**UNIVERSIDAD DE CONCEPCIÓN**
**FACULTAD DE CIENCIAS FORESTALES**
INGENIERÍA EN BIOTECNOLOGÍA VEGETAL

# SENSIBILIZACIÓN DE LA DETECCIÓN Y GENOTIPIFICACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO ÚNICO (SNP) PARA *NOTHOFAGUS ALESSANDRII* Espinosa

Proyecto de Título (Habilitación Profesional) presentado (a) a la Facultad de Ciencias Forestales de la Universidad de Concepción para optar al título profesional de Ingeniero(a) en Biotecnología Vegetal

POR: Valentina Arancibia Acuña

Profesor Guía: Dr. Rodrigo Hasbún

Concepción, Chile 2023

PROPUESTA DE PARÁMETROS PARA EL LLAMADO DE SNP'S EN NOTHOFAGUS ALESSANDRI

Profesor Guía

**Rodrigo Hasbún Zaror**
Profesor Asociado
Ingeniero Forestal, Dr.

Profesor Guía

**Oscar Toro**
Profesor Asociado
Biólogo, Dr.

Calificación de la Habilitación Profesional (o Proyecto de Título):

Rodrigo Hasbún Zaror: 7,0

Oscar Toro: 7,0

## DEDICATORIA

Para todas las mujeres que día a día
Defienden a la naturaleza y a la vida

Por la biodiversidad y por el Ruil
Por la dignidad de los pueblos, no dejamos de resistir
No nos dejamos doblegar, volvemos a rebrotar
Incluso de las cenizas, avanzamos sin prisa
Yo cuando chica, apenas conocía lo que leía, pero allí
Mi corazón se llenó de sueños, y se dejó llevar por la brisa
En medio del vuelo, deje de mirar al cielo
Entonces miré al suelo y vi como lo destruían, vi el saqueo
Del dolor transformé en rabia, y la rabia me dio la razón
Descubrí entonces que los motivos sobraban
Y no podía sentir ni pensar sin razonar
No podía seguir sin ser consciente de los demás
Por eso les dedico esta tesis que, aunque apenas es un paso, me ha enseñado
Que la ciencia debe ser consciente, no es solitaria, es con la gente
No solo con la humanidad, es con todos los seres
Nunca hemos sido ajenos, nuestra vida es solo un pestañeo,
Antes y después, somos, y juntos siempre hemos estado,
Viviendo en la misma tierra, por eso espero que aprendamos
A respetarnos, compartamos nuestros saberes, y construyamos
Un nuevo mundo, para convivir con buenvivir.

## AGRADECIMIENTOS

Para a todos a quienes me he encontrado en mi camino y me han ayudado a construirme; desde mi primera infancia, a mi familia por apoyar mis sueños y creatividad. A mi Mamá, mi tía Angélica, Pablo, Tia Sandra, los quiero de todo corazón, su amor y su apoyo económico me cobijaron para mantenerme firme (y porfiada) en mis decisiones para seguir adelante, en el que finalmente encontré mi lucha, y mi camino hacia la felicidad.

A mis amigos y amigas incondicionales: Cinthya, Yessy, Karito, Naxita, Anita, Yaris, Daylin y Moreno; quienes han llegado a conocerme más que cualquiera, me han escuchado, aconsejado y acompañado a los momentos más importantes y especiales de mi vida, muchos de los cuales jamás creí experimentar y en la mayoría no tenía ni idea de cómo proceder. Ustedes son mi brújula y me enseñaron lo que realmente es el amor. Les amo profundamente, junto a ustedes planeo seguir construyéndonos, bajo los cimientos de la verdadera amistad, basada en confianza, apoyo y respeto mutuo. Les admiro, y gracias por motivarme cada día a ser mejor persona.

Mención honrosa a Cinthya Vega, compañera de departamento y de la vida, gracias por sacarme de la droga más fea, el amor heterosexual, que si me hubiera quedado en esa relación no terminaba la carrera. Gracias por enseñarme a convivir, lo que significa ser una persona considerada y a el respeto, con apoyo mutuo construimos un hogar sano, un nido de estabilidad emocional en el que puedo llegar todas las noches a sentirme tranquila. Gracias por traer a nuestras vidas a la melisita, que en paz descanse, que fue nuestra salvaguarda, llenando todos los rincones (no solo de sus pelos) de su amor, locuras, e hiperactividad. Gracias melisa por enseñarme otras formas de amor, a mirar más lento y ser la verdadera razón para motivarme a levantarme todos los días a trabajar. Con ustedes, conocí la felicidad de un hogar.

Agradezco a mis profesores, Rodrigo Hasbun, Narciso Aguilera y Oscar Toro, por sus enseñanzas, por su apoyo y guía en todo este proceso para terminar la carrera. Gracias por creer y confiar en mí, que sobró hasta para que yo también lo hiciera. Sin su empuje no hubiera conocido mis verdaderas capacidades para redactar bajo presión.

Agradezco a la misma vida, que a pesar de todo lo que me hizo llorar ayer, hoy podré llamarme una Ingeniera preparada y competente, capaz de afrontar cualquier adversidad, porque ya sé en quienes confiar, y soy capaz de dilucidar en el horizonte un rojo amanecer de esperanza hecho camino que vamos a seguir por un mañana mejor, un día a la vez.

# TABLA DE CONTENIDO

# ÍNDICE DE TABLAS

# ÍNDICE DE ILUSTRACIONES

## RESUMEN

*Nothofagus alessandri* es una especie arbórea del bosque maulino y está amenazada por la continua degradación de su hábitat. Se necesita una reevaluación de su estado de conservación, siendo la estimación de su diversidad genética actual un insumo muy importante. Los marcadores moleculares basados en polimorfimos de un solo nucleótido (SNP) son altamente informativos y en conjunto con la secuenciación de nueva generación (NGS), es muy simple y económico analizarlos. Sin embargo, debido a que esta es una especie poco estudiada y sin genoma de referencia, la detección y genotipificación de SNPs presenta como desafíos evitar SNP's putativos y la alta cantidad de datos perdidos. Se aplicó la tecnología Genotyping-by-Sequencing y baja cobertura de secuenciación, y para afrontar los desafíos anteriormente señalados se realizó un análisis de sensibilidad probando dos parámetros; la tolerancia de error (ETR) y la mínima cantidad de tags por individuo (mintagnum). Se busca probar los efectos de estos dos parámetros sobre la resolución de los resultados e informatividad del análisis genético poblacional. Para cuantificar estos efectos, se utilizaron dendrogramas basados en distancia genética, con sus respectivos índices cofeneticos totales, para representar el nivel de balance de los dendrogramas. Además, se generaron gráficos de Análisis de componentes principales para reflejar la resolución de las matrices, y utilizaron métodos de agrupamiento no supervisado para darle coherencia al número de grupos generado versus las poblaciones reales muestreadas. Se encontró que ambos parámetros influyen significativamente en la cantidad de SNPs detectados y genotipificados, sin embargo, no se observa un efecto claro en la estructura de los datos, ni tampoco en la capacidad resolutiva.

# ABSTRACT

*Nothofagus alessandri* is a tree species from the Malue's forest, which is considered under threat due to the degradation of its habitat. As a result, this species requires a reevaluation of its conservation status, being genetic diversity a pivotal component necessary for a correct assessment. Among molecular markers, Single Nucleotide Polymorfisms (SNP) highlight because their high rates of genetic variation, which have become more accessible thanks to the use of Next Generation Sequencing platforms (NGS). Nonetheless, given that *N. alessandri* lacks of proper referential genomic data, detection and genotyping represents a ubiquitous challenge to avoid putative SNPs and retrieving feasible data under high missing sequencing rates. Given the relevance of addressing its impending conservation threat, this study aimed to determine the effects of filtering settings choice on genotyping, SNP recovery, and resolution of genetic variability among *N. alessandri* populations. For this purpose, Genotyping by Sequencing (GBS) with low sequencing coverage was used to conduct a sensitivity analysis using different combinations of error tolerance (ETR) and minimum number of tags per individual (tagnummin) as filtering choices. The resulting data was used to assess the cophenetic index to evaluate the effects in clade balance and resolution in genetic distance dendrograms. Furthermore, Principal Component Analyses were employed to detect changes in matrices resolution, plus the use of unsupervised clustering methods to elicit the coherence between cluster retrieved and in-site collected populations. The results suggest that the selection of filtering parameters influences on the quantity of SNPs retrieved; yet, neither effect on clade resolution nor data structure were detected.

# I. INTRODUCCIÓN

La zona mediterránea de Chile entre la región de Valparaíso y el Bío-Bío, se destaca por su alta biodiversidad y ha sido clasificado como un punto importante para la conservación con relevancia global (Myers *et al.* 2000). Entre estas regiones también se concentra la mayor densidad poblacional del país, teniendo como consecuencia una mayor explotación de los recursos naturales que ha provocado la transformación y el deterioro del paisaje. Con un auge desde mediados de los 70', los bosques maulinos han sido sucesivamente fragmentados, intervenidos y deteriorados por distintas actividades antrópicas, principalmente por la agricultura, la silvicultura e incendios forestales (Hechenleitner y Gardner 2005, Torres-Diaz *et al.* 2007, Moya *et al.* 2018, Valencia *et al.* 2018).

En esta área la especie *Nothofagus alessandrii*, comúnmente conocido como Ruil, un árbol caducifolio circunscrito a una pequeña área en la cordillera de la costa que no supera los 100km de extensión latitudinal, ha sido de los más afectados por continuas talas y quemas a principio del siglo XX (Grau 1970, Donoso y Landaeta 1983). Sus bosques remanentes están en una situación frágil, muy fragmentados y rodeados en una matriz de plantaciones forestales alóctonas, mayoritariamente de *Pinus radiata*, una especie agresiva capaz de invadir y desplazar los bosques de *N. alessandrii* (San Martín y Ramírez 1984, Bustamante y Castor 1998, Santelices *et al.* 2012, González *et al.* 2022).

De las devastaciones más graves de los últimos años que ha sufrido el *N. alessandrii*, es destacable el incendio forestal de enero y febrero del año 2017, siendo uno de los más grande de la historia del centro-sur del país, cubriendo una superficie de aproximadamente 184.000 ha del bosque maulino, afectado a las poblaciones sobre un 85% moderada o severamente quemadas. (Valencia *et al.* 2018, Gajardo *et al.* 2022, González *et al.* 2022).

Desde el 2005 *N. alessandrii* se encuentra en *peligro crítico* (Hechenleitner y Gardner 2005), e internacionalmente la UICN lo clasifica en peligro de extinción (Barstow 2017). Las especies en peligro de extinción suelen tener pequeñas poblaciones en declive que significan

pérdidas constantes en la diversidad genética, por lo tanto, pérdidas de adaptabilidad de la población ante cualquier cambio del ambiente. Ante la grave situación de *N. alessandrii,* es urgente conocer su diversidad genética para orientar un plan de reforestación y restauración de sus bosques, con el fin aumentar el flujo génico. Según Torres-Diaz, *et al.* en 2007, *N. alessandrii*, mantiene una diversidad genética relativamente alta, dando a entender que esta, a pesar de su restringido rango geográfico, no ha sido (aún) severamente afectado por la deriva génica o la endogamia. Por otra parte, el estudio indica que la diversidad genética entre poblaciones es baja indicando que los fragmentos remanentes de la especie no están diferenciados. Sin embargo, otros estudios sugieren que existen al menos dos poblaciones genéticamente diferentes (Pineda Bravo 1998, Santelices Moya *et al.* 2009), tal contradicción hace necesario hacer mayores estudios para consensuar la información, ya que al no existir claridad puede significar un riesgo de contaminación genética introducir individuos de distintas localidades (Moya *et al.* 2018).

Además, el estudio se basó en marcadores moleculares proteicos (isoenzimas), marcadores con baja cobertura que no reflejan todos los cambios que el ADN puede tener y también se pueden ver afectados por factores ambientales en su expresión, por lo tanto, los resultados deben ser interpretados con cautela. Igualmente, ya han pasado más de quince años desde su investigación, y es necesario reevaluar la situación actual de *N. alessandrii* con el uso de técnicas más modernas de marcadores moleculares basados en ADN.

Un marcador molecular es un carácter o secuencia de ADN en un lugar identificable del genoma cuya herencia es rastreable. Estos se clasifican como co-dominantes o dominantes (esto se refiere a si son capaces o no de identificar si un carácter es homocigoto o heterocigoto), los que se distinguen entre ellos según su sensibilidad, cobertura y reproducibilidad (Sunnucks 2000, Alcántara 2007). A partir de las cualidades mencionados anteriormente, un marcador molecular ideal para evaluar la diversidad genética es aquel que sea capaz de cubrir todo el genoma y que contengan loci que no se encuentren bajo presiones de las fuerzas evolutivas (Höglund 2009)

Los marcadores SNPs (o polimorfismo de nucleótido único) son mutaciones en la secuencia de ADN de una sola base nitrogenada; deseables para estudios de genética de poblaciones ya que: están distribuidos en el genoma de muchas especies en zonas codificantes y no-codificantes, su evolución se asemeja a modelos de mutación simples, acumula menos errores de replicación por la ADN polimerasa en comparación a otros marcadores, entre otros (Höglund 2009). Para trabajar con marcadores SNPs se necesitan dos etapas, el "descubrimiento" de los loci, y luego genotipificarlos (Morin *et al.* 2004). Ambos pasos ocurren de forma paralela utilizando la técnica de GBS (Genotyping-by-Sequencing) (Elshire *et al.* 2011). Debido a que GBS está basado en enzimas de restricción y la secuenciación de nueva generación (NGS), el descubrimiento y genotipado de SNPs se ha vuelto un sistema sencillo incluso para especies sin un genoma de referencia, los cuales suelen ser problemáticos ya que se vuelven necesarios mayores cuidados para los criterios a considerar en el llamado de SNPs y su filtrado para la eliminación de loci putativos (Ekblom y Galindo 2011, Narum *et al.* 2013, Torkamaneh *et al.* 2016).

La secuenciación masiva puede producir decenas, cientos o miles de gigabases (Gb) en datos de secuencia, por lo cual se utilizan pipelines para filtrar, ordenar y alinear dichas secuencias. Cada pipeline está diseñado para contextos distintos como la ploidía de la especie problema, el material utilizado o el ensamblaje. El nivel de filtrado de datos depende de los objetivos del estudio ya que pueden ser sensibles a los datos perdidos y al tamaño de muestra de los individuos. Variadas publicaciones recientes muestran como los parámetros para identificar loci *de novo* pueden afectar considerablemente los resultados (Andrews *et al.* 2016). Por ejemplo, en la búsqueda de genes de resistencia de *Medicago sativa L*. contra *Verticillium Wilt*, utilizando diferentes pipelines encontraron diferencias en la cantidad de marcadores descubiertos, densidad, heterogocidad y los loci dentro de los distintos cromosomas (Yu *et al.* 2017).

