

OPINIONES

Mejoramiento genético forestal acelerado mediante selección genómica

Accelerated forest genetic breeding using genomic selection

Jaime Zapata-Valenzuela ^{}, Rodrigo Hasbun Zaror ^b***Autor de correspondencia: ^aNorth Carolina State University, Tree Improvement Program, Raleigh, USA, tel./fax: 1-919-539-2859, jazapata@ncsu.edu^bGenómica Forestal S.A., Concepción, Chile.

SUMMARY

Chilean forest industry should not pursue development without incorporating advances from current biotechnology and genomics tools. Traditional forest breeding programs have been founded in phenotypic selection based on field testing which is time consuming, and where the prediction of breeding values is subjected to the influence of the environment. The use of marker-assisted selection has been focused on discovering quantitative trait loci using markers related to the expression of commercial traits, although a small fraction of explained genetic variance is usually achieved. In animal breeding, a recent revolutionary technology called genomic selection, defined as the use of dense single nucleotide polymorphism markers to predict genomic breeding values, has allowed taking advantage of all possible marker information in estimating genomic values which are highly correlated with true breeding values. The main aspects of this technology and its potential implications in forest breeding are discussed. While main factors that are controlling the performance of genomic selection should be carefully defined, such as the linkage disequilibrium level, trait heritability and the analytical procedures of estimation and validation of genomic values, it is expected that incorporation of genomic selection in forest breeding programs would allow the early capture of the best genotypes for propagation, selection and deployment in breeding of commercial forest traits.

Key words: population genetics, molecular marker, *Pinus radiata*.

RESUMEN

La industria silvícola chilena no debiera desarrollarse al margen de las nuevas herramientas biotecnológicas. Los actuales programas de mejoramiento genético forestal se han sustentado en la selección de progenitores basándose en ensayos de progenie evaluados a largo plazo, en que la estimación de los valores genéticos está afectada por el ambiente. La selección asistida por marcadores moleculares ha facilitado detectar regiones genómicas cuantitativas asociadas a características productivas, aunque explica solo una pequeña proporción de la variación genética. En mejoramiento animal, la incorporación de marcadores de alta resolución en la predicción de valores genómicos, estrategia denominada selección genómica, ha permitido utilizar simultánea y eficientemente la información de miles de marcadores ahorrando tiempo en evaluación y selección. En este trabajo se revisan los principales aspectos de la selección genómica con el objeto de evaluar sus posibles implicancias en el mejoramiento genético forestal. La aplicación de esta estrategia requiere desarrollar un gran número de marcadores para capturar gran parte de la variación genética y definir los factores involucrados en los análisis, tales como el nivel de desequilibrio de ligamiento de la población, la heredabilidad de la característica, los métodos estadísticos y de validación de las predicciones. Se concluye que la implementación de selección genómica permitiría la captura temprana de los mejores individuos, obtener una mayor ganancia genética anual y permitir la plantación operacional del mejor material genético en forma más eficiente y productiva.

Palabras clave: genética poblacional, marcadores moleculares, *Pinus radiata*.

INTRODUCCIÓN

El manejo de la variación genética y su traspaso a la progenie son factores fundamentales para un programa de mejoramiento forestal. El uso de estrategias convencionales de mejoramiento permite estimar el componente genético asociado a una característica y utilizarlo con fines de selección. Generalmente, posterior a siete años en ensayos de terreno, para el caso de *Pinus radiata* D. Don, el fe-

notipo es evaluado y se escogen los mejores padres que permiten generar semilla mejorada vía cruzamientos controlados. A mediano plazo, se seleccionan los mejores clones a nivel intra-familiar, aumentando el rendimiento operacional de las plantaciones y la calidad de los productos industriales (White *et al.* 2007, Grattapaglia y Kirst 2008). No obstante, para la mayoría de las especies forestales, un ciclo de mejoramiento requiere entre 10 y 15 años para seleccionar los mejores fenotipos, representando un alto

costo en logística, evaluación y análisis de ensayos.

La selección temprana facilitaría un proceso más eficiente, reduciendo los ciclos y logrando que el material valioso esté tempranamente disponible para la industria. La posibilidad de selección asistida por marcadores moleculares ha sido explorada en especies forestales, pero su implementación es difícil debido a un reducido desequilibrio de ligamiento (unión no aleatoria de dos loci en el genoma) propio del genoma de las coníferas, limitado número de genes candidatos conocidos y baja reproducibilidad de resultados.

