

RAFAEL SIMÕES TOMAZ

**MÉTODO DE HASEMAN-ELSTON E SUAS MODIFICAÇÕES NO ESTUDO
DE GENES CONTROLADORES DE CARACTERÍSTICA QUANTITATIVA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

T591m
2012

Tomaz, Rafael Simões, 1983-

Método de Haseman - Elston e suas modificações no estudo de genes controladores de característica quantitativa / Rafael Simões Tomaz. – Viçosa, MG, 2012.
viii, 63f. : il. ; 29cm.

Inclui anexos.

Orientador: Cosme Damião Cruz.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 57-60

1. Locos de caracteres quantitativos – Métodos estatísticos.
2. Genética quantitativa. I. Universidade Federal de Viçosa.
II. Título.

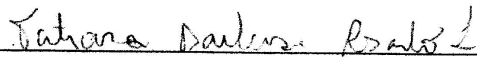
CDD 22. ed. 572.8633

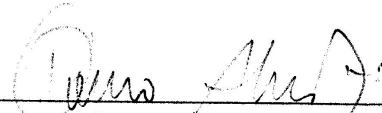
RAFAEL SIMÕES TOMAZ


Método de Haseman-Elston e suas modificações no estudo de genes controladores de característica quantitativa

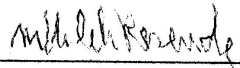
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 27 de janeiro de 2012.


Dra. Tatiana Barbosa Rosado Laviola


Prof. Dr. Dário Ahnert


Prof. Dr. Pedro C. Souza Carneiro
(Coorientador)


Prof. Dr. Marcos Deon Vilela de Resende
(Coorientador)


Prof. Dr. Cosme Damião Cruz
(Orientador)

Sempre que chamados à crítica, respeitemos o esforço nobre dos semelhantes. Para construir, são necessários amor e trabalho, estudo e competência, compreensão e serenidade, disciplina e devotamento. Para destruir, porém, basta, às vezes, uma só palavra.

(ANDRÉ LUIZ)

AGRADECIMENTOS

Deixo meus sinceros agradecimentos àqueles sem os quais o presente trabalho não seria possível:

Primeiramente, a Deus, meu guia;

Aos meus pais Gabriel Tomaz e Isabel Simões Tomaz, exemplos de perseverança e determinação;

A Universidade Federal de Viçosa, a CAPES e ao CNPq, pela oportunidade de concluir a minha formação;

Ao Prof. Cosme Damião Cruz, meu orientador, pelo seu exemplo como pessoa e pesquisador, pelo estímulo, paciência e pelas contribuições valiosíssimas sem as quais este trabalho seria impossível;

Ao Prof. Dr. Pedro Crescêncio Souza Carneiro e ao professor Marcos Deon, pelos valiosos ensinamentos e pela significativa contribuição para minha formação;

À Dra. Tatiana Barbosa Rosado Laviola, e ao professor Dário Ahnert, pela oportunidade de tê-los presente na minha banca de tese.

Aos professores de departamento de Estatística, Paulo Ceccon, Adair José Regazi, Fabiano Fonseca e Luiz Alexandre Peterneli, pelo imenso aprendizado e estímulo.

À Prof. Riva de Oliveira, da Universidade Federal de Ouro Preto, pelo estímulo e imenso aprendizado.

Aos companheiros(as) e amigos(as) do Laboratório de Bioinformática: Caio, Eliel, Lívia, Moisés, Felipe, Márcio, Danielle, Rafael Mauri, Digner, Pedro, e aos afiliados, Talles, Moisés, Carol, Gilson; e a todos que de alguma forma contribuíram, tanto para o meu crescimento, tanto como pessoa e quanto como pesquisador;

A todos os funcionários da Universidade Federal de Viçosa, em especial as funcionárias da secretaria de genética e melhoramento Rita e Edna; ao funcionário da Pró-reitoria de Pós-Graduação Gilcemir, pela imensa paciência; e os funcionários do BIOAGRO;

Finalmente, a quem estiver valorizando este trabalho, por meio de sua leitura ou utilização de qualquer forma.

BIOGRAFIA

Rafael Simões Tomaz, filho de Gabriel Tomaz e Isabel Simões Tomaz, nasceu em Cachoeiro de Itapemirim, Espírito Santo, no dia 27 de Fevereiro de 1983.

Ingressou no curso de Bioquímica da Universidade Federal de Viçosa (UFV) em abril de 2002.

Iniciou estágio no laboratório de Genética de Microrganismos em novembro de 2004, sendo contemplado com bolsa de iniciação científica no período de março de 2005 a fevereiro de 2006.

Ainda em Fevereiro de 2006, foi contemplado com bolsa de iniciação pela Sociedade de Investigações Florestais.

Graduou-se bacharel em Bioquímica em outubro de mesmo ano.