UNEAK (Universal Network Enabled Analysis Kit) (Lu *et al.* 2013), es una pipeline hecha para especies sin genoma de referencia y genotipificación de baja cobertura. Esta pipeline se encarga de generar un mapa de haplotipos (HapMap) a partir de los datos en FASTQ

entregados por la NGS. Los archivos son procesados para obtener una lista de *tags*, (alelos) de 64 pb, los cuales se alinean y se le entregan datos de posición en el genoma (de existir un genoma de referencia; no siendo un paso obligado), obteniendo la matriz en formato HapMap. En estos procesos, los parámetros fundamentales son la enzima de restricción utilizada; *ETR*, la tolerancia de error, que se refiere a los errores de secuenciación que podría resultar del NGS, típicamente predeterminado en 0.03 por la tasa de error de illumina; y el *minTagnum*, que es la mínima cantidad de veces que un tag debe estar presente entre los individuos, para evitar posibles tags resultantes de un error de secuenciación y suele estar predeterminado en 5 o 10 (Lu *et al.* 2013, Glaubitz *et al.* 2014).

Algunos cambios en las variables mencionadas, como un alto ETR causan un mayor llamado de SNPs, pero también mayor cantidad de SNPs putativos y de datos perdidos. Caso contrario, un ETR más cercano a 0 disminuye el llamado de SNPs y por lo tanto la cobertura de SNPs en el genoma, sin embargo, esto generará una estimación poco realista de la real diversidad genómica. En el caso de minTagnum, valores bajos aumentan la cobertura de SNPs, tanto reales como putativos, en cambio altos valores causará el caso contrario. (Lu *et al.* 2013)

Se ha reportado que filtros tolerantes para aumentar la cantidad de detección de SNPs a pesar de tener una gran cantidad de datos perdidos, aumenta también la resolución para arboles filogenéticos tanto en datos simulados (Huang y Knowles 2016), como también para datos reales, como los peces cíclidos del lago victoria (Wagner *et al.* 2013), y para las aves terrestres galliformes (Hosner *et al.* 2016). Por esto, es importante probar diversos filtros considerando la cantidad de individuos del análisis, cobertura de SNPs esperada, y los objetivos de la investigación. (Lu *et al.* 2012)

Pregunta de Investigación

¿Cómo influyen los parámetros de detección y genotipificación de SNPs usando GBS en la resolución de los resultados e informatividad del análisis genético poblacional en *Nothofagus alessandrii*?

Hipótesis:

El aumento de la tolerancia en los parámetros de detección y genotipificación de SNPs usando GBS genera un aumento en la resolución de ordenamiento de las poblaciones de *Nothofagus alessandrii.*

Objetivo general

Sensibilizar los parámetros de detección y genotipificación de SNPs usando GBS en un análisis genético poblacional en *Nothofagus alessandrii.*

Objetivos específicos:

Comparar parámetros estrictos y tolerantes para la detección y genotipificación de SNPs de las poblaciones de *N. alessandrii*.

Construir, analizar y comparar dendrogramas y PCA de las poblaciones de *N. alessandrii*.

# II.METODOLOGÍA

## 2.1. Material vegetal

Se muestrearon 30 individuos de *Nothofagus alessandrii* de 10 localidades en la región del Maule: “Agua Buena” (AB), “El Desprecio” (ED), “Huelon” (HUE), “La Bodega” (LB), “Lo Ramirez” (LR), “Macal” (MAC), “Porvenir” (POR), “Quivolgo” (QUI), “RNAC El Corte” (REC) y “RNAC El Fin” (REF). Las zonas en donde fueron recolectados se pueden observar en la Figura 1. El material vegetal se obtuvo de los brotes después del incendio del 2017, de los cuales se extrajo el ADN en la Laboratorio de Epigenética Vegetal de la Universidad de Concepción y se mandaron a procesar por Genotyping-by-Sequencing (GBS) en Wisconsine University (EEUU).


Figura 1: Mapa de las zonas de muestreo de *Nothofagus alessandrii* en la Región del Maule

## 2.2. Procesamiento de datos brutos

Con Illimina, se utilizó la enzima *Ape*KI y obtuvo 42.654.563 lecturas en archivos FASTQ. A partir de estos archivos, se realizó el análisis con la pipeline UNEAK (Lu *et al.* 2012), con los parámetros recomendados a excepción de la tasa de tolerancia de error (ETR) y el tagnumber. Se sensibilizaron los parámetros de ETR probando 0,01, 0,03 y 0,05, y tagnum usando los valores 2, 3, 4 y 5, obteniéndose así 12 archivos HapMap.hmp.

El funcionamiento y los cambios en la pipeline UNEAK se describen como:

### 2.2.1. Tags y filtrado de datos de secuencias brutos.

Primero se ingresa a la pipeline un "Key file" con los códigos de barra (barcodes) de cada individuo, UNEAK lee todas los archivos FASTQ y busca coincidencias con uno de los barcodes esperados seguidos de los remanentes de los sitios de corte (CAGC o CTGC para la enzima *Ape*KI) y los trunca en 64 pares de base (pb), dejando el sitio de corte remanente y eliminando los barcodes. Las lecturas que contengan datos perdidos (N) dentro de las primeras 64 pb después de los barcodes son eliminados a modo de filtro inicial. Las lecturas que contengan un sitio de corte incompleto (por una incompleta digestión de la enzima o formación de quimeras) o que contentan el comienzo de un adaptador común (por fragmentos de restricción menores a 64pb) son rellenados con hasta 64pb con polyA y se truncan apropiadamente. La longitud real de las lecturas truncadas es registrada en un archivo TagCount de salida, que además contienen un nuevo código de barras con el ID de cada muestra a la que corresponda. Estos archivos además contienen el número de veces que fue un tag observado y son ordenados por su secuencia.

Posteriormente, se vuelve a remover aquellos tags únicos por posibles errores de secuenciación, de las secuencias alineadas se genera una red (network) en el que se descartan networks muy "complicados", este filtro es regulado por la tasa de tolerancia de error (ETR por sus siglas en inglés). Luego se define la mínima cantidad de veces que un tag debe estar presente en el archivo TagCount, (minTagNumber, o minTG). Finalmente, los tags recíprocamente alineados que se diferencian en una sola base son definidas como un SNP.

#### 2.2.2. Análisis de parámetros

##### 2.2.2.1. MinTagnumber

Se refiere a que un tag debe estar presente al menos a una cierta cantidad de individuos. Se probaron los MinTagnum (TG) 2, 3, 4 y 5. La configuración 5 o 10 es típica, pero estos fueron escogidos para probar matrices con la mayor cantidad de SNPs posibles y por la cantidad de individuos disponibles. Tags que no cumplían estos requisitos fueron eliminados por el pipeline.

##### 2.2.2.2. Tasa de tolerancia de error

La tasa de tolerancia de error (ETR) se refiere a los errores de secuenciación de Illumina. ETR más altos genera más SNPs, pero también más SNPs putativos. Cuando ETR es igual a 0 significa que solo se mantendrán aquellos tags que son puramente recriprocos y que no hubo ningún error de secuenciación, siendo este es el valor más estricto. El valor predeterminado es de 0,03, ya que se sabe que este es el error promedio de Illumina. Este parámetro depende de la cobertura de la enzima, y debido a que no tenemos un genoma de referencia, decidimos evaluar un ETR inferior (0,01) y uno mayor (0,05), que son los valores mínimos y máximos recomendados.

Los archivos FASTQ se ingresaron al pipeline con todos los parámetros mencionados generando un total de 12 archivos HapMap.hmp, los que fueron ingresados y procesados con TASSEL V5.0 (Glaubitz *et al.* 2014)

### 2.3. Análisis exploratorio de datos

#### 2.3.1. TASSEL V5.0

Con los archivos HapMap obtenidos se transformaron a archivos .VCF (Danecek *et al.* 2011), (por sus siglas en inglés de Variant Call Format, o llamado de variantes, es un formato de texto que almacena las variantes de varias secuencias de genes) y se obtuvieron tablas resumen de las matrices, cantidad de SNPs, datos perdidos, entre otros.

### 2.3.2. R4.2.1

Con la interfaz de RStudio 2022.07.02 (Team 2016), se ingresaron cada una de las matrices en formato .vcf, utilizando el paquete vcfR (Knaus y Grünwald 2017), junto al paquete adegenet (Jombart *et al.* 2018), y se le agregó información de las localidades donde fueron recolectados cada individuo.

#### 2.3.2.1. Dendrogramas

Utilizando el paquete ape (Paradis *et al.* 2019), para cada archivo se calculó una matriz de distancia utilizando la función bit.wise que utiliza la distancia de Hamming. Se generaron dendogramas con el algoritmo de Neighbor Joining, del cual se calcularon 999 pseudoreplicas con Bootstrap para medir el soporte de los grupos obtenidos. También se utilizó el paquete RColorBrewer (Neuwirth y Brewer 2014), para agregar colores según la localidad y facilitar la visualización de los dendrogramas.

Este tipo de agrupamiento no supervisado ignora de que localidad viene cada individuo y no requiere que se especifique de antemano la cantidad de grupos que presentan los datos, ya que los ordena jerárquicamente en base a su distancia genética. Este tipo de agrupamiento, gráficamente son arboles enraizados con hojas etiquetadas, en la que las hojas representan los individuos, los nodos miden la distancia que hay entre las hojas, y ayudan a ver su orden jerárquicamente.

Para medir y comparar la resolución de la información de los dendrogramas se utilizó el paquete TotalCopheneticIndex (Mir y Rosselló 2013), este utilizando el índice de cofenético total estima el balance de un árbol, calculando la suma de todos los pares de hoja diferentes, sobre la profundidad de su ancestro común más bajo. Si bien este índice está hecho para comparar árboles filogenéticos, es utilizado para medir resolución ya que está basado no solo en la topología de los árboles, sino que también considera el largo de las ramas, entre otras cosas.

Los valores más bajos de este índice indican un mayor balance, por lo tanto, una mejor resolución de los datos y son para los dendrogramas más resueltos que todos sus nodos internos se bifurcan en dos nodos hijos hasta llegar a las hojas. Los valores más altos son para los dendrogramas tipo “oruga” (que son árboles en el que sus nodos internos bifurcan

en hijos hojas o nodos terminales) o tipo “estrella” (árboles en que enraiza a variadas nodos hojas).

#### 2.3.2.2. Análisis de componentes principales

Otros análisis de agrupamiento no supervisado fue en Análisis de Componentes principales, utilizando el paquete Ade4 (Chessel *et al.* 2004), para visualizar la estructura de los datos y sus distancias genéticas. También los individuos fueron coloreados según su localidad.

#### 2.3.2.3. Análisis discriminante basado Componentes Principales

Para responder la pregunta de cuantos grupos hay en los datos, o a que grupo pertenece cada individuo se utilizaron los algoritmos de agrupamiento K-medias y un Análisis discriminante de componentes principales (DAPC, por sus siglas en inglés) utilizando el paquete adegenet 2.0 (Jombart y Collins 2015). Para orientar el análisis, primero debe aproximarse un agrupamiento *apriori* de los grupos presentes en el set de datos, lo que se calculó usando un criterio de K-medias. Posteriormente, los grupos definidos fueron testeados usando un análisis discriminante usando los de componentes principales más informativos, los que fueron complementados usando un gráfico de barras con la probabilidad posterior de pertenencia para cada genoma. Dado que el número de grupos (k) puede estar influenciado por el nivel de grupos seleccionados apriori, se probó este test con diversos números de agrupamiento. Con estos tres análisis, se hizo con el paquete ggpubr (Kassambara y Kassambara 2020) un gráfico multipanel (número de agrupamientos, DAPC de grupos obtenidos y grafico de barras con probabilidad posterior de pertenencia) que justifican la cantidad de poblaciones por cada matriz.

#### 2.3.2.4. Factorización de matriz no negativa dispersa

Finalmente, para complementar los resultados del análisis anterior, se realizó un nuevo análisis de agrupamiento basado en la factorización de una matriz no negativa dispersa (sNMF por sus siglas en inglés) utilizando el paquete LEA (Frichot *et al.* 2014). El programa calcula el coeficiente de ascendencia, que es la proporción de los genomas individuales provenientes de K poblaciones ancestrales (pool gene). Es similar a los programas ADMIXTURE y STRUCTURE, pero este algoritmo fue seleccionado por su rapidez,

eficiencia y porque es más acertado en presencia de endogamia (Frichot *et al.* 2014). Para escoger la cantidad de conglomerados, este algoritmo utiliza el criterio de entropía cruzada (cross-entropy criterion) en el que valores más bajos usualmente significan mejores ejecuciones del programa.

Para comprobar la sensibilidad de los resultados, se hizo un dendrograma a partir de las distancias de las proporciones del resultado anterior para cada localidad con el paquete stats (Team 2016).

Así, se definió como mejor matriz aquella que tenga un menor índice cofenetico, menor perdida de datos, y que ordene y distribuya mejor los datos, según poblaciones y distribución geográfica real.

