UNIVERSIDAD DE TALCA

INSTITUTO DE CIENCIAS BIOLÓGICAS

PREDICCIÓN GENÓMICA BASADA EN LOCI EN DESEQUILIBRIO DE LIGAMIENTO, EN DOS ESPECIES DE *EUCALYPTUS* QUE DIFIEREN EN SU HISTORIAL DE SELECCIÓN Y ESTRUCTURA GENÉTICA

PAULINA ANDREA BALLESTA MUÑOZ

Tesis para optar al grado académico de

Doctor en Ciencias mención en

Ingeniería Genética Vegetal

DIRECTOR DE TESIS

DR. FREDDY MORA POBLETE

SEPTIEMBRE 2020

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el

autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma

total o parcial e ilimitada del mismo.

Talca, 2022

PREDICCIÓN GENÓMICA BASADA EN LOCI EN DESEQUILIBRIO DE LIGAMIENTO EN DOS ESPECIES DE *EUCALYPTUS* QUE DIFIEREN EN SU HISTORIAL DE SELECCIÓN Y ESTRUCTURA GENÉTICA

HAPLOTYPE-BASED GENOMIC PREDICTION IN TWO *EUCALYPTUS* SPP. WITH DIFFERENT GENETIC STRUCTURE AND SELECTION HISTORY

PAULINA ANDREA BALLESTA MUÑOZ

Fecha de inicio de la Tesis: 13 de marzo de 2016

Fecha de término de la Tesis: 10 de marzo de 2020

PROFESOR GUÍA

DR. FREDDY MORA POBLETE

Universidad de Talca

Instituto de Ciencias Biológicas

fmora@utalca.cl

Rodrigo Hasbún

Ingeniero Forestal, Dr. _____

Francisco Zamudio

Ingeniero Forestal, Dr. _____

Christian Figueroa

Licenciado en Ciencias Biológicas, Dr. _____

Agradecimientos

Deseo agradecer a mi pareja, amigas, amigos, familia por todo el apoyo que he recibido de ellos durante estos últimos 4 años. En particular, quiero agradecer a Don Freddy Mora, quien ha dado un voto de confianza en mi durante años, y le debo la mayor parte de mi formación profesional, además de años de amistad. Quiero agradecer a Sergio, mi pareja, por su contención y comprensión en momentos complicados. A mi hermana, quien siempre ha estado pendiente de mi bienestar, a pesar de la distancia. A mis amigas y amigos, de quienes recibí apoyo emocional e intelectual.

ÍNDICE DE MATERIAS

RESUMEN

Las especies del género *Eucalyptus* L'Hér son económicamente relevantes en el contexto nacional y mundial, las cuales han sido mejoradas genéticamente para diferentes características en diversos ambientes. Los avances en las técnicas de secuenciación de genomas han posibilitado la identificación de diversos factores genéticos que determinan la variación fenotípica de diferentes características económicamente relevantes, los cuales son utilizados en selección asistida por marcadores moleculares (MAS) y en la selección genómica (GS) en diferentes especies de *Eucalyptus*. La eficacia de los modelos predictivos de GS ha sido variable entre diferentes poblaciones de mejoramiento, dependiendo de la naturaleza de la característica estudiada. Por ejemplo, en características relacionadas al crecimiento de los árboles, el poder predictivo de los modelos genómicos ha sido relativamente moderado, a pesar de contar con una alta densidad de marcadores (>10000 polimorfismos de nucleótido único; SNPs). En términos generales, la precisión de los modelos de GS depende de los supuestos analíticos de los modelos predictivos, aspectos intrínsecos de las poblaciones (por ejemplo, diversidad y estructura genética), y de la arquitectura genética de los rasgos que se desean predecir. En este sentido, diversos estudios han sido realizados con el objetivo de aumentar el poder predictivo en características complejas. En cultivos agrícolas, por ejemplo, se ha observado que la GS basada en loci en desequilibrio de ligamiento (bloques de haplotipos) podría mejorar la precisión de los métodos de predicción genómica en características de baja heredabilidad en sentido amplio. Por otra parte, no existen estudios que aborden el enfoque de haplotipos en la predicción de rasgos complejos de árboles forestales.

En base a lo descrito anteriormente, el presente trabajo tuvo como objetivo determinar regiones genómicas en desequilibrio de ligamiento y su influencia en la predicción genómica de caracteres poligénicos en dos especies de *Eucalyptus* (*E. globulus* y *E. cladocalyx*) que difieren en su historial de selección y estructura genética. Para ello, en una primera etapa, se estudió la tasa de transferencia de marcadores SNPs en ambas especies, usando un arreglo de ADN de alta densidad (60K), desarrollado en *E. grandis* y en otras especies del género.

La transferibilidad de los marcadores SNP fue mayor en *E. globulus* (~14000 SNPs) que en *E. cladocalyx* (~3800 SNPs), lo cual puede ser explicado por el hecho de que *E. cladocalyx*

pertenece a una sección genéticamente distante de las secciones de *Eucalyptus* utilizadas en el desarrollo del arreglo de SNP. Consecuentemente, un menor número de bloques de haplotipos fueron identificados en la población de *E. cladocalyx* (~108) que en la población de *E. globulus* (~1137). En características de baja heredabilidad ($h^2<0,1$), los bloques de haplotipos mejoraron el poder predictivo, y estimaron mayores valores de heredabilidad (genómica) comparado con las estimaciones basadas en pedigrí. Por otro lado, los modelos GS tuvieron una mejor respuesta en *E. globulus*. En este sentido, la baja densidad de marcadores encontrada en *E. cladocalyx* podría haber influido en estos resultados. Para mejorar la predicción genómica en esta especie, se propuso un enfoque de predicción que combina los modelos de GS con el mapeo de loci de características cuantitativas (QTLs) y los antecedentes de estructura genética de la población (genealogía).

Se concluye que el enfoque de haplotipos podría ser especialmente beneficioso para predecir características de baja heredabilidad de *Eucalyptus*, en un escenario donde la base genética ha sido reducida debido a la intensidad de selección. Por otro lado, en poblaciones que no poseen un historial de selección artificial, el uso combinado de los antecedentes de estructura genética y la información genómica contribuyen a una mejor predicción fenotípica. Los hallazgos de este estudio contribuyen a optimizar el proceso de selección de árboles en la industria forestal basada en *Eucalyptus*.

Palabras clave: Bloques de haplotipos, Pedigrí, Poder predictivo, Selección genómica, Variabilidad genética.

ABSTRACT

Eucalyptus are economically important species in the local and worldwide context, which have been genetically improved for different traits under several environmental conditions. Advances in genome sequencing techniques have allowed to identify several genetic factors explaining the phenotypic variation of different economic traits. Currently, molecular markers (MM) are used in genomic selection (GS) to predict the breeding value of an individual. Different *Eucalyptus* spp. have been studied with the principles of GS. The efficacy of the GS models has been variable among different breeding populations, depending strongly on the genetic architecture of the studied trait. For instance, the ability of GS models to predict growth-related traits has been relatively moderate, despite having a high density of markers (> 10,000 single nucleotide polymorphisms; SNPs). In general, the accuracy of GS depends on: the analytical assumptions of the prediction models, marker density, linkage disequilibrium pattern, intrinsic features of the populations (instance, diversity and genetic structure), and the genetic architecture of the target traits. In this sense, several studies have been carried out for increasing the predictive ability in complex traits. In crops, it has been reported that GS based on loci in linkage disequilibrium (haplotype blocks) could improve the accuracy of GS methods, due to these genomic regions have a greater ability to predict low heritable traits than SNPs. On the other hand, there are not studies that address the haplotype approach in the prediction of complex traits of forest trees.

Based on this information, the aim of the present study was to identify genomic regions in linkage disequilibrium (LD) and their influence on the genomic prediction of polygenic traits in two *Eucalyptus* spp. (*E. globulus* and *E. cladocalyx*) that differ in their selection history and genetic structure. For this, the transferability of SNPs markers from a high-density DNA array (more than 60,000 SNPs) to the studied *Eucalyptus* spp. was evaluated. The transferability rate in *E. globulu*s (~ 14,000 SNPs) was higher than in *E. cladocalyx* (~ 3,800 SNPs), which can be explained by the fact that *E. cladocalyx* belongs to a genetically distant section from the *Eucalyptus* sections used in the preparation of the SNPs array. Consequently, fewer haplotype blocks were identified in *E. cladocalyx* population (~ 108) than *E. globulus* population (~ 1137). According to the results, the inclusion of haplotypes as predictor variables in GS models improved the predictive ability of low heritability (<0.1) traits. Additionally, the genomic

x

heritability based on haplotypes was higher than those based on pedigree for these traits. On the other hand, the GS models had a better ability to predict the phenotypic traits evaluated in *E. globulus* than *E. cladocalyx*. In this sense, the low density of markers could be influencing the predictive ability of phenotypic traits of *E. claodcalyx*. A prediction approach was implemented that combines the benefits of GS, the quantitative trait loci mapping (QTLs) and the genetic structure of the population (genealogical antecedents), to obtain a greater predictive ability of the traits evaluated in *E. cladocalyx*.

In conclusion, the haplotype approach could be especially beneficial to predict low heritability traits of *Eucalyptus*, in a context of the limited genetic base of these traits, due to the artificial selectionintensity. On the other hand, in populations that have not been subjected to artificial selection, the combined use of genetic structure and genomic information contribute to a better phenotypic prediction. The findings of this study could optimize the tree selection process and be useful for the forestry industry.

Keywords: Artificial selection, Genetic variability, Genomic selection, Haplotype blocks, Pedigree, Predictive ability.

I. INTRODUCCION

1.1 Presentación del problema

La metodología de modelos mixtos de Herdenson ha sido frecuentemente utilizada para determinar el mérito genético de plantas y animales, en el contexto de selección y mejoramiento de características de interés económico, considerando la mejor predicción lineal insesgada (BLUP; *Best Linear Unbiased Prediction*) como método estándar de predicción. El mérito genético es predicho de acuerdo al grado de parentesco entre individuos, bajo la suposición de que los individuos con un linaje común son fenotípicamente cercanos. Recientemente, este procedimiento de predicción ha sido complementado con información genómica, cuyo método combinado es conocido como selección genómica o predicción de genoma amplio (GS; Meuwissen et al., 2001). En el contexto del mejoramiento genético de *Eucalyptus*, a pesar de contar con una densidad de marcadores relativamente alta (superior a 10000 SNPs), algunas características relacionadas al crecimiento, que tuvieron una heredabilidad basada en pedigrí relativamente baja (h^2<0,2), y que son de interés económico, los modelos GS han tenido un poder de predicción considerado como moderado (inferior a 0,6; Tan et al., 2017; Müller et al., 2017; Suontama et al., 2019).

Tanto en los métodos tradicionales como GS, la precisión predictiva del fenotipo de un individuo depende de que tan precisa es la estimación de las relaciones genéticas entre los individuos (Goddard et al., 2011; Scutari et al., 2016). El poder predictivo de un modelo GS puede incrementarse si los individuos que se utilizan como referencia para estimar los efectos aditivos (de los loci), están genéticamente relacionados a los individuos que se desean predecir fenotípicamente (Ly et al., 2013; Lee et al., 2017; Thistlethwaite et al. 2020). En este contexto, para un buen uso de GS, se deben considerar varios factores genéticos que son intrínsecos de la población de estudio (Norman et al., 2018), tales como los patrones de desequilibrio de ligamiento (DL), la estructura genética de la población, la diversidad genética, entre otros (Wientjes et al., 2013; Wolc et al., 2015; Sun et al., 2016; Schopp et al., 2017); y de esta manera optimizar las tasas de ganancia genética. Las relaciones de parentesco derivadas de pedigrí representan la proporción (en forma teórica) del genoma que es compartido entre individuos. Como consecuencia, la matriz de parentesco generada por los antecedentes genealógicos es una

estimación no sesgada de las relaciones de parentesco entre los genes que controlan el fenotipo (Goddard et al. 2011; Velazco et al., 2019). No obstante, en algunas poblaciones no es posible realizar una buena reconstrucción de la estructura genética que define a la población (por ejemplo, debido a contaminación de polen, ensayos de polinización abierta, entre otros), lo cual conlleva a obtener una menor precisión en la estimación de parámetros genéticos (Klápšťe et al., 2017). Inclusive, los eventos en la historia evolutiva de la población podrían ser ignorados, lo cual dificultaría la estimación precisa de los parámetros genéticos que son claves en la selección de árboles superiores. Las estimaciones del parentesco basadas en regiones genómicas son una mejor aproximación de la proporción real del genoma que comparten dos individuos. Cabe destacar que para determinar con precisión las relaciones genéticas entre individuos se requiere una alta densidad de marcadores (Voorrips et al., 2016). Actualmente, la determinación del parentesco entre individuos es basada en polimorfismos de nucleótido únicos (SNPs) o haplotipos (Edwards, 2015; Howard et al., 2017; Mathew et al., 2018). De acuerdo con Edwards (2015), los haplotipos pueden establecer relaciones genealógicas más precisas y confiables que una matriz de parentesco construida netamente por SNPs, debido a que permiten examinar la identidad por descendencia que existe entre los individuos de una población. Interesantemente, los haplotipos han sido utilizados en modelos GS, abarcando regiones extensas del genoma de algunas plantas anuales (como trigo y maíz), lo cual ha permitido aumentar el poder predictivo de características fenotípicas complejas (Calus et al., 2008; Matias et al., 2017, He et al., 2019; Sallam et al., 2020; Lan et al., 2020). Interesantemente, los estudios han demostrado que el enfoque de haplotipos podría ser especialmente beneficioso para predecir características con una relativa baja heredabilidad. Algunos autores sugieren que estos resultados pueden ser explicados por el hecho de que el uso de haplotipos en GS permite acceder a componentes genéticos heredables que no pueden ser examinados por los SNPs (Villumisen et al., 2009; Matias et al., 2017). A pesar de estos posibles beneficios demostrados in silico y experimentalmente, aún existe escasa información respecto al uso de haplotipos en GS, sobre todo en plantas alógamas, tales como *Eucalyptus* y otras especies de árboles.

1.2 Propuesta de Estudio

Teniendo en cuenta de que el conocimiento de las relaciones genéticas entre individuos y otros factores genéticos intrínsecos de una población determinan el grado de precisión que se pueda obtener al predecir una característica (basado en genómica y/o fenotipos), surge las siguientes preguntas de investigación: ¿La historia de selección y mejoramiento de una población influye en la estimación de parámetros genéticos mediante predicción genómica? Considerando que los haplotipos permiten establecer si dos individuos son idénticos por descendencia con un mayor grado de exactitud que los polimorfismos de nucleótido único, ¿Estas huellas moleculares podrían tener una mayor capacidad predictiva que polimorfismos de nucleótido único en características fenotípicas de especies alógamas como *Eucalyptus*? Por sus características biológicas, ¿Los haplotipos acceden a componentes heredables que no pueden ser estimados por la genealogía, o incluso por los polimorfismos de nucleótido único en características de bajo control genético en *Eucalyptus*? ¿Los haplotipos podrían estimar una mayor variabilidad genética (heredabilidad) para este tipo de características, que los antecedentes de pedigrí, o incluso que los polimorfismos de nucleótido único?

Para responder a estas preguntas, la presente propuesta tiene como objetivo determinar regiones genómicas en desequilibrio de ligamiento y su influencia en la predicción genómica de caracteres de herencia poligénica en dos especies de *Eucalyptus* (*E. globulus* y *E. cladocalyx*) que difieren en su historial de selección y en su estructura genética. En relación al historial de selección, la población de *E. globulus* es producto de al menos dos ciclos previos de selección artificial, basada en el volumen de los árboles, mientras que los árboles de *E. cladocalyx* provienen de poblaciones naturales distribuidas en el sur de Australia. En relación a la estructura genética, la población de *E. globulus* se conforma principalmente de árboles provenientes de polinización controlada (familias de hermanos completos y medios hermanos, mientras que la población de *E. cladocalyx* se constituye de árboles de los cuales sólo se conoce su linaje materno (familias de medios hermanos).

Este proyecto propone un enfoque integral de cómo abordar la predicción genómica en árboles forestales, particularmente en *Eucalyptus*, el cual permite obtener un mejor conocimiento para la gestión de los recursos naturales. En relación a la predicción genómica basada en haplotipos, este proyecto propone un enfoque de estudio genómico no abordado

previamente en especies forestales. La construcción de un mapa de haplotipos para cada especie de *Eucalyptus*, y su utilización en los modelos de predicción genómica, será un gran aporte a las ciencias biológicas, especialmente en rasgos complejos, de bajo control genético y con significativa influencia ambiental. Los resultados de este trabajo podrían contribuir a una gestión y producción más rentable de los recursos genéticos de las especies del género *Eucalyptus*.

1.3 Género *Eucalyptus* en el contexto mundial y nacional

El género *Eucalyptus* se compone de más de 700 especies, las cuales son nativas de Australia e islas cercanas (Tahir et al., 2016). Estas especies son plantadas en una extensa variedad de ambientes, tales como mediterráneos, tropicales, subtropicales y templados (Drake et al., 2015), debido a su rápido crecimiento, corta rotación y adaptabilidad. Tanto las especies del género como sus híbridos, se encuentran entre las principales fuentes de biomasa a nivel mundial y son las principales maderas duras utilizadas para la producción de pulpa y madera (Paiva et al., 2011). En el contexto nacional, las especies de *Eucalyptus* proporcionan la principal fuente de chips de madera de la industria chilena, lo cual representa un 54 % de la producción total (Morales et al., 2015). *E. globulus* es la especie del género mayormente plantada en el territorio nacional, abarcando el 20 % de las plantaciones (alrededor de 45000 hectáreas; Schmit et al., 2015). Adicionalmente, en la década de los 60s, varias especies del género fueron introducidas al país, entre las cuales se destaca *E. cladocalyx*, considerada un buen modelo de especies adaptadas a áreas con baja disponibilidad hídrica (Gleadow y Woodrow, 2002; Mora et al., 2009; Bush y Thumma, 2013; Bush et al., 2015; Arriagada et al., 2018). En adición, *E. cladocalyx* puede ser usada para fines melíferos, madereros (madera de alta durabilidad natural) y ornamentales (Bush et al., 2011; Cané-Retamales et al., 2011; Bush et al., 2015; Arriagada et al., 2018).

1.4 Mejoramiento genético forestal para la selección de árboles fenotípicamente superiores

En el mejoramiento genético forestal, usualmente se realizan evaluaciones fenotípicas y se determinan las relaciones de parentesco para identificar los árboles que son superiores para una característica de interés (Zapata-Valenzuela y Hasbún, 2011; Klápstě et al., 2017; Almeida-Filho et al., 2019). En general, las evaluaciones se realizan en ensayos de progenies, tales como medios hermanos y/o hermanos completos, o bien a través de silvicultura clonal (White et al., 2007). La selección fenotípica ha contribuido significativamente al incremento de las ganancias genéticas de *E. globulus* y *Pinus radiata* durante las últimas décadas en Chile. No obstante, la mantención de los ensayos de progenie involucra una importante inversión de recursos, logística y extensos ciclos de selección antes de generar resultados aprovechables (Zapata-Valenzuela y Hasbún, 2011).

Los genotipos parentales son evaluados en base al desempeño de su progenie, de tal forma que aquellas combinaciones dialélicas que generan una descendencia con un desempeño superior, son considerados como genéticamente superiores y son usados para avanzar en los sucesivos ciclos de mejoramiento. Se requiere, por lo tanto, de una estimación o predicción precisa del "valor genético" de un individuo (*breeding value*). Por otro lado, es frecuente que los resultados de un ensayo de competición forestal sean complicados de analizar debido a que: (1) usualmente sólo un subconjunto de los parentales está representado en cada ensayo de progenie; (2) los parentales están representados en diferente número de ensayos de progenie; (3) los ensayos son evaluados en diferentes edades de los árboles, y (4) los antecedentes del parentesco entre individuos pueden estar errados, debido al manejo de un huerto semillero, o inclusive debido a los procesos endogámicos (Klápstě et al., 2017). La histórica innovación a los modelos de estimación fue considerar a los efectos dados por los parentales como efectos aleatorios, en lugar de fijos (Henderson 1963; 1973; 1977; 1984). Este método analítico, conocido como la mejor predicción lineal (*Best Linear Prediction*; BLP), y ajustado a la mejor predicción lineal no sesgada (*Best Linear Unbiased Prediction;* BLUP), permite maximizar la precisión o la correlación entre los valores predichos y reales

de los méritos genéticos, el cual incorpora la información del pedigrí de los individuos evaluados (Viana et al., 2010).

1.5 El descubrimiento de polimorfismos de nucleótido único (SNPs) habilita diferentes campos de estudios en plantas

Los marcadores moleculares han sido ampliamente usados en estudios de genética y mejoramiento de plantas (Kumar et al., 2012; Hayward et al., 2015; Nadeem et al., 2018; Cobb et al., 2019). Los Polimorfismos de Nucleótido Único (*Single Nucleotide Polymorphisms*: SNPs) son actualmente los marcadores mayormente elegidos por los investigadores, debido a su amplia distribución en los genomas y prácticamente están presentes en cualquier población. Los marcadores SNP han sido aplicados a diversas áreas del conocimiento, como la ciencia forense y diagnóstico en humanos, acuicultura, selección asistida por marcadores en ganado, mejoramiento de cultivos agrícolas y estudios de conservación (Brenner y Weir, 2003; Seddon et al., 2005; Yu et al., 2011; Adhikari et al., 2017; Garrido-Cardenas et al., 2018).

En el año 2011, el número de genomas de plantas secuenciados se duplicó en comparación a la década anterior (http://phytozome.net), lo cual es gracias al creciente rendimiento en las metodologías de secuenciación. Las plataformas de secuenciación de segunda y tercera generación (NGS: *Next Generation Sequencing*), tales como Illumina, piro-secuenciación 454 (Roche), SOLID (Invitrogen) y Ion Torrent (Invitrogen) tienen la capacidad de obtener resultados a partir de una gran cantidad de secuencias que pueden ser usadas para descubrir nuevos marcadores moleculares en forma viable y a bajo costo (Harismendy et al., 2009; Paszkiewicz y Studholme, 2012). Estas técnicas han sido usadas a gran escala para el descubrimiento de SNPs en un representativo conjunto de individuos de varias especies, tales como arroz (Kharabian-Masouleh et al., 2011), trigo (Chandra et al., 2017), pino (Durán et al., 2019), y *Eucalyptus* (Silva-Junior et al., 2015). Debido al importante aporte que ha tenido el descubrimiento de marcadores SNP para áreas de estudio como la genómica, transcriptómica, genética de poblaciones, mapeo de loci de característica cuantitativa (QTL), entre otros, algunas empresas han optado por desarrollar herramientas de análisis que le permitan al investigador identificar la presencia o ausencia de polimorfismos

conocidos para una especie. A través del conocimiento del genoma y la localización física de estos polimorfismos, se han confeccionado arreglos de SNPs para diferentes especies tales como cacao (Livingstone et al., 2015), cerezo (Peace et al., 2012), trigo (Sun et al., 2020) y arroz (Singh et al., 2015). Ante la necesidad de obtener herramientas moleculares que apoyen los estudios genómicos de *Eucalyptus* spp., Silva-Junior et al. (2015) desarrollaron un arreglo de SNPs de alta densidad (EUChip60K), el cual es transferible en las doce taxas del género *Eucalyptus*, con un máximo de ~60000 SNPs informativos y polimórficos, y alrededor de 50000 SNPs para otras dos especies relacionadas. Este arreglo de SNPs de alta densidad es una nueva y prometedora plataforma de información para diferentes campos de estudios en *Eucalyptus spp.,* tales como diversidad y genética de poblaciones, selección genómica, estudios de asociación, entre otros.

1.6 Selección genómica, una extensión de los métodos de BLUP para maximizar el poder de predicción de características fenotípicas

Tras el surgimiento a gran escala de los marcadores moleculares, y el descenso en los costos asociados a estas herramientas, los mejoradores propusieron utilizar marcadores de ADN para apoyar en los ciclos de selección. En general, el mejoramiento asistido por genómica puede estar basado en Selección asistida por marcadores (MAS) o por la predicción o selección genómica (*Genomic Selection*; GS; Meuwissen et al., 2001; Crossa et al., 2010; Liu et al., 2016). La eficiencia de cada uno de los métodos varía de acuerdo con la arquitectura genética que subyace a la característica en estudio. La metodología MAS propone que los valores genotípicos de los individuos son estimados en base a los efectos de marcadores moleculares seleccionados, y presentan mayor efectividad en características fenotípicas que posee una arquitectura genética oligogénica (Liu et al., 2016). Mientras que GS es un método preferible cuando se estudian características complejas, las cuales son afectadas por un gran número de genes y son altamente influenciados por el ambiente. En el caso de *Eucalyptus*, varios loci de características cuantitativas (QTL) relacionados a diversas características de interés han sido mapeados y obtenidos en estudios de asociación, tales como crecimiento (Freeman et al., 2009; Thavamanikumar et al., 2014; Arriagada et al., 2018), componentes de floración (Missiaggia et al., 2005; Bundock et al., 2008; Arriagada et al., 2018), habilidad

pulpable (Thumma et al., 2010), propiedades de la madera (Thumma et al., 2005; Gion et al., 2011; Thavamanikumar et al., 2014; Valenzuela et al., 2019), entre otros. Sin embargo, en la práctica, la aplicación de MAS en características con herencia poligénica ha sido limitada. En este sentido, vale la pena mencionar que las características importantes para una plantación forestal, como el crecimiento, son controladas por un gran número de genes que aportan pequeños efectos a la variación fenotípica (Mora y Serra, 2014; Bartholomé et al., 2020).

GS es un método propuesto por Meuwissen et al. (2001), el cual apuntó a incrementar la eficiencia en los programas de mejoramiento de ganado lechero. GS nace como una alternativa de BLUP, incorporando datos genómicos. A diferencia de MAS, en GS se realiza una predicción de los efectos de miles de marcadores simultáneamente, a pesar de que éstos no sean significativos en forma individual para una característica. De acuerdo con Daetwyler et al. (2013), GS puede incrementar los rangos de ganancia genética, ya que los méritos genéticos individuales son estimados con mayor precisión. A pesar de que GS no permite identificar la función de los posibles genes controlando una característica, los modelos predictivos proporcionan un criterio de selección a corto plazo de aquellos individuos que poseen un mejor rendimiento. Más aun, GS ha servido para mejorar el entendimiento de la arquitectura genética de características fenotípicas, e incluso para implementar planes de restauración ecológica (Supple et al., 2018). Los métodos de GS mayormente conocidos son las estimaciones bayesianas: Bayes A, Bayes B, Bayes Cπ, Bayesian LASSO (*Least Absolute Shrinkage and Selection Operator* (Tibshirani, 1996; Meuwissen et al., 2001; Habier et al., 2011), Genomic-BLUP (GBLUP; VanRaden, 2008) y Regresión *Ridge Regression BLUP* (RR-BLUP; Meuwissen et al., 2001). En relación a los supuestos de análisis, RR-BLUP y GBLUP asumen que los marcadores poseen la misma varianza y cada marcador aporta un efecto pequeño al modelo de predicción (modelo infinitesimal). La predicción vía GBLUP es realizada similarmente a BLUP, con la diferencia de que la matriz de pedigrí de BLUP es remplazada por una matriz de parentesco construida a partir de marcadores moleculares. Por su parte, RR-BLUP es un método de regresión múltiple, en el cual los marcadores son miles de regresores que explican la variación de una característica fenotípica. En el contexto del método RR-BLUP, el valor genético (genómico) de cada individuo es definido por la siguiente fórmula (Resende et

al., 2012b): $GEBV_j = \sum_i^n Z_{ij}\hat{m}_i$, donde n corresponde al número total de marcadores, $\hat{m}_i$ es el efecto estimado para el i-ésimo marcador, y Z corresponde a la matriz de diseño asociada al vector de los efectos de los marcadores, la cual contiene codificado el genotipo del j-ésimo individuo, para el i-ésimo marcador. La sigla GEBV viene del inglés Genomic Estimated Breeding Value. Contrariamente, los métodos Bayes A, Bayes B y Bayes C asumen que cada marcador tiene su propia varianza, y la varianza fenotípica es explicada por loci con efectos de diferente magnitud (Wang et al., 2018). Estos métodos bayesianos se diferencian por las distribuciones a priori que son establecidas, y el grado de ajuste que se emplea. Una descripción más detallada de cada método puede ser encontrada en Heslot et al. (2012). Por ejemplo, el método Bayesian LASSO asume que los efectos de los marcadores se distribuyen a priori de acuerdo a una doble exponencial (DE): $p(m_i|\lambda, \sigma_e^2) = DE((m_i|0, \lambda/\sigma_e^2)$ donde λ corresponde a un parámetro de regularización. La distribución de DE genera una fuerte contracción (cercana a cero) para estimar los efectos de los marcadores. BRR es un método bayesiano que se basa en que los regresores del modelo (sean SNPs u otros marcadores) poseen una varianza común (σ_m^2), de tal manera, que aquellos regresores con una misma frecuencia alélica explican la misma proporción de la varianza aditiva, y posee un mismo efecto de contracción (Gianola 2013). El efecto de marcadores (m$_i$) se distribuye de la siguiente manera: $m_i|\sigma_m^2 \sim N(0, \sigma_m^2); \sigma_m^2|v_m, S_m \sim \chi^{-2}(v_m, S_m)$. En el modelo Bayes A, supone que cada marcador (m$_i$) sigue una distribución a priori normal independiente con media 0 y varianza $\sigma_{m_i}^2$, mientras que la varianza de cada uno de ellos se asume que se distribuyen $\sigma_{m_i}^2|v, S^2 \sim \chi^{-2}(v, S^2)$, con S^2 y v son un parámetros de escala y grados de libertad, respectivamente. Por su parte, el método Bayes-B usa una distribución mezclada con una masa hasta cero, tal que la distribución a priori de los efectos de los marcadores es dado por (Pérez-Rodríguez et al. 2012):

$$\beta_j|\sigma_j^2, \pi = \begin{cases} 0 \text{ con probabilidad } \pi \\ N(0, \sigma_j^2) \quad \text{con probabilidad } 1-\pi \end{cases}$$

El prior asignado para σ_j^2, j=1,…, p es igual para todos los marcadores, por ejemplo, una distribución chi-cuadrado invertida $\chi^{-2}(df_\beta, s_\beta,)$, con df_β como los grados de libertad, y s_β es un parámetro de escala. En el método Bayes Cπ, se considera todos los marcadores poseen una varianza común (σ_m^2) y promueve la selección de variables al igual que Bayes B. En todo estudio de selección genómica se recomienda probar varios métodos disponibles (Du et al., 2018), y

éstos deben ser contrarrestados en cuanto a sus valores de precisión o poder predictivo. No obstante, si el investigador posee una aproximación de cuántos loci podrían explicar la variación de una característica, podría utilizar un método en particular. Por ejemplo, el modelo Bayes B basa sus supuestos analíticos en características altamente heredables y cuya variación es explicada por loci de grandes efectos (Wang et al., 2018). Por su parte, Bayes A representa una opción para características que son controladas por una moderada cantidad de genes. Algunos estudios han demostrado que los métodos bayesianos suelen ser más precisos que GBLUP cuando las poblaciones de entrenamiento y validación están débilmente relacionadas desde el punto de vista genético (Gao et al., 2013; Wu et al., 2015).

En el contexto del mejoramiento forestal, GS fue originalmente propuesta para el análisis de características complejas como el crecimiento de los árboles y características relacionadas a la madera. El principal parámetro que revela si un modelo GS es adecuado para la estimación de los méritos genéticos es la precisión y el poder predictivo, el cual expresa el grado de correlación que existe entre los valores genéticos predichos por el modelo GS y los valores fenotípicos ajustados (o bien los valores genéticos ajustados) (Resende et al., 2012a; Weber et al., 2012; Pryce et al., 2014). En el estudio de Beaulieu et al. (2014), se obtuvo una precisión de hasta 0,435 al generar modelos GS para predecir características relacionadas a propiedades de la madera en abeto. En el caso de *Pinus taeda*, Resende et al. (2012b) obtuvieron un poder predictivo de 0,37 a 0,77 para características de diferente naturaleza. Debido a la importancia económica que tienen las *Eucalyptus* spp. a nivel mundial, diferentes especies del género han sido estudiadas con los principios de la predicción genómica (Resende et al., 2012a; Denis et al., 2013; Durán et al., 2017; Tan et al., 2017; Müller et al., 2017; Resende et al., 2017; Suontama et al., 2019). En estos estudios, se utilizaron marcadores SNP y Diversity Arrays (DArts) para predecir características relacionadas al crecimiento, calidad de madera y habilidad pulpable. En el contexto de características relacionadas con la química de la madera, el poder predictivo de los modelos implementados alcanzó valores por sobre 0,7, indicando que este tipo de caracteres pueden ser predichos con alta precisión por herramientas genómicas. Desafortunada, para algunas características relacionadas al crecimiento de los árboles, los modelos implementados tuvieron un poder de predicción moderado (menor que 0,6), a excepción del incremento medio anual (Resende et al., 2012; Tan et al., 2017; Müller et al., 2017; Resende et al., 2017; Cappa et

al., 2019). En este contexto, los métodos de predicción genómica han sido limitados para predecir estos componentes en *Eucalyptus* debido a su naturaleza compleja (Müller et al., 2017; Resende et al., 2017). Por otro lado, las herramientas genómicas han permitido acceder a componentes heredables que no pueden ser examinados mediante las relaciones genealógicas entre individuos (Müller et al., 2017; Klápstein et al., 2017; Suontama et al., 2019). Por ejemplo, Müller et al. (2017) determinaron que la heredabilidad del diámetro a la altura de pecho en una población de *E. pellita* fue inestimable (extremadamente baja) mediante el método de pedigrí, no obstante, la heredabilidad genómica basada en SNPs alcanzó un valor de 0,55 (método Bayes B).

1.7 Factores que determinan la precisión de los modelos de predicción genómica

Anteriormente, se discutió como influye la precisión de los métodos de predicción genómica (Por ejemplo, Bayes A, Bayes B, entre otros) de acuerdo a los supuestos analíticos que sustentan cada método. No obstante, factores genéticos que son propios de las poblaciones determinan la efectividad de la selección asistida por genómica. El patrón de desequilibrio de ligamiento (DL) dentro de una población es uno de los principales factores determinando el poder de las herramientas genómicas para predecir características fenotípicas, debido a que la extensión del DL a lo largo del genoma de una especie determina la densidad de marcadores que es necesitada para una predicción precisa (Sun et al., 2016; Schopp et al., 2017). En estricto, si el DL se extiende dentro de distancias genómicas relativamente grandes, una menor densidad de marcadores podría ser necesitada, debido a que aumenta la probabilidad de detectar marcadores en DL con loci de característica cuantitativa (QTLs). Por otro lado, si el DL disminuye dentro de una distancia genómica relativamente corta, una mayor densidad de marcadores tendría que ser empleada para obtener una predicción mayormente precisa.

La magnitud del DL en una población procede de la historia y la dinámica de la población (Brown et al., 2012). En general, el DL en poblaciones de plantas de polinización cruzada (por ejemplo, especies forestales) disminuye rápidamente conforme aumenta la distancia física entre marcadores, debido a que el número de recombinaciones efectivas es relativamente más alto que en especies de autopolinización. Por ejemplo, el DL deja de ser significativo entre los 100-150 pb en poblaciones de *Pinus taeda*, y a los 500 pb en vid; mientras

que, en especies autógamas, tales como soya y arroz, DL puede ser significativo (en algunas poblaciones) a una distancia entre loci de 100000 pb y 250000 pb, respectivamente (revisado por Gupta et al., 2005). Adicionalmente, las poblaciones de plantas alógamas como árboles tienden a ser más heterocigotas en sus loci (Fiil et al., 2011; Rodríguez et al., 2012), debido a los mecanismos reproductivos que subyacen a estas especies y poseen gran tamaño efectivo, alta diversidad genética y una baja diferenciación genética intra-poblacional (Neale y Kremer, 2011).

Las poblaciones naturales de especies forestales presentan valores de DL relativamente más bajos que poblaciones cultivadas y mejoradas (Olson et al., 2010; Thavamanikumar et al., 2011; Kelleher et al., 2012; Guerra et al., 2012; Larsson et al., 2013; Lu et al., 2016; Müller et al., 2017; Durán et al., 2017). Por ejemplo, el desequilibrio de ligamiento en una población natural de *Populus* puede disminuir dentro de los 750-1000 pb (Olson et al., 2010; Kelleher et al., 2012), no obstante, en una población mejorada puede extenderse hasta los 2500 pb (Guerra et al., 2012). En el caso de poblaciones de coníferas, el desequilibrio puede llegar a disminuir dentro de distancias genómicas muchos más cortas que para otras especies de árboles (<1000 pb) (Larsson et al., 2013; Thistlethwaite et al., 2020). En el caso de poblaciones de mejoramiento de *Eucalyptus*, el DL puede disminuir rápidamente sobre los 3000 pb y los 25000 pb (Müller et al., 2017; Durán et al., 2017; Suontama et al., 2019), mientras que en poblaciones que no han sido sometidas a selección, se ha encontrado que el patrón de desequilibrio disminuye dentro de los 500 pb (Thavamanikumar et al., 2011).

Otro de los factores que afectan la precisión de GS es el tamaño efectivo poblacional (Lee et al., 2017; Stejskal et al., 2018). De acuerdo con Grattapaglia y Resende (2011), el tamaño efectivo es inversamente proporcional a nivel de precisión de los modelos GS. Por ejemplo, si se desea predecir un fenotipo controlado por 50 QTLs, con un tamaño efectivo de 100 individuos y una densidad de 2 marcadores/cM, se estima un valor predictivo de 0,36, el cual puede ser incrementado a 0,8 si se tiene una densidad de 20 marcadores/cM. No obstante, con un tamaño efectivo de 10 individuos, bajo estas mismas condiciones, el poder predictivo puede ser incrementado de 0,73 a 0,88, lo cual demuestra que tamaños efectivos pequeños permiten incrementar el poder predictivo de un modelo, especialmente con una baja densidad de marcadores. En este mismo contexto, tanto en los métodos GS, como la predicción basada en

23

pedigrí, la precisión predictiva del fenotipo de un individuo depende de qué tan precisa es la estimación de las relaciones genéticas entre los individuos de la población (Goddard et al., 2011; Scutari et al., 2016). El poder predictivo de un modelo GS puede ser incrementado si los individuos que se utilizan como referencia para estimar los efectos aditivos de los loci, están genéticamente relacionados a los individuos que se desean predecir fenotípicamente (Ly et al., 2013; Lorenz y Smith, 2015; Lee et al., 2017; Thistlethwaite et al. 2020). En este mismo contexto, se ha demostrado que la estructura genética de una población es un factor que influye en el grado de precisión que pueda tener un modelo de predicción genómica (Lorenz et al., 2012; Thorwarth et al., 2017; Tan et al., 2017; Klápstein et al., 2017; Sapkota et al., 2020). La estructura genética de una población es definida por el grado de parentesco que existe entre los individuos que componen a la población, el cual puede ser establecido por los antecedentes genealógicos, o bien, mediante datos genómicos. En el caso de individuos que provienen de una población natural, la estructura genética puede estar dada por los antecedentes genealógicos y por la existencia de sub-poblaciones (grupos genéticamente diferenciados) dentro de una misma población. Por ejemplo, Thorwarth et al. (2017) demostraron que el poder de 9000 marcadores SNPs para predecir cinco características fenotípicas en cebada puede ser sobrestimado si la estructura de población no es considerada en el modelo de predicción. Lorenz et al. (2012) reportaron que el poder predictivo para la resistencia a *Fusarium* en cebada podría ser disminuido si los modelos de predicción genómica son entrenados con individuos que pertenecen a colecciones de germoplasmas genéticamente muy distantes. Esto implica que el poder de predicción fue incrementado cuando los individuos de entrenamiento y validación del modelo pertenecieron a un mismo grupo genéticamente diferenciado.

La determinación de las relaciones de parentesco entre individuos u organismos, ha sido un aspecto importante en varios campos del conocimiento, tales como ciencia forense, genética de conservación, y mejoramiento genético animal y vegetal (Edwards et al., 2015). Las estimaciones del parentesco entre individuos se basan tradicionalmente en datos de pedigrí, en los cuales se asume que aquellos individuos que actúan como fundadores de poblaciones (por ejemplo, familias) no se encuentran genéticamente relacionados (Gienapp et al., 2017). Si los individuos de una población son genotipados, los loci polimórficos que tengan en común pueden

ser idénticos por estado (*Identical By State;* IBS) o bien idénticos por descendencia (*Identical By Descent;* IBD), dependiendo si estos loci fueron heredados desde un mismo ancestro. Se espera que individuos que comparten las mismas líneas parentales (hermanos completos) posean en promedio un 50% de alelos que son IBD. No obstante, el porcentaje del genoma que comparten está sometido a variación debido a sucesos aleatorios que pueden presentarse. En este sentido, las relaciones genéticas basadas en pedigrí son arbitrarias y teóricas, por lo tanto, no son necesariamente un reflejo de la manera real en que los genomas son heredados. A pesar de este inconveniente, los datos de pedigrí han sido ampliamente usados para establecer las relaciones entre individuos de una población en los estudios de genética cuantitativa (Gienapp et al., 2017). En base a esta problemática, VanRaden (2008) propuso realizar predicciones vía BLUP utilizando marcadores moleculares (SNPs) para determinar las relaciones de parentesco dentro de un grupo de individuos, en lugar de la matriz de pedigrí (lo cual posteriormente fue conocido como Genomic BLUP; GBLUP). Adicionalmente, Yu et al. (2006) propuso utilizar matrices de parentesco basadas en marcadores como covariables en los modelos de asociación genética. En el contexto de GS, el método GBLUP está basado en el uso de matrices de parentesco construidas a partir de marcadores. Cabe destacar que los marcadores bialélicos SNPs no contienen suficiente información a nivel individual, por lo tanto, algunos autores sugieren que se requiere una gran densidad de marcadores para poder establecer relaciones genealógicas confiables (Yamamoto et al., 2010; Voorrips et al., 2016). Adicionalmente, en poblaciones con múltiples fundadores, generaciones, y pocos descendientes, los errores en el genotipado son difíciles de detectar (Pikunova et al., 2014; Di Guardo et al., 2015). Una alternativa para los análisis de parentesco es la identificación de haplotipos (Mathew et al., 2018; Edwards ,2015; Howard et al., 2017). Cuando dos o más loci tiene una baja probabilidad de recombinación entre ellos, se forman combinaciones de alelos llamados haplotipos (Nordborg y Tavaré, 2002; Machiela et al., 2015), los cuales son regiones genómicas dentro de un cromosoma que tienden a heredarse en forma conjunta entre generaciones (Andersen y Lübberstedt, 2003; Hodgkinson y Pullman, 2010; Sun et al., 2015). En este contexto, Edwards (2015) propuso evaluar el parentesco entre individuos basándose en la construcción de haplotipos presentes en la población de estudio. De acuerdo con este mismo autor, estas relaciones de parentesco pueden estimar relaciones genealógicas más precisas que la matriz de parentesco de VanRaden (2008). Cabe destacar que si se desea conocer la coancestría intra-

poblacional, lo más recomendable es utilizar haplotipos que sean IBD. no obstante, los haplotipos pueden establecer relaciones tanto de IBD como de IBS.

1.8 La detección de alelos en estrecho desequilibrio de ligamiento (DL) permite optimizar la precisión los modelos de predicción genómica

Existen dos principales modalidades de cómo efectuar los análisis de GS. En el modelo más simple, se asume que cada alelo de un locus tiene un efecto en la característica de interés, de tal manera que una regresión simple puede ser usada para predecir los méritos genéticos. El otro enfoque involucra asociar un conjunto de dos o más alelos a los valores fenotípicos, de tal manera que la sumatoria de estos conjuntos permite obtener la predicción de los méritos genéticos. Diversos estudios han evaluado el poder predictivo de características fenotípicas en modelos de GS que incluyen haplotipos provenientes de arreglos de SNPs (Villumsen et al., 2009; De Roos et al., 2011; Boichard et al., 2012; Edriss et al., 2013; Cuyabano et al., 2014; Jónás et al., 2017). Una de las ventajas de utilizar haplotipos en GS es la capacidad para detectar mutaciones (Cuyabano et al., 2014). De acuerdo con Curtis et al. (2001), cuando se han producido mutaciones, es posible que las frecuencias de los alelos permanezcan (casi) inalteradas. Sin embargo, cuando se analizan los haplotipos, las mutaciones en diferentes loci tienden a provocar cambios importantes en las frecuencias de haplotipos. Por lo tanto, un QTL que no se encuentra en completo DL con un marcador individual, puede estar en completo DL con un haplotipo en específico. Más aun, los enfoques basados en haplotipos pueden incluir los efectos epistáticos, además de los efectos aditivos (Jiang et al., 2018), lo cual es beneficioso en mejoramiento en plantas. Adicionalmente, el uso de haplotipos, en lugar de marcadores individuales, reduce los grados de libertad en los modelos de predicción o de asociación genómica, lo cual contribuye a una mayor precisión en la detección de QTL (Yu et al. 2006). Un aspecto también relevante es el tamaño de los haplotipos encontrados en una determinada población. Mientras más largos sean los haplotipos (mayor número de SNPs en DL), el número de efectos que deben estimarse será menor, lo cual conduce a estimaciones más precisas (Hickey et al., 2014) y son más sencillos de implementar computacionalmente.

Cabe destacar que el poder predictivo de los haplotipos y de los marcadores (en forma individual), podría ser dependiente de la característica que se busca predecir. En el contexto del mejoramiento genético animal, algunos autores sugieren que el enfoque de haplotipos puede ser especialmente beneficioso para predecir características de una heredabilidad relativamente alta (Calus et al., 2008; Cuyabano et al., 2014; Ødegård et al., 2014), no obstante, en el contexto de plantas, la predicción basada en haplotipos ha sido especialmente beneficiosa para predecir características de baja heredabilidad (Matias et al., 2017: He et al., 2019; Lan et al., 2020; Sallam et al., 2020). Por ejemplo, Matias et al., (2017) reportaron que la precisión predictiva de los modelos basados en haplotipos fue superior la precisión de los SNPs (no agrupados en haplotipos) para predecir el rendimiento de granos de maíz, no obstante, este resultado no fue observado en la predicción genómica de la altura de planta. De acuerdo con estos mismos autores, el rendimiento de granos de maíz tiene un bajo control genético comparado con la altura de planta. En otro estudio, Lan et al. (2020) desarrollaron modelos de predicción genómica para diferentes características en lino, los cuales se basaron en la identificación de SNPs y haplotipos asociados a QTLs. Los resultados de este estudio revelaron que el modelo basado en haplotipos tuvo un poder predictivo superior a los modelos basados en SNPs, para predecir el rendimiento de granos, el contenido de proteínas y número de días de maduración de granos, mientras que los modelos que incluyeron SNPs asociados a QTLs tuvieron mayor precisión predictiva para el contenido de diferentes tipos de ácidos grasos de semillas. Interesantemente, el rendimiento de granos, el contenido de proteínas y el número de días de maduración son características que tienen una menor heredabilidad en sentido amplio (H^2= 0,2-0,44), comparado con el contenido de los diferentes ácidos grasos en semillas (H^2= 0,69-0,81).

Villumisen y Janss et al. (2009) demostraron mediante datos simulados que las predicciones genómicas basadas en haplotipos son especialmente beneficiosas para características de bajo control genético (heredabilidad relativamente baja), en el cual haplotipos conformados por 5 SNPs en DL tienen una mejor bondad de ajuste y poder predictivo que los modelos basados en marcadores no agrupados en haplotipos. De acuerdo con Villumisen et al. (2009), el uso de haplotipos permite una evaluación a nivel multi-alélico, lo cual conduce a una mejor representación de la variabilidad asociada a las características de baja heredabilidad, las cuales generalmente están siendo controladas por varios QTLs de un efecto relativamente

pequeño. En este contexto, los haplotipos permitirían acceder a componentes heredables de ciertas características fenotípicas que no puede ser capturados por los SNPs. En humanos, el enfoque de haplotipos ha permitido capturar una mayor proporción de la variabilidad genética de la esquizofrenia, en comparación con la predicción genómica basada en SNPs (Bhatia et al., 2015). De acuerdo Bhatia et al (2015), la esquizofrenia es altamente heredable (h^2=0,7-0,8) y una gran cantidad de estudios han reportado loci significativamente asociados con esta enfermedad, no obstante, la heredabilidad genómica de estos loci ha sido sustancialmente más baja que la heredabilidad esperada para esta enfermedad (h^2=0,03). En este contexto, los autores desarrollaron modelos de predicción genómica para estimar la heredabilidad de la esquizofrenia basada en SNPs y haplotipos, utilizando datos experimentales y simulados. De acuerdo a sus resultados, la heredabilidad genómica estimada por los haplotipos fue dos veces superior a la estimación basada en SNPs. Cabe destacar que la efectividad del enfoque de haplotipos en GS depende de cómo son definidos los haplotipos en la población de estudio (Cuyabano et al., 2014; Matias et al., 2017). Por ejemplo, un haplotipo puede ser definido de acuerdo a un número determinado de loci polimórficos, un tamaño definido (en términos de distancia genómica), o de acuerdo a un valor umbral de DL. Es común que los haplotipos sean definidos por un número determinado de SNPs (Calus et al., 2008; Villumisen et al., 2009), no obstante, este enfoque no considera el DL entre los marcadores y los eventos de recombinación históricos de la población, por lo tanto, algunos antecedentes genéticos podrían no ser considerados para cuantificar la variabilidad entre los individuos (Cuyabano et al., 2014).

Algunos estudios sugieren que GS basada netamente en marcadores moleculares puede conducir a una pérdida de variabilidad genética, lo cual conlleva a un aumento de la tasa de endogamia (Rutkoski et al., 2015; Lin et al., 2016, 2017; Eynard et al., 2018; Doublet et al., 2019). Rutkoski et al. (2015) demostraron que las ganancias genéticas para la resistencia a la roya en trigo que se obtienen por GS y selección fenotípica son equivalentes, no obstante, GS genera una reducción más rápida de la diversidad genética (por año) que la selección fenotípica. En un estudio simulado, Lin et al. (2016) reportaron que GS permitiría duplicar y triplicar las ganancias genéticas de la persistencia y del rendimiento de *Lolium perenne*, respecto a la selección fenotípica. Sin embargo, GS condujo a una mayor tasa de endogamia por ciclo de selección que la selección fenotípica. Varios estudios han propuesto diferentes

estrategias para establecer un balance entre las ganancias genéticas y la diversidad conservada post-selección (Sonesson et al., 2012; Daetwyler et al., 2015; Müller et al., 2018). De acuerdo con Daetwyler et al. (2015), la GS basada en haplotipos en líneas doble haploides de trigo permitió conservar un mayor grado de diversidad que la GS basada en SNPs. Estos autores propusieron que GS basada sólo en SNPs podría generar la pérdida de ciertos alelos deletéreos (o bien, que aparentemente no tienen efecto en el fenotipo) en la población, no obstante, los haplotipos permiten gestionar una selección basada en alelos que tienen o no efecto, pero permanecen en DL. En este sentido, GS basada en haplotipos podría además contribuir a una buena gestión de los recursos genéticos, de tal manera que se puedan obtener ganancias genéticas, sin sacrificar la variabilidad genética.

Particularmente, la GS basada en haplotipos ha sido principalmente implementada en cultivos agrícolas y en plantas autógamas (por ejemplo, trigo; He et al., 2019; Sallam et al., 2020), en donde extensos valores de DL pueden ser encontrados en sus genomas, lo cual favorece la identificación de haplotipos en una población. Mientras que, en plantas alógamas, tales como especies forestales, el DL suele disminuir a cortas distancias genómicas, lo cual permite identificar haplotipos de menor tamaño, y éstos se componen de una menor cantidad de alelos. No obstante, los métodos de predicción basados en haplotipos no han sido explorados en el contexto de las especies forestales.

1.9 HIPOTESIS DE ESTUDIO

Considerando que: i) La industria forestal puede ser potenciada por el uso de herramientas biotecnológicas para asistir los programas de mejoramiento genético, ii) Los avances recientes en tecnologías de secuenciación han generado información genómica de alto rendimiento, lo cual ha facilitado la implementación de Predicción o Selección Genómica (GS), iii) GS es uno de los métodos biotecnológicos más recientes en la era genómica, el cual que busca incrementar la eficiencia de los programas de mejoramiento y las ganancias genéticas, iv) No obstante, el poder de los modelos GS para predecir características poligénicas, tales como caracteres asociados al crecimiento, ha sido relativamente moderado o bajo en especies forestales económicamente relevantes como *Eucalyptus*, v) En GS, la precisión depende de los supuestos analíticos de los modelos de predicción y de características intrínsecas de las poblaciones, tales como la estructura genética, el patrón de desequilibrio de ligamiento, entre otros, vi) En el contexto de GS de plantas, loci en desequilibrio de ligamiento (haplotipos) pueden estimar variabilidad genética que no puede ser estimada por los polimorfismos de nucleótido único y por los antecedentes genealógicos, vii) En GS, el enfoque de haplotipos ha tenido una mayor precisión predictiva que los polimorfismos de nucleótido único, para predecir características de baja heredabilidad en plantas, y finalmente viii) El enfoque haplotipos en GS no ha sido investigado en especies forestales o árboles; se propone las siguiente hipótesis de trabajo:

Hipótesis: Loci en desequilibrio de ligamiento, que tienden a heredarse en forma conjunta entre generaciones, explican una mayor variación genética que los polimorfismos de nucleótido único y la genealogía poblacional, además de tener una mayor precisión predictiva en características de bajo control genético en *Eucalyptus*, lo cual depende del historial de selección y la estructura genética de los individuos.

1.10 OBJETIVOS

1.10.1 Objetivo General

Determinar regiones genómicas en desequilibrio de ligamiento y su influencia en la predicción genómica de caracteres poligénicos en dos especies de *Eucalyptus* (*E. globulus* y *E. cladocalyx*) que difieren en su historial de selección y estructura genética.

1.10.2 Objetivos Específicos

1.- Evaluar la representatividad de un arreglo de SNPs de alta densidad (60K), en dos especies de *Eucalyptus* plantadas en ensayos de progenies de medios hermanos (*E. cladocalyx*), y en un ensayo combinado de hermanos completos y medios hermanos (*E. globulus*).

2.- Determinar y caracterizar regiones genómicas en desequilibrio de ligamiento en los dos ensayos de progenie de *Eucalyptus* que difieren en su estructura genética e historial de selección.

3.- Determinar la heredabilidad genómica de características fenotípicas en ambos ensayos de *Eucalyptus*, basado en haplotipos y polimorfismos de nucleótido único.

4.- Evaluar modelos de predicción genómica basados en regiones en desequilibrio de ligamiento para predecir características fenotípicas de árboles de *Eucalyptus*.

2. CAPÍTULO I: SELECCIÓN GÉNOMICA BASADA EN HAPLOTIPOS Y SNPS, DE DOS POBLACIONES DE *EUCALYPTUS* QUE DIFIEREN EN SU ESTRUCTURA GENÉTICA E HITORIAL DE SELECCIÓN

2.1 INTRODUCCION

Los árboles del género *Eucalyptus* L'Hér son reconocidos por su alta producción de biomasa, rápido crecimiento, gran adaptación a diferentes condiciones ambientales y excelente calidad de madera para la producción de papel y productos derivados (Mora y Serra, 2014; Schmit et al., 2015; Arriagada et al., 2018). Particularmente, *Eucalyptus globulus* es la segunda especie leñosa más importante en la industria forestal chilena, la cual representa el 21% de las plantaciones nacionales (Morales et al., 2015). Debido a su importancia económica, *E. globulus* ha sido objetivo en varios programas de mejoramiento genético y manejo silvicultural (Carocha et al., 2015; Carbonari et al., 2016). Por otro lado, en razón de diversificar los recursos forestales, varias especies del género fueron introducidas en Chile en la década de los 60s, entre las cuales se destaca *E. cladocalyx*. Esta última ha sido considerada como un buen modelo de especies adaptadas a áreas con baja disponibilidad hídrica (Gleadow y Woodrow, 2002; Mora et al., 2009; Bush y Thumma, 2013; Bush et al., 2015; Arriagada et al., 2018). En adición, se ha planteado que *E. cladocalyx* puede proporcionar varios tipos de recursos, tales como melíferos, madereros (madera de alta durabilidad natural) y ornamentales (Bush et al., 2011; Cané-Retamales et al., 2011; Bush et al., 2015; Arriagada et al., 2018).

Tradicionalmente, el periodo de rotación de *Eucalyptus* varía entre los ocho y 12 años, no obstante, el manejo forestal y el mejoramiento genético ha permitido optimizar la producción, reduciendo el periodo de rotación hasta cinco años (Morales et al., 2015). En el contexto actual, el número de programas de mejoramiento que utilizan los principios de selección genómica (GS) ha aumentado considerablemente en los últimos años. Durante la última década, varias investigaciones han ilustrado cómo incorporar los principios de GS en los programas de mejora genética de animales y plantas (Gianola, 2006; VanRaden, 2008; Pérez et al., 2010; Habier et al., 2011; Resende et al., 2012a, 2012b; Gianola, 2013; Azevedo et al., 2015). Los polimorfismos de nucleótido único (SNPs) han sido una herramienta poderosa en los programas

de mejoramiento para diferentes cultivos agrícolas (Contreras-Soto et al., 2017; Rasheed et al., 2017; Battenfield et al., 2018; Li et al., 2018; Maldonado et al., 2019). Los polimorfismos de nucleótido único (SNPs) tienen múltiples aplicaciones en plantas, incluyendo los estudios de asociación de genoma amplio, el mapeo de loci de características cuantitativas (QTL) y la determinación del parentesco genético entre individuos. No obstante, el poder de los modelos GS, basados en SNPs o en Diversity arrays), para predecir características poligénicas, tales como caracteres asociados al crecimiento, ha sido relativamente moderado y bajo (<0,6) en especies económicamente relevantes como *Eucalyptus* spp (Resende et al., 2012a; Müller et al., 2017; Tan et al., 2017; Suontama et al., 2019).

La precisión de los modelos GS depende de los supuestos analíticos de los modelos de predicción, la densidad de marcadores, y características intrínsecas de las poblaciones, lo cual incluye el tamaño efectivo de población, los patrones de desequilibrio de ligamiento (DL), el historial de selección, entre otros (Schopp et al., 2017; Klápště et al. 2017; Stejskal et al., 2018). Por ejemplo, se ha demostrado que la estimación de parámetros genéticos es más precisa cuando los individuos que se utilizan como referencia para estimar los efectos aditivos de los loci, están genéticamente relacionados a los individuos que se desean predecir fenotípicamente (Grattapaglia y Resende, 2011; Ly et al., 2013; Lorenz y Smith, 2015; Lee et al., 2017; Thistlethwaite et al. 2020). En este sentido, la precisión predictiva del fenotipo de un individuo depende de que tan precisa es la estimación de las relaciones genéticas entre los individuos (Goddard et al., 2011; Scutari et al., 2016).

GS puede implementarse utilizando SNPs u otros marcadores, o bien mediante marcadores que se encuentran en desequilibrio de ligamiento (haplotipos). Matias et al. (2017) reportaron que el enfoque de haplotipos es un 20% más poderoso que el enfoque de SNP para predecir características de baja heredabilidad en maíz. Además, algunos estudios sugieren que el enfoque de haplotipos podría ser especialmente adecuado para predecir caracteres con una relativa baja heredabilidad, debido a que permiten examinar componentes heredables que no pueden ser estimados por los antecedentes genealógicos, e incluso los SNPs (Villumisen et al., 2009; Matias et al., 2017; Bhatia et al., 2015). Existen escasos estudios que hayan evaluado el enfoque de haplotipos en GS, plantas (Matias et al., 2017; He et al., 2019; Lan et al., 2020;

Sallam et al., 2020). Más aun, este enfoque de predicción no ha sido puesto a prueba en árboles y plantas de interés forestal.

El objetivo del presente estudio fue determinar regiones genómicas en desequilibrio de ligamiento y su influencia en la predicción genómica de caracteres poligénicos en dos especies de *Eucalyptus* (*E. globulus* y *E. cladocalyx*), que difieren en su historial de selección y estructura genética. En relación al historial de selección, la población en estudio de *E. globulus* es producto de al menos dos ciclos de selección artificial, basada en el volumen de los árboles, mientras que los árboles que constituyen a la población de *E. cladocalyx* provienen de poblaciones naturales distribuidas en el sur de Australia. En relación a la estructura genética, la población de *E. globulus* se conforma principalmente de árboles de familias de hermanos completos y medios hermanos, mientras que la población de *E. cladocalyx* se constituye de árboles de los cuales sólo se conoce su linaje materno (familias de medios hermanos).

2.2 MATERIALES Y MÉTODOS

2.2.1 Características y condiciones de sitio del ensayo *E. globulus*

El estudio fue conducido en un ensayo de progenie, establecido en el año 2012, compuesto por 62 familias de hermanos completos y tres familias de medios hermanos de *Eucalyptus globulus* (~1860 árboles), localizado en sector La Poza, provincia de Purranque, Región de Los Lagos, Chile. Los árboles provienen de un programa de mejoramiento avanzado, en el cual se han realizado al menos dos ciclos de selección basado en el volumen de los árboles (ANEXO I, Figura S1). La población base de este programa es desconocida. La estructura genética de este ensayo es representada por árboles provenientes de cruzamientos controlados entre 15 y 21 árboles actuando como líneas maternas y paternas, respectivamente. Como resultado, se obtuvieron las 62 familias de hermanos completos. Estos árboles provienen de un huerto semillero de árboles plus (Las Violetas). Adicionalmente, desde este mismo huerto se tomaron semillas de dos árboles plus y se plantaron en el ensayo (dos familias de medios hermanos), y semillas de un árbol plus desde un huerto semillero independiente (Los Boldos). Las condiciones locales de La Poza se encuentran detalladas en la Tabla 1. Los árboles fueron distribuidos de acuerdo a un diseño de bloques completos al azar (30 bloques), considerando un

árbol por parcela (diseño *single tree plot*), y una densidad de plantación de 2.5 m entre cada árbol dentro de cada bloque.

Tabla 1. Condiciones de sitio de La Poza, Purranque, Provincia de Osorno, Región de los Lagos, Chile.

Condiciones de sitio	Mediciones
Coordenadas	40°57' S, 73°30' O
Tipo de clima	Oceánico o marítimo
Temperatura anual	13°C
Temperatura media en meses fríos	6°C
Temperatura media en meses cálidos	16°C
Precipitación anual acumulada	1282 mm
Altitud	326 m

2.2.2 Mediciones fenotípicas evaluadas en *E. globulus*

Los árboles se evaluaron a los seis años de edad para las siguientes características: diámetro a la altura de pecho (DAP), altura total del árbol (ALT), rectitud del fuste (RF), densidad de madera (DM) y calidad de ramas (CR). La DM fue medida indirectamente usando un Pilodyn 6J Forest (PROCEQ, Zurich, Switzerland), con dos repeticiones (± 2 mm). RF fue evaluada en los primeros 2/3 de la altura total del árbol de acuerdo a una escala ordinal (siete niveles), donde el valor 0 corresponde a árboles que poseen una curvatura en el primer tercio de la altura total del árbol, y 6, en el caso de árboles que podían presentar una curvatura leve en el tercio superior de la altura del árbol, sin afectar la productividad. CR fue evaluado de acuerdo a diferentes criterios que definen la calidad (diámetro, ángulo y distribución de ramas en el árbol), mediante una escala ordinal de seis niveles, en la cual un valor de 1 es asignado a árboles con una extrema deficiencia en el diámetro de ramas y cualquier otra variable, y un valor de 6 corresponde a árboles que poseen una óptima combinación de todas las variables de calidad, sin generar pérdida de productividad.

2.2.3 Características y condiciones de sitio del ensayo *E. cladocalyx*

La población de estudio consistió de un ensayo de progenie de 49 familias de medios hermanos de *Eucalyptus cladocalyx* F. Muell, establecido en la localidad de Los Vilos, Provincia de Choapa, Región de Coquimbo, Chile (Tabla 2), en el año 2001 (Ensayo Hacienda Las Caracas). Los árboles provienen de semillas recolectadas de árboles nativos del sur de Australia. Cuarenta y siete familias provienen de diferentes puntos de la distribución natural de la especie (ANEXO I, Figura S2; Tabla S1): 16 familias de Mount Remarkable (Flinders Ranges), 10 familias de Cowell (Península Eyre), 4 familias de Marble Range (Península Eyre), 9 familias de Wirrabara (Flinders Ranges), y 8 familias de Flinders Chase (Isla Kangoroo). Las dos familias restantes provienen de una colección local en la Provincia de Choapa en Chile. Al igual que la población de *E. globulus*, la estructura genética está definida por el parentesco que existe entre los árboles. No obstante, se ha documentado que esta población de *E. cladocalyx* se divide en subpoblaciones debido al grado de diferenciación genética que ocurre entre las regiones de procedencia (Mora et al., 2017; Arriagada et al., 2018). Por lo tanto, el efecto de esta estructura genética también fue considerado en los modelos de predicción (Ver punto 2.2.7). Los árboles fueron organizados de acuerdo a un diseño de bloques completos al azar (30 bloques). Se estableció un representante de cada familia en cada bloque (Diseño de árbol único por parcela), resultando en un total de 1470 árboles en el ensayo. Los árboles fueron plantados con densidad de ~1667 árboles por ha^{-1}.

Tabla 2. Condiciones de sitio de ensayo Hacienda Caracas, Los Vilos, Provincia de Choapa, Región de Coquimbo, Chile.

Condiciones de sitio	Mediciones
Coordenadas	31°54' S; 71°27'O
Tipo de clima	Estepárico costero y predominantemente árido*
Temperatura anual	16.9 °C
Temperatura media en meses fríos	13 °C
Temperatura media en meses cálidos	18 °C
Precipitación anual acumulada	211 mm
Altitud	167 m

*De acuerdo a Índice de aridez de Martonne (Arriagada et al., 2018).

2.2.4 Mediciones fenotípicas evaluadas en *E. cladocalyx*

Los árboles fueron evaluados a los 17 años de edad para las siguientes características fenotípicas: Diámetro a la altura de pecho (DAP), altura total del árbol (ALT), altura de la primera bifurcación (APB), densidad de madera (DM) y rectitud del fuste (RF). La APB fue medida en una escala categórica de 5 niveles, de tal manera que (modificado de Bush et al., 2015): un valor de 1 fue asignado a árboles con una pérdida del eje central en el primer quinto de la altura total; un valor de 2 fue asignado a árboles con una pérdida del eje central en el segundo quinto de la altura total; un valor de 3 fue asignado a árboles con una pérdida del eje central en el tercer quinto de la altura total; un valor de 4 fue asignado a árboles con una pérdida del eje central en el cuarto quinto de la altura total; y un valor de 5 fue asignado a árboles con una pérdida del eje central en el último quinto de la altura total. La DM fue indirectamente medida en términos de penetrancia mediante un Pilodyn 6J Forest (PROCEQ, Zurich, Switzerland). La DM fue medida de 2 a 3 veces por árbol a la altura de pecho (~1,3 m), hasta obtener una diferencia entre medidas de hasta 2 mm. La RF fue medida en una escala de cuatro niveles, dentro de los dos primeros tercios de la altura del árbol: se asignó un valor de 0 a árboles severamente torcidos respecto al eje central; un valor 1 para árboles con un moderado nivel de torción; un valor de 2 para árboles con fuste sutilmente curvados; y un valor de 3 fue asignado

a árboles completamente rectos. Adicionalmente, se evaluó la intensidad de floración (IF) de los árboles a los 18 años de edad, en una escala de cuatro niveles (Arriagada et al., 2018): se asignó un valor de 0 para aquellos árboles que no presentaron estructuras florales (flores, brotes o cápsulas); un valor de 1 fue asignado a árboles que presentaron estructuras florales en forma dispersa en una parte pequeña de la copa del árbol; un valor de 2 fue asignado a árboles con un 50% de su copa cubierto con estructuras florales; y un valor de 3 fue asignado a aquellos árboles con presencia de estructuras florares en la totalidad de la copa.

2.2.5 Genotipado de árboles mediante polimorfismos de nucleótido único (SNPs)

Se aisló ADN genómico de hojas de 646 y 480 individuos de *E. globulus* y *E. cladocalyx*, respectivamente, aleatoriamente seleccionados (aproximadamente 10 individuos por familia). El protocolo de extracción de ADN empleado siguió los trabajos de Doyle y Doyle (1990) y Porebsky et al. (1997). Los individuos fueron genotipados usando el arreglo comercial de SNPs EUChip60K SNP system (GeneSeek, Lincoln, NE, USA), desarrollado por Silva-Junior et al. (2015). La calidad del genotipado de las muestras fue evaluado en el programa Genome Studio software (Illumina, San Diego, CA). Los marcadores SNP monomórficos y aquellos con un Call Rate < 90% fueron descartados de los análisis. Posteriormente, se eliminaron aquellos SNPs con frecuencia mínima alélica (MAF) < 0,05, y aquellos con más de 10% de datos perdidos. Los valores perdidos fueron imputados de acuerdo al método del genotipo k-ésimo más cercano (k-nearest neighbor genotype imputation method; LD-kNNi) en el programa TASSEL versión 5.2 (Bradbury et al., 2007).

Con el fin de establecer una relación entre el historial de selección de cada población y los datos genómicos, se aplicó la prueba de neutralidad de D Tajima (Tajima, 1989), para cada una de las poblaciones estudiadas. Este análisis fue implementado en el programa TASSEL versión 5.2 (Bradbury et al., 2007). Además, se estimaron los coeficientes de parentesco basados en haplotipos (Ver punto 2.2.6) para establecer una relación entre los datos genómicos y la estructura genética de cada población. Los coeficientes de parentesco fueron calculados de acuerdo a Endelman y Jannik (2012; Centered Identity By Descendent Matrix), en el programa TASSEL versión 5.2 (Bradbury et al., 2007).

