

FREDDY LUIS MORA POBLETE

**Abordagem bayesiana na avaliação genética de plantas perenes  
e modelos lineares generalizados aplicados na seleção de  
cultivares e no mapeamento de QTLs**

MARINGÁ  
PARANÁ – BRASIL  
OUTUBRO DE 2008

FREDDY LUIS MORA POBLETE

**Abordagem bayesiana na avaliação genética de plantas perenes  
e modelos lineares generalizados aplicados na seleção de  
cultivares e no mapeamento de QTLs**

Tese apresentada à Universidade Estadual de Maringá como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de Doutor.

MARINGÁ  
PARANÁ – BRASIL  
OUTUBRO DE 2008

Dados Internacionais de Catalogação-na-Publicação (CIP)  
(Biblioteca Central - UEM, Maringá – PR., Brasil)

M827a Mora Poblete, Freddy Luis  
Abordagem bayesiana na avaliação genética de plantas perenes e modelos lineares generalizados aplicados na seleção de cultivares e no mapeamento de QTLs / Freddy Luis Mora Poblete. -- Maringá, 2008.

151 f. : il.

Orientador : Prof. Dr. Carlos Alberto Scapim.  
Tese (doutorado) - Universidade Estadual de Maringá, Programa de Pós-graduação em Genética e Melhoramento, 2008.

1. Bioestatística - Modelos Lineares Generalizados. 2. Bioestatística - Avaliação genética. 3. Plantas - Melhoramento - Mapeamento QTLs. 4. Plantas - Melhoramento - Inferência Bayesiana. 5. Plantas - Melhoramento - Análise longitudinal. 6. Bioestatística - Inferência Bayesiana. I. Universidade Estadual de Maringá. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 21.ed. 576.58

FREDDY LUIS MORA POBLETE

**Abordagem bayesiana na avaliação genética de plantas perenes  
e modelos lineares generalizados aplicados na seleção de  
cultivares e no mapeamento de QTLs**

Tese apresentada à Universidade Estadual de Maringá, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, na área de concentração em Genética e Melhoramento, para obtenção do título de Doutor.

APROVADA em 23 de outubro de 2008.



Prof. Dr. Alexandre Pio Viana



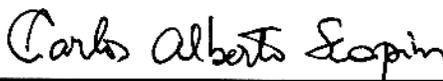
Prof.<sup>a</sup> Dr.<sup>a</sup> Terezinha Aparecida Guedes



Prof. Dr. Elias Nunes Martins



Prof. Dr. Ronald José Barth Pinto



Prof. Dr. Carlos Alberto Scapim  
Orientador

## **DEDICO**

Aos meus filhos, Javiera Ignacia e Felipe Andrés, cujas especiais presenças me motivam a tentar ser cada dia melhor.

A Juanita Bruneau, pelo fundamental apoio nesta iniciativa.

Aos meus pais, José Mora e Silvia Poblete, e ao meu irmão, Carlitos, por fazerem parte de minha vida.

## AGRADECIMENTOS

À Universidade Estadual de Maringá (UEM) e ao Programa de Pós-Graduação em Genética e Melhoramento (PGM), pela oportunidade de realização do Curso de Doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelo apoio financeiro concedido durante o curso.

Ao professor doutor Carlos Alberto Scapim, pelas idéias centrais das temáticas abordadas nesta tese e por eficientemente orientar os trabalhos, permitindo-me ir muito além dos objetivos propostos.

Ao professor doutor Elias Nunes Martins, cuja colaboração foi crucial para melhorar o meu entendimento sobre a avaliação genética, usando a inferência Bayesiana e a metodologia dos Modelos Lineares Mistos. Sou muito grato por ele ter aceitado minha participação no grupo de melhoramento do Programa de Pós-Graduação em Zootecnia (PPZ) e também como aluno em suas disciplinas, o que possibilitou o enriquecimento de minhas experiências.

Ao diretor de pesquisa Raúl Meneses Rojas e ao pesquisador Francisco Tapia Contreras, do Instituto de Investigaciones Agropecuárias da Região de Coquimbo (INIA-Intihuasi), Ministério de Agricultura do Chile, por fornecer gentilmente os dados experimentais e as informações dos programas de melhoramento de *Olea europaea*, conduzidos no sul do Deserto de Atacama, os quais foram pontos-chave para a realização dos Capítulos II, IV e VI desta tese.

À pesquisadora Sandra Perret Durán, do Instituto de Investigación Forestal, Ministerio de Agricultura do Chile, pelo apoio, incentivo e por conceder as informações e dados dos ensaios provenientes dos programas de melhoramento florestal conduzidos no norte do Chile, contribuindo grandemente com a elaboração dos Capítulos V e VII.

À professora doutora Roslyn Gleadow, da Monash University, Austrália, por suas contribuições na elaboração do Capítulo VII.

À professora doutora Kathrin F. Stock, University of Veterinary Medicine Hannover, Alemanha, pela valiosa troca de informações relativas ao mapeamento de QTLs e à avaliação genética Bayesiana.

Ao professor doutor Ivan Schuster, da COODETEC – Cascavel PR, pelos ensinamentos na área de estatística genômica.

À professora doutora Rosangela Santana, do Departamento de Estatística por me aceitar como aluno na disciplina “Modelos Lineares Generalizados”, cujo aprendizado muito enriqueceu minhas experiências profissionais.

Aos amigos que contribuíram desinteressadamente com esta tese: Alexandre Leseur dos Santos, pelos ensinamentos dos programas de avaliação genética MTDFREML e MTGSAM; e Cristiane Zocatelli Ribeiro, Elisangela Mendes dos Santos, Lorena de Souza e Alexandra Inês dos Santos, que contribuíram grandemente fazendo traduções dos capítulos para a Língua Portuguesa.

Aos editores e revisores anônimos das Revistas Euphytica (Holanda), Ciencia e Investigación Agraria (Chile), Agrociencia (México), Crop Breeding and Applied Biotechnology (Brasil) e Journal of Central European Agriculture (Croácia), cujas sugestões contribuíram para a melhoria dos artigos resultantes desta tese.

Aos professores doutores membros da banca de defesa de tese, Ronald José Barth Pinto, Terezinha Aparecida Guedes, Elias Nunes Martins e Alexandre Pio Viana, pelo apoio e pelos importantes conselhos para esta pesquisa.

Ao Fondo de Desarrollo e Innovación, Innova Chile e Fondo de Desarrollo Regional, pertencentes ao Governo do Chile, que viabilizaram financeiramente os projetos de pesquisa dos Capítulos III, IV, V, VI e VII.

Aos amigos e empresários chilenos, Enso Ferrario e Héctor Troncoso Reyes, os quais me apoiaram com recursos financeiros para iniciar minha viagem e para o sustento dos primeiros oito meses do curso.

Finalmente, aos amigos, cuja lista de nomes ultrapassam o espaço disponível neste material, meu sincero muito obrigado. Sem o apoio das pessoas que passaram por mim, deixando profundas marcas pessoais e profissionais, eu não teria chegado ao fim desta tão sonhada jornada.

## BIOGRAFIA

Graduou-se em Engenharia florestal pela Facultad de Ciencias Forestales da Universidad de Talca, no Chile, com tema de pesquisa baseado na avaliação genética de árvores florestais da espécie *Pinus radiata*, utilizando dois experimentos genéticos da empresa florestal MININCO S.A.

Foi pesquisador e co-diretor de projetos de pesquisa no Instituto de Investigaciones Forestales (INFOR), pertencente ao Ministério de Agricultura do Chile, onde desenvolveu estudos e programas relacionados ao melhoramento genético de plantas, visando à seleção de plantas para incrementar a tolerância ao déficit hídrico, na parte sul do Deserto de Atacama, no Chile. O trabalho de pesquisa desenvolvido também incluiu a utilização de métodos quantitativos úteis para aprimorar as análises da experimentação agropecuária e florestal e das etapas de avaliação genética nos programas de melhoramento. Trabalhou em projetos de pesquisa relacionados aos recursos genéticos de zonas áridas e semi-áridas do Chile, no Instituto de Investigaciones Agropecuarias (INIA), pertencente ao Ministério de Agricultura do Chile. Atuou como professor da Universidad Del Mar, La Serena, Chile, no Curso de Graduação em Agronomia, ministrando a disciplina Genética Vegetal.

Concluiu o Curso de Pós-graduação, em nível de Mestrado em Ciências Biológicas, na Universidad de La Serena, no Chile, na área de Genética Quantitativa. Para a pesquisa, utilizou a metodologia dos modelos lineares mistos na avaliação de árvores/progênes de espécies vegetais úteis para a região árida do Chile (Métodos da Máxima Verossimilhança Restrita e da Melhor Predição Linear não Tendenciosa – REML/BLUP).

A pesquisa de Doutorado, realizada na Universidade Estadual de Maringá, na área de Genética e melhoramento, foi focalizada na utilização da inferência bayesiana na avaliação genética de plantas e nos modelos lineares generalizados aplicados à seleção de cultivares e mapeamento de QTLs. Neste trabalho, confirmou-se a efetividade e ampla aplicabilidade do método generalizado na área de melhoramento de plantas. A abordagem bayesiana proveu úteis informações para propósitos de melhoramento e para entender como as características quantitativas de interesse são controladas geneticamente em diversas espécies de plantas.

## ÍNDICE

RESUMO.....	x
ABSTRACT.....	xii
1. INTRODUÇÃO GERAL.....	1
2. REVISÃO DE LITERATURA.....	5
2.1. Modelos lineares e não lineares.....	5
2.2. Modelos lineares generalizados.....	6
2.3. Modelos lineares generalizados e dados longitudinais.....	9
2.4. Inferência Bayesiana.....	11
3. REFERÊNCIAS BIBLIOGRÁFICAS.....	16
CAPÍTULO I ANÁLISE MULTIVARIADA NO MAPEAMENTO DE <i>LOCI</i> DE CARACTERÍSTICAS QUANTITATIVAS (QTL), USANDO MODELOS MISTOS E EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS.....	20
RESUMO.....	20
ABSTRACT.....	21
1. INTRODUÇÃO.....	22
2. MATERIAL E MÉTODOS.....	24
2.1. Estudo de simulação.....	24
2.2. QTL controlando características de distribuição normal.....	25
2.3. QTL controlando características de distribuição binomial.....	26
3. RESULTADOS E DISCUSSÃO.....	27
4. CONCLUSÕES.....	33
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	34
CAPÍTULO II REGRESSÃO LINEAR GENERALIZADA NO MAPEAMENTO DE <i>LOCI</i> DE CARACTERÍSTICAS QUANTITATIVAS (QTL) POR INTERVALOS COMPOSTOS.....	37
RESUMO.....	37
ABSTRACT.....	38
1. INTRODUÇÃO.....	39
2. MATERIAL E MÉTODOS.....	41
3. RESULTADOS E DISCUSSÃO.....	44
4. CONCLUSÕES.....	49
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	50
CAPÍTULO III MODELOS LINEARES GENERALIZADOS NA AVALIAÇÃO DE <i>Olea europaea</i> CV. ARBEQUINA, PICUAL E FRANTOIO NO SUL DO DESERTO DE ATACAMA.....	52
RESUMO.....	52
ABSTRACT.....	53
1. INTRODUÇÃO.....	54
2. MATERIAIS E MÉTODOS.....	57
2.1. Ensaio de campo.....	57
2.2. Variáveis analisadas.....	57
2.3. Análise estatística para a produção de frutos.....	58
2.4. Análise estatística para sobrevivência e precocidade.....	59
3. RESULTADOS E DISCUSSÃO.....	61
3.1. Produção de frutos.....	61
3.2. Precocidade.....	65
3.3. Sobrevivência.....	66
4. CONCLUSÕES.....	68

5. REFERÊNCIAS BIBLIOGRÁFICAS .....	69
CAPÍTULO IV AVALIAÇÃO DE SEIS CULTIVARES DE OLIVEIRA USANDO EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS, MODELOS MISTOS E MODELOS GENERALIZADOS .....	
RESUMO .....	72
ABSTRACT .....	73
1. INTRODUÇÃO .....	74
2. MATERIAL E MÉTODOS .....	76
2.1. Localização do estudo .....	76
2.2. Delineamento experimental e medições .....	77
2.3. Análise estatística .....	77
3. RESULTADOS E DISCUSSÃO .....	80
4. CONCLUSÕES .....	86
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	87
CAPÍTULO V VARIAÇÃO DE PROCEDÊNCIAS DE <i>Acacia saligna</i> NA ZONA ÁRIDA DO CHILE, USANDO UMA ABORDAGEM BAYESIANA .....	
RESUMO .....	90
ABSTRACT .....	91
1. INTRODUÇÃO .....	92
2. MATERIAIS E MÉTODOS .....	94
2.1. Descrição do local e o ensaio .....	94
2.2. Variáveis e genótipos .....	94
2.3. Abordagem Bayesiana .....	94
3. RESULTADOS E DISCUSSÃO .....	97
4. CONCLUSÕES .....	102
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	103
CAPÍTULO VI AVALIAÇÃO GENÉTICA DE CLONES DE OLIVEIRA NO DESERTO DO ATACAMA, CHILE, USANDO INFERÊNCIA BAYESIANA .....	
RESUMO .....	105
ABSTRACT .....	106
1. INTRODUÇÃO .....	107
2. MATERIAIS E MÉTODOS .....	109
3. RESULTADOS E DISCUSSÃO .....	113
4. CONCLUSÕES .....	119
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	120
CAPÍTULO VII HERDABILIDADE DO FLORESCIMENTO PRECOCE, CRESCIMENTO E SOBREVIVÊNCIA DE <i>Eucalyptus cladocalyx</i> , NA REGIÃO ÁRIDA DO CHILE: UMA ABORDAGEM BAYESIANA .....	
RESUMO .....	124
ABSTRACT .....	125
1. INTRODUÇÃO .....	126
2. MATERIAIS E MÉTODOS .....	129
2.1. Localização do estudo .....	129
2.2. Ensaio e genótipos .....	129
2.3. Características .....	130
2.4. Análise Bayesiana .....	130
2.5. Estimativas de parâmetros genéticos .....	131
3. RESULTADOS E DISCUSSÃO .....	133
4. CONCLUSÕES .....	141
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	142

6. CONCLUSÕES GERAIS.....	147
6.1. Modelos Lineares Generalizados.....	147
6.2. Contexto Bayesiano.....	147
APÊNDICE.....	148

## RESUMO

MORA, Freddy. D. Sc. Universidade Estadual de Maringá, outubro de 2008. **Abordagem Bayesiana na avaliação genética de plantas perenes e Modelos Lineares Generalizados aplicados na seleção de cultivares e no mapeamento de QTLs.** Professor Orientador: Dr. Carlos Alberto Scapim. Professores Conselheiros: Dr. Elias Nunes Martins e Dr. Ronald José Barth Pinto.

Os procedimentos Bayesianos e a teoria dos Modelos Lineares Generalizados (GLM) têm sido indicados como metodologias de inferência, apropriadas para a análise genética. Estas têm sido utilizadas recentemente, por exemplo, na avaliação de diversidade, em estudos de filogenia e no mapeamento de *loci* de características quantitativas (QTL). A inferência Bayesiana é também considerada como método alternativo aos Modelos Lineares Mistos clássicos. Entretanto, o uso destes procedimentos nos programas de melhoramento de plantas é ainda relativamente restrito. Assim, o objetivo do presente estudo foi examinar diferentes experimentos de melhoramento de plantas, usando tais metodologias e incluindo algumas variantes dos GLM, como os GLM-mistos e as Equações de Estimção Generalizada (GEE). O estudo confirmou a efetividade e ampla aplicabilidade dos GLM em experimentos relacionados ao melhoramento vegetal. Por meio desta técnica, foi possível: 1) o mapeamento de QTLs que controlam características binárias, medidas repetidamente sobre a mesma unidade experimental (análise longitudinal); 2) o mapeamento de QTLs que controlam características contínuas com viés da distribuição normal, usando o princípio de intervalos compostos; e 3) avaliação de experimentos com cultivares de oliveira para variáveis com distribuição Gama (produção de frutos) e Binomial (produção precoce e sobrevivência). Confirmou-se que a informação acerca da distribuição dos dados agrônômicos deve ser considerada nos programas de melhoramento de plantas, para melhorar a confiabilidade das inferências de avaliação genética. A abordagem Bayesiana foi útil para propósitos de melhoramento e para entender como as características quantitativas são controladas geneticamente em oliveira, acácia e eucalipto. Características binárias, como a sobrevivência das plantas, a

precocidade da produção e o florescimento precoce, também foram possíveis de serem incluídos nos programas de melhoramento, usando o conceito do modelo de limiar (threshold) e os diferentes métodos de simulação de Monte Carlo, via Cadeias de Markov (MCMC). A habilidade da inferência Bayesiana para prever os valores genéticos de clones, procedências, famílias e plantas individuais (modelo animal) foi considerada muito valiosa na avaliação genética de plantas.

Palavras-chave: Modelos Lineares Generalizados, avaliação genética, mapeamento de QTL, inferência Bayesiana, análise longitudinal.

## ABSTRACT

MORA, Freddy. D. Sc. Universidade Estadual de Maringá. October, 2008. **Bayesian approach in the genetic evaluation of perennial plants and Generalized Linear Models applied to cultivar selection and QTLs mapping.** Supervisor: Dr. Carlos Alberto Scapim. Assistant professors: Dr. Elias Nunes Martins, Dr. Ronald José Barth Pinto.

Bayesian procedures and Generalized Linear Models theory (GLM) have been indicated as inference methodologies appropriated for genetic analysis. These have been recently used, for instance: for diversity and phylogeny studies, and mapping of Quantitative Trait Loci (QTL). Bayesian inference is also considered as an alternative method of the classic Mixed Linear Models. However, the use of these procedures in the plant breeding programs has been relatively restricted. Thus, the analytical objective of the present study was to examine several experiments about plant breeding using such methodologies, and including variants of the GLM approach: Mixed-GLM and Generalized Estimating Equations (GEE). The study confirmed both the effectiveness and broad applicability of GLM, in experiments about plant breeding. In the current study, by using GLM, it was possible: 1. Mapping of QTLs controlling binary traits, which were measured repeatedly on the same subject (longitudinal analysis), 2. Mapping of QTLs controlling continuous trait, but biased from the normal distribution, and considering the principle of composite interval method, and 3. Evaluating field trials with olive cultivars, in situations where the response variables follow the Gama (fruit production) and Binomial (early fruit production and survival) distributions. It was confirmed that the information about agronomical data distribution should be stressed in the plant breeding programs, to improve the reliability of the genetic inference. The Bayesian approach was found to provide practical information useful for breeding purposes and for understanding how quantitative traits are controlled genetically in olive, acacia and eucalyptus. Binary traits, which were used to analyze plant survival, production precocity and early flowering, were also included in the breeding programs by using threshold models and the Monte Carlo Markov Chain (MCMC) variants. The ability of Bayesian inference to predict breeding values of clones, provenances, families and

individual plants (animal model) was found to be very valuable for the genetic evaluation of plants.

Key words: Generalized Linear Models, genetic evaluation, QTL mapping, Bayesian inference, longitudinal analysis.

## 1. INTRODUÇÃO GERAL

A estimação de componentes de variância e a predição dos valores genéticos são procedimentos freqüentes no melhoramento de características de importância econômica. A teoria dos modelos lineares mistos, aplicada dentro do contexto do melhoramento genético vegetal e animal, tem tido um importante desenvolvimento e uma ampla utilização nos programas de melhoramento e na pesquisa aplicada (Nogueira et al., 2003; Resende 2002; De Paula, 2006), o que tem possibilitado a avaliação de genótipos em diferentes programas de seleção e o melhoramento de espécies de interesse econômico em uma grande variedade de condições ambientais.

No contexto da metodologia de modelos lineares mistos, os efeitos aleatórios usualmente são preditos por meio do método da melhor predição linear não tendenciosa (BLUP; Best Linear Unbiased Prediction), usando a maximização da função de densidade conjunta que considera tanto os efeitos fixos como os aleatórios do modelo. Por sua vez, a melhor estimação linear não tendenciosa (BLUE; Best Linear Unbiased Estimation) dos efeitos fixos pode ser obtida por meio do sistema de equações de modelos mistos.

A estimação precisa dos parâmetros genéticos é de fundamental importância nos programas de melhoramento genético, pois permite predizer o valor genético dos genótipos e a identificação daqueles geneticamente superiores. Em consequência das diferenças genéticas da população, de ambiente, do tipo de análise e do método de estimação de componentes de (co) variância, entre outros, estas estimativas podem variar consideravelmente (De Paula, 2006). Neste sentido, a predição dos efeitos aleatórios do modelo misto, por exemplo, a predição dos valores genéticos das plantas, depende das estimações dos componentes de variância, os quais têm sido estimados freqüentemente usando o procedimento das esperanças de quadrados médios (ou tipo I de Henderson). Entretanto, quando não existe um balanceamento no conjunto de dados agrônômicos, o método da máxima verossimilhança restrita – REML (Restricted Maximum Likelihood) tem sido comumente preferido (Resende, 2002).

Entretanto, o procedimento Bayesiano, usando diferentes métodos de Monte-Carlo, via Cadeias de Markov (MCMC) tal como o algoritmo de Gibbs, é

indicado como alternativa ao método de estimação/predição REML/BLUP (Cappa e Cantet, 2006; Resende, 2002). De acordo com Van Tasell e Van Vleck (1996), no contexto Bayesiano as distribuições *a posteriori* dos parâmetros de interesse são impossíveis de se obter pela usual via computacional. Devido a isso, são utilizados os métodos MCMC para resolver as integrações numéricas e permitir a análise das distribuições *a posteriori* (Cadena-Meneses e Castillo-Morales, 2002). Isso tem permitido que as técnicas Bayesianas sejam utilizadas com maior frequência nas análises genéticas (Blasco, 2001; Cadena-Meneses e Castillo-Morales, 2002; Cerón-Rojas e Sahagún-Castellanos, 2007). Diversos trabalhos no melhoramento vegetal (Cappa e Cantet, 2006; Soria et al., 1998) têm enfatizado as diferenças entre o enfoque Bayesiano e Frequentista (abordagem clássica) para a obtenção dos parâmetros genéticos, os quais têm utilizado essencialmente o algoritmo de Gibbs como técnica auxiliar da inferência Bayesiana.

Os métodos MCMC permitem a análise de complexas superfícies de verossimilhança e o cálculo de distribuições *a posteriori* Bayesianas (Gonçalves-Vidigal et al., 2008). Por esta razão, Walsh (2001) afirmou que nos próximos anos será incrementado notavelmente o uso dos métodos Bayesianos, substituindo a sua contraparte da verossimilhança. O algoritmo de Gibbs e os métodos MCMC em geral permitem uma análise estatística relativamente mais fácil em modelos em que os cálculos da verossimilhança são extremamente difíceis (Walsh, 2001).

O algoritmo de Gibbs é atualmente o método MCMC mais utilizado na avaliação genética, dentro dos programas de melhoramento animal e de plantas (Blasco, 2001; Cadena-Meneses e Castillo-Morales, 2002; Cappa e Cantet, 2006; Soria et al., 1998; Van Tasell e Van Vleck, 1996). Alguns estudos têm mostrado pequenas diferenças entre as estimativas dos componentes de variâncias pelo REML e o algoritmo de Gibbs (Waldmann e Ericsson, 2006), sugerindo-se a realização de estudos adicionais com as distribuições *a priori* dos componentes de variância.

Outra metodologia estatística que está sendo utilizada na experimentação agropecuária refere-se aos Modelos Lineares Generalizados (GLM; Generalized Linear Models), desenvolvidos por Nelder e Wedderburn (1972). Este procedimento tem sido útil em várias situações, nas quais o pressuposto de normalidade da variável resposta não é alcançado (Demetrio, 2001; Dobson,

2001), como, por exemplo, em estudos cuja variável resposta é representada por meio do sucesso ou fracasso de algum evento de interesse, a qual pode ser medida como uma resposta binária. Análises de sobrevivência em plantas, resistência a doenças e estudos do florescimento precoce são alguns dos exemplos específicos de variáveis binárias. O problema com as análises de variáveis que representam contagem, proporções ou dados binários é a possibilidade de violar um ou mais supostos da análise de variância e, portanto, a inferência estatística pode afetar o resultado da pesquisa (Sokal e Rohlf, 2003).

A metodologia GLM permite que a distribuição de densidade e probabilidade da variável resposta seja qualquer distribuição que pertença à família exponencial e utiliza métodos análogos do modelo linear com dados de distribuição normal (Myers et al., 2002; Nelder e Wedderburn, 1972). O modelo linear generalizado pode ser visto como uma unificação dos modelos de regressão linear e não linear, o qual incorpora uma ampla família de distribuições da variável resposta, seja normal (Gaussiana) ou não (Myers et al., 2002).

Tanto os procedimentos Bayesianos como os princípios da modelagem linear generalizada têm sido relativamente pouco utilizados nos programas de melhoramento de plantas (Gonçalves-Vidigal et al., 2008). Os procedimentos Bayesianos têm sido aplicados recentemente, por exemplo, para examinar e avaliar diversidade genética (Holsinger, 1999, Kothera et al., 2007), examinar filogenia (Delgado-Salinas et al., 2006; Papini et al., 2007), proceder o mapeamento de locos de características quantitativas (Quantitative Trait Loci – QTL; Heath, 1997; Sillanpaa e Arjas, 1998; Ball, 2001; Broman e Speed, 2002) e realizar a avaliação genética de plantas (Gonçalves-Vidigal et al., 2008).

Estudos acerca do mapeamento de QTLs usando Modelos Lineares Generalizados são também raros. Recentemente, Yin e Zhang (2006) e Yin et al. (2005), embora no melhoramento animal, discutiram a utilidade desta ferramenta no mapeamento de QTLs que controlam características de distribuição discreta. Spyrides-Cunha et al. (2000), em plantas, também mencionam as vantagens da utilização de um modelo linear generalizado no mapeamento de QTLs, considerando características que apresentam distribuição binomial (regressão logística).

Assim, a proposta geral do presente estudo foi analisar diferentes experimentos de melhoramento de plantas, usando os seguintes princípios de

inferência científica: 1) Modelos Lineares Mistos clássicos; 2) Procedimentos Bayesianos, por meio dos algoritmos de Gibbs e de Cadeias Independentes (métodos MCMC); 3) Metodologia dos Modelos Lineares Generalizados (fixos ou mistos) e 4) Equações de Estimação Generalizada (extensão dos modelos lineares generalizados para respostas correlacionadas). Os experimentos de melhoramento considerados neste estudo incluíram: 1) avaliação genética de genótipos de eucalipto, oliveira e acácia; 2) avaliação e seleção de cultivares de oliveira; 3) mapeamento de QTLs para características de distribuição não Gaussiana e 4) mapeamento de QTLs para características de distribuição normal e binomial, medidas repetidamente no tempo (análise longitudinal).

## 2. REVISÃO DE LITERATURA

### 2.1. Modelos lineares e não lineares

Um modelo pode ser definido como uma simples abstração da realidade, o qual fornece uma aproximação de um fenômeno relativamente mais complexo. Os modelos podem ser classificados como determinísticos e probabilísticos (Myers et al., 2002). Nos modelos determinísticos o sistema de respostas ou resultados é definido com exatidão freqüentemente por um conjunto de equações. Nos modelos probabilísticos, o sistema de respostas ou resultados exhibe variabilidade porque o modelo inclui elementos aleatórios ou é influenciado em alguma medida por forças aleatórias.

O modelo linear probabilístico é definido como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

em que  $y$  representa a variável resposta;  $x_1, x_2, \dots, x_k$  formam um conjunto de variáveis preditoras;  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  representa o conjunto de parâmetros desconhecidos; e  $\varepsilon$  é o termo do erro aleatório. Assume-se que o erro  $\varepsilon$  tem um valor esperado de zero  $E[\varepsilon] = 0$ , e:

$$E[y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

O modelo é linear porque a resposta média é uma função linear dos parâmetros desconhecidos  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ , o qual significa que os modelos seguintes são todos modelos lineares (Myers et al., 2002):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

$$y = \beta_0 + \beta_1 \sin\left(\frac{2\pi x}{12}\right) x_1 + \beta_2 \cos\left(\frac{2\pi x}{12}\right) + \varepsilon$$

Existem situações nas quais um modelo de regressão linear não é apropriado. Qualquer modelo em que a resposta média não seja uma função linear dos parâmetros desconhecidos é um modelo de regressão não linear, como, por exemplo, o modelo seguinte (Myers et al., 2002; Resende, 2002), que não é linear nos parâmetros  $\beta_1$  e  $\beta_2$ :

$$y = \beta_1 e^{\beta_2 x} + \varepsilon$$

Em geral, um modelo de regressão não linear pode ser escrito como:

$$y = f(x, \beta) + \varepsilon$$

em que  $\beta$  é um vetor ( $p \times 1$ ) dos parâmetros desconhecidos e  $\varepsilon$  é o termo do erro aleatório com  $E(\varepsilon) = 0$  e  $\text{var}(\varepsilon) = \sigma^2$ . Assume-se que os erros são normalmente distribuídos, como em regressão linear.  $f(x, \beta)$  é a função de esperança para o modelo de regressão não linear:

$$E(y) = E[f(x, \beta) + \varepsilon] = f(x, \beta)$$

Em um modelo de regressão não linear, pelo menos uma das derivadas parciais da função de esperança do modelo, i.e.  $f(x, \beta)$ , com respeito aos parâmetros, depende pelo menos de um dos parâmetros desconhecidos.

Considerando o seguinte modelo não linear:

$$y = f(x, \beta) + \varepsilon = \beta_1 e^{\beta_2 x} + \varepsilon$$

As derivadas parciais da função  $f(x, \beta)$  com respeito de  $\beta_1$  e  $\beta_2$  são:

$$\frac{\partial f(x, \beta)}{\partial \beta_1} = e^{\beta_2 x}$$

$$\frac{\partial f(x, \beta)}{\partial \beta_2} = \beta_1 e^{\beta_2 x}$$

Essas derivadas parciais constituem em uma função linear dos parâmetros desconhecidos  $\beta_1$  e  $\beta_2$  e, portanto, o modelo é não linear.

## 2.2. Modelos lineares generalizados

Nos modelos de regressão linear e não linear, a distribuição normal da variável aleatória  $y$  tem um papel fundamental nos procedimentos de inferência estatística. Entretanto, existem várias situações nas quais o pressuposto de normalidade não é atingido (Demetrio, 2001; Dobson, 2001). Isso ocorre, por exemplo, quando a variável resposta é uma variável discreta, tal como contagem. Outra situação refere-se aos estudos em que a variável resposta seja sucesso ou fracasso (i.e. 0 ou 1), a qual é uma variável de resposta binária. Existem outras situações em que a variável resposta é contínua, porém o pressuposto de normalidade não é atingido (Myers et al., 2002).

O Modelo Linear Generalizado (GLM, do inglês Generalized Linear Model) foi desenvolvido por Nelder e Wedderburn (1972) como um método alternativo para a análise de dados para os quais a pressuposição de normalidade não é atingida. GLM é uma extensão natural dos modelos lineares clássicos e fornece uma teoria de estrutura conceitual unificada para diversos métodos estatísticos utilizados freqüentemente (Dobson, 2001). A metodologia GLM é interessante para a experimentação biológica, em situações nas quais a variável resposta segue uma distribuição pertencente à chamada família exponencial de distribuições (Demetrio, 2001; Myers et al., 2002). Esta família exponencial inclui as distribuições de Poisson, Binomial, Gama, Binomial Negativa, Exponencial, Geométrica, Inversa Qui-quadrado, Gama Inversa, Multinomial e Inversa Gaussiana. Nestas distribuições, a função de densidade e probabilidade para a resposta observada  $y$  pode ser expressa na seguinte forma:

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

em que  $a(\phi)$ ,  $b(\theta)$  e  $c(y, \phi)$  são chamadas funções específicas; o parâmetro  $\theta$  é o parâmetro de posição natural e  $\phi$  é o freqüentemente chamado de parâmetro de dispersão. Para alguns membros comuns da família exponencial (como a Poisson e Binomial),  $\phi = 1$ , exceto em situações de superdispersão (Dobson, 2001; Myers et al., 2002).

Consideremos o caso da distribuição normal. A função de densidade e probabilidade para uma variável aleatória  $y$  com parâmetros  $\mu$  e  $\sigma^2$  é dada por:

$$\begin{aligned} f(y; \mu; \sigma) &= \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi\sigma}} \\ &= \exp\left\{\frac{(y\mu - \mu^2/2)}{\sigma^2} - \frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\} \end{aligned}$$

Esta função de densidade é da mesma forma exponencial apresentada anteriormente com:

$$\theta = \mu$$

$$b(\theta) = \mu^2 / 2$$

$$a(\phi) = \phi = \sigma^2$$

$$c(y; \phi) = -\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$$

Como esperado,  $\mu$  representa o parâmetro de posição e  $\sigma^2$  o parâmetro escalar natural (Demetrio, 2001; Dobson, 2001; Myers et al., 2002).

McCullagh e Nelder (1987) detalharam a estrutura do GLM. Basicamente, o modelo é construído por meio de um preditor linear:

$$\eta = \mathbf{x}'\boldsymbol{\beta} = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

O modelo também utiliza uma função de ligação, que liga a média com o preditor linear:

$$\eta_i = g(\mu_i), \quad i = 1, 2, \dots, n$$

Na regressão linear múltipla, por exemplo, tem-se o modelo (Myers et al., 2002):

$$\mu_i = \eta_i = \mathbf{x}_i'\boldsymbol{\beta}$$

em que  $\mu_i = g(\mu_i)$ , e, portanto, a função de ligação usada é a chamada função de ligação identidade.

Alguns exemplos de funções de ligação canônicas para as distribuições Normal, Binomial, Poisson, Exponencial e Gama, correspondem, respectivamente, a:

$$\eta_i = \mu_i \text{ (ligação identidade)}$$

$$\eta_i = \ln(P / 1 - P_i) \text{ (ligação logística)}$$

$$\eta_i = \ln(\mu_i) \text{ (ligação logarítmica)}$$

$$\eta_i = 1 / \mu_i \text{ (ligação recíproca)}$$

$$\eta_i = 1 / \mu_i \text{ (ligação recíproca)}$$

O método da máxima verossimilhança é teoricamente o procedimento base para a estimação dos efeitos fixos (parâmetros) do modelo linear generalizado. A função do logaritmo da verossimilhança na forma exponencial é (Myers et al., 2002; Dobson, 2001):

$$\log L = l(y, \beta) = \sum_{i=1}^n \{ [y_i \theta_i - b(\theta_i)] / a(\phi) + c(y_i, \phi) \}$$

Para a função de ligação canônica, temos:  $\eta_i = g[\mu_i] = \mathbf{x}_i^T \boldsymbol{\beta}$ , portanto:

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\beta}} &= \frac{\partial L}{\partial \theta_i} \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^n \frac{1}{a(\phi)} \left[ y_i - \frac{db(\theta_i)}{d\theta_i} \right] \mathbf{x}_i \\ &= \sum_{i=1}^n \frac{1}{a(\phi)} (y_i - \mu_i) \mathbf{x}_i \end{aligned}$$

Assim, as estimações de máxima verossimilhança dos parâmetros são obtidas resolvendo o seguinte sistema de equações para  $\boldsymbol{\beta}$ :

$$\sum_{i=1}^n \frac{1}{a(\phi)} (y_i - \mu_i) \mathbf{x}_i = 0$$

Na maioria dos casos,  $a(\phi)$  é uma constante (Myers et al., 2002), portanto as equações (denominadas equações score da máxima verossimilhança) tornam-se:

$$\sum_{i=1}^n (y_i - \mu_i) \mathbf{x}_i = 0$$

e na forma matricial:

$$\mathbf{X}'(\mathbf{y} - \boldsymbol{\mu}) = 0$$

em que  $\boldsymbol{\mu}' = [\mu_1, \mu_2, \dots, \mu_p]$ .

Segundo Resende (2002) a técnica de GLM permite a generalização ou flexibilização dos modelos lineares clássicos de variáveis contínuas, de forma que toda a estrutura para a estimação e predição em modelos lineares normais pode ser estendida para os modelos não lineares. Os modelos lineares clássicos são considerados casos especiais de modelos lineares generalizados.

### 2.3. Modelos lineares generalizados e dados longitudinais

A modelagem do GLM é baseada no uso da máxima verossimilhança para estimação dos parâmetros desconhecidos (Dobson, 2001). Em linhas gerais, os pressupostos básicos para esta modelagem podem ser resumidos em dois: 1) as distribuições das respostas são assumidas ser da família exponencial, e 2) existe independência entre as observações. Entretanto, existem diversos casos nos quais são realizadas medições repetidas à mesma unidade experimental, e,

portanto, tal informação forma um *cluster* (agrupamento) de observações correlacionadas (Myers et al., 2002). O exemplo clássico ocorre quando um estudo longitudinal é feito com variáveis medidas repetidamente sobre um mesmo indivíduo (ou unidade experimental) no tempo (Keuls e Garretsen, 1982; Khattree e Naik, 1999; Yang et al., 2006).

Os caracteres de interesse no melhoramento de plantas podem se expressar mais de uma vez em cada indivíduo (ex. em plantas perenes) gerando dados longitudinais ou medidas repetidas. Portanto, dados dessas características podem apresentar estrutura correlacionada através do tempo, safras ou medições (Resende, 2002).

A metodologia das Equações de Estimação Generalizadas (GEE; *Generalized Estimating Equations*) foram desenvolvidas por Liang e Zeger (1986) e Zeger e Liang (1986) como um procedimento para testar hipóteses que consideram a influência de fatores sobre variáveis distribuídas exponencialmente (família exponencial, ex. Poisson, Gama, binomial negativa, binomial) e que foram medidas no tempo, nas mesmas unidades experimentais. Esta metodologia é considerada uma extensão dos GLM, a qual facilita as análises de regressão sobre variáveis dependentes que não apresentam distribuição normal (Nelder e Wedderburn, 1972; McCullagh e Nelder, 1989).

