

Aplicación de técnicas bayesianas en el análisis genético de árboles forestales

Applications of bayesian techniques in forest tree breeding

Freddy Mora^{a*} y Sandra Perret^b

*Autor de correspondencia: ^aUniversidad Estatal de Maringá-Paraná, Centro de Ciencias Agrarias, Av. Colombo 5790, bloco 05, sala 01, Maringá – PR, Brasil, tel.: 55 44 3261-4552, fax: 55 44 3262-7336, fmora@universiabrasil.net, morapoblete@gmail.com

^bInstituto de Investigación Forestal (INFOR), Huérfanos 554, Santiago, Chile

SUMMARY

This study was undertaken to investigate bayesian methods to scientific inference in forest tree breeding. The Independence Chain (IC) and Gibbs sampling (GS) algorithms were applied on data sets from a field trial of *Eucalyptus cladocalyx* aiming to predict random family and additive genetic effects, respectively. The trial was carried out in the coastal area of Coquimbo Region, northern Chile. Diameter and height growth rate (TCD and TCA) were the data sets collected in a period of 30 months. Bayesian procedures were compared with the ranking obtained by Best Linear Unbiased Prediction (BLUP). BLUP and IC had similar family ranking, although the family selection based on credible intervals of the genotypic effects (bayesian framework) was clearly more intensive. Spearman rank correlation coefficients were high and significant ($I'_s = 0.99$) between BLUP and GS. The estimates of heritability were moderate: $h^2 = 0.42$ (TCA) and $h^2 = 0.43$ (TCD). The traits had a significant correlation ($I'_s = 0.7$). The estimated genetic gains, in relation to the experimental average, ranged from 17% to 28%. The greatest genetic gain and genetic diversity of the original diversity of the experimental populations may be achieved and maintained if the trees are selected for TCD. It was concluded that the bayesian inference may be a useful tool in the genetic evaluation of forest trees, since it incorporates the variability of the genetic parameters by using posterior distributions.

Key words: *Eucalyptus*, Gibbs sampling, heritability, independence chain, REML/BLUP.

RESUMEN

El objetivo de este estudio fue investigar los métodos bayesianos como una alternativa de inferencia científica aplicada a la evaluación genética forestal. Se usaron los algoritmos de Cadenas Independientes (IC) y de Gibbs (GS) en un conjunto de datos provenientes de un ensayo de *Eucalyptus cladocalyx*, para la predicción del efecto familiar e individual, respectivamente. El ensayo fue establecido en el sector costero de la Región de Coquimbo, norte de Chile. Se evaluó la tasa de crecimiento promedio de la altura (TCA) y el diámetro (TCD), medidas en un periodo de 30 meses. Los procedimientos se compararon con la mejor predicción lineal insesgada (BLUP). Se confirmó una significativa asociación entre el ranking familiar de BLUP e IC, aunque se evidenciaron mayores intensidades de selección al utilizar las regiones de credibilidad de los efectos genotípicos (enfoque bayesiano). Se obtuvieron altas y significativas correlaciones de Spearman ($I'_s = 0,9$) entre BLUP y GS. Se obtuvieron moderadas heredabilidades individuales: $h^2 = 0,42$ (TCA) y $h^2 = 0,43$ (TCD). Las características se correlacionaron significativamente entre sí ($I'_s = 0,7$). Las ganancias genéticas, en relación al promedio del ensayo, variaron de 17% a 28%. Al seleccionar para TCD, mayor ganancia y diversidad genética puede ser alcanzada y mantenida que la diversidad encontrada en este ensayo de progenie. Se concluyó que la inferencia bayesiana puede ser una herramienta metodológica útil en la evaluación genética forestal, ya que permitió incorporar la variación de los parámetros genéticos a través de las distribuciones *a posteriori*.

Palabras clave: algoritmo de Gibbs, cadenas independientes, *Eucalyptus*, heredabilidad, REML/BLUP.

INTRODUCCIÓN

La teoría de modelos lineales mixtos aplicada dentro del contexto del mejoramiento genético forestal ha tenido un importante desarrollo y una amplia utilización en la investigación aplicada en Chile, lo que ha permitido la evaluación genética en diferentes programas de selección

y mejoramiento de especies de interés económico, y en un amplio rango de condiciones ambientales (Ipinza *et al.* 1998, Apiolaza *et al.* 2000, Gutiérrez e Ipinza 2000, De Veer *et al.* 2001, Zamudio *et al.* 2002, Matheson *et al.* 2006, Perret *et al.* 2006).

En los modelos mixtos, los parámetros de efectos fijos pueden ser estimados a través del mejor estimador lineal

insesgado (BLUE = *best linear unbiased estimation*), donde son considerados los efectos ambientales identificables en el diseño del experimento. A su vez, la predicción de los valores genéticos o de los efectos aleatorios del modelo puede ser obtenida a través de la mejor predicción lineal insesgada (BLUP = *best linear unbiased prediction*) (White y Hodge 1989). Para la obtención de BLUE y BLUP es necesario definir o decidir entre los distintos métodos de estimación de los componentes de varianza. El método de la máxima verosimilitud restringida (REML = *restricted maximum likelihood*) (Searle *et al.* 1992) es en la actualidad un procedimiento ampliamente utilizado para la estimación de componentes de varianza, y posterior solución de los efectos fijos y aleatorios del modelo, BLUE y BLUP, con observaciones fenotípicas no balanceadas (o con datos perdidos) (Lu *et al.* 1999).

No tan sólo la obtención de la mejor estimación y/o predicción lineal insesgada dependen de los componentes de varianza o covarianza, sino también las estimaciones de parámetros genéticos, como la heredabilidad, o el grado de control genético de una característica de interés (Arnold *et al.* 1991), y las correlaciones genéticas entre características, ya sea aquellas consideradas dentro de un mismo árbol (correlación Tipo A), o entre árboles emparentados que crecen en ambientes diferentes (correlación Tipo B) (Burdon 1977, Lu *et al.* 1999). Estos parámetros genéticos están directamente relacionados a la idoneidad de la estimación de los componentes de (co)varianza.

Recientemente, Cappa y Cantet (2006) mencionan que el método de la máxima verosimilitud restringida (REML) aplicado en genética forestal para la estimación de los componentes de varianza, a partir de modelos de árboles individuales multivariados (con características múltiples), es afectado por la pérdida de datos (observaciones) en cualquier característica e individuo, en la forma de la distribución de las estimaciones de parámetros genéticos vía REML. Los autores también proponen como una alternativa a la estimación vía REML la utilización de técnicas bayesianas a través de los métodos de Monte Carlo vía cadenas de Markov (MCMC = Markov-Chain Monte Carlo).

