

Universidad Nacional de Córdoba
Facultad de Ciencias Exactas, Físicas y Naturales



Tesina de Grado para optar por el título de Bióloga

Efectos de la pérdida de hábitat sobre la diversidad y estructura genética en poblaciones de *Euphorbia acerensis* Boiss. (Euphorbiaceae) en el distrito Chaco Serrano (Córdoba, Argentina)

Tesista: Elisa Beatriz Rodriguez

Firma:

Director: Dr. Leonardo David Amarilla

Firma:

Co-director: Dr. Gabriel Grilli

Firma:

Cátedra de Diversidad Biológica III (FCEFyN, Universidad Nacional de Córdoba)

Instituto Multidisciplinario de Biología Vegetal (CONICET-UNC)

Marzo, 2022

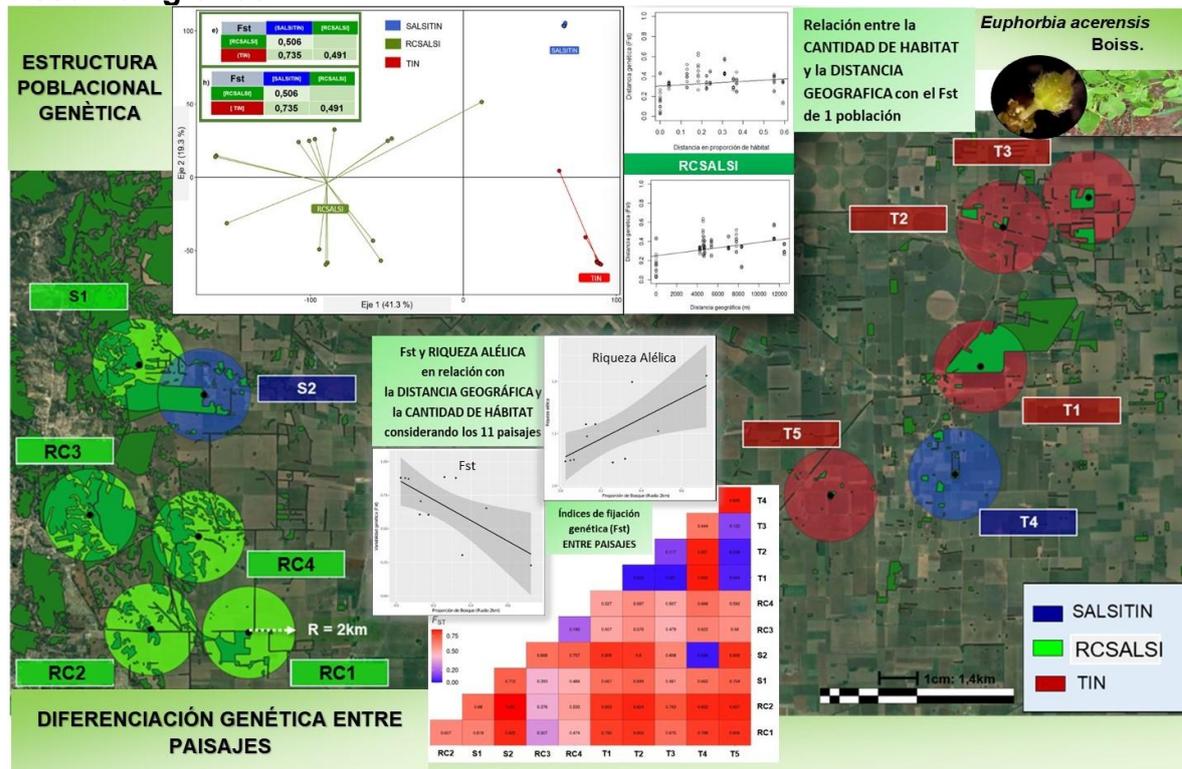
Efectos de la pérdida de hábitat sobre la diversidad y estructura genética en poblaciones de *Euphorbia acerensis* Boiss. (Euphorbiaceae) en el distrito Chaco Serrano (Córdoba, Argentina)

Resumen

Los cambios en el uso de la tierra promueven una conversión de áreas naturales en tierras agrícolas o urbanas debido a la acción del hombre. La pérdida y la fragmentación de hábitat *per se* son los procesos que ocurren a escala de paisaje provocando la degradación del ambiente natural. La nueva configuración, cantidad y composición de los componentes del paisaje influyen en procesos y patrones ecológicos y genéticos de los organismos. El objetivo del este trabajo fue evaluar la influencia de la cantidad de hábitat seminatural y la distancia geográfica sobre la diversidad y estructura genética de una morfoespecie de planta nativa y de habito anual (*Euphorbia acerensis*) que crece en el Bosque Chaqueño Serrano. Se utilizó la secuenciación de próxima generación de ADN para obtener marcadores tipo RADseq-SNPs (polimorfismos de un sólo nucleótido asociado a sitios de restricción) en 32 individuos de *E. acerensis* recolectados en 11 paisajes con distinta cantidad de hábitat seminatural. Los paisajes estuvieron incluidos en las localidades geopolíticas de Río Ceballos (RC), Salsipuedes (SALSI) y Tinoco (TIN). Se estimaron parámetros poblacionales y se analizó la escala del paisaje a la cual las variables genéticas se relacionan más fuertemente con la cantidad de hábitat seminatural presente en los paisajes. Los marcadores moleculares nucleares tipo RADSeq-SNP permitieron obtener una buena cantidad de loci informativos para estimar índices de diversidad genética e índices de fijación/estructura poblacional. Todos los paisajes presentaron diferentes cantidades de hábitat y la escala de análisis adecuada resultó en un buffer de paisaje de dos kilómetros de radio. La riqueza alélica varió positivamente con la cantidad de hábitat. El Fis indicó que los genotipos incluidos en todos los paisajes están en desequilibrio de Hardy-Weinberg por exceso ($Fis < 0$) o déficits de heterocigotos ($Fis > 0$). Las variaciones del Fis resultaron independientes a la cantidad de hábitat. El flujo de genes está restringido entre los paisajes (Fst elevados) y tanto la cantidad de hábitat como la distancia geográfica influyen en esta restricción. Podemos indicar que se recuperaron efectos significativos de la cantidad de hábitat seminatural y la distancia geográfica sobre la diversidad y estructura genética en todos los paisajes. Los individuos incluidos en este estudio

pueden formar tres poblaciones genéticas ([RCSALSI] [TIN] [SALSITIN]) o bien, pueden formar dos poblaciones genéticas (una de ellas con dos subpoblaciones): i- [RCSALSI], ii- [TIN-SALSITIN]). La cantidad de hábitat y la distancia geográfica resultaron influyentes en la estructura genética de [RCSALSI]. Se recuperaron efectos significativos de la cantidad de hábitat seminatural y la distancia geográfica sobre la diversidad y estructura genética en una de las poblaciones genéticas recuperadas.

Resumen gráfico



Se muestra a la morfoespecie (*E. acerensis*) considerada en este estudio y los 11 paisajes (áreas circulares) en los que se coleccionaron los ejemplares botánicos. Se muestra que la cantidad de hábitat seminatural y la distancia geográfica influyen sobre la diversidad y estructura genética en todos los paisajes (regresiones y heatmap). Se muestran las tres poblaciones genéticas recuperadas y la influencia de la cantidad de hábitat seminatural y la distancia geográfica sobre la diversidad y estructura genética en una de ellas (PCA, AMOVAs y regresiones).

Palabras Clave: genética de poblaciones, marcadores moleculares nucleares, pérdida de hábitat, secuenciación de próxima generación.

Introducción

Los cambios en el uso de la tierra están promoviendo la pérdida y degradación del hábitat natural en la mayoría de los ecosistemas del planeta y se constituyen así, en una de las principales causas de pérdida de biodiversidad en esta era, el Antropoceno (Crutzen & Stoermer, 2000; Rico, 2019). Dentro de estas causas se puede reconocer a la “*fragmentación de hábitat*”, el cual ha sido utilizado como un concepto paraguas que involucra a dos procesos diferentes: **la fragmentación de hábitat *per se*** (en adelante fragmentación del hábitat) y la **pérdida de hábitat**. La fragmentación del hábitat implica la subdivisión de un hábitat natural en parches de menor tamaño sin que ocurra, necesariamente, pérdida significativa del hábitat. Por el contrario, la pérdida de hábitat implica una reducción significativa en la superficie de hábitat natural o seminatural sin subdivisión del área remanente (Fahrig, 2003; 2013). Sin embargo, muchos estudios que evalúan los posibles efectos de la “*fragmentación de hábitat*” sobre la biodiversidad, no consideran la evaluación independiente de estos procesos (Fletcher et.al., 2018). La fragmentación del hábitat es un proceso que ocurre a escala de paisaje, y no de fragmentos individuales (Fahrig, 2003; 2017). Históricamente se han extrapolado erróneamente patrones y mecanismos evaluados a escala de parche para explicar la fragmentación de hábitat (Figura 1). En la actualidad, se sabe que evaluar los efectos de la fragmentación a escala de parche no es apropiado ya que se confunden en gran medida con la pérdida de hábitat, debido a que existe una alta correlación entre la reducción en el área de los parches o un aumento en el aislamiento de estos con la cantidad de hábitat presente a su alrededor. Los estudios que han abordado al proceso de la fragmentación *per se* (i.e. subdivisión del hábitat sin que ocurra pérdida) a escala de paisaje han mostrado efectos positivos o nulos de este proceso sobre la biodiversidad (Fahrig, 2017 y referencias allí citadas). Por ende, tanto la fragmentación, como la pérdida de hábitat son procesos que deben ser abordados a escala de paisaje (Figura 1).

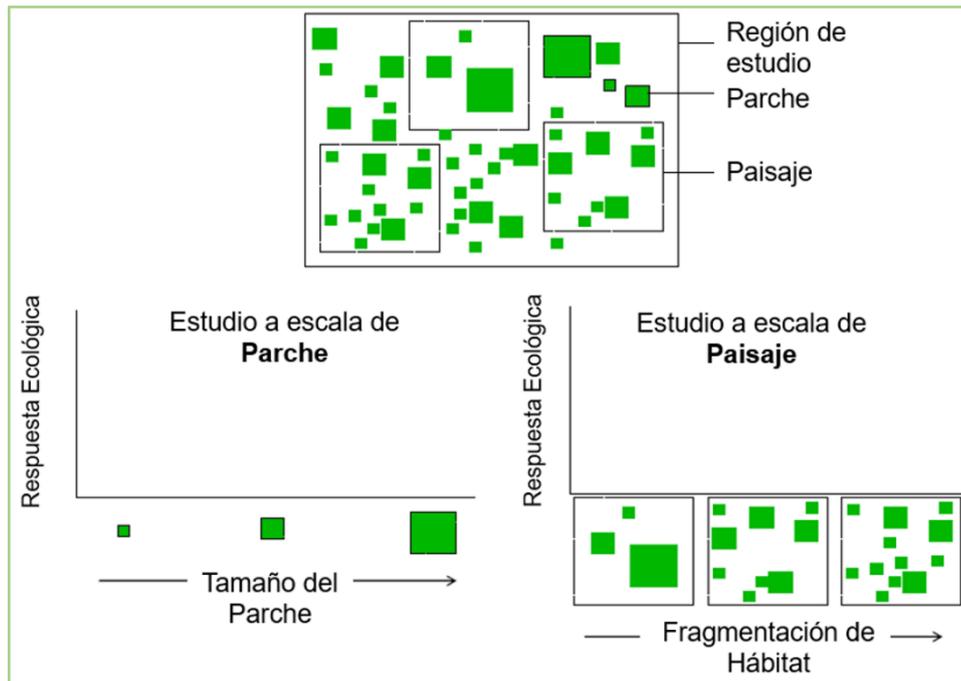


Figura 1. Esquema de diferenciación entre estudios a escala de parche y a escala de paisaje. En una escala de parches, las muestras son seleccionadas a lo largo de gradientes de tamaño de parche y/o aislamiento. Estos gradientes no son medidas de fragmentación del hábitat *per se*. En un estudio a escala de paisaje, se seleccionan múltiples paisajes a lo largo de un gradiente de fragmentación (de parches). Estos se pueden seleccionar para evitar una correlación con la cantidad de hábitat, lo que permite que la estimación de los efectos de la fragmentación del hábitat sea independiente de los efectos producidos por la cantidad de hábitat. *Tomado y modificado de Fahrig, 2018.*

En consecuencia, se ha propuesto la hipótesis de la cantidad de hábitat, la cual postula que los procesos y patrones ecológicos y genéticos observados en un paisaje responden a la cantidad de hábitat presente en un área circundante independientemente del tamaño del parche focal donde se midan estos procesos y de la configuración espacial que posea ese hábitat remanente (Fahrig, 2013; Saura, 2020). En particular, se ha visto que la cantidad de hábitat, y no la configuración de este, sería un mejor predictor de los patrones genéticos de las poblaciones en paisajes fragmentados (Jackson y Fahrig, 2016). Esto se debe a que el tamaño poblacional resulta importante en la estructuración genética, principalmente cuando la población es pequeña o se encuentran múltiples subpoblaciones en el paisaje circundante. Por todo lo anterior, en esta propuesta nos concentraremos en la idea de cantidad de hábitat.

Un paisaje tiene características a múltiples escalas espaciales (e.g., gradientes, distancias, distribución, conectividad) y temporales, y naturalmente afectan tanto

la estructura genética espacial como las inferencias posibles acerca del flujo génico. Para evaluar posibles efectos de la/s característica/s del paisaje sobre atributos/procesos genéticos, se debe seleccionar una escala determinada de análisis y para ello es importante considerar todos los aspectos posibles de la especie en estudio (i.e., ciclo de vida, estrategias reproductivas, polinización, estrategias de dispersión, distribución) y, la estrategia de muestreo de genotipos más adecuada (incluidas las opciones sobre métodos analíticos y la selección de marcadores genéticos). La definición/delimitación del paisaje y el muestreo de genotipos introducen diferentes características que pueden afectar a las inferencias sobre el flujo de genes, por lo tanto, se debe tener cuidado, especialmente cuando la extensión espacial del área de estudio es limitada. Por ejemplo, el sitio de muestreo debe ser más pequeño que el tamaño promedio del área de distribución o la distancia de dispersión del organismo de estudio (Dale & Fortin, 2014) y, para los datos ráster, simplificar la resolución en clases discretas puede resultar en una baja potencia para detectar efectos en el flujo de genes (Anderson et. al., 2010).

Los métodos utilizados para caracterizar el paisaje circundante a los sitios de muestreo pueden ser un determinante principal de la escala espacial en la que se aplican los resultados analíticos. A su vez, el uso de una sola escala de muestreo para un método estadístico en particular puede llevar a pasar por alto factores importantes que afectan el flujo de genes (Keitt & Urban, 2005). La técnica analítica utilizada para correlacionar los datos del paisaje y los datos genéticos también puede influir en los resultados particularmente cuando la velocidad de cambio del paisaje es más rápida de lo que pueden resolver algunos marcadores moleculares comunes (Anderson et. al., 2010 y referencias allí citadas).

La extinción de una especie es el resultado más notorio de la pérdida de hábitat, sin embargo, y debido a que la configuración y la cantidad de hábitat configura la estructura espacial de la variación genética poblacional (Templeton et al., 1990; Manel & Holderegger. 2013), mucho antes de que la extinción ocurra, la diversidad genética resulta afectada (i.e., pérdida o disminución de la diversidad genética o (re)estructuración espacial de la diversidad genética) (Manel & Holderegger, 2013). Es así como los grupos de individuos que ocupan paisajes con cambios en la cantidad de hábitats, con el tiempo pueden diferenciarse genéticamente, lo que resulta en una estructuración genética (Amarilla et al.,

2015; Zhao et al., 2020). Esta estructuración puede deberse a la disminución o interrupción del flujo génico, las fluctuaciones en el tamaño poblacional (limitado por la cantidad de hábitat) y/o a los efectos de la deriva genética y consecuentes cambios en la frecuencia alélica (Amarilla et al., 2015; Jackson y Fahrig, 2016; Zhao et al., 2020).

