

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Arquitetura genética de componentes periódicos de crescimento de *Hevea
brasiliensis***

Rafael Tassinari Resende

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2014**

Rafael Tassinari Resende
Engenheiro Florestal

**Arquitetura genética de componentes periódicos de crescimento de *Hevea
brasiliensis***

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Genética e
Melhoramento de Plantas

**Piracicaba
2014**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Resende, Rafael Tassinari

Arquitetura genética de componentes periódicos de crescimento de *Hevea brasiliensis* /
Rafael Tassinari Resende. - - versão revisada de acordo com a resolução CoPGr 6018 de
2011. - - Piracicaba, 2014.
111 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.
Bibliografia.

1. Modelo multi-caracteres-multi-períodos 2. Mapeamento de QTLs 3. Equações
estruturais I. Título

CDD 633.895
R433a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Aos meus pais, Cezar e Auxiliadora, e à minha tia Adélia,
dedico.

“Sou tudo, por ter a convicção de ser nada
Sou a onda no mar, sou o pó da estrada
Vivo, nos meus sonhos, a grande ilusão
Ora sendo acalanto, ora um abraço.”

Tio Venerando Luiz Tassinari

(Do poema Tudo e Nada - 09/06/2010)

“Ethical axioms are found and tested not very differently from the axioms of science.
Truth is what stands the test of experience.”

Albert Einstein

“Don’t gain the world and lose your soul
Wisdom is better than silver and gold.”

Bob Marley

AGRADECIMENTOS

Em ordem: (1) Cronológica; (2) Alfabética:

A DEUS.

Aos meus maravilhosos pais, Cezar e Auxiliadora, à minha adorável tia Adélia e aos meus irmãos Léo e Leandro.

À Ana, que despertou minha paixão pela genética.

Ao meu orientador Augusto, que foi mais que um chefe, foi um exemplo como profissional e pessoa, a quem muito admiro.

Aos meus colegas de laboratório, e também àqueles que passaram por lá nesses dois anos: Rodrigo, Marcelo, Maria Marta, Luciano, Jane, Renato, Adriana, Guilherme, Carina, João, Thiago, Luis Felipe, Letícia Lara, Marianella, Gabriel, Amanda, e os meninos da iniciação científica: Kateto, Izabel e Letícia. Pelos ótimos momentos, trabalhos, risadas e todo aprendizado.

Em especial: à Adriana e ao Guilherme, pelos estudos nas disciplinas; ao Renato pelas dicas logo quando cheguei; ao Guilherme pelo apoio no início e pelo convívio na república; ao João, Carina e Adriana que são pessoas admiráveis, com quem aprendi muito e muito me incentivaram; ao Marcelo pelas conversas, toda ajuda e esclarecimentos; à Jane, uma figura fantástica que muito me ajudou e motivou; ao Rodrigo, que foi um segundo orientador pra mim, pela paciência e amparo no meu trabalho; à Jane e ao Gabriel pela ajuda com modelos mistos; à Adriana pelos estudos em *SEM*; e à Letícia Lara, uma grande companheira e amiga.

Aos meus colegas de república Lesma-Lerda-pg, àqueles que passaram por lá e agregados: Guilherme, Mateuzim, Thiaguim, Vanessa, Yuri, Naiara, Vítor, Hugo, Lucas, Cabelo, Gustavo, Alex, Luis Felipe, Ju, Dona Elza e o cão Negão.

Aos professores da ESALQ, que tanto contribuíram para minha formação.

À professora Anete e à Lívia da Unicamp, por terem gentilmente cedido os dados, em especial à Lívia ter sido sempre solícita no esclarecimento das dúvidas.

Ao Elias Chaibub Neto da *Sage Bionetworks*, pelas sugestões e considerações no trabalho.

Aos funcionários da *Michelin LTDA*, em especial ao Saulo Cardoso por esclarecimentos sobre o experimento.

Aos funcionários da genética Marcão, Berdan, Sr. Antônio, Waldir, Léia, Sr. Natálio, as moças da limpeza e aos demais colegas de trabalho e amigos que fiz durante o mestrado.

Aos amigos: Yuri, Thiagão, Gentil, Victor e Gabriel, pela companhia e boas conversas.

SUMÁRIO

RESUMO	11
ABSTRACT	13
1 INTRODUÇÃO	15
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Importância e Melhoramento Genético da Seringueira	19
2.2 Modelos Mistos	20
2.2.1 Modelo Misto para Análise de Dados Longitudinais	24
2.2.2 Estruturas Alternativas de Matriz de Variâncias e Covariâncias para \mathbf{G}^P e \mathbf{R}_0	25
2.2.3 Modelos para Múltiplos Caracteres em Múltiplos Períodos	28
2.3 Seleção de Modelos	29
2.4 Mapeamento de QTLs	31
2.4.1 Métodos de Mapeamento de QTLs	32
2.5 Análise de Equações Estruturais (<i>SEM</i>)	34
2.5.1 Diagramas Causais e Terminologias	36
2.5.2 O Método	37
2.5.3 Modelos Recursivos	43
2.5.4 Modelos Estruturais Generalizados	44
2.5.5 <i>SEM</i> no Cenário de Mapeamento de QTLs	47
3 MATERIAL E MÉTODOS	51
3.1 Material	51
3.1.1 Descrição dos Dados Experimentais	51
3.1.2 Descrição dos Dados Moleculares e Grupos de Ligação	53
3.2 Métodos	53
3.2.1 Modelo Misto: Dados Fenotípicos	53
3.2.2 Modelo para Características Individuais	54
3.2.3 Modelo Multi-Caracteres-Multi-Períodos	57
3.3 Mapeamento dos QTLs	59
3.4 Modelos de Equações Estruturais	60
3.4.1 Modelo Estrutural Fenotípico	61
3.4.2 Modelo Estrutural Genético	62
4 RESULTADOS	65
4.1 Análise Exploratória dos Dados Experimentais	65

4.2	Modelos de Ajuste dos Dados Fenotípicos	67
4.2.1	Modelo para Caracteres Individuais	68
4.2.2	Modelo Multi-Caracteres-Multi-Períodos	72
4.3	Análise Fenotípica	73
4.3.1	Modelo Estrutural Fenotípico	73
4.4	Análise Genotípica	78
4.4.1	Mapeamento dos QTLs	78
4.4.2	Modelo Estrutural Genético	81
5	DISCUSSÃO	87
6	CONCLUSÃO	97
	REFERÊNCIAS	99

RESUMO

Arquitetura genética de componentes periódicos de crescimento de *Hevea brasiliensis*

Nas metodologias de mapeamento de QTLs tradicionais, a relação de causalidade entre os caracteres fenotípicos e QTLs normalmente não são consideradas. O desenvolvimento deste trabalho contou com a utilização de dados longitudinais de crescimento de progênies oriundas do cruzamento entre os parentais PB217 e PR255 de um plantio de seringueira, localizado em uma área com dois períodos bem definidos ao longo do ano (altas e médias temperaturas; altas e baixas taxas precipitação). O experimento contém 4 medidas de incremento em diâmetro e altura, que são componentes periódicos do crescimento total da cultura, mensurados em um intervalo de dois anos (entre os 18 aos 52 meses de idade das plantas), sendo dois períodos em estação climática favorável ao desenvolvimento e dois em estação desfavorável, intercalados. Dessa forma foram estudados os parâmetros de relacionamento fenotípico e genético com objetivo de construir um diagrama de arquitetura genética que pondere relações de causalidade. Para modelar os dados fenotípicos foi realizado um elaborado modelo multi-caracteres que contemplou a variação espacial das parcelas experimentais e a variação entre os períodos de medição. Para tanto, foram ajustadas matrizes de variância-covariância (VCOV) adequadas à realidade dos dados, e incorporados dados meteorológicos que descrevessem cada um dos períodos. A partir destes modelos, os valores genotípicos ajustados foram utilizados na detecção dos QTLs. Posteriormente, fenótipos e genótipos foram articulados em um diagrama causal estrutural capaz de inferir sobre padrões genéticos de comportamento de crescimento da cultura. Foram mapeados um total de 13 QTLs, sendo que dois deles foram coincidentes para componentes periódicos de diâmetro nos períodos de estação desfavorável. Foi possível identificar efeitos aditivos e devido à dominância interessantes para o desenvolvimento em períodos de menores temperaturas, apontar o parental PR255 como portador de alelos importantes no desenvolvimento em clima adverso, estimar efeitos indiretos de QTLs não mapeados para determinadas características e explicar o padrão comportamental de crescimento no período em que as progênies foram avaliadas. Esta abordagem demonstrou-se proficiente para utilização em programas melhorando genético assistido por marcadores, por agregar informações pertinentes à seleção dos melhores materiais genéticos.

Palavras-chave: Modelo multi-caracteres-multi-períodos; Mapeamento de QTLs; Equações estruturais

ABSTRACT

Genetic architecture of periodic growth components of *Hevea brasiliensis*

In traditional methodologies of QTL mapping, the causal relationship between phenotypic characters and QTLs are usually not considered. The development of this work involved the use of longitudinal growth data of progenies from parental PR255 and PB217 of a rubber tree plantation, located in an area with two periods of high and medium temperature and low and high precipitation rates well defined throughout the year. The experiment contains four measures of increment of diameter and height, which are periodic growth components of the total crop growing at an interval of two years (from 18 to 52 months old plants), two periods in a favorable climate station and two in a adverse station, intercalated. Was studied the parameters of phenotypic and genetic relationships in order to construct a diagram of genetic architecture to examine these causal relationships. A multi-trait-multi-occasion model that take into consideration spatial variation and climatic variation was developed. It also contains a variation-covariation matrix with appropriate to the reality of the data were adjusted and incorporated meteorological data was conducted to describe each of the periods. From these models the adjusted genotypic values were used in the detection of QTLs and later phenotypes and genotypes were linked in a structural causal diagram to infer about the genetic patterns of behavior. A total of 13 QTLs were mapped to the periodic growth components and total growth. The genetic architecture was able to identify additive effects and effects due dominance interesting to the development in periods of lower temperatures and drought, pointing parental PR255 as carrier of important alleles to development in adverse weather, estimating indirect effects of QTLs that were not mapped to certain characteristics and explain the physiological behavior pattern of growth in the period in which progenies were evaluated. This approach proved to be proficient to use in breeding programs aiming to implement marker assisted selection.

Keywords: Multi-trait-multi-occasion model; QTL mapping; Structural equations modeling

1 INTRODUÇÃO

Programas de melhoramento genético de espécies florestais são baseados em seleção recorrente (simples ou recíproca), hibridação interespecífica e, em muitos casos, seleção clonal (FONSECA et al., 2010). Para aprimorar as estratégias de melhoramento genético florestal é interessante explorar abordagens alternativas para a construção dos mapas de ligação, e posteriormente incorporar novas técnicas às metodologias tradicionais. O mapeamento de QTLs (“*Quantitative Trait Loci*”) é uma ferramenta muito poderosa para o processo de seleção assistida por marcadores, dado o conhecimento prévio acurado dos padrões de segregação e recombinação dos alelos relacionados às características de importância agrônômica, e seu comportamento perante as alterações ambientais e interações intraespecíficas (COLLARD et al., 2005).

Entretanto, as abordagens tradicionais de mapeamento de QTLs ignoram o relacionamento causal presente na relação entre fenótipos e genótipos (ROSA et al., 2011). A construção de redes de causalidades entre estes dois eixos da genética clássica permite ao pesquisador a compreensão de novos parâmetros de interação entre genes e a manifestação das características, gerando suporte na inferência de efeitos, sentidos e magnitudes de interações, componentes fundamentais na elaboração e entendimento da arquitetura genética dos caracteres estudados (CHAIBUB-NETO et al., 2008).

Os valores fenotípicos de diferentes características são muitas vezes correlacionados entre si, como por exemplo: o número de vagens e a produção de grãos da soja (ARSHAD; ALI; GHAFOR, 2006); intervalo entre florescimentos e tolerância à seca em milho (RIBAUT et al., 1996); e a altura e o diâmetro de árvores (CURTIS, 1967). Fatores genéticos e ambientais são as causas fundamentais para essa correlação. Dessa maneira, podemos decompor a correlação fenotípica em correlação genética e correlação ambiental, sendo que as correlações genéticas podem ser oriundas tanto de causas pleiotrópicas (genes com influência em mais de uma característica), como pelo desequilíbrio de fase gamética entre genes que afetam diferentes características. Por conseguinte, o entendimento das causas de correlações é muito importante para o melhoramento.

Quando lidamos com uma progênie constituída por irmãos completos, gerada por parentais com estado de homozigose desconhecido, o padrão de segregação varia muito se comparado com as clássicas populações de autofecundação (F2), retrocruzamento ou populações originadas de linhagens endogâmicas recombinantes (“*recombinant inbred lines*” ou RILs). A segregação presente em progênies desse tipo de cruzamento pode aparecer de diversas formas, previstas

pelo perfil alélico dos locos presentes nos parentais. Por exemplo, se examinarmos um loco dentro de uma população diploide que possui três alelos distintos, sendo dois deles codominantes entre si e dominantes sobre o terceiro, gametas com três tipos de alelos serão formados (respectivamente: A , B e o) e estes darão origem a três genótipos em homozigose (AA , BB e oo) e três em heterozigose (AB , Ao e Bo). Se ratificarmos que esse gene está presente no grupo de genes responsáveis por uma característica quantitativa, que a contribuição de seus genótipos para fenótipo seja incomensurável e eles sejam cruzados aleatoriamente, seriam possíveis 21 cruzamentos diferentes e a progênie F1 segregante apresentaria um cenário de genótipos com diferentes padrões de segregação (WU et al., 2002).

Por isso, em análises de ligação de espécies de melhoramento baseadas na F1 segregante (“*outcrossing*”), como as seringueiras, é importante a utilização de marcadores moleculares codominantes e multialélicos. Os microssatélites, que são repetições de sequencias simples (“*Simple Sequence Repeats ou SSR*”), consistem de múltiplas cópias repetidas em tandem (subsequentes e ininterruptas) de motivos de mono, bi, tri ou tetranucleotídeos que fornecem informações ideais na caracterização do polimorfismo presente nesses tipos de cruzamentos (WU et al., 2002).

Os caracteres quantitativos são altamente influenciados pelo ambiente, possuem baixa herdabilidade e por serem formados por muitos genes, interações epistáticas podem estar presentes na manifestação de seus fenótipos (LYNCH; WALSH et al., 1998). As características de crescimento, como diâmetro do fuste e a altura da árvore, tipicamente são de natureza quantitativa e apresentam respostas positivas ou estáveis de incremento ao longo do tempo de acordo com os benefícios e restrições impostas pelo meio. Via de regra, condições de umidade, temperatura e a presença de agentes causadores de doenças são de grande importância no desenvolvimento dos vegetais (TAIZ; ZEIGER, 2010), e essas exigências específicas variam conforme a evolução adaptação das espécies. Em particular as plantas de origem tropical, por estarem estabelecidas em locais com umidade abundante e temperaturas mais elevadas durante todo ano, se submetidas à outras circunstâncias ambientais, certamente manifestarão respostas depreciadas de crescimento, mas com algum nível de tolerância às novas condições.

Muito já foi feito a respeito da identificação e contabilização dos genes quantitativos que correspondem às características de interesse agrônomo. Porém, as abordagens estatísticas tradicionais de mapeamento de QTLs negligenciam os estágios de desenvolvimento da formação característica. Uma abordagem aproximada para detectar os efeitos genéticos dependentes do tempo para essas características dinâmicas tem sido associar marcadores com os diferentes fenótipos ao longo do tempo ou etapas de desenvolvimento e comparar essas diferenças entre esses

estádios.

Embora as bases das resistências a doenças sejam relativamente bem estudadas na seringueira (LESPINASSE et al., 2000; GUEN et al., 2003, 2011), pouco é conhecido sobre as bases genéticas da tolerância ao frio e ao déficit de água sobre o desenvolvimento de características relacionadas ao crescimento nessa cultura. Souza et al. (2013) realizaram um importante trabalho, desenvolvendo um mapa de ligação e mapeamento de QTLs para caracteres de crescimento, considerando duas estações climáticas de verão e inverno.

Este trabalho teve por objetivo estudar a arquitetura genética de caracteres de crescimento relacionados à tolerância à menores temperaturas e à seca da espécie *Hevea brasiliensis*. Para tanto foram mapeadas regiões cromossômicas que representem respostas significativas ao maior crescimento da seringueira em períodos de condições climáticas favoráveis e desfavoráveis. Com a utilização de modelos estruturais (ROSA et al., 2011) foram calculados coeficientes de correlações e de causalidades entre os diferentes componentes periódicos de crescimento das progênies e os períodos/estações definidas com intuito de agregar informações pertinentes aos programas de melhoramento.

2 REVISÃO BIBLIOGRÁFICA

2.1 Importância e Melhoramento Genético da Seringueira

Hevea brasiliensis é uma espécie da família das Euphorbiaceae que apresenta grande relevância na produção do látex utilizado na fabricação da borracha. O gênero *Hevea* é nativo da região Amazônica e possui 11 espécies em sua constituição taxonômica (PIRES; SECCO; IVANIR, 2002). As seringueiras são árvores perenes, de vida econômica útil de 30 a 35 anos, e, como uma espécie de origem tropical, possuem crescimento e desenvolvimento ótimo em regiões de climas com ampla distribuição da precipitação e menor flutuação de condições de temperatura e de umidade ao longo do ano (RAO; SARASWATHYAMMA; SETHURAJ, 1998).

Nas regiões em que a espécie é nativa, é comum a presença do fungo “*Microcyclus ulei*”, que confere a doença Mal Das Folhas (MDF), um fator determinístico ao desenvolvimento e na produção de látex pelas plantas. Devido ao MDF e questões agronômicas e econômicas, o cultivo da seringueira expandiu-se para outras fronteiras ao redor do mundo (ONOKPISE, 2004), como para o nordeste da Índia, algumas regiões do Vietnã, sul da China (RAJ et al., 2005) e no Brasil, nas regiões Centro Oeste, Nordeste, Sudeste e Noroeste da região Sul (COSTA et al., 2008).

Dessa forma, a exploração da cultura em outras áreas é uma forma interessante de contornar os danos gerados pelo MDF devido às condições ambientais desfavoráveis à ocorrência da doença (COSTA et al., 2008). Porém, a questão é que essas áreas apresentam climas contrastantes com o de origem da *H. brasiliensis*.

Embora as novas áreas de produção apresentem climas satisfatórios para o crescimento das árvores e produção do látex, essas regiões possuem épocas com condições de estresse às plantas, como baixas temperaturas e períodos mais secos (PRIYADARSHAN; GONCALVES, 2003). Além do decréscimo das taxas de crescimento da planta, as produções de látex ficam praticamente interrompidas em aproximadamente 1 a 3 meses por ano nessas áreas (RAO; SARASWATHYAMMA; SETHURAJ, 1998).

O melhoramento de características de produção das seringueiras tem sido feito baseado em resultados envolvendo estatística e genética quantitativa para determinar os melhores genótipos que serão utilizados como novos cultivares. Uma espécie de ciclo longo requer um tempo maior no programa de melhoramento, que pode durar 30 anos ou mais (VERARDI; JUNIOR; GONCALVES, 2011). Diversos trabalhos que visam maximizar os ganhos genéticos da seringueira

são realizados em diferentes locais ao redor do mundo, como no Sri Lanka (JAYASEKERA, 1983), Indonésia (DASLIN et al., 1986), Nigéria (ONOKPISE; OLAPADE; MEKAKO, 1986), Índia (BASHA; SUJATHA, 2007) e também no Brasil (GONCALVES et al., 1998; COSTA et al., 2000; GONCALVES et al., 2005; FURLANI et al., 2005) e, desde 1927, numerosas pesquisas tem sido feitas para investigar a resistência de *H. brasiliensis* ao MDF por institutos de pesquisa e empresas como a Ford, Goodyear, Firestone e Pirelli (RIVANO et al., 2013; SOUZA et al., 2013).

Com o advento dos marcadores moleculares, novas possibilidades para caracterização dos genótipos estão sendo propostas, bem como análises de diversidade genética, estabelecimento de relação entre características de interesse agrônômico com regiões no genoma (QTLs) e a identificação de genes de interesse (BESSE et al., 1994; LOW et al., 1996; VARGHESE et al., 1997; LESPINASSE et al., 2000; GUEN et al., 2003; ROY; NAZEER; SAHA, 2004; SAHA; ROY; NAZEER, 2005; SOUZA et al., 2009; GUEN et al., 2011; YU et al., 2011; SOUZA et al., 2013).

Mapas genéticos são ferramentas muito úteis em programas de melhoramento em muitas espécies. Para árvores perenes como a seringueira, existe a dificuldade de obtenção de linhas puras, devido à depressão por endogamia e principalmente pelo seu longo período de juvenilidade. Por isso, os métodos tradicionais usados para construir mapas de ligação para populações F2 e populações de retrocruzamento não podem ser usados para essas espécies (GRATTAPAGLIA; SEDEROFF, 1994). Tradicionalmente, a construção de mapas genéticos para populações de irmãos-completos (populações F1) envolvem estratégias de duplo “*pseudo-testcross*”, onde somente os marcadores que segregam na proporção de 1:1 são considerados. No caso da seringueira, o primeiro mapa genético foi gerado usando a estratégia de “*pseudo-testcross*” (LESPINASSE et al., 2000) e foi obtido pela saturação com muitos marcadores. Esse mapa foi um importante passo para obtenção de ligações genéticas saturadas em *H. brasiliensis*. Recentemente, Souza et al. (2013) desenvolveram um importante mapa genético adotando a abordagem proposta por Gazaffi (2013) que é capaz de utilizar marcadores com todos os padrões de segregação (1:1, 1:2:1, 1:1:1:1 e 3:1) e resultou em 23 grupos de ligação.

2.2 Modelos Mistos

O objetivo de um modelo estatístico é explicar as respostas de uma variável dependente em função dos diferentes fatores que são atribuídos a uma série de variáveis independentes (SEARLE; CASELLA; MCCULLOCH, 1992). Ao contrário dos modelos lineares tradicionais em

que somente o resíduo é de natureza aleatória, nos modelos lineares mistos, qualquer fator, exceto o intercepto (μ) e o resíduo, pode ser classificado como fixo ou aleatório, obedecendo os objetivos da pesquisa. Os modelos que apresentam apenas fatores de efeitos aleatórios, fora o intercepto μ , que é sempre fixo, é denominado modelo aleatório (SEARLE; CASELLA; MCCULLOCH, 1992).

Quando um determinado fator é atribuído a um conjunto finito de tratamentos ou níveis específicos de fatores experimentais, tal fator é dito de efeito fixo, e as conclusões, neste caso, são restritas aos níveis dos fatores estudados (SEARLE; CASELLA; MCCULLOCH, 1992; MCCULLOCH; SEARLE; NEUHAUS,). Por outro lado, considera-se que o fator possa assumir um número infinito de níveis e que o conjunto estudado seja uma amostra aleatória de uma população de níveis deste fator, o fator é dito de efeito aleatório e é possível inferir sobre a população da qual os níveis do fator foram amostrados (SEARLE; CASELLA; MCCULLOCH, 1992).

Fundamentalmente, o argumento utilizado para auxiliar na decisão sobre a natureza dos efeitos de um modelo está em como os níveis de um fator são amostrados em uma população, ou se foram definidos de acordo com o interesse do pesquisador (SEARLE, 1971), dessa maneira, comumente, quando os níveis de um determinado fator não são satisfatoriamente representativos àquela fonte de variação, ele é designado como de efeito aleatório, mas, se tais níveis são os de objetivo da pesquisa, com o intuito de determinar valores de significância, ele é designado como de efeito fixo. Smith, Cullis e Thompson (2005) afirmaram que essa escolha depende do objetivo da análise e de considerações a respeito das propriedades dos dois tipos de procedimentos de estimação, isto é, a predição linear não viesada, para os efeitos aleatórios, e a estimação linear não viesada, para os efeitos fixos.

PASTINA (2010) elucidou a questão de definição da natureza dos efeitos, com o cenário de seleção de variedades, em que o objetivo é ranquear os efeitos das estimativas genóticas o mais próximo possível dos valores genéticos verdadeiros. Nesse caso, as estimativas dos efeitos de variedades devem predizer os efeitos verdadeiros com maior acurácia possível, e nesse caso o efeito de variedade deve ser considerado como aleatório. Nesse sentido, Burgueno et al. (2012) mostraram que a probabilidade de ocorrer uma alteração no ranqueamento de genótipos está associada ao coeficiente de regressão de Pearson (r), e quando a correlação entre os efeitos estimados e os efeitos verdadeiros é igual a um, a probabilidade de ocorrer uma alteração no ranqueamento é zero (FREITAS, 2013). Quando o objetivo é determinar diferenças entre pares específicos de genótipos, PASTINA (2010) acrescenta que o método de predição é inadequado, pois a predição de uma diferença específica é viesada. Nesse caso, o efeito de variedade deve ser considerado como fixo. Vale destacar que o uso de efeitos aleatórios de variedades ou genó-

tipos apresenta como vantagem adicional a vantagem de permitir análises de conjunto de dados históricos combinados ao longo de vários anos (SMITH; CULLIS; THOMPSON, 2005).

Frequentemente, a aplicação de modelos mistos no melhoramento de plantas tem dado ênfase a estimação de componentes de variância e identificação apropriada do erro experimental para testar as hipótese dos efeitos fixos. E, raramente, tem sido usado uma proposta geral com a modelagem de estrutura da matriz de variância e covariância (VCOV) genética e previsões dos efeitos aleatórios (BALZARINI, 2002).

Uma forma geral de modelo linear misto para o melhoramento genético de plantas, situação em que é muito comum em avaliações ao longo do tempo, pode ser representada matricialmente como (CROSSA; YANG; CORNELIUS, 2004):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}$$

em que $\mathbf{Y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_T)'$ é o vetor de respostas fenotípicas, tomadas em I indivíduos ou genótipos $i = (1, 2, \dots, I)$ avaliados em T períodos ($t = 1, 2, \dots, T$); \mathbf{X} e \mathbf{Z} são matrizes de delineamentos de posto completo, associadas ao vetor $\boldsymbol{\beta}$ de parâmetros fixo, $\boldsymbol{\beta} = (\beta'_1, \beta'_2, \dots, \beta'_T)'$ e o vetor $\mathbf{g} = (\mathbf{g}'_1, \mathbf{g}'_2, \dots, \mathbf{g}'_T)'$ de efeito genético aleatório, respectivamente; e $\boldsymbol{\varepsilon} = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_T)'$ é o vetor de erro aleatório. Os vetores aleatórios, \mathbf{g} e $\boldsymbol{\varepsilon}$, assumem distribuição normal com média zero, ou seja, $E(\mathbf{g}) = 0$ e $E(\boldsymbol{\varepsilon}) = 0$, e estruturas de variâncias e covariâncias (VCOV) expressadas da seguinte forma:

$$\begin{bmatrix} \mathbf{g} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right),$$

com as matrizes de VCOV para o vetor de resíduo $\boldsymbol{\varepsilon}$ dada por \mathbf{R} e de efeitos genéticos aleatórios expressa por \mathbf{G} . Assim desde que $E(\mathbf{g}) = 0$ e $E(\boldsymbol{\varepsilon}) = 0$ por definição $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$.

Desta forma, tem-se que:

$$\mathbf{V} = Var(\mathbf{Y}) = Var(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R},$$

em que o primeiro termo explica a contribuição dos efeitos genéticos aleatórios, enquanto o segundo apresenta a variância devido ao efeito residual. Caso o procedimento usual de análise seja aplicado, assume-se que a variância residual é constante (homogeneidade de variância) e não correlacionada; isto é, o modelo tradicional de ANOVA. Então, \mathbf{R} neste caso é uma matriz diagonal, com $\mathbf{R} = \sigma_e^2 \mathbf{I}$.

Assumindo ainda que \mathbf{V} é não singular,

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta}$$

implica que,

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$$

No modelo linear misto para predição de valores genotípicos, a matriz de VCOV genética pode ser decomposta, de forma que $\mathbf{G} = \mathbf{G}^P \otimes \mathbf{A}$, em que \mathbf{G}^P é a matriz $T \times T$ de informação de covariância genética de genótipos avaliadas em diferentes períodos, atribuída ao efeito genético principal e de interação $\mathbf{G} \times \mathbf{P}$; \otimes denota o produto de Kronecker (ou produto direto) entre as duas matrizes; e $\mathbf{A} = \{a(i, i')\}$ é uma matriz $I \times I$ de parentesco genético aditivo, conhecida como *numerator relationship matrix*, e seus elementos são duas vezes o coeficiente de parentesco, $2\Phi_{i,i'}$, entre os pares de indivíduos. Ainda, a $Cov(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}') = \mathbf{R} = \mathbf{I}_I \otimes \mathbf{R}_0$ e a $Cov(\mathbf{g}, \boldsymbol{\varepsilon}) = \mathbf{0}$. A matriz de covariância global \mathbf{G} , dada por $Cov(\mathbf{g}, \mathbf{g}') = \mathbf{G}^P \otimes \mathbf{A}$ é comumente representada por (BURGUENO et al., 2012):

$$\mathbf{G}^P \otimes \mathbf{A} = \begin{bmatrix} \sigma_{a_1}^2 & \rho_{12}\sigma_{a_1}\sigma_{a_2} & \cdots & \rho_{1T}\sigma_{a_1}\sigma_{a_T} \\ \rho_{21}\sigma_{a_1}\sigma_{a_2} & \sigma_{a_2}^2 & \cdots & \cdot \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{T1}\sigma_{a_T}\sigma_{a_1} & \cdot & \cdots & \sigma_{a_T}^2 \end{bmatrix} \otimes \mathbf{A},$$

em que o t -ésimo elemento da diagonal da matriz \mathbf{G}^P é a variância genética aditiva $\sigma_{a_T}^2$ no t -ésimo período, sendo o elemento $t't$ a covariância genética aditiva ($\rho_{t't}\sigma_{a_{t'}}\sigma_{a_t}$) entre os períodos t' e t ; assim, $\rho_{t't}$ é a correlação dos efeitos genéticos aditivo entre os períodos t' e t .

A matriz de VCOV residual $\mathbf{R}_0 = \{Cov(\varepsilon_{it}, \varepsilon_{it'})\}$ tem dimensão $T \times T$ e modela a correlação espacial entre parcelas. É possível acomodar a heterogeneidade de variância residual substituindo \mathbf{I}_I do produto direto, $\mathbf{I}_I \otimes \mathbf{R}_0$, por uma matriz diagonal, $\mathbf{D} = \text{Diag}\{\sigma_t^2\}$.

Com essas pressuposições, e seguindo as propriedades de distribuição normal multivariada, a densidade marginal dos dados é normal multivariada e pode ser expressa por (HENDERSON, 1984):

$$[\mathbf{Y}|\boldsymbol{\beta}, \mathbf{R}_0, \mathbf{G}^P] \sim \mathbf{N}_{MV}[\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}(\mathbf{G}^P \otimes \mathbf{A})\mathbf{Z}' + \mathbf{I}_I \otimes \mathbf{R}_0]$$

As estimativas dos parâmetros, $\hat{\boldsymbol{\beta}}$, e a predição dos efeitos genéticos aleatórios, $\hat{\mathbf{g}}$, são as soluções do sistema de equações do modelo misto de acordo com (HENDERSON, 1984).

A abordagem mais empregada para estimação do parâmetros de variância e covariância de

um modelo misto é baseada no método da máxima verossimilhança restrita (*Residual Maximum Likelihood, REML*), que apresenta a vantagem sobre o método da máxima verossimilhança (*Maximum Likelihood, ML*), por eliminar o viés das estimativas dos componentes de variância (PATTERSON; THOMPSON, 1971). O método *REML* envolve a maximização do logaritmo de verossimilhança, que é descrito como:

$$l_R = -\frac{1}{2}\{\log|\mathbf{H}| + \log|\mathbf{X}'\mathbf{H}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y}\}$$

em que: $\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}$. A maximização da verossimilhança do vetor de parâmetros de variância usualmente requer um processo iterativo, tal como o algoritmo *Fisher Scoring* (PATTERSON; THOMPSON, 1971), *EM (Expectation-Maximization)* (DEMPSTER; LAIRD; RUBIN, 1977) e o algoritmo *Average Information* (GILMOUR; THOMPSON; CULLIS, 1995).

Se os parâmetros contidos em \mathbf{H} são conhecidos, $\hat{\beta}$ é o melhor estimador linear não viesado (*Best Linear Unbiased Estimator, BLUE*) de β e $\tilde{\mathbf{u}}$ é o melhor preditor linear não viesado (do inglês, *Best Linear Unbiased Predictor, BLUP*) de \mathbf{u} . No entanto, geralmente os parâmetros em \mathbf{H} são desconhecidos, e são substituídos pelas suas estimativas obtidas pelo *REML*. As estimativas resultantes dos efeitos fixos e aleatórios são então ditas *BLUE* e *BLUP* empíricos, ou *E-BLUE* e *E-BLUP*, respectivamente (SMITH; CULLIS; THOMPSON, 2005)

Quando o método de modelos mistos é utilizado na predição de efeitos aleatórios, como por exemplo as médias ajustadas genotípicas, as estimativas *BLUPs* dos efeitos aleatórios são menores do que se a estimação fosse feita para efeitos fixos. Isso ocorre, devido à ponderação das médias por pesos que são funções de razões de componentes de variância. Por este motivo, as estimativas *BLUP* de efeitos aleatórios são chamadas estimativas “encolhidas” (BALZARINI, 2002). As estimativas dos efeitos fixos incluem uma correção para a existência de variâncias heterogêneas e de covariâncias, já que a matriz invertida \mathbf{H}^{-1} aparece na expressão do estimador $\hat{\tau}$. Dessa forma, mesmo a parte fixa do modelo é afetada pela inclusão dos termos aleatórios, fato que torna as estimativas mais realistas do que as obtidas pelos mínimos quadrados utilizados em modelos fixos (PINHEIRO et al., 2007).

2.2.1 Modelo Misto para Análise de Dados Longitudinais

O modelo simplificado para análise de dados longitudinais pode ser representado como

(LITTELL; HENRY; AMMERMAN, 1998; VERBEKE; MOLENBERGHS, 2009):

$$y_{ijt} = \mu + \underline{g}_i + b_j + p_t + \underline{gp}_{it} + \underline{\varepsilon}_{ijt}$$

em que: y_{ijt} é a resposta fenotípica do genótipo i no bloco j e no período t ; μ é o intercepto; \underline{g}_i é o efeito genético aleatório do i -ésimo genótipo; b_j é o efeito fixo do j -ésimo bloco/repetição experimental; p_t é o efeito fixo do t -ésimo período; \underline{gp}_{it} é efeito de interação genótipo \times período; $\underline{\varepsilon}_{ijt}$ é o erro aleatório. No contexto do melhoramento genético de plantas, o modelo com o efeito aleatório de genótipo em cada período pode ser escrito como:

$$y_{ijt} = \mu + \underline{g}_{it} + b_j + p_t + \underline{\varepsilon}_{ijt}$$

em que \underline{g}_{it} é o efeito do genótipo i no período t . Na matriz \mathbf{G}^P , os elementos da diagonal são a variância genética em cada período (reflete a magnitude da variação entre os genótipos em cada período), e fora da diagonal, estão colocadas as covariâncias genéticas entre os pares de períodos. Diferentes estruturas de VCOV, \mathbf{G}^P para \underline{g}_{it} , podem ser consideradas (Tabela 1).