Ainda em Outubro 2006, iniciou seu Mestrado em Genética e Melhoramento na Universidade de Viçosa (UFV), sob orientação da Professora Elza Fernandes de Araújo, e co-orientação dos professores Cosme Damião Cruz e Pedro Crescêncio Souza Carneiro, submetendo-se à defesa de dissertação em julho de 2008.

No segundo semestre do mesmo ano, ingressou no programa de Doutorado em Genética e Melhoramento na mesma instituição sob orientação do professor Cosme Damião Cruz e sob orientação dos professores Pedro Crescêncio Souza Carneiro e Marcos Deon Vilela de Resende; submetendo-se à defesa de tese em janeiro de 2011.

Durante o curso de doutorado, teve ainda a oportunidade de trabalhar como professor substituto em nível de assistente, na Universidade Federal de Ouro Preto, onde lecionou disciplinas de Genética (Básica, Molecular e de Populações) em nível de graduação para estudantes dos cursos de Biologia (bacharelado e licenciatura), Medicina e Nutrição.

SUMÁRIO

RESUMO.....	vi
ABSTRACT	viii
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	2
2.1. O PROCESSO DE IDENTIFICAÇÃO DE GENES	2
2.2. ANÁLISE DE LIGAÇÃO	6
2.2.1. REGRESSÃO DE HASEMAN-ELSTON (oHE)	9
2.2.2. MODIFICAÇÕES DO MÉTODO ORIGINAL.....	17
3. MATERIAL E MÉTODOS.....	19
3.1. SIMULAÇÃO	19
3.1.1. CENÁRIOS SIMULADOS.....	19
3.1.2. SIMULAÇÃO DA CARACTERÍSTICA FENOTÍPICA.....	20
3.2. DETECÇÃO DE QTL.....	21
3.2.1. REGRESSÃO DE HASEMAN-ELSTON (oHE)	21
3.2.2. REVISTED HASEMAN-ELSTON (rHE).....	23
3.2.3. EXTREME DISCORDANT AND CONCORDANT (EDAC) STRATEGY	24
3.3. INTERPRETAÇÃO DOS RESULTADOS.....	24
4. RESULTADOS E DISCUSSÃO.....	26
4.1. AVALIAÇÃO DO PODER ESTATÍSTICO DOS MÉTODOS.....	26
4.2. ESTIMATIVA DE PARÂMETROS GENÉTICOS.....	35
4.3. IBD – IDENTICAL BY DESCENDENT	55
5. CONCLUSÕES	56
6. BIBLIOGRAFIA	57
7. ANEXOS	61

RESUMO

TOMAZ, Rafael Simões, D.Sc., Universidade Federal de Viçosa, janeiro de 2012. **Método de Haseman-Elston e suas modificações no estudo de genes controladores de característica quantitativa.** Orientador: Cosme Damião Cruz. Coorientadores: Marcos Deon Vilela de Resende e Pedro Crescêncio Souza Carneiro.

O mapeamento de QTLs permite estudar os locos quantitativos envolvidos na herança complexa, bem como determinar suas localizações cromossômicas. Para tanto, um conjunto de métodos bastante úteis para o estudo de características quantitativas são aqueles baseados na regressão de Haseman-Elston, muito empregados na genética humana. Estes constituem métodos robustos de regressão para detecção de locos ligados ou associados com a característica fenotípica de interesse. Dada a facilidade de genotipagem de marcadores SNP dentro das regiões codificadoras dos mais diversos organismos, tais metodologias tornam-se mais uma alternativa para associar polimorfismos às características fenotípicas de interesse. Tais métodos passam a ter potencial importância por constituírem uma maneira eficiente e rápida de avaliar os dados genéticos provenientes de cruzamentos exogâmicos. Faz-se necessário, entretanto, melhor entendimento das propriedades de tais metodologias, no que tange do processo de detecção, bem como seu poder. Desta forma, este trabalho se propõe a, por meio de simulação de dados, investigar as metodologias baseadas em pares de irmãos, Haseman-Elston original (oHE) com e sem seleção, e revisited Haseman-Elston (rHE) com o intuito fornecer embasamento teórico para que populações exogâmicas possam ser melhor utilizadas em estudos de identificação de genes. Para tanto, foram simulados, por meio do software GQMOL, QTLs com diferentes ações gênicas (1, 3, 5, 7, 10, 13, 16, 20, 25, 27, 30 e 43%) e características fenotípicas com diferentes herdabilidades (35, 50, 65 e 80%) sob diferentes tamanhos populacionais (25, 50, 75, 100 e 200 indivíduos) com o intuito de avaliar o poder estatístico dos métodos. A análise de dados foi realizada por meio do software livre R para Estatística Computacional. Os métodos avaliados mostraram-se pouco eficientes no estudo de características quantitativas, mas apresentaram grande potencialidade para estudo de características

oligogênicas. O método rHE apresentou grande potencial para detecção de genes de efeito maior, em populações de tamanho superior a 100 indivíduos e para características de herdabilidade superior a 50%. Todos os métodos, porém, apresentaram-se pouco eficientes na detecção de genes de baixo efeito.