Recientemente estudiada y validada en mejoramiento animal, la selección genómica ha surgido como una alternativa eficiente y que podría ser aplicada para acelerar el mejoramiento forestal. Esta tecnología permite seleccionar basándose en el valor genómico del individuo, predicho mediante miles de marcadores moleculares distribuidos uniformemente a lo largo del genoma. El objetivo de este trabajo es revisar los principales factores relacionados con selección genómica que permitan evaluar su factibilidad como herramienta de selección asistida en mejoramiento genético forestal.

MEJORAMIENTO GENÉTICO FORESTAL CONVENCIONAL

El método tradicional de mejoramiento utiliza la evaluación fenotípica y las relaciones de parentesco como principales herramientas. En silvicultura familiar, se generan familias de medios hermanos o hermanos completos para establecer ensayos de terreno y evaluar su progenie en el tiempo. A través de silvicultura clonal, se realiza selección intra-familiar para transferir la varianza genética total (aditiva y no aditiva) hacia los mejores individuos y propagarlos para su posterior plantación (White *et al.* 2007). Aplicando genética cuantitativa, se usan modelos lineales mixtos para descomponer la varianza fenotípica en sus dos componentes: genético y ambiental. El objetivo es predecir valores genéticos aditivos (proporción de alelos favorables tras pasados desde los padres hacia su progenie). Estos se tratan como variables aleatorias en el cálculo de (co)varianzas, junto a variables fijas como sitio o bloque. El resultado es el mejor predictor lineal insesgado del valor genético para cada individuo (Lynch y Walsh 1998), valor utilizado por el genetista en seleccionar los mejores fenotipos para plantaciones o iniciar un nuevo ciclo de mejoramiento.

Una limitante de esta metodología es que la estimación del genotipo se sustenta únicamente en el fenotipo, el que es afectado por la variación ambiental y variación interna debida a efectos de dominancia y epistasia (White *et al.* 2007). Aun así, la selección fenotípica ha logrado incrementar las ganancias volumétricas de *P. radiata* y *Eucalyptus globulus* Labill. durante las últimas décadas en Chile (INFOR 2009). Además, la silvicultura clonal, masificada rápidamente en nuestro país, es considerada una es-

trategia efectiva al generar mayor ganancia genética comparado con material familiar. En el caso de *Pinus taeda* L. en el sur de Estados Unidos, se han identificado clones con un 70 % de ganancia por sobre las mejores familias de polinización abierta seleccionadas en primera generación¹. No obstante, la mantención de cualquier estrategia de mejoramiento basada en ensayos familiares o clonales exige una importante inversión en recursos, logística y extensos ciclos de selección antes de lograr resultados significativos, dependiendo en gran medida de la heredabilidad (h^2 ; fracción de varianza fenotípica atribuida a varianza genética aditiva) de las características a mejorar (Lynch y Walsh 1998).

SELECCIÓN ASISTIDA POR MARCADORES Y GENES CANDIDATOS

El interés por acelerar los programas de mejoramiento mediante selección asistida por marcadores moleculares (MAS: *marker-assisted selection*) ha sido objeto de varios proyectos en biotecnología forestal. Estos se fundamentan en que los modelos lineales pueden incorporar regiones genómicas cuantitativas (QTL *quantitative trait loci*) como variable predictora. Las regiones genómicas cuantitativas son segmentos del genoma que contienen uno o más genes estadísticamente asociados a una característica cuantitativa. La mayoría de ellas, segregantes en familias de polinización cruzada, siguen una distribución *Gamma*, donde un gran número de efectos pequeños contribuye a la variación de la característica (MacKay *et al.* 2009).

La construcción de mapas genéticos ha permitido identificar marcadores ligados a regiones genómicas cuantitativas reguladoras del crecimiento, propiedades de la madera y tolerancia a enfermedades (Grattapaglia y Kirst 2008). No obstante, su implementación en especies forestales ha sido ineficiente, fundamentalmente porque: la asociación región genómica-marcador es válida para una estructura poblacional específica (intra-familiar); su detección requiere una población segregante (retrocruzas o generación F2); existe un limitado desequilibrio de ligamiento, característico de coníferas con genomas complejos (Neale 2007, Grattapaglia y Resende 2010).

Otra alternativa es la selección basada en asociación genética, mediante estudios de mapeo asociativo. El principal objetivo es detectar asociaciones significativas entre marcadores polimórficos de un solo nucleótido (SNP: *single nucleotide polymorphism*) y observaciones fenotípicas. Estos marcadores, estables y abundantes en el genoma, permiten cubrir mayores segmentos cromosómicos. Debido al reducido desequilibrio de ligamiento, la gran diversidad nucleotídica y la inexistencia de genomas de referencia, los estudios de asociación genética forestal han identificado unas pocas decenas de genes candidatos específicos, donde la magnitud de cada asociación escasamente

¹ Steve McKeand. North Carolina State University. Comunicación personal.

representa el 5 % de la varianza fenotípica, incluso, para características de alta heredabilidad (Neale 2007, MacKay *et al.* 2009).