En el mejoramiento genético forestal, la estimación de los componentes de varianza y la predicción de los valores genéticos son aspectos claves en el análisis genético-cuantitativo de características de importancia económica (Schenone *et al.* 2002, Zamudio *et al.* 2002, Perret *et al.* 2006), los cuales pueden ser abordados por medio de métodos bayesianos, incorporando las diferentes variantes de los procedimientos MCMC (Mora y Arnhold 2006, Mora *et al.* 2006b). Tales procedimientos son utilizados para determinar las distribuciones marginales *a posteriori* de los parámetros, dando paso a la aplicación del teorema de Bayes a través de algoritmos que convergen a tales densidades marginales. La metodología bayesiana es considerada una herramienta importante en la evaluación genética, ya que lleva en consideración la variabilidad existente en todos los parámetros del modelo (Wright

et al. 2000, Nogueira *et al.* 2003), en el sentido de que es posible caracterizarlos a través de la moda, mediana o el promedio de la distribución *a posteriori* de dicho parámetro, así como la obtención de intervalos de confianza (llamadas de regiones de credibilidad en el caso bayesiano) tanto para los parámetros como para la predicción de los efectos genéticos.

El presente trabajo tuvo como objetivo investigar las técnicas bayesianas como una alternativa de inferencia científica aplicada en el mejoramiento genético forestal. Para ello se utilizó un conjunto de datos provenientes de un ensayo de polinización abierta de *Eucalyptus cladocalyx* F. Muell, establecido en el norte de Chile, donde se aplicaron dos variantes de los procedimientos MCMC: el algoritmo de cadenas independientes, para la predicción/selección basada en el promedio de las progenies, y el algoritmo de Gibbs, para la predicción/selección de árboles individuales. Con los resultados de este análisis se discute la utilidad del enfoque bayesiano en la teoría del mejoramiento forestal y los resultados de la evaluación genética de árboles de *Eucalyptus spp.* en condiciones de baja disponibilidad hídrica.

MÉTODOS

Descripción del ensayo. El ensayo se localizó en el sector costero de la Región de Coquimbo, comuna de Los Vilos, en la hacienda agrícola Caracas (31°55'05" S, 71°27'10" O, 167 m de altitud). La población total consistió de 49 familias de medios hermanos de *E. cladocalyx*, establecida en un diseño de bloques al azar, con treinta bloques y una planta por parcela. De estas familias, 47 eran provenientes de cinco localidades de Australia (procedencias) y las restantes de plantaciones nacionales (fuente de semilla nacional). Informaciones adicionales de las procedencias y del ensayo pueden ser encontradas en Mora (2004, 2006). Las tasas de crecimiento absoluto de la altura y del diámetro se consideraron como las variables de interés, calculadas en un período de 30 meses, desde el inicio de la plantación, en septiembre de 2001, hasta marzo de 2004.

Se realizó un análisis a través de técnicas bayesianas, considerando la predicción y/o selección de familias, a través del algoritmo de cadenas independientes, y de árboles individuales, mediante el algoritmo de Gibbs; ambos procedimientos se compararon con los métodos clásicos de estimación/predicción.

Selección de familias. Para este análisis, se consideró el siguiente modelo lineal:

$$y = X\beta + Zf + e \quad [1]$$

donde: y = vector de las observaciones (tasas de crecimiento), β = vector de los efectos fijos bloques (30) y poblaciones (6), f = vector de los efectos aleatorios de las

familias (49), X y Z = matrices de incidencia para cada efecto identificable del modelo, e = vector de los efectos residuales, aleatorios.

En el contexto bayesiano, todos los parámetros del modelo son considerados variables aleatorias, o sea, no existe una distinción entre los efectos fijos y aleatorios, aunque el modelo lineal mixto es utilizado dentro de esta inferencia científica (Blasco 2001).

Para la aplicación del teorema de Bayes, generalmente se especifican las distribuciones *a priori* de los parámetros, a no ser que sea utilizada una distribución no informativa (o de Jeffreys). En el presente análisis, las distribuciones *a priori* de los parámetros se basaron en el estudio realizado por Mora y Arnhold (2006), donde, para facilidad, el vector β sigue una distribución *a priori* uniforme o constante (*flat*), los vectores de los efectos aleatorios siguen una distribución normal y los componentes de varianza una distribución *a priori* gamma invertida.

El modelo [1] se basó en un análisis unicarácter utilizando el algoritmo de cadenas independientes (IC) (Wolfinger y Kass 2000) para la obtención de las distribuciones *a posteriori* de la predicción de los efectos aleatorios. Para esto se generaron diez mil muestras aleatorias utilizando el procedimiento PROC MIXED del software SAS, con la opción PRIOR (Mora y Arnhold 2006, Mora *et al.* 2007). Se obtuvieron intervalos de confianza (90% de probabilidad) de cada efecto genotípico a través de las distribuciones *a posteriori*, utilizando el procedimiento PROC UNIVARIATE de SAS.

Los resultados obtenidos de la predicción bayesiana del efecto familiar se compararon con los obtenidos del procedimiento de estimación/predicción REML/BLUP (máxima verosimilitud restringida/mejor predicción lineal insesgada), usando el procedimiento PROC MIXED.

Selección de árboles individuales. Árboles individuales de la especie se analizaron en función de los efectos genéticos aditivos. El análisis bayesiano se basó en el algoritmo de Gibbs, considerando el siguiente modelo lineal:

$$y = X\beta + Za + e \quad [2]$$

donde: y = vector de las observaciones (tasas de crecimiento), β = vector de los efectos fijos debido al bloque y población, a = vector de los efectos aditivos de los árboles individuales (incluyendo los árboles genitores identificables), X y Z = matrices de incidencia del modelo, e = vector de los efectos residuales.

La distribución invertida Wishart (IW) se utilizó como distribución *a priori* de los componentes de (co)varianza. El modelo [2] se basó en un análisis multicarácter (considerando ambas características en el modelo). Debido a su simplicidad, la distribución IW es usualmente asumida en la definición de las distribuciones *a priori* de los componentes de varianza, para un análisis multicarácter (Van Tassell y Van Vleck 1996).

Para la implementación de las distribuciones *a posteriori* de la predicción de los efectos genéticos aditivos, se utilizó el programa MTGSAM (*Multi-Trait Gibbs Sampler for Animal Model*) desarrollado por Van Tassell y Van Vleck (1996).