Al comparar especies arbóreas y herbáceas anuales que crecen en paisajes con diferente cantidad de hábitat y, sobre la base de una serie de marcadores moleculares, se ha demostrado que las plantas herbáceas anuales resultarían más “sensibles” a los fines de estudiar los efectos de la pérdida de hábitat sobre la diversidad y estructura genética, sobre todo si se quiere abordar una escala temporal ecológica (Hamrick & Godt, 1996; Nybom, 2004; Mona et al., 2014). Cabe mencionar que diversidades genéticas bajas o altas, pueden deberse a procesos tanto históricos como recientes, que pueden ser causados o no por las fluctuaciones en la cantidad de hábitat, o que pueden ser intrínsecos de cada linaje. Recientemente se demostró que, para detectar posibles efectos de la pérdida de hábitat sobre la variabilidad genética, se necesita un muestreo exhaustivo a nivel local. Resulta menester mencionar que el concepto de hábitat es inherente a los requerimientos de la especie, por ejemplo, existen especies asociadas exclusivamente a los bordes o al interior de los fragmentos de bosque y, cuando las especies no presentan una asociación específica determinada a un sector del bosque se las denomina generalistas. De esta forma, en especies generalistas, la cantidad de hábitat de vegetación seminatural remanente en el paisaje coincide con la cantidad de hábitat disponible para la especie (Jackson y Fahrig, 2016).

Con bases en lo antedicho, *Euphorbia acerensis* Boiss (Euphorbiaceae) se constituye en un buen sistema de estudio para abordar esta problemática ya que; es una hierba ruderal, anual, nativa y generalista de las Provincias Fitogeográficas Chaqueña y del Espinal. En el Distrito del Chaco Serrano crece en paisajes con pérdida de hábitat seminatural debido a cambios en el uso del suelo (Cabrera, 1976; Subils, 1977; Grilli et al., 2013). Los ciatios de *E. acerensis* son visitados principalmente por diferentes morfoespecies de moscas y hormigas, y secundariamente por abejas nativas y por *Apis mellifera* (Obs. Pers, Grilli & Amarilla). Los frutos son esquizocárpicos dehiscentes y las semillas carunculadas y mucilaginosas (Subils, 1977). En un estudio previo, 11 marcadores moleculares

tipo microsatélites (STR) resultaron no informativos en *E. acerensis* (Amarilla et al., 2010). Se ha reportado que la reducción en el tamaño de los fragmentos de bosque afecta negativamente tanto el desarrollo y el éxito reproductivo de *E. acerensis*, como a su interacción con simbiontes fúngicos radicales (Grilli et al., 2013). Nos interesa conocer las relaciones que pudieran existir entre la variación de la cantidad de hábitat seminatural y atributos genéticos (tales como diversidad genética, flujo génico, estructura poblacional) en poblaciones de *E. acerensis* que crecen en paisajes con diferentes cantidades de hábitat.

El objetivo general de este trabajo fue explorar posibles efectos de la variación de la cantidad de hábitat seminatural y la distancia sobre la diversidad y estructura genética en paisajes y poblaciones de *E. acerensis* que crecen en las tierras bajas del Chaco Serrano.

Los objetivos específicos fueron:

- a- estimar la cantidad de hábitat seminatural en 11 paisajes con diferentes grados de pérdida de hábitat y donde crezca *E. acerensis*.
- b- estimar índices de diversidad genética e índices de fijación genética utilizando marcadores moleculares provenientes del genoma nuclear de *E. acerensis*.
- c- definir la escala de paisaje más apropiada para evaluar posibles relaciones entre los índices de diversidad y fijación genética y la de cantidad de hábitat seminatural en los paisajes y la distancia geográfica.
- d- explorar posibles patrones de estructuración genética poblacional para establecer posibles relaciones entre la cantidad de hábitat seminatural y la distancia geográfica con la estructura poblacional recuperada.

Hipótesis: La cantidad de hábitat y la distancia geográfica en los paisajes seminaturales afectan al flujo genético provocando cambios en la diversidad genética y en los índices de fijación poblacional. En el mismo sentido, la estructura genética poblacional a diferentes niveles jerárquicos está modelada por la cantidad de hábitat y por la distancia geográfica.

Predicciones:

- a- los índices de diversidad genética se relacionan positivamente con la cantidad de hábitat en todos los paisajes. Por su parte, los índices de

fijación se relacionan negativamente con la cantidad de hábitat y positivamente con la distancia geográfica.

- b- los genotipos incluidos en este estudio se agrupan en poblaciones genéticas y la estructuración poblacional a diferentes niveles jerárquicos se relaciona con la cantidad de hábitat y la distancia geográfica.

Materiales y métodos (se presentan según los objetivos específicos planteados)

Estimar la cantidad de hábitat seminatural en 11 paisajes con diferentes grados de pérdida de hábitat y donde crezca E. acerensis.

Área de estudio y estimación de la cantidad y proporción de hábitat seminatural en los paisajes

El distrito del Chaco Serrano (perteneciente a la Provincia Fitogeográfica Chaqueña; Cabrera, 1976) incluye a una zona de tierras bajas y otra de montañas. Durante el siglo XX y especialmente en los últimos 50-30 años, la zona de tierras bajas, ocupado por los bosques de Chaco y Espinal, se ha transformado en un agroecosistema con altas tasas de pérdida de hábitat seminatural por deforestación (Boletta et al., 2006; Zak et al., 2008). En la zona de las tierras bajas se delimitaron 11 paisajes (Figura 2). Considerando el centro de cada paisaje, los mismos estuvieron separados por al menos cuatro kilómetros lineales (Figura 2).

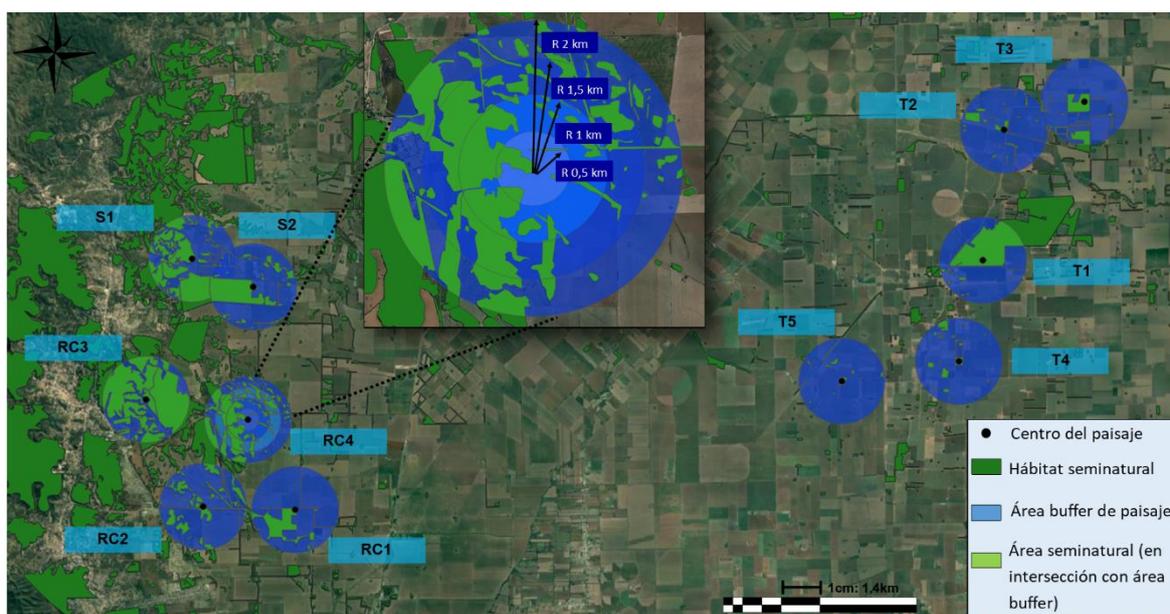


Figura 2. Se muestran los 11 paisajes (Departamento Colón, Córdoba) con las diferentes proporciones de hábitat incluidos en ellos a los 2 km de buffer. En uno de ellos se detallan los diferentes radios considerados (0,5, 1, 1,5 y 2km).

Con el objetivo de estimar para cada paisaje la cantidad de hábitat seminatural remanente, se definieron zonas buffers a cuatro escalas; 0,5, 1, 1,5 y 2 km de radio (Figura 2). Las zonas buffers se superpusieron con el área de hábitat seminatural, previamente identificado a través de imágenes satelitales (Figura 2). Consideramos como “hábitat seminatural” a cualquier área con cobertura de

vegetación leñosa (remanente de bosque nativo), porque conocemos la vegetación asociada a la presencia de *E. acerensis* (Subils, 1977; Grilli et al., 2013) Entonces, las áreas de hábitat seminatural estuvieron conformadas por remanentes de bosque (> 1 árbol), bordes de caminos y cortinas forestales/vegetales de vegetación principalmente nativa e identificadas a través de imágenes satelitales (Figura 2). De esta forma, a través de la interacción entre las zonas buffer y la cantidad de hábitat seminatural, se calculó la proporción de hábitat seminatural por cada uno de los buffers (Tabla 2, ver sección resultados). Las métricas del paisaje se calcularon con el plugin Group Stats para capas vectoriales. Todas las mediciones se realizaron utilizando el software QGIS (QGIS Development Team, 2016).

Estimar índices de diversidad genética y de fijación genética utilizando marcadores moleculares provenientes del genoma nuclear de E. acerensis.

Colección de material vegetal

En el centro de 10 paisaje (Tabla 1) se colectaron tres individuos de *E. acerensis* y en un paisaje se coleccionaron dos individuos (Se incluyeron 32 individuos en total). En todos los casos, los individuos estuvieron separados por al menos 10 metros. Un ejemplar de referencia para cada paisaje fue depositado en el herbario CORD. **Tabla 1.** Coordenadas geográficas del centro de los 11 paisajes incluidos en este estudio. Se detalla además las localidades geopolíticas más cercanas. T= Tinoco, RC= Rio Ceballos, S= Salsipuedes (el número que acompaña a cada letra indica cada paisaje incluido en la localidad correspondiente (e.g., T1; paisaje uno en la localidad de Tinoco).

Paisaje	Longitud	Latitud	Localidad
T1	-4417591	-6559356	Tinoco
T2	-4417499	-6563633	Tinoco
T3	-4421185	-6564927	Tinoco
T4	-4415434	-6552946	Tinoco
T5	-4410071	-6552030	Tinoco
RC1	-4385014	-6546098	Rio Ceballos
RC2	-4380797	-6546240	Rio Ceballos
RC3	-4378176	-6551183	Rio Ceballos
RC4	-4382854	-6550252	Rio Ceballos
S1	-4380291	-6557678	Salsipuedes
S2	-4383091	-6556384	Salsipuedes

Extracción, evaluación de calidad y cuantificación del ADN

El material vegetal se deshidrató y almacenó en bolsas herméticas con gel de sílice. Previo a su procesamiento, se realizó un secado en estufa a 30 °C por 12 h. A los fines de obtener ADN genómico en la calidad y la cantidad requerida para la construcción de las bibliotecas genómicas y la secuenciación por estrategia de próxima generación, en esta tesina se modificó el protocolo reportado por Kumar et. al. (2018). El nuevo protocolo se detalla en el **Anexo I**.

Para evaluar la integridad del ADN extraído se realizaron electroforesis a 90V por 30 min en gel de agarosa al 1% con buffer TAE y 5µl de colorante SYBR Green, y se utilizó un marcador de peso molecular Lambda/HindIII. La visualización se realizó en un trans-iluminador de luz UV. Las fotografías de los geles con los extractos (bandas) de ADN se presentan en el **Anexo II**. La concentración y la pureza del ADN se midió en un espectrofotómetro BioSpec-nano. La concentración de las muestras de ADN se determinó en ng/µl y se calculó teniendo en cuenta la absorbancia a una longitud de onda de 260nm. Mientras que la relación de absorbancias A260/280 y A260/230 se utilizó para evaluar la pureza de las muestras. La concentración y puridades obtenidas pueden verse en el **Anexo III**. Para el proceso de secuenciación se seleccionaron aquellos extractos con pureza entre 1,6-1,8 y se realizaron diluciones para uniformizar la concentración de ADN en las muestras (500-800 ng/µl) y evitar así, diferencias en las profundidades de lecturas durante el proceso de secuenciación de próxima generación.

Construcción de bibliotecas genómicas

Fragmentación/Digestión del ADN genómico por método enzimático; se utilizaron las enzimas de restricción SphI y MboI. Posteriormente se seleccionaron las bandas de 250pb, adecuado según la estrategia de secuenciación utilizada.

Enriquecimiento de las regiones de interés: En este caso se utilizó enriquecimiento basado en amplicones obtenidos mediante PCR multiplex (Figura 3).

Multiplexado de los amplicones enriquecidos: En este paso, se adicionaron oligonucleótidos que actúan como identificadores únicos, denominados “identificadores moleculares” (molecular barcodes, unique identifiers o primer IDs en inglés), los cuales permiten la identificación de cada fragmento y el individuo al cual pertenecen, además incluyen una secuencia de anclaje, específica para la adherencia a la superficie bidimensional (“celda de flujo”) del soporte Illumina. De

esta manera, en un mismo ensayo se combinaron múltiples muestras sin riesgo de entrecruzamiento entre las mismas (Yohe & Thyagarajan, 2017). El paso de hibridación con los ID también permite que las lecturas con errores vinculados a la polimerasa puedan eliminarse bioinformáticamente, y que además se puedan filtrar las duplicaciones de lecturas (Jauk, 2019). Todos los ensayos se realizaron en la Unidad Genómica del Instituto de Biotecnología – CICVyA (INTA) (Hurlingham, Buenos Aires) y con la supervisión y colaboración de la Dra. Andrea Puebla (directora de la unidad).

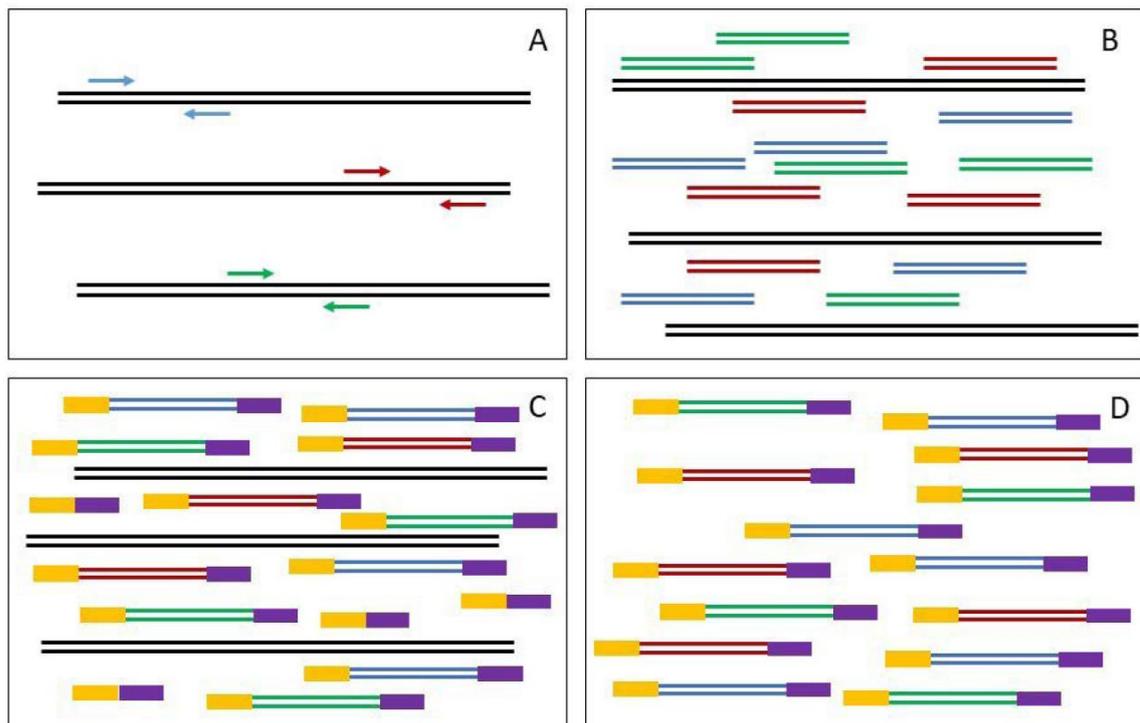


Figura 3. Enriquecimiento del blanco basado en amplicones (PCR multiplex): Las regiones de interés son amplificadas a partir de ADN genómico (doble hebra, color negro) por medio de cebadores específicos (flechas de color azul, rojo, verde) (A). Se obtienen así múltiples copias de las regiones amplificadas (doble hebra color azul, rojo, verde) (B). Luego se añaden los adaptadores para la construcción de la biblioteca (amarillo y violeta) (C). Posteriormente, la biblioteca es purificada para eliminar ADN genómico y los dímeros de adaptadores (D). Tomado y modificado de Jauk, F. (2019, October).