Sin desconocer los importantes avances alcanzados en estudios de mapeo y asociación genética, el descubrimiento de genes asociados a características específicas no ha logrado impactar el mejoramiento de características poligénicas propias de especies forestales. Alternativamente, el mejoramiento genómico que analiza el genoma en forma amplia, representa un inexplorado potencial de aplicación. Un ejemplo es la selección genómica o selección basada en el uso simultáneo de miles de marcadores dispersos en el genoma.

PRINCIPALES ASPECTOS DE SELECCIÓN GENÓMICA

La selección genómica se define como la selección de individuos basada en su valor genómico, propuesto por Meuwissen *et al.* (2001). Los valores genómicos se predicen en función de los efectos de marcadores moleculares distribuidos en todo el genoma del individuo. Con una adecuada densidad de marcadores, los genes que influyen la característica deberían estar en desequilibrio de ligamiento (estrechamente unidos) con al menos un marcador. Los efectos de dichos marcadores se modelan primero en una población de referencia en conjunto con datos fenotípicos. Con esta información se genera una ecuación para predecir el valor genómico en una población de validación parcial, independiente de la población de referencia. Es fundamental “entrenar” la ecuación de predicción con algunos datos fenotípicos para revisar su poder predictivo. Finalmente, se requiere una población de validación en que solo se miden marcadores para realizar el cálculo del valor genómico, sin necesidad de datos fenotípicos (Meuwissen *et al.* 2001, Resende *et al.* 2008).

El parámetro cuantitativo de selección genómica es la correlación ($0 < r < 1$) entre el valor genómico predicho y el valor genético verdadero, con un óptimo ideal de $r = 1$ (Schaeffer 2006). El nivel de desequilibrio de ligamiento es considerado el factor más determinante de la confiabilidad del valor r . Aumentar ésta correlación implica cuantificar la herencia de múltiples genes con efectos pequeños y explotar el desequilibrio de ligamiento entre marcadores genotipificados dentro de segmentos cromosómicos específicos. En otras palabras, se aprovecha la existencia de asociaciones no aleatorias entre alelos favorables que son heredados por los individuos de una población y que se espera sean detectadas en todo el genoma. Grattapaglia y Resende (2010) establecen que el desequilibrio de ligamiento puede ser manejado controlando el tamaño efectivo de la población de referencia. Se sugieren tamaños efectivos de población entre 20 y 40 individuos élite, abarcando el rango de los actuales programas de mejoramiento forestal. Además, el nivel de desequilibrio de ligamiento capturado aumenta con una mayor densidad de marcadores (Solberg

et al. 2008). Por lo tanto, una vez completado el mapa genómico de las coníferas comerciales, junto al avance bioinformático y al progresivo menor costo en genotipificación masiva de marcadores, se podrán incorporar mapas de marcadores de alta densidad en estrategias de selección genómica aumentando su eficiencia (Neale 2007).

La heredabilidad es otro factor determinante en selección genómica, inherente a la población y que escapa al control del genetista. En general, se espera que la superioridad relativa de la selección genómica en relación a la selección fenotípica aumente con características de baja heredabilidad ($h^2 < 0,3$). Según (Grattapaglia y Resende 2010), la heredabilidad tendría un impacto relativo pequeño en la funcionalidad real de selección genómica en especies forestales, y los parámetros más relevantes son el tamaño efectivo de la población de referencia, número de genes controlando la característica y magnitud del desequilibrio de ligamiento. En resumen, la selección genómica está revolucionando el mejoramiento en animales productores de leche durante la última década y podría tener un gran impacto en mejoramiento forestal (Hayes *et al.* 2008, Resende *et al.* 2008). A la fecha, algunos resultados empíricos experimentales evaluados en *Eucalyptus grandis* Hill ex Maiden (Grattapaglia *et al.* 2011) y *Pinus taeda* (Resende *et al.* 2011, Zapata-Valenzuela *et al.* 2011) confirman la efectividad de la selección genómica en el mejoramiento forestal, aunque estudios adicionales son requeridos.