2.2.6 Estimación del desequilibrio de ligamiento y construcción de bloques de haplotipos

Se estimaron los patrones de desequilibrio de ligamiento (DL) y se construyeron bloques de haplotipos por separado para cada población de *Eucalyptus*. El patrón de DL fue expresado en términos de coeficiente de correlación alélica (r^2) entre pares de marcadores. Considerando dos loci (A y B), bialélicos (A$_1$, A$_2$, B$_1$ y B$_2$), los valores de r^2 fueron calculados por la siguiente fórmula:

$$r^2 = \frac{(p_{A_1B_1} - p_{A_1}p_{B_1})^2}{p_{A_1}(1-p_{A_1})p_{B_1}(1-p_{B_1})} \tag{1}$$

donde, p_{A_1} corresponde a la frecuencia del alelo *A₁* del locus A. p_{B_1} corresponde a la frecuencia del alelo *B₁* del locus B. $p_{A_1B_1}$ es la frecuencia de individuos que poseen el alelo *A₁* del locus A y el alelo *B₁* del locus B. Las frecuencias de los alelos A$_2$ y B$_2$ son contabilizados por la expresión $1 - p_{A_1}$ y $1 - p_{B_1}$, respectivamente. Los valores de r^2 fueron corregidos por los coeficientes de parentesco (entre individuos en la biblioteca de R LDcorSV versión 1.3.2 (Mangin et al., 2012). La curva de disminución de DL fue ajustada de acuerdo a Hill y Weir (1988). Las posiciones físicas de cada SNP fueron establecidas de acuerdo al mapa consenso del genoma de *E. grandis* (Myburg et al., 2014). El valor crítico de r^2 fue calculado de acuerdo al método propuesto por Breseghello y Sorells (2006).

Para la construcción de los bloques de haplotipos, el DL entre pares de marcadores SNPs fue estimado de acuerdo a los valores del coeficiente de desequilibrio (D') entre pares de SNPs pertenecientes a un mismo cromosoma. Los bloques de haplotipos fueron definidos de acuerdo al algoritmo de intervalo de confianza Gabriel et al. (2002) en el programa Haploview v. 4.2 (Barret et al., 2005). Los pares de SNPs fueron considerados en fuerte DL si el límite superior del intervalo de confianza del 95% del valor de D' es superior a 0,98 y si el límite inferior tiene un valor mínimo de 0,7. El valor de D' entre dos loci (A y B), con dos alelos (A$_1$, A$_2$, B$_1$, B$_2$, respectivamente) fue calculado como:

$$D'_{AB} = D/D_{MAX} \tag{2}$$

donde D es calculado como: $D = p_{A_1B_1}p_{A_2B_2} - p_{A_1B_2}p_{A_2B_1}$. El valor de D_{MAX} es estimado como:

$$D_{MAX} = \begin{cases} -min\{p_{A_1}p_{B_1}, p_{A_2}p_{B_2}\}, \text{cuando } D < 0 \\ min\{p_{A_1}p_{B_2}, p_{A_2}p_{B_1}\}, \text{cuando } D \geq 0 \end{cases} \tag{3}$$

2.2.7 Predicción basada en pedigrí y datos genómicos

En el modelo basado en información de pedigrí, los méritos genéticos de los árboles y los componentes de varianza fueron predichos mediante un modelo lineal generalizado implementado en el paquete de R v.3.6.1, MCMCglmm (Markov Chain Monte Carlo—Generalized Linear Mixed Model; Hadfield, 2010). Este análisis fue llevado a cabo mediante el siguiente modelo:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \boldsymbol{\varepsilon} \tag{4}$$

donde $\mathbf{y}$ corresponde al vector de datos fenotípicos, $\boldsymbol{\beta}$ es el vector de los efectos fijos de bloque, $\mathbf{a}$ es el vector de los efectos genéticos aditivos, donde $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$, $\mathbf{A}$ es la matriz de los coeficientes de parentesco de Wright (calculados dado la estructura de pedigrí), y σ_a^2 corresponde a la varianza genética aditiva. $\mathbf{X}$ y $\mathbf{Z}$ corresponden a las matrices de incidencia relacionadas con los vectores $\boldsymbol{\beta}$ y $\mathbf{a}$, respectivamente. $\boldsymbol{\varepsilon}$ corresponde al vector de efectos residuales, los cuales $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_e^2)$. $\mathbf{I}$ es una matriz de identidad y σ_e^2 corresponde a varianza residual. El modelo bayesiano fue implementado considerando 1000000 de iteraciones, con un periodo de quema de 100000 iteraciones y un ancho de muestra de 50.

Los métodos de predicción genómica basados en SNP y/o haplotipos usados en este estudio fueron los siguientes: Bayes A (BA), Bayes B (BB), Bayes C (BC), y Regresión Contraída Bayesiana (Bayesian Ridge Regression; Región Bayesiana Contraída; BRR). Todas las predicciones genómicas fueron realizadas de acuerdo al siguiente modelo:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zm} + \boldsymbol{\varepsilon} \tag{5}$$

donde $\mathbf{y}$ corresponde al vector de datos fenotípicos, $\boldsymbol{\beta}$ es el vector de los efectos fijos de bloque, $\mathbf{m}$ es el vector de los efectos de los marcadores SNP y/o haplotipos). Las distribuciones a priori

de los efectos y las varianzas de marcadores son establecidas de acuerdo al método de predicción seleccionado (BA, BB, BC, o BRR; Ver marco teórico). $\boldsymbol{\varepsilon}$ corresponde al vector de efectos residuales, los cuales $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_e^2)$. $\mathbf{I}$ es una matriz de identidad y σ_e^2 corresponde a varianza residual. X y Z corresponden a las matrices de incidencia que relacionan al vector de observaciones $\mathbf{y}$ con los $\boldsymbol{\beta}$ y $\mathbf{m.}$ De acuerdo con Mora et al. (2017) y Arriagada et al. (2018), la población de *E. cladocalyx* está genéticamente diferenciada en sub-poblaciones, por lo tanto, se consideró el efecto de la estructura genética de población, la cual fue incluida como una covariable contenida en el vector de los efectos fijos del modelo. La estructura genética fue inferida de acuerdo al método de agrupamiento bayesiano, a través del algoritmo de cadenas de Markov de Monte Carlo en el programa STRUCTURE v.2.3.4 (Pritchard et al. 2000). Se evaluó una estructura de población compuesta de 2 a 8 grupos genéticamente homogéneos. El número óptimo de grupos fue determinado de acuerdo a Evanno et al. (2005). Los modelos fueron implementados usando 10 simulaciones independientes, las cuales consistieron de 100000 iteraciones y un periodo de quema de 10000 iteraciones.

Las matrices de marcadores SNP y/o haplotipos fueron codificadas por los valores 0, 1 y 2. En caso de los SNPs, el valor de 0 representa el genotipo homocigoto para el alelo de menor frecuencia en la población genotipada para el i-ésimo SNP (i=1,…n), un valor de 1 representa a los genotipos heterocigotos, y un valor de 2 representa el genotipo homocigoto para el alelo de mayor frecuencia en la población genotipada. A diferencia de los marcadores SNP, los bloques de haplotipos pueden ser multi-alélicos, por lo tanto, los valores de 0, 1 y 2 representa el número de copias de cada haplotipo (alelo del bloque de haplotipo) que posee cada individuo. En el caso de los haplotipos, un valor de 0 es asignado a aquellos individuos que no presentaron el j-ésimo haplotipo del i-ésimo bloque de haplotipo, un valor de 1 implica que el individuo posee una copia del j-ésimo haplotipo del i-ésimo bloque de haplotipo, y un valor de 2 implica que el individuo posee dos copias del j-ésimo haplotipo del i-ésimo bloque de haplotipo. Esta forma de codificación para los haplotipos fue basada en el trabajo de Cuyabano et al. (2014) y He et al. (2019). En el presente trabajo, se pusieron a prueba tres tipos de modelos: 1) modelo de predicción basado en SNPs individuales (expresado en el texto como SNP); 2) modelo de predicción basado en haplotipos (expresado en el texto como HAP); y 3) modelo de predicción

basado en haplotipos y aquellos SNP no asignados a ningún haplotipo particular (expresado en el texto como HAP-SNP).

2.2.8 Estimación de parámetros genéticos

La heredabilidad en sentido estricto de todas las características estudiadas, basada en los antecedentes de pedigrí fue obtenida con la siguiente fórmula:

$$\hat{h}_a^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} \tag{6}$$

donde $\hat{\sigma}_a^2$ y $\hat{\sigma}_e^2$ corresponden a la varianza genética aditiva y residual, respetivamente. Las heredabilidades genómicas de SNP y/o haplotipos ($\hat{h}_g^2$) fueron estimadas usando las distribuciones marginales a posteriori de cada parámetro estimado (Torres et al., 2018; Volpato et al., 2019; Mora et al., 2019). La varianza genética aditiva ($\hat{\sigma}_g^2$) para cada modelo de predicción fue calculada de acuerdo las siguientes fórmulas (Silva et al., 2018):

Para los modelos basados en los métodos BRR y BC:

$$\hat{\sigma}_g^2 = 2\hat{\sigma}_m^2 \sum_{i=1}^n p_i(1 - p_i) \tag{7}$$

Para los modelos basados en los métodos BA y BB:

$$\hat{\sigma}_g^2 = 2 \sum_{i=1}^n p_i(1 - p_i)\hat{\sigma}_{m_i}^2 \tag{8}$$

donde, p_i es la frecuencia del alelo de menor frecuencia del i-ésimo marcador y $\hat{\sigma}_m^2$ es una varianza común para todos los marcadores en el caso de los modelos BC y BRR. En los métodos BA y BB, se estima una varianza individual para cada marcador ($\hat{\sigma}_{m_i}^2$). Tanto $\hat{\sigma}_m^2$ como $\hat{\sigma}_{m_i}^2$ fueron estimadas por la expresión (5).

2.2.9 Comparación del poder predictivo y bondad de ajuste de los modelos de predicción genómica

Para definir el mejor modelo de predicción para cada una de las características fenotípicas evaluadas, se consideró el siguiente flujo: 1) Primero, se identificaron los modelos con mejor bondad de ajuste para predecir cada una de las características evaluadas en cada población; 2) Segundo, se identificaron aquellos modelos con mayor poder predictivo; 3) Tercero, se identificaron los modelos que estimaron mayor variabilidad genética para cada una de las características, en términos de heredabilidad genómica.

La bondad de ajuste de los modelos fue evaluada de acuerdo al Criterio de Información de Desviación (Deviance Criterion Information; DIC; Spiegelhalter et al., 2002). Una diferencia de 10 entre los valores de DIC para cada modelo se consideró como una fuerte evidencia de que existen diferencias en términos de bondad de ajuste. Una diferencia de 5 entre los valores de DIC se consideró como una diferencia significativa entre modelos, y un valor < 5 se consideró que no existen diferencias entre los modelos, en términos de bondad de ajuste.

El poder predictivo (PP) de cada modelo fue calculado como el valor de correlación de Pearson entre los valores fenotípicos predichos ($\hat{y}*$) por la expresión (5) y los valores fenotípicos ajustados por los efectos fijos del modelo ($\hat{y}*$; Tan et al., 2018). Los modelos fueron entrenados con el 90% de los individuos de cada población (muestreados al azar) y el 10% restante fue considerado como población de validación. El PP fue informado como el promedio de coeficientes de correlación para 100 ciclos de validación cruzada.

2.3 RESULTADOS

2.3.1 Representatividad del arreglo SNP60K en *E. globulus* y construcción de bloques de haplotipos

Un total de 14422 SNPs informativos (MAF>0,05; call rate>90%) fueron reportados en esta la población La Poza de *E. globulus*. Un promedio de 1356 SNPs fueron detectados por cromosoma, con una frecuencia de 1 SNP por cada 4000 pb (Figura 1). Estos marcadores fueron usados para determinar el patrón de disminución de desequilibrio de ligamiento (DL) a través de todo el genoma, y construir bloques de haplotipos mediante el método de Gabriel et al. (2002). El DL disminuyó rápidamente dentro de los ~ 35 Kpb a través de todo el genoma (Figura

2; r^2 crítico= 0,15). Un total de 1137 bloques de haplotipos y 3279 haplotipos fueron identificados en todos los cromosomas de la especie de estudio. Sólo el 14,4 % del total de SNPs fueron agrupados en bloque es haplotipos. En promedio, se construyeron 300 bloques de haplotipos por cromosoma. Los bloques de haplotipos fueron construidos con un máximo promedio de ~7 SNPs en desequilibrio de ligamiento. En específico, en el cromosoma Chr11 se detectó hasta 12 SNPs en desequilibrio de ligamiento.
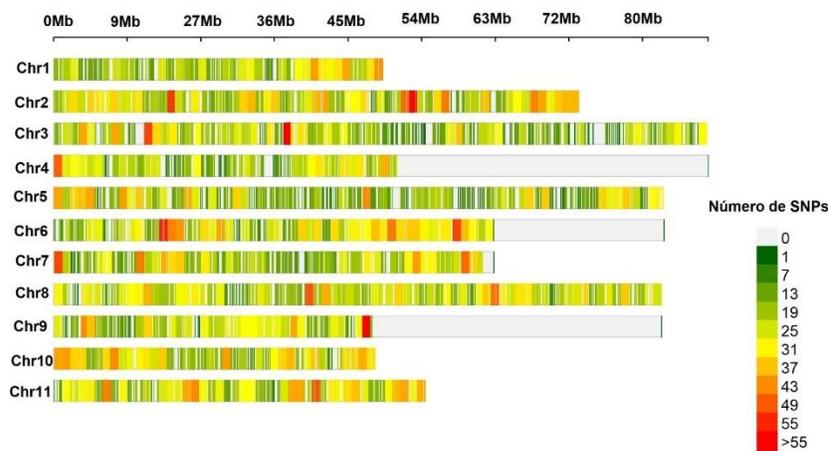
Figura 1. Ideograma representando densidad de SNPs en un rango de 1 Mpb para cada cromosoma (Chr=11) en la población estudiada de *E. globulus*.
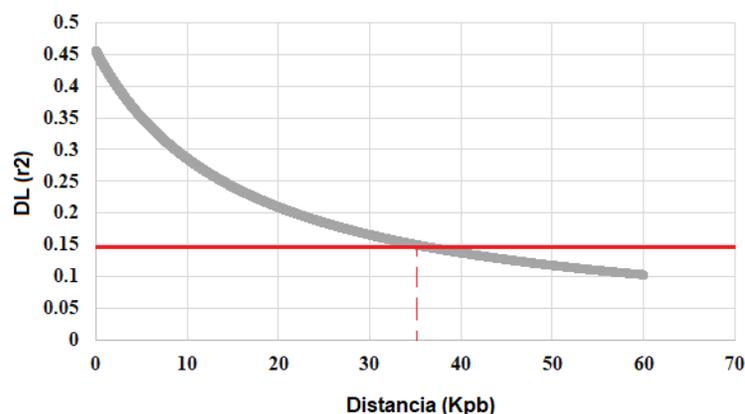
Figura 2. Disminución del desequilibrio de ligamiento (DL) a traves de todo el genoma para la población en estudio de *E. globulus*. Los valores de DL entre pares de marcadores fueron expresados en términos de r^2. La curva de dimisnucion de DL fue ajustada de acuerdo a Hill y Weir (1988). La distancia entre SNP fue expresada en Kpb. El valor de umbral de disminución de DL fue de $r^2= 0,15$, el cual fue representado por una línea continua color rojo. El punto donde disminuye el DL a nivel de distancia genómica es representado por una línea vertical discontinua color rojo.

2.3.2 Representatividad del arreglo SNP60K en *E. cladocalyx* y construcción de bloques de haplotipos

Un total de 3879 SNPs informativos (MAF>0,05; call rate>90%) fueron reportados en la población evaluada de *E. cladocalyx*. Un promedio de 353 SNPs fueron detectados por cromosoma, con una frecuencia de 1 SNP por cada 11600 pb (Figura 3). Estos marcadores fueron usados para determinar el patrón de disminución de desequilibrio de ligamiento (DL) a través de todo el genoma, y construir bloques de haplotipos mediante el método de Gabriel et al. (2002). El DL disminuyó rápidamente dentro de los ~ 32 Kpb a través de todo el genoma (r^2 crítico= 0,14; Figura 4). Un total de 108 bloques de haplotipos y 311 haplotipos fueron identificados en todos los cromosomas de la especie de estudio. El 94 % (3636 SNPs) no fueron agrupados en bloques de haplotipos. Los bloques de haplotipos fueron construidos con un máximo promedio de ~2 SNPs en desequilibrio de ligamiento. En específico, en el cromosoma Chr6 se detectó hasta 5 SNPs en desequilibrio de ligamiento.

Figura 3. Ideograma representando densidad de SNPs en un rango de 1 Mpb para cada cromosoma (Chr=11) en la población estudiada de *E. cladocalyx*.
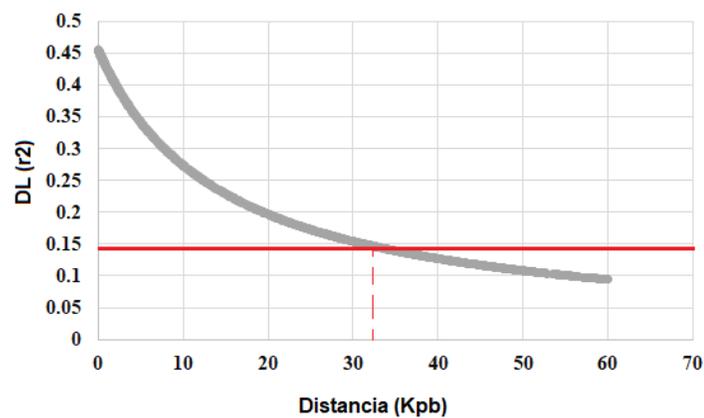
Figura 4. Disminución del desequilibrio de ligamiento (DL) a través de todo el genoma para la población en estudio de *E. cladocalyx*. Los valores de DL entre pares de marcadores fueron expresados en términos de r^2. La curva de dimisnucion de DL fue ajustada de acuerdo a Hill y Weir (1988). La distancia entre SNPs fue expresada en Kpb. El valor de umbral de disminución de DL fue de $r^2= 0,14$, el cual fue representado por una color rojo. El punto donde disminuye el DL a nivel de distancia genómica es representado por una línea vertical discontinua color rojo.

Para la población de *E. globulus*, los valores D (D_T) variaron entre 3,14 (Cromosoma Chr3) y 3,47 (Cromosoma Chr11) (p<0,05), con un promedio de 3,31, mientras que los valores D en la población de *E. cladocalyx* fluctuaron entre 1,42 (Cromosoma Chr5) y 1,74 (Cromosoma Chr6) , con un promedio de 1,65. Los parentescos genómicos se encuentran representados mediante mapas de calor en la Figura 5 para ambas poblaciones de *Eucalyptus*. Para ambas poblaciones, el parentsco genómico tuvo un promedio de cero. Los coeficientes de parentesco basados en haplotipos para la población de *E. globulus* tuvieron un promedio de cero y variaron entre -0,22 y 1,43. Para la población de *E. cladocalyx*, los coeficientes de parentesco basados en haplotipos variaron entre -0,18 y 0,57.
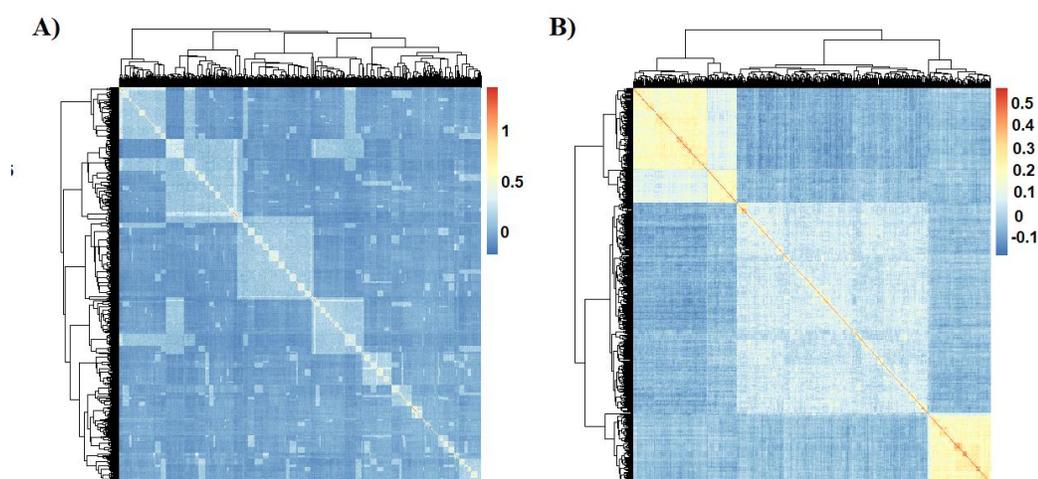
Figura 5. Mapas de calor de los coeficientes de parentesco génomico basado en haplotipos, entre los individuos que componen ambas poblaciones de *Eucalyptus*. En A) y B) son mostrados los valores de parentesco en las poblaciones de *E. globulus* y *E. cladocalyx*, respectivamente.

2.3.3 Estimación de parámetros genéticos basados en pedigrí y datos genómicos de *E. globulus*

Los componentes de varianza y heredabilidad fueron estimados para todos los modelos de predicción basados en genómica e información de pedigrí. La heredabilidad en sentido estricto basada en pedigrí varió entre 0,04 a 0,46 para todas las características evaluadas en *E. globulus* (Tabla 3). La densidad de madera (DM) fue una de las características mayormente

heredables, mientras que el diámetro a la altura de pecho (DAP), rectitud de fuste (RF) y calidad de ramas (CR) tuvieron una heredabilidad relativamente baja. Los métodos de predicción genómica fueron comparados con las estimaciones basadas en pedigrí utilizando los valores promedio de las distribuciones marginales posteriores de cada parámetro estimado. En la mayoría de los casos, las estimaciones de heredabilidad genómica fueron más altas que la estimación basada en pedigrí para la población evaluada de *E. globulus*. Para la altura de árboles (ALT), las estimaciones de heredabilidad genómica variaron entre 0,11 (BA-SNP) y 0,36 (BC-SNP y BC-HAP-SNP). Las estimaciones basadas en SNP y HAP-SNP fueron estadísticamente más altas que la heredabilidad basada en pedigrí. En el caso del DAP, las estimaciones de heredabilidad genómica basadas en BC y BRR variaron entre 0,26–0,32 y 0,12–0,16, respectivamente, y fueron significativamente más altas que estimación de heredabilidad basada en pedigrí. Para la RF, las estimaciones de heredabilidad variaron entre 0,09 (BA-HAP) y 0,34 (BC-SNP y BC-HAP-SNP). Particularmente, las estimaciones de heredabilidad basadas en los métodos BB, BC y BRR fueron más altas que la heredabilidad basada en pedigrí. Las estimaciones de heredabilidad genómica de la CR variaron entre 0,04 (BA-SNP) y 0,29 (BC-SNP y BC-HAP-SNP). Para esta misma característicalas heredabilidades basadas en los métodos BC y BRR, fueron más altas que la heredabilidad basada en pedigrí. Contrariamente a los resultados anteriores, las estimaciones de heredabilidad genómica basadas en cualquier modelo fueron estadísticamente más bajas que la heredabilidad basada en pedigrí de la DM. Las estimaciones de heredabilidad genómica de DM variaron entre 0,05 (BA-HAP) y 0,34 (BC-SNP).

Tabla 3. Estimaciones de heredabilidad basadas en pedigrí (h_a^2) y genómica (h_g^2) para cada característica evaluada en *E. globulus*. Los métodos de predicción se basaron en: pedigrí (PBP), SNPs (SNP), haplotipos (HAP), y haplotipos en conjunto con SNPs no asignados a haplotipos (HAP-SNP). BA, BB, BC y BRR corresponden a Bayes A, Bayes B, Bayes C y regresión contraída bayesiana, respectivamente.

Característica/Modelo	Pedigrí	SNP	HAP	HAP-SNP
	h_a^2 [RC]	h_g^2	h_g^2	h_g^2
Altura de árbol				
PBP	0,15 [0,01–0,28]	-	-	-
BA	-	0,11	0,06	0,12
BB	-	0,27	0,11	0,29
BC	-	0,36	0,28	0.36
BRR	-	0,19	0,14	0,20
Diámetro a la altura de pecho				
PBP	0,04 [<0,01–0,10]	-	-	-
BA	-	0,08	0,04	0,07
BB	-	0,19	0,09	0,14
BC	-	0,31	0,26	0,32
BRR	-	0,16	0,12	0,16
Rectitud de fuste				
PBP	0,06 [<0,01–0,14]	-	-	-
BA	-	0,11	0,09	0,09
BB	-	0,26	0,18	0,28
BC	-	0,34	0,30	0,34
BRR	-	0,18	0,15	0,20
Calidad de ramas				
PBP	0,05 [<0,01–0,11]	-	-	-
BA	-	0,04	0,04	0,05
BB	-	0,12	0,10	0,08

BC	-	0,29	0,25	0,29
BRR	-	0,15	0,12	0,15
Densidad de madera				
PBP	0,46 [0,22–0,69]	-	-	-
BA	-	0,07	0,05	0,08
BB	-	0,17	0,12	0,16
BC	-	0,34	0,26	0,33
BRR	-	0,16	0,12	0,17

RC: Región de credibilidad del 90% de las distribuciones marginales a posterior.

2.3.4 Estimación de parámetros genéticos basados en pedigrí y datos genómicos de *E. cladocalyx*

Los componentes de varianza y heredabilidad fueron estimados para todos los modelos de predicción basados en genómica e información de pedigrí. La heredabilidad en sentido estricto basada en pedigrí varió entre 0,29 y 0,52 para todas las características evaluadas (Tabla 4). Todas las características evaluadas fueron de modera a altamente heredables. RF y ALT de los fueron las características mayormente heredables (0,52 y 0,44, respectivamente), mientras que DAP y APB tuvieron una menor estimación de heredabilidad (0,29 y 0,3, respetivamente). En la mayoría de los casos, las estimaciones de heredabilidad genómica fueron iguales o más bajas que aquellas basadas en pedigrí. En el caso de ALT, la mayoría de las heredabilidades genómicas (ya sean basadas en SNP, HAP o HAP-SNP) fueron estadísticamente iguales a la estimación basada en pedigrí. Particularmente, la estimación de heredabilidad basada en HAP y el método BB fue más alta que para la estimación basada en pedigrí. El DAP tuvo una heredabilidad moderada en la población estudiada de acuerdo al método de predicción basado en pedigrí. Por su parte, las estimaciones de heredabilidad genómica fluctuaron entre los valores de 0,14 y 0,69, de las cuales, la mayoría fue estadísticamente más alta que la estimación basada en pedigrí. Más aun, todas las estimaciones de heredabilidad basadas en HAP fueron más altas que aquellas basadas en pedigrí. La heredabilidad basada en pedigrí de la RF fue altamente heredable (>0,5) y fue más alta que para cualquiera de los métodos basados en genómica probados en este estudio (las cuales fluctuaron entre los valores 0,18 y 0,55). En promedio, las

estimaciones de heredabilidad basadas en SNP fueron las más bajas estimadas para RF. De acuerdo al modelo basado en pedigrí, la DM fue moderadamente heredable y las estimaciones basadas en genómica fueron estadísticamente iguales a aquellas basadas en pedigrí. La IF tuvo una heredabilidad moderada (0,41), y ésta fue más alta o igual que aquellas estimadas por los métodos genómicos, exceptuando al modelo basado en HAP y el método BC. La mayoría de las estimaciones de heredabilidad genómica para APB fueron estadísticamente iguales a la estimación basada en pedigrí. Por otro lado, las estimaciones de heredabilidad basadas en HAP y/o HAP-SNP y el método BB fueron más altas que la estimación basada en pedigrí, mientras que la estimación basada en SNP, para este mismo método, fue estadísticamente inferior (más del 50%) a la heredabilidad basada en pedigrí.

Tabla 4. Estimaciones de heredabilidad basadas en pedigrí (h_a^2) y genómica (h_g^2) para cada característica evaluada en *E. cladocalyx*. Los métodos de predicción se basaron en: pedigrí (PBP), SNPs (SNP), haplotipos (HAP), y haplotipos en conjunto con SNPs no asignados a haplotipos (HAP-SNP). BA, BB, BC y BRR corresponden a Bayes A, Bayes B, Bayes C y regresión contraída bayesiana, respectivamente.

Característica/Modelo	Pedigrí	SNP	HAP	HAP-SNP
	h_a^2 [RC]	h_g^2	h_g^2	h_g^2
Altura del árbol				
PBP	0,44 [0,31-0,57]	-	-	-
BA	-	0,24	0,41	0,25
BB	-	0,45	0,58	0,38
BC	-	0,45	0,57	0,44
BRR	-	0,24	0,39	0,26
Diámetro a la altura de pecho				
PBP	0,3 [0,23-0,37]	-	-	-
BA	-	0,14	0,40	0,43
BB	-	0,37	0,61	0,37
BC	-	0,39	0,57	0,44

BRR	-	0,24	0,38	0,20
Rectitud de fuste				
PBP	0,28 [0,14-0,31]	-	-	-
BA	-	0,32	0,28	0,25
BB	-	0,31	0,60	0,43
BC	-	0,30	0,55	0,45
BRR	-	0,30	0,18	0,20
Densidad de madera				
PBP	0,36 [0,13-0,59]	-	-	-
BA	-	0,28	0,31	0,38
BB	-	0,5	0,5	0,46
BC	-	0,45	0,52	0,44
BRR	-	0,42	0,19	0,23
Intensidad de floración				
PBP	0,41 [0,34-0,48]	-	-	-
BA	-	0,04	0,19	0,16
BB	-	0,41	0,34	0,10
BC	-	0,46	0,58	0,48
BRR	-	0,26	0,33	0,31
Altura de la primera bifurcación				
PBP	0,29 [0,16-0,42]	-	-	-
BA	-	0,05	0,18	0,17
BB	-	0,12	0,45	0,44
BC	-	0,27	0,50	0,27
BRR	-	0,14	0,24	0,24

RC: Región de credibilidad del 90% de las distribuciones marginales a posterior.

2.3.5 Poder predictivo de modelos de predicción genómica en *E. globulus*

Se evaluó el poder predictivo (PP) de SNPs, haplotipos (HAP) y haplotipos en conjunto con SNPs que no fueron asignados a un haplotipo (HAP-SNP) (Tabla 5). Estos modelos fueron

basados en diferentes métodos de predicción: Bayes A (BA), Bayes B (BB), Bayes C (Bayes C), y regresión contraída bayesiana (BRR). Los valores de PP dependieron del método (BA, BB, BC o BRR) y marcador empleado (SNP, HAP o HAP-SNP). En promedio, el método BRR y BC tuvieron un mayor valor de PP para predecir la mayoría de las características evaluadas. Particularmente, los modelos genómicos que incluyeron el efecto de haplotipos (ya sea HAP o HAP-SNP) tuvieron un mayor PP para características con una menor heredabilidad, tales como RF, DAP y CR.

Los valores de PP para ALT variaron entre 0,21 y 0,44. Los modelos basados en el método BC tuvieron en promedio un valor de PP (PP= 0,36) mayor que el resto de los modelos. Particularmente, el modelo con mayor habilidad para predecir ALT en árboles de *E. globulus* fue el modelo BC basado en SNP (PP= 0,44), seguido por el modelo BC basado en HAP-SNP (PP= 0,38). Para el DAP, los valores de PP de los modelos de predicción fluctuaron entre 0,19 y 0,46. En general, el PP de los modelos basados en los métodos BC y BRR fue mayor que para el resto de los modelos. El modelo BC basado en HAP-SNP tuvo el mayor PP para DAP (PP= 0,46), seguido por el modelo BRR basado en SNP. El PP para la RF en *E. globulus* varió de 0,38 a 0,58. Los modelos con el PP más alto para RF fueron el modelo BRR basado en HAP (PP=0,58) y el modelo BC basado en SNP (PP= 0,54). El PP de la CR de *E. globulus* fluctuó entre 0,13 y 0,33. Entre los modelos basados en SNP, BC tuvo el mayor valor de PP para la CR (PP = 0,31). Para los modelos basados en HAP, el valor de PP más alto fue obtenido por un modelo BC (PP = 0,28). En el caso de los modelos basados en HAP-SNP, BC y BRR tuvieron valores de PP más altos (PP = 0,31 y 0,33, respectivamente) que los demás modelos de predicción. Finalmente, el PP de DM de *E. globulus* varió entre 0,26 y 0,46.

Tabla 5. Estimación del poder predictivo (PP) de SNPs, haplotipos (HAP) y SNP en conjunto con SNP no asignados a haplotipos (HAP-SNP). BA, BB, BC y BRR corresponden a Bayes A, Bayes B, Bayes C y regresión contraída bayesiana, respectivamente, para predecir características cuantitativas en *E. globulus*. Los valores de PP expuesto en tabla corresponden al valor promedio de PP para 100 ciclos de validación cruzada.

Característica/Marcador	Método			
	BA	BB	BC	BRR
Altura de árbol				
SNP	0,31	0,32	0,44	0,30
HAP	0,21	0,28	0,25	0,35
HAP-SNP	0,21	0,31	0,38	0,33
Diámetro a la altura de pecho				
SNP	0,35	0,34	0,39	0,45
HAP	0,26	0,21	0,33	0,36
HAP-SNP	0,28	0,19	0,46	0,37
Rectitud de fuste				
SNP	0,38	0,52	0,54	0,48
HAP	0,40	0,42	0,50	0,58
HAP-SNP	0,38	0,49	0,52	0,52
Calidad de ramas				
SNP	0,22	0,13	0,16	0,31
HAP	0,20	0,14	0,28	0,22
HAP-SNP	0,19	0,18	0,31	0,33
Densidad de madera				
SNP	0,26	0,30	0,46	0,32
HAP	0,32	0,31	0,39	0,41
HAP-SNP	0,34	0,29	0,32	0,44

2.3.6 Poder predictivo de modelos de predicción genómica en *E. cladocalyx*

Los valores de PP de los modelos que predijeron la ALT de *E. cladocalyx* variaron entre 0,26 y 0,33 (Tabla 6). Los modelos basados en el SNP tuvieron en promedio un valor de PP (PP= 0,32) mayor que el resto de los modelos. Particularmente, el modelo tuvo un mayor PP para ALT de E. cladocalyx (PP=0,33), seguido por el resto de los modelos basados en SNP (PP= 0,32). Para el DAP, los valores de PP de los modelos de predicción variaron entre 0,17 y 0,24. En general, el mayor poder predictivo para el DAP de *E. cladocalyx* fue obtenido por los modelos basados en SNP y HAP. Particularmente, el modelo BC basado en SNP tuvo el mayor PP para DAP (PP= 0,24), seguido por el resto de los modelos basados en SNP. La capacidad de los modelos para predecir la RF varió de 0,39 a 0,44. Los modelos con el valor de PP más alto para RF fueron los modelos BA y BB basados en HAP-SNP (PP=0,44), seguidos por los modelos BC y BRR basados en HAP-SNP (PP= 0,43). El poder predictivo para APB en *E. cladocalyx* varió entre 0,14 y 0,19. En promedio, los modelos basados en SNP y HAP tuvieron un valor de PP más alto que los modelos basados en HAP-SNP. El poder predictivo para la DM de *E. cladocalyx* varió entre 0,22 y 0,25. Los modelos basados en SNP tuvieron un valor de PP (PP= 0,25) más alto que otros modelos para DM. Finalmente, el PP para la IF de *E. cladocalyx* fue relativamente más bajo que para el resto de las características. El valor de PP varió de 0,06 a 0,1. Particularmente, los modelos BC y BRR basados en SNP tuvieron un mayor poder predictivo que otros modelos.

Tabla 6. Estimación del poder predictivo (PP) de SNPs, haplotipos (HAP) y SNP en conjunto con SNP no asignados a haplotipos (HAP-SNP). BA, BB, BC y BRR corresponden a Bayes A, Bayes B, Bayes C y regresión contraída bayesiana, respectivamente, para predecir características cuantitativas en *E. cladocalyx* Los valores de PP expuesto en tabla corresponden al valor promedio de PP para 100 ciclos de validación cruzada.

Característica/Marcador	Método			
	BA	BB	BC	BRR
Altura de árbol				
SNP	0,32	0,32	0,32	0,33
HAP	0,29	0,28	0,29	0,29
HAP-SNP	0,26	0,26	0,26	0,27
Diámetro a la altura de pecho				
SNP	0,23	0,23	0,24	0,23
HAP	0,22	0,22	0,21	0,23
HAP-SNP	0,17	0,18	0,18	0,18
Rectitud de fuste				
SNP	0,40	0,40	0,40	0,39
HAP	0,41	0,41	0,41	0,40
HAP-SNP	0,44	0,44	0,43	0,43
Densidad de madera				
SNP	0,25	0,25	0,25	0,25
HAP	0,23	0,22	0,23	0,23
HAP-SNP	0,22	0,22	0,22	0,22
Altura de la primera bifurcación				
SNP	0,18	0,18	0,18	0,19
HAP	0,17	0,19	0,19	0,17
HAP-SNP	0,15	0,14	0,14	0,15
Intensidad de floración				
SNP	0,07	0,09	0,10	0,10
HAP	0,07	0,07	0,08	0,06

HAP-SNP	0,07	0,08	0,08	0,08

2.3.7 Evaluación del mejor modelo de predicción genómica para caracteristicas fenótipicas en *E. globulus*

De acuerdo a los valores de DIC (ANEXO III, Tabla S2), los modelos basados en SNP y HAP-SNP tuvieron mejor bondad de ajuste que los modelos basados en HAP para la ALT de *E. globulus*. En particular, los modelos BC basados en SNP y HAP-SNP, y el modelo BB basado en HAP-SNP tuvieron las mejores bondades de ajuste que el resto de los modelos. En términos de poder predictivo, el modelo BC basado en SNP fue superior a los modelos basados en HAP-SNP. Interesantemente, la estimación de heredabilidad genómica del modelo BC basado en SNP tuvo un valor más alto que la heredabilidad genómica estimada por los modelos BB y BC basados en HAP-SNP. Adicionalmente, la heredabilidad genómica estimada por el modelo BC basado en SNP tuvo un valor más alto que la heredabilidad basada en pedigrí. En el caso del DAP de los árboles de *E. globulus*, todos los modelos tuvieron una misma bondad de ajuste. El modelo BC basado en HAP-SNP y el modelo BRR basado en SNP tuvieron un mayor PP que el resto de los modelos para el DAP de *E. globulus*. No obstante, la estimación de heredabilidad genómica del modelo BC basado en HAP-SNP fue superior a la estimación de heredabilidad obtenida a través del modelo BRR basado en SNP. En adición, la heredabilidad genómica estimada por el modelo BC basado en HAP-SNP tuvo un valor 7 veces más alto que la heredabilidad estimada por el modelo basado en pedigrí. Todos los modelos genómicos para la RF de los árboles de *E. globulus* tuvieron la misma bondad de ajuste. Por otro lado, el PP del modelo BRR basado en HAP fue mayor al resto de los modelos. La estimación de heredabilidad genómica del modelo BRR basado en HAP fue ~ dos veces más alta que la heredabilidad basada en pedigrí. En general, los modelos basados en HAP o HAP-SNP tuvieron mejor bondad de ajuste que los modelos basados en SNP para predecir la CR de *E. globulus*. En particular, los modelos BA, BB y BC basados en HAP y los modelos BA, BC y BRR basados en HAP-SNP fueron superiores al resto de los modelos en términos de bondad de ajuste. Por otro lado, los modelos BC basados en HAP y HAP-SNP, y el modelo BB basado en HAP-SNP, tuvieron un mayor PP que el resto de los modelos (con una misma bondad de ajuste). No obstante, la

57

heredabilidad genómica del modelo BC basado en HAP-SNP fue más alta que las estimaciones basadas en los otros modelos, e incluso que la estimación de heredabilidad basada en pedigrí. En el caso de la DM de los árboles de *E. globulus*, todos los modelos tuvieron una misma bondad de ajuste. Los modelos BC basados en SNP y HAP, y el modelo BRR basado en HAP-SNP tuvieron un mayor PP que el resto de los modelos para la DM. No obstante, la estimación de heredabilidad genómica del modelo BC basado en SNP fue superior a la estimación de heredabilidad obtenida a través del modelo BRR basado en HAP-SNP y el modelo BC basado en HAP. Por otro lado, la heredabilidad genómica estimada mediante los modelos BC basados en SNP y HAP fueron iguales a la estimación basada en pedigrí, mientras que la heredabilidad genómica del modelo BRR basado en HAP-SNP fue inferior a estimación basada en pedigrí.

2.3.8 Evaluación del mejor modelo de predicción genómica para caracteristicas fenótipicas en *E. cladocalyx*

En términos de bondad de ajuste, los modelos BA fueron los menos adecuados para predecir la ALT de *E. cladocalyx* (ANEXO III, Tabla S3). El modelo BB basado en SNP tuvo la mejor bondad de ajuste entre todos los modelos de predicción genómica. El poder predictivo de este mismo modelo fue de 0,32. En adición, la heredabilidad genómica estimada mediante el modelo BB basado en SNP fue equivalente a la heredabilidad basada en pedigrí. En general, los modelos basados en SNP o HAP tuvieron mejor bondad de ajuste que los modelos basados en HAP-SNP para predecir el DAP de árboles de *E. cladocalyx*. En particular, los modelos BB, BC y BRR basados en SNP y los modelos BB, BC y BRR basados en HAP fueron superiores al resto de los modelos en términos de bondad de ajuste. En promedio, el PP de los modelos basados en SNP fue sutilmente más alto que el poder de los modelos basados en HAP. La heredabilidad genómica de los modelos basados en SNP varió entre 0,14 y 0,39. La heredabilidad estimada por el modelo BC basado en SNP para DAP fue superior a la estimación de heredabilidad dada por los antecedentes de pedigrí, mientras que la heredabilidad genómica estimada por los métodos BB y BRR fueron estadísticamente iguales a la heredabilidad basada en pedigrí.

En términos de bondad de ajuste, los modelos basados en HAP-SNP fueron superiores al resto de los modelos para predecir la RF de los árboles de *E. cladocalyx,* exceptuando al modelo BRR basado en HAP-SNP. El PP fue similar entre los modelos basados en HAP-SNP, no obstante, los modelos BB y BC tuvieron una estimación de heredabilidad genómica más alta que la estimación basada en BA. La heredabilidad genómica estimada por el modelo BB basado en HAP-SNP fue estadísticamente más pequeña que la estimación basada en pedigrí, mientras que el modelo BC basado en HAP-SNP estimó un valor de heredabilidad equivalente a la heredabilidad basada en pedigrí. En el caso de la APB de los árboles de *E. cladocalyx*, los modelos BB, BC y BRR basados en SNP fueron superiores al resto de los modelos en términos de bondad de ajuste. El PP fue similar entre estos modelos, no obstante, la heredabilidad genómica estimada por el modelo BC basado en SNP fue más alta que la estimada por los métodos BB y BRR. No obstante, la heredabilidad genómica estimada mediante el modelo BC basado en SNP fue estadísticamente inferior a la heredabilidad basada en pedigrí.

En el caso de la DM de los árboles de *E. cladocalyx*, el modelo BB basado en SNP y los modelos BC basadas en HAP y HAP-SNP tuvieron una mejor bondad de ajuste que el resto de los modelos. El poder predictivo del modelo BB basado en SNP fue levemente superior al poder predictivo de los modelos BC basados en HAP y HAP-SNP. Al igual que en *E. globulus*, ninguna de las estimaciones de heredabilidad genómica fue estadísticamente más alta que la heredabilidad basada en pedigrí para la DM de *E. cladocalyx*. De acuerdo a los valores de DIC, los modelos basados en SNP (exceptuando BA) tuvieron mejor bondad de ajuste que el resto de los modelos para predecir la IF de *E. cladocalyx*. En particular, los modelos basados en los métodos BC y BRR tuvieron un poder predictivo levemente superior al método BB. La heredabilidad genómica estimada por el modelo BC basado en SNP fue casi dos veces más alta que la heredabilidad estimada por el modelo BRR basado en SNP. No obstante, la heredabilidad estimada por el modelo BC basado en SNP fue estadísticamente igual a la heredabilidad basada en pedigrí.

2.4 DISCUSION

2.4.1. Evaluación de la representatividad del arreglo SNP60K, en las poblaciones de *Eucalyptus* en estudio.

La transferibilidad del arreglo de SNP60K en *E. globulus* fue mayor de lo esperado. Anteriormente, Durán et al. (2017) reportaron una transferencia de ~12000 SNPs en una población conformada por 310 clones provenientes de 53 familias de *E. globulus*. Contrario al caso de *E. globulus*, la transferibilidad del arreglo de SNP 60K a *E. cladocalyx* fue relativamente baja (~ sólo el 6 % del total de marcadores). *E. cladocalyx* es una especie relativamente distante, en términos filogenéticos, a otras especies de *Eucalyptus* (Steane et al., 2011). De acuerdo con Brooker (2000) y Boland et al. (2006), *E. cladocalyx* ha sido clasificado en la sección monofilética *Senjuntae*. Debido a que el panel de genotipado usado en este estudio ha sido desarrollado principalmente para *Eucalyptus* spp. de las secciones *Maidenaria*, *Exsertaria* y *Latoangulatae*, es posible que la distancia que existe entre secciones pueda explicar la baja representatividad del arreglo de SNP en *E. cladocalyx*. Si bien el arreglo de SNPs 60K ha sido eficientemente usado en múltiples especies de *Eucalytpus* (Torres-Dini et al., 2016; Klápšťe et al., 2017; Durán et al., 2017; Suontama et al., 2019), algunas especies del género podrían estar pobremente representadas en este panel de marcadores (Aguirre et al., 2019).

Los valores de neutralidad, expresados en D-Tajima, revelaron que ambas poblaciones podrían estar sometidas a una selección positiva o balanceadora, debido a que estos valores fueron positivos (Biswas y Akey, 2006; Lachance y Tishkoff, 2013; Fusari et al., 2017; Li et al., 2017). Particularmente, los valores D-Tajima fueron más altos en la población de *E. globulus* que *E. cladocalyx*. Esto podría ser explicado por el hecho de que la población de *E. globulus* es producto de al menos dos ciclos de selección artificial, de tal manera que la frecuencia de ciertos alelos ha sido favorecida producto de esta selección. En poblaciones domesticadas, es común encontrar valores de D-Tajima relativamente más altos que en poblaciones silvestres, lo cual es un indicador de una reducida diversidad genética (Hufford et al., 2012; Morrell et al., 2013; Wang et al., 2015). De acuerdo con varios autores, poblaciones naturales de *E. cladocalyx* presentan un relativo alto índice de endogamia (McDonald et al., 2003; Bush y Thumma, 2013; Bush et al., 2015). Interesantemente, Bush y Thumma (2013) y Bush et al. (2015) determinaron

que los índices de endogamia son más elevados en poblaciones naturales que poblaciones cultivadas de *E. cladocalyx*.

El parentesco genómico promedio reveló que la mayoría de los individuos de cada población tienden a estar igualmente relacionados. Se debe tener en cuenta que los valores de parentesco son expresados en relación al promedio poblacional. Esto implica que un valor de cero sugiere que los individuos están genéticamente relacionados, muy cercano a lo que se espera para el total de la población (Endelman y Jannik, 2012). Los valores de parentesco genómico entre los individuos de la población de *E. globulus* fueron menos variables que en *E. cladocalyx*. Estos resultados son consistentes con la estructura genética e historial de selección que posee cada población. Los árboles de *E. globulus* provienen de ciclos de selección artificial, lo cual implica que la base genética ha sido reducida, por lo tanto, varios individuos están genéticamente muy relacionados. En relación a la estructura genética, el ensayo de *E. globulus* se compone de árboles que comparten línea materna, y en algunos casos, comparten la línea materna y paterna. En este contexto, se espera que varios individuos estén altamente relacionados. En el caso de *E. cladocalyx*, los árboles provienen de diferentes puntos de la distribución natural de la especie, lo cual implica que existe una mayor probabilidad de encontrar individuos que estén escasamente relacionados. Adicionalmente, los árboles de *E. cladocalyx* provienen de familias de polinización abierta, por lo tanto, aquellos individuos que comparten una misma línea materna no necesariamente comparten una misma línea paterna. Interesantemente, el patrón de los valores de parentesco genómico de los individuos de *E. cladocalyx* es encontrado en poblaciones genéticamente diferenciadas (Cappa et al., 2013; Raggi et al., 2019; Sallam et al., 2020), lo cual ha sido previamente documentado en *E. cladocalyx* (Mora et al., 2017; Arriagada et al., 2018). En este contexto, la matriz de identidad por descendencia separa individuos estrechamente relacionados (valores extremos positivos), e individuos genéticamente muy distantes (valores extremos negativos). De acuerdo Goudet et al. (2018), los valores negativos de parentesco representan aquellos individuos que tienen un valor de parentesco menor al esperado para la población, lo cual ocurre entre individuos que pertenecen a diferentes grupos genéticos.

2.4.2. Determinación y caracterización de regiones genómicas en desequilibrio de ligamiento en los dos ensayos de progenie de *Eucalyptus*

En el presente estudio, cerca del 24% y 2% de los bloques de haplotipos formados tuvieron una extensión sobre los 10000 y 100000 pb, respectivamente, con un valor de desequilibrio D' sobre 0,7 para la población de *E. globulus*. Basado en el hecho de que los bloques de haplotipos fueron construidos bajo el método de Gabriel et al. (2002), es posible suponer que más de 1300 bloques de haplotipos tuvo una alta probabilidad de ser idénticos por descendencia, y, por lo tanto, son heredados en forma transgeneracional. Un menor número de bloques de haplotipos fueron construidos para *E. cladocalyx*, lo cual podría ser explicado por la relativa baja densidad de marcadores detectados en esta población. El número de bloques de haplotipos conformados para *E. globulus* fue hasta 4 veces mayor que los bloques detectados en *E. cladocalyx*. Más aun, cerca del 80% de los bloques de haplotipos fueron construidos con un máximo de dos SNPs para esta misma especie. En el caso de *E. globulus*, se detectaron bloques de haplotipos conformados por hasta 12 SNPs (Chr11) en desequilibrio de ligamiento.

Los patrones de DL en ambas poblaciones de *Eucalyptus*, reflejaron que el DL disminuye dentro de distancias más cortas en la población de *E. cladocalyx* que en la población de *E. globulus* (difiriendo en ~3Kp), lo cual es esperado debido al historial de selección que tiene cada una de las poblaciones. En general, las poblaciones naturales de especies árboreas exhiben valores de DL entre sus loci relativamente más bajos que poblaciones mejoradas genéticamente (Olson et al., 2010; Thavamanikumar et al., 2011; Kelleher et al., 2012; Guerra et al., 2012; Larsson et al., 2013; Lu et al., 2016; Müller et al., 2017; Durán et al., 2017). En el caso de *Eucalyptus*, en poblaciones de mejoramiento el DL puede disminuir dentro de los 3000 pb y los 25000 pb (Müller et al., 2017; Durán et al., 2017; Suontama et al., 2019), mientras que en poblaciones que no han sido sometidas a selección puede encontrarse un patrón de desequilibrio que disminuye dentro de los 500 pb (Thavamanikumar et al., 2011).

El conocimiento de los patrones de DL permite determinar no sólo el mínimo de marcadores que son requeridos para estudios genómicos de GS, sino que también permite inferir cambios históricos ocurridos en una población (Karimi et al., 2020). Habitualmente, las

poblaciones cultivadas o mejoradas poseen mayores niveles de DL que las poblaciones silvestres, lo cual es consecuencia de la endogamia, adaptación y selección artificial (Noble et al. 2018). Además de los eventos históricos (selección artificial, estructuración, endogamia, entre otros) que han ocurrido en las poblaciones, los mecanismos reproductivos de cada especie también están estrechamente relacionados con los patrones de DL. Como la mayoría de los eucaliptos, *E. globulus* tiene un sistema de apareamiento mixto. En general, la tasa de polinización cruzada en rodales nativos varía entre el 38 y el 100% (Hardner et al., 1996; Patterson et al., 2001; Foster et al., 2007; Potts et al., 2008). De acuerdo con Pound et al. (2002, 2003), *E. globulus* posee un mecanismo de autoincompatibilidad post-cigoto tardía y es común que los eventos de autopolinización generen una menor cantidad de semillas. En este sentido, la extensión del DL y los valores de neutralidad probablemente son reflejo de la selección artificial en lugar de los mecanismos reproductivos de la especie. La ecología reproductiva de *E. cladocalyx* ha sido escasamente estudiada. De acuerdo con Ellis y Sedley (1992), *E. cladocalyx* también cuenta con un mecanicismo de auto-incompatibilidad post-cigótico, por lo tanto, el patrón de DL y los valores de neutralidad en la población estudiada podrían ser consecuencia del aislamiento geográfico entre sus poblaciones, de tal manera que los árboles son polinizados por otros árboles que son genéticamente cercanos (McDonald et al., 2003; Ballesta et al., 2015).

2.4.3 Estimación de la heredabilidad genómica de características fenotípicas y evaluación de los modelos de predicción basados en haplotipos y polimorfismos de nucleótido único

De acuerdo a los resultados de las estimaciones de heredabilidad basadas en pedigrí, las características estudiadas en *E. globulus y E. cladocalyx* fueron de baja a altamente heredables. Interesantemente, los resultados revelaron que los modelos basados en HAP o HAP-SNP tuvieron un mayor poder predictivo que los modelos basados en SNP para la predicción de características con una relativa baja heredabilidad en *E. globulus* (esto es, CR, DAP y RF). No obstante, el poder predictivo de los modelos basados en HAP o HAP-SNP no fueron los más adecuados para predecir las características estudiadas en *E. cladocalyx,* exceptuando a la rectitud del fuste. No obstante, el modelo con mejor bondad de ajuste para esta característica (BB y BC basada en HAP-SNP) capturó una variabilidad genética (en términos de heredabilidad) equivalente al enfoque basado en pedigrí. De acuerdo con Calus et al. (2009), así

como la densidad de marcadores determina la precisión de un modelo GS, el número y tamaño de los haplotipos tiene un efecto sobre el poder predictivo de un modelo. En este contexto, es posible suponer que la baja disponibilidad de regiones detectadas en DL tenga un efecto sobre el poder de los haplotipos para predecir las características estudiadas en *E. cladocalyx*.

En el caso de *E. cladocalyx*, se detectó que el diámetro a la altura de pecho (DAP) tuvo una heredabilidad moderada, lo cual es consistente con estudios previos para poblaciones de *Eucalyptus* de 3 a 15 años de edad (Vargas-Reeve et al. 2013; Sountama et al. 2015; Hung et al. 2015). No obstante, los resultados sugieren que la heredabilidad del DAP en la población de *E. globulus* tiene un control genético aditivo relativamente bajo (medido en términos de pedigrí). Estos resultados son razonables debido a que la población de *E. globulus* es producto de un programa de selección artificial basada en el volumen de los árboles, el cual está altamente correlacionado con el DAP en *Eucalyptus*. Esto implica que la variabilidad para esta característica ha sido reducida de tal manera que alelos favorables para el DAP se han mantenido en la población. Para esta misma característica, el modelo de predicción con mejor bondad de ajuste y poder predictivo en *E. globulus* fue el modelo BC basado en HAP-SNP. Adicionalmente, este modelo capturó una mayor proporción de la variabilidad genética, en términos de heredabilidad, que las estimaciones basadas en pedigrí. Previamente, Tan et al. (2017) reportaron que la estimación de heredabilidad genómica del DAP de híbridos de *Eucalyptus* de 6 años de edad (al igual que la población estudiada de *E. globulus*), basada en los métodos Ridge Regression Best Linear Unbiased Prediction y Reproducing Kernel Hilbert Spaces, fue más alta que la estimación basada en pedigrí. Cabe destacar que esta comparación debe ser vista con precaución debido a que los modelos genómicos y basados en pedigrí poseen diferentes supuestos analíticos. Los supuestos analíticos detrás del método de predicción BC establecen que todas las variables predictoras (marcadores) poseen una varianza común y que ciertas variables pueden tener un efecto igual a cero, de tal manera que sólo las variables con efectos diferente de cero son consideradas para predecir (Habier et al., 2011; Pérez et al., 2014). En una población de *E. pellita*, Müller et al., (2017) determinaron que cinco modelos de predicción genómica (BA, BB, BC, BRR, BLASSO y GBLUP), basados en SNP, fueron capaces de capturar una mayor variabilidad del DAP, que el enfoque de pedigrí. Consistente con los resultados del presente estudio, los modelos BC y BB estimaron una mayor heredabilidad

genómica que el resto de los modelos. Similarmente a *E. globulus*, el modelo BC basado en SNP fue el modelo con mejor bondad de ajuste y poder predictivo para el DAP de *E. cladocalyx*. Cabe destacar que la heredabilidad genómica estimada por este modelo fue sutilmente más alta que la estimación de heredabilidad basada en pedigrí. En contraste con los resultados del presente estudio, Suontama et al. (2019) y Resende et al. (2012a) reportaron que las estimaciones de heredabilidad genómica podrían ser más bajas que aquellas basadas en pedigrí en *E. nitens* e híbridos del mismo género, respectivamente. Consistente con los estudios de estos autores, Müller et al. (2017) reportaron que la heredabilidad genómica del DAP en una población de *E. benthamii* fue relativamente más baja que la heredabilidad estimada mediante antecedentes de pedigrí. En este mismo estudio, Müller et al. (2017) desarrollaron modelos de predicción genómica para el DAP en dos poblaciones de *Eucalyptus*, con diferente historial de selección. Consistente con los resultados del presente trabajo, la población que fue producto de ciclos de selección artificial (*E. benthamii*) exhibió un valor de heredabilidad fenotípica menor a la población con un historial de selección nula (*E. pellita*). En adición, la heredabilidad genómica promedio fue más alta que la estimación de heredabilidad basada en pedigrí para la población de *E. benthamii* Contrariamente, la heredabilidad genómica promedio del DAP fue más baja que la estimación basada en el pedigrí en la población de *E. pellita*. En este sentido, los métodos de selección genómica serían especialmente beneficiosos para capturar componentes heredables que no puede ser estimados por la predicción basada en pedigrí, en el contexto de una población que ha ha reducido su variabilidad genética producto de los ciclos de selección artificial.

La heredabilidad de la ALT en la población de *E. globulus* tuvo un valor moderado (Ballesta et al., 2018; Suontama et al., 2015; Mora et al., 2019), mientras que esta misma característica fue mayormente heredable en la población de *E. cladocalyx*. Similarmente al DAP, la ALT también está altamente correlacionada con el volumen de madera, por lo tanto, también se espera que la base genética de ALT esté más reducida en la población de *E. globulus* que en la población de *E. cladocalyx*. El modelo BC basado en SNP fue el modelo con mejor bondad de ajuste y poder predictivo para ALT de *E. globulus*. Consistentemente, el modelo BB basado en SNP fue el más adecuado, en términos de bondad de ajuste y poder predictivo, para predecir la ALT en *E. cladocalyx*. Interesantemente, Müller et al. (2017) determinaron que los

modelos BB y BC basados en SNPs también capturarían una mayor variabilidad genética que los modelos BRR y BA de la ALT de *E. pellita* y *E. benthamii*.

La RF ha sido previamente reportado como una característica con un control genético aditivo relativamente moderado o bajo en otras poblaciones de *E. globulus* (Lopez et al., 2002; Ballesta et al. 2018; Mora et al. 2019), lo cual es consistente con los hallazgos del presente estudio. No obstante, otros estudios han reportado que la heredabilidad de RF de *E. globulus* podría ser eventualmente de moderada a alta (Callister et al., 2011; Blackburn et al., 2013). En el caso de *E. cladocalyx*, Vargas-Reeve et al. (2013) reportaron la RF puede ser de modera a altamente heredable, lo cual es consistente con los resultados del presente estudio. El modelo BRR basado en HAP tuvo el mayor poder predictivo y bondad de ajuste para predecir la RF en *E. globulus*. En la literatura, existen contradicciones en torno a la arquitectura genética que subyace a la rectitud de los fustes en árboles. Por ejemplo, Bartholomé et al. (2016) reportaron QTLs explicando hasta un 5% de la variación fenotípica del fuste de pino marítimo, mientras que Yang et al. (2015) detectaron regiones genómicas explicando hasta un 15% de la variación de la rectitud de fuste en híbridos de *Pinus*. Adicionalmente, Arriagada et al. (2018) reportaron 5 regiones genómicas asociadas a la rectitud de fuste de *E. cladocalyx*, las cuales estarían explicando entre un 6 y un 14% de la variación fenotípica de esta característica. En este contexto, aun no es bien entendido si unos pocos genes de efecto mayor, o bien una herencia poligénica estaría involucrada en el control genético de la rectitud de fuste en árboles. En el caso de *E. cladocalyx*, los modelos basados en HAP-SNP tuvieron mejor bondad de ajuste y poder predictivo para RF que el resto de los modelos genómicos. No obstante, ninguno de estos modelos estimó una heredabilidad genómica superior al valor de heredabilidad basado en pedigrí. Consistente con estos resultados, Suontama et al. (2019) reportaron que la heredabilidad basada en información de pedigrí de la rectitud de fuste en una población de *E. nitens* fue casi dos veces más alta que aquella estimada mediante marcadores moleculares.

La calidad de ramas (CR) es una característica fenotípica escasamente estudiada y explotada en los programas de mejoramiento de *E. globulus*. Por ejemplo, Callister et al. (2011) y Suontama et al. (2015) reportaron que el grosor de ramas y la calidad de ramas son características con un relativo bajo control genético aditivo en árboles de *E. globulus* y *E.*

regnans de 4 y 3 años de edad, respectivamente (en predicciones basadas en pedigrí). Ballesta et al. (2018) reportaron resultados similares a los anteriores. Los autores informaron que la heredabilidad de la calidad de ramas (medida equivalente al presente estudio) fue menor a un valor de 0,1 en una población de *E. globulus* de 4 años de edad. Por otro lado, en otras especies de árboles, características como la calidad de ramas han sido descritas como moderadamente heredables (Monclus et al., 2012; Song et al., 2017). Nuevamente, es importante destacar que la calidad de ramas es una característica objetivo en los programas de mejoramiento implementados en la población de *E. globulus*, por lo tanto, el bajo control genético aditivo observado puede ser consecuencia de los ciclos de selección. El modelo con mejor bondad de ajuste y con mayor poder predictivo para CR fue el modelo BC basado en HAP-SNP. Adicionalmente, la heredabilidad genómica estimada por este modelo tuvo un valor más alto que la estimación basada en pedigrí. Al igual que el DAP y RF de E. *globulus*, los modelos GS que incluyen haplotipos (ya sea HAP o HAP-SNP) tuvieron una mayor precisión predictiva de la CR, la cual tuvo un valor de heredabilidad basada en pedigrírelativamente bajo. De acuerdo con varios autores, el enfoque de haplotipos tiene una precisión predictiva más alta que los SNP para características con una relativa baja heredabilidad en plantas (Villumisen et al., 2009; Matias et al., 2017; Lan et al., 2020). Estos hallazgos pueden ser explicados por el hecho de que el uso de haplotipos en GS permite acceder a componentes genéticos heredables que no pueden ser examinados por los antecedentes genealógicos, e incluso por los SNPs (Villumisen et al., 2009; Matias et al., 2017).

Varios estudios han establecido que la densidad de madera (DM) es una característica que posee una heredabilidad relativamente alta en *Eucalyptus* y otros árboles (Bartholomé et al., 2016; Suontama et al., 2019; Valenzuela et al., 2019), lo cual es consistente con los resultados del presente estudio para las predicciones basadas en pedigrí en ambas especies estudiadas. Cabe destacar que, tanto para *E. globulus* como *E. cladcalyx*, ninguna de las estimaciones de heredabilidad genómica fue más alta que la heredabilidad basada en la información de pedigrí. En otros estudios, Suontama et al. (2019) reportaron que las estimaciones de heredabilidad de la densidad de madera basadas en pedigrí y datos genómicos poseen valores similares en un ensayo de árboles de *E. nitens*, un aspecto también reportado en el presente estudio para la población estudiada de *E. cladocalyx* (basado en el modelo con mejor bondad de ajuste y poder

de predicción). Consistente con los hallazgos para *E. globulus*, Resende et al. (2012a) informaron que la heredabilidad genómica de la densidad de madera posee un valor relativamente más bajo que la estimación de heredabilidad basada en datos de pedigrí en híbridos de *Eucalyptus*. En otras especies de árboles, Beaulieu et al. (2014) reportaron que la heredabilidad genómica es relativamente más baja que la heredabilidad en sentido estricto de la densidad de madera (basada en pedigrí) de árboles de *Picea glauca*. Beaulieu et al. (2014) propusieron que una de las razones por las cuales la estimación de la heredabilidad basada en pedigrí podría tener un valor más alto que la heredabilidad genómica, es que la información dada por los antecedentes de pedigrí podría capturar el efecto de loci que no han sido considerados en el modelo de predicción genómica, los cuales podrían ser relevantes para el control genético de la densidad de madera en árboles.

La IF de *E. cladocalyx* ha sido previamente reportada como una característica altamente heredable (Cané-Retamales et al., 2011), lo cual es consistente con los hallazgos del presente estudio. En general, los modelos de predicción genómica tuvieron un menor poder predictivo para la IF de *E. cladocalyx* que el resto de las características estudiadas (PP ≤0,1). Adicionalmente, la mayoría de las estimaciones de heredabilidad basadas en datos genómicos tuvieron un valor más bajo que las estimaciones basadas en pedigrí. Hall et al. (2016) han propuesto que características fenológicas en los árboles usualmente son altamente heredables y controladas por regiones puntuales en sus genomas. En este contexto, la baja densidad de marcadores SNP detectados en el presente estudio podría estar limitando la probabilidad de encontrar regiones genómicas claves que controlan la variación de IF. Por esto mismo, los efectos poligénicos (basado en el pedigrí) parecieran ser más convenientes de considerar para predecir la IF.

De acuerdo con Callister et al (2011) y Bush et al. (2015), los patrones de bifurcación y la persistencia del eje central de los árboles, características relacionadas a APB, posee un heredabilidad relativamente baja o moderada en *E. cladocalyx*, lo cual es consistente con los resultados del presente trabajo. El modelo BC basado en SNP tuvo la mejor bondad de ajuste y precisión predictiva de la APB. En este contexto, al igual que IF, los efectos poligénicos podrían ser relevantes para predecir APB en esta población de *E. cladocalyx*. Además, una mejor

cobertura a nivel genómico podría ser necesitada para disponer de marcadores que se encuentren en DL con regiones claves que controlan la variación de APB. Consistentemente, Arriagada et al. (2018) detectaron un QTL de efecto mayor explicando hasta un 12% de la variación fenotípica de la bifurcación de árboles de *E. cladocalyx*.

La precisión en la estimación de parámetros genéticos y el poder predictivo de un modelo depende de varios factores, tales como el número de variables predictoras, la extensión del desequilibrio de ligamiento y el grado de desequilibrio de los marcadores con QTLs, la proporción y relación entre las poblaciones de entrenamiento y validación, entre otros (Wientjes et al., 2013; Lorenz y Smith, 2015; Wang et al., 2017; Zhang et al., 2017). En el contexto de este estudio, el poder predictivo (promedio) para las características fenotípicas evaluadas en *E. globulus* fue relativamente más alto que el poder predictivo para *E. cladocalyx*. Una de las principales razones propuestas para explicar este resultado es la menor densidad de marcadores SNP detectada para *E. cladocalyx*. No obstante, de acuerdo con Arriagada et al. (2018) y Mora et al. (2017), la población estudiada de *E. cladocalyx* se encuentra significativamente estructurada en tres sub-poblaciones. Adicionalmente, la estructura genética basada en el parentesco genómico (mediante haplotipos) y el parentesco teórico en ambas poblaciones de *Eucalytpus*, reveló que los individuos de *E. cladocalyx* están genéticamente menos relacionados que los individuos que componen a la población de *E. globulus*. Tan et al. (2017) reportaron que el poder predictivo para características de crecimiento en *Eucalyptus* puede ser incrementado si la población de entrenamiento esta genéticamente relacionada a la población de validación. En este estudio, los investigadores entrenaron el modelo de predicción utilizando la información genómica de individuos que pertenecen a un mismo grupo genéticamente homogéneo al de la población de validación. Cuando los autores entrenaron el modelo de predicción utilizando individuos genéticamente distantes a la población de validación, el poder predictivo disminuyó sustancialmente. Cabe destacar que el poder de los modelos basados en SNP y/o haplotipos para predecir características relacionadas al crecimiento de *E. cladocalyx* (Esto es, altura y diámetro de árboles) fue equivalente a otros estudios en *Eucalyptus,* a pesar de contar con una menor densidad de marcadores (Tan et al., 2017; Müller et al., 2017), no obstante, para otras características fue extremadamente bajo (Por ejemplo, IF).

.

2.5 CONCLUSION

El presente estudio es el primero en examinar la inclusión de haplotipos en modelos de selección genómica de árboles forestales. Los modelos genómicos que incluyeron el efecto de haplotipo (ya sea HAP o HAP-SNP) aumentaron significativamente el poder predictivo para características de baja heredabilidad. Además, los haplotipos permitieron estimar una mayor proporción de los componentes heredables (heredabilidad genómica) de características de baja heredabilidad, que los antecedentes de pedigrí. En este sentido, el enfoque de haplotipos podría ser especialmente beneficioso para predecir características de baja heredabilidad de *Eucalyptus*, en un escenario donde la base genética de estas características ha sido reducida debido a los procesos de selección artificial. En general, e independiente del enfoque de predicción genómica, el poder predictivo para las características fenotípicas de *E. cladocalyx* fue menor que el poder predictivo para características evaluadas en *E. globulus*. Cabe destacar, que estos resultados pueden ser explicados por el hecho de que la representatividad del arreglo SNP60K en *E. cladocalyx* fue sustancialmente menor que en *E. globulus*. Adicionalmente, de acuerdo con los antecedentes de historial de selección y estructura genética, las diferencias en términos de poder predictivo, entre cada población, también podría ser explicada por el hecho de que los individuos genotipados de la población de *E. globulus* son genéticamente más cercanos que los individuos de la población de *E. cladocalyx*.

