar, A., Wang, Y.X.R. & Song, Y.S., 2015. Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Research*.

Bi, K. et al., 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*.

Bozinovic, F. et al., 2011. Body mass and water economy in the South American olivaceous field mouse along a latitudinal gradient: Implications for climate change. *Journal of Arid Environments*, 75, pp.411–415.

Bozinovic, F. et al., 2003. Seasonal acclimatization in water flux rate, urine osmolality and kidney water channels in free-living degus: molecular mechanisms, physiological processes and ecological implications. *The Journal of experimental biology*, 206(Pt 17), pp.2959–2966.

Bozinovic, F. & Gallardo, P., 2006. The water economy of South American desert rodents: From integrative to molecular physiological ecology. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, 142(3–4 SPEC. ISS.), pp.163–172.

Bryant, D. et al., 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), pp.1917–1932.

Catchen, J. et al., 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology*.

Chen, H., Hey, J. & Chen, K., 2015. Inferring very recent population growth rate from population-scale sequencing data: Using a large-sample coalescent estimator. *Molecular Biology and Evolution*.

Colinvaux, P.A., De Oliveira, P.E. & Bush, M.B., 2000. Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. In *Quaternary Science Reviews*.

Cortes, A. et al., 2000. Water Economy in rodents: evaporative water loss and metabolic water production. *Revista Chilena de Historia Natural*, 6, pp.311–321.

D'Elía, G. et al., 2007. Definition and diagnosis of a new tribe of sigmodontine rodents (Cricetidae: Sigmodontinae), and a revised classification of the subfamily. *Gayana*, 71(2), pp.187–194.

D'Elía, G. & Pardiñas, U.F.J., 2015. Subfamily Sigmodontinae Wagner, 1843. In J. L. Patton, U. F. J. Pardiñas, & G. D'Elía, eds. *Mammals of South America. Volume 2. Rodents.* Chicago: University of Chicago Press, p. 1384.

DeGiorgio, M., Jakobsson, M. & Rosenberg, N.A., 2009. Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), pp.16057–16062.

DeSalle, R., Schierwater, B. & Hadrys, H., 2017. MtDNA : The small workhorse of evolutionary studies. *Frontiers in Bioscience (Landmark Ed)*, 22, pp.873–887.

Esquerré, D. et al., 2019. How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae). *Evolution*.

Excoffier, L. et al., 2013. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, 9(10).

Excoffier, L. & Foll, M., 2011. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*.

Excoffier, L. & Ray, N., 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*.

Ferrando, F., 2002. The Quaternary Glaciations in Chile: A General review. *Revista Geografica de Chile Terra Australis*, 47, pp.129–165.

Folguera, A. et al., 2011. A review of Late Cretaceous to Quaternary palaeogeography of the southern Andes. *Biological Journal of the Linnean Society*.

Gagnaire, P.-A., Normandeau, E. & Bernatchez, L., 2012. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Molecular biology and evolution*, 29, pp.2909–19.

Gallardo, P.A., Cortés, A. & Bozinovic, F., 2005. Phenotypic Flexibility at the Molecular and Organismal Level Allows Desert Dwelling Rodents to Cope with Seasonal Water Availability. *Physiological and Biochemical Zoology*, 78(2), pp.145–152.

Gao, F. & Keinan, A., 2016. Explosive genetic evidence for explosive human population growth. *Current Opinion in Genetics and Development*, 41, pp.130–139.

Gattepaille, L.M., Jakobsson, M. & Blum, M.G.B., 2013. Inferring population size changes with sequence and SNP data: Lessons from human bottlenecks. *Heredity*.

Gautier, M. & Vitalis, R., 2013. Inferring population histories using genome-wide allele frequency data. *Molecular Biology and Evolution*.

Gayral, P. et al., 2013. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate-Invertebrate Gap. *PLoS Genetics*, 9.

Gibson, G., 2012. Rare and common variants: Twenty arguments. *Nature Reviews Genetics*.

Giorello, F.M.F.M. et al., 2014. Characterization of the kidney transcriptome of the South American olive mouse Abrothrix olivacea. *BMC genomics*, 15(1), p.446.

Gutenkunst, R.N. et al., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5.

Hallatschek, O. et al., 2007. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*.

Harvey, M.G. et al., 2016. Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology*, 65(5), pp.910–924.

Heled, J. & Drummond, A.J., 2010. Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*, 27(3), pp.570–580.