Na metodologia GEE, as matrizes de correlação usadas dentro das unidades repetidas são denominadas como matrizes de correlação de trabalho (do inglês: *Working Correlation Matrixes*), as quais possuem as seguintes estruturas pré-definidas:

- Independência (do inglês: *Independence*): as observações repetidas não são correlacionadas.

- Não estruturada (do inglês: *Unstructured* ou *Unspecified*): as correlações entre duas respostas quaisquer são desconhecidas e precisam ser estimadas.

- Simetria composta (do inglês: *Compound Symmetry* ou *Exchangeable*): a correlação é a mesma entre duas medidas repetidas da *i*-ésima unidade experimental.

- Auto-regressiva de primeira ordem (do inglês: *Autoregressive of first order*): assume a mesma correlação entre observações contíguas.

## 2.4. Inferência Bayesiana

Para chegar ao teorema de Bayes, supomos que a probabilidade de que observemos simultaneamente um evento  $\theta$  e um evento  $y$  é dada por:

$$P(\theta \cap y) = P(\theta / y) \cdot P(y)$$

Por outro lado, a probabilidade de que observemos simultaneamente um evento  $\theta$  e um evento  $y$  também pode ser dada por:

$$P(y \cap \theta) = P(\theta \cap y) = P(y / \theta) \cdot P(\theta)$$

Combinando as duas informações temos:

$$P(\theta / y) \cdot P(y) = P(y / \theta) \cdot P(\theta)$$

Isolando a probabilidade condicional  $P(\theta / y)$ :

$$P(\theta / y) = \frac{P(\theta) \cdot P(y / \theta)}{P(y)}$$

em que  $P(\theta / y)$  é a probabilidade *a posteriori* conjunta do evento  $\theta$ , dado o evento  $y$ ;  $P(\theta)$  é denominada a probabilidade *a priori* do evento  $\theta$ ;  $P(y / \theta)$  é a probabilidade condicional do evento  $y$ , dado o evento  $\theta$  (verossimilhança).

Gianola e Fernando (1986), e Van Tassel e Van Vleck (1996) afirmam que os métodos Bayesianos vêm sendo utilizados, desde a década de 80, para a estimação dos componentes de (co) variância e dos parâmetros genéticos para a avaliação genética no melhoramento animal. Estes métodos permitem a análise de grandes conjuntos de dados, não requerem soluções para as equações de modelos mistos e propiciam estimativas diretas e acuradas dos componentes de (co) variância e dos parâmetros genéticos das características de interesse (De Paula, 2006).

As estimações de parâmetros genéticos e, em geral, a análise quantitativa no melhoramento de plantas podem ser conduzidos utilizando também uma abordagem Bayesiana. No melhoramento vegetal, diversos estudos confirmaram uma forte associação entre o enfoque Bayesiano e o procedimento tradicional de estimação / predição, REML/BLUP. Blasco (2001) ratificou que BLUP pode ser considerado um estimador Bayesiano, construído usando uma distribuição a priori uniforme (*flat*) para os efeitos ambientais e uma distribuição *a priori* normal para os efeitos genéticos.

Seguindo o teorema de Bayes, um modelo animal de avaliação genética, pode ser expresso pelo seguinte modelo bi-caráter:

$$y = X\beta + Za + e = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & \emptyset \\ \emptyset & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & \emptyset \\ \emptyset & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

em que  $y_i$  é o vetor dos dados fenotípicos (características  $y_1$  e  $y_2$ );  $\beta_i$  é o  $i$ -ésimo vetor dos efeitos fixos;  $a_i$  é o  $i$ -ésimo vetor dos efeitos genéticos aditivos;  $e_i$  é o  $i$ -ésimo vetor dos efeitos residuais;  $X_i$  e  $Z_i$  são as matrizes de incidência para os efeitos fixos e aleatórios, respectivamente.

Os valores esperados e as matrizes de dispersão de acordo a uma distribuição normal multivariada (MVN) são:

$$\begin{bmatrix} y \\ a \\ e \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} V & ZG & R \\ GZ' & G & \emptyset \\ R & \emptyset & R \end{bmatrix} \right\}$$

em que

$$V = ZGZ' + R$$

$$G = G^* \otimes A$$

$$R = R^* \otimes I$$

$A$  é a matriz de parentesco;  $\otimes$  é o produto direto;  $G^*$  é a matriz de (co) variância genética;  $I$  é a matriz de identidade;  $R^*$  é a matriz de (co) variância residual.  $G^*$  e  $R^*$  têm as seguintes estruturas:

$$G^* = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} \\ \sigma_{a_1 a_2} & \sigma_{a_2}^2 \end{bmatrix} \quad R^* = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix}$$

em que  $\sigma_{a_1}^2$  e  $\sigma_{a_2}^2$  são as variâncias genéticas aditivas para as

características ( $y_1$ ) e ( $y_2$ ), respectivamente;  $\sigma_{a_1a_2}$  é a covariância genética aditiva entre as características ( $y_1$ ) e ( $y_2$ );

$\sigma_{e_1}^2$  e  $\sigma_{e_2}^2$  são as variâncias residuais para ( $y_1$ ) e ( $y_2$ ), respectivamente;

$\sigma_{e_1e_2}$  é a covariância residual entre ambas as características.

No contexto Bayesiano, não existem efeitos fixos, porque todos os parâmetros são considerados variáveis aleatórias (Blasco, 2001). De acordo com Van-Tassell e Van-Vleck (1996), a distribuição Wishart Inversa (IW) pode ser usada como *prior* para os componentes de (co) variância. Assim, a função de densidade posterior conjunta pode ser dada por:

$$f(\beta, a, G, R | y, v_g, G_0, v_r, R_0) \propto f(y | \beta, a, R) \times f(a | G) \times f(G | v_g, G_0) \times f(\beta) \times f(R | v_r, R_0)$$

ou descomposta na forma:

$$\begin{aligned} f(\beta, a, G, R | y, v_g, G_0, v_r, R_0) \propto & |R|^{-1/2} \times \exp\left\{-\frac{1}{2}(y - X\beta - Za)' R^{-1}(y - X\beta - Za)\right\} \times \\ & |G|^{-n/2} \times \exp\left\{-\frac{1}{2} a'(G^{*-1} \otimes A^{-1})a\right\} \times |G|^{-\frac{1}{2}(v_g + m_g + 1)} \times \exp\left\{tr\left(-\frac{v_g^*}{2} G_0 G^{*-1}\right)\right\} \times \\ & |R|^{-\frac{1}{2}(v_r + m_r + 1)} \times \exp\left\{tr\left(-\frac{v_r^*}{2} R_0 R^{*-1}\right)\right\} \end{aligned}$$

em que  $G_0$  e  $R_0$  são as matrizes de parâmetro escala: ( $m_g \times m_g$ ) e ( $m_r \times m_r$ ), respectivamente,  $v_g$  e  $v_r$  são os graus de credibilidade (Van-Tassell e Van-Vleck, 1996).

Gianola e Fernando (1986) propuseram usar o enfoque Bayesiano como método geral de inferência para resolver um grande número de problemas no melhoramento, lineares ou não lineares, ainda em situações nas quais haja incerteza sobre todos os parâmetros de posição e de dispersão. As primeiras aplicações deste paradigma basearam-se nas aproximações gaussianas à distribuição posterior conjunta ou às distribuições posteriores marginalizadas parcialmente, devido às dificuldades técnicas que foram encontradas. O problema geral da inferência Bayesiana consiste, então, em calcular os valores esperados

dos parâmetros genéticos seguindo uma função *a posteriori*. Entretanto, segundo Blasco (2001), quando se aplicou o método de Monte-Carlo, baseado em cadeias de Markov para estimar distribuições marginais *a posteriori*, os procedimentos computacionais foram simplificados e se renovaram os interesses pelos métodos Bayesianos. Existem vários métodos MCMC, porém, sem dúvida, o mais popular tem sido o algoritmo de amostragem de Gibbs.

Os procedimentos MCMC são utilizados para determinar as distribuições marginais *a posteriori* dos parâmetros, dando vez à aplicação do teorema de Bayes por meio dos algoritmos que convergem a tais densidades marginais. A metodologia Bayesiana é considerada ferramenta importante na avaliação genética. Ela leva em consideração a variabilidade existente em todos os parâmetros do modelo (Wright et al., 2000; Nogueira et al., 2003), no sentido de que é possível caracterizá-los por meio da moda, mediana ou da média da distribuição *a posteriori* de tal parâmetro. Também é de particular interesse Bayesiano a obtenção de intervalos de confiança (chamadas de regiões de credibilidade) tanto para os parâmetros como para a predição dos efeitos genéticos.

De acordo com Resende (2002) as seguintes conclusões podem ser relatadas em relação à inferência Bayesiana:

- 1) a análise Bayesiana propicia resultados adicionais àqueles obtidos pela abordagem Freqüentista, destacando-se os intervalos de credibilidade Bayesianos para as estimativas de parâmetros genéticos, valores genéticos e ganhos genéticos;

- 2) as estimativas dos parâmetros genéticos obtidas do amostrador de Gibbs podem ser mais precisas do que pelo método REML/BLUP;

- 3) a análise Bayesiana é uma técnica elegante e flexível que permite a simultânea estimação dos parâmetros genéticos, efeitos “fixos” e valores genéticos de maneira precisa, mesmo para pequenas amostras. Entretanto, o autor afirma que, para a obtenção de uma precisão adequada na abordagem Bayesiana, é necessário que o pesquisador tenha uma quantidade suficiente de informação experimental.

O mesmo autor relata também o problema básico da abordagem Bayesiana referido-se à integração numérica (no espaço do parâmetro) da função de densidade de probabilidade *a posteriori*. Na prática, é impossível conseguir tal

integração por métodos analíticos e, assim, tem sido utilizado o procedimento de simulação estocástica denominada algoritmo de amostragem de Gibbs, para a viabilização da estimação Bayesiana.

Outros algoritmos derivados dos métodos MCMC são utilizados em conjunto com a teoria de modelos lineares mistos. Alguns exemplos são os algoritmos de Cadeias Independentes (da tradução do inglês: Independence Chain: IC) e Amostragem por Importância (Importance Sampling: IS) (SAS Institute, 1996; Wolfinger e Kass, 2000). Estes algoritmos são usados para a construção de intervalos de credibilidade Bayesianos (conhecidos como regiões de credibilidade) dos componentes de variância (Wolfinger e Kass, 2000), os quais são calculados diretamente a partir da distribuição *a posteriori*.

Wolfinger e Kass (2000) examinaram regiões de credibilidade para funções gerais de componentes de variância, usando a variante do método de MCMC, o algoritmo IC. Aqui, a distribuição base para os efeitos aleatórios é considerada exata e, portanto, as amostras MCMC são rejeitadas somente se estas não estão no espaço do parâmetro, o qual pode ocorrer, por exemplo, se um componente de variância individual é estimado com valor negativo. Deste modo, o algoritmo IC é um caso especial de amostragem de rejeição (Tierney, 1994).

No contexto Bayesiano, a distribuição *a posteriori* dos parâmetros genéticos pode ser caracterizada pela média, mediana e moda, os quais são vistos como potenciais estimadores pontuais. Entretanto, a importância fundamental do caso Bayesiano radica no fato de que é possível conformar regiões de credibilidade (intervalos de confiança no contexto Frequentista) para as estimativas dos parâmetros (Gianola e Fernando, 1986). Como consequência, a metodologia Bayesiana é considerada uma ferramenta importante na avaliação genética, porque leva em consideração a variabilidade existente em todos os parâmetros do modelo (Wright et al., 2000, Nogueira et al., 2003).

A exemplo de diversos outros estudos, estudos (Wright et al., 2000; Nogueira et al., 2003; Gonçalves-Vidigal et al., 2008), a abordagem Bayesiana leva em consideração a variabilidade existente em todos os parâmetros do modelo misto, viabilizando a sua caracterização através da moda, mediana ou a média da distribuição *a posteriori*, e a conformação de regiões de credibilidade.

### 3. REFERÊNCIAS BIBLIOGRÁFICAS

BALL, R.D. Bayesian methods for quantitative trait loci mapping based on model selection: Approximate analysis using the Bayesian information criterion. **Genetics**, 59:1351–1364, 2001.

BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, 79:2023–2046, 2001.

BROMAN, K.W.; SPEED, T.P. A model selection approach for the identification of quantitative trait loci in experimental computing time can be intensive within each sweep of crosses. **J R Stat Soc**, 64:641–656, 2002.

CADENA-MENESES, J.A., CASTILLO-MORALES, A. Using MTGSAM and Gibbs sampling in variance components estimation. **Agrociencia**, 36: 345-354, 2002.

CAPPA, E.P.; CANTET, R.J.C. Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. **Canadian Journal of Forest Research**, 36:1276-1285, 2006.

CERÓN-ROJAS, J.J.; Sahagún-Castellanos, J. Estimating QTL biometrics parameters in F2 populations: a new approach. **Agrociencia**, 41: 57-73, 2007.

DE PAULA, M.C. **Heterogeneidade de variâncias e interação genótipo x ambiente para produção de leite em bovinos da raça holandesa no estado do Paraná**. Maringá: Universidade Estadual de Maringá, 2006. 84 p. Tese (Doutorado em Zootecnia).

DELGADO-SALINAS, A.; BIBLER, R.; LAVIN, M. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. **Systematic Botany**, 31:779–791, 2006.

DEMETRIO, C.G.B. **Modelos lineares generalizados em experimentação agrônômica**. Piracicaba: ESALQ, 2001. 113p.

DOBSON, A.J. **An introduction to generalized linear models**. London: Chapman and Hall, 2001. 225p.

GONÇALVES-VIDIGAL, M.C.; MORA, F.; BIGNOTTO, T.S.; MUNHOZ, R.E.F.; SOUZA, L.D. Heritability of quantitative traits in segregating common bean families using a Bayesian approach. **Euphytica**, 164:551-560, 2008.

GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, 63:217-277, 1986.

HEATH, S.C. Markov chain Monte Carlo segregation and linkage analysis of oligogenic models. **Am J Hum Genet**, 61:748–760, 1997.

HOLSINGER, K.E. Analysis of genetic diversity in geographically structured populations: A Bayesian perspective. **Hereditas**, 130:245–255, 1999.

KHATTREE, R.; NAIK D.N. **Applied multivariate statistics with SAS® software**. Cary: SAS Institute Inc., 1999. 338 p.

KEULS, M.; GARRETSEN, F. Statistical analysis of growth curves in plant breeding. **Euphytica**, 31:51-64, 1982.

KOTHERA, L.; RICHARDS, C.M.; CARNEY, S.E. Genetic diversity and structure in the rare Colorado endemic plant *Physaria bellii* Mulligan (Brassicaceae). **Conservation Genetics**, 8:1043-1050, 2007.

LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. **Biometrika**, 73:13-22, 1986.

MCCULLAGH, P.; NELDER, J.A. **Generalized linear models**. London: Chapman and Hall, 1987.

MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. **Generalized linear models, with applications in engineering and the sciences**. New York: John Wiley & Sons, 2002. 342 p.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear model. **Journal of the Royal Statistical Society A**, 135:370-384, 1972.

NOGUEIRA, D.A.; SÁFADI, T.; BEARZOTI, E.; BUENO-FILHO, J.S.S. Análises clássica e Bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência e agrotecnologia**, 27:1614-1624, 2003.

PAPINI, A.; BANCI, F.; NARDI, E. Molecular evidence of polyphyly in the plant genus *Carum* L. (Apiaceae). **Genet Mol Biol**, 30:475–482, 2007.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.

ROEDER, K.; ESCOBAR, M.; KADANE, J.B.; BALAZS I. Measuring heterogeneity in forensic databases using hierarchical Bayes models. **Biometrika**, 85:269–287, 1998.

SAS INSTITUTE. **Statistical analysis system. User's guide**. Cary: SAS Institute, 1996. 956 p.

SILLANPAA, M.J.; ARJAS, E. Bayesian mapping of quantitative trait loci from incomplete inbred line cross data. **Genetics**, 148:1373–1388, 1998.

SOKAL, R.R.; ROHLF, F.J. **Biometry: the principles and practice of statistics in biological research**. New York: Freeman and Company Press, 2003. 850 p.

SORIA, F.; BASURCO, F.; TOVAL, G.; SILIÓ, L.; RODRÍGUEZ, M.C.; TORO, M. An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. **Canadian Journal of Forest Research**, 28:1286-1294, 1998.

TIERNEY, L. Markov chains for exploring posterior distributions (with discussion). **Annals of Statistics**, 23:1701-1762, 1994.

VAN TASSEL, C.P.; VAN VLECK, L.D. Multiple trait Gibbs Sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance components inference. **Journal of Animal Science**, 74:2586-2597, 1996.

WALDMANN, P.; ERICSSON, T. Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. **Theor. Appl. Genet.**, 112:1441-1451, 2006.

WALSH, B. Quantitative genetics in the age of genomics. **Theoret Pop Biol**, 59:175–184, 2001.

WOLFINGER, R.D.; KASS, R.E. Non-conjugate Bayesian analysis of variance component models. **Biometrics**, 56:768-774, 2000.

WRIGHT, D.R.; STERN, H.S.; BERGER, P.J. Comparing traditional and Bayesian analyses of selection experiments in animal breeding. **Journal of Agricultural, Biological, and Environmental Statistics**, 5:240-256, 2000.

YANG, R.; TIAN, Q.; XU, S. Mapping quantitative trait loci for longitudinal traits in line crosses. **Genetics**, 173:2339-2356, 2006.

YIN, Z.J.; ZHANG, Q. Mapping quantitative trait Loci for ordinal traits using the generalized linear model in half-sib designs. **Anim. Res.**, 55:245-255, 2006.

YIN, Z.J.; ZHANG, Q.; ZHANG, J.G.; DING, X.D. Methodology of mapping quantitative trait loci for ordinal traits of disease resistance in livestock. **Acta Genetica Sinica**, 32:1147-55, 2005.

ZEGER, S.L.; LIANG, K.Y. Longitudinal data analysis for discrete and continuous outcomes. **Biometrics**, 42:121-130, 1986.

**CAPÍTULO I**  
**ANÁLISE MULTIVARIADA NO MAPEAMENTO DE *LOC*/DE**  
**CARACTERÍSTICAS QUANTITATIVAS (QTL), USANDO MODELOS MISTOS E**  
**EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS**

**RESUMO**

Os procedimentos estatísticos usados para o mapeamento de loci de caracteres quantitativos (QTL) têm sido amplamente estudados devido a que são essenciais para melhorar a precisão das análises genéticas. O objetivo do presente estudo foi examinar QTL usando métodos multivariados, considerando o princípio de mapeamento por intervalo. Foram simulados dados de marcadores moleculares co-dominantes provenientes de uma população F2. Foi assumido que o QTL controla características de distribuição Normal e Binomial. No caso Normal, cinco modelos de mapeamento de QTL foram avaliados, os quais tiveram as seguintes estruturas de covariâncias na matriz de resíduos: 1. Componentes de variância (VC), 2. Simetria composta (CS), 3. Não estruturada (UN), 4. Diagonal principal bandeada (Banded Main Diagonal; UN1) e 5. Toeplitz heterogênea (TOEPH). O critério de informação de Akaike (AIC) foi utilizado para a comparação entre os modelos. No caso binário, foi utilizado o procedimento das Equações de Estimação Generalizada. A estrutura UN minimizou o valor de AIC no intervalo de maior probabilidade de ocorrência do QTL. No caso binomial, foi ajustada a matriz de correlação (WCM) não independente ( $\rho=0,47$ ). Em ambos os casos, o efeito aditivo do QTL foi significativo ( $p<0,01$ ), sendo não significativo o efeito de dominância ( $p>0,05$ ). O mapeamento de QTL usando uma abordagem multivariada pode ser ferramenta útil dentro dos programas genéticos que buscam melhorar características de interesse agrônomo cujos valores fenotípicos mudam no tempo.

Palavras-chave: QTL, marcadores moleculares, modelação estatística

## ABSTRACT

Statistical procedures for mapping quantitative trait loci (QTL) have been extensively studied because they are the essential for improving the accuracy of genetic analyses. The objective of the present study was to examine QTL using multivariate methods, considering the principle of interval mapping. A co-dominant marker data set from a F2 population was simulated. It was assumed that the QTL control binomial and normal traits. In the normal case, five QTL mapping models were evaluated that had the following residual covariance structures: variance components (VC), compound symmetry (CS), unstructured (UN), banded main diagonal (UN1) and heterogeneous toeplitz. Akaike's information criterion (AIC) was used to select the appropriate structure. In the binary case, the Generalized Estimating Equations (GEE) procedure was used. UN structure minimized the AIC value on the interval that indicated a higher probability of the QTL. In the binomial case, a non-independent working correlation matrix (WCM) was fitted ( $\rho = 0.47$ ). In both cases, the additive effect of QTL was significant ( $p < 0.01$ ), but dominance effects were not ( $p > 0.05$ ). QTL mapping using a multivariate approach may be a useful tool for breeding programs that aim to improve quantitative traits that have phenotypic values that change over time.

Key words: QTL, molecular markers, statistical modeling.

## 1. INTRODUÇÃO

O mapeamento de características quantitativas, realizado por meio da identificação de *loci* de caracteres quantitativos (QTL, Quantitative Trait Loci), é considerado um importante procedimento dentro do melhoramento genético de plantas (Martínez-Gomez et al., 2005; Cerón-Rojas e Sahagún-Castellanos, 2007). Os QTLs são identificados dentro do genoma de uma planta baseando-se no princípio de associação entre os marcadores moleculares polimórficos e o fenótipo dos indivíduos de uma população de melhoramento. Existem vários procedimentos para caracterizar os QTLs (Schuster e Cruz, 2004). Um deles é o mapeamento por intervalos (*Interval mapping*; Lander e Botstein, 1989). Este método considera dois marcadores moleculares adjacentes ao QTL (marcador – QTL – marcador) (Cerón-Rojas e Sahagún-Castellanos, 2007). Todos os métodos de mapeamento de QTLs usados atualmente são baseados na utilização do princípio de mapeamento por intervalos (Schuster e Cruz, 2004).

Conceitualmente, QTL é baseado na existência de *loci* com maior importância para a expressão de características quantitativas (Rocha et al., 2007; Geldermann, 1975). O uso de mapas genéticos de marcadores moleculares para o estudo da herança de características quantitativas permite a identificação de cromossomos (ou grupos de ligação) os quais são importantes para determinar a expressão de um caráter (Schuster e Cruz, 2004).

Os programas de melhoramento genético enfocam a seleção de plantas considerando diversas características de interesse agrônomo e econômico. Nesse sentido, o mapeamento de QTLs usualmente considera uma análise do tipo univariado, de um caráter ou uma característica medida apenas em um determinado tempo (Schuster e Cruz, 2004). Entretanto, atualmente se dispõe de uma importante base em estatística genômica que permite o mapeamento simultâneo do QTL para características medidas repetidamente no tempo. No caso de características que são medidas repetidamente em um mesmo indivíduo, o estudo se estende à análise longitudinal (*Longitudinal traits*). Existem três possíveis métodos disponíveis para a análise de características longitudinais (Yang et al., 2006). O primeiro método trata os valores fenotípicos obtidos em diferentes tempos, como medidas repetidas da mesma característica e analisa esta segundo o conceito de medidas repetidas. O segundo método trata os

fenótipos, medidos em diferentes tempos, como diferentes características e os analisa em função da teoria de análise multivariada (Wu et al., 1999). O terceiro método consiste em ajustar os valores fenotípicos às curvas de crescimento, para os diferentes pontos no tempo, e analisa os parâmetros ajustados segundo a teoria de análise multivariada (Wu et al., 2002). O método de ajuste de curva de crescimento tem duas vantagens: 1) as características medidas no tempo são consideradas diferentes; e 2) considera ordenadamente a correlação gerada pelos pontos no tempo (Yang et al., 2006).

Diferentes estratégias podem ser empregadas na análise de experimentos agrônômicos quando as características não se ajustam à distribuição normal ou quando não se atingem os pressupostos estatísticos básicos (ex. variâncias constantes). Pode-se mencionar a análise não paramétrica (Droste et al., 2005), a transformação da variável resposta (Guzmán et al., 2007) ou o emprego dos Modelos Lineares Generalizados (Mora et al., 2007). Na análise multivariada de QTLs, considerando uma distribuição não normal da variável resposta, Lange e Whittaker (2001) usaram experimentos simulados de populações  $F_2$  de melhoramento para ilustrar a aplicação do procedimento de Equações de Estimação Generalizadas (GEE, *Generalized Estimating Equations*). Este procedimento é amplamente sugerido na análise genética de *loci* de características quantitativas. A metodologia GEE permite a análise estatística baseada em uma ampla base de distribuições probabilísticas da variável resposta e considera que as observações podem ser correlacionadas (formando conjuntos ou *clusters*). Esta metodologia usa procedimentos aproximadamente análogos aos utilizados na análise de dados com distribuição normal multivariada (Nelder e Wedderburn, 1972; Myers et al., 2002).

O presente estudo teve como objetivo examinar *loci* de características quantitativas em uma população  $F_2$  baseada no princípio de mapeamento por intervalos, em um modelo multivariado ou de características medidas repetidamente no tempo (análise longitudinal).

## 2. MATERIAL E MÉTODOS

Neste trabalho, os principais pressupostos de análise incluíram o mapeamento genético utilizando marcadores moleculares co-dominantes, e em uma situação na qual os QTLs controlam características de distribuição normal e binomial (dados de resposta binária).

### 2.1. Estudo de simulação

Os métodos multivariados usados na análise de QTLs foram avaliados empiricamente por meio de dados simulados. Simulou-se um cromossomo de 142,87 centiMorgan (cM) de comprimento (distância de Haldane), considerando uma população  $F_2$  com 178 plantas. Onze marcadores moleculares co-dominantes se distribuíram dentro do cromossomo com distância variável entre os marcadores (variando entre 2,0 e 24,1 cM) (Figura 1). Inicialmente, considerou-se a simulação de uma característica quantitativa com distribuição normal, com presença significativa de um QTL. Esta análise se realizou utilizando o programa QMOL da Universidade Federal de Viçosa, Brasil (Schuster e Cruz, 2004).

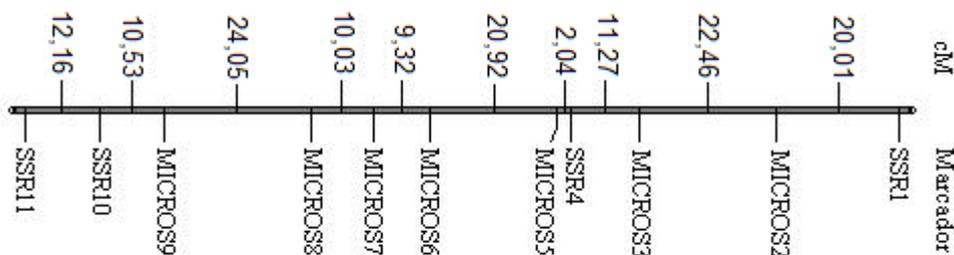


Figura 1 - Grupo de ligação genética com as distancias de Haldane (cM, centiMorgan) para os marcadores moleculares.

Numa segunda etapa, simulou-se uma segunda medição correlacionada significativamente com a característica da simulação inicial, com coeficiente de correlação de Pearson de 0,56 (calculado entre ambas as medições) e com intervalo de confiança Bootstrap de 0,43 a 0,65 (95% de probabilidade). Com este propósito utilizou-se o procedimento SAS SURVEYSELECT (SAS Institute, Cary NC, EUA).

Para o caso da característica quantitativa de distribuição binomial, considerou-se um valor 0 para os fenótipos que apresentaram uma resposta inferior à média do experimento, e 1 em caso contrário.

## 2.2. QTL controlando características de distribuição normal

O modelo geral de mapeamento de QTL, utilizado na análise de regressão, é o seguinte (Schuster e Cruz, 2004):

$$y_{ij} = \mu + ax_{ij} + dz_{ij} + \varepsilon_{ij}$$

em que  $y_{ij}$  corresponde ao valor fenotípico da característica em estudo;  $\mu$  é a média da característica na população;  $a$  é o efeito aditivo do locus que está sendo estudado;  $d$  é o efeito de dominância do locus estudado;  $x_{ij}$  e  $z_{ij}$  são as variáveis condicionadoras e dependentes dos genótipos dos marcadores moleculares que rodeiam o QTL;  $\varepsilon_{ij}$  é o erro aleatório. Embora este modelo seja considerado de efeitos fixos, pode ser visto como um caso especial de um modelo linear misto multivariado (Khattree e Naik, 1999). O modelo descrito é considerado modelo padrão na análise de regressão para o mapeamento de QTLs.

Posterior à caracterização do modelo geral, ajustaram-se as seguintes estruturas de covariâncias na matriz de resíduos: 1) componentes de variância (VC), 2) cimetria composta (CS), 3) não estruturada (UN), 4) diagonal principal bandeada (*Banded Main Diagonal*; UN1) e 5) toeplitz heterogênea (TOEPH). Assim, determinou-se a melhor estrutura para representar a dependência ou independência entre as observações. O critério de informação de Akaike (1974) (AIC) pode ser utilizado para a comparação entre os modelos (Mora e Scapim, 2007; Mora et al., 2007). Os diferentes ajustes do modelo de mapeamento de QTL, considerando medidas repetidas, foram realizados no SAS MIXED com as opções REPEATED e REML (*Restricted Maximum Likelihood*) para a estimação de componentes de (co) variância. O critério AIC também foi utilizado para examinar a posição do QTL dentro do intervalo (marcador – QTL – marcador). Avaliaram-se pontos a cada 2% de frequência de recombinação para a presença

do QTL dentro do cromossomo simulado e na proximidade do QTL considerou-se uma frequência de 1%.

### 2.3. QTL controlando características de distribuição binomial

A metodologia de equações de estimação generalizadas (GEE) tem sido sugerida como uma aplicação do procedimento da quase-verossimilhança (probabilidades quase seguras) para solucionar o problema de ajustar um modelo linear generalizado com medidas repetidas (Dos Santos e Mora, 2007; Liang e Zeger, 1986; Wendderburn, 1974). Baseado neste princípio, o modelo estatístico considerado no presente estudo para o mapeamento de QTL foi o seguinte:

$$y_j = X_j\beta + \varepsilon_j$$

em que  $y_j$  é o vetor que contém os valores fenotípicos de ambas as características correlacionadas;  $X_j$  é a matriz de delineamento experimental;  $\beta$  é o vetor que contém os coeficientes de regressão: efeitos aditivos e de dominância do QTL e  $\varepsilon_j$  é o vetor dos erros aleatórios.

As características quantitativas possuem distribuição pertencente à família exponencial, com função densidade e probabilidade da forma:

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

em que  $a(\phi)$ ,  $b(\theta)$  e  $c(y, \phi)$  são funções específicas conhecidas,  $\theta$  é o parâmetro canônico ou natural e  $\phi$  é o parâmetro de dispersão.

Os parâmetros do modelo genético se estimaram por meio da solução das equações padrão, determinadas por meio do procedimento de Gauss-Newton (Liang e Zeger, 1986). Para esta análise estatística, empregou-se o procedimento SAS GENMOD, com a opção REPEATED e LINK=LOGIT (função de ligação logística). A matriz de correlação independente (inexistência de associação significativa entre as respostas consideradas correlacionadas) apenas foi considerada se o valor da correlação entre elementos contíguos superou em magnitude a 0,3 (Liang e Zeger, 1986).

### 3. RESULTADOS E DISCUSSÃO

As probabilidades condicionais associadas aos genótipos de um possível QTL, em função da população de melhoramento estudada, são apresentadas no Quadro 1. A análise geral do cromossomo, realizada por meio do estatístico de razão da verossimilhança, confirmou a localização do QTL no intervalo compreendido entre os marcadores MICROS8 e MICROS9, para a população  $F_2$  em estudo ( $p < 0,01$ ). A análise específica da cartografia do QTL dentro do intervalo MICROS8-QTL-MICROS9 é apresentada no Quadro 2, considerando os critérios de informação de AIC para cada estrutura de correlação da matriz de resíduos.

Para cada posição do QTL ( $r_a$ ) o melhor ajuste correspondeu à covariância não estruturada (UN). Esta minimizou os valores de AIC. Os resultados indicaram uma maior probabilidade de presença do QTL na posição aproximada de 10 cM (AIC = 4519,7).

Quadro 1 - Probabilidades condicionais associadas aos genótipos de um QTL no intervalo dos marcadores A e B, em uma população  $F_2$ . Não se consideram as duplas recombinações

Genótipo	Genótipo do QTL		
	$QQ$	$Qq$	$qq$
AABB	1	0	0
AABb	(1-p)	(p)	0
AAbb	(1-p) <sup>2</sup>	[2p(1-p)]	(p <sup>2</sup> )
AaBB	(p)	(1-p)	0
AaBb	[qp(1-p)]	[1-2qp(1-p)]	[qp(1-p)]
Aabb	0	(1-p)	(p)
aaBB	(p <sup>2</sup> )	[2p(1-p)]	(1-p) <sup>2</sup>
aaBb	0	(p)	(1-p)
aabb	0	0	1

Fonte: Schuster e Cruz (2004);  $p=r_a/r$ ;  $(1-p)=r_b/r$ ;  $q=r^2/[r^2+(1-r)^2]$ ;  $QQ$ ,  $Qq$  e  $qq$ , representam os genótipos homozigoto dominante, heterozigoto e homozigoto recessivo, respectivamente.

De acordo com a estrutura de covariância UN, no Quadro 3 são apresentadas as estimativas dos parâmetros do modelo de mapeamento de QTL para cada frequência de recombinação considerada neste estudo. O número de iterações para obter a convergência dos parâmetros do modelo genético variou de 40 a 42. Na posição do QTL (10 cM), o mapeamento de QTL determinou um efeito aditivo de 9,356 com dominância de -0,655. O efeito aditivo do QTL foi altamente significativo ( $p < 0,01$ ) e, pelo contrário, o efeito de dominância não foi estatisticamente significativo ( $p > 0,05$ ). O valor verdadeiro do efeito aditivo foi  $a = 9,286$ . Uma vez removidos do modelo, os efeitos não significativos.

Quadro 2 - Critérios de informação de Akaike (1974) para as diferentes estruturas de covariâncias de um modelo genético bi-variado (medidas repetidas em dois períodos), considerando o mapeamento de QTL no intervalo MICROS8-MICROS9 (marcadores microssatélites), da população  $F_2$  simulada

Posição (cM)	Critério de informação de Akaike (AIC)				
	VC <sup>1</sup>	CS <sup>2</sup>	UN <sup>3</sup>	UN1 <sup>4</sup>	TOEPH <sup>5</sup>
0,01	5932,1	5680,6	4531,5	5926,9	5659,6
2,01	5931,8	5680,3	4528,4	5926,6	5659,3
4,01	5931,6	5680,1	4525,3	5926,4	5659,2
6,01	5931,5	5680,1	4522,5	5926,2	5659,1
8,01	5931,4	5680,2	4520,5	5926,1	5659,2
9,01	5931,3	5680,3	4519,9	5926,1	5659,3
10,01	5931,3	5680,5	4519,7	5926	5659,4
11,01	5931,4	5680,6	4519,8	5926,1	5659,6
12,01	5931,4	5680,9	4520,3	5926,1	5659,8
14,01	5931,6	5681,4	4522,7	5926,2	5660,3
16,01	5931,8	5682	4526,5	5926,5	5660,8
18,01	5932,2	5682,6	4531,4	5926,8	5661,4
19,09	5932,4	5683	4534,2	5927	5661,7

<sup>1</sup> VC: Componentes de Variância. <sup>2</sup> CS: Simetria Composta (CS). <sup>3</sup> UN: Não Estruturada. <sup>4</sup> UN1: Diagonal Principal Bandeada. <sup>5</sup> TOEPH: Toeplitz Heterogênea.

Quadro 3 - Média geral ( $\mu$ ), efeitos aditivos ( $a$ ) e de dominância ( $d$ ) do QTL, número de iterações do modelo escolhido (UN, Não Estruturado) e suas respectivas estimativas de variâncias (Var1 e Var2) e covariância (COV)

Posição (cM)	$\mu$	$a$	$d$	Iterações (n°)	Var1	COV	Var2
0,01	27,00	7,260	-0,488	42	732,8	51943	4222878
2,01	27,11	7,808	-0,505	42	726,8	51761	4222507
4,01	27,19	8,345	-0,538	40	722,2	51629	4222220
6,01	27,23	8,824	-0,586	42	719,8	51588	4222144
8,01	27,20	9,184	-0,636	42	720,9	51673	4222352
9,01	27,15	9,298	-0,652	42	722,7	51758	4222445
10,01	27,06	9,356	-0,655	42	726,0	51900	4222744
11,01	26,93	9,349	-0,638	41	730,7	52090	4223174
12,01	26,77	9,273	-0,596	42	736,5	52315	4223620
14,01	26,34	8,905	-0,428	41	751,3	52871	4224718
16,01	25,80	8,275	-0,165	42	769,3	53520	4225990
18,01	25,23	7,456	0,133	41	788,5	54190	4227320
19,09	24,95	6,970	0,281	41	798,4	54524	4227902

O procedimento SAS MIXED (SAS Institute) permite a análise com outras estruturas de covariâncias, além das utilizadas no presente estudo. Ao analisar acima de três medições ou, ao considerar uma característica medida em mais de dois períodos, a análise dos modelos genéticos seria mais ampla, no sentido de que esta incluiria outras estruturas residuais (Khattree e Naik, 1999).