Secuenciación de próxima generación

Se llevó a cabo una secuenciación por síntesis en un secuenciador Miseq de Illumina, de la Unidad de Genómica antes mencionada. En la plataforma Illumina (Figura 4), las bibliotecas fueron cargadas en una celda de flujo. En esta etapa, cada inserto de la biblioteca fue amplificado clonalmente por medio de una PCR-puente, generando un grupo o cluster unido al soporte sólido, que fue utilizado

como templado. Posteriormente se llevó a cabo la secuenciación por síntesis con terminadores reversibles según el procedimiento descrito en la figura 4.

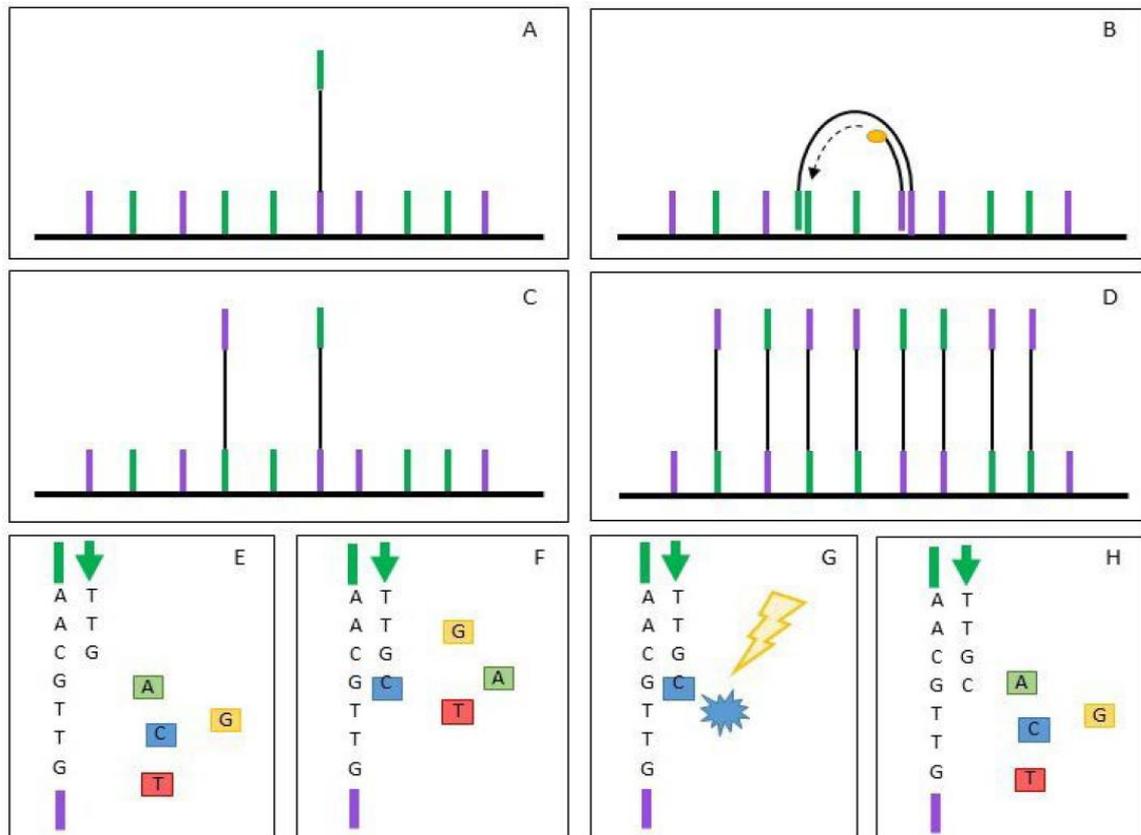


Figura 4. Secuenciación en plataformas Illumina. Luego de la preparación de la biblioteca se realiza amplificación de grupos (clusters) en celda de flujo (A a D). La biblioteca es diluida y cargada en la celda de flujo de modo que cada hebra de ADN se une a la celda de flujo por medio de hibridación de los adaptadores (verde y violeta) con oligonucleótidos complementarios a los mismos (unidos al soporte sólido). Mediante una reacción de polimerización se genera una copia del inserto desde el oligonucleótido unido al soporte sólido (A). Posteriormente se genera la PCR-puente: el fragmento se curva y el adaptador del extremo libre se une a un oligonucleótido complementario anclado al soporte sólido. Mediante otra reacción de polimerización (B) se obtiene una segunda hebra complementaria a la inicial (C). El proceso se repite hasta generar un grupo (cluster) a partir de la hebra primaria. La generación del grupo facilita la detección por medio de fluorescencia. La secuenciación propiamente dicha (secuenciación por síntesis) (E a H) se genera a partir de cebadores específicos para los adaptadores (flecha verde). En cada flujo, los nucleótidos marcados con fluorescencia (terminadores reversibles) compiten por elongar la cadena de ADN (E) y se une a ella sólo aquel complementario a la hebra molde (F). El resto de los nucleótidos es eliminado, y una fuente lumínica excita el fluoróforo del nucleótido añadido (G). La señal es recolectada por un dispositivo CCD para luego ser analizada. El ciclo vuelve a repetirse con un nuevo flujo para incorporar otro nucleótido a la cadena (H). Tomado y modificado de Jauk, F. (2019, October)

Análisis Bioinformático

Procesamiento de datos de secuencias e identificación de los RADseq-SNP

Para la primera parte del procesamiento de las secuencias (demultiplexado y ensamblaje) se utilizó el programa **Stacks** (<http://creskolab.uoregon.edu/stacks/>), un proceso computacional diseñado para trabajar con cualquier información de

genotipificado (GBS por sus siglas en inglés) basada en enzimas de restricción (Catchen et al., 2013). Stacks es computacionalmente robusto, eficiente y flexible y puede ensamblar lecturas cortas de novo o usar datos alineados con un genoma de referencia. El “pipeline” de Stack empleado se puede describir resumidamente de la siguiente manera:

i) Demultiplexación y filtrado de lecturas sin procesar:

El “demultiplexado” es un proceso que consiste en la clasificación del total de lecturas generadas agrupándolas en “pilas” (stacks) de acuerdo con los oligonucleótidos flanqueantes en las mismas. A su vez, se realiza un filtrado de lecturas de acuerdo con la calidad, presencia de etiquetas y sitios de restricción (Rochette & Catchen, 2017).

(ii) Identificación de polimorfismos de un sólo nucleótido (RADseq-SNPs).

Rochette & Catchen (2017) recomiendan explorar los datos con una submuestra y seleccionar así los parámetros de búsqueda e identificación de RADseq-SNPs. Entonces, se calibraron los parámetros con 11 individuos. Este paso se realizó con el objetivo de garantizar que se utilice un número suficientemente grande de lecturas, pero también reducir la proporción de loci, alineaciones y llamadas erróneas. La calibración de parámetros se realiza para la identificación de RADseq-SNPs, y filtrado de estos de acuerdo con la frecuencia genética y la representatividad a través de todas las muestras analizadas. De esta forma se obtiene el mayor número de loci polimórficos, y se evalúa la mayor diversidad genética posible.

Los parámetros que se calibraron son **M** (es el número de errores permitidos entre dos alelos de una muestra heterocigota) y **m** (número de lecturas idénticas necesarias para iniciar un nuevo alelo putativo). Estos parámetros deben ser lo suficientemente relajados para tener en cuenta la variación genética y los errores de secuenciación, pero lo suficientemente estrictos para discriminar entre loci parálogos. Se ejecutó una serie de análisis de novo con diferentes valores de parámetros y se monitoreó el número de loci RAD polimórficos encontrados en el 80% de las muestras o más (Rochette & Catchen, 2017), hasta obtener un conjunto estable de valores para **M** y **m**. Una vez optimizados, los parámetros se aplicaron en el total de las 32 muestras.

(iii) **Exportación del archivo generado para el análisis genético de poblaciones en R (R Core Team, 2020).**

Se compilaron los datos y se generó la matriz básica de datos (MBD).

Estimación de parámetros poblacionales en los paisajes

En cada paisaje se estimó la riqueza alélica, heterocigosis esperada y observada, el Fis (desviación promedio de las proporciones genotípicas de la población del equilibrio de Hardy-Weinberg) y el Fst (medida de la diferenciación de la población debido a la estructura genética). Estos parámetros fueron estimados mediante el paquete *hierfstat* (Goudet & Jombart, 2020) de R (R Core team, 2021). Además, para evaluar la diferenciación genética entre paisajes se calculó una matriz de índices de fijación (MiFst) utilizando la función *genet.dist* del paquete *hierfstat* (Goudet & Jombart, 2020). La matriz se graficó mediante un mapa de calor (“heat map”) con el paquete *ggplot2*.

Definir la escala de paisaje más apropiada para evaluar posibles relaciones entre los índices de diversidad y fijación genética y la de cantidad de hábitat seminatural en los paisajes y la distancia geográfica.

Análisis estadísticos para definir la escala óptima del paisaje

Con el objetivo de encontrar la escala espacial (radio de buffer 0,5; 1; 1,5 o 2 km) que mejor ajuste entre el factor de análisis (proporción de cantidad de hábitat seminatural a una escala en particular) y las variables de respuesta (riqueza alélica, H_o y H_e , Fst, y Fis) se realizó un análisis multiescala (Huais, 2018, y trabajos allí citados) mediante la función *multifit* en R. El código R completamente ejecutable y el manual detallado se proporcionan en <https://github.com/phuais/multifit/>. Específicamente se ejecutó un modelo estadístico para cada escala espacial especificada (factor fijo de análisis: proporción de hábitat seminatural a los 0,5; 1; 1,5 o 2 km, relacionándolo con las variables de respuesta). Se utilizó el criterio de información de Akaike (AIC), para seleccionar el mejor modelo. Una vez calculado el AIC para cada modelo se elige aquel cuyo AIC es mínimo, ya que indica que ese modelo, entre los comparados, es el que presentaría el mejor compromiso entre el ajuste del modelo teórico a los datos y el número de parámetros.

Relación entre índices de diversidad y estructura genética y atributos del paisaje (cantidad de hábitat y distancia geográfica)

Para obtener información acerca de la relación entre la proporción de hábitat natural en la escala seleccionada y cada una de las variables respuesta, se realizaron una serie de regresiones lineales simples utilizando la función *lm* del paquete *stats* (R Core team, 2021), las cuales además son una herramienta útil para predecir una respuesta cuantitativa. Se realizó una prueba de Mantel para analizar, la correlación entre una matriz de distancia geográfica con una matriz de distancia genética conformada por los valores de F_{st} según Nei (1987) entre todas las muestras. Esta prueba se realizó mediante el paquete *ade4* (Dray et al., 2007). Los gráficos fueron realizados en *ggplot2* (Wickham, 2016).

Explorar posibles patrones de estructuración genética poblacional para establecer posibles relaciones entre la cantidad de hábitat seminatural y la distancia geográfica con la estructura poblacional recuperada.

Identificación de poblaciones genéticas: análisis de ordenamiento y AMOVAs

Se implementó un análisis de distancia UPGMA utilizando la distancia de Nei a través de un bootstrap de los loci presentes en todos los individuos muestreados y el resultado se graficó en un fenograma. El UPGMA se estimó en el paquete *ape* (Paradis et. al., 2004) de R. Posteriormente, se realizó un análisis de componentes principales (PCA, por sus siglas en inglés) y se calculó la variabilidad genética (F_{st}) explicada por cada uno de sus ejes calculada en base al marcador RADseq-SNP de todos los individuos muestreados mediante la función *dudi.pca* del paquete *ade4* de R (Dray, Dufour & Chessel 2007). A partir de la distribución geográfica de los genotipos y los resultados obtenidos a partir de los análisis de agrupamiento estadístico UPGMA y PCA (ver más adelante), se pusieron a prueba distintas hipótesis de agrupamiento por medio de AMOVAs (Análisis Molecular de la Varianza). Paralelamente se realizó el cálculo de los F_{st} entre las hipótesis de agrupamiento. Los AMOVAs fueron realizados mediante el paquete *ade4* (Dray, Dufour & Chessel 2007).

Relación entre estructura genética poblacional y atributos del paisaje

Se estimó una MiF_{st} entre los grupos recuperados del AMOVA que más varianza entre poblaciones genéticas explicó. Se realizaron pruebas de Mantel para analizar correlaciones entre una matriz de distancia genética conformada por los valores de F_{st} según Nei (1987) y dos matrices de atributos del paisaje (i- matriz

de distancia geográfica, ii-matriz de distancia de proporción de hábitat seminatural en el área buffer a los 2 km de radio ya que resultó ser la escala con mejor ajuste. Estas pruebas se realizaron mediante el paquete *ade4*.

Resultados (se presentan según los objetivos específicos planteados)

Estimar la cantidad de hábitat seminatural en 11 paisajes con diferentes grados de pérdida de hábitat y donde crezca E. acerensis.

La cantidad de hábitat resultó diferente en los 11 paisajes. En la tabla 2 se detalla para cada uno de los diferentes buffers considerados la proporción de hábitat estimada.

Tabla 2: Se detalla la proporción de hábitat seminatural estimado para cada paisaje y las respectivas áreas buffers.

Paisaje	Radio del buffer (m)	Área buffer (ha)	Área Hábitat seminatural (ha)	Proporción De hábitat seminatural
T1	0,5	78,506	64,121	0,817
T2	0,5	78,506	18,241	0,232
T3	0,5	78,508	47,221	0,601
T4	0,5	78,506	12,107	0,154
T5	0,5	78,504	8,666	0,110
RC1	0,5	78,494	21,872	0,279
RC2	0,5	78,492	28,509	0,363
RC3	0,5	78,491	51,244	0,653
RC4	0,5	78,493	17,587	0,224
S1	0,5	78,492	36,544	0,466
S2	0,5	78,493	35,868	0,457
T1	1,0	314,025	204,8	0,652
T2	1,0	314,026	30,581	0,097
T3	1,0	314,03	80,731	0,257
T4	1,0	314,023	25,814	0,082
T5	1,0	314,016	16,22	0,052
RC1	1,0	313,976	66,856	0,213
RC2	1,0	313,969	54,798	0,175
RC3	1,0	313,964	237,03	0,755
RC4	1,0	313,972	117,207	0,373
S1	1,0	313,968	132,887	0,423
S2	1,0	313,973	129,216	0,412

T1	1,5	706,555	319,651	0,452
T2	1,5	706,558	52,206	0,074
T3	1,5	706,568	100,607	0,142
T4	1,5	706,552	38,043	0,054
T5	1,5	706,536	19,717	0,028
RC1	1,5	706,446	139,687	0,198
RC2	1,5	706,429	119,389	0,169
RC3	1,5	706,418	547,425	0,775
RC4	1,5	706,438	274,486	0,389
S1	1,5	706,427	319,426	0,452
S2	1,5	706,439	222,744	0,315
T1	2,0	1256,098	402,038	0,320
T2	2,0	1256,103	82,756	0,066
T3	2,0	1256,122	157,816	0,126
T4	2,0	1256,093	60,536	0,048
T5	2,0	1256,064	30,047	0,024
RC1	2,0	1255,905	164,048	0,131
RC2	2,0	1255,874	217,319	0,173
RC3	2,0	1255,854	909,035	0,724
RC4	2,0	1255,889	446,094	0,355
S1	2,0	1255,87	608,323	0,484
S2	2,0	1255,891	325,515	0,259

Estimar índices de diversidad genética y de fijación genética utilizando marcadores moleculares provenientes del genoma nuclear de E. acerensis.