As matrizes VCOV \mathbf{R}_0 e \mathbf{G}^P podem ter diferentes estruturas (Tabela 1). Mas quando são do tipo não estruturadas, elas contém $T(T + 1)/2$ parâmetros, sendo T o número de períodos. E o número parâmetros cresce conforme aumenta este número de intervalos temporais. A medida em que T aumenta, os efeitos genéticos e/ou residuais se tornam altamente correlacionados entre os períodos, ao passo que a estimativa de \mathbf{R}_0 ou \mathbf{G}^P torna-se dificultada. O uso de diferentes estruturas de VCOV é a solução a esse impasse, e podem ser ajustadas para \mathbf{R}_0 e/ou \mathbf{G}^P (BURGUENO et al., 2012).

2.2.2 Estruturas Alternativas de Matriz de Variâncias e Covariâncias para \mathbf{G}^P e \mathbf{R}_0

Diversas estruturas de VCOV podem ser adotadas com intuito de selecionar uma mais adequada aos padrões de resposta das observações. Uma breve descrição de cada estrutura é feita na Tabela 1.

As matrizes \mathbf{G}^P e \mathbf{R}_0 com 1's na diagonal e 0's fora da diagonal são estruturas chamadas identidade (ID) e consideram independência e homogeneidade de variância, modelo assumido pela ANOVA tradicional. Esta estrutura consiste de pressuposições pouco realistas no contexto de melhoramento genético, uma vez que, é muito comum haver correlação genética entre pares de períodos, e heterogeneidade de variância genética entre os períodos.

O modelo que adota a matriz diagonal (DIAG) admite independência e heterogeneidade

de variância genética entre os períodos; implica em assumir uma variância genética separada para cada período e ausência de covariância genética entre eles, Neste caso, os períodos são não correlacionados, e isso é similar à análise de cada período separadamente. Patterson et al. (1977) modelaram VCOV para diferentes ambientes e consideram o modelo Simetria Composta (CS) que assume que todos os ambientes tem a mesma variância genética e todos os pares de ambientes tem a mesma covariância. O modelo de Patterson et al. (1977) não modela o efeito de interação, gerando informação apenas de sua magnitude, e também ignora a possibilidade de heterogeneidade de variância dos ambientes. Cullis et al. (1998) ajustaram o modelo de Simetria Composta Heterogênea (CS_{HET}) que assumem heterogeneidade de variância entre ambientes e a existência de mesma correlação genética entre os pares deles, que pode ser análogo ao raciocínio para diferentes períodos.

A matriz auto-regressiva de primeira ordem (AR1) e heterogênea ($AR1_{HET}$) apresentam a pressuposição de correlação genética entre os períodos, com variância homogênea e heterogênea, respectivamente. PASTINA (2010) ressalta a importância dessas estruturas, especialmente para culturas perenes e semi-perenes, tais como a cana-de-açúcar e as florestais como a cultura da seringueira, cujos experimentos normalmente são avaliados em diversos períodos, ou seja, em momentos distintos. É um caso de medidas repetidas no tempo. Para a cultura da cana-de-açúcar, PASTINA (2010) acrescenta que a correlação genética pode diminuir com a distância temporal entre as colheitas, o que confere uma interessante interpretação para este tipo de modelo. Uma justificativa é que genes expressos em um primeiro momento podem não ser expressos em momentos subsequentes (PASTINA, 2010).

Outra estrutura é a fator analítico de primeira ordem (FA1), que é uma aproximação do modelo não estruturado (UNS), com o número de níveis do fator igual a 1 ($k = 1$). Esta pode ser considerada como uma extensão da análise de componentes principais, e estima dois componentes vetoriais que quando multiplicados obtém valores aproximados às variâncias e covariâncias da matriz não estruturada. Muitas pesquisas tem sugerido o modelo fator analítico no melhoramento genético (PIEPHO, 1998; SMITH; CULLIS; THOMPSON, 2001; KELLY et al., 2007; SO; EDWARDS, 2011; BURGUENO et al., 2012).

O modelo UNS assume a estrutura VCOV completa, em que \mathbf{G} e \mathbf{R} apresentam $k(k + 1)/2$ parâmetros para diferentes variâncias genéticas de cada período e diferentes covariância genética entre os pares de períodos. Entretanto, quando o número de períodos avaliados é grande, a estimativa dessa matriz é ineficiente ou não estimável, até mesmo para um número moderado de períodos. Por isso, nesses casos uma estrutura mais parcimoniosa é desejável (SMITH; CULLIS; THOMPSON, 2001).

Tabela 1 – Modelos alternativos de estrutura de variâncias e covariâncias (VCOV) para a matriz G que podem ser consideradas no melhoramento genético de plantas (MARGARIDO, 2011)

Modelo	Descrição	Matriz G
Identidade (ID)		$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & 0 & \cdots & 0 \\ 0 & \sigma_g^2 + \sigma_{ge}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
Diagonal (DIAG)		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{g_T}^2 + \sigma_{ge_T}^2 \end{bmatrix}$
Simetria Composta (CS)		$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \sigma_g^2 & \cdots & \sigma_g^2 \\ \sigma_g^2 & \sigma_g^2 + \sigma_{ge}^2 & \cdots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \cdots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
Simetria Composta Heterogênea (CS _{HET})		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_g^2 & \cdots & \sigma_g^2 \\ \sigma_g^2 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \cdots & \sigma_{g_T}^2 + \sigma_{ge_T}^2 \end{bmatrix}$
Autoregressiva de 1ª Ordem (AR1)		$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \sigma_g^2 \rho_g & \cdots & \sigma_g^2 \rho_g^{d(1,T)} \\ \sigma_g^2 \rho_g & \sigma_g^2 + \sigma_{ge}^2 & \cdots & \sigma_g^2 \rho_g^{d(2,T)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(T,1)} & \sigma_g^2 \rho_g^{d(T,2)} & \cdots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
Autoregressiva de 1ª Ordem heterogênea (AR1 _{HET})		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_g^2 \rho_g & \cdots & \sigma_g^2 \rho_g^{d(1,T)} \\ \sigma_g^2 \rho_g & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & \sigma_g^2 \rho_g^{d(2,T)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(T,1)} & \sigma_g^2 \rho_g^{d(T,2)} & \cdots & \sigma_{g_T}^2 + \sigma_{ge_T}^2 \end{bmatrix}$
Fator Analítico de 1ª Ordem (AF1)		$\begin{bmatrix} \lambda_1^2 + \Psi_1 & \lambda_1 \lambda_2 & \cdots & \lambda_1 \lambda_T \\ \lambda_2 \lambda_1 & \lambda_2^2 + \Psi_1 & \cdots & \lambda_2 \lambda_T \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_T \lambda_1 & \lambda_T \lambda_2 & \cdots & \lambda_T^2 + \Psi_T \end{bmatrix}$
Não Estruturada (UNS)		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_{g_{12}}^2 & \cdots & \sigma_{g_{1T}}^2 \\ \sigma_{g_{21}}^2 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g_{T1}}^2 & \sigma_{g_{T2}}^2 & \cdots & \sigma_{g_T}^2 + \sigma_{ge_T}^2 \end{bmatrix}$

σ_g^2 e σ_{ge}^2 : componentes de variância para o efeito principal de genótipos e de interação genótipos \times períodos, respectivamente; $\sigma_g^2 \rho_g^{d(t,t')}$: correlação genética entre períodos, em que $d(t, t')$ corresponde à distância em tempos entre eles; $\sigma_{g_t}^2$ e $\sigma_{ge_t}^2$: variância genética específica de cada período para o efeito principal de genótipos e de interação genótipos \times períodos, respectivamente; $\sigma_{tt'}$: covariância genética entre os período t e t' ; Ψ_t : variância residual específica de cada período; λ_t e λ'_t : elementos regressores (*loadings*) do fator nos períodos t e t' . Os acrônimos referentes aos vários modelos foram mantidos na forma derivada do inglês e de acordo com os padrões do *software* GenStat.

O modelo do tipo Fator Analítico (FA), apresentados por (SMITH; CULLIS; THOMPSON, 2001), são geralmente preferidos por fornecer maior acurácia de predição dos *E-BLUPs* genotípicos, sendo comumente utilizados para modelar a matriz de VCOV em modelos de genética quantitativa (PIEPHO, 1997, 1998; SMITH; CULLIS; THOMPSON, 2001; RESENDE; THOMPSON, 2004; CROSSA; YANG; CORNELIUS, 2004; CROSSA et al., 2006; BURGUENO et al., 2008, 2011, 2012).

Quando são escolhidas adequadamente as estruturas de VCOV para matrizes $\mathbf{G}_{T \times T}^P$ e \mathbf{R}_0 , aumenta-se a acurácia da predição e o poder preditivo dos valores genotípicos *E-BLUPs* (SMITH; CULLIS; THOMPSON, 2001; CROSSA et al., 2006; BURGUENO et al., 2008; KELLY et al., 2007).

2.2.3 Modelos para Múltiplos Caracteres em Múltiplos Períodos

Além de considerar as diversas condições presentes em diferentes períodos, os experimentos em programas de melhoramento genético comumente também envolvem a avaliação de diversos caracteres simultaneamente, visto que materiais elite devem concentrar alelos favoráveis tanto para produção, como para resistência a doenças, pragas, estresses bióticos, qualidade da produção, entre outros (WELHAM et al., 2010). Além disso, caracteres de baixa herdabilidade ou de difícil mensuração podem ser melhorados através da seleção indireta realizada em caracteres correlacionados e com maior herdabilidade (MALOSETTI et al., 2008).

A prática da utilização de modelos com múltiplos caracteres em múltiplos ambientes, do inglês, *multi-trait multi-environment (MTME)* vem aumentando nos últimos anos, especialmente conjugada com mapeamento de QTLs, onde a abordagem de modelos mistos têm sido amplamente aplicada, em muitas configurações de mapeamento (ALIMI et al., 2013; VERBYLA; CULLIS, 2012; SUKHWINDER et al., 2012; MALOSETTI et al., 2008, 2007). A aplicação destes modelos com dados longitudinais tem escassa aparição na literatura, somente trabalhos relacionados a psicologia (GERE; SCHIMMACK, 2011; BROOKER et al., 2013; ALLEN, 1998) e melhoramento animal (MENENDEZ-BUXADERA et al., 2008; BOLIGON et al., 2010; BUXADERA; AYRADO, 2013).

Como os experimentos de melhoramento genético geralmente possuem dados de múltiplos caracteres em múltiplos momentos, a ocorrência de interação $G \times E$ em dados de vários caracteres é ainda mais desafiadora do que no caso univariado, no que diz respeito à análise e interpretação dos resultados (MALOSETTI et al., 2008). Tal fato é notado pela escassez de trabalhos na literatura e metodologia estatística para o tratamento adequado de dados dessa natureza.

Tradicionalmente os experimentos de melhoramento genético tem sido avaliados para cada caráter individualmente, ajustando-se um modelo para cada um deles em múltiplos períodos (análise longitudinal) ou ambientes (análise conjunta), e através de análises metodológicas convencionais (PIEPHO, 2000). Também são realizadas comparações não-formais, geralmente gráficas, ou alguma forma de análise conjunta é empregada para tirar conclusões a respeito da interação $G \times E$ e da correlação entre caracteres (MARGARIDO, 2011).

Tinker, Mather et al. (1995), citado por (PIEPHO, 2000) enfatizam que quando são feitas análises separadas, a interação $G \times E$ e a heterogeneidade ambiental são desconsideradas. O que torna esta abordagem analogamente útil no caso de um conjunto de períodos sucessivos e heterogêneos.

Segundo (MALOSETTI et al., 2008), a modelagem direta da interação genótipo \times ambientes ($G \times E$) e das correlações genéticas entre caracteres, em uma única análise com múltiplos caracteres em múltiplos ambientes (do inglês, *multi-trait-multi-environment*, *MTME*), aumenta o poder estatístico e torna a análise mais realista. Ainda segundo esses autores, as suposições de variância genética homogênea em todos os ambientes e de ausência de covariâncias genéticas entre ambientes e caracteres são irrealistas, de modo que os modelos mistos representam uma abordagem natural e apropriada para a análise conjunta dos dados. Claramente, para tanto, modelos mais sofisticados se fazem necessários.

2.3 Seleção de Modelos

A seleção do modelo é um tópico complexo, na qual diversas estruturas de VCOV para matriz G podem ser razoáveis. Considerações sobre o problema em questão e objetivos devem estar presentes quando se seleciona um modelo. Depois da seleção de alguns modelos plausíveis, a escolha da estrutura de VCOV pode ser auxiliada por alguns métodos (DIGGLE, 1988; LINDSEY, 1999; BROWN; PRESCOTT, 2006; VERBEKE; MOLENBERGHS, 2009). Várias estratégias tem sido adotadas para orientação, tais como métodos de diagnóstico, gráficos (CHRISTENSEN; PEARSON; JOHNSON, 1992), métodos fundamentados no princípios de verossimilhança (DIGGLE, 1988; OMAN, 1991; WOLFINGER, 1993). Dentre os métodos utilizados na seleção da estrutura de VCOV destacam-se:

i) Teste Assintótico da Razão de Verossimilhança (LR)

$$LR = -2 \times \log \left(\frac{L_R}{L_C} \right), \quad (1)$$

em que L_R é a verossimilhança restrita do modelo reduzido, e L_C a verossimilhança restrita do

modelo completo. Esse teste pode ser utilizado em modelos mistos aninhados. Consiste em comparar dois modelos estimados por máxima verossimilhança, em que um deles é uma versão restrita do outro, ou seja, um modelo completo tem r parâmetros adicionais. Assegurado que a parte fixa do modelo seja a mesma para ambos modelos, apenas o número de parâmetros da estrutura de variâncias e covariâncias é testado. O teste irá verificar se esses parâmetros melhoram significativamente o modelo. A hipótese testada, H_0 , é a de que os dois modelos são equivalentes (os parâmetros extras não diferem de zero). A estatística resultante da diferença ($L_R - L_C$) tem distribuição assintótica Qui-Quadrado (χ_r^2) com r graus de liberdade (MOOD; GRAYBILL; BOES, 1974).

ii) Critério de Informação

Pode-se tomar uma perspectiva na teoria da informação e computar medidas baseadas na informação. Estas são calculadas como um termo de penalização aplicado à função de verossimilhança. Os mais comuns são o Critério de Informação de Akaike (*AIC*) (AKAIKE, 1974) e o Critério de Informação Bayesiano (*BIC*) (SCHWARZ, 1978) para comparação de modelos mistos não aninhados, os quais são definidos como (BOZDOGAN, 1987; WOLFINGER, 1993):

$$AIC = -2 \log(L) + 2 \times n_{PAR} \quad (2)$$

$$BIC = -2 \log(L) + \log(N) \times n_{PAR} \quad (3)$$

em que $\log(L)$ é o logaritmo da função de máxima verossimilhança restrita, n_{PAR} é o número total de parâmetros da matriz de VCOV e N é o número total de observações (HU; SPILKE, 2011). Menores valores desses critérios indicam o melhor modelo, haja vista que o logaritmo da função de máxima verossimilhança tem coeficiente negativo nas duas expressões, de forma que a inclusão de parâmetros aumenta os valores de *AIC* e *BIC*. No entanto, *AIC* tenta escolher o modelo que minimiza a divergência de Kullback-Leibler (BURNHAM; ANDERSON, 2004). Essa divergência está relacionada à informação perdida por se usar um modelo aproximado e não o “real” (frequentemente é conceitualizada com uma “distância” entre modelo “real” completo e outro modelo) (PASTINA, 2010).

Segundo Guerin e Stroup (2000), o *AIC* tende a selecionar modelos mais complexos, que melhor se ajustam aos dados, mas, em compensação, com mais parâmetros. Além disso, possui maior controle da taxa de erro tipo I. O *BIC*, por sua vez, parte do pressuposto da existência de um “modelo verdadeiro”, maximizando a probabilidade para identificá-lo, penalizando mais fortemente o número de parâmetros do modelo. Em síntese, o *LR*, *AIC* e *BIC* são conceitual-

mente diferentes mas utilizam como base o mesmo critério estatístico, o máximo da função de verossimilhança como medida de ajuste. Entretanto, definem valores críticos diferentes (BOZDOGAN, 1987; WOLFINGER, 1993; LITTELL, 2002).

O cálculo desses critérios não envolve os parâmetros referentes à parte fixa do modelo. Desse modo, a significância desses parâmetros especificamente não são avaliadas por esses critérios. Assim, as inferências relativas aos parâmetros de efeitos fixos geralmente são testadas utilizando-se o teste de Wald (VERBEKE; MOLENBERGHS, 2009).

2.4 Mapeamento de QTLs

As regiões cromossômicas onde estão presentes genes ou locos que controlam a variação nos caracteres poligênicos são chamadas de QTLs (“*Quantitative trait loci*”) (FALCONER; MACKAY, 1981). O mapeamento desses QTLs tem como maior importância estimar as relações entre o genótipo e o fenótipo de caracteres quantitativos ao longo do genoma da espécie. À primeira vista, parece uma tarefa simples, especialmente com todo conhecimento de genômica atual e ferramentas computacionais disponíveis (DOERGE, 2002). Mas infelizmente, a tarefa não é nada trivial, isso graças ao grande número de QTLs presentes por caráter, e possíveis interações epistáticas entre eles. Além do mais, muitas outras fontes adicionais de variação podem influenciar os resultados (MACKAY, 2001).

O mapeamento completo deve fornecer informações sobre posição, número, efeitos, interações alélicas entre (epistasia) e dentro de locos (dominância) dos QTLs (GARCIA et al., 2008), permitem também melhor entendimento das bases genéticas da correlação entre caracteres governados por QTLs ligados ou pleiotrópicos (FALCONER; MACKAY, 1981); e a identificação de interações entre QTLs e o ambiente (BOER et al., 2007; MALOSETTI et al., 2004). Porém para poder localizar os QTLs, primeiramente é necessária a elaboração de um mapa de locos marcadores, o qual fornece suporte para a alocação dos QTLs ao longo do genoma. Os marcadores moleculares são ferramentas rápidas e eficazes para estudos genômicos, uma vez que detectam o polimorfismo diretamente ao nível do DNA e não sofrem nenhum tipo de influência ambiental (SOUZA, 2001). São utilizados inúmeros tipos de marcadores moleculares para o estudo de genomas de espécies de interesse agrônomo, ou de diversidade natural.

Existem algumas metodologias genético-estatísticas disponíveis para a análise dos dados provenientes de populações de linhagens endogâmicas, tais como: retrocruzamentos, F2 e RILs. Estas metodologias incluem: a análise para cada marca individualmente (STUBER; EDWARDS; WENDEL, 1987), mapeamento por intervalo (LANDER; BOTSTEIN, 1989), mapeamento por

intervalo composto (ZENG, 1993, 1994) e mapeamento de múltiplos intervalos (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999). Contudo, em espécies sem linhagens endogâmicas à disposição, como é o caso da seringueira, a obtenção de tais populações é impraticável. Nesses casos, a construção de mapas genéticos e o mapeamento de QTLs são realizadas a partir de cruzamentos entre indivíduos não endogâmicos contrastantes (LIN et al., 2003) e o mapeamento é feito diretamente na progênie F1 segregante. Gazaffi (2013), desenvolveu um modelo genético-estatístico para mapeamento de QTLs de populações não endogâmicas para dados de marcadores moleculares utilizando mapeamento por intervalo composto. ANONI (2012) elucidou alguns problemas a serem superados neste caso: a necessidade de estimar a fase de ligação entre os marcadores e os QTL's; a mistura de diferentes padrões de segregação, uma vez que cada loco pode ter número diferentes de alelos segregando; a necessidade de definir corretamente os efeitos genéticos, já que as definições típicas de populações F2 não podem necessariamente ser aplicadas diretamente.

Dentre os resultados de mapeamento genético da cultura da seringueira, destacam-se principalmente trabalhos voltados à identificação de QTLs envolvidos na resistência do fungo *Microcyclus ulei* causador do MDF (LESPINASSE et al., 2000; GUEN et al., 2003, 2007, 2011), e mapeamento para características de importância econômica (SOUZA et al., 2013; GAZAFFI, 2013).

O software OneMap (MARGARIDO; SOUZA; GARCIA, 2007) foi inicialmente desenvolvido para facilitar análises de ligação em espécies sem linhagens endogâmicas (*outcrossing*) utilizando a metodologia proposta por (WU et al., 2002). Embora muitos métodos estatísticos foram desenvolvidos especificamente para o mapeamento de QTL de espécies *outcrossing*, o método de duplo “*pseudo-testcross*” é o mais frequentemente utilizado em estudos de QTL em seringueira (LESPINASSE et al., 2000; GUEN et al., 2007, 2011).

2.4.1 Métodos de Mapeamento de QTLs

A primeira metodologia de mapeamento de QTLs foi a de **Mapeamento por marcas individuais**, do inglês *Single Marker (SM)*, em que o mapeamento de QTLs eram realizados para cada marca individualmente. Nesta abordagem, a existência de associações entre os marcadores e os caracteres quantitativos de interesse, são avaliadas através de testes estatísticos tradicionais como teste *t*, análise de variância, regressão linear simples, regressão linear múltipla e teste de razão de máxima verossimilhanças (WELLER, 1986; DOERGE; ZENG; WEIR, 1997). Nestes testes, a hipótese nula testa se as médias dos caracteres fenotípicos é independente do genótipo

de um determinado marcador. Caso a hipótese nula seja rejeitada, infere-se que um possível QTL está ligado àquele marcador (DOERGE, 2002). Análises de marcas individuais é um método de fácil realização, não necessitando de softwares especializados e ferramentas estatísticas complexas; além disso, não necessita da construção prévia do mapa de ligação. Porém, seus resultados apresentam viés, uma vez que há confundimento entre o efeito do QTL e a sua distância em relação ao marcador, resultando em um baixo poder de detecção de QTLs.

Considerando as limitações do método para análise de marcas individualmente, a abordagem denominada **Mapeamento por Intervalo**, do inglês *Interval Mapping (IM)*, foi desenvolvida por Lander e Botstein (1989). Neste método, os marcadores moleculares são ordenados em um mapa de ligação, e servem como base para inferir sobre a localização dos possíveis QTLs. Aqui, são considerados os intervalos entre pares de marcadores adjacentes, e é testada a existência de um possível QTL dentro do intervalo entre eles. Os resultados dos testes são expressos em *LOD-Score (Logarithm of the odds)* e um maior poder de detecção dos QTLs é observado quando comparado com as análises de marcas individuais. Porém, por não controlar a interferência de QTLs que estão localizados fora dos intervalos analisados, há chances do aparecimento de QTLs falsos-positivos (“QTLs fantasmas”).

Em seguida, foi proposto uma extensão do método IM, trazendo soluções à questão dos “QTLs fantasmas”. Este é chamado de **Mapeamento por Intervalo Composto**, do inglês *Composite Interval Mapping (CIM)* dos autores Zeng (1993, 1994), Jansen e Stam (1994). Este método adota marcadores fora dos intervalos definidos como co-fatores. Deste modo, é possível minimizar a interferência de QTLs que poderiam estar afetando a análise e levando ao aparecimento de QTLs falsos positivos. Aqui, os testes estatísticos são realizados combinando a metodologia do mapeamento por intervalo com análises de regressão múltipla. Pelo fato de considerar múltiplos marcadores no modelo de regressão, a variação genética existente na população pode ser controlada na análise estatística e como resultado, há aumento do poder de detecção dos QTLs verdadeiros.

Kao e Zeng (1997), Kao, Zeng e Teasdale (1999), propuseram o método de **Mapeamento por Múltiplos Intervalos**, do inglês *Multiple Interval Mapping (MIM)*, que utiliza múltiplos intervalos de marcadores simultaneamente para adicionar múltiplos QTLs no modelo de mapeamento. Como vantagem, esta metodologia apresenta maior precisão e poder de mapeamento de QTLs quando comparada com as três abordagens descritas anteriormente. Além disso, há a possibilidade de estimar efeitos de epistasia, valores genotípicos dos indivíduos e herdabilidade dos caracteres quantitativos. Como os componentes da variância genética são estimados para cada QTL individualmente, essa abordagem tem um importante papel como uma técnica para

melhoramento assistido por marcadores.

Baseado no fato de que os modelos de mapeamento de QTLs eram tradicionalmente feitos para populações oriundas de cruzamentos entre linhagens, Lin et al. (2003), propôs um modelo de mapeamento de QTLs para progênies segregantes. Este método, fundamentado no IM, permite a utilização de marcadores com diferentes padrões de segregação (1:1:1:1, 1:2:1, 3:1 e 1:1), havendo a possibilidade de estimativa da fase de ligação entre os QTLs com os marcadores.

Gazaffi (2013) expandiu a abordagem proposta por Lin et al. (2003) para o CIM. Seu modelo, além de localizar QTLs, também determina o padrão de segregação e as fases de ligação entre QTLs, permitindo uma análise simultânea de marcadores e QTLs com diferentes padrões de segregação. Adicionalmente, considera-se a inclusão de co-fatores no modelo, o que possibilita estimativas mais precisas de possíveis QTLs localizados no intervalo do genoma em estudo. O modelo proposto por este autor é baseado em 3 contrastes ortogonais úteis para estimar os efeitos aditivos de cada um dos genitores e os efeitos devido à dominância (interação intra-loco) entre os efeitos aditivos. São incluídos também co-fatores para controle da variação ocasionada por QTLs que estão localizados fora do intervalo de mapeamento.

2.5 Análise de Equações Estruturais (SEM)

Os modelos de equações estruturais (do inglês *structural equation models* ou *simultaneous equation models*, SEM), são uma classe de modelos versáteis que permitem a modelagem complexa de dados multivariados e seus preditores, considerando o grau relacionamento entre eles. O desenvolvimento de métodos de análise envolvendo modelos de equações estruturais com variáveis observadas e latentes fornece aos pesquisadores meios para construir novas teorias, ou ainda, testar e modificar outras bem estudadas e consolidadas pela ciência (ANDERSON; GERBING, 1982).

As variáveis observadas, também chamadas de variáveis indicadoras, medidas ou manifestas, são vetores que contém os valores coletados no experimento em estudo. Porém, quando, por algum motivo, determinado fator não pode ser diretamente medido, entra em cena a utilização das variáveis latentes, que não são diretamente medidas ou observadas, mas que sabidamente surtem algum efeito sobre os fatores avaliados no modelo (SWEENEY, 2009). As variáveis latentes podem ser utilizadas tanto para agrupar um conjunto de informações contidas em um determinada classe de variáveis observadas, sendo indicadas por estas, quanto para simplesmente agregar informação adicional um conjunto de variáveis observadas, sem que necessariamente tenham sido indicadas por outras variáveis observadas.

O método de análise de equações estruturais avalia se uma matriz de covariância amostral é consistente com um possível modelo hipotético (SWEENEY, 2009). Através da construção de um diagrama causal dos dados observados e da utilização de variáveis latentes, alocadas perti-nentemente nesta estrutura causal, os *SEM* são capazes de estimar os clássicos erros de mensura-ção dos preditores e os efeitos de causalidade dos dados multivariados. Embora tal método seja comumente pensado para envolver necessariamente variáveis latentes, também é muito comum encontrar versões envolvendo apenas variáveis observáveis (MARTENS; HAASE, 2006), uma sub-classe de modelos chamados modelos recursivos. Outra sub-classe de *SEM*, os modelos es-truturais generalizados, conhecidos como *LISREL* (FOX, 1989; JÖRESKOG; SORBOM, 1982; FORNELL; LARCKER, 1981), também são sistemas de equações lineares, mas que além de variáveis observadas apresentam em sua estrutura variáveis latentes.

SEM representa uma extensão de técnicas de análises multivariadas baseadas em modelos lineares generalizados, particularmente, análise fatorial, análise de trilha e regressão múltipla (ULLMAN, 2006). Hair et al. (1998) destaca esse ponto definindo *SEM* como uma técnica mul-tivariada que combina aspectos de regressão múltipla com análise fatorial, para estimar séries de inter-correlações simultâneas e dependentes.

Fazendo uma analogia com regressão múltipla, o método de análise estrutural em sua forma mais simplificada, se assemelha com o caso em que se utiliza algumas variáveis independentes mensuradas, empregadas com o intuito de explicar uma única variável dependente correlacio-nada. Sendo assim, com posse de um único modelo, é possível inferir sobre o grau de relaciona-mento entre tais variáveis (GILLARD, 2010). Contudo, de acordo com o aumento do número de variáveis, sejam elas dependentes ou independentes, os modelos vão paralelamente aumentando seu grau de complexidade entre as relações e interações dos dados, e novas equações sistema-ticamente devem ser adotados para contemplar todas as associações presentes na hipótese em estudo (SWEENEY, 2009).

Quando comparada com outras técnicas analíticas estatísticas como análise fatorial e de regressão, *SEM* pode ser considerada uma ferramenta ligeiramente nova. Suas raízes datam meados dos anos 1920, quando Sewall Wright tentou resolver equações simultâneas para dis-cernir influências genéticas ao longo de gerações. No entanto, muitos anos se passaram antes que a técnica recebesse o devido valor merecido; o aumento da complexidade das relações e interações dos temas abordados na ciência, inicialmente nas áreas sociais (WOLFLE, 2003) e posteriormente nas áreas de educação, ciência comportamental (FAN; THOMPSON; WANG, 1999), biologia, economia, medicina, marketing (RAYKOV; MARCOULIDES, 2000), psicolo-gia (BOLLEN; LONG, 1992), demografia e genética populacional (HAIR et al., 1998), geraram

a necessidade da implementação computacional e aperfeiçoamento da técnica pra contemplar a complexidade dessas novas análises.

Diversos softwares com plataforma gráfica, ou mesmos pacotes de ambientes de programação estão disponíveis atualmente para executar tais análises. Narayanan (2012) agrupou oito categorias entre os softwares mais populares: *Amos*, *SAS PROC CALIS*, *LISREL*, *EQS*, *Mplus* e os pacotes *sem*, *lavaan* e *OpenMx*, do R.

2.5.1 Diagramas Causais e Terminologias

Os diagramas são úteis para permitir a visualização das hipóteses sobre as relações entre as variáveis em estudo. Dessa forma, são uma ferramenta excepcional tanto para o pesquisador que gera as hipóteses em estudo, como para um leitor que pode utilizar tais diagramas como suporte para entender os objetivos e pretensões da modelagem estrutural. Para tanto, é necessário que exista correspondência completa entre a estrutura do diagrama com as equações necessárias para a análise.

Os diagramas causais além de serem interessantes para melhorar a compreensão de *SEM*, também podem contribuir substancialmente para as melhorias do modelo estrutural estudado, gerando um suporte visual ao pesquisador sobre quais vias podem ser incluídas ou retiradas do modelo (RAYKOV; MARCOULIDES, 2000).

Para ainda maior clareza, inúmeras convenções foram adotadas para padronizar as representações dos diagramas causais, como, a utilização de letras gregas para representar as relações estruturais e os diferentes tipos de erros envolvidos no sistema (JÖRESKOG; THILLO, 1972; FOX, 2012); os tipos de setas; e as caixas que envolvem cada uma das variáveis. As variáveis observadas são colocadas dentro de caixas quadradas ou retangulares. Variáveis não observadas (também chamadas de variáveis latentes) são colocadas dentro de círculos ou elipses.

As linhas indicam as relações entre as variáveis e representam os parâmetros estruturais presentes nos modelos, podendo ter uma ou duas setas em suas extremidades. Uma linha com uma única seta representa relação direta de causalidade entre duas variáveis no sentido da extremidade nua para a extremidade com a seta. Uma linha com uma seta em ambas extremidades indica a presença de correlação não explicada entre as duas variáveis, obviamente, quando não há uma linha relacionando duas variáveis implica que nenhuma relação *direta* foi constatada entre elas (FOX, 2012), ver Figura 1.

As variáveis exógenas, ou independentes, quando observadas, são representadas pela letra x e quando latentes pela letra ξ (ξ_i); de forma análoga, as variáveis endógenas, ou dependentes,

quando observadas, pela letra y , ou η (eta) quando latentes. As variáveis endógenas normalmente se distinguem das exógenas por terem setas direcionadas voltadas para elas, enquanto as variáveis exógenas aparecem apenas ligadas nas extremidades nuas das setas direcionadas (SHIPLEY, 2002).

As letras γ (gama) e β (beta) são os parâmetros estruturais. Em modelos que contém somente variáveis observadas (modelo recursivo), estas letras são utilizadas para que relacionar uma variável exógena a uma endógena ($x \rightarrow y$) e uma variável endógena a a outra endógena ($y \rightarrow y$), respectivamente (SHIPLEY, 2002). Já em um modelo que contém tanto variáveis observadas como latentes (modelo estrutural generalizado), γ pode ser utilizado tanto para relacionar variáveis exógenas com endógenas, como para relacionar variáveis exógenas com variáveis latentes, e β pode ser utilizado para relacionar endógenas com exógenas e latente com latentes. Normalmente as variáveis exógenas estão dispostas à esquerda das variáveis endógenas, gerando o sentido de causa \rightarrow efeito (no sentido da esquerda para direita), mas esta é uma convenção que nem sempre é possível, dado à complexidade que alguns modelos apresentam (FOX, 2012).

No modelo, λ (lambda) representa o coeficiente de regressão, que relaciona as variáveis observadas com as variáveis latentes. A letra ω (omega) é o parâmetro estrutural que relaciona variáveis endógenas a variáveis latentes, sendo útil para descrever relações de *feedback* entre as causas e respostas.

Os erros do modelo são representados pelas letras ε (epsilon), δ (delta) e ζ (zeta). A letra ε representa o erro de mensuração das variáveis endógenas e quando os erros estão associados com variáveis exógenas a letra utilizada é δ . A letra ζ representa o erro de perturbação eminente sobre as variáveis latentes. Há mais um parâmetro sobre os erros de perturbação, que contém a covariância entre dois ζ 's, que é representado pela letra ψ (psi), as variâncias individuais de cada perturbação ζ são escritas como ψ^2 , mas comumente não são representadas em diagramas causais.

Todas estas convenções descritas acima estão apresentadas em alguma das duas figuras presentes nesta seção (Figuras 1 e 2), que exemplificam respectivamente um modelo recursivo, e um modelo estrutural generalizado.