Para escolher a estrutura de covariância que melhor se ajusta aos dados fenotípicos e genéticos, Khattree e Naik (1999) sugerem o uso dos critérios de informação AIC (Akaike, 1974) e Bayesiano de Schwarz (BIC, Schwarz, 1978). Ambos os procedimentos são considerados igualmente válidos para a comparação entre modelos acoplados (Mora e Scapim, 2007; Mora et al., 2007). No presente estudo, considerando a estrutura de covariância residual selecionada, o número de parâmetros e a dimensão do modelo genético se mantiveram constantes nos diferentes pontos do intervalo de maior probabilidade de existência do QTL. Portanto, também seria igualmente válido escolher aquele

ponto que maximize o valor do logaritmo de verossimilhança (restrita), ou minimize o valor de  $-2 \times$  logaritmo de verossimilhança.

Nesta parte do estudo, os modelos de mapeamento de QTL e os métodos considerados se enquadram dentro de quatro linhas gerais (Khattree e Naik, 1999): 1) somente são considerados modelos lineares; 2) o componente residual (erro) se considerou aditivo; 3) considerou-se uma distribuição normal dos resíduos; e 4) considerou-se uma apropriada estrutura de covariância sobre os erros.

No Quadro 4, pode-se observar o resultado da análise das equações generalizadas (GEE) no intervalo compreendido entre os marcadores que flanqueiam o QTL, para a população  $F_2$  em estudo. O critério escolhido para determinar a maior probabilidade da presença do QTL foi aquele que minimiza o valor da *deviance* (ou desvio). A estrutura de correlação não independente foi a que apresentou melhor ajuste, em cada posição do intervalo. A matriz de correlação mostrou um coeficiente ( $\rho$ ) superior a 0,3 (Myers et al., 2002; Liang e Zeger, 1986), confirmando sua pertinência no modelo ( $\rho$  variou de 0,4740 a 0,4885; Quadro 4). O valor da deviance dividido pelos graus de liberdade do modelo genético foi próximo ao valor 1, indicando um ajuste adequado. O mesmo foi confirmado pelo valor do  $\chi^2$  de Pearson (Myers et al., 2002).

Quadro 4 - Análise das Equações de Estimação Generalizadas (GEE) no intervalo de mapeamento compreendido entre os marcadores moleculares MICROS8 e MICROS9, para a população  $F_2$  simulada, considerando uma característica de tipo binário em um modelo multivariado

Posição (cM)	WCM <sup>1</sup>	Deviance	Log L <sup>2</sup>	$\frac{Deviance}{GL}$ <sup>3</sup>	$\frac{\chi^2 \text{ Pearson}}{GL}$
0,01	0,4885	464,2932	-232,147	1,3153	1,0085
1,01	0,487	463,409	-231,705	1,3128	1,008
2,01	0,4855	462,5236	-231,262	1,3103	1,0074
3,01	0,4839	461,6512	-230,826	1,3078	1,007
4,01	0,4824	460,8096	-230,405	1,3054	1,0065
5,01	0,4808	460,0205	-230,01	1,3032	1,0062
6,01	0,4793	459,3094	-229,655	1,3012	1,0059

Quadro 4, Cont.

9,01	0,4756	457,9418	-228,971	1,2973	1,0059
10,01	0,4747	457,8432	-228,922	1,297	1,006
11,01	0,4742	457,9662	-228,983	1,2974	1,0062
12,01	0,474	458,3251	-229,163	1,2984	1,0065
14,01	0,4746	459,7473	-229,874	1,3024	1,007
16,01	0,4768	461,9801	-230,99	1,3087	1,0077
18,01	0,4802	464,7461	-232,373	1,3166	1,0082
19,09	0,4825	466,3486	-233,174	1,3211	1,0085

<sup>1</sup> WCW: Coeficiente da matriz de correlação. <sup>2</sup> Log L: Logaritmo da verossimilhança. <sup>3</sup> GL: graus de liberdade.

De acordo com o esperado, a análise do intervalo compreendido entre os marcadores MICROS8 e MICROS9 indicou a maior probabilidade de existência do QTL na região localizada a 10 cM, a qual evidenciou a menor deviance ( $D = 457,8432$ ) e correlação entre conjuntos de  $\rho = 0,4747$ .

Dada a maior probabilidade de localização do QTL ( $r_a = 10$  cM) dentro do intervalo em estudo, no Quadro 5 apresentam-se os resultados do ajuste da equação generalizada (parâmetros genéticos do QTL). De igual forma, o efeito aditivo do QTL foi altamente significativo ( $p < 0,01$ ), ao contrário do efeito de dominância ( $p > 0,05$ ). Descartando do modelo genético aqueles parâmetros não significativos, a equação de estimação generalizada foi:

$$\text{logit}(p) = \eta = \ln\left(\frac{p}{1-p}\right) = 1,1078x$$

em que  $\eta$  é a função de ligação (*link function*) que neste caso corresponde a *logit* (Mora et al., 2007);  $p$  é a probabilidade de sucesso e  $(1-p)$  fracasso;  $x$  é a variável condicionadora do efeito aditivo do QTL (único parâmetro significativo no modelo genético).

A utilização da técnica GEE (no contexto de medidas repetidas sobre uma mesma unidade experimental) envolve uma análise do tipo multivariado, similar ao caso da distribuição normal (Myers et al., 2002). A diferença fundamental radica no uso de distribuições diferentes à normal, que incluem as distribuições

pertencentes à família exponencial (Dos Santos e Mora, 2007). Alguns exemplos de distribuições contínuas desta família são: Gama, Qui-quadrado, Exponencial, Inversa Gaussiana (ou Inversa Normal), e Geométrica; e de distribuições discretas: Binomial, Poisson, Multinomial, e Binomial Negativa. Segundo Thomson (2003), há vários métodos para detectar *locus* de característica quantitativa, variando desde os métodos de regressão simples e o método dos momentos, até os métodos da máxima verossimilhança e Monte Carlo – Cadeias de Markov (MCMC). A maioria dos procedimentos de mapeamento de QTL considera essencialmente uma distribuição contínua e normal da variável resposta, apesar de que muitas características de interesse científico e econômico seguem distribuição diferente à normal.

Quadro 5 - Resultados do ajuste da equação generalizada na localização do QTL (10 cM), e estimativa dos parâmetros do modelo (efeitos aditivos e de dominância do QTL)

Parâmetro	Estimativa	Erro Padrão	Limites de confiança (95%)		Z	p >  Z
			Inferior	Superior		
Intercepto	0,2333	0,2379	-0,2331	0,6996	0,98	0,3269
Efeito aditivo	1,1128	0,2347	0,6528	1,5728	4,74	<0,0001
Efeito de dominância	-0,0312	0,3635	-0,7436	0,6812	-0,09	0,9316

Análises longitudinais (no tempo ou espaço) ou com respostas correlacionadas (ex. multicaráter), são comuns dentro da experimentação agrônômica (Coppo et al., 2003; Dos Santos e Mora, 2007; Rodríguez et al., 2006). A seleção de plantas incluindo diversas características de interesse econômico é entendida como uma estratégia apropriada de melhoramento genético. De acordo com Khattree e Naik (1999) o enfoque principal da análise multivariada refere-se à análise estatística de dados coletados sobre mais que uma variável resposta. Estas variáveis podem ser correlacionadas (como no presente estudo) e sua dependência estatística (neste caso, genética) deve ser estudada e considerada dentro da avaliação genética – quantitativa.

#### **4. CONCLUSÕES**

A presente análise de QTLs que controlam características de distribuição normal e binomial, realizada por meio de técnicas multivariadas, tem validade para a análise de características longitudinais (medidas repetidas sobre a mesma unidade experimental). De acordo com os resultados, é possível concluir que o mapeamento de QTLs usando um enfoque multivariado pode ser considerada ferramenta útil dentro dos programas genéticos que buscam melhorar características de interesse agrônômico cujos valores fenotípicos mudam no tempo.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, 19(6):716–723, 1974.

CERÓN-ROJAS, J.J.; SAHAGÚN-CASTELLANOS, J. Estimating QTL biometrics parameters in  $F_2$  populations: a new approach. **Agrociencia**, 41:57-73, 2007.

COPPO, J.A.; MUSSART, N.B.; REVIDATTI, M.A.; CAPELLARI, A. Absence of biochemically demonstrable stress in early weaned half-bred zebu calves. **Cien. Inv. Agr.**, 30:69-146, 2003.

DOS SANTOS, A.; MORA, F. Análisis experimental de tratamientos floculantes de residuos orgánicos derivados de la producción porcina. **Cien. Inv. Agr.**, 34:49-56, 2007.

DROSTE, A.; SILVA, A.M.; MATOS, A.V.; ALMEIDA, J.W. In vitro culture of *Vriesea gigantea* and *Vriesea philippocoburgii*: two vulnerable bromeliads native to Southern Brazil. **Braz. Arch. Biol. Technol.**, 48:717-722, 2005.

GELDERMANN, H. Investigations on inheritance of quantitative characters in animals by genes markers. **Theor. Appl. Genet.**, 46:310-330, 1975.

GUZMÁN, G.; LATORRE, B.A.; TORRES, R.; WILCOX, W.F. Relative susceptibility of peach rootstocks to crown gall and Phytophthora root and crown rot in Chile. **Cien. Inv. Agr.**, 34:31-40, 2007.

KHATTREEE, R.; NAIK, D.N. **Applied Multivariate Statistics with SAS® Software**. Cary: SAS Institute Inc., 1999. 338p.

LANDER, E.S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, 121:185-199, 1989.

LANGE, C.; WHITTAKER, J.C. Mapping quantitative trait loci using Generalized Estimating Equations. **Genetics**, 159:1325-1337, 2001.

LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. **Biometrika**, 73:13-22, 1986.

MARTÍNEZ-GÓMEZ, P.; SÁNCHEZ-PÉREZ, R.; RUBIO, M.; DICENTA, F.; GRADZIEL, T.M.; SOZZI, G.O. Application of recent biotechnologies to Prunus tree crop genetic improvement. **Cien. Inv. Agr.**, 32:73-96, 2005.

MORA, F.; SCAPIM, C.A. Predicción de valores genéticos del efecto de poblaciones de maíz evaluadas en Brasil y Paraguay. **Agricultura Técnica**, 67:139-146, 2007.

MORA, F.; PERRET, S.; SCAPIM, C.A.; MARTINS, E.N.; MOLINA, M.P. Source-dependent blooming variability of *Eucalyptus cladocalyx* in the Region of Coquimbo, Chile. **Cien. Inv. Agr.**, 34:131-139, 2007.

MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. **Generalized Linear Models, with Applications in Engineering and the Sciences**. New York: John Wiley and Sons Press, 2002. 342p.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear model. **Journal of the Royal Statistical Society**, 135:370-384, 1972.

ROCHA, R.B.; BARROS, E.G.; CRUZ, C.D.; ROSADO, A.M.; ARAÚJO, E.F. Mapping of QTLs related with wood quality and developmental characteristics in hybrids (*Eucalyptus grandis* x *Eucalyptus urophylla*). **Arvore**, 31:13-24, 2007.

RODRÍGUEZ, G.R.; PRATTA, G.R.; ZORZOLI, R.; PICARDI, L.A. Evaluación de caracteres de planta y fruto en líneas recombinantes autofecundadas de tomate obtenidas por cruzamiento entre *Lycopersicon esculentum* y *L. pimpinellifolium*. **Cien. Inv. Agr.**, 33:111-118, 2006.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, 6:461–464, 1978.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. Viçosa: UFV, 2004. 568p.

THOMSON, P.C. A generalized estimating equations approach to quantitative trait locus detection of non-normal traits. **Genet. Sel. Evol.**, 35:257-280, 2003.

WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, 61:439-447, 1974.

WU, W.R.; LI, W.M.; TANG, D.Z.; LU, H.R.; WORLAND, A.J. Time-related mapping of quantitative trait loci underlying tiller number in rice. **Genetics**, 151:297-303, 1999.

WU, W.; ZHOU, Y.; LI, W.; MAO, D.; CHEN, Q. Mapping of quantitative trait loci based on growth models. **Theor. Appl. Genet.**, 105:1043-1049, 2002.

YANG, R.; TIAN, Q.; XU, S. Mapping quantitative trait loci for longitudinal traits in line crosses. **Genetics**, 173:2339-2356, 2006.

**CAPÍTULO II**  
**REGRESSÃO LINEAR GENERALIZADA NO MAPEAMENTO DE *LOCI* DE**  
**CARACTERÍSTICAS QUANTITATIVAS (QTL) POR INTERVALOS**  
**COMPOSTOS**

**RESUMO**

Em estudos genéticos, a maioria dos métodos de mapeamento de loci de caracteres quantitativos (QTL) pressupõe que a característica contínua segue uma distribuição Normal (Gaussiana). Entretanto, muitas características agronômicas economicamente importantes possuem distribuição não normal. O Modelo Linear Generalizado (GLM) é flexível extensão do método de regressão linear, o qual é útil para variáveis distribuídas não normalmente. O objetivo foi mapear QTL usando uma modelagem linear generalizada, considerando uma situação na qual a característica agronômica não é normalmente distribuída. Para este propósito se simularam dados de marcadores moleculares co-dominantes de uma população F<sub>2</sub>, considerando a existência de um QTL que controla uma característica de distribuição Gama. Tanto GLM como o mapeamento assumindo erroneamente distribuição normal detectaram a presença significativa do QTL ( $P < 0,01$ ), porém a posição do QTL diferiu em 5 centiMorgan entre as metodologias, e foi localizado em diferentes intervalos. As diferenças podem ser relativamente pequenas, porém foi enfatizado o fato de que os resultados apresentados neste estudo são provenientes de um simples conjunto de dados moleculares (1 cromossomo). O uso do GLM pode oferecer melhorada eficiência na análise de QTL que controlam características de distribuição não-Gaussiana.

Palavras-chave: mapeamento de QTL, distribuição não normal, modelos lineares generalizados

## ABSTRACT

In genetic studies, most Quantitative Trait Loci (QTL) mapping methods presuppose that the continuous trait of interest follows a Normal (Gaussian) distribution. However, many economically important traits of agricultural crops have a non-normal distribution. The Generalized Linear Model (GLM) is a flexible extension of lineal regression method that has been especially useful for non-normally distributed variables. The objective of this study was to map QTL by using a generalized linear regression modeling approach, in a situation where the agronomical trait is non-normally distributed. Data of co-dominant molecular markers from a F2 population were simulated considering the existence of a QTL controlling an agronomical trait with Gamma distribution. Both the GLM and wrongly assuming a normal distribution, detected the presence of a significant QTL ( $P < 0.01$ ), but the QTL position differed by 5 centiMorgan between the approaches, and was located at different marker intervals. The differences may seem relatively small, but we emphasize the fact that the results presented here are based on an extremely simple molecular data set (1 chromosome). The use of GLM methodology can offers improved efficiency in the analysis of QTLs controlling traits of non-Gaussian distribution.

Key words: QTL mapping, non-normal distribution, generalized linear models.

## 1. INTRODUÇÃO

A grande maioria das características de importância econômica resulta da ação conjunta de vários genes, e o seu fenótipo geralmente apresenta variação contínua. A expressão de tais características quantitativas pode ser controlada por *loci* denominados QTL (*Quantitative Trait Loci*). Em geral, existem duas categorias básicas de marcadores moleculares (Ovesná et al., 2002): 1) marcadores que segregam e determinam a presença de um único gene, dominante ou recessivo; 2) marcadores associados a um QTL. É muito mais fácil e econômico desenvolver marcadores para uma característica que é determinada por um único gene do que aqueles baseados em *loci* de características quantitativas. Por este motivo, a detecção e a análise de QTLs devem ser realizadas com a melhor precisão possível.

A associação entre um loco marcador e a região genômica de um QTL e o mapeamento da suas localizações dentro do cromossomo fornecem uma importante ferramenta para os programas de melhoramento genético. Os genótipos de QTLs conhecidos podem ser somados à informação de desempenho dos indivíduos e, juntamente com as informações genealógicas, podem ser utilizados para aumentar a precisão da predição dos valores genéticos nos métodos tradicionais de melhoramento (Gonçalves-Vidigal et al., 2008), como o caso da Melhor Predição Linear Não Tendenciosa (BLUP - Best Linear Unbiased Prediction).

A estratégia para verificar se um marcador e um QTL estão ligados é a análise estatística da associação entre a variação fenotípica da característica e dos marcadores. Os métodos de regressão simples, da máxima verossimilhança e aqueles baseados nos métodos de Monte Carlo - Cadeias de Markov têm sido freqüentemente utilizados para detectar QTLs, os quais têm se focalizado essencialmente em uma distribuição contínua (Normal ou Gaussiana) da característica de interesse (Thomson, 2003). Entretanto, diversos estudos relacionados ao melhoramento de plantas têm mostrado que várias características de interesse econômico e científico não apresentam distribuição Gaussiana (Mora et al., 2007; Mora et al., 2008; Spyrides-Cunha et al., 2000; Ribeiro et al., 2005).

Neste sentido, na década de setenta foi desenvolvida a metodologia dos Modelos Lineares Generalizados (GLM: Generalized Linear Models), a qual é aplicada a dados com distribuições da família exponencial, e utiliza métodos análogos aos métodos lineares tradicionais, para dados com distribuição normal (Myers et al., 2002). As distribuições tais como Gama, Poisson, Binomial, Multinomial e Normal são alguns exemplos de distribuições da família exponencial, as quais são freqüentemente encontradas nos experimentos biológicos (Mora et al., 2008; Dos Santos e Mora, 2007). GLM permite a análise paramétrica dispensando a pressuposição de normalidade, necessária para a análise convencional.

O presente estudo tem sido motivado pela existência de características de interesse econômico que apresentam distribuições diferentes à Gaussiana; tais como Bernoulli (ausência e presença de um evento biológico), Poisson (dados de contagem), Gama, Multinomial, dentre outras. Assim, o presente trabalho teve como objetivo analisar os efeitos da distribuição da variável resposta (característica quantitativa sob seleção) na detecção e localização de um loci de característica quantitativa. O conhecimento acerca das possíveis implicações do método de análise no mapeamento de QTL permitirá melhorar a precisão das análises genético-quantitativas e melhorar o entendimento acerca das regiões genômicas que controlam características de interesse agrônomo.

## 2. MATERIAL E MÉTODOS

A metodologia da Regressão Linear Generalizada foi utilizada para o mapeamento de um QTL que controla uma característica de distribuição não Gaussiana. Para tal propósito, simularam-se dados de onze marcadores moleculares co-dominantes distribuídos dentro de um cromossomo de 105,6 centiMorgan (cM, distância de Haldane) usando o programa QMOL (Schuster e Cruz, 2004). A população de melhoramento consistiu em uma população  $F_2$  com 252 indivíduos.

Assumiu-se que a característica sob seleção é geneticamente controlada por infinito número de locos aditivos, cada um com efeito infinitesimal (poligenes), mais um QTL bi-alelo único (alelos Q e q). A característica foi obtida dos ensaios do Banco de Germoplasma do Instituto de Investigación Agropecuaria (INIA), do Ministério da Agricultura do Chile, correspondendo a valores de produção precoce de azeitonas (expressadas em  $\text{kg}\cdot\text{árvore}^{-1}$ ) e, posteriormente, foi ajustada aos dados dos marcadores moleculares.

De acordo com os pressupostos prévios, a característica quantitativa usada no presente estudo (produção de azeitonas) possui uma distribuição pertencente à família exponencial (Dobson, 2001). A função de densidade e probabilidade para a resposta observada  $y$ , pode ser expressa na seguinte forma:

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

em que  $a(\phi)$ ,  $b(\theta)$  e  $c(y, \phi)$  são chamadas funções específicas; o parâmetro  $\theta$  é o parâmetro de posição natural (*natural location parameter*), e  $\phi$  é o freqüentemente chamado de parâmetro de dispersão (Dobson, 2001; Myers et al., 2002).

O modelo de regressão, considerando o mapeamento de QTL por intervalo composto (Haley e Knott, 1992), é o seguinte:

$$y_j = [u + ax_j^* + dz_j^*] + \sum b_k c_i + e_j$$

em que  $y_j$  corresponde ao valor da característica no indivíduo  $j$ ;  $\mu$  é a média geral;  $a$  é o efeito aditivo do locus que está sendo estudado;  $d$  é o efeito de dominância do locus estudado;  $x_j^*$  e  $z_j^*$  são as variáveis condicionadoras e dependentes dos genótipos dos marcadores moleculares que flanqueiam o QTL;  $\sum b_k c_i$  corresponde aos co-fatores;  $e_j$  é o erro aleatório. No Quadro 1 se apresentam as probabilidades condicionais associadas aos genótipos de um possível QTL, dado um intervalo específico.

O Modelo Linear Generalizado (GLM) para o mapeamento do QTL, com distribuição pertencente à família exponencial, é o seguinte:

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

em que  $\eta_i = g(\mu_i)$  corresponde à função de ligação.  $\beta_0 + \sum_{i=1}^k \beta_i x_i$ , são os parâmetros do modelo (o intercepto, efeitos aditivos e de dominância do QTL, e os co-fatores).

Quadro 1 - Probabilidades condicionais associadas aos genótipos de um QTL no intervalo dos marcadores A e B, em uma população  $F_2$ . Não se consideram as duplas recombinações

Genótipo	Genótipo do QTL		
	<i>QQ</i>	<i>Qq</i>	<i>qq</i>
AABB	1	0	0
AABb	(1-p)	(p)	0
AAbb	(1-p) <sup>2</sup>	[2p(1-p)]	(p <sup>2</sup> )
AaBB	(p)	(1-p)	0
AaBb	[qp(1-p)]	[1-2qp(1-p)]	[qp(1-p)]
Aabb	0	(1-p)	(p)
aaBB	(p <sup>2</sup> )	[2p(1-p)]	(1-p) <sup>2</sup>
aaBb	0	(p)	(1-p)
aabb	0	0	1

Fonte: Schuster e Cruz (2004);  $p=r_a/r$ ;  $(1-p)=r_b/r$ ;  $q=r^2/[r^2+(1-r)^2]$ ; *QQ*, *Qq* e *qq*, representam os genótipos homozigoto dominante, heterozigoto e homozigoto recessivo, respectivamente.

O procedimento GENMOD do SAS (SAS-Institute, 1996) foi utilizado para o mapeamento do QTL, de acordo com Myers et al. (2002), para ajustar o modelo de regressão linear generalizado. O teste da razão da verossimilhança (LRT; *Likelihood Ratio Test*) foi utilizado para comparar os modelos de cada intervalo de mapeamento. Estes resultados foram comparados com uma análise assumindo uma distribuição normal da característica de interesse agrônômico.

### 3. RESULTADOS E DISCUSSÃO

Na Figura 1 (a), se apresenta o resultado do grupo de ligação (cromossomo) com suas respectivas distâncias entre os marcadores moleculares co-dominantes. Os valores de distância de Haldane variaram de 5,5 (estimada entre os marcadores SSR6 e SSR7) a 19,1 cM (entre os marcadores SSR2 E SSR3). A distribuição observada e esperada da característica de produção precoce de azeitonas ( $\text{kg árvore}^{-1}$ ) é apresentada na Figura 1 (b). O relativamente pequeno viés dos dados em relação à distribuição normal foi confirmado pelos testes estatísticos Shapiro-Wilk e Lilliefors ( $P < 0,01$ ). A análise gráfica confirma que a variável resposta possui distribuição Gama, de acordo com Freund (1992).

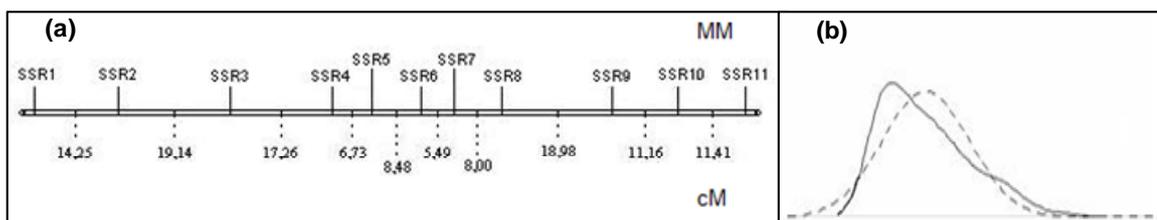


Figura 1 - (a) Esquematização do cromossomo (grupo de ligação) construído com onze marcadores moleculares (MM); na parte inferior do cromossomo, são mostradas as distâncias de Haldane em cM; (b) Gráfico de distribuição esperada (Gaussiana: linha pontuada) e observada (Gama: linha contínua) da característica quantitativa.

Considerando que a característica apresenta uma distribuição Gama, a função de densidade para  $y$  é dada por:

$$f(y) = \frac{1}{\Gamma(r)} \left( \frac{1}{\lambda} \right)^r e^{-y/\lambda} y^{r-1}, \quad y \geq 0, r > 0, \lambda > 0$$

Se esta função de densidade é colocada dentro da forma exponencial, temos:

$$\theta = -\frac{1}{\lambda r} = -\frac{1}{\mu}$$

$$\mu = r\lambda$$

$$\text{var}(y) = \frac{\mu^2}{r} \Rightarrow \text{var}(y) = r\lambda^2$$

$$b(\theta) = -\ln(-\theta)$$

$$a(\phi) = r^{-1}$$

$$c(\phi) = r \ln r - \ln \Gamma(r) + (r-1) \ln y$$

em que  $r$  é o parâmetro escalar (*scale*). Quando  $r=1$ , a distribuição altera-se para uma distribuição exponencial. No presente estudo, para a regressão linear generalizada de mapeamento do QTL, foi utilizada a função de ligação logarítmica de acordo com Myers et al. (2002).

No Quadro 2 se apresenta o resultado da análise realizada para identificar aqueles marcadores que estão associados com a característica produção de azeitonas. Este procedimento é chave para a identificação de co-fatores que serão incorporados no mapeamento por intervalo composto. A regressão linear generalizada identificou aos marcadores SSR3 e SSR11 como significativos ( $P < 0,05$ ). Uma análise similar realizada usando regressão de Stepwise confirmou estas marcas moleculares, porém também evidenciou significância estatística para o marcador SSR9 ( $P < 0,05$ ). Embora os resultados do mapeamento por intervalos compostos não evidenciassem diferenças em relação ao procedimento de intervalos simples (dados não mostrados), devido principalmente à simplicidade dos dados genômicos, é importante destacar que as diferenças encontradas na seleção de co-fatores podem afetar a detecção de QTLs (Schuster e Cruz, 2004).

Quadro 2 - Regressão Linear Generalizada usada para a seleção de marcadores moleculares como co-fatores para o mapeamento de QTL por intervalo composto. Em negrito, os marcadores moleculares significativos ( $P < 0,05$ )

Marcador	Estimativa	Erro Padrão	Limites de Confiança Wald (95%)		$\chi^2$	$P > \chi^2$
SSR1	-0,021	0,1595	-0,3336	0,2916	0,02	0,8952
SSR2	0,1213	0,1845	-0,2403	0,4828	0,43	0,511
SSR3	0,4454	0,1553	0,1409	0,7498	8,22	0,0041
SSR4	0,1952	0,1954	-0,1878	0,5782	1,00	0,3179

SSR5	0,0605	0,2715	-0,4717	0,5927	0,05	0,8236
SSR6	-0,152	0,2813	-0,7034	0,3994	0,29	0,589
SSR7	0,2954	0,255	-0,2044	0,7952	1,34	0,2467
SSR8	-0,2948	0,2304	-0,7464	0,1567	1,64	0,2007
SSR9	-0,0398	0,1703	-0,3736	0,2941	0,05	0,8155
SSR10	-0,1234	0,1796	-0,4754	0,2286	0,47	0,492
SSR11	0,3671	0,1586	0,0562	0,678	5,36	0,0207

Na Figura 2 se apresenta um gráfico com a localização do QTL dependendo do método de análise utilizada: 1. Ignorando a distribuição real da característica, isto é, assumindo erroneamente uma distribuição normal (QTL1) e 2. Assumindo a distribuição verdadeira da variável resposta (Gama) e, portanto, utilizando o princípio dos Modelos Lineares Generalizados para o mapeamento do QTL (QTL2).

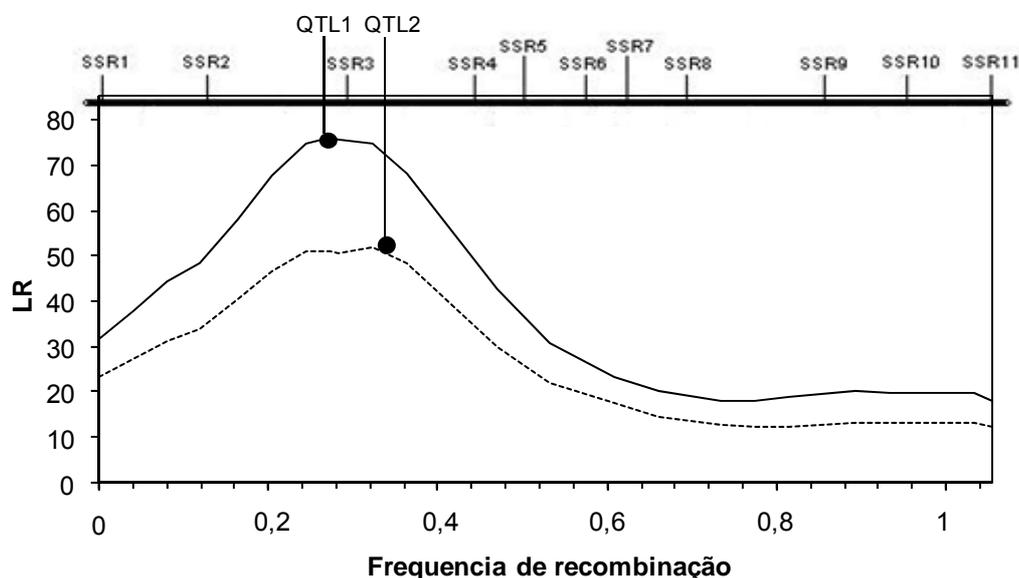


Figura 2 - Localização do QTL assumindo erroneamente uma distribuição normal (QTL1) e assumindo a verdadeira distribuição da variável resposta (Gama) em um modelo de regressão linear generalizado (QTL2). LR é o teste da razão de verossimilhança; SSR1, SSR2,..., e SSR11 são os nomes dos marcadores moleculares co-dominantes.

No presente estudo, pôde-se observar que a análise de mapeamento de QTL, quando a rejeição de normalidade foi ignorada, resultou relativamente robusta no sentido de que ambos os enfoques de análise detectaram uma presença significativa do QTL ( $P < 0,01$ ) dentro do cromossomo analisado. Entretanto, a localização do QTL, entre uma metodologia e outra, deferiu em 5 cM, apesar do pequeno viés na distribuição da característica. Outro resultado importante é que o QTL foi detectado em intervalos distintos.

O modelo linear generalizado, que foi ajustado aos dados do presente estudo, considerou a função de ligação logarítmica. Tecnicamente, tal ligação não é a função canônica (natural) da distribuição Gama, porém é freqüentemente utilizada (Myers et al., 2002) porque permite superar inconvenientes matemáticos com o uso da função recíproca, que seria a ligação canônica da distribuição Gama.

A modelagem linear generalizada forneceu importantes resultados no mapeamento do QTL. Considerando que 1 cM é equivalente a aproximadamente 1 milhão de pares de bases, no presente estudo, as diferenças encontradas entre um procedimento e outro podem ser, portanto, da ordem de 5 milhões de pares de base. Além disso, se é considerado o mapeamento do QTL ignorando a pressuposição de normalidade, pode significar na perda de importantes recursos ao momento de aumentar a saturação do mapa entre os intervalos dos marcadores SSR2 e SSR3, com o intuito de aumentar a precisão da localização do QTL.

Originalmente, a teoria dos GLM foi considerada como um método alternativo para a análise de dados em que o pressuposto de normalidade é irrealista; como o apresentado no presente estudo. O uso da metodologia GLM pode ser interessante e eficaz no mapeamento de QTLs, em situações onde a variável resposta segue uma distribuição pertencente à família exponencial de distribuições (Dobson, 2001). Embora, existam vários métodos para detectar QTLs que controlam uma característica quantitativa, desde os métodos de regressão simples até os métodos Bayesianos (Thomson, 2003), a maioria dos procedimentos considera uma distribuição normal das características.

Muitas características de interesse tanto científico como econômico apresentam distribuição não-normal (Thomson, 2003). Por exemplo, dados binários são freqüentemente encontrados quando se deseja melhorar

características como *status* de doenças (Park et al., 2002; Setiawan et al., 2000), mortalidade ou sobrevivência (Pesqueira et al., 2006; Azadi e Khosh-Khui, 2007), florescimento (Mora et al., 2008; Missiaggia et al., 2005), entre outras características. Neste sentido, o presente estudo confirmou que o mapeamento de QTLs que controlam características de distribuição não Gaussiana pode ser aperfeiçoado por meio da utilização da teoria dos Modelos Lineares Generalizados.

#### **4. CONCLUSÕES**

A precisão no mapeamento de QTL dependeu da abordagem metodológica utilizada. A informação acerca da distribuição dos dados agronômicos deve ser estudada e incluída nos programas de melhoramento genético, para melhorar a confiabilidade do mapeamento de QTL. Neste sentido, a modelagem linear generalizada pode ser usada eficientemente no mapeamento de QTLs que controlam características de distribuição não-Gaussiana.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

AZADI, P.; KHOSH-KHUI, M. Micropropagation of *Lilium ledebourii* (Baker) Boiss as affected by plant growth regulator, sucrose concentration, harvesting season and cold treatments. . **J. Biotechnol.**, 10:582-591, 2007.

DOBSON, A.J. **An introduction to generalized linear models**. London: Chapman and Hall, 2001. 225p.

DOS SANTOS, A.; MORA, F. Análisis experimental de tratamientos floculantes de residuos orgánicos derivados de la producción porcina. **Cien. Inv. Agr.**, 34:49-56, 2007.

FREUND, J.E. **Mathematical statistics**. New Jersey: Prentice-Hall, 1992. 658p.

GONÇALVES-VIDIGAL, M.C.; MORA, F.; BIGNOTTO, T.S.; MUNHOZ, R.E.F.; SOUZA, L.D. Heritability of quantitative traits in segregating common bean families using a Bayesian approach. **Euphytica**, 164:551-560, 2008.

HALEY, C.S., KNOTT, S.A. A simple regression method for mapping quantitative trait loci in linecrosses using flanking markers. **Heredity**, 69:315-324, 1992.

MISSIAGGIA, A.A.; PIACEZZI, A.L.; GRATTAPAGLIA, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. **Tree Genetics & Genomes**, 1:79-84, 2005.

MORA, F.; PERRET, S.; SCAPIM, C.A.; MARTINS, E.N.; MOLINA, M.P. Source-dependent blooming variability of *Eucalyptus cladocalyx* in the Region of Coquimbo, Chile. **Cien. Inv. Agr.**, 34:131-139, 2007.

MORA, F.; TAPIA, F.; SCAPIM, C.A.; MARTINS, E.N.; PINTO, R.J.B.; IBACACHE, A. Early performance of *Olea europaea* cv. Arbequina, Picual and Frantoio in southern Atacama Desert. **Crop Breeding and Applied Biotechnology**, 8:30-38, 2008.

MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. Generalized linear models, with applications in engineering and the sciences. New York: John Wiley and Sons Press, 2002. 342p.

OVESNÁ, J., POLÁKOVÁ, K. LEIŠOVÁ, L. DNA analyses and their applications in plant breeding. **Czech J. Genet. Plant Breed.**, 38:29-40, 2002.

PARK, S.O.; COYNE, D.P.; STEADMAN, J.R.; SKROCH, P.W. Mapping of QTL for resistance to white mold disease in common bean. **Crop Science**, 41:1253-1262, 2001.

PESQUEIRA, J.; GARCIA, M.D.; STALTARI, S.; MOLINA, M.C. NaCl effects in *Zea mays* L. x *Tripsacum dactyloides* (L.) L. hybrid calli and plants. **Electron. J. Biotechnol.**, 9:286-290, 2006.

RIBEIRO, A.O.; BEARZOTI, E.; SÁFADI, T. QTL mapping of Poisson traits: a simulation study. **Crop Breeding and Applied Biotechnology**, 5:310-317, 2005.

SAS-INSTITUTE. **Statistical analysis system user's guide**. Cary: SAS Institute, 1996. 956p.

SETIAWAN, A.; KOCH, G.; BARNES, S.R.; JUNG, C. Mapping quantitative trait loci (QTLs) for resistance to *Cercospora* leaf spot disease (*Cercospora beticola* Sacc.) in sugar beet (*Beta vulgaris* L.). **Theoretical and Applied Genetics**, 100:1176-1182, 2000.

SCHUSTER, I.; CRUZ, C.D. Estatística genómica aplicada a populações derivadas de cruzamentos controlados. Viçosa: Universidade Federal de Viçosa, 2004. 568p.

SPYRIDES-CUNHA, M.H.; DEMETRIO, C.G.B.; CAMARGO, L.E.A. Proportional odds model applied to mapping of disease resistance genes in plants. **Genetics and Molecular Biology**, 23:223-227, 2000.

THOMSON, P.C. A generalized estimating equations approach to quantitative trait locus detection of non-normal traits. **Genet. Sel. Evol.**, 35:257-280, 2003.