La predicción genética de los árboles individuales obtenida a través de técnicas bayesianas se comparó con el *ranking* obtenido a través del método de estimación/predicción REML/BLUP, utilizando un análisis gráfico y las estimaciones de correlación de Spearman entre *ranking*. Se incluyeron intervalos de confianza *bootstrap* (Efron 1979) (95% de probabilidad) para los coeficientes de correlación, a través de la generación de cinco mil muestras aleatorias con reemplazo, utilizando los procedimientos SURVEYSELECT y CORR de SAS (SAS Institute 1996), de acuerdo con Mora y Arnhold (2006).

Para la predicción de los efectos genéticos aditivos vía procedimiento REML/BLUP se utilizó el programa MTDFREML (*Multi-Trait Derivative Free REML*), metodología basada en la maximización de la función de verosimilitud sin usar derivada (Boldman *et al.* 1995).

Adicionalmente, utilizando el análisis de árboles individuales, a través del algoritmo de Gibbs, se estimaron los valores promedio (incluyendo los intervalos de confianza) de la distribución *a posteriori* de los siguientes parámetros:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad [3]$$

$$r_g = \frac{\sigma_{a12}}{\sqrt{\sigma_{a1}^2 \sigma_{a2}^2}} \quad [4]$$

donde: h^2 = heredabilidad individual, σ_a^2 y σ_e^2 = componentes de varianza aditiva y residual, respectivamente; r_g = correlación genética entre las características; σ_{a12} = covarianza aditiva entre las características; σ_{a1}^2 y σ_{a2}^2 = varianzas aditivas de la tasa de crecimiento en altura y diámetro, respectivamente.

Adicionalmente, se estimaron la ganancia genética al seleccionar árboles individuales (considerando el valor promedio de todo el ensayo como línea base) y la diversidad genética mantenida después de la selección (de la diversidad genética originalmente encontrada en el ensayo de progenie):

$$GG\% = \frac{\bar{X}_{EG_{Sel}}}{\bar{X}_{Exp}} * 100 \quad [5]$$

$$D = \frac{(\sum k_f)^2 / \sum k_f^2}{N_p} \quad [6]$$

donde: $GG\%$ = ganancia genética (en %), $\bar{X}_{EG_{Sel}}$ = valor promedio de los efectos genéticos de los árboles seleccionados

nados, \bar{X}_{Exp} = línea base de la selección (valor promedio de todo el ensayo), D = diversidad genética mantenida después de la selección, k_f = número de árboles individuales seleccionados por progenie, N_p = número de progenies originales que tiene el ensayo (49).

RESULTADOS

En el cuadro 1 se presentan los efectos genotípicos (EG) para las 11 mejores familias de *E. cladocalyx* (equivalentes al 22,45% del total de familias), seleccionadas en función de las características de interés. Se presentan también los límites de confianza, inferior y superior, para la predicción bayesiana (90% de probabilidad), los cuales son provenientes de las distribuciones *a posteriori* de cada efecto genotípico. Aunque la probabilidad asociada al intervalo de confianza es menor que la frecuentemente utilizada, algunos efectos genotípicos evidenciaron límites de confianza con valores negativos. Los efectos genotípicos fueron similares entre BLUP y el promedio de la distribución *a posteriori* (vía algoritmo de cadenas independientes).

En el cuadro 2 se muestran los valores de la correlación de Spearman, incluyendo los intervalos de confianza *bootstrap* (95% de probabilidad). Este coeficiente de correlación se estimó entre los ranking de árboles individuales, obtenidos a través del algoritmo de Gibbs y de la mejor predicción lineal insesgada (BLUP), con estimación de componentes de varianza vía máxima verosimilitud restringida (REML). Se muestran también las correlaciones de Spearman calculadas entre las características de crecimiento, dentro de cada método. Altos y significativos valores de la correlación de Spearman se obtuvieron entre los distintos

métodos de análisis y entre las características de crecimiento ($P < 0,01$). La estimativa puntual de la correlación genética entre ambas características de crecimiento tuvo un valor de $r_g = 0,775$ y confirma la estrecha relación entre estos caracteres, con límites de confianza aproximados de 0,738 y 0,819 (provenientes de la distribución posterior). En la figura 1 se muestran los valores de la predicción de los efectos genéticos aditivos, para ambas características de crecimiento, basados en el algoritmo de Gibbs. La línea recta es una medida de comparación entre ambas características, y también confirma la asociación entre los efectos aditivos.

Cuadro 2. Correlación de Spearman e intervalos de confianza *bootstrap* (95% de probabilidad) calculados para determinar la equivalencia entre los procedimientos bayesiano, vía algoritmo de Gibbs (GS), y la mejor predicción lineal insesgada (BLUP), y entre las características de crecimiento.

Spearman rank correlation coefficients and bootstrap confidence intervals (95% of probability) calculated to determine the association between Bayesian, via Gibbs sampling, and the Best Linear Unbiased Prediction (BLUP), and between growth traits.

Parámetro	GS-BLUP	GS-BLUP	Gibbs	BLUP
	TCA	TCD	TCA-TCD	TCA-TCD
Correlación de Spearman	0,9996	0,9995	0,7268	0,7422
Límite de confianza inferior	0,9995	0,9994	0,6995	0,7160
Límite de confianza superior	0,9997	0,9996	0,7542	0,7684

TCA y TCD: son las tasas de crecimiento en altura y diámetro.

Cuadro 1. Efectos genotípicos (EG) para las 11 mejores familias, de acuerdo con BLUP e inferencia bayesiana (vía algoritmo de cadenas independientes). CI y CS son los límites de confianza, inferior y superior (90% de probabilidad), provenientes de la distribución *a posteriori* de cada efecto genotípico.

Genotypic effects (EG) based on BLUP and Bayesian procedures (using Independence Chain algorithm) for the best families. CI and CS are the lower and upper cutoff for 90% credible sets, from posterior distributions of each genotypic effect.