2.5.2 O Método

SEM é um método de análise de componentes da correlação entre os dados, sendo esta subdividida em efeitos diretos e indiretos com informação de séries de causalidades das variáveis sobre outras variáveis. Por isso a qualidade dos dados utilizados aumenta consideravelmente a

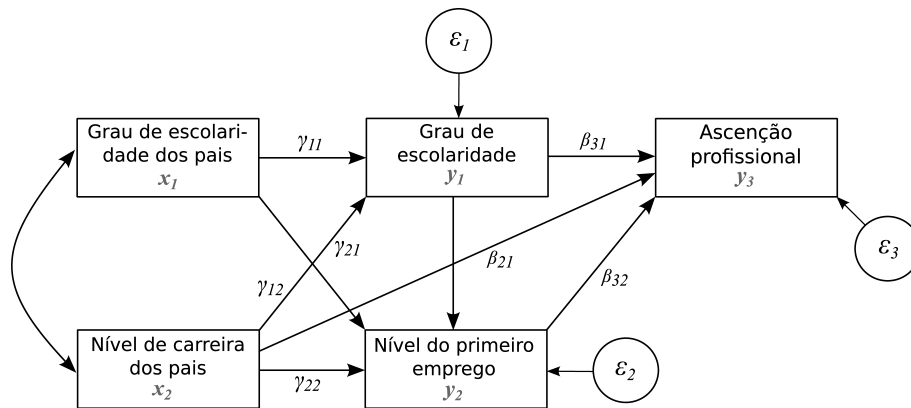


Figura 1 – Modelo estrutural com dados observados (Modelo Recursivo). Um exemplo para demonstrar as terminologias utilizadas em de SEM: os γ 's e os β 's são os parâmetros estruturais que relacionam as variáveis exógenas com as endógenas ($x \rightarrow y$) e endógenas com endógenas ($y \rightarrow x$), respectivamente, ε 's são os erros de mensuração das variáveis observadas y . Figura adaptada de (FOX, 2012)

confiabilidade dos resultados, sendo dessa maneira evitadas as variáveis que apresentem variância e covariância duvidosa, comumente presente naquelas que contém dados discrepantes ou *outliers* e apresentam ausência de distribuição normal residual. De acordo com Schumacker e Lomax (1996), a triagem dos dados é um primeiro passo muito importante na modelagem de equações estruturais, o tamanho amostral e os tipos de variáveis são fatores cruciais a serem avaliados para alcançar sucesso nas análises (JACKSON, 2003).

O tamanho amostral é uma questão muito discutida entre os pesquisadores. Darlington (1990) afirma que as amostras devem ter no mínimo 50 observações, Hoyle (1995) e Kling (2001) enfatizam que esse valor deve ser maior que 100, enquanto Kelloway (1998) recomenda amostras maiores que 200 observações. Hair et al. (1998) diz que o tamanho da amostra não é uma regra clara, e deve variar conforme a amplitude dos erros de mensuração ou desvios de normalidade, esta escolha também é afetada pelo método de estimativa escolhido. O método de máxima verossimilhança, por exemplo, exige um tamanho mínimo da amostra de 100 a 150 observações, porém o método torna-se pouco sensível para amostras maiores que 400 observações, e assim pouca ou nenhuma diferença é detectada.

Variáveis instrumentais são variáveis não correlacionadas com o erro, que são utilizadas para fornecer estimativas dos parâmetros estruturais de SEM (KLEIBERGEN, 2002). Neste contexto, as variáveis exógenas são empregadas como variáveis instrumentais nos ajustes dos modelos estruturais, que seminalmente podem ser descritos como:

$$y = \mathbf{X}\theta + \varepsilon$$

em que: y é o vetor $n \times 1$ da variável resposta da equação; X é uma matrix $n \times p$ contendo p preditores endógenos e exógenos, onde é incluído uma coluna com 1's para o intercepto; θ é o vetor $p \times 1$ de parâmetros, contendo os γ 's e os β 's da equação estrutural; e finalmente ε é o vetor de resíduos $n \times 1$.

Z é uma matrix $n \times p$ contendo as variáveis instrumentais e novamente incluindo uma coluna de 1's. E em seguida multiplicando a equação estrutural por Z' , temos:

$$Z'y = Z'X\theta + Z'\varepsilon$$

Utilizando o operador *P-lim* de limite de probabilidade, $\frac{1}{n}Z'\varepsilon$ converge para 0, devido a não-correlação entre as variáveis instrumentais com o erro (FOX, 2012). Consequentemente o estimador de variáveis instrumentais pode ser escrito como:

$$\hat{\theta} = (Z'X)^{-1}Z'y$$

A partir daí, temos que número de variáveis instrumentais é igual ao número de preditores na equação estrutural e que os produtos cruzados da matriz $Z'X$ é não-singular. Outra maneira de $Z'X$ ser não-singular é quando as variáveis instrumentais são correlacionadas com os preditores, mas quando isso ocorre, deve-se verificar a existência de colinearidade entre os dados para evitar problemas de convergência.

Com relação ao número de variáveis instrumentais comparadas com o número de preditores, os modelos podem ter três tipos de identificação, podem ser: identificados, sub-identificados e sobre-identificados. Quando o número de variáveis instrumentais é igual a p , a equação estrutural é chamada equação identificada:

$$Z'y = Z'X\hat{\theta} \Rightarrow y = X\hat{\theta}$$

Assim como há um parâmetro para cada estimativa, há zero graus de liberdade no modelo (MARUYAMA, 1998). Porém se houver menos variáveis instrumentais que preditores (coeficientes estruturais), as estimativas das equações são insuficientemente determinadas (KAPLAN, 2000), e a equação estrutural torna-se sub-identificada pois o modelo apresenta graus de liberdade negativos (MARUYAMA, 1998; KELLOWAY, 1998).

No caso de haver mais variáveis instrumentais que preditores, as equações de estimação serão sobre-determinadas. Sendo assim, inúmeras soluções podem ser encontradas e uma saída plausível seria escolher aquelas que fazem algum sentido científico com alguma margem de

erro (KELLOWAY, 1998), a equação estrutural é dessa forma chamada de sobre-identificada, e neste caso há graus de liberdades positivos. Diante destas questões, para obter estimativas consistentes, uma saída seria descartar algumas variáveis instrumentais, o quê, possivelmente descaracterizaria as hipóteses em estudo. Contudo há outras alternativas para contornar o excesso de identificação. Uma delas é a utilização do método de *mínimos quadrados em dois estágios* (*Two-stage least squares, 2SLS*).

O método de *2SLS*, pode ser assim chamado, devido a junção de duas regressões por *mínimos quadrados ordinários* (*ordinary least squares, OLS*), que é um método para estimar parâmetros desconhecidos em um modelo de regressão linear. Este método minimiza a soma de quadrados das distâncias verticais entre as variáveis respostas observadas no conjunto de dados e as respostas previstas pela aproximação linear. O estimador resultante pode ser expresso por uma fórmula simples, especialmente no caso de um único regressor. O estimador *OLS* é consistente quando os regressores são exógenos, não há multi-colinearidade entre os preditores e os resíduos apresentam homocedasticidade e correlação expressiva. Sob essas condições, o método fornece estimativas de variância mínima e média não viesada. Sob a hipótese de distribuição normal residual, o *OLS* coincide com o estimador de *máxima verossimilhança* (HILLIER; KINAL; SRIVASTAVA, 1984).

Para o método *2SLS*, no *primeiro estágio*, os preditores \mathbf{X} entram no modelo de regressão juntamente com as variáveis instrumentais \mathbf{Z} , para obter os valores dos preditores ajustados (HILLIER; KINAL; SRIVASTAVA, 1984):

$$\hat{\mathbf{X}} = \mathbf{X}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$$

No *segundo estágio*, a variável resposta y entra num segundo modelo de regressão, que têm a matriz \mathbf{X} substituída pela matriz $\hat{\mathbf{X}}$ obtida na regressão do primeiro estágio. Esta nova regressão é útil para fornecer os estimadores dos parâmetros estruturais θ :

$$\hat{\theta} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'y$$

Os dois estágios do *2SLS* podem ser combinados algebricamente, produzindo a seguinte equação para os estimadores:

$$\hat{\theta} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'y$$

A matriz de estimativas de covariâncias assintóticas dos coeficientes, também chamada de

C, é:

$$\hat{V}(\hat{\theta}) = C = \hat{s}^2 [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$$

em que s^2 é a variância residual estimada do modelo de equação estrutural, que é, a soma de quadrados dividido pelos graus de liberdade do resíduo:

$$\hat{s}^2 = \frac{(y - \mathbf{X}\hat{\theta})'(y - \mathbf{X}\hat{\theta})}{n - p}$$

Há cinco etapas básicas envolvidas nas análises *SEM*: especificação, identificação, estimativa, testes e modificação do modelo (SCHUMACHER et al., 2002), estas são descritas como:

- **Especificação do modelo**

Esta etapa envolve o desenvolvimento teórico de um modelo primário. Este processo deve ser guiado pela combinação de fundamentos teóricos, resultados empíricos *a priori* e a inclusão das hipóteses a serem analisadas (HAIR et al., 1998; FORNELL; LARCKER, 1981; BOLLEN; LONG, 1992).

Inicialmente são incluídas somente as variáveis de relevância comprovada, e esse modelo deve ser consistente com o histórico teórico atestado por pesquisas já realizadas. Caso este modelo não seja condizente com a realidade, ele é chamado *modelo mal especificado* (FOX, 2012). Isso pode acontecer caso não sejam incluídas as variáveis verdadeiramente importantes, ou estas estejam saturadas com erro. O passo final desta etapa é a construção do diagrama causal.

- **Identificação do modelo**

Em *SEM* é fundamental que o pesquisador resolva a questão da identificação antes da estimativa dos parâmetros (SCHUMACHER et al., 2002). Em termos gerais, aqui devem ser escolhidas as variáveis instrumentais que sejam suficientes para se obter uma solução única para os parâmetros a serem estimados (DIAMANTOPOULOS; SIGUAW; SIGUAW, 2000; KELLOWAY, 1998). A identificação determina se é possível encontrar valores possíveis para os parâmetros do modelo especificado (BOLLEN; LONG, 1992) e garante a correspondência entre a informação observada (variância e covariância observadas) com a informação a ser gerada (parâmetros a serem estimados) (HOYLE, 1995).

- **Estimativa do modelo**

O propósito da estimação é gerar valores numéricos para os parâmetros a serem estimados, de tal maneira que produza uma matriz de covariância C com valores paramétricos o mais próximos o possível da matriz de covariância amostral observada (S). O processo de estimação envolve a seleção de uma uma função que minimiza as diferenças entre C e S .

Muitas metodologias de ajuste são tradicionalmente utilizadas como *mínimos quadrados não ponderados (UWLS)*, *máxima verossimilhança (ML)*, *mínimos quadrados generalizados (GLS)* e *mínimos quadrados ponderados (WLS)* (RAYKOV; MARCOULIDES, 2000). O método *UWLS* foi inicialmente a técnica mais comum, mas foi rapidamente substituído pelo método *ML* (HAIR et al., 1998).

• Testes do modelo

Uma vez que as estimativas dos parâmetros são obtidas por um modelo de *SEM*, o próximo passo para o pesquisador é determinar o quão bem os dados foram ajustados pelo modelo (SCHUMACHER et al., 2002) e avaliar se o modelo em teste é capaz de se adequar corretamente para os dados. Esta é uma das etapas mais importantes de *SEM* (YUAN, 2005), uma vez que é nela que o pesquisador determina se o modelo a ser testado deve ou não ser rejeitado.

Ajuste do modelo refere-se à extensão em que o modelo é consistente com os dados observados (DIAMANTOPOULOS; SIGUAW; SIGUAW, 2000). O ajuste do modelo é definido por Hair et al. (1998), como o grau em que a matriz de entrada observada é prevista pela estimativa do modelo. Um modelo é dito adequado quando a matriz de covariância ajustada se aproxima da matriz de covariância observada (HOYLE, 1995).

A questão da avaliação de ajuste tem sido um assunto de grande interesse e debate desde o início do *SEM* (BOLLEN; LONG, 1992). É uma área muito ampla e complexa (RAYKOV; MARCOULIDES, 2000) e representa um grande desafio para os desenvolvedores de teoria e investigadores.

A rejeição ou não rejeição dos modelos pode ser feita por índices de critérios de seleção como *Akaike Information Criteria (AIC)*, (AKAIKE, 1974) *Bayesian Information Criteria (BIC)* (SCHWARZ, 1978) e/ou pela probabilidade de não rejeição do modelo que segue distribuição de *qui-quadrado* (χ^2) (HOYLE, 1995; FOX, 2012; SCHUMACHER et al., 2002).

• Modificação do modelo

Se o ajuste do modelo teórico implícito não for adequado (o que é frequentemente com os modelos iniciais), então o próximo passo é modificar o modelo e, subsequentemente, avaliar este modelo modificado.

Este representa um dos aspectos mais controversos entre as etapas (KELLOWAY, 1998). As modificações podem ser feitas empregando ou extraíndo vias do modelo, tanto ligando uma variável observada a uma variável latente como o inverso, criando novas ligações entre as variáveis observadas, permitindo ou restringindo as correlações entre os erros de medição e as correlações entre as variáveis latentes (DIAMANTOPOULOS; SIGUAW; SIGUAW, 2000).

Esta etapa tem como maior objetivo melhorar o modelo, o que implica um modelo de melhor ajuste e um modelo mais parcimonioso que seja também mais interpretável. Mas as alterações devem ser fiéis à realidade observada e de interpretação clara e bem fundamentada.

2.5.3 Modelos Recursivos

Os modelos recursivos são modelos de única direção, sem vias de *feedback* que não apresentam variáveis latentes em sua estrutura (MARTENS; HAASE, 2006).

Analisando o exemplo da Figura 1 (FOX, 2012), que utiliza dados de estudos sociais (BLAU; DUNCAN, 1967), vemos cinco variáveis observadas, sendo duas exógenas (x) e três endógenas (y). Em modelos de regressão convencionas, esse tipo de análise não pode ser feita, pois há mais variáveis endógenas que exógenas. Além disso, utilizando modelos estruturais é possível inferir sobre as possíveis causalidades que cada variável é ou não capaz de exercer sobre as outras, algo que não é possível com regressões lineares de nenhuma ordem.

Com uma cadeia de relações, as hipóteses sobre o problema são arquitetadas no sentido de se relacionarem de forma lógica, e assim poder estimar todos os parâmetros de forma conjunta. Dessa maneira, três equações para este modelo (Figura 1) são formadas (FOX, 2012):

$$\begin{cases} y_{1i} = \gamma_{10} + \gamma_{11}x_{1i} + \gamma_{12}x_{2i} + \varepsilon_{1i} \\ y_{2i} = \gamma_{20} + \gamma_{21}x_{1i} + \gamma_{22}x_{2i} + \beta_{21}y_{1i} + \varepsilon_{2i} \\ y_{3i} = \gamma_{30} + \gamma_{32}x_{2i} + \beta_{31}x_{2i} + \beta_{32}y_{2i} + \varepsilon_{3i} \end{cases}$$

em que: γ_{10} , γ_{20} e γ_{30} são os interceptos dos modelos, γ 's os parâmetros estruturais dos modelos, x o vetor de dados observados e ε o erro de mensuração das variáveis. Ao escrever as equações estruturais de um diagrama, é comum omitir os parâmetros dos interceptos, para os quais não há linhas no diagrama. Para justificar essa prática, podemos expressar todas as variáveis como desvios de suas esperanças (na amostra como desvios de suas médias), eliminando o intercepto de cada equação regressora (FOX, 2012).

Esses tipos de modelos, constituídos somente de dados observados são chamados de modelos recursivos. Tais modelos apresentam duas características peculiares: a não existência de

caminhos recíprocos, ou seja, vias de *feedback* entre duas ou mais variáveis e os erros das variáveis resposta são independentes entre eles. Dessa maneira não há a vinculação de correlação entre os erros ε , portanto, estes modelos não apresentam o parâmetro ψ em sua estrutura.

Como consequência destas duas propriedades, os preditores das equações estruturais dos modelos recursivos sempre são independentes do erro da equação e a equação estrutural pode ser estimada por regressão *OLS*. Dessa maneira, a estimação de modelos recursivos são simplesmente a aplicação em sequência de modelos lineares simples.

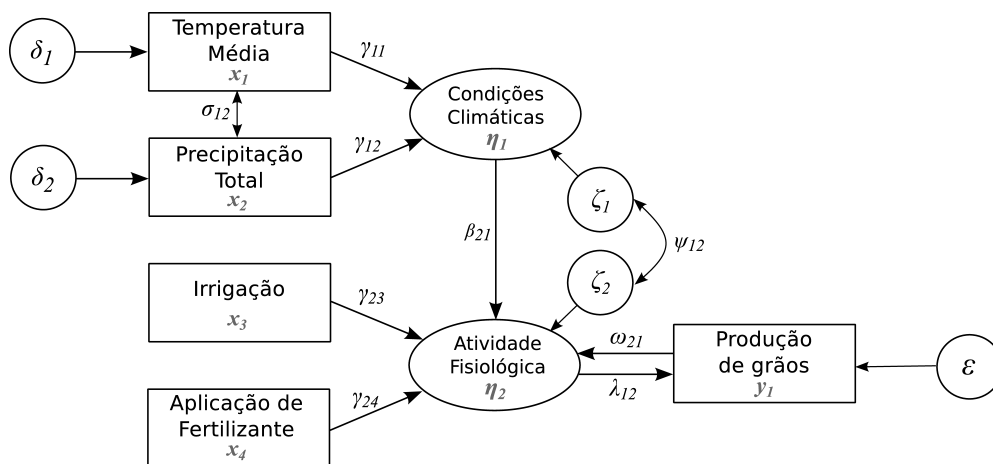


Figura 2 – Modelo hipotético com dados observados e latentes (modelo estrutural generalizado), exemplificando a metodologia *SEM* em dados biológicos. As letras gregas γ 's e o β são os parâmetros estruturais que relacionam as variáveis exógenas (x) com as latentes (η) e latentes com latentes, respectivamente. δ 's são os erros de mensuração das variáveis exógenas observadas; ε é o erro da variável endógena observada, ζ 's são os erros de perturbação das variáveis latentes e ψ a covariância entre os ζ 's, λ é o parâmetro estrutural que relaciona uma variável latente com uma variável endógena e ω o parâmetro que relaciona a variável endógena com a variável latente. Figura adaptada de Fox (2012)

2.5.4 Modelos Estruturais Generalizados

Modelos de equações estruturais generalizados incluem variáveis exógenas e endógenas não observáveis (também chamadas de variáveis latentes), além dos erros de perturbação entre elas (FOX, 1989; SHIPLEY, 2002). Estes modelos normalmente são chamados de *LISREL*, isso graças ao famoso *software* computacional que leva este mesmo nome (JÖRESKOG; THILLO, 1972). *LISREL* é um acrônimo para relações lineares estruturais (*linear structural relations*).

Duas características são comuns nesses tipos de modelos: A existência de um sub-modelo

estrutural relativo às variáveis endógenas para exógenas e endógenas para endógenas; e um modelo de medição, relacionando variáveis latentes com as variáveis indicadoras observadas.

As equações do modelo estrutural e do modelo de medição são escritas como (FOX, 2012):

$$\begin{cases} \eta_i = B\eta_i + \Gamma\xi_i + \varsigma_i \\ y_i = \Lambda_y\eta_i + \varepsilon_i \\ x_i = \Lambda_x\xi_i + \delta_i \end{cases}$$

em que: η_i a variável latente exógena para a observação i ; B o parâmetro estrutural que relaciona as variáveis latentes; Γ é o parâmetro estrutural que relaciona às variáveis latentes endógenas às variáveis exógenas; ξ_i são as variáveis latentes exógenas ς_i é o erro de perturbação estrutural; y_i são as variáveis observadas que estão relacionadas às variáveis latentes endógenas Λ_y é o coeficiente de regressão que relaciona as variáveis observadas às variáveis latentes endógenas; ε_i é o erro amostral das variáveis observadas; x_i são as variáveis observadas que estão relacionadas às variáveis latentes exógenas; Λ_x é o coeficiente de regressão que relaciona as variáveis observadas às variáveis latentes exógenas; δ_i é o erro de mensuração das variáveis observadas exógenas.

Uma notação muito usual na abordagem de modelos estruturais generalizados é a *RAM Formulation* (do inglês *reticular action model*) (MCARDLE, 1980). O modelo *RAM* inclui dois conjuntos de variáveis: \mathbf{v} , que contem as variáveis exógenas e endógenas observadas e as variáveis latentes endógenas e exógenas utilizadas; e o vetor \mathbf{u} , que contém os erros sobre as variáveis exógenas, endógenas e os erros de perturbação, esses dois conjuntos de variáveis são agrupadas na modelo (FOX, 2012):

$$\mathbf{v} = \mathbf{A}\mathbf{v} + \mathbf{u}$$

Nessa equação, além dos vetores \mathbf{v} e \mathbf{u} , existe a matriz \mathbf{A} , que inclui os coeficientes estruturais γ e β , os coeficientes de regressão γ e o fator de *feedback* ω .

Para o exemplo da Figura 2, tal modelo com seus respectivos parâmetros é dado por:

$$\begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ y1 \\ \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{12} \\ \gamma_{11} & \gamma_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{23} & \gamma_{24} & \omega_{21} & 0 & \beta_{21} \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ y1 \\ \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ x_3 \\ x_4 \\ \varepsilon \\ \zeta_1 \\ \zeta_2 \end{bmatrix}$$

Este é um típico em que A é uma matriz esparsa, que contém muitos 0's. Note que há um tratamento especial às variáveis x_3 e x_4 que são unidades sem erro, portanto aparecem tanto no vetor \mathbf{v} como no \mathbf{u} .

O componente final é a matriz de covariância P do erro u . Partindo do pressuposto que todas as variáveis de erro tem esperança igual a zero, e que todas as outras variáveis foram expressas como desvios de suas esperanças, temos:

$$\mathbf{P} = \mathbf{E}(\mathbf{uu}')$$

, e:

$$\mathbf{P} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 & 0 & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{44} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\varepsilon} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_{11} & \psi_{12} \\ 0 & 0 & 0 & 0 & 0 & \psi_{21} & \psi_{22} \end{bmatrix}$$

em que: σ_{11} , σ_{12} , σ_{21} , σ_{22} , σ_{33} e σ_{44} , são as variâncias e covariâncias entres variáveis exógenas observadas x ; σ_{ε} é a variância dos erros de medição da variável y ; e ψ_{11} , ψ_{11} , ψ_{11} e ψ_{11} são as variâncias e covariâncias entre os erros de perturbação.

A chave para estimar os parâmetros do modelo é a relação que existe entre as covariâncias das variáveis observadas, as quais podem ser calculadas diretamente a partir dos dados amostrais, e os parâmetros da matriz \mathbf{A} e \mathbf{P} . Seja \mathbf{m} o números de variáveis observadas em \mathbf{v} e \mathbf{n} o número de observações totais. A matrix \mathbf{J} de ordem $m \times m$ é definida para distinguir as variáveis observadas, sendo (FOX, 2012):

$$\mathbf{J} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$$

onde I_n é uma matriz identidade de ordem n , e os 0's são matrizes nulas que preenchem o restante da matriz \mathbf{J} . O seguinte modelo indica a matriz de covariância \mathbf{C} entre as variáveis observadas.

$$\mathbf{C} = \mathbf{E}(\mathbf{J}'\mathbf{v}\mathbf{v}'\mathbf{J}') = \mathbf{J}(\mathbf{I}_m - \mathbf{A})^{-1}\mathbf{P}(\mathbf{I}_m - \mathbf{A})^{-1}'\mathbf{J}'$$

Seja \mathbf{S} a matriz de variâncias e covariâncias obtida a partir de todos os dados amostrais, o ajuste do modelo se dá por meio da estimação dos parâmetros contidos em \mathbf{A} e \mathbf{P} de forma que os valores da matriz-implícita de covariâncias \mathbf{C} se tornem os mais próximos possíveis dos valores da matriz \mathbf{S} . Sobre as suposições de que os erros e as variáveis latentes são multivariadamente normalmente distribuídos, a máxima verossimilhança das estimativas dos parâmetros de \mathbf{A} e \mathbf{P} são obtidos pela minimização da equação:

$$\mathbf{F}(\mathbf{A}, \mathbf{P}) = \text{tr}(\mathbf{S}\mathbf{C}^{-1}) - n + \log_e \det \mathbf{C} - \log_e \det \mathbf{S}$$

em que $\text{tr}(\mathbf{S}\mathbf{C}^{-1})$ é o traço da matriz $\mathbf{S}\mathbf{C}^{-1}$, ou seja, o somatório dos elementos de sua diagonal, e $\log_e \det \mathbf{C}$ e $\log_e \det \mathbf{S}$ os logaritmos na base neperiana dos determinantes das matrizes \mathbf{C} e \mathbf{S} , respectivamente.

2.5.5 SEM no Cenário de Mapeamento de QTLs

Com dados experimentais coletados em experimentos bem delineados, e com representação amostral satisfatória, é possível desenvolver modelos de equações estruturais para se obter os valores de parâmetros estruturais presentes no modelo. As populações genéticas experimentais que segregam naturalmente com variabilidade alélica entre os indivíduos, podem fornecer base para a inferência de redes de associações causais entre locos genéticos e fenótipos fisiológicos, como por exemplo os de produção vegetal ou resistências a doenças. Em particular, a aleatorização dos alelos que ocorre durante a meiose apresenta uma configuração que é análoga a um delineamento experimental randomizado, de tal modo que pode ser inferida a causalidade dentro da estrutura da estatística clássica fisheriana (THOMAS; CONTI, 2004).

Schadt et al. (2005) propuseram um processo para inferir sobre o relacionamento causal entre duas características fenotípicas e um QTL comum, destacando as relações de causalidade

envolvidas na expressão de um gene em particular, com a atuação direta do QTL, e características quantitativas, para determinar se eles estão relacionados uns com os outros seguindo um modelo causal, reativo ou independente (Modelos a, b e c da Figura 3, respectivamente). Nestes modelos, as variáveis *QTL*, *y1* e *y2* representam a ação do gene, a manifestação da primeira e da segunda característica, respectivamente. O modelo causal demonstrado na Figura 3a refere-se à relação causal mais simples no que diz respeito à manifestação da primeira característica, em que as variações alélicas do QTL causam alteração em *y2* indiretamente pela influência direta do QTL na característica *y1*. O modelo reativo demonstrado na Figura 3b demonstra que a expressão de *y1* é modulada pela característica *y2*. Finalmente, o modelo independente (Figura 3c), representa uma situação em que o QTL controla *y1* e *y2* independentemente.

Li et al. (2006) apresentou uma metodologia alternativa para analisar o relacionamento entre genótipo → fenótipo (genótipo causando fenótipo), considerando múltiplos locos, seu método vai além do anterior, não só pela maior quantidade de informação colocada no modelo, mas também por outras formas de relações causais possíveis entre os dados, de tal forma que forneça uma melhor explanação da arquitetura genética entre caracteres quantitativos. Por exemplo, mesmo se houver apenas um QTL e duas características, diversas outras relações causais são possíveis além das três propostas por Schadt et al. (2005), descritas na Figura 3.

O método de Li et al. (2006), compreende cinco passos: inicialmente é feito o mapeamento de QTLs para cada característica individualmente utilizando teste LOD. Em seguida, é feito novo mapeamento de QTLs utilizando outras características como co-variáveis. A escolha das co-variáveis é feita com base em relações biológicas conhecidas, ou um diagrama causal previamente construído com os dados fenotípicos. Nessa situação, os caracteres que causam efeito direto sobre a característica em pauta, são os utilizados como co-variáveis condicionantes. A necessidade da utilização dessas co-variáveis reside no fato de que determinado QTL pode estar sendo detectado na característica *y2* por exemplo, não simplesmente por ele surtir efeito sobre ela, mas sim, por surgir efeito em uma característica correlacionada *y1* que por sua vez, tem relação causal com a dada característica *y2* em questão. Esse processo pode ser chamado condicionado, e quando tais características são não correlacionadas, o processo é dito incondicionado, ambos são representados pelas Figuras 3i e 3b, respectivamente. A limitação da utilização de caracteres como co-variáveis no mapeamento é que, quando determinado caractere é modelado como uma variável independente no mapeamento da variável dependente, os QTLs em comum entre as duas variáveis (dependente e independente) são perdidos, sendo detectados somente aqueles que são exclusivos à variável sendo mapeada.

Quando um QTL é detectado para duas características, *y1* e *y2* por exemplo, elas estão

inter-conectadas, como nos modelos da Figura 3c, 3d e 3e, ou seja, as correlações entre elas serão todas significativamente diferentes de zero, mas com diferentes valores, e possivelmente sinais, para os parâmetros estruturais correspondentes (ROSA et al., 2011).

O terceiro passo do método de Li et al. (2006), refere-se à construção de um diagrama causal e suas respectivas equações estruturais. Nesta estrutura gráfica, cada característica é representada por um nó, incluindo os QTLs identificados na primeira e segunda etapa. Setas devem ser colocadas entre os QTLs e as características, e também devem ser colocadas entre as características com relações causais existentes.

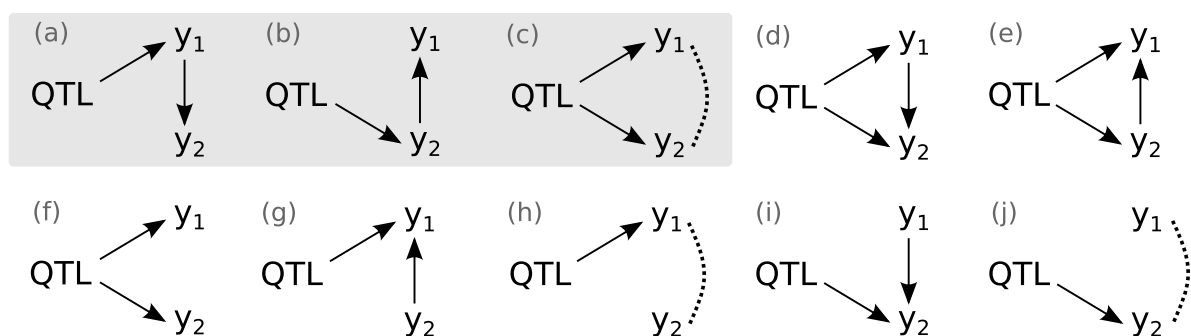


Figura 3 – Possíveis relações estruturais existentes entre um QTL e dois caracteres quantitativos (y_1 e y_2). (a), (b) e (c) correspondem às relações: causal, reativa e independente, respectivamente, propostas por Schadt et al. (2005). Além dessas três, as demais correspondem às relações propostas por Li et al. (2006), em que redes de maior complexidade são admissíveis nos modelos estruturais. As setas representam causalidade, e os pontilhados, correlação não explicada. Adaptada de Rosa et al. (2011)

Outro método para inferir sobre relações causais entre fenótipo e genótipo foi proposto por Liu, Fuente e Hoeschele (2008). Este método que envolve níveis de regulação genética envolvidas com um conjunto de estruturas esparsas dentro de uma rede dirigida de relações causais. Este é um método semelhante ao método proposto por Li et al. (2006), com a diferença que estruturas mais complexas são possíveis dado à não utilização de co-variáveis de características condicionadas. Novas interações podem ser identificadas e ponderadas na análise, além de dados de bioinformática contendo informações sobre o nível de expressão de um número elevado de genes. O método compreende três etapas. Primeiramente, técnicas de detecção de QTLs são utilizadas para identificar regiões cromossômicas envolvidas com a expressão dos caracteres quantitativos. Em seguida, os pares de reguladores-alvo são identificados, ou seja, são relacionados com setas entre os QTLs e as características em que estão envolvidas. Em seguida o modelo é refinado, utilizando os convencionais passos de ajuste de *SEM*.

Chaibub-Neto et al. (2008) propuseram uma metodologia que utiliza tanto informação de QTLs como a de fenótipos, composta por duas etapas principais. A primeira é a construção de um diagrama estrutural contendo uma rede de associações. Em seguida são feitos testes *LOD-score* para determinar a direção causal para cada relação que conecta um par de fenótipos, condicionados pelo efeito dos QTLs encontrados. Esse método foi avaliado utilizando simulações computacionais, que mostraram que ele é capaz de inferir sobre as relações causais com taxas de confiabilidade elevadas. No entanto, embora este método possa ser aplicado em populações F1-segregantes, ele depende muito da disponibilidade de informações confiáveis de QTLs (ROSA et al., 2011).

Uma vez que as abordagens de mapeamento de QTLs tradicionais não levam em consideração a estrutura causal fenotípica, tais análises podem identificar QTLs individualmente para cada característica, que posteriormente possam ser colocados em uma rede causal, para que os efeitos diretos e indiretos de cada um sobre os caracteres quantitativos possam ser obtidos. Por exemplo, considere o gráfico demonstrado por (ROSA et al., 2011), em que cinco caracteres (y_1 , y_2 , y_3 , y_4 e y_5) e três QTLs ($QTL1$, $QTL2$ e $QTL3$) são colocados em um modelo estrutural (Figura 4a). Neste podemos ver que, alguns QTLs têm efeitos diretos sobre as características, que são aqueles que justamente foram encontrados pelo método utilizado de mapeamento de QTLs. Contudo, outros QTLs que não foram significativos pela metodologia de mapeamento empregada, também, podem surtir influência indireta nos caracteres não diretamente ligados a eles, por intermédio da correlação existente entre os caracteres fenotípicos.

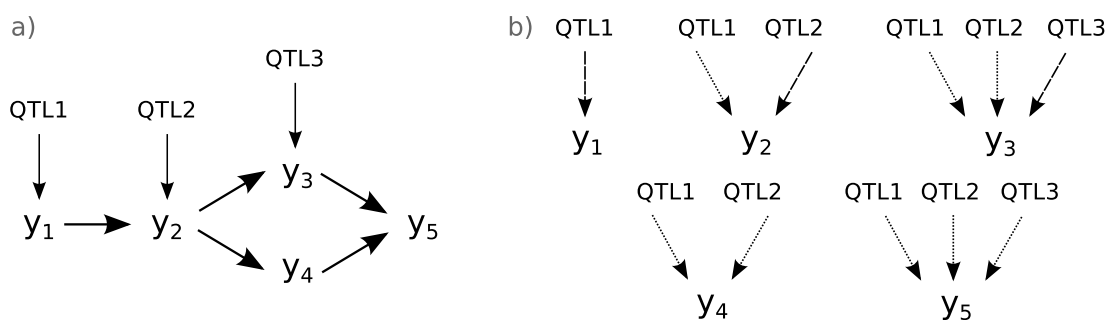


Figura 4 – Modelo estrutural hipotético com dados fenotípicos (y_1 , y_2 , y_3 , y_4 e y_5) e QTLs ($QTL1$, $QTL2$ e $QTL3$). (a) Diagrama causal. (b) relações diretas e indiretas existentes entre genótipos e fenótipos. As setas tracejadas correspondem às relações diretas e as pontilhadas às relações indiretas. Adaptada de Rosa et al. (2011)

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Descrição dos Dados Experimentais

O plantio foi instalado no ano de 2006, em uma área da empresa *Michelin LTDA* chamada *Michelin Edouard Plantation*, localizada no Mato Grosso, Brasil, com as coordenadas $17^{\circ}23'59,60''S$ e $54^{\circ}44'53,93''W$, em altitude de aproximadamente 519 metros e relevo fortemente plano.

Foram utilizados 263 progênies de irmãos completos como tratamentos, oriundas do cruzamento entre dois clones geneticamente divergentes e comumente empregados em plantios comerciais de empresas brasileiras produtoras de látex (PB217 e PR255). O clone PB217 apresenta alto potencial de produção de látex com elevadas concentrações de sacarose em sua composição, característica essa, que é expressa durante todo o ciclo produtivo do genótipo. O Clone PR255 tem como característica principal a resistência a danos e ao frio e maior crescimento.

Além da progênie, foram utilizados outros materiais genéticos como testemunhas: meio-irmãos, clones comerciais e os próprios parentais. Os tratamentos e testemunhas foram distribuídos em delineamento experimental de blocos casualizados com quatro repetições (blocos). Dessa maneira, cada bloco consistiu de 280 parcelas (263 progênies de irmãos completos, os 2 parentais e as demais 15 testemunhas), sendo que cada uma continha quatro replicatas do mesmo genótipo (clones), geradas por enxertia, pela técnica de borbulhia, caracterizando assim, parcelas quadradas de 2×2 plantas cada. O experimento total ocupou 7,2 hectares e as parcelas foram dispostas em 40 linhas e 28 colunas e os procedimentos de manejo de campo foram os mesmos utilizados para fins comerciais.

