
**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
DOCTOR EN CIENCIAS
(MENCIÓN EN INGENIERÍA GENÉTICA VEGETAL)**

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*.

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. globulus* ($\sim 14,000$ SNPs) was higher than in *E. cladocalyx* ($\sim 3,800$ 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 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. cladocalyx*. 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 selection intensity. 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.