Los resultados de este estudio proporcionan perspectivas adicionales para la implementación de la selección genómica en los programas de mejoramiento de *Eucalyptus*, lo cual podría ser especialmente beneficioso para mejorar caracteres de baja heredabilidad en el caso de *E. globulus*.

3 CAPITULO III: Predicción genómica mediante el uso de una baja densidad de marcadores e información de pedigrí en *Eucalyptus cladocalyx*

3.1 INTRODUCCION

Uno de los principales enfoques de la genética cuantitativa moderna es poder evaluar la asociación entre marcadores polimórficos y variaciones fenotípicas en características complejas. Los estudios de asociación entre genotipo y fenotipo generalmente requieren de paneles de marcadores genéticos de alta densidad, es decir, un gran número de marcadores distribuidos a lo largo de todo el genoma, y grandes tamaños de población para obtener suficiente precisión en la predicción (Viana et al., 2018; Singh et al., 2019). El desarrollo de diferentes plataformas de genotipado, tales como los arreglos de polimorfismo de nucleótido único (SNPs) o el genotipado por secuenciación (GBS), ha permitido la identificación de loci de características cuantitativas (QTL) para diferentes caracteres, los cuales son objetivo en los programas de mejoramiento de varias especies vegetales (Contreras-Soto et al., 2017; Maldonado et al., 2019; Senhorinho et al., 2019; Mafra et al., 2019). Silva-Junior et al (2015), por ejemplo, desarrollaron un arreglo de SNPs 60K para múltiples especies de *Eucalytpus*, el cual ha sido eficientemente usado en estudios genómicos de una amplia variedad de especies económicamente importantes y sus híbridos, incluidos *E. grandis*, *E. urophylla*, *E. nitens* y *E. globulus* (Torres-Dini et al., 2016; Klápšťe et al., 2017; Durán et al., 2017; Suontama et al., 2019). Sin embargo, algunas especies del género podrían estar pobremente representadas en este arreglo de SNPs (Aguirre et al., 2019). En el Capítulo II del presente trabajo, la transferibilidad del arreglo de SNP 60K a *E. cladocalyx* fue relativamente baja (~ sólo el 6 % del total de marcadores). *E. cladocalyx* es una especie relativamente distante, en términos filogenéticos, a otras especies de *Eucalyptus* (Steane et al., 2011). De acuerdo con Brooker (2000) y Boland et al. (2006), *E. cladocalyx* ha sido clasificado en la sección monofilética *Senjuntae*. Debido a que el panel de genotipado usado en el Capítulo II ha sido desarrollado principalmente para *Eucalyptus* spp. de las secciones *Maidenaria*, *Exsertaria* y *Latoangulatae,* es posible que la distancia que existe entre secciones pueda explicar estos resultados.

De acuerdo con Pryce et al. (2014), el uso de paneles marcadores de baja densidad puede afectar la precisión en la detección de QTL en estudios de asociación y de predicción genómica. Por otro lado, Müller et al. (2017) encontraron que los modelos de predicción basados en una baja densidad de marcadores (un subconjunto de ~ 5000 SNPs) proporcionaron capacidades predictivas casi equivalentes a modelos basados en el uso de más de 12000 SNPs en *Eucalyptus* spp. Sin embargo, concluyeron que aún no está claro si el uso de un subconjunto de SNPs más pequeño podría garantizar una implementación a largo plazo de selección genómica en *Eucalyptus*. Cabe destacar que, los paneles de marcadores podrían considerarse de baja o alta densidad, dependiendo del tamaño del genoma, la extensión del desequilibrio de ligamiento (DL) y las características estudiadas. Para genomas relativamente grandes, y con una disminución rápida del desequilibrio de ligamiento (DL), se requeriría disponer de una mayor densidad de SNPs para detectar un QTL. Varios autores describen una alta o baja densidad de marcadores en forma arbitraria, de modo que las comparaciones entre organismos deben hacerse con precaución. En el ganado bovino (tamaño del genoma de 3 Gb), por ejemplo, un arreglo de SNPs de 50 K puede considerarse como un panel de baja densidad (Pryce et al., 2014), mientras que en *E. nitens* (640 Mb), un chip de 60 K ha sido considerado como de alta densidad por Suontama et al. (2019). El DL en el genoma bovino domesticado disminuye rápidamente (Porto-Neto et al., 2014), el cual es más extenso que en poblaciones naturales de *Eucalyptus*, por lo tanto, la comparación en cuanto a densidad de marcadores sigue siendo compleja.

En el mejoramiento genético animal, el uso de paneles de SNPs de baja densidad es considerado como una alternativa para reducir el costo de los paneles SNP de alta densidad, lo cual permite que los estudios genómicos sean rentables (Bolormaa et al., 2015; Wu et al., 2016). En este contexto, recientemente Silva et al. (2018) implementaron el enfoque denominado "LALDA", el cual combina los beneficios de los análisis de ligamiento (Linkage Analysis; LA) y de desequilibrio de ligamiento (Linkage Disequilibrium Analysis; LDA) para fines de predicción genómica con una baja disponibilidad de marcadores. En esta experiencia, los autores demostraron una ligera superioridad de LALDA por sobre los modelos genómicos tradicionales. Los autores plantearon que estos resultados podrían ser explicados por el hecho de que LALDA permite explorar regiones genómicas relevantes que no son consideradas en el enfoque tradicional de predicción genómica.

En el Capítulo II, el poder predictivo promedio de los modelos basados en SNP, HAP y/o HAP-SNP fue extremadamente bajo para algunas características (Por ejemplo, IF ≤0,1). Adicionalmente, ninguno de los modelos con mejor bondad y poder predictivo capturó mayor variabilidad genética que los antecedentes de pedigrí, a excepción del modelo BC basado en SNP (para DAP). Teniendo en cuenta estos antecedentes, el presente capítulo tuvo como objetivo evaluar si el uso combinado de información genómica y de la estructura genética (teórica) de población, permite incrementar el poder predictivo de características complejas en *E. cladocalyx*.

3.2 MATERIALES Y MÉTODOS

3.2.1 Población de estudio

La población de *Eucalyptus cladocalyx* F. Muell utilizada en el presente estudio ha sido previamente descrita en el Capítulo II, sección 2.2.3. El método de genotipado fue descrito en esta misma sección. Las características fenotípicas evaluadas fueron previamente descritas en el Capítulo II, sección 2.2.4.

3.2.2 Modelos de predicción basados en SNPs e información de pedigrí

Se utilizaron cuatro métodos de predicción para estimar los efectos de los marcadores SNP, los componentes de varianza y la heredabilidad genómica de las características previamente examinadas en *E. cladocalyx*: 1) Bayes A, Bayes B, Bayes C y la regresión contraída bayesiana (BRR). Debido a la baja densidad de marcadores, se probaron tres enfoques de predicción: 1) Predicción genómica tradicional (GS), 2) Predicción genómica + información de pedigrí (GSp), y 3) Predicción genómica con previa selección de variables + información de pedigrí (GSq). El modelo GSp fue definido por la siguiente fórmula:

$$\mathbf{y}^* = \mathbf{1}\mu + \sum_{i=1}^{m} \mathbf{x}_i \mathbf{m}_i + \mathbf{Za} + \boldsymbol{\varepsilon} \tag{11}$$

donde, $\mathbf{y}^*$ corresponde al valor fenotípico de cada árbol ajustado por los efectos fijos de bloque y estructura genética (Tan et al., 2018). $\mathbf{1}$ y $\boldsymbol{\mu}$ corresponden a un vector de unos y la media

general del modelo, respectivamente. $\mathbf{m_i}$ corresponde al efecto aditivo de i-ésimo SNP, con m como el número de marcadores. $\mathbf{x_i}$ es el vector de incidencia de cada marcador, el cual fue codificado como AA = 0, AB = 1 y BB = 2. La varianza de $\mathbf{m}$ Var ($\mathbf{m}$) depende del método de predicción implementado en cada modelo Bayesiano. $\mathbf{a}$ corresponde a los efectos poligénicos (aditivos basados en pedigrí), el cual $\mathbf{a}$~N(0, σ_a^2A), y Z es la matriz de incidencia relacionada a los efectos poligénicos. Para estimar los efectos poligénicos, los coeficientes de parentesco entre individuos fueron establecidos de acuerdo con Bush et al. (2015). Finalmente, $\mathbf{\varepsilon}$ corresponde al vector de residuos, el cual $\mathbf{\varepsilon}$~N(0, σ_e^2I).

El enfoque GSq consistió de dos pasos. En un primer paso, se evaluó si cada uno de los 3879 SNPs tuvo un efecto significativo en la variación de cada una de las características estudiadas (p<0,001), basado en un estudio de asociación marcador-característica (Yu et al., 2006), de acuerdo a la siguiente expresión:

$$\mathbf{y}^* = \mathbf{Sa} + \mathbf{Zu} + \mathbf{\varepsilon} \tag{12}$$

donde, $\mathbf{y}^*$ corresponde al valor fenotípico de cada árbol ajustado por los efectos fijos de bloque y estructura genética. $\mathbf{S}$ y $\mathbf{Z}$ corresponden a las matrices de incidencia de los vectores $\mathbf{a}$ y $\mathbf{u}$, respectivamente. $\mathbf{a}$, $\mathbf{u}$ y $\mathbf{\varepsilon}$ corresponden a los vectores de los efectos de los SNPs (de efecto fijo), de los efectos poligénicos (de efecto aleatorio) y de residuos del modelo, respectivamente. Las varianzas de $\mathbf{u}$ y $\mathbf{\varepsilon}$ son calculadas como Var ($\mathbf{u}$) =$\mathbf{2K}\sigma_g^2$ y Var ($\mathbf{\varepsilon}$) =$\mathbf{R}\sigma_e^2$, respectivamente, donde $\mathbf{K}$ corresponde a una matriz de coeficientes de parentesco genómico, calculados mediante los marcadores SNP en el programa TASSEL versión 5.2 (Bradbury et al., 2007), de acuerdo a la metodología propuesta por Endelman y Jannik (2012).

En un segundo paso, teniendo identificado aquellos SNPs significativos de acuerdo a la expresión 12, esta información fue incluida en el modelo GSp (expresión 11), resultando en la siguiente expresión para el modelo GSq:

$$\mathbf{y}^* = \mathbf{1\mu} + \sum_{i=1}^{m} \mathbf{x_i m_i} + \mathbf{Za} + \mathbf{Zq} + \mathbf{\varepsilon} \tag{13}$$

donde, $\mathbf{y}^*$ corresponde al valor fenotípico de cada árbol ajustado por los efectos fijos de bloque y estructura genética. $\mathbf{1}$ y $\mathbf{\mu}$ corresponden a un vector de unos y la media general del modelo, respectivamente. En el caso del modelo GSq, $\mathbf{m_i}$ corresponde al efecto aditivo de i-ésimo SNP,

el cual no fue significativo de acuerdo a la expresión 12, y $\mathbf{x_i}$ es el vector de incidencia de cada uno de estos marcadores, los cuales fueron codificado como AA = 0, AB = 1 y BB = 2. La varianza de $\mathbf{m}$ Var ($\mathbf{m}$) depende del método de predicción implementado en cada modelo (BA, BB, BC o BRR). $\mathbf{Z}$ es la matriz de incidencia asociada a los efectos aditivos poligénicos ($\mathbf{a}$) y de los SNP significativos ($\mathbf{q}$) de acuerdo a la expresión 12, asumiendo $\mathbf{a}$~N(0, $\sigma_a^2\mathbf{A}$) y $\mathbf{q}$~N(0, $\sigma_q^2\mathbf{Q}$). $\mathbf{Q}$ corresponde a una matriz de covarianza, en la cual los elementos corresponden a probabilidades de identidad por descendencia de los marcadores significativos para la expresión (12), de acuerdo a Endelman y Jannik (2012).

Las predicciones genómicas fueron realizadas en el paquete de BGLR de R (Pérez y De Los Campos, 2014), utilizando 1000000 de cadenas de Markov Monte Carlo con un periodo de quema de 100000 cadenas.

3.2.3 Estimaciones de heredabilidad

Para el modelo GS, la heredabilidad genómica fue calculada de acuerdo a la siguiente fórmula:

$$h_m^2 = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_e^2} \tag{14}$$

donde σ_m^2 y σ_e^2 corresponden a la varianza del efecto de todos los marcadores SNP (n=3879) y a la varianza residual del modelo, respectivamente. En el caso del modelo GSp, se estimó la heredabilidad basada en información de pedigrí y marcadores SNP:

$$h_m^2 = \frac{\sigma_m^2}{\sigma_a^2 + \sigma_m^2 + \sigma_e^2} \tag{15}$$

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_m^2 + \sigma_e^2} \tag{16}$$

donde, h_m^2 y h_a^2 corresponden a la heredabilidad genómica basada en los 3879 SNPs y en información del pedigrí de los árboles, respectivamente. Los términos σ_m^2, σ_a^2 y σ_e^2 corresponden a las varianzas del efecto de todos los marcadores SNP, de los efectos poligénicos aditivos (basados en pedigrí), y de los residuos del modelo, respectivamente. En el modelo GSq, se calcularon los siguientes valores de heredabilidad:

$$h_m^2 = \frac{\sigma_m^2}{\sigma_a^2 + \sigma_m^2 + \sigma_q^2 + \sigma_e^2} \tag{17}$$

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_m^2 + \sigma_q^2 + \sigma_e^2} \tag{18}$$

$$h_q^2 = \frac{\sigma_q^2}{\sigma_a^2 + \sigma_m^2 + \sigma_q^2 + \sigma_e^2} \tag{19}$$

donde, h_m^2 corresponde a la heredabilidad genómica basada en todos aquellos SNPs que no fueron significativos de acuerdo a la expresión 12. El término h_q^2 corresponde a la heredabilidad genómica de los marcadores SNP detectados como significativos de acuerdo a la expresión 12, y h_a^2 corresponde a la heredabilidad dada por los efectos poligénicos aditivos (basados en pedigrí). σ_m^2 y σ_q^2 correspoden a las varianzas de los marcadores no significativos y significativos para la expresión 12, respectivamente. Tanto para los modelos GS, GSp y GSq, las varianzas de marcadores fueron calculadas dependiendo del método bayesiano implementado. En el caso de los métodos BC y BRR, la varianza total de todos los marcadores fue calculada como (Beaulieu et al., 2014; Silva et al., 2018):

$$\sigma_m^2 = 2\sigma_{SNP}^2 \sum_{i=1}^{n} p_i(1 - p_i) \tag{20}$$

donde σ_{SNP}^2 corresponde a una varianza común para todos los marcadores SNP y p_i corresponde a la frecuencia del alelo de menor presencia en la población de estudio, del i-ésimo marcador. Los métodos de BA y BB asumen heterogeneidad en las varianzas de los marcadores, por lo tanto, la varianza total de los marcadores SNP fue calculada como:

$$\sigma_m^2 = 2 \sum_{i=1}^{n} p_i(1 - p_i)\sigma_{SNP_i}^2 \tag{21}$$

donde $\sigma_{SNP_i}^2$ corresponde a la varianza del i-ésimo marcador SNP.

3.2.4 Evaluación de la bondad de ajuste de los modelos

Este análisis fue realizado para determinar el verdadero número de parámetros que son necesarios para predecir y obtener un buen ajuste del modelo de predicción. Para cada característica fenotípica y método, se compararon los valores de bondad de ajuste de cada

modelo (GS, GSp y GSq) en términos de Criterio de Información de Desviación (Deviance Criterion Information; DIC). Una diferencia de 10 entre los valores de DIC para cada modelo se consideró como una fuerte evidencia de que existen diferencias en términos de bondad de ajuste. Una diferencia de 5 entre los valores de DIC se consideró como una diferencia significativa entre modelos, y un valor < 5 se consideró que no existen diferencias entre los modelos, en términos de bondad de ajuste.

3.2.5 Poder predictivo de los modelos de predicción

El poder predictivo de cada modelo fue calculado como el valor de correlación de Pearson entre las observaciones fenotípicas (previamente ajustadas por los efectos fijos; $\mathbf{y}^*$) de un conjunto de validación de individuos y los valores predichos por el modelo ($\hat{\mathbf{y}}^*$). Un total de 20 subconjuntos de validación cruzada fueron usados para obtener los valores de poder predictivo de todos los modelos. Los valores de $\hat{\mathbf{y}}^*$ para los modelos GS, GSp y GSq fueron calculados como:

$$\hat{\mathbf{y}}^*_{GS} = \mathbf{1}\hat{\mu} + \sum_{i=1}^{m} \mathbf{x_i}\hat{\mathbf{m}}_\mathbf{i} \tag{22}$$

$$\hat{\mathbf{y}}^*_{GSp} = \mathbf{1}\hat{\mu} + \sum_{i=1}^{m} \mathbf{x_i}\hat{\mathbf{m}}_\mathbf{i} + \mathbf{Z}\hat{\mathbf{u}} \tag{23}$$

$$\hat{\mathbf{y}}^*_{GSq} = \mathbf{1}\hat{\mu} + \sum_{i=1}^{m} \mathbf{x_i}\hat{\mathbf{m}}_\mathbf{i} + \mathbf{Z}\hat{\mathbf{u}} + \mathbf{Z}\hat{\mathbf{q}} \tag{24}$$

donde, $\hat{\mu}$ corresponde a una media general estimada para el modelo, $\hat{\mathbf{m}}_\mathbf{i}$ es el efecto predicho para el i-ésimo marcador, $\hat{\mathbf{u}}$ corresponde al efecto genético aditivo (basado en pedigrí) y $\hat{\mathbf{q}}$ es el efecto de los marcadores significativos dado por la expresión 12. En el contexto del modelo GSq, $\hat{\mathbf{m}}_\mathbf{i}$ es el efecto predicho para el i-ésimo marcador que no fue significativo de acuerdo a la expresión 12.

3.3 RESULTADOS

3.3.1 SNP significativamente asociados a características fenotípicas en *E. cladocalyx*

De acuerdo a la expresión 12, se identificaron 64 marcadores SNPs que fueron significativamente relacionados con todas las características estudiadas ($p > 0,001$). Particularmente, 11, 16, 5, 10, 5 y 17 SNPs fueron significativos para las características de ALT, DAP, RF, DM, IF y APB, receptivamente. En la figura 6, son representados los efectos de todos los SNPs para cada una de las características estudiadas. Los efectos de todos los SNPs fluctuaron valores entre -0,02 y 0,025; -0,045 y 0,04; -0,07 y 0,01; -0,06 y 0,06; -0,006 y 0,006; -0,007 y 0,005; para la ALT, DAP, RF, DM, IF y APB, respectivamente.
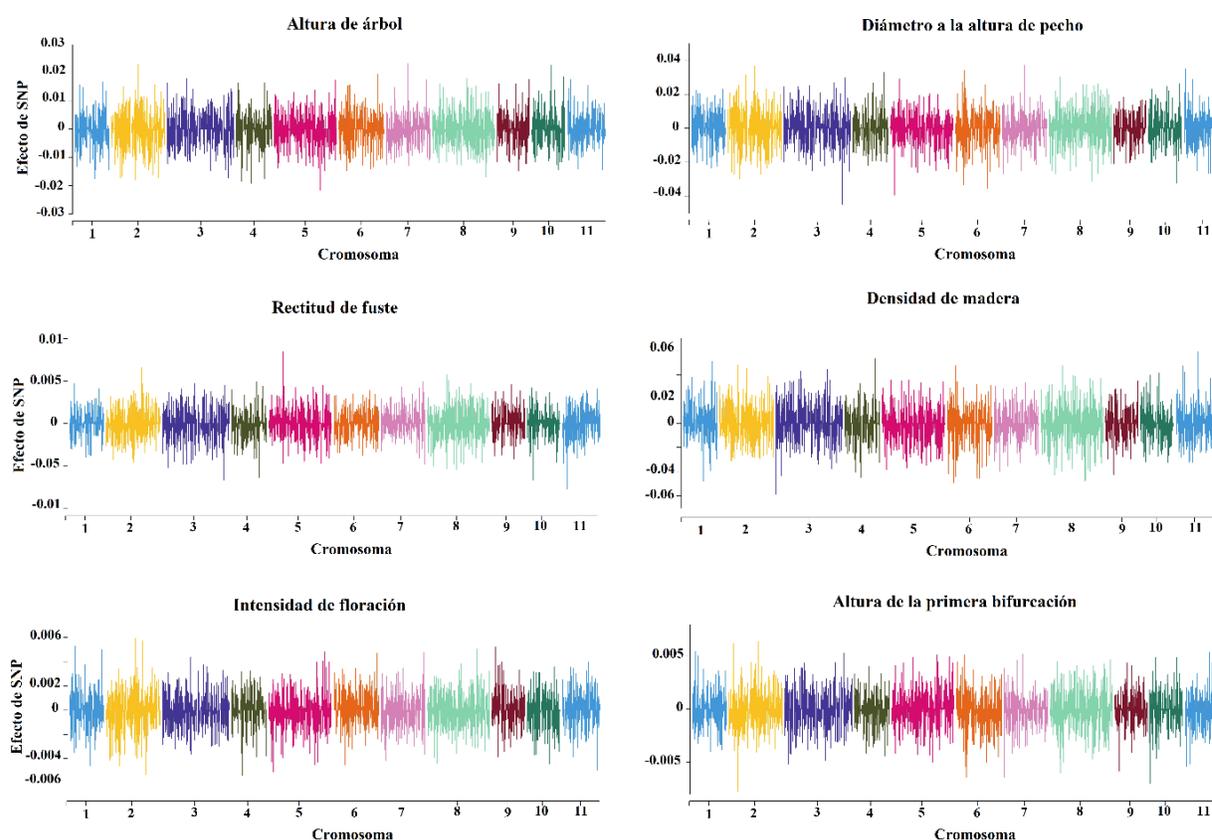

Figura 6. Gráficas de los efectos de los 3879 SNPs informativos para *E. cladocalyx*, para cada una de las características fenotípicas estudiadas.

3.3.2 Evaluación de la bondad de ajuste y poder predictivo de los modelos de predicción

El Criterio de Información de Desviación (DIC) fue usado para comparar los modelos GS, GSp y GSq en términos de bondad de ajuste. Los valores de DIC y poder predictivo de todos los modelos son mostrados en la Tablas 7 y 8, respectivamente. Los modelos GSp y GSq tuvieron una mejor bondad de ajuste que los modelos GS para todas las características evaluadas. Los valores del poder predictivo (PP) de los modelos GSq fueron más altos que aquellos basado en el enfoque GSp cuando más de 10 SNPs estadísticamente significativos fueron incluidos en el modelo de predicción GSq. Esta condición fue observada al predecir las características de ALT, DAP, APB y DM, donde los valores de PP fluctuaron entre 0,19-0,34 y 0,38-0,45 para los modelos GSp y GSq, respectivamente. En la mayoría de los casos, los modelos con un mayor valor de PP también tuvieron una mejor bondad de ajuste. De acuerdo a los valores de DIC, el modelo GSp basado en el método Bayes B tuvo mejor bondad de ajuste que el modelo GSq basado en este mismo método, para predecir la altura de árboles. No obstante, los modelos GSq basados en los métodos Bayes A, Bayes C y regresión bayesiana contraída (BRR) tuvieron una mejor bondad de ajuste que aquellos basados en el enfoque GSp. Por otro lado, el enfoque GSq tuvo un mayor poder predictivo que todos los modelos basados en el enfoque GSp. En el caso del DAP de los árboles, todos los modelos GSq fueron superiores a los modelos GSp, en términos de bondad de ajuste y poder predictivo. Más aun, el poder predictivo del DAP basado en los modelos GSq fue hasta dos veces mayor que el obtenido mediante los modelos GSp. Contrariamente, los modelos basados en el enfoque GSp tuvieron valores de DIC más bajos que aquellos basados en GSq (significativamente en el caso de BC y BRR), mientras que el poder predictivo de los modelos basados en GSp y GSq fue similar. En el caso de la DM, la mayoría de los modelos basados en el enfoque GSq tuvieron valores de DIC más bajos que aquellos modelos basados en el enfoque GSp. El poder predictivo para la DM fue más alto en aquellas predicciones basadas en los modelos GSq que en los modelos basados en el enfoque GSp. Los modelos con mejor ajuste para IF (en términos de valores de DIC) estuvieron basados en el enfoque GSp, mientras que el poder predictivo de ambos enfoques fue similar y fluctuaron entre 0,23 y 0,25. En el caso de APB de los árboles, los modelos basados en el enfoque GSq fueron superiores a los modelos basados en GSp en términos de bondad de ajuste. Consistentemente, el poder predictivo de los modelos basados en

el enfoque GSq fue hasta dos veces más alto que el poder predictivo alcanzado por los modelos GSp.

Tabla 7. Criterio de información de desviación (DIC) de las predicciones genómicas en *E. cladocalyx* basadas en (i) el efecto de los 3879 SNPs (GS), (ii) el efecto de los 3879 SNPs, en conjunto con los efectos poligénicos (GSp), y (iii) los efectos poligénicos, el efecto de SNPs significativos y el efecto de los SNPs no significativo (GSq).

Característica/Modelo	Bayes A	Bayes B	Bayes C	BRR [b]
Altura total del árbol				
GS	2015,1	2000,4	2006,3	2007,5
GSp	1968,9	1959,5	1968,6	1965,4
GSq	1951,2	1971,3	1941,3	1941,2
$\Delta DIC_{GSp-GSq}$ [a]	17,7 **	11,8 **	27,3 **	24,2 **
Diámetro a la altura de pecho				
GS	2562,1	2553,2	2554,6	2557,5
GSp	2556,5	2544,7	2539,8	2538,2
GSq	2490,4	2480,7	2480,2	2473,3
$\Delta DIC_{GSp-GSq}$ [a]	66,1 **	64,0 **	59,6 **	64,9 **
Rectitud del fuste				
GS	969,7	968,5	967,7	966,5
GSp	947,7	941,4	932,7	935,3
GSq	947,0	944,7	947,2	946,2
$\Delta DIC_{GSp-GSq}$ [a]	0,7	3,3	14,5 **	10,8 **
Densidad de madera				

GS	2139,6	2131,1	2143,1	2136,2
GSp	2094,3	2082,4	2101,8	2098,0
GSq	2067,1	2075,9	2067,6	2070,3
$\Delta DIC_{GSp-GSq}$ [a]	27,2 **	6,5 *	34,3 **	28,4 **

Intensidad de floración

GS	1368,2	1359,1	1358,2	1355,2
GSp	1293,9	1301,2	1282,3	1285,9
GSq	1306,0	1309,7	1301,3	1295,2
$\Delta DIC_{GSp-GSq}$ [a]	12,1 **	8,4 *	19,0 **	9,4 *

Altura de la primera bifurcación

GS	1508,2	1503,4	1504,2	1502,1
GSp	1491,9	1490,6	1487,3	1485,8
GSq	1426,6	1424,9	1424,9	1423,5
$\Delta DIC_{GSp-GSq}$ [a]	65,3 **	65,7 **	62,4 **	62,3 **

[a] Diferencia entre los valores de DIC de los modelos GSp y GSq. [b] Regresión Bayesiana contraída. * Diferencia sustancialmente significativa entre los modelos GSp y GSq. ** Fuerte evidencia de diferencia significativa entre los modelos GSp y GSq.

Tabla 8. Poder predictivo (PP) de todas las características estudiadas en *E. cladocalyx* de acuerdo a los modelos basados en: (i) el efecto de los 3879 SNPs (GS), (ii) el efecto de los 3879 SNPs, en conjunto con los efectos poligénicos (GSp), y (iii) los efectos poigénicos, el efecto de SNPs significativos y el efecto de los SNPs no significativo (GSq). Los valores de PP de cada modelo corresponde al promedio de los valores de PP para 20-sets de validación.

Característica/Modelo	Bayes A	Bayes B	Bayes C	BRR[a]	X_{PP}[b]
Altura total del árbol					
GS	0,32	0,32	0,32	0,33	0,32
GSp	0,33	0,32	0,33	0,34	0,33
GSq	0,45	0,44	0,44	0,45	0,44
Diámetro a la altura de pecho					
GS	0,23	0,23	0,24	0,23	0,23
GSp	0,21	0,23	0,22	0,22	0,22
GSq	0,41	0,41	0,41	0,42	0,41
Rectitud del fuste					
GS	0,40	0,40	0,40	0,39	0,40
GSp	0,39	0,39	0,39	0,39	0,39
GSq	0,40	0,40	0,40	0,39	0,40
Densidad de madera					
GS	0,25	0,25	0,25	0,25	0,25
GSp	0,27	0,27	0,27	0,28	0,27
GSq	0,43	0,43	0,43	0,43	0,43
Intensidad de floración					
GS	0,07	0,09	0,10	0,10	0,09
GSp	0,25	0,25	0,24	0,23	0,24
GSq	0,25	0,25	0,25	0,24	0,25
Altura de la primera bifurcación					
GS	0,18	0,18	0,19	0,19	0,19
GSp	0,19	0,20	0,20	0,19	0,19

GSq	0,38	0,38	0,39	0,39	0,38

[a] Regresión Bayesiana contraída. [b] Promedio de los valores de PP de los métodos BA, BB, BC y BRR.

3.3.3 Estimaciones de heredabilidad

En razón de los valores de bondad de ajuste de los modelos, se estimaron los componentes de varianza exclusivamente para los modelos GSp y GSq (Tabla 9). La estimación de heredabilidad basada en los efectos poligénicos fue más alta para los modelos basados en GSp que en aquellos basados en GSq para todas las características, mientras que las estimaciones de heredabilidad genómica (basados en GSp y GSq) fueron dependiente de los métodos bayesianos de predicción. Teniendo en cuenta los modelos con una mejor bondad de ajuste (GSp y GSq), la heredabilidad de ALT varió entre 0,13 (h_a^2) y 0,21, 0,14 y 0,45 (h_m^2), y 0,29 y 0,34 (h_q^2). En el caso de DAP, las estimaciones de heredabilidad basadas en los antecedentes de pedigrí y en los marcadores significativos para DAP fluctuaron entre los valores 0,05 y 0,17, mientras que la heredabilidad del total de marcadores, o en el caso de GSq de los marcadores no significativos para DAP, varió entre 0,4 y 0,45. En el caso de la RF, las estimaciones de heredabilidad basadas en pedigrí (mediante el enfoque GSp) variaron entre 0,16 y 0,23, mientras que aquellas basadas en datos genómicos fluctuaron entre 0.3 y 0.32. Basado el enfoque GSq, las estimaciones de heredabilidad basadas en pedigrí, SNP significativos y SNPs no significativamente asociados a la DM fluctuaron entre 0,14-0,17, 0,27-0,31 y 0,12-0,24, respectivamente. Las estimaciones de heredabilidad de IF basadas en información de pedigrí, y bajo un enfoque GSp, variaron de 0,27 a 0,34, mientras que la heredabilidad genómica fluctuó entre valores de 0,07 y 0,29. Finalmente, las estimaciones de heredabilidad basadas en pedigrí de la APB (en el contexto de un enfoque GSq) fueron de 0,08 para todos los métodos de predicción estudiados, mientras que la heredabilidad de los marcadores significativamente relacionados con esta característica y de aquellos que no fueron significativos, variaron de 0,41-0,45 y 0,04 – 0,13, respectivamente.

Tabla 9. Estimación de los valores de heredabilidad genéticas y genómicas para las características evaluadas en *E. cladocalyx*, basada en los modelos que incluyen (i) el efecto de los 3879 SNPs (GS), (ii) el efecto de los 3879 SNPs, en conjunto con los efectos poligénicos (GSp), y (iii) los efectos poligénicos, el efecto de SNPs significativos y el efecto de los SNPs no significativo (GSq).

Característica/ Modelo	Bayes A			Bayes B			Bayes C			BRR[a]		
	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$
Altura de árbol												
GS	-	0,24	-	-	0,45	-	-	0,4	-	-	0,24	-
GSp	0,28	0,24	-	0,21	0,45	-	0,22	0,40	-	0,29	0,24	-
GSq	0,16	0,14	0,32	0,18	0,12	0,29	0,13	0,27	0,29	0,14	0,17	0,34
Diámetro a la altura de pecho												
GS	-	0,14	-	-	0,37	-	-	0,39	-	-	0,24	-
GSp	0,20	0,14	-	0,15	0,37	-	0,15	0,39	-	0,19	0,24	-
GSq	0,11	0,05	0,44	0,09	0,15	0,42	0,09	0,17	0,40	0,10	0,11	0,45
Rectitud de fuste												
GS	-	0,32	-	-	0,31	-	-	0,3	-	-	0,3	-
GSp	0,23	0,32	-	0,18	0,31	-	0,16	0,30	-	0,21	0,30	-
GSq	0,18	0,18	0,01	0,14	0,37	0,01	0,14	0,32	0,01	0,18	0,19	0,013
Densidad de madera												
GS	-	0,28	-	-	0,5	-	-	0,45	-	-	0,42	-
GSp	0,25	0,28	-	0,18	0,50	-	0,19	0,45	-	0,21	0,42	-
GSq	0,17	0,13	0,31	0,17	0,18	0,27	0,14	0,24	0,27	0,17	0,12	0,31

Intensidad de floración

GS	-	0,04	-	-	0,41	-	-	0,46	-	-	0,26	-
GSp	0,34	0,07	-	0,32	0,10	-	0,27	0,29	-	0,33	0,13	-
GSq	0,30	0,06	0,00	0,29	0,06	0,00	0,27	0,20	0,00	0,31	0,10	0,002

Altura de la primera bifurcación

GS	-	0,05	-	-	0,12	-	-	0,27	-	-	0,14	-
GSp	0,20	0,05	-	0,19	0,12	-	0,16	0,27	-	0,19	0,14	-
GSq	0,08	0,04	0,44	0,08	0,11	0,42	0,08	0,13	0,41	0,08	0,06	0,45

[a] Regresión bayesiana contraída.

3.4 DISCUSIÓN

3.4.1 Regiones asociadas a caracteres fenotípicos en *E. cladocalyx*

Un total del 64 SNPs tuvieron un efecto significativo en el fenotipo de todas las características evaluadas. Los SNPs significativamente relacionados a la ALT y DAP de los árboles se localizaron principalmente en los cromosomas Chr2 (5 y 4 SNPs relacionados a ALT y DAP, respectivamente), Chr6 (3 relacionados a DAP), y Chr8 (1 y 2 SNPs relacionados a ALT y DAP, respectivamente). Estudios previos han reportado varias regiones genómicas explicando la variación de características relacionadas al crecimiento en *E. cladocalyx* (Ballesta et al., 2015; Arriagada et al., 2018; Maldonado et al., 2018; Valenzuela et al., 2019). En concordancia con los resultados del presente estudio, Maldonado et al. (2018) detectaron una región en el cromosoma Chr6 explicando un 27% de la variación fenotípica del DAP de árboles de *E. cladocalyx*. Adicionalmente, Arriagada et al. (2018) reportaron regiones microsatélites ubicadas en los Chr6, Chr7 y Chr8, las cuales explicaron hasta un 8% de la variación fenotípica de la ALT y DAP en *E. cladocalyx*. Particularmente, los componentes de floración han sido objetivo

en los programas de mejoramiento en algunas especies de *Eucalyptus*, incluyendo *E. cladocalyx*, con perspectivas de potenciar la producción de miel (Cané-Retamales et al., 2011; De Lange et al., 2013; Arriagada et al., 2018). En el presente estudio, la mayoría de los SNPs significativamente relacionados a la IF en *E. cladocalyx* fueron localizados en el cromosoma Chr2 (80% de los SNPs significativos para IF). Consistentemente, Missiagia et al. (2005) reportaron un loci de característica cuantitativa (QTL), llamado como locus *Eef1*, relacionado a la precocidad de florecimiento en *E. grandis,* localizado en el grupo de ligamiento 2. En *E. cladocalyx*, estudios previos han reportado que la intensidad y precocidad de la floración estarían positivamente correlacionadas y controladas por QTLs que explican la variación fenotípica de ambas características (Cané-Retamales et al., 2011; Contreras-Soto et al., 2016). Los SNPs identificados como significativos para aquellas características relacionadas con la calidad de la madera (tales como DM, RF y AP) fueron localizados principalmente en los cromosomas Chr2 (DM), Chr5 (APB) y Chr8 (RF, DM y APB). En concordancia con estos resultados, Valenzuela et al. (2019) detectaron un QTL en el cromosoma Chr2 que explicó un 8% de la variación fenotípica de la DM de *E. cladocalyx.*

3.4.2 Comparación entre modelos de predicción

Una importante cantidad de estudios han explorado los beneficios de la selección genómica en especies forestales (Ratcliffe et al., 2015; Gamal-Dien et al., 2015; Lenz et al., 2017; Chen et al., 2018; Ballesta et al., 2018). De acuerdo con los resultados del presente trabajo, los modelos GSq y GSp tuvieron una mejor bondad de ajuste que los modelos basados en sólo marcadores (GS), lo cual reveló la importancia de incluir la estructura genética en los modelos de predicción que incluyen una relativa baja densidad de variables predictoras (SNPs), un aspecto también discutido por Silva et al. (2018). Más aun, el poder predictivo de los modelos GSq y/o GSp fue hasta ~ 4 veces mayor que el poder predictivo de los modelos GS (por ejemplo, para predecir IF). En el contexto de este trabajo, el uso de la información de pedigrí compensó el hecho de la baja disponibilidad de marcadores SNP, de tal manera que permitió un incremento en la precisión predictiva de las características fenotípicas estudiadas en *E. cladocalyx*. Cabe destacar que la evidencia empírica sugiere que los métodos de predicción basados

exclusivamente en genómica superan, términos de poder predictivo, a aquellos basados en información de pedigrí. No obstante, en concordancia con los hallazgos del presente estudio, Burgueño et al. (2012) reportaron que la combinación de los antecedentes de pedigrí y marcadores moleculares incrementan la precisión predictiva de diferentes características agronómicas en trigo. Velazco et al. (2019) demostraron que la información de pedigrí combinada con la genómica permite optimizar la predicción de características de baja heredabilidad en sorgo, en un contexto donde los marcadores moleculares no pueden capturar toda la variabilidad genética de estas características. Interesantemente, estos autores realizaron los estudios genómicos con una densidad de marcadores relativamente baja para sorgo (~4800 DArts; Nelson et al. 2011). Más aun, algunos autores sugieren que combinar las relaciones de parentesco basadas en genómica y pedigrí, puede incrementar el poder predictivo de un modelo (Legarra et al., 2009; Jensen et al., 2012; Tiezzi et al., 2015; Cappa et al., 2019). Interesantemente, Cericola et al. (2017) evaluaron la precisión predictiva de modelos equivalente a GS, GSp y GSq para predecir diferentes características en trigo. Estos autores concluyeron que la inclusión de información genealógica permite incrementar el poder predictivo de un modelo, aun con una densidad de marcadores no superior a 1000 SNPs. Por lo tanto, basado en los resultados y en estos antecedentes, parece recomendable utilizar ambos tipos de información, genómico y genealógico, para obtener mejores resultados predictivos, en el contexto de una baja disponibilidad de marcadores moleculares.

De acuerdo con De Los Campos et al. (2013), los métodos Bayes A, Bayes B, Bayes C y BRR pueden incrementar el poder predictivo en evaluaciones basadas en genómica, no obstante, estos métodos podrían tener problemas de sobreajuste en los modelos cuando la proporción entre el número de marcadores e individuos es superior a un valor de 50 (González-Recio et al., 2014). Para superar esta dificultad, el uso de matrices de parentesco entre individuos dentro de los modelos de predicción podría ayudar a capturar información general y reducir el problema de dimensionalidad y colinealidad de las variables predictoras (Solberg et al., 2009; Du et al., 2018). Interesantemente, en el presente estudio se encontró que los modelos GSq tuvieron una mejor bondad de ajuste que los modelos GSp, y al mismo tiempo tuvieron un mayor poder predictivo de DAP, APB y DM. Cabe destacar que estos resultados sólo fueron detectados

cuando más de 10 de SNPs fueron significativamente asociados con las características estudiadas. De acuerdo con Silva et al. (2018), el mejor rendimiento de los modelos GSq, por sobre los modelos GSp, puede ser explicado por el hecho de que los modelos GSq permiten explorar regiones genómicas relevantes no directamente consideradas en los modelos GSp y/o GS, y de esta manera, incrementar la precisión predictiva de características fenotípicas. Adicionalmente, estudios previos han informado que una preselección de variables (SNPs) o incluir los principios del mapeo asociativo en los modelos de predicción genómica, puede incrementar el poder predictivo de un modelo (Macciotta et al., 2009; Ballesta et al., 2018; Arojju et al., 2018).

3.4.3 Estimaciones de heredabilidad

Las estimaciones de heredabilidad basadas en información de pedigrí, para todas las características, revelaron un control genético de bajo a moderado, lo cual es esperado para características relacionadas al crecimiento, floración y calidad de madera en especie forestales, incluyendo *E. cladocalyx* (Mora et al., 2009; Bush et al., 2011; Cané-Retamales et al., 2011; Vargas-Reeve et al., 2013; Bush et al., 2015; Valenzuela et al., 2019). En el contexto de las estimaciones basadas en genómica, en términos de bondad de ajuste, el modelo GSq fue superior al modelo GSp para las características de APB, DAP, y DM, mientras que, para las características de RF y IF, el modelo GSp tuvo un mejor desempeño. En adición, para el caso de las características de ALT, DAP, DM y APB, las estimaciones de heredabilidad genómica para los SNPs significativos fueron más altas que las estimaciones basadas en los SNP no significativamente asociado a alguna característica. Por otro lado, las estimaciones de heredabilidad genómica basadas en lo SNPs no significativamente asociados, considerando los métodos Bayes B y Bayes C, fueron más altas que aquellas estimaciones basadas en los métodos Bayes A y BRR para todas las características estudiadas, exceptuando la ALT. Los modelos Bayes B y Bayes C involucran procedimientos de selección de variables, los cuales favorecen la selección de aquellos marcadores con efectos relativamente más importantes en las características (Habier et al., 2011; Wolfe et al., 2017). De acuerdo con De Los Campos et al.

(2015), en algunas ocasiones los métodos Bayes B y Bayes C podrían sobreestimar los valores de heredabilidad, por lo tanto, estos resultados deben ser interpretados con precaución.

Contrariamente al caso anterior, las estimaciones de heredabilidad basadas en los marcadores significativamente asociados a RF y IF fueron más bajas que aquellas basadas en información de pedigrí y /o en los SNP no significativos para estas características. En árboles, varios estudios han reportado regiones genómicas explicando un alto porcentaje de la variación fenotípica de la rectitud del fuste y de la intensidad de floración. Por ejemplo, Arriagada et al. (2018) reportaron un QTL explicando hasta una 15% de la variación fenotípica de la rectitud de fuste en *E. cladocalyx*. En un meta-análisis, Hall et al. (2016) determinaron que características relacionadas con la fenología de los árboles, en general, son altamente heredables y controladas por regiones claves en sus genomas. Más aun, Missiaggia et al. (2005) identificaron una región clave en el cromosoma Chr2 (*Eef1*) controlando la precocidad de floración de *E. grandis*. En este contexto, la baja densidad de marcadores SNP detectados en el presente estudio podría estar limitando la probabilidad de encontrar regiones genómicas controlando la variación de IF en *E. cladocalyx*. Adicionalmente, estos resultados son respaldados por el hecho de que los modelos GSp tuvieron una mejor bondad de ajuste que aquellos basados en el enfoque GSq, lo cual implica que la variación fenotípica de IF y RF es mejor explicada por un modelo con los componentes de varianza de los efectos poligénicos y de todos los marcadores en forma conjunta.

3.5 CONCLUSION

En este estudio, se evaluó el desempeño de modelos genómicos bayesianos que incluyen información de pedigrí e información de regiones significativamente asociadas, lo cual se implementó en un contexto de una baja densidad de marcadores. En este capítulo, se demostró que la estructura genética puede ser un recurso adicional de información, que permite aumentar el poder predictivo de características poligénicas en *E. cladocalyx*, en el contexto de una baja densidad de marcadores. Adicionalmente, el poder predictivo de los modelos GSq fueron superiores al resto de los modelos cuando la proporción de la varianza explicada por los SNPs

significativos fue mayor que la variación explicada por los efectos poligénicos y por los marcadores SNP que no se encontraron significativamente asociados con una característica. Tanto el número de SNP significativos, como el porcentaje de variación explicado por estos mismos, son determinantes en la efectividad del método GSq. Este enfoque puede ser particularmente útil para especies de árboles u otras plantas pobremente representadas en arreglos de SNP comerciales, los cuales han sido desarrollados para especies económicamente importantes.

4 CONCLUSIONES GENERALES

El presente trabajo es uno de los primeros estudios que examinan la inclusión de haplotipos en los modelos de predicción genómica en árboles. De acuerdo a los resultados, la representatividad del arreglo de SNP fue sustancialmente mayor en *E. globulus* que en *E. cladocalyx,* lo cual es explicado por el hecho de *E. cladocalyx* es una especie genéticamente más distante que *E. globulus,* a las secciones de *Eucalyptus* usadas para el desarrollo del arreglo de ADN. Consecuentemente, se detectó un menor número de haplotipos en *E. cladoclayx* que en *E. globulus*. Los modelos genómicos que incluyeron el efecto de haplotipo (ya sea HAP o HAP-SNP) tuvieron una mejor bondad de ajuste y poder predictivo de características de baja heredabilidad en el ensayo de *E. globulus*, un aspecto relevante para el mejoramiento genético de características con una herencia compleja y altamente influenciadas por el ambiente. Adicionalmente, las estimaciones de heredabilidad genómica de las características de baja heredabilidad basadas en el enfoque de haplotipos tuvieron un valor más alto que las estimaciones basadas en pedigrí, lo cual implica que los haplotipos permitieron cuantificar una variabilidad genética no explicada por el enfoque de pedigrí. En general, e independiente del enfoque de predicción genómica, el poder predictivo de las características fenotípicas de *E. cladocalyx* fue menor que para *E. globulus*. Cabe destacar, que estos resultados pueden ser explicados por el hecho de que la representatividad del arreglo de ADN fue menor en *E. cladocalyx*. En este sentido, existen especies que no están bien representadas en los arreglos comerciales de SNP de alta densidad. Basado en los antecedentes de historial de selección y estructura genética, las diferencias en el poder predictivo para cada población también podrían ser explicadas por el hecho de que los individuos de la población de *E. globulus* son genéticamente más cercanos que los individuos de la población de *E. cladocalyx*.

En el presente trabajo se demostró que los antecedentes genealógicos pueden ser un recurso de información valioso en los métodos de predicción genómica, en un contexto en donde se dispone de una baja densidad de marcadores. Los modelos basados en genómica y en la estructura genética de población tuvieron una mejor bondad de ajuste y precisión predictiva de características fenotípicas en *E. cladocalyx*, que los modelos netamente basados en datos genómicos.

En síntesis, el enfoque de haplotipos podría ser especialmente adecuado para predecir características de baja heredabilidad de *Eucalyptus*, en un escenario donde la base genética de estas características ha sido reducida debido a los procesos de selección artificial. Por otro lado, en poblaciones que no poseen un nulo historial de selección artificial, y que han sido genotipadas con una baja densidad de polimorfismos de nucleótido único, el uso combinado de los antecedentes de estructura genética y la información genómica, contribuye a una mejor predicción fenotípica.

La **hipótesis** de trabajo fue demostrada por el hecho de que los modelos basados en haplotipos (ya sea HAP o HAP-SNP) tuvieron una mejor bondad de ajuste para características de baja heredabilidad en *E. globulus*, que los polimorfismos de nucleótido único. Además, estos modelos explicaron una mayor variación genética (en términos de heredabilidad genómica) en características de baja heredabilidad, que los antecedentes genealógicos. Algunos modelos basados en SNP exhibieron valores de heredabilidad más altos que los modelos basados en HAP o HAP-SNP, no obstante, tuvieron una menor bondad de ajuste que los modelos basados en HAP o HAP-SNP. Cabe destacar que esto resultados solo fueron observados en el contexto de una población con un historial de selección artificial, en el cual la base genética de algunos caracteres ha sido sustancialmente reducida (Por ejemplo, el DAP, REC y CR de *E. globulus*). Adicionalmente, en este trabajo se demostró que los haplotipos tienen una mayor precisión predictiva para características de baja heredabilidad en *E. globulus*, que los polimorfismos de nucleótido único.

Finalmente, el presente estudio proporciona nuevas perspectivas para la implementación de la selección basada en genómica en *Eucalyptus,* y en otras especies de árboles, lo cual podría ser especialmente adecuado para mejorar caracteres de baja heredabilidad, o bien, cuando existe una baja disponibilidad de marcadores. Los hallazgos del presente estudio podrían ayudar a optimizar el proceso selectivo de árboles, y de esta manera contribuir a una industria forestal más rentable. Es importante destacar que se requiere de nuevos estudios que examinen cómo se desempeñan los métodos planteados a través de nuevas generaciones y en otros ambientes, con el fin de evaluar su aplicación durante los ciclos de selección.

5. REFERENCIAS BIBLIOGRÁFICAS

Adhikari, S., Saha, S., Biswas, A., Rana, T. S., Bandyopadhyay, T. K., Ghosh, P. (2017). Application of molecular markers in plant genome analysis: a review. The Nucleus **60**(3): 283-297.

Aguirre, N.C., Filippi, C.V., Zaina, G., Rivas, J.G., Acuña, C.V., Villalba, P.V., García, M.N., González, S., Rivarola, M., Maetínez, M.C. (2019). Optimizing ddRADseq in non-model species: A case study in *Eucalyptus dunnii* Maiden. Agronomy **9**: 484.

Almeida Filho, J. E., Guimarães, J. F. R., e Silva, F. F., de Resende, M. D. V., Muñoz, P., Kirst, M., & de Resende Júnior, M. F. R. (2019). Genomic Prediction of Additive and Non-additive Effects Using Genetic Markers and Pedigrees. G3: Genes, Genomes, Genetics, 9(8), 2739-2748.

Andersen, J., Lübberstedt, T. (2003). Functional markers in plants. Trends Plant Sci. **8**: 554–560.

Arojju, S.K., Conaghan, P., Barth, S., Milbourne, D., Casler, M.D., Hodkinson, T.R., Michel, T., Byrne, S.L. (2018). Genomic prediction of crown rust resistance in *Lolium perenne*. BMC Genet. **19**: 35

Arriagada, O., Mora, F., Amaral-Junior, A.T. (2018). Thirteen years under arid conditions: exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. Breeding Sci. **68**(3): 367-374

Asoro, F.G., Newell, M.A., Beavis, W.D., Scott, M.P., Jannink, J.-L. (2011). Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome-US **4**: 132–144.

Azevedo, C.F., de Resende, M.D.V., e Silva, F.F., Viana, J.M.S., Valente, M.S.F., Resende, M.F.R., Muñoz, P. (2015). Ridge, Lasso and Bayesian additive-dominance genomic models. BMC Genet. **16**: 105.

Ballesta, P., Serra, N., Guerra, F. (2018). Genomic prediction of growth and stem quality traits in *Eucalyptus globulus* Labill. at its southernmost distribution limit in Chile. Forests **9**: 779.

Ballesta, P., Mora, F., Contreras-Soto, R. I., Ruiz, E., Perret, S. (2015). Analysis of the genetic diversity of *Eucalyptus cladocalyx* (sugar gum) using ISSR markers. Acta Sci-Agron. **37**(2): 133-140.

Barrett, J.C., Fry, B., Maller, J.D.M.J., Daly, M.J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

Bartholomé, J., Mabiala, A., Burlett, R., Bert, D., Leplé, J. C., Plomion, C., Gion, J. M. (2020). The pulse of the tree is under genetic control: eucalyptus as a case study. The Plant Journal.

Bartholomé, J., Bink, M.C., van Heerwaarden, J., Chancerel, E., Boury, C., Lesur, I., Isik, F., Bouffier, L., Plomion, C. (2016). Linkage and association mapping for two major traits used in the maritime pine breeding program: Height growth and stem straightness. Plos one **11**: e0165323.

Battenfield, S.D., Sheridan, J.L., Silva, L.D., Miclaus, K.J., Dreisigacker, S., Wolfinger, R.D., Peña, R.J., Singh, R.P., Jackson, E.W., Fritz, A.K., Guzmán, C., Poland, J.A. (2018). Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. Plos one **13**: e0204757.

Beaulieu, J., Doerksen, T., Clément, S., MacKay, J., Bousquet, J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity **113**(4): 343.

Bhatia, G., Gusev, A., Loh, P. R., Vilhjálmsson, B. J., Ripke, S., Purcell, S., ... O'Donovan, M. C. (2015). Haplotypes of common SNPs can explain missing heritability of complex diseases. BioRxiv, 022418.

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. TRENDS in Genetics, 22(8), 437-446.

Blackburn, D.P., Hamilton, M.G., Harwood, C.E., Baker, T.G., Potts, B.M. (2013). Assessing genetic variation to improve stem straightness in *Eucalyptus globulus*. Ann. For. Sci. **70**: 461-470

Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M.N., Boscher, M.Y., Druet, T., Genestout, L., Colleau, J.J., Journaux, L., Ducrocq, V., Fritz, S. (2012). Genomic selection in French dairy cattle. Anim. Prod. Sci. **52**(3): 115-120.

Bolormaa, S., Gore, K., van der Werf, J.H.J., Hayes, B.J.; Daetwyler, H.D. (2015). Design of a low density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Anim. Genet. **46:** 544–556.

Boland, D.J., Brooker, M.I.H.; McDonald, M.W.; Chippendale, G.M. (2006). Forest Trees of Australia. Csiro Publishing, Australia, p. 278

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics **23**: 2633–2635.

Breseghello, F., Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics **172**(2): 1165-1177.

Brenner, C. H., Weir, B. S. (2003). Issues and strategies in the DNA identification of World Trade Center victims. Theor. Popul. Biol. **63**(3):173–178.

Brooker, M.I.H. (2000). A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). Aust. Syst. Bot. **13**: 79–148.

Brown, M. D., Glazner, C. G., Zheng, C., Thompson, E.A. (2012). Inferring coancestry in population samples in the presence of linkage disequilibrium. Genetics **190**(4): 1447-1460.

Bundock, P. C., Potts, B. M., Vaillancourt, R. E. (2008). Detection and stability of quantitative trait loci (QTL) in Eucalyptus globulus. Tree Genet. Genomes **4**: 85-95.

Burgueño, J., de los Campos, G., Weigel, K., Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype$\times$ environment interaction using pedigree and dense molecular markers. Crop Sci. **52**(2): 707-719.

Bush, D., Kain, D., Kanowski, P., Matheson, C. (2015). Genetic parameter estimates informed by a marker-based pedigree: a case study with Eucalyptus cladocalyx in southern Australia. Tree Genet. Genomes **11**(1): 798.

Bush, D., Thumma, B. (2013). Characterizing a Eucalyptus cladocalyx breeding population using SNP markers. Tree Genet. Genomes **9**(3): 741-752.

Bush, D., McCarthy, K., Meder, R. (2011). Genetic variation of natural durability traits in Eucalyptus cladocalyx (sugar gum). Ann. For. Sci. **68**(6): 1057.

Callister, A.N.; England, N.; Collins, S. (2011). Genetic analysis of Eucalyptus globulus diameter, straightness, branch size, and forking in Western Australia. Can. J. For. Res. **41**: 1333 1343.

Calus, M.P., Meuwissen, T.H., Windig, J.J., Knol, E.F., Schrooten, C., Vereijken, A.L., Veerkamp, R.F. (2009). Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. Genet. Sel. Evol. **41**: 11.

Calus, M.P.L., T.H.E. Meuwissen, A.P.W. De Roos, R.F. Veerkamp. (2008). Accuracy of genomic selection using different methods to define haplotypes. Genetics **178**: 553–561.

Cané-Retamales, C., Mora, F., Vargas-Reeve, F., Perret, S., Contreras-Soto, R. (2011). Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in Eucalyptus cladocalyx under arid conditions. Euphytica **178**(2): 177-183.

Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. Plant Sci. **284**: 9-15.

Cappa, E. P., El-Kassaby, Y. A., Garcia, M. N., Acuña, C., Borralho, N. M., Grattapaglia, D., & Poltri, S. N. M. (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. PloS one **8**(11): e81267.

Carbonari, C. A., Miranda, L. G., Gomes, G. L. G. C., Picoli Junior, G. J., Matos, A. K. A. Velini, D. E. (2016). Differential tolerance of eucalyptus clones to sulfentrazone applied in different soil textures. Sci. For. **44**: 9-18.

Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevereiro, P., Myburg, A. A., Paiva, J. A. Grima-Pettenati, J. (2015). Genome-wide analysis of the lignin toolbox of Eucalyptus grandis. New Phytol. **206**: 1297-1313.

Cericola, F., Jahoor, A., Orabi, J., Andersen, J. R., Janss, L. L., Jensen, J. (2017). Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in advanced wheat breeding lines. PloS one **12**(1): e0169606.

Chandra, S., Singh, D., Pathak, J. (2017). SNP discovery from next-generation transcriptome sequencing data and their validation using KASP assay in wheat (Triticum aestivum L.). Mol. Breed. **37**(7): 92

Chen, Z.Q., Baison, J., Pan, J., Karlsson, B., Andersson, B., Westin, J., García-Gil, M.R., Wu, H.X. (2018). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genom. **19**: 946.

Cobb, J. N., Biswas, P. S., Platten, J. D. (2019). Back to the future: revisiting MAS as a tool for modern plant breeding. Theor. Appl. Genet. **132**(3): 647-667.

Contreras-Soto, R.I., Mora, F., De Oliveira, M.A.R., Higashi, W., Scapim, C.A., Schuster, I. A. (2017). Genome-wide association study for agronomic traits in soybean using SNP markers and SNP based haplotype analysis. Plos one **12**: e0171105.

Contreras-Soto, R.; Ballesta, P.; Ruiz, E.; Mora, F. (2016). Identification of ISSR markers linked to flowering traits in a representative sample of *Eucalyptus cladocalyx*. J. For. Res. **27**: 239–245.

Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., Braun, H-J. (2010).

Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics **186**(2): 713-724.

Curtis, D., North, B.V., Sham, P.C. (2001). Use of an artificial neural network to detect association between a disease and multiple marker genotypes. Ann. Hum. Genet. **65**: 95–107.

Cuyabano, B.C., Su, G., Lund, M.S. (2014). Genomic prediction of genetic merit using LD based haplotypes in the Nordic Holstein population. BMC Genom. **15**: 1171.

Daetwyler, H. D., Hayden, M. J., Spangenberg, G. C., Hayes, B. J. (2015). Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. Genetics **200**(4), 1341-1348.

Daetwyler, H. D., Calus, M. P., Pong-Wong, R., De Los Campos, G., Hickey, J.M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics **193**(2): 347365.

De Lange, W.J., Veldtman, R., Allsopp, M.H. (2013). Valuation of pollinator forage services provided by *Eucalyptus cladocalyx*. J. Environ. Manag. **125**: 12–18.

De los Campos, G., Sorensen, D., Gianola, D. (2015). Genomic heritability: What is it? Plos Genet. **11**: e1005048.

De los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M.P.L. (2013). Whole Genome regression and prediction methods applied to plant and animal breeding. Genetics **193**: 327–345.

Denis, M., Bouvet, J. M. (2013). Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus breeding*. Tree Genet. Genomes **9**(1): 37-51.

De Roos, A. P. W., Schrooten, C., Druet, T. (2011). Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. J. Dairy Sci. **94**(9): 4708-4714.

Di Guardo, M., Micheletti, D., Bianco, L., Koehorst-van Putten, H.J., Longhi, S., Costa, F., Aranzana, M.J., Velasco, R., Arús, P., Troggio, M., Van De Weg, E.W. (2015). ASSIsT: an automatic SNP scoring tool for in- and outbreeding species. Bioinformatics **31**(23): 3873-3874.

Doublet, A. C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., ... & Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. Genetics Selection Evolution, 51(1), 52.

Doyle, J. J., Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. Focus **12**: 13-15.

Drake, J. E., Aspinwall, M. J., Pfautsch, S., Rymer, P.D., Reich, P.B., Smith, R.A., Crous, K.Y., Tissue, D.T., Ghannoum, O., Tjoelker, G. (2015). The capacity to cope with climate warming declines from temperate to tropical latitudes in two widely distributed *Eucalyptus* species. Glob. Change Biol. **21**(1): 459-472.

Doublet, A. C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., ... Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. Genet. Sel. Evol. **51**(1): 52.

Du, C., Wei, J., Wang, S., Jia, Z. (2018). Genomic selection using principal component regression. Heredity **121**(1): 12-23.

Durán, R., Rodriguez, V., Carrasco, A., Neale, D., Balocchi, C.,Valenzuela, S. (2019). SNP discovery in radiata pine using a de novo transcriptome assembly. Trees **33**(5): 1505-1511.

Durán, R.; Isik, F.; Zapata-Valenzuela, J.; Balocchi, C.; Valenzuela, S. (2017). Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. Tree Genet. Genomes **13**: 74.

Edriss, V., Fernando, R. L., Su, G., Lund, M.S., Guldbrandtsen, B. (2013). The effect of using genealogy-based haplotypes for genomic prediction. Genet. Sel. Evol. **45**(1): 5.

Edwards, D. (2015). Two molecular measures of relatedness based on haplotype sharing. BMC Bioinformatics **16**(1): 383.

Ellis, M. F., Sedgley, M. (1992). Floral morphology and breeding system of three species of Eucalyptus, section Bisectaria (Myrtaceae). Aust. J. Bot. **40**(3): 249-262.

Endelman, J. B., & Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. G3: Genes, Genomes, Genetics, 2(11), 1405-1413.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology, 14(8), 2611-2620.

Eynard, S. E., Croiseau, P., Laloë, D., Fritz, S., Calus, M. P., Restoux, G. (2018). Which individuals to choose to update the reference population? Minimizing the loss of genetic diversity in animal genomic selection programs. G3: Genes Genom. Genet. **8**(1): 113-121.

Fiil, A., Lenk, I., Petersen, K., Jensen, C.S., Nielsen, K.K., Schejbel, B., Andersen, J.R., Lübberstedt, T. (2011). Nucleotide diversity and linkage disequilibrium of nine genes with putative effects on flowering time in perennial ryegrass (Lolium perenne L.). Plant Sci. **180**(2): 228-237.

Freeman, J. S., Whittock, P. S., Brad, B. M., Vaillancourt, R.E. (2009) QTL influencing growth and wood properties in *Eucalyptus globulus*. Tree Genet. Genomes **5**: 713-722.

Fusari, C. M., Kooke, R., Lauxmann, M. A., Annunziata, M. G., Enke, B., Hoehne, M., ... Stitt, M. (2017). Genome-wide association mapping reveals that specific and pleiotropic regulatory mechanisms fine-tune central metabolism and growth in Arabidopsis. The Plant Cell **29**(10): 2349-2373.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., Altshuler, D. (2002). The structure of haplotype blocks in the human genome. Science **296**: 2225–2229.

Gamal-Dien, O., Ratcliffe, B., Klápštˇe, J., Chen, C., Porth, I., El-Kasaby, Y.A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by sequencing. BMC Genomes **16**: 370.

Gao, H., Su, G., Janss, L., Zhang, Y.,Lund, M.S. (2013). Model comparison on genomic predictions using high density markers for different groups of bulls in the Nordic Holstein population. J. Dairy Sci. **96**: 4678–4687.

Garrido-Cardenas, J. A., Mesa-Valle, C., Manzano-Agugliaro, F. (2018). Trends in plant research using molecular markers. Planta **247**(3), 543-557.

Gleadow, R. M., Woodrow, I. E. (2002). Defense chemistry of cyanogenic *Eucalyptus cladocalyx* seedlings is affected by water supply. Tree Physiol. **22**(13): 939-945.

Gianola, D. (2013). Priors in whole-genome regression: The Bayesian alphabet returns. Genetics **194**: 573–596.

Gianola, D. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics **173**: 1761–1776.

Gienapp, P., Fior, S., Guillaume, F., Lasky, J.R., Sork, V.L., Csilléry, K. (2017). Genomic quantitative genetics to study evolution in the wild. Trends Ecol. Evol. **32**(12): 897-908.

Gion, J. M., Carouche, A., Deweer, S., Bedon, F., Pichavant, F., Charpentier, J.P., Baillères, H., Rozenberg, P., Carocha, V., Ognouabi, N., Verhaegen, D., Grima-Pettenati, J., Vigneron, P., Plomion, C. (2011). Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. BMC Genomics **12**: 301.

Goddard, M. E., Hayes, B. J., Meuwissen, T. H. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. Journal of animal breeding and genetics **128**(6): 409-421.

González-Recio, O.; Rosa, G.J.; Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livest. Sci. **166**: 217–231

Gorjanc, G., Gaynor, R. C., & Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theoretical and applied genetics, 131(9), 1953-1966.

Grattapaglia, D., Resende, M. D. (2011). Genomic selection in forest tree breeding. Tree Genet. Genom. **7**(2): 241-255.

Gupta, P.K., Pawan, S., Kulwal, P.L. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. **57**: 461–485.

Guerra, F. P., Wegrzyn, J. L., Sykes, R., Davis, M. F., Stanton, B. J., Neale, D. B. (2013). Association genetics of chemical wood properties in black poplar (Populus nigra). New Phytologist. **197**(1): 162-176.

Habier, D., Fernando, R.L., Kizilkaya, K., Garrick, D.J. (2011). Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics **12**: 186.

Hadfield, J.D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. J. Stat. Softw. **33**: 1–22

Hall, D., Hallingbäck, H.R., Wu, H.X. (2016). Estimation of number and size of QTL effects in forest tree traits. Tree Genet. Genomes **12**: 110.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S., Frazer, K.A. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol **10**: R32.

Hardner, C. M., Vaillancourt, R. E., Potts, B. M. (1996). Stand density influences outcrossing rate and growth of ppen-pollinated families of Eucalyptus globulus. Silvae Genet. **45**(4).

Hayward, A. C., Tollenaere, R., Dalton-Morgan, J., Batley, J. (2015). Molecular marker applications in plants. In Plant Genotyping, Humana Press, New York, NY. 13-27. pp.

He, S., Thistlethwaite, R., Forrest, K., Shi, F., Hayden, M. J., Trethowan, R., Daetwyler, H. D. (2019). Extension of a haplotype-based genomic prediction model to manage multi-environment wheat data using environmental covariates. Theor. Appl. Genetics **132**(11): 3143-3154.

He, S., Reif, J.C., Korzun, V., Bothe, R., Ebmeyer, E., Jiang, Y. (2017). Genomewide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. Theor. Appl. Genet. **130**:635–647

Henderson, C. R. (1984). Applications of linear models in animal breeding. University of Guelph, Ontario, Canada, 462 p.

Henderson, C. R. (1963) Selection index and expected genetic advance. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. NAS-NRC Publ No. 982, Washington/DC, pp. 141-163 pp.

Henderson, C. R. (1973). Sire evaluation and genetic trends. In: Anim breed genet symp in honor of J. Lush. Anim Sci Assoc Am, Champaign/IL. pp 10-41.

Henderson, C. R. (1977) Prediction of future records. In: Pollack E, Kempthorne O, Bailey T (eds) Proc Int Conf Quan Genet. Iowa State University Press, Ames/IA. pp 615-638.

Heslot, N., Yang, H. P., Sorrells, M. E., Jannik, J-L. (2012). Genomic selection in plant breeding: a comparison of models. Crop Sci. **52**(1): 146-160.

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B.B., Grondona, M., Zambelli, A., Windhausen, V.S., Mathews, K., Gorjac, G. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. **54**(4): 1476-1488.

Hill, W.G., Weir, B.S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. **33**: 54–78.

Hodgkinson, K., Pullman, D. (2010). Duty to warn and genetic disease. Can. J. Cardiovasc. Nurs. **20**(1):12-5.

Howard, N. P., Van De Weg, E., Bedford, D. S., Peace, C.P., Vanderzande, M.D.C., Clark, M.D., Teh, S.L., Cai, L., Luby, J.J. (2017). Elucidation of the 'Honeycrisp'pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (*Malus× domestica*) pedigree-connected SNP data set. Hortic. Res. **4**: 17003.

Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., ... & Lai, J. (2012). Comparative population genomics of maize domestication and improvement. Nature genetics **44**(7): 808-811.

Hung, T. D., Brawner, J. T., Meder, R., Lee, D. J., Southerton, S., Thinh, H. H., & Dieters, M. J. (2015). Estimates of genetic parameters for growth and wood properties in Eucalyptus pellita F. Muell. to support tree breeding in Vietnam. Annals of Forest Science, **72**(2), 205-217.

INFOR (Instituto Forestal, CL) (2009) El sector forestal chileno 2009. Instituto Forestal. Consultado el 20 sept. 2017. Disponible en http://www.infor.cl.

Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., ... Bouffier, L. (2016). Genomic selection in maritime pine. Plant Sci. **242**: 108-119.

Jensen, J., Su, G., Madsen, P. (2012). Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. BMC genetics **13**(1): 44.

Jiang, Y., Schmidt, R.H., Reif, J.C. (2018). Haplotype-based genome-wide prediction models exploit local epistatic interactions among markers. G3 Genes Genomes Genet. **8**:1687–1699.

Jónás, D., Ducrocq, V., Croiseau, P. (2017). The combined use of linkage disequilibrium– based haploblocks and allele frequency–based haplotype selection methods enhances genomic evaluation accuracy in dairy cattle. J. Dairy Sci. **100**(4): 2905-2908.