ANÁLISIS, VALIDACIÓN Y APLICACIÓN DE SELECCIÓN GENÓMICA

Goddard y Hayes (2007) describen en detalle las principales etapas del análisis estadístico para calcular el valor genómico. Lo primero es determinar el nivel de desequilibrio de ligamiento existente entre los marcadores disponibles o calcular directamente las correlaciones en una población de validación. Una alta correlación con el valor fenotípico de los individuos validados infiere una magnitud positiva de desequilibrio de ligamiento existente en la población (Resende *et al.* 2008). Luego, es importante decidir la distribución probabilística más adecuada para los efectos de los marcadores en los individuos de la población experimental. Tradicionalmente, se asumen efectos aleatorios muestreados desde una distribución normal con varianza constante en todos los segmentos cromosómicos, utilizando el método del mejor predictor lineal insesgado para predecir el valor genómico. Al incluir un factor poligénico para capturar la variación no explicada por los marcadores, las predicciones pueden ser mejoradas (Hayes *et al.* 2008). Otra alternativa es usar métodos bayesianos, asumiendo los efectos de los marcadores como aleatorios, con tamaño diferente y varianza heterogénea asociada a una distribución *a priori* específica (Goddard y Hayes 2007).

Una comparación entre diferentes métodos de pre-

dicción fue simulada por Meuwissen *et al.* (2001) para un genoma de 1.000 cM^2 y con 50.000 haplotipos. Los métodos bayesianos obtuvieron correlaciones entre 0,74 y 0,85 dependiendo del espaciamiento entre marcadores, demostrando que los valores genómicos pueden ser predichos con exactitud, facilitando una selección más eficiente. Evaluaciones en experimentos reales de mejoramiento animal se han desarrollado en Nueva Zelanda, Australia y Estados Unidos, usando 4.500 progenitores como población de referencia y 50.000 marcadores, obteniendo correlaciones entre 0,20 y 0,67 para individuos sin información de progeñe. Actualmente, la validez de selección genómica en mejoramiento animal es superior a la fenotípica, permitiendo que algunas empresas utilicen esta tecnología rutinariamente para selección de progenitores, reduciendo sus costos operacionales hasta un 92 % (Hayes *et al.* 2008). Similares expectativas en mejoramiento forestal debieran ser analizadas con precaución, al considerar importantes diferencias con el modelo animal: menor número de marcadores disponibles para especies leñosas, genomas de mayor tamaño, limitado desarrollo de marcadores codominantes del tipo polimórficos de un solo nucleótido y tamaños efectivos de población variables.

¿Cómo escoger el mejor método analítico para la selección? La respuesta implica disponer de un método de validación práctico de los valores predichos, etapa que incluye seleccionar individuos sin información fenotípica, ordenados únicamente por valor genómico y confirmar posteriormente sus respuestas en terreno. Por ejemplo, Hayes *et al.* (2008) utilizaron efectos predichos para asignar valores genómicos a una población de toros recién nacidos y luego de algunos años, dichos valores se correlacionaron positivamente con los valores genéticos verdaderos calculados en censos de progeñe. Legarra *et al.* (2008) evaluaron la eficiencia de selección genómica ajustando modelos en una población animal usando validación cruzada al dividir equitativamente la población original: una mitad para calcular el valor genómico y la otra para validar los efectos predichos. En ambos casos, la correlación alcanzó un valor máximo de 0,67 dependiendo de la característica y heredabilidad evaluadas.

Como destacan White *et al.* (2007) y Legarra *et al.* (2008), los programas de mejoramiento en cualquier especie están diseñados para tomar decisiones sobre acciones futuras y deben ser analizados rigurosamente. En el caso forestal, la validación representa una etapa crucial en selección genómica, considerando que evaluar ensayos genéticos requiere varios años para permitir resultados confiables. Algunas alternativas se proponen como métodos de validación: dividir la población disponible y validar en un subgrupo de ella incluyendo repeticiones aleatorias para reducir error de muestreo; en casos de ensayos que contienen familias/clones repetidos en diversos sitios o

²cM = centimorgan: unidad de distancia entre dos loci, utilizada en mapas de ligamiento genético. Equivale a la distancia que produce una fracción de recombinación del 1% entre dos loci.

ambientes, una fracción de ellos podría ser utilizada como población de referencia y el grupo restante destinarlo a validación. Esto contribuiría a explorar el tema de interacción genotipo-ambiente, aspecto relevante en la adaptabilidad del material genético a diversas condiciones de suelo/clima existentes en Chile.