### 2.3.3. QGIS

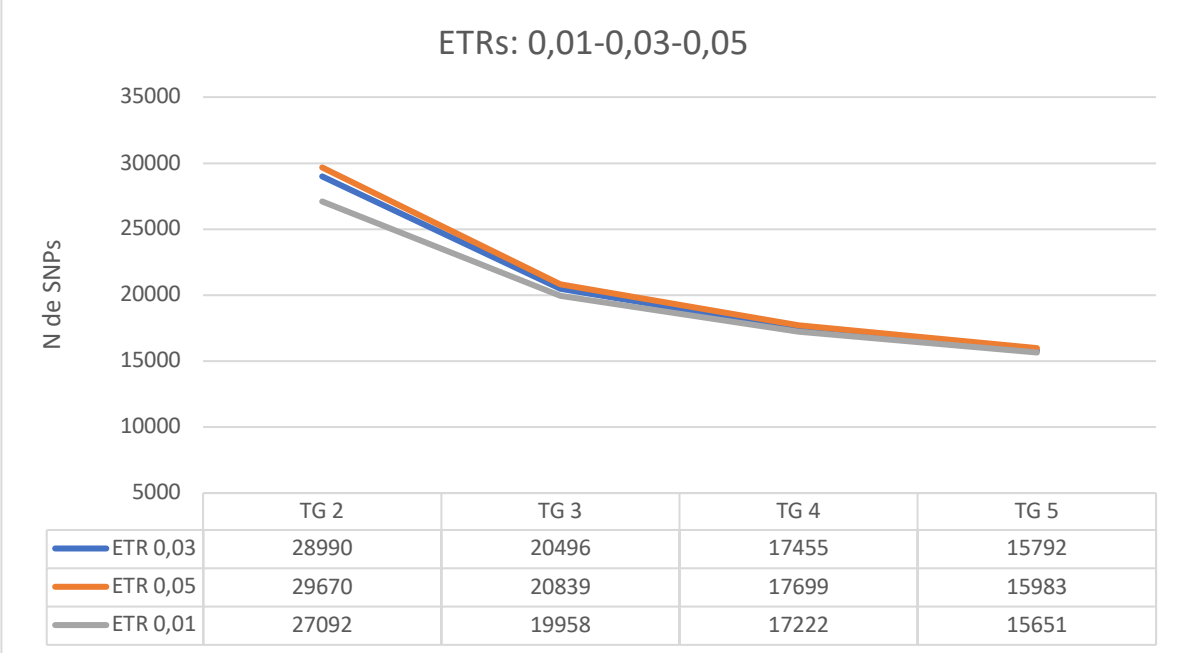
Finalmente, utilizando QGIS 3.28.2 (QGIS 2015), se introdujeron los valores de los coeficientes de ancestros obtenidos del análisis anterior y se graficaron en el mapa de las localidades muestreadas con gráficos de torta.

# III.RESULTADOS Y DISCUSIÓN

## 3.1. Análisis de Matrices

De las 12 matrices generadas con el Software TASSEL, se hizo una tabla resumen de la cantidad de SNPs que albergó cada matriz. Tal como se observa en la Tabla 1, la caída de SNPs se mantiene entre los distintos ETR, indicando que la disminución de los marcadores depende del TG seleccionado.

Tabla 1: Número de SNPs según Tagnumbers con distintos ETR.




| | TG 2 | TG 3 | TG 4 | TG 5 |
|---|---|---|---|---|
| ETR 0,03 | 28990 | 20496 | 17455 | 15792 |
| ETR 0,05 | 29670 | 20839 | 17699 | 15983 |
| ETR 0,01 | 27092 | 19958 | 17222 | 15651 |

De igual forma, el porcentaje de datos perdidos varía solo según TG, como se observa en la Figura 2, y no tiene grandes diferencias entre los ETR.

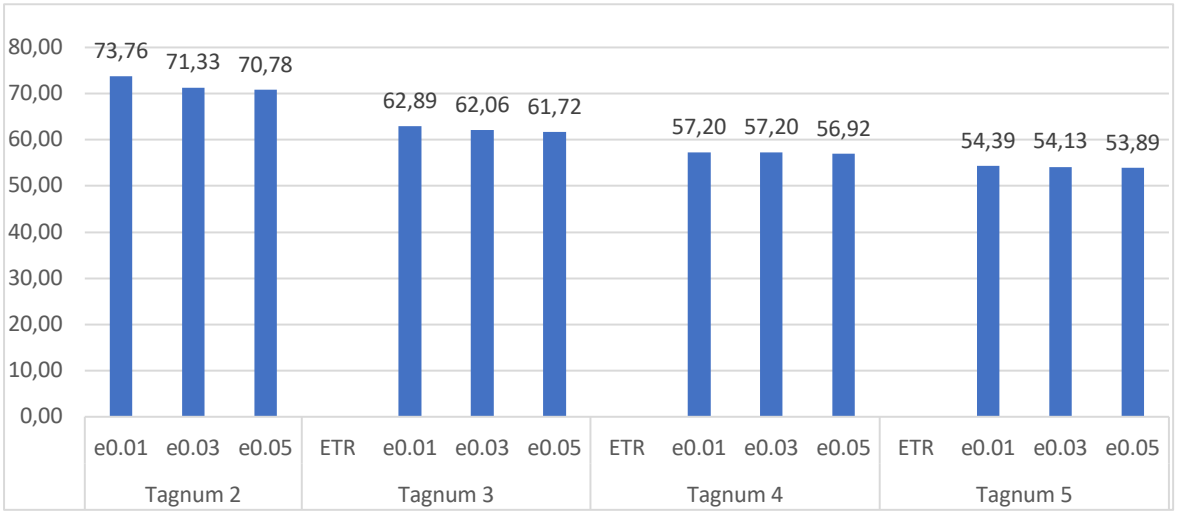



Figura 2: Porcentaje de datos perdidos según Tagnum para ETR 0.01, 0.03 y 0.05

## 3.2. Agrupamiento no supervisado de datos, Jerárquico

### 3.2.1. Dendrogramas

En la Figura 3 comparamos los índices cofenéticos totales con la cantidad de SNPs descubiertos para cada matriz. Las matrices con valores más bajos del índice cofenético total son ETR0.03TG4, ETR0.03TG5, ETR0,05TG2 y ETR0,05TG3, matrices que contienen un variado número de SNPs, desde 17455 a 29670, sugiriendo que el índice cofenético no es afectado por la cantidad de SNPs, o que estas matrices contienen una menor cantidad de loci putativos.

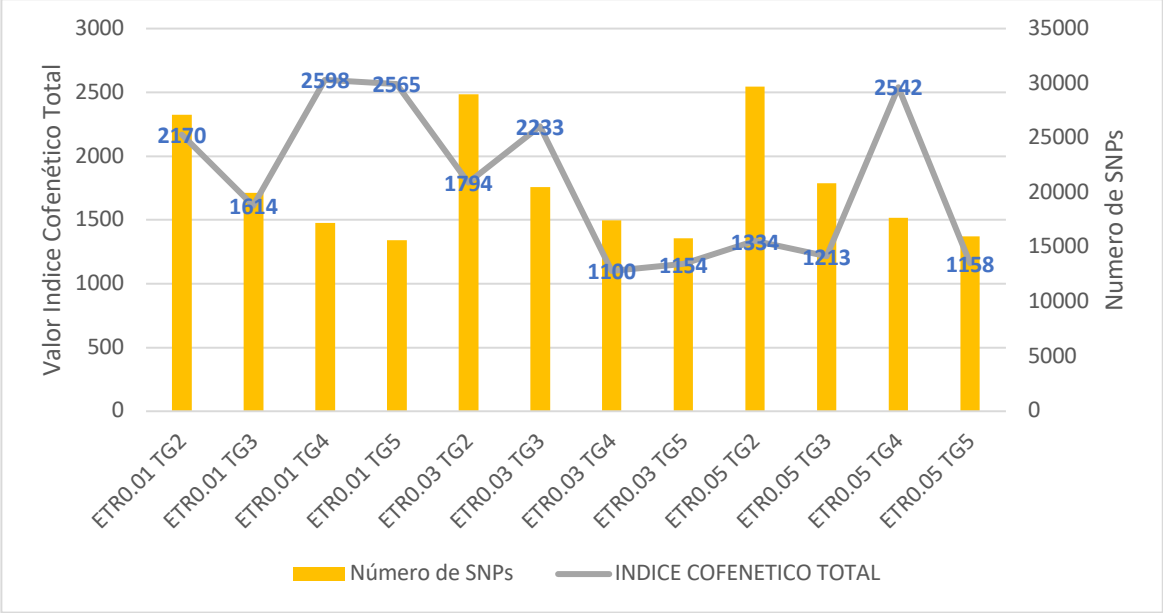



Figura 3: Valores del índice cofenético total para cada uno de los dendrogramas generados, comparándose con la cantidad de SNPs analizados.

En la Figura 4 se compara el número de SNPs con el índice cofenético total de todos los filtros aplicados en un diagrama de dispersión. En él se puede confirmar que no existe una relación entre estas variables, y que el filtro de ETR 0.05 presenta un mayor éxito en resolución con valores más bajos de índice cofenético total, junto a los filtros TG 3 y 5.

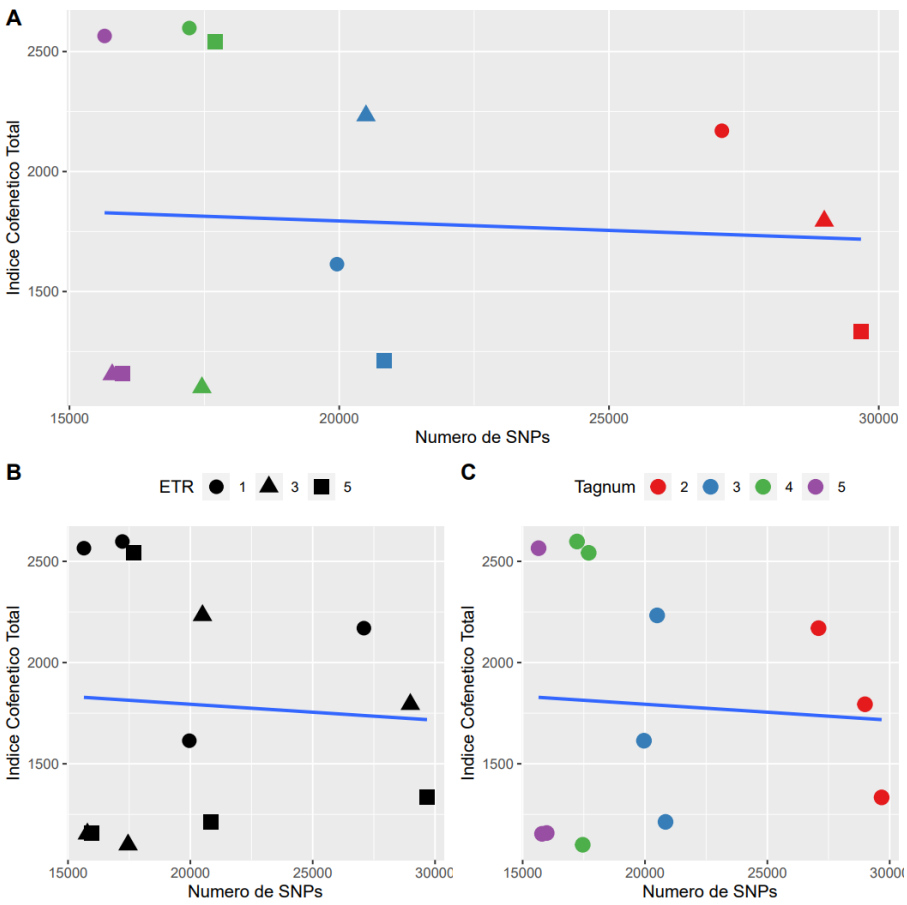



Figura 4:Diagrama de dispersión de los filtros ETR y TG comparando sus índice cofenético total y numero de SNPs llamados.

En la Figura 5 podemos observar los dendrogramas con los índices cofenéticos totales más altos, obtenidos de las matrices ETR0,05 TG4 Y ETR0,01 TG4, en ellos sí se puede diferenciar entre los grupos de norte y sur, sin embargo, quedan algunos individuos como grupos externos y se observan “orugas”, indicando un bajo balance del árbol.

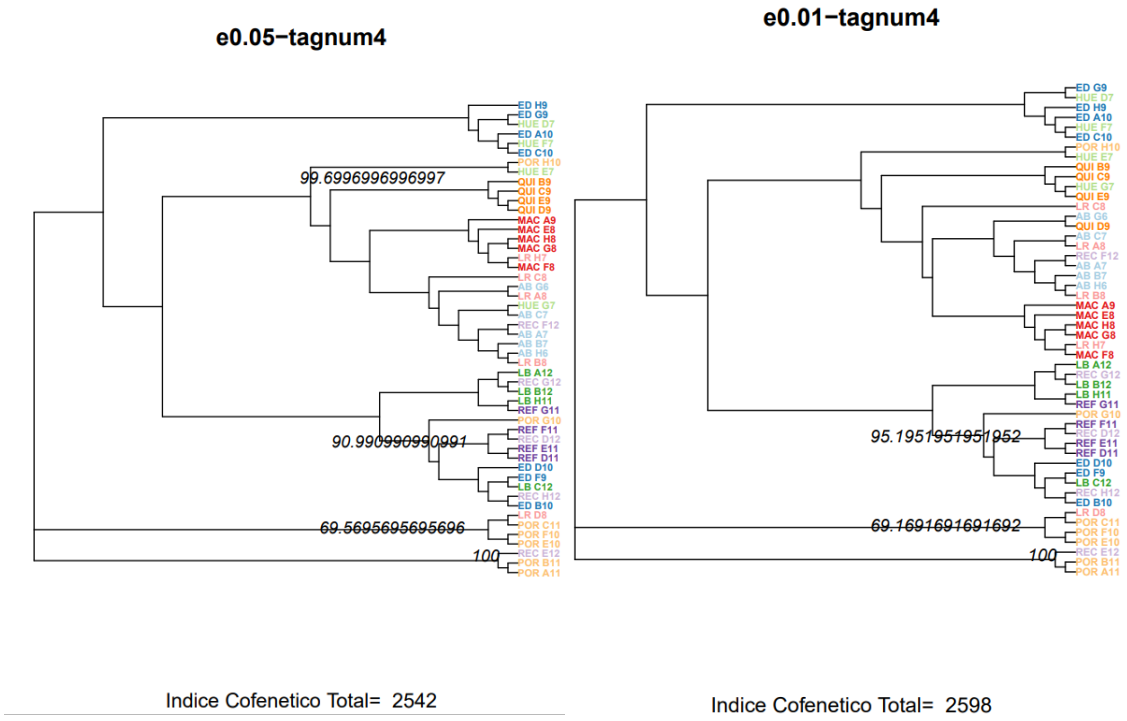



Figura 5: Dendograma para ETR0.05 TG 4 y ETR 0.01 TG 4 de izquierda a derecha

En la Figura 6 podemos observar dendrogramas generados a partir de las matrices E0.03TG4 y E0.03TG5. Estos gráficos fueron los que presentaron índices cofenéticos totales más bajos, y se logra diferenciar a los grupos de norte a sur, los que fueron divididos en tres grupos. En el dendrograma generado por E0.03T4 se tiene por un lado a “ED”, “HUE”, “REC”, “REF” de la zona sur en un grupo, a los individuos de “POR” junto a un individuo de “REC” en un segundo grupo, y a los individuos muestreados en el norte en un tercer grupo. En el tercer grupo también podemos observar un subgrupo externo que aglomera a individuos muestreados de la zona sur. En el dendrograma generado por ETR0.03TG5 también se formaron tres grupos, en el primer grupo de encuentra a “ED”, “HUE, “POR” y un individuo de “REC” y uno de “LR”. En el segundo grupo se ordenan juntos a “ED”, “REC”, “REF” y a un individuo de “LB”, y en el tercer grupo a individuos de la zona norte, como “QUI”, “HUE”, “AB” “MAC”, “LR”, y un subgrupo externo de individuos de la zona sur.