Karimi, K., Farid, A. H., Sargolzaei, M., Myles, S., Miar, Y. (2020). Linkage Disequilibrium, Effective Population Size and Genomic Inbreeding Rates in American Mink Using Genotyping-by-Sequencing Data. Front. Genet. **11**: 223.

Kelleher, C. T., Wilkin, J., Zhuang, J., Cortés, A. J., Quintero, Á. L. P., Gallagher, T. F., ... Ritland, K. (2012). SNP discovery, gene diversity, and linkage disequilibrium in wild populations of Populus tremuloides. Tree Genet. Genom. **8**(4): 821-829.

Kharabian-Masouleh, A., Waters, D.L.E., Reinke, R.F. Henry, R.J. (2011). Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. Plant Biotechnol. J. **9**: 1074–1085.

Klápšťe, J., Suontama, M., Telfer, E., Graham, N., Low, C., Stovold, T., McKinley, R., Dumgey, H. (2017). Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. Plos one **12**: e0185137.

Kumar, S., Banks, T. W., Cloutier, S. (2012). SNP Discovery through Next-Generation Sequencing and Its Applications. Int. J. Plant Genomics **2012**: 831460.

Lachance, J., Tishkoff, S. A. (2013). Population genomics of human adaptation. Annual review of ecology, evolution, and systematics **44**: 123-143.

Lan, S., Zheng, C., Hauck, K., McCausland, M., Duguid, S. D., Booker, H. M., Cloutier, S., You, F.M. (2020). Genomic Prediction Accuracy of Seven Breeding Selection Traits Improved by QTL Identification in Flax. Int. J. Mol. Sci. **21**(5): 1577.

Larsson, H., Källman, T., Gyllenstrand, N., Lascoux, M. (2013). Distribution of long-range linkage disequilibrium and Tajima's D values in Scandinavian populations of Norway spruce (Picea abies). G3: Genes Genom. Genet., **3**(5): 795-806.

Lee, S. H., Clark, S., Van Der Werf, J. H. (2017). Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. PloS one **12**(12): e0189775.

Legarra, A., Aguilar, I., Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. J. Dairy Sci. **92**(9): 4656-4663.

Lenz, P.R., Beaulieu, J., Mansfield, S.D., Clément, S., Desponts, M., Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced breeding population of black spruce (*Picea mariana*). BMC Genomes **18**: 335.

Lin, Z., Shi, F., Hayes, B. J., & Daetwyler, H. D. (2017). Mitigation of inbreeding while preserving genetic gain in genomic breeding programs for outbred plants. Theoretical and applied genetics, 130(5), 969-980.

Lin, Z., Cogan, N. O., Pembleton, L. W., Spangenberg, G. C., Forster, J. W., Hayes, B. J., & Daetwyler, H. D. (2016). Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. The Plant Genome, 9(1), 1-12.

Li, C.X.; Xu, W.G.; Guo, R.; Zhang, J.Z.; Qi, X.L.; Hu, L.; Zhao, M.Z. (2018). Molecular marker assisted breeding and genome composition analysis of Zhengmai 7698, an elite winter wheat cultivar. Sci. Rep. **8**: 322.

Li, Y., Ruperao, P., Batley, J., Edwards, D., Davidson, J., Hobson, K., & Sutton, T. (2017). Genome analysis identified novel candidate genes for ascochyta blight resistance in chickpea using whole genome re-sequencing data. Frontiers in plant science **8**: 359.

Lin, Z., Shi, F., Hayes, B. J., Daetwyler, H. D. (2017). Mitigation of inbreeding while preserving genetic gain in genomic breeding programs for outbred plants. Theor. Appl. Genet. **130**(5): 969-980.

Lin, Z., Cogan, N. O., Pembleton, L. W., Spangenberg, G. C., Forster, J. W., Hayes, B. J., Daetwyler, H. D. (2016). Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. The Plant Genom. **9**(1): 1-12

Liu, G., Zhao, Y., Gowda, M., Longin, F.H., Reif, J.C., Mette, M.F. (2016). Predicting Hybrid Performances for Quality Traits through Genomic-Assisted Approaches in Central European Wheat. PloS one **11**(7): e0158635.

Livingstone, D., Royaert, F., Stack, C., Mockaitis, K., May, G., Farmer, A., Saski, C., Schnell, R., Kuhn, D., Motamayor, J.C. (2015). Making a chocolate chip: development and evaluation of a 6K SNP array for *Theobroma cacao*. DNA Res. **22**(4): 279–291.

Lopez, G.A., Potts, B.M., Dutkowski, G.W. (2002). Genetic variation and inter-trait correlations in Eucalyptus globulus base population trials in Argentina. For. Genet. **9**: 217–231

Lorenz, A. J., Smith, K. P. (2015). Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci. **55**(6): 2657-2667.

Lorenz, A.J.; Smith, K.P.; Jannink, J.L. (2012). Potential and optimization of genomic selection for *Fusarium* head blight resistance in six-row barley. Crop Sci. **52**: 1609–1621.

Ly, D., Hamblin, M., Rabbi, I., Melaku, G., Bakare, M., Gauch, H. G., Okechukwu, R., Dixon, A.G.O., Kulakow, P., Jannink, J-L. (2013). Relatedness and genotype× environment

interaction affect prediction accuracies in genomic selection: a study in cassava. Crop Sci. **53**(4): 1312-1325.

Macciotta, N.P., Gaspa, G., Steri, R., Pieramati, C., Carnier, P., Dimauro, C. (2009). Pre-selection of most significant SNPS for the estimation of genomic breeding values. BMC Proc. **3**: 14

Machiela, M.J., Chanock, S.J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics **31**: 3555–3557.

Mafra, G.S.; Do Amaral Júnior, A.T.; Almeida, F.J.E.D., Vivas, M., Araújo Diniz-Santos, P.H., Saltires-Santos, J., Ferreira-Pena, G., De Lima, V.J., Kamphorst, S.H., De Oliveira, F.T., Pequeño de Souza, Y., Schwantes, I.A., De Oliveira-Santos, T., Rosimeiri-Barbosa, B., Maldonado, C., Mora, F. (2019). SNP-based mixed model association of growth- and yield-related traits in popcorn. Plos one **14**: e0218552.

Maldonado, C., Mora, F., Scapim, C.A., Coan, M. (2019). Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: hapLA4. Plos one **14**: e0212925.

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity **108**: 285.

Mathew, B., Léon, J., Sillanpää, M. J. (2018). A novel linkage-disequilibrium corrected genomic relationship matrix for SNP-heritability estimation and genomic prediction. Heredity **120**(4): 356.

Matias, F.I., Galli, G., Correia Granato, I.S., Fritsche-Neto, R. (2017). Genomic prediction of autogamous and allogamous plants by SNPs and haplotypes. Crop Sci. **57**: 2951–2958.

McDonald, M. W., Rawlings, M., Butcher, P. A., Bell, J. C. (2003). Regional divergence and inbreeding in Eucalyptus cladocalyx (Myrtaceae). Australian Journal of Botany, **51**(4): 393-403.

Meuwissen, T. H., Odegard, J., Andersen-Ranberg, I., Grindflek, E. (2014). On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. Genet. Sel. Evol. **46**(1): 1-8.

Meuwissen, T.H., Hayes, B.J., Goddard, M.E. (2001). Prediction of total genetic value using genome wide dense marker maps. Genetics **157**: 1819–1829.

Missiaggia, A. A., Piacezzi, A. L., Grattapaglia, D. (2005) Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. Tree Genet. Genomes **1**: 79-84.

Monclus, R., Leplé, J.C., Bastien, C., Bert, P.F., Villar, M., Marron, N., Brignolas, F., Jorge, V. (2012). Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp. BMC Plant Biol **12**: 173.

Mora, F., Ballesta, P., Serra, N. (2019). Bayesian analysis of growth, stem straightness and branching quality in full-sib families of *Eucalyptus globulus*. Bragantia **78**(3): 328-336.

Mora, F., Arriagada, O., Ballesta, P., Ruiz, E. (2017). Genetic diversity and population structure of a drought-tolerant species of Eucalyptus, using microsatellite markers. Journal of Plant Biochemistry and Biotechnology, **26**(3): 274-281.

Mora, F., Serra, N. (2014). Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. Tree Genet. Genomes **10**: 711-719.

Mora, F., Gleadow, R., Perret, S., Scapim, C.A. (2009). Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. Euphytica **169**(3): 335-344.

Morales, M., Aroca, G., Rubilar, R., Acuña, E., Mola-Yudego, B., González-García, S. (2015). Cradle-to-gate life cycle assessment of *Eucalyptus globulus* short rotation plantations in Chile. J. Clean. Prod. **99**: 239-249.

Morrell, P. L., Gonzales, A. M., Meyer, K. K., & Clegg, M. T. (2014). Resequencing data indicate a modest effect of domestication on diversity in barley: a cultigen with multiple origins. Journal of Heredity **105**(2): 253-264.

Müller, D., Schopp, P., Melchinger, A. E. (2018). Selection on expected maximum haploid breeding values can increase genetic gain in recurrent genomic selection. G3: Genes Genom. Genet. **8**(4): 1173-1181.

Müller, B.S., Neves, L.G., De Almeida, F.J.E., Resende, M.F.R., Jr., Muñoz, P.R., Dos Santos, P.E.T., Paludzyszyn Filho, E., Kirst, M., Grattapaglia, D. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. BMC Genomes **18**: 524.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., …, Schmutz, J. (2014). The genome of *Eucalyptus grandis*. Nature **510**: 356.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., Baloch, F.S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol. Biotech. Eq. **32**(2): 261-285.

Neale, D.B., Kremer, A. (2011). Forest tree genomics: growing resources and applications. Nature Rev. Genetics **12**: 111–122.

Nelson, J. C., Wang, S., Wu, Y., Li, X., Antony, G., White, F. F., Yu, J. (2011). Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. BMC Genom. **12**(1): 352.

Noble, T. J., Tao, Y., Mace, E. S., Williams, B., Jordan, D. R., Douglas, C. A., Mundree, S. G. (2018). Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. Front. Plant Sci. **8**: 2102.

Nordborg, M.; Tavaré, S. (2002). Linkage disequilibrium: What history has to tell us. Trends Genet. **18**: 83–90.

Norman, A., Taylor, J., Edwards, J., & Kuchel, H. (2018). Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. G3: Genes Genom. Genet. **8**(9): 2889-2899.

Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., Meuwissen, T. H. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). Front. Genetics **5**: 402.

Olson, M. S., Robertson, A. L., Takebayashi, N., Silim, S., Schroeder, W. R., Tiffin, P. (2010). Nucleotide diversity and linkage disequilibrium in balsam poplar (Populus balsamifera). New Phytol. **186**(2): 526-536.

Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U.R., Stegmeir, T., Sebolt, A., Gilmore, B., Lawley, C., Mockler, T.C., Bryant, D.W., Wilhelm, L., Lezzoni, A. (2012). Development and Evaluation of a GenomeWide 6K SNP Array for Diploid Sweet Cherry and Tetraploid Sour Cherry. Plos one **7**(12): e48305.

Paiva, J. A., Prat, E., Vautrin, S., Santos, M.D., San-Clemente, H.S., Brommonschenkel, S., Fonseca, P.G.S., Grattapaglia, D., Song, X., Ammiraju, J.S.S., Kudrna, D., Wing, R.A., Freitas, A.T., Bergès, H., Grima-Pettenati, J. (2011). Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. BMC genomics **12**(1): 137.

Paszkiewicz, K., Studholme, D.J. (2012) High-throughput sequencing data analysis software: current state and future developments. In Bioinformatics for high throughput sequencing (Rodriguez-Ezpeleta, N., Hackenberg, M. and Aransay, A.M., eds), Springer Science, New York, pp. 231– 248.

Patterson, B., Potts, B., Vaillancourt, R. (2001). Variation in outcrossing rates within the canopy of Eucalyptus globulus. In Proceedings Developing the Eucalypts of the Future. IUFRO Int Symp (pp. 10-15).

Pérez, P., de Los Campos, G. (2014). Genome-wide regression & prediction with the BGLR statistical package. Genet. Soc. Am. **198**: 483-495

Pérez, P., de los Campos, G., Crossa, J., Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. Plant Genome **3**: 106–116.

Pikunova, A., Madduri, M., Sedov, E., Noordijk, Y. Peil, A., Troggio, M., Bus, V.G.M., Visser, R.G.F., Van De Weg, E. (2014). "Schmidt"s Antonovka' is identical to "Common Antonovka", an apple cultivar widely used in Russia in breeding for biotic and abiotic stresses. Tree Genet. Genomes **10**:261–271.

Porebski, S., Bailey, L. G., Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rep. **15**(1): 8-15.

Porto-Neto, L.R., Kijas, J.W., Reverter, A. (2014). The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genet. Sel. Evol. **46**: 22.

Potts, B. M., McGowen, M. H., Williams, D. R., Suitor, S., Jones, T. H., Gore, P. L., Vaillancourt, R. E. (2008). Advances in reproductive biology and seed production systems of Eucalyptus: the case of Eucalyptus globulus. J. For. Sci. **70**(2): 145-154.

Pound, L. M., Wallwork, M. A. B., Potts, B. M., Sedgley, M. (2002). Early ovule development following self-and cross-pollinations in Eucalyptus globulus Labill. ssp. globulus. Ann. Bot. **89**(5): 613-620.

Pound, L. M., Patterson, B., Wallwork, M. A. B., Potts, B. M., Sedgley, M. (2003). Pollen competition does not affect the success of self-pollination in Eucalyptus globulus (Myrtaceae). Aust. J. Bot. **51**(2): 189-195.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Pryce, J.E., Johnston, J., Hayes, B.J., Sahana, G., Weigel, K.A., McParland, S., Spurlock, D., Krattenmacher, N., Spelman, R.J., Wall, E., Calus, M.P.L. (2014). Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research

herds in Europe, North America, and Australasia using 2 reference populations. J. Dairy Sci. **97**: 1799–1811.

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R.K., He, Z. (2017). Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. Mol. Plant **10**: 1047–1064.

Ratcliffe, B., El-Dien, O.G., Klápšťe, J., Porth, I., Chen, C., Jaquish, B., El-Kasaby, Y.A. (2015). A comparison of genomic selection models across time in interior spruce (*Picea engelmannii × glauca*) using unordered SNP imputation methods. Heredity **115**: 547–555.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. Heredity **119**(4): 245-255.

Resende, M.D., Resende, M.F., Sansaloni, C.P., Petroli, C.D., Missiaggia, A.A., Aguiar, A.M., Abad, J.M., Takahashi, E.K., Rosado, A.M., Faria, D.A., Pappas, G.J. Jr., Kilian, A., Grattapaglia, D. (2012a). Genomic selection for growth and wood quality in *Eucalyptus*: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol. **194**: 116–128.

Resende, M. F., Muñoz, P., Resende, M. D., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., Kirst, M. (2012b). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics **190**(4): 1503-1510.

Rodríguez, M., Rau, D., O'Sullivan, D., Brown, A.H., Papa, R., Attene, G. (2012). Genetic structure and linkage disequilibrium in landrace populations of barley in Sardinia. Theor. Appl. Genet. **125**(1): 171-184.

Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., Sorrells, M. E. (2015). Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. The Plant Genom. **8**(2): 1-10.

Sallam, A. H., Conley, E., Prakapenka, D., Da, Y., Anderson, J. A. (2020). Improving Prediction Accuracy Using Multi-allelic Haplotype Prediction and Training Population Optimization in Wheat. G3: Genes Genom. Genet. **10**(7): 2265-2273.

Sapkota, S., Boyles, R., Cooper, E., Brenton, Z., Myers, M., Kresovich, S. (2020). Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. Crop Sci. **60**(1): 132-148.

Schmit, R., Mora, F., Emhart, V. I., Rubilar, R. (2015). Longitudinal analysis in the selection of *Eucalyptus globulus* clones under contrasting water availability conditions. Sci. For. **43**(105): 217-224.

Schopp, P., Müller, D., Technow, F., Melchinger, A. E. (2017). Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. Genetics **205**(1): 441-454

Scutari, M., Mackay, I., & Balding, D. (2016). Using genetic distance to infer the accuracy of genomic prediction. PLoS Genet. **12**(9): e1006288.

Seddon, J. M., Parker, H. G., Ostrander, E. A., Ellegren, H. (2005). SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Mol. Ecol. **14**(2): 503–511.

Senhorinho, H.J.C., Coan, M.M.D., Marino, T.P., Kuki, M.C., Barth-Pinto, R.S., Scapim, C.A., Holland, J.B. (2019). Genomic-Wide Association Study of Popping Expansion in Tropical Popcorn and Field Corn Germplasm. Crop Sci. **59**: 2007–2019.

Silva, F.F., Jerez, E.A.Z., De Resende, M.D.V., Soriano-Viana, M., Ferreira-Azevedo, C., Lopes, P.S., Nascimento, M., Oliveira de Lima, R., Facioni-Guimaraes, S.E. (2018). Bayesian model combining linkage and linkage disequilibrium analysis for low density-based genomic selection in animal breeding. J. Appl. Anim. Res. **46**: 873–878.

Silva-Junior, O.B., Faria, D.A., Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. New Phytol. **206**: 1527–1540.

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R.P., Poland, J. (2019). High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. Front. Plant Sci. **10**: 394.

Singh, N., Kumar, P. J., Panda, K., Mandal, P., Kumar, V., Singh, B., Mishra, S., Singh, Y., Singh, R., Rai, V., Gupta, A., Sharma, T.R., Kumar-Singh, N. (2015). Single-copy gene based 50K SNP chip for genetic studies and molecular breeding in rice. Sci. Rep. **5**: 11600.

Song, J., Brendel, O., Bodénès, C., Plomion, C., Kremer, A., Colin, F. (2017). X-ray computed tomography to decipher the genetic architecture of tree branching traits: Oak as a case study. Tree Genet. Genomes **13**: 5

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L., Cregan, P.B. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. Plos one **8**: e54985.

Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Meuwissen, T.H.E. (2009). Reducing dimensionality for prediction of genome-wide breeding values. Genet. Sel. Evol. **41**: 29.

Sonesson, A. K., Woolliams, J. A., Meuwissen, T. H. (2012). Genomic selection requires genomic control of inbreeding. Genet. Sel. Evol. **44**(1), 27.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. J. Roy. Stat. Soc **64**(4): 583-639.

Steane, D. A., Nicolle, D., Sansaloni, C. P., Petroli, C. D., Carling, J., Kilian, A., Myburg, A.A., Grattapaglia, D., Vaillancourt, R. E. (2011). Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. Mol. Phylogenet. Evol. **59**(1), 206-224.

Stejskal, J., Lstibůrek, M., Klápště, J., Čepl, J., & El-Kassaby, Y. A. (2018). Effect of genomic prediction on response to selection in forest tree breeding. Tree Genetics & Genomes, 14(5), 74.

Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., Chen, F. (2020). The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnol. J.

Sun, X., Fernando, R., Dekkers, J. (2016). Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. Genet. Sel. Evol. **48**(1): 77.

Sun, Y., Lu, Y., Xie, L., Deng, Y., Li, S., Qin, X. (2015). Interferon gamma polymorphisms and hepatitis B virus-related liver cirrhosis risk in a Chinese population. Cancer Cell Int. **15**(1): 35.

Suontama, M., Klápštˇe, J., Telfer, E., Graham, N., Stovold, T., Low, C., McKinley, R., Dungey, H. (2019). Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. Heredity **122**: 370.

Suontama, M., Low, C. B., Stovold, G. T., Miller, M. A., Fleet, K. R., Li, Y., Dungey, H. S. (2015). Genetic parameters and genetic gains across three breeding cycles for growth and form traits of Eucalyptus regnans in New Zealand. Tree Genet. Genomes **11**(6): 133.

Supple, M. A., Bragg, J. G., Broadhurst, L. M., Nicotra, A. B., Byrne, M., Andrew, R. L., ... Borevitz, J. O. (2018). Landscape genomic prediction for restoration of a Eucalyptus foundation species under climate change. Elife 7: e31835.

Tahir, M. A., Bhatti, H. N., Iqbal, M. (2016). Solar red and brittle blue direct dyes adsorption onto *Eucalyptus angophoroides* bark: equilibrium, kinetics and thermodynamic studies. J. Environ. Chem. Eng. **4**(2): 2431-2439.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**(3): 585-595.

Tan, B., Grattapaglia, D., Wu, H. X., & Ingvarsson, P. K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid Eucalyptus. Plant science **267**: 84-93.

Tan, B., Grattapaglia, D., Martins, G.S., Ferreira, K.Z., Sundberg, B., Ingvarsson, P.K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. BMC Plant Biol. **17**: 110.

Thavamanikumar, S., McManus, L.J., Ades, P.K., Bossinger, G., Stackpole, D.J., Kerr, R., Hadjigol, S., Freeman, J.S.; Vaillancourt, R.E., Zhu, P., Tibbits, J.F.G. (2014). Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp. *globulus* Labill identifies nine stable marker-trait associations for seven traits. Tree Genet. Genomes **10**: 1661–1678.

Thistlethwaite, F. R., Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., ... El-Kassaby, Y. A. (2020). Linkage disequilibrium vs. pedigree: Genomic selection prediction accuracy in conifer species. Plos one **15**(6): e0232201.

Thorwarth, P., Ahlemeyer, J., Bochard, A. M., Krumnacker, K., Blümel, H., Laubach, E., ... Schmid, K. J. (2017). Genomic prediction ability for yield-related traits in German winter barley elite material. Theor. Appl. Genet. **130**(8): 1669-1683.

Thumma, B. R., Baltunis, B. S., Bell, J. C., Emebiri, L.C., Moran, G.F., Southerton, S.G. (2010) Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. Tree Genet. Genomes **6**: 877-889.

Thumma, B. R., Nolan, M. R., Evans, Moran, G.F. (2005). Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics **171**: 1257-1265.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. R. Stat. Soc. **58**(1): 267–288

Tiezzi, F., Parker-Gaddis, K. L., Cole, J. B., Clay, J. S., Maltecca, C. (2015). A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. Plos one **10**(2): e0114919.

Torres, L.G., Rodrigues, M.C., Lima, N.L., Trindade, T.F.H., e Silva, F.F., Azevedo, C.F., DeLima, R.O. (2018). Multi-trait multi-environment Bayesian model reveals G × E interaction for nitrogen use efficiency components in tropical maize. Plos one **13**: e0199492.

Torres-Dini, D., Nunes, A.C.P., Aguiar, A., Nikichuk, N., Centurión, C., Cabrera, M., Moraes, M.L.T., Resende, M.D.V., Sebbenn, A.M. (2016). Clonal selection of *Eucalyptus*

grandis x *Eucalyptus globulus* for productivity, adaptability, and stability, using SNP markers. Silvae Genet. **65**: 30–38.

Valenzuela, C.E.; Ballesta, P.; Maldonado, C.; Baettig, R.; Arriagada, O.; Sousa Mafra, G.; Mora, F. (2019). Bayesian mapping reveals large-effect pleiotropic QTLs for wood density and slenderness index in 17-year-old trees of *Eucalyptus cladocalyx*. Forests **10**: 241.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci. **91**: 4414–4423.

Vargas-Reeve, F., Mora, F., Perret, S., Scapim, C.A. (2013). Heritability of stem straightness and genetic correlations in Eucalyptus cladocalyx in the semi-arid region of Chile. Crop Breed. Appl. Biot. 13(2): 107-112.

Velazco, J. G., Malosetti, M., Hunt, C. H., Mace, E. S., Jordan, D. R., Van Eeuwijk, F. A. (2019). Combining pedigree and genomic information to improve prediction quality: an example in sorghum. Theor. Appl. Genet. **132**(7): 2055-2067.

Viana, J., Pereira, H.D., Mundim, G.B., Piepho, H.-P., Fonseca e Silva, F. (2018). Efficiency of genomic prediction of non-assessed single crosses. Heredity **120**: 283–295.

Viana, J. M. S., Sobreira, F. M., De Resende, M. D. V., Faria, V.R. (2010). Multi- trait BLUP in half- sib selection of annual crops. Plant Breed. **129**(6): 599-604.

Villumsen, T. M., Janss, L. (2009). Bayesian genomic selection: the effect of haplotype length and priors. BMC Proc. 3: 1-11.

Villumsen, T. M., Janss, L., Lund, M. S. (2009). The importance of haplotype length and heritability using genomic selection in dairy cattle. J. Anim. Breed. Genet. **126**(1): 3-13. Volpato, L., Alves, R.S., Teodoro, P.E., de Resende, M.D.V., Nascimento, M., Nascimento, A.C.C., Ludke, W.H., da Silva, F.L., Borém, A. (2019). Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. Plos one **14**: e0215315.

Voorrips, R. E., Bink, M. C., Kruisselbrink, J. W., Koehorst-van Putten, H.J.J., Van De Weg, W.E. (2016). PediHaplotyper: software for consistent assignment of marker haplotypes in pedigrees. Mol. Breed. **36**(8): 119.

Wang, X., Xu, Y., Hu, Z., Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. The Crop J. **6**(4): 330-340.

Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F., Xiang, J. (2017). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp Litopenaeus vannamei. BMC Genet. **18**(1): 1-9.

Wang, Y., Ren, X., Sun, D., & Sun, G. (2015). Origin of worldwide cultivated barley revealed by NAM-1 gene and grain protein content. Frontiers in plant science **6**: 803.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. 7(2): 256-276.

Weber, K. L., Thallman, R. M., Keele, J. W., Snelling, W.M., Bennett, G.L., Smith, T.P., McDaneld, T.G., Allan, M.F., Van Eenennaam, A.L., Kuehn, L.A. (2012). Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. J. Anim. Sci. **90**(12): 4177-4190.

Wientjes, Y. C., Veerkamp, R. F., Calus, M. P. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics **193**(2): 621-631.

White, T. L., Adams, W. T., Neale, D. B. (2007) Forest genetics. CABI Publishing CAB International, Cambridge, USA. 682 p.

Wolc, A., Zhao, H. H., Arango, J., Settar, P., Fulton, J. E., O'sullivan, N. P., ... & Garrick, D. J. (2015). Response and inbreeding from a genomic selection experiment in layer chickens. Genet. Sel. Evol. **47**(1): 59.

Wolfe, M.D., Del Carpio, D.P., Alabi, O., Ezenwaka, L.C., Ikeogu, U.N., Kayondo, I.S., Lozano, R., Okeke, U.G., Ozimati, A.A., Williams, E., Egesi, C., Kawuki, R.S., Kulakow, P., Rabbi, I.Y., Jannink, J.L. (2017). Prospects for genomic selection in cassava breeding. Plant Genome **10**: 1–9.

Wu, X.-L., Xu, J., Feng, G., Wiggans, G.R., Taylor, J.F., He, J., Qian, C., Qiu, J., Simpson, B., Walker, J., Bauck, S. (2016). Optimal Design of Low-Density SNP Arrays for Genomic Prediction: Algorithm and Applications. Plos one **11**: e0161719.

Wu, X., Lund, M.S. Sun, D. Zhang, Q., Su, G. (2015). Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. J. Anim. Breed. Genet. **132**: 366–375.

Yang, H., Liu, T., Xu, B. (2015). QTL detection for growth and form traits in three full-sib pedigrees of *Pinus elliottii* var. *elliottii* × *P. caribaea* var. *hondurensis* hybrids. Tree Genet. Genomes **11**: 130.

Yamamoto, T., Nagasaki, H., Yonemaru, J. I., Ebana, K., Nakajima, M., Shibaya, T., Yano, M. (2010). Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single nucleotide polymorphisms. BMC Genomics **11**(1): 267.

Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., Xiao, J., Zhang, Q. (2011). Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. Plos one **6**(3): e17595.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**(2): 203.

Zapata-Valenzuela, J., Whetten, R. W., Neale, D., (2012). Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. G3: Genes, Genomes, Genet. **3**(5): 909-916.

Zapata-Valenzuela, J., Hasbun Zaror, R. (2011). Accelerated forest genetic breeding using genomic selection. Bosque **32**(3): 209-213

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., ... Olsen, M. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. Front. Plant Sci. **8**: 1916.

6. ANEXO I: ESQUEMATIZACION DE LOS ENSAYOS DE PROGENIE *E. GLOBULUS Y E. CLADOCALYX*
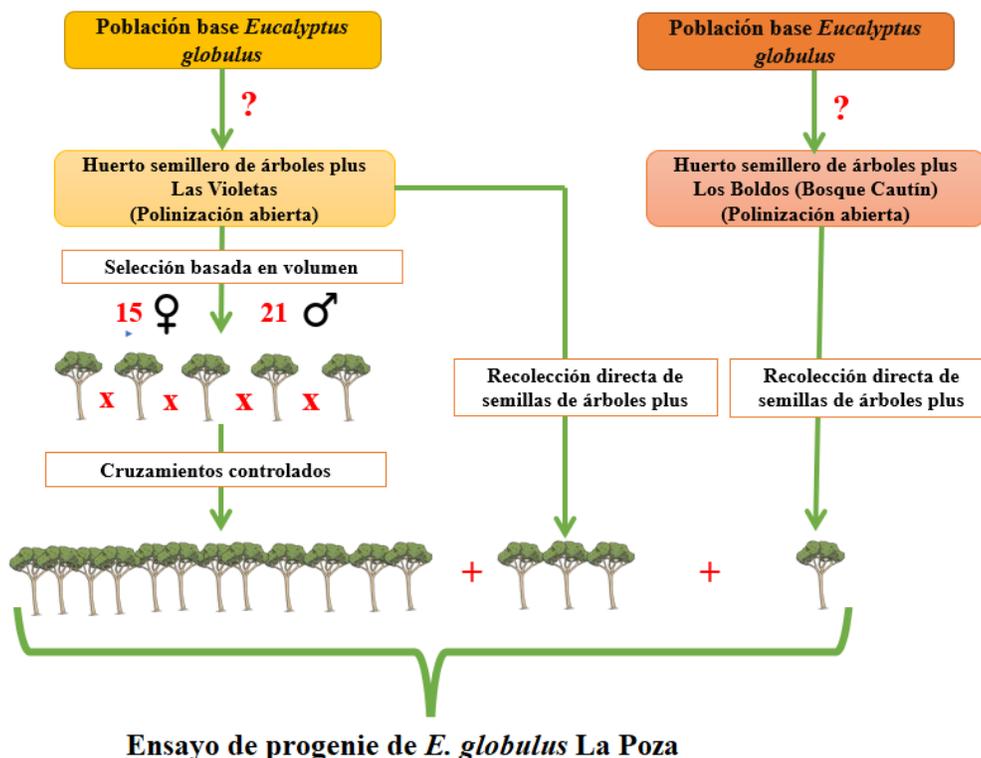
Figura S1. Historial de selección y estructura genética del ensayo La Poza, de árboles de *E. globulus*. Los árboles plantados provienen de tres principales fuentes: 1) Cruzamiento controlados entre 15 y 21 líneas maternas y paternas, respectivamente, que provienen de un huerto semillero de árboles plus (Las Violetas), los cuales representan a 62 familias de hermanos completos. 2) Semillas de recolección directa desde árboles plus del huerto semillero Las Violetas, las cuales representan a 2 dos familias de medios hermanos. 3) Semillas de recolección directa desde árboles plus del huerto semillero Los Boldos, las cuales representan a una familia de medios hermanos.
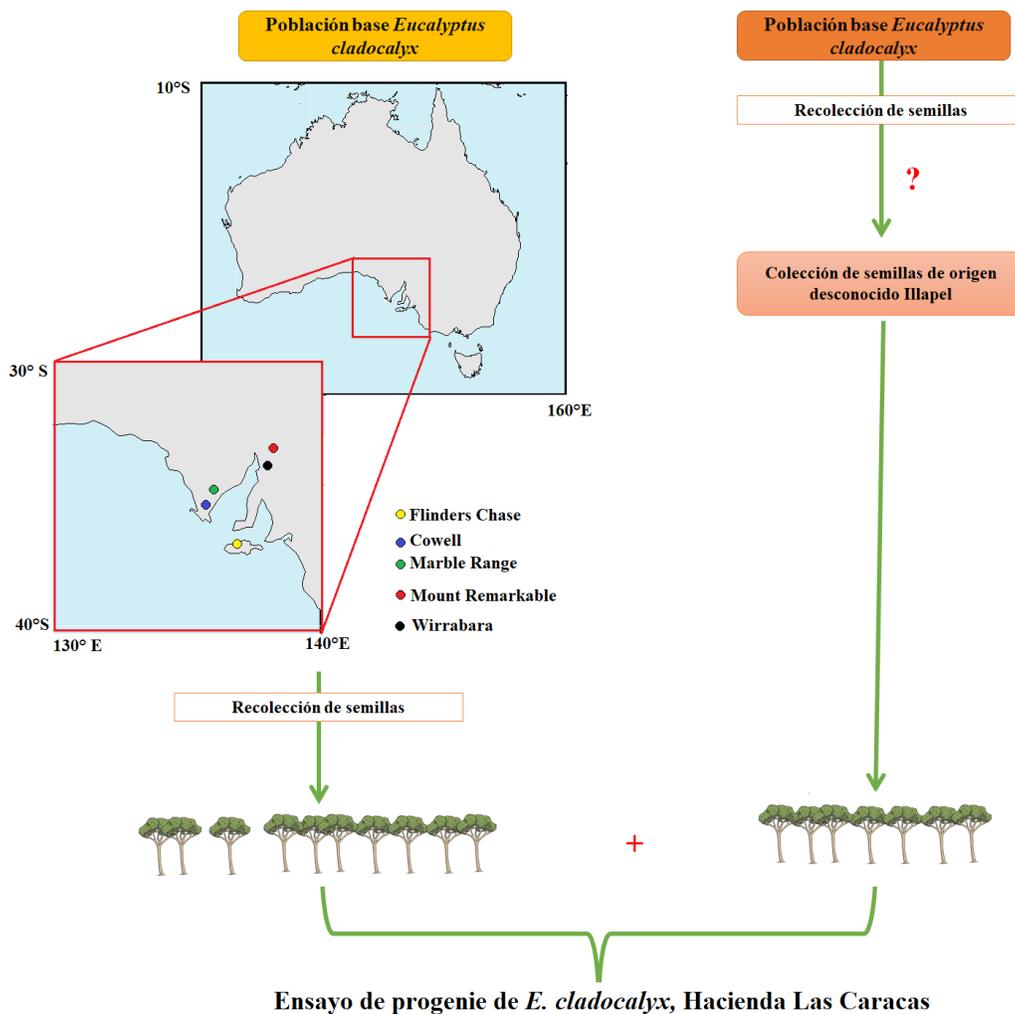
Figura S2. Historial de selección y estructura genética del ensayo Hacienda Las Caracas de *E. cladocalyx*. Los árboles plantados provienen principalmente de dos fuentes: 1) Recolección directa de semillas de árboles (de línea paterna desconocida) provenientes de cinco procedencias al sur de Australia, las cuales son Flinders Chase (Isla Kangaroo), Cowell (Península Eyre), Marble Range (Península Eyre), Mount Remarkable (Flinder Ranges) y Wirrabara (Flinder Ranges). Los árboles provenientes de esta fuente representan a 47 familias de medios hermanos. 2) Semillas recolectadas de árboles previamente plantados en Illapel, Provincia de Choapa, Chile (origen natural desconocido). Los árboles provenientes de esta fuente representan a dos familias de medios hermanos.

Tabla S1. Descripción del origen de los árboles de *E. cladocalyx* y el número de familias que representan a cada procedencia, plantados en el ensayo Las Caracas, Los Vilos, Chile.

Procedencia	Región	Latitud	Longitud	Número de Familias
Mount Remarkable	Flinder Ranges	32°43'S	138°06'E	16
Cowell	Península Eyre	33°38'S	136°40'E	10
Marble Range	Península Eyre	34°30'S	135°30'E	4
Wirrabara	Flinder Ranges	33°06'S	138°14'E	9
Flinders Chase	Isla Kangaroo	35°57'S	136°42'E	8
Illapel	Choapa, Chile	31°40'S	71°14'O	2

7. ANEXO II: DESCRIPCIÓN SUPLEMENTARIA DE MATERIALES Y MÉTODOS

La expresión matricial (4) es presentada en forma escalar por la expresión (4.1):

$$y_{ij} = \mu + b_i + a_j + e_{ij} \tag{4.1}$$

donde y_{ij} es el fenotipo del i-ésimo bloque y del j-ésimo árbol. y_{ij} depende del efecto de la media general (μ), del efecto del i-ésimo bloque (b_i) y del efecto genético aditivo del j-ésimo árbol. El término e_{ij} corresponde a al residuo para cada observación en y_{ij}.

La expresión matricial (5) es presentada en forma escalar por la expresión (5.1):

$$y_{ij} = \beta_0 + \beta_i + Z_1 m_1 + Z_2 m_2 + Z_3 m_3 + \cdots Z_n m_n + e_{ij} \tag{5.1}$$

donde y_{ij} es el fenotipo del i-ésimo bloque y del j-ésimo marcador. y_{ij} depende de una media general (β_0), del efecto del i-ésimo bloque (β_i) y del efecto aditivo del j-ésimo marcador de 1,…,n (m_n).

8. ANEXO III: BONDAD DE AJUSTE DE LOS MODELOS DE PREDICCIÓN GENÓMICA EN *E. GLOBULUS* Y *E. CLADOCALYX*

Tabla S2. Criterio de Información de la Devianza (DIC) para cada modelo de predicción genómica en *E. globulus*, basados en: haplotipos (HAP), polimorfismos de nucleótido único (SNP) y haplotipos en conjunto con SNPs no asignados a haplotipos.

Característica/Modelo	Bayes A	Bayes B	Bayes C	BRR*
Altura del árbol				
SNP	2737,3	2736,0	2736,9	2734,7
HAP	2747,6	2746,7	2744,3	2743,8
HAP-SNP	2739,1	2735,0	2737,0	2737,4
Diámetro a la altura de pecho				
SNP	3277,0	3277,0	3276,4	3276,8
HAP	3278,1	3278,2	3278,9	3279,2
HAP-SNP	3276,8	3277,0	3277,0	3277,1
Rectitud del fuste				
SNP	1669,1	1668,7	1668,7	1668,1
HAP	1668,3	1668,7	1667,5	1667,5
HAP-SNP	1668,1	1667,7	1667,9	1667,8
Calidad de ramas				
SNP	1995,2	1994,8	1994,8	1997,1
HAP	1989,3	1990,6	1990,3	1995,3
HAP-SNP	1993,7	1995,0	1993,0	1993,7
Densidad de madera				
SNP	2665,2	2664,9	2665,6	2664,7
HAP	2667,1	2668,0	2667,1	2668,1
HAP-SNP	2666,2	2665,9	2665,3	2666,1

*Regresión Bayesiana contraída

Tabla S3. Criterio de Información de la Devianza (DIC) para cada modelo de predicción genómica en *E. cladocalyx*, basados en: haplotipos (HAP), polimorfismos de nucleótido único (SNP) y haplotipos en conjunto con SNPs no asignados a haplotipos.

Característica/Modelo	Bayes A	Bayes B	Bayes C	BRR*
Altura del árbol				
SNP	2015,1	2000,4	2006,3	2007,5
HAP	2016.7	2007,8	2006,8	2006,8
HAP-SNP	2011,7	2008,4	2008,6	2009,3
Diámetro a la altura de pecho				
SNP	2562,1	2553,2	2554,6	2552,5
HAP	2560,0	2554,0	2555,6	2550,6
HAP-SNP	2563,8	2560,4	2561,5	2560,1
Rectitud del fuste				
SNP	969,7	968,5	967,7	966,5
HAP	964,8	961,6	963,4	967,5
HAP-SNP	931,7	933,9	933,1	936,9
Altura de la primera bifurcación				
SNP	1508,2	1503,4	1504,2	1502,1
HAP	1510,3	1508,8	1508,9	1510,1
HAP-SNP	1511,4	1511,6	1513,4	1510,4
Densidad de madera				
SNP	2139,6	2131,1	2143,1	2136,2
HAP	2138,6	2136,1	2133,4	2137,2
HAP-SNP	2140,2	2136,4	2135,0	2137,6
Intensidad de floración				
SNP	1368,2	1359,1	1358,2	1355,2
HAP	1366,0	1366,0	1365,9	1365,8
HAP-SNP	1360,7	1364,0	1361,7	1361,4

*Regresión Bayesiana contraída

9. ANEXO IV: PUBLICACION RELACIONADAS AL TRABAJO DE TESIS

plants

MDPI

SNP and Haplotype-Based Genomic Selection of Quantitative Traits in *Eucalyptus globulus*

Paulina Ballesta [1], Carlos Maldonado [1], Paulino Pérez-Rodríguez [2] and Freddy Mora [1,*]

1   Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile
2   Colegio de Postgraduados, Statistics and Computer Sciences, Montecillos, Edo. de México 56230, Mexico
*   Correspondence: fmora@utalca.cl; Tel.: +56-71-2200280

check for updates

Abstract: *Eucalyptus globulus* (Labill.) is one of the most important cultivated eucalypts in temperate and subtropical regions and has been successfully subjected to intensive breeding. In this study, Bayesian genomic models that include the effects of haplotype and single nucleotide polymorphisms (SNP) were assessed to predict quantitative traits related to wood quality and tree growth in a 6-year-old breeding population. To this end, the following markers were considered: (a) ~14 K SNP markers (SNP), (b) ~3 K haplotypes (HAP), and (c) haplotypes and SNPs that were not assigned to a haplotype (HAP-SNP). Predictive ability values (PA) were dependent on the genomic prediction models and markers. On average, Bayesian ridge regression (BRR) and Bayes C had the highest PA for the majority of traits. Notably, genomic models that included the haplotype effect (either HAP or HAP-SNP) significantly increased the PA of low-heritability traits. For instance, BRR based on HAP had the highest PA (0.58) for stem straightness. Consistently, the heritability estimates from genomic models were higher than the pedigree-based estimates for these traits. The results provide additional perspectives for the implementation of genomic selection in *Eucalyptus* breeding programs, which could be especially beneficial for improving traits with low heritability.

Keywords: genomic prediction; haplotype blocks; predictive ability; Bayesian models

1. Introduction

The number of breeding programs that use the principles of genomic selection (GS) has increased considerably in recent years. Over the last decade, several investigations have illustrated how to incorporate the principles of genomic prediction in the genetic improvement programs of complex traits [1–10]. In this sense, the use of GS principles has been facilitated by the development of genotyping techniques such as genotyping by sequencing (GBS) and DNA arrays (or chips) of high density in various crops including self-pollinated plants such as soybean; wheat; barley; rice [10–13]; and outcrossing plants such as fruit trees, forest trees, and corn, among others [14–17]. Single nucleotide polymorphic markers (SNP) have been a powerful tool in breeding programs for different agricultural crops [18–23]. SNP markers have multiple applications in plants, including positional cloning, whole genome association studies, the mapping of quantitative trait loci (QTL), and the determination of genetic relationships between individuals.

The non-random association between two or more loci due to a low probability of recombination between them (linkage disequilibrium: LD) in a given population constitutes haplotypes [24,25], which correspond to sets of genomic regions within a chromosome that tend to be inherited together [26]. In this context, GS can be implemented not only using individual SNPs but also haplotypes, or a combination of both, using haplotypes in conjunction with SNPs not assigned to a haplotype. For example, Cuyabano et al. [27] presented a haplotype approach for genomic prediction using high-density data in dairy cattle as an alternative to individual marker methods, demonstrating

that the haplotypes improved prediction accuracy compared to an individual SNP. On the other hand, Calus et al. [28] determined that the inclusion of haplotypes in genomic prediction models was beneficial for low-heritability traits. Matias et al. [29] found that the use of haplotypes in the prediction of complex traits of maize increased the predictive ability by 20%. From these and other studies, it appears that the haplotype approach emerges as a methodological variant that can improve not only the predictive abilities but also the precision in the detection of genomic regions in association studies [30–33]. Moreover, given that a haplotype is defined as a set of nearby SNPs, which are in strong LD [19], this analytical approach would take the natural dependence that exists between SNPs into account, which becomes more relevant when considering high-density DNA arrays.

Several studies have emphasized the use of haplotypes (from high-density DNA arrays) to estimate the predictive ability of different GS models [27,34–38]. One of the advantages of using haplotypes in GS is the ability to detect and include mutations as genomic information. According to Curtis et al. [39], when mutations have occurred, it is possible that the frequencies of the alleles remain (almost) unchanged. However, when analyzing haplotypes, mutations at different loci tend to cause significant changes in haplotype frequencies. Therefore, a QTL that is not in complete LD with an individual marker can be in complete LD with a particular haplotype [27]. Additionally, the use of haplotypes reduces the degrees of freedom in the models of prediction or genomic association (reduction of dimensionality), which contributes to greater precision in the detection of QTL [40]. It should be noted that there are few studies that have evaluated the combined use of GS and haplotypes in plants [29,41,42]. In particular, these methods have been implemented in predominantly self-pollinated (autogamous) species—for example, soybean and wheat [41,42]—in which extensive LD values can be found in their genomes, which favors the identification of haplotypes, while in outcrossing species (allogamous) such as *Eucalyptus* and most species of forest interest, LD usually decays at short genomic distances [43,44], which allows the identification of smaller haplotype blocks and those conformed by a smaller number of alleles.

The objectives of this study were to (i) determine and characterize haplotype blocks in *Eucalyptus globulus* from a high-quality SNP array (EUChip60K), (ii) investigate the possible benefits of using haplotype information to predict complex traits in trees, and (iii) estimate genomic parameters (e.g., genomic heritability and genetic gains) using Bayesian whole-regression models (Bayesian Ridge Regression, Bayes A, Bayes B, Bayes C and Bayesian Least Absolute Shrinkage and Selection Operator) that include the haplotype/SNP effect on a 6-year-old breeding population of *E. globulus*.

2. Results

2.1. Haplotype Block Construction

The final dataset used for the haplotype block construction and genomic predictions consisted of 14,422 SNPs. An average of 1356 SNPs per chromosome and an average frequency of one SNP every 4000 bp were found. Chromosome 8 contained 12.5% (1811 SNPs) of the total SNPs, while chromosome 4 presented only 6.2% (893 SNPs). Total values of 1137 haplotype blocks and 3279 haplotypes (Table 1) were identified in all chromosomes of the species under study. In total, 14.5% of the total SNPs (2092 SNPs) were grouped into haplotype blocks. The largest number of haplotype blocks was determined by combinations of SNPs located on chromosome 8 ($n = 152$), while a smaller amount was constructed by SNPs located on chromosome 4 ($n = 71$). Of the total blocks formed by the chromosomes, 300 haplotypes were obtained (on average) per chromosome, while each block had three haplotype variants on average. The number of SNPs within each haplotype block varied from two to 12. The smallest blocks had an extension of 36 bp, while the largest haplotypes had a length of 482 kbp (chromosome 8). About 24% of the total haplotype blocks had a size greater than 10 kbp, while 2% had a size above 100 kbp. Particularly, on chromosomes 4 and 9, no haplotype blocks with an extension higher than 100 kbp were detected. The genome-wide average linkage disequilibrium decay estimated across all chromosomes

is shown in Figure S1 (Supplementary Material). On a genome-wide average, LD decayed within ~10–12 Kbp to a level below $r^2 = 0.14$ (critical r^2 value).

Table 1. Summary of information on haplotypes and haplotype blocks determined in a breeding population of *E. globulus*. Ch corresponds to the chromosome number and single nucleotide polymorphic markers (SNPs) to the number of SNPs detected; HAP-Blocks is the number of haplotype blocks constructed; HAPs is the number of haplotypes; Max (kbp) corresponds to the maximum size (in kbp) for the haplotype blocks; Min (bp) corresponds to the minimum size (in bp) for the haplotype blocks; and Min (SNPs) and Max (SNPs) correspond to the maximum and minimum number of SNPs forming the haplotype blocks, respectively.

Ch	SNPs	HAP-Blocks	HAPs	Max (kb)	Min (bp)	Max (SNPs)	Min (SNPs)
1	924	75	219	381	61	6	2
2	1766	121	370	357	36	6	2
3	1587	99	299	123	30	11	2
4	893	71	207	31	31	5	2
5	1500	83	238	279	49	8	2
6	1474	144	407	343	63	6	2
7	1220	87	248	356	121	5	2
8	1811	152	418	482	70	6	2
9	946	89	249	94	34	5	2
10	1065	103	295	318	49	10	2
11	1236	113	329	250	75	12	2
Total	14,422	1137	3279	-	-	-	-
Mean	1311	103	298	274	54	7	2

2.2. Estimates of Genetic Parameters Based on Pedigree

Estimates of variance components, heritability and genetic gains based on pedigree information are shown in Table 2. The heritability estimates varied from 0.04 to 0.46. Wood density (WD) was the most heritable trait ($h_a^2 = 0.46$), while diameter at breast height (DBH), stem straightness (ST), and branch quality (BQ) had the lowest heritability estimates in this breeding population ($h_a^2 = 0.04, 0.06$, and 0.05, respectively). Additionally, the estimated genetic gains based on pedigree information were 7.7%, 2.9%, 4.4%, 3.9% and 9.7% for tree height (HT), DBH, ST, BQ and WD, respectively, considering an intensity of selection of ~10% (10.06%; $n = 65$).

2.3. Prediction Based on Genomic Data

The genomic prediction methods were compared using the average values of marginal posterior distributions of each estimated parameter. Genomic heritability, genetic gains and predictive ability (PA) values, obtained for each trait and prediction model, are shown in Tables 2 and 3 and Figure 1. PA values were dependent on the genomic prediction models and marker type (SNP markers, haplotypes (HAP) and haplotypes in conjunction with SNPs that were not assigned to a haplotype (HAP-SNP)). PA values for HT varied between 0.19 and 0.44. For Bayes B (BB) and Bayesian Ridge Regression (BRR) methods, the PA values based on SNP, HAP and HAP-SNP were not statistically different from each other, but Bayes A (BA), Bayesian Least Absolute Shrinkage and Selection Operator (Bayesian LASSO or BL) and Bayes C (BC) methods based on the three markers (SNP, HAP and HAP-SNP) showed significant differences in terms of PA values. Based on the comparison among the assessed models, BC had the highest predictive ability (PA = 0.44 (SNP) and 0.38 (HAP-SNP)) for HT in most cases, and this method consistently gave one of the highest values for genomic heritability ($\hat{h}_g^2 = 0.36$) and genetic gain (GG = 8%). In addition, the genomic heritability based on this method (either SNP or HAP-SNP) was statistically higher than the pedigree-based heritability ($\hat{h}_a^2 = 0.15$ [0.01–0.28]). On the other hand, the genetic gains for HT based on pedigree information and BC methods were not statistically different.

For DBH, PA values ranged between 0.17 and 0.46. The PA varied significantly among models based on SNP, HAP and HAP-SNP. In general, the models based on SNP had the highest PA

values. Among the models based on SNP, BRR had the highest predictive ability for DBH (PA = 0.45), while HAP-SNP-based BC had the highest predictive ability (PA = 0.46). The genomic heritability estimates based on BC and BRR ranged between 0.26–0.32 and 0.12–0.16, respectively, and were significantly higher than the pedigree-based heritability ($\hat{h}_a^2 = 0.04$ [<0.01–0.10]). Consistently, the genetic gains based on the selection with these genomic models were statistically higher than those based on pedigree information (GG = 2.9% [1.4–4.5]).

The predictive ability of ST varied from 0.20 to 0.58. On the other hand, the PA values of ST were not statistically different among markers (SNP, HAP or HAP-SNP) in most cases. BB and BC had the highest PA among the models based on SNP (PA = 0.52 and 0.54, respectively) and HAP-SNP (PA = 0.52 for BB and BC), while BRR had the highest predictive ability (PA = 0.58) among the HAP-based models. Consistently, all genomic heritability estimates based on BB, BC and BRR ($\hat{h}_g^2 = 0.15$–0.34) were higher than the pedigree-based heritability ($\hat{h}_a^2 = 0.06$ [<0.01–0.14]). In addition, only BC and BRR had genetic gain estimates statistically higher than those based on pedigree information for ST.

The predictive ability of BQ ranged between 0.06 and 0.33, which did not vary significantly between models based on SNP, HAP or HAP-SNP in most cases. Among the models based on SNP, BC had the highest predictive ability of BQ (PA = 0.31). For the HAP-based models, the highest PA was obtained by BC (PA = 0.28). In accordance with this, BC and BRR had the highest PA of BQ among the HAP-SNP-based models (PA = 0.31 and 0.33, respectively). All genomic heritability estimates based on BC and BRR ($\hat{h}_g^2 = 0.12$–0.29) were higher than the pedigree-based heritability estimate ($\hat{h}_a^2 = 0.05$ [<0.01–0.11]). However, only BRR had genetic gain estimates statistically higher than those based on pedigree information.

The predictive ability of WD ranged between 0.24 and 0.46, which varied significantly between models based on SNP, HAP and HAP-SNP in most cases. Among SNP-based models, BC had the highest predictive ability of WD (PA = 0.46). For HAP-based models, the highest predictive ability was obtained by BC and BRR (PA = 0.41 and 0.39, respectively). In addition, for HAP-SNP-based models, BRR had the highest predictive ability (PA = 0.44). The heritability of WD based on any genomic model ($\hat{h}_g^2 = 0.04$–0.26) was statistically lower than the pedigree-based heritability ($\hat{h}_a^2 = 0.46$ [0.22–0.69]). Consistently, the genetic gains for WD based on any genomic prediction model did not exceed 5% and were statistically lower than those based on pedigree information.

The regressions between pedigree-based Estimated Breeding Values (EBVs) and genomic-based EBVs (GEBVs considering models with the highest predictive ability for each trait) are shown in Figure 2. All coefficients of determination (R^2) between EBVs and GEBVs for all traits were above 0.98.

130

Table 2. Estimates of pedigree-based heritability ($\hat{h}_a^2$), genomic heritability ($\hat{h}_g^2$) and genetic gain (GG; percentage) for each method of prediction based on pedigree (PBP), SNP markers (SNP), haplotype (HAP), and haplotypes and SNPs that were not assigned to a haplotype (HAP-SNP). BA, BB, BC, BL, and BRR correspond to Bayes A, Bayes B, Bayes C, Bayesian Least Absolute Shrinkage, and Selection Operator and Bayesian Ridge Regression, respectively.

Trait/Model	Pedigree		SNP		HAP		HAP-SNP	
	$\hat{h}_a^2$ [CR]	GG [CR]	$\hat{h}_g^2$	GG	$\hat{h}_g^2$	GG	$\hat{h}_g^2$	GG
Tree height								
PBP	0.15 [0.01–0.28]	7.7 [5.4–10.3]	-	-	-	-	-	-
BA	-	-	0.11	5.6	0.06	4.2 *	0.10	5.2 *
BB	-	-	0.27	6.0	0.11	3.6 *	0.29 *	6.3
BC	-	-	0.36 *	7.9	0.28	6.6	0.36 *	7.8
BL	-	-	0.07	4.2 *	0.04	3.1 *	0.06	3.6 *
BRR	-	-	0.19	8.7	0.14	7.6	0.20	8.6
Diameter at breast height								
PBP	0.04 [<0.01–0.10]	2.9 [1.4–4.5]	-	-	-	-	-	-
BA	-	-	0.08	5.0 *	0.04	3.8	0.07	4.2
BB	-	-	0.19 *	4.9 *	0.09	3.4	0.14 *	3.3
BC	-	-	0.31 *	7.2 *	0.26 *	6.6 *	0.32 *	7.2 *
BL	-	-	0.05	3.6	0.05	4.1	0.05	3.4
BRR	-	-	0.16 *	8.3 *	0.12 *	7.8 *	0.16 *	8.2 *
Stem straightness								
PBP	0.06 [<0.01–0.14]	4.4 [1.9–7.1]	-	-	-	-	-	-
BA	-	-	0.10	4.4	0.09	4.7	0.09	4.1
BB	-	-	0.26 *	5.7	0.18 *	4.7	0.28 *	5.5
BC	-	-	0.34 *	7.1	0.30 *	7.5 *	0.34 *	7.2 *
BL	-	-	0.05	2.7	0.04	2.8	0.07	3.2
BRR	-	-	0.18 *	7.6 *	0.15 *	7.9 *	0.20 *	8.0 *
Branch quality								
PBP	0.05 [<0.01–0.11]	3.9 [1.4–6.0]	-	-	-	-	-	-
BA	-	-	0.04	2.2	0.04	2.7	0.05	2.3
BB	-	-	0.12 *	2.4	0.10	2.7	0.08	1.8
BC	-	-	0.29 *	5.0	0.25 *	5.4	0.29 *	4.9
BL	-	-	0.03	2.0	0.03	2.3	0.04	2.0
BRR	-	-	0.15 *	6.2 *	0.12 *	6.4 *	0.15 *	6.1 *
Wood density								
PBP	0.46 [0.22–0.69]	9.7 [7.5–12]	-	-	-	-	-	-
BA	-	-	0.07 *	2.0 *	0.05 *	2.0 *	0.08 *	2.1 *
BB	-	-	0.17 *	2.2 *	0.12 *	1.9 *	0.16 *	2.1 *
BC	-	-	0.34	3.2 *	0.26	3.0 *	0.33	3.1 *
BL	-	-	0.06 *	1.7 *	0.04 *	1.7 *	0.06 *	1.8 *
BRR	-	-	0.16 *	3.6 *	0.12 *	3.4 *	0.17 *	3.5 *

Numbers with asterisks are statically different from pedigree-based estimates (90% Bayesian credible sets). CR: 90% credible region from marginal posterior distributions.
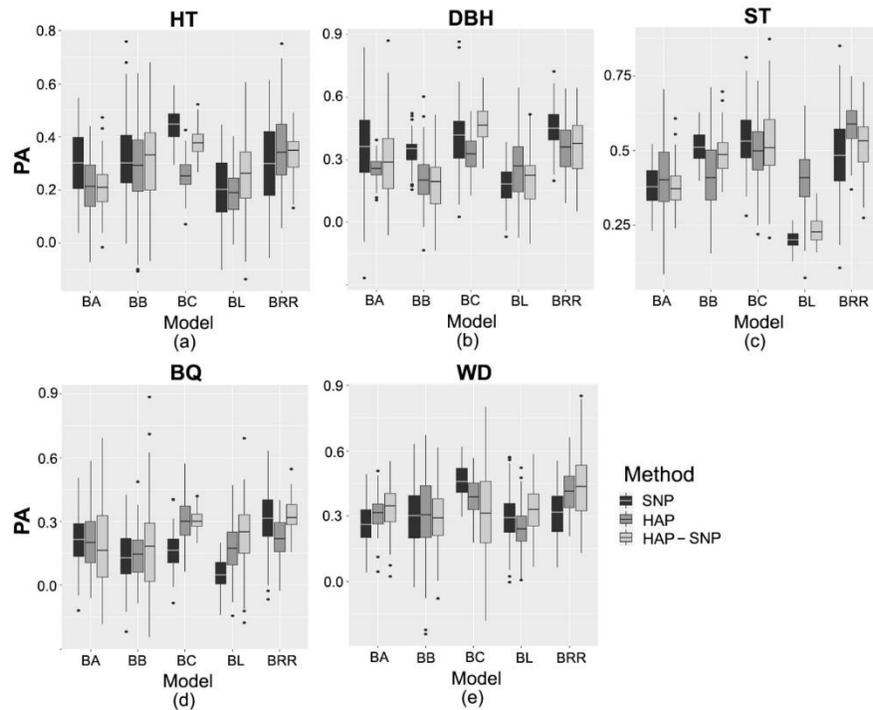
Figure 1. Predictive ability (PA) of (**a**) tree height (HT), (**b**) diameter at breast height (DBH), (**c**) stem straightness (ST), (**d**) branch quality (BQ), and (**e**) wood density (WD). Models based on SNP markers (SNP), haplotypes (HAP), and haplotypes with SNPs that were not assigned to a haplotype (HAP-SNP) are represented by black, dark gray and light gray bars, respectively. BA, BB, BC, BL, and BRR correspond to Bayes A, Bayes B, Bayes C, Bayesian Least Absolute Shrinkage and Selection Operator, and Bayesian Ridge Regression, respectively. Each box-plot represents the distribution of PA values for 100 cycles of cross-validation.

Table 3. Estimates of predictive ability (average of 100 cross-validation cycles) of Bayesian models based on SNPs (SNP), haplotypes (HAP) and haplotypes with SNPs that were not assigned to a haplotype (HAP-SNP) for each studied trait.

Trait/Markers	Genomic Model				
	BA	BB	BC	BL	BRR
Tree height					
SNP	0.31 bA	0.32 bA	**0.44** aA	0.21 cB	0.30 bA
HAP	0.21 cdB	0.28 bA	0.25 bcC	0.19 dB	0.35 aA
HAP-SNP	0.21 dB	0.31 bA	0.38 aB	0.26 cA	0.33 bA
Diameter at breast height					
SNP	0.35 bA	0.34 bA	0.39 bB	0.17 cB	0.45 aA
HAP	0.26 bcB	0.21 cB	0.33 aC	0.26 bA	0.36 aB
HAP-SNP	0.28 cAB	0.19 dB	**0.46** aA	0.20 dB	0.37 bB
Stem straightness					
SNP	0.38 cA	0.52 abA	0.54 aA	0.20 dC	0.48 bB
HAP	0.40 cA	0.42 cB	0.50 bA	0.40 cA	**0.58** aA
HAP-SNP	0.38 bA	0.49 aA	0.52 aA	0.23 cB	0.52 aB
Branch quality					
SNP	0.22 bA	0.13 cA	0.16 cB	0.06 dC	0.31 aA
HAP	0.20 bA	0.14 cA	0.28 aA	0.17 bcB	0.22 bB
HAP-SNP	0.19 bA	0.18 bA	0.31 aA	0.24 bA	0.33 aA
Wood density					
SNP	0.26 bB	0.30 bA	**0.46** aA	0.29 bA	0.32 bB
HAP	0.32 bA	0.31 bA	0.39 aB	0.24 cB	0.41 aA
HAP-SNP	0.34 bA	0.29 bA	0.32 bC	0.33 bA	0.44 aA

BA: Bayes A; BB: Bayes B; BC: Bayes C; BL: Bayesian Least Absolute Shrinkage and Selection Operator; BRR: Bayesian Ridge Regression. Statistical significance between different genomic models (BA, BB, BC, BL and BRR) is noted by lowercase letters, while that between different markers (SNP, HAP and HAP-SNP) is shown by upper case letters. Different letters show the statistical significance at $p < 0.01$ using the Tukey–Kramer test. Numbers in bold show the highest PA estimates considering both approaches: genomic models and marker types (SNP, HAP or HAP-SNP).
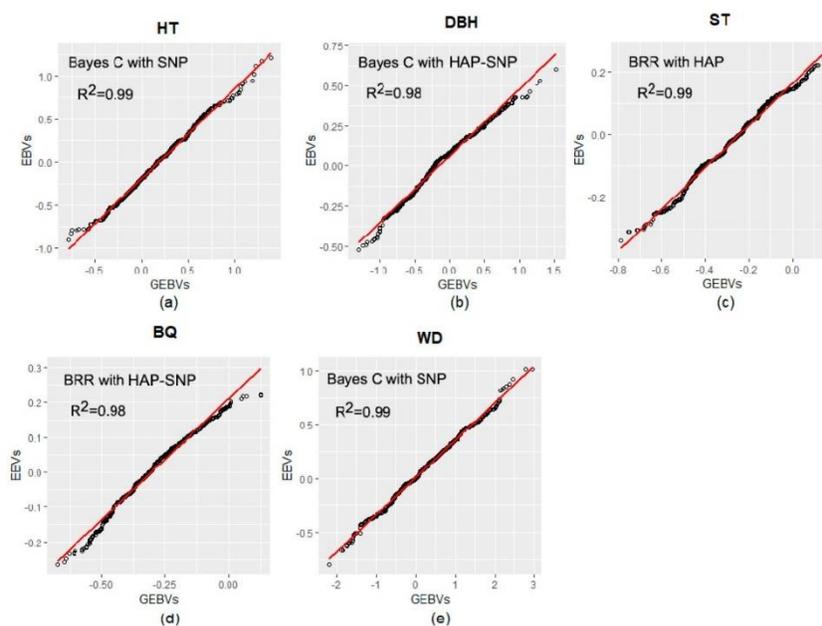
Figure 2. Linear regression plots relating estimated breeding values (pedigree-based Estimated Breeding Values; EBVs) and genomic-based EBVs (GEBVs). (**a**) EBVs and GEBVs for tree height (HT); (**b**) EBVs and GEBVs for diameter at breast height (DBH); (**c**) EBVs and GEBVs for stem straightness (ST); (**d**) EBVs and GEBVs for branch quality (BQ); and (**e**) EBVs and GEBVs for wood density (WD).

3. Discussion

3.1. Haplotype Blocks Construction

Previously, Durán et al. [45], Thavamanikumar et al. [46] and Cappa et al. [47] determined that the linkage disequilibrium (LD) in natural and controlled populations of *E. globulus* decreases rapidly in the range of 1000–4000 bp. However, in this study, several genomic regions were detected in strong disequilibrium above 10,000 bp. About 24% and 2% of the haplotype blocks formed had an extension over 10,000 bp and 100,000 bp, respectively, with a disequilibrium coefficient (D′) value higher than 0.7. The construction of the haplotype blocks was based on the confidence interval algorithm of Gabriel et al. [48], which establishes that those pairs of SNPs that are in a strong linkage disequilibrium have a D′ value between 0.7 and 0.98, considering a confidence interval of 95%. These D′ values could reveal that, historically, the probability of recombination between both loci is quite low. In this study, more than 1300 haplotype blocks that would meet this condition were detected. Therefore, it is possible to assume that the SNPs grouped in blocks are in strong LD and they are possibly inherited together across generations.

In breeding populations of self-pollinated plants and controlled crosses, it is expected that LD decays over long distances. For example, in soybean and rice, LD can be significant at a distance of 100,000 bp and 250,000 bp between loci, respectively [49], while in the genomes of outcrossing plants, LD is expected to decay at short distances due to the reproductive mechanisms that underlie these species, and their loci tend to be heterozygous [50,51]. In the context of this study, the genotyped individuals were sampled from a breeding population formed by full- and half-sib families, which could increase the probability of the occurrence of regions that form haplotype blocks.

3.2. Performance of Pedigree and Genomic Prediction Models

According to the pedigree and genomic prediction models, the traits studied in this population were found to be weakly to highly heritable. The results indicated that DBH, ST and BQ are under relatively low genetic control. However, previous studies have reported that the heritability of DBH in *Eucalyptus* spp. can be greater than 0.1 in populations from 3 to 15 years of age [52–55]. For DBH, the genomic prediction models based on BC and BRR had higher heritability values than those based on pedigree, with a genetic gain up to ~8%. Previously, Tan et al. [56] determined that the genomic heritability of DBH, based on Ridge Regression Best Linear Unbiased Prediction (BLUP) and Reproducing Kernel Hilbert Spaces, was higher than the pedigree-based heritability estimate in *Eucalyptus* hybrids at 6 years of age. In this study, SNP-based BRR and HAP-SNP-based BC had the highest predictive ability for DBH, but the genomic heritability estimated by BC (HAP-SNP) was two times higher than the BRR model based on SNP markers. In BC, the prior assumptions of all SNP effects have a common variance, and the method assigns a nonnull prior probability for the marker effect to be equal to zero [5,57]. Due to this assumption, Bayes C has been used to identify QTLs with large effects [58]. Similarly, BRR assumes a common variance for all SNP markers, but all SNP effects are shrunk to a similar extent [7]. In contrast, Suontama et al. [59] and Resende et al. [6] reported that the genomic heritability of DBH was not superior to the heritability estimated by BLUP in *Eucalyptus nitens* and hybrids of *Eucalyptus urophylla* × *E. grandis*, respectively. In another study, Müller et al. [60] reported that the genomic heritability of DBH was lower than the estimated heritability by BLUP in *E. benthamii*. Interestingly, Müller et al. [60] performed a prediction for DBH using 13,787 and 10,460 SNPs that were not in LD, which could confirm our hypothesis that the inclusion of SNPs in strong LD (such as a BC model based on HAP-SNP) could be beneficial for genetic parameter estimation.

The heritability of ST has been previously reported as low [54,61], which is consistent with the pedigree-based model and some genomic prediction models assessed in the present study. However, other studies have found that the heritability of ST in *E. globulus* can be moderate [62] and even high ($\hat{h}_a^2 > 0.3$, [53]). Almost all genomic-based heritability estimates of ST were statistically higher than those estimated by the pedigree-based method. Suontama et al. [59] found that the pedigree-based heritability of stem straightness in *E. nitens* increased approximately twice when using marker-based models. In the context of our study, the highest predictive ability for ST was obtained by BC based on SNP, BRR based on HAP, and BRR and BC models based on HAP-SNP. BC and BRR models had higher heritability than the estimate based on pedigree. However, BC models (based on either SNP or HAP) had a higher genomic heritability than other prediction models. There have been conflicting reports on the genetic architecture of stem straightness in trees. For example, Bartholomé et al. [63] detected QTLs that explained up to 5% of the total phenotypic variation of ST in maritime pine. Yang et al. [64] found QTLs that explained up to 15% of the total phenotypic variation of the stem straightness of Pinus hybrids. Additionally, Arriagada et al. [65] reported five QTLs for ST, explaining a total of 6–14% of the total proportion of trait variation in *Eucalyptus cladocalyx*.

BQ has been a scarcely evaluated trait in breeding programs of *E. globulus*. For instance, Callister et al. [53] reported a range in branching quality (measured as branch thickness) of 0–0.16 in full-sib families of E. globulus (at age 3.5). Ballesta et al. [43] reported that the pedigree-based heritability estimate of BQ in *E. globulus* at 4 years of age was less than 0.1. In other tree species, traits related to branch quality have been described to be moderately heritable [66,67]. In the present study, among the models based on SNP and HAP markers, BC had the highest predictive ability of branch quality. On the other hand, BC and BRR had the highest PA values among the models based on HAP-SNP. Interestingly, all genomic heritability estimates based on BC and BRR were higher than the pedigree-based heritability estimates. However, this result should be interpreted with caution due to Bayes C performing variable selection and shrinkage procedures, which could mean that the polygenic background may not be taken into account, favoring a selection based on major effect genes [68]. This could explain why the genetic gains based on BC were lower than those based on BRR.