Os componentes de crescimento periódicos utilizados foram: o incremento em diâmetro (medido a 1,0 metro do solo) e o incremento em altura. Os componentes foram calculados com base nas medidas de outubro de 2007 à outubro de 2009, quando as plantas apresentam 18, 24, 30, 36 e 42 meses de idade. Sendo assim, as variáveis utilizadas nesse trabalho foram definidas como a diferença de crescimento entre as medições feitas em cada idade, totalizando 4 medidas de período (entre outubro a abril e entre abril a outubro) e uma medida de incremento total, que

é todo crescimento das plantas ao longo dos dois anos de medição. Dessa forma, foram adotados os códigos *D1*, *D2*, *D3*, *D4*, e *DT* para os componentes de diâmetro nos períodos 1, 2, 3, 4 e total, e *H1*, *H2*, *H3*, *H4* e *HT* para altura para os componentes de altura para os mesmos períodos.

Estes dois períodos durante o ano foram definidos com os termos estação favorável e estação adversa, por apresentarem condições tipicamente vantajosas e desvantajosas, respectivamente, para a seringueira.

As informações sobre o experimento estão resumidas na Figura 5:

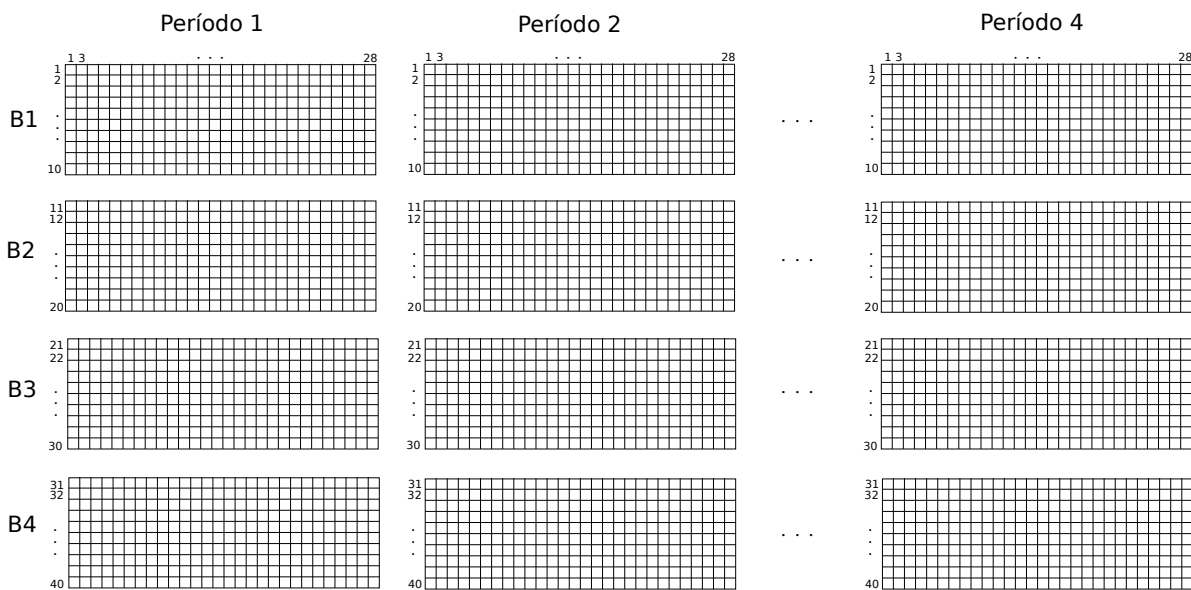


Figura 5 – Croqui do experimento: Delineamento em blocos casualizados, com 4 repetições (B1, B2, B3 e B4). Cada bloco contém 280 parcelas quadradas de 4 plantas, distribuídas em 40 linhas e 28 colunas. Foram obtidas 4 medidas de incremento, correspondentes a 4 períodos consecutivos

Margeando o experimento, foram incluídas duas bordaduras, uma localizada acima da linha 1, e a outra abaixo da linha 40, contendo clones de seus genitores. Entre cada bloco, foi instalado aceiros com tamanho de 6 metros, para facilitar o manejo, o tráfego de pessoal e o acesso às parcelas.

Na conjuntura de condições climáticas, a área de implantação do experimento é caracterizada por clima subtropical com estação de seca e temperatura mais baixa, que dura quatro meses de maio a setembro (SOUZA et al., 2013).

A seringueira é uma árvore de origem tropical, ou seja, adaptada às regiões de clima quente e úmido o ano inteiro (RAO; SARASWATHYAMMA; SETHURAJ, 1998). Quando cultivada em outros locais que não seu centro de origem, a espécie pode ser submetida a diferentes condições

de solo e clima os quais não é naturalmente adaptada.

Para entender melhor o comportamento dos genótipos quando submetidos a diferentes condições climáticas, alguns dados meteorológicos como temperatura máxima, mínima e média, em graus celsius ($^{\circ}\text{C}$), e os índices pluviométricos, em milímetros por metro quadrado (mm/m^2 , ou simplesmente mm), foram diariamente coletados, durante os dois anos de medições dos dados fenotípicos que foram utilizados neste trabalho. Dessa maneira foi possível empregar os dados de média de temperatura, amplitude de temperatura e precipitação como covariáveis no modelo fenotípico.

3.1.2 Descrição dos Dados Moleculares e Grupos de Ligação

As análises químico-laboratoriais, como as extrações de DNA e a definição dos marcadores microssatélites, bem como a construção do mapa de ligação, foram feitas por Souza et al. (2013). Os marcadores apresentaram padrões de segregação de 1:1:1:1, 1:2:1, 3:1 e 1:1 e foram atribuídos a 23 grupos de ligação com tamanho total de 2792,8 centi-Morgan (cM), corrigidos pela função de mapeamento *Kosambi*.

Os grupos de ligação construídos por Souza et al. (2013), foram organizados de acordo com o mapa previamente desenvolvido por Lespinasse et al. (2000) e informações de outros mapas (não publicados) serviram como suporte para identificar marcadores sintênicos. Além disso, devido à razoável densidade de ocupação dos marcadores nos 23 grupos de ligação, os grupos foram classificados e numerados de 1 a 19, em que, os grupos 2, 6, 11 e 17 foram divididos em subgrupos "a" e "b", o que deixou o mapa mais realista, visto que, é esperado o número de 18 cromossomos homólogos para a espécie ($2n = 36$).

Os polimorfismos de um total de 603 marcadores microssatélites de diferentes origens foram testadas tanto para o parental PB217, como para PR255. Dos 603 marcadores testados, entre SSR's e EST-SSR's, 284 foram segregantes na população de mapeamento, sendo utilizados no mapeamento de QTLs.

3.2 Métodos

3.2.1 Modelo Misto: Dados Fenotípicos

A partir dos dados coletados nos experimentos e os dados climáticos, a primeira etapa do trabalho consistiu em ajustar um modelo que explicasse o comportamento dos genótipos nas condições de desenvolvimento. Para tanto, foram ajustados e avaliados diferentes modelos de

estruturas de variância-covariância (VCOV) para os efeitos genéticos dos caracteres dentro dos períodos, e as estruturas VCOV que corrigissem da maneira mais apropriada os resíduos experimentais.

O ajuste e comparação destes modelos foi possível através da utilização da 16^a edição do software *Genstat* (PAYNE; WELHAM; HARDING, 2013), que foi utilizado para a realização das análises com modelos mistos, através dos algoritmos presentes no domínio chamado análise REML do software.

Dentro de cada parcela experimental, foram feitas médias entre os 4 clones presentes, ou eventualmente, entre os sobreviventes. Não foram estimados os erros dentro das parcelas.

Inicialmente foram testados dois modelos independentes para as características diâmetro e altura e posteriormente um único modelo multi-caracteres, capaz de ajustar as médias genotípicas conjuntamente para as duas características (MALOSETTI et al., 2008). Os métodos e estratégias para escolher o melhor modelo serão abordadas a seguir.

A seleção de modelos foi feita com base no menor valor do critério de informação bayesiano (*BIC*) (SCHWARZ, 1978). O Critério de informação de Akaike (*AIC*) (AKAIKE, 1974) serviu como suporte na escolha dos mesmos, porém não foi determinante para a seleção dos melhores modelos.

3.2.2 Modelo para Características Individuais

Para melhor entender o comportamento das variâncias e covariâncias entre os níveis das fontes de variação sistemáticas e aleatórias, foram testados modelos para cada característica separadamente. Como se tratam de medidas repetidas no tempo, um modelo de blocos com parcelas subdivididas no tempo foi pensado como um possível candidato de melhor ajuste aos dados (Equação 4). Contudo, a presença de informações de linhas e colunas entre e dentro dos blocos (Figura 5), e diferentes dados meteorológicos entre os períodos destacando a heterogeneidade climática entre eles (Figura 7), fizeram com que o modelo de parcelas subdivididas fosse gradativamente adaptado, tornando-o mais realista.

A inclusão de linhas e colunas no modelo torna a fonte de variação do bloco ambígua, pois tal controle espacial contém toda informação presente nos blocos, e como bônus, acrescenta dados de posição de cada árvore no experimento, fazendo com que o modelo seja capaz de captar pequenas variações ambientais dentro do bloco (PAYNE; WELHAM; HARDING, 2013). Por isso, com a adição destas duas fontes de variação ao modelo (Equação 5), a manutenção dos dados de bloco é redundante.

No modelo clássico de parcelas subdivididas no tempo, há a indexação de um fator contendo informação de períodos, tempo ou o momento em que cada variável foi mensurada (LITTELL; HENRY; AMMERMAN, 1998). Este fator é utilizado para estimar os efeitos de interação entre os genótipos com os períodos. No modelo aqui utilizado, essa indexação, que tipicamente é feita de forma qualitativa (discreta), foi feita quantitativamente, acrescentando as informações climáticas de cada período. Esse tipo de informação é útil para aproximar estimativas de períodos mais semelhantes, e não simplesmente considerar todos os períodos como se fossem completamente distintos.

As variáveis temperatura média, amplitude da temperatura e precipitação total foram incluídas como covariáveis de efeitos fixos nos modelos (Equação 6). Foi avaliado se estas variáveis eram capazes de explicar estatisticamente toda variação contida na variável discreta períodos. Os efeitos que foram significativos foram mantidos nos modelos. De forma análoga ao controle ambiental de linhas e colunas, o fator período foi retirado do modelo por conter as informações fornecidas pelas covariáveis climáticas.

Cada período apresentou apenas um valor climático para cada covariável testada, ou seja, foi assumido que todas as árvores dentro de um mesmo período são representadas por valores correspondentes à média de temperatura, amplitude térmica e precipitação total. Dessa forma, a variação presente nas alterações meteorológicas está contida entre os períodos e não dentro deles.

A transição entre o modelo de parcelas subdivididas no tempo para o modelo individual final descrito acima está sucintamente apresentado nas seguintes 3 equações:

$$y_{ijt} = \mu + \beta_1 B_j + \underline{G}_i + \underline{BG}_{ji} + \beta_2 P_t + \underline{PG}_{ti} + \underline{\varepsilon}_{tji} \quad (4)$$

em que: y_{ijt} é valor medido do i -ésimo genótipo no j -ésimo bloco no t -ésimo período; μ é o intercepto; $\beta_1 B_j$ é o efeito fixo de bloco com respectivo parâmetro β ; \underline{G}_i é efeito aleatório do i -ésimo genótipo; \underline{BG}_{ji} é o erro entre parcelas, ou erro A; $\beta_2 P_t$ é o efeito fixo do t -ésimo período com seu respectivo parâmetro; \underline{PG}_{ti} é a interação genótipo \times período; $\underline{\varepsilon}_{tji}$ é o erro geral do experimento, ou erro B.

$$y_{imnt} = \mu + \underline{L}_m + \underline{C}_n + \underline{LC}_{mn} + \underline{G}_i + \underline{LCG}_{mni} + \beta P_t + \underline{PG}_{ti} + \underline{\varepsilon}_{tmni} \quad (5)$$

em que: y_{imnt} é o valor medido do i -ésimo genótipo no m -ésima linha e n -ésima coluna no t -ésimo período; μ é o intercepto; \underline{L}_m é o efeito aleatório da m -ésima linha; \underline{C}_n é o efeito aleatório da n -ésima coluna; \underline{LC}_{mn} é a interação espacial linha \times coluna; \underline{G}_i é o efeito aleatório

do i -ésimo genótipo; \underline{LCG}_{mni} é o erro entre parcelas, ou erro A; βP_t é o efeito fixo do t -ésimo período; \underline{PG}_{ti} é a interação genótipo \times período; $\underline{\varepsilon}_{tmni}$ é o erro geral do experimento, ou erro B.

$$y_{imnt} = \mu + \underline{L}_m + \underline{C}_n + \underline{LC}_{mn} + \underline{G}_i + \underline{LCG}_{mni} + \sum_1^k \beta_k Q_{ti} + \underline{PG}_{ti} + \underline{\varepsilon}_{tmni} \quad (6)$$

em que: y_{imnt} é o valor medido do i -ésimo genótipo na m -ésima linha e n -ésima coluna no t -ésimo período; μ é o intercepto; \underline{L}_m é o efeito aleatório da m -ésima linha; \underline{C}_n é o efeito aleatório da n -ésima coluna; \underline{LC}_{mn} é a interação espacial linha \times coluna; \underline{G}_i é o efeito aleatório do i -ésimo genótipo; \underline{LCG}_{mni} é o erro entre parcelas, ou erro A; $\sum_1^k \beta_k Q_{it}$ é o efeito fixo do t -ésimo dado meteorológico da k -ésima covariável sobre o i -ésimo genótipo; \underline{PG}_{ti} é a interação genótipo \times período; $\underline{\varepsilon}_{tmni}$ é o erro geral do experimento, ou erro B.

Para refinar o controle ambiental espacial, buscou-se estruturas de VCOV que explicassem melhor a covariância incremental presente na interação entre linhas e colunas. Por exemplo, espera-se que plantas mais próximas umas às outras tenham efeito ambiental mais semelhante que plantas mais distantes. Seguindo este raciocínio, há maior covariância entre plantas próximas, e essa vai diminuindo gradativamente de acordo com o distanciamento das plantas. Para verificar essa hipótese, foram testadas as estruturas de VCOV diagonal (DIAG), auto-regressiva homogênea (AR), auto-regressiva heterogênea (AR_{HET}) e não estruturada (UNS).

A matriz de estrutura DIAG apenas estima as variâncias heterogêneas presentes em cada uma das linhas e colunas, assumindo ausência de correlações entre elas. A matriz AR contém covariâncias entre as variáveis de controle espacial, que vão reduzindo a medida que os índices forem se distanciando, e possuem variância comum na diagonal. A matriz AR_{HET} além das covariâncias presentes na AR, contém as variâncias correspondentes a cada nível do fator na diagonal. Por fim, a matriz UNS, apresenta todas variâncias e covariâncias entre os níveis sem restrição de número de parâmetros. Entretanto, a utilização desta estrutura de VCOV se tornou impraticável, por existirem muitos níveis de linhas e colunas, elevando exorbitantemente a quantidade de parâmetros a serem estimados.

Algumas outras estruturas de VCOV plausíveis também foram utilizadas na tentativa de melhor explicar a covariância existente entre os períodos. Porém, a escolha destas estruturas de VCOV foi gravemente comprometida pelo fato de períodos semelhantes serem intercalados, e não sequenciais, o que seria um forte argumento para utilização de alguma estrutura da família das auto-regressivas. Portanto estruturas adotadas foram a DIAG, fator analítico (FA) e UNS. O critério para seleção destas matrizes foi simplesmente fato de serem estruturas amplas que

consideram a variação particular de cada fator, mas não apresentam relações progressivas em suas células matriciais.

Os procedimentos adotados até o momento são de natureza residual, ou seja, estas melhorias do modelo e avaliações de estruturas de VCOV, estão contidas na parte **R** da matriz aleatória do modelo misto. Porém, como o ajuste do modelo, tem como finalidade principal o cálculo das médias genóticas ajustadas dos materiais genéticos utilizados, a parte **G** da matriz também foi modelada.

A parte **G** do modelo contém a interação genótipo \times período, refletindo o fato que os diferentes materiais genéticos podem ter diferentes padrões de crescimento, não só pelas diferentes capacidades genético-adaptativas ao meio, como também, pelas capacidades genético-fisiológicas de desenvolvimento ao longo do tempo. Por exemplo, alguns indivíduos tem maior arranque inicial. Porém, em função do gasto de reservas, estagnam seu crescimento com o passar do tempo (TAIZ; ZEIGER, 2010). Em contrapartida, outros podem ter menor desenvolvimento inicial, e se destacarem em momentos posteriores.

Assim, há um forte indício sobre a existência de variância e covariância genética implícita dentro de cada período e entre períodos. Para testar essa hipótese, foram adotadas as estruturas de VCOV **DIAG**, que considera somente a variância genotípica dentro de cada período na diagonal, assumindo ausência de correlação entre os períodos; e a estrutura **FA**, que é um sistema multiplicativo que permite heterogeneidade de variâncias e covariâncias genéticas entre os períodos. A **FA**, tenta simular à estrutura **UNS**, porém com a vantagem de requerer um número menor de parâmetros (OMAN, 1991).

3.2.3 Modelo Multi-Caracteres-Multi-Períodos

Ao assumirmos independência entre as características, a correlação genética existente entre elas não é considerada. E se tratando de dois caracteres altamente correlacionados, como a altura e o diâmetro, a manutenção desta relação entre elas é de fundamental importância não somente para a construção do modelo estrutural foco deste trabalho, como também para aumentar a acurácia de predição dos valores genotípicos (MALOSETTI et al., 2008). Isso ocorre pelo fato do modelo utilizar os dois conjuntos de dados como “repetições” dos valores, porém, obviamente, considerando a variância individual de cada um dos caracteres e a covariância presente entre eles (PIEPHO, 2000).

Por todos esses motivos, foi feita a última expansão do modelo, que foi a transformação do modelo para as características individuais (Equação 6) no modelo multi-caracteres abaixo:

$$y_{imnt} = \mu + \underline{L}_{cm} + \underline{C}_{cn} + \underline{LC}_{cmn} + \underline{G}_{ci} + \underline{LCG}_{cmni} + \sum_1^k \beta_{ck} Q_{cti} + \underline{PG}_{cti} + \underline{\varepsilon}_{ctmni} \quad (7)$$

em que: y_{imnt} é o valor medido do i -ésimo genótipo no m -ésima linha e n -ésima coluna no t -ésimo período para o c -ésimo caráter; μ é o intercepto; \underline{L}_{cm} é o efeito aleatório da m -ésima linha para o c -ésimo caráter; \underline{C}_{cn} é o efeito aleatório da n -ésima coluna para o c -ésimo caráter; \underline{LC}_{cmn} é a interação espacial linha \times coluna para o c -ésimo caráter; \underline{G}_{ci} é o efeito aleatório do i -ésimo genótipo para o c -ésimo caráter; \underline{LCG}_{cmni} é o erro entre parcelas, ou erro A; $\sum_1^k \beta_{ck} Q_{cti}$ é o efeito fixo do t -ésimo dado meteorológico da k -ésima covariável sobre o i -ésimo genótipo para o c -ésimo caráter; \underline{PG}_{cti} é a interação genótipo \times período para cada caráter c ; $\underline{\varepsilon}_{ctmni}$ é o erro geral do experimento, ou erro B.

A inclusão da informação da existência de dois caracteres no mesmo modelo é feita de modo que todos os efeitos presentes no modelo individual sejam aninhados quanto à informação de caráter, no formato de interação. Este aninhamento é de ordem tanto genética, como residual. Logo, outras duas estruturas VCOV são necessárias no ajuste do modelo, uma para a variância e covariância genética para cada um dos caracteres, e outra para a variância e covariância residual entre os dois.

Estas duas estruturas adicionais foram testadas, inicialmente considerando apenas a variância particular presente em cada um dos caracteres, utilizando a estrutura VCOV diagonal, e em seguida, considerando a variância e covariância entre elas, adotando a estrutura UNS. A adoção de estruturas VCOV alternativas plausíveis nesse caso é nula, pois se tratando de somente dois níveis em um fator, outras matrizes como por exemplo AR, AR_{HET} , simetria composta homogênea (CS) ou heterogênea (CS_{HET}) e a FA, seriam reproduções das matrizes DIAG ou UNS.

Embora o modelo multi-caracteres utilizado seja uma transformação matemática do modelo individual, tratam-se de modelos diferentes, com números de parâmetros diferentes. Assim, as estruturas das matrizes **G** e **R** candidatas foram novamente testadas. Como para o modelo individual foram testadas 64 combinações possíveis, se estas mesmas estruturas fossem combinadas com as estruturas VCOV candidatas para as interações entre os caracteres do modelo multi-caracteres, o número de combinações possíveis seria 3 vezes maior, ou seja, 192 novas combinações necessitariam ser testadas. Como o volume de dados é aproximadamente o dobro, a avaliação deste número de modelos torna-se inexecutável computacionalmente. Dessa maneira, as estruturas dos modelos individuais que foram coincidentes para os dois modelos individuais

foram mantidas, reduzindo o espaço de busca.

Os modelos candidatos foram comparados utilizando os critérios *AIC* (AKAIKE, 1974) e *BIC* (SCHWARZ, 1978).

3.3 Mapeamento dos QTLs

Das 263 progênies disponíveis, 252 indivíduos foram utilizados para compor a população de mapeamento devido à contaminações nos cruzamentos e/ou perda de parcelas.

O mapeamento de QTLs foi feito por meio da utilização de um conjunto de funções do *R* elaboradas por Gazaffi (2013), que utiliza os conceitos de Wu et al. (2002) para lidar com progênies de irmãos completos originadas pelo cruzamento de indivíduos *outcrossing* (Populações do tipo F1-segregante). As informações de marcadores com padrões de segregação 1:1:1:1, 1:2:1, 3:1 e 1:1 são incluídas nas análises, permitindo de estimar os efeitos aditivos dos dois parentais e o efeito de dominância para cada loco marcador.

A metodologia adotada por Gazaffi (2013) foi o mapeamento por intervalo composto (*composite interval mapping, CIM*). Foram estimadas as probabilidades de segregações para marcadores virtuais a cada cM entre os marcadores reais utilizados, como descrito por Zeng (1994). Os próprios marcadores significativos foram incluídos como co-fatores nas análises para controlar a variação devido à presença de QTLs fora do intervalo de mapeamento que podem levar à detecção de QTLs fantasmas.

As médias genotípicas ajustadas dos caracteres diâmetro e altura para cada período provenientes do modelo misto multi-caracteres foram utilizadas separadamente no mapeamento de QTLs. Foram considerados os efeitos aditivos dos pais e os efeitos devido à dominância genética dos caracteres, conforme Gazaffi (2013).

A estimativa dos preditores genéticos ao longo de todo o genoma foi feita com base no cálculo da probabilidade condicional do genótipo dos QTLs considerando o mapa estimado por Souza et al. (2013).

Para identificação dos efeitos aditivos e dominantes dos genitores, foi adotado o modelo (Equação 8) proposto por Gazaffi (2013). Este modelo inclui os co-fatores para controle da variação causada por QTLs fora do intervalo mapeado, o que aumenta o poder estatístico do mapeamento.

$$y_j = Z_j\gamma + \alpha_p^*x_{pj}^* + \alpha_q^*x_{qj}^* + \delta_{pq}^*x_{pj}^*x_{qj}^* + e_j \quad (8)$$

em que, y_j é o fenótipo do j -ésimo indivíduo ($j = 1, \dots, n$); Z_j é j -ésima linha da matriz indicadora

Z_j , com dimensão $n \times (1 + 3c)$, que contém uma coluna de 1's e variáveis indicadoras dos genótipos dos c cofatores (marcadores moleculares fora da região sendo mapeada), de acordo com os contrastes; γ é o vetor de parâmetros $(1 + 3c) \times 1$ contendo o intercepto (μ) e coeficientes de regressão linear múltipla (α_p, α_q e δ_{pq}) para os cofatores; α_p^* e α_q^* são os efeitos aditivos para o QTL nos genitores P e Q, respectivamente; δ_{pq}^* é o efeito da interação intra-loco (dominância) entre os efeitos aditivos; e_j erro.

Os dois primeiros contrastes (α_p e α_q) utilizados no modelo, representam os efeitos aditivos dos QTLs nos genitores P e Q (são os parentais PB217 e PR255), enquanto que o terceiro contraste (δ_{pq}) representa o efeito de dominância ou interação intra-loco entre os efeitos aditivos de cada genitor. Para simplificar, os contrastes α_p, α_q e δ_{pq} neste trabalho foram chamados de α_1, α_2 e δ , respectivamente.

A seleção dos cofatores neste trabalho, foi feita com base na inclusão dos marcadores via *stepwise forward* (HOCKING, 1976), em que, cada marcador, um a um, foi incluído e avaliado estatisticamente, ao nível de significância de 95% de probabilidade, até que todos os marcadores utilizados fossem significativos ao mesmo nível de significância.

Para definir os valores *LOD-score* referentes ao limiar de inclusão de QTL (*threshold*) foi feito teste de permutação com 1000 iterações, adoção da janela entre marcadores (*window size*) de 20 cM e probabilidade $\alpha = 95\%$. O estabelecimento do *threshold* foi feito com base na metodologia de (CHURCHILL; DOERGE, 1994).

Uma vez definido os QTLs para todos os componentes periódicos de crescimento e os crescimentos totais para as duas variáveis, os locos significativos encontrados nas regiões de interseção dos intervalos de confiança para dois ou mais componentes ou totais, tanto para diâmetro como para altura, foram definidos como um mesmo QTL. Esse procedimento foi adotado devido às estimativas das probabilidades dos locos nos intervalos condicionados aos marcadores adjacentes serem muito parecidas de acordo com a proximidade destes locos. Tal semelhança ocasionaria um efeito colinearidade na análise estrutural genotípica, possivelmente viesando o modelo.

3.4 Modelos de Equações Estruturais

Neste trabalho, a técnica *SEM* foi utilizada essencialmente para estimar as relações causais em dois momentos, inicialmente entre os valores genotípicos estimados pelo modelo misto descrito acima, gerando suporte para criação de um diagrama causal o mais parcimonioso o quanto fosse possível mas que ainda assim indicasse as direções das causalidades e seus efeitos, na forma

de um modelo estrutural. Posteriormente esse modelo foi utilizado para servir como esqueleto genético para incorporação de QTLs significativos encontrados através da metodologia de mapeamento por intervalo composto (*CIM*) para espécies *outcrossing* (GAZAFFI, 2013).

3.4.1 Modelo Estrutural Fenotípico

Para realizar essa parte do trabalho, além dos *e-BLUPs* genotípicos ajustados pelo modelo fenotípico, utilizados como variáveis endógenas observadas, também foram incluídas variáveis latentes que fizessem a conexão entre os períodos climáticos mais semelhantes. A primeira etapa da construção do modelo estrutural foi elaborada a partir de um modelo primário hipotético com informação teórica prévia que refletia conceitos biológicos bem conhecidos. Esta etapa, conhecida como *Especificação do Modelo*, serviu como ponto de partida para a modelagem *SEM* e as pressuposições utilizadas estão citadas a seguir:

- O crescimento em altura e diâmetro possuem relações causais bem conhecidas na pesquisa científica florestal, como a relação hipsométrica entre estas duas variáveis (CURTIS, 1967; KNOWE, 1994; NIGH; LOVE, 1999; EERIKAINEN, 2003). Essa relação é baseada na alta correlação presente entre estes fenótipos e por representar um padrão biológico. Diversos fatores a influenciam, como por exemplo a estrutura da floresta, idade das árvores, material genético e a qualidade do local de plantio.
- O crescimento em altura se dá a partir do desenvolvimento do meristema primário da planta, enquanto que o crescimento em diâmetro se dá pelo desenvolvimento do meristema secundário (TAIZ; ZEIGER, 2010). Inicialmente a árvore investe suas reservas em crescimento vertical, com o objetivo de atingir o mais breve possível o topo da floresta e se destacar perante às suas vizinhas competidoras. Em seguida, após a estabilização do porte, ela emprega suas forças no processo de crescimento em diâmetro (SADIQ; SMITH, 1983). Embora isso gere a interpretação de que as correlações entre a altura e o diâmetro sejam significativas somente a partir de um estágio determinado de desenvolvimento, o sentido de causalidade altura \rightarrow diâmetro está presente em todas as etapas do ciclo de vida da planta, isso porque, apesar de indivíduos com maior estatura terem maior estabilidade para crescer em diâmetro, ao mesmo tempo, eles precisam de suporte anatômico do diâmetro para crescer em altura (TAIZ; ZEIGER, 2010).
- Um estágio atual pode causar o próximo, e assim sucessivamente. E essas relações são diretas e não bifurcam, com um momento atual causando um momento futuro, não direta-

mente ligado. Por exemplo, um primeiro momento causa o segundo que por sua vez causa o terceiro e assim por diante, mas nunca o primeiro causa diretamente o terceiro, e sim, indiretamente por meio do segundo. Isto caracteriza um processo markoviano (RABINER, 1989).

- Cada período de desenvolvimento tem um sentido biológico diferente para o vegetal, tanto na essência fisiológica da direção momento juvenil para maturidade, como nas respostas aos variados estímulos ambientais presente em cada um dos variados períodos (TAIZ; ZEIGER, 2010). Este último pôde ser amparado com a adição de parâmetros (variáveis endógenas latentes) que incorporem ao modelo a semelhança residual entre períodos congêneres, e não simplesmente os declarem como quatro épocas distintas.

Estas relações foram consideradas para construção do primeiro diagrama estrutural, que serviu como suporte para derivação do modelo definitivo.

No próximo passo de elaboração do modelo estrutural, foi verificado o número de parâmetros utilizados em relação ao número de variáveis observadas. Esta etapa é conhecida como *Identificação do Modelo*, onde o número de parâmetros a serem estimados devem ser compatíveis com número variáveis instrumentais utilizadas.

As duas etapas seguintes são: *Estimativa do Modelo*; e *Testes do Modelo*. Nestas etapas efeitos diretos e indiretos e os valores de probabilidade de não rejeição das relações estruturais e do modelo completo foram estimadas utilizado o pacote *sem* do R feito por Fox (2006). Este pacote contém funções para os ajuste de modelos de equações estruturais generalizadas, tanto com variáveis observadas como com variáveis não observadas (latentes). Ele utiliza a abordagem de *Reactive approach motivation (RAM)* para compor os modelos de equações estruturais e utiliza a metodologia de mínimos quadrados parciais em dois estágios (*2SLS*).

A etapa final gera uma cadeia de *feedback* com as duas etapas anteriores. Na *Modificação do Modelo*, os efeitos não significativos obtidos, foram substituídos ou alterados, fundamentado em algum argumento biológico, até que o modelo final fosse concluído.

No processo de seleção dos modelos descrito acima, foram considerados os Critérios de seleção *AIC* (AKAIKE, 1974) e *BIC* (SCHWARZ, 1978).

3.4.2 Modelo Estrutural Genético

Depois do modelo estrutural fenotípico definido, a última parte deste trabalho foi a incorporação dos efeitos dos QTLs ao modelo selecionado. Cada QTL contém três contrastes referentes

aos efeitos genéticos aditivos de cada um dos dois parentais, e um efeito devido à dominância do loco. Estes efeitos foram utilizados como variáveis exógenas observadas no modelo estrutural.

Na primeira etapa, a de *especificação do modelo*, todos os efeitos aditivos para ambos os pais e os efeitos de dominância, encontrados através do mapeamento de QTLs, foram colocados como agentes causadores de variância genética das características de crescimento periódico.

Na etapa de *identificação do modelo*, foram verificados o número de graus de liberdades do modelo, e também a compatibilidade entre o número de variáveis instrumentais com o número de parâmetros. O ajuste deste modelo estrutural também contou com a utilização do pacote *sem* do R (FOX, 2006), nas etapas de *estimativa do modelo* e *testes do modelo*.

Na última etapa, a de *modificações do modelo*, aqueles contrastes dos QTLs que não foram significativos foram retirados gradativamente do modelo pela técnica de *stepwise backward* (HOCKING, 1976). Dessa forma, somente aqueles efeitos genéticos parentais e de dominância que surtiram efeito significativo genético foram mantidos.

Tal como no modelo estrutural fenotípico, o processo de seleção dos modelos descrito acima foi feito com base nos critérios de seleção *AIC* (AKAIKE, 1974) e *BIC* (SCHWARZ, 1978).

4 RESULTADOS

4.1 Análise Exploratória dos Dados Experimentais

Entre os quatro períodos de crescimento utilizados neste estudo, o primeiro e o terceiro, e o segundo e o quarto, foram os mais semelhantes entre si no que diz respeito à média e variância de incremento diâmetro e altura das árvores.

Fazendo uma análise entre estes períodos, tanto para a variável diâmetro, como para a variável altura, os períodos 3 e 4 (que são os de idade mais avançada entre os quatro) obtiveram médias superiores aos seus respectivos períodos semelhantes 1 e 2.

Os períodos de maior crescimento foram também aqueles que apresentaram maiores variâncias, ou seja, houve maior amplitude de incremento entre os indivíduos destes dois períodos, que entre os mesmos indivíduos medidos nos períodos 2 e 4.

As testemunhas não apresentaram diferença significativa, ao nível de 95% de probabilidade, de crescimento quando comparadas aos irmãos completos utilizados como população de mapeamento. O parental PR255 foi superior ao parental PB217 nas duas características de crescimento em todos os períodos, exceto para o caráter altura no período 3 (Figura 6).

Analisando os valores meteorológicos mensurados diariamente durante os períodos, foi possível constatar que o primeiro e o terceiro, e o segundo e o quarto período apresentaram semelhanças tanto nas medidas de temperatura como nas medidas de precipitação.

Os períodos entre outubro a abril apresentaram temperaturas mais elevadas. Nestes períodos sendo de 24, 82 e 25, 90°C em média, para cada período respectivamente. Os valores de 2, 78 e 6, 13°C de desvio padrão que representam a amplitude de temperatura para os referidos períodos. Já os períodos entre abril a outubro, apresentaram temperaturas moderadamente mais baixas, sendo que, os períodos 2 e 4 obtiveram médias de 23, 05 e 23, 44°C com amplitude de 9, 36 e 8, 37°C.

Com relação à precipitação, os períodos 2 e 4 apresentaram ambos baixos índices pluviométricos, porém o período 2 foi marcado por ter tido praticamente nenhuma chuva nos meses de junho e julho. Embora estes dois períodos apresentem condições de precipitação semelhantes, o quarto período ocorreu seca consideravelmente menos intensa que o segundo (Figura 7).