Tasa de crecimiento en altura						Tasa de crecimiento en diámetro					
BLUP		Bayes				BLUP		Bayes			
Familia	EG	Familia	EG	CI	CS	Familia	EG	Familia	EG	CI	CS
32	0,8084	32	0,8089	0,4917	1,1249	31	0,1444	31	0,1448	0,0665	0,2261
3	0,5875	3	0,5921	0,2759	0,9033	32	0,1222	32	0,1214	0,0471	0,1971
4	0,5784	4	0,5798	0,2798	0,8811	6	0,1024	6	0,1014	0,0281	0,1784
6	0,4415	6	0,4423	0,1446	0,7372	39	0,0867	39	0,0870	0,0074	0,1659
39	0,3507	39	0,3519	0,0338	0,6731	13	0,0827	13	0,0818	0,0119	0,1564
15	0,3284	15	0,3301	0,0422	0,6180	43	0,0740	43	0,0744	-0,0038	0,1531
16	0,3013	16	0,3063	0,0131	0,5950	19	0,0703	19	0,0716	-0,0097	0,1519
19	0,2749	19	0,2743	-0,0561	0,6132	42	0,0601	23	0,0599	-0,0237	0,1475
28	0,2588	28	0,2616	-0,0711	0,5720	23	0,0596	42	0,0596	-0,0183	0,1378
2	0,2249	13	0,2276	-0,0698	0,5144	3	0,0588	3	0,0576	-0,0144	0,1341
13	0,2248	2	0,2262	-0,0698	0,5241	7	0,0497	7	0,0497	-0,0224	0,1229

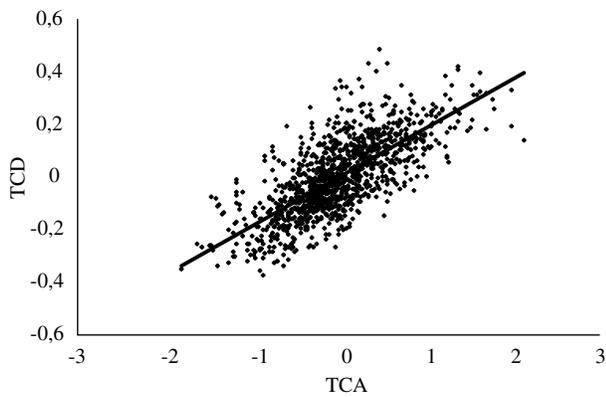


Figura 1. Gráfico de los efectos genéticos aditivos para las tasas de crecimiento en diámetro (TCD) y altura (TCA) del total de árboles analizados. La línea recta entrega una medida de relación entre ambas características.

Graphic of the additive genetic effects for diameter (TDC) and height (TCA) growth rates of the total trees under analysis. The line provides an indication of relationship between traits.

En la figura 2 se muestran gráficos con la comparación del ranking de los árboles padres, basados en BLUP y técnicas bayesianas, a través del algoritmo de Gibbs. Aunque estos individuos genitores no presentan información dentro del ensayo, es posible evaluarlos genéticamente sobre la base de sus descendientes. La figura 3 muestra los gráficos con los valores de la predicción de los efectos genéticos aditivos de un conjunto de árboles seleccionados (intensidad de selección de 8,16%) para la tasa de crecimiento en altura y diámetro, utilizando las metodologías REML/BLUP y algoritmo de Gibbs. En ambas figuras, las líneas rectas entregan una medida de equivalencia entre ambos procedimientos.

En el cuadro 3 se muestran los resultados de las estimaciones promedio de la distribución *a posteriori* de los componentes de varianza y la heredabilidad, estimados a través de inferencia bayesiana, usando el algoritmo de Gibbs. Los límites de confianza inferior y superior también son presentados, los cuales se determinaron a partir de la distribución *a posteriori* de los parámetros. Los valores promedio de las heredabilidades fueron moderados en ambas

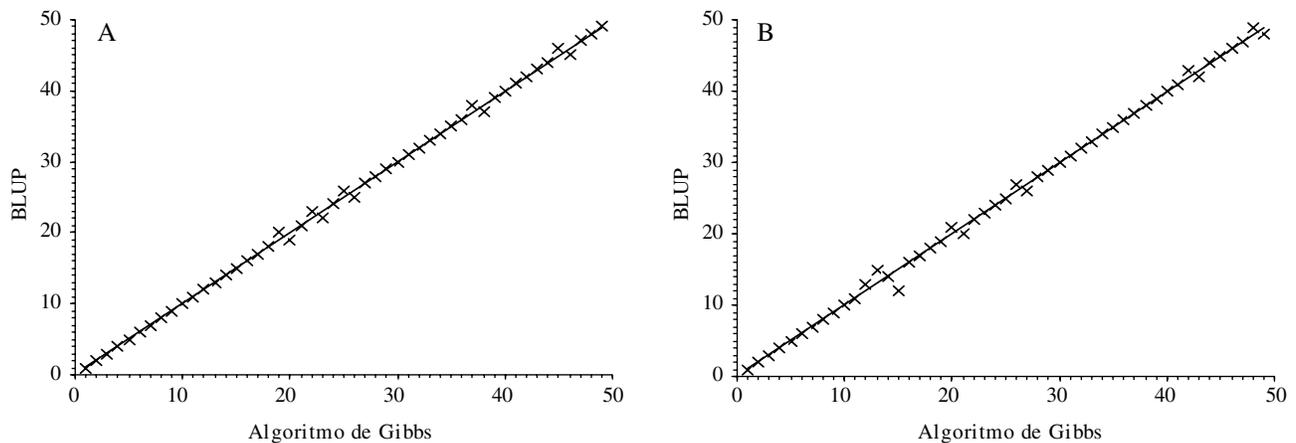


Figura 2. Gráfico de comparación entre los valores del ranking de los árboles genitores (47 árboles creciendo en Australia y 2 árboles en Chile) basado en las informaciones de sus descendientes, para las tasas de crecimiento en altura (A) y diámetro (B). La línea recta entrega una medida de relación entre los procedimientos de la mejor predicción lineal insesgada (BLUP) y el algoritmo de Gibbs, contexto bayesiano.

Graphic with the ranking values for the progenitors (47 trees growing in Australia and 2 trees growing in Chile) based on information of the progenies, for height (A) and diameter (B) growth rate. The line provides an indication of relationship between Best Linear Unbiased Prediction (BLUP) and Gibbs sampling, Bayesian framework.

Cuadro 3. Valores promedio de la distribución *a posteriori* de los componentes de varianza y la heredabilidad, estimados a través de inferencia bayesiana, usando el algoritmo de Gibbs. CI y CS son los límites de confianza inferior y superior (95%) determinados a partir de la distribución de los parámetros.

Bayesian posterior means for variance components and heritability, using Gibbs sampling. CI and CS are the lower and upper cutoff for 95% credible sets, from the posterior distributions of the parameters.