**CAPÍTULO III**  
**MODELOS LINEARES GENERALIZADOS NA AVALIAÇÃO DE *Olea europaea***  
**CV. ARBEQUINA, PICUAL E FRANTOIO NO SUL DO DESERTO DE**  
**ATACAMA**

**RESUMO**

*Olea europaea* cv. Arbequina, Picual e Frantoio são cultivados em vários países Mediterrâneos. Em 1999, estes cultivares foram plantados em três locais da Região de Coquimbo; uma zona árida Mediterrânea do Chile. Uma modelagem linear generalizada foi conduzida porque o conjunto de dados agronômicos não teve distribuição Normal. Produção de frutos (medida nas safras 2002-2003), precocidade (2002) e a sobrevivência das árvores (quatro anos após a plantação) diferiram significativamente entre os cultivares. Arbequina e Picual tiveram um positivo efeito sobre a produção. Picual teve alta precocidade em dois locais. Frantoio teve alta sobrevivência nos três locais (90-100%) ao contrario de Picual (56-83%). Os Modelos Lineares Generalizados foram úteis em uma situação onde o pressuposto de normalidade não foi atendido. A seleção de cultivares tem importante potencial na região árida do Chile, mas sua efetividade dependerá da seleção de genótipos que respondem bem sobre um determinado local.

Palavras-chave: modelos lineares generalizados, produção de azeitonas, ambientes áridos

## ABSTRACT

The cultivars Arbequina, Picual and Frantoio of *Olea europaea* are cultivated in several Mediterranean countries. In 1999, these cultivars were planted at three locations in the region of Coquimbo, an arid Mediterranean-like area in Chile. A generalized linear modeling approach was used in view of the non-normal distribution of the agronomic data sets. Fruit yield (harvested of 2002-2003), precocity (2002) and tree survival (after four growing seasons) differed significantly between the cultivars. Arbequina and Picual had a positive effect on the yield. Picual was the earliest cultivar at two sites. The survival rate of Frantoio was high at the three sites (90-100%), as opposed to Picual (56-83%). The approach of Generalized Linear Models was particularly useful where the assumption of normality was not satisfied. The selection of cultivars is promising in this arid region of Chile, while the success will depend on the selection of well-adapted genotypes to a particular location.

Key words: generalized linear models, olive production, arid environments

## 1. INTRODUÇÃO

As variedades de oliveira (*Olea europaea* L.) Arbequina, Picual e Frantoio têm muitas aplicações práticas e são cultivares muito importantes para a produção de azeite virgem ou extravirgem. Picual e Arbequina são preferidas em diversas regiões Mediterrâneas da Espanha devido a sua precocidade, alta produtividade de azeite e por outras vantagens agronômicas (Tous et al., 1998). O azeite de oliva proveniente de Picual representa cerca de 25% da produção mundial de azeite (Brenes et al., 2002). Espanha, Itália e Grécia produzem 75% da produção mundial de azeite, sendo a Espanha responsável pela metade das colheitas do fruto da oliveira. A produção mundial de azeite de oliva é aproximadamente 2 milhões de toneladas (Brenes et al., 2002). Arbequina, originada da Espanha, também conhecida como Alberquina, de acordo com o banco de dados da coleção de germoplasma de oliva da FAO, é um dos cultivares mais amplamente usado para a produção de azeite (Lopes et al., 2004).

Na bacia do mar Mediterrâneo existem muitas variedades de árvores de oliveira, sendo a região responsável pela produção de 99% e pelo consumo de 87% do azeite (Loumou e Giourga, 2003). Conseqüentemente, a produção de azeite de oliva e seu cultivo são freqüentemente referidos às condições de clima mediterrâneo, onde centenas de cultivares de oliveira têm sido selecionadas durante séculos em função da sua adaptação a variados micro-ambientes e tipos de solo. A região de Chile central tem clima de tipo mediterrâneo e durante a década passada foram empreendidos diversos projetos de pesquisa, financiados pelas agências governamentais desse país, para avaliar materiais genéticos incluindo as cultivares de Arbequina, Picual e Frantoio.

Quando as cultivares de oliveira são avaliados à idade precoce, com o intuito de obter genótipos precoces (ou seja, melhorar a habilidade de uma cultivar para começar a produzir a uma idade precoce) é muito comum encontrar árvores sem frutos e, por outra parte, árvores com maior produção. Além disso, as cultivares de oliveira, ou árvores individuais, podem evidenciar alternância da produção (do inglês *alternate bearing*), a qual é devida principalmente a uma inibição da indução floral (Rallo et al., 1994). Esta alternância não necessariamente ocorre simultaneamente em todas as cultivares, e algumas vezes em árvores individuais dentro de uma mesma cultivar (Tous et al., 1998).

Em conseqüência disso, a pressuposição de normalidade da variável resposta pode ser completamente irrealista. Outro exemplo é quando se tem uma resposta de tipo binária, como por exemplo, na análise da sobrevivência das árvores, a qual é registrada com valores 0 e 1. A precocidade das árvores frutíferas pode também ser estudada considerando-a como uma resposta de tipo binária. Um agricultor pode estar interessado em prever se uma cultivar de oliveira produz ou não precocemente sob condições ambientais específicas.

Nelder e Wedderburn (1972) propuseram a teoria dos Modelos Lineares Generalizados (GLM; do inglês *Generalized Linear Models*) como um método alternativo para a análise de dados na qual a pressuposição de normalidade não é atingida. GLM é uma extensão natural dos modelos lineares clássicos e fornece uma teoria de estrutura conceitual unificada para diversos métodos estatísticos utilizados freqüentemente (Dobson, 2001). O uso da metodologia GLM é interessante para a experimentação biológica, em situações onde a variável resposta segue uma ampla classe de distribuições, chamada de família exponencial (Demetrio, 2001; Soares et al. 2005).

O modelo linear generalizado assume que as observações são independentes e não considera correlações entre a resposta das observações. Existem muitas situações na quais as medidas de respostas repetidas são feitas sobre uma mesma unidade experimental. Esta análise requer o uso de técnicas estatísticas especiais porque as respostas tendem a ser fortemente correlacionadas (Myers et al. 2002). A alternância da produção nas árvores de oliveira pode ser claramente um bom exemplo disto, sendo um sério problema no cultivo de numerosas espécies de árvores frutíferas, incluindo as oliveiras (Goldschmidt, 2005; Levin e Lavee, 2005). Um típico exemplo de respostas correlacionadas é um estudo longitudinal onde cada unidade experimental é seguida durante um período de tempo, e onde são registradas observações repetidas da variável resposta e as co-variáveis relevantes do experimento. Quando as medições são realizadas repetidamente sobre uma mesma unidade experimental, as respostas observadas são geralmente correlacionadas (Pan e Connett, 2002).

Liang e Zeger (1986) propuseram o procedimento estatístico das Equações de Estimação Generalizadas (GEE; do inglês: *Generalized Estimating Equations*) como uma aplicação da quase-verossimilhança (Wedderburn, 1974)

para superar o problema de ajustar um modelo linear generalizado a dados correlacionados. A metodologia GEE é considerada uma intuitiva extensão do Modelo Linear Generalizado.

O presente estudo foi empreendido para investigar a resposta precoce das variedades de Arbequina, Picual e Frantoio sob condições ambientais áridas do norte do Chile. Neste trabalho se reportam os primeiros resultados dos estudos conduzidos com essas variedades na Região Administrativa de Coquimbo. Devido à ausência de normalidade no conjunto de dados, foi realizada uma modelagem linear generalizada. O entendimento e utilização deste procedimento estatístico poderiam conduzir ao aprimoramento das inferências dos resultados dos experimentos com as cultivares de oliveira, naquelas situações nas quais a pressuposição de normalidade não é atingida.

## 2. MATERIAIS E MÉTODOS

### 2.1. Ensaio de campo

A avaliação do experimento foi conduzida em três locais na zona árida do norte do Chile: dois locais da Província de Limarí (Limarí-1: 30° 35' S, 71° 28' W, 200 m de altitude, e Limarí-2: 30° 45' S, 70° 44' W, 820 m de altitude) e um local da Província de Choapa (Choapa: 31° 50' S, 70° 51' W, 580 m de altitude). Estes locais estão em uma área árida de transição localizada entre o deserto de Atacama e as terras semi-áridas do Chile central (Ormazábal, 1991; Squeo et al., 2006).

O delineamento experimental foi o inteiramente ao acaso com três tratamentos (cultivares de oliveira: Arbequina, Picual e Frantoio), considerando um esquema fatorial, com dois fatores: cultivar e local (fatorial 3x3), e variável número de repetições (58 em Limarí-1; 52 em Limarí-2; 48 em Choapa). As árvores foram plantadas a 8 x 4 m. Os protocolos de manejo da cultura e a caracterização específica dos locais têm sido já descritos em artigos prévios (Astorga e Mora, 2005) onde foram avaliados diversos experimentos contíguos que consideraram outras cultivares de oliveira.

### 2.2. Variáveis analisadas

Os locais foram plantados manualmente em 1999. A produção de frutos por árvore ( $\text{kg árvore}^{-1}$ ) foi avaliada nas temporadas agrícolas 2002 e 2003. Com esta coleção de dados, duas características quantitativas foram consideradas:

1. Produção de frutos ( $\text{kg árvore}^{-1}$ ), avaliada por meio de um estudo longitudinal (ambas as temporadas agrícolas).

2. Precocidade, a qual foi medida como a ocorrência ou não ocorrência de um determinado nível de produção à idade de 3 anos, na temporada agrícola 2002. Esta resposta binária foi baseada na seguinte pressuposição (Tous et al. 1998): valor 0 para aquela árvore que não produziu frutos, de um ponto de vista econômico (ou seja, uma produção menor que  $5 \text{ kg árvore}^{-1}$ ), e valor 1 para uma produção individual maior que (ou igual a)  $5 \text{ kg árvore}^{-1}$ .

Adicionalmente, a sobrevivência foi medida quatro anos depois do plantio, na temporada agrícola 2003. Esta variável resposta segue uma distribuição Binomial, a qual foi codificada como 1 para sobrevivência, e 0 em caso contrário.

### 2.3. Análise estatística para a produção de frutos

As pressuposições prévias de normalidade para a produção de frutos de oliveira foram checadas por meio de uma inspeção visual de gráfico de normalidade e utilizando os seguintes testes estatísticos: Shapiro-Wilk, Qui-quadrado e Lilliefors (Silva et al., 2006). O gráfico de probabilidade normal e o teste de Shapiro-Wilk foram obtidos usando os procedimentos UNIVARIATE e SAS-INSIGHT (SAS-Institute, 1996). Os testes de Qui-quadrado e Lilliefors foram testados no programa GENES (Cruz, 2006).

Na presença de problemas com as pressuposições de normalidade, a avaliação experimental foi conduzida usando a metodologia dos modelos lineares generalizados (Nelder e Wedderburn, 1972). Como ambas as temporadas agrícolas foram conjuntamente avaliadas (análise longitudinal), as análises dos dados agronômicos foram conduzidas usando as Equações de Estimação Generalizadas (Liang e Zeger, 1986), uma extensão do procedimento GLM para ajustar um modelo linear generalizado a dados correlacionados.

GEE e GLM assumem que as funções de densidade probabilidade para uma característica observada podem ser expressas como uma distribuição geral chamada de família exponencial de distribuições (Demétrio, 2001; Dobson 2001; Myers et al., 2002).

As árvores individuais medidas repetidamente durante duas temporadas agrícolas foram referidas como o sujeito de análise (*subject*) e, portanto as observações no mesmo *subject* foram consideradas correlacionadas, derivando o problema para uma análise multivariada (Myers et al. 2002). Para um específico *subject* ( $j$ ), tem-se o seguinte modelo linear:

$$y_j = X_j \beta + \varepsilon_j$$

em que  $y_j$  é o vetor que contém as observações da variável resposta sobre o  $j$ -ésimo *subject*,  $X_j$  é a matriz de delineamento que definem as

contribuições ao modelo dos efeitos da cultivar, local e interação,  $\beta$  é um vetor com os coeficientes de regressão ou parâmetros do modelo,  $\varepsilon_j$ , é um vetor de erros aleatórios. A distribuição básica marginal para a produção de frutos de oliveira *per se* é desconhecida. Subseqüentemente, a distribuição aproximada da variável resposta foi confirmada por meio de uma inspeção visual gráfica da resposta usando o programa **R** de acordo a Dos Santos e Mora (2007). O modelo foi também testado usando o procedimento GENMOD do SAS (SAS-Institute, 1996).

A abordagem GEE estima  $\beta$  resolvendo as seguintes equações de estimação (Liang e Zeger, 1986):

$$\sum_{j=1}^n D_j' V_j^{-1} (y_j - \mu_j) = 0$$

em que  $D$  é a matriz das derivadas parciais  $\partial\mu_i / \partial\beta_j$ ;  $V$  é a matriz de variância - covariância das observações sobre o  $j$ -ésimo *subject*. A expressão anterior é chamada freqüentemente função quase-escore por que envolve a quase-verossimilhança. As estimações de quase-verossimilhança foram consideradas envolvendo o uso dos quadrados mínimos generalizados.

A abordagem das equações de estimação generalizadas foi conduzida usando o procedimento GENMOD, com a opção REPEATED (SAS-Institute, 1996). A opção TYPE foi usada para escolher uma apropriada estrutura de correlação para as estimativas dos parâmetros das GEE (Horton e Lipsitz, 1999; Myers et al. 2002).

#### 2.4. Análise estatística para sobrevivência e precocidade

A precocidade das cultivares e a sobrevivência das árvores (resposta binária) foram analisadas usando Modelos Lineares Generalizados (GLM). Nesta análise foi assumido que a variável resposta tem uma distribuição Binomial com parâmetro  $n$  e  $P$ , com a seguinte função de densidade de probabilidade:

$$f(y; P; \theta) = \exp \left\{ y \ln(P/1-P) - n \ln(1 + e^\theta) + \ln \left( \frac{n}{y} \right) \right\}$$

em que  $\theta$  denota um parâmetro natural de locação. A distribuição Binomial pertence à família exponencial de distribuições que pode ser expressa na forma:

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

em que  $a(\phi) = 1$ ,  $b(\theta) = n \ln(1 + e^\theta)$ ,  $\theta = \ln(P/1 - P)$ , e  $c(\cdot) = \ln(n/y)$ . A função de ligação *logit* foi empregada usando o procedimento GENMOD (SAS-Institute, 1996), a qual ajusta o modelo linear generalizado aos dados por meio da estimativa da máxima verossimilhança do vetor de parâmetros  $\beta$ , com a seguinte equação escore de máxima verossimilhança:

$$X'(y - \mu) = 0$$

### 3. RESULTADOS E DISCUSSÃO

#### 3.1. Produção de frutos

O gráfico de probabilidade da distribuição normal e os testes usados para checar as pressuposições de normalidade indicaram que não é razoável supor a distribuição normal para a característica de interesse ( $P < 0,01$ ). A análise conduzida nos programas R e GENES indicou falta de simetria na distribuição dos dados. A moda (1,63 kg árvore<sup>-1</sup>, estimada usando o método de estimação de densidade de Kernel) foi mais baixa que a média (13,24 kg árvore<sup>-1</sup>), confirmando a assimetria. Conseqüentemente, como alternativa de análise, foi usado um modelo linear generalizado com distribuição Gama, de acordo com Freund (1992) e função de ligação logarítmica. A distribuição Gama tem aplicações potenciais em estudos em que a variável resposta é contínua e a variância não é constante, porém proporcional ao quadrado da média (Myers et al., 2002) como ocorre no presente estudo.

Neste estudo, 29% do total das árvores não tiveram produção de frutos considerando ambas as temporadas agrícolas. Essas observações são consideradas valores de resposta inválidos para a distribuição Gama (Freund, 1992), porém, este problema foi resolvido por meio da adição de um valor arbitrário (i.e. 0,5) às observações. Estudos prévios (Myers et al., 2002) indicaram que tal abordagem não afeta a distribuição dos dados (um fato confirmado neste estudo).

A análise das equações de estimação generalizada é mostrada no Quadro 1, admitindo uma combinação satisfatória entre a distribuição da variável resposta (Gama) e a função de ligação *Log* (Demetrio, 2001). Tecnicamente a função *Log* não é a ligação canônica. Entretanto, estudos prévios (Myers et al., 2002) a recomendam para superar certos problemas matemáticos com o uso da ligação recíproca (canônica). Como somente duas temporadas agrícolas foram analisadas, são estimados os mesmos valores de correlação para qualquer estrutura de correlação não-independente no procedimento GENMOD (Myers et al. 2002). Conseqüentemente, as comparações foram conduzidas entre os modelos de independência e não-independência de acordo com Myers et al. (2002) e Liang e Zeger (1986). A correlação foi suficientemente pequena

( $\rho = -0,20$ ) indicando que o uso da estrutura de correlação independente deveria ser mais adequado. Logo, foi usado o modelo independente, i.e. matriz de identidade, para as estimativas dos parâmetros da GEE. Liang e Zeger (1986) discutiram este ponto e afirmaram que, quando todas as correlações são menores que 0,3 em magnitude, freqüentemente o uso da estrutura de correlação independente parece ser satisfatória. O procedimento GENMOD fornece ambos os erros: do modelo base e o padrão empírico dos coeficientes de regressão. Estimativas do erro padrão empírico são comumente preferidas porque permitem a obtenção de estimativas robustas (Myers et al., 2002).

Quadro 1 - Análise geral das Equações de Estimação Generalizadas (GEE) para a produção de frutos das cultivares de oliveira: Arbequina, Picual e Frantoio, avaliadas sob condições Mediterrâneas áridas do norte do Chile, nas temporadas 2002-2003. Usou-se a matriz de independência

Parâmetro	Estimativa	SE	95% Limite de Confiança		Z	P >  Z
			Inferior	Superior		
Intercepto	2,3802	0,0565	2,2695	2,4908	42,16	< 0,0001
Arbequina (A)	0,3191	0,076	0,1702	0,468	4,2	< 0,0001
Picual (P)	0,3039	0,0898	0,1278	0,4799	3,38	0,0007
Frantoio (F)	0	0	0	0	.	.
Limarí-1 (L1)	0,4306	0,073	0,2874	0,5737	5,9	< 0,0001
Limarí-2 (L2)	-2,6903	0,2033	-3,0888	-2,2919	-13,24	< 0,0001
Choapa (C)	0	0	0	0	.	.
AxL1	-0,0137	0,0968	-0,2033	0,176	-0,14	0,8876
AxL2	1,9803	0,2336	1,5224	2,4381	8,48	< 0,0001
AxC	0	0	0	0	.	.
PxL1	-0,3344	0,1171	-0,564	-0,1048	-2,85	0,0043
PxL2	2,5549	0,2327	2,0989	3,011	10,98	< 0,0001
PxC	0	0	0	0	.	.
FxL1	0	0	0	0	.	.
FxL2	0	0	0	0	.	.
FxC	0	0	0	0	.	.

GEE  $y = e^{2,3802+0,3191A+0,3039P+0,4306L1-2,6903L2+1,9803AxL2-0,3344PxL1+2,5549PxL2}$

SE: resíduo padrão empírico.

Arbequina e Picual foram significativos e o sinal positivo nas estimativas dos parâmetros GEE sugere um efeito positivo (incremento) sobre a produção de frutos. O local de Limarí-1 também teve um positivo efeito e foi estatisticamente significativo. Por outro lado, o local de Limarí-2 foi significativo, porém, com um efeito negativo (redução) sobre a média total de produção.

Os componentes dos efeitos da interação foram significativos para AxL2, PxL1 e PxL2, e estão indicando a necessidade de uma análise por local. Este resultado foi confirmado por meio do teste quase-escore da análise GEE Type3, rodada no procedimento GENMOD, no qual os efeitos da cultivar, local e interação cultivar local foram significativos ( $P < 0.01$ ).

As equações de estimação generalizadas (GEE) dentro dos locais são mostradas no Quadro 2. Neste contexto, as análises de contrastes são úteis para determinar diferenças específicas entre as cultivares. As cultivares de oliveira responderam diferentemente dependendo dos locais (Quadro 2). Arbequina e Picual responderam bem em Choapa, porém na Província de Limarí suas respostas foram diferentes. Tous et al. (1998) verificou diferenças significativas na produção de frutos de oliveira em um grupo de cinco variedades estabelecidas na Espanha, sendo que Arbequina e Picual tiveram respostas similares. Picual não respondeu bem em Limarí-2 e Choapa, e em Limarí-1 esta variedade produziu menos que Arbequina.

Quadro 2 - Análises de contrastes e Equações de Estimação Generalizadas (GEE) dentro dos locais para a produção de frutos das cultivares de oliveira: Arbequina, Picual e Frantoio, avaliadas sob condições Mediterrâneas áridas do norte do Chile, nas temporadas 2002-2003

Item	GL	Limarí-1 **		Limarí-2 **		Choapa **	
		$\chi^2$	$p > \chi^2$	$\chi^2$	$p > \chi^2$	$\chi^2$	$p > \chi^2$
<b>Contraste</b>							
Arbequina - Picual	1	19,58	< 0,0001	12,01	0,0005	0,03	0,8600
Arbequina - Frantoio	1	20,83	< 0,0001	27,1	< 0,0001	14,58	0,0001
Picual - Frantoio	1	0,13	0,7227	41,15	< 0,0001	9,14	0,0025

Produção de frutos média (kg árvore <sup>-1</sup> ).			
Arbequina (A)	22,06 a	6,81 b	14,37 a
Picual (P)	15,46 b	12,29 a	14,14 a
Frantoio (F)	16,12 b	0,23 c	10,31 b
GEE	$y = e^{2,8108 + 0,3054 A}$	$y = e^{-0,3102 + 2,2993 A + 2,8588 P}$	$y = e^{2,3802 + 0,3191 A + 0,3039 P}$
WCM	$\begin{bmatrix} 1 & -0,43 \\ -0,43 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0,00 \\ 0,00 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0,74 \\ -0,74 & 1 \end{bmatrix}$

$\chi^2$  Qui-quadrado; p: probabilidade; GL: Graus de liberdade; \*\* Cultivares evidenciaram diferenças significativas dentro dos locais de acordo com o teste quase-escore ( $P < 0,01$ ). WCM: Matriz de correlação de trabalho.

Ben-Rouina et al. (2002) afirmaram que o crescimento e a produtividade das árvores de oliveira são influenciados por muitos fatores relacionados às condições climáticas, solo, disponibilidade de água e nutrientes. Astorga e Mora (2005) evidenciaram diferenças significativas entre outras seis cultivares de oliveira em relação à produtividade e produção de frutos acumulada, considerando ambientes análogos ao presente experimento. León et al. (2004) estudaram, durante três anos, a variabilidade genética de plântulas provenientes de cruzamentos entre as cultivares de Arbequina, Frantoio e Picual, que evidenciaram uma ampla variabilidade em algumas características de produção. Eles também estudaram coeficientes de correlação entre os anos e evidenciaram uma alta correlação ano a ano, o qual poderia ser um indicador de confiança dos valores obtidos nos anos subseqüentes.

As matrizes de correlação de trabalho (*Working correlation matrices*) dependeram do local, o qual evidenciou a mais alta correlação em Choapa. Em Limarí-1 a correlação foi moderada e em Limarí-2 foi a mais baixa (zero). Pan e Connett (2002) discutiram sobre a importância da escolha da matriz de correlação de trabalho, a qual poderia influenciar a eficiência das estimativas, e em geral, será mais eficiente se é utilizada uma matriz de correlação que é mais próxima à verdadeira correlação. Myers et al. (2002) mostrou como uma escolha errônea da estrutura de correlação produz uma grande discrepância entre as estruturas de (co)variância do modelo base e empíricas.

### 3.2. Precocidade

Houve diferenças significativas na precocidade entre as cultivares ( $P < 0,05$ ; Quadro 3). A precocidade (medida como a presença/ausência de um determinado nível de produção aos 3 anos; uma característica binária) foi também influenciada significativamente pelo ambiente e a interação cultivar x local ( $P < 0,01$ ). Picual foi mais precoce que Arbequina e Frantoio na Província de Limarí (locais de Limarí-1 e Limarí-2), porém esta variedade evidenciou carência de precocidade no local da Província de Choapa (Quadro 4). Arbequina e Frantoio responderam identicamente nos três locais. Essas variedades evidenciaram também carência de precocidade em Limarí-2. A precocidade pode ser definida como a habilidade de uma cultivar específica para começar a produzir a uma idade precoce, e é uma característica altamente desejável em árvores de oliveira (Pritsa et al., 2003). É importante considerar que geralmente as árvores de oliveira são cultivadas em áreas de alto valor econômico. Considerando o alto custo de estabelecimento de um horto de oliveiras, seria interessante o cultivo de uma variedade que não apresente um período improdutivo tão extenso (Pritsa et al., 2003; Tous et al., 1998).

Quadro 3 - Valores de deviance e Qui-quadrado para sobrevivência e precocidade das cultivares de oliveira: Arbequina, Picual e Frantoio, avaliadas sob condições Mediterrâneas áridas do norte do Chile, nas temporadas 2002

Fonte de Variação	Sobrevivência			Precocidade	
	GL	Deviance	Qui-quadrado	Deviance	Qui-quadrado
Intercepto		389,3	--	473,0	--
Cultivar	2	354,4	34,85 **	465,4	7,59 *
Local	2	323,3	31,10 **	335,5	129,82 **
Cultivar x Local	4	297,0	26,31 **	299,4	36,16 **

GL: Graus de liberdade do delineamento experimental. \*\*  $P < 0,01$ . \*  $P < 0,05$ .

Se a necessidade de um pequeno agricultor é recuperar os custos de estabelecimento tão logo quanto possível, Picual deveria ser considerado na Província de Limarí. Embora, a seleção de Picual para plantações diretas em

Limarí-1 implique em perda de produtividade por árvore, devido ao fato de Arbequina ter evidenciado a maior produtividade média. Page (2004) afirmou que o melhoramento da precocidade em variedades comerciais beneficiaria aos produtores da cultura, porque isto conduziria a um mais rápido retorno do investimento e, durante um período fixo, maior produtividade acumulada que cultivares menos precoces com similar produtividade. Portanto, uma análise econômica deveria ser conduzida para avaliar a apropriada decisão.

Quadro 4 - Precocidade média para as cultivares de oliveira: Arbequina, Picual e Frantoio, avaliadas sob condições Mediterrâneas áridas do norte do Chile, na temporada 2002, e análise de contrastes dentro de cada local para o efeito da cultivar

Contraste	GL	Limarí-1 **		Limarí-2 **		Choapa **	
		$\chi^2$	p > $\chi^2$	$\chi^2$	p > $\chi^2$	$\chi^2$	p > $\chi^2$
Arbequina – Picual	1	7,43	0,0064	9,48	0,0021	10,1	0,0015
Arbequina – Frantoio	1	0,11	0,7365	2,13	0,1443	0,03	0,8517
Picual – Frantoio	1	6,16	0,0131	23,38	< 0,0001	9,08	0,0026
Precocidade média							
Arbequina		0,609 b		0,048 b		0,178 a	
Picual		0,854 a		0,400 a		0,000 b	
Frantoio		0,642 b		0,000 b		0,163 a	

$\chi^2$  Qui-quadrado; p: probabilidade; GL: Graus de liberdade; \*\* Cultivares evidenciaram diferenças significativas dentro dos locais de acordo com o teste escore (P < 0,01).

### 3.3. Sobrevivência

A sobrevivência, medida após 4 temporadas agrícolas, diferiu significativamente entre as cultivares (P < 0,01; Quadro 3). Os efeitos devido à interação cultivar local e do local foram significativos (P < 0,01). A média da sobrevivência variou de 53,8% a 100%, de 55,8% a 82,8% e de 89,6% a 100% em Arbequina, Picual e Frantoio, respectivamente (Quadro 5). Frantoio apresentou alta sobrevivência nos três locais, porém sua produção foi baixa, como já discutido nas análises prévias. Arbequina teve também alta sobrevivência em Limarí-1 e Choapa, indicando uma possibilidade para obter uma cultivar com

ambas as características, porém a interação cultivar local poderia ser o fator limitante para atingir este objetivo.

Quadro 5 - Sobrevivência média para as cultivares de oliveira: Arbequina, Picual e Frantoio, depois de quatro temporadas agrícolas, e análise de contrastes dentro dos locais para o efeito da cultivar

Contraste	GL	Limarí-1 **		Limarí-2 **		Choapa **	
		$\chi^2$	p > $\chi^2$	$\chi^2$	p > $\chi^2$	$\chi^2$	p > $\chi^2$
Arbequina – Picual	1	12,52	0,0004	0,03	0,8552	6,78	0,0092
Arbequina – Frantoio	1	1,24	0,2653	30,44	< 0,0001	0,55	0,4580
Picual – Frantoio	1	8,65	0,0033	31,5	< 0,0001	3,59	0,0581
Sobrevivência média (%)							
Arbequina		100,0 a		53,8 b		93,8 a	
Picual		82,8 b		55,8 b		75,0 b	
Frantoio		98,1 a		100,0 a		89,6 a	

$\chi^2$  Qui-quadrado; p: probabilidade; GL: Graus de liberdade; \*\* Cultivares evidenciaram diferenças significativas dentro dos locais de acordo com o teste escore (P<0,01).

A sobrevivência é característica importante para a produção de oliveiras manejadas intensivamente. A sobrevivência não tem sido incluída como uma característica em programas de seleção de árvores de oliveira (e.g. Tous et al., 1998), embora seu impacto sobre a produtividade pode ser significativo. A sobrevivência das árvores é uma característica importante para os sistemas agro-florestais com déficit hídrico (Abdelkdair e Schultz, 2005; Abebe, 1994). Estudos prévios (Rawat e Banerjee, 1998) mencionaram que as regiões semi-áridas e litorais são afetadas pela salinização; um fato que pode ter um impacto significativo sobre a produtividade e/ou taxas de sobrevivência total.

#### **4. CONCLUSÕES**

As seleções de cultivares de oliveira têm potencial nos sistemas agrícolas manejados intensivamente na zona árida do norte do Chile. Sua efetividade dependerá da seleção de genótipos que respondem bem sobre um local em particular.

Os resultados demonstraram a efetividade da metodologia dos modelos lineares generalizados, incluindo sua extensão para análise longitudinal: o método das Equações de Estimação Generalizadas, para atender as necessidades de avaliação precoce de cultivares de oliveira, em uma situação onde o pressuposto de normalidade não é alcançado.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ABEBE, T. Growth performance of some multipurpose trees and shrubs in the semi-arid areas of Southern Ethiopia. **Agroforestry Systems**, 26:237-248, 1994.

ABDELKDAIR, A.; SCHULTZ, R.C. Water harvesting in a 'runoff-catchment' agroforestry system in the dry lands of Ethiopia. **Agroforestry Systems**, 63:291-298, 2005.

ASTORGA, M.; MORA, F. Componentes de varianza e interacción variedad-sitio del vigor, producción y productividad de *Olea europaea*, en Chile. **Cerne**, 11:25-33, 2005.

BEN-ROUINA, B.; TRIGUI, A.; BOUKHRIS, M. Effect of tree growth and nutrients status of "Chemlali de Sfax" olive trees and their productivity. **Acta Hort.**, 586:349-352, 2002.

CRUZ, C.D. **Programa Genes - Estatística Experimental e Matrizes**. Viçosa: Universidade Federal de Viçosa, 2006. 285p.

BRENES, M.; GARCÍA, A.; RIOS, J.J.; GARCÍA, P.; GARRIDO, A. Use of 1-acetoxypinoresinol to authenticate Picual olive oils. **International Journal of Food Science & Technology**, 37:615-625, 2002.

DEMETRIO, C.G.B. **Modelos lineares generalizados em experimentação agrônômica**. Piracicaba: ESALQ, 2001. 113p.

DOBSON, A.J. **An introduction to generalized linear models**. London: Chapman and Hall, 2001. 225p.

DOS SANTOS, A.L.; MORA, F. Experimental analysis of flocculant treatments of organic waste from swine production. **Ciencia e Investigación Agraria**, 34:47-54, 2007.

FREUND, J.E. **Mathematical statistics**. New Jersey: Prentice-Hall, 1992. 658p.

GOLDSCHMIDT, E.E. Regolazione dell'alternanza di produzione negli alberi da frutto. **Italus Hortus**, 12:11-17, 2005.

- HORTON, N.J.; LIPSITZ, S.R. Review of software to fit generalized estimating equation regression models. **The American Statistician**, 53:160-169, 1999.
- LEÓN, L.; RALLO, L.; DEL RÍO, C.; MARTÍN, L.M. Variability and early selection on the seedling stage for agronomic traits in progenies from olive crosses. **Plant Breeding**, 123:73-78, 2004.
- LEVIN, A.G.; LAVEE, S. The influence of girdling on flower type, number, inflorescence density, fruit set, and yields in three different olive cultivars (Barnea, Picual, and Sourì). **Australian Journal of Agricultural Research**, 56:827–831, 2005.
- LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models, **Biometrika**, 73:13-22, 1986.
- LOPES, M.S.; MENDONÇA, D.; SEFC, K.M.; GIL, F.S.; MACHADO, A.C. Genetic evidence of intra-cultivar variability within Iberian olive cultivars. **HortScience**, 39:1562-1565, 2004.
- LOUMOU, A.; GIOURGA, C. Olive groves: the life and identity of the Mediterranean. **Agriculture and Human Values**, 20:87-95, 2003.
- MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. **Generalized linear models, with applications in engineering and the sciences**. New York: John Wiley and Sons Press, 2002. 342p.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear model. **Journal of the Royal Statistical Society A**, 135:370-384, 1972.
- ORMAZÁBAL, C.S. Silvopastoral systems in arid and semiarid zones of northern Chile. **Agroforestry Systems**, 14:207-217, 1991.
- PAGE, T. **Muntries: The domestication and improvement of *Kunzea pomifera* (F.Muell)**. Kingston: RIRDC, 2004. 95p.
- PAN, W.; CONNETT, J.E. Selecting the working correlation structure in generalized estimating equations with application to the lung health study. **Statistica Sinica**, 12:475-490, 2002.

- PRITSA, T.S.; VOYIATZIS, D.G.; VOYIATZI, C.J.; SOTIRIOU, M.S. Evaluation of vegetative growth traits and their relation to time to first flowering of olive seedlings. **Australian Journal of Agricultural Research**, 54:371-376, 2003.
- RAWAT, J.S.; BANERJEE, S.P. The influence of salinity on growth, biomass production and photosynthesis of *Eucalyptus camaldulensis* Dehnh. and *Dalbergia sissoo* Roxb. seedlings. **Plant and Soil**, 205:163-169, 1998.
- RALLO, L.; TORREÑO, P.; VARGAS, A.; ALVARADO, J. Dormancy and alternate bearing in olive. **Acta Hort.**, 356:127-136, 1994.
- SAS-INSTITUTE. **Statistical analysis system user's guide**. Cary: SAS Institute, 1996. 956p.
- SILVA, G.O.; SOUZA, V.Q.; PEREIRA, A.S.; CARVALHO, F.I.F.; FRITSCHENETO, R. Early generation selection for tuber appearance affects potato yield components. **Crop Breeding and Applied Biotechnology**, 6:73-78, 2006.
- SQUEO, F.A.; TRACOL, Y.; LÓPEZ, D.; GUTIÉRREZ, J.R.; CORDOVA, A.M.; EHLERINGER, J.R. ENSO effects on primary productivity in Southern Atacama Desert. **Advances in Geosciences**, 6:273-277, 2006.
- TOUS, J.; ROMERO, A.; PLANA, J. Comportamiento agronómico y comercial de cinco variedades de olivo en Tarragona. **Investigación Agraria, Producción y Protección Vegetal**, 13: 97-109, 1998.
- SOARES, P.C.; MELO, P.G.S.; MELO, L.C.; SOARES, A.A. Genetic gain in an improvement program of irrigated rice in Minas Gerais. **Crop Breeding and Applied Biotechnology**, 5:142-148, 2005.
- WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, 61:439-447, 1974.

**CAPÍTULO IV**  
**AVALIAÇÃO DE SEIS CULTIVARES DE OLIVEIRA USANDO EQUAÇÕES DE**  
**ESTIMAÇÃO GENERALIZADAS, MODELOS MISTOS E MODELOS**  
**GENERALIZADOS**

**RESUMO**

A sobrevivência, produção precoce de frutos, crescimento vegetativo e alternância da produção foram analisados em seis cultivares de oliveira (Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino) sob condições agronômicas intensivas no sul do deserto de Atacama, Região administrativa de Coquimbo, Chile. Os cultivares foram avaliados durante quatro anos sucessivos, e evidenciaram uma alta sobrevivência (93%) confirmando o potencial para esta região árida. A produção de frutos (medida durante as temporadas 2002-2003), crescimento vegetativo (2000-2003) e a alternância da produção diferiram significativamente entre cultivares ( $p < 0,01$ ). Seleção de cultivares de oliveira em plantações manejadas intensivamente na parte sul do deserto de Atacama, depende da alocação de cultivares específicos a locais nos quais eles respondam melhor.

Palavras-chave: modelos lineares generalizados, produção de azeitonas, ambientes áridos

## ABSTRACT

Tree survival, early fruit production, vegetative growth and alternate bearing were examined in six different olive cultivars (Barnea, Biancolilla, Coratina, Empeltre, Koroneiki and Leccino) under intensive agronomic conditions in southern Atacama Desert, in the Coquimbo Region of Chile. The cultivars were evaluated over four successive years and had a high survival rate (93%) confirming their potential for these dry-lands. Fruit production (recorded over the growing seasons 2002-2003), vegetative growth (2000-2003) and alternate bearing differed significantly among cultivars ( $p < 0.01$ ). Olive selection in intensively managed planting at the southern part of the Atacama Desert depends on matching specific cultivars to sites on which they perform the best.

Key words: generalized linear models, olive production, arid environments

## 1. INTRODUÇÃO

As condições de clima mediterrâneo são consideradas vantajosas para o estabelecimento de plantações de oliveira (*Olea europaea* L.). A área compreendida na bacia do mar Mediterrâneo produz cerca de 90% da produção mundial de azeitonas. A oliveira é uma das espécies mais cultivada nessa região, e é economicamente importante para vários países de clima subtropical mediterrâneo, como Croácia, Grécia, Itália, Espanha, Portugal e Turquia (Aydinalp et al., 2004). A cultura da oliveira também está sendo considerada importante alternativa produtiva em países tais como Argentina, Austrália, Brasil, Chile, África do Sul e EUA. Esta espécie foi naturalizada em diversas regiões da América, onde é utilizada para a indústria da azeitona e do azeite.