Several studies have shown that WD is a highly heritable trait in *Eucalyptus* spp. [59,62,69], which is consistent with the estimates from the pedigree-based model and BC based on SNP, HAP and HAP-SNP. None of the methods based on genomic data exhibited higher heritability (and genetic gain) for this trait than the pedigree-based method. In other studies, Suontama et al. [59] reported that genomic-based and pedigree-based heritability values were similar for the basic wood density in *E. nitens*. Consistent with our findings, Resende et al. [6] reported that the genomic heritability of WD is subtly lower than the pedigree-based heritability in hybrids of *E. urophylla* and *E. grandis*. In other tree species, Beaulieu et al. [70] reported that the heritability of the wood density in trees of *Picea glauca* is lower when estimated by genomics than pedigree. As expected, the highest value of genetic gain was obtained for WD, while the lowest value was obtained for DBH. In this sense, the selection based on genomics for traits such as DBH, ST and BQ is highly justified and particularly attractive because of the potential of enhancing selection accuracy for low heritability traits and increasing the genetic gains for these traits [6,71]. On the other hand, Beaulieu et al. [70] stated that one of the reasons why genomic selection might not be as effective for predicting wood density is that pedigree information makes it possible to capture loci that have not been considered in a genomic prediction model, which can be important for the genetic control of the trait. According to our results, the BC model had a higher predictive ability and genomic heritability in most cases, supporting the idea of an oligogenic architecture for WD. Contrarily, Durán et al. [45] reported that BC model had a lower PA than the BL model.

The genomic heritability for HT, estimated by a BC model (SNP and HAP-SNP), was higher than the pedigree-based estimate. However, the HAP-SNP-based BC model had a predictive ability lower than the model based on SNP markers. In accordance with this, Tan et al. [56] determined that the genomic heritability of HT, based on Ridge Regression BLUP and Reproducing Kernel Hilbert Spaces, was superior to the pedigree-based heritability in *Eucalyptus* hybrids. However, the genetic gains estimated using the pedigree and BC methods were not statistically different. Previously, Lenz et al. [72] and Beaulieu et al. [70] reported that the genetic gains based on genomic prediction could be lower than the gains based on the pedigree method for HT in other tree species. As mentioned, the BC model (based on SNP and HAP-SNP) had the highest predictive ability for HT in the most cases. Contrarily, Müller et al. [60] found no differences in the predictive ability of growth traits (HT and DBH) between the BA, BB, BC, BL, and BRR models.

Several studies in animals have emphasized the use of haplotypes (from high-density DNA arrays) to estimate the predictive ability of different GS models [27,35–38]. In plants, Matias et al. [29] demonstrated that the use of haplotypes in prediction studies in maize increased its predictive ability by 20%. It should be noted that in outcrossing plants, such as *Eucalyptus* and other forest species, LD usually decays at short genomic distances, which allows the identification of smaller haplotype blocks and with fewer variants (haplotypes). In this study, we found haplotype blocks formed by SNPs in a strong LD (D′ > 0.7) with a size above 300 kbp. According to Cuyabano et al. [27], haplotypes that are constructed by SNPs in a LD of D′ > 0.45 can significantly increase the predictive ability of a genomic selection model. Our results revealed that predictive ability values were mainly dependent on the Bayesian methods assessed (i.e., BA, BB, BC, BL and BRR) more than on the marker type (SNP, HAP or HAP-SNP). However, genomic models that included the haplotype effect (either HAP or HAP-SNP) significantly increased the PA of low-heritability traits. These results, nevertheless, should be interpreted with caution due to the age of the trees, and we therefore emphasize that further studies are needed to evaluate the performance of the genomic models. On the other hand, the development of techniques to select trees at early growth stages may greatly increase the genetic gain per unit time, and thus, substantially accelerate tree-breeding programs.

4. Materials and Methods

4.1. Trial Conditions and Phenotyping

In 2018, 6-year-old trees of *Eucalyptus globulus* from a progeny trial consisting of a mix of half-sib and full-sib families located in the La Poza sector, Purranque, Chile [43], were evaluated according to the following traits (description in Table S1 and Figure S2): wood density (WD), stem straightness (ST), branch quality (BQ), diameter at breast height (DBH), and tree height (HT). This location (40°57′S, 73°30′W; 326 m.a.s.l.) has an Oceanic or Marine climate type with an annual accumulated rainfall of 1282 mm and an average annual temperature of 13 °C. The WD was measured indirectly according to Valenzuela et al. [69]. ST was evaluated in the first 2/3 of the total height of the tree according to an ordinal scale (seven levels). The value 0 corresponds to trees that have a curvature in the first third of the total height of the tree and 6 in the case of trees that could present a slight curvature in the upper third of the total height of the tree without affecting productivity. BQ was evaluated according to different criteria that define quality (diameter, angle and distribution of branches in the tree) by means of an ordinal scale of six levels, in which a value of 0 is assigned to trees with an extreme deficiency in the diameter of branches and any other variable, and a value of 6 corresponds to trees that have an optimal combination of all quality variables without generating loss of productivity. The trees were distributed in a randomized complete block design with 30 blocks, considering single-tree plots (each family is represented by only one tree in each block) and a planting density of 2.5 m between each tree within each block.

4.2. Genotyping, Linkage Disequilibrium and Haplotype Blocks

Genomic DNA was isolated from the leaves of 646 randomly selected individuals of *E. globulus* (~10 individuals per family). The DNA extraction protocol followed the work of Ballesta et al. [43]. Individuals were genotyped using the EUChip60K SNP system (GeneSeek, Lincoln, NE, USA) [17]. The genotyping quality of the samples was evaluated in Genome Studio software (Illumina, San Diego, CA). Monomorphic SNP markers and those with a call rate <90% were removed. Subsequently, those SNPs with a minor allele frequency (MAF) <0.05 were eliminated. A total of 14,442 remaining SNPs was retained for the 646 individuals.

The haplotype blocks were defined according to the confidence interval algorithm developed by Gabriel et al. [48] using the software Haploview v. 4.2 [73]. The pairs of SNPs were considered to be in strong linkage disequilibrium (LD) if the upper limit of the 95% confidence interval of the value of normalized disequilibrium coefficient (D′) was higher than 0.98 and if the lower limit had a minimum value of 0.7. The D′ between A and B loci, was calculated as follows:

$$D'_{AB} = D/D_{MAX} \tag{1}$$

where D is calculated as $D = p_{A_1B_1}p_{A_2B_2} - p_{A_1B_2}p_{A_2B_1}$, and D_{MAX}:

$$D_{MAX} = \begin{cases} -\min\left\{p_{A_1}p_{B_1}, p_{A_2}p_{B_2}\right\}, \text{ when } D < 0 \\ \min\left\{p_{A_1}p_{B_2}, p_{A_2}p_{B_1}\right\}, \text{ when } D \geq 0 \end{cases} \tag{2}$$

The physical positions of each SNP were established according to the consensus map of the genome of *E. grandis* [74]. The extent of LD was also estimated as the squared allele frequency correlation (r^2). The critical r^2 value was calculated according to the method used by Breseghello and Sorells [75].

4.3. Prediction Models Based on Pedigree and Genomic Data

In this study, prediction models based on pedigree and genomic data from an array of SNP markers were used. In the pedigree-based model, individual breeding values were predicted using a Bayesian generalized linear model implemented in the MCMCglmm (Markov Chain Monte Carlo—Generalized

Linear Mixed Model) library [76] of R 3.6.1 [77] This Bayesian analysis was carried out using the following base model:

$$y = X\beta + Za + \varepsilon \tag{3}$$

where y corresponds to the phenotypic data vector, β is the vector of block effects, a is the vector of the additive genetic effects $a \sim N(0, A\sigma_a^2)$, A corresponds to the matrix of Wright's coefficients (pedigree information), and σ_a^2 is the additive genetic variance. X and Z correspond to the known incidence matrices that relate the observation vector (y) to vectors β and a, respectively, and ε corresponds to the vector of residual effects, $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$, where I is an identity matrix, and σ_ε^2 is the residual variance. The Bayesian models were run with 1,000,000 iterations, a burn-in period of 100,000 and a thin of 50.

The prediction models based on SNPs/haplotypes were the following: Bayesian Least Absolute Shrinkage and Selection Operator (Bayesian LASSO or BL, [78,79]), Bayesian Ridge Regression (BRR, [7]), Bayes A (BA, [1]), Bayes B (BB, [1]) and Bayes Cπ (BC, [5]). All whole-regression models can be expressed in matrix form as follows:

$$y = X\beta + Zm + \varepsilon \tag{4}$$

where y corresponds to the phenotypic data vector, β is the vector of block effects, m corresponds to marker effects (SNPs and/or haplotypes), and depending on the model, different prior distributions are assigned; for instance, a double exponential distribution in the case of BL and a Gaussian distribution in the case of BRR, among others (see Refs. [7,57]). ε corresponds to the vector of residuals, $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$. X and Z correspond to the incidence matrices that relate the observation vector (y) to vectors β and m, respectively.

The matrix of SNP markers or haplotypes was coded by the numbers 0, 1 and 2. In the case of SNP markers, 0 represents the homozygous genotype of the allele with the lowest frequency for the *i*-th marker (*i* = 1, … , n), 1 represents the heterozygous genotype for the *i*-th marker, and 2 represents the homozygous genotype of the allele with the highest frequency for the *i*-th marker. In the case of haplotypes, since one haplotype block can have more than two allelic variants, the values of 0, 1 and 2 represent the number of copies for each variant (haplotype), in which a value of 0 was assigned for those individuals who did not present any copy of the *j*-th haplotype and *i*-th block, a value of 1 was assigned if they presented a copy of the *j*-th haplotype and *i*-th block, and a value of 2 was assigned if they presented two copies of the *j*-th haplotype and *i*-th block [27]. The models based on SNP markers, haplotypes and haplotypes in conjunction with SNPs (that were not assigned to a haplotype) are identified by SNP, HAP and HAP-SNP, respectively.

The BL method assumes that the marker effects are distributed a priori according to a double exponential (DE), $p(m_i|\lambda, \sigma_\varepsilon^2) = DE(m_i|0, \lambda, \sigma_\varepsilon^2)$, where λ corresponds to a regularization parameter. The distribution of DE generates a strong contraction (close to zero) to estimate the effects of the markers. BRR is a Bayesian method based on the fact that model regressors (SNPs and/or haplotypes) have a common variance (σ_m^2); those regressors with the same allelic frequency explain the same proportion of the additive variance and have the same contraction effect [7]. The marker effect (m_i) is distributed as follows: $m_i|\sigma_m^2 \sim N(0, \sigma_m^2)$; and the common variance $p(\sigma_m^2) \sim$ scaled inverse Chi-squared ($\sigma_m^2|v_m, S_m \sim \chi^{-2}(v_m, S_m)$), with degree of freedom and scale parameters v_m and S_m, respectively. In BA, the marginal distribution of marker effects is a scaled-t density, in which, for computational convenience, this density is implemented as an infinite mixture of scaled-normal densities (see Ref. [57]). The variance of each marker is assumed to be distributed scaled inverse Chi-squared. BB uses a mixed distribution with a mass at zero, such that the prior distribution of the effects of the all markers is given by [51]

$$m_i|\sigma_{m_i}^2, \pi = \begin{cases} 0 & \text{with probability } \pi \\ N(0, \sigma_{m_i}^2) & \text{with probability } 1 - \pi \end{cases} \tag{5}$$

A scaled inverse Chi-square prior distribution $\chi^{-2}(v_m, S_m)$ is assumed for $\sigma^2_{m_i}$ ($i = 1, \ldots, n$), which is equal for all markers. In BC, all markers are considered to have a common variance (σ^2_m) and promote the selection of variables such as Bayes B. The marker effects are assumed to be $m_i | \sigma^2_{m_i} \sim N(0, \sigma^2_m)$ with a probability of $1 - \pi = 0$.

All the genomic-based Bayesian methods were implemented in the library BGLR (Bayesian Generalized Linear Regression) [57] of R 3.6.1 [77]. Variance components were estimated with a total of 1,000,000 iterations, a burn-in period of 100,000, and a thin of 50. The predictive ability (PA) of each model was measured as the correlation between the Genomic Estimated Breeding Values (GEBVs) obtained by Equation (2) and genomic breeding values predicted by cross-validation, which considered 90% of the individuals as the training population and the remaining 10% as the validation population. The PA was reported as the average of correlation coefficients for 100 cycles of the cross-validation.

4.4. Heritability and Genetic Gain

The prediction methods (genomic and pedigree-based) were also compared by the values of heritability and genetic gains. In the case of the pedigree-based method [80], the heritability ($\hat{h}^2_a$) in a narrow sense was calculated as follows:

$$\hat{h}^2_a = \frac{\hat{\sigma}^2_a}{\hat{\sigma}^2_a + \hat{\sigma}^2_\varepsilon} \tag{6}$$

where $\hat{\sigma}^2_a$ and $\hat{\sigma}^2_\varepsilon$ correspond to the additive genetic and residual variances, respectively. In the case of Bayesian genomic prediction models (SNP/haplotypes), genomic heritability ($\hat{h}^2_g$), genomic variance ($\hat{\sigma}^2_g$) and the residual variance ($\hat{\sigma}^2_\varepsilon$) were estimated using the marginal posterior distributions of each estimated parameter [81–83]. The genomic variance was estimated for each model as follows:

For BRR and BC:

$$\hat{\sigma}^2_g = 2\hat{\sigma}^2_m \sum_{i=1}^{n} p_i(1 - p_i) \tag{7}$$

For BA and BB:

$$\hat{\sigma}^2_g = 2\sum_{i=1}^{n} p_i(1 - p_i)\hat{\sigma}^2_{m_i} \tag{8}$$

For BL:

$$\hat{\sigma}^2_g = 2\sum_{i=1}^{n} \tau^2_i \hat{\sigma}^2_\varepsilon p_i(1 - p_i) \tag{9}$$

where p_i is the MAF of ith marker, $\hat{\sigma}^2_m$ is the variance of markers, and τ^2_i was assumed to be the exponential of λ, $\tau^2_i \sim \text{Exp}(\lambda^2)$, in which λ was assumed to belong to Gamma distribution $\lambda^2 \sim G(\varphi_1, \varphi_2)$. BRR and BC models assume that all markers have the same variance ($\hat{\sigma}^2_m$), while BB and BA models assume a variance for each ith marker ($\hat{\sigma}^2_{m_i}$). BB, BC and BL models assume the selection of variables; however, BL uses regularization parameter λ that directs markers with irrelevant effects close to zero.

The genetic gain (GG) was estimated for each prediction method using the following expression [54,84]:

$$GG = \frac{(\bar{y}_{sel} - \bar{y}_{pop})}{\bar{y}_{phe}} * 100 \tag{10}$$

where $\bar{y}_{sel}$ and $\bar{y}_{pop}$ correspond to the estimated posterior mean of the breeding values of selected trees (with a selection intensity of ~ 10.06%) and the estimated posterior population mean of the breeding values, respectively, and $\bar{y}_{phe}$ is the phenotypic mean for each trait. For ordinal traits, the $\bar{y}_{phe}$ term was calculated according to Burdon et al. [85].

5. Conclusions

To our knowledge, this study is one of the first to examine the inclusion of haplotypes in genomic selection models of *Eucalyptus*. In general, genomic heritability estimates were higher than those based on pedigree information for most of the studied traits. On the other hand, the predictive ability values were dependent on the genomic prediction models and marker type. On average, the homocedastic methods (BRR and Bayes C) had the highest predictive ability for the majority of traits. Notably, genomic models that included the haplotype effect (either HAP or HAP-SNP) significantly increased the PA of traits with low heritability. The results of this study provide additional perspectives for the implementation of genomic selection in *Eucalyptus* breeding programs, which could be especially beneficial for improving low heritability traits.

Supplementary Materials: The following are available online at http://www.mdpi.com/2223-7747/8/9/331/s1, Figure S1: Genome-wide average linkage disequilibrium (LD) decay plot estimated across all chromosomes of *Eucalyptus* for the studied population. The LD values correspond to the average of the correlation between alleles at two loci (r^2) and the normalized disequilibrium coefficient (D′) for each 1 Kpb. The LD threshold of $r^2 = 0.14$ is indicated with a red line; Figure S2: Distributions and histograms of the studied traits: (a) tree height (HT) in m, (b) diameter at breast height (DBH) in cm, (c) stem straightness (ST), (d) branch quality (BQ) and (e) wood density (WD) in mm; Table S1: Summary of phenotypic information for quantitative traits related to wood quality and tree growth measured in a six-year-old breeding population of *E. globulus*.

Author Contributions: Conceptualization, P.B. and F.M.; methodology, P.B. and F.M.; software, C.M.; validation, P.B., F.M. and C.M.; formal analysis, C.M., F.M. and P.B.; investigation, P.B.; resources, F.M.; data curation, C.M., F.M.; writing original draft preparation, P.B.; writing review and editing, F.M., P.P.-R.; supervision, F.M.; project administration, F.M.; funding acquisition, F.M.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829.
2. Gianola, D. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* **2006**, *173*, 1761–1776. [CrossRef]
3. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
4. Pérez, P.; de los Campos, G.; Crossa, J.; Gianola, D. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* **2010**, *3*, 106–116. [CrossRef]
5. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef]
6. Resende, M.D.; Resende, M.F.; Sansaloni, C.P.; Petroli, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; et al. Genomic selection for growth and wood quality in *Eucalyptus*: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* **2012**, *194*, 116–128. [CrossRef]
7. Gianola, D. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* **2013**, *194*, 573–596. [CrossRef]
8. Azevedo, C.F.; Silva, F.F.; de Resende, M.D.V.; Lopes, M.S.; Duijvesteijn, N.; Guimarães, S.E.F.; Lopes, P.S.; Kelly, M.J.; Viana, J.M.S.; Knol, E.F. Supervised independent component analysis as an alternative method for genomic selection in pigs. *J. Anim. Breed. Genet.* **2014**, *131*, 452–461. [CrossRef]
9. Azevedo, C.F.; de Resende, M.D.V.; e Silva, F.F.; Viana, J.M.S.; Valente, M.S.F.; Resende, M.F.R.; Muñoz, P. Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genet.* **2015**, *16*, 105. [CrossRef]
10. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE* **2013**, *8*, e54985. [CrossRef]

11. Chen, H.; Xie, W.; He, H.; Yu, H.; Chen, W.; Li, J.; Yu, R.; Yao, Y.; Zhang, W.; He, Y.; et al. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant* **2014**, *7*, 541–553. [CrossRef]

12. Avni, R.; Nave, M.; Eilam, T.; Sela, H.; Alekperov, C.; Peleg, Z.; Dvorak, J.; Korol, A.; Distelfeld, A. Ultra-dense genetic map of durum wheat× wild emmer wheat developed using the 90K iSelect SNP genotyping assay. *Mol. Breed.* **2014**, *34*, 1549–1562. [CrossRef]

13. Bayer, M.M.; Rapazote-Flores, P.; Ganal, M.; Hedley, P.E.; Macaulay, M.; Plieske, J.; Ramsay, L.; Russell, J.; Shaw, P.D.; Thomas, W.; et al. Development and evaluation of a barley 50k iSelect SNP array. *Front. Plant Sci.* **2017**, *8*, 1792. [CrossRef]

14. Verde, I.; Bassil, N.; Scalabrin, S.; Gilmore, B.; Lawley, C.T.; Gasic, K.; Micheletti, D.; Rosyara, U.R.; Cattonaro, F.; Vendramin, E.; et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS ONE* **2012**, *7*, e35668. [CrossRef]

15. Bianco, L.; Cestaro, A.; Sargent, D.J.; Banchi, E.; Derdak, S.; Di Guardo, M.; Salvi, S.; Jansen, J.; Viola, R.; Gut, I.; et al. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus× domestica* Borkh). *PLoS ONE* **2014**, *9*, e110377. [CrossRef]

16. Unterseer, S.; Bauer, E.; Haberer, G.; Seidel, M.; Knaak, C.; Ouzunova, M.; Meitinger, T.; Strom, T.M.; Fries, R.; Pausch, H.; et al. A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomes* **2014**, *15*, 823.

17. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D. Flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* **2015**, *206*, 1527–1540. [CrossRef]

18. Mora, F.; Quitral, Y.A.; Matus, I.; Russell, J.; Waugh, R.; Del Pozo, A. SNP-based QTL mapping of 15 complex traits in barley under rain-fed and well-watered conditions by a mixed modeling approach. *Front. Plant Sci.* **2016**, *7*, 909. [CrossRef]

19. Contreras-Soto, R.I.; Mora, F.; de Oliveira, M.A.R.; Higashi, W.; Scapim, C.A.; Schuster, I. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS ONE* **2017**, *12*, e0171105. [CrossRef]

20. Rasheed, A.; Hao, Y.; Xia, X.; Khan, A.; Xu, Y.; Varshney, R.K.; He, Z. Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Mol. Plant* **2017**, *10*, 1047–1064. [CrossRef]

21. Battenfield, S.D.; Sheridan, J.L.; Silva, L.D.; Miclaus, K.J.; Dreisigacker, S.; Wolfinger, R.D.; Peña, R.J.; Singh, R.P.; Jackson, E.W.; Fritz, A.K.; et al. Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS ONE* **2018**, *13*, e0204757. [CrossRef]

22. Li, C.X.; Xu, W.G.; Guo, R.; Zhang, J.Z.; Qi, X.L.; Hu, L.; Zhao, M.Z. Molecular marker assisted breeding and genome composition analysis of Zhengmai 7698, an elite winter wheat cultivar. *Sci. Rep.* **2018**, *8*, 322. [CrossRef]

23. Maldonado, C.; Mora, F.; Scapim, C.A.; Coan, M. Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: hapLA4. *PLoS ONE* **2019**, *14*, e0212925. [CrossRef]

24. Nordborg, M.; Tavaré, S. Linkage disequilibrium: What history has to tell us. *Trends Genet.* **2002**, *18*, 83–90. [CrossRef]

25. Machiela, M.J.; Chanock, S.J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **2015**, *31*, 3555–3557. [CrossRef]

26. Andersen, J.; Lübberstedt, T. Functional markers in plants. *Trends Plant Sci.* **2003**, *8*, 554–560. [CrossRef]

27. Cuyabano, B.C.; Su, G.; Lund, M.S. Genomic prediction of genetic merit using LD based haplotypes in the Nordic Holstein population. *BMC Genom.* **2014**, *15*, 1171. [CrossRef]

28. Calus, M.P.; Meuwissen, T.H.; Windig, J.J.; Knol, E.F.; Schrooten, C.; Vereijken, A.L.; Veerkamp, R.F. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet. Sel. Evol.* **2009**, *41*, 11. [CrossRef]

29. Matias, F.I.; Galli, G.; Correia Granato, I.S.; Fritsche-Neto, R. Genomic prediction of autogamous and allogamous plants by SNPs and haplotypes. *Crop Sci.* **2017**, *57*, 2951–2958. [CrossRef]

30. Sun, C.; Wang, B.; Yan, L.; Hu, K.; Liu, S.; Zhou, Y.; Guan, C.; Zhang, Z.; Li, J.; Zhang, J.; et al. Genome-wide association study provides insight into the genetic control of plant height in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* **2016**, *7*, 1102. [CrossRef]

31. Nimmakayala, P.; Abburi, V.L.; Saminathan, T.; Alaparthi, S.B.; Almeida, A.; Davenport, B.; Nadimi, M.; Davidson, J.; Tonapi, K.; Yadav, L.; et al. Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum Annuum*, L. *Sci. Rep.* **2016**, *6*, 38081. [CrossRef]

32. Vinholes, P.; Rosado, R.; Roberts, P.; Borém, A.; Schuster, I. Single nucleotide polymorphism-based haplotypes associated with charcoal rot resistance in Brazilian soybean germplasm. *Agron. J.* **2018**, *111*, 182–192. [CrossRef]

33. Nyine, M.; Wang, S.; Kiani, K.; Jordan, K.; Liu, S.; Byrne, P.; Haley, S.; Baenziger, S.; Chao, S.; Bowden, R.; et al. Genotype imputation in winter wheat using first generation haplotype map SNPs improves genome-wide association mapping and genomic prediction of traits. *G3 Genes Genomes Genet.* **2019**, *9*, 125–133. [CrossRef]

34. Calus, M.P.L.; De Roos, A.P.W.; Veerkamp, R.F. Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **2008**, *178*, 553–561. [CrossRef]

35. De Roos, A.P.W.; Schrooten, C.; Druet, T. Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. *J. Dairy Sci.* **2011**, *94*, 4708–4714. [CrossRef]

36. Boichard, D.; Guillaume, F.; Baur, A.; Croiseau, P.; Rossignol, M.N.; Boscher, M.Y.; Druet, T.; Genestout, L.U.C.I.E.; Colleau, J.J.; Journaux, L.; et al. Genomic selection in French dairy cattle. *Anim. Prod. Sci.* **2012**, *52*, 115–120. [CrossRef]

37. Edriss, V.; Fernando, R.L.; Su, G.; Lund, M.S.; Guldbrandtsen, B. The effect of using genealogy-based haplotypes for genomic prediction. *Genet. Sel. Evol.* **2013**, *45*, 5. [CrossRef]

38. Jónás, D.; Ducrocq, V.; Croiseau, P. The combined use of linkage disequilibrium–based haploblocks and allele frequency–based haplotype selection methods enhances genomic evaluation accuracy in dairy cattle. *J. Dairy Sci.* **2017**, *100*, 2905–2908. [CrossRef]

39. Curtis, D.; North, B.V.; Sham, P.C. Use of an artificial neural network to detect association between a disease and multiple marker genotypes. *Ann. Hum. Genet.* **2001**, *65*, 95–107. [CrossRef]

40. Yu, J.; Pressoir, G.; Briggs, W.H.; Bi, I.V.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203. [CrossRef]

41. Jarquín, D.; Kocak, K.; Posadas, L.; Hyma, K.; Jedlicka, J.; Graef, G.; Lorenz, A. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomes* **2014**, *15*, 740. [CrossRef]

42. Habyarimana, E. Genomic prediction for yield improvement and safeguarding of genetic diversity in CIMMYT spring wheat (*Triticum aestivum* L.). *Aust. J. Crop. Sci.* **2016**, *10*, 127.

43. Ballesta, P.; Serra, N.; Guerra, F. Genomic prediction of growth and stem quality traits in *Eucalyptus globulus* Labill. at its southernmost distribution limit in Chile. *Forests* **2018**, *9*, 779. [CrossRef]

44. Thavamanikumar, S.; McManus, L.J.; Ades, P.K.; Bossinger, G.; Stackpole, D.J.; Kerr, R.; Hadjigol, S.; Freeman, J.S.; Vaillancourt, R.E.; Zhu, P.; et al. Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp. *globulus* Labill identifies nine stable marker-trait associations for seven traits. *Tree Genet. Genomes* **2014**, *10*, 1661–1678.

45. Durán, R.; Isik, F.; Zapata-Valenzuela, J.; Balocchi, C.; Valenzuela, S. Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genet. Genomes* **2017**, *13*, 74. [CrossRef]

46. Thavamanikumar, S.; McManus, L.J.; Tibbits, J.F.; Bossinger, G. The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Aust. For.* **2011**, *74*, 23–29. [CrossRef]

47. Cappa, E.P.; El-Kassaby, Y.A.; Garcia, M.N.; Acuña, C.; Borralho, N.M.; Grattapaglia, D.; Poltri, S.N.M. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. *PLoS ONE* **2013**, *8*, e81267. [CrossRef]

48. Gabriel, S.B.; Schaffner, S.F.; Nguyen, H.; Moore, J.M.; Roy, J.; Blumenstiel, B.; Higgins, J.; DeFelice, M.; Lochner, A.; Faggart, M.; et al. The structure of haplotype blocks in the human genome. *Science* **2002**, *296*, 2225–2229. [CrossRef]

49. Gupta, P.K.; Pawan, S.; Kulwal, P.L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol. Biol.* **2005**, *57*, 461–485. [CrossRef]

50. Fiil, A.; Lenk, I.; Petersen, K.; Jensen, C.S.; Nielsen, K.K.; Schejbel, B.; Andersen, J.R.; Lübberstedt, T. Nucleotide diversity and linkage disequilibrium of nine genes with putative effects on flowering time in perennial ryegrass (*Lolium perenne* L.). *Plant Sci.* **2011**, *180*, 228–237. [CrossRef]

51. Pérez-Rodríguez, P.; Gianola, D.; González-Camacho, J.M.; Crossa, J.; Manès, Y.; Dreisigacker, S. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes Genomes Genet.* **2012**, *2*, 1595–1605. [CrossRef]

52. Costa e Silva, J.C.; Hardner, C.; Potts, B.M. Genetic variation and parental performance under inbreeding for growth in *Eucalyptus globulus*. *Ann. For. Sci.* **2010**, *67*, 606. [CrossRef]

53. Callister, A.N.; England, N.; Collins, S. Genetic analysis of *Eucalyptus globulus* diameter, straightness, branch size, and forking in Western Australia. *Can. J. For. Res.* **2011**, *41*, 1333–1343. [CrossRef]

54. Mora, F.; Serra, N. Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. *Tree Genet Genomes* **2014**, *10*, 711–719. [CrossRef]

55. Resende, R.T.; Resende, M.D.V.; Silva, F.F.; Azevedo, C.F.; Takahashi, E.K.; Silva-Junior, O.B.; Grattapaglia, D. Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* **2017**, *119*, 245. [CrossRef]

56. Tan, B.; Grattapaglia, D.; Martins, G.S.; Ferreira, K.Z.; Sundberg, B.; Ingvarsson, P.K. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol.* **2017**, *17*, 110. [CrossRef]

57. Pérez, P.; De Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [CrossRef]

58. Van den Berg, I.; Fritz, S.; Boichard, D. QTL fine mapping with Bayes C (π): A simulation study. *Genet. Sel. Evol.* **2013**, *45*, 19. [CrossRef]

59. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dungey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2018**, *122*, 370–379. [CrossRef]

60. Müller, B.S.; Neves, L.G.; de Almeida Filho, J.E.; Resende, M.F.; Muñoz, P.R.; dos Santos, P.E.; Paludzyszyn Filho, E.; Kirst, M.; Grattapaglia, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomes* **2017**, *18*, 524. [CrossRef]

61. Lopez, G.A.; Potts, B.M.; Dutkowski, G.W. Genetic variation and inter-trait correlations in *Eucalyptus globulus* base population trials in Argentina. *For. Genet.* **2002**, *9*, 217–231.

62. Blackburn, D.P.; Hamilton, M.G.; Harwood, C.E.; Baker, T.G.; Potts, B.M. Assessing genetic variation to improve stem straightness in *Eucalyptus globulus*. *Ann. For. Sci.* **2013**, *70*, 461–470. [CrossRef]

63. Bartholomé, J.; Bink, M.C.; van Heerwaarden, J.; Chancerel, E.; Boury, C.; Lesur, I.; Isik, F.; Bouffier, L.; Plomion, C. Linkage and association mapping for two major traits used in the maritime pine breeding program: Height growth and stem straightness. *PLoS ONE* **2016**, *11*, e0165323.

64. Yang, H.; Liu, T.; Xu, B. QTL detection for growth and form traits in three full-sib pedigrees of *Pinus elliottii* var. *elliottii* × *P. caribaea* var. *hondurensis* hybrids. *Tree Genet. Genomes* **2015**, *11*, 130. [CrossRef]

65. Arriagada, O.; Mora, F.; Amaral Junior, A.T. Thirteen years under arid conditions: Exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breed Sci.* **2018**, *68*, 367–374. [CrossRef]

66. Song, J.; Brendel, O.; Bodénès, C.; Plomion, C.; Kremer, A.; Colin, F. X-ray computed tomography to decipher the genetic architecture of tree branching traits: Oak as a case study. *Tree Genet. Genomes* **2017**, *13*, 5. [CrossRef]

67. Monclus, R.; Leplé, J.C.; Bastien, C.; Bert, P.F.; Villar, M.; Marron, N.; Brignolas, F.; Jorge, V. Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp. *BMC Plant Biol.* **2012**, *12*, 173. [CrossRef]

68. Wolfe, M.D.; Del Carpio, D.P.; Alabi, O.; Ezenwaka, L.C.; Ikeogu, U.N.; Kayondo, I.S.; Lozano, R.; Okeke, U.G.; Ozimati, A.A.; Williams, E.; et al. Prospects for genomic selection in cassava breeding. *Plant Genome* **2017**, *10*, 1–9. [CrossRef]

69. Valenzuela, C.E.; Ballesta, P.; Maldonado, C.; Baettig, R.; Arriagada, O.; Sousa Mafra, G.; Mora, F. Bayesian mapping reveals large-effect pleiotropic QTLs for wood density and slenderness index in 17-year-old trees of *Eucalyptus cladocalyx*. *Forests* **2019**, *10*, 241. [CrossRef]

70. Beaulieu, J.; Doerksen, T.; Clément, S.; MacKay, J.; Bousquet, J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* **2014**, *113*, 343. [CrossRef]

71. Makowsky, R.; Pajewski, N.M.; Klimentidis, Y.C.; Vazquez, A.I.; Duarte, C.W.; Allison, D.B.; de Los Campos, G. Beyond missing heritability: Prediction of complex traits. *PLoS Genet.* **2011**, *7*, e1002051. [CrossRef]
72. Lenz, P.R.; Beaulieu, J.; Mansfield, S.D.; Clément, S.; Desponts, M.; Bousquet, J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genome* **2017**, *18*, 335. [CrossRef]
73. Barrett, J.C.; Fry, B.; Maller, J.D.M.J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [CrossRef]
74. Myburg, A.A.; Grattapaglia, D.; Tuskan, G.A.; Hellsten, U.; Hayes, R.D.; Grimwood, J.; Jenkins, J.; Lindquist, E.; Tice, H.; Bauer, D.; et al. The genome of *Eucalyptus grandis*. *Nature* **2014**, *510*, 356. [CrossRef]
75. Breseghello, F.; Sorrells, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* **2006**, *172*, 1165–1177. [CrossRef]
76. Hadfield, J.D. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
77. R Core Team. *R: A Language and Environment for Statistical Computing, 3.6.1*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
78. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **1996**, *58*, 267–288. [CrossRef]
79. Legarra, A.; Robert-Granié, C.; Croiseau, P.; Guillaume, F.; Fritz, S. Improved LASSO for genomic selection. *Genet. Res.* **2011**, *93*, 77–87. [CrossRef]
80. Mora, F.; Ballesta, P.; Serra, N. Bayesian analysis of growth, stem straightness and branching quality in full-sib families of *Eucalyptus globulus*. *Bragantia* **2019**, *78*, 1–9. [CrossRef]
81. Torres, L.G.; Rodrigues, M.C.; Lima, N.L.; Trindade, T.F.H.; e Silva, F.F.; Azevedo, C.F.; DeLima, R.O. Multi-trait multi-environment Bayesian model reveals G × E interaction for nitrogen use efficiency components in tropical maize. *PLoS ONE* **2018**, *13*, e0199492. [CrossRef]
82. Volpato, L.; Alves, R.S.; Teodoro, P.E.; de Resende, M.D.V.; Nascimento, M.; Nascimento, A.C.C.; Ludke, W.H.; da Silva, F.L.; Borém, A. Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. *PLoS ONE* **2019**, *14*, e0215315. [CrossRef]
83. Mora, F.; Zúñiga, P.E.; Figueroa, C.R. Genetic variation and trait correlations for fruit weight, firmness and color parameters in wild accessions of *Fragaria chiloensis*. *Agronomy* **2019**, *9*, 506. [CrossRef]
84. Baltunis, B.S.; Huber, D.A.; White, T.L.; Goldfarb, B.; Stelzer, H.E. Genetic gain from selection for rooting ability and early growth in vegetatively propagated clones of loblolly pine. *Tree Genet. Genomes* **2007**, *3*, 227–238. [CrossRef]
85. Burdon, R.D. Short note: Coefficients of variation in variables with bounded scales. *Silvae Genet.* **2008**, *57*, 179–180. [CrossRef]

Genomic Predictions Using Low-Density SNP Markers, Pedigree and GWAS Information: A Case Study with the Non-Model Species *Eucalyptus cladocalyx*

Paulina Ballesta [1], David Bush [2] , Fabyano Fonseca Silva [3] and Freddy Mora [1,*

[1]   Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile; pballesta@utalca.cl
[2]   CSIRO–Australian Tree Seed Centre, Acton 2601, Australia; David.Bush@csiro.au
[3]   Department of Animal Science, Universidade Federal de Viçosa, Viçosa 36570-900, Brazil;
      fabyanofonseca@ufv.br
*   Correspondence: fmora@utalca.cl

Abstract: High-throughput genotyping techniques have enabled large-scale genomic analysis to precisely predict complex traits in many plant species. However, not all species can be well represented in commercial SNP (single nucleotide polymorphism) arrays. In this study, a high-density SNP array (60 K) developed for commercial *Eucalyptus* was used to genotype a breeding population of *Eucalyptus cladocalyx*, yielding only ~3.9 K informative SNPs. Traditional Bayesian genomic models were investigated to predict flowering, stem quality and growth traits by considering the following effects: (i) polygenic background and all informative markers (GS model) and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model). The estimates of pedigree-based heritability and genomic heritability varied from 0.08 to 0.34 and 0.002 to 0.5, respectively, whereas the predictive ability varied from 0.19 (GS) and 0.45 (GSq). The GSq approach outperformed GS models in terms of predictive ability when the proportion of the variance explained by the significant marker-trait associations was higher than those explained by the polygenic background and non-significant markers. This approach can be particularly useful for plant/tree species poorly represented in the high-density SNP arrays, developed for economically important species, or when high-density marker panels are not available.

Keywords: Bayesian models; deviance information criterion; marker-trait associations; predictive ability

1. Introduction

A major focus of modern quantitative genetics is on assessing the association between polymorphic markers with phenotypic variations of complex traits. In this sense, genotype–phenotype studies for quantitative traits at the genome level usually require high-density genetic marker panels, i.e., a large number of markers throughout the genome and large population sizes to obtain sufficient power and prediction resolution [1,2]. The development of several genotyping platforms through high-density single nucleotide polymorphism (SNP) arrays, such as genotyping-by-sequencing (GBS) or SNP chips, has enabled the identification of quantitative trait loci (QTL) for different target traits in various plant species [3–6]. Silva-Junior et al. [7], for instance, developed a genome-wide SNP chip for multiple species of *Eucalyptus*, which has been effective for genomic studies in a wide variety of economically important eucalypt species and their hybrids, including *Eucalyptus grandis*, *Eucalyptus urophylla*, *Eucalyptus nitens* and *Eucalyptus globulus* [8–12]. However, despite the versatility of this SNP

array, it does not perform as well in terms of genome coverage or number of available SNPs for species which are more-distantly related to those for which the chip was developed [13].

According to Pryce et al. [14], the use of low-density marker panels will inevitably affect the precision of QTL detection in genome-wide association studies (GWAS) and the accuracy of genomic prediction of target traits to some degree. On the other hand, Müller et al. [15] found that prediction models using a low-density marker panel (a subset of ~5000 SNPs) provided predictive abilities almost equivalent to using all available SNPs in *Eucalyptus* spp., however, they concluded that it is not yet clear whether the use of smaller SNP subsets is warranted for the long-term implementation of genomic selection in *Eucalyptus*, an aspect that remains still unknown. Marker panels could be considered as either low- or high-density depending on the genome size, the extent of linkage disequilibrium (LD) and the traits of interest. Larger genomes and rapid breakdown of LD would imply that a higher density of SNP loci would be needed to detect QTL. However, many authors quite arbitrarily describe panels as high or low density such that comparisons among organisms should be carefully made. In cattle (genome size 3 Gb), for instance, a DNA array of 50 K SNPs is considered a low-density panel [14], while in *Eucalyptus nitens* (640 Mb) the 60 K SNP chip was considered as a high-density marker panel by Suontama et al. [11]. However, LD in the domesticated bovine genome, though it decays rapidly [16], is more extensive than in nearly-wild eucalypts, thus comparison of SNP density is further complicated. Müller et al. [15] considered the subset of ~5 K SNPs that they used as a low-density marker panel in *Eucalyptus benthamii* and *Eucalyptus pellita*. Notably, low-density SNP chips are considered as a way to reduce the cost of high-density SNP panels in animal breeding and would enable cost-effective implementation of genomic studies [17,18]. In accordance with this, recently Silva et al. [19] used linkage (LA) and linkage disequilibrium (LDA) analyses (termed 'LALDA') for low density-based genomic selection (GS) purposes in animal breeding. In a Bayesian framework, the authors evaluated several GS models and verified the slight superiority of the LALDA models in comparison to traditional LDA models, concluding that the best performance evidenced by the LALDA approach can be due mainly to the small number of markers used, since it enabled to exploit relevant genomic regions that were not directly considered in the LDA, in which this extra information may have contributed to the improvement of the model performance. Similarly, other studies have explored the benefits of including fixed-effect covariates tagging peak genome-wide association study (GWAS) signals [20]. Based on their results, the authors conclude that the universal implementation of GS + GWAS for predicting the breeding values of all possible traits should be investigated on a trait-by-trait basis. On the other hand, Bernardo [21] determined whether explicitly modeling the effects of known major genes affects the response to genomic selection and showed that specifying a fixed effect for a single major gene was never disadvantageous except with a gene explains <10% of genetic variance. In that case, it should be included as a covariate in the traditional ridge-regression best linear unbiased prediction (RR-BLUP) model. Therefore, the objectives of this study were (i) to investigate the possible benefits of including marker-trait associations (from a GWAS analysis) and pedigree information into traditional GS model to predict complex traits in trees of *E. cladocalyx* using low-density SNP markers, and (ii) to assess the efficiency of Bayesian whole-genome regression models (Bayes A, Bayes B, Bayes Cπ and Bayesian Ridge Regression) in terms of predictive ability of complex traits and goodness-of-fit measures in the presence of QTL-genotype effects obtained from a marker-trait association analysis.

2. Results

2.1. SNP Data and Comparison of Genomic Prediction Models

In the present study, we genotyped a breeding population of *E. cladocalyx* using the 60 K SNP chip, yielding a subset of only ~3.9 K informative SNPs, due to the low number of polymorphic loci found at locations where they might have been expected. The ~3.9 K SNPs (~6% of the total SNP array) that were retained after filtering for minor allele frequency (MAF) and missing data were located in all eleven chromosomes of *Eucalyptus*, with an average of ~353 SNPs per chromosome, a density of 6

SNPs per 1 Mb and distributed with an average distance between SNPs of 11,600 bp (Table S1 and Figure S1). The genetic diversity of the genotyped population (evaluated with the 3879 SNPs) was 0.28 (in terms of expected heterozygosity) and the observed heterozygosity was 0.22. According to the population structure analysis, the studied population was strongly differentiated in three groups, with the following pairwise F_{ST} values: F_{ST} 1 = 0.086, F_{ST} 2 = 0.28 and F_{ST} 3 = 0.25.

The association analysis identified a total of 90 significant marker-trait associations (MTAs), which were also distributed across all 11 chromosomes, of which 11, 16, 5, 26, 10, 5 and 17 MTAs were identified for total tree height (HT), diameter at breast height (DBH), stem straightness (STR), slenderness index (SLD), wood density (WD), flowering intensity (FI) and first bifurcation height (BHT), respectively (Figure 1). The MTAs explained a relatively low proportion of the total phenotypic variation, with values of 2% to 4%, 3% to 6%, 3% to 4%, 3% to 5%, 2% to 10%, 3% to 7% and 3% to 4% for HT, DBH, STR, BHT, SLD, WD and FI, respectively. The SNPs involved in these MTAs were posteriorly considered as relationship matrices in the prediction models that include the QTL-genotype effects (GSq model).
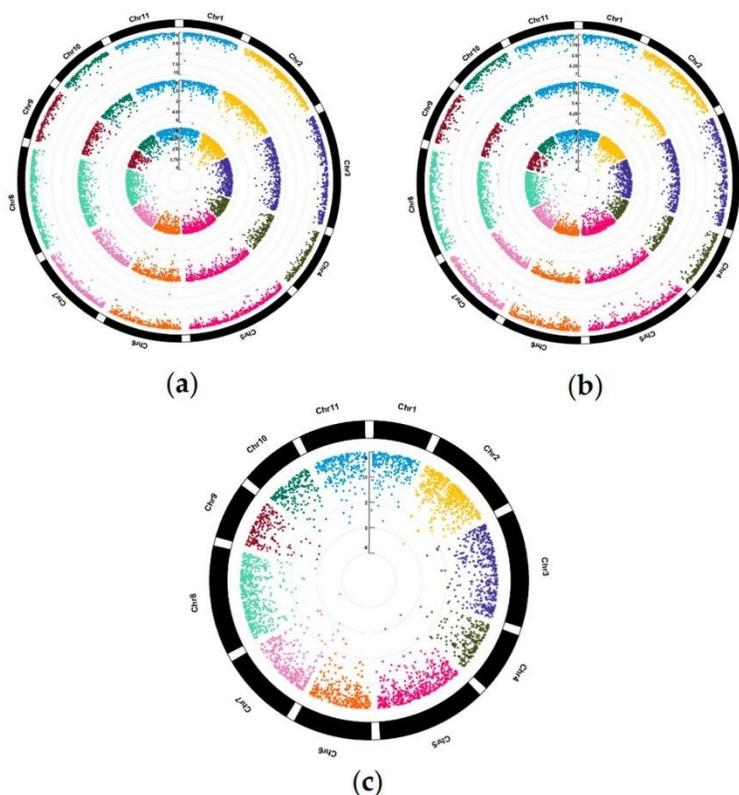

Figure 1. Manhattan plot for (**a**) growth-related traits (tree height, diameter at breast height and slenderness index; displayed from inside to outside), (**b**) stem quality traits (stem straightness, wood density and first bifurcation height; displayed from inside to outside) and (**c**) flowering intensity.

The prediction models that include the MTA information (GSq) and traditional genomic prediction models (GS) were compared in terms of goodness of fit through the Deviance Information Criterion

147

(DIC) and predictive ability (PA). The DIC and PA values of all fitted models are shown in Tables 1 and 2, respectively. GSq models outperformed the GS model in terms of PA when more than 10 significant MTAs were included in GSq predictions (i.e., the following traits HT, DBH, SLD, BHT and WD), whose values ranged between 0.19–0.39 (GS) and 0-24-0.45 (GSq). This result was consistent with the goodness-of-fit measures in most of cases. For HT, GS model outperformed the GSq model for Bayes B (BB), according to the ΔDIC value, while the GSq model was significantly superior in Bayes A (BA), Bayes C (BC) and Bayesian Ridge Regression (BRR) methods (ΔDIC > 10). The GSq approach had a higher predictive ability than the GS approach for all Bayesian models. For DBH, in contrast, the best performance (in terms of goodness-of-fit and PA) was obtained by GSq models based on any Bayesian prediction method (BA, BB, BC or BRR; with ΔDIC > 50). In addition, the predictive ability of DBH based on GSq model was two times higher than those based on traditional GS. For STR, GS models presented significantly lower DIC values than GSq models for BC and BRR (ΔDIC > 10), while the PA values for both approaches were similar. For SLD, goodness-of-fit measures for GSq models were better than traditional models based on any Bayesian genomic model (ΔDIC > 20). The predictive ability of SLD varied between 0.20–0.21 and 0.31–0.32 for GS and GSq models, respectively. For WD, the most of GSq models had better goodness-of-fit measures compared with GS models (ΔDIC > 5) for all Bayesian methods. The PAs of WD ranged between 0.27 and 0.43, which were higher for all GSq models than those based on the GS approach. For FI, GS models presented lower DIC values than GSq models for BA, BB, BC and BRR prediction methods (ΔDIC > 5). Predictive ability values for FI based on both approaches were similar and varied from 0.23 to 0.25. For BHT, there was a strong superiority of the GSq models over GS in all Bayesian methods in terms of goodness-of-fit measures (ΔDIC > 60). Moreover, the PA of BHT varied between 0.19–0.20 and 0.38–0.39 for GS and GSq models, respectively. Consistently with ΔDIC values, the predictive ability of BHT based on GSq models was two times higher than those based on GS models.

Table 1. Deviance information criterion (DIC) of genomic prediction in *Eucalyptus cladocalyx* based on (i) polygenic background (pedigree information) and all informative markers (GS model) and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model).

Trait/Model	Bayes A	Bayes B	Bayes C	BRR [b]
Tree height				
GS	1968.9	1959.5	1968.6	1965.4
GSq	1951.2	1971.3	1941.3	1941.2
ΔDIC [a]	17.7 **	11.8 **	27.3 **	24.2 **
Diameter at breast height				
GS	2556.5	2544.7	2539.8	2538.2
GSq	2490.4	2480.7	2480.2	2473.3
ΔDIC	66.1 **	64.0 **	59.6 **	64.9 **
Stem straightness				
GS	947.7	941.4	932.7	935.3
GSq	947.0	944.7	947.2	946.2
ΔDIC	0.7	3.3	14.5 **	10.8 **
Slenderness index				
GS	4302.5	4299.5	4294.3	4290.5
GSq	4268.1	4268.5	4264.3	4261.3
ΔDIC	34.4 **	31.0 **	30.0 **	29.2 **
Wood density				
GS	2094.3	2082.4	2101.8	2042.0
GSq	2067.1	2075.9	2067.6	2070.3
ΔDIC	27.2 **	6.5 *	34.3 **	28.4 **

Table 1. *Cont.*

Trait/Model	Bayes A	Bayes B	Bayes C	BRR [b]
Flowering intensity				
GS	1293.9	1301.2	1282.3	1285.9
GSq	1306.0	1309.7	1301.3	1295.2
ΔDIC	12.1 **	8.4 *	19.0 **	9.4 *
First bifurcation height				
GS	1491.9	1490.6	1487.3	1485.8
GSq	1426.6	1424.9	1424.9	1423.5
ΔDIC	65.3 **	65.7 **	62.4 **	62.3 **

[a] Difference between DIC values of GSq and GS models. [b] Bayesian Ridge Regression. * Substantial statistical difference between GSq and GS models. ** Strong evidence of statistical difference between GSq and GS models.

Table 2. Predictive ability (PA) of all studied traits in *Eucalyptus cladocalyx* according to (i) polygenic background (pedigree information) and all informative markers (GS model) and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model). The PA values for each method correspond to the mean of PA values for 20-folds of cross-validation.

Trait/Model	Bayes A	Bayes B	Bayes C	BRR [a]	$\bar{X}_{PA}$ [b]
Tree height					
GS	0.33	0.32	0.33	0.34	0.33
GSq	0.45	0.44	0.44	0.45	0.44
Diameter at breast height					
GS	0.21	0.23	0.22	0.22	0.22
GSq	0.41	0.41	0.41	0.42	0.41
Stem straightness					
GS	0.39	0.39	0.39	0.39	0.39
GSq	0.40	0.40	0.40	0.39	0.40
Slenderness index					
GS	0.20	0.20	0.21	0.21	0.21
GSq	0.32	0.32	0.31	0.31	0.32
Wood density					
GS	0.27	0.27	0.27	0.28	0.27
GSq	0.43	0.43	0.43	0.43	0.43
Flowering intensity					
GS	0.25	0.25	0.24	0.23	0.24
GSq	0.25	0.25	0.25	0.24	0.25
First bifurcation height					
GS	0.19	0.20	0.20	0.19	0.19
GSq	0.38	0.38	0.39	0.39	0.38

[a] Bayesian Ridge Regression. [b] Corresponds to the average of PA values.

2.2. Heritability Estimates

The heritability estimates of the studied traits based on all genomic prediction models are shown in Table 3. The estimates of pedigree-based heritability ($\hat{h}_a^2$) for GS were higher than GSq for all traits, whereas the values of genomic heritability ($\hat{h}_m^2$ and $\hat{h}_q^2$) were dependent on the Bayesian prediction

method for both models (GS and GSq). Based on the best fitted models (in terms of ΔDIC values), the heritability estimates of HT varied between 0.13 and 0.21 ($\hat{h}_a^2$), 0.14 and 0.45 ($\hat{h}_m^2$), and 0.29 and 0.34 ($\hat{h}_q^2$). Based on the models with the lowest DIC values, the heritability estimates of DBH based on pedigree and MTAs ($\hat{h}_q^2$) were similar, which ranged between 0.05 and 0.17, while the heritability estimates based on SNPs ($\hat{h}_m^2$) varied between 0.4 and 0.45. In the case of STR, the estimates of pedigree-based heritability for GS models varied between 0.16 and 0.23, while the estimates of genomic heritability ranged between 0.3 and 0.32. For SLD, the heritability estimates based on pedigree information varied between 0.08 and 0.10. The heritability estimates based on MTAs varied between 0.33 and 0.39, while the heritability estimates based on SNPs ($\hat{h}_m^2$) varied between 0.02 and 0.17. The heritability estimates of WD based on pedigree information (GSq models) ranged between 0.14 and 0.17. The heritability estimates of no-significant QTLs by MTAs analysis varied between 0.12 and 0.24, and those for significant QTLs ranged between 0.27 and 0.31. The heritability estimates of FI based on pedigree with a better goodness-of-fit measure (i.e., the GS model) varied between 0.27 and 0.34, while the genomic heritability estimates ranged between 0.07 and 0.29. Based on the models with the lowest DIC values (i.e., all GSq models), the pedigree-based heritability of BHT was 0.08 in the context of all Bayesian methods (BA, BB, BC and BRR). The heritability estimates based on non-significant SNPs by MTAs analysis varied between 0.04 and 0.13, while those based on the significant QTL varied between 0.41 and 0.45.

Table 3. Estimates of heritability of the studied traits for each Bayesian genomic model and effect (i) polygenic background (pedigree information) and all informative markers (GS model), and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model). $\hat{h}_a^2$ corresponds to the pedigree-based estimated heritability. $\hat{h}_m^2$ is the heritability estimate based on a set of markers that were not found to be significantly associated with a trait (GSq) or all SNP markers (GS), $\hat{h}_q^2$ represents the heritability estimates based on a set of SNPs significantly associated with a trait.

Trait/Model	Bayes A			Bayes B			Bayes C			BRR [a]		
	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$
Tree height												
GS	0.28	0.24	-	0.21	0.45	-	0.22	0.40	-	0.29	0.24	-
GSq	0.16	0.14	0.32	0.18	0.12	0.29	0.13	0.27	0.29	0.14	0.17	0.34
Diameter at breast height												
GS	0.20	0.14	-	0.15	0.37	-	0.15	0.39	-	0.19	0.24	-
GSq	0.11	0.05	0.44	0.09	0.15	0.42	0.09	0.17	0.40	0.10	0.11	0.45
Stem straightness												
GS	0.23	0.32	-	0.18	0.31	-	0.16	0.30	-	0.21	0.30	-
GSq	0.18	0.18	0.01	0.14	0.37	0.01	0.14	0.32	0.01	0.18	0.19	0.013
Slenderness index												
GS	0.19	0.12	-	0.16	0.27	-	0.15	0.33	-	0.18	0.21	-
GSq	0.09	0.02	0.39	0.08	0.05	0.36	0.08	0.17	0.33	0.09	0.10	0.35
Wood density												
GS	0.25	0.28	-	0.18	0.50	-	0.19	0.45	-	0.21	0.42	-
GSq	0.17	0.13	0.31	0.17	0.18	0.27	0.14	0.24	0.27	0.17	0.12	0.31
Flowering intensity												
GS	0.34	0.07	-	0.32	0.10	-	0.27	0.29	-	0.33	0.13	-
GSq	0.30	0.06	0.00	0.29	0.06	0.00	0.27	0.20	0.00	0.31	0.10	0.002
First bifurcation height												
GS	0.20	0.05	-	0.19	0.12	-	0.16	0.27	-	0.19	0.14	-
GSq	0.08	0.04	0.44	0.08	0.11	0.42	0.08	0.13	0.41	0.08	0.06	0.45

[a] Bayesian Ridge Regression.

150

3. Discussion

3.1. Marker-Trait Associations for All Studied Traits

Eucalyptus cladocalyx is not a close relative of other eucalypts. Brooker [22] placed *E. cladocalyx* in the monophyletic section *Sejunctae* [23]. As the marker panel we used had been developed for the most widely-planted species which all fall within sections *Maidenaria*, *Exsertaria* and *Latoangulatae*, a lower rate of cross-species amplification of SSR (Simple Sequence Repeats) can be expected in *E. cladocalyx* [24] due to differentiation among widely-distant sections. Despite the low availability of SNP markers, the genetic diversity values were similar to other previous studies of natural populations of *E. cladocalyx* [23,25]. The strong genetic differentiation in three clusters had also been previously reported [23,25,26].

Ninety significant marker-trait associations (MTAs) were detected for the seven target traits, which were subsequently used for the GSq approach. In accordance with this, several studies have previously identified genomic regions explaining part of the phenotypic variation of growth-related traits in *E. cladocalyx* (e.g., Ballesta et al. [25], Arriagada et al. [26], Maldonado et al. [27], Valenzuela et al. [28]). For HT, DBH and SLD, the MTAs were mainly located on chromosomes Chr2 (5 and 4 MTAs for HT and DBH, respectively), Chr6 (3 and 6 MTAs for DBH and SLD, respectively), Chr8 (4 MTAs for SLD) and Chr10 (4 MTAs for SLD). In agreement with these results, Maldonado et al. [27] detected one QTL, based on Simple Sequence Repeat (SSR) markers, located on linkage group LG6 explaining up to 27% of the phenotypic variation of DBH. In addition, Arriagada et al. [26] reported SSR markers associated with HT and DBH, located on the linkage groups LG6, LG8 and LG10, explaining up to 23% of the phenotypic variation.

Flowering components are target traits in breeding programs of some species of *Eucalyptus*, as the flowers provide a reliable source for honey production [26,29]. In dry regions of Chile and South Africa, prolific flowering from forest plantations of *E. cladocalyx* and other eucalypts is particularly advantageous for the supply of honey [30]. In the present study, almost all associations (4/5 MTAs) for flowering intensity were located on chromosome Chr2, which is in accordance with Missiaggia et al. [31], who reported a major QTL located on chromosome Chr2 (*Eef1*) controlling the early flowering in *Eucalyptus grandis*. In *E. cladocalyx*, previous studies have reported that the flowering intensity and early flowering have a positive genetic correlation and common QTLs controlling the phenotypic variation of both traits [29,32]. Interestingly, according to linkage disequilibrium analyses, only two significant SNPs (MTAs for FI) were in disequilibrium, which covered a genomic region of 17,849 bp. For WD, BHT and STR, the MTAs were mainly located on chromosomes Chr2 (WD), Chr5 (BHT) and Chr8 (STR, WD and BHT). In accordance whit this, Valenzuela et al. [28] detected a QTL located on Chr2 explaining 8% of the phenotypic variation of WD in *E. cladocalyx*.

3.2. Comparison between Genomic Prediction Models

Several studies have explored the potential of the selection based on genomic tools in forest species [33–36], including *E. cladocalyx* [12,37,38]. According to the results, GSq models outperformed traditional GS models in terms of predictive ability when at least ten significant marker-trait associations were included in GSq. In addition, another important finding of this study was that the GS and GSq models that include the pedigree information (i.e., pedigree information as a relationship matrix), outperformed the model based solely on SNP marker effects (in terms of goodness-of-fit), revealing the importance of polygenic effects in the prediction model based on low-density markers; an aspect emphasized by Silva et al. [19]. For instance, the predictive ability of flowering intensity based on only the SNP marker panel was three times lower than those based on GS or GSq.

According to De Los Campos et al. [39], Bayes A, Bayes B, Bayes C and BRR methods can improve the predictive ability in genome-based evaluations, but these prediction models could have overfitting problems when the ratio of number of markers and individuals is over 50 [40]. To overcome this, the use of genomic relationship matrices between individuals into genomic prediction models could

beneficially capture general information and reduce the dimensionally problem [41,42], while exploring regions in linkage disequilibrium with QTLs [43]. Interestingly, we found that the GSq models had a better fit than GS models and, at the same time increased the predictive ability of BHT, DBH, SLD and WD. Notably, these benefits were only detected for the prediction of traits with greater than ten MTAs, while for the traits with a lower number of MTAs (i.e., HT, FI and STR), the GS models had a better fit than GSq models. The predictive ability for STR and FI was similar between GSq and GS models. It is worth mentioning that the MTAs detected by the classical linkage analysis explained relatively low values of the total phenotypic variation; which is in accordance with the genetic architecture of quantitative traits. According to Silva et al. [19], the superior performance of the models that include QTL information compared with GS model can be due to GSq model exploiting relevant regions not directly considered in GS model, improving the performance of the prediction model. Additionally, previous studies have reported that a pre-selection of SNPs or the use of genome-wide association analyses to identify and rank markers could increase predictive ability [38,44,45].

Although the analytical assumptions differ among studied Bayesian genomic models, the predictive ability of the studied traits was not severely different among them. In accordance with our findings, several studies have reported that the predictive ability did not differ between methods in forest tree species, especially for growth and wood quality related traits [15,33,36,46–48]. In the present study, the main differences in predictive ability values were observed between the GSq and GS methods, so that the superiority of GSq (or GS) to predict the studied traits was conserved for any Bayesian genomic model.

3.3. Heritability Estimates

According to the goodness-of-fit measures, the GSq model outperformed the GS model for BHT, DBH, SLD and WD, whereas GS offered better model fit compared with GSq for STR and FI. Additionally, for HT, DBH, SLD, WD and BHT, the genomic heritability estimates based on MTAs were higher than the estimates based on SNPs (not significantly associated with a trait). On the other hand, the genomic heritability estimates based on SNP markers (not associated) considering BB and BC method, were higher than those based on BA and BRR methods (for all traits, except HT). BB and BC models involve variable selection procedures, which favor the selection based on major effect markers/genes [49]. Notably, other studies have confirmed that some whole-regression methods could overestimate the heritability values [50], and therefore, these findings should be interpreted with caution. Overall, all target traits had pedigree-based heritability estimates from low to moderate ($h^2 = 0.08$–0.34), which are in accordance with the range usually expected for forest tree growth, flowering and stem quality related traits, including *E. cladocalyx* [28,29,51–54].

The heritability estimates based on MTAs for STR and FI were lower than those based on pedigree information or non-significant SNPs from the MTA analysis. In tree species, several studies had reported genomic regions explaining a high percent of phenotypic variation of STR and FI. For example, Arriagada et al. [26] reported one QTL explaining up to 15% of the total variation of STR in *E. cladocalyx*. In a meta-analysis, Hall et al. [55] confirmed that phenological traits are highly heritable and controlled by key genomic regions in trees. In fact, Missiaggia et al. [31] reported a key region on chromosome Chr2 (*Eef1*) controlling the early flowering in *E. grandis*. In this context, low-density SNP panels could limit the probability to detect key regions explaining the phenotypic variation of these traits. In addition, these results are supported by the fact that the GS model outperformed the GSq model for STR and FI (in terms of goodness-of-fit), which means that the total variation of STR and FI is better explained by the classic prediction model (i.e., σ_a^2 and σ_m^2).

4. Materials and Methods

4.1. Plant Material and Phenotypic Evaluation

A genomic selection study was performed in a long-term open-pollinated progeny trial comprising 49 families of *E. cladocalyx* established in 2001 in northern Chile (locality of Los Vilos; 31°54′ S; 71°27′ W; 167 m.a.s.l). The climate is classified as predominantly arid, according to the De Martonne aridity index [26]. Trees were arranged in a randomized complete block design with 30 blocks and single-tree plots (total of 1470 trees). Trees were planted at 2 m spacing within rows and 3 m between rows (~1667 trees ha^{-1}). The following quantitative traits were measured in 17 years-old trees: diameter at breast height (DBH), total tree height (HT), first bifurcation height (BHT), wood density (WD), stem straightness (STR) and slenderness index (SLD). The BHT was rated on a scale of five levels, in which a value of 1 is assigned to trees with a loss of the central axis in the first fifth of the tree's height, a value of 2 indicates that a loss of the central axis occurs in the second fifth of the tree, a value of 3 indicates that a loss of the central axis occurs in the third fifth of the tree, a value of 4 implies a loss of the central axis in the fourth fifth of the tree's height, and a value of 5 implies that a loss of the central axis in the last fifth of the tree height or does not show loss of the apical axis (modified scale by Bush et al. [54]). WD was measured indirectly according to Valenzuela et al. [28]. STR was measured in the first two-thirds of the total height of tree and was considered as ordinal variables with four levels [53], in which a value of 0 if the stem was strongly twisted, a value of 1 if the stem presents moderate levels of curvature, a value of 2 if the stem was slightly curved, and 3 if the stem was completely straight. The SLD was calculated as the ratio between HT (m) and DBH (m). Additionally, flowering intensity (FI) was measured using a scale that ranged from 0 to 3, in 18 years-old trees, according to Arriagada et al. [26], where a value of 0 means absence of flowers, buds and/or capsules, a value of 1 sparse flowers on a small part of crown, a value of 2 if the flowers/capsules/buds were covering the half of the crown, and a value of 3 for trees with numerous flowers on the whole crown.

4.2. DNA Extraction and Tree Genotyping

DNA for tree genotyping was isolated from leaf tissues of 480 individuals according to Porebsky et al. [56] and Doyle and Doyle [57]. On average, 10 individuals per family were randomly selected to be genotyped using the Illumina Infinium EUChip60K SNP array [7]. The SNP data were filtered for SNP call rate score >0.7 and minor allele frequency (MAF) >0.05. The missing data were imputed using the LD-kNNi method in Tassel 5.2 [58]. Linkage disequilibrium between marker pairs (MTAs at the same chromosome) was calculated using TASSEL version 5.2.

4.3. Genomic Prediction Models

The following four Bayesian whole-genome regression models were used for the estimation of SNP marker effects, variance components and genomic heritability: Bayes A ([59]; BA), Bayes B ([59]; BB), Bayes Cπ ([60]; BC) and Bayesian Ridge Regression ([61]; BRR). Due to the low density of markers found in this studied population, the following two approaches were used to predict the studied traits: traditional genomic prediction model (GS), considering all informative markers (~3.8 K SNP) from the commercial SNP panel, and a combination of traditional GS and QTL information (GSq). The GS model is defined as:

$$\mathbf{y}* = \mathbf{1}\mu + \sum_{i=1}^{m} \mathbf{x_i}\mathbf{m_i} + \mathbf{Za} + \varepsilon \tag{1}$$

where $\mathbf{y}*$ is the vector of phenotypic records pre-corrected for the effects of block and genetic structure [62,63]. Respectively, $\mathbf{1}$ and μ are vectors of ones and overall mean. $\mathbf{m_i}$ corresponds to the additive genetic effect of the *i*-th marker, with m as the number of markers. $\mathbf{x_i}$ is the incidence vector of each marker (codified as: AA = 0, AB = 1 and BB = 2). $\mathbf{a}$ corresponds to polygenic effects, with $\mathbf{a} \sim \mathbf{N}(0, \hat{\sigma}_a^2 \mathbf{A})$, and $\mathbf{Z}$ is the incidence matrix related to polygenic effects. The coefficients of relationship matrix

were adjusted according to Bush et al. [54]. Finally, ε is the residual vector, $\varepsilon \sim \mathbf{N}(0, \hat{\sigma}_{\varepsilon}^2 \mathbf{I_n})$. The genomic predictions were performed using BGLR package in R [64]. Specifically, 1,000,000 iterations of Markov Chain Monte Carlo simulations were used in all genomic prediction models, with a burn-in period of 100,000.

The GSq model consisted of two steps, in which the first step involves a marker-trait association (MTA) analysis for QTL detection according to Silva et al. [19]. The MTA analysis was conducted using the following mixed linear model:

$$\mathbf{y*} = \mathbf{Sa} + \mathbf{Qv} + \mathbf{Zu} + \varepsilon \tag{2}$$

where **y*** is a vector of adjusted phenotypic observations (by the block effect). **S**, **Q** and **Z** correspond to the incidence matrices for **a**, **v** and **u**, respectively. **a** and **v** are the vectors of SNP effects (fixed) and population structure effects (fixed). **u** corresponds to a vector of polygenic effects (random), and ε is the residual vector. The variances of **u** and ε are $\mathrm{Var}(\mathbf{u}) = 2\mathbf{K}\sigma_g^2$ and $\mathrm{Var}(\varepsilon) = \mathbf{R}\sigma_e^2$, respectively, where **K** is the kinship coefficient matrix, which was estimated using the program TASSEL [58]. For this analysis, the significance threshold of 0.001 and a false discovery rate (FDR) <10% were applied to test for significant MTAs according to Uchiyama et al. [65]. The GWAS analysis was carried out using rrBLUP package in R v. 3.5 [66]. The population structure was assessed by Bayesian model-based clustering in STRUCTURE software v. 2.3.4 [67]. In addition, the SNP data was used to estimate the genetic diversity of the genotyped population using GenAlex v.6.5 [68].

The second step of the GSq approach consists of the compilation of traditional GS model (Equation (1)) and QTL-genotype effect (**q**) from the association analysis, which is expressed as:

$$\mathbf{y*} = \mathbf{1}\mu + \sum_{i=1}^{m} \mathbf{x_i m_i} + \mathbf{Za} + \mathbf{Zq} + \varepsilon \tag{3}$$

where **y*** is the vector of phenotypic records pre-corrected for the effects of block and genetic structure, and **1** and μ are vectors of ones and overall mean, respectively. $\mathbf{m_i}$ corresponds to the additive genetic effect of the *i*-th marker that were not found to be significantly associated with a trait. $\mathbf{x_i}$ is the incidence vector of each marker (AA = 0, AB = 1 and BB = 2). The variance of m ($\hat{\sigma}_m^2$) depends on the Bayesian model implemented. **Z** is the incidence matrix of polygenic (**a**) and QTL-genotype (**q**) effects assuming $\mathbf{a} \sim \mathbf{N}(0, \hat{\sigma}_a^2 \mathbf{A})$ and $\mathbf{q} \sim \mathbf{N}(0, \hat{\sigma}_q^2 \mathbf{Q})$, respectively. The **Q** is a covariance matrix, in which the elements are the probabilities that individuals are identical by descent based on significant SNP markers according to the MTA analysis.