Parámetro	Tasa de crecimiento en altura			Tasa de crecimiento en diámetro		
	Valor	CI	CS	Valor	CI	CS
Varianza aditiva	0,7387	0,5800	0,8600	0,0566	0,0500	0,0600
Varianza residual	1,0022	0,9000	1,1300	0,0755	0,0700	0,0800
Varianza fenotípica	1,7408	1,6800	1,8000	0,1321	0,1300	0,1400
Heredabilidad	0,4215	0,3415	0,4831	0,4266	0,3846	0,4628

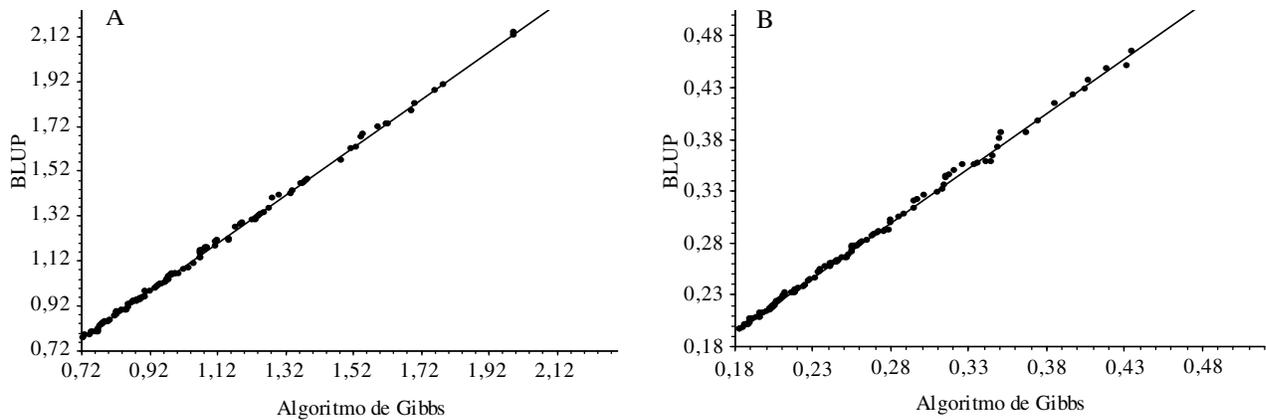


Figura 3. Comparación de los valores genéticos predichos, de una muestra de árboles seleccionados (intensidad de selección de 8,16%) para la tasa de crecimiento en altura (A) y diámetro (B), usando algoritmo de Gibbs y la mejor predicción lineal insesgada (BLUP). La línea recta entrega una medida de equivalencia entre ambos procedimientos.

Comparison of predicted breeding values of selected trees (selection intensity: 8.16%) for height (A) and diameter (B) growth rates using Gibbs sampling and Best Linear Unbiased Prediction (BLUP). The line provides an indication of association between both procedures.

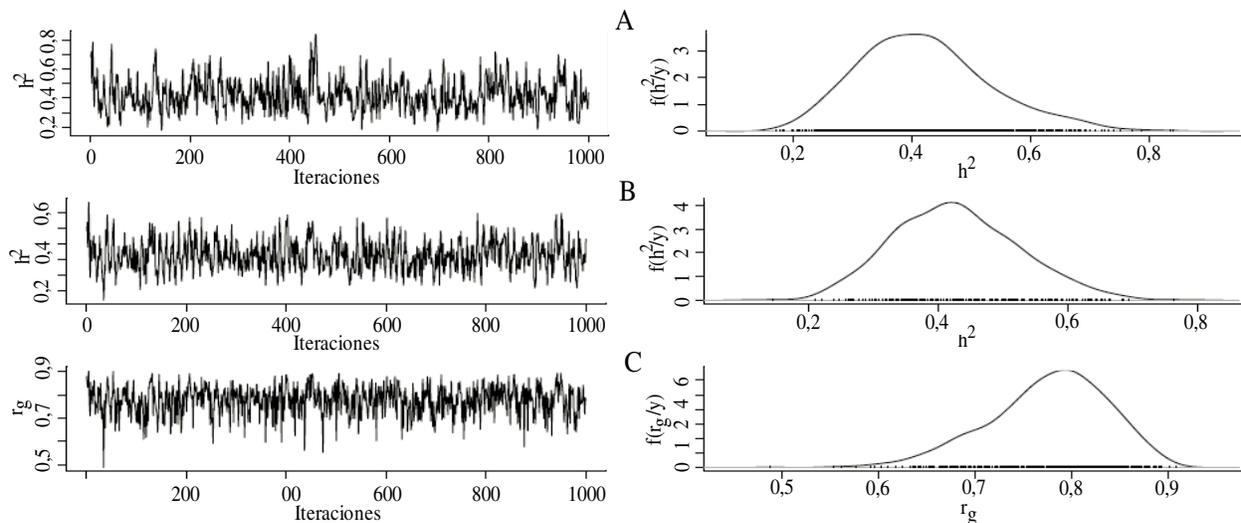


Figura 4. Distribución de densidad *a posteriori* (derecha) de la heredabilidad, para la tasa de crecimiento en altura (A) y crecimiento en diámetro (B); y correlación genética entre ambas características (C). A la izquierda se presenta el resultado de las iteraciones del algoritmo de Gibbs para cada parámetro genético.

Marginal posterior distributions (right) of heritabilities for height (A) and diameter (B) growth rates, and genetic correlation between both traits (C). To the left are presented the iterations of Gibbs chain for each genetic parameter.

características, con valores aproximados de $h^2 = 0,4215$ y $0,4266$, para la tasa de crecimiento en altura y diámetro, respectivamente. Las estimaciones por intervalo de la heredabilidad, llamados en el caso bayesiano de “regiones de credibilidad” (Blasco 2001), incluyen los valores puntuales entre ambas características. Estas estimaciones por intervalo asumen un 95% de probabilidad, y son calculadas a través de las distribuciones *a posteriori*.

Las distribuciones *a posteriori* de las heredabilidades y de la correlación genética entre ambas características son mostradas en la figura 4, así como gráficos con las iteraciones de las cadenas del algoritmo de Gibbs. Al contrario del algoritmo de cadenas independientes (Wolfinger y Kass 2000) las cadenas generadas por el algoritmo de Gibbs deben evidenciar convergencia para la obtención de valo-

res fidedignos de los parámetros (Van Tassell y Van Vleck 1996). La convergencia de una cadena MCMC puede ser evidenciada a través del método desarrollado por Heidelberger y Welch (1983). Basado en este principio, en el presente estudio las cadenas de Gibbs consiguieron la convergencia al considerar un total de 85.000 muestras, con 5.000 de periodo de descarte (*burn-in*) para evitar la influencia de los valores iniciales, y con un intervalo de 80 muestras para la conformación de la cadena definitiva. Este último intervalo es realizado para evitar la dependencia (o correlación) entre muestras contiguas (Van Tassell y Van Vleck 1996). Existe una fuerte asociación entre la complejidad del modelo estadístico asociado al experimento y los resultados de la convergencia de la cadena de Gibbs, de un parámetro de interés. En general, a mayor complejidad del modelo mayor

es la dificultad para alcanzar la convergencia de la cadena, la cual puede estar relacionada con la alta correlación de los elementos de ésta (Nogueira *et al.* 2003).