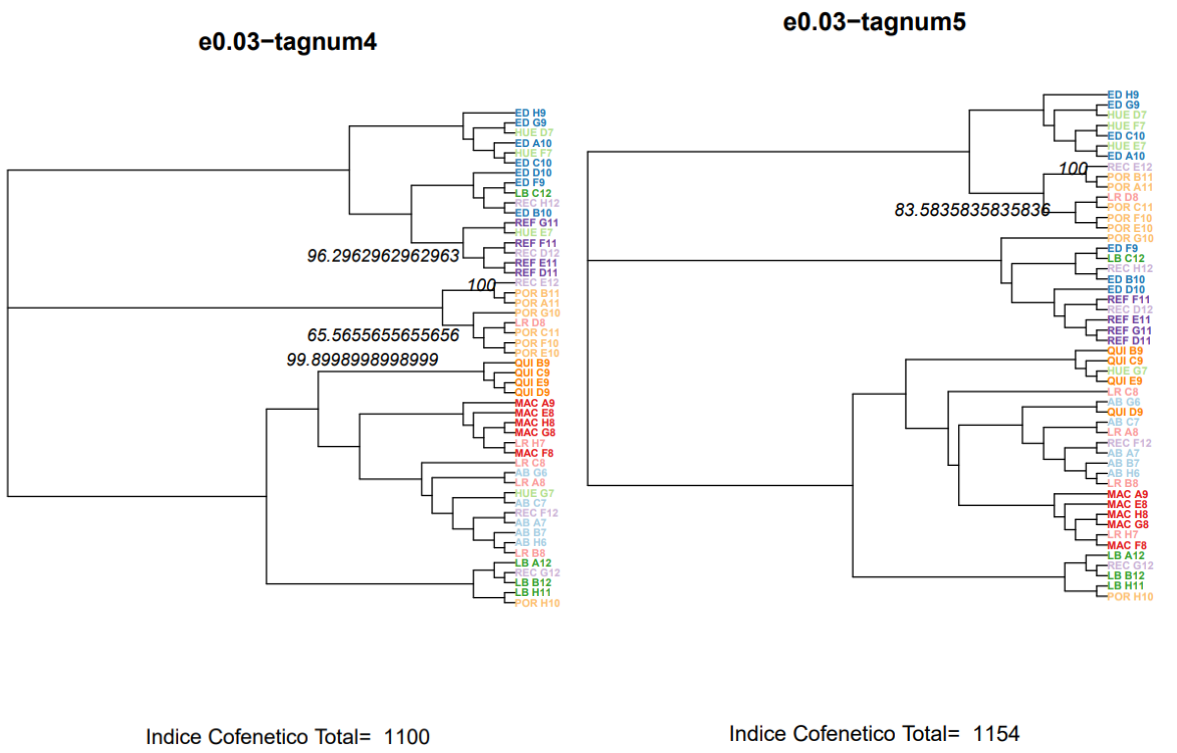



Figura 6: Dendograma para ETR0.03 TG 4 y ETR 0.03 TG 5 de izquierda a derecha

Estos subgrupos externos de individuos muestreados en la zona sur pero que son agrupados junto a individuos de la zona norte, podría ser algo completamente normal, si se considera el flujo génico en una especie con distribución restringida como *Nothofagus alessandrii* o bien por una traslocación por acción humana de alguna reforestación pasada sin consideraciones genéticas por la poca información que se tiene de la especie (San Martín y Ramírez 1984). Otros dendrogramas pueden ser revisados en el apéndice B

### 3.3. Análisis de componentes principales

Para evaluar la estructura poblacional, se realizaron PCA a cada una de las matrices (que pueden ser revisados en el apéndice A), en las que no se observaron grandes diferencias entre los ETR, pero si se puede observar una diferencia en la estructuración de los datos entre los TG. Con estos resultados se reafirma que TG es un parámetro crucial para la estructura y resolución de los datos.

Todos los PCA realizados logran diferenciar entre los componentes a las localidades de la zona norte y de la zona sur, como se puede observar en la Figura 7, que en el lado izquierdo del componente 1, se encuentran los individuos de las localidades sur “REF”, “LB”, “POR”, “ED”. Los individuos de la localidad “REC” se encuentran más cercanos al centro de origen, pero se siguen agrupando cercanos a las localidades del sur. También se encuentra a “QUI” a la izquierda del componente 1, pero está diferenciado de los individuos de la localidad sur por el componente 2. A la derecha del componente 1 se encuentra “AB” muy alejado de los otros individuos, a “MAC” y a “LR” más cerca del origen. En particular los individuos de la localidad “AB” están muy apartados de las demás localidades si se los compara con el componente 1, pero si se les compara con el componente 2, está más bien diferenciado con los individuos de “QUI” y “MAC”. Haciendo que tenga más sentido los dendrogramas en los que “AB” se encontraban agrupados junto a individuos de “HUE”, “LR” Y “REC” que en los PCA, son las localidades más cercanas al origen.

También se pueden destacar en particular a 3 localidades. Quivolgo, representado con un naranjo saturado, con parámetros menos estrictos se encuentra alejado de las demás muestras, pero según se van restringiendo los parámetros, estos individuos se acercan al origen. Con Macal, representado con rojo saturado, al contrario, con TG 2 (menos estricto) se encuentra más cerca del origen, y cuando se restringe los parámetros con TG 5, estos individuos se alejan de los demás. Y con AguasBuenas, con mayor restricción de parámetros los individuos se acercaron más entre ellos y se mantuvieron a la misma distancia de los demás. Revisar apéndice A para mayor información.

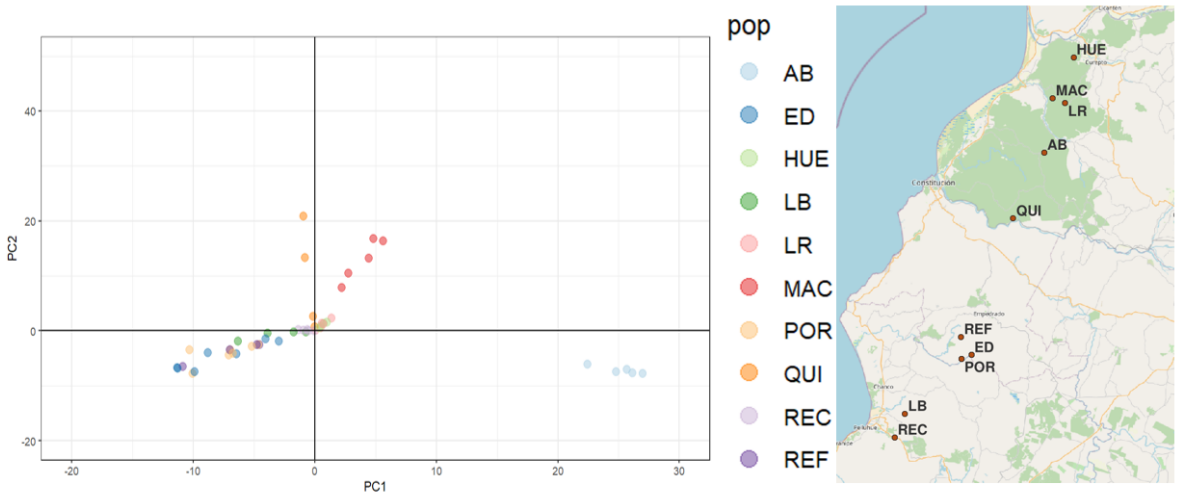



Figura 7: PCA generado a partir de las distancias genéticas de la matriz ETR0.03 TG 4, comparándose con el mapa de la distribución de las localidades muestreadas a la izquierda

### 3.4. K-medias y Análisis discriminante de componentes principales.

En la Figura 8, se compara cuatro matrices de los distintos minTagNumber para ETR 0.03, con gráficos multi-panel de las varianzas de las K-medias para cada grupo, un análisis discriminante de los componentes principales, y un gráfico de barras de la probabilidad posterior de a qué grupo pertenecería cada individuo en dinstios K-grupos.

En esta comparación, los 4 gráficos son casi iguales, debido a que este tipo de análisis optimiza la separación de individuos en grupos pre-definidos basados en la discriminación de las combinaciones lineales provenientes del PCA, en este caso se retienen solo dos componentes y se utilizó solo la probabilidad posterior de pertenencia, excluyéndose información valiosa para conocer la estructura genética de los individuos (Jombart y Collins 2015).

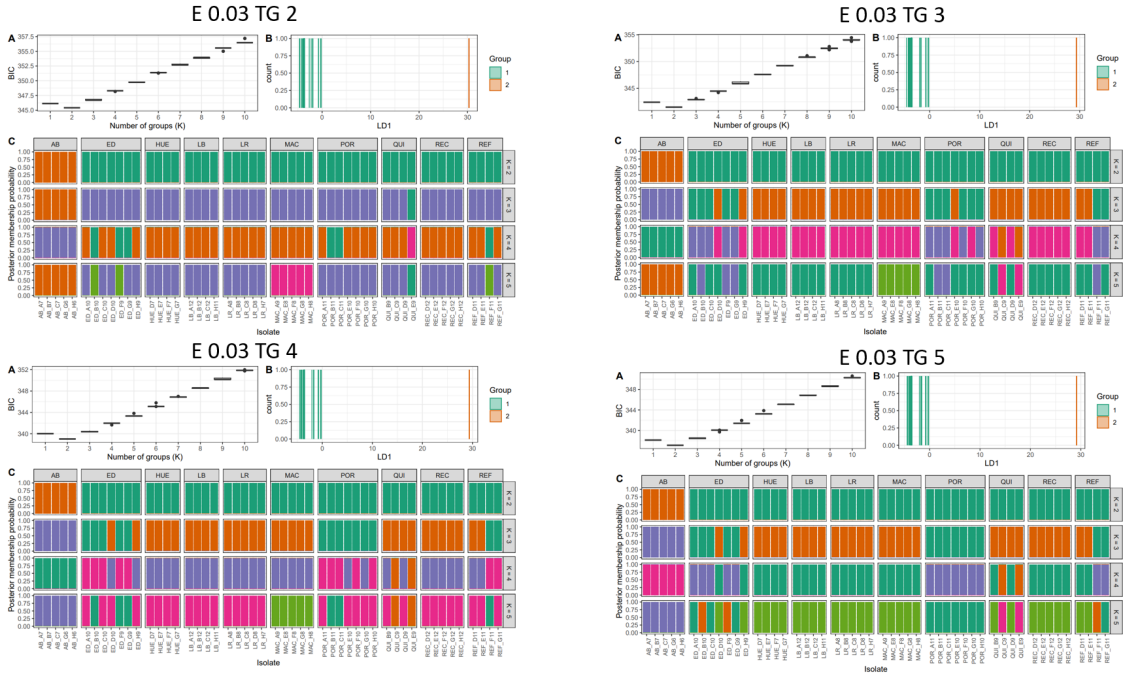



Figura 8: Comparación entre los minTagnumber para ETR 0.03 con gráficos multi-panel de tres partes, **A:** Varianza de las K-medias. Representa en su eje X el número de grupos, y en el eje Y, la varianza de las distancias intra de los grupos. **B:** Análisis discriminante de componentes principales para dos grupos. **C:** Gráfico de barras de la probabilidad posterior a que grupo pertenece cada individuo en distintos K-grupos.

En todas las matrices, que pueden ser revisadas en el apéndice C, el análisis de K-medias tiene su menor varianza intra-grupos con K = 2, y como se puede observar en la parte C de los gráficos de la figura 10, la localidad “AB” se diferencia tanto de los otros grupos, que hace que el resto de las localidades se agrupen todas juntas. Esto sucedió de igual forma para las matrices ETR 0.01 y ETR 0.05.

### 3.4.1. Proporción de coeficientes de ascendencia.

Debido a que no hubo mayor sensibilidad en el análisis anterior, se hizo un nuevo análisis de agrupamiento no supervisado particionado, basado en la factorización de una matriz no negativa dispersa (sMNF por sus siglas en inglés).

En la Figura 9 se observa que, según el criterio de entropía cruzada, el número más probable de grupos en estos sets de datos son tres poblaciones y este resultado se mantiene igual para el resto de las matrices ETR 0.01 Y ETR 0.05, que pueden observarse en el apéndice D

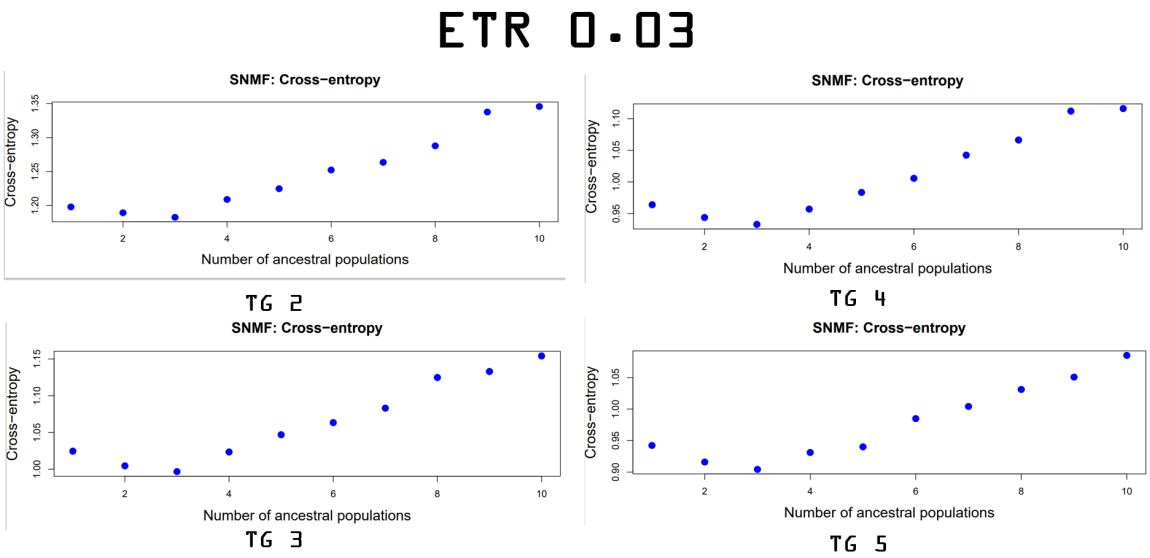



Figura 9: Gráficos para la validación cruzada de entropía para K = 1-10, para los distintos minTagNumber de las matrices de ETR 0.03

En la Figura 10, se grafican los coeficientes de ascendencia de tres ancestros para las matrices de ETR 0,03 para cada TG con los individuos ordenados de norte a sur, en los que no se observa una gran diferencia entre ellos, siendo igual para el resto de las matrices, que pueden ser revisadas en el apéndice E. En todas las matrices se observa que AB se diferencia de las demás localidades con un ancestro único representado con color rosa. También muestran que las localidades del norte son híbridos de los tres ancestros, con proporciones más homogéneas entre ellas que los individuos del sur, ya que muestran coeficientes de ascendencia similares con una gran proporción de verde oscuro y siempre algo de blanco y rosado, en cambio los individuos del sur, que también son híbridos, presentan menor proporción del ancestro rosado (o no presentan) y diferentes proporciones entre los ancestros verde oscuro y blanco dado cuenta de una posible especialización. Alguno de los individuos de más al sur presenta ancestros rosas del norte, que podrían ser migrantes o individuos traslocados en restauraciones pasadas.