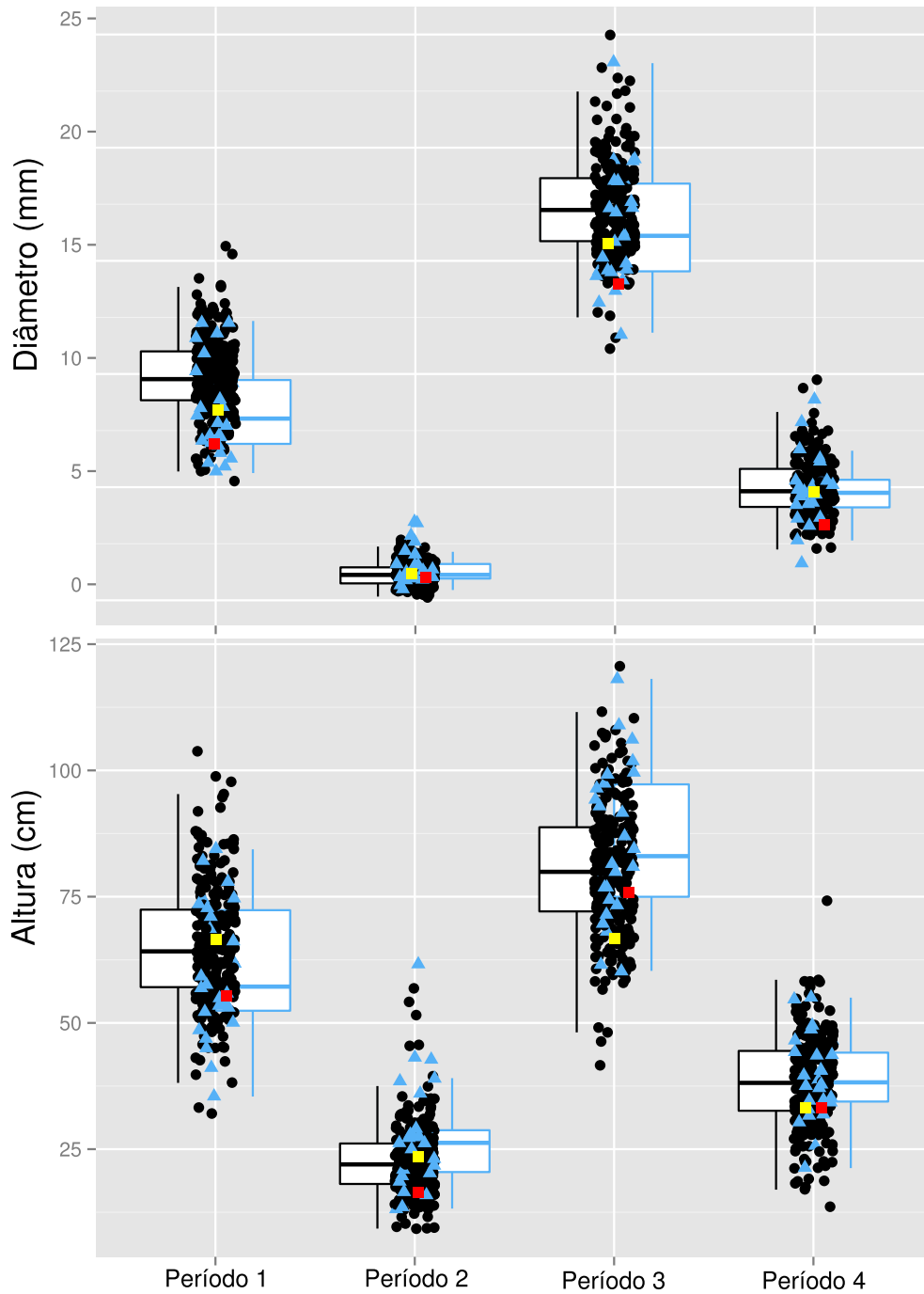


Figura 6 – Distribuição dos dados fenotípicos ao longo dos períodos. Nas ordenadas dos gráficos estão o incremento em diâmetro e altura (figura superior e inferior, respectivamente). Na abscissa está a variação temporal, representada pelos quatro períodos. Os pontos em preto são os valores de cada genótipo de irmãos-completos da população de mapeamento, os triângulos azuis e os quadrados vermelhos e amarelos são as testemunhas, sendo que os quadrados vermelhos representam o parental PB217 e os amarelos o parental PR255. Os *boxplots* em preto são referentes à população de mapeamento e os azuis às testemunhas, em cada período

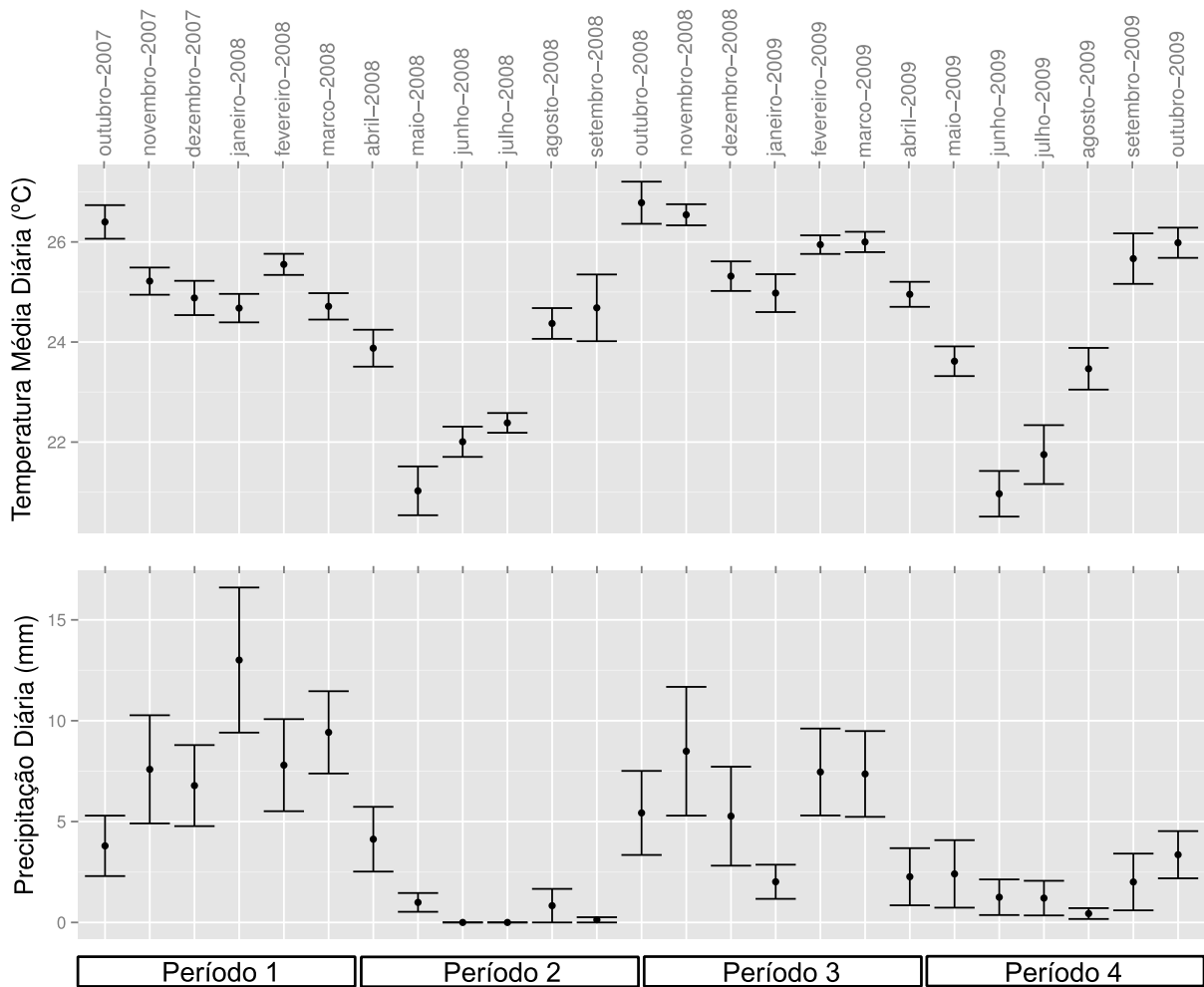


Figura 7 – Distribuição dos valores climáticos nos meses dos períodos avaliados de crescimento da seringueira. Na figura superior, estão os valores de temperatura média (graus Celsius) e na figura inferior estão os valores de precipitação total (milímetros por metro quadrado). Em ambos os gráficos, para cada mês, estão representadas as medidas diárias destas variáveis meteorológicas na forma de média (ponto central) e erro padrão da média (eixos superior e inferior)

4.2 Modelos de Ajuste dos Dados Fenotípicos

Como melhorias no modelo de parcelas subdivididas no tempo, foram incluídos o efeito de linhas e colunas para maior controle ambiental no experimento. Com o mesmo objetivo, foram incluídos como covariáveis, os dados climáticos medidos ao longo dos períodos de medições para substituir quantitativamente a variável discreta de períodos. Posteriormente os modelos foram analisados conjuntamente para os dois caracteres avaliados (diâmetro e altura das árvores).

A respeito das variáveis climáticas, os valores utilizados foram as médias de temperatura para cada período, sendo 24,82, 23,05, 25,90 e 23,44°C, amplitude de temperatura, sendo

2, 78, 9, 36, 6, 13 e 8, 37 e precipitação total, sendo 1457, 5, 92, 8, 1072, 6 e 311, 9 *mm* para os períodos 1, 2, 3 e 4 respectivamente.

4.2.1 Modelo para Caracteres Individuais

Inicialmente foram testados dois modelos, um para cada caráter, assumindo independência entre estas duas características.

Para ambos os caracteres, as covariáveis meteorológicas incluídas foram capazes de substituir perfeitamente a variável período, explicando toda variação contida nesta variável discreta. Além disso todas elas foram estatisticamente significativas (p -valor = 0, 001) pelo teste *Wald* de parâmetros fixos.

Para cada um dos dois modelos, dentro da matriz **G** estão as variâncias genótípicas dentro dos períodos, e as covariâncias genótípicas entre eles. Já dentro da matriz **R** estão três estruturas, uma para as VCOV residuais dentro e entre períodos, outra para as VCOV residuais dentro e entre linhas e a última para VCOV residuais dentro e entre colunas.

Em ambos os modelos, o de menor *BIC* foi o que empregava a combinação entre a estrutura FA para matriz **G**, e as estruturas UNS, AR e AR para as estruturas de VCOV residuais de dentro e entre períodos, linhas e colunas respectivamente, com valores de 16171, 53 e 34915, 47, nesta ordem, para os caracteres diâmetro e altura.

Em relação ao *AIC*, duas diferentes combinações de estruturas de VCOV apresentaram os menores valores, isto é, não houve a mesma equivalência como quando utilizado o critério *BIC*; dessa maneira, para o caráter diâmetro, as melhores estruturas condicionadas a este critério foram: FA para matriz **G** e UNS, AR_{HET} e AR_{HET} para as três estruturas da matriz **R** (períodos, linhas e colunas respectivamente). Para o caráter altura, as estruturas selecionadas foram: UNS para matriz **G** e UNS, AR_{HET} E AR_{HET} UNS, AR_{HET} e AR_{HET} para as três estruturas da matriz **R** (períodos, linhas e colunas respectivamente). Os dois modelos apresentaram valores de *AIC* iguais a 16002,02 e 34683,05, para as variáveis diâmetro e altura respectivamente.

Utilizando o *AIC*, para ambos os caracteres, as estruturas para linhas e colunas que se destacaram foram AR_{HET} e AR_{HET}, porém, a magnitude dos valores deste critério quando comparados com a estruturas AR e AR foi pequena, em contrapartida, utilizando o critério *BIC*, as estruturas AR e AR obtiveram valores consideravelmente menores quando comparados com os valores deste critério para as estruturas AR_{HET} e AR_{HET}.

As estruturas de variância e covariância (VCOV) testadas para as matrizes aleatórias **G** e **R** e os respectivos valores de *AIC* e *BIC* estão apresentadas nas Tabelas 2 e 3 que contém os

resultados para os caracteres diâmetro e altura, respectivamente.

Tabela 2 – Diferentes modelos para a matriz de variâncias-covariâncias genéticas **G** e residual **R** e valores correspondentes de *AIC* (Critério de Informação de Akaike) e *BIC* (Critério de Informação Bayesiano) para os modelos com múltiplos períodos para o caráter diâmetro

(Continua)					
	Matriz G	Matriz R	n_{PAR}	<i>AIC</i>	<i>BIC</i>
1)	DIAG	DIAG \otimes AR \otimes AR	16	16490,75	16554,04
2)	FA	DIAG \otimes AR \otimes AR	20	16276,41	16365,01
3)	ID	DIAG \otimes AR \otimes AR	13	16664,60	16708,89
4)	UNS	DIAG \otimes AR \otimes AR	22	16261,65	16362,91
5)	DIAG	FA \otimes AR \otimes AR	20	16153,63	16242,23
6)	FA	FA \otimes AR \otimes AR	24	16065,36	16179,27
7)	ID	FA \otimes AR \otimes AR	17	16228,70	16298,32
8)	UNS	FA \otimes AR \otimes AR	26	16052,80	16179,37
9)	DIAG	ID \otimes AR \otimes AR	13	17669,61	17713,91
10)	FA	ID \otimes AR \otimes AR	17	17474,17	17543,79
11)	ID	ID \otimes AR \otimes AR	10	17964,05	17989,36
12)	UNS	ID \otimes AR \otimes AR	19	17364,02	17446,29
13)	DIAG	UNS \otimes AR \otimes AR	22	16138,59	16239,85
14)	FA	UNS \otimes AR \otimes AR	26	16044,96	16171,53
15)	ID	UNS \otimes AR \otimes AR	19	16213,13	16295,40
16)	UNS	UNS \otimes AR \otimes AR	28	16040,03	16179,26
17)	DIAG	DIAG \otimes AR _{HET} \otimes AR _{HET}	82	16446,93	16927,89
18)	FA	DIAG \otimes AR _{HET} \otimes AR _{HET}	86	16236,53	16742,81
19)	ID	DIAG \otimes AR _{HET} \otimes AR _{HET}	79	16616,12	17078,10
20)	UNS	DIAG \otimes AR _{HET} \otimes AR _{HET}	88	16217,12	16736,05
21)	DIAG	FA \otimes AR _{HET} \otimes AR _{HET}	86	16103,35	16609,63
22)	FA	FA \otimes AR _{HET} \otimes AR _{HET}	90	16020,53	16552,12
23)	ID	FA \otimes AR _{HET} \otimes AR _{HET}	83	16183,17	16670,46
24)	UNS	FA \otimes AR _{HET} \otimes AR _{HET}	92	16002,50	16546,75
25)	DIAG	ID \otimes AR _{HET} \otimes AR _{HET}	79	17587,90	18049,88
26)	FA	ID \otimes AR _{HET} \otimes AR _{HET}	83	17392,78	17880,07
27)	ID	ID \otimes AR _{HET} \otimes AR _{HET}	76	17896,60	18339,59
28)	UNS	ID \otimes AR _{HET} \otimes AR _{HET}	85	17565,28	18065,23
29)	DIAG	UNS \otimes AR _{HET} \otimes AR _{HET}	88	16089,98	16608,91
30)	FA	UNS \otimes AR _{HET} \otimes AR _{HET}	92	16002,02	16546,27
31)	ID	UNS \otimes AR _{HET} \otimes AR _{HET}	85	16169,50	16669,45
32)	UNS	UNS \otimes AR _{HET} \otimes AR _{HET}	94	15992,45	16549,36
33)	DIAG	DIAG \otimes DIAG \otimes DIAG	80	16625,25	17093,56
34)	FA	DIAG \otimes DIAG \otimes DIAG	84	16413,71	16907,33
35)	ID	DIAG \otimes DIAG \otimes DIAG	77	16785,57	17234,89
36)	UNS	DIAG \otimes DIAG \otimes DIAG	86	16387,29	16893,57
37)	DIAG	FA \otimes DIAG \otimes DIAG	84	16257,56	16751,18
38)	FA	FA \otimes DIAG \otimes DIAG	88	16181,81	16700,74

Tabela 2 – Diferentes modelos para a matriz de variâncias-covariâncias genéticas \mathbf{G} e residual \mathbf{R} e valores correspondentes de AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano) para os modelos com múltiplos períodos para o caráter diâmetro

		(Conclusão)			
	Matriz G	Matriz R	n_{PAR}	AIC	BIC
39)	ID	FA \otimes DIAG \otimes DIAG	81	16331,36	16806,00
40)	UNS	FA \otimes DIAG \otimes DIAG	90	16161,10	16692,69
41)	DIAG	ID \otimes DIAG \otimes DIAG	77	17904,37	18353,69
42)	FA	ID \otimes DIAG \otimes DIAG	81	17700,47	18175,10
43)	ID	ID \otimes DIAG \otimes DIAG	74	18209,25	18639,58
44)	UNS	ID \otimes DIAG \otimes DIAG	83	36504,08	36991,44
45)	DIAG	UNS \otimes DIAG \otimes DIAG	86	16241,89	16748,17
46)	FA	UNS \otimes DIAG \otimes DIAG	90	16162,12	16693,71
47)	ID	UNS \otimes DIAG \otimes DIAG	83	16315,02	16802,31
48)	UNS	UNS \otimes DIAG \otimes DIAG	92	16150,57	16694,82
49)	DIAG	DIAG \otimes ID \otimes ID	14	16661,90	16712,53
50)	FA	DIAG \otimes ID \otimes ID	18	16448,84	16524,78
51)	ID	DIAG \otimes ID \otimes ID	11	16832,75	16864,39
52)	UNS	DIAG \otimes ID \otimes ID	20	16426,97	16515,57
53)	DIAG	FA \otimes ID \otimes ID	18	16311,75	16387,69
54)	FA	FA \otimes ID \otimes ID	22	16230,95	16332,20
55)	ID	FA \otimes ID \otimes ID	15	16385,30	16442,26
56)	UNS	FA \otimes ID \otimes ID	24	16216,73	16330,64
57)	DIAG	ID \otimes ID \otimes ID	11	17990,96	18022,60
58)	FA	ID \otimes ID \otimes ID	15	17787,99	17844,94
59)	ID	ID \otimes ID \otimes ID	8	18280,01	18292,67
60)	UNS	ID \otimes ID \otimes ID	17	19017,44	19087,06
61)	DIAG	UNS \otimes ID \otimes ID	20	16294,65	16383,25
62)	FA	UNS \otimes ID \otimes ID	24	16209,52	16323,43
63)	ID	UNS \otimes ID \otimes ID	17	16367,30	16436,91
64)	UNS	UNS \otimes ID \otimes ID	26	16203,61	16330,18

\mathbf{G} : matriz de variâncias-covariâncias genéticas ($\mathbf{G} = \mathbf{G}_{P \times P}^{Per}$); \mathbf{R} : matriz de variâncias-covariâncias não-genéticas ou residual ($\mathbf{R} = \mathbf{R}_{P \times P}^{Per} \otimes \mathbf{R}_{L \times L}^{Lin} \otimes \mathbf{R}_{C \times C}^{Col}$); *Carac*, *Per*, *Lin* e *Col* são os índices das matrizes referentes à T caracteres, P períodos, L linhas e C colunas, respectivamente. ID: Identidade; DIAG: diagonal; FA: fator analítico; AR: auto-regressiva de primeira ordem; AR_{HET} : auto-regressiva de primeira ordem heterogênea; UNS: não-estruturada; n_{PAR} : número total de parâmetros no modelo. Os menores valores de AIC e BIC estão evidenciados em negrito.

Tabela 3 – Diferentes modelos para a matriz de variâncias-covariâncias genéticas **G** e residual **R** e valores correspondentes de *AIC* (Critério de Informação de Akaike) e *BIC* (Critério de Informação Bayesiano) para os modelos com múltiplos períodos para o caráter altura

(Continua)					
	Matriz G	Matriz R	n_{PAR}	<i>AIC</i>	<i>BIC</i>
1)	DIAG	DIAG \otimes AR \otimes AR	16	34864,19	34927,48
2)	FA	DIAG \otimes AR \otimes AR	20	34855,41	34944,01
3)	ID	DIAG \otimes AR \otimes AR	13	34873,89	34918,19
4)	UNS	DIAG \otimes AR \otimes AR	22	34840,42	34941,67
5)	DIAG	FA \otimes AR \otimes AR	20	34833,71	34922,31
6)	FA	FA \otimes AR \otimes AR	24	34820,50	34934,41
7)	ID	FA \otimes AR \otimes AR	17	51924,81	51994,42
8)	UNS	FA \otimes AR \otimes AR	26	34818,58	34945,15
9)	DIAG	ID \otimes AR \otimes AR	13	35237,13	35281,42
10)	FA	ID \otimes AR \otimes AR	17	35229,14	35298,75
11)	ID	ID \otimes AR \otimes AR	10	35319,88	35345,19
12)	UNS	ID \otimes AR \otimes AR	19	35213,11	35295,38
13)	DIAG	UNS \otimes AR \otimes AR	22	34823,35	34924,60
14)	FA	UNS \otimes AR \otimes AR	26	34809,41	34935,97
15)	ID	UNS \otimes AR \otimes AR	19	34833,20	34915,47
16)	UNS	UNS \otimes AR \otimes AR	28	34806,60	34945,83
17)	DIAG	DIAG \otimes AR _{HET} \otimes AR _{HET}	82	34733,41	35214,38
18)	FA	DIAG \otimes AR _{HET} \otimes AR _{HET}	86	34728,58	35234,86
19)	ID	DIAG \otimes AR _{HET} \otimes AR _{HET}	79	34750,32	35212,30
20)	UNS	DIAG \otimes AR _{HET} \otimes AR _{HET}	88	34711,14	35230,08
21)	DIAG	FA \otimes AR _{HET} \otimes AR _{HET}	86	34705,59	35211,86
22)	FA	FA \otimes AR _{HET} \otimes AR _{HET}	90	34695,66	35227,25
23)	ID	FA \otimes AR _{HET} \otimes AR _{HET}	83	34886,59	35373,88
24)	UNS	FA \otimes AR _{HET} \otimes AR _{HET}	92	34691,01	35235,26
25)	DIAG	ID \otimes AR _{HET} \otimes AR _{HET}	79	35114,25	35576,23
26)	FA	ID \otimes AR _{HET} \otimes AR _{HET}	83	35376,44	35863,73
27)	ID	ID \otimes AR _{HET} \otimes AR _{HET}	76	35213,46	35656,45
28)	UNS	ID \otimes AR _{HET} \otimes AR _{HET}	85	35091,77	35591,72
29)	DIAG	UNS \otimes AR _{HET} \otimes AR _{HET}	88	34698,25	35217,18
30)	FA	UNS \otimes AR _{HET} \otimes AR _{HET}	92	34688,18	35232,43
31)	ID	UNS \otimes AR _{HET} \otimes AR _{HET}	85	34714,99	35214,94
32)	UNS	UNS \otimes AR _{HET} \otimes AR _{HET}	94	34683,05	35239,95
33)	DIAG	DIAG \otimes DIAG \otimes DIAG	80	34882,49	35350,80
34)	FA	DIAG \otimes DIAG \otimes DIAG	84	34874,61	35368,23
35)	ID	DIAG \otimes DIAG \otimes DIAG	77	34898,88	35348,20
36)	UNS	DIAG \otimes DIAG \otimes DIAG	86	34853,48	35359,76
37)	DIAG	FA \otimes DIAG \otimes DIAG	84	34848,20	35341,82
38)	FA	FA \otimes DIAG \otimes DIAG	88	35007,63	35083,57
39)	ID	FA \otimes DIAG \otimes DIAG	81	34863,43	35338,06
40)	UNS	FA \otimes DIAG \otimes DIAG	90	34826,11	35357,70
41)	DIAG	ID \otimes DIAG \otimes DIAG	77	35256,38	35705,70

Tabela 3 – Diferentes modelos para a matriz de variâncias-covariâncias genéticas \mathbf{G} e residual \mathbf{R} e valores correspondentes de AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano) para os modelos com múltiplos períodos para o caráter diâmetro

					(Conclusão)
	Matriz G	Matriz R	n_{PAR}	AIC	BIC
42)	FA	ID \otimes DIAG \otimes DIAG	81	67842,53	68317,23
43)	ID	ID \otimes DIAG \otimes DIAG	74	35363,07	35793,41
44)	UNS	ID \otimes DIAG \otimes DIAG	83	35221,47	35708,76
45)	DIAG	UNS \otimes DIAG \otimes DIAG	86	34845,33	35351,61
46)	FA	UNS \otimes DIAG \otimes DIAG	90	34898,72	35430,31
47)	ID	UNS \otimes DIAG \otimes DIAG	83	34860,98	35348,27
48)	UNS	UNS \otimes DIAG \otimes DIAG	92	34820,22	35364,47
49)	DIAG	DIAG \otimes ID \otimes ID	14	35019,05	35069,67
50)	FA	DIAG \otimes ID \otimes ID	18	35007,63	35083,57
51)	ID	DIAG \otimes ID \otimes ID	11	35026,53	35058,17
52)	UNS	DIAG \otimes ID \otimes ID	20	34988,41	35077,00
53)	DIAG	FA \otimes ID \otimes ID	18	68431,89	68508,08
54)	FA	FA \otimes ID \otimes ID	22	77775,89	77877,47
55)	ID	FA \otimes ID \otimes ID	15	34991,08	35048,03
56)	UNS	FA \otimes ID \otimes ID	24	34957,83	35071,75
57)	DIAG	ID \otimes ID \otimes ID	11	35387,45	35419,09
58)	FA	ID \otimes ID \otimes ID	15	51924,81	51994,42
59)	ID	ID \otimes ID \otimes ID	8	35472,05	35484,71
60)	UNS	ID \otimes ID \otimes ID	17	35365,11	35415,74
61)	DIAG	UNS \otimes ID \otimes ID	20	34975,28	35063,88
62)	FA	UNS \otimes ID \otimes ID	24	34967,01	35080,92
63)	ID	UNS \otimes ID \otimes ID	17	34982,94	35052,55
64)	UNS	UNS \otimes ID \otimes ID	26	34946,60	35073,17

\mathbf{G} : matriz de variâncias-covariâncias genéticas ($\mathbf{G} = \mathbf{G}_{P \times P}^{Per}$); \mathbf{R} : matriz de variâncias-covariâncias não-genéticas ou residual ($\mathbf{R} = \mathbf{R}_{P \times P}^{Per} \otimes \mathbf{R}_{L \times L}^{Lin} \otimes \mathbf{R}_{C \times C}^{Col}$); *Carac*, *Per*, *Lin* e *Col* são os índices das matrizes referentes à T caracteres, P períodos, L linhas e C colunas, respectivamente. ID: Identidade; DIAG: diagonal; FA: fator analítico; AR: auto-regressiva de primeira ordem; AR_{HET}: auto-regressiva de primeira ordem heterogênea; UNS: não-estruturada; n_{PAR} : número total de parâmetros no modelo. Os menores valores de AIC e BIC estão evidenciados em negrito.

4.2.2 Modelo Multi-Caracteres-Multi-Períodos

O modelo multi-caracteres representa uma fusão dos dois modelos individuais, contudo, novas estruturas de VCOV foram testadas para cada uma das matrizes \mathbf{G} e \mathbf{R} , com exceção daquelas estruturas que foram criticamente superiores para ambos os modelos individuais.

A estrutura AR foi fixada tanto para linhas como para colunas neste modelo, porém, foram novamente testadas as estruturas ID, DIAG e UNS para as VCOV genóticas dentro e entre os caracteres.

Considerando a matriz de VCOV genotípica dentro e entre os períodos as mesmas estruturas testadas nos modelos individuais (ID, DIAG, FA e UNS) foram testadas no modelo multi-caracteres. Para a matriz de VCOV residual somente as estruturas FA e UNS foram testadas dentro e entre períodos, por terem sido superiores quando comparadas com as outras estruturas para ambas as características.

O modelo com menor *BIC* entre todos foi aquele que combinou as estruturas UNS e UNS para a estruturas de caráter e período da matriz **G** e UNS, FA, AR e AR para as estruturas caráter, período, linhas e colunas da matriz **R**, com valor de 51318,86. Este foi o modelo selecionado (Tabela 4).

Com relação ao *AIC*, este modelo foi o que obteve o segundo menor valor entre todos (51146,41). No que diz respeito à quantidade de parâmetros, ele foi o que apresentou o segundo maior número, sendo inferior apenas àquele que apresentou matrizes UNS para todas as estruturas VCOV, exceto para linhas e colunas.

As correlações fenotípicas dentro e entre as variáveis mensuradas foram mantidas através da utilização deste modelo. Os maiores valores de correlação foram observados entre o crescimento dos indivíduos dentro de cada estação, sendo que para a variável diâmetro o maior valor foi verificado entre os períodos de estação favorável *D1* e *D3* (0,85), e para a variável altura, o maior valor foi verificado entre os períodos da estação adversa *H2* e *H4* (0,81) (Figura 8).

Para ambas as variáveis de crescimento, os componentes periódicos que mais apresentaram correlação com o crescimento total foram os de estação favorável, sendo que, entre estes os maiores valores de correlação foram observados nos períodos de maior idade (0,94 para a variável diâmetro e 0,80 para a variável altura) (Figura 8).

Entre as variáveis de crescimento, as correlações observadas entre elas, em cada período foram de 0,37, 0,19, 0,46 e 0,27 para os períodos 1, 2, 3 e 4 respectivamente, e 0,45 para ambas no período total (Figura 8).

4.3 Análise Fenotípica

4.3.1 Modelo Estrutural Fenotípico

Seguindo as etapas fundamentais na construção de modelos estruturais, algumas etapas foram necessárias para que o modelo fenotípico mais parcimonioso fosse proposto. Inicialmente um modelo considerando as relações biológicas admissíveis foi testado (Figura 9a), progressivamente este modelo foi sendo alterado visando deixá-lo mais condizente com os dados obser-

vados.

Tabela 4 – Diferentes modelos para a matriz de variâncias-covariâncias genéticas (**G**) e residual (**R**) e valores correspondentes de *AIC* (Critério de Informação de Akaike) e *BIC* (Critério de Informação Bayesiano) para o modelos com múltiplos caracteres em múltiplos períodos

	Matriz G	Matriz R	n_{PAR}	<i>AIC</i>	<i>BIC</i>
a)	ID ⊗ ID	ID ⊗ FA ⊗ AR ⊗ AR	18	64043,64	64120,88
b)	ID ⊗ ID	ID ⊗ UNS ⊗ AR ⊗ AR	20	64020,63	64111,91
c)	ID ⊗ DIAG	ID ⊗ FA ⊗ AR ⊗ AR	21	64018,63	64116,93
d)	ID ⊗ DIAG	ID ⊗ UNS ⊗ AR ⊗ AR	23	63994,80	64107,15
e)	ID ⊗ FA	ID ⊗ FA ⊗ AR ⊗ AR	25	63983,49	64109,88
f)	ID ⊗ FA	ID ⊗ UNS ⊗ AR ⊗ AR	27	63958,73	64099,17
g)	ID ⊗ UNS	ID ⊗ FA ⊗ AR ⊗ AR	27	63977,21	64117,64
h)	ID ⊗ UNS	ID ⊗ UNS ⊗ AR ⊗ AR	29	63950,72	64105,19
i)	DIAG ⊗ ID	DIAG ⊗ FA ⊗ AR ⊗ AR	22	51533,49	51624,78
j)	DIAG ⊗ ID	DIAG ⊗ UNS ⊗ AR ⊗ AR	24	51512,05	51617,38
k)	DIAG ⊗ DIAG	DIAG ⊗ FA ⊗ AR ⊗ AR	25	51465,82	51578,16
l)	DIAG ⊗ DIAG	DIAG ⊗ UNS ⊗ AR ⊗ AR	27	51443,28	51569,67
m)	DIAG ⊗ FA	DIAG ⊗ FA ⊗ AR ⊗ AR	29	51350,55	51490,98
n)	DIAG ⊗ FA	DIAG ⊗ UNS ⊗ AR ⊗ AR	31	51340,86	51495,33
o)	DIAG ⊗ UNS	DIAG ⊗ FA ⊗ AR ⊗ AR	31	51334,53	51489,01
p)	DIAG ⊗ UNS	DIAG ⊗ UNS ⊗ AR ⊗ AR	33	51323,87	51492,39
q)	UNS ⊗ ID	UNS ⊗ FA ⊗ AR ⊗ AR	24	51349,29	51454,61
r)	UNS ⊗ ID	UNS ⊗ UNS ⊗ AR ⊗ AR	26	51332,66	51452,03
s)	UNS ⊗ DIAG	UNS ⊗ FA ⊗ AR ⊗ AR	27	51281,72	51408,11
t)	UNS ⊗ DIAG	UNS ⊗ UNS ⊗ AR ⊗ AR	29	51264,58	51405,01
u)	UNS ⊗ FA	UNS ⊗ FA ⊗ AR ⊗ AR	31	51169,58	51324,06
v)	UNS ⊗ FA	UNS ⊗ UNS ⊗ AR ⊗ AR	33	51164,82	51333,34
x)	UNS ⊗ UNS	UNS ⊗ FA ⊗ AR ⊗ AR	33	51150,34	51318,86
z)	UNS ⊗ UNS	UNS ⊗ UNS ⊗ AR ⊗ AR	35	51146,41	51328,97

G: matriz de variâncias-covariâncias genéticas ($\mathbf{G} = \mathbf{G}_{T \times T}^{Carac} \otimes \mathbf{G}_{P \times P}^{Per}$); **R**: matriz de variâncias-covariâncias não-genéticas ou residual ($\mathbf{R} = \mathbf{R}_{T \times T}^{Carac} \otimes \mathbf{R}_{P \times P}^{Per} \otimes \mathbf{R}_{L \times L}^{Lin} \otimes \mathbf{R}_{C \times C}^{Col}$); *Carac*, *Per*, *Lin* e *Col* são os índices das matrizes referentes à *T* caracteres, *P* períodos, *L* linhas e *C* colunas, respectivamente. ID: Identidade; DIAG: diagonal; FA: fator analítico; AR: autoregressiva de primeira ordem; UNS: não-estruturada ; n_{PAR} : número total de parâmetros no modelo. Os menores valores de *AIC* e *BIC* estão evidenciados em negrito.