Bajo el contexto operacional del mejoramiento forestal, una aplicación a corto plazo sería selección genómica temprana de clones en vivero o de embriones somáticos en laboratorio, en que hayan sido predichas sus características de volumen, resistencia a enfermedades o adaptabilidad a factores abióticos, reduciendo los costos y tiempo de espera para obtener información similar desde ensayos de terreno. A largo plazo, el resultado de selección genómica se extiende a ganancia operacional (R) a través de la ecuación $R = ir\sigma$, en que i es la intensidad de selección, r es la correlación entre el valor genómico predicho y el valor genético aditivo verdadero, y σ es la raíz de la varianza aditiva (Legarra *et al.* 2008). Al ponderar la ecuación anterior por un reducido ciclo de "mejoramiento genómico", sería posible obtener una mayor ganancia (R) anual.

CONCLUSIONES

La selección genómica constituye una tecnología probada en programas de mejoramiento animal que posibilita al genetista seleccionar individuos en función de una correlación (r) precisa entre el valor genómico y valor genético. Su implementación al mejoramiento forestal depende del ajuste correcto de un modelo estadístico-cuantitativo que incluya parámetros genéticos inherentes a la población, datos de marcadores y observaciones fenotípicas. El resultado esperado es una alta correlación que sumado a una reducción del ciclo de mejoramiento, permitiría obtener una mayor ganancia operacional anual (R), maximizar la razón eficiencia-costo del ciclo de selección forestal, ahorrando tiempo y recursos en la evaluación y análisis de ensayos genéticos. En consecuencia, la actual selección fenotípica podría ser restringida a poblaciones de menor tamaño utilizadas únicamente para ajustar las estimaciones de los valores genómicos obtenidos desde marcadores moleculares. Como cualquier otra tecnología productiva emergente, se requieren y recomiendan evaluaciones empíricas para confirmar las expectativas generadas y discutidas en estudios previos de simulación.

REFERENCIAS

- Goddard ME, BJ Hayes. 2007. Genomic selection. *Journal of Animal Breeding and Genetics* 124: 323-330.
- Grattapaglia D, M Kirst. 2008. Eucalyptus applied genomics: from gene sequences to breeding tools. *New Phytologist* 179(4): 911-929.
- Grattapaglia D, M Resende. 2010. Genomic selection in forest tree breeding. *Tree Genetics and Genomes* 7(2): 241-255.
- Grattapaglia D, MDV Resende, MFR Resende, CP Sansaloni,

- CD Petroli, AA Missiaggia, EK Takahashi, KC Zamprogno, A Kilian. 2011. High realized accuracies of genomic selection for volume growth and wood density in *Eucalyptus* breeding populations with contrasting effective sizes. Consultado 30 mayo 2011. Disponible en http://www.intl-pag.org/19/abstracts/W35_PAGXIX_235.html
- Hayes BJ, PJ Bowman, AJ Chamberlain, ME Goddard. 2008. Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92: 433-443.
- INFOR (Instituto Forestal, CL). 2009. El sector forestal chileno 2009. Instituto Forestal. Consultado 20 oct. 2010. Disponible en <http://www.infor.cl>
- Legarra A, C Robert-Granie, E Manfredi, JM Elsen. 2008. Performance of genomic selection in mice. *Genetics* 180: 611-618.
- Lynch M, B Walsh. 1998. Genetics and analysis of quantitative traits. Sunderland, USA. Sinauer Associates. 980 p.
- Mackay TFC, EA Stone, JF Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10: 565-577.
- Meuwissen THE, BJ Hayes, ME Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.
- Neale DB. 2007. Genomics to tree breeding and forest health. *Current Opinion in Genetics & Development* 17: 539-544.
- Resende MFR, JJ Acosta, PR Muñoz Del Valle, T Quesada, JM Davis, MDV Resende, M Kirst. 2011. Genomic selection and next-generation genotyping to hyper-accelerate pine breeding. Consultado 30 mayo 2011. Disponible en http://www.intl-pag.org/19/abstracts/W35_PAGXIX_236.html
- Resende MDV, PS Lopes, RL Silva, IE Pires. 2008. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesquisa Florestal Brasileira* 56: 63-78.
- Schaeffer LR. 2006. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics* 123: 218-223.
- Solberg TR, AK Sonesson, JA Woolliams, THE Meuwissen. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science* 86: 2447-2454.
- White TL, WT Adams, DB Neale. 2007. Forest genetics. Cambridge, USA. CABI Publishing CAB International. 682 p.
- Zapata-Valenzuela J, F Isik, C Maltecca, JL Wegrzyn, DB Neale, RW Whetten. 2011. Genomic selection in a cloned population of loblolly pine. Consultado 30 mayo 2011. Disponible en http://www.intl-pag.org/19/abstracts/W35_PAGXIX_237.html

Recibido: 01.08.10

Aceptado: 07.06.11