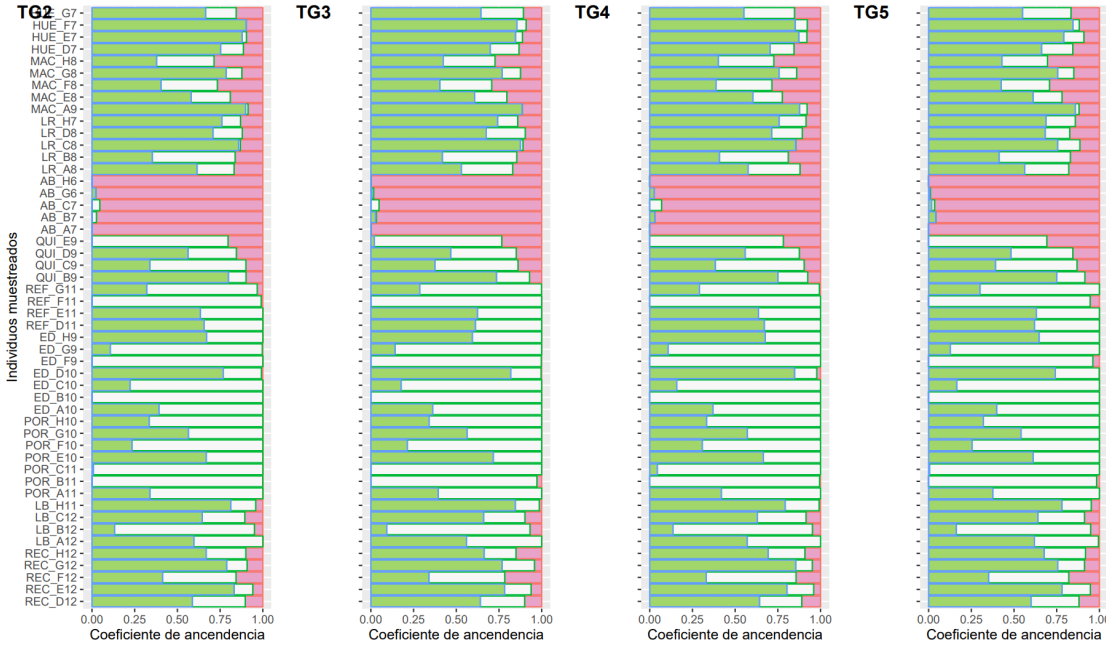



Figura 10: Gráfico de los coeficientes de ascendencia para las matrices ETR 0.03 para cada TG, con los individuos ordenados de norte a sur en la izquierda.

Debido a la dificultad de comparar las proporciones visualmente, se realizaron dendrogramas para cada población de las localidades muestreadas. En las que se pudo observar que en general se ordenan juntas por filtro, indicándonos que no hay una mayor diferencia entre las proporciones de los coeficientes. En la Figura 11 se ejemplifica un dendrograma para las proporciones de la localidad ED, en el que los individuos se ordenan en general por colores, significando que independiente del filtro los individuos tienen coeficientes de ascendencia similar.

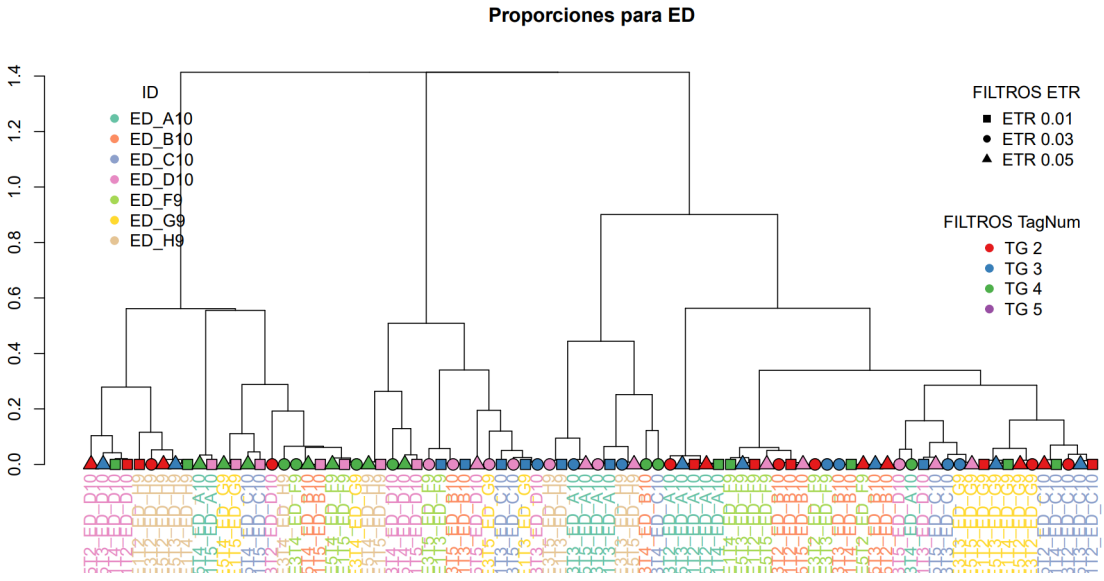



Figura 11: Dendrograma basado en distancia para los individuos de los filtros ETR 0,01, ETR 0,03, ETR 0,05 Y TG 2,3,4 y 5 para la localidad de ED.

En la Figura 12 se grafica en el mapa de las zonas muestreadas los coeficientes de ancestros en gráficos de torta para la matriz ETR0,03 TG4, en la que se reafirma la diferencia entre las localidades del norte y sur, mostrando 3 posibles grupos, con diferentes proporciones de ancestros.

La alta diversidad genética reportada por estudios anteriores, a pesar de la pequeña distribución de *N. alessandrii*, puede ser explicada porque la degradación de sus bosques es relativamente reciente en comparación a la antigüedad de la especie con individuos que superan los 200 años (Torres-Diaz *et al.* 2007). Además, los individuos de las localidades del sur como REC y LB que se agrupaban junto a las localidades del norte explica los resultados reportados anteriormente de un ecotipo diferente (Santelices Moya *et al.* 2009)

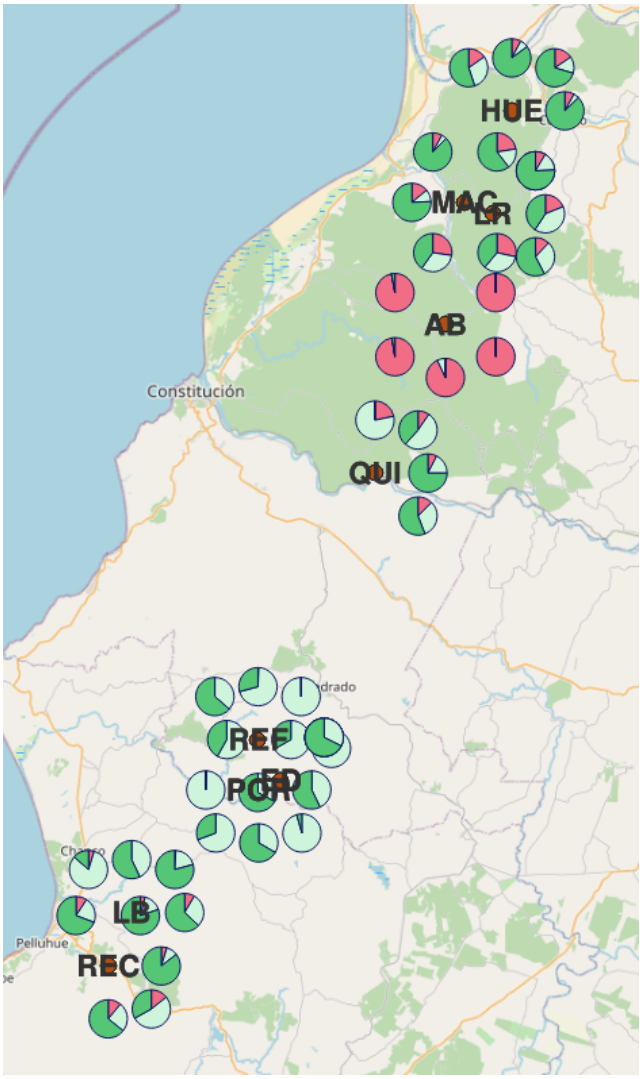



Figura 12: Mapa de las zonas de muestreo de las localidades son sus coeficientes de ancestros representados en gráficos de torta, de la matriz ETR 0,03 TG4.

# IV. CONCLUSIONES

- El parámetro de minTagnum es crucial para los datos pérdidos y para la cantidad de SNPs en su descubrimiento y genotipificación. En cambio, ETR si bien influencia la cantidad de SNPs detectados, no se encontraron diferencias entre los distintos análisis realizados.

-La estructura de datos poblacionales y su resolución en su informatividad no es influenciada directamente por la cantidad de SNPs ni con filtros más tolerantes. En el futuro, deberían probarse nuevos parámetros como filtrar por taxa, o la mínima cantidad de locus por individuo.

-Los mejores parámetros para una buena resolutividad basada en los dendrogramas de distancia se encontraron en ETR 0.05, con Tagnum 2, 3 y 5, y para ETR 0.03, con Tagnum 5 y 4.

-Se encontraron tres posibles ancestros para *Nothofagus alessandrii*, diferenciándose en la proporción de estos según la geografía norte a sur. La localidad aguas buenas parece ser una población diferente, es necesario hacer nuevas pruebas para corroborar esta información.

## V.REFERENCIAS BIBLIOGRÁFICAS:

1. Alcántara M.R.J.E.m. 2007. Breve revisión de los marcadores moleculares. 541-566.

2. Andrews K.R., J.M. Good, M.R. Miller, G. Luikart, P.A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17(2): 81-92.

3. Barstow M.E., C.; Baldwin, H. y Rivers, M. C. 2017. Nothofagus alessandrii (amended version of 2017 assessment).The IUCN Red List of Threatened Species 2020: e.T32033A177350927.

4. Bustamante R.O., C. Castor. 1998. The decline of an endangered temperate ecosystem: the ruil (Nothofagus alessandrii) forest in central Chile. Biodiversity & Conservation 7: 1607-1626.

5. Chessel D., A.B. Dufour, J. Thioulouse. 2004. The ade4 package-I-One-table methods. R news 4(1): 5-10.

6. Danecek P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15): 2156-2158.

7. Donoso C., E. Landaeta. 1983. Ruil (Nothofagus alessandri), a threatened Chilean tree species. Environmental Conservation 10(2): 159-162.

8. Ekblom R., J. Galindo. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107(1): 1-15.

9. Elshire R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. 6(5): e19379.

10. Frichot E., F. Mathieu, T. Trouillon, G. Bouchard, O. François. 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics 196(4): 973-983.

11. Gajardo J., M. Yáñez, S. Espinoza, M. Carrasco-Benavides, Y. Ormazábal, C. Mena, et al. 2022. Comparison of the absolute and relative difference spectral indices to estimate burn severity: the case of endangered nothofagus alessandrii (ruil). Ecological Restoration 40(3): 191-202.

12. Glaubitz J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. 9(2): e90346.
13. González M.E., M. Galleguillos, J. Lopatin, C. Leal, C. Becerra-Rodas, A. Lara, et al. 2022. Surviving in a hostile landscape: Nothofagus alessandrii remnant forests threatened by mega-fires and exotic pine invasion in the coastal range of central Chile. Oryx: 1-11.
14. Grau P.C. 1970. Factores en la destrucción del paisaje chileno: recolección, caza y tala coloniales. Investigaciones Geográficas(20): ág. 235-264.
15. Hechenleitner V., M.F. Gardner. 2005. Plantas amenazadas del centro-sur de Chile. Universidad Austral de Chile y Real Jardín Botánico de Edimburgo.
16. Höglund J. 2009. Evolutionary Conservation Genetics. 1-200 p.
17. Hosner P.A., B.C. Faircloth, T.C. Glenn, E.L. Braun, R.T. Kimball. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Molecular biology and evolution 33(4): 1110-1125.
18. Huang H., L.L. Knowles. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic biology 65(3): 357-365.
19. Jombart T., C. Collins. 2015. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.
20. Jombart T., Z.N. Kamvar, C. Collins, R. Lustrik, M.-P. Beugin, B.J. Knaus, et al. 2018. Package 'adegenet'. Github Repository Available online: https://github com/thibautjombart/adegenet (accessed on 11 November 2021).
21. Kassambara A., M.A. Kassambara. 2020. Package 'ggpubr'. R package version 01 6(0).
22. Knaus B.J., N.J. Grünwald. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular ecology resources 17(1): 44-53.
23. Lu F., J. Glaubitz, J. Harriman, T. Casstevens, R.J.W.P. Elshire. 2012. TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation. 2012: 1-12.
24. Lu F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, et al. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS Genetics 9(1): e1003215.

25. Mir A., F. Rosselló. 2013. A new balance index for phylogenetic trees. Mathematical biosciences 241(1): 125-136.

26. Morin P.A., G. Luikart, R.K. Wayne. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution 19(4): 208-216.

27. Moya R.S., S.E. Meza, A.C. Ariza, C.M. Díaz. 2018. Risk management as a tool for the conservation of the forests of nothofagus alessandrii, an endangered species of central Chile. Interciencia 43(2): 144-150.