As cultivares Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino têm muitas aplicações práticas para a indústria de azeite virgem, porém seu desempenho agrônômico é pouco documentado no sul do deserto de Atacama, no Chile. Esta zona árida tem um clima de tendência Mediterrânea e, na última década, foram empreendidos diversos projetos financiados pelos programas de fomento do governo chileno, com o intuito de atender às necessidades dos agricultores.

As árvores de oliveira são capazes de suportar períodos prolongados sem precipitações, condição ambiental comum do clima mediterrâneo com tendência árida ou semi-árida (Lavee, 1996). A zona árida do norte do Chile tem potencial agrônômico para o cultivo e desenvolvimento agroindustrial de variedades melhoradas de oliveiras. Como uma consequência, os agricultores estão começando a estabelecer plantações de oliveira manejadas intensivamente, e as instituições de pesquisa estão conduzindo programas de melhoramento.

A avaliação de cultivares de oliveira em condições ambientais distintas, ao longo do tempo, é considerada parte de uma estratégia empreendida por instituições de pesquisa regionais, com o intuito de examinar as respostas genóticas em uma diversidade de condições ambientais. A precocidade da produção, visando um retorno mais rápido do investimento, a ausência da alternância de produção de frutos (*alternate bearing*) e a elevada produção por

árvore são algumas das características freqüentemente procuradas nos programas de melhoramento de oliveira estabelecidas sob condições áridas.

Os dados de medidas repetidas aplicados à mesma unidade experimental são freqüentemente desbalanceados e fortemente correlacionados (Myers et al., 2002). A alternância da produção das oliveiras pode ser um adequado exemplo de dados com respostas correlacionadas. A alternância de produção ocorre quando as cultivares evidenciam rendimentos altos em um ano particular, seguido por rendimentos baixos no ano seguinte (Pearce e Dobersek-Urbanc, 1967). Assim, a alternância de produção pode ser um fator limitante na maximização dos rendimentos de produção de frutos.

É crescente o interesse dos pesquisadores em analisar dados que provêm de estudos longitudinais (medidas repetidas) e havendo correlação entre observações de uma mesma unidade experimental (Horton e Lipsitz, 1999). A produção de azeitonas e o crescimento vegetativo medidos sobre a mesma árvore (*subject*) no tempo são dois exemplos de estudos longitudinais.

O presente estudo tem como objetivo principal examinar o crescimento vegetativo e a produção de frutos durante quatro e dois anos agrícolas, respectivamente, em seis diferentes cultivares de oliveira: Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino. Também foram examinados a sobrevivência das árvores (monitorada durante quatro períodos agrícolas) e o índice de alternância de produção (que utiliza séries de dados de produção de frutos). O entendimento acerca do desempenho agrônômico precoce destas cultivares de oliveira será crucial para maximizar a produtividade de frutos sob condições agrônômicas intensivas no sul do Deserto de Atacama.

## 2. MATERIAL E MÉTODOS

### 2.1. Localização do estudo

Seis diferentes variedades de oliveira (Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino) foram plantadas manualmente em três locais da região administrativa de Coquimbo, Chile: Illapel, Ovalle e Monte Patria (Quadro 1).

Quadro 1 - Locais da região Mediterrânea árida do Chile, considerados na avaliação de seis cultivares de oliveira

Locais	Província	Latitude (S)	Longitude (O)	Altitude (m)
Ovalle	Limarí	30° 35'	71° 28'	200
Illapel	Choapa	31° 50'	70° 51'	580
Monte Patria	Limarí	30° 45'	70° 44'	820

Esta região apresenta um clima mediterrâneo árido, em que as precipitações estão concentradas principalmente durante os meses de inverno (junho a setembro), com precipitação média anual que varia de 132 mm a 208 mm (Quadro 2). O local de Ovalle (localizado na área rural de Cerrillos) apresenta uma temperatura média relativamente mais baixa, devido à influência da neblina do oceano Pacífico, que está diretamente associada à corrente fria de Humboldt (Ormazábal, 1991). Illapel possui uma temperatura média relativamente mais baixa que Monte Patria devido a sua latitude mais elevada.

Quadro 2 - Dados climáticos das três localidades incluídas neste estudo, correspondentes à região Mediterrânea árida do Chile

Localidades	Horas-frio (0-12,5°C)	Umidade relativa (%)	Evaporação (mm ano <sup>-1</sup> )	Precipitação (mm ano <sup>-1</sup> )	Temperatura média (°C)
Ovalle	2194	69,92	1702	132	14,61
Illapel	2029	59,70	1807	208	15,40
Monte Patria	1918	53,11	2050	152	15,99

Os ensaios foram estabelecidos na primavera de 1999. Em cada experimento de campo foi estabelecido um sistema de irrigação, em que as exigências da água da cultura foram estimadas de acordo com Orgaz e Fereres (1999) e Doorenbos e Pruitt (1977), baseadas na reposição da evaporação de bandeja.

## 2.2. Delineamento experimental e medições

O delineamento experimental foi inteiramente ao acaso com seis tratamentos (cultivares), com seis árvores por parcela e a um espaçamento entre plantas de 8 x 4 m.

A sobrevivência total das árvores foi registrada após quatro anos agrícolas, na primavera de 2003. Esta característica foi medida como uma resposta binária. A taxa de crescimento absoluta do perímetro do tronco e a produção de frutos foram registradas nas temporadas agrícolas 2000-2003 e 2002-2003, respectivamente. As séries de dados de produção de frutos foram usadas para examinar o índice de alternância da produção (ABI), o qual foi calculado usando a seguinte expressão proposta por Pearce e Dobersek-Urbanc (1967):

$$ABI = \left( \frac{1}{(n-1)} \right) \left\{ \frac{|(a_2 - a_1)|}{a_1 + a_2} + \frac{|(a_3 - a_2)|}{a_2 + a_3} \dots + \frac{|(a_n - a_{n-1})|}{a_{n-1} + a_n} \right\}$$

em que  $n$  corresponde ao número de anos da avaliação;  $a_1, a_2, a_3, \dots, a_{n-1}, a_n$  simbolizam as produções de frutos por árvore ( $\text{kg árvore}^{-1}$ ) medidas no período respectivo da análise. Neste caso, a análise corresponde a dois períodos agrícolas (2002-2003).

## 2.3. Análise estatística

A sobrevivência total das árvores foi analisada usando a metodologia dos modelos lineares generalizados (GLM) (McCullagh e Nelder, 1989) porque a variável resposta foi considerada binária, possuindo distribuição de Bernoulli. Esta análise foi realizada usando o procedimento GENMOD do programa SAS (SAS-

Institute, 1996). As comparações de médias foram realizadas usando os contrastes ortogonais ( $\alpha = 0,05$ ).

A produção precoce de frutos também foi avaliada usando os modelos lineares generalizados devido à ausência de normalidade. Como foram considerados dois períodos conjuntamente (estudo longitudinal), a análise de dados agronômicos foi realizada usando a metodologia de Equações de Estimação Generalizadas (Liang e Zeger, 1986), uma extensão dos GLM para ajustar um modelo linear generalizado a dados correlacionados (Wedderburn, 1974). As suposições prévias de normalidade para a característica produção de frutos foram conferidas visualmente (gráficos de densidade e probabilidade), usando os seguintes testes estatísticos: Shapiro-Wilk, Qui-quadrado e Lilliefors (Dos Santos e Mora, 2007). O gráfico de probabilidade e os testes estatísticos foram obtidos no SAS-INSIGHT e no procedimento UNIVARIATE do SAS (SAS-Institute, 1996). A distribuição aproximada desta variável foi julgada visualmente em um gráfico da variável resposta obtido no programa R, de acordo com Freund (1992).

A suficiência estatística do modelo linear generalizado foi verificada por meio do procedimento GENMOD do SAS (SAS-Institute, 1996). Foi realizada uma comparação entre as matrizes de correlação de trabalho (WCM: *working correlation matrix*), se independente ou dependente, de acordo com Myers et al. (2002) e Liang e Zeger (1986).

A taxa de crescimento absoluta do perímetro do tronco (variável aleatória com distribuição normal) medido de 2000 a 2003 foi analisada usando o procedimento MIXED do SAS (SAS-Institute, 1996), considerando ao ano como a medida repetida no tempo. Sete estruturas diferentes de correlação foram checadas para a matriz de (co) variância residual: Componentes de Variância (VC), Toeplitz (TOEP2 e TOEP3), Auto-regressiva (AR1), Simetria composta (CS), Huynh-Feldt (HF) e Não Estruturada (UN), e foram testadas usando os critérios de informação: Bayesiano de Schwarz (BIC) e Akaike (AIC). As diferenças entre as médias foram testadas usando o teste de contrastes ( $\alpha = 0,05$ ).

O índice de alternância da produção (variável aleatória com distribuição normal) foi analisado usando o procedimento GLM de SAS (SAS-Institute, 1996) e as comparações de médias foram realizadas usando o teste de Tukey ( $\alpha = 0,05$ ).

### 3. RESULTADOS E DISCUSSÃO

A sobrevivência não diferiu significativamente entre as cultivares de oliveira após quatro anos de crescimento ( $P > 0,05$ ; Quadro 3) e atingiu um valor de 93%, confirmando o potencial de estas variedades para o estabelecimento de plantações intensivamente manejadas no sul do Deserto de Atacama. A interação cultivar local e os efeitos do local também não foram significativos ( $P > 0,05$ ).

A produção de frutos diferiu significativamente entre as cultivares ( $P < 0,01$ ; Quadro 3). Os efeitos do local e da interação cultivar local também foram significativos para esta característica ( $P < 0,01$ ). As análises conduzidas no programa R confirmaram a falta de simetria na distribuição dos dados; a moda, calculada pelo método da estimação de densidade de Kernel, no SAS- Insight, foi de 3 kg árvore<sup>-1</sup>, e foi muito menor do que a média geral (14 kg árvore<sup>-1</sup>). No presente estudo o quadrado da média foi proporcional à variância ( $\sigma^2 = 189,4$ ). Assim, na análise estatística foi usado um modelo linear generalizado com distribuição Gama (Freund, 1992) e função de ligação logarítmica. A distribuição Gama tem aplicações potenciais em problemas de regressão em que a variável resposta é contínua e a variância não é constante, porém é proporcional ao quadrado da média (Myers et al., 2002).

Quadro 3 - Significância estatística para todas as características agrônômicas avaliadas em seis diferentes cultivares de oliveira, no sul do deserto de Atacama, região administrativa de Coquimbo, Chile

Fonte de variação	GL	S	FP	AGP	AB
		$\chi^2$	$\chi^2$	F	F
Cultivar	5	3,36	44,40 **	7,74 **	16,58 **
Local	2	0,31	40,70 **	64,70 **	2,74
Cultivar x Local	10	6,60	37,35 **	3,23 **	7,00 **

\*\* Significativo a  $P < 0,01$ . GL: grau de liberdade. S: sobrevivência aos quatro anos de crescimento. FP: produção de frutos (períodos de crescimento 2002-2003). AGP: taxa de crescimento absoluto do perímetro do tronco (2000-2003). AB: alternância da produção (2002-2003).  $\chi^2$ : Qui-quadrado.

Na abordagem das Equações de Estimação Generalizadas, a matriz WCM mostrou uma correlação entre períodos acima do valor |0,3| (o valor

verdadeiro, rho, da correlação foi -0,32) considerada correlação apropriada para a estrutura de dependência (Myers et al., 2002; Liang e Zeger, 1986).

Foi considerada uma análise por local, devido ao fato de que o efeito da interação cultivar local foi significativo. Esta análise é apresentada no Quadro 4. As matrizes WCM variaram com o local estudado, evidenciando alta correlação em Ovalle ( $\rho = -0,88$ ), moderada em Illapel ( $\rho = -0,45$ ) e correlação nula em Monte Patria (matriz de correlação independente).

Os parâmetros do modelo (usando a metodologia das equações de estimação generalizada) são também apresentados no Quadro 4. As equações generalizadas para produção de frutos são as seguintes:

$$P_{Illapel} = EXP(2,4653 + 0,5641 KO + 0,649 LE + 0,9359 BA + 1,038 CO - 2,1652 BI) \}$$

$$P_{Ovalle} = EXP(2,3918 + 0,3295 KO + 0,8177 LE + 0,5752 BA + 0,4276 CO)$$

$$P_{Monte Patria} = EXP(1,0761 + 1,7027 KO + 1,4032 BA + 0,9186 CO)$$

em que P é a média da produção de frutos (kg árvore<sup>-1</sup>) em cada local; KO, LE, BA, CO e BI são as cultivares Koroneiki, Leccino, Barnea, Coratina e Biancolilla, respectivamente. Os modelos têm forma exponencial devido à função de ligação logarítmica.

Neste caso, os modelos das equações de estimação generalizadas representam a contribuição de cada parâmetro (ou neste caso cultivar) sobre a média da produção de frutos. Os parâmetros com sinais positivos indicam que uma variedade específica incrementa a média da produção de frutos, e sinais negativos indicam que aquela cultivar contribui para reduzir a média da produção de frutos em um local específico. Por exemplo, em Illapel; Koroneiki, Leccino, Barnea e Coratina têm um efeito positivo e significativo sobre a produção média de frutos, baseado na estimativa dos parâmetros das equações de estimação generalizadas. Ao contrário, Biancolilla tem um efeito significativo, porém negativo, reduzindo a média da produção de frutos no mesmo local. A mesma análise pode ser feita para cada local.

O efeito da interação cultivar local também pode ser avaliado com os modelos das equações generalizadas, ajustado em cada local. Leccino respondeu relativamente bem no local de Ovalle, porém em Illapel e Monte Patria, Coratina e

Koroneiki, respectivamente, foram as cultivares que melhor responderam, e Leccino não foi significativo sobre a média da produção de frutos em Monte Patria.

Quadro 4 - Análise das equações de estimação generalizadas para a produção de frutos (medida nos anos de crescimento 2002-2003) em seis cultivares de oliveira (Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino) avaliada em três locais do norte do Chile

Local	Parâmetro	Estimativa	95% Limites de confiança		Z	P >  Z
			Inferior	Superior		
Illapel	Intercepto	2,465	2,083	2,848	12,63	< 0,0001
	Koroneiki	0,564	0,120	1,018	2,49	0,0128
	Leccino	0,649	0,169	1,129	2,65	0,008
	Barnea	0,936	0,508	1,364	4,29	< 0,0001
	Coratina	1,038	0,626	1,449	4,94	< 0,0001
	Biancolilla	-2,165	-3,188	-1,142	-4,15	< 0,0001
	Empeltre	0	0	0	.	.
	$\rho$	-0,45				
Ovalle	Intercepto	2,392	2,327	2,456	72,63	< 0,0001
	Koroneiki	0,329	0,085	0,574	2,64	0,0083
	Leccino	0,818	0,648	0,987	9,47	< 0,0001
	Barnea	0,575	0,476	0,675	11,32	< 0,0001
	Coratina	0,428	0,297	0,558	6,43	< 0,0001
	Biancolilla	-0,003	-0,399	0,393	-0,02	0,988
	Empeltre	0	0	0	.	.
	$\rho$	-0,88				
Monte Patria	Intercepto	1,076	0,405	1,747	3,14	0,0017
	Koroneiki	1,703	1,021	2,384	4,9	< 0,0001
	Leccino	-0,461	-1,866	0,944	-0,64	0,5202
	Barnea	1,403	0,643	2,164	3,62	0,0003
	Coratina	0,919	0,071	1,766	2,12	0,0336
	Biancolilla	-0,029	-1,033	0,975	-0,06	0,9551
	Empeltre	0	0	0	.	.
	$\rho$	0				

$\rho$ : valores Rho da matriz de correlação de trabalho (WCM).

Embora Koroneiki seja uma cultivar que pode evidenciar um maior rendimento, como no local de Monte Patria, Stefanoudaki et al. (2001) afirmaram que a qualidade do azeite desta cultivar pode ser influenciada pela disponibilidade de água. Porém, a dependência da disponibilidade de água na qualidade do azeite não é uma característica particular de Koroneiki. Trabalhando com a cultivar Ascolana, Proietti e Antognozzi (1996) afirmaram que uma adequada umidade do solo é também fator importante para obter um tamanho satisfatório do fruto, o fator fortemente correlacionado com os valores comerciais. Em condições de clima mediterrâneo, nas quais as precipitações de verão são moderadas, como nas regiões áridas e semi-áridas, a irrigação pode melhorar o valor comercial do fruto de oliveira.

Houve diferenças significativas do crescimento entre as cultivares de oliveira ( $P < 0,01$ ; Quadro 3). Esta característica também foi influenciada significativamente pela interação cultivar local e pelo efeito do ambiente ( $P < 0,01$ ). A estrutura de covariâncias não estruturada (UN) foi aquela que minimizou os valores dos critérios de informação considerados no estudo (Bayesiano de Schwarz, BIC, e Akaike, AIC).

Koroneiki mostrou baixos valores de taxas de crescimento nos três locais (Quadro 5), os quais parecem estar correlacionados inversamente com os valores de produção de frutos. Coratina evidenciou um efeito positivo sobre a média da produção de frutos em todos os locais e respondeu bem para a taxa de crescimento. Coratina é uma variedade amplamente cultivada no sul da Itália a qual tem mostrado excelente rendimento. Também foram relatadas propriedades antioxidantes (Aldini et al., 2006), embora Coratina pudesse ser mais sensível às baixas temperaturas quando comparado com outras cultivares de oliveira (Fiorino e Mancuso, 2000).

Leccino evidenciou baixo desempenho para o perímetro médio do tronco em todos os locais, porém em Ovalle, esta variedade mostrou o maior efeito sobre a média da produção de frutos. Leccino é considerado uma cultivar promissora para ambientes áridos devido a suas características agrônomicas (Bartolini et al., 2002), qualidade do azeite (Gucci et al., 2004), e sua relativa tolerância à alta salinidade (Demiral, 2005). Estes fatores são muito importantes e requeridos para o estabelecimento de plantações de oliveiras intensivamente manejadas no sul do Deserto de Atacama.

Quadro 5 - Valores médios da taxa de crescimento absoluto do perímetro do tronco para as cultivares de oliveira: Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino, as quais foram avaliadas sob condições Mediterrâneas áridas no norte do Chile, durante os anos 2000-2003

Cultivar	Taxa de crescimento do perímetro do tronco (cm)		
	Illapel	Ovalle	Monte Patria
Koroneiki	11,53 bc	9,13 bc	9,09 c
Leccino	12,31 bc	9,36 b	10,34 bc
Barnea	11,35 c	9,28 bc	11,24 ab
Coratina	13,13 ab	10,92 a	12,11 a
Biancolilla	14,03 a	8,79 c	10,59 ab
Empeltre	13,03 ab	10,21 b	11,07 ab

<sup>1</sup>Médias seguidas pelas mesmas letras nas colunas não são significativamente diferentes a  $P = 0,05$ , segundo o teste de contrastes.

A alternância da produção de frutos, medida no período 2002-2003, evidenciou diferenças estatísticas significativas entre as cultivares de oliveira ( $P < 0,01$ ). A interação cultivar local foi significativa ( $P < 0,01$ ), porém o efeito do local não ( $P > 0,05$ ). O índice médio da alternância de produção variou de 0,17 a 1 (Quadro 6). A alternância de produção é ainda problema sério no cultivo de numerosas espécies frutíferas, incluindo as árvores de oliveira (Levin e Lavee, 2005).

Na alternância de produção foi considerada uma análise de separações médias para cada local, individualmente, devido ao fato de que o efeito da interação cultivar local foi significativo (Quadro 6). Coratina, Leccino e Koroneike tiveram os valores mais baixos do índice da alternância em Illapel, Ovalle e Monte Patria, respectivamente. As estimativas do índice de alternância foram 0,18, 0,17 e 0,23 para Coratina, Leccino e Koroneike, respectivamente. Foi encontrada uma forte associação entre produção de frutos e o índice de alternância. As três cultivares com menor alternância de produção, também evidenciaram os maiores efeitos na produção de frutos. Ou seja, se são escolhidas aquelas cultivares que minimizam a alternância de produção, indiretamente seriam selecionadas aquelas cultivares que maximizam a produção média de frutos em cada local. Rallo et al. (1994) afirmaram que várias cultivares de oliveira podem ter diferentes graus de alternância de produção, devido principalmente a uma inibição da indução floral.

Esta alternância não ocorre necessariamente em todas as cultivares simultaneamente, e às vezes nas árvores individuais dentro de uma mesma cultivar (Tous et al., 1998).

Quadro 6 - Índice da alternância de produção para as seis cultivares de oliveira (Barnea, Biancolilla, Coratina, Empeltre, Koroneiki e Leccino) avaliada em três locais do sul do Deserto de Atacama, Chile

Cultivar	Índice de alternância de produção <sup>1</sup>		
	Illapel	Ovalle	Monte Patria
Barnea	0,48 bc	0,45 bc	0,62 b
Biancoli	1,00 a	1,00 a	0,81 ab
Coratina	0,18 c	0,47 b	1,00 a
Empeltre	1,00 a	1,00 a	1,00 a
Koroneike	0,51 bc	0,70 a	0,23 c
Leccino	0,56 ab	0,17 c	0,79 ab
Média geral	0,62	0,63	0,74

<sup>1</sup>Médias seguidas pelas mesmas letras nas colunas não são significativamente diferentes a P = 0,05, segundo o teste de Tukey.

#### 4. CONCLUSÕES

A mortalidade das árvores de oliveira não foi um problema no Sul do Deserto de Atacama; a taxa total de sobrevivência (acima de 90%) confirmou o potencial das cultivares para produção de azeitonas no norte do Chile.

A alternância de produção é crucial para melhorar a produtividade em oliveiras. Entretanto, é importante também identificar aquelas cultivares que apresentaram conjuntamente boa produção de frutos e baixa alternância de produção. Coratina, Leccino e Koroneike foram as cultivares que minimizaram a alternância de produção e maximizaram a produção de frutos, em Illapel, Ovalle e Monte Patria, respectivamente.

O aumento da produtividade em plantações de oliveira manejadas intensivamente, no sul do deserto de Atacama, dependerá da seleção de cultivares que respondem bem em um local particular.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ALDINI, G.; PICCOLI, A.; BERETTA, G.; MORAZZONI, P.; RIVA, A.; MARINELLO, C.; MAFFEI-FACINO, R. Antioxidant activity of polyphenols from solid olive residues of c.v. Coratina, **Fitoterapia**, 77:121-8, 2006.

AYDINALP, C.; CRESSER, M.S.; MCCLEAN, C. Characterization of some important agricultural soils under olive trees. **J. Cent. Eur. Agric.**, 5:101-108, 2004.

BARTOLINI, S.; SAPONARI, M.; GUERRIERO, R.; LORETI, F. Two new clones of cultivar 'Leccino'. **Acta Hort.**, 586:225-228, 2002.

DEMIRAL, M.A. Comparative response of two olive (*Olea europaea* L.) cultivars to salinity. **Turkish Journal of Agriculture and Forestry**, 29:267-274, 2005.

DOORENBOS, J.; PRUITT, W.O. **Las necesidades de agua de los cultivos, estudio de Riego y Drenaje**. Roma: FAO, 1977. 194p.

DOS SANTOS, A.L.; MORA, F. Experimental analysis of flocculant treatments of organic waste from swine production. **Ciencia e Investigación Agraria**, 34:47-54, 2007.

FIORINO, P.; MANCUSO, S. Differential thermal analysis, supercooling and cell viability in organs of *Olea europaea* at subzero temperatures. **Adv. Hort. Sci.**, 14:23-27, 2000.

FREUND, J.E. **Mathematical statistics**. New Jersey: Prentice-Hall, 1992. 658 p.

GUCCI, R.; SERVILI, M.; ESPOSTO, S.; SELVAGGINI, R. Oil quality of olive cv. 'Leccino' grown under irrigated or dry-farmed conditions. **Acta Hort.**, 664:297-302, 2004.

HORTON, N.J.; LIPSITZ S.R. Review of software to fit generalized estimating equation regression models. **The American Statistician**, 53:160-169, 1999.

LAVEE, S. **Enciclopedia Mundial del Olivo: Biología y fisiología del olivo**. Madrid: Consejo Oleícola Internacional, 1996. 479p.

- LEVIN, A.G.; LAVEE, S. The influence of girdling on flower type, number, inflorescence density, fruit set, and yields in three different olive cultivars (Barnea, Picual, and Souri). **Aust. J. Agr. Res.**, 56:827-831, 2005.
- LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. **Biometrika**, 73:13-22, 1986.
- MCCULLAGH, P.; NELDER, J.A. **Generalized linear models**. New York: Chapman and Hall/CRC, 1989. 511p.
- MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. **Generalized linear models, with applications in engineering and the sciences**. New York: John Wiley and Sons Press, 2002. 342p.
- ORGAZ, F.; FERERES, E. Riego. In: BARRANCO, N.; FERNÁNDEZ-ESCOBAR, D.; RALLO, R.L. (eds.). **El Cultivo del Olivo**. Madrid: Junta de Andalucía y Ediciones Mundi-Prensa, 1999. p. 267-288.
- ORMAZÁBAL, C.S. Silvopastoral systems in arid and semiarid zones of northern Chile. **Agroforestry Systems**, 14:207-217, 1991.
- PEARCE, S.C.; DOBERSEK-URBANC, S. The measurements of irregularity in growth and cropping. **J. Hort. Sc.**, 42:295-305, 1967.
- PROIETTI, P.; ANTOGNOZZI, E. Effect of irrigation on fruit quality of table olives (*Olea europaea*), cultivar 'Ascolana tenera'. **New. Zeal. J. Crop Hort. Science**, 24:175-181, 1996.
- RALLO, L.; TORREÑO, P.; VARGAS, A.; ALVARADO, J. Dormancy and alternate bearing in olive. **Acta Hortic.**, 356:127-136, 1994.
- SAS-INSTITUTE. **Statistical analysis system user's guide**. Cary: SAS Institute, 1996. 956p.
- STEFANOUDAKI, E.; CHARTZOULAKIS, K.; KOUTSAFTAKIS, A.; KOTSIFAKI, F. Effect of drought stress on qualitative characteristics of olive oil of cv Koroneiki. **Grasas y Aceites**, 52:202-206, 2001.

TOUS, J.; ROMERO, A.; PLANA, J. Comportamiento agronómico y comercial de cinco variedades de olivo en Tarragona. **Inv. Agr. Prod. Prot. Veg.**, 13:97-109, 1998.

WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, 61:439-447, 1974.

**CAPÍTULO V**  
**VARIAÇÃO DE PROCEDÊNCIAS DE *Acacia saligna* NA ZONA ÁRIDA DO**  
**CHILE, USANDO UMA ABORDAGEM BAYESIANA**

**RESUMO**

As espécies de acácia possuem apreciadas características agronômicas, tais como tolerância a condições ambientais desfavoráveis. O objetivo do presente estudo foi analisar a variabilidade no diâmetro da copa, altura, diâmetro basal do caule e sobrevivência em quatorze populações naturais de acácia azul (*Acacia saligna*) em dois locais (Cuz-Cuz e Tangué) no sul do deserto de Atacama, Chile. As características quantitativas foram avaliadas em arbustos de 5 anos. Uma abordagem Bayesiana, usando o algoritmo de Cadeias Independentes, foi empregada na análise estatística. A sobrevivência foi de 87% e 82% em Cuz-Cuz e Tangué, respectivamente. A correlação intra-classe para a sobrevivência atingiu o maior valor (0,47). A mesma correlação variou de 0,03 a 0,29 para as características de crescimento. Os resultados sugerem que a variação entre procedências existe, a qual pode ser usada para propósitos de melhoramento sob condições áridas do sul do deserto de Atacama.

Palavras-chave: seca, abordagem Bayesiana, ambientes áridos

## ABSTRACT

Acacia species have valuable agronomic characteristics, such as tolerance to unfavorable environmental conditions. The objective of the present study was to examine the crown diameter, height, basal trunk diameter and survival variability in fourteen natural populations of blue-leaf wattle (*Acacia saligna*) at two arid sites (Cuz-Cuz and Tangué) in the southern Atacama Desert, Chile. The quantitative traits were assessed in 5 years old shrubs. A Bayesian approach, implemented using Independence Chain algorithm, was used in data analyses. The survival rates were 87% and 82% in Cuz-Cuz and Tangué, respectively. The intra-class correlation for plant survival achieved the highest magnitude (0.47). The same correlation varied from 0.03 up to 0.29 for the growth characters. The results suggest that provenance variation exists among these populations, which may be used for breeding purposes under the dryland conditions of southern Atacama Desert.

Key words: drought, Bayesian approach, arid environments

## 1. INTRODUÇÃO

*Acacia saligna* (Labill) é um recurso genético valioso para o estabelecimento de plantações de arbustos forrageiros nos ambientes áridos (McDonald e Maslin, 2000), considerando que esta espécie tem desenvolvido certa tolerância ao estresse hídrico (Nativ et al., 1999; McDonald e Maslin, 2000). Esta leguminosa arbustiva tem sido plantada para produção de forragem e madeira, como planta ornamental, e também para reabilitação de terras e conservação de solos (Degen et al., 1995; McDonald e Maslin, 2000). O estabelecimento de sistemas com cultivos múltiplos (*inter-crop systems*) usando plantas de *Acacia saligna* é extremamente útil para a proteção do solo e a cultura agrícola associada, fixando nitrogênio, fornecendo recursos madeireiros aos agricultores e forragem de alta qualidade no meio da estação seca (Tiedeman e Johnson, 1992).

A parte sul do deserto de Atacama, no Chile, tem uma crítica e delicada produtividade agrônoma, devido aos contínuos períodos de seca (Squeo et al., 2006; Vidiella et al., 1999). Estes fenômenos climáticos estão afetando as atividades sociais e econômicas dos agricultores, uma vez que sua principal atividade é o gado caprino, os quais apresentam freqüentemente insuficientes ganhos de peso devido à carência de pastagens naturais. Os estudos prévios realizados no Chile (Ormazábal, 1991) indicaram que os sistemas silvo-pastorais são eficientes alternativas para superar os problemas de sustentabilidade econômica das zonas áridas e semi-áridas do país.

O estabelecimento de ensaios de procedências é o primeiro passo em qualquer programa de melhoramento de plantas, porque eles são ferramenta útil para a identificação dos melhores recursos genéticos, com o intuito de reabilitação dos solos e produção agro-florestal (Souza et al., 1999). Uma variabilidade genética significativa em genótipos de *Acacia saligna* é chave para a obtenção de uma melhor produtividade e aceitáveis níveis de sobrevivência nos períodos de seca (George et al., 2006; Mora e Meneses, 2004).

O objetivo analítico do presente estudo foi examinar o grau de variação para o diâmetro da copa, sobrevivência, altura da planta e diâmetro basal, em quatorze populações naturais de *Acacia saligna*, plantadas em dois ambientes do

sul do deserto de Atacama no Chile. Para este objetivo, foi empregada uma abordagem Bayesiana, via algoritmo de Cadeias Independentes.

## 2. MATERIAIS E MÉTODOS

### 2.1. Descrição do local e o ensaio

O estudo foi conduzido em dois locais do sul do Deserto de Atacama, na região administrativa de Coquimbo, norte do Chile. O local do Tangué está localizado na Província do Elqui (30°45', latitude sul; 71°47', longitude oeste); o local de Cuz-Cuz está localizado na Província do Choapa (31°63', latitude sul; 71°22', longitude oeste). O regime de precipitações médio está frequentemente na faixa dos 100 e 200 mm·ano<sup>-1</sup> no Tangué e maior que 200 mm·ano<sup>-1</sup> no Cuz-Cuz. Os registros de precipitações durante o período do experimento (1999-2004) foram: média de 121,6 e 217,6 mm·ano<sup>-1</sup>, mínima de 65 e 103 mm·ano<sup>-1</sup>, e máxima de 192 e 394 mm·ano<sup>-1</sup> para o Tangué e Cuz-Cuz, respectivamente.

O delineamento experimental utilizado foi em blocos completos ao acaso, contituídos por quatorze procedências da espécie, em dois locais, com 3 blocos de 20 plantas espaçadas a 2 x 3 m. No plantio, as plantas foram irrigadas com 5 litros de água, para contornar o problema do forte déficit hídrico causado pelo fenômeno da La Niña em 1999 (Squeo et al., 2006).

### 2.2. Variáveis e genótipos

Os ensaios incluíram as seguintes procedências Australianas provenientes da coleção do *Commonwealth Scientific and Industrial Research Organisation* (CSIRO): Ravensthorpe-1, Kelmscott, Sanford River, Greenough River, Geraldton, Murchison River, Mingenew, Lake Indoon, Moora, Muntadgin Rock, Lake Muir, Boyatup Hill, Lancelin e Ravensthorpe-2. As plantas foram medidas à idade de 5 anos, para as características diâmetro da copa, altura, diâmetro basal do tronco e sobrevivência (medida como uma característica binária).

### 2.3. Abordagem Bayesiana

Os componentes de variância, correlações intra-classe, e a predição dos efeitos de procedências foram pesquisados usando técnicas Bayesianas, via algoritmo de Cadeias Independentes (*Independent Chain algorithm*), de acordo

com Wolfinger e Kass (2000). Esta análise foi realizada no programa SAS (SAS INSTITUTE, 1999). O modelo geral associado ao ensaio de procedências foi o seguinte:

$$y = X\beta + Z_1p + Z_2w + \varepsilon$$

em que  $y$  é o vetor das respostas observadas,  $X$  é a matriz de delineamento que corresponde ao vetor de parâmetros do efeito do bloco ( $\beta$ );  $Z_1$  e  $Z_2$  são as matrizes de incidência correspondentes aos vetores de parâmetros do efeito genotípico (ou procedências),  $p$ , e efeito da parcela,  $w$ ;  $\varepsilon$  é o vetor de resíduos. Assume-se que  $p$ ,  $w$  e  $\varepsilon$  são independentes e com distribuição normal univariada, com média 0 e matriz de variância  $G$ ,  $W$  e  $R$ , respectivamente.

No contexto Bayesiano, definimos  $\theta$  como o vetor de componentes de variância. A função de densidade conjunta de  $(\beta, \gamma, \theta)$  é dada por:

$$f(\beta, p, \theta | y) = f(\beta, p | \theta, y) f(\theta | y)$$

A especificação de cada distribuição *a priori* dos elementos do modelo é dado por:

$$f(y | \beta, p, w, \sigma_e^2, \sigma_p^2, \sigma_w^2) \sim N(X\beta + Z_1p + Z_2w, R)$$

$$f(\beta) \propto 1$$

$$f(p | \sigma_p^2) \sim N(0, G)$$

$$f(w | \sigma_w^2) \sim N(0, W)$$

$$f(\sigma_p^2 / \sigma_{p0}^2, \nu_{p0}) \sim IG(\sigma_{p0}^2, \nu_{p0})$$

$$f(\sigma_w^2 / \sigma_{w0}^2, \nu_{w0}) \sim IG(\sigma_{w0}^2, \nu_{w0})$$

$$f(\sigma_e^2 / \sigma_{e0}^2, \nu_{e0}) \sim IG(\sigma_{e0}^2, \nu_{e0})$$

Os componentes de variância da procedência ( $\sigma_p^2$ ), da parcela ( $\sigma_w^2$ ) e do resíduo ( $\sigma_e^2$ ) têm distribuição Gama Inversa ( $\sim IG$ ), com parâmetros *a priori* ( $\sigma_{-0}^2, \nu_{-0}$ ) obtidos por meio das densidades base, utilizando distribuição não informativa de Jeffreys. A função de densidade *a priori* do parâmetro de efeito fixo é assumida como uniforme (opção FLAT).

A distribuição *a posteriori* conjunta é definida como:

$$f(\beta, p, \sigma_e^2, \sigma_p^2 | y) \propto f(\beta)f(p | \sigma_p^2)f(w | \sigma_w^2)f(\sigma_p^2)f(\sigma_w^2) \\ f(\sigma_e^2)f(y | \beta, p, w, \sigma_e^2, \sigma_p^2, \sigma_w^2)$$

As amostras *a posteriori* foram analisadas utilizando PROC AUTOREG e UNIVARIATE do SAS (SAS-Institute, 1999).

A moda de cada distribuição *a posteriori* da herdabilidade foi determinada de acordo ao procedimento de densidade de Kernel, calculada no aplicativo SAS/INSIGHT do SAS.

O subprograma GLIMMIX% (macro) foi usado para a análise da sobrevivência das plantas (usando um modelo logito). Em todos os casos de análise, a distribuição *a posteriori* teve 11000 amostras. As correlações de Pearson foram calculadas para determinar a relação genética entre a sobrevivência e as características de crescimento.

### 3. RESULTADOS E DISCUSSÃO

A média, mediana e moda da correlação intra-classe, e os valores médios *a posteriori* dos componentes de variância e seus respectivos intervalos de credibilidade Bayesianos são apresentados no Quadro 1. A magnitude das correlações intra-classe variou de baixa a moderada. No Cuz-Cuz a sobrevivência das plantas teve a maior correlação (0,474). No Tangué esta correlação foi moderada (0,198). A altura total teve uma moderada correlação em ambos os locais: 0,286 e 0,204, no Cuz-Cuz e Tangué, respectivamente. O diâmetro basal e da copa tiveram os menores valores de correlação. Em geral, a média *a posteriori* foi maior que a estimativa da moda dos coeficientes da correlação intra-classe, indicando assimetria da distribuição *a posteriori* para este parâmetro genético.