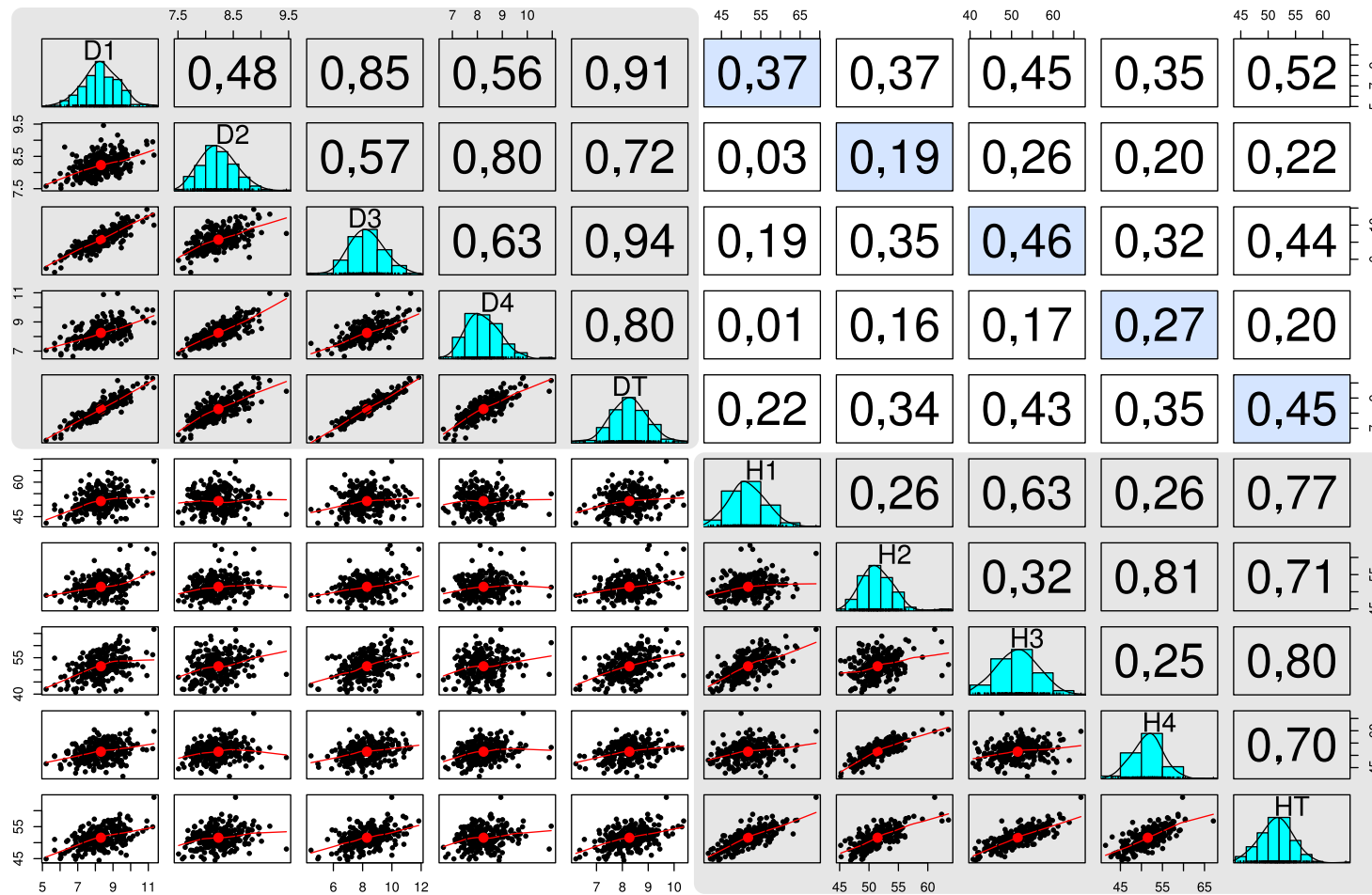


Figura 8 – Correlação existente entre os componentes periódicos de crescimento e crescimentos totais. D e H são o diâmetro e a altura nos períodos 1, 2, 3, 4 e total, respectivamente. Na diagonal estão os gráficos de distribuição normal para os valores ajustados de cada variável; na parte inferior os gráficos de dispersão das variáveis tomadas duas a duas; na parte superior estão os valores das correlações. No canto superior direito estão os gráficos e correlações entre a variável diâmetro; no canto inferior esquerdo estão os gráficos e correlações entre a variável altura. Estão enfatizadas em azul claro as correlações entre as variáveis nos períodos 1, 2, 3, 4 e total

Uma medida adicional foi a inclusão de variáveis latentes climáticas que conectasse os períodos mais semelhantes. Para isso duas alternativas foram adotadas: a primeira foi a inclusão de somente duas variáveis latentes que conectasse os períodos climaticamente semelhantes (clima favorável e adverso) para ambos os períodos; a segunda foi a inclusão de variáveis com a suposição de que tais condições ambientais pudessem afetar de maneira distinta cada um dos caracteres, ou seja, dois pares de variáveis latentes climáticas foram colocadas representando períodos com condições meteorológicas favoráveis e adversas para cada um dos dois caracteres respectivamente (η_1, η_2, η_3 e η_4) da Figura 9b.

Ao longo do processo de seleção, os modelos foram avaliados com base nos seus valores de *AIC* e *BIC*. Por fim, somente as relações estatisticamente significativas pelo teste *z*, foram deixadas no modelo. A exclusão dos parâmetros foram feitas através do método *stepwise backward*, onde os parâmetros foram retirados na ordem do mais não-significativo para o menos. Tais parâmetros foram combinados de tal maneira que prevalecesse o modelo com menores valores dos critérios de seleção utilizados.

Valores de *AIC* e *BIC* para os diversos modelos testados foram obtidos nesse processo, mas os quatro principais (1º: modelo com todas relações estruturais plausíveis (Figura 9a); 2º: modelo com 2 variáveis latentes climáticas; 3º: modelo com 4 variáveis latentes climáticas (Figura 9b); e 4º: modelo com 4 variáveis latentes climáticas e somente relações estruturais significativas (Figura 9c)) obtiveram *AIC* de 912,50, 514,22, 106,5019 e 101,66 e *BIC* de 787,56, 396,33, 9,80 e -12,69, respectivamente. Com base nesses resultados, o modelo escolhido foi o 4º (Figura 9c). Tal modelo apresentou valor de probabilidade (*p-valor*) igual a $3,14 \times 10^{-7}$, probabilidade esta que segue distribuição aproximada de chi-quadrado (χ^2).

Os valores dos efeitos e as respectivas significâncias estatísticas dos parâmetros do modelo são apresentados na Tabela 5. Por este modelo, as relações β 's entre variáveis endógenas observadas (*y*) foram negativas nos seus efeitos de causalidade entre as variáveis diâmetro no período 1 causando o diâmetro no período 2 (β_{21}), altura no período 1 causando altura no período 2 (β_{65}) e altura no período 3 causando diâmetro no período 4 (β_{47}), e positivas nos demais ($\beta_{34}, \beta_{34}, \beta_{15}, \beta_{26}, \beta_{21}, \beta_{37}$ e β_{48}).

As covariâncias entre as variáveis latentes foram todas significativas, exceto a correlação entre as variáveis latentes de clima favorável sobre o diâmetro ($FAV(D)$) e a de clima favorável sobre a altura ($FAV(H)$).

Os efeitos dos erros de perturbação (ζ) das variáveis latentes climáticas sobre a variável altura foram mais expressivos, sendo que o efeito de ζ dos climas favoráveis sobre esta variável apresentou o maior valor entre todos. O menor efeito de ζ foi o de clima adverso sobre a variável

diâmetro, sendo este aproximadamente 66 vezes menor que o maior ζ .

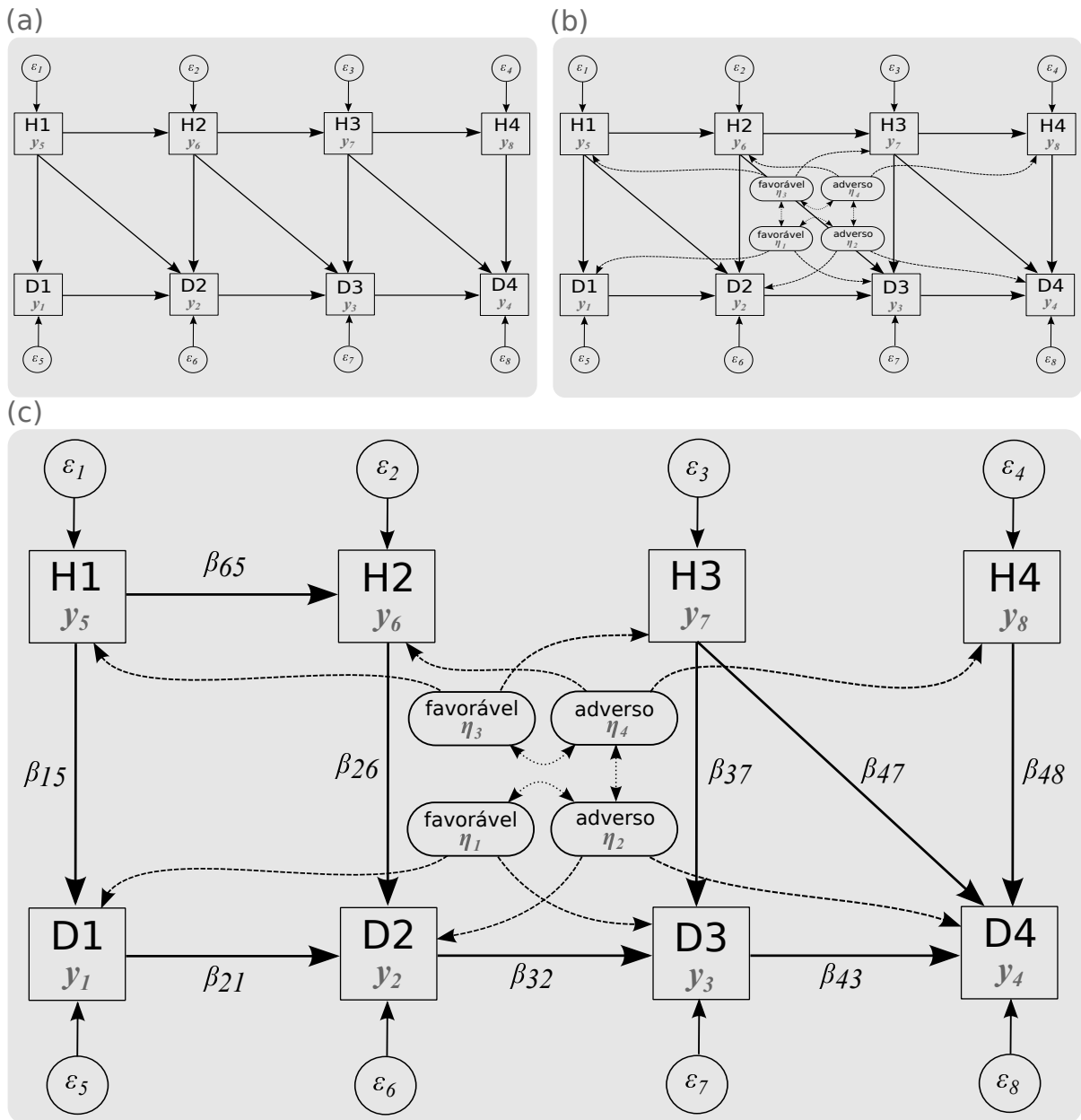


Figura 9 – Modelos estruturais para os dados fenotípicos. (a) modelo inicial com todas as possíveis relações causais plausíveis. (b) modelo inicial com acréscimo de variáveis latentes (η) conectando períodos com condições climáticas semelhantes com seus respectivos efeitos sob cada variável. (c) modelo final considerando somente as relações estruturais significativas. D e H são o diâmetro e a altura, respectivamente, nos períodos 1, 2, 3 e 4; y é o indicador das variáveis instrumentais observadas endógenas; β é efeito das relações estruturais entre as variáveis y ; ε é efeito residual sob cada variável y . Entre as setas, as setas de aresta lisa são referentes aos efeitos diretos β 's e aos residuais ε , as setas tracejadas são referentes aos efeitos λ das variáveis latentes sobre as variáveis observadas e as pontilhadas são as covariâncias ψ entre as variáveis latentes

Tabela 5 – Efeitos dos parâmetros das relações estruturais do modelo estrutural fenotípico com os respectivos P-valores de significância testados pelo *teste-Z* e as respectivas direções destas relações

Parâmetro	Efeito	Erro Padrão	estatística z	P-valor	Relação Estrutural
β_{21}	-0,1216	0,0320	-3,7994	$1,45 \times 10^{-4}$	D2 \leftarrow D1
β_{32}	0,3027	0,1187	2,5498	$1,08 \times 10^{-2}$	D3 \leftarrow D2
β_{43}	0,1510	0,0313	4,8269	$1,39 \times 10^{-6}$	D4 \leftarrow D3
β_{65}	-0,0653	0,0272	-2,3993	$1,64 \times 10^{-2}$	H2 \leftarrow H1
β_{15}	0,0766	0,0077	9,9281	$3,14 \times 10^{-23}$	D1 \leftarrow H1
β_{26}	0,0734	0,0126	5,8445	$5,08 \times 10^{-9}$	D2 \leftarrow H2
β_{37}	0,0615	0,0085	7,2586	$3,91 \times 10^{-13}$	D3 \leftarrow H3
β_{47}	-0,0193	0,0051	-3,7676	$1,65 \times 10^{-4}$	D4 \leftarrow H3
β_{48}	0,0575	0,0098	5,8943	$3,76 \times 10^{-9}$	D4 \leftarrow H4
ε_1	0,0628	0,0252	2,4896	$1,28 \times 10^{-2}$	D1 \leftrightarrow D1
ε_2	-0,0256	0,0063	-4,0670	$4,76 \times 10^{-5}$	D2 \leftrightarrow D2
ε_3	0,2451	0,0326	7,5080	$6,01 \times 10^{-14}$	D3 \leftrightarrow D3
ε_4	0,1533	0,0149	10,2783	$8,83 \times 10^{-25}$	D4 \leftrightarrow D4
ε_5	7,4481	1,1705	6,3632	$1,98 \times 10^{-10}$	H1 \leftrightarrow H1
ε_6	-0,4487	0,2861	-1,5687	$1,17 \times 10^{-1}$	H2 \leftrightarrow H2
ε_7	8,9840	1,2535	7,1673	$7,65 \times 10^{-13}$	H3 \leftrightarrow H3
ε_8	4,9200	0,5226	9,4146	$4,75 \times 10^{-21}$	H4 \leftrightarrow H4
ζ_1	0,8337	0,0803	10,3774	$3,14 \times 10^{-25}$	FAV(D) \leftrightarrow FAV(D)
ζ_3	13,6274	1,6237	8,3928	$4,75 \times 10^{-17}$	FAV(H) \leftrightarrow FAV(H)
ζ_2	0,2050	0,0294	6,9619	$3,36 \times 10^{-12}$	ADV(D) \leftrightarrow ADV(D)
ζ_4	8,8516	0,8552	10,3499	$4,19 \times 10^{-25}$	ADV(H) \leftrightarrow ADV(H)
ψ_{24}	-0,5104	0,1136	-4,4916	$7,07 \times 10^{-6}$	ADV(H) \leftrightarrow ADV(D)
ψ_{12}	0,2494	0,0406	6,1463	$7,93 \times 10^{-10}$	ADV(D) \leftrightarrow FAV(D)
ψ_{34}	4,9790	0,9340	5,3306	$9,79 \times 10^{-8}$	ADV(H) \leftrightarrow FAV(H)

Os β 's são efeitos das relações estruturais entre as variáveis endógenas observadas (y); ε é efeito residual sob cada variável y ; ζ são os erros de perturbação sobre as variáveis latentes; e ψ são as covariâncias entre elas. D e H são o diâmetro e a altura, respectivamente, nos períodos 1, 2, 3 e 4; $FAV(D)$, $FAV(H)$, $ADV(D)$, $ADV(H)$ são as variáveis latentes climáticas específicas de cada caráter de crescimento.

4.4 Análise Genotípica

4.4.1 Mapeamento dos QTLs

Dos 263 indivíduos da população de mapeamento, foram detectados 11 contaminantes com base nos padrões de segregação alélica dos genótipos, e por isso 252 foram utilizados no mapeamento de QTLs.

De acordo com a abordagem de determinação de threshold por permutação proposta por Churchill e Doerge (1994), os valores em *LOD-score* ao nível de 95% de probabilidade, refe-

rentes aos caracteres diâmetro e altura nas medições 1, 2, 3 e 4 ($D1$, $D2$, $D3$, $D4$, $H1$, $H2$, $H3$, $H4$) foram: 5, 13, 4, 66, 4, 90, 5, 12, 4, 79, 4, 99, 4, 83 e 5, 05, respectivamente. E os valores dos *threshold's* para o diâmetro total e altura total dos quatro períodos de incremento (DT e HT) foram 5, 19, 4, 49, respectivamente.

Com posse desses valores de *threshold*, foram mapeados um total de 13 QTLs considerando as duas características ao longo dos quatro períodos e no período total. Em que, 6 foram do caráter diâmetro e 7 do caráter altura (Figura 10).

Os QTLs do caráter diâmetro foram encontrados nos grupos de ligação 3, 5, 6b, 10, 16; enquanto que os do caráter altura, nos grupos 1, 5, 8, 10, 13. Nos fenótipos de incremento em diâmetro no período 3 e o incremento total em altura não foram detectados QTLs significativos para o *threshold* utilizado.

No caráter diâmetro, dos 6 QTLs encontrados, 2 QTLs foram coincidentes nos períodos 2 ($D2$) e 4 ($D4$), e 2 QTLs foram mapeados para o incremento total considerando os quatro períodos avaliados.

Para o caráter altura, todos os 7 QTLs mapeados estão relacionados com os incrementos periódicos. Nenhum foi QTL encontrado para o incremento total dos quatro períodos.

Quando consideradas as duas características, nenhum QTL foi encontrado exercendo efeitos pleiotrópicos entre elas, embora haja indicativo de que os QTLs 5 e 12 sejam um mesmo QTL, isso por estarem muito próximos no grupo de ligação 10, estando ambos contidos em um intervalo menor que 6 cm.

Considerando os 13 QTLs totais, o grupo de ligação que foi encontrado mais QTLs foi o grupo 10, com 5 QTLs ($QTL4$, $QTL5$, $QTL6$, $QTL8$ e $QTL12$), os demais grupos apresentaram somente um QTL. Os QTLs apresentaram diferentes padrões de segregação para as diferentes caracteres de crescimento periódico. Esses padrões variaram entre 1:1, 3:1, 1:1:1:1 e 1:2:1. Os QTLs $QTL4$, $QTL7$, $QTL11$, $QTL12$ e $QTL13$ foram coincidentes com marcadores, enquanto que os demais foram encontrados em intervalos entre marcadores (Tabela 6)

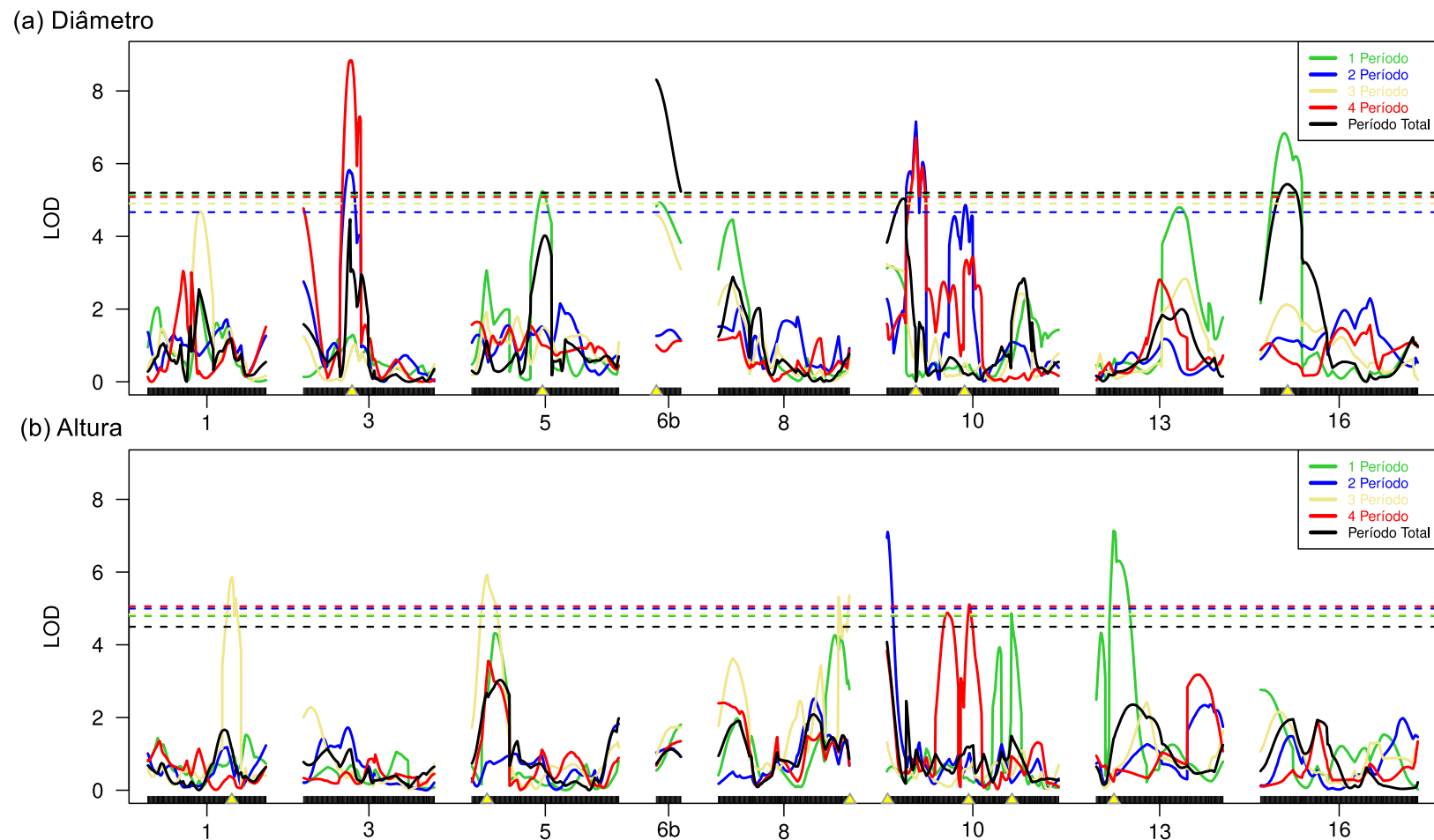


Figura 10 – Dentre os 23 grupos de ligação mapeados, foram encontrados um total 13 QTLs (localizados acima de cada triângulo amarelo) para o diâmetro e altura ao longo dos períodos e no período total de crescimento, distribuídos em 8 grupos. As cores: verde, azul, bege, vermelho e preto representam os incrementos nos períodos 1, 2, 3 e 4 e o incremento total nos quatro períodos para os dois caracteres. As linhas tracejadas nos gráficos correspondem aos valores *LOD-score* do *threshold* para cada caráter em cada período calculados com base na abordagem de Churchill e Doerge (1994)

Tabela 6 – QTLs mapeados para os componentes periódicos de crescimento (diâmetro e altura) e períodos totais para ambas características

Fenótipo	QTLs	Marcador próximo	GL	Posição (cM)	Segregação	Fase
D1	QTL1	A2317-HB155	5	94,00	1:2:1	$P^2 P^1 \times Q^1 Q^2$
	QTL2	TAs2225-HBE64	16	32,00	3:1	$P^1 P^2 \times Q^1 Q^2$
D2	QTL3	TAs2264-HB24	3	61,00	1:2:1	$Q^1 Q^2$
	QTL4	T2617	10	38,72	1:2:1	$P^2 P^1 \times Q^1 Q^2$
	QTL5	T2612-HBE94	10	104,00	1:2:1	$P^1 P^2 \times Q^2 Q^1$
D4	QTL3	TAs2264-HB24	3	61,00	1:1:1:1	$P^1 P^2 \times Q^1 Q^2$
	QTL4	T2617	10	38,72	1:2:1	$P^2 P^1 \times Q^1 Q^2$
H1	QTL6	HBE204	10	165,72	1:2:1	$P^1 P^2$
	QTL7	T2597	13	23,04	3:1	$P^2 P^1 \times Q^2 Q^1$
H2	QTL8	A2481-HB151	10	1,00	3:1	$P^2 P^1 \times Q^2 Q^1$
H3	QTL9	TAs2557-HB197	1	112,00	3:1	$P^1 P^2 \times Q^2 Q^1$
	QTL10	T2636-HB1	5	20,00	1:2:1	$P^1 P^2 \times Q^1 Q^2$
	QTL11	A2525	8	174,23	3:1	$P^1 P^2 \times Q^2 Q^1$
H4	QTL12	HBE94	10	109,76	1:1	$P^1 P^2$
DT	QTL2	TAs2225-HBE64	16	32,00	1:2:1	$P^1 P^2 \times Q^1 Q^2$
	QTL13	T2449	6b	0,00	1:2:1	$P^1 P^2 \times Q^2 Q^1$

D1, D2, D4, DT, H1, H2, H3, H4 são os incrementos em diâmetro e altura nos períodos 1, 2, 4, total, 1, 2, 3 e 4 respectivamente. GL são os grupos de ligação em que os QTLs foram encontrados; cM é a unidade de distância no GL em *centi-Morgan*; P é o parental PB217, Q é o parental PR255; P¹ e Q¹ apresentam efeito positivo (aumentam o valor fenotípico) e P² e Q² apresentam efeito negativo (reduzem o valor fenotípico)

4.4.2 Modelo Estrutural Genético

Neste momento, os QTLs significativos obtidos pela metodologia CIM, foram incorporados ao modelo estrutural fenotípico, com o objetivo de se obter as estimativas dos efeitos causais genéticos sobre os fenótipos de crescimento estudados.

Do total de 12 QTLs mapeados para as 8 variáveis de crescimento periódico em diâmetro e altura, 8 obtiveram efeitos significativos, sendo eles: QTL3, QTL4, QTL5, QTL6, QTL9, QTL10, QTL11 e QTL12 (Figura 11).

Os efeitos aditivos dos parentais PB217 e PR255 (α_1 e α_2) e o efeito devido à dominância intra-loco (δ) presentes em cada QTL evidentemente apresentaram valores de significância di-

ferentes, sendo que, somente alguns destes efeitos específicos de cada QTLs permaneceram no modelo, os outros foram retirados por não demonstrarem efeito estatisticamente significativo.

Considerando o crescimento em diâmetro, somente permaneceram no modelo os efeitos dos QTLs sobre o crescimento em períodos de menores temperaturas e menores índices pluviométricos (Período 2 e 4), além disso, somente os efeitos do parental PR255 ($\gamma_{23}(\alpha_2)$, $\gamma_{43}(\alpha_2)$, $\gamma_{24}(\alpha_2)$, $\gamma_{44}(\alpha_2)$ e $\gamma_{25}(\alpha_2)$) revelaram efeitos causais sobre estes fenótipos. Os QTLs que apresentaram relação causal sobre o diâmetro nestes dois períodos foram o *QTL3*, *QTL4* e o *QTL5*.

Considerando o crescimento em altura, permaneceram no modelo predominantemente os efeitos aditivos do parental PB217 ($\gamma_{79}(\alpha_1)$, $\gamma_{711}(\alpha_1)$, $\gamma_{812}(\alpha_1)$ e somente o $\gamma_{710}(\alpha_2)$) e os efeitos devido à dominância ($\gamma_{56}(\delta)$, $\gamma_{79}(\delta)$, $\gamma_{812}(\delta)$). Os QTLs que apresentaram relação causal nesta característica foram o *QTL6*, *QTL9*, *QTL10*, *QTL11* e o *QTL12* (Tabela 7).

Quando verificamos a magnitude dos efeitos, os QTLs com relação causal sobre as variáveis periódicas de altura apresentam valores maiores que os QTLs com relação causal sobre as variáveis periódicas de diâmetro, porém estes QTLs também obtiveram maiores valores de erro padrão da média.

Todos os efeitos aditivos de ambos parentais forneceram valores positivos, com exceção do *QTL5* que apresentou efeito significativo negativo para o parental PB255 (*QTL5_{α2}*). Todos os efeitos devido à dominância δ , sem exceção, apresentaram valores negativos (Tabela 7).

Os valores dos parâmetros estruturais β entre os variáveis endógenas e os efeitos residuais ε destas variáveis, assim como, os erros de perturbação ζ das variáveis latentes e as estimativas de covariância (ψ) entre os erros de perturbação destas, permaneceram bastante semelhantes aos valores obtidos no modelo estrutural fenotípico (Tabelas 5 e 7).

Alguns efeitos indiretos dos QTLs sobre determinados fenótipos foram detectados (Tabela 8). Mesmo aqueles fenótipos que não possuem relação causal direta com QTLs podem sofrer a influência deles por meio de outras vias. Os caracteres de crescimento em diâmetro nos períodos 1 e 3 e altura no período 2 são casos em que não foi possível observar ação significativa de QTLs, mas apresentam efeitos indiretos de alguns QTLs.

Também foi possível observar que alguns QTLs exerceram ação indireta sobre os fenótipos com magnitude pequena, porém muito próxima à algumas relações causais diretas significativas, como é o caso dos efeitos dos QTLs *QTL9_{α1}*, *QTL9_δ*, *QTL10_{α2}*, *QTL11_{α1}* sobre o caráter diâmetro no período 3 e do *QTL6_δ* no caracteres diâmetro no período 1 e altura no período 2 (Tabela 8).

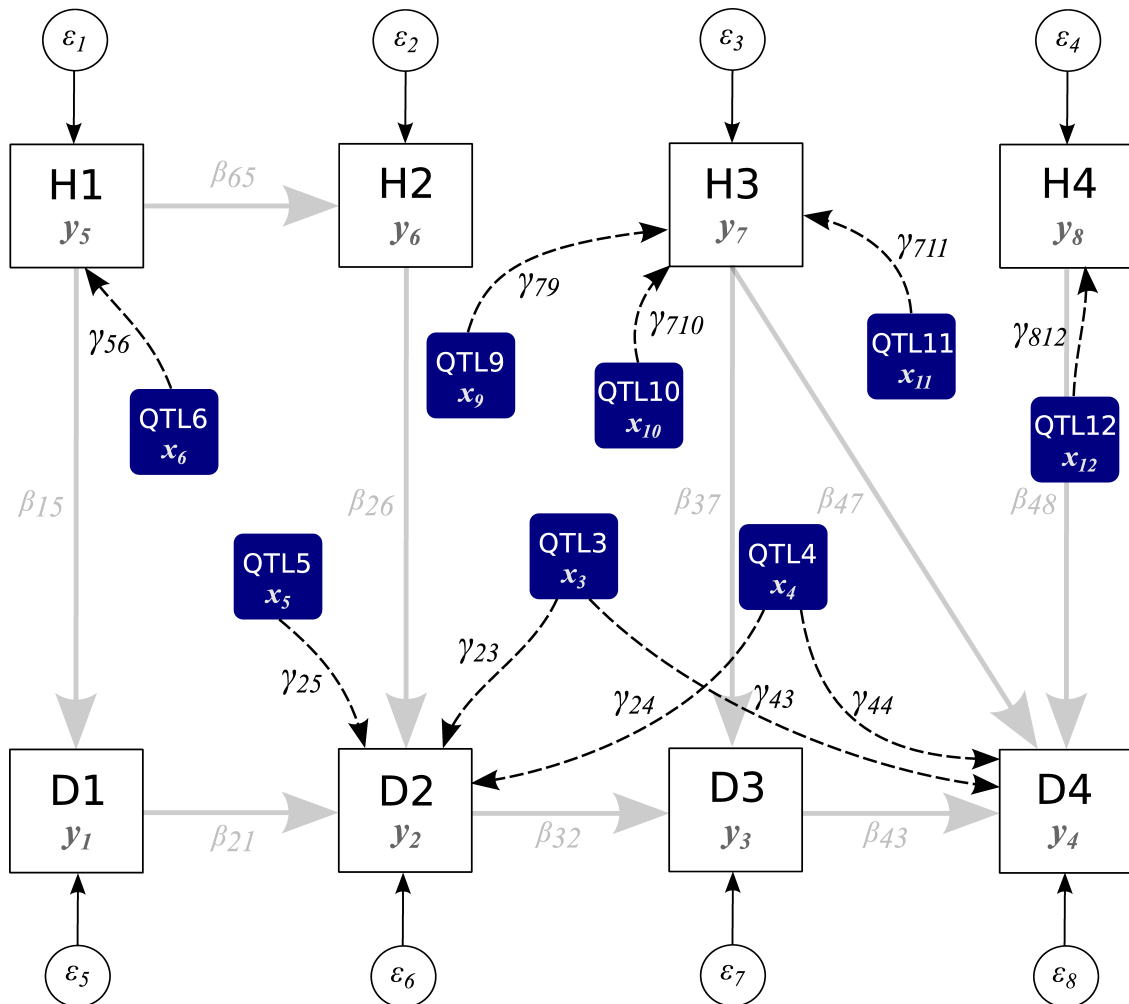


Figura 11 – Modelo estrutural genotípico com ação dos QTLs significativos (caixas azul marinho). $D1, D2, D3, D4, H1, H2, H3$ e $H4$ são as variáveis diâmetro e altura nos períodos 1, 2, 3 e 4 respectivamente (caixas sem preenchimento); As setas opacas de aresta lisa, representadas pelos β 's, são os efeitos das relações estruturais entre as variáveis endógenas observadas (y); ε 's são os efeitos residuais sob cada variável y . As setas de aresta tracejada, representadas pelos γ 's, são efeitos das variáveis exógenas observadas (x) sobre y ; ε são os resíduos sobre cada variável y . As variáveis latentes climáticas (η) e seus respectivos parâmetros estruturais (λ, ζ e ψ) do modelo estrutural fenotípico estão presentes neste modelo, porém, não foram representados na figura para maior clareza

Tabela 7 – Efeitos dos parâmetros das relações estruturais do modelo estrutural genotípico com os respectivos P-valores de significância testados pelo *teste-z* e as respectivas direções destas relações

Parâmetro	Efeito	Erro Padrão	estatística z	P-valor	Relação Estrutural
$\gamma_{23}(\alpha_2)$	0,0449	0,0168	2,6747	$7,48 \times 10^{-3}$	D2 \leftarrow QTL3 $_{\alpha_2}$
$\gamma_{43}(\alpha_2)$	0,1053	0,0321	3,2757	$1,05 \times 10^{-3}$	D4 \leftarrow QTL3 $_{\alpha_2}$
$\gamma_{24}(\alpha_2)$	0,0807	0,0190	4,2490	$2,15 \times 10^{-5}$	D2 \leftarrow QTL4 $_{\alpha_2}$
$\gamma_{44}(\alpha_2)$	0,0840	0,0309	2,7202	$6,52 \times 10^{-3}$	D4 \leftarrow QTL4 $_{\alpha_2}$
$\gamma_{25}(\alpha_2)$	-0,0451	0,0156	-2,8894	$3,86 \times 10^{-3}$	D2 \leftarrow QTL5 $_{\alpha_2}$
$\gamma_{56}(\delta)$	-0,5484	0,2209	-2,4822	$1,31 \times 10^{-2}$	H1 \leftarrow QTL6 $_{\delta}$
$\gamma_{79}(\alpha_1)$	0,5033	0,2322	2,1679	$3,02 \times 10^{-2}$	H3 \leftarrow QTL9 $_{\alpha_1}$
$\gamma_{79}(\delta)$	-0,5717	0,2360	-2,4228	$1,54 \times 10^{-2}$	H3 \leftarrow QTL9 $_{\delta}$
$\gamma_{710}(\alpha_2)$	0,5446	0,2247	2,4239	$1,54 \times 10^{-2}$	H3 \leftarrow QTL10 $_{\alpha_2}$
$\gamma_{711}(\alpha_1)$	0,5331	0,2202	2,4207	$1,55 \times 10^{-2}$	H3 \leftarrow QTL11 $_{\alpha_1}$
$\gamma_{812}(\alpha_1)$	0,3817	0,1317	2,8993	$3,74 \times 10^{-3}$	H4 \leftarrow QTL12 $_{\alpha_1}$
$\gamma_{812}(\delta)$	-0,3072	0,1331	-2,3088	$2,10 \times 10^{-2}$	H4 \leftarrow QTL12 $_{\delta}$
β_{21}	-0,1129	0,0312	-3,6158	$2,99 \times 10^{-4}$	D2 \leftarrow D1
β_{32}	0,3079	0,1230	2,5029	$1,23 \times 10^{-2}$	D3 \leftarrow D2
β_{43}	0,1441	0,0310	4,6413	$3,46 \times 10^{-6}$	D4 \leftarrow D3
β_{65}	-0,0692	0,0262	-2,6459	$8,15 \times 10^{-3}$	H2 \leftarrow H1
β_{15}	0,0765	0,0077	9,9091	$3,80 \times 10^{-23}$	D1 \leftarrow H1
β_{26}	0,0741	0,0121	6,1377	$8,37 \times 10^{-10}$	D2 \leftarrow H2
β_{37}	0,0614	0,0085	7,1855	$6,69 \times 10^{-13}$	D3 \leftarrow H3
β_{47}	-0,0170	0,0052	-3,2656	$1,09 \times 10^{-3}$	D4 \leftarrow H3
β_{48}	0,0602	0,0096	6,3021	$2,94 \times 10^{-10}$	D4 \leftarrow H4
ε_1	0,0709	0,0258	2,7484	$5,99 \times 10^{-3}$	D1 \leftrightarrow D1
ε_2	-0,0219	0,0059	-3,7272	$1,94 \times 10^{-4}$	D2 \leftrightarrow D2
ε_3	0,2368	0,0325	7,2922	$3,05 \times 10^{-13}$	D3 \leftrightarrow D3
ε_4	0,1486	0,0144	10,3209	$5,67 \times 10^{-25}$	D4 \leftrightarrow D4
ε_5	7,7291	1,1423	6,7662	$1,32 \times 10^{-11}$	H1 \leftrightarrow H1
ε_6	-0,3957	0,2724	-1,4528	$1,46 \times 10^{-1}$	H2 \leftrightarrow H2
ε_7	7,8158	1,1470	6,8142	$9,48 \times 10^{-12}$	H3 \leftrightarrow H3
ε_8	4,6591	0,4959	9,3952	$5,71 \times 10^{-21}$	H4 \leftrightarrow H4
ζ_1	0,8291	0,0802	10,3344	$4,93 \times 10^{-25}$	FAV(D) \leftrightarrow FAV(D)
ζ_2	13,3514	1,5773	8,4646	$2,57 \times 10^{-17}$	FAV(H) \leftrightarrow FAV(H)
ζ_3	0,1816	0,0267	6,8088	$9,84 \times 10^{-12}$	ADV(D) \leftrightarrow ADV(D)
ζ_4	8,8972	0,8532	10,4280	$1,85 \times 10^{-25}$	ADV(H) \leftrightarrow ADV(H)
ψ_{24}	-0,5053	0,1080	-4,6806	$2,86 \times 10^{-6}$	ADV(H) \leftrightarrow ADV(D)
ψ_{12}	0,2280	0,0386	5,9072	$3,48 \times 10^{-9}$	ADV(D) \leftrightarrow FAV(D)
ψ_{34}	5,2204	0,9143	5,7095	$1,13 \times 10^{-8}$	ADV(H) \leftrightarrow FAV(H)

Os parâmetros γ são referentes aos efeitos dos QTLs nas variáveis de crescimento periódico; α_1 , α_2 e δ são os efeitos aditivos do parental PB217, parental PR255 e de dominância no loco, respectivamente; β 's são efeitos das relações estruturais entre as variáveis endógenas observadas (y); ε é efeito residual sob cada variável y ; ζ são os erros de perturbação sobre as variáveis latentes; e ψ são as covariâncias entre elas. D e H são o diâmetro e a altura, respectivamente, nos períodos 1, 2, 3 e 4; $FAV(D)$, $FAV(H)$, $ADV(D)$, $ADV(H)$ são as variáveis latentes climáticas específicas de cada caráter de crescimento.