28. Myers N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. Da Fonseca, J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature 403(6772): 853-858.

29. Narum S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, P.A.J.M.e. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. 22(11): 2841-2847.

30. Neuwirth E., R.C. Brewer. 2014. ColorBrewer palettes. R package version 1: 4.

31. Paradis E., S. Blomberg, B. Bolker, J. Brown, J. Claude, H.S. Cuong, et al. 2019. Package 'ape'. Analyses of phylogenetics and evolution, version 2(4): 47.

32. Pineda Bravo G.E. 1998. Determinación de los patrones de variabilidad genética en poblaciones de raulí (Nothofagus alpina (Poepp. et Endl.) Oerst.) y ruil (Nothofagus alessandrii Espinosa), por medio de electroforesis horizontal en geles de almidón.

33. QGIS D.T. 2015. QGIS geographic information system. Open source geospatial Foundation project.

34. San Martín J.F., Heriberto, C. Ramírez. 1984. Fitosociología de los bosque s de mil (Nothofagus alessandri Espinoza) en Chile Central. Revista Chilena de Histori a Natural 57: 171-200.

35. Santelices Moya R., R.M. Navarro Cerrillo, F. Drake Aranda. 2009. Caracterización del material forestal de reproducción de cinco procedencias de nothofagus alessandrii espinosa, una especie en peligro de extinción. Interciencia 34: 113-120.

36. Santelices R., F. Drake, C. Mena, R. Ordenes, R.M. Navarro-Cerrillo. 2012. Current and potential distribution areas for Nothofagus alessandrii, an endangeredtree species from central Chile. Ciencia e investigación agraria 39(3): 521-531.

37. Sunnucks P. 2000. Efficient genetic markers for population biology. Trends in Ecology & Evolution 15(5): 199-203.

38. Team R.C. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www R-project org/.

39. Torkamaneh D., J. Laroche, F. Belzile. 2016. Genome-Wide SNP Calling from Genotyping by Sequencing (GBS) Data: A Comparison of Seven Pipelines and Two Sequencing Technologies. Plos One 11(8).

40. Torres-Diaz C., E. Ruiz, F. Gonzalez, G. Fuentes, L.A. Cavieres. 2007. Genetic diversity in Nothofagus alessandrii (Fagaceae), an endangered endemic tree species of the coastal maulino forest of central chile. Annals of Botany 100(1): 75-82.

41. Valencia D., J. Saavedra, J. Brull, R. Santelices. 2018. Fire severity damages caused on Nothofagus alessandrii forest on the maule region of chile. Gayana - Botanica 75(1): 531-534.

42. Wagner C.E., I. Keller, S. Wittwer, O.M. Selz, S. Mwaiko, L. Greuter, et al. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology 22(3): 787-798.

43. Yu L.X., P. Zheng, S. Bhamidimarri, X.P. Liu, D. Main. 2017. The impact of genotyping-by-sequencing pipelines on SNP discovery and identification of markers associated with verticillium wilt resistance in autotetraploid alfalfa (Medicago sativa L.). Frontiers in Plant Science 8(FEBRUARY).

# VI. APÉNDICE

A. PCA realizados para las distintas matrices ETR 0,01, 0,03 y 0,05, y para los TG 2,3,4 y 5

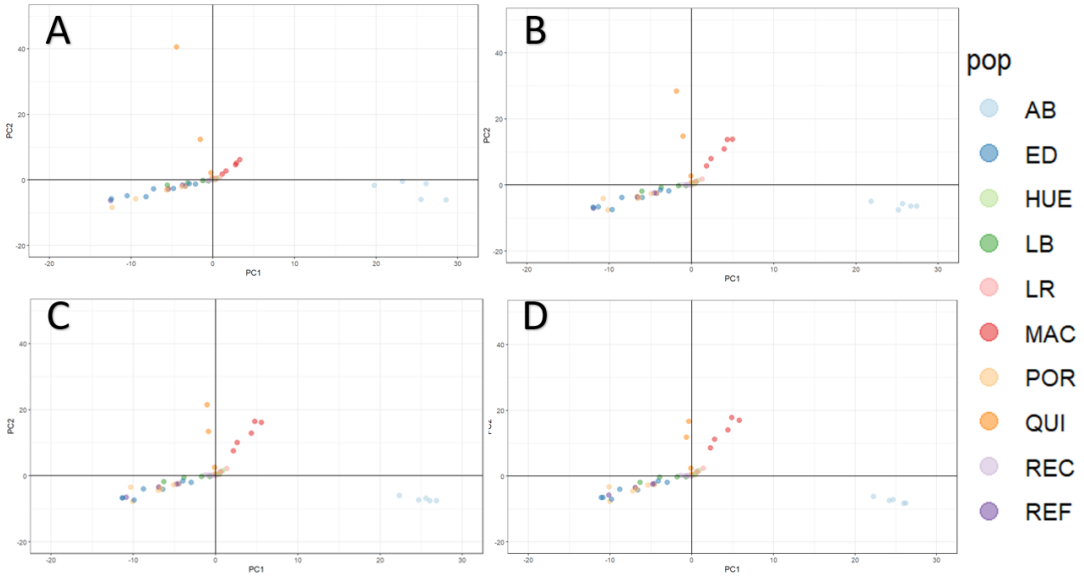



Figura A.1: Análisis de componentes principales para las matrices de ETR0.01, A: Tagnum 2; B: Tagnum 3; C: Tagnum 4; D: Tagnum 5

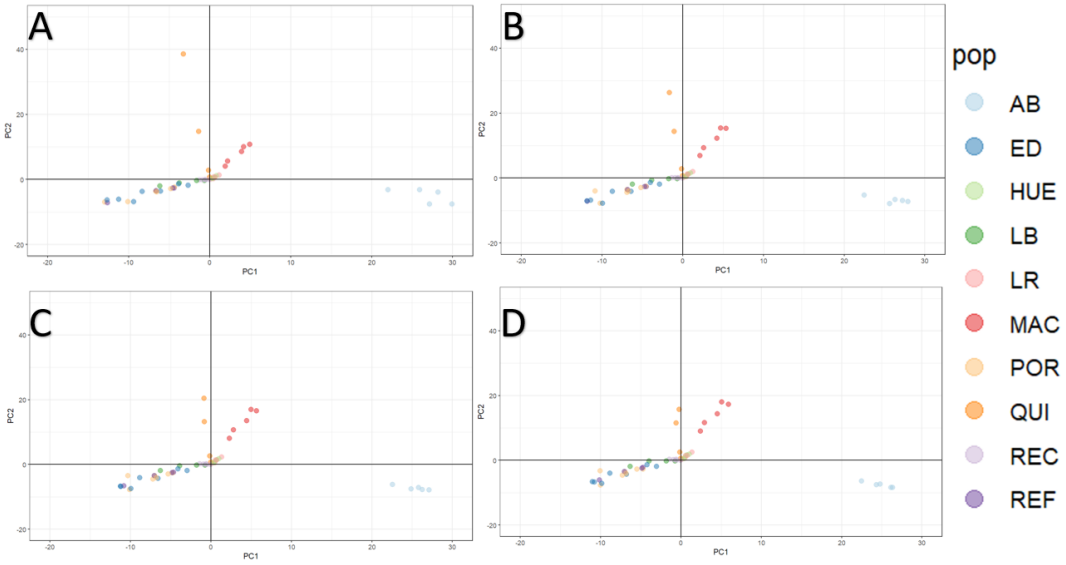



Figura A.2: Análisis de componentes principales para las matrices de ETR0.03, A: Tagnum 2; B: Tagnum 3; C: Tagnum 4; D: Tagnum 5

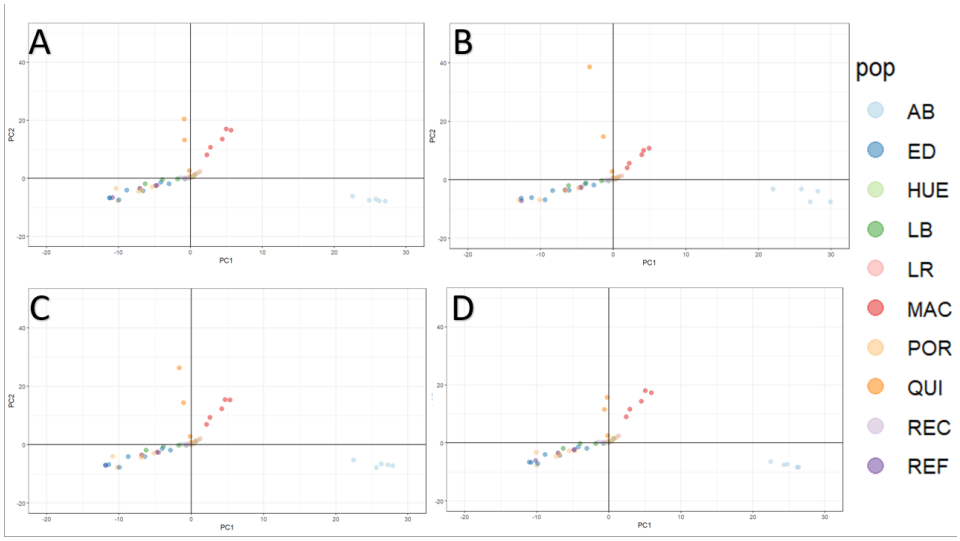



Figura A.3: Análisis de componentes principales (PCA) para las matrices de ETR0.05, A: Tagnum 2; B: Tagnum 3; C: Tagnum 4; D: Tagnum 5

B. Dendrogramas realizados las distintas matrices ETR 0,01, 0,03 y 0,05, y para los TG 2,3,4 y 5

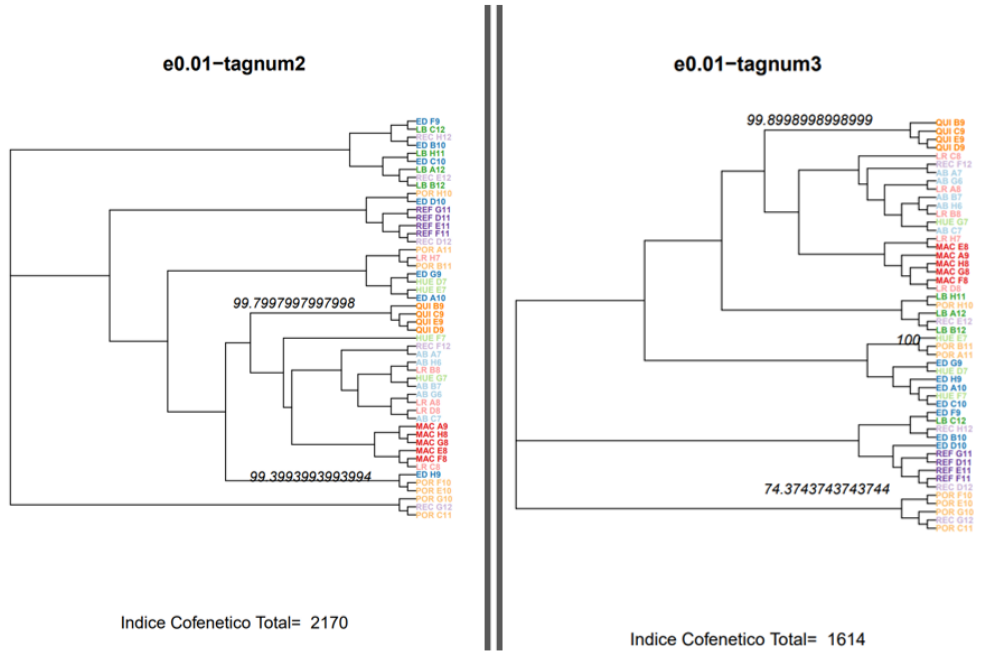



Figura B.1: Dendrograma basado en distancia para E0,01, a la izquierda TG2 y a la derecha TG3

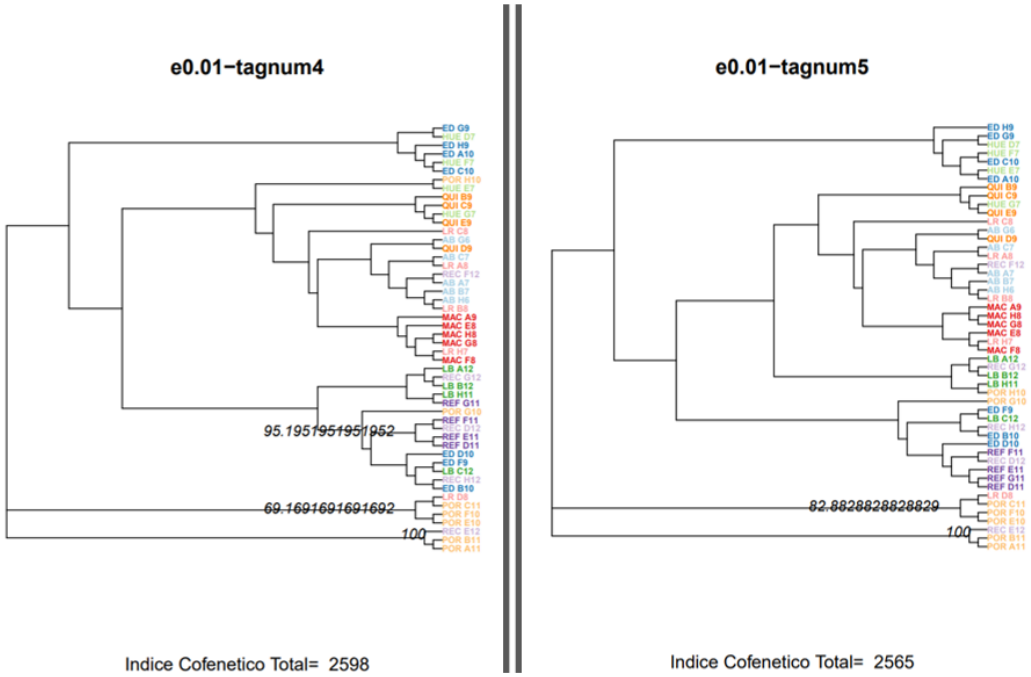



Figura B.2: Dendrograma basado en distancia para E0,01, a la izquierda TG4 y a la derecha T5