Identificación de RADseq-SNPs

Se obtuvieron 200 millones de lecturas y se retuvieron el 98% de las mismas (196.085.248 lecturas) para los análisis posteriores. Los parámetros para la identificación de los RADseq-SNPs se estabilizaron en $M = 3$ y $m = 3$. Para M , tres diferencias explican el 80% de los loci polimórficos compartidos en las muestras (Figura 5a). Para m , tres es el número mínimo de lecturas para considerar un nuevo alelo putativo (Figura 5b). Se obtuvieron un total de 888.756 loci, el número de loci retenidos son aquellos RAD loci que pasaron el filtrado con los parámetros especificados (i.e. $M = 3$, $m = 3$ y loci presente en al menos el 80% de las muestras). De esta forma, se consideraron y dejaron finalmente 46.238 loci

disponibles para los análisis posteriores, excluyendo los loci problemáticos o inciertos.

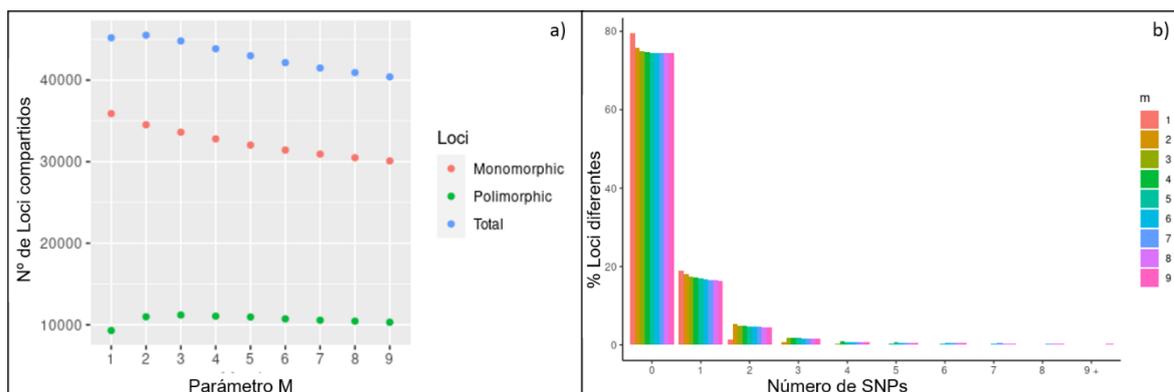


Figura 5. 5a. Número de Loci compartidos en el 80% de las muestras en función del parámetro M. **5b.** Porcentaje de loci diferentes en función al número de RADseq-SNPs.

Parámetros poblacionales

Los índices de diversidad y de fijación para cada paisaje, en base a los genotipos obtenidos en los respectivos lugares, se detallan en la tabla 3. La heterocigosis esperada varió entre los 0,04 y 0,25 mientras que la observada entre los 0,07 y 0,1. La riqueza alélica varió entre 1,04 y 1,2. El Fis varió entre -0,8 y 0,6 y el Fst entre 0,2 y 0,89.

Tabla 3: Índices poblacionales de diversidad y de fijación para los 11 paisajes seleccionados. Medida de diferenciación poblacional debido a estructura genética (Fst). Heterocigosis esperada (He), Heterocigosis observada (Ho). Desviación promedio de las proporciones genotípicas de la población del equilibrio de Hardy-Weinberg (Fis). Riqueza Alélica (RA).

Paisaje	Fst	He	Ho	Fis	RA
T1	0,88	0,04	0,07	-0,83	1,052
T2	0,87	0,04	0,08	-0,72	1,05
T3	0,61	0,13	0,07	0,29	1,118
T4	0,88	0,04	0,07	-0,73	1,049
T5	0,88	0,04	0,07	-0,78	1,047
RC1	0,71	0,10	0,08	0,05	1,095
RC2	0,61	0,13	0,07	0,30	1,118
RC3	0,23	0,25	0,07	0,63	1,211
RC4	0,30	0,22	0,10	0,41	1,199
S1	0,65	0,12	0,07	0,26	1,105
S2	0,89	0,04	0,07	-0,74	1,045

Los valores de Fst obtenidos por la matriz MiFst se muestran en la figura 6 (heatmap, o mapa de color) y variaron entre los 0,02 y 0.8.

Definir la escala de paisaje más apropiada para evaluar posibles relaciones entre los índices de diversidad y fijación genética y la de cantidad de hábitat seminatural en los paisajes y la distancia geográfica.

El mejor ajuste se obtuvo para la relación entre la cantidad de hábitat a los 2 km y la riqueza alélica (Figura 7a), la He (Figura 8a) y el Fst (Figura 9a). El Fis y la Ho, no fueron explicados por el factor fijo de análisis en ninguna escala (no se muestra). Las regresiones lineales indican que la riqueza alélica ($t = 2,87$, p valor= 0,02, Figura 7b) y la He ($t = 2,95$, p valor= 0,02, Figura 8b) se relacionaron positivamente con la cantidad de hábitat seminatural. El Fst se relacionó negativamente ($t = 2,95$, p valor= 0,02 Figura 9b) con la cantidad de hábitat seminatural.

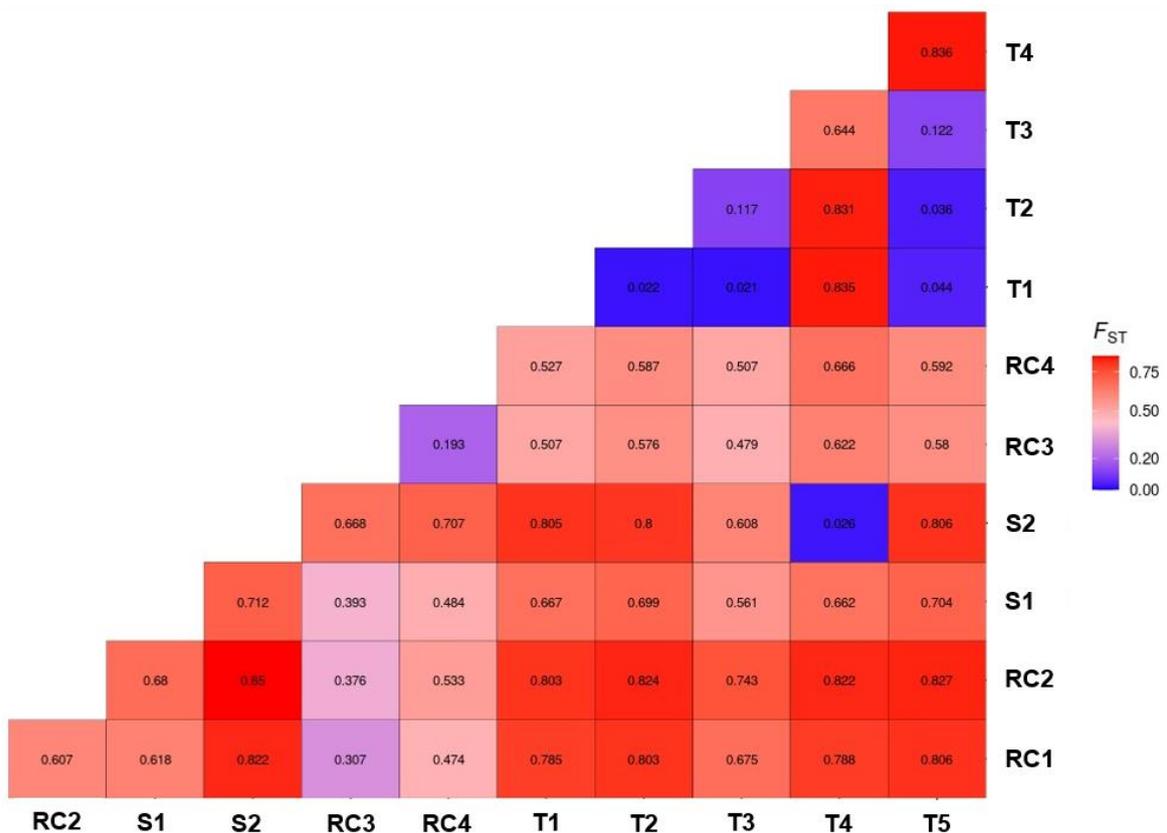


Figura 6. Mapa de calor (heatmap) mostrando los Fst entre los 11 sitios de muestreo. Los tonos rojos indican valores de Fst más alejados de cero y los tonos azules corresponden a valores más cercanos a cero.

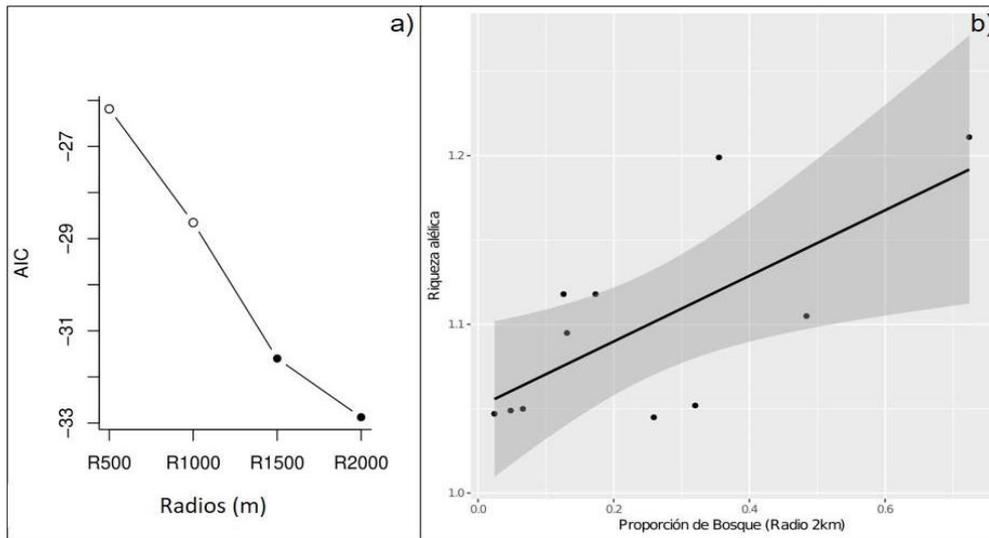


Figura 7. 7a, Ajuste de los modelos estadísticos en cada escala espacial especificada según el AIC para la riqueza alélica. Los puntos blancos representan modelos con coeficientes no significativos y los negros los significativos. **7b**, Curva de regresión de la riqueza alélica sobre la proporción de hábitat seminatural a los 2km.

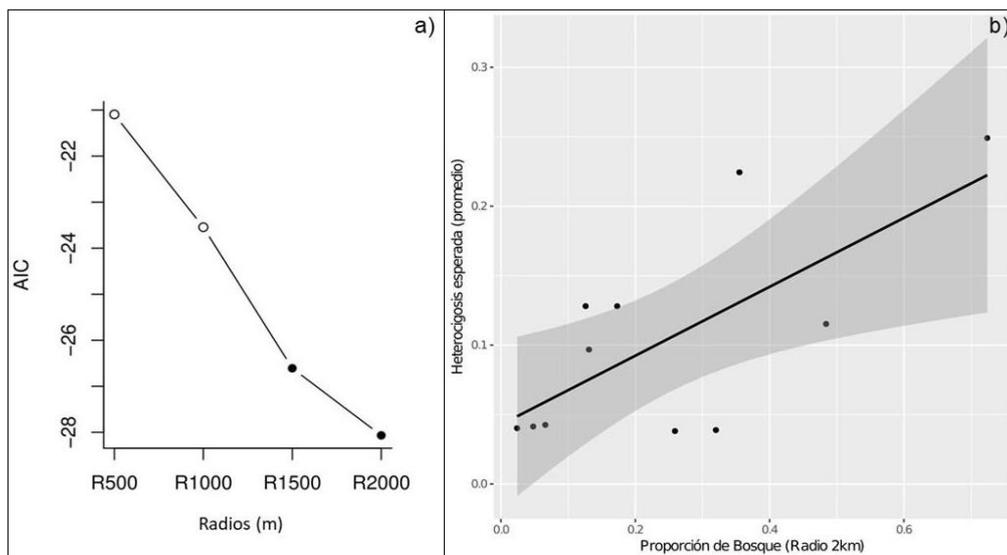


Figura 8. 8a, Ajuste de los modelos estadísticos en cada escala espacial especificada según el AIC para la heterocigosis esperada. Los puntos blancos representan modelos con coeficientes no significativos y los negros los significativos. **8b**, Curva de egresión de la He sobre la proporción de hábitat seminatural a los 2km.

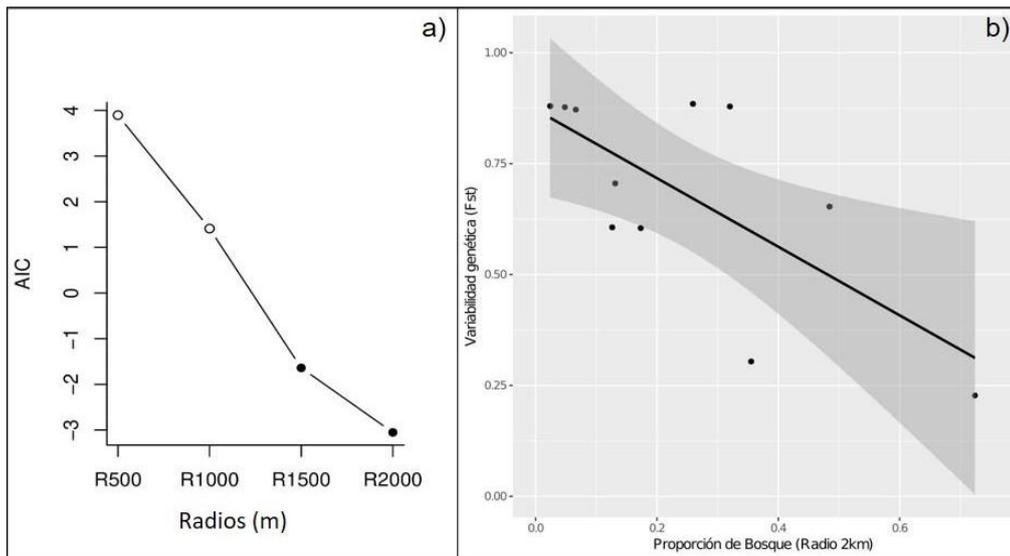


Figura 9. 9a, Ajuste de los modelos estadísticos en cada escala espacial especificada según el AIC para el Fst. Los puntos blancos representan modelos con coeficientes no significativos y los negros los significativos. **9b**, Curva de egresión de los Fst sobre la proporción de hábitat seminatural a los 2km.

La prueba de mantel entre distancia genética (obtenida por una matriz de Fst) y distancia geográfica entre todos los individuos muestreados resultó significativa ($r = 0,45$, $p = 0,001$), indicando un aislamiento por distancia geográfica (Figura 10).

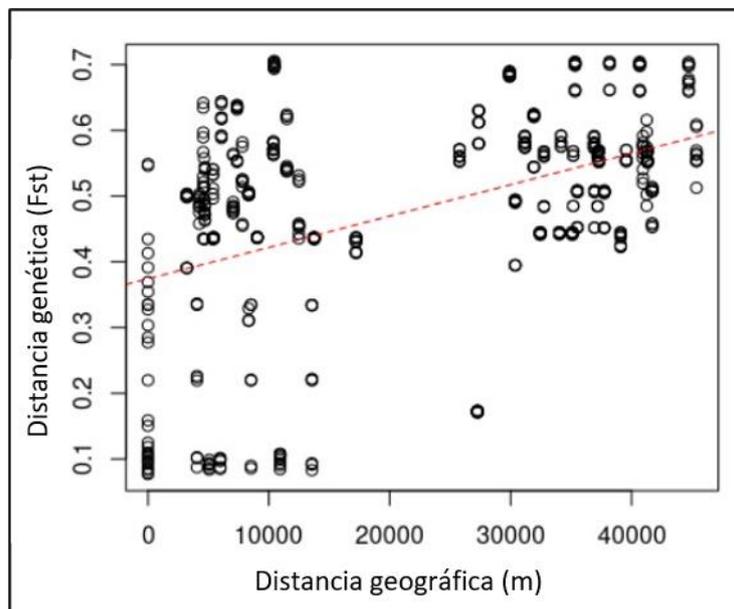


Figura 10. Prueba de mantel. Correlación entre distancia genética (obtenida por una matriz de Fst) y distancia geográfica entre todos los individuos muestreados.