Los valores de la moda de las heredabilidades de la tasa de crecimiento en diámetro y altura (datos no mostrados) fueron similares a los valores de las estimativas puntuales del promedio, evidenciando una relativa simetría en las distribuciones *a posteriori* de dicho parámetro genético (figura 4). Los resultados de las estimaciones de las ganancias genéticas predichas y la diversidad genética después de la selección son mostrados en el cuadro 4, considerando el promedio del ensayo como línea base. En el presente estudio se consideraron dos estrategias de mejoramiento de acuerdo con Furlani *et al.* (2005): a) de corto plazo, que contempla una selección para generar una población de producción de semillas, cuya intensidad de selección es de 8,16%; y b) de largo plazo, para generar una población de mejoramiento, con intensidad de selección de 17,01%. En ambas estrategias se observó que tanto la ganancia genética como la diversidad genética después del proceso de selección resultaron ser mayores en la tasa de crecimiento en diámetro.

Cuadro 4. Resultados de las estimaciones de ganancia genética predicha, en relación al promedio del experimento, y diversidad genética después de la selección, por característica y estrategia de mejoramiento (corto o largo plazo) (Furlani *et al.* 2005).

Predicted genetic gain in relation to the experimental average, and genetic diversity after selection, per trait and breeding strategy (short and long term) (Furlani *et al.* 2005).

Parámetro	Tasa de crecimiento en altura		Tasa de crecimiento en diámetro	
	Corto plazo	Largo plazo	Corto plazo	Largo plazo
Ganancia genética (%)	22,89	17,31	27,69	21,73
Diversidad genética	0,361	0,441	0,415	0,550

DISCUSIÓN

La predicción del efecto familiar a través de técnicas bayesianas, utilizando el algoritmo de cadenas independientes, una variante de los métodos de Monte Carlo vía cadenas de Markov (MCMC), puede ser considerada equivalente al procedimiento de estimación/predicción REML/BLUP. Este resultado coincide con otros estudios de evaluación genética (Mora y Arnhold 2006, Mora *et al.* 2007) donde se estimaron altas y significativas correlaciones entre BLUP y este algoritmo en la predicción de efectos aleatorios en un modelo mixto. A su vez, basado en el ordenamiento de diferentes genotipos (*ranking*) de *Acacia saligna* (Labill.) Wendl., Mora *et al.* (2006a) también muestran equivalencias entre los procedimientos en ensayos establecidos en el norte de Chile.

Los límites de confianza en el contexto bayesiano (o regiones de credibilidad) pueden ser usados como una alternativa de estimación de los parámetros genéticos, incluyendo los efectos genotípicos, los cuales son obteni-

dos directamente de la distribución *a posteriori* (Wright *et al.* 2000, Nogueira *et al.* 2003). A modo de ejemplo, en el contexto clásico (REML/BLUP) no existe ninguna evidencia estadística que imposibilite seleccionar las 11 mejores familias (equivalentes al 22,45%), en función de los efectos genotípicos de la tasa de crecimiento en altura (resultados mostrados en el cuadro 1). Alternativamente, si se desea trabajar utilizando estimaciones por intervalo en el contexto bayesiano, las familias 19, 28, 13 y 2 (cuadro 1) podrían no ser seleccionadas ya que dentro de la región de los límites de confianza se incluyen valores negativos de predicción, indicando que los valores de las predicciones podrían ser inferiores al promedio del ensayo (línea base). Consecuentemente, las siete primeras familias serían seleccionadas (equivalentes a 14,3%) en función del valor genético de la tasa de crecimiento en altura, estimado a través de la utilización de las regiones de credibilidad. A su vez, para la tasa de crecimiento en diámetro, la intensidad de selección sería más severa, pasando a 10,2% (cinco familias), al adoptar esta técnica de predicción bayesiana. Intervalos de confianza en el contexto tradicional pueden ser construidos a través de técnicas semiparamétricas de *bootstrap*, como es definido por Efron (1979). Sin embargo, en el caso de BLUP, Morris (2002) aclara que algunas propiedades deseadas de este procedimiento podrían ser perdidas.

El alto coeficiente de correlación de Spearman estimado a partir del ranking de árboles individuales está indicando que el ordenamiento es similar en ambas características. La fuerte asociación entre éstas es confirmada por el valor promedio *a posteriori* de la correlación genética, determinada a través del algoritmo de Gibbs. Desde el punto de vista del mejoramiento, la selección de árboles de *E. cladocalyx* con 30 meses de edad, basado en la tasa de crecimiento en altura, podría tener un significativo efecto genético sobre la tasa de crecimiento en diámetro, o viceversa. Tales características probablemente estén bajo la influencia de genes aditivos en común (Wu *et al.* 2003). La fuerte asociación entre características de crecimiento vegetativo, en árboles forestales, ha sido frecuente e históricamente reportada en la literatura (White y Hodge 1989), aunque la relación significativa entre la tasa de crecimiento en diámetro y altura podría ser perdida en edades más avanzadas.

En el presente estudio, una alta correlación (asociación) entre los procedimientos BLUP y algoritmo de Gibbs es observada en la predicción de los efectos genéticos aditivos de los árboles individuales. En el mejoramiento vegetal, diversos estudios confirman esta fuerte asociación entre el enfoque bayesiano, vía Gibbs, y el procedimiento tradicional de estimación/predicción. Soria *et al.* (1998), por ejemplo, mencionan que los valores genéticos determinados a través de BLUP o por técnicas bayesianas, utilizando el algoritmo de Gibbs, no presentan diferencias significativas para predecir la superioridad genética de árboles selectos de *Eucalyptus globulus* Labill. Aunque también pueden existir ciertas discrepancias entre tales metodologías al estimar parámetros genéticos (Soria *et al.* 1998), y se atribuyen al ajuste de las estimaciones de máxima verosimilitud residual. También recientemente Waldmann y Ericsson (2006) encontraron

pocas diferencias entre REML y el algoritmo de Gibbs en un análisis con datos simulados de un ensayo genético con *Pinus sylvestris* L., aunque concluyen que estudios adicionales de simulación serían necesarios para evaluar el efecto de las diferentes distribuciones *a priori* (*prior*) de los componentes de varianza en un modelo genético de árboles individuales. Finalmente, Blasco (2001) demostró que BLUP puede ser considerado un estimador bayesiano, construido usando una distribución *a priori* uniforme (*flat*) para los efectos ambientales, y una distribución *a priori* normal para los efectos genéticos.