Quadro 1 - Estimativas de parâmetros usando inferência Bayesiana, para um conjunto de populações de *Acacia saligna* amostradas de diferentes localidades de Austrália, e avaliadas no sul do deserto de Atacama

Parâmetros	Cuz-Cuz				Tangué			
	BD	H	CD	S	BD	H	CD	S
Componentes de variância								
Procedência ( $\sigma_p^2$ )	78,22	2233,97	605,42	2,09	14,54	633,15	63,69	1,16
Parcela ( $\sigma_w^2$ )	694,51	2141,87	3335,05	1,70	86,54	484,68	187,32	4,14
Erro ( $\sigma_e^2$ )	526,56	3217,80	2479,82	0,44	481,27	1970,99	1735,21	0,37
Correlação intra-classe								
Média	0,059	0,286	0,092	0,474	0,025	0,201	0,032	0,198
Mediana	0,045	0,282	0,079	0,486	0,006	0,195	0,026	0,180
Moda	0,014	0,273	0,033	0,504	0,018	0,175	0,009	0,105
Desvio Padrão	0,051	0,092	0,067	0,158	0,024	0,071	0,026	0,131
Limite inferior	0,004	0,142	0,009	0,192	0,001	0,093	0,002	0,021
Limite superior	0,158	0,442	0,222	0,711	0,070	0,325	0,083	0,443

BD: diâmetro basal; H: altura; CD: diâmetro da copa; S: sobrevivência.

A predição dos efeitos de procedências é apresentada na Figura 1. Os resultados indicaram que as populações de Lancelin e Ravensthorpe-2 tiveram os

maiores valores preditos para as características de crescimento em Cuz-Cuz. A população Ravensthorpe-2 teve um valor predito negativo para a sobrevivência ao contrário de Lancelin, Kelmscott e Geraldton, que foram os melhores genótipos para estas características. Lancelin teve um valor predito alto para todas as características avaliadas em Cuz-Cuz, entretanto, no Tangué, o desempenho foi diferente.

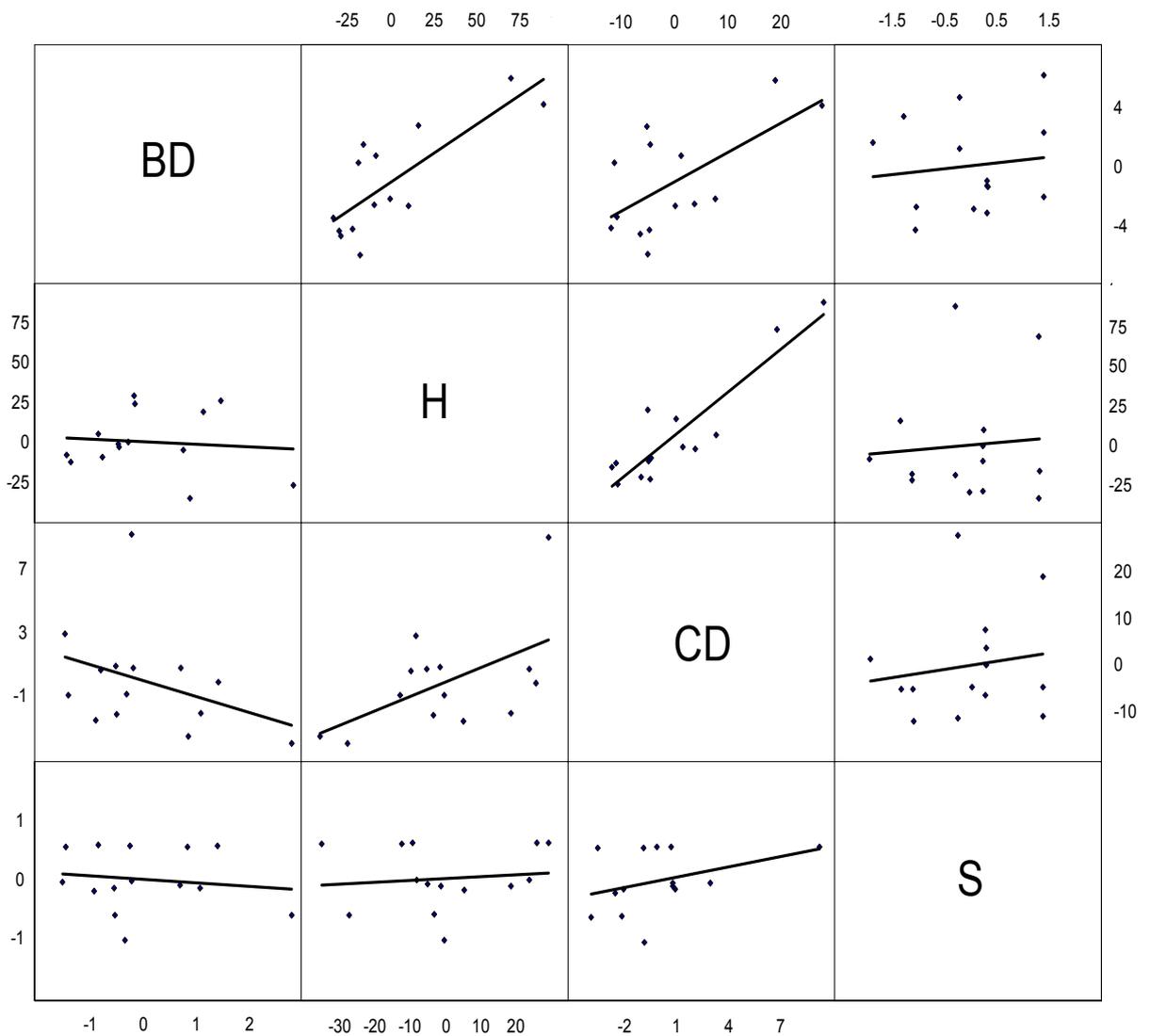


Figura 1 - Matrizes de dispersão com os valores de predição Bayesiana do efeito de procedências para as características: diâmetro basal (BD), altura total (H), diâmetro de copa (CD) e sobrevivência (S), avaliadas em populações de *Acacia saligna* no sul do deserto de Atacama; as linhas fornecem uma idéia da associação entre as características.

No Tanguê, as procedências de Murchison River e Greenough River se desempenharam bem para o diâmetro da copa, a qual é uma característica significativamente correlacionada com a produção de forragem (Mora e Meneses, 2004) embora Greenough River evidenciasse um baixo valor predito para a sobrevivência.

As estimativas dos coeficientes de correlação de Pearson calculados entre os valores preditos da sobrevivência e as demais características (Quadro 2) variaram desde -0,130 a 0,367 e não foram significativamente diferentes de zero. Os coeficientes de correlação de Pearson calculados entre os valores preditos das características de crescimento foram positivos e significativamente diferentes de zero em Cuz-Cuz ( $P < 0,01$ ; Quadro 2); esta forte relação foi também corroborada no gráfico de dispersão da Figura 1. Por outro lado, no Tanguê houve uma relação não significativa entre o diâmetro basal, a altura total e o diâmetro da copa (Quadro 2; Figura 1). Entretanto, esta correlação foi significativa e positiva entre a altura total e o diâmetro da copa. O gráfico de dispersão corrobora a associação não significativa entre os valores preditos da sobrevivência e das características de crescimento em ambos os locais.

Quadro 2 - Correlações de Pearson calculadas entre os valores de predição Bayesiana do efeito de procedências. Na diagonal (em negrito) são apresentadas as estimativas para a mesma característica medida em ambos os ambientes; acima da diagonal são apresentados os coeficientes de correlação para o local de Cuz-Cuz e abaixo da diagonal os mesmos coeficientes para o local do Tanguê

	BD	H	CD	S
BD	<b>0,303 ns</b>	0,789 *	0,635 *	0,125 ns
H	-0,097 ns	<b>0,378 ns</b>	0,905 *	0,082 ns
CD	-0,362 ns	0,545 *	<b>0,189 ns</b>	0,160 ns
S	-0,130 ns	0,113 ns	0,367 ns	<b>-0,013 ns</b>

BD: diâmetro basal; H: altura; CD: diâmetro da copa; S: sobrevivência; \*: coeficiente de correlação foi significativamente diferente de zero ( $P < 0,05$ ); ns: a correlação não foi significativamente diferente de zero ( $P > 0,05$ ) segundo o teste t.

A colheita anual de forragem em plantas de *Acacia saligna* pode reduzir a quantidade de forragem disponível para alimento do gado e produção de madeira. Pode também reduzir o vigor, a produtividade e o tempo de vida da planta

(Tiedeman e Johnson, 1992). Estas respostas podem também variar entre genótipos da espécie. George et al. (2006) encontraram uma grande variação genética em uma amostra de populações naturais de *Acacia saligna*, embora dois grupos de plantas morfologicamente diferentes apresentassem o mesmo padrão genético. Esta alta variação genética entre e dentro de grupos de arbustos de *Acacia saligna* sugeriu a existência de uma ampla base genética, a qual é favorável para os programas de melhoramento da espécie. A variabilidade entre procedências depende das características sob estudo, entre as quais a sobrevivência e altura apresentaram as maiores respostas.

Nos ambientes áridos, a competência por recursos essenciais para as plantas, como a água, nutrientes, e luz, é fator chave que afeta as comunidades de plantas (Krueger-Mangold et al., 2004). No sul do deserto de Atacama, o evento de oscilação climática do El Niño (Squeo et al., 2006) tem uma tendência a aumentar os níveis de precipitações médias entre as latitudes de 30° e 35° S, na estação de inverno (Montecinos e Aceituno, 2003). Esta situação ocorreu em 2002 e, por conseguinte, as sementes provenientes das plantas locais germinaram abundantemente. Posteriormente, suas respectivas plântulas deram início a fortes processos de competição (Jara et al., 2006). Apesar disto, no presente estudo foram encontradas altas taxas de sobrevivência em ambos os locais: 87,3 e 81,8% para Cuz-Cuz e Tangué, respectivamente.

O grau de variação entre procedências confirma a existência de moderadas possibilidades para melhorar a sobrevivência das plantas e o crescimento em altura, por meio da utilização destas populações naturais. Tal informação é muito útil para os propósitos de melhoramento da espécie. Embora *Acacia saligna* seja altamente tolerante às condições de seca (Nativ et al., 1999), a variabilidade significativa do efeito de procedências para a sobrevivência pode ser usada para melhorar esta tolerância para as condições áridas.

Os coeficientes de correlação de Pearson não significativos (Quadro 2, diagonal em negrito) indicam uma fraca associação entre os valores de predição do efeito de procedências de ambos os locais. Por exemplo, Lancelin teve bom desempenho em Cuz-Cuz, porém a correlação de Pearson não significativa, estimada para a mesma característica sobre diferentes locais, indica que o efeito de interação procedência-local é significativo. A sobrevivência das plantas em diferentes locais pode ser determinada por várias causas: geadas temporais ou

prolongadas, seca, vento, competição e doenças. Para avaliar a importância das diferentes causas de mortalidade, a magnitude e a importância das interações genótipo-ambiente devem ser previamente conhecidos (Chambers et al., 1996).

A associação genética entre as taxas de crescimento e resistência à seca encontrada na literatura tem sido muito baixa (Coyle et al., 2006). No presente estudo, a sobrevivência das plantas teve uma fraca associação com as características de crescimento. No entanto, os valores de predição do efeito de procedências para Lancelin no Cuz-Cuz e Murchinson River no Tangué indicam que métodos de seleção indireta poderiam ser aplicados usando estas populações naturais.

#### 4. CONCLUSÕES

As altas taxas de sobrevivência demonstraram o potencial das plantas de *Acacia saligna* para crescer nos ambientes áridos ao sul do deserto de Atacama, embora os níveis de sobrevivência (determinados por meio dos valores preditos) dependam da população. Os resultados permitiram a identificação de certas procedências com desempenho satisfatório em sobrevivência e o crescimento, porém indicaram que a seleção baseada apenas na sobrevivência pode não ter efeito significativo sobre o crescimento. Pelos resultados obtidos, conclui-se que existe variação genética entre procedências, o que possibilita a seleção para as condições áridas do sul do Deserto de Atacama.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- CHAMBERS, P.G.S.; BORRALHO, N.M.G.; POTTS B.M. Genetic analysis of survival in *Eucalyptus globulus* ssp. *globulus*. **Silvae Genetica**, 45:107-112, 1996.
- COYLE, D.R.; COLEMAN, M.D.; DURANT, J.A.; NEWMAN, L.A. Survival and growth of 31 *Populus* clones in South Carolina. **Biomass and Bioenergy**, 30:750-758, 2006.
- DEGEN, A.A.; BECKER, K.; MAKKAR, H.P.S.; BOROWY, N. *Acacia saligna* as a fodder tree for desert livestock and the interaction of its tannins with fiber fractions. **Journal of the Science of Food and Agriculture**, 68:65-71, 1995.
- GEORGE, N.; BYRNE, M.; MASLIN, B.; YAN, G. Genetic differentiation among morphological variants of *Acacia saligna* (Mimosaceae). **Tree Genetics and Genomes**, 2:109-119, 2006.
- HOLMGREN, M.; LÓPEZ, B.C.; GUTIÉRREZ, J.R.; SQUEO, F.A. Herbivory and plant growth rate determine the success of El Niño Southern Oscillation-driven tree establishment in semiarid South America. **Global Change Biology**, 12:2263-2271, 2006.
- JARA, P.A.; ARANCIO, G.; MORENO, R.; CARMONA, M.R. Factores abióticos que influyen la germinación de seis especies herbáceas de la zona árida de Chile. **Revista Chilena de Historia Natural**, 79:309-319, 2006.
- KRUEGER-MANGOLD, J.; SHELEY, R.; ENGEL, R.; JACOBSEN, J.; SVEJCAR, T.; ZABINSKI, C. Identification of the limiting resource within a semi-arid plant association. **Journal of Arid Environments**, 58:309-320, 2004.
- MCDONALD, M.W.; MASLIN, B.R. Taxonomic revision of the Salwoods: *Acacia aulacocarpa* Cunn. ex Benth. and its allies (Leguminosae: Mimosoideae: section Juliflorae). **Australian Systematic Botany**, 13:21-78, 2000.
- MONTECINOS, A.; ACEITUNO, P. Seasonality of the ENSO-Related Rainfall Variability in Central Chile and Associated Circulation Anomalies. **Journal of Climate**, 16:281-296, 2003.

- MORA, F.; MENESES, R. Comportamiento de procedencias de *Acacia saligna* (Labill.) H. L. Wendl. en la región de Coquimbo, Chile. **Ciência Florestal**, 14:103-109, 2004.
- NATIV, R.; EPHRATH, J.E.; BERLINER, P.R.; SARANGA, Y. Drought resistance and water use efficiency in *Acacia saligna*. **Australian Journal of Botany**, 47:577-586, 1999.
- ORMAZÁBAL, C.S. Silvopastoral systems in arid and semiarid zones of northern Chile. **Agroforestry Systems**, 14:207-217, 1991.
- SAS-INSTITUTE. **SAS/STAT®8.0 User's Guide**. Cary: SAS Institute, 1999. 373p.
- SOUZA, G.M.; GONÇALVES, A.N.; ALMEIDA, M. Water deficit in relation to leaf and stem anatomy of *Eucalyptus camaldulensis* Dehn. shoots cultivated in vitro. **Scientia Agrícola**, 56:723-731, 1999.
- SQUEO, F.A.; TRACOL, Y.; LÓPEZ, D.; GUTIÉRREZ, J.R.; Cordova, A.M.; Ehleringer, J.R. ENSO effects on primary productivity in Southern Atacama Desert. **Advances in Geosciences**, 6:273-277, 2006.
- TIEDEMAN, J.A.; JOHNSON, D.E. *Acacia cyanophylla* for forage and fuelwood in North Africa. **Agroforestry Systems**, 17:169-180, 1992.
- VIDIELLA, P.E.; ARMESTO, J.J.; GUTIÉRREZ, J.R. Vegetation changes and sequential flowering after rain in the southern Atacama Desert. **Journal of Arid Environments**, 43:449-458, 1999.
- WOLFINGER, R.D.; KASS, R.E. Non-conjugate Bayesian analysis of variance component models. **Biometrics**, 56:768-774, 2000.

**CAPÍTULO VI**  
**AVALIAÇÃO GENÉTICA DE CLONES DE OLIVEIRA NO DESERTO DO**  
**ATACAMA, CHILE, USANDO INFERÊNCIA BAYESIANA**

**RESUMO**

A oliveira (*Olea europaea* L.) é espécie economicamente importante para países com clima mediterrâneo e tem sido alternativa apropriada e útil para sistemas agrícolas intensivos das zonas áridas do Chile. O objetivo do presente trabalho foi determinar os componentes de variância, herdabilidade no sentido amplo e a predição dos valores genéticos usando técnicas Bayesianas, em vinte e oito clones de oliveira avaliados no sul do Deserto de Atacama. Usou-se o algoritmo de cadeias independentes para estimar as distribuições a posteriori dos parâmetros. As estimativas da herdabilidade e os intervalos de credibilidade (parênteses) associados foram  $H^2=0,50$  (0,38-0,61), 0,49 (0,38-0,61), 0,81 (0,74-0,86), 0,53 (0,42-0,64), 0,26 (0,14-0,40) e 0,43 (0,17-0,65) para a produção de frutos (2006), produção total de frutos (2003-2006), área da secção transversal do tronco (2006), produtividade da árvore (2006), alternância da produção (2003-2006) e precocidade (2003). O ganho genético por seleção direta variou de 16 a 55% com intensidade de seleção de 28,6%. A moderada herdabilidade para produção de frutos, alternância de produção e precocidade sugere uma resposta positiva à seleção de clones de oliveira nas condições ambientais áridas.

Palavras-chave: herdabilidade, inferência Bayesiana, ambientes áridos.

## ABSTRACT

The olive (*Olea europaea* L.) is an economically important species for countries that have Mediterranean climate and has been an appropriate, useful alternative for intensive farming systems in the arid zones of Chile. The objective of this study was to determine: variance components, broad-sense heritability, and prediction of genetic values using the Bayesian techniques on twenty-eight olive clones in the southern part of the Atacama Desert. The independence chains algorithm was used to estimate posterior parameter distributions. Estimations of heritability and associated credibility intervals (parentheses) were  $H^2=0.50$  (0.38-0.61), 0.49 (0.38-0.61), 0.81 (0.74-0.86), 0.53 (0.42-0.64), 0.26 (0.14-0.40), and 0.43 (0.17-0.65) for fruit production (2006), total fruit production (2003-2006), area of trunk cross section (2006), tree productivity (2006), alternation of production (2003-2006), and precocity (2003). Genetic advance by direct selection varied from 16 to 55% with a selection intensity of 28.6%. Moderate heritability for fruit production, alternation of production, and precocity suggests a positive response to olive selection under the arid environmental conditions.

Key words: heritability, Bayesian inference, arid zones.

## 1. INTRODUÇÃO

A oliveira (*Olea europaea* L.) é cultivada em grande escala em diversos países de clima Mediterrâneo subtropical, e representa uma das árvores frutíferas mais importantes do ponto de vista econômico. Estima-se que 90% da produção mundial de azeitonas provêm da bacia do mar Mediterrâneo (Aydinalp et al., 2004; La Rosa et al., 2003), mas seu cultivo também tem se expandido como uma alternativa agrônômica em países tais como Argentina, Austrália, Brasil, Chile, México, África do Sul e Estados Unidos, entre outros. Na América, as oliveiras foram introduzidas para a produção de azeitonas de mesa, extração de azeite e para propósitos ornamentais.

As árvores da oliveira possuem uma alta tolerância às condições de desidratação (Tognetti et al., 2004) e aos ambientes caracterizados por longos períodos de ausência de chuvas. As oliveiras têm sido cultivadas em diversas condições climáticas, apresentando ampla adaptabilidade, desde terras secas com diferentes condições de solo, a montanhas com alta umidade (Sofo et al., 2004) embora apresentem diferenças locais para a produção de frutos (Lavee et al., 2003; Levin e Lavee, 2005).

No Chile, as zonas desérticas, áridas e semi-áridas abrangem uma extensão de 30 milhões de ha, correspondentes a aproximadamente 40% da área total do país. Embora estas zonas sejam consideradas extremadamente pobres em produtividade do ponto de vista agrônômico (Ormazábal, 1991), o cultivo da oliveira tem sido uma alternativa apropriada e útil para os sistemas agrícolas manejados intensivamente. A indústria chilena de azeite e azeitonas, provenientes desta zona do país, está sendo sustentada por meio de plantações irrigadas, para assegurar os níveis adequados de produção econômica e comercial.

Os agricultores da região estão considerando aumentar o número de plantações comerciais manejadas intensivamente, com o objetivo de desenvolver a indústria olivícola, com um mercado em constante crescimento. Como consequência, a estação experimental do Instituto de Investigaciones Agropecuárias (INIA), localizada na região administrativa do Atacama, tem coletado e conservado material genético de oliveira (clones, variedades,

sementes) os quais são avaliados periodicamente para atender as necessidades técnicas dos agricultores.

Nos programas de avaliação genética, a estimativa dos componentes de variância e a predição dos valores genéticos são aspectos chaves na análise genético-quantitativa de características de importância econômica (Cadena-Meneses e Castillo-Morales, 2000a, 2000b). Estatisticamente, as estimações dos componentes de variância podem ser obtidas por meio do método da Máxima Verossimilhança Restrita (REML), e as predições genéticas por meio da Melhor Predição Linear não Tendenciosa (BLUP), as quais são metodologias preferidas quando o conjunto de dados apresenta informações perdidas (Cadena-Meneses e Castillo-Morales, 2000; Domínguez-Viveros et al., 2003; Sánchez-Vargas et al., 2004).

Atualmente, os métodos Bayesianos também estão sendo utilizados em diferentes programas de melhoramento para estimar parâmetros genéticos de características de interesse econômico (Cappa e Cantet, 2006; Mora e Arnhold 2006; Mora et al., 2007; Wright et al., 2000). As técnicas Bayesianas utilizam diferentes métodos derivados das Cadeias de Markov (Monte-Carlo Markov Chain: MCMC) tal como o algoritmo de Gibbs, que é o método mais utilizado nos programas de melhoramento (Van Tassell e Van Vleck, 1996). Os procedimentos Bayesianos são considerados uma alternativa aos métodos de estimação/predição REML/BLUP (Blasco, 2001; Cadena-Meneses e Castillo-Morales, 2000b).

Uma das estratégias adotadas dentro dos programas de melhoramento na região administrativa de Atacama é a avaliação de material propagado vegetativamente proveniente do banco de germoplasma do INIA. O entendimento do controle genético de diversas características agronômicas e a associação entre os caracteres (com fins de seleção genética) poderia gerar um conjunto de dados agronômicos que seriam úteis para o avanço do atual programa de melhoramento da oliveira, conduzido na região administrativa de Atacama. Por isto, o objetivo do presente trabalho foi examinar parâmetros genéticos, por meio de técnicas Bayesianas, em vinte e oito clones de oliveira avaliados para seis características quantitativas, no Sul do Deserto do Atacama.

## 2. MATERIAIS E MÉTODOS

O ensaio de campo se estabeleceu na estação experimental de Huasco, pertencente ao Instituto de Investigaciones Agropecuárias (INIA), localizada na Província de Huasco (28°34' Latitude Sul, 70°47' Longitude Oeste, e 453 m de altitude) na região administrativa de Atacama, norte de Chile. De acordo com a classificação de Köppen, o clima desta parte sul do Deserto de Atacama é classificado como tipo Bw, com uma precipitação média anual de 21 mm, que ocorrem durante o inverno. A temperatura média é de 16 °C, e com 63% de umidade relativa.

As árvores se originaram de estacas enraizadas, as quais foram obtidas aleatoriamente do banco de germoplasma do INIA, considerando o estoque de material agrônômico apropriado para a produção comercial de azeitonas.

O ensaio se estabeleceu na temporada agrícola 2000 usando um modelo de blocos completos ao acaso com vinte e oito tratamentos (clones), e dezesseis réplicas em esquema de uma árvore por parcela. O material vegetal foi plantado manualmente a um espaçamento de 8 x 4 m. Foi utilizado um sistema de irrigação por gotejamento para a aplicação de nutrientes e água (Orgaz e Fereres, 1999; Chartzoulakis et al., 1999). As necessidades de água foram determinadas usando dados locais de evaporação de bandeja.

Durante um período de seis anos avaliaram-se as seguintes características de campo:

- Produção de frutos ( $\text{kg árvore}^{-1}$ ), avaliada na temporada agrícola de 2006 (P2006).

- Produção acumulada de frutos ( $\text{kg árvore}^{-1}$ ), avaliada durante as temporadas agrícolas de 2003, 2004, 2005 e 2006 (PA).

- Área da secção transversal do tronco ( $\text{cm}^2$ ) avaliadas em 2006 (ASTT). A ASTT é uma medida do vigor das árvores (Tous et al., 1998) a qual foi calculada usando o perímetro do tronco a 20 cm do nível do solo.

- Produtividade da árvore ( $\text{kg cm}^{-2}$ ), medida entre o quociente de PA e ASTT (PA/ASTT), de acordo com Tous et al. (1998).

- Alternância da produção (AP), cujo valor por árvore foi medido utilizando os dados de produção das temporadas agrícolas de 2003 a 2006. AP foi medido por meio do seguinte índice:

$$AP = \left( \frac{1}{(n-1)} \right) \left\{ \frac{|(a_2 - a_1)|}{a_1 + a_2} + \frac{|(a_3 - a_2)|}{a_2 + a_3} \dots + \frac{|(a_n - a_{n-1})|}{a_{n-1} + a_n} \right\}$$

onde n é número de anos de avaliação;  $a_1, a_2, a_3, \dots, a_{n-1}$  é a produção de frutos por árvore ( $\text{kg árvore}^{-1}$ ) avaliados durante quatro temporadas agrícolas.

- Precocidade da produção (P), avaliada na temporada inicial de produção (2003). P foi medido como uma característica binária, considerando a ausência ou presença de um determinado nível de produção na idade de 3 anos. Esta característica binária foi baseada na seguinte pressuposição: valor 0, para as árvores que não produziram frutos, do ponto de vista comercial (ou seja, uma produção mais baixa que  $5 \text{ kg árvore}^{-1}$ ) e valor 1, para aquelas árvores cuja produção foi superior (ou igual) a  $5 \text{ kg árvore}^{-1}$  (Tous et al., 1998).

Foi utilizado o método Bayesiano para examinar componentes de variância, valores genéticos e a herdabilidade (sentido amplo) para as características quantitativas. Para a obtenção das distribuições *a posteriori* dos parâmetros foi utilizado o algoritmo de Cadeias Independentes (Wolfinger e Kass, 2000), variante dos métodos MCMC (Markov-Chain Monte Carlo). Cada distribuição *a posteriori* da herdabilidade foi caracterizada em função da média, mediana, moda e os intervalos de credibilidade Bayesianos. Os componentes de variância e valores genéticos foram caracterizados por meio da média *a posteriori*. Para esta análise foram utilizados os procedimentos PROC MIXED do SAS, com a opção PRIOR (SAS Institute, 1996). Adicionalmente, foi utilizada a macro GLIMMIX% (Wolfinger e O'Connell, 1993) para ajustar um modelo logístico (distribuição binomial função de ligação *logit*) na característica precocidade da produção.

O modelo geral associado ao experimento foi o seguinte:

$$y = X\beta + Z\gamma + \varepsilon$$

em que  $y$  é o vetor das respostas observadas,  $X$  é a matriz de delineamento que corresponde ao vetor de parâmetros do efeito do bloco ( $\beta$ );  $Z$  é a matriz de incidência que corresponde ao vetor de parâmetros do efeito genotípico ( $\gamma$ ); e  $\varepsilon$  é o vetor de resíduos. Assume-se que  $\gamma$  e  $\varepsilon$  são independentes

e com distribuição normal univariada, com média 0 e matriz de variância G e R, respectivamente.

Definimos  $\theta$  como o vetor de componentes de variância. A função de densidade conjunta de  $(\beta, \gamma, \theta)$  é dada por:

$$f(\beta, \gamma, \theta | y) = f(\beta, \gamma | \theta, y) f(\theta | y)$$

A especificação de cada distribuição *a priori* dos elementos do modelo é dado por:

$$f(y | \beta, \gamma, \sigma_e^2, \sigma_g^2) \sim N(X\beta + Z\gamma, R)$$

$$f(\beta) \propto 1$$

$$f(\gamma | \sigma_g^2) \sim N(0, G)$$

$$f(\sigma_g^2) \sim IG(a, b)$$

$$f(\sigma_e^2) \sim IG(c, d)$$

$$f(\sigma_g^2 / \sigma_{g0}^2, \nu_{g0}) \sim IG(\sigma_{g0}^2, \nu_{g0})$$

$$f(\sigma_e^2 / \sigma_{e0}^2, \nu_{e0}) \sim IG(\sigma_{e0}^2, \nu_{e0})$$

Os componentes de variância genética ( $\sigma_g^2$ ) e do resíduo ( $\sigma_e^2$ ) têm distribuição Gama Inversa ( $\sim IG$ ), Cadena-Meneses e Castillo-Morales (2000b), com parâmetros *a priori* ( $\sigma_{-0}^2, \nu_{-0}$ ) obtidos por meio das densidades base, utilizando distribuição não informativa de Jeffreys. A função de densidade *a priori* do parâmetro de efeito fixo é assumida como uniforme (opção FLAT).

A distribuição *a posteriori* conjunta é definida como:

$$f(\beta, \gamma, \sigma_e^2, \sigma_g^2 | y) \propto f(\beta) f(\gamma | \sigma_g^2) f(\sigma_g^2) f(\sigma_e^2) f(y | \beta, \gamma, \sigma_e^2, \sigma_g^2)$$

As amostras *a posteriori* foram analisadas utilizando PROC AUTOREG e UNIVARIATE do SAS (SAS Institute, 1996).

A moda de cada distribuição *a posteriori* da herdabilidade foi determinada de acordo ao procedimento de densidade de Kernel, calculada no aplicativo SAS/INSIGHT do SAS.

A herdabilidade em sentido amplo ( $H^2$ ) foi calculada para cada característica, usando o quociente entre a variância genética e fenotípica (Dhanai et al., 2003):

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

em que  $\sigma_g^2$  e  $\sigma_e^2$  correspondem à variância genética e residual, respectivamente. Para examinar a significância estatística do efeito genotípico (clones) empregou-se o Critério de Informação de Deviance (*Deviance Information Criteria*: DIC):

$$DIC = \bar{D} + pD = D(\bar{\theta}) + 2pD$$

em que  $\bar{D}$  corresponde à média *a posteriori* de  $-2 \times$  Logaritmo (Verossimilhança);  $pD$  é o número efetivo de parâmetros do modelo. Para esta análise Bayesiana foi utilizado o programa WinBugs 1.4 (Spiegelhalter et al., 2002).

Foram calculados coeficientes de correlação de Pearson entre os valores genéticos (médias *a posteriori*) de cada característica ( $n = 28$ ) para examinar a magnitude de associação genética de acordo com Agrawal et al. (2004). Foram construídas matrizes de dispersão com as classificações de clones para estabelecer as diferenças no ordenamento genotípico, de acordo com Gonçalves et al. (2005).

### 3. RESULTADOS E DISCUSSÃO

As estimativas provenientes das distribuições *a posteriori* dos componentes de variância e a herdabilidade (sentido amplo;  $H^2$ ) são apresentados no Quadro 1. O modelo que incorpora o efeito devido aos clones evidenciou o menor valor do Critério de Informação de Deviance (DIC), indicando a existência de variabilidade genética em todas as características estudadas. Este resultado confirma que a estratégia adotada no atual programa de melhoramento, resultará em avanços significativos devido à seleção de clones de oliveira.

Quadro 1 - Estimativas *a posteriori* dos componentes de variância (genética:  $\sigma_g^2$ , residual:  $\sigma_\varepsilon^2$  e fenotípica:  $\sigma_p^2$ ) e herdabilidade em sentido amplo ( $H^2$ ), para seis características agrônômicas avaliadas em vinte e oito clones de oliveira, no Deserto do Atacama, Chile

Variáveis	Componentes de variância			Herdabilidade ( $H^2$ )				
	$\sigma_g^2$	$\sigma_\varepsilon^2$	$\sigma_p^2$	Média	Mediana	Moda	LI	LS
P	0,8310	0,9717	1,8027	0,432	0,444	0,475	0,171	0,646
PA	2451,4	2447,1	4898,5	0,494	0,495	0,491	0,378	0,608
ASTT	9822,4	2281,6	12104,0	0,806	0,809	0,812	0,742	0,863
AP	0,0109	0,0305	0,0413	0,259	0,257	0,240	0,135	0,395
PA/ASTT	0,0423	0,0371	0,0793	0,526	0,527	0,525	0,418	0,636
P2006	473,83	466,70	940,52	0,497	0,498	0,497	0,381	0,611

P: precocidade da produção; PA: produção acumulada (período 2003-2006); ASTT: área da secção transversal do tronco; AP: alternância da produção; PA/ASTT: produtividade da árvore; P2006: produção de azeitonas no ano de 2006. LI e LS: são os limites inferiores e superiores do intervalo de credibilidade Bayesiano.

As distribuições *a posteriori* dos valores de herdabilidade são mostradas na Figura 1. No presente estudo, foi confirmada a convergência das cadeias estocásticas geradas pelo algoritmo de cadeias independentes (Heidelberger e Welch, 1983). Este resultado era esperado uma vez que segundo o recente estudo de Wolfinger e Kass (2000) o algoritmo de cadeias independentes é bastante flexível para dar convergência à cadeia de amostras *a posteriori*, sendo

vantajoso quando comparado aos métodos MCMC tradicionais, como o algoritmo de Gibbs.

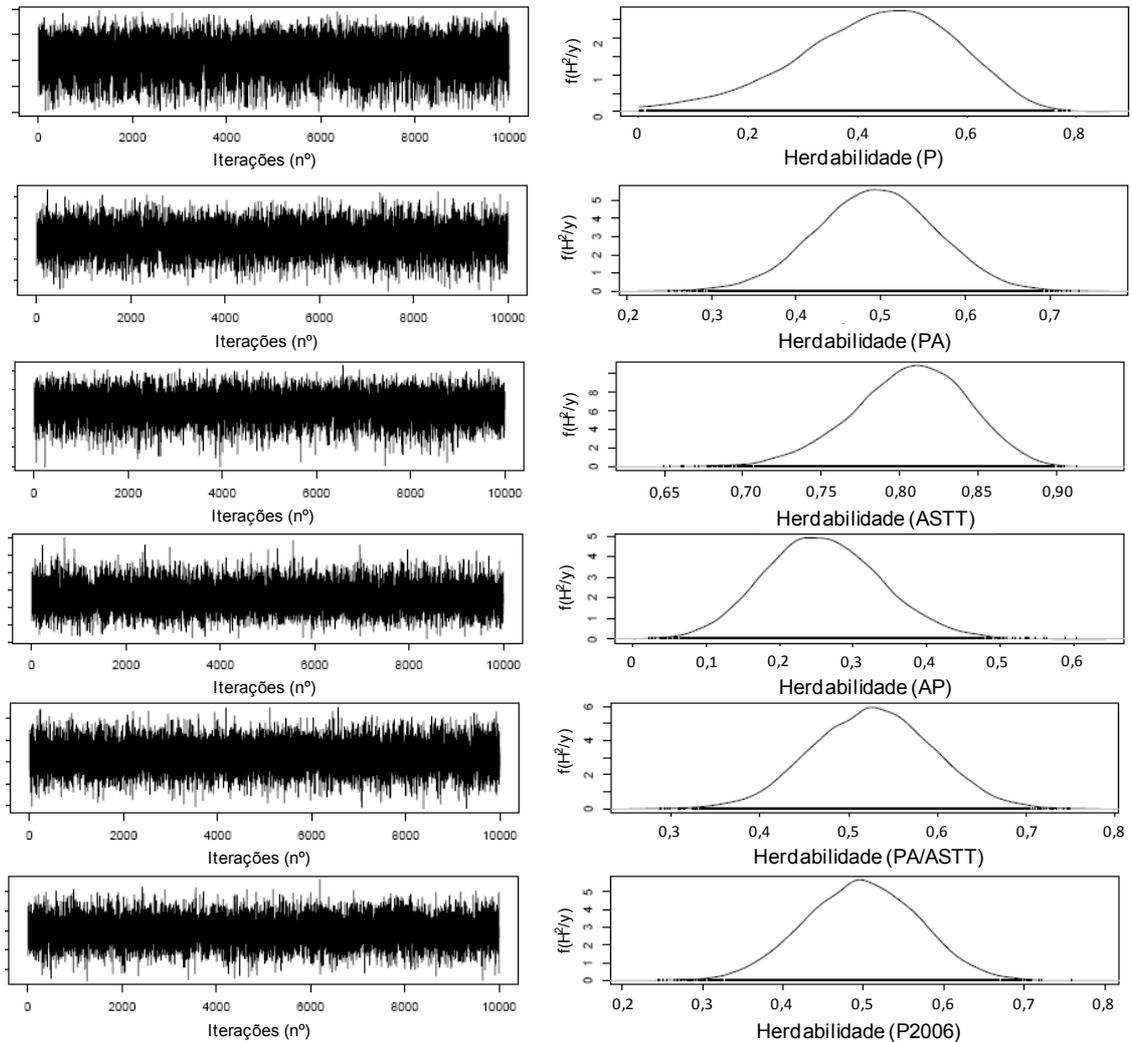


Figura 1 - Distribuições *a posteriori* das herdabilidades em sentido amplo (direta), para as características de precocidade (P), produção acumulada (PA), área da secção transversal do tronco (ASTT), alternância da produção (AP), produtividade (PA/ASTT) e produção do ano de 2006 (P2006), avaliado em clones de oliveiras, no Deserto do Atacama, Chile. À esquerda é mostrado o processo de iteração do algoritmo de cadeias independentes (análise Bayesiana).