Explorar posibles patrones de estructuración genética poblacional para establecer posibles relaciones entre la cantidad de hábitat seminatural y la distancia geográfica con la estructura poblacional recuperada.

Agrupamiento por distancia (UPGMA y PCA)

El resultado del UPGMA se muestra mediante un fenograma en la figura 11a. Con una distancia genética de hasta 0.03 se diferencian 2 grupos: **A**- conformado por todos los genotipos provenientes de los paisajes ubicados al Este del área de estudio (localidad de Tinoco) y tres genotipos provenientes del Oeste del área de estudio (localidad de Salsipuedes); **B**- conformado por todos los genotipos restantes que provienen de los paisajes del Oeste (localidades de Rio Ceballos y Salsipuedes). Considerando el valor de 0.10 de distancia genética, el grupo **A** se diferencia en dos, a- uno que incluye a 11 genotipos de Tinoco y b- que incluye a tres genotipos de Tinoco y tres de Salsipuedes. A la par, considerando el valor de 0.10 de distancia genética el grupo **B** se subdivide en 6 subgrupos, incluyendo siempre los genotipos del Oeste de área de estudio (Rio Ceballos y Salsipuedes) (Figura 11a). El PCA que resume en cada uno de sus ejes el mayor porcentaje de varianza de los datos en esa dimensión se muestra en la figura 11b. En conjunto, los dos primeros ejes explican 60% de la variabilidad genética total. El eje 1 explica el 41.3% y permitió discriminar dos grupos, separando los genotipos de Rio Ceballos y tres genotipos de Salsipuedes (en adelante grupo RCSALSI) del resto de los genotipos. El eje 2 explicó el 19,3% y separó dos grupos: uno con tres genotipos de Salsipuedes y tres de Tinoco (en adelante SALSITIN) y otro, con los genotipos de Tinoco (en adelante TIN).

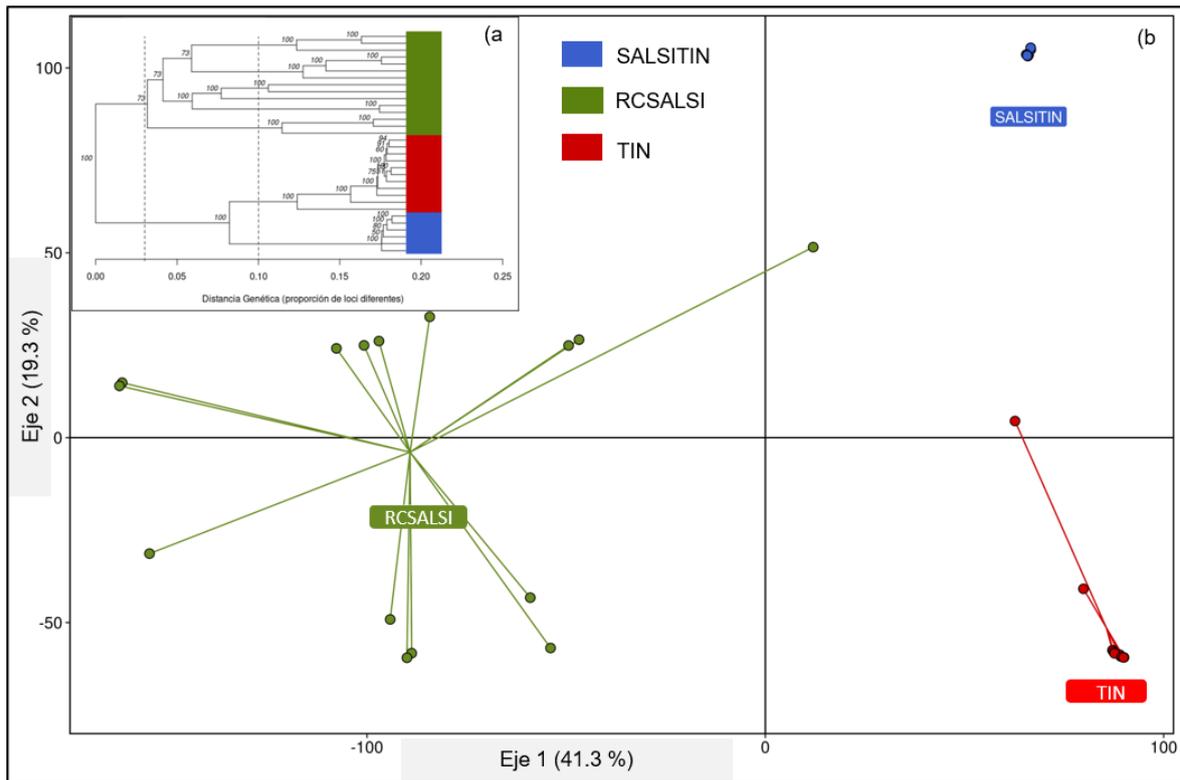


Figura 11: 11a, Fenograma; Los números sobre las ramas indican el soporte Bootstrap de cada grupo. Las líneas punteadas verticales indican las distancias genéticas a 0,03 y 0,1 basada en la proporción de loci diferentes. En verde los genotipos de Rio Ceballos y Salsipuedes. En rojo genotipos de Tinoco. En azul genotipos de Tinoco y Salsipuedes. **11b, PCA;** En el grafico se observa el primer y segundo eje de las componentes principales: PC1 = 41.3% y PC2 = 19.3%. En verde los genotipos de Rio Ceballos y Salsipuedes (grupo RCSALSI). En rojo genotipos de Tinoco (grupo TIN). En azul genotipos de Tinoco y Salsipuedes (grupo SALSITIN)

Estructura genética poblacional y sus relaciones con atributos del paisaje

Los resultados de todos los AMOVAs y los F_{st} correspondientes se muestran en la Tabla 4, Figura 12 y **Anexo IV**). Los grupos RCSALSI, TIN y SALSITIN (en igual composición a la recuperada en el PCA), valores más altos de F_{st} (Figura 12h). Los grupos TIN y SALSITIN también pueden ser considerados como subpoblaciones de un grupo mayor (Figura 12e). La matriz MiF_{st} y su heat map asociado, indicó mayor diferenciación entre TIN y SALSITIN (Figura 13).

Tabla 4. AMOVAs con agrupamientos alternativos para evaluar posible estructuración genética entre los genotipos de *E. acerensis* incluidos en este estudio. Entre corchetes [] se distinguen las [poblaciones] y entre paréntesis () se distinguen las (subpoblaciones) dentro de cada población. El + separa subpoblaciones. La / separa genotipos de paisajes distintos que fueron incluidos en la misma población o subpoblación.

	Obs	Std.Obs	P valor
a) Dos Poblaciones, (1 con 3 subpoblaciones)			
[RC] vs [(T)+(S1)+(S2/T4)]			
Variación dentro de las muestras	743,099	-11,562	0,01
Variación entre las muestras	736,924	5,509	0,01
Variaciones entre las Subpoblaciones	1807,872	13,297	0,01
Variaciones entre las poblaciones	674,837	1,298	0,27
b) Dos poblaciones, (ambas con 2 subpoblaciones)			
[(RC)+(S1)] vs [(T)+(S2/T4)]			
Variación dentro de las muestras	743,099	-9,426	0,01
Variación entre las muestras	736,924	5,531	0,01
Variaciones entre las Subpoblaciones	1555,214	13,482	0,01
Variaciones entre las poblaciones	988,861	1,673	0,01
c) Dos poblaciones			
[RC/S1] vs [T/S2]			
Variación dentro de las muestras	743,099	-11,232	0,01
Variación entre las muestras	1388,284	9,494	0,01
Variaciones entre las poblaciones	1899,202	18,185	0,01
d) Dos poblaciones			
[RC/S] vs [T]			
Variación dentro de las muestras	743,099	- 10,3821	0,01
Variación entre las muestras	1.640,892	7,8734	0,01
Variaciones entre las poblaciones	1.424,615	10,4312	0,01
e) Dos poblaciones, (1 con 2 subpoblaciones)			
[RC/S1] vs [(T) +(S2/T4)]			
Variación dentro de las muestras	743,099	- 11,407	0,01
Variación entre las muestras	944,972	7,724	0,01
Variaciones entre las Subpoblaciones	1712,797	12,622	0,01
Variaciones entre las poblaciones	605,378	1,350	0,01
f) Dos poblaciones, (1 con 2 subpoblaciones)			
[RC] vs [(T)+(T4/S)]			

Variación dentro de las muestras	743,099	-11,540	0,01
Variación entre las muestras	976,857	6,540	0,01
Variaciones entre las Subpoblaciones	1409,504	10,631	0,01
Variaciones entre las poblaciones	877,259	1,198	0,01
<i>g) Dos poblaciones, (1 con 2 subpoblaciones)</i>			
[RC] vs [(T)+(S)]			
Variación dentro de las muestras	743,099	-10,635	0,01
Variación entre las muestras	132,185	1,996	0,02
Variaciones entre las Subpoblaciones	1913,681	13,822	0,01
Variaciones entre las poblaciones	1344,634	2,015	0,01
<i>h) Tres poblaciones</i>			
[RC/S1] + [T] + [S2/T4]			
Variación dentro de las muestras	743,099	-11,619	0,01
Variación entre las muestras	944,972	7,841	0,01
Variaciones entre las poblaciones	2193,705	16,681	0,01

a)	<table border="1"><tr><th>Fst</th><th>(S1)</th><th>(S2/T4)</th><th>[SD]</th></tr><tr><td>(S2/T)</td><td>0,739</td><td></td><td></td></tr><tr><td>[RC]</td><td>0,287</td><td>0,570</td><td></td></tr><tr><td>(T)</td><td>0,693</td><td>0,735</td><td>0,547</td></tr></table>	Fst	(S1)	(S2/T4)	[SD]	(S2/T)	0,739			[RC]	0,287	0,570		(T)	0,693	0,735	0,547	b)	<table border="1"><tr><th>Fst</th><th>(S2/T4)</th><th>(RC)</th><th>[(RC) + (S1)]</th></tr><tr><td>(RC)</td><td>0,570</td><td></td><td></td></tr><tr><td>[(RC) + (S1)]</td><td>0,740</td><td>0,288</td><td></td></tr><tr><td>(T)</td><td>0,735</td><td>0,547</td><td>0,693</td></tr></table>	Fst	(S2/T4)	(RC)	[(RC) + (S1)]	(RC)	0,570			[(RC) + (S1)]	0,740	0,288		(T)	0,735	0,547	0,693
Fst	(S1)	(S2/T4)	[SD]																																
(S2/T)	0,739																																		
[RC]	0,287	0,570																																	
(T)	0,693	0,735	0,547																																
Fst	(S2/T4)	(RC)	[(RC) + (S1)]																																
(RC)	0,570																																		
[(RC) + (S1)]	0,740	0,288																																	
(T)	0,735	0,547	0,693																																
c)	<table border="1"><tr><th>Fst</th><th>[RC/S1]</th></tr><tr><td>[T/S2]</td><td>0,460</td></tr></table>	Fst	[RC/S1]	[T/S2]	0,460	d)	<table border="1"><tr><th>Fst</th><th>[T]</th></tr><tr><td>[RC/S1]</td><td>0,341</td></tr></table>	Fst	[T]	[RC/S1]	0,341	e)	<table border="1"><tr><th>Fst</th><th>(S2/T4)</th><th>[RC/S1]</th></tr><tr><td>[RC/S1]</td><td>0,506</td><td></td></tr><tr><td>(T)</td><td>0,735</td><td>0,491</td></tr></table>	Fst	(S2/T4)	[RC/S1]	[RC/S1]	0,506		(T)	0,735	0,491													
Fst	[RC/S1]																																		
[T/S2]	0,460																																		
Fst	[T]																																		
[RC/S1]	0,341																																		
Fst	(S2/T4)	[RC/S1]																																	
[RC/S1]	0,506																																		
(T)	0,735	0,491																																	
f)	<table border="1"><tr><th>Fst</th><th>(T4/S)</th><th>[RC]</th></tr><tr><td>[RC]</td><td>0,442</td><td></td></tr><tr><td>(T)</td><td>0,542</td><td>0,547</td></tr></table>	Fst	(T4/S)	[RC]	[RC]	0,442		(T)	0,542	0,547	g)	<table border="1"><tr><th>Fst</th><th>[(T) + (S)]</th></tr><tr><td>[RC]</td><td>0,446</td></tr></table>	Fst	[(T) + (S)]	[RC]	0,446	h)	<table border="1"><tr><th>Fst</th><th>[S2/T4]</th><th>[RC/S1]</th></tr><tr><td>[RC/S1]</td><td>0,506</td><td></td></tr><tr><td>[T]</td><td>0,735</td><td>0,491</td></tr></table>	Fst	[S2/T4]	[RC/S1]	[RC/S1]	0,506		[T]	0,735	0,491								
Fst	(T4/S)	[RC]																																	
[RC]	0,442																																		
(T)	0,542	0,547																																	
Fst	[(T) + (S)]																																		
[RC]	0,446																																		
Fst	[S2/T4]	[RC/S1]																																	
[RC/S1]	0,506																																		
[T]	0,735	0,491																																	

Figura 12 (a-h). Tablas de Fst correspondientes a los agrupamientos puestos a prueba mediante los AMOVAs (ver tabla 4). En recuadro verde se resalta la estructuración e y h, que son igualmente consistentes con los resultados obtenidos por análisis de distancia y que presentan iguales valores de Fst. En **12h** se presenta la combinación equivalente a los grupos RCSALSI, TIN y SALSITIN. En **12e** se presentan los valores de Fst al considerar a TIN y SALSITIN como subpoblaciones de una población inclusiva.

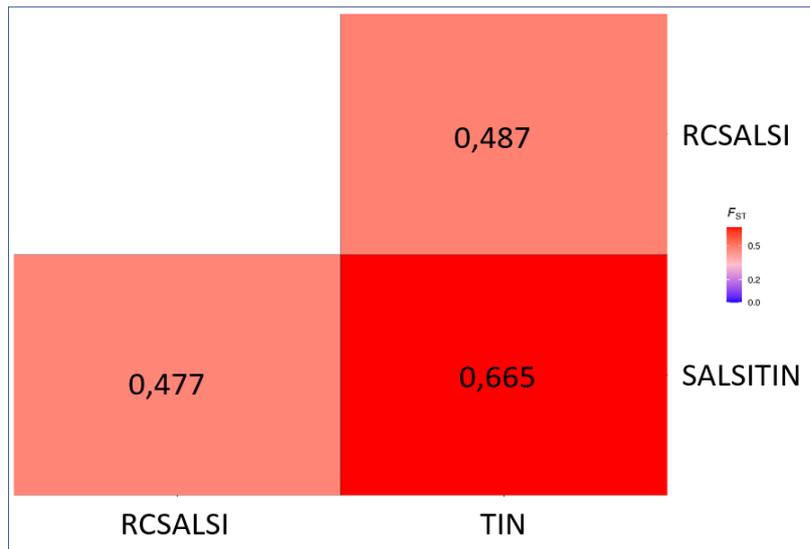


Figura 13. Mapa de calor (heatmap) muestra los F_{ST} entre las tres poblaciones: RCSALSI, TIN y SALSITIN. Los tonos rojos indican valores de F_{ST} más altos y los tonos azulados corresponden a valores más cercanos a cero.