Los valores relativamente altos de las heredabilidades individuales, o grado de control genético aditivo, estimados en este estudio, sugieren que métodos simples de selección pueden ser implementados. Minvielle (1990) menciona que cuando el valor estimado de la heredabilidad es alto, indica que la correlación entre el fenotipo y genotipo de los dos individuos también es alta, pudiendo realizarse una selección con base en el desempeño del propio individuo. Ginwal *et al.* (2004) estimaron altas heredabilidades del crecimiento en familias de *Eucalyptus camaldulensis* Dehn. en condiciones semiáridas de la India, mostrando ser una característica altamente heredable cuando los árboles se plantan en condiciones de estrés hídrico.

Cabe destacar las severas condiciones de clima encontradas en esta región del país, con un marcado déficit hídrico (balance entre las precipitaciones y la evapotranspiración) donde es demostrado que otras especies del mismo género que requieren mayor disponibilidad hídrica muestran ser escasamente heredables para características de crecimiento (Chambers *et al.* 1996).

Contrariamente a las estimativas de heredabilidad, para la correlación genética entre las características en análisis, la moda es marcadamente diferente del promedio (figura 4), evidenciando una mayor asimetría en su distribución *a posteriori*, tal como en Van Tassell *et al.* (1998). Distribuciones marginales *a posteriori* asimétricas de los parámetros genéticos, basadas en el algoritmo de Gibbs, son frecuentemente reportadas en la literatura (Blasco 2001, Nogueira *et al.* 2003, Waldmann y Ericsson 2006). Wright *et al.* (2000) afirman que la moda, como una medida de posición de las distribuciones *a posteriori*, sería el punto donde se obtienen las mayores frecuencias, y por lo tanto podría ser un indicador eficiente en la determinación de parámetros puntuales de la distribución *a posteriori*.

Una mayor diversidad genética puede ser mantenida de la diversidad original de la población experimental, al seleccionar árboles individuales basados en la tasa de crecimiento en diámetro, con un valor que puede llegar a 55% en la estrategia de mejoramiento de largo plazo. Las ganancias genéticas son superiores para la tasa de crecimiento del diámetro basal. Ambas características presentan ganancias predichas consideradas significativas, siendo un resultado altamente promisorio, considerando que el ensayo se encuentra en una zona prioritaria para la forestación con la especie dados los requerimientos de clima y suelo (Mora y Perret 2002), abarcando gran parte de la comuna de Los Vilos, lugar donde se condujo el presente ensayo. La utilización de la variabilidad genética, encontrada en

esta población de *E. cladocalyx*, no sólo permite lograr un avance genético directo, a través de plantaciones forestales operacionales, sino que también sirve como base genética para el desarrollo de programas de hibridaciones interespecíficas, para mejorar características como la tolerancia al déficit hídrico (Mora *et al.* 2005), resistencia a *Phoracantha semipunctata* Fab. (Coleoptera: Cerambycidae). (Hanks *et al.* 1994), o tolerancia a suelos salinos (Marcar *et al.* 1995) en otras especies de *Eucalyptus* de interés económico (Teixeira *et al.* 1995, Geldres y Schlatter 2004, Santelices 2005). Para ello, material seleccionado podría ser utilizado en estos programas de mejoramiento.

CONCLUSIONES

La variabilidad genética asociada a la tasa de crecimiento de los genotipos de *E. cladocalyx* puede expresar una importante ganancia si las plantaciones se establecen con el material seleccionado de esta población de mejoramiento. Árboles que expresan una rápida ocupación del espacio disponible es un aspecto fundamental en ambientes desfavorables.

La predicción de los efectos genotípicos utilizando las técnicas bayesianas, ya sea considerando el efecto familiar o los efectos genéticos aditivos, es equivalente al procedimiento tradicional de estimación/predicción REML/BLUP, pudiendo ser una alternativa metodológica efectiva para la evaluación genética de árboles forestales, ya que considera la variabilidad de los parámetros genéticos a través de las distribuciones *a posteriori*.

AGRADECIMIENTOS

Los autores agradecen a la hacienda Caracas, Los Vilos, por el aporte de infraestructura y cuidados del ensayo, y al Fondo de Desarrollo e Innovación – FDI, por el aporte financiero de la presente investigación.

REFERENCIAS

- Apiolaza LA, AR Gilmour, DJ Garrick. 2000. Variance modeling of longitudinal data from progeny tests. *Canadian Journal of Forest Research* 30:645-654.
- Arnold R, J Carmona, C Balocchi. 1991. Programa de propagación clonal para *Eucalyptus globulus* en Chile. *Bosque* 12(1):3-9.
- Blasco A. 2001. The Bayesian controversy in animal breeding. *Journal of Animal Science* 79:2023-2046.
- Burdon RD. 1977. Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genetica* 26: 168-175.
- Boldman KG, LA Kriese, LD Van Vleck, CP Van Tassell, SD Kachman. 1995. A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances. Lincoln, United States. USDA and ARS. 120 p.
- Cappa EP, RJC Cantet. 2006. Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. *Canadian Journal of Forest Research* 36(5):1276-1285.