De acordo com os resultados obtidos, as características agrônômicas evidenciaram herdabilidade de moderada à alta. O crescimento vegetativo (medido por meio da ASTT) teve o maior valor de herdabilidade, com média e intervalo de credibilidade *a posteriori* de  $H^2=81\%$  e (74% - 86%), ao contrário de AP, que teve o valor mais baixo, indicando herdabilidade moderada, com média *a posteriori* e intervalo de credibilidade de  $H^2=26\%$  e (14% - 40%). Para a

precocidade da produção foi também encontrada herdabilidade moderada,  $H^2=43\%$  e (17% - 65%). O parâmetro de dispersão do modelo logístico ajustado para a análise Bayesiana foi de 0,97 indicando um ajuste adequado (Myers et al., 2002). P2006, PA e PA/ASST tiveram valores similares de herdabilidade (cerca de 50%), as quais se mostraram características de moderada herdabilidade. A magnitude da herdabilidade encontrada nesta população de melhoramento revela uma situação favorável à seleção de clones de oliveira, que poderia significar em ganhos genéticos importantes se este material selecionado é usado nas futuras plantações com a espécie. Os valores de herdabilidade obtidos em um único local são úteis para entender a arquitetura genética da população de melhoramento, cuja variabilidade e resposta à seleção são estudadas nas condições ambientais específicas (Zamudio et al., 2002).

O programa de seleção e melhoramento de oliveira está desenvolvendo clones que respondam com adaptabilidade específica para as condições ambientais do Deserto de Atacama. A seleção clonal tem sido importante e valiosa estratégia, mas o conjunto de dados agronômicos provenientes dos ensaios clonais é usualmente desbalanceado. Considerando este fato, as estratégias de avaliação que incorporam as metodologias REML/BLUP, são consideradas ótimos métodos para um conjunto de dados com observações perdidas (Cadena-Meneses e Castillo-Morales, 2000a e 2000b; Mora e Scapim, 2007) e a metodologia Bayesiana oferece também importante ferramenta de inferência científica na avaliação genética, uma vez que também leva em consideração a variabilidade existente em todos os parâmetros (Wright et al., 2000; Nogueira et al., 2003; Cadena-Meneses e Castillo-Morales, 2000a).

A predição dos efeitos genéticos (aditivos e não aditivos) para os melhores oito clones de oliveira (28,6% de intensidade de seleção) é mostrada no Quadro 2. A classificação clonal realizada em função da média *a posteriori* de cada valor genético (usando a predição Bayesiana, via algoritmo de Cadeias Independentes) dependeu da característica em análise. Com base nestes resultados, foi possível identificar clones que respondem bem para diversas características agronômicas; por exemplo, o clone I39 teve uma boa resposta para P, P2006, PA, ASTT e AP. Outros clones com resposta satisfatória, ao considerar as características de produção, foram I10 e I8. A predição dos ganhos genéticos, considerando a seleção direta dos oito melhores clones desta

população de melhoramento, variou de 16,3% a 55,4%. A precocidade da produção apresentou o maior avanço genético, e a alternância da produção o menor. Importantes avanços podem ser obtidos considerando as características de produção de frutos (P2006 e PA).

Quadro 2 - Resultados do avanço genético (GG) por seleção direta (28,6% de intensidade de seleção) e a predição de valores genéticos para os melhores oito clones de oliveira, avaliados na região administrativa de Atacama, norte do Chile

Clone	P	Clone	PA	Clone	ASTT	Clone	AP	Clone	PA/ASTT	Clone	P2006
I24	0,91	I10	111,09	I24	240,85	I39	-0,14	I1	0,53	I10	47,34
I39	0,78	I8	68,66	I8	185,50	I1	-0,12	I44	0,24	I8	30,38
I1	0,77	I1	62,44	I15	143,83	I8	-0,09	I45	0,21	I39	25,90
I44	0,58	I39	42,20	I10	116,56	I24	-0,08	I41	0,20	I23	23,80
I35	0,45	I41	37,31	I29	66,14	I29	-0,07	I42	0,17	I32	12,79
I45	0,36	I29	27,88	I39	62,74	I25	-0,07	I10	0,13	I11	12,19
I29	0,30	I15	25,42	I23	48,98	I35	-0,06	I11	0,09	I46	9,95
I17	0,29	I40	24,77	I28	36,68	I15	-0,06	I35	0,09	I4	9,39
Média	0,63		141,37		246,58		0,53		0,61		62,84
GG(%)	55,4		35,3		45,7		16,3		34,1		34,2

P: Precocidade da produção; PA: Produção acumulada (2003-2006); ASTT: Área da seção transversal do tronco; AP; Alternância da produção; PA/ASTT: Produtividade da árvore; P2006: Produção de azeitonas no ano 2006.

Os gráficos de dispersão com a classificação genética (*ranking*) para as características em análise são mostrados na Figura 2. Este *ranking* foi realizado utilizando a predição Bayesiana dos valores genéticos dos clones. As linhas ajustadas à dispersão dos dados representam uma medida de associação entre o ordenamento dos genótipos (*ranking*). Os valores de correlação de Pearson variaram de -0,67 a 0,72. AP apresentou uma correlação (associação) negativa com a maioria das características; as estimativas significativas de correlação variaram de -0,41 a -0,67 ( $p < 0,05$ ). Esta característica apresentou um valor negativo de correlação com PA/ASTT (-0,25), mas esta não foi significativamente diferente de zero ( $p > 0,05$ ).

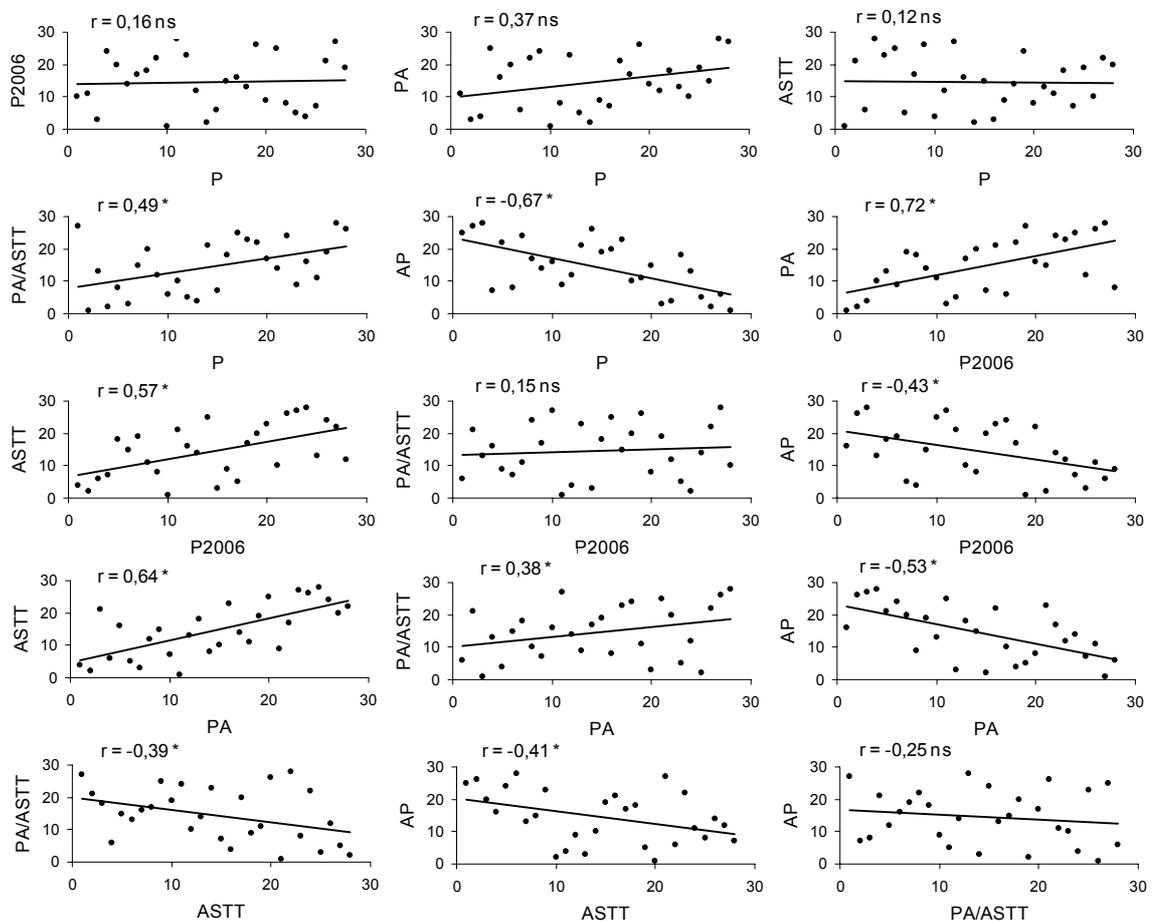


Figura 2 - Gráficos de dispersão com o ordenamento dos clones (*ranking*), realizado via predição Bayesiana, para as características em análise. \* Correlação de Pearson ( $r$ ) é significativamente diferente de zero a  $p=0,05$  (*t-student*); ns: não significativa. P: Precocidade da produção; PA: Produção acumulada (2003-2006); ASTT: Área da seção transversal do tronco; AP; Alternância da produção; P/ASTT: Produtividade da árvore; P2006: Produção de azeitonas no ano 2006.

As correlações de Pearson entre P e as características de produção (P2006 e PA) e crescimento vegetativo (ASTT) não foram estatisticamente diferentes de zero (Figura 2). Embora haja uma variabilidade genética importante para a precocidade ( $H^2=0,432$ ), a baixa associação genética com as características de produção pode apresentar desvantagem, uma vez que as plantações usualmente possuem elevada produção em conjunto com indivíduos precoces. Estes resultados são coincidentes com o estudado por Pritsa et al. (2003) e Tous et al. (1998), os quais também coincidem na importância de

encontrar genótipos de oliveira com habilidade para produzir frutos a uma idade precoce. Por sua vez, o genótipo da oliveira deveria ser preferencialmente precoce, embora economicamente viável (Tous et al., 1998) uma vez que o custo de estabelecimento de uma plantação com este cultivo é normalmente alto, e as árvores de oliveira são cultivadas em locais com custo elevado.

Apesar da fraca associação entre precocidade e produção, de acordo com os resultados da predição genética de clones (Quadro 2) existem clones que poderiam ser valiosos para a obtenção de indivíduos precoces e com máxima produção. A seleção combinada seria coincidente com o realizado por León et al. (2004) os quais promoveram o estabelecimento de sistemas agrícolas custo-efetivo, por meio da seleção de genótipos precoces com alta produção acumulada (León et al., 2004).

Os valores de correlação de Pearson estimados entre AP e P, P2006, PA, e ASTT são importantes para a seleção indireta de características de interesse econômico, ou seja, a correlação negativa entre AP e PA indica que um índice de alternância de produção alta está associado com produção baixa, e vice versa. A alternância de produção é medida da tendência de um genótipo (árvore) para alternar anos de alta e baixa produtividade, a qual se considera um fator limitante nas plantações comerciais de oliveira (Levin et al., 2005). Finalmente, o presente estudo enfatiza o fato que a precocidade e a alta produção de frutos, em conjunto com uma baixa a moderada tendência de alternância de produção, são os componentes mais importantes para o estabelecimento de plantações eficientes em ambientes com limitações hídricas.

#### **4. CONCLUSÕES**

No presente estudo, a produção de frutos, a alternância de produção e a precocidade foram moderadamente herdáveis, permitindo a seleção direta de clones quanto ao seu valor genotípico. Esta é uma informação útil para dar continuidade ao atual programa de melhoramento genético conduzido no sul do Deserto de Atacama.

A seleção genética baseada em características econômicas e a utilização da inferência Bayesiana na predição dos valores genéticos dos clones foram suficientes para o presente programa de melhoramento. A abordagem Bayesiana utilizada como ferramenta de análise estatística foi útil para propósitos de melhoramento de oliveira e para entender como as características são controladas geneticamente.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, A.A.; CONNER, J.K.; STINCHCOMBE, J.R. Evolution of plant resistance and tolerance to frost damage. **Ecology Letters**, 7:1199-1208, 2004.

AYDINALP, C.; CRESSER, M.S.; MCCLEAN, C. Characterization of some important agricultural soils under olive trees. **J. Cent. Eur. Agric.**, 5:101-108, 2004.

BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, 79:2023–2046, 2001.

CADENA-MENESES, J.A.; CASTILLO-MORALES, A. A comparison of SAS and Harvey in variance component estimation in mixed models. **Agrociencia**, 34:57-68, 2000a.

CADENA-MENESES, J.A.; CASTILLO-MORALES, A. Variance component estimation, a comparison supported by simulation. **Agrociencia**, 34:343-352, 2000b.

CAPPA, E.P.; CANTET, R.J.C. Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. **Canadian Journal of Forest Research**, 36:1276-1285, 2006.

CHARTZOULAKIS, K.; PATAKAS, A.; BOSABALIDIS, A.M. Changes in water relations, photosynthesis and leaf anatomy induced by intermittent drought in two olive cultivars. **Environ. Exp. Bot.**, 42:113–120, 1999.

DHANAI, C.S.; UNİYAL, A.K.; TODARIA, N.P. Source variation in *Albizia chinensis* (Osbeck) Mer.: seed and seedling characteristics. **Silvae Genética**, 52:259-266, 2003.

DOMÍNGUEZ-VIVEROS, J.; NÚÑEZ-DOMÍNGUEZ, R.; RAMÍREZ-VALVERDE, R.; RUIZ-FLORES, A. Genetic evaluation of growth traits in tropicane cattle: I. Model selection. **Agrociencia**, 37:323-335, 2003.

GONÇALVES, P.S.; BORTOLETTO, N.; CARDINAL, A.B.B.; GOUVÊA, L.R.L.; COSTA, R.B.; MORAES, M.L.T. Age-age correlation for early selection of rubber

tree genotypes in São Paulo State, Brazil. **Genetics and Molecular Biology**, 28(4): 758-764, 2005.

HEIDELBERGER, P.; WELCH, P.D. Simulation run length control in the presence of an initial transient. **Operations Research**, 31:1109-1114, 1983.

LA ROSA, R.; ANGIOLILLO, A.; GUERRERO, C.; PELLEGRINI, M.; RALLO, L.; BESNARD, G.; BERVILLÉ, A.; MARTÍN, A.; BALDÓN, L. A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. **Theor. Appl. Genet.**, 106:1273-1282, 2003.

LAVEE, S.; AVIDAN, B.; MENI, Y. "Askal", una nueva variedad de almazara sobresaliente por su comportamiento agronómico para olivares intensivos y superintensivos. **Olivae**, 97:53-59, 2003.

LEVIN, A.G.; LAVEE, S. The influence of girdling on flower type, number, inflorescence density, fruit set, and yields in three different olive cultivars (Barnea, Picual, and Souri). **Aust. J. Agr. Res.**, 56:827-831, 2005.

LEÓN, L.; RALLO, L.; DEL RÍO, C.; MARTÍN, L.M. Variability and early selection on the seedling stage for agronomic traits in progenies from olive crosses. **Plant Breeding**, 123:73-78, 2004.

MORA, F.; ARNHOLD, E. Inferencia Bayesiana y metodología de modelos lineales mixtos aplicados al mejoramiento de maíz. **Ciencia e Investigación Agraria**, 33:217-223, 2006.

MORA, F.; PUPIM-JUNIOR, O.; SCAPIM, C.A. Predicción del efecto de cultivares de algodón en la presencia de interacción genotipo-ambiente. **Ciencia e Investigación Agraria**, 34:11-19, 2007.

MORA, F.; SCAPIM, C.A. Predicción de valores genéticos del efecto de poblaciones de maíz evaluadas en Brasil y Paraguay. **Agricultura Técnica**, 67:139-146, 2007.

MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. **Generalized linear models, with applications in engineering and the sciences**. New York: John Wiley & Sons, 2002. 342p.

ORGAZ, F.; FERERES, E. Riego. In: BARRANCO N.; FERNÁNDEZ-ESCOBAR, D.; RALLO, R.L. (eds.) **El cultivo del olivo**. 3ra edición. Madrid: Junta de Andalucía y Ediciones Mundi-Prensa, 1999. 364p.

ORMAZÁBAL, C.S. Silvopastoral systems in arid and semiarid zones of northern Chile. **Agroforestry Systems**, 14:207-217, 1991.

PRITSA, T.S.; VOYIATZIS, D.G.; VOYIATZI, C.J.; SOTIRIOU, M.S. Evaluation of vegetative growth traits and their relation to time to first flowering of olive seedlings. **Australian Journal of Agricultural Research**, 54:371-376, 2003.

SÁNCHEZ-VARGAS, N.M.; VARGAS-HERNÁNDEZ, J.J.; RUIZ-POSADAS, L.M.; LÓPEZ-UPTON, J. Repeatability of genetic parameters in a clonal test of *Eucalyptus urophylla* S. T. Blake in southeastern México. **Agrociencia**, 38:465-475, 2004.

SOFO, A.; DICHIO, B.; XILOYANNIS, C.; MASIA, A. Lipoxygenase activity and proline accumulation in leaves and roots of olive trees in response to drought stress. **Physiologia Plantarum**, 121:58-65, 2004.

SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P.; VAN DER LINDE, A. Bayesian Measures of Model Complexity and Fit. **Journal of the Royal Statistical Society Series B**, 64:583-639, 2002.

TOGNETTI, R.; D'ANDRIA, R.; MORELLI, G.; CALANDRELLI, D.; FRAGNITO, F. Irrigation effects on daily and seasonal variations of trunk sap flow and leaf water relations in olive trees. **Plant and Soil**, 263:249-264, 2004.

TOUS, J.; ROMERO, A.; PLANA, J. Comportamiento agronómico y comercial de cinco variedades de olivo en Tarragona. **Investigación Agraria: Producción e Protección Vegetal**, 13:97-109, 1998.

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. **Journal of Animal Science**, 74:2586-2597, 1996.

WOLFINGER, R.; O'CONNELL, M. Generalized Linear Mixed Models: A pseudo-likelihood approach. **Journal of Statistical Computation and Simulation**, 4:233–243, 1993.

WRIGHT, D.R.; STERN, H.S.; BERGER, P.J. Comparing traditional and Bayesian analyses of selection experiments in animal breeding. **Journal of Agricultural, Biological, and Environmental Statistics**, 5:240-256, 2000.

ZAMUDIO, F.; BAETTYG, R.; VERGARA, A.; GUERRA, F.; ROZENBERG, P. Genetic trends in wood density and radial growth with cambial age in a radiata pine progeny test. **Ann. For. Sci.**, 59:541–549, 2002.

**CAPÍTULO VII**  
**HERDABILIDADE DO FLORESCIMENTO PRECOCE, CRESCIMENTO E**  
**SOBREVIVÊNCIA DE *Eucalyptus cladocalyx*, NA REGIÃO ÁRIDA DO CHILE:**  
**UMA ABORDAGEM BAYESIANA**

**RESUMO**

O objetivo deste estudo foi examinar o controle genético do florescimento precoce em trinta e nove famílias originadas de polinização aberta de *Eucalyptus cladocalyx*, no sul do deserto de Atacama, Chile, visando à seleção de árvores que floresçam precocemente, porém com um mínimo impacto sobre a sobrevivência e o crescimento. Uma abordagem Bayesiana, usando o algoritmo de Gibbs, foi empregada nas análises estatísticas. A precocidade foi medida como uma resposta binária em árvores de três anos, a qual foi altamente herdável. A estimativa da média *a posteriori* da herdabilidade e o intervalo de credibilidade de 95% foram 0,51 e (0,46-0,56) respectivamente. A estimativa do ganho genético variou de 36,15% a 41,22% (intensidades de seleção de 14,29% e 7,48%). A altura total, diâmetro basal e a sobrevivência (medida aos cinco anos) tiveram uma positiva e relativamente moderada correlação com o florescimento precoce ( $r=0,23$ ,  $0,29$  e  $0,11$ ) indicando que a seleção para o florescimento precoce poderia ter um moderado impacto sobre o crescimento e a sobrevivência. O florescimento precoce de *Eucalyptus cladocalyx* foi altamente herdável nesta população de melhoramento, a qual pode ser usada para melhorar a produção de mel em situações onde as flores de espécies nativas não são encontradas em quantidades suficientes.

Palavras-chave: herdabilidade, inferência Bayesiana, ambientes áridos

## ABSTRACT

The aim of this study was to examine the genetic control of flowering precocity in thirty-nine open-pollinated families of *Eucalyptus cladocalyx* in southern Atacama Desert, Chile, with the view to the selection of trees that flower earlier but with minimal impact on survival and growth. A Bayesian approach, implemented using Gibbs sampling algorithm, was used in data analyses. Precocity was assessed as a binary response at age three and was found to be highly heritable. The estimates of posterior mean and of the 95% posterior interval of the heritability were 0.51 and (0.46-0.56), respectively. Estimated genetic gain ranged from 36.15% to 41.22% (selection intensities of 14.29% and 7.48%). Total height, basal diameter and survival (measured at age 5) had a positive and relatively moderate genetic correlation with early flowering ( $r=0.23$ , 0.29 and 0.11) indicating that selection for early flowering alone would have a moderate impact on growth and survival. Early flowering of *Eucalyptus cladocalyx* was found to be highly heritable in this breeding population, which could be used to improve the honey production in situations where flowers from native species are not available in sufficient quantities.

Key words: arid environmental conditions, breeding values, Gibbs sampling

## 1. INTRODUÇÃO

As espécies arbóreas de eucalipto têm sido cultivadas comercialmente para diversos propósitos como produção de madeira, óleos essenciais, lenha, polpa e mel (He e Barr, 2004; YueMing et al., 2004; Hajari et al., 2006). Muitas espécies de eucalipto são particularmente bem adaptadas a condições climáticas áridas (Cohen et al., 1997; Wildy et al., 2004; Akhter et al., 2005) em tanto que outras podem também tolerar solos salinos (Rawat e Banerjee, 1998; Akilan et al., 1997; Marcar et al., 2002). Dado que, pelo menos 50% da superfície mundial terrestre são classificadas como de zonas áridas (Williams, 2000), resulta obvio o enorme potencial que tem estas espécies para o estabelecimento de plantações extensivas.

No Chile, por exemplo, as zonas áridas e semi-áridas abrangem 30 milhões de hectares, ou 40% da superfície total do país (Ormazábal, 1991). Grandes áreas de outros países de América do Sul são similarmente áridas, tais como Argentina, Brasil e Peru. As práticas agro-florestais que integram plantações de árvores e arbustos com cultivos agrícolas e/ou forragem são reconhecidas alternativas à agricultura convencional nestas regiões áridas. No norte do Chile, por exemplo, a região administrativa de Coquimbo tem uma limitada superfície produtiva do ponto de vista agrônômico (Ormazábal, 1991) e a salinidade dos solos áridos é percebida como um severo problema ambiental, econômico e social (Mora et al., 2007; Mora et al., 2008). Deste ponto de vista, as plantações florestais que permitam diferentes aplicações práticas são extremamente importantes, os quais podem contribuir ao uso sustentável da terra, além de fornecer um produto.

Uma razão do porque a espécie *Eucalyptus cladocalyx* F. Muell é cultivada amplamente no Chile, é devido a sua utilização tanto para produção de mel como para produção de madeira (Manikis e Thrasivoulou, 2001; Mora, 2006). Esta espécie é preferida a outras espécies devido a sua moderada tolerância à salinidade (Marcar et al., 1995), sua habilidade para crescer em ambientes áridos (Hanks et al., 1994; Gleadow e Woodrow, 2002; Woodrow et al., 2002; McDonald et al., 2003) e sua relativa resistência a *Phoracantha semipunctata* (Hanks et al., 1994). A produção de mel representa fonte potencial de renda extra para os agricultores locais.

As flores de *Eucalyptus cladocalyx* têm aptidão para a produção de mel de qualidade (Manikis e Thrasivoulou, 2001), complementando a escassa produção das espécies nativas e endêmicas da região de Coquimbo (Montenegro et al., 2003). Entretanto, as árvores de *Eucalyptus* não produzem flores antes dos 3 a 6 anos de idade e sua produção não é contínua nem uniforme. O florescimento precoce tem sido considerado importante característica nos programas de melhoramento de eucalipto (Chambers et al., 1997; Wiltshire et al., 1998) e as pesquisas relacionadas ao florescimento têm sido realizadas com o intuito de antecipar os resultados dos ensaios de progênie e para acelerar a produção de sementes melhoradas geneticamente (Meilan, 1997). Entretanto, a inclusão de genes para o florescimento em populações de eucalipto destinadas a produzir madeira pode ter efeitos indesejáveis no longo prazo. De fato, Chambers et al. (1997) afirmaram que a expressão contínua desta característica em sucessivas gerações poderia conduzir à redução significativa da produtividade, devido à perda de energia envolvida na produção de flores e frutos.

A produção precoce de flores, ou floração precoce, é freqüentemente medida como a presença/ausência de cápsulas e ou brotos de flores (Chambers et al., 1997). Estatisticamente, tal característica tem uma distribuição de Bernoulli, na qual são registrados somente dois distintos fenótipos não contínuos: presença ou ausência de eventos florais. Uma alternativa de análise é a utilização de métodos Bayesianos, os quais são considerados mais precisos para predizer o valor genético de tais características de limiar (Van Tassell et al., 1998). Na avaliação genética de árvores florestais, a abordagem Bayesiana tem sido utilizada, por exemplo, nos estudos de Sorya (1998), Cappa e Cantet (2006) e Waldmann e Ericsson (2006). Diferentes métodos de Monte Carlo – Cadeias de Markov (MCMC: Monte Carlo Markov Chain) tal como o algoritmo de Gibbs, têm mostrado ser apropriados para a análise de dados categóricos em programas de melhoramento. Além disto, os métodos MCMC permitem a análise em superfícies de verossimilhança complexas e permitem a obtenção das distribuições *a posteriori* Bayesianas (Melchinger et al., 2004; Bink et al., 2008).

O objetivo do presente estudo, portanto, foi determinar o controle genético do florescimento precoce em *Eucalyptus cladocalyx* visando a seleção de árvores que florescem precocemente, mas com um mínimo impacto sobre a sobrevivência e crescimento. Nas análises de parâmetros genéticos, foi considerada uma

abordagem Bayesiana, bem como na predição de valores genéticos para as características de florescimento, sobrevivência e crescimento, avaliadas em 39 famílias de *E. cladocalyx* cultivadas sob condições áridas no sul do deserto de Atacama, norte do Chile.

## 2. MATERIAIS E MÉTODOS

### 2.1. Localização do estudo

O experimento foi conduzido na Província de Choapa, na comunidade rural de Tunga Norte, localizada ao sul do Deserto de Atacama, na região administrativa de Coquimbo, Chile (31°38' S de Latitude; 71°19' W de Longitude; e altitude de 297 m). 75% da superfície da região são compostas de vegetação arbustiva xerófila, predominantemente, numa área de baixa altitude (altitude de 0 a 1000 m), sendo que, no entanto apenas 3% da superfície é representada por áreas destinadas à agricultura (Squeo et al., 2006).

A região tem um clima Mediterrâneo árido (Montenegro et al., 2003), na qual existem importantes variações em disponibilidade de água entre os anos de La Niña (seco) e El Niño (chuvoso) (Squeo et al., 2006). Durante os episódios de El Niño existe uma tendência a aumentar a precipitação média entre as latitudes 30° e 35° S (no inverno: meses de Junho a Agosto), entanto que ocorre um efeito contrário durante os eventos de La Niña (Montecinos e Aceituno, 2003). A precipitação média anual é de 180 mm (1973–2004, medida na estação meteorológica de Illapel, capital da Província de Choapa), variando de 513 a 15,8 mm. Os registros de precipitações durante o período experimental (2001-2006) foram: média de 226 mm ano<sup>-1</sup>, variando de 103 mm ano<sup>-1</sup> (2003) a 394 mm ano<sup>-1</sup> (2002; um ano de El Niño).

### 2.2. Ensaio e genótipos

O experimento foi estabelecido no ano 2001. O delineamento experimental foi de blocos completos com os tratamentos ao acaso, com 39 famílias de meios-irmãos, como os tratamentos principais, aleatorizadas em 30 blocos, e com parcelas conformadas por uma única árvore. A densidade de plantio foi de, aproximadamente, 1.667 árvores ha<sup>-1</sup>; equivalentes a um espaçamento de 2 x 3 m. Trinta e sete famílias foram provenientes da população base do Centro Australiano de Sementes e Árvores (CSIRO) e duas famílias de meio-irmãos provenientes de plantações previamente estabelecidas em uma localidade da Província de Choapa (semente procedente do Chile; 31°40' de Latitude Sul; 71°14' de Longitude Oeste). O lote de sementes Australianas foi

proveniente de quatro locais (procedências) do centro de origem da espécie (Ruthrof et al., 2003): Flinders Chase NP (35°57' de Latitude Sul; 136°42' de Longitude Oeste; 637,9 mm·ano<sup>-1</sup>), MT Remarkable (32°43' de Latitude Sul; 138°06' de Longitude Oeste; 242,8 mm·ano<sup>-1</sup>), Marble Range (34°30' de Latitude Sul; 135°30' de Longitude Oeste; 485,1 mm·ano<sup>-1</sup>), e Wirrabara SF (33°06' de Latitude Sul; 138°14' de Longitude Oeste; 256,6 mm·ano<sup>-1</sup>).

### **2.3. Características**

O florescimento precoce foi medido aos 3 anos de idade (ano 2004). Esta característica foi avaliada quanto à presença ou ausência de cápsulas, flores e/ou brotos. Estudos de precocidade em eucaliptos são freqüentemente tratados como sendo de resposta binária (Chambers et al., 1997). Por conseguinte, a variável resposta florescimento precoce foi tratada como binária, codificada com o valor 1 para a presença, e 0 para a ausência dos eventos de florescimento. A altura das árvores e o diâmetro basal foram medidos quando as árvores tinham 5 anos de idade, durante o ano 2006. O diâmetro basal foi medido aos 10 cm da base do solo para minimizar o erro associado ao nível do solo. A sobrevivência das árvores foi também considerada como sendo de resposta binária em 2006.

### **2.4. Análise Bayesiana**

Neste estudo, o método Bayesiano foi empregado com o intuito de minimizar o problema de viés, gerando uma distribuição de probabilidade para os parâmetros genéticos. De acordo com Ruzzante et al. (2004) esta distribuição pode então ser caracterizada pela média, mediana e moda, as quais podem ser utilizadas como estimativas pontuais.

Foi assumido um modelo Bayesiano de estimação de componentes de variância utilizando o algoritmo da amostragem de Gibbs. Para os dados de florescimento e de sobrevivência (que se mediram como variáveis discretas com resposta binária) foram ajustados modelos de limiar de acordo com Van Tassell et al. (1998). Modelos lineares foram ajustados para as características de crescimento. Para essas análises Bayesianas foi utilizado o programa MTGSAM (Van Tassell e Van Vleck 1996; Van Tassell et al., 1998). No presente estudo, os efeitos “fixos” correspondem aos efeitos de bloco e de procedência (ou sub-raça),

que seguiram a distribuição *flat prior* (uniforme). Foi assumido que os efeitos aleatórios são normalmente distribuídos e que os componentes de variância possuem distribuição de Qui-quadrado Invertida (Van Tassell e Van Vleck, 1996). Para mais detalhes sobre a densidade posterior conjunta, ver Van Tassell et al. (1998) e Van Tassell e Van Vleck (1996).

A convergência das cadeias geradas pelo algoritmo de Gibbs foi checada por meio do teste de Heidelberger e Welch's (1983) incorporado no programa R com a rotina CODA (*Convergence Diagnosis and Output Analysis*).

As significâncias do efeito de população (origem da semente ou procedência) foram testadas pelo procedimento MIXED do SAS para características de crescimento, e a macro %GLIMMIX (modelo *logit*; função de ligação canônica) para as características de florescimento precoce e sobrevivência (SAS Institute, 1996). Estes resultados foram também corroborados por meio dos intervalos de credibilidade (95%) construídos a partir das distribuições *a posteriori* dos valores médios das populações, estimado no programa WinBugs (Bayesian inference Using Gibbs Sampling; Spiegelhalter et al., 2003).

## 2.5. Estimativas de parâmetros genéticos

A herdabilidade individual (estrito senso) das plantas ( $h_a^2$ ) ou o grau de controle genético foi calculado para cada característica da seguinte forma:

$$h_a^2 = \frac{\sigma_a^2}{(\sigma_a^2 + \sigma_e^2)}$$

em que  $\sigma_a^2$  e  $\sigma_e^2$  são as variâncias aditiva e residual, respectivamente.  $(\sigma_a^2 + \sigma_e^2)$ , é a variância fenotípica ( $\sigma_p^2$ ). O progresso genético, ou ganho genético aditivo, foi posteriormente estimado para cada característica por meio dos valores genéticos das árvores individuais (Furlani et al., 2005).

A diversidade genética foi estimada, para cada característica, de acordo com Wei e Lindgren (1996):

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

em que  $D$  é a diversidade mantida após a seleção; o numerador  $\left[ (\sum k_f)^2 / \sum k_f^2 \right]$  é o número efetivo de progênies selecionadas;  $k_f$  é o número de indivíduos selecionados por progênie;  $N_p$  é o número original de progênies (trinta e nove). As estimativas de  $D$  variam dentro do intervalo  $0 < D \leq 1$ . Neste caso, um valor de diversidade genética próximo de 1 indica que quase toda a variabilidade da população de referência foi mantida, e uma estimativa próxima de zero indica que quase toda a variabilidade da população original foi extinta (Furlani et al., 2005).

Foram calculados os coeficientes de correlação de Pearson entre as características, usando as estimativas dos valores genéticos como medida de associação genética entre duas características. Intervalos de confiança Bootstrap com 95% de probabilidade para os valores de correlação de Pearson foram estimados, usando o procedimento SURVEYSELECT (SAS Institute, 1996) com 1000 amostras aleatórias e com substituição.

### 3. RESULTADOS E DISCUSSÃO

O florescimento precoce, o crescimento e a sobrevivência diferiram significativamente entre as procedências ( $p < 0,05$ ; Quadro 1). A magnitude das diferenças entre a melhor e a pior procedência variou entre 43%, na altura, 19% para o diâmetro, 9% para a sobrevivência e 82% para o florescimento precoce. Após um período de cinco anos, a média de sobrevivência das árvores foi alta (92,65%) e variou de 88% em Flinders Chase NP a 96% em Wirrabara SF. O florescimento precoce variou de 5% em Flinders Chase NP a 30% em Illapel-Chile, com média de 21%. Após cinco estações de cultivo, a altura das árvores variou de 1,98 m em Marble Range a 3,47 m em Wirrabara SF. De modo similar, a Marble Range evidenciou o menor crescimento de diâmetro (55 mm) e Wirrabara SF o maior (68 mm).

Quadro 1 - Médias populacionais para o crescimento (altura e diâmetro basal medidos à idade de 5 anos), florescimento precoce (medido à idade de 3 anos) e sobrevivência (medida depois de 5 anos de crescimento) avaliadas no sul do Deserto de Atacama em 39 famílias de meios-irmãos de *Eucalyptus cladocalyx*

População	Árvores avaliadas (número)	Famílias avaliadas (número)	Florescimento precoce (%)	Altura (m)	Diâmetro (mm)	Sobrevivência (%)
MT Remarkable *	480	16	25 a	3,15 b	62,22 b	94 ab
Marble Range *	120	4	23 a	1,98 c	55,01 c	90 ab
Wirrabara SF *	270	9	25 a	3,47 a	67,95 a	96 a
Flinders Chase NP *	240	8	05 b	3,27 ab	62,11 b	88 b
Illapel-Chile	60	2	30 a	3,03 b	58,87 bc	90 ab

\* Sementes provenientes do CSIRO Australian Tree Seed Centre. Valores nas colunas com a mesma letra indicam que as procedências não são diferentes significativamente a  $\alpha = 0,05$  para cada característica (intervalos de credibilidade da distribuição *a posteriori* das diferenças entre médias).

As estimativas dos componentes de variância e da herdabilidade das distribuições posteriores Bayesianas são mostradas no Quadro 2. As cadeias de Markov (amostragem de Gibbs) atingiram a convergência para todos os

parâmetros (componentes de variância e herdabilidade) empregando um total de 230.000 ciclos para cada análise. Os 30.000 ciclos iniciais foram descartados como *burn-in* ou período de inicialização (Van-Tassell et al., 1998). As distribuições *a posteriori* da herdabilidade podem ser visualizadas na Figura 1. A estimativa de média *a posteriori* da herdabilidade ( $h^2$ ) foi similar à estimativa da moda *a posteriori* para o modelo de limiar do florescimento precoce. No caso Bayesiano, a distribuição posterior dos parâmetros genéticos pode ser caracterizada pela média, mediana e moda; esses valores centrais são vistos como potenciais estimadores de ponto (Ruzzante et al., 2004), porém Wright et al. (2000) recomenda estimadores modais para descrever a distribuição posterior dos parâmetros. Esta é a região de alta densidade, ou seja, o valor que maximiza a função  $f(h^2|y)$ , ou o valor mais provável de  $h^2$  (Blasco, 2001). Para as características de crescimento (modelo linear) e sobrevivência as diferenças entre a média e a moda refletem falta de simetria nas distribuições posteriores estimadas (Van Tassell et al., 1998; ver Figura 1).