Tabela 8 – Efeitos Indiretos dos QTLs sobre os fenótipos

Fenótipo	QTL	Efeito Indireto
D1	QTL6 $_{\delta}$	-0.0419
D2	QTL6 $_{\delta}$	0,0075
D3	QTL6 $_{\delta}$	0,0023
	QTL3 $_{\alpha_2}$	0,0138
	QTL5 $_{\alpha_2}$	-0,0139
	QTL4 $_{\alpha_2}$	0,0248
	QTL9 $_{\alpha_1}$	0,0309
	QTL11 $_{\alpha_1}$	0,0327
	QTL10 $_{\alpha_2}$	0,0334
	QTL9 $_{\delta}$	-0,0351
D4	QTL6 $_{\delta}$	0,0003
	QTL3 $_{\alpha_1}$	0,0020
	QTL5 $_{\alpha_2}$	-0,0020
	QTL4 $_{\alpha_2}$	0,0036
	QTL9 $_{\alpha_1}$	-0,0041
	QTL11 $_{\alpha_1}$	-0,0044
	QTL10 $_{\alpha_2}$	-0,0045
	QTL9 $_{\delta}$	0,0047
	QTL12 $_{\delta}$	-0,0185
QTL12 $_{\alpha_1}$	0,0230	
H2	QTL6 $_{\delta}$	0,0380

D2, *D3* e *D4* são as variáveis de diâmetro nos períodos 2, 3 e 4; *H2* é a variável altura no período 3; α_1 , α_2 e δ são os efeitos aditivos do parental PB217, parental PR255 e de dominância no loco, respectivamente; os efeitos indiretos em cada componente periódico foram ordenados pelo módulo dos valores em ordem crescente.

5 DISCUSSÃO

O crescimento ao longo do tempo traz implicações na expressão genética do organismo. Os estádios de desenvolvimento envolvem a ativação e repressão de genes e a manifestação ou supressão de hormônios e enzimas, que estão intimamente relacionados com a maturação dos indivíduos. Outras variações que podem ocorrer são alterações ambientais, que comumente são de natureza edafo-climáticas e também podem alterar o comportamento genético do organismo.

Os dados utilizados nesse trabalho apresentam medidas ao longo de dois anos de crescimento da seringueira. Como essas medidas foram feitas aproximadamente de seis em seis meses, dois momentos climáticos particulares foram identificados em cada ano, um entre outubro e abril, que é característico por ter temperaturas razoavelmente mais elevadas e quantidade de chuvas maiores; e outro entre abril e outubro, marcado por temperaturas mais amenas e escassez de chuvas. Estes dois momentos foram aqui classificados como estações ou períodos favoráveis e adversos para o desenvolvimento vegetal.

Como era esperado, o comportamento de crescimento, em medidas de incremento, tanto para o diâmetro, como para altura, foi relevantemente superior nos períodos favoráveis, comportamento este que está evidentemente ligado aos benefícios e recursos benéficos que esses períodos proporcionam. Por outro lado, nos períodos menos favoráveis, ou adversos, o desenvolvimento das árvores foi gravemente comprometido, devido à carência de água, essencial na absorção de nutrientes do solo e relações hídricas, e às baixas temperaturas. Contudo se observarmos dentro das estações favoráveis e adversas, é facilmente visto que os períodos em que as plantas possuem idades mais avançadas, no caso os dois últimos períodos avaliados, as árvores apresentam valores de incremento superiores aos respectivos valores das duas primeiras medições correspondentes para cada estação.

Uma explicação para essa diferença seria pelas variações entre os dois períodos de uma mesma estação (favorável ou adversa), e devido à essas variações, o crescimento dos indivíduos terem valores diferentes mesmo em condições climáticas semelhantes. Mas isso não foi observado, ocorrendo inclusive, valores climáticos ligeiramente melhores justamente nos períodos iniciais e não nos dois períodos finais como era esperado (Figura 7).

Dessa forma, esse comportamento provavelmente está relacionado à curva de crescimento fisiológico vegetal que é dividida em três momentos principais. No primeiro, o desenvolvimento é pouco acelerado, sendo chamado de período adaptativo ou *fase lag*. Em seguida, os indivíduos entram no período de desenvolvimento exponencial, também chamado de *fase log*, onde é

durante esse momento que acontece o ponto de máximo de crescimento de toda vida do organismo. Por último ocorre a redução e estagnação do desenvolvimento, também chamada de *fase estacionária* (TAIZ; ZEIGER, 2010).

Nesse trabalho, há indicativo que os quatro períodos de incremento encontram-se entre as duas primeiras fases de desenvolvimento (fase lag e fase log), porém sofrendo forte interferência das estações com condições climáticas divergentes em que estão submetidos. Essa suposição faz sentido se pensarmos que caso os períodos estivessem somente em meio da primeira ou da terceira fase, os valores de crescimento estariam apenas sofrendo efeito do ambiente, assim não apresentando variações pronunciadas entre os períodos favoráveis ou entre os períodos adversos. Seguindo a mesma lógica, caso os quatro períodos de incremento estivessem contidos na segunda e terceira fase, os incrementos nas primeiras duas idades medidas seriam certamente maiores que os incrementos das duas idades finais, evidentemente respeitando a influência das condições climáticas e ambientais sobre os indivíduos.

O material genético utilizado neste trabalho continha indivíduos aparentados. O grau de parentesco variou desde irmãos completos irmãos completos, que foram utilizados como tratamentos na população de mapeamento de QTLs, meio-irmãos, parentes próximos e os dois parentais, que foram utilizados como testemunhas no experimento.

As testemunhas de maneira geral não apresentaram diferença significativa entre os tratamentos. Isso pode ser explicado pelo fato de serem materiais congêneres geneticamente, todavia, ao observarmos o comportamento dos parentais, e comparamos com o comportamento da sua progênie, vemos uma notável superioridade dos filhos em relação aos pais (Figura 6), principalmente nos períodos favoráveis.

Como a seringueira é uma espécie alógama, a diferença presente entre o desenvolvimento dos pais e da progênie pode ser explicada pela heterose frequente em espécies dessa natureza. Devido ao cruzamento do tipo divergente entre dois indivíduos não aparentados com diferenças marcantes para as duas características mensuradas, a capacidade de combinação genética dos alelos parentais proporcionou a manifestação de vigor híbrido na população de mapeamento, tendo assim, a progênie com valores de crescimento superiores aos dos pais. Essa superioridade foi mais acentuada entre as progênies e o parental PB217, este parental é caracterizado por ter maior produção de látex e com maiores concentrações de sacarose na sua constituição, porém é um indivíduo com baixos valores de crescimento em diâmetro e altura se comparado com o parental PR255, que é caracterizado por ter maior tolerância à danos causados pelo frio. Por esse motivo espera-se que os alelos do parental PR255 sejam mais efetivos na manifestação do crescimento em diâmetro e altura nos períodos adversos.

Para contemplar esses comportamentos genético-fisiológicos ao longo dos períodos, foram criados e testados alguns modelos estatísticos. Neste estudo em específico, foi interessante a utilização da abordagem de modelos mistos, pela capacidade de definição de alguns fatores como aleatórios, como por exemplo o material genético, e assim, ser possível a obtenção das estimativas de médias genóticas ajustadas (*E-BLUPs*) dos materiais utilizados, bem como, o emprego de estruturas de variância e co-variância (VCOV) hábeis em melhor explicar as relações presentes entre níveis de um mesmo fator.

Basicamente três modelos gerais foram testados: um para o caráter diâmetro, um para o caráter altura, e um último para ambos os caracteres simultaneamente (modelo multi-caracteres). Para tornar tais modelos mais realistas, nenhuma informação no que diz respeito a descrição ambiental foi negligenciada, sendo utilizadas desde as informações climáticas as quais as árvores foram submetidas, até as informações espaciais da localização de cada parcela no campo.

Os dados climáticos foram importantes na caracterização dos períodos, sendo úteis para descrever cada período como único, mas também agregar informações quantitativas meteorológicas particulares de cada um deles (Figura 7). Os dados espaciais utilizados foram as linhas e colunas de cada parcela no experimento (Figura 5). Essa informação foi útil para incorporar controle micro-ambiental que é superior ao macro-ambiental, como seria o caso da utilização de blocos experimentais. Este controle ambiental por linhas e colunas traz informações de vizinhança, tornando possível a absorção de diferenças presentes entre parcelas e não entre blocos.

Apesar de todas essas vantagens do modelo, a integração destes novos fatores ainda assume independência entre os níveis dos mesmos, por isso, foram testadas algumas estruturas de VCOV que incorporasse o grau de relacionamento que cada um apresenta.

O modelo adotado neste trabalho foi o de multi-caracteres (Tabela 4x), que ajusta as médias genóticas conjuntamente para os dois caracteres. Como são duas variáveis distintas, elas apresentam parâmetros próprios de média e variância, e possivelmente uma covariância significativa em comum. Para validar essa pressuposição, foram testadas as estruturas VCOV diagonal (DIAG), que contém a variância de cada variável em particular na diagonal da matriz, e a não estruturada (UNS), que contém além das variâncias na diagonal a co-variância entre tais variáveis.

Dado a alta correlação presente entre estes dois caracteres, como era de se esperar, a estrutura que melhor se ajustou aos dados foi a UNS, e esta matriz foi empregada tanto na parte residual (**R**) como na parte genética (**G**) da matriz aleatória do modelo. É possível observar pela Figura 8 que as correlações entre as médias genóticas ajustadas foram mantidas, algo que não seria possível caso tivesse sido utilizado um modelo independente para cada variável

individualmente.

Para selecionar matrizes de VCOV que melhor corrigissem os efeitos residuais ambientais, duas estratégias foram utilizadas: a primeira foi integrar as relações existentes entre linhas e colunas que estivessem espacialmente mais próximas, e conseqüentemente tais relações fossem diminuindo de acordo com o distanciamento espacial das parcelas contidas nas linhas e colunas. Ou seja, parcelas próximas tendem a ter influências ambientais mais parecidas que parcelas mais distantes; a segunda estratégia foi entender a relação existente entre os resíduos espaciais ao longo do tempo. Neste contexto as matrizes que melhor se adaptaram foram as auto-regressivas homogêneas de primeira ordem tanto para linha como para coluna (AR) e a fator analítico (FA) para os resíduos ao longo dos períodos (Tabela 4). Estas duas matrizes demonstraram sentido biológico bastante coerente. Na primeira estratégia, as matrizes AR foram interessantes pois como o experimento está alocado em local plano, assim, as variâncias entre linhas e colunas puderam perfeitamente serem declaradas como homogêneas, havendo neste caso pouca influência de relevo, como por exemplo: gradiente de fertilidade, capacidade de retenção de água no solo e/ou diferenças de luminosidade. Na segunda estratégia, a matriz FA foi consistente pois os períodos são intercalados com estações diferentes, ou seja, não seguiram uma progressão residual linear entre eles. Assim uma matriz que agrupasse a informação de co-variâncias entre os períodos, e sem excesso do número de parâmetros graças ao método de análise fatorial foi o modelo que garantiu o melhor ajuste.

Na correção dos efeitos residuais, as estruturas AR eleitas para linhas e colunas dos modelos individuais foram novamente empregadas. Esse critério foi utilizado seguindo dois argumentos: primeiro, a matriz AR foi superior em ambos os modelos individuais, gerando um forte indicio de ser o melhor no multi-caracteres; segundo, computacionalmente seria impraticável testar o modelo dado o volume de dados e a quantidade de combinações a serem testadas neste modelo conjunto.

Na inferência dos efeitos genéticos a matriz que obteve melhor ajuste para a VCOV genotípica ao longo dos períodos foi a UNS. Isso pode ser justificado pela complexidade genética presente na interação dos componentes genéticos de variâncias e covariâncias periódicas. Assim foi essencial a estimação de todos os parâmetros, não cabendo a aproximação da matriz FA. Além disso, a variação genética ao longo dos períodos também não segue uma progressão linear, graças à intercalação de períodos favoráveis e adversos.

A relação existente entre diâmetro e altura de árvores é amplamente conhecida e discutida na literatura e algumas sub-relações também estão contidas entre estas duas variáveis. Estas sub-relações podem ser decompostas em dados longitudinais de crescimento, como são o caso

dos dados utilizados neste trabalho.

Os vegetais apresentam fundamentalmente dois tipos de desenvolvimento, o do meristema primário que implica no crescimento vertical, e o do meristema secundário que implica no crescimento horizontal (TAIZ; ZEIGER, 2010). No caso das árvores, os indivíduos arbóreos necessitam conquistar um espaço no docéu da floresta, para terem acesso à luz para na produção de fotossimilados, tendo assim, que drenar seus recursos principalmente para o desenvolvimento do meristema apical em seus estágios iniciais, ganhando potencial adaptativo perante suas vizinhas competidoras.

Após atingir o estado de climáx, em que há uma estagnação do crescimento em altura, a árvore pode então drenar vigorosamente seus recursos no desenvolvimento secundário, isto é, em crescimento em diâmetro. Todavia, essa comutação de crescimento não ocorre de forma brusca, sendo que há evidentemente investimento em ambos os desenvolvimentos (primário e secundário) em todos os estádios de amadurecimento da planta, justificando a correlação presente entre estes dois caracteres em todos os períodos de desenvolvimento vegetal (Figura 8).

Consequentemente, as correlações presentes nos componentes periódicos de crescimento em diâmetro ao longo do tempo tendem a aumentar. Inicialmente como investimento no incremento dessa variável é limitado, os genótipos apresentam flutuação de quantidades de recursos empregados, e diferentes taxas de incremento em diâmetro são observadas. Posteriormente, o investimento nesta variável torna-se abundante. Desta maneira, os genótipos apresentam uniformidade de crescimento ao longo do tempo, justificando assim o aumento da correlação.

De forma inversa, as correlações presentes nos componentes periódicos de crescimento em altura ao longo do tempo tendem a diminuir. Neste caso, o raciocínio é semelhante ao dos componentes do diâmetro, com a diferença de que para essa variável, há um determinado momento no desenvolvimento do vegetal em que o crescimento é interrompido.

Todas estas relações são influenciadas pelas estações, de forma que, nas estações favoráveis os genótipos tendem a ter liberdade para expressar seu desenvolvimento tanto em altura como em diâmetro, explicando assim a intercalação de correlações entre estações mas sempre no sentido de aumentar ou diminuir-las. Por exemplo, as correlações existentes entre a altura e diâmetro ao longo do tempo foram 0,37, 0,19, 0,46 e 0,27, para os períodos 1, 2, 3 e 4, respectivamente (Figura 8). À primeira vista estes valores não aparentam ter uma ordenação crescente ou decrescente lógica, mas se olharmos com cuidado vemos que os valores dos períodos favoráveis (1 e 3) e dos períodos adversos (2 e 4) são crescentes, gerando a suposição de que as correlações entre diâmetro e altura aumentam com o passar do tempo, sendo afetadas pelas estações.

Cada momento no desenvolvimento pode ter uma relação diferente com o crescimento to-

tal para ambas as variáveis. Esta é uma análise muito interessante do ponto de vista do melhoramento genético, por proporcionar a possibilidade de realizar seleções indiretas precoces, evitando que o melhorista espere um tempo maior para realizar a seleção. Nesta análise, a influência da estação e do estágio de maturação podem ser determinantes no sucesso do programa de melhoramento, pois estágios de máximo crescimento certamente irão influenciar de forma mais crítica o crescimento total, que pode ser considerado como o momento definitivo de coleta do material para utilização comercial. Entretanto, a influência do clima pode mascarar os resultados da seleção precoce, por alterar o verdadeiro potencial de crescimento dos materiais genéticos desejados.

O modelo estrutural fenotípico (Figura 9) incluiu quatro variáveis latentes climáticas que conectavam os períodos de mesma estação climática. A opção pela utilização de quatro variáveis é justificada pela suposição de que o ambiente pode afetar de maneira diferenciada cada uma das variáveis, pois cada uma apresenta valores de herdabilidades genéticas diferentes (GONCALVES et al., 2005; FURLANI et al., 2005; COSTA et al., 2008, 2010), dessa maneira tendo diferentes padrões de interações ambientais. Esta hipótese foi corroborada levando em consideração que o ajuste deste modelo foi o melhor, pelos critérios de AIC e BIC empregados na avaliação de avaliação de modelos.

Para ambas variáveis de crescimento, ocorreu efeito de causalidade negativa entre os períodos 1 e 2 (β_{21} e β_{65}). Esse evento possivelmente aconteceu devido à inversão do potencial genético de crescimento entre esses dois períodos, indicando que os indivíduos que se desenvolveram com maior desenvoltura no primeiro período não foram os mesmos que mais cresceram no segundo. Uma explicação para esse fato pode ser pelo gasto excessivo de reservas por alguns genótipos no primeiro momento favorável, e quando no momento seguinte o meio se tornou adverso à todos os genótipos, àqueles que guardavam reservas sobressalentes tiveram a oportunidade de se desenvolver mais.

Todos os efeitos causais posteriores para a variável de crescimento em diâmetro (β_{32} e β_{43}) obtiveram valores positivos, indicando que os genótipos que cresceram mais em um momento tiveram a oportunidade de continuar crescendo nos dois momentos subsequentes. Já para a variável de crescimento em altura, a relação causal foi interrompida logo no momento seguinte, entre os períodos 2 e 3 (apenas se manteve o β_{65}), indicativo de que a estagnação do crescimento em altura logo será pronunciada.

Todas as relações diretas de causalidade entre os componentes periódicos de crescimento para um mesmo período foram comprovadas estatisticamente (β_{15} , β_{26} , β_{37} e β_{48}). Isto faz sentido se pensarmos que independentemente do desenvolvimento do meristema primário ou

secundário, a correlação entre essas duas variáveis é sempre existente em todos os estádios de desenvolvimento do vegetal. Afinal, embora seja a partir de um determinado momento que genótipos de maior porte tenham condições de se desenvolver mais em diâmetro, o desenvolvimento do meristema secundário é necessário em todos os estágios de amadurecimento da planta, por fornecer aporte anatômico para o crescimento em altura. Estes quatro efeitos são de ordem decrescente, no sentido do primeiro período para o último, indicando de que os genótipos com crescimento em altura atrasado, precisaram despender reservas com este crescimento nos momentos posteriores, deixando de empregá-las no crescimento em diâmetro.

O efeito do componente de crescimento em altura no período 3 apresentou relação causal negativa no componente de crescimento em diâmetro no período 4 (β_{47}). Este fato provavelmente ocorreu devido ao descompasso entre a desaceleração de crescimento em altura e a aceleração de crescimento em diâmetro nesses momentos, indicando que os genótipos com maiores crescimento em altura no período 3 não foram efetivamente os mesmo que mais cresceram em diâmetro no período 4.

Com base no modelo estrutural fenotípico, um modelo estrutural genotípico (Figura 11) foi pensado com o intuito de entender as relações causais entre os genes e os componentes periódicos de crescimento. Para tanto, inicialmente foram identificados QTLs que apresentaram efeito significativo sobre os caracteres estudados através da abordagem proposta por Gazaffi (2013). Como se tratava de uma população de mapeamento proveniente de um cruzamento biparental, sendo assim sendo uma progênie F1-segregante de irmãos completos, cada QTL apresentou efeitos genéticos aditivos específicos para cada parental PB217 e PR255 (α_1 e α_2), e o efeito de dominância (δ) intra-alélica que a combinação destes efeitos aditivos foram capazes de ocasionar.

O mapeamento de QTLs forneceu a informação de 16 locos significativos à expressão dos componentes periódicos de crescimento e para os componentes totais de crescimento em diâmetro e altura, que foram agrupados em 13 QTLs, seguindo a pressuposição de que dois ou mais QTLs, entre os caracteres utilizados, presentes em um mesmo intervalo de confiança são o mesmo QTL (Tabela 6). Desta maneira os 12 QTLs (13 totais, exceto o *QTL13* exclusivo para o fenótipo diâmetro no período total) para os componentes periódicos de crescimento foram utilizadas na montagem de um modelo capaz de inferir sobre a arquitetura genética destes caracteres de crescimento.

Os QTLs mapeados apresentaram diferentes padrões de segregação, e não somente o padrão 1:1 como é considerado pela abordagem Pseudo-Testcross (GRATTAPAGLIA; SEDEROFF, 1994). Isto representa uma vantagem no que diz respeito a capacidade de detecção de QTLs,

pois agrega a possibilidade de mapear QTLs para quaisquer padrão de segregação possível em uma população F1-segregante.

No modelo estrutural genotípico, todos os efeitos de cada QTL foram endereçados a causar, no modelo estrutural fenotípico, o fenótipo particular para o qual ele foi mapeado. E estes efeitos foram manejados de forma que somente os com valores de causalidade significativos fossem mantidos no modelo.

Dentre os 36 efeitos aditivos e de dominância para os 12 QTLs utilizados, somente os efeitos $QTL3_{\alpha_2}$, $QTL4_{\alpha_2}$, $QTL5_{\alpha_2}$, $QTL6_{\delta}$, $QTL9_{\alpha_1}$, $QTL9_{\delta}$, $QTL10_{\alpha_2}$, $QTL11_{\alpha_1}$, $QTL12_{\alpha_1}$ e $QTL12_{\delta}$ permaneceram no modelo, sendo que os três primeiros foram para a variável diâmetro nos períodos 2 e 4, e os 7 últimos para a variável altura nos períodos 1, 3 e 4. Embora tipicamente a altura em árvores seja um caráter de maior herdabilidade (NARAYANAN; MYDIN, 2011), um número mais elevado de QTLs foram mapeados para esta característica. Porém, como é possível observar, os efeitos de causalidade sobre os fenótipos dos mesmos também foram maiores (Tabela 7). Possivelmente este é o motivo para a detecção de um maior número de efeitos de QTLs para os componentes periódicos de crescimento em altura se comparados com o número dos componentes periódicos de crescimento em diâmetro.

Os valores dos parâmetros β 's de relações causais entre as variáveis fenotípicas no modelo estrutural genotípico mantiveram-se bastante semelhantes ao do modelo estrutural fenotípico, indicativo que o esqueleto do modelo estrutural genotípico é firme e consistente, ou seja não se alterou com a inclusão das informações genéticas. Dessa maneira, para o modelo estrutural genotípico, os parâmetros em comum com o modelo fenotípico têm mesma interpretação.

Analisando os efeitos dos QTLs sobre a variável diâmetro ao longo do tempo, somente foram detectados efeitos causais nos períodos de estação adversa e apenas efeitos aditivos do parental PR255 foram significativos ($\gamma_{23}(\alpha_2)$, $\gamma_{43}(\alpha_2)$, $\gamma_{24}(\alpha_2)$, $\gamma_{44}(\alpha_2)$ e $\gamma_{25}(\alpha_2)$). Sabendo-se que o parental PR255 é característico por ser mais tolerante a períodos mais frios, e entre os dois parentais foi o de melhor adaptação no local do experimento utilizado (Figura 6), estes resultados permitem a interpretação de que os alelos destes QTLs em particular, presentes neste parental, são determinantes no desenvolvimento dos caracteres de crescimento em diâmetro em períodos em que as condições climáticas são menos favoráveis.

Dos cinco efeitos mencionados acima, apenas o $\gamma_{25}(\alpha_2)$ apresentou valor negativo. Este efeito refere-se ao QTL_{α_2} sobre a variável diâmetro no período 2. Por representar um estágio de desenvolvimento precoce, em um período submetido à condições climáticas desfavoráveis, é possível especular que este QTL contenha genes de repressão de crescimento, interessante para frear o investimento de reservas por parte do vegetal em um momento de condições não

propícias ao desenvolvimento, ou ainda genes de manutenção de juvenildade.

Analisando os efeitos dos QTLs sobre a variável altura ao longo do tempo, foi possível observar que todos os valores dos efeitos de aditividade, independente do parental, obtiveram valores positivos, enquanto que todos os valores de dominância apresentaram valores negativos, o que caracteriza efeitos de sub-dominância ou dominância parcial no sentido do alelo recessivo e/ou repressor.

O efeito negativo de dominância do $QTL6_{\delta}$ sobre a altura no período 1, possivelmente é a supressão dos alelos vigorosos do parental PR255 por parte de alguns alelos do parental PB217 causando um efeito restritivo sobre vigor híbrido da progênie.

Nesta mesma variável, os períodos mais tardios 3 e 4, foram os que apresentaram um maior número de efeitos. Esses efeitos podem ser interessantes desde para iniciar a desaceleração da expressão desta variável em períodos de maturidade, como também serem genes respostas às diversidades ambientais.

O $QTL9$ apresentou tanto efeitos positivos de α_1 , como efeitos negativos de δ sobre a altura no período 3. Se observamos a Figura 6, vemos que neste período, atipicamente, o parental PB217 é superior ao parental PR255. Esse fato pode ser condicionado a tentativa desde genótipo em repor o crescimento não alcançado nos períodos anteriores. Neste momento há dois efeitos positivos aditivos de dois QTLs diferentes deste parental ($\gamma_{79}(\alpha_1)$ e $\gamma_{711}(\alpha_1)$) e um positivo do parental PR255 ($\gamma_{710}(\alpha_2)$) estimulando o crescimento em altura, porém, no caso do $QTL9$, há o efeito negativo de dominância que pode estar reprimindo essa atividade nos indivíduos da progênie.

Finalmente, no último período de crescimento em altura há o $QTL12$, que apresenta efeito aditivo positivo do parental PB217 e efeito devido à dominância negativo. Neste momento os efeitos aditivos provavelmente estão condicionando um maior crescimento em altura, caracterizando a possível estabilidade de crescimento em altura deste parental ao longo dos períodos, porém o efeito de dominância está regulando o crescimento nesse momento devido ao clima adverso.

Com base nestes resultados com relação aos componentes periódicos em altura, cria-se a suposição de que embora o parental PB217 seja inferior ao parental PR255 para esta variável, ele apresenta maior estabilidade ao longo dos períodos. Essa hipótese pode ser reforçada pelo fato de que este é um genótipo caracterizado por apresentar produção constante de látex durante todo ciclo extrativo, com concentrações mais elevadas de sacarose. Essa característica possivelmente está atrelada à necessidade de indivíduos mais altos e conseqüentemente com maior potencial fotossintetizante, e não necessariamente indivíduos de maior diâmetro.

A partir do modelo estrutural genotípico, também foi possível inferir sobre os efeitos indiretos que um QTL não diretamente relacionado à uma característica pode exercer sobre ela. Esses valores especialmente dão um direcionamento ao pesquisador sobre quais efeitos indiretos podem ou não existir, o sentido dos mesmos (positivo ou negativo) e uma percepção comparativa entre a magnitude dos valores.

Dessa maneira, embora não tenha sido encontrado QTLs pleiotrópicos entre as duas variáveis de crescimento, há indícios que o efeito de dominância do $QTL6_{\delta}$ que apresenta relação direta de valor negativo sobre o caráter altura no período 1, também exerça efeito indireto negativo no diâmetro no período 1. Além deste, os efeitos aditivos dos QTLs $QTL9_{\alpha_1}$, $QTL10_{\alpha_2}$ e $QTL11_{\alpha_1}$ exerçam além de efeitos diretos sobre a altura no período 3, também podem exercer efeitos indiretos positivos sobre o diâmetro nesse mesmo período, e o $QTL9_{\delta}$, efeito negativo indireto de dominância sobre esta mesma variável. Da mesma maneira, os efeitos de aditividade e dominância do QTL12 ($QTL12_{\alpha_1}$ e $QTL12_{\delta}$), que possuem influência direta sobre a variável altura no período 4, também podem exercer algum efeito sobre a variável diâmetro no período 4 (Tabela 8).

A respeito destes efeitos indiretos, é interessante comentar que, mesmo os caracteres que não apresentaram QTLs associados a sua expressão fenotípica, podem ter QTLs exercendo alguma influência em suas manifestações, como é o caso dos caracteres diâmetro no período 1 e 3, e a altura no período 2. Os resultados dos efeitos indiretos, de forma geral, podem ser de interesse do melhorista, por gerar informações sobre a seleção de genes que podem ter efeitos indiretos indesejáveis em outros caracteres. Abrindo assim uma possibilidade de analisar as vantagens e desvantagens em selecionar determinados marcadores para integrarem um programa de melhoramento vegetal auxiliado por marcadores genéticos.

Portanto é possível afirmar que os resultados da arquitetura genética dos componentes de crescimento ao longo do tempo, são úteis para enriquecer um programa de melhoramento assistido por marcadores, pois a arquitetura traz informações de mapeamento funcional sobre a atividade, relações e interações entre os genótipos e fenótipos. Estas informações são convenientes na seleção direcionada dos marcadores com finalidades específicas, sejam elas à respeito de selecionar indivíduos com explosão de crescimento inicial, materiais adaptados às condições climáticas adversas com objetivo de escape espacial à doença mal-das-folhas (MDF), potenciais produtores de látex, materiais com diâmetro elevado para fins madeireiros ou ainda genótipos mais adaptados ou mesmo com maior estabilidade ao longo do tempo.

6 CONCLUSÃO

Modelos com ajuste de matrizes de variância e covariância espacial e entre períodos e com a inclusão de co-variáveis de dados meteorológicos obtiveram melhores ajustes quando comparados com outros.

A abordagem utilizada foi capaz de indentificar efeitos aditivos e devido à dominância interessantes para o desenvolvimento da cultura em períodos de menores temperaturas e escassez de água, apontar o parental PR255 como portador de alelos importantes para o crescimento em diâmetro e altura, explicar o padrão genético-comportamental de crescimento no período em que as progênies foram avaliadas e estimar efeitos indiretos de QTLs não mapeados para determinadas características.

Tal abordagem demonstrou-se muito promissora para utilização em programas de melhoramento genético assistido por marcadores, por agregar informações pertinentes à arquitetura genética de caracteres importantes.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **The Institute of Electrical and Electronics Engineers Transaction and Automatic Control**, Notre Dame, v. 19, p. 716–723, 1974.
- ALIMI, N. A.; BINK, M. C. A. M.; DIELEMAN, J. A.; MAGAN, J. J.; WUBS, A. M.; PALLOIX, A.; EEUWIJK, F. A. van. Multi-trait and multi-environment qtl analyses of yield and a set of physiological traits in pepper. **Theoretical and Applied Genetics**, New York, v. 126, p. 2597-2625, 2013.
- ALLEN, H. Anticipating market demand: tracking enrollee satisfaction and health over time. **International Journal for Quality in Health Care**, Oxford, v. 10, p. 521–530, 1998.
- ALWALA, S.; KIMBENG, C.; HENRY, R.; KOLE, C. et al. Molecular genetic linkage mapping in saccharum: strategies, resources and achievements. **Genetics, genomics and breeding of sugarcane**, Austin, p. 69–96, 2010.
- ANDERSON, J. C.; GERBING, D. W. Some methods for respecifying measurement models to obtain unidimensional construct measurement. **Journal of Marketing Research**, Chicago, v. 19, p. 453–460, 1982.
- ANONI, C. **Mapeamento de QTLs em progênies de irmãos-completos de cana-de- açúcar e utilizando imputação múltipla**. 2012. 107 p. Tese (Doutorado) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2012.
- ARSHAD, M.; ALI, N.; GHAFOOR, A. Character correlation and path coefficient in soybean glycine max (l.) merrill. **Pakistan Journal of Botany**, Peshawar, v. 38, p. 121–130, 2006.
- BALZARINI, M. 23 applications of mixed models in plant breeding. **Quantitative genetics, genomics, and plant breeding**, New York, p. 353, 2002.
- BASHA, S. D.; SUJATHA, M. Inter and intra-population variability of jatropha curcas (l.) characterized by rapd and issr markers and development of population-specific scar markers. **Euphytica**, Wageningen, v. 156, p. 375–386, 2007.
- BESSE, P.; SEGUIN, M.; LEBRUN, P.; CHEVALLIER, M. H.; NICOLAS, D.; LANAUD, C. Genetic diversity among wild and cultivated populations of hevea-brasiliensis assessed by nuclear rflp analysis. **Theoretical and Applied Genetics**, Berlin, v. 88, p. 199–207, 1994.
- BLAU, P. M.; DUNCAN, O. D. The american occupational structure. **New York Vilcy**, New York, ERIC, 1967.
- BOER, M. P.; WRIGHT, D.; FENG, L.; PODLICH, D. W.; LUO, L.; COOPER, M.; EEUWIJK, F. A. van. A mixed-model quantitative trait loci (qtl) analysis for multiple-environment trial data using environmental covariables for qtl-by-environment interactions, with an example in maize. **Genetics**, Austin, v. 177, p. 1801–1813, 2007.
- BOLIGON, A. A.; MERCADANTE, M. E. Z.; FORNI, S.; LOBO, R. B.; ALBUQUERQUE, L. G. Covariance functions for body weight from birth to maturity in nellore cows. **Journal of Animal Science**, Albany, v. 88, p. 849–859, 2010.