- Chambers PGS, NMG Borralho, BM Potts. 1996. Genetic analysis of survival in *Eucalyptus globulus* spp. *globulus*. *Silvae Genetica* 45: 107-112.
- De Veer C, C Balocchi, P Crespell, G Hodge, F Droppelmann. 2001. BIOCAT: Un programa para el análisis de ensayos genéticos forestales. In Simposio internacional IUFRO. Valdivia, Chile, 10 al 15 Septiembre 2001.
- Efron B. 1979. Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7:1-26.
- Furlani RCM, MLT Moraes, MDV Resende, E Furlani-Junior, PS Gonçalves, WV Valério-Filho, JR Paiva. 2005. Estimation of variance components and prediction of breeding values in rubber tree breeding using the REML/BLUP procedure. *Genetics and Molecular Biology* 28:271-276.
- Geldres E, JE Schlatter. 2004. Crecimiento de las plantaciones de *Eucalyptus globulus* sobre suelos rojo arcillosos de la provincia de Osorno, Décima Región. *Bosque* 25(1):95-101.
- Ginwal HS, P Kumar, VK Sharma, AK Mandal. 2004. Seed source variation in growth performance of *Eucalyptus camaldulensis* Dehnh. of Australian origin in India. *Silvae Genetica* 53(4):182-186.
- Gutiérrez B, R Ipinza. 2000. Evaluación de parámetros genéticos en *Nothofagus*. In Ipinza R, B Gutiérrez, V Emhart eds. Domesticación y Mejora Genética de Raulí y Roble. Valdivia, Chile. Exion. p. 371-390.
- Heidelberger P, PD Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31:1109-1114.
- Ipinza R, B Gutiérrez, V Emhart. 1998. Mejora Genética Forestal Operativa. Valdivia, Chile. Universidad Austral de Chile, Facultad de Ciencias Forestales. 422 p.
- Lu PX, TL White, DA Huber. 1999. Estimating type B genetic correlations with unbalanced data and heterogeneous variances for half-sib experiments. *Forest Science* 45:562-572.
- Marcar NE, DF Crawford, PM Leppert, T Jovanovic, R Floyd, R Farrow. 1995. Trees for saltland: a guide to selecting native species for Australia. Melbourne, Australia. CSIRO. 72 p.
- Matheson AC, ME Devey, TL Gordon, W Werner, DR Vogler, C Balocchi, MJ Carson. 2006. Heritability of response to inoculation by pine pitch canker of seedlings of radiata pine. *Australian Forestry Journal* 70(2):101-106.
- Minvielle F. 1990. Principes d'amélioration génétique des animaux domestiques. Paris, Francia. INRA. 211 p.
- Mora F. 2004. Análisis preliminar de la heredabilidad e interacción genotipo-ambiente del crecimiento, en dos ensayos de progenie de *Eucalyptus cladocalyx* (F. Muell) establecidos en la zona árida de Chile. Tesis de Magíster en Ciencias Biológicas. La Serena, Chile. Facultad de Ciencias, Universidad de La Serena. 71 p.
- Mora F. 2006. Heredabilidad y valor genético (REML/BLUP) en genotipos de un eucalipto tolerante a la sequía, en el norte de Chile. *Ciencia Florestal* 16:145-151.
- Mora F, E Arnhold. 2006. Inferencia bayesiana y metodología de modelos lineales mixtos aplicados al mejoramiento de maíz. *Ciencia e Investigación Agraria* 33(3):217-223.
- Mora F, P Jara-Seguel, C Palma-Rojas. 2005. Comparación del cariotipo de *Eucalyptus globulus* y *Eucalyptus cladocalyx* (Myrtaceae). *Agricultura técnica* 65(1):20-25.
- Mora F, S Perret. 2002. *Eucalyptus cladocalyx* en la Región de Coquimbo: especie maderera para el negocio frutícola. *Revista Tierra Adentro* 46:28-29.
- Mora F, S Perret, CA Scapim, EN Martins, MP Molina. 2006a. Inferencia bayesiana, vía algoritmo de cadenas independientes, en el análisis genético de especies perennes. In Proceedings of second Latin American IUFRO congress. La Serena, Chile, October 23 to 27.
- Mora F, CA Scapim, A Minuzzi, R Meneses, S Perret, F Tapia. 2006b. Enfoque bayesiano na avaliação genética de milho pipoca, via algoritmo de cadeias independentes. *Arquivos do Mudi* 10(2):174.
- Mora F, O Pupim-Junior, CA Scapim. 2007. Predicción del efecto de cultivares de algodón en la presencia de interacción genotipo-ambiente. *Ciencia e Investigación Agraria* 34:11-19.
- Morris JS. 2002. The BLUPs are not "best" when it comes to bootstrapping. *Statistics & Probability Letters* 56(4):425-430.
- Nogueira DA, T Sáfiadi, E Bearzoti, JSS Bueno-Filho. 2003. Análises clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. *Ciênc. agrotec* 27: 1614-1624.
- Perret S, F Mora, MP Molina. 2006. Predicción de valores genéticos vía REML/BLUP en familias de *Eucalyptus cladocalyx* establecidas en el norte de Chile. In Proceedings of second Latin American IUFRO congress. La Serena, Chile, October 23 to 27.
- Santelices R. 2005. Desarrollo de una plantación de *Eucalyptus globulus* establecida en primavera con diferentes tratamientos de riego. *Bosque* 26(3):105-112.
- Schenone RA, L Storck, MP Martins-Corder. 2002. Variación genética en procedencias/progenies de polinización abierta de *Eucalyptus dunnii* Maiden. *Bosque* 23(1):79-89.
- Searle SR, G Casella, CE McCulloch. 1992. Variance components. New York, United States. John Wiley. 501 p.
- Soria F, F Basurco, G Toval, L Silió, MC Rodríguez, M Toro. 1998. An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. *Canadian Journal of Forest Research* 28(9):1286-1294.
- Teixeira P, P Leles, N de Barros, R Ferreira. 1995. Nutrición potásica y relaciones hídricas en mudas de *Eucalyptus* spp. *Bosque* 16(1):61-68.
- Van Tassell CP, LD Van Vleck. 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.
- Van Tassell CP, LD Van Vleck, HE Gregory. 1998. Bayesian analysis of twinning and ovulation rates using a multiple-trait threshold model and Gibbs sampling. *Journal of Animal Science* 76:2048-2061.
- Waldmann P, T Ericsson. 2006. Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. *Theor. Appl. Genet.* 112:1441-1451.
- White TL, GR Hodge. 1989. Predicting breeding values with applications in forest tree improvement. Dordrecht, The Netherlands. Kluwer. 367 p.
- Wolfinger RD, RE Kass. 2000. Non-conjugate Bayesian analysis of variance component models. *Biometrics* 56: 768-774.
- Wright DR, HS Stern, PJ Berger. 2000. Comparing traditional and Bayesian analyses of selection experiments in animal breeding. *Journal of Agricultural, Biological, and Environmental Statistics* 5:240-256.
- Wu R, CX Ma, W Zhao, G Casella. 2003. Functional mapping for quantitative trait loci governing growth rates: a parametric model. *Physiol. Genomics* 14:241-249.
- Zamudio F, R Baettyg, A Vergara, F Guerra, P Rozenberg. 2002. Genetic trends in wood density and radial growth with cambial age in a radiata pine progeny test. *Ann. For. Sci.* 59:541-549.