De acuerdo con los resultados obtenidos, en la figura 14 presentamos la composición y distribución de las tres poblaciones genéticas recuperadas.



Figura 14. Composición y distribución de las poblaciones recuperadas en este estudio: RCSALSI, TIN y SALSITIN.

Al correlacionar la matriz de F_{ST} , entre individuos de cada población genética (RCSALSI y TIN), con la matriz de proporción de cantidad de hábitat a los 2 km, se encontraron relaciones marginalmente significativas ($p = 0,05$, permutaciones = 999) en la población RCSALSI (Figura 15a). Para la población TIN no se encontró relaciones significativas entre las matrices (Figura 15c). Al correlacionar la matriz

de F_{st} entre individuos de cada población genética (RCSALSI, TIN) con la matriz de distancia geográfica, se encontraron relaciones significativas ($p = 0,002$, permutaciones = 999) en la población RCSALSI (Figura 15b). No se encontraron relaciones significativas entre estas matrices en las poblaciones de TIN (Figura 15d). No se incluyó a la población SALSITIN en estos análisis por que sólo dos paisajes conforman esta población (datos insuficientes).

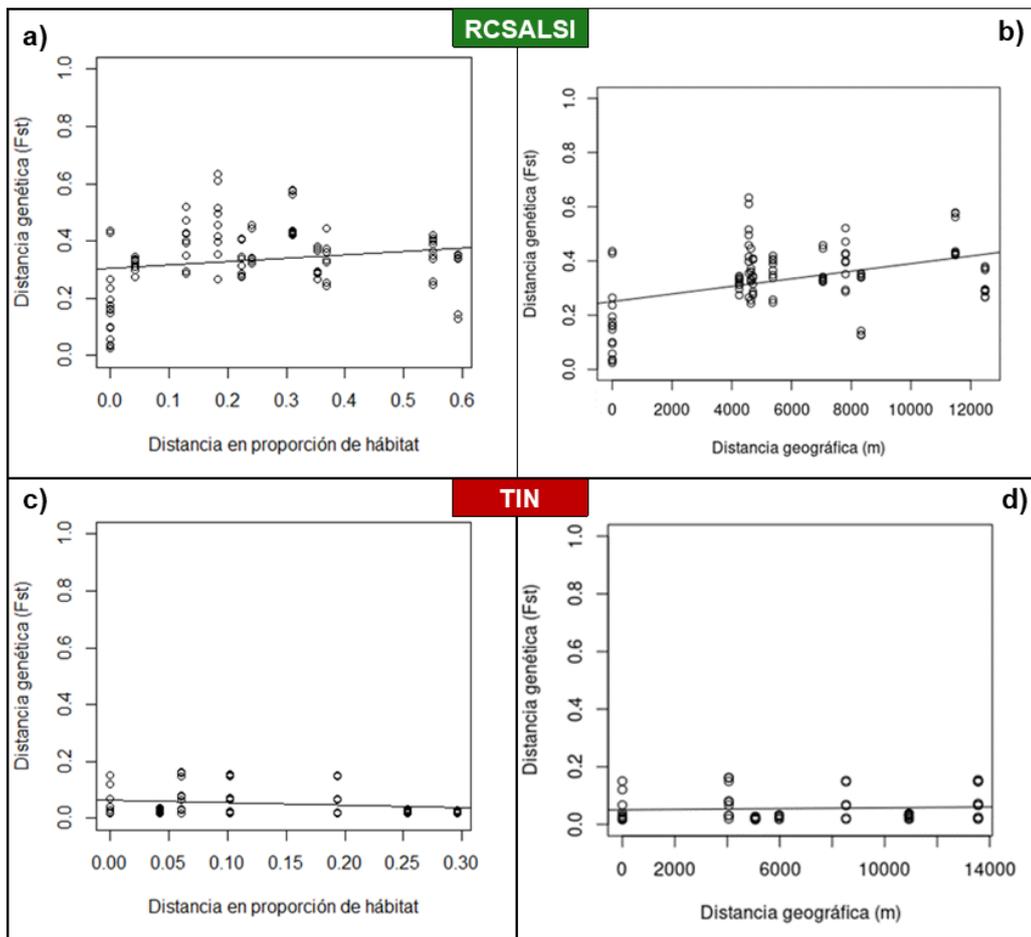


Figura 15. Relaciones entre matrices de F_{st} para cada población genética y proporción de cantidad de hábitat y distancia geográfica.

Discusión

Los paisajes incluidos en este estudio presentaron diferentes cantidades de hábitat y esta variación espacial se habría originado durante los últimos 100 años como productos de diferentes actividades antrópicas (Boletta et al., 2006; Zak et al., 2008). La elección de los marcadores moleculares utilizados en los estudios de genética del paisaje tiene una fuerte influencia en las escalas espacial y temporal sobre las que es posible hacer inferencias sobre el flujo de genes. Por ejemplo, los microsatélites son eficientes para estudios de estructura genética a escala fina y en tiempo ecológico (Cavers et al., 2005). Sin embargo, estos marcadores no resultaron informativos en *Euphorbia acerensis* (Amarilla et al., 2010). Por ello, en este trabajo se ha optado por utilizar marcadores tipo RADseq-SNP (polimorfismos de un solo nucleótido).

Los RADseq-SNP son utilizados en organismos no modelo, y para estudios de genética del paisaje (e.g. Manel et al., 2010; Anderson et al., 2010; Stojanova et al., 2020) y, cuando se logra acceder a grandes conjuntos de datos, los RADseq-SNP permiten detectar incluso estructura genética incipiente, especialmente en especies con ciclo de vida anual (Anderson et. al., 2010). El número de lecturas obtenidas y filtradas, el número de loci polimórficos encontrados y el valor de los parámetros **M** y **m** estimados, son consistentes con reportes previos para organismos vegetales no modelo (e.g., para identificar marcadores informativos específicos) (Blanc-Jolivet et. al., 2020), para mejoramiento vegetal (Holdsworth et. al., 2017), y para estudios de estructura poblacional (Mehravi et. al., 2021) y/o de genética del paisaje (Xu et. al., 2017; Toczydlowski & Waller 2019).

Los valores de Heterocigosis y de Fis indican que los grupos de genotipos coleccionados en cada paisaje no están en equilibrio de Hardy-Weinberg (H-W), ya sea por exceso (paisajes en Tinoco) o déficits de heterocigotas (paisajes en Río Ceballos). Esto puede deberse a múltiples causas, entre ellas: -la sexualidad y/o el sistema de incompatibilidad de *E. acerensis*, ya que muchas morfoespecies de *Euphorbia* L. presentan protoginia y algunas son auto-incompatibles, estrategias que favorece la fecundación cruzada/exogamia (Narbona et al., 2011) y consecuentemente podría aumentar la heterosis ($Fis < 0$), como se observa en general en los paisajes ubicados en Tinoco; -procesos poblacionales como la deriva genética y consecuentemente un aumento en los homocigotas ($Fis > 0$) como se observa en los paisajes de Río Ceballos (Storz et al., 2001; Stoeckel et.

al., 2006; Parreira et al. 2020). Estudios sobre la biología reproductiva de la *E. acerensis* y análisis que contemplen los procesos poblacionales antes mencionados son necesarios para analizar y comprender mejor los resultados obtenidos (un estudio de adaptación local involucraría un diseño experimental particular).

Las características del paisaje existen en múltiples escalas espaciales y temporales, y naturalmente afectan la estructura genética espacial y nuestra capacidad para hacer inferencias sobre el flujo génico (Anderson et. al., 2010). Sin embargo, no es común conocer a priori la escala espacial adecuada a la que se debe examinar una determinada variable de respuesta, principalmente debido a la falta de una comprensión completa de una respuesta biológica dada (Jackson & Fahrig, 2015). La selección de una escala espacial para una determinada especie suele justificarse en la literatura bajo ciertas hipótesis, como la movilidad de las especies, su biología reproductiva, mecanismos de dispersión o los efectos de la densidad de población, etc., sin embargo, en muchos casos dichas hipótesis no han sido probadas empíricamente (Miguet et al., 2016) o, no se cuenta con toda esa información, como es en el caso de *E. acerensis*. Para estos casos es recomendable evaluar estadísticamente *a priori* la escala más adecuada para analizar una variable de respuesta (Huais, 2018). En este trabajo se evaluó *a priori* la escala a la que se deben analizar las variables de respuesta que aquí se analizaron y resultó significativo el modelo que incluyó la proporción de cantidad de hábitat a los 2 km.

Nuestros resultados indican que la riqueza alélica es mayor en los paisajes con mayores proporciones de hábitat. Esto podría deberse a que en estos paisajes con mayor cantidad de hábitat (al compararlos con los de menor cantidad de hábitat) podrían ocurrir con más frecuencia los procesos de polinización y fecundación cruzadas y/o dispersión de semillas, los que en conjunto favorecerían el flujo génico. *E. acerensis* tiene semillas carunculadas y con mucílago, las que suelen ser dispersadas por hormigas, y una inflorescencia que es visitada por moscas, hormigas y abejas. En otras especies vegetales dispersadas por animales, la pérdida de hábitat amenaza la dispersión de sus semillas (Cazetta & Fahrig, 2022). La heterocigosis no resultó explicada por el factor fijo de análisis. Esto puede deberse a que, previo al proceso de pérdida de hábitat de la zona, se hayan fijado grupos de genes ligados y heterocigotas.

En nueve de los 11 paisajes estudiados, los valores de F_{st} , estuvieron por encima de los 0,55 y los 11 están relacionados negativamente con la cantidad de hábitat, es decir, los paisajes con menores proporciones de cantidad de hábitat poseen mayores valores de F_{st} . Esto podría deberse a la fijación alternativa de alelos por restricción en el flujo de genes entre paisaje y la consecuente diferenciación entre ellos. Los dos km de radio en el que se definió el umbral de radio del paisaje pueden estar relacionado con la distancia de movilidad de los dispersores de semillas y la distancia de movilidad de los polinizadores (polinizadores nativos tienen radio de movilidad de 0,5 y 2 km, Beekma & Ratnieks, 2000). Además, los paisajes con mayor proporción de hábitat podrían sostener mayor densidad de individuos de *E. acerensis* y procesos de polinización mediada por polinizadores y dispersión de semillas mediada por animales más frecuentes y, en consecuencia, mayor flujo génico dentro del paisaje (valores de F_{st} más bajos).

Los valores de F_{st} calculados mediante la matriz MiF_{st} y la prueba de Mantel indican un aislamiento entre paisajes aledaños y un aislamiento por distancia. Esto no se cumple en el caso de los genotipos pertenecientes a los paisajes S2 y T4 (figura 7, $F_{st} = 0,026$), que conforman el grupo SALSITIN, y puede deberse a eventos de dispersión antrópicas reciente (la dispersión por medio de las máquinas agrícolas podría explicar la afinidad entre estos genotipos de paisajes distantes.). Un mayor esfuerzo de muestreo en Salsipuedes o a distancias intermedias entre las dos áreas (Salsipuedes – Tinoco) de muestreo incluidas en este estudio serían necesarios para verificar este agrupamiento.

Los análisis de agrupamiento por distancia recuperan tres grupos: una denominada RCSALSI e incluye a todos los genotipos coleccionados al Oeste de la zona de estudio (estancia Santo Domingo) y una del Noroeste (Salsipuedes), otra denominada TIN e incluye a los genotipos del Este (menos a tres, Tinoco), y la tercera denominada SALSITIN que incluye a tres genotipos provenientes del Noreste (Salsipuedes) y tres genotipos del Este (Tinoco). Los AMOVAs indican estructuración poblacional y los grupos RCSALSI, SALSITIN, y TIN se constituirían en poblaciones genéticas. Alternativamente SALSITIN y TIN podrían considerarse como dos subpoblaciones de una población más inclusiva, siempre distinta a RCSALSI. El MiF_{st} recuperó un F_{st} elevado al comparar SALSITIN y TIN. Preliminarmente consideraremos que los genotipos incluidos en este estudio están agrupados en tres poblaciones genéticas (Figura 14). Sin embargo, mayor

esfuerzo de muestreo, o análisis complementarios tales como una red de alelos y/o una red de haplotipos (construida con marcadores cloroplásticos) serían necesario para sustentar esta propuesta. Las redes nos ayudarían a determinar si los individuos de la población SALSITIN están conectadas por eventos de colonización reciente o por compartir polimorfismos ancestrales. Estudios demográficos también podrían ayudar a descifrar la historia genética reciente de estas agrupaciones. La diferenciación genética entre los genotipos de la población RCSALSI, es explicada tanto por distancia geográfica como por las diferencias en la cantidad de hábitat. Por lo que la cantidad de hábitat y la distancia geográfica entre los paisajes influyen en el flujo génico y la estructuración genética poblacional. Resultados e interpretaciones similares fueron reportados en *Ricinocarpos brevis* (Euphorbiaceae) (Krauss & Anthony, 2018). La diferenciación genética entre los genotipos de la población TIN, no son explicadas por la distancia geográfica y ni por las diferencias en la cantidad de hábitat. Resulta necesario incluir análisis de correlación entre la distancia geográfica y la cantidad de hábitat, o bien modelar regresiones múltiples para conocer la influencia conjunta de los factores de análisis en las variables de respuesta. A la par, se podrían incluir más paisajes en el estudio y que los mismos estén en un gradiente de distancia intermedia entre las localidades de Salsipuedes y Tinoco.

Conclusiones

Los marcadores moleculares nucleares tipo RADseq-SNP utilizados en *E. acerencis* permitieron obtener una buena cantidad de loci informativos para estimar índices de diversidad genética e índices de fijación poblacional. Todos los paisajes presentaron diferentes cantidades de hábitat y la escala de análisis adecuada resultó en un buffer de paisaje de dos kilómetros de radio. La riqueza alélica varió positivamente con la cantidad de hábitat. El Fis indicó que los genotipos incluidos en todos los paisajes están en desequilibrio de Hardy-Weinberg por exceso ($F_{is} < 0$) o déficits de heterocigotos ($F_{is} > 0$). Las variaciones del Fis resultaron independientes a la cantidad de hábitat. El flujo de genes está restringido entre los paisajes (F_{st} elevados) y tanto la cantidad de hábitat como la distancia geográfica influyen en esta restricción. Podemos indicar que se recuperaron efectos significativos de la cantidad de hábitat seminatural y la distancia geográfica sobre la diversidad y estructura genética en todos los paisajes. Los individuos incluidos en este estudio pueden formar tres poblaciones genéticas ([RCSALSI] [TIN] [SALSITIN]) o bien, pueden formar dos poblaciones genéticas (una de ellas con dos subpoblaciones): i- [RCSALSI], ii- [TIN-SALSITIN]). La cantidad de hábitat y la distancia geográfica resultaron influyentes en la estructura genética de [RCSALSI]. Se recuperaron efectos significativos de la cantidad de hábitat seminatural y la distancia geográfica sobre la diversidad y estructura genética en una de las poblaciones genéticas recuperadas.