Quadro 2 - Componentes de variância (médias posteriores) e herdabilidade no sentido restrito para o florescimento precoce, sobrevivência e crescimento, avaliado em trinta e nove famílias de *Eucalyptus cladocalyx* no sul do Deserto de Atacama

Parâmetro	Florescimento Precoce	Altura	Diâmetro	Sobrevivência
Variância aditiva	0,13	0,19	36,76	0,02
Variância residual	0,12	0,50	218,31	0,05
Variância fenotípica	0,25	0,69	255,07	0,07
Herdabilidade				
Média	0,52	0,28	0,14	0,22
Mediana	0,52	0,27	0,13	0,21
Moda	0,52	0,25	0,10	0,18
Limite inferior	0,48	0,21	0,04	0,09
Limite superior	0,57	0,37	0,28	0,38

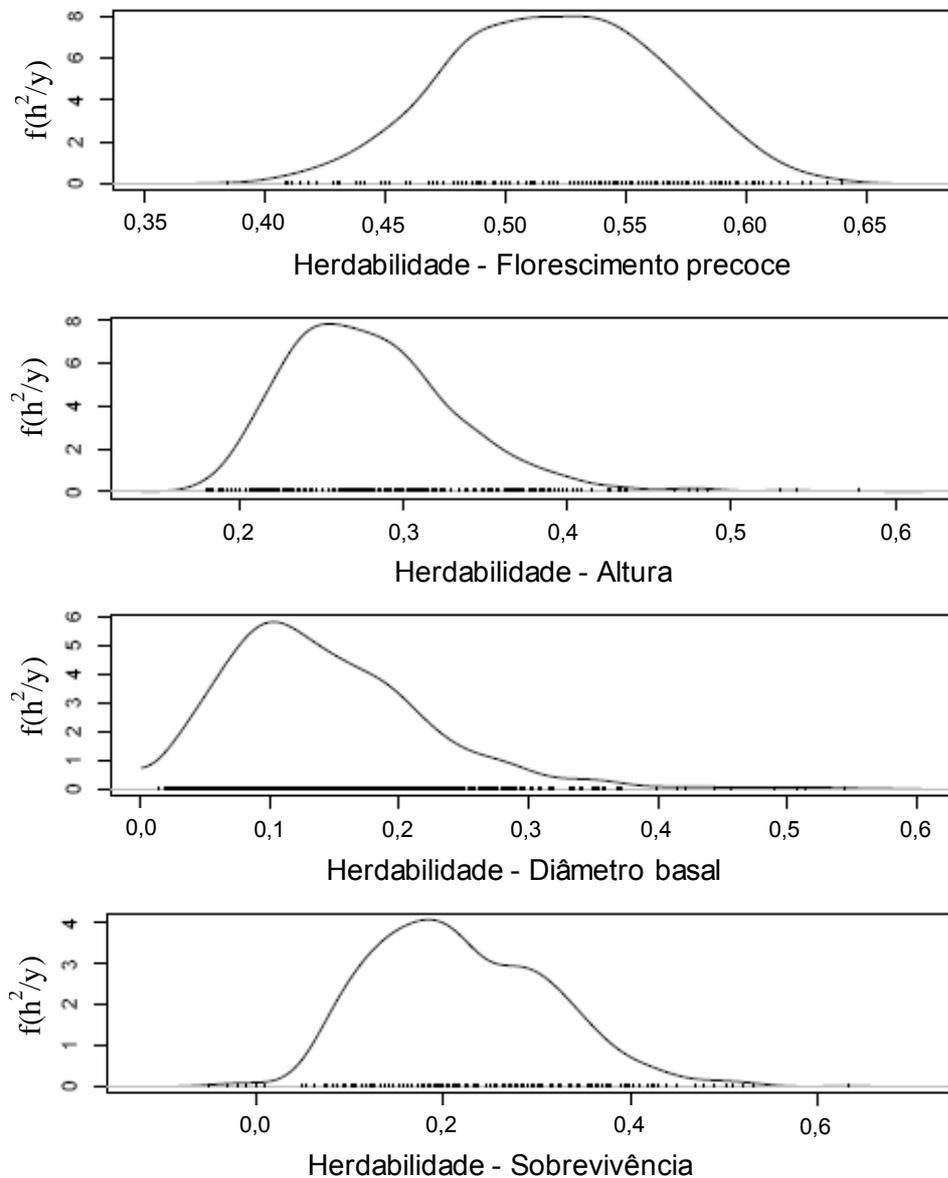


Figura 1 - Distribuição marginal *a posteriori* da herdabilidade ( $h^2$ ) para florescimento precoce (3 anos depois da plantação), altura, diâmetro basal e sobrevivência (medidos à idade de 5 anos), avaliados em 1.170 árvores individuais de *Eucalyptus cladocalyx* no sul do Deserto de Atacama.

Os resultados indicam que a precocidade é uma característica altamente herdável, sendo o maior valor da herdabilidade, com médias posteriores e intervalos de credibilidade de:  $h^2=0,52$  (0,48-0,57). A altura da árvore, o diâmetro basal e a sobrevivência da árvore tiveram valores de herdabilidade moderada, com médias posteriores e intervalos de credibilidade de  $h^2=0,28$  (0,21-0,37),  $h^2=0,14$  (0,04-0,28) e  $h^2=0,22$  (0,09-0,38), respectivamente. O intervalo de

credibilidade, ao nível de 95%, para a herdabilidade do florescimento precoce não incluíram o valor médio da herdabilidade para as características de crescimento e sobrevivência.

A predição dos efeitos genéticos aditivos para as árvores, individualmente, é dada na Figura 2. As linhas fornecem indicação da correlação entre as características, que confirma relação moderada entre o florescimento precoce e as características de crescimento. Essa associação moderada foi também confirmada pela correlação Pearson (ver Quadro 3) com médias e intervalos Bootstrap de 0,23 (0,18; 0,27) e 0,29 (0,25; 0,34). Adicionalmente, as correlações de Pearson entre os valores genéticos de florescimento precoce e de crescimento foram significativamente diferentes de zero ( $p > 0,05$ ). A associação entre o florescimento precoce e a sobrevivência foi menor, com valor médio e intervalo Bootstrap de 0,11 (0,06; 0,15).

Quadro 3 - Coeficientes de correlação de Pearson (e intervalos de confiança Bootstrap; 95%, N=1000 amostras) calculadas entre os valores genéticos das árvores de *Eucalyptus cladocalyx* (N=1170) provenientes de Austrália e Chile, as quais estão sendo avaliadas no sul do deserto de Atacama.

Características	Florescimento	Altura	Diâmetro
Altura	0,23 (0,18-0,27)	---	---
Diâmetro	0,29 (0,25-0,34)	0,49 (0,45-0,53)	---
Sobrevivência	0,11 (0,06-0,15)	0,09 (0,04-0,13)	0,28 (0,23-0,32)

A correlação genética entre altura da árvore e diâmetro basal foi alta, com média *a posteriori* e intervalo de credibilidade de 0,49 (0,45; 0,53) e as linhas de regressão (Figura 2) confirmam esse resultado. Outros estudos (Momen et al., 2004) têm encontrado similares resultados, indicando uma forte relação genética entre altura e diâmetro da árvore.

O florescimento precoce teve o maior ganho genético predito com valores de 41,22% e 36,15% para as intensidades de seleção de 7,48 e 14,29%, respectivamente (Quadro 4). O florescimento precoce e a altura da árvore apresentaram os maiores valores de diversidade genética (D), enquanto a sobrevivência da árvore apresentou o menor valor.

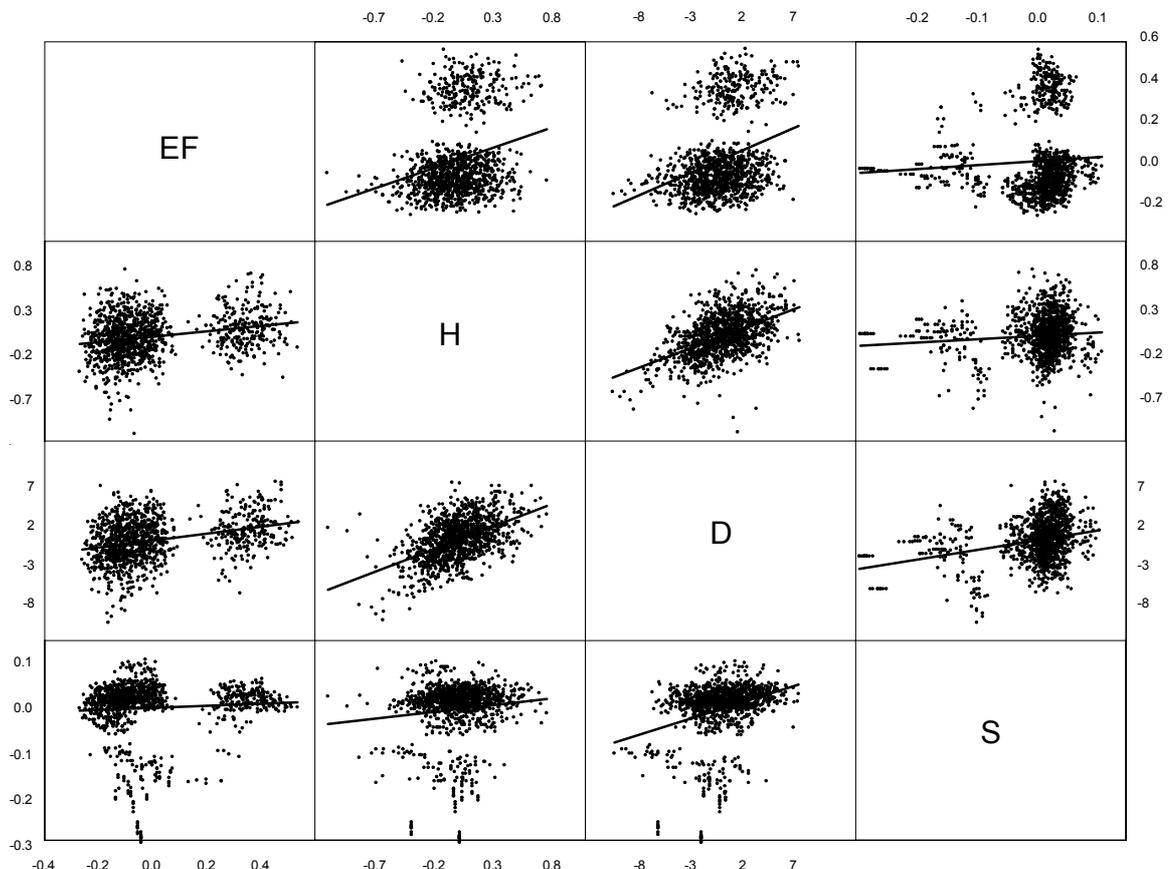


Figura 2 - Matrizes de dispersão dos valores genéticos das árvores para o florescimento precoce (EF), altura da árvore (H), diâmetro basal (D) e sobrevivência (S); as linhas indicam a associação entre características.

Os programas de plantações florestais freqüentemente buscam altas produtividades de biomassa combinando altas taxas de sobrevivência e crescimento (Chambers et al., 1997). Uma alta sobrevivência das árvores em um estágio de desenvolvimento precoce da plantação florestal deve, portanto, ser considerada uma característica relevante para os programas de seleção genética. Nas regiões áridas do Chile, o florescimento precoce também é desejável para os sistemas de produção de mel, cujo objetivo é suplementar o errático florescimento das plantas nativas. Neste estudo foi demonstrado que o florescimento precoce é uma característica altamente herdável, mas os negativos impactos sobre o crescimento das árvores devem ser considerados se as árvores são cultivadas para propósito múltiplo.

A sobrevivência das árvores é importante característica para os sistemas agro-florestais com déficit hídrico, e a alta taxa encontrada neste estudo confirma o potencial das árvores de *Eucalyptus cladocalyx* para crescer nos ambientes

áridos do sul do deserto de Atacama. Esta característica evidenciou a mais baixa magnitude das diferenças entre a melhor e a pior procedência, ao contrário que o florescimento precoce, que evidenciou as mais altas magnitudes. Condições ambientais no sul do deserto durante 2002 (segundo ano de experimentação) foram muito bons para a produtividade primária (Squeo et al., 2006). A precipitação média registrada nesse ano (394 mm) foi similar à precipitação média encontrada na Austrália (405 mm). Esse foi um ano associado ao evento do El Niño – Southern Oscillation (ENSO) (Montecinos e Aceituno 2003). Este efeito é especialmente importante em ecossistemas desérticos onde o recurso hídrico é o principal fator limitante para a produtividade da planta (Squeo et al., 2006).

Quadro 4 - Estimativas de ganhos genéticos preditos (GG%) e diversidade genética (GD) desde a seleção para o florescimento precoce, sobrevivência e características de crescimento, em trinta e nove famílias de *Eucalyptus cladocalyx* famílias (1.170 árvores individuais) avaliados no sul do deserto de Atacama durante os anos 2001 e 2006

Características	Intensidade de seleção					
	7,48%			14,29%		
	N <sub>SF</sub>	GG%	GD	N <sub>SF</sub>	GG%	GD
Florescimento	22	41,22	0,314	35	36,15	0,607
Altura	30	13,55	0,358	34	10,95	0,499
Diâmetro	17	7,52	0,200	25	6,14	0,338
Sobrevivência	11	6,38	0,169	23	5,41	0,287

N<sub>SF</sub>: número de famílias selecionadas.

Chambers et al. (1997) encontrou que o florescimento precoce nas árvores de *Eucalyptus globulus*, de quatro anos de idade, variou de 2 a 25%, dependendo da localidade, e mencionou que a idade na qual as árvores de eucalipto produzem flores varia marcadamente entre espécies e indivíduos dentro de espécies. Neste estudo, foram encontrados similares valores de florescimento (5% ou 30%, dependendo da população; ver Quadro 1), com um florescimento mais precoce (detectado em árvores de 3 anos de idade). O florescimento precoce seria vantajoso na região de Coquimbo, pois os agricultores locais têm

baixíssima produtividade (Ormazábal, 1991) e a produção de mel proveniente das árvores de *Eucalyptus cladocalyx* é uma fonte potencial de renda extra.

Neste estudo, a herdabilidade estimada ( $h^2$ ), ou o grau de controle genético aditivo, confirma uma significativa variação genética. O controle genético aditivo encontrado nestas populações permite a seleção de árvores para o estabelecimento de sistemas agro-florestais custo-efetivo. Adicionalmente, os eventos de florescimento precoce são particularmente importantes para os programas de melhoramento de árvores florestais e de reflorestamento, pois os indivíduos geneticamente superiores podem ser identificados e multiplicados mais rapidamente (Meilan, 1997).

A precocidade do florescimento tem sido incluída em outros importantes programas de melhoramento de *Eucalyptus* (Chambers et al., 1997; Wiltshire et al., 1998), porém, até o momento, é conhecido muito pouco acerca do controle genético desta característica em *E. cladocalyx*. No presente estudo, foi encontrado florescimento precoce sendo altamente herdável. Este resultado está de acordo com Chambers et al. (1997) onde o controle genético do florescimento precoce em *Eucalyptus globulus* ssp. *globulus* também foi altamente herdável ( $h^2=0,47$  ou  $0,59$ ; com média de 3 ou 4 ensaios, respectivamente). Um aspecto importante concluído no estudo de Chambers et al. (1997) foi que o florescimento precoce poderia ser considerado única característica nos seus ensaios, uma vez que foi encontrada uma não significativa interação genótipo x ambiente (Chambers et al., 1997).

Resultados conflitantes têm sido reportados sobre o impacto que a precocidade poderia ter nas características de crescimento. Chambers et al. (1997) encontrou que o florescimento precoce não foi correlacionado geneticamente com o diâmetro basal em *Eucalyptus globulus* (média de 0,04), e foi concluído que a seleção para unicamente esta característica teria pouco impacto sobre o crescimento precoce. Wiltshire et al. (1998), por sua vez, encontrou um moderado controle genético para o florescimento precoce, mas só uma relação fraca entre esta característica e o crescimento, dentro e entre populações de *Eucalyptus*. Williams et al. (2003) também encontrou que o crescimento precoce poderia afetar o crescimento e o desenvolvimento das árvores, embora, no seu estudo, o florescimento tenha sido artificialmente promovido. Os dados apresentados neste estudo confirmam que a seleção para

precocidade pode afetar significativamente tanto ao crescimento como a sobrevivência das árvores (Quadro 3).

A espécie *Eucalyptus cladocalyx* produz flores só quando as árvores atingem a maturidade, que é durante o desenvolvimento heteroblástico (Meilan, 1997). Portanto, provavelmente os eventos de florescimento detectados neste estudo, em plantas de 3 anos de idade, são um resultado de um estágio juvenil curto. Esta fase é importante porque o estágio juvenil é também a etapa do crescimento mais rápido. Não obstante, as características de crescimento foram medidas mais tarde, na metade da idade considerada como data de colheita (Mora, 2006). Portanto, uma moderada relação entre o florescimento precoce e o crescimento pode ser mantida no tempo porque a fase de crescimento rápido ocorre no estágio juvenil.

Para o florescimento precoce, o ganho genético e a manutenção dos altos níveis de diversidade são importantes para a produção de mel no sul do Deserto de Atacama, nas situações em que as espécies nativas não são encontradas em quantidades suficientes (Montenegro et al., 2003). Embora *Eucalyptus cladocalyx* tenha uma distribuição restrita ao sul da Austrália (Ruthrof et al., 2003) (Flinders Ranges, Eyre Peninsula e Kangaroo Island; Gleadow e Woodrow 2000; McDonald et al., 2003), a variabilidade genética encontrada no presente estudo pode ser empregada para atender programas de melhoramento em regiões similares onde as condições ambientais são fator limitante para o estabelecimento de outras espécies de eucaliptos.

#### 4. CONCLUSÕES

O florescimento precoce de *Eucalyptus cladocalyx* teve um alto grau de controle genético aditivo. As árvores precoces selecionadas na idade de 3 anos podem ter efeitos significativos sobre o crescimento e a sobrevivência das árvores. A informação gerada a respeito do controle genético do florescimento precoce pode ser utilizada por melhoristas de plantas no intuito de melhorar a produção de mel no sul do Deserto de Atacama, ou em situações onde as espécies nativas não são encontradas em quantidades suficientes. Árvores precoces de *Eucalyptus cladocalyx* têm muitas aplicações práticas nas zonas áridas do norte do Chile e o ganho genético confirmou potencial para os apicultores e agricultores que estão procurando árvores precoces com uma larga distribuição de flores durante a estação seca.

A inferência Bayesiana, juntamente com a técnica do algoritmo de Gibbs, forneceu informações práticas para propósitos de melhoramento e para entender como as características quantitativas são geneticamente controladas em *Eucalyptus cladocalyx*.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

AKHTER, J.; MAHMOOD, K.; TASNEEM, M.A.; MALIK, K.A.; NAQVI, M.H.; HUSSAIN, F.; SERRAJ, R. Water-use efficiency and carbon isotope discrimination of *Acacia ampliceps* and *Eucalyptus camaldulensis* at different soil moisture regimes under semi-arid conditions. **Biologia Plantarum**, 49:269-272, 2005.

AKILAN, K.; MARSHALL, J.K.; MORGAN, A.L.; FARRELL, R.C.C.; BELL, D.T. Restoration of catchment water balance: responses of clonal river red gum (*Eucalyptus camaldulensis*) to waterlogging. **Restoration Ecology**, 5:101-108, 1997.

BINK, M.C.A.M.; BOER, M.P.; TER BRAAK, C.J.F.; CANSEN, J.; VOORRIPS, R.E.; VAN DE WEG, W.E. Bayesian analysis of complex traits in pedigreed plant populations. **Euphytica**, 161:85-96, 2008.

BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, 79:2023–2046, 2001.

CAPPA, E.P.; CANTET, R.J.C. Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. **Canadian Journal of Forest Research**, 36:1276-1285, 2006.

CHAMBERS, P.G.S.; POTTS, B.M.; TILYARD, P.A. The genetic control of flowering precocity in *Eucalyptus globulus* ssp. *globulus*. **Silvae Genetica**, 46:207-214, 1997.

COHEN, Y.; ADAR, E.; DODY, A.; SCHILLER, G. Underground water use by *Eucalyptus* trees in an arid climate. **Trees, Structure and Function**, 11:356-362, 1997.

FURLANI, R.C.M.; MORAES, M.L.T.; RESENDE, M.D.V.; FURLANI-JUNIOR, E.; GONÇALVES P.S.; VALÉRIO-FILHO, W.V.; PAIVA, J.R. 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, 2005.

GLEADOW, R.M.; WOODROW, I.E. Polymorphism in cyanogenic glycoside content and cyanogenic beta-glucosidase activity in natural populations of *Eucalyptus cladocalyx*. **Australian Journal of Plant Physiology**, 27:693-699, 2000.

GLEADOW, R.M.; WOODROW, I.E. Defence chemistry of *Eucalyptus cladocalyx* seedlings is affected by water supply. **Tree Physiology**, 22: 939-945, 2002.

GONÇALVES-VIDIGAL, M.C.; MORA, F.; BIGNOTTO, T.S.; MUNHOZ, R.E.F.; SOUZA, L.D. Heritability of quantitative traits in segregating common bean families using a Bayesian approach. **Euphytica**, 164:551-560, 2008.

HAJARI, E.; WATT, M.P.; MYCOCK, D.J.; MCALISTER, B. Plant regeneration from induced callus of improved Eucalyptus clones. **South African Journal of Botany**, 72:195-201, 2006.

HANKS, L.M.; PAINE, T.D.; MILLAR, J.G.; HOM, J.L. Variation among Eucalyptus species in resistance to eucalyptus longhorned borer in Southern California. **Entomologia Experimentalis et Applicata**, 74:185-194, 1994.

HE, D.; BARR, C. China's pulp and paper sector: an analysis of supply-demand and medium term projections. **International Forestry Review**, 6:254-266, 2004.

HEIDELBERGER, P.; WELCH, P.D. Simulation run length control in the presence of an initial transient. **Operations Research**, 31:1109-1114, 1983.

MANIKIS, I.; THRASIVOULOU, A. The relation of physicochemical characteristics of honey and the crystallization sensitive parameters. **Apiacta**, 36:106-112, 2001.

MARCAR, N.E.; CRAWFORD, D.F.; LEPPERT, P.M.; JOVANOVIC, T.; FLOYD, R.; FARROW, R. **Trees for saltland: A guide to selecting native species for Australia**. Melbourne: CSIRO Publications, 1995. 72p.

MARCAR, N.E.; ZOHAR, Y.; GUO, J.; CRAWFORD, D.F. Effect of NaCl and high pH on seedling growth of 15 *Eucalyptus camaldulensis* Dehnh. provenances. **New Forests**, 23:193-206, 2002.

- MCDONALD, M.W.; RAWLINGS, M.; BUTCHER, P.A.; BELL, J.C. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). **Australian Journal of Botany**, 51:393-403, 2003.
- MEILAN, R. Floral induction in woody angiosperms. **New Forests**, 14:179-202, 1997.
- MELCHINGER, A.E.; UTZ, H.F.; SCHÖN, C.C. QTL analyses of complex traits with cross validation, bootstrapping and other biometric methods. **Euphytica**, 137:1–11, 2004.
- MOMEN, B.; ANDERSON, P.D.; SULLIVAN, J.H.; HELMS, J.A. A multivariate statistical approach to predicting mature tree performance based on seedling characteristics. **New Forest**, 27: 303-313, 2004.
- MONTECINOS, A.; ACEITUNO, P. Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies. **J Climate**, 16:281–296, 2003.
- MONTENEGRO, G.; PIZARRO, R.; ÁVILA, G.; CASTRO, R.; RÍOS, C.; MUÑOZ, O.; BAS, F.; GÓMEZ, M. Origen botánico y propiedades químicas de las mieles de la región mediterránea árida de Chile. **Ciencia e Investigación Agraria**, 30:161-174, 2003.
- MORA, F. 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, 2006.
- MORA, F.; TAPIA, F.; SCAPIM, C.A.; MARTINS, E.N.; PINTO, R.J.B.; IBACACHE, A. Early performance of *Olea europaea* cv. Arbequina, Picual and Frantoio in southern Atacama Desert. **Crop Breeding and Applied Biotechnology**, 8:30-38, 2008.
- MORA, F.; TAPIA, F.; SCAPIM, C.A.; MARTINS, E.N. Vegetative growth and early production of six olive cultivars, in southern Atacama desert, Chile. **Journal of Central European Agriculture**, 8:269-276, 2007.

ORMAZÁBAL, C.S. Silvopastoral systems in arid and semiarid zones of northern Chile. **Agroforestry Systems**, 14:207-217, 1991.

RAWAT, J.S.; BANERJEE, S.P. The influence of salinity on growth, biomass production and photosynthesis of *Eucalyptus camaldulensis* Dehnh. and *Dalbergia sissoo* Roxb. seedlings. **Plant and Soil**, 205:163-169, 1998.

RUTHROF, K.X.; LONERAGAN, W.A.; YATES, C.J. Comparative population dynamics of *Eucalyptus cladocalyx* in its native habitat and as an invasive species in an urban bushland in south-western Australia. **Diversity & Distributions**, 9:469-483, 2003.

RUZZANTE, D.E.; HANSEN, M.M.; MELDRUP, D.; EBERT, K.M. Stocking impact and migration pattern in an anadromous brown trout (*Salmo trutta*) complex: where have all the stocked spawning sea trout gone? **Molecular Ecology**, 13:1433-1445, 2004.

SAS-INSTITUTE. **Statistical analysis system user's guide**. Cary: SAS Institute, 1996. 956p.

SORYA, F.; BASURCO, F.; TOVAL, G.; SILIÓ, L.; RODRÍGUEZ, M.C.; TORO, M. An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. **Canadian Journal of Forest Research**, 28:1286-1294, 1998.

SQUEO, F.A.; TRACOL, Y.; LÓPEZ, D.; GUTIÉRREZ, J.R.; CORDOVA, A.M.; EHLERINGER, J.R. ENSO effects on primary productivity in Southern Atacama desert. **Advances in Geosciences**, 6:273-277, 2006.

SPIEGELHALTER, D.J.; THOMAS, A.; BEST, N.G.; LUNN, D. **WinBugs Version 1.4 Users Manual**. Cambridge: MRC Biostatistics Unit, 2003.

WALDMANN, P.; ERICSSON, T. Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. **Theor. Appl. Genet.**, 112:1441-1451, 2006.

- WEI, R.P.; LINDGREN, D. Effective family number following selection with restrictions. **Biometrics**, 52:198-208, 1996.
- WILDY, D.T.; PATE, J.S.; BARTLE, J.R. Budgets of water use by *Eucalyptus kochii* tree belts in the semi-arid wheatbelt of Western Australia. **Plant and Soil**, 262:129-149, 2004.
- WILLIAMS, D.R.; POTTS, B.M.; SMETHURST, P.J. Promotion of flowering in *Eucalyptus nitens* by paclobutrazol was enhanced by nitrogen fertilizer. **Canadian Journal of Forest Research**, 33:74-8, 2003.
- WILLIAMS, W.D. Dryland lakes. **Lakes and Reservoirs: Research and Management**, 5:207-212, 2000.
- WILTSHIRE, R.J.E.; REID, J.B.; POTTS, B.M. Genetic control of reproductive and vegetative phase change in the *Eucalyptus risdonii*-*E. tenuiramis* complex. **Australian journal of botany**, 46:45-63, 1998.
- WOODROW, I.E.; SLOCUM, D.; GLEADOW, R.M. Influence of water stress on cyanogenic capacity in *Eucalyptus cladocalyx*. **Australian Journal of Plant Physiology**, 29:103-110, 2002.
- WRIGHT, D.R.; STERN, H.S.; BERGER, P.J. Comparing traditional and Bayesian analyses of selection experiments in animal breeding. **Journal of Agricultural, Biological, and Environmental Statistics**, 5:240-256, 2000.
- 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. **Journal of Animal Science**, 74:2586-2597, 1996.
- VAN-TASSELL, C.P.; VAN-VLECK, L.D.; GREGORY, K.E. Bayesian analysis of twinning and ovulation rates using a multiple-trait threshold model and Gibbs sampling. **Journal of Animal Science**, 76:2048-2061, 1998.
- YUEMING, J.; LIHU, Y.; SINGANUSONG, R.; DATTA, N.; RAYMONT, K. Phenolic acids and abscisic acid in Australian Eucalyptus honeys and their potential for floral authentication. **Food Chemistry**, 86:169-177, 2004.

## 6. CONCLUSÕES GERAIS

### 6.1. Modelos Lineares Generalizados

O presente estudo confirmou a efetividade e a eficiente aplicabilidade dos Modelos Lineares Generalizados em experimentos relacionados ao melhoramento vegetal. Por meio desta técnica foi possível: 1. O mapeamento de QTLs que controlam características binárias, realizada em características medidas repetidamente sobre a mesma unidade experimental (análise longitudinal), 2. O mapeamento de QTLs que controlam características contínuas com viés da distribuição normal, e 3. Avaliação de experimentos com cultivares de oliveira para variáveis repostas com distribuição Gama (produção de frutos), e Binomial (produção precoce e sobrevivência). Confirma-se que a informação acerca da distribuição dos dados agronômicos deve ser considerada nos programas de melhoramento de plantas, para melhorar a confiabilidade das inferências dos resultados dos experimentos de avaliação genética.

### 6.2. Contexto Bayesiano

A abordagem Bayesiana utilizada como ferramenta científica de análise estatística foi útil para propósitos de melhoramento e para entender como as características são controladas geneticamente. O método Bayesiano permitiu a avaliação genética de genótipos em diferentes programas de melhoramento, como oliveira, acácia e eucalipto, e baseado em diferentes situações de pressupostos de análise. Características binárias, como a sobrevivência das plantas, precocidade da produção, e florescimento precoce, também são possíveis de incluir no programas de melhoramento usando o conceito do modelo de limiar (*threshold*) e os diferentes métodos de Monte Carlo, via Cadeias de Markov (MCMC). A seleção baseada em características econômicas e a habilidade da inferência Bayesiana para prever os valores genéticos de clones, procedências, famílias e plantas individuais (modelo animal) são suficientes para os programas de avaliação genética de plantas.

## APÊNDICE

Rotinas do SAS consideradas nas diferentes análises do estudo.

Exemplo de rotinas para a seleção de co-fatores no mapeamento de QTLs (Intervalo Composto) nas linhas clássica e generalizadas.

```
PROC STEPWISE DATA=QTL;
MODEL Q=M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 / SLENTRY=0.05
SLSTAY=0.05 STEPWISE DETAILS;
RUN;
PROC GENMOD DATA=QTL;
MODEL Q=M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 / TYPE1
DISTRIBUTION=GAMMA LINK=LOG;
RUN; /*utilizar o quadro das estimativas dos parâmetros e suas significâncias,
incluindo intervalos de confiança de WALD */
/*LINK=LOG não é a canônica para D=GAMMA, mas é preferida devido a sua
interpretação e simplicidade matemática*/
```

Exemplo de rotina para a abordagem Bayesiana

```
/*Distribuições a priori informativas*/
DATA PRIOR;
INPUT DENSITY TYPE $ PARM1 PARM2;
CARDS;
1 IG 18.336 16251
2 IG 508.02 124516
; /*Data prior é próprio para cada conjunto de dados, o qual é obtido por meio do
quadro de densidades base, gerado a partir de uma análise prévia usando
JEFFREYS*/
PROC MIXED DATA=EUCA;
CLASS FAM BLOCK;
MODEL D = BLOCK;
RANDOM FAM / S;
PRIOR DATA=PRIOR / OUT=BAYES2 NSAMPLE=50000;
RUN;
PROC UNIVARIATE DATA=BAYES2; WHERE SAMPLE >1000;
VAR COVP1 COVP2 H GAM1 GAM2 GAM3 GAM4 GAM5 GAM6 GAM7 GAM8
GAM9 GAM10 GAM11 GAM12 GAM13 GAM14 GAM15 GAM16 GAM17 GAM18
GAM19 GAM20 GAM21 GAM22 GAM23 GAM24 GAM25 GAM26 GAM27 GAM28
GAM29 GAM30 GAM31 GAM32 GAM33 GAM34 GAM35 GAM36 GAM37 GAM38
GAM39;
RUN;

/* Distribuições a priori não informativas */
PROC MIXED DATA=EUCA;
CLASS FAM BLOCK;
MODEL D = BLOCK;
```

```

RANDOM FAM / S;
PRIOR JEFFREYS / OUT=BAYES2 NSAMPLE=50000;
RUN;
PROC UNIVARIATE DATA=BAYES2; WHERE SAMPLE >1000;
VAR COVP1 COVP2 H GAM1 GAM2 GAM3 GAM4 GAM5 GAM6 GAM7 GAM8
GAM9 GAM10 GAM11 GAM12 GAM13 GAM14 GAM15 GAM16 GAM17 GAM18
GAM19 GAM20 GAM21 GAM22 GAM23 GAM24 GAM25 GAM26 GAM27 GAM28
GAM29 GAM30 GAM31 GAM32 GAM33 GAM34 GAM35 GAM36 GAM37 GAM38
GAM39;
RUN;

```

```

/*Abordagem Bayesiana usando um Modelo Linear Generalizado Misto*/
%GLIMMIX(DATA=VER, PROCOPT=METHOD=REML, STMTS=%STR(
  CLASS FAM BLOCK;
  MODEL D = BLOCK;
  RANDOM FAM / SOLUTION;
  PRIOR / OUT=BAYES3 NSAMPLE=50000;),
  ERROR=BINOMIAL,
  LINK=LOGIT);

```

```

RUN; /*precisa rodar previamente a Macro, obtida do site do SAS*/
PROC UNIVARIATE DATA=BAYES3; WHERE SAMPLE >1000;
VAR COVP1 COVP2 H GAM1 GAM2 GAM3 GAM4 GAM5 GAM6 GAM7 GAM8
GAM9 GAM10 GAM11 GAM12 GAM13 GAM14 GAM15 GAM16 GAM17 GAM18
GAM19 GAM20 GAM21 GAM22 GAM23 GAM24 GAM25 GAM26 GAM27 GAM28
GAM29 GAM30 GAM31 GAM32 GAM33 GAM34 GAM35 GAM36 GAM37 GAM38
GAM39;
RUN;

```

Exemplo de rotina para intervalos de confiança Bootstrap para os coeficientes de correlação de Spearman.

```

PROC SURVEYSELECT DATA=BVALUES METHOD=URS REP=1000
  N=200 OUT=BOOT SEED=999155 OUTHITS;
PROC CORR DATA=BOOT NOPRINT OUTS=SPEARMAN;
VAR Q IS;
BY REPLICATE;
DATA ALLMEANS;
SET SPEARMAN;
IF _TYPE_ = 'CORR' AND _NAME_ = 'IS' THEN OUTPUT;
PROC UNIVARIATE DATA=ALLMEANS CIBASIC;
VAR Q;
RUN;
PROC CORR Spearman DATA=M_I;
VAR Q IS;
RUN;

```

Exemplo de rotina para o mapeamento de QTL por intervalo composto

```

/*a rotina considera apenas uma posição do genoma*/
PROC GLM;

```

```

MODEL C = ADITIVO DOMINANCIA COFATOR;
RUN;
PROC MIXED METHOD=ML;
MODEL C = ADITIVO DOMINANCIA COFATOR;
RUN;
PROC GENMOD;
MODEL C = ADITIVO DOMINANCIA COFATOR / D=GAMMA LINK=LOG TYPE1;
RUN;
/*para obter resultados sem efeitos*/
PROC MIXED METHOD=ML;
MODEL C = ;
RUN;
PROC GENMOD;
MODEL C = / D=GAMMA LINK=LOG TYPE1;
RUN;

```

Exemplo de rotina para o mapeamento de QTL (intervalos simples) para medidas repetidas e dados de distribuição binomial (Equações de estimação Generalizadas).

/\*a rotina considera apenas uma posição do genoma e alguns exemplos de estruturas de (co) variâncias\*/

```

PROC GENMOD DESC;
CLASS SUBJ;
MODEL BIN = AD DM / D=B TYPE3;
REPEATED SUBJECT=SUBJ / TYPE=AR CORRW;
RUN;
PROC GENMOD DESC;
CLASS SUBJ;
MODEL BIN = AD DM / D=B TYPE3;
REPEATED SUBJECT=SUBJ / TYPE=EXCH CORRW;
RUN;
PROC GENMOD DESC;
CLASS SUBJ;
MODEL BIN = AD DM / D=B TYPE3;
REPEATED SUBJECT=SUBJ / TYPE=UN CORRW;
RUN;
PROC GENMOD DESC;
CLASS SUBJ;
MODEL BIN = AD DM / D=B TYPE3;
REPEATED SUBJECT=SUBJ / TYPE=IND CORRW;
RUN;

```

Exemplo de rotina para o mapeamento de QTL (intervalos simples) para medidas repetidas e distribuição normal.

/\*a rotina considera apenas uma posição do genoma e alguns exemplos de estruturas de (co) variâncias\*/

```

PROC MIXED;
CLASS SUBJ;

```

```

MODEL C = AD DM /S;
REPEATED / SUBJECT=SUBJ TYPE=VC R RCORR;
RUN;
PROC MIXED;
CLASS SUBJ;
MODEL C = AD DM /S;
REPEATED / SUBJECT=SUBJ TYPE=CS R RCORR;
RUN;
PROC MIXED;
CLASS SUBJ;
MODEL C = AD DM /S;
REPEATED / SUBJECT=SUBJ TYPE=UN R RCORR;
RUN;
PROC MIXED;
CLASS SUBJ;
MODEL C = AD DM /S;
REPEATED / SUBJECT=SUBJ TYPE=UN(1) R RCORR;
RUN;
PROC MIXED;
CLASS SUBJ;
MODEL C = AD DM /S;
REPEATED / SUBJECT=SUBJ TYPE=TOEP R RCORR;
RUN;
PROC MIXED;
CLASS SUBJ;
MODEL C = AD DM /S;
REPEATED / SUBJECT=SUBJ TYPE=TOEPH R RCORR;
RUN;

```