Hodges, E. et al., 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols*.

Hohenlohe, P.A. et al., 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*.

Holsinger, K.E., 2012. *Lecture Notes in Population Genetics* C. I. P. Platform, ed., Available at: http://darwin.eeb.uconn.edu/eeb348/lecture-notes/book.pdf.

Hooghiemstra, H. & Van Der Hammen, T., 1998. Neogene and Quaternary development of the neotropical rain forest: The forest refugia hypothesis, and a literature overview. *Earth Science Reviews*.

Hoorn, C. et al., 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*.

Hoorn, C. et al., 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. *Geology*.

Iglesias, A., Artabe, A.E. & Morel, E.M., 2011. The evolution of Patagonian climate and vegetation from the Mesozoic to the present. *Biological Journal of the Linnean Society*.

Kamm, J.A. et al., 2018. Efficiently inferring the demographic history of many populations with allele count data. *bioRxiv*.

Kaplan, M.R. et al., 2009. Can glacial erosion limit the extent of glaciation? *Geomorphology*.

Leigh, E.G., O'Dea, A. & Vermeij, G.J., 2014. Historical biogeography of the isthmus of panama. *Biological Reviews*, 89(1), pp.148–172.

Lemoine, M. & Pradeu, T., 2018. Dissecting the Meanings of "Physiology" to Assess the Vitality of the Discipline. *Physiology*, 33(4), pp.236–245.

Lessa, E.P., D'Elía, G. & Pardiñas, U.F.J., 2010. Genetic footprints of late Quaternary climate change in the diversity of Patagonian-Fueguian rodents. *Molecular Ecology*, 19(15), pp.3031–3037.

Li, H. & Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), pp.493–496.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*.

Liu, X. & Fu, Y.X., 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics*.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), p.550.

De Maio, N., Schrempf, D. & Kosiol, C., 2015. PoMo: An allele frequency-based approach for species tree estimation. *Systematic Biology*, 64(6), pp.1018–1031.

Mallo, D. & Posada, D., 2016. Multilocus inference of species trees and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702).

Mann, G., 1978. Los pequeños mamíferos de Chile. *Gayana, Zoologia*, 40, pp.1–342.

Marshall, L.G., 1979. A model for paleobiogeography of South American cricetine rodents. *Paleobiology*.

McCoy, R.C. et al., 2014. Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Molecular ecology*, 23, pp.136–50.

McNab, B.. & Brown, J.., 2002. *Physiological Ecology of Vertebrates: A View from Energetics* First Edic., Ithaca, United States: Cornell University Press.

Melville, J. et al., 2017. Identifying hybridization and admixture using SNPs: Application of the DArTseq platforminphylogeographic research on vertebrates. *Royal Society Open Science*.

Montgomery, D.R., Balco, G. & Willett, S.D., 2001. Climate, tectonics, and the morphology of the Andes. *Geology*.

Morris, A.B. & Shaw, J., 2018. Markers in time and space: A review of the last decade of plant phylogeographic approaches. *Molecular Ecology*, 27(10), pp.2317–2333.

Morrone, J.J., 2006. Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annual review of entomology*, 51, pp.467–494.

Morrone, J.J., 2004. Panbiogeografía, componentes bióticos y zonas de transición. *Revista Brasileira de Entomologia*.

Myers, N., 2003. Biodiversity hotspots revisited. *Bioscience Biotechnology and Biochemistry*.

Nielsen, R. et al., 2017. Tracing the peopling of the world through genomics. *Nature*, 541(7637), pp.302–310.

Oshlack, A., Robinson, M.D. & Young, M.D., 2010. From RNA-seq reads to differential expression results. *Genome biology*, 11, p.220.

Palamara, P.F. & Pe'er, I., 2013. Inference of historical migration rates via haplotype sharing. In *Bioinformatics*.

Park, J. et al., 2018. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science*, 2131 pp.758–763.

Pavey, S.A. et al., 2010. The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, 1206, pp.110–129.

Peter, B.M. & Slatkin, M., 2013. Detecting range expansions from genetic data. *Evolution*, 67(11), pp.3274–3289.

Pickrell, J.K. & Pritchard, J.K., 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*.

Pizarro, R. et al., 2012. Latitudinal analysis of rainfall intensity and mean annual precipitation in chile. *Chilean Journal of Agricultural Research*, 72(2), pp.252–261.