Referencias Bibliográficas

- Amarilla LD, Anton AM, Chiapella JO, Manifesto M M, Angulo DF & Sosa V (2015) *Munroa argentina*, a grass of the South American transition zone, survived the Andean uplift, aridification and glaciations of the Quaternary. PloS one, 10(6), e0128559. <https://doi.org/10.1371/journal.pone.0128559>
- Amarilla LD, Grilli G, Chiapella J & Galetto L (2010) XIV Congreso Latinoamericano de Genética/ XXXIX Congreso de la Sociedad Argentina de Genética (SAG). Viña del Mar, Chile. "Estimación de parámetros poblacionales mediante Microsatélites (STR) en poblaciones fragmentadas de dos especies de *Euphorbia*".
- Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James PM, Rosenberg MS, ... & Spear S (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. Molecular ecology, 19(17), 3565-3575. <https://doi.org/10.1111/j.1365-294X.2010.04757.x>
- Beekman M, & Ratnieks FL W (2000) Long-range foraging by the honey-bee, *Apis mellifera* L. Functional Ecology, 14(4), 490-496. <https://doi.org/10.1046/j.1365-2435.2000.00443.x>
- Blanc-Jolivet C, Bakhtina S, Yanbaev R, Yanbaev Y, Mader M, Guichoux, E & Degen B (2020) Development of new SNPs loci on *Quercus robur* and *Quercus petraea* for genetic studies covering the whole species' distribution range. Conservation Genetics Resources, 12(4), 597-600. <https://doi.org/10.1007/s12686-020-01141-z>
- Boletta PE, Ravelo AC, Planchuelo AM & Grilli M (2006) Assessing deforestation in the Argentine Chaco. Forest Ecology and Management, 228(1-3), 108-114. <https://doi.org/10.1016/j.foreco.2006.02.045>
- Cabrera AL (1976) Regiones fitogeográficas argentinas.
- Catchen JM, Amores A, Hohenlohe P, Cresko W & Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes|genomes|genetics, 1(3), 171-182. <https://doi.org/10.1534/g3.111.000240>
- Cavers S, Degen B, Caron H, Lemes MR, Margis R, Salgueiro F & Lowe AJ (2005) Optimal sampling strategy for estimation of spatial genetic structure in tree populations. Heredity, 95(4), 281-289. <https://doi.org/10.1038/sj.hdy.6800709>
- Cazetta E & Fahrig L (2022) The effects of human-altered habitat spatial pattern on frugivory and seed dispersal: a global meta-analysis. Oikos, 2022(2). <https://doi.org/10.1111/oik.08288>
- Crutzen PJ & Stoermer EF (2000) Global change newsletter. The Anthropocene, 41, 17-18. https://doi.org/10.1007/3-540-26590-2_3
- Dale MR & Fortin MJ (2014) Spatial analysis: a guide for ecologists. Cambridge University Press.
- Dray S, Dufour AB & Chessel D (2007) The ade4 package-II: Two-table and K-table methods. R news, 7(2), 47-52.
- Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annual review of ecology, evolution, and systematics, 34(1), 487-515. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132419>
- Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. Journal of Biogeography, 40(9), 1649-1663. <https://doi.org/10.1111/jbi.12130>
- Fahrig L (2017) Ecological responses to habitat fragmentation per se. Annual Review of Ecology, Evolution, and Systematics, 48, 1-23. <https://doi.org/10.1146/annurev-ecolsys-110316-022612>
- Fahrig L (2018) Habitat fragmentation: A long and tangled tale. Global Ecology and Biogeography, 28(1), 33-41. <https://doi.org/10.1111/geb.12839>
- Fletcher Jr RJ, Didham RK, Banks-Leite C, Barlow J, Ewers RM, Rosindell J, ... & Melo FP (2018) Is habitat fragmentation good for biodiversity?. Biological conservation, 226, 9-15. <https://doi.org/10.1016/j.biocon.2018.07.022>
- Goudet J & Jombart T (2020) hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22. 2015. <https://www.r-project.org/>

- Grilli G, Urcelay C & Galetto L (2013) Linking mycorrhizal fungi and soil nutrients to vegetative and reproductive ruderal plant development in a fragmented forest at central Argentina. *Forest ecology and management*, 310, 442-449. <https://doi.org/10.1016/j.foreco.2013.08.052>
- Hamrick JL & Godt MW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1345), 1291-1298. <https://doi.org/10.1098/rstb.1996.0112>
- Holdsworth WL, Gazave E, Cheng P, Myers JR, Gore MA, Coyne CJ, ... & Mazourek M (2017) A community resource for exploring and utilizing genetic diversity in the USDA pea single plant plus collection. *Horticulture research*, 4. <https://doi.org/10.1038/hortres.2017.17>
- Huais PY (2018) multifit: an R function for multi-scale analysis in landscape ecology. *Landscape Ecology*, 33(7), 1023-1028. <https://doi.org/10.1007/s10980-018-0657-5>
- Jackson ND & Fahrig L (2016) Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. *Landscape Ecology*, 31(5), 951-968. <https://doi.org/10.1007/s10980-015-0313-2>
- Jauk F (2019). Secuenciación masiva paralela (NGS): conceptos básicos y aplicaciones. In *Hematología: Volumen 23-Extraordinario XXIV Congreso Argentino* (Vol. 4, p. 21). Sociedad Argentina de Hematología.
- Keitt TH & Urban DL (2005) Scale-specific inference using wavelets. *Ecology*, 86(9), 2497-2504.
- Krauss SL & Anthony J (2018) Genetic impacts of habitat loss on the rare Banded Ironstone Formation endemic *Ricinocarpos brevis* (*Euphorbiaceae*). *Australian Journal of Botany*, 67(3), 183-193. <https://doi.org/10.1071/BT18131>
- Kumar DA, Anup C, Santan B, Ajay T & Mishra RJ (2018) DNA extraction and molecular characterization of *Acacia pseudoeburnea*—An endemic species. *Research Journal of Biotechnology* Vol, 13, 8.
- Manel S & Holderegger R (2013) Ten years of landscape genetics. *Trends in ecology & evolution*, 28(10), 614-621. <https://doi.org/10.1016/j.tree.2013.05.012>
- Manel S, Poncet BN, Legendre P, Gugerli F & Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular ecology*, 19(17), 3824-3835. <https://doi.org/10.1111/j.1365-294X.2010.04716.x>
- Mehravi S, Ranjbar GA, Mirzaghaderi G, Severn-Ellis AA, Scheben A, Edwards D & Batley J (2021) De novo SNP discovery and Genotyping of iranian *Pimpinella* species Using ddRAD sequencing. *Agronomy*, 11(7), 1342. <https://doi.org/10.3390/agronomy11071342>
- Miguet P, Jackson HB, Jackson ND, Martin AE & Fahrig L (2016) What determines the spatial extent of landscape effects on species?. *Landscape ecology*, 31(6), 1177-1194. <https://doi.org/10.1007/s10980-015-0314-1>
- Mona S, Ray N, Arenas M & Excoffier L (2014) Genetic consequences of habitat fragmentation during a range expansion. *Heredity*, 112(3), 291-299. <https://doi.org/10.1038/hdy.2013.105>
- Narbona E, Ortiz PL & Arista M (2011) Linking self-incompatibility, dichogamy, and flowering synchrony in two Euphorbia species: alternative mechanisms for avoiding self-fertilization?. *PloS one*, 6(6), e20668. <https://doi.org/10.1371/journal.pone.0020668>
- Nei M (1987) *Molecular Evolutionary Genetics*.
- Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular ecology*, 13(5), 1143-1155. <https://doi.org/10.1111/j.1365-294X.2004.02141.x>
- Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289-290. <https://doi.org/10.1093/bioinformatics/btg412>
- Parreira B, Quéméré E, Vanpé C, Carvalho I, & Chikhi L (2020) Genetic consequences of social structure in the golden-crowned sifaka. *Heredity*, 125(5), 328-339. <https://doi.org/10.1038/s41437-020-0345-5>

- QGIS Development Team (2016) QGIS geographic information system. Open source geospatial foundation project.
- Rico Y (2019) Landscape genetics of Mexican biodiversity: A review. *Acta Universitaria* 29, e1894. doi. <http://doi.org/10.15174.au.2019.1894>
- Rochette NC & Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12), 2640-2659. <https://doi.org/10.1038/nprot.2017.123>
- Saura S (2020) The Habitat Amount Hypothesis implies negative effects of habitat fragmentation on species richness. *Journal of Biogeography*. <https://doi.org/10.1111/jbi.13958>
- Stojanova B, Šurinová M, Zeisek V, Münzbergová Z & Pánková H (2020) Low genetic differentiation despite high fragmentation in the endemic serpentinophyte *Minuartia smejkalii* (*M. verna* agg., *Caryophyllaceae*) revealed by RADSeq SNP markers. *Conservation Genetics*, 21(2), 187-198. <https://doi.org/10.1007/s10592-019-01239-4>
- Stoeckel S, Grange J, Fernández-Manjarres JF, Bilger I, Frascaria-Lacoste N & Mariette S (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species—*Prunus avium* L. *Molecular Ecology*, 15(8), 2109-2118. <https://doi.org/10.1111/j.1365-294X.2006.02926.x>
- Storz JF, Bhat HR & Kunz TH (2001) Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*. I. Inbreeding, outbreeding, and population subdivision. *Evolution*, 55(6), 1215-1223. <https://doi.org/10.1111/j.0014-3820.2001.tb00642.x>
- Subils R (1977) Las especies de *Euphorbia* de la República Argentina. *Kurtziana* 10:83–248.
- R Core Team (2020) R: the R project for statistical computing. 2019. Accessed Feb, 28.
- Templeton AR, Shaw K, Routman E & Davis SK (1990) The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden*, 13-27.s
- Toczydlowski RH & Waller DM (2019). Drift happens: Molecular genetic diversity and differentiation among populations of jewelweed (*Impatiens capensis* Meerb.) reflect fragmentation of floodplain forests. *Molecular ecology*, 28(10), 2459-2475. <https://doi.org/10.1111/mec.15072>
- Wickham H, Chang W & Wickham MH (2016). Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version, 2(1), 1-189. <https://github.com/hadley/ggplot2>
- Xu B, Sun G, Wang X, Lu J, Wang IJ & Wang Z (2017) Population genetic structure is shaped by historical, geographic, and environmental factors in the leguminous shrub *Caragana microphylla* on the Inner Mongolia Plateau of China. *BMC plant biology*, 17(1), 1-12. <https://doi.org/10.1186/s12870-017-1147-7>
- Yohe S & Thyagarajan B (2017) Review of clinical next-generation sequencing. *Archives of pathology & laboratory medicine*, 141(11), 1544-1557. <https://doi.org/10.5858/arpa.2016-0501-RA>
- Zak MR, Cabido M, Cáceres D & Díaz S (2008) What drives accelerated land cover change in central Argentina? Synergistic consequences of climatic, socioeconomic, and technological factors. *Environmental management*, 42(2), 181-189. <https://doi.org/10.1007/s00267-008-9101-y>
- Zhao W, Sun YQ, Pan J, Sullivan AR, Arnold ML, Mao JF & Wang XR (2020) Effects of landscapes and range expansion on population structure and local adaptation. *New Phytologist*. <https://doi.org/10.1111/nph.16619>

PROTOCOLO DE EXTRACCIÓN según Kumar et al. 2018 (*DNA extraction and molecular characterization of Acacia pseudobaena – An endemic species*)

ADAPTADO Y UTILIZADO PARA *Euphorbia acerensis*

El procesamiento de las muestras vegetales se realizó por tandas de hasta 15 muestras. El protocolo se desarrolló en el transcurso de 3 días:

[Día 1] Separación y pesaje de las muestras, la rotura mecánica y posterior incubación con búfer de extracción en baño húmedo a 65°C por una noche (entre 10 y 12 hs) [paso 1].

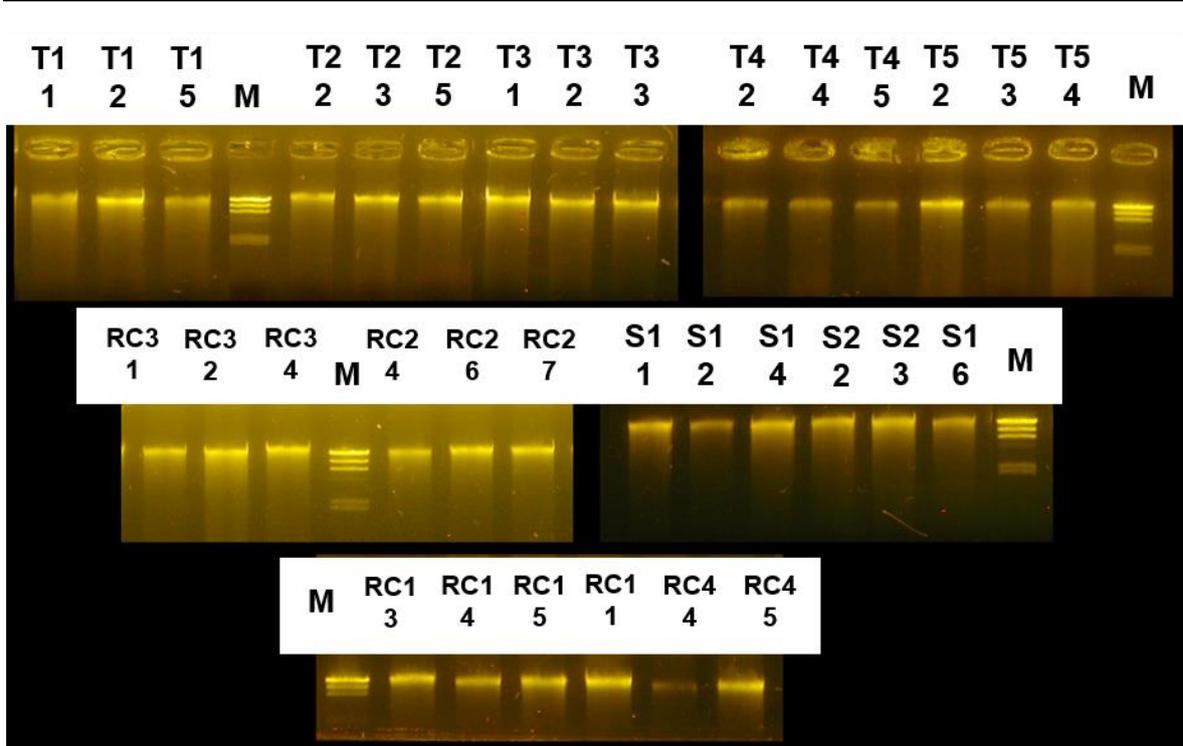
[Día 2] Tratamiento químico, centrifugación y precipitación final durante toda la noche [pasos 2 a 6]

[Día 3] Lavado del pellet y resuspensión de las muestras de ADN en buffer TE [pasos 7 a 12].

Composición de buffer de extracción: 3% CTAB, 100mM Tris-HCL (pH 8), 20mM EDTA (pH 8), 1,4M NaCl y 4% PVP.

1. Moler finamente hojas limpias y secas (25 mg) en un tubo eppendorf de 2 ml. Incubar a 65°C (baño húmedo) durante toda una noche (18 hs) en buffer de extracción: **1000 µl de buffer de extracción y 2 µl 1% β-mercaptoethanol**. [Si las muestras se encuentran en tubo de 1,5 ml, transferir muestras a tubos de 2 ml.]
2. Luego de la incubación, dejar enfriar las muestras a temperatura ambiente por 15 minutos, y agregar **1000 µl de chloroform-isoamyl alcohol (24:1)**.
3. Vortexear. Mezclar la solución por 20 minutos para formar una emulsión (shaker orbital a 200 rpm y a temperatura ambiente).
4. Centrifugar la emulsión a 13.000 rpm por 10 minutos a temperatura ambiente.
5. Extraer con cuidado el sobrenadante (alrededor de 700 µl con pipeta) y transferirlo a tubos de centrifugación estériles de 2 ml. Añadir **500 µl de NaCl 5M** y mezclar bien (vortex).
6. Añadir el mismo volumen (Alrededor de 700 µl) de **Isopropanol (-20 °C)** y llevar la mezcla a 4 °C durante toda la noche (12-18 hs), o a -20 por 2-3 horas.
Cercano al paso siguiente, pre-enfriar centrífuga a 4°C
7. Centrifugar a 10.000 rpm por 15 minutos a 4 °C para la formación del pellet.
8. Descartar el sobrenadante (a mano o con pipeta si el pellet no es visible) y limpiar el pellet con buffer de limpieza: **1000 µl de etanol 76% frío y 2µl de acetato de amonio 5 M**.
9. Añadir 2,5 µl de RNAsa A, mezclar todo por inversión y mantener a 37 °C por 1 hora.
Mientras se espera, ir haciendo el gel para la electroforesis si se requiere.
10. Centrifugar a 13.000 rpm por 5 minutos (10 minutos para que se pegue al fondo del tubo) a 4°C. Remover el sobrenadante.
11. Limpiar el pellet con **500µl de etanol 70%**, centrifugar a 4°C por 15 minutos a 10.000 rpm y remover el sobrenadante.
12. Secar muestras en termoblock por 10 minutos a 37°C. Resuspender el pellet en **50µl de buffer TE autoclavado (pH 8.0)**, a 37°C en termoblock. Guardar el DNA extraído a -20 °C.