4.4. Heritability Estimates

For the GSq model, the heritability estimates were obtained as follows:

$$\hat{h}_a^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_m^2 + \hat{\sigma}_q^2 + \hat{\sigma}_e^2} \tag{4}$$

$$\hat{h}_m^2 = \frac{\hat{\sigma}_m^2}{\hat{\sigma}_a^2 + \hat{\sigma}_m^2 + \hat{\sigma}_q^2 + \hat{\sigma}_e^2} \tag{5}$$

$$\hat{h}_q^2 = \frac{\hat{\sigma}_q^2}{\hat{\sigma}_a^2 + \hat{\sigma}_m^2 + \hat{\sigma}_q^2 + \hat{\sigma}_e^2} \tag{6}$$

where $\hat{h}_a^2$, $\hat{h}_m^2$ and $\hat{h}_q^2$ correspond to heritability estimates based on pedigree information, a set of markers that were not found to be significantly associated with a trait and a set of SNPs significantly associated with a trait (MTAs), respectively. $\hat{\sigma}_a^2$ is the variance due to the additive polygenic effect, $\hat{\sigma}_m^2$ corresponds

to the marker effect variance, $\hat{\sigma}_q^2$ is the variance of markers significantly associated with a trait, and $\hat{\sigma}_e^2$ is the residual variance.

For traditional GS model, the term $\hat{h}_m^2$ correspond to the heritability estimates based on all SNPs (n = 3879). In the cases of BC and BRR methods, $\hat{\sigma}_m^2$ was calculated as: $\hat{\sigma}_m^2 = 2\hat{\sigma}_{SNP}^2 \sum_{i=1}^n \hat{p}_i(1-\hat{p}_i)$, in which $\hat{\sigma}_{SNP}^2$ corresponds to a common variance for SNP markers and $\hat{p}_i$ is the MAF of the i-th marker. The $\hat{\sigma}_m^2$ term for BA and BB models was estimated as $\hat{\sigma}_m^2 = 2\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i)\hat{\sigma}_{SNP_i}^2$, in which $\hat{\sigma}_{SNP_i}^2$ corresponds to the variance due to the i-th marker.

4.5. Comparison between Genomic Prediction Models

The Bayesian prediction models were compared in terms of goodness-of-fit, by using Deviance Information Criterion (DIC) [69], and the predictive ability (PA). The DIC is defined by the following expression:

$$DIC = \overline{D} + pD \tag{7}$$

where D is a Bayesian measure of model fit, which is defined as the posterior expectation of the deviance $\left(\overline{D} = E_{\theta/y}[-2 \cdot \ln f(y/\theta)]\right)$; pD is the effective number of parameters. A *DIC* difference of more than 10 between two competitive models (GSq and GS models) was considered to be supported against a model with higher *DIC*; a *DIC* difference between 5 and 10 was considered as substantial difference between models, and a difference less than 5 was considered as not significant.

The predictive ability of each model was calculated as the correlation between the pre-corrected observations ($\mathbf{y}^*$) from validation dataset and the estimated breeding value ($\hat{y}^*$). A total of 20-fold cross-validation was used to evaluate the predictive ability of all models. The $\hat{y}^*$ for GS and GSq models was calculated as $\hat{y}^* = \sum_{i=1}^n \mathbf{x}_i \hat{m}_i + \mathbf{Z}\hat{\mathbf{u}}$ and $\hat{y}^* = \sum_{i=1}^n \mathbf{x}_i \hat{m}_i + \mathbf{Z}\hat{\mathbf{u}} + \mathbf{Z}\hat{\mathbf{q}}$, respectively.

5. Conclusions

In this study, we evaluated the performance of Bayesian genomic models that include the genetic background (pedigree) and QTL information from GWAS analysis, which was specially implemented in the context of a low-density SNP markers. Importantly, predictive abilities were superior in GSq when the proportion of the variance explained by the significant marker-trait associations (MTAs) was higher than those explained by the polygenic background and SNP markers (that were not found to be significantly associated with a trait). Therefore, we emphasized and hypothesized that both the number of associations and/or the percentage of variation explained by the MTAs are determinants in the effectiveness of the GSq method. This approach can be particularly useful for plant/tree species poorly represented in the high-density SNP arrays, developed for economically important species or when high-density marker panels are not available.

References

1. Viana, J.; Pereira, H.D.; Mundim, G.B.; Piepho, H.-P.; Fonseca e Silva, F. Efficiency of genomic prediction of non-assessed single crosses. *Heredity* **2018**, *120*, 283–295. [CrossRef] [PubMed]

2. Singh, D.; Wang, X.; Kumar, U.; Gao, L.; Noor, M.; Imtiaz, M.; Singh, R.P.; Poland, J. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Front. Plant Sci.* **2019**, *10*, 394. [CrossRef] [PubMed]

3. Contreras-Soto, R.I.; Mora, F.; De Oliveira, M.A.R.; Higashi, W.; Scapim, C.A.; Schuster, I. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS ONE* **2017**, *12*, e0171105. [CrossRef]

4. Maldonado, C.; Mora, F.; Scapim, C.A.; Coan, M. Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: hapLA4. *PLoS ONE* **2019**, *14*, e0212925. [CrossRef] [PubMed]

5. Senhorinho, H.J.C.; Coan, M.M.D.; Marino, T.P.; Kuki, M.C.; Barth-Pinto, R.S.; Scapim, C.A.; Holland, J.B. Genomic-Wide Association Study of Popping Expansion in Tropical Popcorn and Field Corn Germplasm. *Crop Sci.* **2019**, *59*, 2007–2019. [CrossRef]

6. Mafra, G.S.; Do Amaral Júnior, A.T.; Almeida, F.J.E.D.; Vivas, M.; Araújo Diniz-Santos, P.H.; Saltires-Santos, J.; Ferreira-Pena, G.; De Lima, V.J.; Kamphorst, S.H.; De Oliveira, F.T.; et al. SNP-based mixed model association of growth-and yield-related traits in popcorn. *PLoS ONE* **2019**, *14*, e0218552. [CrossRef]

7. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* **2015**, *206*, 1527–1540. [CrossRef]

8. Torres-Dini, D.; Nunes, A.C.P.; Aguiar, A.; Nikichuk, N.; Centurión, C.; Cabrera, M.; Moraes, M.L.T.; Resende, M.D.V.; Sebbenn, A.M. Clonal selection of *Eucalyptus grandis* x *Eucalyptus globulus* for productivity, adaptability, and stability, using SNP markers. *Silvae Genet* **2016**, *65*, 30–38. [CrossRef]

9. Klápště, J.; Suontama, M.; Telfer, E.; Graham, N.; Low, C.; Stovold, T.; McKinley, R.; Dumgey, H. Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. *PLoS ONE* **2017**, *12*, e0185137. [CrossRef]

10. Durán, R.; Isik, F.; Zapata-Valenzuela, J.; Balocchi, C.; Valenzuela, S. Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genet Genomes* **2017**, *13*, 74. [CrossRef]

11. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dungey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2019**, *122*, 370. [CrossRef] [PubMed]

12. Ballesta, P.; Maldonado, C.; Pérez-Rodríguez, P.; Mora, F. SNP and Haplotype-Based Genomic Selection of Quantitative Traits in *Eucalyptus globulus*. *Plants* **2019**, *8*, 331. [CrossRef] [PubMed]

13. Aguirre, N.C.; Filippi, C.V.; Zaina, G.; Rivas, J.G.; Acuña, C.V.; Villalba, P.V.; García, M.N.; González, S.; Rivarola, M.; Maetínez, M.C. Optimizing ddRADseq in non-model species: A case study in *Eucalyptus dunnii* Maiden. *Agronomy* **2019**, *9*, 484. [CrossRef]

14. Pryce, J.E.; Johnston, J.; Hayes, B.J.; Sahana, G.; Weigel, K.A.; McParland, S.; Spurlock, D.; Krattenmacher, N.; Spelman, R.J.; Wall, E.; et al. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *J. Dairy Sci.* **2014**, *97*, 1799–1811. [CrossRef] [PubMed]

15. Müller, B.S.; Neves, L.G.; De Almeida, F.J.E.; Resende, M.F.R., Jr.; Muñoz, P.R.; Dos Santos, P.E.T.; Paludzyszyn Filho, E.; Kirst, M.; Grattapaglia, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomes* **2017**, *18*, 524.

16. Porto-Neto, L.R.; Kijas, J.W.; Reverter, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Sel. Evol.* **2014**, *46*, 22. [CrossRef]

17. Bolormaa, S.; Gore, K.; van der Werf, J.H.J.; Hayes, B.J.; Daetwyler, H.D. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Anim. Genet.* **2015**, *46*, 544–556. [CrossRef]

156

18. Wu, X.-L.; Xu, J.; Feng, G.; Wiggans, G.R.; Taylor, J.F.; He, J.; Qian, C.; Qiu, J.; Simpson, B.; Walker, J.; et al. Optimal Design of Low-Density SNP Arrays for Genomic Prediction: Algorithm and Applications. *PLoS ONE* **2016**, *11*, e0161719. [CrossRef]

19. Silva, F.F.; Jerez, E.A.Z.; De Resende, M.D.V.; Soriano-Viana, M.; Ferreira-Azevedo, C.; Lopes, P.S.; Nascimento, M.; Oliveira de Lima, R.; Facioni-Guimaraes, S.E. Bayesian model combining linkage and linkage disequilibrium analysis for low density-based genomic selection in animal breeding. *J. Appl. Anim. Res.* **2018**, *46*, 873–878. [CrossRef]

20. Rice, B.; Lipka, A.E. Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome* **2019**, *12*. [CrossRef]

21. Bernardo, R. Genomewide selection when major genes are known. *Crop Sci.* **2014**, *54*, 68–75. [CrossRef]

22. Brooker, M.I.H. A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Aust. Syst. Bot.* **2000**, *13*, 79–148. [CrossRef]

23. McDonald, M.W.; Rawlins, M.; Butchet, P.A.; Bell, C. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Aust. J. Bot.* **2003**, *51*, 393–403. [CrossRef]

24. Mora, F.; Arriagada, O.; Ballesta, P.; Ruiz, E. Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *J. Plant Biochem. Biotechnol.* **2017**, *26*, 274–281. [CrossRef]

25. Ballesta, P.; Mora, F.; Contreras-Soto, R.I.; Ruiz, E.; Perret, S. Analysis of the genetic diversity of *Eucalyptus cladocalyx* (sugar gum) using ISSR markers. *Acta Sci. Agron.* **2015**, *37*, 133–140. [CrossRef]

26. Arriagada, O.; Do Amaral Junior, A.T.; Mora, F. Thirteen years under arid conditions: Exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breed. Sci.* **2018**, *68*, 367–374. [CrossRef]

27. Maldonado, C.; Contreras-Soto, R.I.; Gerhardt, I.F.S.; Do Amaral Júnior, A.T.; Mora, F. Stable marker-trait associations for growth across different ages in *Eucalyptus cladocalyx* through the use of microsatellites. *Sci. For.* **2018**, *46*, 367–376. [CrossRef]

28. Valenzuela, C.E.; Ballesta, P.; Maldonado, C.; Baettig, R.; Arriagada, O.; Mafra, G.S.; Mora, F. Bayesian Mapping Reveals Large-Effect Pleiotropic QTLs for Wood Density and Slenderness Index in 17-Year-Old Trees of *Eucalyptus cladocalyx*. *Forests* **2019**, *10*, 241. [CrossRef]

29. Cané-Retamales, C.; Mora, F.; Vargas-Reeve, F.; Perret, S.; Contreras-Soto, R. Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. *Euphytica* **2011**, *178*, 177–183. [CrossRef]

30. De Lange, W.J.; Veldtman, R.; Allsopp, M.H. Valuation of pollinator forage services provided by *Eucalyptus cladocalyx*. *J. Environ. Manag.* **2013**, *125*, 12–18. [CrossRef]

31. Missiaggia, A.A.; Piacezzi, A.L.; Grattapaglia, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genet. Genomes* **2005**, *1*, 79. [CrossRef]

32. Contreras-Soto, R.; Ballesta, P.; Ruiz, E.; Mora, F. Identification of ISSR markers linked to flowering traits in a representative sample of *Eucalyptus cladocalyx*. *J. For. Res.* **2016**, *27*, 239–245. [CrossRef]

33. Ratcliffe, B.; El-Dien, O.G.; Klápště, J.; Porth, I.; Chen, C.; Jaquish, B.; El-Kasaby, Y.A. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity* **2015**, *115*, 547–555. [CrossRef] [PubMed]

34. Gamal-Dien, O.; Ratcliffe, B.; Klápště, J.; Chen, C.; Porth, I.; El-Kasaby, Y.A. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genom.* **2015**, *16*, 370.

35. Lenz, P.R.; Beaulieu, J.; Mansfield, S.D.; Clément, S.; Desponts, M.; Bousquet, J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genom.* **2017**, *18*, 335. [CrossRef]

36. Chen, Z.Q.; Baison, J.; Pan, J.; Karlsson, B.; Andersson, B.; Westin, J.; García-Gil, M.R.; Wu, H.X. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genom.* **2018**, *19*, 946. [CrossRef]

37. Bush, D.; Thumma, B. Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genet. Genomes* **2013**, *9*, 741–752. [CrossRef]

38. Ballesta, P.; Serra, N.; Guerra, F.; Hasbún, R.; Mora, F. Genomic prediction of growth and stem quality traits in *Eucalyptus globulus* Labill at its southernmost distribution limit in Chile. *Forests* **2018**, *9*, 779. [CrossRef]

39. De los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. Whole-Genome regression and prediction methods applied to plant and animal breeding. *Genetics* **2013**, *193*, 327–345. [CrossRef]

40. González-Recio, O.; Rosa, G.J.; Gianola, D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* **2014**, *166*, 217–231. [CrossRef]

41. Solberg, T.R.; Sonesson, A.K.; Woolliams, J.A.; Meuwissen, T.H.E. Reducing dimensionality for prediction of genome-wide breeding values. *Genet. Sel. Evol.* **2009**, *41*, 29. [CrossRef] [PubMed]

42. Du, C.; Wei, J.; Wang, S.; Jia, Z. Genomic selection using principal component regression. *Heredity* **2018**, *121*, 12–23. [CrossRef] [PubMed]

43. Habier, D.; Fernando, R.L.; Dekkers, J.C. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **2007**, *177*, 2389–2397. [CrossRef] [PubMed]

44. Macciotta, N.P.; Gaspa, G.; Steri, R.; Pieramati, C.; Carnier, P.; Dimauro, C. Pre-selection of most significant SNPS for the estimation of genomic breeding values. *BMC Proc.* **2009**, *3*, 14. [CrossRef] [PubMed]

45. Arojju, S.K.; Conaghan, P.; Barth, S.; Milbourne, D.; Casler, M.D.; Hodkinson, T.R.; Michel, T.; Byrne, S.L. Genomic prediction of crown rust resistance in *Lolium perenne*. *BMC Genet.* **2018**, *19*, 35. [CrossRef]

46. Resende, M.F., Jr.; Muñoz, P.; Resende, M.D.; Garrick, D.J.; Fernando, R.L.; Davis, M.J.; Jokela, E.J.; Martin, T.A.; Peter, G.F.; Kirst, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **2012**, *190*, 1503–1510. [CrossRef]

47. Beaulieu, J.; Doerksen, T.K.; MacKay, J.; Rainville, A.; Bousquet, J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genom.* **2014**, *15*, 1048. [CrossRef]

48. Thistlethwaite, F.R.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Stoehr, M.U.; El-Kassaby, Y.A. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genom.* **2017**, *18*, 930. [CrossRef]

49. Wolfe, M.D.; Del Carpio, D.P.; Alabi, O.; Ezenwaka, L.C.; Ikeogu, U.N.; Kayondo, I.S.; Lozano, R.; Okeke, U.G.; Ozimati, A.A.; Williams, E.; et al. Prospects for genomic selection in cassava breeding. *Plant Genome* **2017**, *10*, 1–9. [CrossRef]

50. De los Campos, G.; Sorensen, D.; Gianola, D. Genomic heritability: What is it? *PLoS Genet.* **2015**, *11*, e1005048. [CrossRef]

51. Mora, F.; Gleadow, R.; Perret, S.; Scapim, C.A. Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. *Euphytica* **2009**, *169*, 335–344. [CrossRef]

52. Bush, D.; McCarthy, K.; Meder, R. Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). *Ann. For. Sci.* **2011**, *68*, 1057. [CrossRef]

53. Vargas-Reeve, F.; Mora, F.; Perret, S.; Scapim, C.A. Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. *Crop Breed. Appl. Biotechnol.* **2013**, *13*, 107–112. [CrossRef]

54. Bush, D.; Kain, D.; Kanowski, P.; Matheson, C. Genetic parameter estimates informed by a marker-based pedigree: A case study with *Eucalyptus cladocalyx* in southern Australia. *Tree Genet. Genomes* **2015**, *11*, 798. [CrossRef]

55. Hall, D.; Hallingbäck, H.R.; Wu, H.X. Estimation of number and size of QTL effects in forest tree traits. *Tree Genet. Genomes* **2016**, *12*, 110. [CrossRef]

56. Porebski, S.; Bailey, L.G.; Baum, B.R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **1997**, *15*, 8–15. [CrossRef]

57. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.

58. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

59. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829.

60. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef]

61. Gianola, D. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* **2013**, *194*, 573–596. [CrossRef] [PubMed]

62. Asoro, F.G.; Newell, M.A.; Beavis, W.D.; Scott, M.P.; Jannink, J.-L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* **2011**, *4*, 132–144. [CrossRef]

63. Lorenz, A.J.; Smith, K.P.; Jannink, J.L. Potential and optimization of genomic selection for *Fusarium* head blight resistance in six-row barley. *Crop Sci.* **2012**, *52*, 1609–1621. [CrossRef]
64. Pérez, P.; De Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [CrossRef] [PubMed]
65. Uchiyama, K.; Iwata, H.; Moriguchi, Y.; Ujino-Ihara, T.; Ueno, S.; Taguchi, Y.; Tsuboruma, M.; Mishima, K.; Iki, T.; Watanabe, A.; et al. Demonstration of genome-wide association studies for identifying markers for wood property and male strobili traits in *Cryptomeria japonica*. *PLoS ONE* **2013**, *8*, e79866. [CrossRef] [PubMed]
66. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **2011**, *4*, 250–255. [CrossRef]
67. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
68. Peakall, R.O.D.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [CrossRef]
69. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc.* **2002**, *64*, 583–639. [CrossRef]

159

forests

MDPI

Article

Genomic Prediction of Growth and Stem Quality Traits in *Eucalyptus globulus* Labill. at Its Southernmost Distribution Limit in Chile

Paulina Ballesta [1], Nicolle Serra [2], Fernando P. Guerra [1], Rodrigo Hasbún [3] and Freddy Mora [1,*]

[1] Instituto de Ciencias Biológicas, Universidad de Talca, 2 Norte 685, Talca 3460000, Chile; paballesta@gmail.com (P.B.); fguerra@utalca.cl (F.P.G.)

[2] Semillas Imperial SpA, Av. Las Industrias 13320, Los Ángeles 4440000, Chile; nserra@semillasimperial.cl

[3] Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, Concepción 4070386, Chile; rodrigohasbun@udec.cl

* Correspondence: fmora@utalca.cl; Tel.: +56-71-2200280

check for updates

Abstract: The present study was undertaken to examine the ability of different genomic selection (GS) models to predict growth traits (diameter at breast height, tree height and wood volume), stem straightness and branching quality of *Eucalyptus globulus* Labill. trees using a genome-wide Single Nucleotide Polymorphism (SNP) chip (60 K), in one of the southernmost progeny trials of the species, close to its southern distribution limit in Chile. The GS methods examined were Ridge Regression-BLUP (RRBLUP), Bayes-A, Bayes-B, Bayesian least absolute shrinkage and selection operator (BLASSO), principal component regression (PCR), supervised PCR and a variant of the RRBLUP method that involves the previous selection of predictor variables (RRBLUP-B). RRBLUP-B and supervised PCR models presented the greatest predictive ability (PA), followed by the PCR method, for most of the traits studied. The highest PA was obtained for the branching quality (~0.7). For the growth traits, the maximum values of PA varied from 0.43 to 0.54, while for stem straightness, the maximum value of PA reached 0.62 (supervised PCR). The study population presented a more extended linkage disequilibrium (LD) than other populations of *E. globulus* previously studied. The genome-wide LD decayed rapidly within 0.76 Mbp (threshold value of $r^2 = 0.1$). The average LD on all chromosomes was $r^2 = 0.09$. In addition, the 0.15% of total pairs of linked SNPs were in a complete LD ($r^2 = 1$), and the 3% had an r^2 value >0.5. Genomic prediction, which is based on the reduction in dimensionality and variable selection may be a promising method, considering the early growth of the trees and the low-to-moderate values of heritability found in the traits evaluated. These findings provide new understanding of how develop novel breeding strategies for tree improvement of *E. globulus* at its southernmost range limit in Chile, which could represent new opportunities for forest planting that can benefit the local economy.

Keywords: Bayesian methods; stem straightness; branching quality; predictive accuracy; principal components regression

1. Introduction

The genus *Eucalyptus* L'Her. comprises more than 700 species that are distributed mainly in Australia, and these species are planted in a wide variety of environments, such as the Mediterranean, tropical, subtropical and temperate climates [1,2]. Both the species and hybrids of the genus are among the main sources of biomass worldwide and among the main hardwoods used for the production of pulp and wood [3]. For example, *Eucalyptus globulus* Labill. is one of the most planted hardwood

species for industrial uses in various countries with temperate zones (e.g., Portugal, Spain, Uruguay and Chile), which has been the target of breeding in several programs around the world to improve economically important traits such as tree growth and wood quality. *E. globulus* is naturally distributed in coastal south-eastern Australia [4], between 38° and 43° S latitude, where oceanic and subpolar oceanic climates are predominant. However, the species has been extensively planted under temperate conditions [1,2]. Notably, *E. globulus* exhibits physiological plasticity and has been successfully grown in a broad range of environmental conditions that are characterized by moderate abiotic stresses [5–7]. In Chile, for instance, several studies have been conducted to provide an understanding of the mechanisms through which *E. globulus* trees respond to cold conditions [8–10]. In fact, severe cold winter periods restrict the majority of the *E. globulus* plantations to coastal and central Chile [7]. In fact, the southern limit of *E. globulus* in Chile occurs in central Los Lagos (administrative region) at approximately latitude 41° S, accounting for only 4% of the total *E. globulus* plantations in Chile.

Individual selection based on phenotypic data of the traits of interest, which includes pedigree records, has been the most commonly used strategy to genetically improve forest tree species. In this sense, the estimation of quantitative genetic parameters in tree breeding programs is important to ensure efficient selection [11]. On the other hand, the growth rate is relatively long, and low juvenile-mature correlations in forest trees have stimulated interest in marker-assisted selection (MAS) to accelerate breeding through early selection [12]. In MAS, individual tree breeding values are predicted based on the effects of selected markers, and is most effective for traits controlled by few quantitative trait loci (QTLs), each of which controls a relatively large proportion of the total phenotypic variation. In this context, efforts to increase genetic gains using MAS in different breeding programs of *Eucalyptus* have been performed [11–17]. However, the benefits of MAS may be diminished by the low genetic variance that can be explained by a given QTL [18–20]. This principle is particularly important when the selection processes focus mainly on complex traits (controlled by multiple genes), such as stem straightness, wood volume, wood quality and growth traits [21–23]. Moreover, many QTLs underlying a complex trait are useful within the same or related families and environments (e.g., Ukrainetz et al. [24], Mamani et al. [25]), and the use of small population sizes and conventional statistical methods have been inadequate to accurately detect small effects of QTLs [26].

To overcome the limitations of MAS in polygenic traits, Meuwissen et al. [27] proposed the genomic prediction/selection (GS) method; in which environmental factors are important, and it focuses on traits that are affected by a large number of genes. In their study, these researchers compared several frequentist analyses (e.g., best linear unbiased prediction (BLUP) and least squares) and their Bayesian counterpart models based on the predictive ability of breeding values. The GS principles agree with the infinitesimal model, which is based on the fact that breeding values result from the small and additive effects of alleles in a large number of loci [28]. GS is based on the simultaneous prediction of the effects of thousands of DNA markers (e.g., SNPs) scattered throughout the genome of an organism, which disregards the use of significance tests for individual markers. Daetwyler et al. [29] highlighted that GS can increase genetic gain through greater precision in the estimation of breeding values, the reduction in generational intervals and an optimal use of available genetic resources. In the context of the genetic improvement of forest tree species, GS was originally proposed as a promising way to capture a greater proportion of the genetic variation for growth traits [30]. Moreover, according to Suontama et al. [31], Grattapaglia [32], Iwata et al. [33] and Zhong et al. [34], genomic selection is expected to enhance the genetic improvement of plants, including forest tree species by providing more accurate estimates of breeding values compared with pedigree-based methodologies.

Several models and methodological variants have been proposed for a better use of the benefits of GS because this method requires the development of statistical models that usually contain a massive number of markers (predictor variables) and a limited number of phenotypic data. The statistical methods commonly used in GS correspond to Bayes-A, Bayes-B, Bayes Cπ, Bayesian Least Absolute Shrinkage and Selection Operator (LASSO; BLASSO) [27,35,36], Genomic-BLUP (GBLUP) [36] and Ridge Regression BLUP (RRBLUP) [27]. In each GS model, genetic merits are predicted under

different analytical assumptions. Therefore, there is no universally optimal statistical method for all the traits in every population. For instance, Suontama et al. [31] and Durán et al. [37] showed a significant improvement in breeding value accuracy for wood properties in *Eucalyptus* by implementing GBLUP method compared to pedigree based prediction. On the other hand, Resende-Junior et al. [38] determined the SNP effects based on five predictive models in a training population of *Pinus taeda* L., in which the Bayes Cπ, Bayes-A and RRBLUB-B methods (using a subset of selected markers) showed greater predictive ability than RRBLUP and BLASSO methods. In turn, some studies have shown that the predictive accuracy of GS models can be enhanced by previously selecting a subset of predictor variables based on the individual effects of each marker on the study traits [38–40]. Likewise, Long et al. [41], Solberg et al. [42], Du et al. [43] and Azevedo et al. [44] demonstrated the potential to combine the dimension reduction and variable selection for accurate and cost-effective prediction of genomic breeding values. Among the methods proposed for these objectives, the principal component regression (PCR), partial least squares (PLS) and their extensions supervised PCR and sparse PLS stand out. The reduction in dimensionality becomes more relevant when considering the new genomic data platforms, which have allowed for high-density genomic data to be obtained, thereby increasing the possibility of finding genomic regions controlling the variation of a trait. Considering the large size of the SNP platforms currently available in *Eucalyptus* (e.g., EUchip60K), a preselection of SNP and/or the use of dimensionality reduction methods may improve the prediction ability of the traits of interest for the forestry industry.

The present study was undertaken to examine the ability of different GS models to predict growth traits (diameter at breast height, tree height and wood volume), stem straightness and branching quality of *E. globulus* trees using 60,000 SNP markers, in one of the southernmost progeny trials of the species, close to its southern distribution limit in Chile. A better understanding of how develop novel breeding strategies for tree improvement of *E. globulus* could represent new opportunities for forest planting at its southernmost range limit in Chile.

2. Materials and Methods

2.1. Genetic Material and Phenotypic Measurements

The breeding population consisted of a mixture of *E. globulus* Labill. families composed of 62 full-sib and 3 half-sib families (1968 individuals), established in 2012 in La Poza, municipality of Purranque, in the administrative region of Los Lagos; the southernmost distribution of *E. globulus* in Chile [7]. The local conditions of La Poza are shown in Table 1. A randomized complete block design was used in this experiment, with 30 blocks, single-tree plots, and a spacing of 2.5 m between the trees within a block. A total of 1968 trees were measured at four years of age for the following five phenotypic traits: total tree height (H), diameter at breast height (DBH), total wood volume (VOL), stem straightness (ST) and branching quality (BQ). The ST was evaluated in the first two-thirds of the total height of the tree using a categorical scale of 6 levels (1 represents trees with curvature in the first stretch of the total height of the tree; and 6 represents trees without problems that may show a slight curvature in the upper third of the tree without loss of productivity). BQ was evaluated according to different criteria that define the quality of branches (diameter, angle and distribution in the tree) using a categorical scale of 6 levels (1 represents trees with extreme deficiency in branch diameter and any other variable; and 6 represents trees that present an optimal combination of all the variables that qualify the quality of branches without loss of productivity).

162

Table 1. Site conditions of La Poza, municipality of Purranque, Chilean administrative region of Los Lagos.

Site Conditions	Metrics
Coordinates	40°57′ S, 73°30′ W
Climate Type	Oceanic or marine
Annual Temperature	13 °C
Average temperature in coldest months	6 °C
Average temperature in warmest months	16 °C
Annual accumulated rainfall	1282 mm
Altitude	326 masl

2.2. Genotyping and Estimation of Linkage Disequilibrium (LD)

Nuclear DNA was isolated from leaf tissue of 647 individuals randomly selected (approximately 10 individuals per family) from the breeding population according to Doyle and Doyle [45] and Porebsky et al. [46]. The sample was genotyped using the EUChip60K SNP system (GeneSeek, Lincoln, NE, USA) developed by Silva-Junior et al. [47]. The genotyping quality was evaluated using Genome Studio software (Illumina, San Diego, CA, USA). Subsequently, the genotyping matrix was filtered considering a minor allele frequency (MAF) ≥0.05 and a maximum proportion of 10% of lost data. The genotyped sample was used to estimate the LD pattern in the breeding population studied. The LD between marker pairs (expressed in terms of r^2) was corrected by relatedness and estimated using LDcorSV package version 1.3.2 [48]. The LD decay curve was plotted using R software version 3.4 according to the method of Hill and Weir [49], and it was based on the physical distance of the genome of *Eucalyptus grandis* [50].

2.3. Estimation of Pedigree-Based Breeding Values

Estimated breeding values (EBV) were obtained using the Best Linear Unbiased Prediction (BLUP) with estimates of variance components based on the Restricted Maximum Likelihood (REML) method by the ASREML program version 4 [51]. This pedigree-based model was as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{\varepsilon} \tag{1}$$

where $\mathbf{y}$ represents the vector of phenotypic data, $\mathbf{\beta}$ is the vector of fixed effects (general mean and block effect), $\mathbf{a}$ is the vector of additive effects, which $\sim N(0, \mathbf{A}\sigma_a^2)$ where $\mathbf{A}$ is the average numerator relationship matrix from pedigree information and σ_a^2 is the additive genetic variance, $\mathbf{X}$ and $\mathbf{Z}$ correspond to the incidence matrices associated with the fixed and random effects, respectively, and $\mathbf{\varepsilon}$ represents the vector of residual effects, which $\sim N(0, \mathbf{I}\sigma_e^2)$ where $\mathbf{I}$ is an identity matrix and σ_e^2 is the residual variance. Due to the ordinal nature of the traits BQ and ST, their estimated breeding values (EBVs) were obtained using a Generalized Linear Mixed Model, evaluated with a logistic regression model in the ASREML program v4 [51]. Additionally, coefficients of additive genetic variation (*CVa*) were calculated using variance component estimates and the means of each trait ($\overline{X}$) as:

$$CVa = \frac{\sqrt{\hat{\sigma}_a^2}}{\overline{X}} \tag{2}$$

2.4. Genomic Prediction Models

We compared the following statistical methods with regard to their predictive performance: Ridge Regression BLUP (RRBLUP), Bayes-A, Bayes-B, Bayesian Least Absolute Shrinkage and Selection Operator (BLASSO), PCR, supervised PCR and a variant of RRBLUP, which involves the previous selection of predictive variables (RRBLUP-B) [27,35,38,43]. The genotypic information is usually fitted

using a basic linear model that includes a vector β (the overall mean) and a vector **m** with the marker effects [38].

In RRBLUP and RRBLUP-B, a mixed model was fitted in which the vector of fixed effect (β), i.e., the overall mean, and the vector of random marker effects (**m**) were estimated simultaneously using the following mixed model equations:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \sigma_e^2/\sigma_m^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \tag{3}$$

where σ_m^2 and σ_e^2 correspond to the variance components (estimated by REML) of marker and residual effects, respectively, and the variance ratio σ_e^2/σ_m^2 corresponds to the shrinkage parameter for the random SNP marker effects. RRBLUP-B corresponds to a variant of the RRBLUP method [38], which utilizes a selected subset of markers. The number of markers in the subset was defined according to the criteria of Resende-Junior et al. [38]. In a first step, the marker effects from the RRBLUP were ranked in decreasing order by their absolute values and grouped in multiples of 50 markers (50, 100, 150 . . . , n). The group that maximized the predictive ability was selected as the optimum number of marker effects to be used in predictive model. Finally, markers effects were re-estimated in a new RRBLUP analysis, using selected markers (more details in Resende-Junior et al. [38]).

The methods Bayes A, Bayes B and BLASSO relax the assumption of common prior variance to all marker effects [30]. In the Bayes-A method [27], it is assumed that each ith marker effect ($\mathbf{m}_i$) follows a normal prior distribution $\sim N(0, \mathbf{I}\sigma_{m_i}^2)$. The prior distribution of marker variances is assumed to be scaled inverted chi-square, $\sigma_{m_i}^2 \sim \chi^{-2}(v, S^2)$, where S^2 and v are a scale parameter and the number of degrees of freedom, respectively. The Bayes-B method [27] uses a prior distribution that has a high density, π, at $\sigma_{m_i}^2 = 0$, and an inverted chi-square distribution for $\sigma_{m_i}^2 > 0$ with probability $1-\pi$. The BLASSO method, a double exponential prior distribution (DE) is assumed for marker effects, $p(m_i|\lambda, \sigma_e^2) = DE(m_i|0, \lambda/\sigma_e^2)$, where λ corresponds to a regularization parameter. The DE distribution generates a strong contraction (closer to zero) to estimate the effect of the markers [52].

PCR is a statistical method that addresses the problem of multicollinearity when there are a large number of explanatory variables in a prediction model [53]. Importantly, dimension reduction techniques have recently gained much attention in the analysis of high dimensional genomic data [54–56]. This method consists of finding combinations of the predictor variables that best represent the variability of the data (main components). In the first step, the components that mainly explain the variation of the response variable (e.g., deregressed EBV) are identified. Subsequently, the model is validated by reducing the matrix dimension of the predictor variables in such a way that only relevant components are incorporated in the prediction model. Considering that matrix **X** (matrix of markers or predictor variables) is centered and scaled (**X***) with n $\times$ m dimensions (n is the sample size and m is the number of markers), its singular decomposition value is as follows [54]:

$$\mathbf{X}^* = \mathbf{UDV^T} \tag{4}$$

where **U**, **D** and **V^T** are matrices of $n \times m$, $m \times m$ and $m \times p$, respectively, n is the number of observations ($n = 647$), p is the total number of predictors (SNPs); m is the number of X "ranked" elements; **D** is a diagonal matrix of singular values d_j; and U contains the main components (u$_1$, u$_2$ . . . u$_m$) in the order of $d_1 \geq d_2 \geq \ldots \geq 0$. The PCR model corresponds to a regression expressed as follows:

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}^* \beta^* + \varepsilon \tag{5}$$

where β^* corresponds to the regression coefficients that allow selecting the main components that explain the variation of **y**. The matrix $\mathbf{V}_{m \times m}$ is also known as the load matrix of **X**. When the number of components (p) is defined, the load matrix is truncated to a matrix of $m \times p$. Therefore, the matrix

containing the selected components is represented by $\mathbf{T}_{nxp} = \mathbf{X}_{nxm}\mathbf{V}_{mxp}$. The regression model using these new terms is $\mathbf{y} = \mathbf{Tb} + \mathbf{e}$. Therefore, the estimated score coefficients can be expressed as follows:

$$\hat{\mathbf{b}} = (\mathbf{T}^{\mathbf{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathbf{T}}\mathbf{y} \tag{6}$$

(further details in Du et al. [49]). The supervised PCR method consists of two steps [44]. In the first step, the individual coefficients of each predictor variable (effects of the SNPs) are obtained to "rank" each SNP according to the magnitude of its effects. Subsequently, a set of variables ranked in the first positions are selected to perform a PCR analysis. The variable selection in supervised PCR model was based upon association with the phenotype of each SNP, using a single-marker regression (Long et al. [41]). A value of significance $p < 0.01$ for the regression between SNP and phenotypic response was considered as threshold for SNP selection in the supervised PCR method.

The optimal number of components for the PCR and supervised PCR methods was determined by selecting the minimum prediction error value (predicted residual error sum of squares—PRESS) for each component. These methods were implemented using the PLS package in R [57]. The RRBLUP, Bayes-A, Bayes-B and BLASSO methods were implemented in R software using the RRBLUP version 4.6 [58] and BLR version 1.5 [59] packages.

2.5. Cross-Validation and Prediction Ability

The sample of 647 trees was divided in two subsets of individuals. A total of 582 individuals (~90%) were used to train the models and estimate the effects of the markers, while the rest of the individuals (~10%) were used to evaluate and validate the prediction values. In the case of the RRBLUP, BAYES-A, BAYES-B, BLASSO and RRBLUP-B methods, 100 cycles were used to obtain the average prediction ability of each method. The PCR and supervised PCR methods were validated according to Du et al. [43] using 10 folds in the cross-validation analysis. The prediction ability of all the methods was calculated as the correlation coefficient between genomic estimated breeding values (GEBVs) and EBVs of the validation subsample.

2.6. Validation of Pedigree Data

In an additional analysis, a genomic relationship matrix was calculated to evaluate the consistency between the relatedness coefficient based on pedigree information and genomic information based on SNPs data. The genomic relationship matrix was obtained in R software using the RRBLUP version 4.6 [58] package.

3. Results

3.1. Estimates of Variance Components and Heritability of Growth Traits, Branching Quality and Stem Straightness

The estimates of the variance components (REML) and heritability for branching quality (BQ), stem straightness (ST), total tree height (H), diameter at breast height (DBH) and wood volume (VOL) are presented in Table 2 (BLUP analysis). According to these results, the study traits showed low-to-moderate heritability, with estimates varying from 0.07 to 0.19 for BQ and DBH. The heritability values of all traits were consistent with other breeding populations of *E. globulus* [60–63]. Furthermore, these findings were supported by the coefficients of additive genetic variation, which were similar to those found in other studies with *E. globulus* and *Eucalyptus* spp. [63–65]. The results evidenced that growth traits such (H, DBH, VOL) and stem form traits (ST and BQ) had a low additive genetic control in early stages of *E. globulus*, which can be usual for populations younger than 5 years [61,62]. In this context, the low heritability found is supporting the use of genomic prediction methods that have the potential to capture a greater proportion of the phenotypic variation during selection cycles.

Table 2. Restricted Maximum Likelihood (REML) estimates of variance components and narrow-sense heritability (standard error in parenthesis) for stem straightness (ST), branching quality (BQ), total tree height (H), diameter at breast height (DBH) and stem volume (VOL), evaluated in full- and half-sib families of *Eucalyptus globulus* Labill. REML: Restricted Maximum Likelihood, CVa: coefficients of additive genetic variation.

REML Estimates	ST	BQ	H	DBH	VOL
Additive variance	0.194	0.073	0.107	0.693	0.00001
Residual variance	1.000	1.000	0.777	3.01	0.00006
Heritability	0.162 (0.03)	0.068 (0.02)	0.121 (0.04)	0.187 (0.06)	0.156 (0.05)
CVa	17.6 *	11.0 *	5.03	8.5	18.7

* Corrected according to Burdon [66].

In this study, the consistence between the frequency distribution of coefficients of genomic and pedigree-based relationship data was additionally evaluated. As expected, the analysis of genomic data revealed that the most of individuals are not related. This result is in agreement with the analysis of pedigree data (Figure 1). However, both relationship matrix (based on genomic and pedigree), showed that several individuals are related with others, and the coefficients were between 0.2–0.3, which reflects the mating design implemented in this breeding population of *E. globulus*.
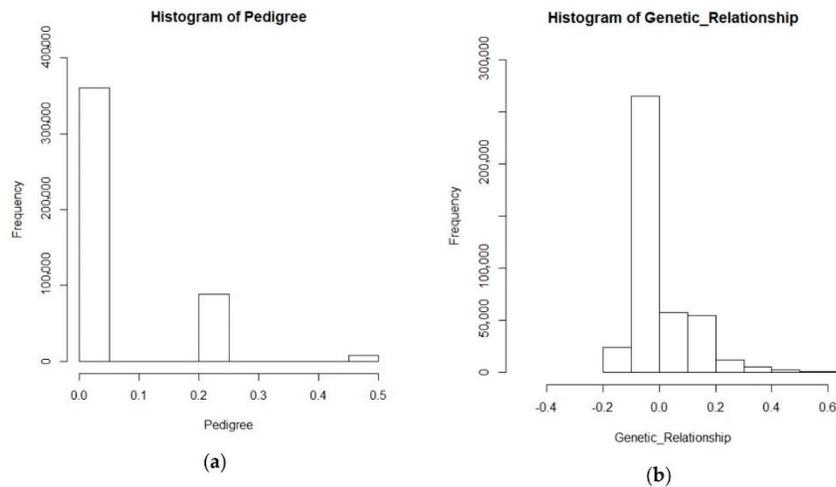

Figure 1. Distribution of pedigree (**a**) and genetic relationship values (**b**) for the breeding population studied.

3.2. Final Set of Qualified SNPs and Linkage Disequilibrium Decay

After applying the corresponding filters, the final set of markers consisted of 14,442 high-quality SNPs, which were distributed on the 11 chromosomes of *E. globulus* with an average of 1356 SNPs per chromosome and an average distance of approximately 4000 bp. The genome-wide linkage disequilibrium (LD) decay pattern for the study population is shown in Figure 2. Decay of LD showed a clear nonlinear trend with physical distance. According to these results, the genome-wide LD decayed rapidly within 0.76 Mbp, regarding a threshold value of $r^2 = 0.1$ ($p < 0.05$). The average LD on all chromosomes was $r^2 = 0.09$. In addition, the 0.15% of total pairs of linked SNPs were in a complete LD ($r^2 = 1$), and the 3% had an r^2 value >0.5. Particularly, the LD of chromosome 2 decayed faster than the other linkage groups, and chromosome 11 had the greatest extension of LD, decaying at 1 Mbp (r^2

= 0.11) (Table 3). Approximately 3.8% of the marker pairs of each chromosome registered a value of LD (r^2) > 0.5. Particularly, 5.6% of the marker pairs of chromosome 11 were found in strong LD (r^2 > 0.5).
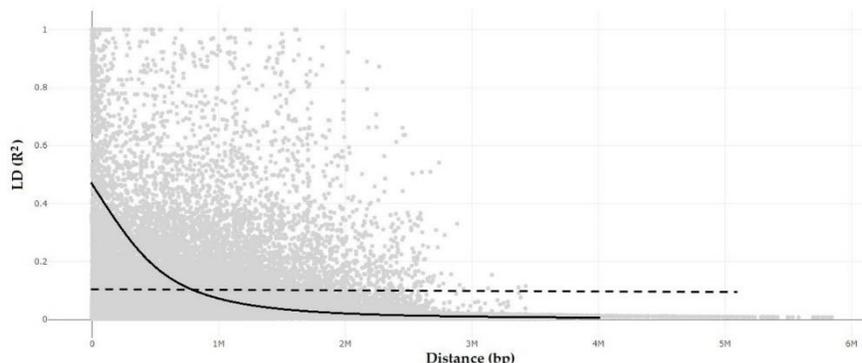

Figure 2. Genome-wide linkage disequilibrium (LD) decay pattern for the study population of *Eucalyptus globulus*. Labill. Pairwise LD values expressed by r^2; LD (R^2), as adjusted according to Hill and Weir [53] (curve colored by black). Distance (bp) is the distance between SNP pairs. The LD threshold of r^2 = 0.1 is indicated with a dotted line.

Table 3. LD estimates (corrected by the relatedness) for each chromosome. Distance (Mbp) and r^2 express the average values of genomic distance and LD between pairs of markers. Max. r^2 and Min. r^2 correspond to the maximum and minimum LD values, respectively, found between each pair of markers. Dist. Max and Dist. Min represent the maximum and minimum distance between markers found in LD, respectively. CH is the number of chromosome of *Eucalyptus*.

CH	r^2	Max. r^2	Min. r^2	Distance (Mbp)	Dist. Max (Mbp)	Dist. Min (Mbp)
1	0.03	0.37	0.0057	0.96	3.65	0.00003
2	0.03	0.37	0.0057	0.78	3.19	0.00003
3	0.09	0.36	0.0057	1.11	5.85	0.00003
4	0.09	0.37	0.0057	0.99	5.31	0.000031
5	0.09	0.36	0.0057	1.04	5.5	0.00003
6	0.09	0.36	0.0137	0.81	3.19	0.000039
7	0.09	0.37	0.0137	0.97	4.59	0.000033
8	0.09	0.36	0.0137	0.9	3.85	0.000036
9	0.09	0.37	0.0137	0.93	3.92	0.00003
10	0.09	0.37	0.0137	0.81	3.4	0.000049
11	0.11	0.37	0.0217	0.83	3.43	0.000038

3.3. Predictive Ability of Frequentist, Bayesian and Dimension Reduction Methods

In Table 4 and Figure 3, the predictive ability (PA) values of all the methods under study and each evaluated trait are shown. As expected, PA varied between the seven prediction methods explored here, as well as between the traits under study [43]. The highest PA values were obtained by means of methods of dimension reduction and variable selection (RRBLUP-B and supervised PCR). In the case of RRBLUP-B, PA was increased in all traits by reducing between 94 and 97% the total number of SNP markers (Figure 4). The highest predictive ability values for DBH, VOL, H, ST and BQ were obtained using models based on 450, 900, 850, 800 and 950 SNPs, respectively. In particular, the maximum PA values reached by the RRBLUP, Bayes-A, Bayes-B and BLASSO methods were significantly lower than the rest of the methods, with maximum values corresponding to 0.14 (ST), 0.17 (BQ and ST), 0.24 (BQ) and 0.21 (BQ), respectively. According to the results, more parsimonious models provided a higher prediction of growth and stem form than other predictive models in this population of *E. globulus*. On the other hand, the RRBLUP and BAYES-A methods performed a greater predictive ability than the

167

BAYES-B and BLASSO methods for growth traits (VOL, DHB and H), and vice versa for stem form traits (BQ and ST).

Table 4. Predictive ability of seven prediction methods (with cross-validation) for branching quality (BQ), stem straightness (ST), wood volume (VOL), diameter at breast height (DBH) and tree height (H) in *Eucalyptus globulus* families, evaluated under oceanic climate conditions. NPV corresponds to number of predictor variables, SNPs or Components (PCR and Supervised PCR (S PCR)) used.

| Trait/NPV	RRBLUP	RRBLUP-B	BAYES-A	BAYES-B	BLASSO	PCR	S PCR
BQ	0.17	0.68	0.21	0.28	0.23	0.1	0.69
NPV	14,422	950	14,422	14,422	14,422	573	71
ST	0.14	0.59	0.17	0.14	0.1	0.16	0.62
NPV	14,422	800	14,422	14,422	14,422	579	95
VOL	0.13	0.42	0.07	0.04	0.04	0.35	0.35
NPV	14,422	900	14,422	14,422	14,422	575	148
DBH	0.04	0.43	0.02	0.01	0.01	0.35	0.43
NPV	14,422	450	14,422	14,422	14,422	579	62
H	0.04	0.5	0.05	0.03	0.04	0.21	0.54
NPV	14,422	850	14,422	14,422	14,422	570	338
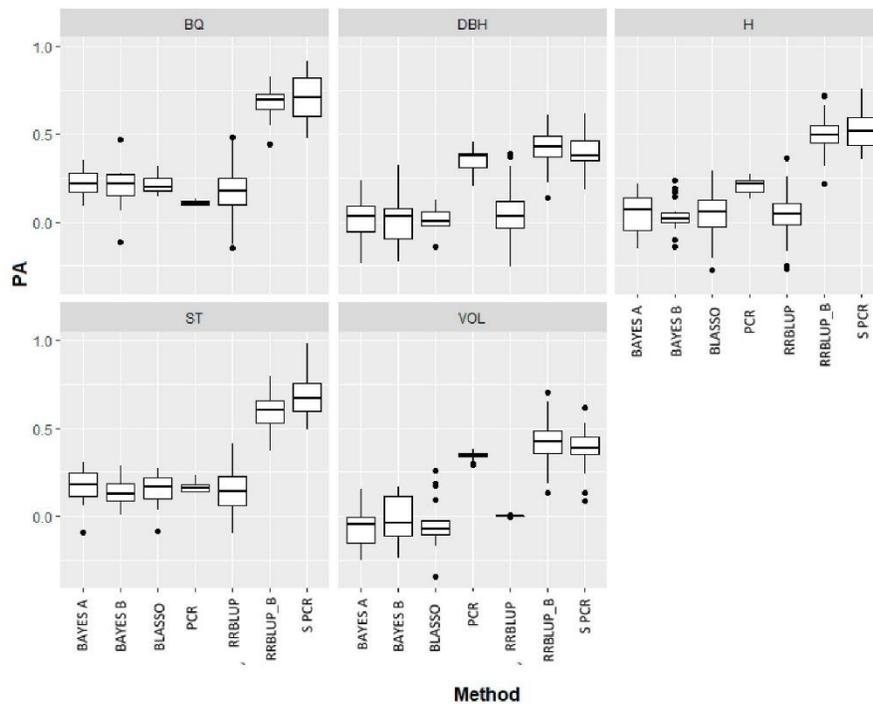
Figure 3. Box plot of adjusted means for the predictive ability (PA) of seven GS methods for branching quality (BQ), stem straightness (ST), wood volume (VOL), diameter at breast height (DBH) and tree height (H).
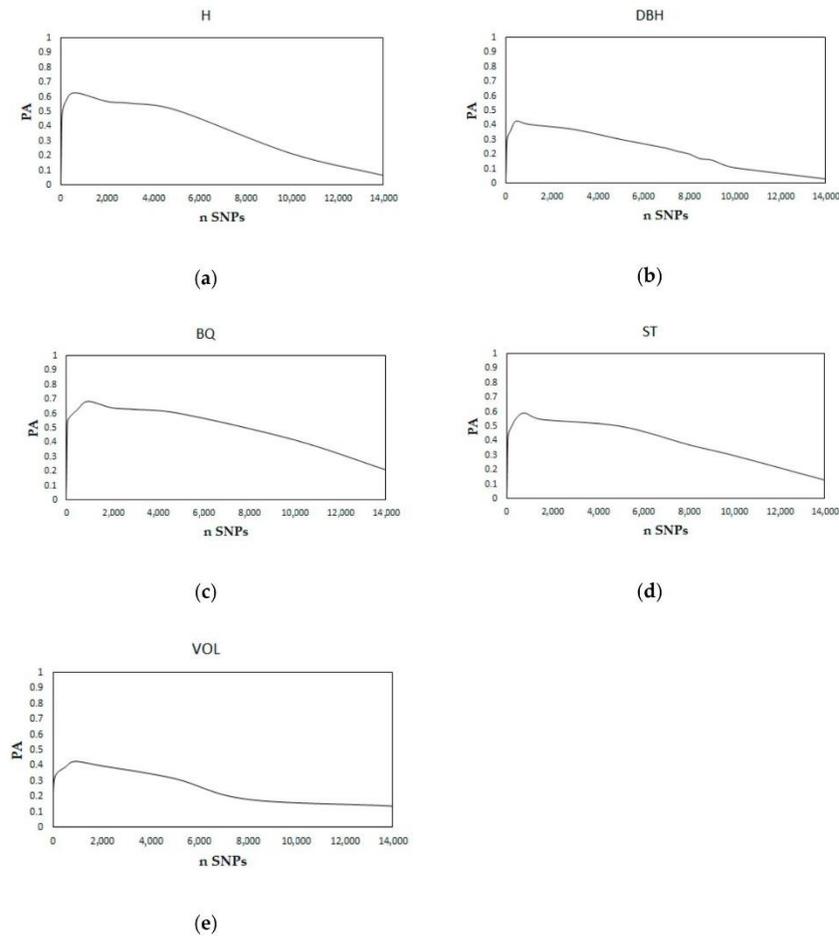
Figure 4. Predictive ability (PA) of the RRBLUP-B method (without cross-validation), where n SNPs corresponds to a subset of markers for (**a**) tree height (H), (**b**) diameter at breast height (DBH), (**c**) branching quality (BQ), (**d**) stem straightness (ST) and (**e**) stem volume (VOL).

The importance of each SNP on the studied traits was evaluated according to their effects and individual regression coefficients (Figure 5). The 950, 800, 900, 450 and 850 SNPs that were used to predict the traits under study through the RRBLUP-B model were located mainly in chromosomes 2 (BQ), 2 (ST), 6 (VOL), 11 (DBH), and 8 (H), respectively. The variability of SNPs effects was higher for BQ and ST, and lower for H and DBH. The SNP effects for H and DBH yielded a more homogeneous distribution than the other studied traits, while some SNPs generated an effect at least 75% above the mean of SNP effects for BQ and ST.

169
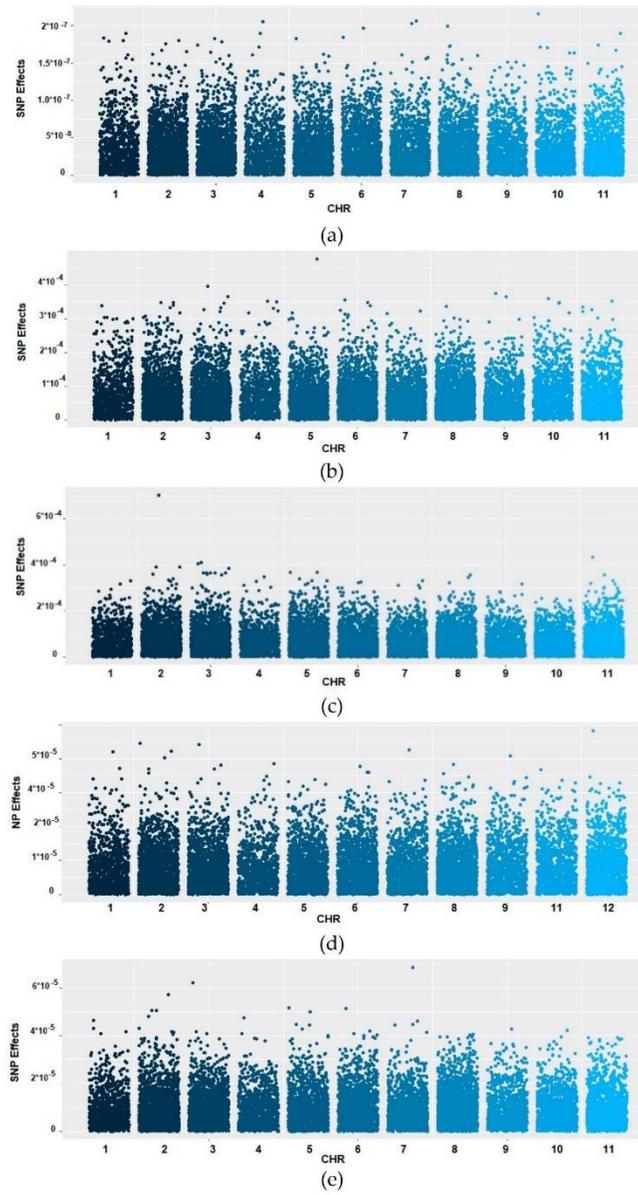
Figure 5. Absolute values of the SNP effects (*n* = 14,442) estimated by RRBLUP (without cross-validation) for (**a**) wood volume, (**b**) stem straightness, (**c**) branching quality, (**d**) diameter at breast height, and (**e**) tree height.

4. Discussion

The study population presented a more extended LD than other populations of *E. globulus* previously studied. For example, Durán et al. [37] found that LD decays within 3000 bp in a clonal population of the species. In contrast, Thavamanikumar et al. [67] found that the LD decays rapidly within 1000 bp in natural populations of *E. globulus*. The high LD found in the present study may be explained by the fact that most of the families included in the studied population (62/65) were derived from controlled crosses (from an incomplete factorial mating design). The genomic relationship matrix showed that several individuals were tightly related with others; as revealed by the values of relatedness coefficients (estimated between 0.2–0.3), which reflect the mating design implemented in this population of *E. globulus*. The LD pattern is one of the main factors that contribute to the success of breeding programs that involve GS [27,68,69]. The greater extent of LD in a population results in fewer markers that are needed to find QTLs associated with phenotypic traits, and a more accurate genomic prediction can be obtained. According to Liu et al. [70], populations that present extensive LD in long genomic distances allow for predictions with higher stability. In this context, genomic prediction models may benefit from the genetic structure of the studied population in such a way that the high LD values in their chromosomes could contribute in obtaining models with a relatively high predictive ability.

The results of the present study revealed that the reduction in dimensionality and selection of variables could contribute to obtain predictive models with greater predictive accuracy. Several studies have shown that a subset of selected variables can reasonably increase PA and the reliability of BV estimates predicted by GS [40,71,72]. For example, Long et al. [71], Usai et al. [72] and Weigel et al. [73] found that the maximum accuracy value of a predictive model was obtained by using only 1% to 3% of the total variables. In agreement with present results, Liu et al. [70] reported that closer genetic relationships between the training and validation populations as well as relatively high LD result in the need for fewer markers to achieve greater predictive accuracy. Moreover, the genetic architecture that underlies the study trait also determines the most appropriate method for predicting BV [74,75]. Growth-related traits, such as H, DBH and VOL, have been described as less heritable ($h^2 < 0.3$; in narrow-sense) than other complex traits in *E. globulus* (e.g., wood quality and chemical properties; $h^2 >$ 0.5 [76,77]), which explain the small number of genomic regions or QTLs detected for these traits, in contrast to other traits.

The predictive ability for growth-related traits was maximized by dimension reduction and variable selection methods, while all tested Bayesian methods performed similarly, with the lowest PA values for these traits. Bayes-B and BLASSO are also considered to be methods of variable selection, in which the predictor variables with an effect value closer to zero are removed from the prediction model [59]. Unexpectedly, the Bayes-A method was slightly superior to the previous methods, suggesting that a large number of markers have an effect on the study trait (similar to RRBLUP). In general, the BLASSO and Bayes-B methods have been suggested for the prediction of traits that are controlled by a small number of QTLs and relatively large effects [78,79], while Bayes-A method exploits the prior knowledge that many SNPs have small individual effects on the trait [80]. For the present study, the BLASSO and Bayes-B methods would not be adequate to predict breeding values of DBH, H and VOL because these traits are better represented by an infinitesimal model where the effects are relatively small and homogeneous. On the other hand, the predictive ability of DBH, VOL and H obtained by the PCR method was superior to the Bayesian methods. The PCR method develops the components and/or latent variables identifying linear combinations that present a correlation degree [41], which would explain the response variable considerably. Therefore, these methods can deal with the multicollinearity effects of the predictor variables. Based on the LD pattern results of the study population, approximately 3% of the predictor variables were highly correlated, which may indicate that the main components consisted of SNPs found in LD because their frequencies in the population would be correlated. Notably, the SNPs that play a relatively important role for growth-related traits (RRBLUP-B) were located on all 11 chromosomes, which was consistent with

171

the fact that QTLs detected for growth traits, such as DBH and H of *Eucalyptus*, have previously been identified in different linkage groups [81,82]. For instance, Bundock et al. [83], Thumma et al. [82] and Gion et al. [84] identified QTLs for DBH and H on chromosomes 2, 5, 6, 8 and 11 in *Eucalyptus* trees of 4 and 5 years of age, which is consistent with the findings of the present study. In addition, when we ranked the top ten SNPs explaining the phenotypic variation of each trait, three SNPs performed a high effect on DBH and VOL, which is consistent with previous studies of QTL for growth traits in *Eucalyptus* (e.g., Arriagada et al. [13]). These background and results confirm that the methods involving dimensionality reduction and variable selection (RRBLUP-B, supervised PCR) perform a better predictive ability than the PCR method.

The PA for BQ and ST traits was significantly increased with the use of predictive methods based on dimension reduction and variable selection (RRBLUP-B, supervised PCR) as well as Bayes-B and BLASSO. Bayes-B and BLASSO methods assume that the effects of some SNPs on phenotypic variation tend to be zero; therefore, these SNPs could be removed from regression model [27]. The effects of the markers (expressed in absolute value) of BQ and ST were highly variable, indicating that the variation of these traits may be explained by a combination of QTLs of greater effects and others of lesser effects. Further, the BLASSO and Bayes-B methods were slightly superior to RRBLUP and Bayes-A, supporting the previous hypothesis of that some SNPs with an effect value closer to zero are not important to explain the phenotypic variation of these traits, and they can be removed from the prediction model. To date, no study on *Eucalyptus* species has reported QTLs associated with BQ. However, in rice [85] and, more recently, in peach [86], the angle of branches and spikes has been found to be controlled by a QTL of greater effect known as *TAC1*, which provides an estimate of the possible oligogenic inheritance of BQ. Regarding ST, a limited number of QTL studies associated with this trait have been reported for *Eucalyptus* [13,14]. In agreement with the present findings, Arriagada et al. [13] in a breeding population of *Eucalyptus cladocalyx*, reported microsatellite markers associated with ST, which were located on chromosome 11.

The results of the genomic prediction of the present study were comparable to previous reports in *Eucalyptus* species. For instance, in hybrids of *E. gradis* Hill ex Maiden, *E. urophylla* S.T. Blake and *E. globulus* of 3–4 years of age, Resende et al. [30] obtained PA values higher than 0.5 in growth traits using the RRBLUP method. However, the present results involved a smaller number of predictor variables to obtain maximum accuracy. In other *Eucalyptus* species, Müller et al. [87] obtained a PA for growth-related traits (DBH and VOL) lower than 0.5 in open pollinated families of *E. benthamii* Maiden & Cambage and *E. pellita* F. Muell. Regarding other traits relevant to wood production, such as BQ and ST, Isik et al. [88] in another species of interest for the forestry industry (*Pinus pinaster* Aiton), reported a PA for stem sweep over 0.55, and they obtained the maximum predictive accuracy by non-Bayesian methods.

According to the present results, the methods based on dimensionality reduction in the matrix of predictor variables (i.e., PCR) performed better than traditional prediction methods, such as RRBLUP and Bayesian methods. An advantage of PCR is that no assumptions of prior distribution of marker affects are made [42]. This quality allows PCR to have broad applications in the prediction of phenotypic traits with different genetic architectures. Furthermore, supervised methods tend to be computationally more efficient than unsupervised methods [44] and allow the removal of variables correlated with other predictor variables, but they are irrelevant to explain the variation of the variable response. Prediction models based on variable reduction have been widely used in animal genomic prediction, but their use in plants is scarce [38,89]. Therefore, taking advantage of prediction models that involve dimensionality reduction and variable selection is beneficial.

The results of this study provide new knowledge to assist breeding programs of *E. globulus* based on growth and wood form traits. We evaluated a progeny trial established under unusual conditions for *E. globulus* to generate accurate genomic prediction models. In fact, the southern limit of *E. globulus* in Chile occurs close to the study location, at the central part of the Chilean administrative region of Los Lagos; a region that account for only 4% of the total *E. globulus* plantations in Chile. In the

global context, several regions worldwide will be affected by the climate change, and therefore, it is expected a significant reduction of temperate forests [90,91]. The diversification of plantation areas could represent new opportunities for forest planting that can benefit the local economy.

5. Conclusions

The present GS analysis performed on 4-year-old trees of *Eucalyptus globulus* found that the dimensionality reduction and variable selection showed a strong impact on the predictive ability. In this sense, the RRBLUP-B and supervised PCR methods provided the highest PA values for all the traits studied. On the other hand, we confirmed that the prediction method depends on the genetic architecture of the trait under study, in which the BLASSO and Bayes-B methods performed better for BQ and ST, while RRBLUP and Bayes-A were more suitable for growth traits. RRBLUP-B as well as supervised PCR allowed increases in PA regardless of the genetic nature of the traits.

The present results are promising in terms of early genetic improvement of *E. globulus* in unusual regions for the development of the species because the present study was conducted in one of the southernmost sites for the species. Finally, the genomic prediction approach based on dimensionality reduction and variable selection has been scarcely explored in woody species, such as *Eucalyptus*, and it was shown to be a promising method, considering the early growth and the low heritability values.

References

1. Drake, J.E.; Aspinwall, M.J.; Pfautsch, S.; Rymer, P.D.; Reich, P.B.; Smith, R.A.; Crous, K.Y.; Tissue, D.T.; Ghannoum, O.; Tjoelker, M.G. The capacity to cope with climate warming declines from temperate to tropical latitudes in two widely distributed *Eucalyptus* species. *Glob. Chang. Biol.* **2015**, *21*, 459–472. [CrossRef] [PubMed]
2. Mora, F.; Arriagada, O.; Ballesta, P.; Ruiz, E. Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *J. Plant Biochem. Biotechnol.* **2017**, *26*, 274–281. [CrossRef]
3. Paiva, J.A.; Prat, E.; Vautrin, S.; Santos, M.D.; San Clemente, H.; Brommonschenkel, S.; Fonseca, P.G.S.; Grattapaglia, D.; Song, X.; Ammiraju, J.S.S.; et al. Advancing *Eucalyptus* genomics: Identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC Genom.* **2011**, *12*, 137. [CrossRef] [PubMed]
4. Foster, S.A.; McKinnon, G.E.; Steane, D.A.; Potts, B.M. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytol.* **2007**, *175*, 370–380. [CrossRef] [PubMed]
5. Dutkowski, G.W.; Potts, B.M. Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Aust. J. Bot.* **1999**, *47*, 237–263. [CrossRef]
6. Tibbits, W.N.; White, T.L.; Hodge, G.R.; Borralho, N.M. Genetic variation in frost resistance of *Eucalyptus globulus* ssp. *globulus* assessed by artificial freezing in winter. *Aust. J. Bot.* **2006**, *54*, 521–529. [CrossRef]
7. Lanfranco, D.; Dungey, H.S. Insect damage in Eucalyptus: A review of plantations in Chile. *Aust. Ecol.* **2001**, *26*, 477–481. [CrossRef]
8. Navarrete-Campos, D.; Bravo, L.A.; Rubilar, R.A.; Emhart, V.; Sanhueza, R. Drought effects on water use efficiency, freezing tolerance and survival of *Eucalyptus globulus* and *Eucalyptus globulus* × *nitens* cuttings. *New For.* **2013**, *44*, 119–134. [CrossRef]
9. Fernández, M.; Valenzuela, S.A.; Arora, R.; Chen, K. Isolation and characterization of three cold acclimation-responsive dehydrin genes from *Eucalyptus globulus*. *Tree Genet. Genom.* **2012**, *8*, 149–162. [CrossRef]

10. Castillo, R.; Otto, M.; Freer, J.; Valenzuela, S. Multivariate strategies for classification of *Eucalyptus globulus* genotypes using carbohydrates content and NIR spectra for evaluation of their cold resistance. *J. Chem. Soc.* **2008**, *22*, 268–280.

11. Tambarussi, E.V.; Pereira, F.B.; Da Silva, P.H.M.; Lee, D.; Bush, D. Are tree breeders properly predicting genetic gain? A case study involving *Corymbia* species. *Euphytica* **2018**, *214*, 150. [CrossRef]

12. Grattapaglia, D.; Resende, M.D. Genomic selection in forest tree breeding. *Tree Genet. Genom.* **2011**, *7*, 241–255. [CrossRef]

13. Arriagada, O.; Mora, F.; Amaral Junior, A.T. Thirteen years under arid conditions: Exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breed. Sci.* **2018**, *68*, 367–374. [CrossRef] [PubMed]

14. Ballesta, P.; Mora, F.; Ruiz, E.; Contreras-Soto, R. Marker-trait associations for survival, growth, and flowering components in *Eucalyptus cladocalyx* under arid conditions. *Biol. Plant.* **2015**, *59*, 389–392. [CrossRef]

15. Bartholomé, J.; Salmon, F.; Vigneron, P.; Bouvet, J.M.; Plomion, C.; Gion, J.M. Plasticity of primary and secondary growth dynamics in *Eucalyptus* hybrids: A quantitative genetics and QTL mapping perspective. *BMC Plant Biol.* **2013**, *13*, 120–133. [CrossRef] [PubMed]

16. Cappa, E.P.; El-Kassaby, Y.A.; Garcia, M.N.; Acuña, C.; Borralho, N.M.G.; Grattapaglia, D.; Marcucci-Poltri, S.N. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. *PLoS ONE* **2013**, *8*, e81267. [CrossRef] [PubMed]

17. Carocha, V.; Soler, M.; Hefer, C.; Cassan-Wang, H.; Fevereiro, P.; Myburg, A.A.; Paiva, J.A.; Grima-Pettenati, J. Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol.* **2015**, *206*, 1297–1313. [CrossRef] [PubMed]

18. Isik, F. Genomic selection in forest tree breeding: The concept and an outlook to the future. *New For.* **2014**, *45*, 379–401. [CrossRef]

19. Beaulieu, J.; Doerksen, T.; Clément, S.; MacKay, J.; Bousquet, J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* **2014**, *113*, 343. [CrossRef] [PubMed]

20. Heffner, E.L.; Lorenz, A.J.; Jannink, J.L.; Sorrells, M.E. Plant breeding with genomic selection: Gain per unit time and cost. *Crop Sci.* **2010**, *50*, 1681–1690. [CrossRef]

21. Neale, D.B.; Kremer, A. Forest tree genomics: Growing resources and applications. *Nat. Rev. Genet.* **2011**, *12*, 111. [CrossRef] [PubMed]

22. Beaulieu, J.; Doerksen, T.K.; MacKay, J.; Rainville, A.; Bousquet, J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genom.* **2014**, *15*, 1048. [CrossRef] [PubMed]

23. Ratcliffe, B.; El-Dien, O.G.; Klápště, J.; Porth, I.; Chen, C.; Jaquish, B.; El-Kassaby, Y.A. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii× glauca*) using unordered SNP imputation methods. *Heredity* **2015**, *115*, 547. [CrossRef] [PubMed]

24. Ukrainetz, N.K.; Ritland, K.; Mansfield, S.D. Identification of quantitative trait loci for wood quality and growth across eight full-sib coastal Douglas-fir families. *Tree Genet. Genom.* **2008**, *4*, 159–170. [CrossRef]

25. Mamani, E.M.; Bueno, N.W.; Faria, D.A.; Guimarães, L.M.; Lau, D.; Alfenas, A.C. Positioning of the major locus for *Puccinia psidii* rust resistance (Ppr1) on the *Eucalyptus* reference map and its validation across unrelated pedigrees. *Tree Genet. Genom.* **2010**, *6*, 953–962. [CrossRef]

26. Heffner, E.L.; Jannink, J.L.; Sorrells, M.E. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genom.* **2011**, *4*, 65–75. [CrossRef]

27. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [PubMed]

28. De Los Campos, G.; Sorensen, D.; Gianola, D. Genomic Heritability: What Is It? *PLoS Genet.* **2015**, *11*, e1005048. [CrossRef] [PubMed]

29. Daetwyler, H.D.; Calus, M.P.L.; Pong-Wong, R.; De Los Campos, G.; Hickey, J.M. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. *Genetics* **2013**, *193*, 347–365. [CrossRef] [PubMed]

30. Resende, M.D.; Resende, M.F.; Sansaloni, C.P.; Petroli, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; et al. Genomic selection for growth and wood quality in *Eucalyptus*: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* **2012**, *194*, 116–128. [CrossRef] [PubMed]

31. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dungey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2018**. [CrossRef]

32. Grattapaglia, D. Breeding forest trees by genomic selection: Current progress and the way forward. *Genom. Plant Genet. Resour.* **2014**, *1*, 651–682.