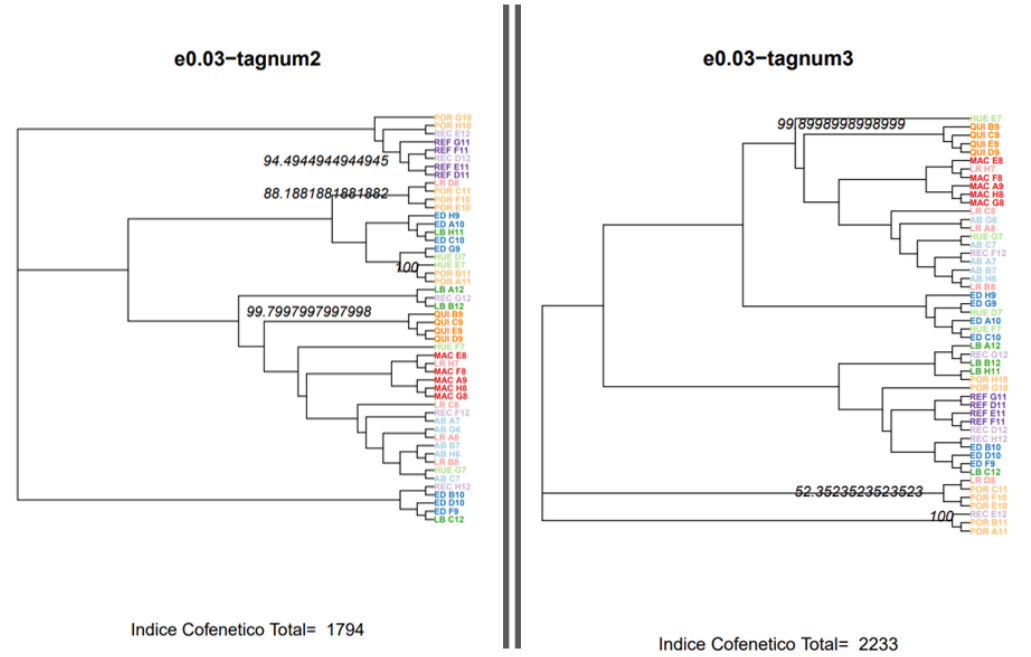



Figura B.3 Dendrograma basado en distancia para E0,03, a la izquierda TG2 y a la derecha TG3

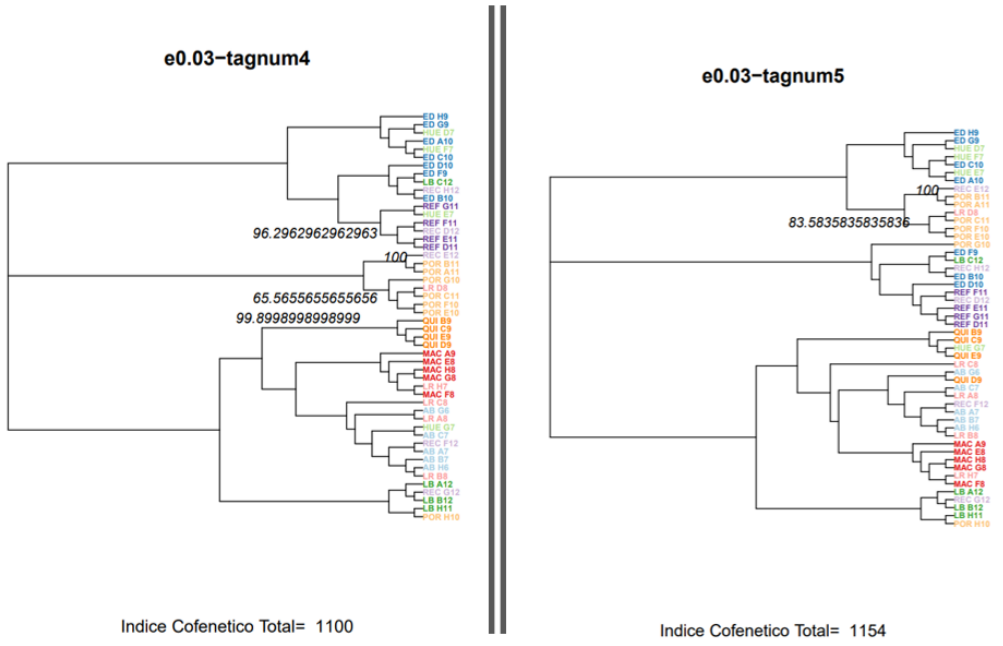



Figura B.4 Dendrograma basado en distancia para E0,03, a la izquierda TG4 y a la derecha T5

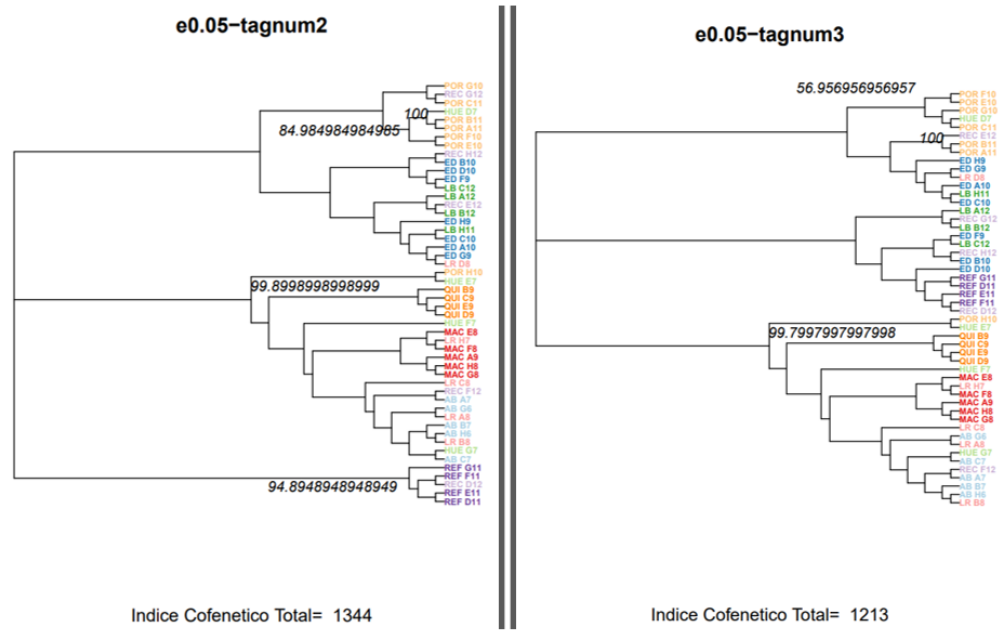



Figura B.5: Dendrograma basado en distancia para E0,05, a la izquierda TG2 y a la derecha TG3

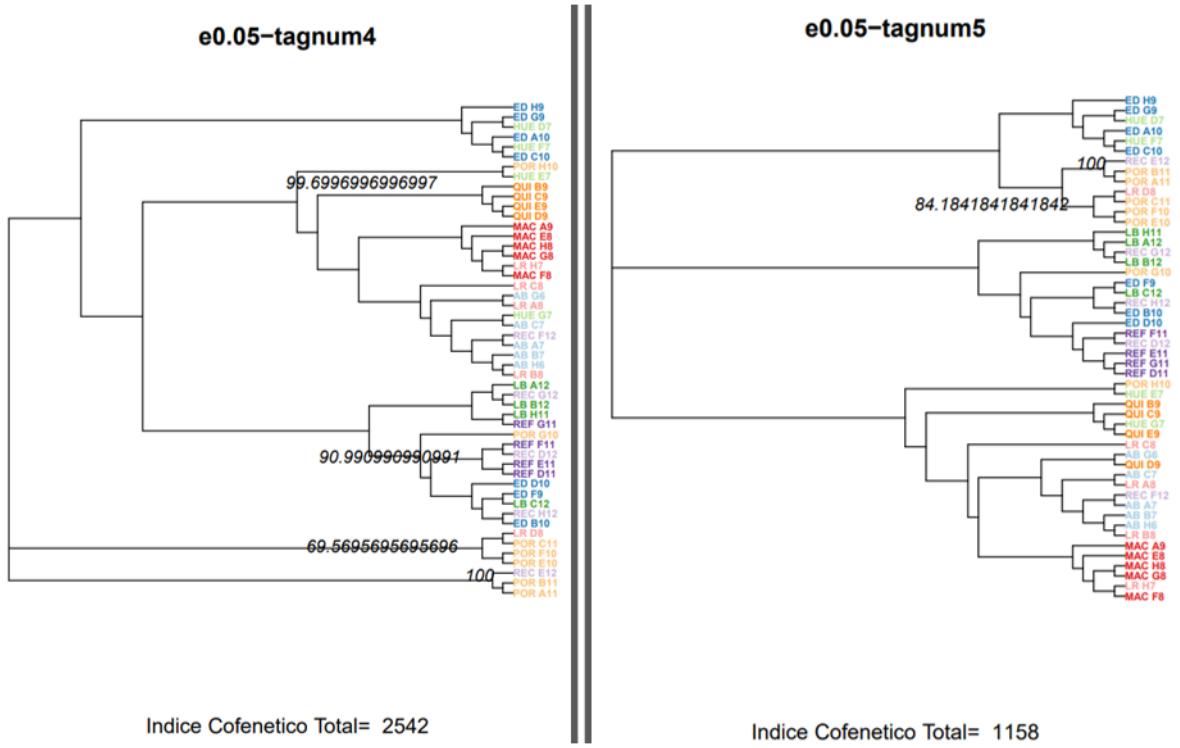



Figura B.6: Dendrograma basado en distancia para E0,05, a la izquierda TG4 y a la derecha T5

C. DAPC realizados para las distintas matrices ETR 0,01, 0,03 y 0,05, y para los TG 2,3,4 y 5

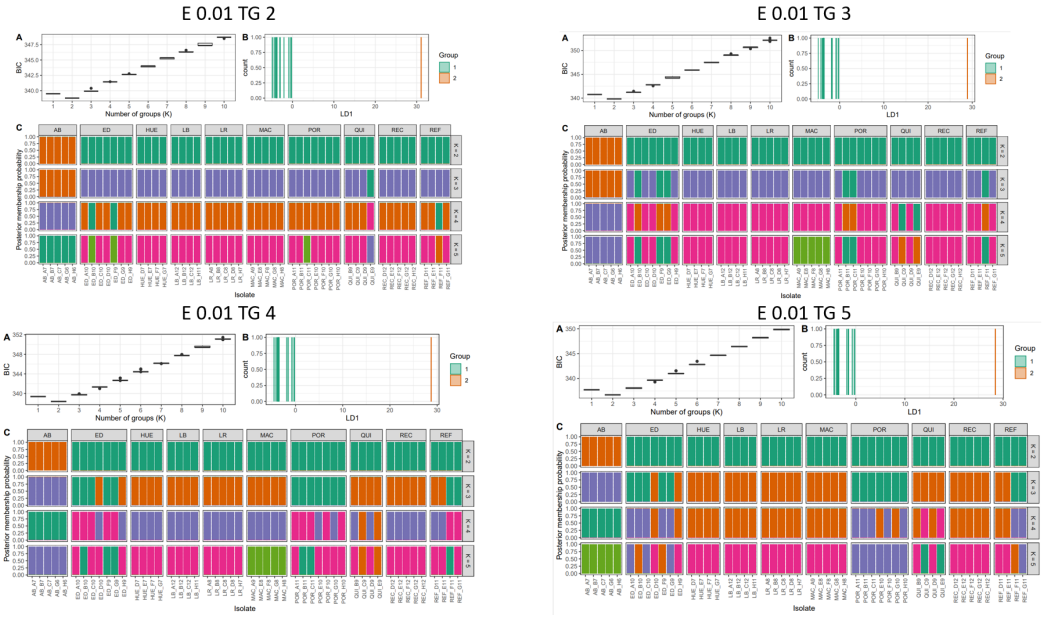



Figura C.1 DAPC para ETR 0,01, arriba a la izquierda TG2, arriba a la derecha TG3, abajo a la izquierda T4 y abajo a la derecha TG5.

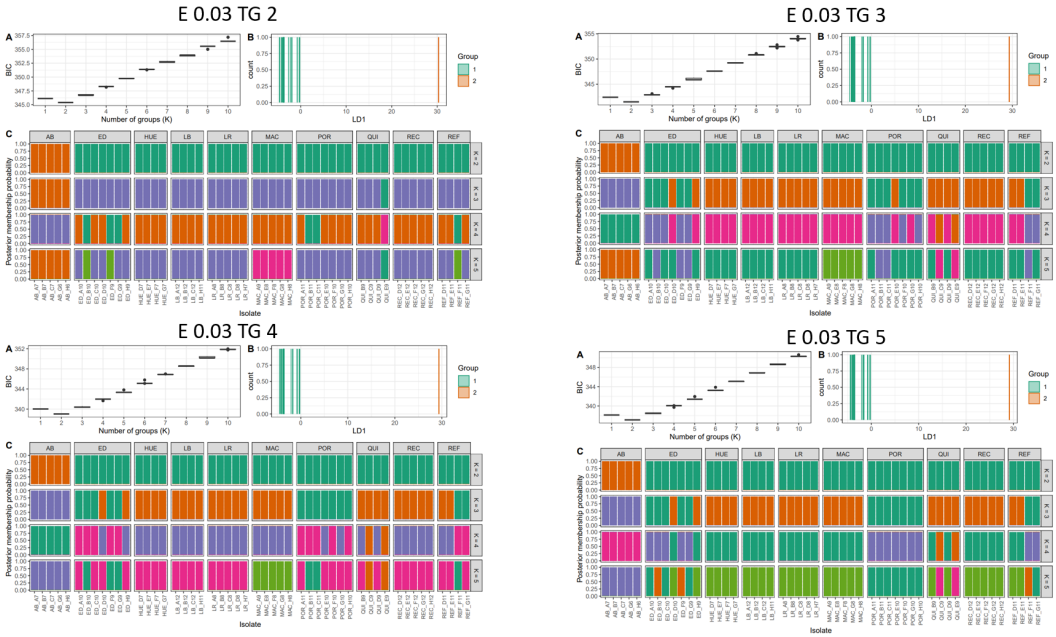



Figura C.2 DAPC para ETR 0,03, arriba a la izquierda TG2, arriba a la derecha TG3, abajo a la izquierda T4 y abajo a la derecha TG5.