Anexo II



Geles de agarosa mostrando la integridad del ADN genómico obtenido.
M = Marcador de Peso Molecular

Tabla de medición de concentración de ADN y pureza

La relación A260/280 es muy estable y se considera que un ADN de pureza óptima tiene un valor entre 1.8-2.0. Un ADN de pureza aceptable debe tener al menos una relación A260/280 > 1.6. Un valor A260/280 < 1.6 indica una posible contaminación por compuestos aromáticos como fenoles y proteínas. Un radio A260/280 > 2.1 podría deberse a la presencia de ARN en la muestra.

A 230nm absorben contaminantes como sales caotrópicas, fenoles o carbohidratos. En general, se considera que el ADN es puro cuando el ratio A260/230 se sitúa en torno 1.8-2.2. Una ratio menor de 1.8 se relaciona con presencia de contaminantes en la muestra. Cuanto menor sea esta ratio la presencia de contaminantes en la muestra será mayor. Un ratio < 1.5 estaría indicando una impureza relevante en la muestra que podría comprometer su funcionalidad.

Técnica de análisis	Análisis	Parámetros	Criterios de validez
Espectrometría	A260/280	Pureza	≥1.8-2.1 = Pureza óptima
			≥1.6-1.7 = Pureza aceptable
			<1.6 = ADN contaminado con compuestos aromáticos
			>2.1 = ADN contaminado con sales, fenol, hidratos de carbono...

Individuos	Nucleic Acid Conc(n g/μL)	OD260 /280	OD260 /230	OD260	OD280	OD230	OD320
T1 1	946,58	1,98	1,41	19,029	9,657	13,552	0,098
T1 2	1394,51	2,07	1,86	27,98	13,595	15,104	0,089
T1 5	1179,97	2,03	1,62	23,861	11,898	14,793	0,261
T2 2	920,69	2,01	1,53	18,437	9,18	12,082	0,023
T2 3	971,45	2	1,74	22,908	13,17	14,659	3,479
T2 5	1269,64	2,01	1,57	26,129	13,377	16,955	0,737
T3 1	1310,72	2,07	1,82	26249	12707	14409	34
T3 2	1455,25	1,94	1,24	33,497	19,429	27,939	4,393
T3 3	1329,38	1,93	1,36	28,588	15,779	21,479	2

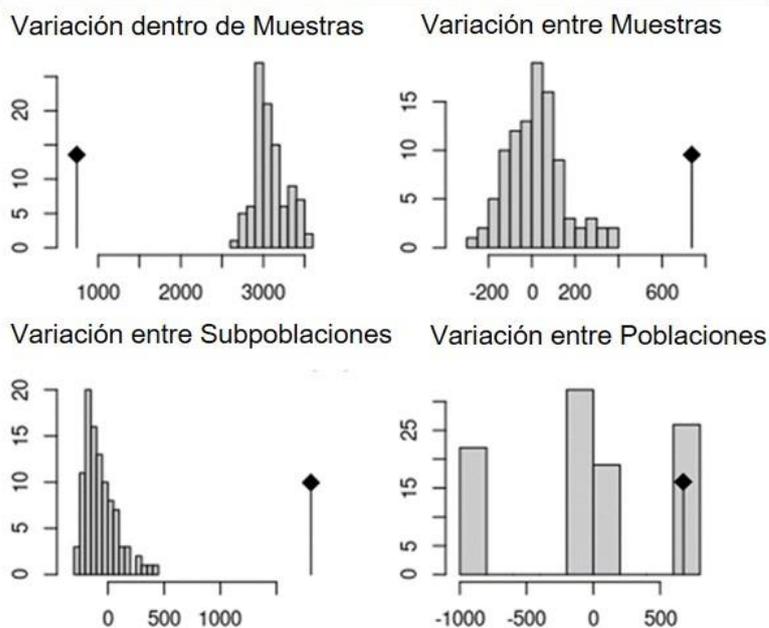
T4 2	1734,85	1,84	0,98	43,176	27,34	43,901	8,479
T4 4	759,12	1,98	1,33	15,644	8,133	11,858	0,462
T4 5	1476,14	1,73	0,77	41,654	29,243	50,72	12,131
T5 2	1103,34	2,06	1,71	22,354	11,013	13,183	0,287
T5 3	1616,92	1,97	1,39	35,548	19,667	26,402	3,21
T5 4	623,06	1,94	1,16	13,137	7,093	11,387	0,676
RC1 3	2364,32	2,12	2,02	47,438	22,5	23,523	0,152
RC1 4	1654,27	2,09	1,96	33,402	16,127	17,221	0,317
RC1 5	2713,08	2,13	2,03	54,494	25,729	26,932	0,232
RC2 4	1114,06	2,06	1,82	22,433	10,978	12,374	0,152
RC2 6	854,14	2,03	1,72	17,48	8,804	10,304	0,397
RC2 7	737,74	2,02	1,57	14,864	7,416	9,483	0,109
RC3 1	1344,08	2,04	1,6	27,507	13,798	17,42	0,625
RC3 2	2961,77	2,15	2,06	59,883	28,205	29,413	0,647
RC3 4	2185,9	2,1	1,95	44,073	21,191	22,768	0,355
RC4 1	2068,87	2,08	1,86	42,086	20,581	22,936	0,709
RC4 4	2208,81	2,08	1,78	45,129	22,194	25,828	0,953
RC4 5	2876,44	2,14	2,01	58,082	27,443	29,226	0,554
S1 1	1187,48	2,07	2,04	23,723	11,471	11,594	-0,027
S1 2	641,33	2,02	1,53	12,741	6,275	8,319	-0,086
S1 4	846,13	2,06	1,94	16,972	8,256	8,794	0,05
S2 2	1046,65	2	1,49	21,876	11,389	15,005	0,943
S2 3	1224,92	2,06	1,75	24,621	12,008	14,133	0,123
S2 6	882,05	2,08	2	17,756	8,612	8,948	0,115

Anexo IV

Salidas correspondientes a las AMOVAs realizadas para las distintas hipótesis de agrupamiento.

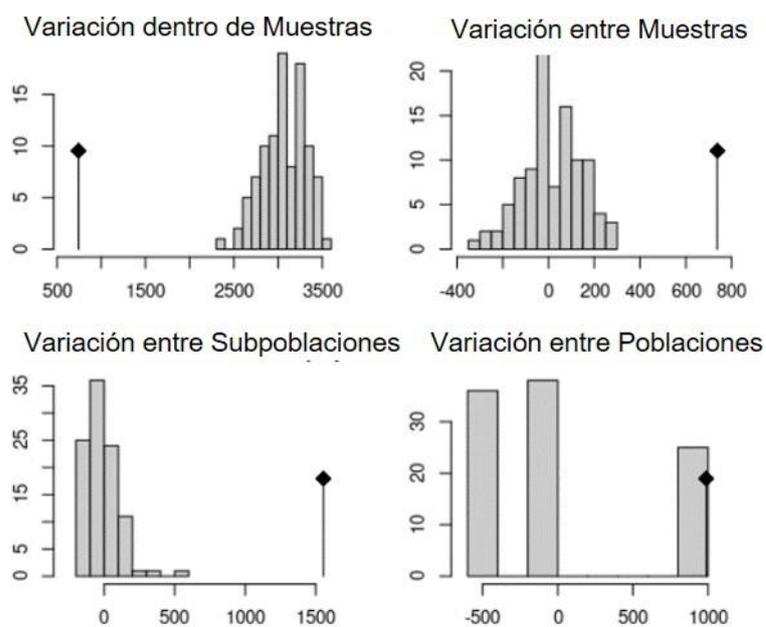
a) *Dos Poblaciones, (1 con 3 subpoblaciones)*

[RC] vs [(T)+(S1)+(S2/T4)]



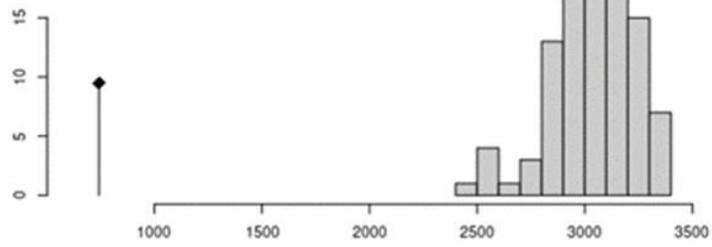
b) *Dos poblaciones, ambas con 2 subpoblaciones*

[(RC)+(S1)] vs [(T)+(S2/T4)]

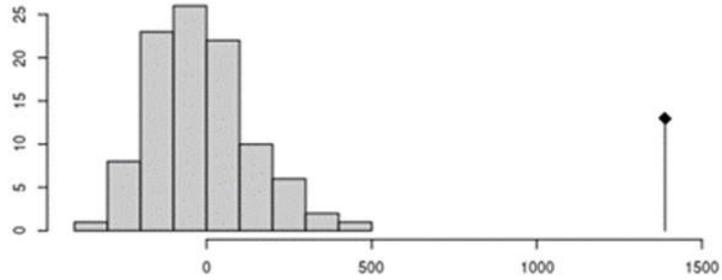


c) Dos poblaciones
[RC/S1] vs [T/S2]

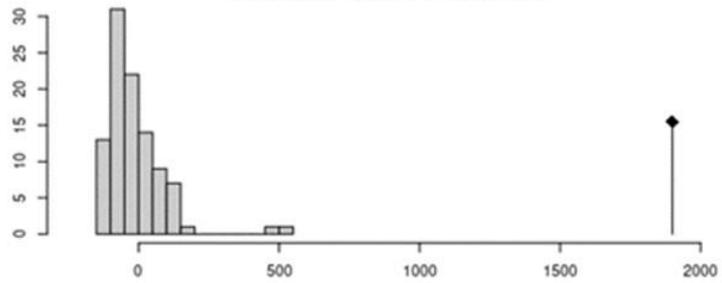
Variación dentro de Muestras



Variación entre Muestras

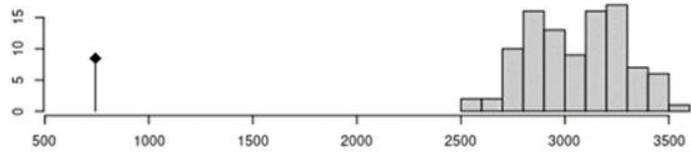


Variación entre Poblaciones

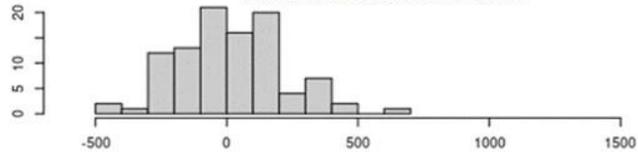


d) Dos poblaciones
[RC/S] vs [T]

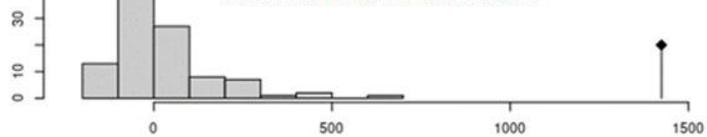
Variación dentro de Muestras



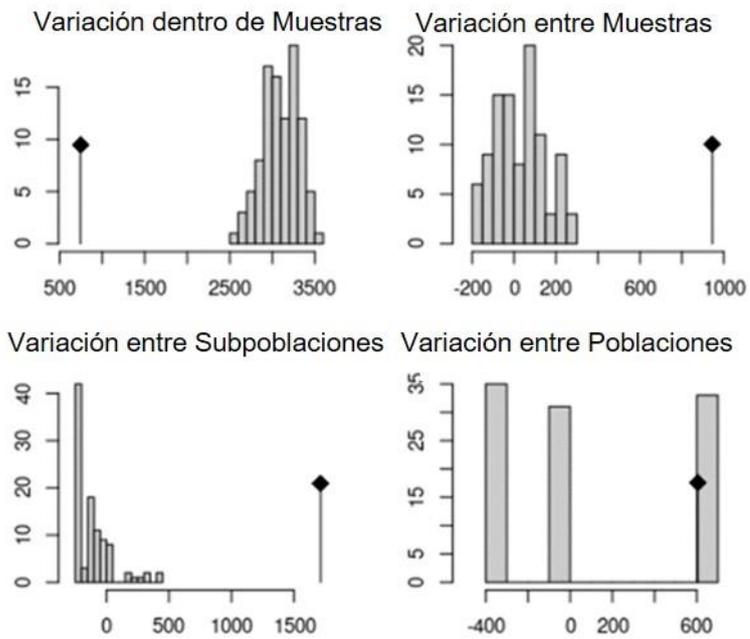
Variación entre Muestras



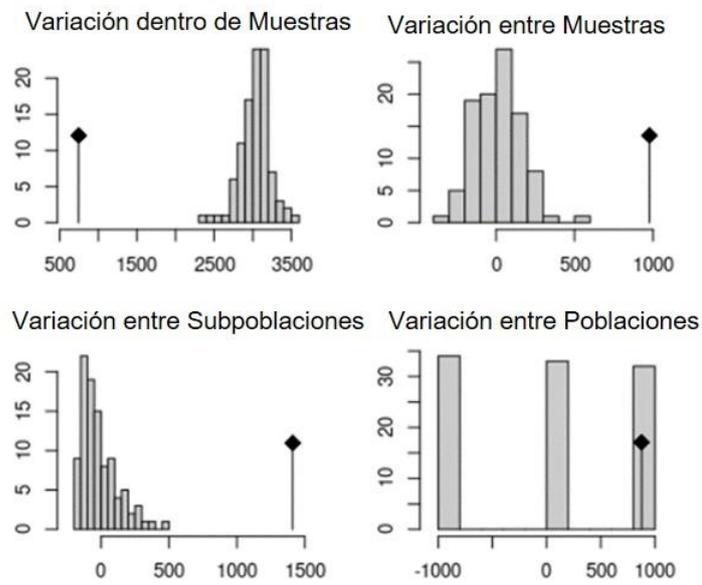
Variación entre Poblaciones



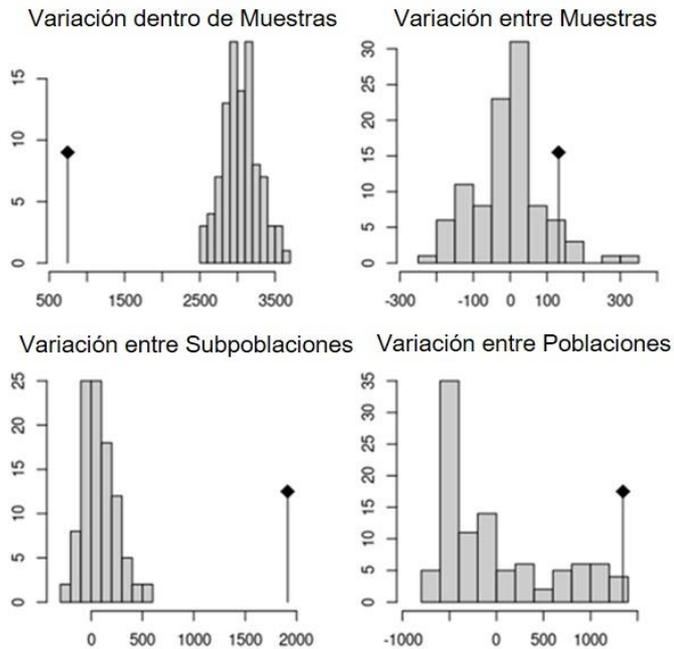
e) *Dos poblaciones, 1 con 2 subpoblaciones*
[RC/S1] vs [(T) +(S2/T4)]



f) *Dos poblaciones, 1 con 2 subpoblaciones*
[RC] vs [(T)+(T4/S)]



g) Dos poblaciones, 1 con 2 subpoblaciones
[RC] vs [(T)+(S)]



h) Tres poblaciones
[RC/S1] + [T] + [S2/T4]