- BOLLEN, K. A.; LONG, J. S. Tests for structural equation models - introduction. **Sociological Methods & Research**, London, v. 21, p. 123–131, 1992.
- BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W. Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics**, Chicago, v. 32, p. 314–331, 1980.
- BOZDOGAN, H. Model selection and akaike information criterion (aic) - the general-theory and its analytical extensions. **Psychometrika**, Colorado, v. 52, p. 345-370, 1987.
- BROMAN, K. W.; SEN, S. A guide to qtl mapping with r/qtl introduction. **Guide to Qtl Mapping with R/Qtl**, Oxford, 2009.
- BROOKER, R. J.; BUSS, K. A.; LEMERY-CHALFANT, K.; AKSAN, N.; DAVIDSON, R. J.; GOLDSMITH, H. H. The development of stranger fear in infancy and toddlerhood: normative development, individual differences, antecedents, and outcomes. **Applied Developmental Science**, Oxford, v. 16, p. 864-878, 2013.
- BROWN, H.; PRESCOTT, R. Applied mixed models in medicine. **Technometrics**, Washington, v. 42, p. 442–443, 2006.
- BURGUENO, J.; CAMPOS, G. de los; WEIGEL, K.; CROSSA, J. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. **Crop Science**, Madison, v. 52, p. 707–719, 2012.
- BURGUENO, J.; CROSSA, J.; CORNELIUS, P. L.; YANG, R.-C. Using factor analytic models for joining environments and genotypes without crossover genotype x environment interaction. **Crop Science**, Madison, v. 48, p. 1291–1305, 2008.
- BURGUENO, J.; CROSSA, J.; COTES, J. M.; VICENTE, F. S.; DAS, B. Prediction assessment of linear mixed models for multienvironment trials. **Crop Science**, Madison, v. 51, p. 944–954, 2011.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference - understanding aic and bic in model selection. **Sociological Methods & Research**, London, v. 33, p. 261–304, 2004.
- BUTCHER, P. A.; WILLIAMS, E. R.; WHITAKER, D.; LING, S.; SPEED, T. P.; MORAN, G. F. Improving linkage analysis in outcrossed forest trees - an example from acacia mangium. **Theoretical and Applied Genetics**, Berlin, v. 104, p. 1185–1191, 2002.
- BUXADERA, A. M.; AYRADO, Y. Genetic (co)variance for sire fertility estimated by additive, non-additive and longitudinal models in holstein-zebu cross-bred cows. **Cambridge Journals: Animal**, Cambridge, v. 7, p. 721–728, 2013.
- CHAIBUB-NETO, E.; FERRARA, C. T.; ATTIE, A. D.; YANDELL, B. S. Inferring causal phenotype networks from segregating populations. **Genetics**, Austin, v. 179, p. 1089–1100, 2008.
- CHRISTENSEN, R.; PEARSON, L. M.; JOHNSON, W. Case-deletion diagnostics for mixed models. **Technometrics**, Washington, v. 34, p. 38–45, 1992.
- CHURCHILL, G. A.; DOERGE, R. W. Empirical threshold values for quantitative trait mapping. **Genetics**, Austin, v. 138, p. 963–971, 1994.

- COLLARD, B.; JAHUFER, M.; BROUWER, J.; PANG, E. An introduction to markers, quantitative trait loci (qtl) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica**, Wageningen, v. 142, p. 169–196, 2005.
- COSTA, R. B. d.; RESENDE, M. D. V. d.; ARAUJO, A. J. d.; GONÇALVES, P. d. S.; HIGA, A. R. Selection and genetic gain in rubber tree (hevea) populations using a mixed mating system. **Genetics and Molecular Biology**, Ribeirão Preto, v. 23, p. 671–679, 2000.
- COSTA, R. B. d.; RESENDE, M. D. V. d.; GONÇALVES, P. d. S.; CHICHORRO, J. F.; ROA, R. A. R. Variabilidade genética e seleção para caracteres de crescimento da seringueira. **Bragantia**, Campinas, v. 67, p. 299–305, 2008.
- COSTA, R. B. d.; RESENDE, M. D. V. d.; GONÇALVES, P. d. S.; ROA, R. A. R.; FEITOSA, K. C. d. O. Predição de parâmetros e valores genéticos para caracteres de crescimento e produção de látex em progênies de seringueira. **Bragantia**, Campinas, v. 69, p. 49–56, 2010.
- CROSSA, J.; BURGUENO, J.; CORNELIUS, P. L.; MCLAREN, G.; TRETOWAN, R.; KRISHNAMACHARI, A. Modeling genotype x environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Science**, Madison, v. 46, p. 1722–1733, 2006.
- CROSSA, J.; YANG, R. C.; CORNELIUS, P. L. Studying crossover genotype x environment interaction using linear-bilinear models and mixed models. **Journal of Agricultural Biological and Environmental Statistics**, Alexandria, v. 9, p. 362–380, 2004.
- CULLIS, B.; GOGEL, B.; VERBYLA, A.; THOMPSON, R. Spatial analysis of multi-environment early generation variety trials. **Biometrics**, Washington, v. 54, p. 1–18, 1998.
- CURTIS, R. O. Height-diameter and height-diameter-age equations for second-growth douglas-fir. **Forest Science**, Lawrence, v. 13, p. 365, 1967.
- DARLINGTON, R. B. **Regression and linear models**. 6th edition. New York: McGraw-Hill, 1990. 480 p.
- DASLIN, A.; BAIHAKI, A.; DANAKUSUMA, T.; HAERUMAN, M. Genotypes x environment interaction in rubber and their implications in clonal selection. **Bull. Perkaretan**, Indonesia, v. 4, p. 23–28, 1986.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society**, Oxford, p. 1–38, 1977.
- DIAMANTOPOULOS, A.; SIGUAW, J. A.; SIGUAW, J. A. **Introducing LISREL: A guide for the uninitiated**. Thousand Oaks: Sage, 2000. 171 p.
- DIGGLE, P. J. An approach to the analysis of repeated measurements. **Biometrics**, Washington, v. 44, p. 959–971, 1988.
- DOERGE, R. W. Mapping and analysis of quantitative trait loci in experimental populations. **Nature Reviews Genetics**, New York, v. 3, p. 43–52, 2002.
- DOERGE, R. W.; ZENG, Z. B.; WEIR, B. S. Statistical issues in the search for genes affecting quantitative traits in experimental populations. **Statistical Science**, Ithaca, v. 12, p. 195–219, 1997.

- EERIKAINEN, K. Predicting the height-diameter pattern of planted pinus kesiya stands in zambia and zimbabwe. **Forest Ecology and Management**, Amsterdam, v. 175, p. 355–366, 2003.
- FALCONER, D. S; MACKAY T.F. C. **Introduction to Quantitative Genetics**. 4th edition, New York: Ronald Press Co, 1981. 386 p.
- FAN, X.; THOMPSON, B.; WANG, L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. **Structural Equation Modeling**, Los Angeles, v. 6, p. 56–83, 1999.
- FERREIRA, M.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa-Cenargen, Brasília, 1996.
- FONSECA, S.; RESENDE, M.D.V.; ALFENAS, A.; GUIMARÃES, L.; ASSIS, T.; GRATTAPAGLIA, D. **Manual prático de melhoramento genético de eucalipto**. [S.l.]: Ed. UFV, Viçosa, 2010. 219 p.
- FORNELL, C.; LARCKER, D. F. Structural equation models with unobservable variables and measurement error - algebra and statistics. **Journal of Marketing Research**, Chicago, v. 18, p. 382–388, 1981.
- FOX, J. Structural equation modeling with lisrel - essentials and advances - hayduk,la. **Canadian Journal of Sociology-Cahiers Canadiens De Sociologie**, Edmonton, v. 14, p. 257–259, 1989.
- FOX, J. Structural equation modeling with the sem package in R. **Structural Equation Modeling-a Multidisciplinary Journal**, Los Angeles, v. 13, p. 465–486, 2006.
- FOX, J. Structural Equation Modeling in R with the sem Package. **An Appendix to An R Companion to Applied Regression**, Hamilton, v. 1, p. 1–54, 2012.
- FREITAS, E. G. **Uso de informações de parentesco e modelos mistos para avaliação e seleção de genótipos de cana-de-açúcar**. 2013. 105 p. Tese (Doutorado) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2013.
- FURLANI, R. C. M.; MORAES, M. L. T. de; RESENDE, M. D. V. de; FURLANI, E.; GONCALVES, P. D. S.; VALERIO, W. V.; PAIVA, J. R. de. Estimation of variance components and prediction of breeding values in rubber tree breeding using the reml/blup procedure. **Genetics and Molecular Biology**, Ribeirão Preto, v. 28, p. 271–276, 2005.
- GARCIA, A. A. F.; KIDO, E. A.; MEZA, A. N.; SOUZA, H. M. B.; PINTO, L. R.; PASTINA, M. M.; LEITE, C. S.; SILVA, J. A. G. da; ULIAN, E. C.; FIGUEIRA, A.; SOUZA, A. P. Development of an integrated genetic map of a sugarcane (saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **Theoretical and Applied Genetics**, Berlin, v. 112, p. 298–314, 2006.
- GARCIA, A. A. F.; WANG, S.; MELCHINGER, A. E.; ZENG, Z.-B. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. **Genetics**, Austin, v. 180, p. 1707–1724, 2008.
- GAZAFFI, R; MARGARIDO, G. R. A.; PASTINA, M. M.; MOLLINARI, M; GARCIA, A. A. F. A model for quantitative trait loci mapping, linkage phase and segregation pattern estimation for a full-sib progeny. **Tree Genetics and Genomes**, Berlin, 2013.

GERE, J.; SCHIMMACK, U. A multi-occasion multi-rater model of affective dispositions and affective well-being. **Journal of Happiness Studies**, Dordrecht, v. 12, p. 931–945, 2011.

GILLARD, J. An overview of linear structural models in errors in variables regression. **Revstat-Statistical Journal**, Lisboa, v. 8, p. 57–80, 2010.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Washington, p. 1440–1450, 1995.

GONCALVES, P. D. S.; BATAGLIA, O. C.; SANTOS, W. R. dos; ORTOLANI, A. A.; SEGNINI, I.; SHIKASHO, E. H. Growth trends, genotype-environment interaction and genetic gains in six-year-old rubber tree clones (hevea) in sao paulo state, brazil. **Genetics and Molecular Biology**, Ribeirão Preto, v. 21, p. 115–122, 1998.

GONCALVES, P. d. S.; MORAES, M. L. T.; BORTOLETTO, N.; COSTA, R. B. da; GONCALVES, E. C. P. Genetic variation in growth traits and yield of rubber trees (hevea brasiliensis) growing in the brazilian state of sao paulo. **Genetics and Molecular Biology**, Ribeirão Preto, v. 28, p. 765–772, 2005.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic-linkage maps of eucalyptus-grandis and eucalyptus-urophylla using a pseudo-testcross - mapping strategy and rapd markers. **Genetics**, London, v. 137, p. 1121–1137, 1994.

GUEN, V. L.; GARCIA, D.; MATTOS, C. R. R.; DOARE, F.; LESPINASSE, D.; SEGUIN, M. Bypassing of a polygenic microcyclus ulei resistance in rubber tree, analyzed by qtl detection. **New Phytologist**, Cambridge, v. 173, p. 335–345, 2007.

GUEN, V. L.; GAY, C.; XIONG, T. C.; SOUZA, L. M.; RODIER-GOUD, M.; SEGUIN, M. Development and characterization of 296 new polymorphic microsatellite markers for rubber tree (hevea brasiliensis). **Plant Breeding**, Berlin, v. 130, p. 294–296, 2011.

GUEN, V. L.; LESPINASSE, D.; OLIVER, G.; RODIER-GOUD, M.; PINARD, F.; SEGUIN, M. Molecular mapping of genes conferring field resistance to south american leaf blight (microcyclus ulei) in rubber tree. **Theoretical and Applied Genetics**, Berlin, v. 108, p. 160–167, 2003.

GUERIN, L.; STROUP, W. A simulation study to evaluate proc mixed analysis of repeated measures data. In: **Proceedings of the 2000 Conference on Applied Statistics in Agriculture**, Manhattan, 2000. p. 170–203.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; WILLIAM, C. **Multivariate data analysis**. 7th edition. Saddle River: Prentice Hall, 1998. 816 p.

HAYASHI, F. Econometrics. 2000. **Princeton University Press. Section**, Princeton, v. 1, p. 60–69, 2000.

HEINZ, D.; TEW, T. Hybridization procedures. **Sugarcane improvement through breeding**. Elsevier, Amsterdam, v. 3, p. 313–342, 1987.

HENDERSON, C. **Applications of linear models in animal breeding**. 3th edition. Ontario: University of Guelph, 1984. 462 p.

HILLIER, G. H.; KINAL, T. W.; SRIVASTAVA, V. K. On the moments of ordinary least-squares and instrumental variables estimators in a general structural equation. **Econometrica**, Chicago, v. 52, p. 185–202, 1984.

HOCKING, R. R. A biometrics invited paper. the analysis and selection of variables in linear regression. **Biometrics**, Washington, v. 32, p. 1–49, 1976.

HOYLE, R. **SEM: concepts, issues, and applications**. 3th edition. California: Sage, 1995. 289.

HU, X.; SPILKE, J. Variance-covariance structure and its influence on variety assessment in regional crop trials. **Field Crops Research**, Amsterdam, v. 120, p. 1–8, 2011.

JACKSON, D. L. Revisiting sample size and number of parameter estimates: Some support for the $n : q$ hypothesis. **Structural Equation Modeling**, Los Angeles, v. 10, p. 128–141, 2003.

JAYASEKERA, N. E. M. A basis for selecting hevea clones stable to unpredictable agro-climatic variability. **Silvae Genetica**, Frankfurt, v. 32, p. 181–184, 1983.

JANSEN, R. C.; STAM, P. High resolution of quantitative traits into multiple loci via interval mapping. **Genetics**, Austin, v. 136, p. 1447–1455, 1994.

JIANG, C. J.; ZENG, Z. B. Multiple-trait analysis of genetic-mapping for quantitative trait loci. **Genetics**, Austin, v. 140, p. 1111–1127, 1995.

JÖRESKOG, K.; THILLO, M. van. LISREL: A general computer program for estimating a linear structural equations system involving indicators of unmeasured variables. **Educational Resources Information Center**, Princeton, v. 53, p. 1–73, 1972.

JÖRESKOG, K. G.; SORBOM, D. Recent developments in structural equation modeling. **Journal of Marketing Research**, Chicago, v. 19, n. 4, p. 404–416, 1982.

KAO, C. H.; ZENG, Z. B. General formulas for obtaining the mles and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the em algorithm. **Biometrics**, Washington, v. 53, p. 653–665, 1997.

KAO, C. H.; ZENG, Z. B.; TEASDALE, R. D. Multiple interval mapping for quantitative trait loci. **Genetics**, Austin, v. 152, p. 1203–1216, 1999.

KAPLAN, D. **Structural equation modelling: Advanced methods and applications**. 2th edition. Los Angeles: Sage, 2000. 215 p.

KELLOWAY, E. K. **Using LISREL for structural equation modeling: A researcher's guide**. 3th edition. Thousand Oaks: Sage, 1998. p. 160.

KELLY, A. M.; SMITH, A. B.; ECCLESTON, J. A.; CULLIS, B. R. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. **Crop Science**, Madison, v. 47, p. 1063–1070, 2007.

KLEIBERGEN, F. Pivotal statistics for testing structural parameters in instrumental variables regression. **Econometrica**, Chicago, v. 70, p. 1781–1803, 2002.

KLING, R. **Principles and practices of structural equation modelling**. 3th edition. New York: Guilford Press, 2001. 315 p.

- KNOWE, S. A. Effect of competition control treatments on height age and height diameter relationships in young douglas-fir plantations. **Forest Ecology and Management**, Amsterdam, v. 67, p. 101–111, 1994.
- KULLBACK, S. An application of information theory to multivariate analysis. **Annals of Mathematical Statistics**, Ann Arbor, v. 23, p. 88–102, 1952.
- LANDER, E. S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using rflp linkage maps. **Genetics**, Austin, v. 121, p. 185–199, 1989.
- LESPINASSE, D.; GRIVET, L.; TROISPOUX, V.; RODIER-GOUD, M.; PINARD, F.; SEGUIN, M. Identification of qtls involved in the resistance to south american leaf blight (microcyclus ulei) in the rubber tree. **Theoretical and Applied Genetics**, Berlin, v. 100, p. 975–984, 2000.
- LI, R.; TSAIH, S.-W.; SHOCKLEY, K.; STYLIANOU, I. M.; WERGEDAL, J.; PAIGEN, B.; CHURCHILL, G. A. Structural model analysis of multiple quantitative traits. **Plos Genetics**, San Francisco, v. 2, p. 1046–1057, 2006.
- LIN, M.; LOU, X. Y.; CHANG, M.; WU, R. L. A general statistical framework for mapping quantitative trait loci in nonmodel systems: Issue for characterizing linkage phases. **Genetics**, Austin, v. 165, p. 901–913, 2003.
- LINDSEY, J. K. **Models for repeated measurements** 2th edition. Oxford: Oxford University Press, 1999. 536 p.
- LITTELL, R. C. Analysis of unbalanced mixed model data: A case study comparison of anova versus reml/gls. **Journal of Agricultural Biological and Environmental Statistics**, Alexandria, v. 7, p. 472–490, 2002.
- LITTELL, R. C.; HENRY, P. R.; AMMERMAN, C. B. Statistical analysis of repeated measures data using sas procedures. **Journal of Animal Science**, Albany, v. 76, p. 1216–1231, 1998.
- LIU, B.; FUENTE, A.; HOESCHELE, I. Gene network inference via structural equation modeling in genetical genomics experiments. **Genetics**, Austin, v. 178, p. 1763–1776, 2008.
- LOW, F.; ATAN, S.; JAAFAR, H.; TAN, H. Recent advances in the development of molecular markers for hevea studies. **Journal of Natural Rubber Research**, Kuala Lumpur, v. 11, 1996.
- LU, Q.; CUI, Y. H.; WU, R. L. A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. **BMC Genetics**, London, v. 5, 2004.
- LYNCH, M.; WALSH, B. et al. **Genetics and analysis of quantitative traits**. Sinauer Sunderland, p. 980, 1998.
- MACKAY, T. F. C. The genetic architecture of quantitative traits. **Annual Review of Genetics**, Palo Alto, v. 35, p. 303–339, 2001.
- MALIEPAARD, C.; JANSEN, J.; OOIJEN, J. W. V. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetical Research**, London, v. 70, p. 237–250, 1997.

MALOSETTI, M.; RIBAUT, J. M.; VARGAS, M.; CROSSA, J.; BOER, M. P.; EEUWIJK, F. A. V. Multi-trait multi-environment QTL modelling for drought-stress adaptation in maize. **Euphytica**, Wageningen, v. 149, p. 25–36, 2007.

MALOSETTI, M.; RIBAUT, J. M.; VARGAS, M.; CROSSA, J.; EEUWIJK, F. A. van. A multi-trait multi-environment qtl mixed model with an application to drought and nitrogen stress trials in maize (*zea mays* l.). **Euphytica**, Wageningen, v. 161, p. 241–257, 2008.

MALOSETTI, M.; VOLTAS, J.; ROMAGOSA, I.; ULLRICH, S. E.; EEUWIJK, F. A. van. Mixed models including environmental covariables for studying qtl by environment interaction. **Euphytica**, Wageningen, v. 137, p. 139–145, 2004.

MARGARIDO, G. **Mapeamento de QTLs em múltiplos caracteres e ambientes em cruzamento comercial de cana-de-açúcar usando modelos mistos**. 2011. 108 p. Tese (Doutorado) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2011.

MARGARIDO, G. R. A.; SOUZA, A. P.; GARCIA, A. A. F. Onemap: software for genetic mapping in outcrossing species. **Hereditas**, Lund, v. 144, p. 78–79, 2007.

MARTENS, M. P.; HAASE, R. F. Advanced applications of structural equation modeling in counseling psychology research. **Counseling Psychologist**, Newbury Park, v. 34, p. 878–911, 2006.

MARUYAMA, G. **Basics of structural equation modelling**. 1th edition. Thousand Oaks: Sage, 1998. 328 p.

MCARDLE, J. J. Causal modeling applied to psychonomic systems simulation. **Behavior Research Methods and Instrumentation**, Austin, v. 12, p. 193–209, 1980.

MCCULLOCH, C.; SEARLE, S.; NEUHAUS, J. **Generalized, linear, and mixed models**. 2th edition. New York: Wiley, 2001. 384 p.

MENENDEZ-BUXADERA, A.; CARLEOS, C.; BARO, J. A.; VILLA, A.; CANON, J. Multi-trait and random regression approaches for addressing the wide range of weaning ages in asturiana de los valles beef cattle for genetic parameter estimation. **Journal of Animal Science**, Albany, v. 86, p. 278–286, 2008.

MOOD, A. M.; GRAYBILL, F.; BOES, D. **Introduction to the theory of statistics**. 3th edition. New York: McGraw-Hill, 1974. 577 p.

NARAYANAN, A. A review of eight software packages for structural equation modeling. **American Statistician**, Washington, v. 66, p. 129–138, 2012.

NARAYANAN, C.; MYDIN, K. K. Heritability of yield and secondary traits in two populations of para rubber tree (*hevea brasiliensis*). **Silvae Genetica**, Frankfurt, v. 60, p. 132–139, 2011.

NIGH, G. D.; LOVE, B. A. A model for estimating juvenile height of lodgepole pine. **Forest Ecology and Management**, Amsterdam, v. 123, p. 157–166, 1999.

OMAN, S. D. Multiplicative effects in mixed model analysis of variance. **Biometrika**, Oxford, v. 78, p. 729–739, 1991.

ONOKPISE, O.; OLAPADE, O.; MEKAKO, H. Genotype x environment interaction in hevea brasiliensis (mull. arg.). **Indian Journal of Genetics and Plant Breeding**, New Delhi, v. 46, 1986.

ONOKPISE, O. U. Natural rubber, hevea brasiliensis (willd. ex a. juss.) mull. arg., germplasm collection in the amazon basin, brazil: A retrospective. **Economic Botany**, Bronx, v. 58, p. 544–555, 2004.

PASTINA, M. **Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar usando a abordagem de modelos mistos**. 2010. 89 p. Tese (Doutorado) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2010.

PATTERSON, H.; SILVEY, V.; TALBOT, M.; WEATHERUP, S. Variability of yields of cereal varieties in uk trials. **Journal of Agricultural Science**, Cambridge, v. 89, p. 239, 1977.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, Oxford, v. 58, p. 545–554, 1971.

PAYNE, R.; WELHAM, S.; HARDING, S. **A Guide to REML in GenStat**. 16th edition. Rothamsted: VSN International, 2013. 97 p.

PEREIRA, M.; PEREIRA, T. **Marcadores moleculares no pré-melhoramento de plantas**. 1.ed. Viçosa: Editora UFV, 2006. 106 p.

PIEPHO, H. P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, Washington, v. 53, p. 761–766, 1997.

PIEPHO, H. P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, Berlin, v. 97, p. 195–201, 1998.

PIEPHO, H. P. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. **Genetics**, Austin, v. 156, p. 2043–2050, 2000.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D. Linear and nonlinear mixed effects models. **R package version**, New York, v. 3, p. 57, 2007.

PIRES, J.; SECCO, R.; IVANIR, J. Taxonomia e fitogeografia das seringueiras (hevea spp., euphorbiaceae). **Embrapa Amazônia Oriental**, Belém, p. 103, 2002.

PRIYADARSHAN, P. M.; GONCALVES, P. D. S. Hevea gene pool for breeding. **Genetic Resources and Crop Evolution**, Dordrecht, v. 50, p. 101–114, 2003.

RABINER, L. R. A tutorial on hidden markov-models and selected applications in speech recognition. **Proceedings of the Institute of Electrical and Electronics Engineers**, London, v. 77, p. 257–286, 1989.

RAJ, S.; DAS, G.; POTHEN, J.; DEY, S. K. Relationship between latex yield of hevea brasiliensis and antecedent environmental parameters. **International Journal of Biometeorology**, Lisse, v. 49, p. 189–196, 2005.

- RAO, P. S.; SARASWATHYAMMA, C. K.; SETHURAJ, M. R. Studies on the relationship between yield and meteorological parameters of para rubber tree (*hevea brasiliensis*). **Agricultural and Forest Meteorology**, Amsterdam, v. 90, p. 235–245, 1998.
- RAYKOV, T.; MARCOULIDES, G. A. A method for comparing completely standardized solutions in multiple groups. **Structural Equation Modeling-a Multidisciplinary Journal**, Los Angeles, v. 7, p. 292–308, 2000.
- RESENDE, M. D. V.; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Rev. Mat. Estat**, São Paulo, v. 22, p. 31–52, 2004.
- RIBAUT, J. M.; HOISINGTON, D. A.; DEUTSCH, J. A.; JIANG, C.; GONZALEZDELEON, D. Identification of quantitative trait loci under drought conditions in tropical maize .11. flowering parameters and the anthesis-silking interval. **Theoretical and Applied Genetics**, Berlin, v. 92, p. 905–914, 1996.
- RIDOUT, M. S.; TONG, S.; VOWDEN, C. J.; TOBUTT, K. R. Three-point linkage analysis in crosses of allogamous plant species. **Genetical Research**, London, v. 72, p. 111–121, 1998.
- RITTER, E.; GEBHARDT, C.; SALAMINI, F. Estimation of recombination frequencies and construction of rflp linkage maps in plants from crosses between heterozygous parents. **Genetics**, Austin, v. 125, p. 645–654, 1990.
- RIVANO, F.; MATTOS, C. R. R.; CARDOSO, S. E. A.; MARTINEZ, M.; CEVALLOS, V.; GUEN, V. L.; GARCIA, D. Breeding *hevea brasiliensis* for yield, growth and salb resistance for high disease environments. **Industrial Crops and Products**, Washington, v. 44, p. 659–670, 2013.
- ROSA, G. J. M.; VALENTE, B. D.; CAMPOS, G. de los; WU, X.-L.; GIANOLA, D.; SILVA, M. A. Inferring causal phenotype networks using structural equation models. **Genetics Selection Evolution**, Les Ulis, v. 43, p. 436–449, 2011.
- ROY, C. B.; NAZEER, M. A.; SAHA, T. Identification of simple sequence repeats in rubber (*hevea brasiliensis*). **Current Science**, Bangalore, v. 87, p. 807–811, 2004.
- SADIQ, R. A.; SMITH, V. G. Estimation of individual tree volumes with age and diameter. **Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere**, Edmonton, v. 13, p. 32–39, 1983.
- SAHA, T.; ROY, C. B.; NAZEER, M. A. Microsatellite variability and its use in the characterization of cultivated clones of *hevea brasiliensis*. **Plant Breeding**, Berlin, v. 124, p. 86–92, 2005.
- SCHADT, E. E.; LAMB, J.; YANH, X.; ZHU, J.; EDWARDS, S.; GUHATHAKURTA, D.; SIEBERTS, S. K.; MONKS, S.; REITMAN, M.; ZHANG, C. An integrative genomics approach to infer causal associations between gene expression and disease. **Nature Reviews Genetics**, New York, v. 37, p. 710–717, 2005.
- SCHUMACHER, M. A.; BASHOR, C. J.; SONG, M. H.; OTSU, K.; ZHU, S.; PARRY, R. J.; ULLMAN, B.; BRENNAN, R. G. The structural mechanism of gtp stabilized oligomerization and catalytic activation of the *toxoplasma gondii* uracil phosphoribosyltransferase. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 99, p. 78–83, 2002.

SCHUMACKER, R.; LOMAX, R. **A beginner's guide to structural equation**. 1th edition. Mahwah: Lawrence Erlbaum, 1996. 550 p.

SCHWARZ, G. Estimating the dimension of a model. **Annals of statistics**, Hayward, v. 6, p. 461–464, 1978.

SEARLE, S.; CASELLA, G.; MCCULLOCH, C. **Variance components**. 1th edition. New York: John Wiley and Sons, 1992. 536 p.

SEARLE, S. R. **Linear models**. [S.l.]: New York, Wiley & Sons, 1971.

SHIPLEY, B. **Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference**. [S.l.]: Cambridge University Press, Cambridge, 2002.

SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, Washington, v. 57, p. 1138–1147, 2001.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **Journal of Agricultural Science**, Cambridge, v. 143, p. 449–462, 2005.

SO, Y.-S.; EDWARDS, J. Predictive ability assessment of linear mixed models in multi-environment trials in corn. **Crop Science**, Madison, v. 51, p. 542–552, 2011.

SOUZA, A. d. Biologia molecular aplicada ao melhoramento. **Recursos genéticos e melhoramento de plantas**, Rondonópolis, v.1, p. 939–966, 2001.

SOUZA, L. M.; GAZAFFI, R.; MANTELLO, C. C.; SILVA, C. C.; GARCIA, D.; GUEN, V. L.; CARDOSO, S. E. A.; GARCIA, A. A. F.; SOUZA, A. P. Qtl mapping of growth-related traits in a full-sib family of rubber tree (*hevea brasiliensis*) evaluated in a sub-tropical climate. **Plos One**, Cambridge, v. 8, p. 1–10, 2013.

SOUZA, L. M.; MANTELLO, C. C.; SANTOS, M. O.; GONCALVES, P. de S.; SOUZA, A. P. Microsatellites from rubber tree (*hevea brasiliensis*) for genetic diversity analysis and cross-amplification in six *hevea* wild species. **Conservation Genetics Resources**, Berlin, v. 1, p. 75–79, 2009.

STUBER, C. W.; EDWARDS, M. D.; WENDEL, J. F. Molecular marker-facilitated investigations of quantitative trait loci in maize .2. factors influencing yield and its component traits. **Crop Science**, Madison, v. 27, p. 639–648, 1987.

SUKHWINDER, S.; HERNANDEZ, M. V.; CROSSA, J.; SINGH, P. K.; BAINS, N. S.; SINGH, K.; SHARMA, I. Multi-trait and multi-environment qtl analyses for resistance to wheat diseases. **Plos One**, Cambridge, v. 7, p. 38008, 2012.

SWEENEY, L. **A study of current practice of corporate social responsibility (csr) and an examination of the relationship between csr and financial performance using structural equation modelling (sem)**. Unpublished doctoral thesis, Dublin Institute of Technology, Dublin, 2009.

TAIZ, L.; ZEIGER, E. **Plant physiology, 782p**. 5th edition. Sunderland: Sinauer Associates Inc., 2010. 782 p.

- TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic dna markers. **Nucleic Acids Research**, Oxford, v. 17, p. 6463–6471, 1989.
- THOMAS, D.; CONTI, D. Commentary: The concept of mendelian randomization. **International Journal Epidemiology**, Oxford, v. 33, p. 21–25, 2004.
- TINKER, N.; MATHER, D. et al. Methods for qtl analysis with progeny replicated in multiple environments. **Journal of Agricultural Genomics**, Washington, v. 1, p. 112–124, 1995.
- ULLMAN, J. B. Structural equation modeling: Reviewing the basics and moving forward. **Journal of Personality Assessment**, Hillsdale, v. 87, p. 35–50, 2006.
- VARGHESE, Y. A.; KNAAK, C.; SETHURAJ, M. R.; ECKE, W. Evaluation of random amplified polymorphic dna (rapd) markers in hevea brasiliensis. **Plant Breeding**, Berlin, v. 116, p. 47–52, 1997.
- VERARDI, C. K.; JUNIOR, E. J. S.; GONCALVES, P. d. S. Prediction of direct and indirect genetic gains and genotypic correlations in rubber tree progenies. **Pesquisa Agropecuaria Brasileira**, Campinas, v. 46, p. 1076–1081, 2011.
- VERBEKE, G.; MOLENBERGHS, G. **Linear Mixed Models for Longitudinal Data**. 1th edition. New York: Springer, 2009. 568 p.
- VERBYLA, A. P.; CULLIS, B. R. Multivariate whole genome average interval mapping: Qtl analysis for multiple traits and/or environments. **Theoretical and Applied Genetics**, Berlin, v. 125, p. 933–953, 2012.
- WELHAM, S. J.; GOGEL, B. J.; SMITH, A. B.; THOMPSON, R.; CULLIS, B. R. A comparison of analysis methods for late-stage variety evaluation trials. **Australian & New Zealand Journal of Statistics**, Oxford, v. 52, p. 125–149, 2010.
- WELLER, J. I. Maximum-likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic-markers. **Biometrics**, Washington, v. 42, p. 627–640, 1986.
- WILLIAMS, J. G. K.; KUBELIK, A. R.; LIVAK, K. J.; RAFALSKI, J. A.; TINGEY, S. V. Dna polymorphisms amplified by arbitrary primers are useful as genetic-markers. **Nucleic Acids Research**, Oxford, v. 18, p. 6531–6535, 1990.
- WOLFINGER, R. Covariance structure selection in general mixed models. **Communications in statistics-Simulation and computation**, New York, v. 22, p. 1079–1106, 1993.
- WOLFLE, L. M. The introduction of path analysis to the social sciences, and some emergent themes: An annotated bibliography. **Structural Equation Modeling**, Los Angeles, v. 10, p. 1–34, 2003.
- WU, R. L.; LIN, M. Opinion - functional mapping - how to map and study the genetic architecture of dynamic complex traits. **Nature Reviews Genetics**, New York, v. 7, p. 229–237, 2006.
- WU, R. L.; MA, C. X.; PAINTER, I.; ZENG, Z. B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. **Theoretical Population Biology**, New York, v. 61, p. 349–363, 2002.

YU, F.; WANG, B.-H.; FENG, S.-P.; WANG, J.-Y.; LI, W.-G.; WU, Y.-T. Development, characterization, and cross-species/genera transferability of ssr markers for rubber tree (*hevea brasiliensis*). **Plant Cell Reports**, Berlin, v. 30, p. 335–344, 2011.

YUAN, K. H. Fit indices versus test statistics. **Multivariate Behavioral Research**, Fort Worth, v. 40, p. 115–148, 2005.

ZABEAU, M. Genomics and post-genomics research in the european research area. **Outlook on Agriculture**, New York, v. 30, p. 97–99, 2001.

ZENG, Z. B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 90, p. 10972–10976, 1993.

ZENG, Z. B. Precision mapping of quantitative trait loci. **Genetics**, Austin, v. 136, p. 1457–1468, 1994.