33. Iwata, H.; Hayashi, T.; Tsumura, Y. Prospects for genomic selection in conifer breeding: A simulation study of *Cryptomeria japonica*. *Tree Genet. Genom.* **2011**, *7*, 747–758. [CrossRef]

34. Zhong, S.; Dekkers, J.C.; Fernando, R.L.; Jannink, J.L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. *Genetics* **2009**, *182*, 355–364. [CrossRef] [PubMed]

35. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **1996**, *58*, 267–288. [CrossRef]

36. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef] [PubMed]

37. Durán, R.; Isik, F.; Zapata-Valenzuela, J.; Balocchi, C.; Valenzuela, S. Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genet. Genom.* **2007**, *13*, 74. [CrossRef]

38. Resende-Junior, M.F.R.; Muñoz, P.; Resende, M.D.V.; Garrick, D.J.; Fernando, R.L.; Davis, J.M.; Jokela, E.J.; Martin, T.A.; Peter, G.F.; Kirst, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **2012**, *190*, 1503–1510. [CrossRef] [PubMed]

39. Macciotta, N.P.; Gaspa, G.; Steri, R.; Pieramati, C.; Carnier, P.; Dimauro, C. Pre-selection of most significant SNPS for the estimation of genomic breeding values. *BMC Proc.* **2009**, *3*, 14. [CrossRef]

40. Arojju, S.K.; Conaghan, P.; Barth, S.; Milbourne, D.; Casler, M.D.; Hodkinson, T.R.; Michel, T.; Byrne, S.L. Genomic prediction of crown rust resistance in *Lolium perenne*. *BMC Genet.* **2018**, *19*, 35. [CrossRef] [PubMed]

41. Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A. Dimension reduction and variable selection for genomic selection: Application to predicting milk yield in Holsteins. *J. Anim. Breed. Genet.* **2011**, *128*, 247–257. [CrossRef] [PubMed]

42. Solberg, T.R.; Sonesson, A.K.; Woolliams, J.A.; Meuwissen, T.H. Reducing dimensionality for prediction of genome-wide breeding values. *Genet. Sel. Evol.* **2009**, *41*, 29. [CrossRef] [PubMed]

43. Du, C.; Wei, J.; Wang, S.; Jia, Z. Genomic selection using principal component regression. *Heredity* **2018**, *121*, 12–23. [CrossRef] [PubMed]

44. Azevedo, C.F.; Silva, F.; Resende, M.D.; Lopes, M.S.; Duijvesteijn, S.E.; Lopes, P.S.; Kelly, M.J.; Viana, J.M.; Knol, E.F. Supervised independent component analysis as an alternative method for genomic selection in pigs. *J. Anim. Breed. Genet.* **2014**, *131*, 452–461. [CrossRef] [PubMed]

45. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.

46. Porebski, S.; Bailey, L.G.; Baum, B.R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **1997**, *15*, 8–15. [CrossRef]

47. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D.A. Flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* **2015**, *206*, 1527–1540. [CrossRef] [PubMed]

48. Mangin, B.; Siberchicot, A.; Nicolas, S.; Doligez, A.; This, P.; Cierco-Ayrolles, C. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* **2012**, *108*, 285. [CrossRef] [PubMed]

49. Hill, W.G.; Weir, B.S. Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* **1988**, *33*, 54–78. [CrossRef]

50. Myburg, A.A.; Grattapaglia, D.; Tuskan, G.A.; Hellsten, U.; Hayes, R.D.; Grimwood, J.; Jenkins, J.; Lindquist, E.; Tice, H.; Bauer, D.; et al. The genome of *Eucalyptus grandis*. *Nature* **2014**, *510*, 356. [CrossRef] [PubMed]

51. Gilmour, A.R.; Gogel, B.J.; Cullis, B.R.; Welham, S. ASReml User Guide Release 4.1 Structural Specification. 2015. Available online: https://www.vsni.co.uk/downloads/asreml/release4/UserGuideStructural.pdf (accessed on 13 April 2018).

52. Wang, X.; Yang, Z.; Xu, C. A comparison of genomic selection methods for breeding value prediction. *Sci. Bull.* **2015**, *60*, 925–935. [CrossRef]

53. Pant, S.; Schenkel, F.S.; Verschoor, C.P.; You, Q. A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle. *Genomics* **2010**, *95*, 176–182. [CrossRef] [PubMed]

54. Bair, E.; Hastie, T.; Paul, D.; Tibshirani, R. Prediction by supervised principal components. *J. Am. Stat. Assoc.* **2006**, *101*, 119–137. [CrossRef]

55. Chun, H.; Keleş, S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. B* **2010**, *72*, 3–25. [CrossRef] [PubMed]

56. Lê Cao, K.A.; Rossouw, D.; Robert-Granié, C.; Besse, P. A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. B* **2008**, *7*, 37. [CrossRef] [PubMed]

57. Mevik, B.H.; Wehrens, R. The pls package: Principal component and partial least squares regression in R. *J. Stat. Softw.* **2007**, *18*, 1–24. [CrossRef]

58. Endelman, J.B. Ridge regression and other kernels for genomic selection with R. package rrBLUP. *Plant Genome* **2011**, *4*, 250–255. [CrossRef]

59. Pérez, P.; de los Campos, G.; Crossa, J.; Gianola, D. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* **2010**, *3*, 106–116. [CrossRef] [PubMed]

60. Li, Y.; Dutkowski, G.W.; Apiolaza, L.A.; Pilbeam, D. The genetic architecture of a *Eucalyptus globulus* full-sib breeding population in Australia. *For. Genet.* **2007**, *12*, 167–179.

61. Costa e Silva, J.C.; Hardner, C.; Potts, B.M. Genetic variation and parental performance under inbreeding for growth in *Eucalyptus globulus*. *Ann. For. Sci.* **2010**, *67*, 606. [CrossRef]

62. Callister, A.N.; England, N.; Collins, S. Genetic analysis of *Eucalyptus globulus* diameter, straightness, branch size, and forking in Western Australia. *Can. J. For. Res.* **2011**, *41*, 1333–1343. [CrossRef]

63. Mora, F.; Serra, N. Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. *Tree Genet. Genom.* **2014**, *10*, 711–719. [CrossRef]

64. Blackburn, D.; Farrell, R.; Hamilton, M.; Volker, P. Genetic improvement for pulpwood and peeled veneer in *Eucalyptus nitens*. *Can. J. For. Res.* **2012**, *42*, 1724–1732. [CrossRef]

65. Blackburn, D.P.; Hamilton, M.G.; Harwood, C.E.; Baker, T.G. Assessing genetic variation to improve stem straightness in *Eucalyptus globulus*. *Ann. For. Sci.* **2013**, *70*, 461–470. [CrossRef]

66. Burdon, R.D. Short note: Coefficients of variation in variables with bounded scales. *Silvae Genet.* **2008**, *57*, 179–180. [CrossRef]

67. Thavamanikumar, S.; McManus, L.J.; Tibbits, J.F.; Bossinger, G. The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Aust. For.* **2013**, *74*, 23–29. [CrossRef]

68. Brito, F.V.; Neto, J.B.; Sargolzaei, M.; Cobuci, J.A. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genet.* **2011**, *12*, 80. [CrossRef] [PubMed]

69. Habier, D.; Fernando, R.; Garrick, D. Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* **2013**, *194*, 597–607. [CrossRef] [PubMed]

70. Liu, H.; Zhou, H.; Wu, Y.; Li, X.; Zhao, J.; Zuo, T.; Zhang, X.; Zhang, Y.; Liu, S.; Shen, Y.; et al. The impact of genetic relationship and linkage disequilibrium on genomic selection. *PLoS ONE* **2015**, *10*, e0132379. [CrossRef] [PubMed]

71. Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A. Machine learning classification procedure for selecting SNPs in genomic selection: Application to early mortality in broilers. *J. Anim. Breed. Genet.* **2007**, *124*, 377–389. [CrossRef] [PubMed]

72. Usai, M.G.; Goddard, M.E.; Hayes, B.J. LASSO with cross-validation for genomic selection. *Genet. Res.* **2009**, *91*, 427–436. [CrossRef] [PubMed]

73. Weigel, K.A.; de los Campos, G.; González-Recio, O.; Naya, H.; Wu, X.L.; Long, N.; Rosa, G.J.; Gianola, D. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J. Dairy Sci.* **2009**, *92*, 5248–5257. [CrossRef] [PubMed]

74. Spindel, J.; Begum, H.; Akdemir, D.; Virk, P.; Collard, B.; Redoña, E.; Atlin, G.; Jannink, J.L.; McCouch, S.R. Genomic selection and association mapping in rice (Oryza sativa): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* **2015**, *11*, e1004982.

75. Wimmer, V.; Lehermeier, C.; Albrecht, T.; Auinger, H.J.; Wang, Y.; Schon, C.C. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* **2013**, *113*, 573–587. [CrossRef] [PubMed]

76. Stackpole, D.J.; Vaillancourt, R.E.; de Aguigar, M.; Potts, B.M. Age trends in genetic parameters for growth and wood density in *Eucalyptus globulus*. *Tree Genet. Genom.* **2010**, *6*, 179–193. [CrossRef]

77. Stackpole, D.J.; Vaillancourt, R.E.; Alves, A.; Rodrigues, J.; Potts, B.M. Genetic variation in the chemical components of *Eucalyptus globulus* wood. *G3: Genes Genom. Genet.* **2011**, *1*, 151–159. [CrossRef] [PubMed]

78. Daetwyler, H.D.; Pong-Wong, R.; Villanueva, B.; Woolliams, J.A. The impact of 539 genetic architecture on genome-wide evaluation methods. *Genetics* **2010**, *185*, 1021–1031. [CrossRef] [PubMed]

79. Jannink, J.L.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From 566 theory to practice. *Brief. Funct. Genom.* **2010**, *9*, 166–177. [CrossRef] [PubMed]

80. Colombani, C.; Croiseau, P.; Fritz, S.; Guillaume, F.; Legarra, A.; Ducrocq, V.; Robert-Granié, C. Comparison of partial least squares (PLS) and sparse PLS regressions in genomic selection in French dairy cattle. *J. Dairy Sci.* **2012**, *95*, 2120–2131. [CrossRef] [PubMed]

81. Freeman, J.S.; Potts, B.M.; Downes, G.M.; Pilbeam, D. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytol.* **2013**, *198*, 1121–1134. [CrossRef] [PubMed]

82. Thumma, B.R.; Baltunis, B.S.; Bell, J.C.; Emebiri, L.C. Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. *Tree Genet. Genom.* **2010**, *6*, 877–889. [CrossRef]

83. Bundock, P.C.; Potts, B.M.; Vaillancourt, R.E. Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*. *Tree Genet. Genom.* **2008**, *24*, 85–95. [CrossRef]

84. Gion, J.M.; Carouché, A.; Deweer, S.; Bedon, F.; Pichavant, F.; Charpentier, J.P.; Bailleres, H.; Rozenberg, P.; Carocha, V.; Ognouabi, N.; et al. Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. *BMC Genom.* **2011**, *12*, 301. [CrossRef] [PubMed]

85. Yu, B.; Lin, Z.; Li, H.; Li, X.; Li, J.; Wang, Y.; Zhang, X.; Zhu, Z.; Zhai, W.; Wang, X.; et al. *TAC1*, a major quantitative trait locus controlling tiller angle in rice. *Plant J.* **2007**, *52*, 891–898. [CrossRef] [PubMed]

86. Dardick, C.; Callahan, A.; Horn, R.; Ruiz, K.B.; Zhebentyayeva, T.; Hollender, C.; Whitaker, M.; Abbott, A.; Scorza, R. PpeTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species. *Plant J.* **2013**, *75*, 618–630. [CrossRef] [PubMed]

87. Müller, B.S.; Neves, L.G.; de Almeida Filho, J.E.; Resende, M.F.; Muñoz, P.R.; Dos Santos, P.E.T.; Filho, E.P.; Kirst, M.; Grattapaglia, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genom.* **2017**, *18*, 524. [CrossRef] [PubMed]

88. Isik, F.; Bartholomé, J.; Farjat, A.; Chancerel, E.; Raffin, A.; Sanchez, L.; Plomion, C.; Bouffier, L. Genomic selection in maritime pine. *Plant Sci.* **2016**, *242*, 108–119. [CrossRef] [PubMed]

89. Iwata, H.; Ebana, K.; Uga, Y.; Hayashi, T. Genomic prediction of biological shape: Elliptic fourier analysis and kernel partial least squares (PLS) regression applied to grain shape prediction in rice (*Oryza sativa* L.). *PLoS ONE* **2015**, *10*, e0120610. [CrossRef] [PubMed]

90. Estrada-Contreras, I.; Equihua, M.; Castillo-Campos, G.; Rojas-Soto, O. Climate change and effects on vegetation in Veracruz, Mexico: An approach using ecological niche modelling. *Acta Bot. Mex.* **2015**, *112*, 73–93. [CrossRef]

91. Woillez, M.N.; Kageyama, M.; Combourieu-Nebout, N.; Krinner, G. Simulating the vegetation response in western Europe to abrupt climate changes under glacial background conditions. *Biogeosciences* **2013**, *10*, 1561–1582. [CrossRef]

Bayesian analysis of growth, stem straightness and branching quality in full-sib families of *Eucalyptus globulus*

Freddy Mora[1], Paulina Ballesta[1,*], Nicolle Serra[2]

1.Universidade de Talca - Instituto de Ciências Biológicas - Talca, Chile.
2.Semillas Imperial SpA - Los Ángeles, Chile.

ABSTRACT: *Eucalyptus globulus* is one of the most commonly planted hardwood species for industrial use in various temperate regions around the world. The present study aimed to evaluate 62 full-sib families of *E. globulus* in one of the southernmost progeny trials of the species in the south of Chile. Estimates of genetic parameters for stem straightness, branching quality and growth traits were based on a Bayesian modelling approach using Gibbs sampling. A Bayes Factor (BF) analysis supported the hypothesis of significant additive genetic variation for all traits under study. Conversely, the BF supported a model with significant dominance effects for the diameter at breast height and stem volume, which explained up to 25% of the phenotypic variation. The greatest narrow-sense heritability estimates were found for the tree height and stem straightness, which were 0.15 (0.08 to 0.26) and 0.18 (0.10 to 0.28), respectively (mean of posterior distributions and 90% credible sets). In turn, the branching quality had a low heritability (narrow-sense) that varied from 0.05 to 0.10 (90% Bayesian credible region). The mean posterior estimate of genetic correlation between both quality traits was 0.22 (0.01 to 0.63, 90% credible set from a bi-trait threshold model), which indicates that stem straightness is positively related to branching quality. Our findings reveal that the study population responds to common patterns of breeding populations of *E. globulus*. This information is valuable for the development of improved seeds in the southern zone of Chile.

Key words: bayes factor, credible region, gibbs sampling, quality traits, threshold models.

178

INTRODUCTION

Trees of the genus *Eucalyptus* L'Hér. are recognized for their high biomass production, rapid growth rate, good adaptation to diverse environmental conditions and excellent wood quality for the production of paper and products derived from solid wood (Schmit et al. 2015; Mora and Arriagada 2016). In particular, *E. grandis, E. urophylla* and *E. globulus* are the most commonly planted hardwood species for industrial uses in tropical (or subtropical) and temperate zones, respectively (Carocha et al. 2015; Quang et al. 2010). Therefore, they stand out as the targets of multiple breeding programs and silvicultural management (Rosado et al. 2010; Carocha et al. 2015; Carbonari et al. 2016).

Eucalyptus globulus Labill. is a genetically diverse species with different geographic races (Foster et al. 2007) that allow it to tolerate environments with a certain degree of drought and/or low temperatures. Naturally, *E. globulus* is dominant in the coastal forests of southeastern Australia (Foster et al. 2007), where the climate is oceanic. Nevertheless, the species has been successfully and extensively planted in several temperate regions worldwide (e.g., Chile, Portugal, Australia, Spain) (Thavamanikumar et al. 2011; Águas et al. 2014; Mora and Serra 2014; Larcombe et al. 2013). Notably, *E. globulus* has been positively grown in a widespread range of environmental conditions that are adverse for plant establishment (Dutkowski and Potts 1999; Tibbits et al. 2006). Particularly in Chile, several studies have been focused on the mechanisms through which *E. globulus* trees respond to abiotic stresses (e.g. Navarrete-Campos et al. 2013; 2017; Aguayo et al. 2016).

E. globulus is the second most important woody species in the Chilean forest industry (Ballesta et al. 2018), where it comprises approximately 21% of national plantations (Morales et al. 2015). Traditionally, the species has experienced a rotation period that varies between eight and 12 years. However, breeding and intensive forest management have allowed production optimization, reducing the rotation time to five years (Morales et al. 2015). Due to the economic importance of *Eucalyptus*, several tree breeding programs have achieved substantial gains in various traits, such as growth (Blackburn et al. 2013; Mora and Serra 2014; Hamilton et al. 2015), wood properties (Stackpole et al. 2010; Hamilton et al. 2010), flowering-related traits (Ballesta et al. 2015; Contreras-Soto et al.

2016) pulp yield (Stackpole et al. 2010; Hamilton et al. 2010) and stem quality traits (Mora and Serra 2014; Arriagada et al. 2018). Estimation of the individual breeding value of a trait is closely related to the degree of precision with which the (co)variance components are estimated. In fact, lower precision is directly proportional to a greater deviation of the real genetic value of a genotype (Faria et al. 2007). In this context, the Bayesian method is a useful alternative for scientific inference of the genetic merit of trees because it considers levels of uncertainty in the estimated parameters and, generally, the credibility regions are more accurate than the confidence intervals obtained with frequentist inference (Gazola et al. 2016). Bayesian inference has been increasingly used in plant breeding and genetic studies in general (Fresnedo-Ramírez et al. 2017; Torres et al. 2018) by Markov Chain Monte Carlo (MCMC) methods, which use Markov sequences to effectively simulate complex (or not mathematically addressable) distributions. For instance, the joint density distribution that considers known environmental effects (e.g., experimental design), (co)variance components and additive and non-additive genetic effects, among others. An advantage of Bayesian procedures is the use of prior information in the analysis, which is particularly important when data are scarce (Cappa et al. 2012; Amaral Junior et al. 2016; Mora et al. 2016). Therefore, this method of statistical inference has been used in the analysis and estimation of genetic parameters in different tree species, including *Eucalyptus* (Vargas-Reeve et al. 2013; Mora and Serra 2014), *Pinus* (Cappa and Cantet 2006; 2008) and *Populus* (Wang et al. 2014).

The present study aimed to genetically evaluate full-sib families in one of the southernmost progeny trials of the species in the south of Chile. The estimation of genetic parameters (heritability, additive genetic correlations and variance components) of growth, stem straightness and branching quality (or branch-related defects) was based on Bayesian principles, including the selection of genetic models based on the Bayes Factor as a method of statistical inference.

MATERIALS AND METHODS
Field trial and phenotypic measurements

The study included 62 full-sib families resulting from an incomplete factorial mating design that involved 15 and 21

179

male and female parental trees, respectively. The trial was conducted in 2012 in the Chilean administrative region of Los Lagos, district of Purranque (40°58' S, 73°30' W, 326 m above sea level), under a randomized complete block design with 30 blocks and single-tree plots. The site features an oceanic climate with an average annual temperature of 13 °C; the average temperatures in the coldest and warmest months are 6 °C and 16 °C, respectively. The area was plowed to a depth of 0.60 m, and was fertilized with Di-Ammonium Phosphate (15% N, 15% P, and 15% K; Basacote® Plus).

The following traits of interest were evaluated in 4-year-old trees in 2016: tree height (H), diameter at breast height (DBH), wood volume (VOL), stem straightness (STR) and branching quality (BQ, or branch-related defects). STR was measured initially on a categorical 6-level scale according to Cameron et al. (2012) (1 = very crooked to 6 = completely straight without loss of productivity). BQ was also measured on a categorical 6-level scale (1 = tree with serious limitations to 6 = tree with all branching variables in good condition without loss of productivity). In addition, an analysis of tree survival was included, which was recorded as a binary response at 4 years old.

Bayesian genetic analysis

For estimation of the variance components and genetic parameters, a Bayesian analysis was carried out using the following base model (complete model) (Eq. 1):

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{s} + \mathbf{\varepsilon} \tag{1}$$

where: $\mathbf{y}$ corresponds to the observed values for a specific trait (phenotype) and $\mathbf{X}$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the known incidence matrices that relate the observation vector ($\mathbf{y}$) to vectors β, $\mathbf{a}$ and $\mathbf{s}$, respectively; B is the vector of the block effect; $\mathbf{a}$ is the additive effect vector of the individual trees; $\mathbf{s}$ is the effect vector of the full-sib families (accounts for dominance); and ε is the residual effect vector.

The Bayes Factor (BF) was used to test the significance of the effects of vectors $\mathbf{a}$ and $\mathbf{s}$ by model comparison, i.e., models M1 versus M2 in each of the following cases:

To evaluate the significance of the additive effects (Eqs. 2 and 3):

$$\text{M1:} \qquad \mathbf{y} = \mathbf{X\beta} + \mathbf{\varepsilon} \tag{2}$$

$$\text{M2:} \qquad \mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{\varepsilon} \tag{3}$$

To evaluate the significance of the effects of the full-sib families (dominance) (Eqs. 4 and 5):

$$\text{M1:} \qquad \mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{\varepsilon} \tag{4}$$

$$\text{M2:} \qquad \mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{s} + \mathbf{\varepsilon} \tag{5}$$

The BF based on the null hypothesis H0: $\mathbf{a}$ = 0 or H0: $\mathbf{s}$ = 0 comparing models M1 or M2, respectively, corresponds to Eq. 6:

$$\text{BF} = \frac{f(M_1/y)}{f(M_2/y)} \cdot \left[\frac{f(M_1)}{f(M_2)}\right]^{-1} \tag{6}$$

where: $f(M_1/y)$ and $f(M_2/y)$ correspond to the relative posterior probabilities of models M1 and M2, respectively; and $f(M_1)$ and $f(M_2)$ correspond to the prior probabilities for the competing models M1 and M2, respectively.

Subsequently, the posterior distributions of the variance components of the selected models were determined using the Gibbs sampling algorithm implemented in the MTGSAM program (Van Tassell and Van Vleck 1996). For the traits measured on the categorical scale (STR and BD), the likelihood-based threshold version of MTGSAM was used (Van Tassell et al. 1998). The following model (Eq. 7) was considered for the analysis of genetic associations between pairs of traits:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{12} \end{bmatrix}\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \\ + \begin{bmatrix} \mathbf{Z}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{\varepsilon}_1 \\ \mathbf{\varepsilon}_2 \end{bmatrix} \tag{7}$$

whose (co)variance components assumed an Inverted Wishart (IW) prior distribution. A uniform prior distribution (Flat) was considered for known environmental effects (blocks), whereas a normal distribution was used for the vector of observed responses, additive genetic effects and full-sib family effects. The convergence and auto-correlation of the Gibbs chains were assessed using the tests available in the CODA library of the R program, version 2.6.2 (R Development Core Team 2011).

180

Parameter estimates

Posterior estimates for individual-tree heritability (h^2), coefficient of additive genetic variation (CVa), dominance ratio (d^2) and coefficient of genetic variation of dominance (CVd) were estimated from the uni-trait analysis using the following expressions (Eqs. 8 to 11):

$$\hat{h}^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_s^2 + \hat{\sigma}_\varepsilon^2} \tag{8}$$

$$CVa = \frac{\sqrt{\hat{\sigma}_a^2}}{\bar{x}} \cdot 100 \tag{9}$$

$$\hat{d}^2 = \frac{4 \cdot \hat{\sigma}_s^2}{\hat{\sigma}_a^2 + \hat{\sigma}_s^2 + \hat{\sigma}_\varepsilon^2} \tag{10}$$

$$CVd = \frac{\sqrt{4 \cdot \hat{\sigma}_s^2}}{\bar{x}} \cdot 100 \tag{11}$$

where $\hat{\sigma}_a^2$, $\hat{\sigma}_s^2$ and $\hat{\sigma}_\varepsilon^2$ correspond to the additive genetic, full-sib family and residual variances, respectively. If the effect of full-sib families was not significant according to the BF, only the heritability and the corresponding coefficient of additive genetic variation were considered in the analysis. The additive genetic association between each pair of traits measured in the same tree was calculated from a bi-trait model using Eq. 12:

$$\hat{r} = \frac{\hat{\sigma}_{a_{XY}}}{\sqrt{\hat{\sigma}_{a_X}^2 + \hat{\sigma}_{a_Y}^2}} \tag{12}$$

in which $\hat{\sigma}_{a_{XY}}$ corresponds to the estimation of the additive covariance component between two pairs of traits X and Y (posterior mode values), $\hat{\sigma}_{a_X}^2$ and $\hat{\sigma}_{a_Y}^2$ are the posterior variances for each pair of traits in the analysis.

RESULTS AND DISCUSSION

Significant additive genetic variation was observed for all traits under study (Table 1). Specifically, each

trait had a BF < 0.01, or Log_e BF < 0, which provided decisive evidence against the model that considered an additive effect equal to zero (i.e., $\hat{\sigma}_a^2 = 0$). Conversely, the BF indicated a significant family effect (i.e., dominance variance) for the DBH and the VOL, whose dominance ratios were moderate and were superior to that of heritability (Table 2) for both traits; that is, VOL ($d^2 = 0.23$, $CVd = 23.7\%$) and DBH ($d^2 = 0.25$, $CVd = 9.7\%$). The coefficients of experimental variation were 14.8% (H), 18.5% (DBH), 47.4% (VOL), 35.5% (STR) and 20.4% (BQ), with H and DBH showing moderate values according to the proposed classification by Mora and Arriagada (2016). Additionally, no significant difference for tree survival was detected among families (BF > 1).

The Gibbs sampling chains converged for all posterior distributions of the parametric estimates using 10,000 burn-in iterations and a total of 50,000 Gibbs sampling rounds, in which 4,000 samples were withdrawn to estimate the marginal posterior distributions. Based on the models selected according to the BF (Table 1), the traits under study ranged from low to moderate heritability (estimates from the posterior distributions) (Casell 2009; Bush et al. 2011; Poke et al. 2006; Costa e Silva et al. 2009) with variation between 0.03 (VOL) and 0.15 (STR), and the coefficients of additive variation ranged between $CVa = 4.4\%$ and $CVa = 14.2\%$ for DBH and STR, respectively (Table 2). In a genetic evaluation of twelve-month-old $E.$ $globulus$ clones also in southern Chile, Mora et al. (2013) found moderate to high values of heritability of tree height with a range from 0.12 to 0.41 (mode value of the posterior distribution of heritability). In a frequentist approach, Costa e Silva et al. (2005) reported that the heritability of height growth for $E.$ $globulus$ in early stages (at 3 years old) ranged from 0.15 to 0.4 (evaluated under six different environmental conditions), which is also consistent with our findings. Previous studies and our results revealed that the height of plants of $E.$ $globulus$ is controlled by an additive genetic control and, in a minor proportion, by dominance effects. Contrarily, our results suggested that VOL and DBH could be moderately controlled by dominance effects, a result also reported by Hamilton et al. (2015) in $E.$ $globulus$ at age 10 (DBH: $d^2 = 0.13$; VOL: $d^2 = 0.24$).

The stem straightness and branch quality (or the proportion of defective branches) are important traits

181

Table 1. Model assessment based on Bayes factor for the total height (H), diameter at breast height (DBH), volume (VOL), stem straightness (STR) and branching quality (BQ) in full-sib families of *Eucalyptus globulus*.

Competitive models	H		DBH		VOL		STR		BD	
	BF	LogBF	BF	LogBF	BF	LogBF	BF	LogBF	BF	LogBF
M1: $y = X\beta + Z_1a + \varepsilon$	>1	1.6	<0.01	−6.8	<0.01	−5.5	>1	63.4	>1	39.9
M2: $y = X\beta + Z_1a + Z_2s + \varepsilon$										
M1: $y = X\beta + \varepsilon$	<0.01	−7.6	<0.01	−18.9	<0.01	−15.5	<0.01	−32.8	<0.01	−23.9
M2: $y = X\beta + Z_1a + \varepsilon$										

According to Jeffreys (1961), if BF > 1 or LogBF > 0, model M1 is supported. If BF < 0.01 or LogBF < 0, decisive evidence against model M1.

Table 2. Bayesian estimates of the individual-tree heritability (h^2), coefficient of additive genetic variation (*CVa*), dominance ratio (d^2) and coefficient of genetic variation of dominance (*CVd*) with lower and upper cutoffs for 90% credible sets the total height (H), diameter at breast height (DBH), volume (VOL), stem straightness (STR) and branching quality (BQ) in full-sib families of *Eucalyptus globulus*.

Estimates	H	DBH		VOL		STR	BQ
	h^2	h^2	d^2	h^2	d^2	h^2	h^2
Mean	0.15	0.07	0.28	0.06	0.25	0.18	0.06
Median	0.14	0.06	0.27	0.04	0.24	0.17	0.05
Mode	0.12	0.05	0.25	0.03	0.23	0.15	0.05
Standard Deviation	0.06	0.05	0.08	0.05	0.07	0.06	0.03
Lower cutoff	0.08	0.01	0.16	0.01	0.12	0.10	0.05
Upper cutoff	0.26	0.16	0.41	0.14	0.38	0.28	0.10
CVa (%)	5.2	4.4		8.6		14.2	4.7
CVd (%)	0.0	9.7		23.7		0.0	0.0

for the production of timber of *E. globulus*, whose quality is sensitive to the market (Blackburn et al. 2013; Mora and Serra 2014). In the present study, STR showed a moderate heritability ranging from 0.10 to 0.28 (90% Bayesian credible set). These results matched the values obtained by Mora and Serra (2014) in a trail of half-sib families of *E. globulus* of 15 years old, in which they found low to moderate heritability varying from 0.03 to 0.21 (90% credible set). Conversely, Callister et al. (2011) found relatively higher estimates of heritability in full-sib families of *E. globulus* in Australia at an age similar to that of the present study (3.5 years) with a range from 0.10 to 0.46 and a mean of 0.28 (the upper limit of the 90% credible set of the posterior distribution found in the present study). In turn, the branch-related defects had a low heritability that varied from 0.05 to 0.1 (*CVa* = 4.7%). Although this finding indicated that the genetic progress for this trait would be modest in the present breeding population, we could emphasize that only 2.2% of the trees in the test were in the lower BQ category, with the great majority showing an intermediate level of quality (74%). Callister et al. (2011) determined that the branch

thickness (a trait related to BQ) featured low genetic control ($h^2 = 0.05$), which was consistent with our results.

The estimates of the additive genetic correlations between pairs of traits (i.e., the point estimate of the posterior distribution, mode and 90% credible regions) are shown in Table 3. According to the 90% credible set, 6 of the 10 estimates were significantly different from zero. As expected, the growth traits showed a positive association and were significantly different from zero, which agreed with previous studies of *E. globulus* (Hamilton et al. 2010; Mora and Serra 2014). Genetic correlation values between growth traits suggest that the genetic selection based on one growth trait would have a positive impact on other trait in early stages of *E. globulus*. Several studies have reported quantitative trait loci (QTLs) associated with DBH and H (Arriagada et al. 2018; Thumma et al. 2010; Ballesta et al. 2015), supporting to the hypothesis of genetic relation between these growth traits.

The mean posterior estimate of genetic correlation between both timber quality traits was 0.22 (0.01 to 0.63, 90% credible set for the threshold model), which indicates that stem straightness is positively related to branching quality.

182

Table 3. Posterior means of genetic correlations (above the diagonal) and 90% credible sets from marginal posterior distributions (below the diagonal), calculated between pairs of traits for the total height (H), diameter at breast height (DBH), volume (V), stem straightness (STR) and branching quality (BQ).

Traits	H	DBH	VOL	STR	BD
H	1	0.45	0.61	−0.43	0.30
DBH	[0.14, 0.71]	1	0.94	−0.20	−0.34
VOL	[0.36, 0.82]	[0.89, 0.98]	1	−0.24	−0.22
STR	[−0.67, −0.14]	[−0.50, 0.11]	[−0.55, 0.10]	1	0.22
BD	[0, 0.71]	[−0.70, −0.01]	[−0.65, 0.24]	[0.01, 0.63]	1

Cameron et al. (2012) found a similar genetic correlation between stem straightness and branching habit (scored for quality) in *Picea sitchensis* (r = 0.28). Our findings could be promise in the context of an early selection program of *E. globulus*. According to the results, stem straightness was more heritable than branching quality at age 4, but both traits were genetically related. Therefore the selection based on stem straightness could promote the selection of other important economic traits with low heritability, such as branching quality.

The stem straightness presented a negative association with the total height (−0.51), whose value was significantly different from zero (−0.67 to −0.14), whereas the associations between straightness with VOL and DBH were not significantly different from zero even though the point estimates were negative. These findings were in agreement with those of Callister et al. (2011) in *E. globulus*, who found an additive genetic correlation between DBH and STR with an average value of −0.18 and variation from −0.71 to 0.33. Mora and Serra (2014) also found negative additive genetic correlations between growth (DBH, H and VOL) and STR in a 15-year-old *E. globulus* trial, which were not significantly different from zero. In another *Eucalyptus* species, Vargas-Reeve et al. (2013) found that the additive genetic correlations between stem straightness with height and diameter were not significantly different from zero (95% Bayesian credible set: −0.04 to 0.37 and −0.16 to 0.35, respectively) when measured on 9-year-old trees. Coincidentally, Callister et al. (2008) also found non-significant genetic correlations between growth and STR in *E. cladocalyx* trees at 3.5 and 5.5 years of age.

CONCLUSION

Our findings revealed that the study population responds to common patterns of breeding populations of *E. globulus*. The total height of the trees and the stem straightness were the traits showing the greatest narrow-sense heritability, with moderate values and a negligible dominance effect. On other hand, the dominance effect should be considered for future research and breeding prospects based on the wood diameter and/or volume. In addition, the dominance effect was negligible for branch-related defects, but is controlled by minor genetic additive factors at early stages.

ACKNOWLEDGEMENTS

AUTHOR'S CONTRIBUTION

Conceptualization, Mora F.; Methodology, Mora F.; Investigation, Mora F.; Writing – Original Draft, Mora F.; Writing – Review and Editing, Ballesta P.; Funding Acquisition, Mora F.; Resources, Serra N.; Supervision, Mora F. and Ballesta P.

ORCID IDs

F. Mora
https://orcid.org/0000-0002-4845-3549
P. Ballesta
https://orcid.org/0000-0003-1428-0456
N. Serra
https://orcid.org/0000-0002-9889-2355

183

REFERENCES

Águas, A., Ferreira, A., Maia, P., Fernandes, P. M., Roxo, L., Keizer, J., Silva, J. S., Rego, F. C. and Moreira, F. (2014). Natural establishment of Eucalyptus globulus Labill. in burnt stands in Portugal. Forest Ecology and Management, 323, 47-56. https://doi.org/10.1016/j.foreco.2014.03.012

Aguayo, P., Sanhueza, J., Noriega, F., Ochoa, M., Lefeuvre, R., Navarrete, D., Fernández, M. and Valenzuela, S. (2016). Overexpression of an SKn-dehydrin gene from Eucalyptus globulus and Eucalyptus nitens enhances tolerance to freezing stress in Arabidopsis. Trees, 30, 1785-1797. https://doi.org/10.1007/s00468-016-1410-9

Amaral Junior, A. T., Freitas, I. L. J., Guimarães, A. G., Maldonado, C., Arriagada, O. and Mora, F. (2016). Bayesian analysis of quantitative traits in popcorn (Zea mays L.) through four cycles of recurrent selection. Plant Production Science, 19, 574-578. https://doi.org/10.1080/1343943X.2016.1222870

Arriagada, O., Amaral Junior, A. T. and Mora, F. (2018). Thirteen years under arid conditions: exploring marker-trait associations in Eucalyptus cladocalyx for complex traits related to flowering, stem form and growth. Breeding Science, 68, 3. https://doi.org/10.1270/jsbbs.17131

Ballesta, P., Mora, F., Ruiz, E. and Contreras-Soto, R. (2015). Marker-trait associations for survival, growth, and flowering components in Eucalyptus cladocalyx under arid conditions. Biologia Plantarum, 59, 389-392. https://doi.org/10.1007/s10535-014-0459-9

Ballesta, P., Serra, N., Guerra, F. P., Hasbún, R. and Mora, F. (2018). Genomic prediction of growth and stem quality traits in Eucalyptus globulus Labill. at its southernmost distribution limit in Chile. Forests, 9, 779. https://doi.org/10.3390/f9120779

Blackburn, D. P., Hamilton, M. G., Harwood, C. E., Baker, T. G. and Potts, B. M. (2013). Assessing genetic variation to improve stem straightness in Eucalyptus globulus. Annals of Forest Science, 70, 461-470. https://doi.org/10.1007/s13595-013-0277-9

Bush, D., McCarthy, K. and Meder, R. (2011). Genetic variation of natural durability traits in Eucalyptus cladocalyx (sugar gum). Annals of Forest Science, 68(6), 1057. https://doi.org/10.1007/s13595-011-0121-z

Callister, A., Bush, D. J., Collins, S., and Davis, W. (2008). Prospects for genetic improvement of Eucalyptus cladocalyx in Western Australia. New Zealand Journal of Forestry Science, 38, 211-226.

Callister, A. N., England, N. and Collins, S. (2011). Genetic analysis of Eucalyptus globulus diameter, straightness, branch size, and forking in Western Australia. Canadian Journal of Forest Research, 41, 1333-1343. https://doi.org/10.1139/x11-036

Cameron, A. D., Kennedy, S. G. and Lee, S. J. (2012). The potential to improve growth rate and quality traits of stem straightness and branching habit when breeding Picea sitchensis (Bong.) Carr. Annals of Forest Science, 69, 363-371. https://doi.org/10.1007/s13595-011-0167-y

Cappa, E. P. and Cantet, R. J. C. (2006). Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. Canadian Journal of Forest Research, 36, 1276-1285. https://doi.org/10.1139/x06-024

Cappa, E. P. and Cantet, R. J. C. (2008). Direct and competition additive effects in tree breeding: bayesian estimation from an individual tree mixed model. Silvae Genetica, 57, 45-56. https://doi.org/10.1515/sg-2008-0008

Cappa, E. P., Yanchuk, A. D. and Cartwright, C. V. (2012). Bayesian inference for multi-environment spatial individual-tree models with additive and full-sib family genetic effects for large forest genetic trials. Annals of Forest Science, 69, 627-640. https://doi.org/10.1007/s13595-011-0179-7

Carbonari, C. A., Miranda, L. G., Gomes, G. L. G. C., Picoli Junior, G. J., Matos, A. K. A. and Velini, D. E. (2016). Differential tolerance of eucalyptus clones to sulfentrazone applied in different soil textures. Scientia Forestalis, 44, 9-18. https://doi.org/10.18671/scifor.v44n109.01

Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevereiro, P., Myburg, A. A., Paiva, J. A. and Grima-Pettenati, J. (2015). Genome-wide analysis of the lignin toolbox of Eucalyptus grandis. New Phytologist, 206, 1297-1313. https://doi.org/10.1111/nph.13313

Cassell, B. (2009). Using heritability for genetic improvement. Virginia: Virginia Cooperative Extension publication, 404-084.

Contreras-Soto, R., Ballesta, P., Ruiz, E. and Mora, F. (2016). Identification of ISSR markers linked to flowering traits in a representative sample of Eucalyptus cladocalyx. Journal of Forestry Research, 27, 239-245. https://doi.org/10.1007/s11676-015-0149-2

Costa e Silva, J., Borralho, N. M., Araújo, J. A., Vaillancourt, R. E. and Potts, B. M. (2009). Genetic parameters for growth, wood density

184

and pulp yield in *Eucalyptus globulus*. Tree Genetics & Genomes, 5, 291-305. https://doi.org/10.1007/s11295-008-0174-9

Costa e Silva, J. C., Dutkowski, G. W. and Borralho, N. M. (2005). Across-site heterogeneity of genetic and environmental variances in the genetic evaluation of *Eucalyptus globulus* trials for height growth. Annals of forest science, 62, 183-191. https://doi.org/10.1051/forest:2006095

Dutkowski, G. W. and Potts, B. M. (1999). Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. globulus and a revised racial classification. Australian Journal of Botany, 47, 237-263. https://doi.org/10.1071/BT97114

Faria, C. U., Magnabosco, C. U., Reyes, A., Lôbo, R. B., Bezerra, L. A. F. and Sainz, R. D. (2007). Bayesian inference in a quantitative genetic study of growth traits in Nelore cattle (*Bos indicus*). Genetics Molecular Biology, 30, 545-551. https://doi.org/10.1590/S1415-47572007000400007

Foster, S. A., McKinnon, G. E., Steane, D. A., Potts, B. M. and Vaillancourt, R. E. (2007). Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. New Phytologist, 175, 370-380. https://doi.org/10.1111/j.1469-8137.2007.02077.x

Fresnedo-Ramírez, J., Famula, T. R. and Gradziel, T. M. (2017). Application of a Bayesian ordinal animal model for the estimation of breeding values for the resistance to *Monilinia fruticola* (G.Winter) Honey in progenies of peach [*Prunus persica* (L.) Batsch]. Breeding Science, 67, 110-122. https://doi.org/10.1270/jsbbs.16027

Gazola, S., Scapim, C. A., Araujo, Â. M. M., Rossi, R. M., Amaral, A. T. and Vivas, M. (2016). Nonlinear models to describe the maize seed quality during the maturation stage: a Bayesian approach. Australian Journal of Crop Science, 10, 598-603.

Hamilton, M. G., Acuna, M., Wiedemann, J. C., Mitchell, R., Pilbeam, D. J., Brown, M. W. and Potts, B. M. (2015). Genetic control of *Eucalyptus globulus* harvest traits. Canadian Journal of Forest Research, 45, 615-624. https://doi.org/10.1139/cjfr-2014-0428

Hamilton, M. G., Potts, B. M., Greaves, B. L. and Dutkowski, G. W. (2010). Genetic correlations between pulpwood and solid-wood selection and objective traits in *Eucalyptus globulus*. Annals Forest Science, 67, 511-511. https://doi.org/10.1051/forest/2010013

Jeffreys, H. (1961). The Theory of Probability. 3rd ed. Oxford: Oxford Classic Texts in the Physical Sciences.

Larcombe, M. J., Silva, J. S., Vaillancourt, R. E. and Potts, B. M. (2013). Assessing the invasive potential of *Eucalyptus globulus* in Australia:

quantification of wildling establishment from plantations. Biological Invasions, 15, 2763-2781. https://doi.org/10.1007/s10530-013-0492-1

Mora, F. and Arriagada, O. (2016). A classification proposal for coefficients of variation in *Eucalyptus* experiments involving survival, growth and wood quality variables. Bragantia, 75, 263-267. https://doi.org/10.1590/1678-4499.458

Mora, F., Concha, C. M. and Figueroa, C. R. (2016). Bayesian inference of genetic parameters for survival, flowering, fruit set, and ripening in a germplasm collection of Chilean strawberry using threshold models. Journal of the American Society for Horticultural Science, 141, 285-291. https://doi.org/10.21273/JASHS.141.3.285

Mora, F., Rubilar, R., Emhart, V. I. and Saavedra, J. (2013). Predicción bayesiana de parámetros genéticos en clones de *Eucalyptus globulus* bajo condiciones de suplemento hídrico. Ciencia Florestal, 23, 529-536. https://doi.org/10.5902/198050989297

Mora, F. and Serra, N. (2014). Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. Tree Genetics & Genomes, 10, 711-719. https://doi.org/10.1007/s11295-014-0716-2

Morales, M., Aroca, G., Rubilar, R., Acuña, E., Mola-Yudego, B. and González-García S (2015). Cradle-to-gate life cycle assessment of *Eucalyptus globulus* short rotation plantations in Chile. Journal of Cleaner Production, 99, 239-249. https://doi.org/10.1016/j.jclepro.2015.02.085

Navarrete-Campos, D., Bravo, L. A., Rubilar, R. A., Emhart, V. and Sanhueza, R. (2013). Drought effects on water use efficiency, freezing tolerance and survival of *Eucalyptus globulus* and *Eucalyptus globulus* × *nitens* cuttings. New Forests, 44, 119-134. https://doi.org/10.1007/s11056-012-9305-0

Navarrete-Campos, D., Le Feuvre, R., Balocchi, C. and Valenzuela, S. (2017). Overexpression of three novel CBF transcription factors from *Eucalyptus globulus* improves cold tolerance on transgenic *Arabidopsis thaliana*. Trees, 31, 1041-1055. https://doi.org/10.1007/s00468-017-1529-3

Poke, F. S., Potts, B. M., Vaillancourt, R. E. and Raymond, C. A. (2006). Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*. Annals of Forest Science, 63, 813-821. https://doi.org/10.1051/forest:2006080

Quang, T. H., Kien, N. D., von Arnold, S., Jansson, G., Thinh, H. H. and Clapham, D. (2010). Relationship of wood composition to growth traits of selected open-pollinated families of *Eucalyptus urophylla* from a progeny trial in Vietnam. New forests, 39, 301-312. https://doi.org/10.1007/s11056-009-9172-5

185

R Development Core Team (2011). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rosado, T. B., Tomaz, R. S., Ribeiro Junior, M. F., Rosado, A. M., Guimarães, L. M. S., Araujo, E. F., Alfenas, A. C. and Cruz, C. D. (2010). Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic Eucalyptus spp. populations. Crop Breeding and Applied Biotechnology, 10, 321-328. https://doi.org/10.1590/S1984-70332010000400006

Schmit, R., Mora, F., Emhart, V. I. and Rubilar, R. (2015). Longitudinal analysis in the selection of Eucalyptus globulus clones under contrasting water availability conditions. Scientia Forestalis, 43, 217-224.

Stackpole, D. J., Vaillancourt, R. E., Downes, G. M., Harwood, C. E. and Potts, B. M. (2010). Genetic control of kraft pulp yield in Eucalyptus globulus. Canadian Journal of Forest Research, 40, 917-927. https://doi.org/10.1139/X10-035

Thavamanikumar, S., McManus, L. J., Tibbits, J. F. G. and Bossinger, G. (2011). The significance of single nucleotide polymorphisms (SNPs) in Eucalyptus globulus breeding programs. Australian Forestry, 74, 23-29. https://doi.org/10.1080/00049158.2011.10676342

Thumma, B. R., Baltunis, B. S., Bell, J. C., Emebiri, L. C., Moran, G. F. and Southerton, S. G. (2010). Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in Eucalyptus nitens full-sib families. Tree Genetics & Genomes, 6, 877-889. https://doi.org/10.1007/s11295-010-0298-6

Tibbits, W. N., White, T. L., Hodge, G. R. and Borralho, N. M. (2006). Genetic variation in frost resistance of Eucalyptus globulus ssp. globulus assessed by artificial freezing in winter. Australian Journal of Botany, 54, 521-529. https://doi.org/10.1071/BT02061

Torres, L. G., Rodrigues, M. C., Lima, N. L., Trindade, T. F. H., Silva, F. F., Azevedo, C. F. and De Lima, R. O. (2018). Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. Plos One, 13, e0199492. https://doi.org/10.1371/journal.pone.0199492

Van Tassell, C. P., Van Vleck, L. D. and Gregory, K. E. (1998). Bayesian analysis of twinning and ovulation rates using a multiple-trait threshold model and Gibbs sampling. Journal of Animal Science, 76, 2048-2061. https://doi.org/10.2527/1998.7682048x

Van Tassell, C. P. and Van Vleck, L. D. (1996). Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co) variance component inference. Journal of Animal Science, 74, 2586-2597. https://doi.org/10.2527/1996.74112586x

Vargas-Reeve, F., Mora, F., Perret, S. and Scapim, C. A. (2013). Heritability of stem straightness and genetic correlations in Eucalyptus cladocalyx in the semi-arid region of Chile. Crop Breeding and Applied Biotechnology, 13, 107-112.

Wang, Z., Du, S., Dayanandan, S., Wang, D., Zeng, Y. and Zhang, J. (2014). Phylogeny reconstruction and hybrid analysis of Populus (Salicaceae) based on nucleotide sequences of multiple single-copy nuclear genes and plastid fragments. Plos One, 9, e103645. https://doi.org/10.1371/journal.pone.0103645

186

Article

Bayesian Mapping Reveals Large-Effect Pleiotropic QTLs for Wood Density and Slenderness Index in 17-Year-Old Trees of *Eucalyptus cladocalyx*

Camilo E. Valenzuela [1] , **Paulina Ballesta** [1], **Carlos Maldonado** [1] , **Ricardo Baettig** [2] ,
Osvin Arriagada [1] , **Gabrielle Sousa Mafra** [3] and **Freddy Mora** [1,*]

[1]   Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile;
     camilo.vp5@gmail.com (C.E.V.); paballesta@gmail.com (P.B.); cmaldo1782@gmail.com (C.M.);
     arriagada.lagos.o@gmail.com (O.A.)
[2]   Faculty of Forest Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile; rbaettig@utalca.cl
[3]   Universidade Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto Lamego, 2000,
     Campos dos Goytacazes, Rio de Janeiro 28013-602, Brazil; gabrielle.smafra@yahoo.com.br
*   Correspondence: fmora@utalca.cl; Tel.: +56-71-2200280

check for
updates

Abstract: *Eucalyptus cladocalyx* F. Muell is a tree species suitable for low-rainfall sites, even with annual average precipitation as low as 150 mm per year. Its wood is classified as highly durable and its permanence in soil is longer than 25 years, so it can be used for multiple applications. Given that about 41% of the world's land area is classified as drylands, added to the impact of climate change on the availability of water resources, it becomes necessary to use plant species that can tolerate environments with low water availability. In this study, a Bayesian analysis of genetic parameters showed that wood density (WD) was moderately heritable, with a posterior mean of $h^2 = 0.29$ and a Bayesian credibility region (90%) of 0.06–0.74, while the slenderness coefficient (SC) was highly heritable, with a posterior mean of $h^2 = 0.48$ and a Bayesian credibility region (90%) of 0.11–0.87. Through Bayesian regression analysis, we identified four and three significant associations for WD and SC, respectively. Another important finding of the bi-trait Bayesian analysis was the detection of three large-effect pleiotropic QTLs located on LG4 at 52 cM, on LG2 at 125 cM, and on LG6 at 81 cM. Bayesian bi-trait regression and the posterior probability of association indicated that three QTLs presented strong evidence of association with WD and SC. This provides convincing evidence that the loci qtlWD130/qtlSC130, qtlWD195/qtlSC195, and qtlWD196/qtlSC196 have a significant pleiotropic effect. The association mapping based on multivariate Bayesian regression was useful for the identification of genomic regions with pleiotropic effects. These loci can be used in molecular marker-assisted breeding to select trees with better wood density.

Keywords: Bayesian multivariate mapping; pleiotropy analysis; posterior probability of association

1. Introduction

The endemic species *Eucalyptus cladocalyx* F. Muell from southern Australia is naturally distributed in three geographical regions: Kangaroo Island, the Eyre Peninsula, and the Flinders Range [1]. An interesting feature is its natural evolution to tolerate environments where the annual rainfall is lower than 200 mm [2,3]. The trees of *E. cladocalyx* have tremendous potential for the production of wood with multiple applications, due to its straight and poorly branched stems [4]. Lundqvist et al. [5] compared the physical properties of the wood of three genotypes of eucalyptus tolerant to drought: *E. grandis* × *E. camaldulensis* hybrid, *E. gomphocephala* and *E. cladocalyx*, and concluded that *E. cladocalyx*

is one of the most suitable species to produce structural wood, given the high density of wood and rigidity of its microfibrils.

According to the Australian classification of natural durability of wood, *E. cladocalyx* is one of the most resistant to biodegradation [6], with a useful life of 25 years or more, in contact with the soil [7]. Natural durability has been described as the resistance of wood to biological decomposition by fungi, bacteria, insects, or marine organisms without copper/chromium/arsenate treatment to preserve it [8]. Therefore, considering the health problems and environmental pollution associated with the use of chemical preservatives, *E. cladocalyx* has been used to produce high natural durability wood with potential uses in the agricultural and industrial sectors [6,9].

The density of the wood is considered one of the most informative properties in terms of the physical and mechanical behavior of a tree, followed by the production of sawn wood [10]. It has been observed that a higher density of wood also implies greater natural durability [11]. Wood density is related to the structure of the cells of the vessels and the parenchyma [12], in which species resistant to cavitation have a greater density in the wood [13]. In this sense, density is positively correlated with tolerance to drought, so arid environments tend to be dominated by species with high wood density [13].

According to several authors, wood density is moderately to highly heritable in most forest species [10,14,15]. However, other studies have indicated that wood density can vary depending on local environmental conditions [13,16,17]. Bush et al. [6] determined that the wood density of *E. cladocalyx* varies considerably based on the different regions of origin of the species, presenting moderate heritability ($h^2 = 0.41$).

Association studies based on DNA markers provide a comprehensive approach to identify genetic loci associated with the phenotypic response of complex traits. Association studies have identified many quantitative trait loci (QTL) associated with wood density in *E. urophylla*, *E. grandis* [18], and *E. globulus* [19], along with microfibrils angle and total lignin content in *E. nitens* [19]. Wood density and other mechanical properties are highly variable among species and sites, and even between and within the trees themselves [20]. Due to the relevance of this type of trait for the wood industry, several studies have been developed to estimate genetic parameters and identify genes, or regions in the genome, related to the production of structural components of wood (for example, cellulose and lignin) and the spatial arrangement of the microfibrils [21–26]. In general, these studies have been explored with molecular markers and transcriptomic analyzes in the principal forest species. For example, Nakahama et al. [24] showed that genes related to the biosynthesis of lignocellulosic compounds (for example, cellulose synthase, invertase, l-cinnamate-4-hydroxylase and cinnamoyl-CoA reductase) and genes related to the modification of the cell wall (for example, expansins and xyloglucan endo-transglycosylase/hydrolase) are mostly expressed in *Eucalyptus* hybrids that have a high wood density.

On the other hand, a factor that is also relevant in the production of sawn wood is the tree growth stress, which is the product of the stresses of the same growth of the tree when forming a new layer of the trunk during the differentiation of the fibers [27]. One of the best indicators of this type of stress is the slenderness coefficient [28]. This variable reveals the capacity of a tree to sustain itself against these tensions and the tensions of the environment, e.g., wind [29]. Although the slenderness of a tree depends on environmental factors such as the availability of light, wind, and plant density [30–32], it has been concluded that this trait is also controlled by genetic components [33,34]. Moreover, the slenderness of a tree is also influenced by the density of its wood [35].

Due to the potential of *E. cladocalyx* for the production of high-quality wood at low-rainfall sites, and that wood density and the slenderness of trees are traits closely related to the quality of the wood, the following objectives were proposed: (i) to determine Bayesian genetic parameters of wood density, measured indirectly by means of Pilodyn, and the slenderness of *E. cladocalyx* trees grown under arid environmental conditions, (ii) to identify genomic regions associated with density of wood and

slenderness, using a key set of 130 microsatellite markers, and (iii) corroborate the pleiotropy of the loci associated with both traits, using a multivariate Bayesian regression approach.

2. Materials and Methods

2.1. DNA Extraction and Genotyping

The association analysis was focused on a sample of 245 *E. cladocalyx* trees, corresponding to 5 blocks × 49 half-sib families, from a provenance-progeny trial established in 2001 [2]. Forty-seven families come from five natural provenances: Flinders Chase (35°57′ S, 136°42′ E), Marble Range (34°30′ S, 135°30′ E), Mount Remarkable (32°43′ S, 138°06′ E), Cowell (33°38′ S, 136°40′ E) and Wirrabara (33°06′ S, 138°14′ E), distributed geographical in three regions of southern Australia (Eyre Peninsula, Flinders Ranges and Kangaroo Island), and two families from a local seed resource (31°40′ S, 71°14′ W). The trial was situated in a dryland area within the region of Coquimbo, Choapa Province (31°38′ S Latitude; 71°19′ W Longitude; and altitude of 297 m) in the south of the Atacama Desert, Chile. Total genomic DNA was isolated from juvenile leaves using the method CTAB according to da Silva [36] with some modifications. A total of 130 SSR polymorphic markers distributed across the genome of *E. cladocalyx* were used for genotyping the association population [37]. Polymerase chain reaction (PCR) was performed as follows: 0.3 uM of each primer, 40 ng of genomic DNA, 1 U of Taq DNA polymerase, 0.2 mM of each dNTP, 10 mM of Tris-HCl pH 8.3, 50 mM of KCl and 1.5 mM of $MgCl_2$. The amplification was run as follows: initial heat at 95 °C for 5 min followed by 40 cycles of 95 °C for 1 min, annealing temperature of each primer for 1 min, 72 °C for 1 min, and a final extension of 72 °C for 5 min. Finally, the PCR products were separated and visualized according to the methods described by Mora et al. [37].

2.2. Traits Assessment

Wood density (WD) of 17-year-old trees was indirectly estimated after removal of the bark based on Pilodyn penetration using a Pilodyn 6J Forest (PROCEQ, Zurich, Switzerland). Penetration depth (in millimeters) was measured two to three times per tree at breast height, until two identical measurements (at a precision of 2.0 mm) were obtained. The slenderness coefficient (SC) was calculated as follows: SC = height (m)/DBH (m) according to Benomar et al. [38], where DBH corresponds to the diameter at breast height. The measurements of height and DBH were taken from Arriagada et al. [2]. The Tukey-Kramer multiple comparison test was performed to determine significant differences between the means of each provenance for the traits evaluated.

2.3. Genetic Parameters

The subsequent marginal distributions for the parameters, narrow-sense heritability and variance components, were estimated using a Bayesian approach, via Gibbs sampling algorithm, in MTGSAM [39] using the following base model:

$$y = Xb + Zp + Wf + \varepsilon, \tag{1}$$

where y is the vector of observed values of WD and SC; b and p are vectors of fixed effect of block and provenance, respectively; f is the vector of random effects of family, and ε is the residual error. X, Z, and W are known incidence matrices. Posterior estimates for the narrow-sense heritability of traits, were calculated from posterior samples of variance components obtained by the model (1) detailed above, using the expression:

$$\hat{h}^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_\varepsilon^2}, \tag{2}$$

189

where $\hat{\sigma}_a^2$ and $\hat{\sigma}_\varepsilon^2$ are the additive genetic and residual variances. This Bayesian analysis was contrasted with the restricted maximum likelihood (REML) method in ASReml 3.4 [40]. Coefficients of additive genetic variation (*CVa*) were calculated from posterior samples of additive genetic variance as follows:

$$CVa = \frac{\sqrt{\hat{\sigma}_a^2}}{\overline{x}},\tag{3}$$

where $\hat{\sigma}_a^2$ is the additive genetic variance and $\overline{x}$ is the overall mean. Additive genetic correlations between WD and SC was estimated by:

$$r_{xy} = \frac{\hat{\sigma}_{Gxy}}{\hat{\sigma}_{Gx}^2 \cdot \hat{\sigma}_{Gy}^2},\tag{4}$$

where $\hat{\sigma}_{Gxy}$ correspond to posterior distribution samples of genotypic covariance between the traits, and $\hat{\sigma}_{Gx}^2$, $\hat{\sigma}_{Gy}^2$ correspond to posterior mean distribution samples of genotypic variance for WD and SC. Variance and covariance components were assumed to have an Inverted Wishart prior distribution. A uniform prior distribution (Flat) was considered for known environmental effects (blocks), whereas a normal distribution was used for the vector of traits of interest and additive genetic effects. The convergence and autocorrelation of the Gibbs chains were assessed using the tests available in the CODA library of the R program, Coventry, England [41].

2.4. Population Structure and Kinship Analysis

The software STRUCTURE, Stanford [CA], United States 2.3.2 [42] was used to cluster the individuals into a number (K) of genetically homogeneous subpopulations based on an admixture model with correlated allele frequencies between provenances. For each K value (previously K = 1–6), 10 runs were conducted separately each with 100,000 Monte Carlo Markov Chain (MCMC) replicates and a burn-in period of 10,000 iterations. The optimal K value was determined with the highest ΔK method [43]. The membership coefficient (Q) of each individual was used to form the population structure Q matrix. Only markers with allele frequencies of 5% or higher were included in data analyses. The software TASSEL, Ithaca [NY], United States 3.0 [44] was used to calculate the genetic relatedness among all pairs of individuals (kinship matrix).

2.5. Bayesian Association Mapping and Pleiotropy

The residuals from a general linear model for each trait were used as the adjusted phenotypes for the Bayesian association analysis (taking into account the effects of block and genetic structure). This approach is used to reduce the sources of unexplained variability, e.g., the effects of block and genetic structure, which implies a reduction in the error rate and increase the statistical power of association study. The probability that a locus is associated with a given trait was evaluated by the Bayes Factor (BF) and the Posterior Probability of Association (PPA) [45]. The BF is the ratio of marginal likelihoods between the probabilities of the model of association (M1) and a null model of no association (M0) [46]:

$$BF = \frac{P(Data|M_1)}{P(Data|M_0)},\tag{5}$$

where the marginal likelihoods for *M*1 and *M*0 are defined by:

$$P(Data|M_l) = \int \left(\prod_{i=1}^{N} \sum_{k=0}^{2} P(\Phi|G_{ij} = k, \theta) p_{ijk} \right) P(\theta|M_l) d\theta,\tag{6}$$

in which θ denotes the regression parameters, G_i denotes the genotype of the *i*th individual at the *j*th marker, Φ denotes the phenotype of the *i*th individual in a study of *N* samples, and $p_{ijk} = P(G_{ij} = k)$

190

is the probability that the genotype at the *j*th marker of the *i*th individual is k [46]. The Laplace approximation was used to estimate the marginal likelihoods of M1 and M0 [46].

The PPA combines the evidence in the observed association data (BF) with the prior probability (π) that a marker is associated. The π and BF are used to compute the posterior odds (PO):

$$PO = BF \times \frac{\pi}{1 - \pi}, \tag{7}$$

where π is the prior probability that a given marker is associated with a trait. BF was calculated using a Bayesian regression analysis in the SNPTEST software [46]. The PPA was calculated from this PO, as follows:

$$PPA = \frac{PO}{1 + PO}. \tag{8}$$

The possible pleiotropic effect of one loci associated with more of one trait was corroborated by a multivariate Bayesian regression in SNPTEST [46]. Additionally, the association mapping was performed in TASSEL 3.0 software [44], employing the unified mixed linear model (Q + kinship) [47].

3. Results

3.1. Genetic Parameters

The trees from Wirrabara had the highest average depth of penetration of pilodyn at 16.8 mm, in contrast to the individuals from Cowell (13.0 mm), indicating that the latter has a denser wood. In contrast, the mean slenderness coefficient varied between 83.4 and 109.3 for the trees from Wirrabara and Cowell, respectively. Regarding regions, the trees from the Flinders Mountain Range had the highest slenderness coefficient and the lowest wood density, while the individuals of the Kangaroo Island and the Eyre Peninsula presented the highest wood density and the lowest coefficient of slenderness (Table 1).

Table 1. Population means for wood density and slenderness coefficient in the populations of *E. cladocalyx*.

Provenance	Region	N_f	N_t	Wood Density	Slenderness
Wirrabara	Flinders Ranges	9	45	16.8 a	83.4 b
Mount Remarkable	Flinders Ranges	16	80	16.4 a	86.2 b
Illapel	Coquimbo	2	10	16.1 a	92.0 ab
Flinders chase	Kangaroo Island	8	40	13.8 b	101.3 a
Marble Range	Eyre Peninsula	4	20	13.6 b	87.3 ab
Cowell	Eyre Peninsula	10	50	13.0 b	109.3 a

N_f: total number of families evaluated; N_t: total number of trees evaluated. Mean values with the same letter in the same column indicate that the populations are not significantly different according to the Tukey-Kramer test ($p < 0.01$).

Based on subsequent marginal distributions, the density of the wood in *E. cladocalyx* showed a moderate heritability (mean of $\hat{h}_a^2 = 0.29$), with a credibility range (90%) of 0.06 to 0.74; while the heritability for the slenderness coefficient was high $\hat{h}_a^2 = 0.48$ (credibility region from 0.11 to 0.87). The posterior mean of the genetic correlation between wood density and the slenderness coefficient was negative with a value of 0.3 and a credibility range of -0.49 to -0.11. The estimates of heritability via REML were similar to those obtained by the Bayesian method, with values of 0.30 and 0.45, for wood density and slenderness coefficient, respectively. The genetic correlation between both traits was moderate and negative (-0.34). The estimations of variance components, the heritability values and their genetic correlation, based on the point estimates (mean, mode, and median) of the posterior marginal distributions (and from the REML method), are shown in Table 2.

191

Table 2. Variance components, heritability for wood density (WD) and slenderness coefficient (SC), and the genetic correlation between the traits in the population of *E. cladocalyx*.

Trait/Estimates	Bayesian Estimates						REML Estimates
	Mean	Median	Mode *	SD	90% Credible Sets		
					Lower	Upper	
Wood density							
$\hat{\sigma}_a^2$	1.7	1.3	0.6	1.3	0.3	4.5	1.7
$\hat{\sigma}_\varepsilon^2$	4.0	4.2	4.5	1.2	1.6	5.7	4.0
$\hat{h}_a^2$	0.29	0.23	0.12	0.21	0.06	0.74	0.30
CVa (%)	7.9	7.4	6.2	3.1	3.7	13.7	8.5
Slenderness coefficient							
$\hat{\sigma}_a^2$	164.0	155.0	86.1	88.1	37.7	317.0	152.6
$\hat{\sigma}_\varepsilon^2$	174.3	175.2	205.0	79.8	46.6	301.3	183.8
$\hat{h}_a^2$	0.48	0.47	0.34	0.24	0.11	0.87	0.45
CVa (%)	13.4	13.6	14.6	4.0	6.7	19.4	13.5
Genetic correlation	−0.30	−0.30	−0.30	0.12	−0.49	−0.11	−0.34

$\hat{\sigma}_a^2$: additive genetic variance, $\hat{\sigma}_\varepsilon^2$: residual variance, $\hat{h}_a^2$: heritability, *CVa*: additive coefficient of variation, SD: standard deviation. * Kernel density estimates of the mode from marginal posterior distributions.

3.2. Genetic Structure

The trees from the five Australian locations evaluated in this study were grouped into three genetically differentiated groups, which coincide with the three geographical regions from which each population originates. Additionally, individuals from Illapel (local population) were grouped mainly with individuals from the Flinders Mountain Range (Figure 1). Consistent with the above, the mean Fst value of a given cluster is not contained within the credibility regions of the other groups, which confirms the significant genetic differentiation (Table 3).
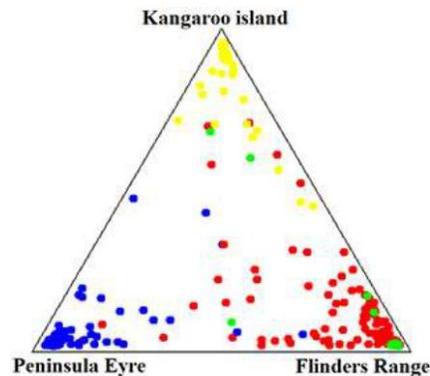

Figure 1. Bayesian clustering approach showing three groups genetically differentiated, according to the geographical origin of *E. cladocalyx* in Australia. The blue dots correspond to the populations from Marble Range and Cowell, the yellow dots correspond to the population from Flinders Chase, the red dots correspond to the populations from Wirrabara and Remarkable, and the green dots correspond to the local population taken from Illapel, Chile.

Table 3. Bayesian point estimation (mode, median, and mean) and 90% credible regions of *Fst* calculated in each group genetically differentiated of *E. cladocalyx*.

Cluster	Individuals	Mean	Median	Mode *	SD	90% Credible Sets	
						Lower	Upper
I	48	0.117	0.116	0.115	0.016	0.092	0.144
II	72	0.072	0.071	0.070	0.015	0.049	0.098
III	125	0.184	0.184	0.182	0.025	0.144	0.217

* Kernel density estimates of the mode from marginal posterior distributions.