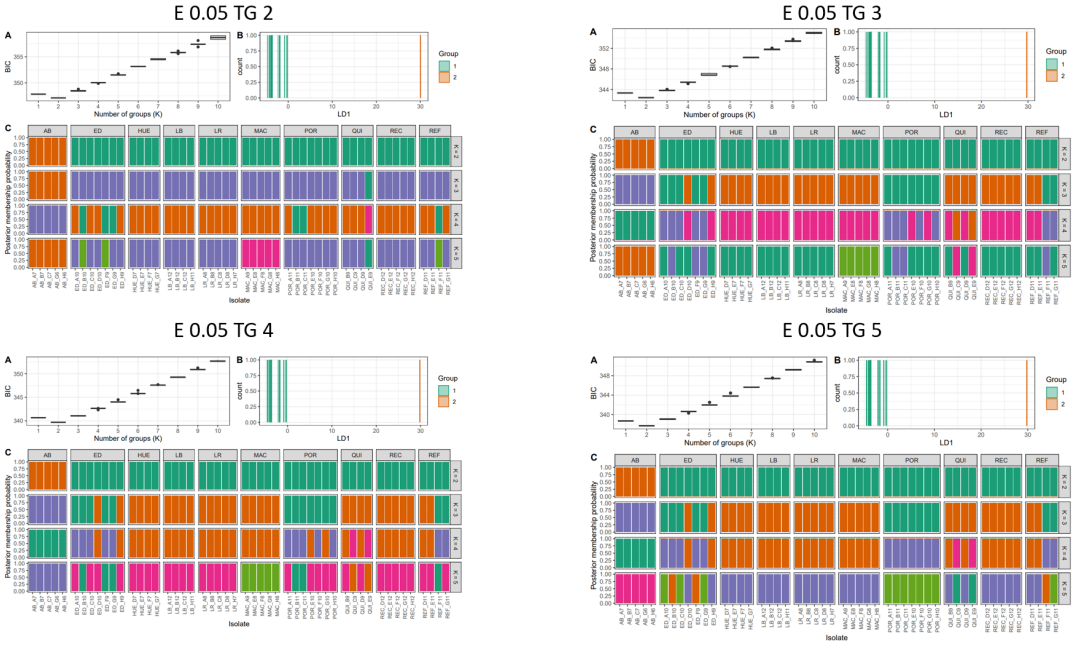



Figura C.3 DAPC para ETR 0,05, arriba a la izquierda TG2, arriba a la derecha TG3, abajo a la izquierda T4 y abajo a la derecha TG5.

D. Gráficos para la validación cruzada de entropía para K = 1-10 para las distintas matrices ETR 0,01 Y 0,05, con TG 2, 3, 4 y 5

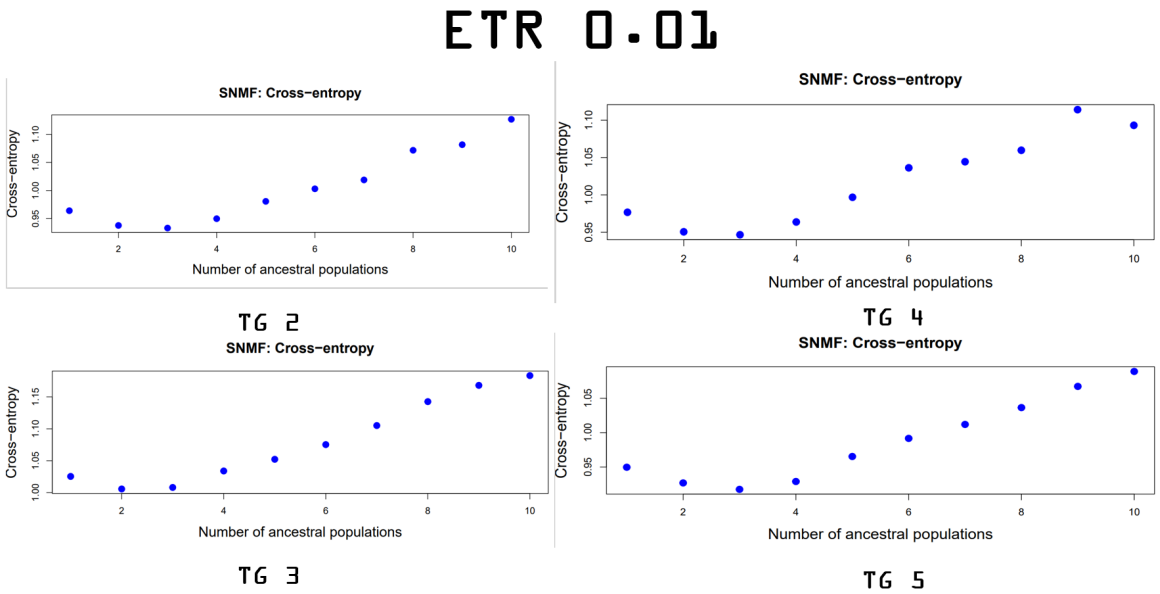



Figura D.1: Gráficos para la validación cruzada de entropía para K = 1-10, para los distintos minTagNumber de las matrices de ETR 0.01

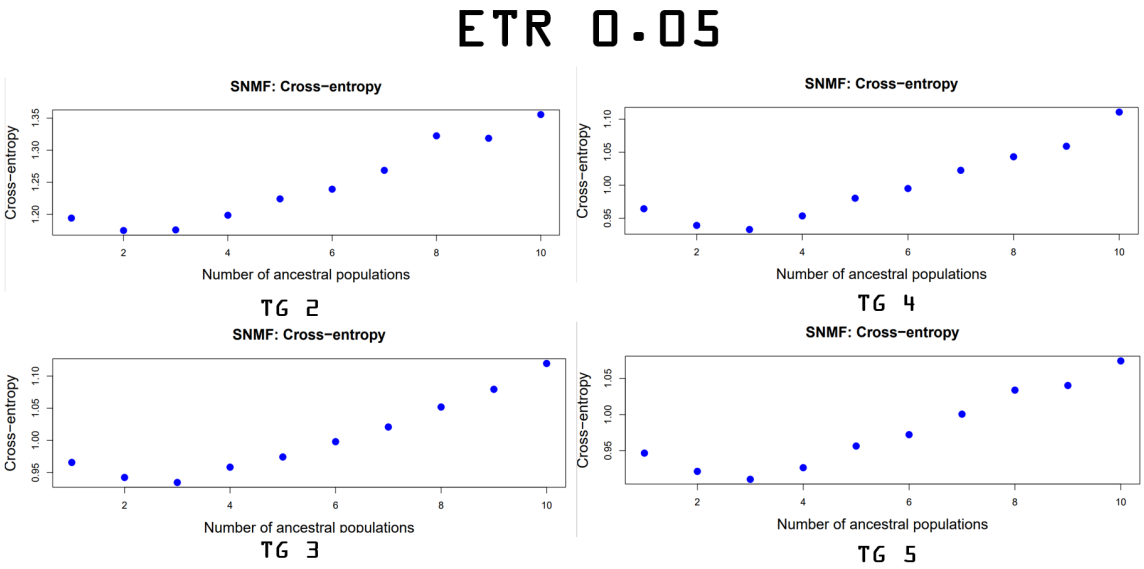



Figura D.2: Gráficos para la validación cruzada de entropía para K = 1-10, para los distintos minTagNumber de las matrices de ETR 0.5

E. SNMF creados para las distintas matrices ETR 0,01, 0,02 y 0,05, y TG 2,3,4 y 5

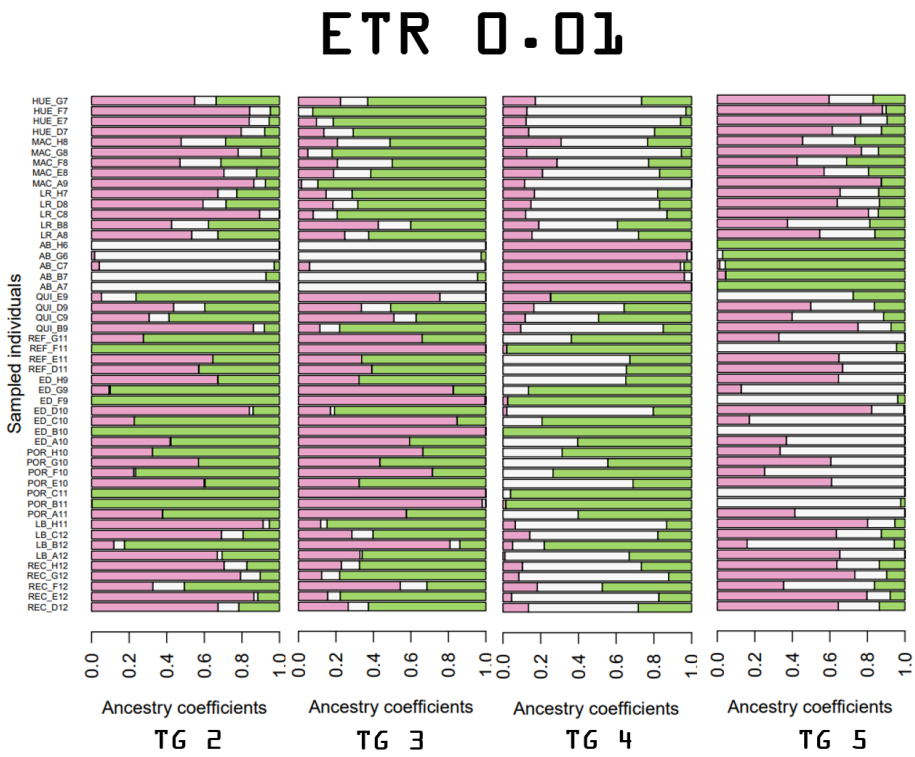



Figura E.1 SNMF con 3 ancestros para ETR 0,01 de izquierda a derecha para los TG 2, 3 4 y 5

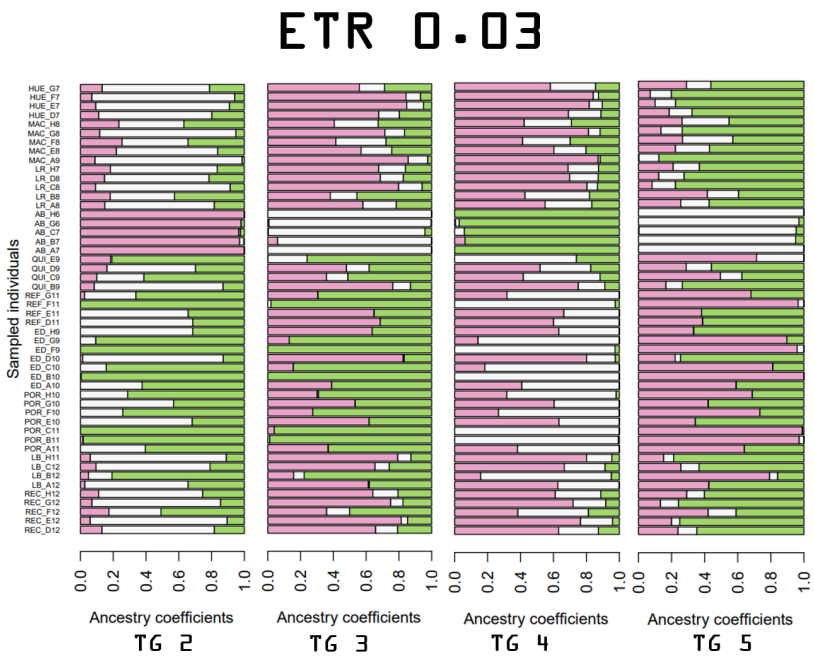



Figura E.2 SNMF con 3 ancestros para ETR 0,03 de izquierda a derecha para los TG 2, 3 4 y 5

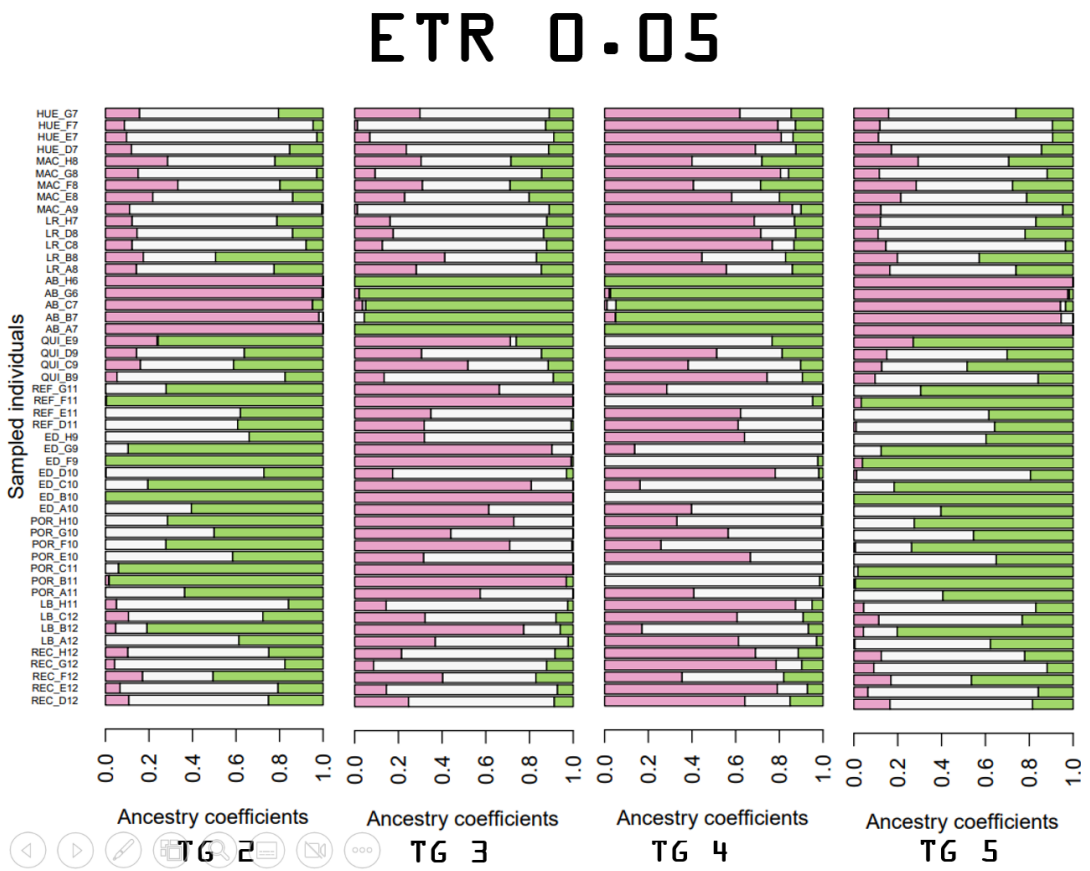



Figura E.3 SNMF con 3 ancestros para ETR 0,05 de izquierda a derecha para los TG 2, 3 4 y 5