Ponce, J.F. et al., 2011. Palaeogeographical evolution of the Atlantic coast of Pampa and Patagonia from the last glacial maximum to the Middle Holocene. *Biological Journal of the Linnean Society*.

Qin, L.-X., Huang, H.-C. & Niu, Y., 2015. Differential Expression Analysis for RNA-Seq: An Overview of Statistical Methods and Computational Software. *Cancer Informatics*, 14(Suppl 1), p.57.

Quattrocchio, M.E. et al., 2011. Changes of the palynobiotas in the Mesozoic and Cenozoic of Patagonia: A review. *Biological Journal of the Linnean Society*.

Rabassa, J., Coronato, A. & Martínez, O., 2011. Late Cenozoic glaciations in Patagonia and Tierra del Fuego: An updated review. *Biological Journal of the Linnean Society*.

Rannala, B. & Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*.

Rapaport, F. et al., 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*.

Renaut, S., Nolte, A.W. & Bernatchez, L., 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (Coregonus spp. Salmonidae). *Molecular ecology*, 19 Suppl 1, pp.115–131.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), pp.139–40.

Robinson, M.D. & Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11(3), p.R25.

Rodríguez-Serrano, E., Cancino, R.A. & Palma, R.E., 2006. Molecular phylogeography of Abrothrix olivaceus (Rodentia: Sigmodontinae) in Chile. *Journal of Mammalogy*, 87(5), pp.971–980.

Rull, V., 2011. Neotropical biodiversity: Timing and potential drivers. *Trends in Ecology and Evolution*, 26(10), pp.508–513.

Rull, V., 1999. Palaeofloristic and palaeovegetational changes across the Paleocene/Eocene boundary in northern South America. *Review of Palaeobotany and Palynology*.

Ruzzante, D.E. & Rabassa, J., 2011. Palaeogeography and palaeoclimatology of Patagonia: effects on biodiversity. *Biological Journal of the Linnean Society*, 103(2), pp.221–228.

Schiffels, S. & Durbin, R., 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*.

Schraiber, J.G. & Akey, J.M., 2015. Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics*, 16(12), pp.727–740.

Shendure, J. & Ji, H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), pp.1135–1145.

Smith, M.F., Kelt, D.A. & Patton, J.L., 2001. Testing models of diversification in mice in the Abrothrix olivaceus/xanthorhinus complex in Chile and Argentina. *Molecular Ecology*, 10(2), pp.397–405.

Soneson, C. & Delorenzi, M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics*.

Spalink, D., MacKay, R. & Sytsma, K.J., 2019. Phylogeography, population genetics and distribution modelling reveal vulnerability of Scirpus longii (Cyperaceae) and the Atlantic Coastal Plain Flora to climate change . *Molecular Ecology*, pp.0–3.

Stange, M. et al., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports miocene closure of the Panamanian Isthmus. *Systematic Biology*.

Steinrücken, M., Kamm, J.A. & Song, Y.S., 2015. Inference of complex population histories using whole-genome sequences from multiple populations. *bioRxiv*.

Tataru, P., Nirody, J.A. & Song, Y.S., 2014. DiCal-IBD: Demography-aware inference of identity-by-descent tracts in unrelated individuals. *Bioinformatics*.

Todd, E. V., Black, M.A. & Gemmell, N.J., 2016. The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology*, 25(6), pp.1224–1241.

Trénel, P. et al., 2008. Landscape genetics, historical isolation and cross-Andean gene flow in the wax palm, Ceroxylon echinulatum (Arecaceae). *Molecular Ecology*.

Turchetto-Zolet, A.C. et al., 2013. Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, 22(5), pp.1193–1213.

Valdez, L. et al., 2015. Characterization of the kidney transcriptome of the long-haired mouse Abrothrix hirta (Rodentia, Sigmodontinae) and comparison with that of the olive mouse A. olivacea. *PLoS ONE*, 10(4), p.e0121148.

Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), pp.57–63.

Waterhouse, G.R., 1837. Characters of new species of the genus Mus, from collection of Mr. Darwin. *Proceedings of the Zoological Society of London*, 5(1), pp.15-21,27-32.

Wolf, J.B.W. et al., 2010. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular ecology*, 19 Suppl 1, pp.162–175.

Yang, Z., 2014. *Molecular Evolution: A Statistical Approach*, Oxford: Oxford University Press.