3.3. Bayesian Association Mapping and Pleiotropy

A total of 13 putative associations, comprising 9 SSR markers, were identified for wood density and slenderness coefficient by the association mapping based on a unified mixed model (frequentist) (Table 4) [47]. Six loci were associated with WD, where the loci *qtl*WD130 located on chromosome 4 at 52.1 cM, explained the highest proportion of the phenotypic variance (17%). Similarly, seven markers were associated with SC, and the greater proportion of the phenotypic variance was explained by the loci *qtl*SR56 (21%). Four of these loci showed significant associations with WD and SC concomitantly. This suggests a common genetic control between the traits, which could result in possible pleiotropic effects for these loci. On the other hand, less loci were identified in the Bayesian association mapping (four and three loci associated with WD and SC, respectively), which presented a moderate (3 < BF <10) to extremely strong (100 < BF) evidence of association (model M1) (Figure 2 and Table 4), according to the scale presented by Andraszewicz et al. [48]. The posterior likelihood of association (PPA) indicated that loci with moderate evidence of association (3 < BF <10) have a PPA of less than 0.35, while loci with strong evidence of association (10 < BF < 30) showed a PPA between 0.35 and 0.61. Furthermore, loci with extremely strong evidence (100 > BF) have PPA values greater than 0.84 (Table 4).
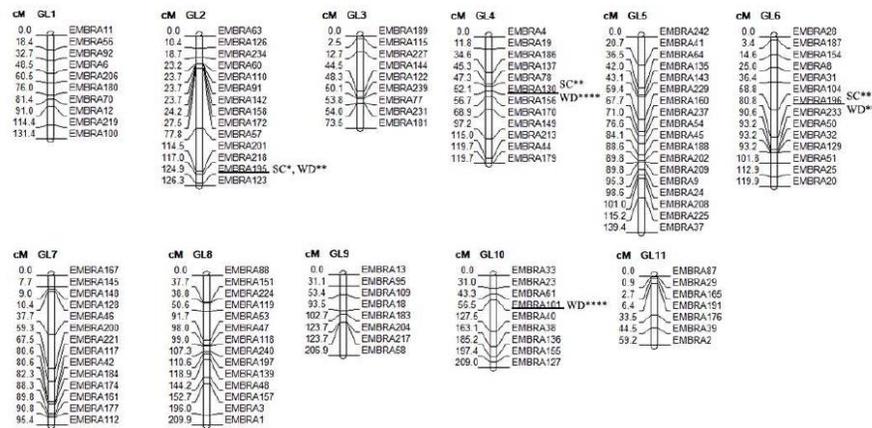

Figure 2. Loci associated with wood density (WD) and slenderness coefficient (SC) across the genome of *E. cladocalyx* (Bayesian multivariate mapping). To the left the distance in centiMorgan (cM) and to the right the name of the molecular markers (SSR, n = 130). LG: Linkage Group. * moderate evidence of association (3 < BF < 10); ** strong evidence of association (10 < BF < 30); ** very strong evidence of association (30 < BF < 100). **** extreme evidence of association (100 < BF). BF: Bayes Factor.

193

Table 4. Summary of loci associated with wood density (WD) and slenderness coefficient (SC) using a unified mixed linear model and Bayesian regression analysis, measured in 17-year-old trees of *E. cladocalyx* in arid conditions.

Traits	Locus	LG	Position (cM)	PV (%)	BF	PO	PPA
WD	*qtl*WD101 *	10	56.5	16.4	1,525,001	80,263	1.000
	*qtl*WD130 *	4	52.1	16.7	1,055	55.55	0.982
	*qtl*WD18	9	93.5	5.2	2.10	0.110	0.099
	*qtl*WD195 *	2	124.9	7.6	15.90	0.837	0.456
	*qtl*WD50	6	93.2	14.5	1.16	0.061	0.057
	*qtl*WD196 *	6	80.8	8.0	15.80	0.832	0.454
SC	*qtl*SC50	6	93.2	18.2	0.91	0.048	0.046
	*qtl*SC195 *	2	124.9	9.2	8.11	0.427	0.299
	*qtl*SC161	7	89.8	5.3	1.26	0.066	0.062
	*qtl*SC56	1	18.4	20.6	0.59	0.031	0.030
	*qtl*SC196 *	6	80.8	7.5	17.53	0.923	0.480
	*qtl*SC130 *	4	52.1	12.8	24.17	1.272	0.560
	*qtl*SC218	2	117	6.1	0.97	0.051	0.048
Pleiotropic loci	*qtl*WD130/*qtl*SC130 *	4	52.1	16.7/12.8	1911	100.58	0.990
	*qtl*WD195/*qtl*SC195 *	2	124.9	7.6/9.2	15.44	0.812	0.448
	*qtl*WD/*qtl*SC50	6	93.2	14.5/18.2	1.30	0.068	0.064
	*qtl*WD196/*qtl*SC196 *	6	80.8	8.0/7.5	24.01	1.264	0.558

PV: Proportion of phenotypic variance explained by each QTL. LG: Linkage group number. BF: Bayes Factor. PO: Posterior Odds. PPA: Posterior Probability of Association. * Loci identified through Bayesian regression analysis.

Three QTL associated with both traits were detected through a multivariate Bayesian analysis [46] on LG2 at 125 cM, on LG4 at 52 cM, and on LG6 at 81 cM. The bi-trait QTL analysis indicated that these loci presented a BF > 15, indicating that the evidence of association is between the categories of strong and extremely strong association. Furthermore, these loci have a PPA of 0.45 to 0.99, indicating a high probability of association (Table 4).

4. Discussion

4.1. Phenotypic Variability and Genetic Parameters

According to the results, wood density (WD) and the slenderness coefficient (SC) varied significantly among the provenances of *E. cladocalyx* evaluated in this study. The morphological variability of *E. cladocalyx* (such as stem straightness, flowering intensity, diameter and tree height) has been previously reported in other studies [49–51]. For example, Mora et al. [49] and Vargas-Reeve et al. [51] determined that trees from the Flinders Range have a diameter of the stem larger than trees from other natural regions of the species. In general, a larger diameter correlates negatively with SC in trees, in which it has been observed that trees with a greater diameter have a lower SC [28,52,53]. Consistently, the trees from Wirrabara and Mount Remarkable showed the lowest SC values, while the trees from Cowell were the slenderest. A high SC value indicates that trees are more susceptible to damage by mechanical pressures [29]. According to Navratil [54], trees with a SC greater than 100 have a high probability of suffering from wind damage. In this sense, the individuals of the Flinders Mountain Range would present a greater mechanical resistance than the trees from the Eyre Peninsula. Interestingly, the Flinders Chase, Cowell and Marble Range populations had the lowest penetration values of pilodyn (WD), which indicates that these trees have a higher wood density than those from the Flinders Range. It should be noted that Bush et al. [6] reported that trees from the Eyre peninsula have a lower wood density (based on the water displacement method, TAPPI [55]) than trees that come from the Flinders Mountain Range and Kangaroo Island. Moreover, these authors also demonstrated that there is a negative correlation between wood density and tree diameter in *E. cladocalyx*. The variability observed among the populations for WD and SC, had similarities with the

natural distribution of the same. Mora et al. [49] indicate that in genetically differentiated groups a reduction of genetic variability is observed. This presents evidence that the location of origin of an individual influenced the expression of their traits. Moreover, in WD the populations belonging to Eyre Peninsula and Kangaroo Island regions did not present significant differences between them (Table 1), indicating a possible genetic flow between these regions that is not shared with the Flinders Range region.

According to REML and Bayesian estimates, the heritability for WD and SC was moderate and high, respectively, indicating that the phenotypic variability observed between provenances for these traits is explained by an important additive genetic component. Bush et al. [6] determined that wood density in *E. cladocalyx* presents a narrow-sense heritability of 0.39, which could be considered moderate to high, while in *E. globulus*, Stackpole et al. [25] determined that wood density is a highly heritable trait, with a value of $h^2 = 0.51$. Our results are subtly lower than in previous studies in *E. cladocalyx* and in other *Eucalyptus* species. In general, the values of heritability, both broad and narrow, are comparatively lower in environments with low water availability [56–58], which would confirm the differences between our results and previous studies.

The SC is not a trait usually evaluated in the genetic studies of *Eucalyptus*. However, it has been determined that this trait has moderate to high heritability in other species. For example, Pastorino et al. [33] determined that the heritability of SC in cypress is 0.15, while Chaendaekattu and Mydin [34] reported a heritability for SC between 0.25 and 0.43 in trees of *Hevea brasiliensis*. It should be noted that the heritability for stem diameter and the height in trees of *E. cladocalyx*, variables related to SC, is low to moderate [49,59], however, the results of the present study suggest that SC is a more inheritable trait than height and diameter in *E. cladocalyx*.

The WD and SC presented a negative genetic correlation. Slender shafts tend to be denser [60], therefore, the degree of penetration of Pilodyn in the stem is expected to be lower in slender trees. A greater Pilodyn penetration implies that the tree is less dense, therefore, the correlation between the density and slenderness of a tree is expected to be negative.

4.2. Genetic Structure

The population of *E. cladocalyx* evaluated in this study is composed of three genetically differentiated groups, which is consistent with the findings of McDonald et al. [61], Mora et al. [37], and Arriagada et al. [2]. The individuals were grouped according to the three geographical regions from which the evaluated populations were taken (Flinders Range, Eyre Peninsula, and Kangaroo Island). In addition, between the determined genetic groups, there is a significant and moderate differentiation according to the values of Fst, which is consistent with previous studies in *E. cladocalyx* [61–63].

The effect of the genetic structure of the population on the analysis of association and detection of QTLs has been widely discussed [64,65]. Particularly, the presence of a significant genetic structure, and inadequate information on kinship among individuals can generate false marker-trait associations [66]. For all QTLs analyses, a full model considering the genetic structure effects was contrasted with a null model not including these effects. The genetic structure effect was significant for all association analyses; therefore, it was used for the detection of QTLs in this study [47]. In several plant species, the effect of the structure and the degree of kinship in the association studies has been evaluated [65,67].

4.3. Bayesian Association Mapping and Pleiotropy

Wood density is one of the most important indicators of sawn wood quality and pulp yield [24]. Several studies have shown that density has a moderate to high genetic control in different species of *Eucalyptus* [25,26], which suggests that this trait could be subject to genetic improvement through selection assisted by molecular markers. In this context, a significant number of studies have been conducted to identify markers associated with wood density in *Eucalyptus*. For example, Bundock et al. [68] detected QTLs associated with wood density in *E. globulus*, on chromosomes 6 and 11, which explained between 8% and 12% of the phenotypic variation, respectively. Thumma et al. [19]

identified QTLs that explained between 3.6% and 5.6% of the variation in wood density of *E. nitens*, using RFLP and SSR markers. These findings are concordant with our results. Some of the QTLs detected in this study explained about 16% of the phenotypic variation of wood density in *E. cladocalyx* (*qtl*WD130 and *qtl*WD101), which can be used as selection criteria for higher trees for this phenotype.

In the present study, QTL *qtl*WD195, located on chromosome 2 at a distance of 124.9 cM, explained 7% of the wood density variation of *E. cladocalyx*. Similarly, in the same chromosome, Gion et al. [18] reported two QTLs, at 123 cM and 124 cM, associated with wood density in hybrids of *E. urophylla* and *E. grandis*, explaining up to 10% of the phenotypic variation. Similarly, the QTL *qtl*WD101, identified on chromosome 10 at 56.5 cM, is consistent with a region indicated on the same chromosome by Freeman et al. [69] in *E. globulus*, which was also stable in several populations and sites. Furthermore, in this same chromosomal region, Hamilton et al. [70] identified QTLs associated with acoustic wave velocity, an indirect measure of wood density in *E. globulus*. According to our results and previous studies, chromosome 10 could be strongly involved in the genetic control of wood density in *E. cladocalyx* and other *Eucalyptus* species.

The SC has been widely used to measure the stability and resistance of the tree to the mechanical pressures of the environment. Moreover, like wood density, the slenderness coefficient is used to evaluate the potential of a tree for the production of paper and sawn wood [28,71]. Another important finding of the bi-trait Bayesian analysis was the existence of three large-effect pleiotropic QTLs located on LG4 at 52 cM, on LG2 at 125 cM, and on LG6 at 81 cM, indicating that both traits are genetically related.

The detection of QTL for the SC has been poorly addressed. However, the results of the present study can be contrasted with the mapping of previously reported QTLs for tree diameter and height variables, since both traits determine the slenderness of a tree. Freeman et al. [69] detected a QTL on chromosome 4 associated with the diameter at breast height (DBH) in *E. globulus*, whose position is close to *qtl*SC130, reported in the present study.

According to Stephens and Balding [45], the BF is comparable to the likelihood ratio, but compares two different models instead of two parameter values in a model. BF is often used as an indicator of the evidence of association of a locus to a trait, in which the moderate or strong categories are considered as definitive evidence against the null model (M0). According to this principle, PPA has been used as a probabilistic measure of posterior evidence, which combines the previous probability of association (π) and the evidence of association in the observed data (BF). Because π is small, the BF must be large to provide convincing evidence of an association.

We determined that a locus with a BF > 8 delivers a moderate evidence to determine that the hypothesis of association with the given trait is more likely than the null hypothesis of non-association. The unified mixed model indicated that 13 loci showed a significant association with traits. However, only six of these loci presented strong evidence of association (BF > 15 and PPA > 0.44) through the Bayesian regression analysis. Additionally, the Bayesian bi-trait regression (pleiotropy analysis) indicated that three of the four loci identified through the unified mixed model presented strong evidence of association with WD and SC. This provides convincing evidence that the loci *qtl*WD130/*qtl*SC130, *qtl*WD195/*qtl*SC195 and *qtl*WD196/*qtl*SC196, have a significant pleiotropic effect, being associated with both traits under study.

5. Conclusions

E. cladocalyx has a recognized potential to produce structural timber under arid environmental conditions. According to our results, wood density and the slenderness of the trees of *E. cladocalyx* presented a moderate and high genetic control, respectively. Several of the associations found for wood density were located in chromosomal regions previously reported for this trait in other *Eucalyptus* species, revealing collinearity of QTLs. In the context of marker-assisted selection, the QTLs associated with WD (4) and SC (2) can be used as selection criteria in breeding programs for the species under arid conditions. The QTLs identified in this study explained 48% and 20% of the observed genetic variance

in WD and SC, respectively. This may be an indication that the coverage of markers in insufficient. Thus, it would be appropriate to include a greater amount of markers in future studies.

Author Contributions: Conceptualization, F.M., O.A., C.E.V., R.B. and C.M.; Data curation, P.B. and C.E.V.; Formal analysis, G.S.M. and C.M.; Funding acquisition, F.M.; Methodology, O.A., C.M., and F.M.; Project administration, F.M.; Resources, O.A. and F.M.; Supervision, C.M., P.B., R.B. and F.M.; Writing—original draft, O.A. and C.E.V.; Writing—review & editing, P.B., C.M. and F.M.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bush, D.; Kain, D.; Kanowski, P.; Matheson, C. Genetic parameter estimates informed by a marker-based pedigree: A case study with *Eucalyptus cladocalyx* in southern Australia. *Tree Genet. Genomes* **2015**, *11*, 798. [CrossRef]
2. Arriagada, O.; Amaral Junior, A.T.; Mora, F. Thirteen years under arid conditions: Exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breed. Sci.* **2018**, *68*, 367–374. [CrossRef] [PubMed]
3. McMahon, L.; Brendan, G.; Hean, R. Eucalyptus cladocalyx. *Primefact* **2010**, *1077*, 1–5.
4. Clarke, B.; Mcleod, I.; Vercoe, T. *Trees for farm Forestry: 22 Promising Species*; RIRDC Press: Canberra, Australia, 2009.
5. Lundqvist, S.O.; Grahn, T.; Olsson, L.; Seifert, T. Comparison of wood, fibre and vessel properties of drought-tolerant eucalypts in South Africa. *Southern For.* **2017**, *79*, 215–225. [CrossRef]
6. Bush, D.; McCarthy, K.; Meder, R. Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). *Ann. For. Sci.* **2011**, *68*, 1057–1066. [CrossRef]
7. Wessels, C.B.; Crafford, P.L.; Du Toit, B.; Grahn, T.; Johansson, M.; Lundqvist, S.O.; Säll, H.; Seifert, T. Variation in physical and mechanical properties from three drought tolerant *Eucalyptus* species grown on the dry west coast of Southern Africa. *Eur. J. Wood Wood Prod.* **2016**, *74*, 563–575. [CrossRef]
8. Li, Y.; Apiolaza, L.A.; Altaner, C. Genetic variation in heartwood properties and growth traits of *Eucalyptus bosistoana*. *Eur. J. For. Res.* **2018**, *137*, 565–572. [CrossRef]
9. Khan, B.I.; Solo-Gabriele, H.M.; Townsend, T.G.; Cai, Y. Release of arsenic to the environment from CCA-treated wood. 1. Leaching and speciation during service. *Environ. Sci. Technol.* **2006**, *40*, 988–993. [CrossRef]
10. Thistlethwaite, F.R.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Stoehr, M.U.; El-Kassaby, Y.A. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genomics* **2017**, *18*, 930. [CrossRef]
11. Sundararaj, R.; Shanbhag, R.R.; Nagaveni, H.C.; Vijayalakshmi, G. Natural durability of timbers under Indian environmental conditions—An overview. *Int. Biodeterior. Biodegrad.* **2015**, *103*, 196–214. [CrossRef]
12. Naidoo, S.; Zbonák, A.; Ahmed, F. The effect of moisture availability on wood density and vessel characteristics of *Eucalyptus grandis* in the warm temperate region of South Africa. In Proceedings of the 5th International Symposium on Wood Structure and Properties, Sielnica, Slovakia, 3–6 September 2006; pp. 117–122.
13. Nabais, C.; Hansen, J.K.; David-Schwartz, R.; Klisz, M.; López, R.; Rozenberg, P. The effect of climate on wood density: What provenance trials tell us? *For. Ecol. Manag.* **2018**, *408*, 148–156. [CrossRef]
14. Isik, F.; Mora, C.R.; Schimleck, L.R. Genetic variation in *Pinus taeda* wood properties predicted using non-destructive techniques. *Ann. For. Sci.* **2011**, *68*, 283–293. [CrossRef]
15. Fundova, I.; Funda, T.; Wu, H.X. Non-destructive wood density assessment of Scots pine (*Pinus sylvestris* L.) using Resistograph and Pilodyn. *PLoS ONE* **2018**, *13*, e0204518. [CrossRef] [PubMed]
16. Woodcock, D.W.; Shier, A.D. Does canopy position affect wood specific gravity in temperate forest trees. *Ann. Bot.* **2003**, *91*, 529–537. [CrossRef] [PubMed]

17. Chave, J.; Muller-Landau, H.C.; Baker, T.R.; Easdale, T.A.; ter Steege, H.; Webb, C.O. Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecol. Appl.* **2006**, *16*, 2356–2367. [CrossRef]

18. Gion, J.M.; Carouché, A.; Deweer, S.; Bedon, F.; Pichavant, F.; Charpentier, J.P.; Baillères, H.; Rozenberg, P.; Carocha, V.; Ognouabi, N.; et al. Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus. BMC Genomics* **2011**, *12*, 301. [CrossRef] [PubMed]

19. Thumma, B.R.; Baltunis, B.S.; Bell, J.C.; Emebiri, L.C.; Moran, G.F.; Southerton, S.G. Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. *Tree Genet. Genomes* **2010**, *6*, 877–889. [CrossRef]

20. Louzada, J.L. Genetic correlations between wood density components in Pinus pinaster Ait. *Ann. For. Sci.* **2003**, *60*, 285–294. [CrossRef]

21. Li, X.; Wu, H.X.; Southerton, S.G. Transcriptome profiling of Pinus radiata juvenile wood with contrasting stiffness identifies putative candidate genes involved in microfibril orientation and cell wall mechanics. *BMC Genomics* **2011**, *12*, 1–16. [CrossRef]

22. Southerton, S.G.; MacMillan, C.P.; Bell, J.C.; Bhuiyan, N.; Dowries, G.; Ravenwood, I.C.; Joyce, K.R.; Williams, D.; Thumma, B.R. Association of allelic variation in xylem genes with wood properties in Eucalyptus nitens. *Aust. For.* **2010**, *73*, 259–264. [CrossRef]

23. Kanberga-Silina, K.; Jansons, A.; Rungis, D. Expression of three phenylpropanoid pathway genes in Scots pine (*Pinus sylvestris* L.) in open-pollinated families with differing relative wood densities during early and late wood formation. *Silvae Genet.* **2015**, *64*, 148–159. [CrossRef]

24. Nakahama, K.; Urata, N.; Shinya, T.; Hayashi, K.; Nanto, K.; Rosa, A.C.; Kawaoka, A. RNA-seq analysis of lignocellulose-related genes in hybrid *Eucalyptus* with contrasting wood basic density. *BMC Plant Biol.* **2018**, *18*, 156. [CrossRef] [PubMed]

25. Stackpole, D.J.; Vaillancourt, R.E.; Alves, A.; Rodrigues, J.; Potts, B.M. Genetic variation in the chemical components of Eucalyptus globulus wood. *G3-Genes Genome Genet.* **2011**, *1*, 151–159. [CrossRef] [PubMed]

26. Stackpole, D.J.; Vaillancourt, R.E.; de Aguigar, M.; Potts, B.M. Age trends in genetic parameters for growth and wood density in Eucalyptus globulus. *Tree Genet. Genomes* **2010**, *6*, 179–193. [CrossRef]

27. Fournier, M.; Baillères, H.; Chanson, B. Tree biomechanics: Growth, cumulative prestresses and re-orientations. *Biomimetics* **1994**, *2*, 229–251.

28. Jullien, D.; Widmann, R.; Loup, C.; Thibaut, B. Relationship between tree morphology and growth stress in mature European beech stands. *Ann. For. Sci.* **2013**, *70*, 133–142. [CrossRef]

29. Eguakun, F.S.; Oyebade, B.A. Linear and nonlinear slenderness coefficient models for *Pinus caribaea* (Morelet) stands in Southwestern Nigeria. *J. Agri. Vet. Sci.* **2015**, *8*, 26–30.

30. Rust, S. Analysis of regional variation of height growth and slenderness in populations of six urban tree species using a quantile regression approach. *Urban For. Urban Green.* **2014**, *13*, 336–343. [CrossRef]

31. Harja, D.; Vincent, G.; Mulia, R.; van Noordwijk, M. Tree shape plasticity in relation to crown exposure. *Trees-Struct. Funct.* **2012**, *26*, 1275–1285. [CrossRef]

32. Watt, M.S.; Kirschbaum, U. Moving beyond simple linear allometric relationships between tree height and diameter. *Ecol. Model.* **2011**, *222*, 3910–3916. [CrossRef]

33. Pastorino, M.J.; Ghirardi, S.; Grosfeld, J.; Gallo, L.A.; Puntieri, J.G. Genetic variation in architectural seedling traits of Patagonian cypress natural populations from the extremes of a precipitation range. *Ann. For. Sci.* **2010**, *67*, 508. [CrossRef]

34. Chaendaekattu, N.; Mydin, K.K. Inheritance pattern and genetic correlations among growth and wood quality traits in Para rubber tree (*Hevea brasiliensis*) and implications for breeding. *Tree Genet. Genomes* **2018**, *14*, 63. [CrossRef]

35. Ducey, M.J. Evergreenness and wood density predict height–diameter scaling in trees of the northeastern United States. *For. Eecol. Manag.* **2012**, *279*, 21–26. [CrossRef]

36. Silva, M.N.D. Extraction of genomic DNA from leaf tissues of mature native species of the cerrado. *Rev. Árvore* **2010**, *34*, 973–978. [CrossRef]

37. Mora, F.; Arriagada, O.; Ballesta, P.; Ruiz, E. Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *J. Plant Biochem. Biotechnol.* **2017**, *26*, 274–281. [CrossRef]

38. Benomar, L.; DesRochers, A.; Larocque, G.R. The effects of spacing on growth, morphology and biomass production and allocation in two hybrid poplar clones growing in the boreal region of Canada. *Trees* **2012**, *26*, 939–949. [CrossRef]

39. Van Tassell, C.P.; Van Vleck, L.D. Multiple-trait Gibbs sampler for animal models: Flexible programs for Bayesian and likelihood-based (co) variance component inference. *J. Anim. Sci.* **1996**, *74*, 2586–2597. [CrossRef]

40. Gilmour, A.R.; Thompson, R.; Cullis, B.R.; Welham, S.J. ASReml estimates variance matrices from multivariate data using the animal model. In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, 19–23 August 2002; Volume 28, pp. 1–2.

41. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014; Available online: http://www.R-project.org/ (accessed on 10 December 2018).

42. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.

43. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]

44. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

45. Stephens, M.; Balding, D.J. Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* **2009**, *10*, 681–690. [CrossRef] [PubMed]

46. Marchini, J.; Band, G. SNPTEST. 2016. Available online: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html (accessed on 15 December 2018).

47. Yu, J.; Pressoir, G.; Briggs, W.H.; Bi, I.V.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [CrossRef] [PubMed]

48. Andraszewicz, S.; Scheibehenne, B.; Rieskamp, J.; Grasman, R.; Verhagen, J.; Wagenmakers, E.J. An introduction to Bayesian hypothesis testing for management research. *J. Manag.* **2015**, *41*, 521–543. [CrossRef]

49. Mora, F.; Gleadow, R.; Perret, S.; Scapim, C.A. Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. *Euphytica* **2009**, *169*, 335–344. [CrossRef]

50. Cané-Retamales, C.; Mora, F.; Vargas-Reeve, F.; Perret, S.; Contreras-Soto, R. Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. *Euphytica* **2011**, *178*, 177–183. [CrossRef]

51. Vargas-Reeve, F.; Mora, F.; Perret, S.; Scapim, C.A. Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. *Crop Breed Appl. Biotechnol.* **2013**, *13*, 107–112. [CrossRef]

52. Díaz-Bravo, S.; Espinosa, M.; Valenzuela, L.; Cancino, J.; Lasserre, J.P. Efecto del raleo en el crecimiento y algunas propiedades de la madera de *Eucalyptus nitens* en una plantación de 15 años. *Maderas-Cienc. Tecnol.* **2012**, *14*, 373–388. [CrossRef]

53. Hallinger, M.; Johansson, V.; Schmalholz, M.; Sjöberg, S.; Ranius, T. Factors driving tree mortality in retained forest fragments. *For. Ecol. Manag.* **2016**, *368*, 163–172. [CrossRef]

54. Navratil, S. Silvicultural systems for managing deciduous and mixedwood stands with white spruce understory. In *Silvicultural of Temperate and Boreal Broadleaf-Conifer Mixture*; Comeau, P.G., Thomas, K.D., Eds.; Ministry of Forests: Victoria, BC, USA, 1996; pp. 35–46.

55. TAPPI. Basic density and moisture content of pulpwood. *TAPPI* **1989**, *258*, 98.

56. Hung, T.D.; Brawner, J.T.; Meder, R.; Lee, D.J.; Southerton, S.; Thinh, H.H.; Dieters, M.J. Estimates of genetic parameters for growth and wood properties in *Eucalyptus pellita* F. Muell. to support tree breeding in Vietnam. *Ann. For. Sci.* **2015**, *72*, 205–217. [CrossRef]

57. Mohammadi, R.; Sadeghzadeh, D.; Armion, M.; Amri, A. Evaluation of durum wheat experimental lines under different climate and water regime conditions of Iran. *Crop Pasture Sci.* **2011**, *62*, 137–151. [CrossRef]

58. Ramírez-Valiente, J.A.; Valladares, F.; Huertas, A.D.; Granados, S.; Aranda, I. Factors affecting cork oak growth under dry conditions: Local adaptation and contrasting additive genetic variance within populations. *Tree Genet. Genomes* **2011**, *7*, 285–295. [CrossRef]

59. Bush, D.; Kain, D.; Matheson, C.; Kanowski, P. Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters—An example using *Eucalyptus cladocalyx*. *Tree Genet. Genomes* **2011**, *7*, 23–35. [CrossRef]

60. Lenz, P.; Auty, D.; Achim, A.; Beaulieu, J.; Mackay, J. Genetic improvement of white spruce mechanical wood traits-early screening by means of acoustic velocity. *Forests* **2013**, *4*, 575–594. [CrossRef]

61. McDonald, M.W.; Rawlins, M.; Butchet, P.A.; Bell, J.C. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Aust. J. Bot.* **2003**, *51*, 393–403. [CrossRef]

62. Ballesta, P.; Mora, F.; Contreras-Soto, R.I.; Ruiz, E.; Perret, S. Analysis of the genetic diversity of *Eucalyptus cladocalyx* (sugar gum) using ISSR markers. *Acta Sci-Agron.* **2015**, *37*, 133–140. [CrossRef]

63. Bush, D.; Thumma, B. Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genet. Genomes* **2013**, *9*, 741–752. [CrossRef]

64. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 29. [CrossRef]

65. Cappa, E.P.; El-Kassaby, Y.A.; Garcia, M.N.; Acuña, C.; Borralho, N.M.; Grattapaglia, D.; Poltri, S.N.M. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. *PLoS ONE* **2013**, *8*, e81267. [CrossRef]

66. Li, M.; Liu, X.; Bradbury, P.; Yu, J.; Zhang, Y.M.; Todhunter, R.J.; Buckler, E.S.; Zhang, Z. Enrichment of statistical power for genome-wide association studies. *BMC Biol.* **2014**, *12*, 73. [CrossRef]

67. Uchiyama, K.; Iwata, H.; Moriguchi, Y.; Ujino-Ihara, T.; Ueno, S.; Taguchi, Y.; Tsubomura, M.; Mishima, K.; Iki, T.; Watanabe, A.; et al. Demonstration of genome-wide association studies for identifying markers for wood property and male strobili traits in *Cryptomeria japonica*. *PLoS ONE* **2013**, *8*, e79866. [CrossRef] [PubMed]

68. Bundock, P.C.; Potts, B.M.; Vaillancourt, R.E. Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*. *Tree Genet. Genomes* **2008**, *4*, 85–95. [CrossRef]

69. Freeman, J.S.; Potts, B.M.; Downes, G.M.; Pilbeam, D.; Thavamanikumar, S.; Vaillancourt, R.E. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytol.* **2013**, *198*, 1121–1134. [CrossRef] [PubMed]

70. Hamilton, M.G.; Freeman, J.S.; Blackburn, D.P.; Downes, G.M.; Pilbeam, D.J.; Potts, B.M. Independent lines of evidence of a genetic relationship between acoustic wave velocity and kraft pulp yield in *Eucalyptus globulus*. *Ann. For. Sci.* **2017**, *74*, 17. [CrossRef]

71. Pirralho, M.; Flores, D.; Sousa, V.B.; Quilhó, T.; Knapic, S.; Pereira, H. Evaluation on paper making potential of nine *Eucalyptus* species based on wood anatomical features. *Ind. Crop Prod.* **2014**, *54*, 327–334. [CrossRef]

SCIENTIA AGRICOLA

Genetics and Plant Breeding

Association mapping of drought tolerance indices in wheat: QTL-rich regions on chromosome 4A

Paulina Ballesta[1], Freddy Mora[1]*, Alejandro Del Pozo[2]

[1]Universidad de Talca/Instituto de Ciencias Biológicas, 2 Norte 685 – 3460000 – Talca – Chile.
[2]Universidad de Talca/Facultad de Ciencias Agrarias/Centro de Mejoramiento Genético y Fenómica Vegetal.
*Corresponding author <morapoblete@gmail.com>

ABSTRACT: Drought is likely the main abiotic stress that affects wheat yield. The identification of drought-tolerant genotypes represents an effective way of dealing with the continuous decrease in water resources as well as the increase in world population. The aim of this study was to identify single nucleotide polymorphisms (SNP) associated with drought tolerance indices in wheat by using a genome-wide association study (GWAS) under fully irrigated and rain-fed conditions. The drought tolerance indices (i.e., Stress Susceptibility Index, Stress Tolerance Index, Tolerance Index and Yield Stability Index) were calculated based on grain yield, 1,000-kernel weight and kernels per spike. The association panel was genotyped using genotyping-by-sequencing (GBS). A total of 175 SNPs exhibited statistical evidence of association with at least one drought tolerance index, explaining up to 6 % of the phenotypic variation. Forty-five SNPs were associated with more than one tolerance index (up to 4 agronomic traits). Most associations were located on chromosome 4A, supporting the hypothesis that this chromosome has a key role in drought tolerance which should be exploited for wheat improvement. In addition, statistical analysis detected SNPs associated with tolerance indices in both growing seasons, providing information about genetic regions with stable effects under different environmental conditions. This GWAS experiment serves as one of the few studies on association mapping for drought tolerance indices in wheat, which could increase the efficiency of rain-fed and irrigated crop production.

Keywords: SNP markers drought, genome-wide association study, rain-fed conditions

Introduction

Wheat (*Triticum aestivum*) is one of the most important crops in the maintaining of the security of the food supply and is the second most consumed cereal worldwide (Galetto et al., 2017; Franco et al., 2018; Oliveira and Pinto-Maglio, 2017). Drought is one of the main constraints affecting wheat production, and is found in virtually all climatic regions, providing a huge challenge to local farming in many countries worldwide (Lobell et al., 2011). However, the challenge posed by water deficit is not unbeatable. In fact, the negative effects of drought could be overcome by the identification and use of drought-tolerant varieties (Van Oosten et al., 2016). Given this solution, the dissection of molecular mechanisms that underlie adaptive traits represents one approach to understanding stress tolerance in plants (Budak et al., 2015; Liu et al., 2017; Arriagada et al., 2017). For example, Merchuk-Ovnat et al. (2016) showed that the introgression of QTLs on chromosomes 1B and 2B of *T. turgidum* into *T. aestivum* can enhance drought tolerance in domesticated wheat. On the other hand, a number of studies have proposed that chromosome 4A has an important role to play in drought tolerance (Alexander et al., 2012; Edae et al., 2014; Kumar et al., 2012; Nezhad et al., 2012). For instance, Edae et al. (2014) found chromosome regions (on 4A) that were associated with drought tolerance related traits such as the drought susceptibility index, leaf senescence, green leaf area and flag leaf traits. Among QTLs detected for

drought susceptibility index, one QTL was found in the same region (on chromosome 4A) for yield-related traits. Kumar et al. (2012) detected a QTL (QGyp.ksu-4A, in spring wheat) for grain yield under drought stress on chromosome 4A, which explained 16 % of the phenotypic variation. Moreover, Edae et al. (2014) found that the chromosomes of spring wheat showed substantial differences in the proportion of marker pairs in significant linkage disequilibrium (LD) from the maximum 62 % for chromosome 4A to the minimum 20 % for chromosome 5A; an important aspect of association studies. Despite the extreme complexity of the wheat genome, the development of molecular marker technology has enabled the discovery of SNP markers, which have increased the chances of identifying genomic regions and explain a quantitative trait on complex genomes (Poland et al., 2012; Sabiel et al., 2017). Additionally, high-density SNP arrays have been developed for economically important crops (Sim et al., 2012; Ps et al., 2017; Contreras-Soto et al., 2017a). Thus, the aim of this study was to identify SNP associated with drought tolerance as measured by different stress tolerance indices of key agronomic traits in wheat.

Materials and Methods

Plant material and field conditions

Cultivars and advanced lines (N = 382) obtained from breeding programs of the Agriculture Research Institute of Chile and Uruguay and the International

Wheat and Maize Improvement Centre (CIMMYT) were evaluated over two growing seasons (2011 and 2012) and tested in a Mediterranean and a humid environment in Chile: Cauquenes (35°58′ S, 72°17′ W; altitude: 518 m above sea level) and Santa Rosa (36°32′ S, 71°55′ W; altitude: 508 m above sea level). Cauquenes is a drought-prone area (rainfed conditions) with annual precipitation varying from 580 mm (2011) to 600 mm (2012), corresponding to the Mediterranean climate type (De Martonne index 20 ≤ IDM < 24) (Baltas, 2007; Croitoru et al., 2013). Santa Rosa exhibits a full irrigation condition with annual precipitation ranging from 736 mm (2011) to 806 mm (2012), corresponding to the humid climate type (De Martonne index 28 ≤ IDM < 35). The trials were arranged using an alpha-lattice experimental design with 20 incomplete blocks, each containing 20 genotypes. Santa Rosa was fully irrigated at the end of tillering (Zadoks stage 21, Zadoks et al., 1974), flag leaf (Z37), heading (Z50) and middle grain filling (Z70). Approximately 936 and 1,006 mm of water (total water supply) were applied in Santa Rosa in 2011 and 2012, respectively.

SNP genotyping and Linkage disequilibrium (LD)

The Genotyping by Sequencing (GBS) technique was employed to construct a library and SNP calling between samples as described by Poland et al. (2012). Genomic DNA was extracted using the DNeasy Plant Maxi Kit (Qiagen). The first step in constructing the library construction required the use of the *Pst*I-*Msp*I GBS protocol for wheat and barley genomes as described by Poland et al. (2012). Sequencing was carried out on an Illumina HiSeq 2000. The sequences were processed in Galaxy (http://galaxy.psu.edu/) to evaluate their quality and distribution in different samples. The Tassel Pipeline (http://maizegenetics.net) was used for SNP calling with modifications for non-reference SNP calling described by Poland et al. (2012). More details about this step are provided in Lado et al. (2013). Heterozygote data were eliminated from the SNP matrix using the inbreeding coefficient in the TASSEL software package (Lado et al., 2013; Song et al., 2015). In addition, alleles with a minor frequency of 0.01 were eliminated (minor allele frequency filter), yielding a total of 2,214 SNP markers.

Genome-wide LD was estimated by calculating r^2 values between all SNP pairs localized on the same chromosome (and genome) using the R package's LDheatmap (Shin et al., 2006), and plotted by the R package's corrplot (Wei et al., 2017). The Bonferroni correction test was performed to correct for multiple testing.

Phenotypic data analysis

Four drought stress indices of three agronomic traits, including grain yield (GY), 1,000-kernel weight (TKW) and kernels per spike (KS), were calculated to perform genome-wide association mapping. GY was determined by harvesting the entire plot, and TKW and KS in 25 spikes were obtained at random. The indices calculated for each trait were as follows: Stress Susceptibility Index (SSI), Stress Tolerance Index (STI), Tolerance Index (TOL) and Yield Stability Index (YSI). The indices were computed according to the following equations:

$$SSI = [1 - (Y_{si} / Y_{pi})]/SI$$

$$TOL = Y_{pi} - Y_{si}$$

$$YSI = Y_{si} / Y_{pi}$$

$$STI = [Y_{si} * Y_{pi}]/Y_p^2$$

where Y_{si} is the yield (trait) for each cultivar in stress condition; Y_{pi} the yield (trait) for each cultivar in normal or productive condition; SI the stress intensity: $SI = 1 - (Y_s/Y_p)$, where Y_s is the total yield (trait) mean in stress condition and Y_p the total yield mean in normal condition. A general linear model was used to evaluate the effect of variety on the indices calculated. The statistical model is described as follows:

$$y_{ijk} = \mu + G_i + S_j + (GS)_{ij} + e_{ijk}$$

where y_{ijk} is the index value (SSI, STI, YSI and TOL) of the *i*th genotype in the *j*th season (2011 or 2012), μ an intercept term, G_i the fixed effect of the *i*th genotype, S_j a fixed effect of the *j*th season, (GS) the effect of genotype × site interaction, and e_{ijk} the residual effect. Data were analyzed using PROC GLM in SAS (Statistical Analysis System, v. 9.2). The abbreviations of each index calculated from agronomic variables are presented in Table 1. In addition, an analysis of stable carbon isotope discrimination ($\Delta^{13}C$) was carried out to supply evidence of the physiological state of cultivars under full irrigation and rainfed conditions. Mature kernels were analyzed in an elemental analyzer coupled with an isotope ratio mass spectrometer.

Association mapping

Genome-wide association mapping was constructed assuming a structured model. Genetic structure anal-

Table 1 – Summary of abbreviations for each studied trait.

Index	Agronomic variable	Abbreviation
	Grain yield	GY-SSI
Stress susceptibility index	Kernels per spike	KS-SSI
	1000-kernel weight	TKW-SSI
	Grain yield	GY-STI
Stress Tolerance Index	Kernels per spike	KS-STI
	1000-kernel weight	TKW-STI
	Grain yield	GY-TOL
Tolerance Index	Kernels per spike	KS-TOL
	1000-kernel weight	TKW-TOL
	Grain yield	GY-YSI
Yield Stability Index	Kernels per spike	KS-YSI
	1000-kernel weight	TKW-YSI

ysis was carried out using the STRUCTURE software program (Pritchard et al., 2000) following Mora et al. (2015). Evanno's method was implemented to define the number of clusters (Evanno et al., 2005). A mixed linear model (MLM) was used to detect associations between SNP markers and stress indices. The assessment was carried out using the TASSEL software package (Bradbury et al., 2007) and the following equation:

$$y = X\beta + Qv + Z\mu + \varepsilon$$

where y is the vector of phenotypic observations (drought stress indices); β a vector of SNP marker effects; v a vector of population effects; μ a vector of random polygene background effects; and ε a vector of residual effects. X, Q and Z are incidence matrices relating y to β, v and μ, respectively. The variation of the μ vector was modeled as Var $(\mu) = 2K\sigma_g^2$, where K is the matrix of pairwise kinship coefficients and σ_g^2 the genetic variance (Yu et al., 2006). Correction for multiple comparisons was made using False Discovery Rate (FDR) analysis in SAS software.

Results and Discussion

According to statistical analyses of fixed effects, the genotype effect gave proof of statistical differences for all tolerance indices ($p < 0.01$), which means that there is an important genetic background which explains the phenotypic variation (in terms of drought tolerance indices). On the other hand, the environment effect (growing seasons 2011 and 2012) showed statistical differences ($p < 0.01$) for the majority of traits, except for all SSI indices and GY-TOL, and the G × S effect gave proof of statistical differences for all indices studied ($p < 0.01$). In accordance with these results, Farshadfar et al. (2012) gave proof that the environment and G × S interaction effects explained an important part of the total variation in tolerance indices (TOL, YSI, SSI and STI) in 16 genotypes of wheat evaluated under both rainfed and irrigated conditions for three years. On the other hand, an analysis of variance revealed that Δ¹³C showed statistical differences between full irrigation (Santa Rosa) and rainfed (Cauquenes) conditions ($p < 0.01$). Under rainfed conditions, the cultivars had lower Δ¹³C values, which is an indicator of better water-use efficiency (Barbou et al., 2010; Brienen et al., 2011). These results are consistent with the De Martonne Index.

Table 2 and Figure 1A and B show the SNP pairs in linkage disequilibrium (LD) per chromosome and genome, respectively. The majority of the SNPs were located on chromosomes of B genome (52 %), followed by the chromosomes of A genome (39 %). At the chromosome level, 7 % of total SNPs were located on chromosome 7B. Only one SNP was found on chromosome 4D. The LD analysis revealed that 29, 27 and 33 % of the SNP pairs on A, B and D genome, respectively, were in LD ($r^2 > 0.03$; $p < 0.05$). r^2 values ranged from 3.8×10^{-7} and

Table 2 – Percentage of SNP pairs in linkage disequilibrium (LD) per chromosome.

Chromosome	Number of total SNP	Number of total SNP pairs	SNP pairs in LD*
			%
1A	103	5253	33
2A	106	5565	29
3A	152	11476	22
4A	126	7875	36
5A	92	4186	27
6A	106	5565	24
7A	188	17578	32
1B	126	7875	27
2B	202	20301	36
3B	206	211115	2
4B	86	3655	35
5B	169	14196	25
6B	138	9453	36
7B	214	22791	30
1D	17	136	29
2D	26	325	34
3D	40	780	46
4D	1	-	0
5D	28	378	27
6D	63	1953	69
7D	25	300	25

*$p > 0.05$ after Bonferroni correction.

1 for all chromosomes. The most linked SNP pairs were located on chromosome 6D, and the most SNPs in high LD ($r^2 > 0.7$) on D genome. Consistent with our results, Mora et al. (2015) and Edae et al. (2014) also reported that LD extended over a longer genetic distance for the D genome than for the A and B genomes.

Genetic population structure analysis identified the presence of two genetically distinct subgroups (Figure 2). Cluster 1 contained 204 genotypes, while Cluster 2 included 178 genotypes. Associations were not detected after correcting for multiple comparisons (false discovery rate – FDR). However, a total of 219 associations were detected at $p < 0.005$, of which 175 SNPs (approximately 8 %) were associated with at least one trait, explaining between 2 % and 6 % of the total phenotypic variation (Table 3). Ninety-nine and 120 associations were detected in the growing seasons of 2011 and 2012, respectively. The difference in number of SNP associations detected between both growing seasons is in accordance with the G × S interaction found in this study (Heidari et al., 2011). However, nine SNPs (located on chromosomes 6D, 3B, 2B, 7D, 6B and 3D) associated with TKW-STI were detected in both growing seasons. Consistently, Mora et al. (2015) reported eight SNPs over the growing seasons associated with TKW in a non-stress site (irrigated site). Interestingly, Saeed et al. (2017) also detected SNPs associated with TKW-STI and GY-STI in more than one growing season in wheat. The early detection of QTL and evaluation of their stability
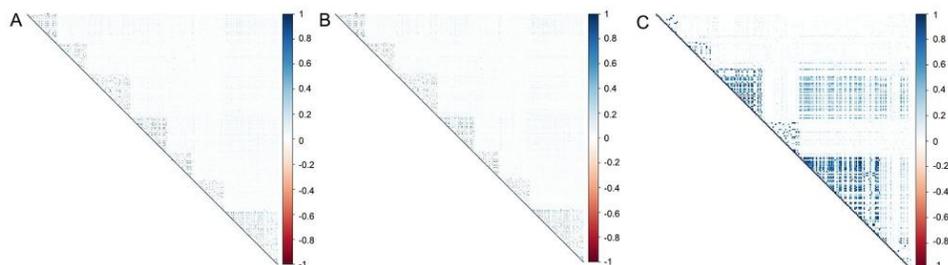
Figure 1 – Linkage disequilibrium (LD) among all SNP pairs calculated for each genome. A, B and C are LD plots for A, B and D genome of wheat, respectively.
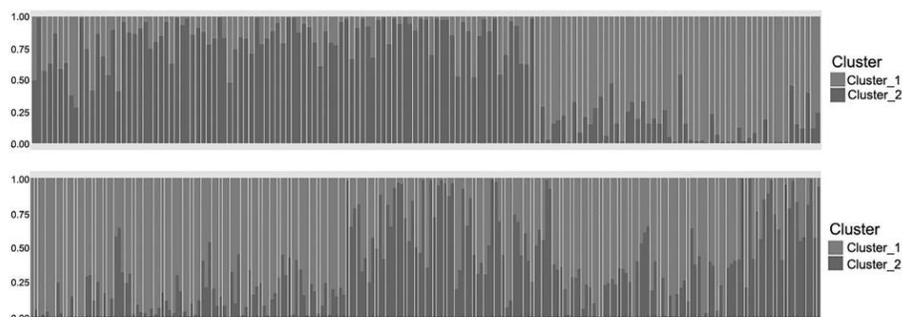

Figure 2 – Population structure of 382 wheat cultivars inferred using the model-based Bayesian algorithm implemented in STRUCTURE. The 382 genotypes are represented in x-axis and the cluster assignment for each cultivar in y-axis (K = 2).

across multiple environmental conditions would allow for the identification of candidate QTL for MAS (Sun et al., 2012). Multiple evaluations may minimize the risk of selecting genotypes that exhibit good performance in particular environments but not under several conditions (El-Soda et al., 2014).

The highest number of trait associations was established with SNP markers located on B genome (43 % of total associations). However, at the chromosome level, chromosome 4A was identified as the most repeated region for SNP-trait associations (26 associations). TKW-STI (evaluated in 2011) was the index with the highest number of associations (31 associations), whereas the fewest number of associations were detected for KS-SSI (one association) and KS-YSI (one association) in the same growing season.

Thirty-two marker-trait associations were detected for GY-SSI, TKW-SSI and KS-SSI and were located mainly on chromosomes 4A, 4B and 6B. In particular, 11 associations were detected for TKW-SSI (located mostly on chromosomes 4B, 7B and 6A), and 14 associations were detected for KS-SSI. These associations were predominantly located on chromosomes 4A and

6B. No SNPs located on genome D were associated with TKW-SSI and KS-SSI indices in both growing seasons. GY-SSI indices (2011 and 2012) were associated with SNPs located on the three genomes and involved seven chromosomes (1A, 2A, 5A, 6A, 2B, 5D and 6B), with one association for each chromosome. One hundred associations were detected for the STI indices, which were recurrent on chromosomes 6D, 4A and 6A. Fifty-six associations were identified for TKW-STI and were located preferentially on chromosomes 6D, 3B and 6A. Thirty-one associations were identified for KS-STI and were mainly located on chromosomes 4A, 6B and 6A. Thirteen associations were detected for GY-STI with SNPs frequently located on chromosome 2D. Forty-six associations were detected for TOL indices, and these associations were frequently noted on chromosomes 2B and 4A. In summary, nine associations were identified for TKW-TOL with SNPs preferentially located on chromosome 6A. Twenty-five associations were detected for KS-TOL and were mainly located on chromosomes 4A and 7A. Twelve associations were detected for GY-TOL with SNPs commonly located on chromosomes 3A and 2B. Finally, 41 associations were detected for the YSI

Table 3 – Chromosome location and number of associations ($p < 0.005$) for tolerance indices: SSI (Stress susceptibility index), STI (Stress Tolerance Index), TOL (Tolerance Index) and YSI (Yield Stability Index) evaluated in the growing seasons of 2011 and 2012.

Index/year	Chromosome (number of associations)	PV%
2011		
TKW-SSI	2A(1), 4B(3), 6B(1), 7B(1)	3-4
TKW-STI	1A(1), 1B(1), 2A(1), 2B(2), 3B(4), 3D(1), 4A(1), 5A(3), 6A(4), 6B(2), 6D(9), 7B(1), 7D(1)	2-4
TKW-TOL	1A(1), 1D(1), 2B(1), 5A(1), 6A(1), 6B(1)	3-5
TKW-YSI	2A(1), 4B(3), 6B(1), 7B(1)	3-4
KS-SSI	4A(1)	3
KS-STI	2B(2), 2D(2), 4A(5), 5B(1), 6A(2), 6B(6), 7A(3), 7B(1)	3-6
KS-TOL	3B(2), 4A(1), 6A(1), 7A(4)	3-4
KS-YSI	4A(1)	3
GY-SSI	2A(1), 5A(1), 6D(1)	3-5
GY-STI	2B(1), 2D(4), 5B(2)	3-4
GY-TOL	4A(1), 6D(2), 7B(2)	4-5
GY-YSI	2A(1), 5A(1), 6D(1)	3-5
2012		
TKW-SSI	3A(1), 3B(1), 6A(2), 7B(1)	3-4
TKW-STI	1B(1), 2B(2), 3A(1), 3B(3), 3D(1), 4A(3), 5B(1), 6A(3), 6B(2), 6D(4), 7B(3), 7D(1)	3-4
TKW-TOL	2B(1), 6A(2)	3-4
TKW-YSI	2A(1), 2B(1), 3A(2), 3B(1), 6D(2), 7B(1)	3-5
KS-SSI	3B(1), 4A(3), 5A(2), 6B(4), 7A(1), 7B(2)	3-5
KS-STI	2A(2), 4A(1), 5B(1), 6A(3), 7B(2)	3-4
KS-TOL	3B(1), 4A(5), 5A(2), 6B(4), 6D(2), 7A(1), 7B(2)	3-5
KS-YSI	2A(1), 2B(1), 4A(2), 5A(3), 6B(4), 7A(2), 7B(2)	3-5
GY-SSI	1A(1), 2B(1), 5D(1), 6A(1)	3-4
GY-STI	3A(1), 3B(1), 4A(2), 7B(2)	3-5
GY-TOL	2B(3), 3A(3), 3B(1)	3-4
GY-YSI	1B(1), 2B(3), 3A(1), 3B(1), 5D(1), 6A(1)	3-5

GY = grain yield; KS = kernels per spike; TKW = 1,000-kernel weight; PV% = the proportion of phenotypic variation explained by SNP markers (in %).

indices, which were recurrent on chromosomes 2B and 6B. Fourteen associations were identified for TKW-YSI with SNPs preferentially located on chromosomes 4B, 3A and 6D. Sixteen associations were detected for KS-YSI and were located mainly on chromosomes 5A and 6B. Eleven associations mainly located on chromosome 2B were identified for GY-YSI.

Few studies of association analyses using stress indices have been conducted on wheat. In fact, the studies available have been conducted using molecular markers, such as SSR (simple sequence repeat), DArT (Diversity Arrays Technology), AFLP (amplified fragment length polymorphism) and RFLP (restriction fragment length polymorphism). Dashti et al. (2007) detected AFLP, RFLP and SSR markers associated with the SSI, STI and TOL indices calculated from GY, explaining 21 %, 15 % and 36 % of the phenotypic variation, respectively. SNP markers associated with GY-SSI were located on chromosomes 7A, 4B and 5B. Associations for GY-STI and GY-TOL were located on chromosomes 1B and 5B, respectively. In addition, Dodig et al. (2012) reported SSR markers associated with GY-SSI and GY-TOL explaining between 15 % and 25 % of phenotypic variation. One SSR located on chromosome 2D was associated with GY-SSI and GY-TOL, and an exclusive SNP for GY-TOL was located on chromosome 2B. Saeed et al. (2017) dis-

covered SSR markers explaining between 7 % and 17 % of phenotypic variation for TKW-STI, TKW-SSI, GY-STI and GY-SSI (GY expressed as GY per plant). SSRs associated with TKW-STI were located on chromosomes 7D and 5D (two QTL detected in two growing seasons), and trait associations for TKW-SSI were detected on chromosomes 7D (one QTL detected in two growing seasons) and 3A. In addition, SSR markers associated with GY-STI were located on chromosome 7D (one QTL detected in two growing seasons), and GY-SSI associations were identified using SSRs located on chromosomes 6D and 7D (one QTL detected in two growing seasons). Kirigwi et al. (2007) reported SSRs associated with GY-SSI (in several growing seasons), which were located on chromosome 4A, explaining between 12 and 41 % of the total phenotypic variation. The authors proposed that chromosome 4A is a region in the wheat genome that contains markers associated with drought tolerance. Interestingly, Edae et al. (2014) reported one DArT marker on chromosome 4A, explaining 4 % of the phenotypic variation for GY-SSI. Consistently, in the present study, most of the SNP markers involved in marker-trait associations were located on chromosome 4A (26/129 associations; approximately 12 % of total associations), supporting the hypothesis that this chromosome plays a central role in the drought tolerance of wheat. On the other hand,

other studies involving QTL analyses have emphasized the importance of genome B in certain drought-related traits. Kumar et al. (2012) for instance, identified QTLs located on chromosomes 2B and 3B explaining up to 56 % and 60 % of the phenotypic variation of potential quantum efficiency of photosystem II (F_v/F_m) and chlorophyll content (Chl), respectively, which were positively correlated with GY-STI (r > 0.95). In the present study, 17 and 16 associations were located on chromosome 2B and 3B respectively. Particularly, two SNPs of these chromosomes were associated with GY-STI, obtaining the highest explained phenotypic variation for GY-STI. The B genome has been recognized by carrying loci controlling water-use efficiency and related traits, and grain yield under water stress conditions (Mohammady et al., 2012; Poersch-Bortolon et al., 2016), which could explain why 43 % of the total associations in the present study were located on the B genome.

There have been conflicting reports on the phenotypic variation explained by the markers (e.g., SSR, RFLP, AFLP and our results with SNP markers) for traits involved in drought stress indices. The findings reported by Saeed et al. (2017) and Edae et al. (2014) are consistent with the present study. On the other hand, Jaganathan et al. (2015) detected one SNP associated with GY-SSI and three associated with GY-STI using GBS technology, explaining between 10 % and 13 %, respectively, of the phenotypic variation in chickpea. PS et al. (2017) reported SNPs associated with SSI and STI (calculated from the % spikelet sterility and yield per plant) in rice, explaining between 6 % and 21 %, respectively, of the phenotypic variation. Curiously, some SNPs are associated with traits related to crop productivity and water-use efficiency, explaining a similar range in phenotypic variation. For instance, we detected one association between the SNP iniaGBS11860 and KS-YSI (2012), which was previously associated with the photosynthetic carbon isotope discrimination in wheat (Mora et al., 2015).

Forty-five SNP markers were associated with more than one stress index. For example, one SNP located on chromosome 3A (i.e., iniaGBS21464) was associated with four indices in 2012: TKW-SSI, TKW-YSI, GY-STI and GY-YSI. Three SNPs located on chromosome 4A (iniaGBS22028, iniaGBS558 and iniaGBS2019) and four SNPs located on chromosome 6B (iniaGBS41345, iniaGBS22659, iniaGBS1579 and iniaGBS22660) were associated with TKW-SSI and TKW-YSI in 2012. Additionally, the SNP iniaGBS44415 was associated with TKW-SSI, TKW-YSI, GY-STI and GY-YSI in 2011. In general, markers are frequently associated with more than one trait (e.g., Contreras-Soto et al., 2017b; Liu et al., 2011; Mora et al., 2015), which could be explained by the linkage between markers or possible pleiotropic effects (Zhu et al., 2014). In the context of genetic improvement, the pleiotropic effects of molecular markers could be used to take advantage of more than one trait of interest. For instance, Ookawa et al. (2010) reported a gene associated with high grain yield and enhanced lodging resistance, which was suggested

as an important pleiotropic gene for the improvement of rice varieties. On the other hand, pleiotropic effects could be harmful for one trait and beneficial for another. QTLs associated with herbicide resistance have exhibited harmful pleiotropic effects on yield in a number of crops (Darmency, 2013).

Conclusions

Wheat culture is strongly affected by drought; therefore, the generation of drought tolerant cultivars is one of the main challenges to genetics and breeders. SNP markers linked to the QTL of drought tolerance indices were identified in a diverse genotype collection, and the phenotypic variation explained by SNP markers (up to 6 %) was within an expected range according to other studies. QTL-rich regions on chromosome 4A were detected, supporting the hypothesis that this chromosome has a key role to play in drought tolerance and should be exploited for wheat improvement. In addition, at the genome level, a high number of SNP-associations were located on the B genome, which have been linked with drought tolerance.

The association analysis found a number of SNP markers associated with drought tolerance indices in both growing seasons, revealing genetic regions with stable effects under different environmental conditions. The use of drought tolerance indices in GWAS provides valuable information for marker-assisted selection in wheat.

Acknowledgments

Authors' Contributions

Conceptualization: Mora, F.; Ballesta, P.; Del Pozo, A. Data acquisition: Del Pozo, A. Data Analysis: Ballesta, P.; Mora, F. Design Methodology: Mora, F.; Ballesta, P.; Del Pozo, A. Writing and Editing: Ballesta, P.; Mora, F.; Del Pozo, A.

References

Alexander, L.M.; Kirigwi, F.M.; Fritz, A.K.; Fellers, J.P. 2012. Mapping and quantitative trait loci analysis of drought tolerance in a spring wheat population using amplified fragment length polymorphism and diversity array technology markers. Crop Science 52: 253-261.

Arriagada, O.; Mora, F.; Quitral, Y.; Del Pozo, A. 2017. Identification of QTL underlying agronomic, morphological and physiological traits in barley under rainfed conditions using SNP markers. Acta Scientiarum Agronomy 39: 321-329.

Baltas, E. 2007. Spatial distribution of climatic indices in northern Greece. Meteorology Applications 14: 69-78.

Barbour, M.M.; Warren, C.R.; Farquhar, G.D.; Forrester, G.U.Y.; Brown, H. 2010. Variability in mesophyll conductance between barley genotypes, and effects on transpiration efficiency and carbon isotope discrimination. Plant, Cell & Environment 33: 1176-1185.

Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Brienen, R.J.; Wanek, W.; Hietz, P. 2011. Stable carbon isotopes in tree rings indicate improved water use efficiency and drought responses of a tropical dry forest tree species. Trees 25: 103-113.

Budak, H.; Hussain, B.; Khan, Z.; Ozturk, N.Z.; Ullah, N. 2015. From genetics to functional genomics: improvement in drought signaling and tolerance in wheat. Frontiers in Plant Science 6: 1012.

Contreras-Soto, R.I.; Mora, F.; Lazzari, F.; Oliveira, M.A.R.; Scapim, C.A.; Schuster, I. 2017a. Genome-wide association mapping for flowering and maturity in tropical soybean: implications for breeding strategies. Breeding Science 67: 435-449.

Contreras-Soto, R.I.; Mora, F.; Oliveira, M.A.R.; Higashi, W.; Scapim, C.A.; Schuster, I. 2017b. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. Plos One 12: e0171105.

Croitoru, A.E.; Piticar, A.; Imbroane, A.M.; Burada, D.C. 2013. Spatiotemporal distribution of aridity indices based on temperature and precipitation in the extra-Carpathian regions of Romania. Theoretical Applied Climatology 112:597-607.

Darmency, H. 2013. Pleiotropic effects of herbicide-resistance genes on crop yield: a review. Pest Management Science 69: 897-904.

Dashti, H.; Yazdi-Samadi, B.; Ghannadha, M.; Naghavi, M.R.; Quarrie, S. 2007. QTL analysis for drought resistance in wheat using doubled haploid lines. International Journal of Agriculture and Biology 9: 98-102.

Dodig, D.; Zoric, M.; Kobiljski, B.; Savic, J.; Kandic, V.; Quarrie, S.; Barnes, J. 2012. Genetic and association mapping study of wheat agronomic traits under contrasting water regimes. International Journal of Molecular Sciences 13: 6167-6188.

Edae, E.A.; Byrne, P.F.; Haley, S.D.; Lopes, M.S.; Reynolds, M.P. 2014. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theoretical and Applied Genetics 127: 791-807.

El-Soda, M.; Malosetti, M.; Zwaan, B.J.; Koornnef, M.; Aarts, M.G.M. 2014. Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends in Plant Science 19: 390-398.

Evanno, G.; Regnaut, S.; Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

Farshadfar, E.; Poursiahbidi, M.M.; Abooghadareh, A.R.P. 2012. Repeatability of drought tolerance indices in bread wheat genotypes. International Journal of Agriculture and Crop Sciences 4: 891-903.

Franco, F.A.; Marchioro, V.S.; Montecelli, T.D.N.; Schuster, I.; Polo, M.; Souza, L.V.; Lima, F.J.A.; Evangelista, A.; Santos, D.A.; Grave, E.L. 2018. CD 1303 - Short stature, high productive potential and industrial quality. Crop Breeding and Applied Biotechnology 18: 123-125.

Galetto, S.L.; Bini, A.R.; Haliski, A.; Scharr, D.A.; Borszowskei, P.R.; Caires, E.F. 2017. Nitrogen fertilization in top dressing for wheat crop in succession to soybean under a no-till system. Bragantia 76: 282-291.

Heidari, B.; Sayed-Tabatabaei, B.E.; Saeidi, G.; Kearsey, M.; Suenaga, K. 2011. Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. Genome 54: 517-527.

Jaganathan, D.; Thudi, M.; Kale, S.; Azam, S.; Roorkiwal, M.; Gaur, P.M.; Kavikishor, P.B.; Nguyen, H.; Sutton, T.; Varshney, R.K. 2015. Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. Molecular Genetics and Genomics 290: 559-571.

Kirigwi, F.M.; van Ginkel, M.; Brown-Guedira, G.; Gill, B.S.; Paulsen, G.M.; Fritz, A. K. 2007. Markers associated with a QTL for grain yield in wheat under drought. Molecular Breeding 20: 401-413.

Kumar, S.; Sehgal, S.K.; Kumar, U.; Prasad, P.V.; Joshi, A.K.; Gill, B.S. 2012. Genomic characterization of drought tolerance-related traits in spring wheat. Euphytica 186: 265-276.

Lado, B.; Matus, I.; Rodríguez, A.; Inostroza, L.; Poland, J.; Belzille, F.; Del Pozo, A.; Quincke, M.; Castro, M.; von Zitzewitz, J. 2013. Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3-Genes Genomes Genetics 3: 2105-2114.

Liu, H.; Able, A.J.; Able, J.A. 2017. Water-deficit stress-responsive microRNAs and their targets in four durum wheat genotypes. Functional & Integrative Genomics 17: 237-251.

Liu, T.; Zhang, Y.; Zhang, H.; Xing, Y. 2011. Quantitative trait loci for the number of grains per panicle dependent on or independent of heading date in rice (Oryza sativa L.). Breeding Science 61: 142-150.

Lobell, D.B.; Schlenker, W.; Costa-Roberts, J. 2011. Climate trends and global crop production since 1980. Science 333: 616-20.

Merchuk-Ovnat, L.; Barak, V.; Fahima, T.; Ordon, F.; Lidzbarsky, G.A.; Krugman, T.; Saranga, Y. 2016. Ancestral QTL alleles from wild emmer wheat improve drought resistance and productivity in modern wheat cultivars. Frontiers in Plant Science 7: 452.

Mohammady, S.; Aminian, R.; Hoshmand, S.; Khodombashi, M. 2012. Genomic analysis of carbon isotope discrimination, photosynthesis rate, stomatal conductance, and grain yield in wheat (Triticum aestivum L.) under water-stressed conditions. Crop & Pasture Science 63: 513-519.

Mora, F.; Castillo, D.; Lado, B.; Matus, I.; Poland, J.; Belzile, F.; von Zitzewitz, J.; del Pozo, A. 2015. Genome-wide association mapping of agronomic traits and carbon isotope discrimination in a worldwide germplasm collection of spring wheat using SNP markers. Molecular Breeding 35: 69.

Nezhad, K.Z.; Weber, W.E.; Röder, M.S.; Sharma, S.; Lohwasser, U.; Meyer, R.C.; Börner, A. 2012. QTL analysis for thousand-grain weight under terminal drought stress in bread wheat (Triticumaestivum L.). Euphytica 186: 127-138.

Oliveira, É.C.D.; Pinto-Maglio, C.A.F. 2017. Cytomolecular characterization of cultivars and landraces of wheat tolerant and sensitive to aluminum toxicity. Bragantia 76: 456-469.

Ookawa, T.; Hobo, T.; Yano, M.; Murata, K.; Ando, T.; Miura, H.; Asano, K.; Ochiai, Y.; Ikeda, M.; Nishitani, R.; Ebitani, T.; Ozaki, H.; Angeles, E.R.; Hirasawa, T.; Matsuoka, M. 2010. New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. Nature Communications 1: 132.

Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. Plos One 7: e32253.

Poersch-Bortolon, L.B.; Pereira, J.F.; Nhani Junior, A.; Gonzáles, H.H.S.; Torres, G.A. M.; Consoli, L.; Arenhart, R.A.; Bodanese-Zanettini, M.H.; Margis-Pinheiro, M. 2016. Gene expression analysis reveals important pathways for drought response in leaves and roots of a wheat cultivar adapted to rainfed cropping in the Cerrado biome. Genetics and Molecular Biology 39: 629-645.

Pritchard, J.K.; Stephens, M.; Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Ps, S.; Sv, A.M.; Prakash, C.; Ramkumar, M.K.; Tiwari, R.; Mohapatra, T.; Singh, N.K. 2017. High resolution mapping of QTLs for heat tolerance in rice using a 5K SNP array. Rice 10: 28.

Sabiel, S.A.I.; Huang, S.; Hu, X.; Ren, X.; Fu, C.; Peng, J.; Sun, D. 2017. SNP-based association analysis for seedling traits in durum wheat (*Triticum turgidum* L. *durum* (Desf.)). Breeding Science 67: 83-94.

Saeed, I.; Chen, X.; Bachir, D.G.; Chen, L.; Hu, Y.G. 2017. Association mapping for photosynthesis and yield traits under two moisture conditions and their drought indices in winter bread wheat (*Triticum aestivum* L.) using SSR markers. Australian Journal of Crop Science 11: 248.

Shin, J.H.; Blay, S.; McNeney, B.; Graham, J. 2006. LDheatmap: an R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. Journal of Statistical Software 16: Code Snippet 3.

Sim, S.C.; Durstewitz, G.; Plieske, J.; Wieseke, R.; Ganal, M.W.; Deyne, A.V.; Hamilton, J.P.; Buell, C.R.; Causse, M.; Wijeratne, S.; Francis, D.M. 2012. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. Plos One 7: e40563.

Song, Q.; Jia, G.; Hyten, D.L.; Jenkins, J.; Hwang, E.Y.; Schroeder, S.G.; Osorno, J.M.; Schmutz, J.; Jackson, S.A.; McClean, P.E.; Cregan, P.B. 2015. SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. G3: Genes, Genomes, Genetics 5: 2285-2290.

Sun, F.D.; Zhang, J.H.; Wang, S.F.; Gong, W.K.; Shi, Y.Z.; Liu, A.Y.; Li, J.W.; Gong, J.W.; Shang, H.H.; Yuan, Y.L. 2012. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. Molecular Breeding 30: 569-582.

Van Oosten, M.J.; Costa, A.; Punzo, P.; Landi, S.; Ruggiero, A.; Batelli, G.; Grillo, S. 2016. Genetics of drought stress tolerance in crop plants. p. 39-70. In: Hossain, M.A.; Wani, S.H.; Bhattachajee, S.; Burritt, D.J.; Tran, L.-S.P., eds. Drought stress tolerance in plants. Springer, Cham, Switzerland.

Wei, T.; Simko, V.; Levy, M.; Xie, Y.; Jin, Y.; Zemla, J. 2017. Package 'corrplot'. Statistician 56: 316-324.

Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh, B.I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; Kresovich, S.; Buckler, E.S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38: 203-208.

Zadoks, J.C.; Chang, T.T.; Konzak, C.F. 1974. A decimal code for the growth stages of cereals. Weed Research 14: 415-421.

Zhu, R.; Gao, Y.; Zhang, Q. 2014. Quantitative trait locus mapping of floral and related traits using an F2 population of *Aquilegia*. Plant Breeding 133: 153-161.