ABSTRACT

TOMAZ, Rafael Simões, D.Sc., Universidade Federal de Viçosa, January, 2012. **Haseman-Elston methodology and modifications on the study genes underlining quantitative trait loci.** Adviser: Cosme Damião Cruz. Co-advisers: Marcos Deon Vilela de Resende and Pedro Crescêncio Souza Carneiro.

QTL mapping allows to study loci that play a role in complex traits, as well to determine their chromosomal locations. For this purpose, a set of methods useful for the study of quantitative traits are those based on Haseman-Elston regression, that is quite used in human genetics. These are robustness regression methods for linkage analysis that can be used in order to associate genetic markers with phenotypic trait. With SNP genetic markers, that can be genotyped within the coding regions, such methods had become an alternative to detect QTLs. Such methods have a potential as they constitute an efficient tool to of evaluating the genetic data from exogamic crosses. However, It is necessary a better understanding of the properties of such methodologies, in terms of the detection process as well as its power. Thus, this study proposes the investigation of sib-pair analysis based on Haseman-Elston regression, by means of simulated data. We evaluated the original Haseman-Elston regression (oHE), in two selective schemes, and revisited Haseman-Elston regression (rHE) in order to provide theoretical basis for using exogamic populations for studies of gene identification. Data simulation procedure was performed with GQMOL software. It was simulated QTLs with different gene actions (1, 3, 5, 7, 10, 13, 16, 20, 25, 27, 30 and 43%) and traits with different heritabilities (35, 50, 65 and 80%) under different size population (25, 50, 75, 100 and 200 sibs) in order to evaluate the statistical power of methods. Data analysis was performed using the free software R for Statistical Computing. All methods have proved ineffective in the study of quantitative traits, but had a great potential for the study of oligogenic traits. The method rHE showed high potential to detect major effect genes in populations larger than a hundred sibs in traits with heritability greater than 50%. However, all methods showed themselves ineffective to detect low effect genes.

1. INTRODUÇÃO

Por razões práticas ou por limitações biológicas, nem sempre é possível empregar para determinadas espécies, delineamentos experimentais para análise de QTLs com base no cruzamento de linhagens endogâmicas, como F2, retrocruzamentos ou RILs. Uma alternativa ao uso desses delineamentos constitui a análise de cruzamentos exogâmicos, nos quais podem ser utilizadas famílias de irmãos completos, famílias de meios-irmãos, ou ainda genealogias mais complexas. A análise e detecção de QTL, ou genes candidatos, baseados em análise de ligação, tem se mostrado uma importante ferramenta na genética humana, mais especificamente na pesquisa em epidemiologia genética, situação na qual essas técnicas são freqüentemente utilizadas. Esse tipo de cruzamento apresenta grande potencial para ser utilizados no melhoramento de plantas, principalmente pela sua aplicação em espécies alógamas.

A análise dessas populações é normalmente realizada em etapas; sendo a primeira, a construção de um mapa genético de marcadores; seguido dos procedimentos de detecção de QTL. Dentre estes, o primeiro passo constitui-se da análise de “associação” dos marcadores genéticos com a característica de interesse visando uma análise exploratória dos dados, que é normalmente realizada por meio de metodologias de marca única. Em seguida, quando de posse do mapa genético, promove-se uma análise por meio de metodologias baseadas em intervalo. Entretanto, quando a população avaliada não apresenta tamanho adequado, muitas vezes não é possível a obtenção de um mapa genético; ou quando este é obtido, ele apresenta viés muito grande, o que pode comprometer as etapas subseqüentes da pesquisa.

Neste contexto, as metodologias de análise de ligação, mais especificamente aquelas baseadas na regressão de Haseman-Elton, que são baseadas na análise de pares de irmãos, passam a ter potencial importância por constituírem uma maneira eficiente e rápida de avaliar os dados genéticos. Entretanto, faz-se necessário um melhor entendimento das propriedades de dessas metodologias, no que tange o poder e a eficiência do processo de detecção. Desta forma, este trabalho propõe a investigar as metodologias baseadas em pares de irmãos, regressão de Haseman-Elston original e

modificações, e discutir sua potencial aplicação em populações exogâmicas associadas ao melhoramento vegetal, utilizando processos de simulação de dados.

2. REVISÃO DE LITERATURA

2.1. O PROCESSO DE DETECÇÃO E IDENTIFICAÇÃO DE GENES

Para uma característica simples determinada por um único loco com dois alelos, pode-se observar a segregação Mendeliana com dois ou três fenótipos na progênie. Para características complexas, no entanto, esta relação direta entre genótipos e fenótipos não pode ser realizada. Seja em espécies vegetais ou animais, as características complexas são influenciadas por inúmeros processos bioquímicos e fisiológicos, nos quais, fatores ambientais podem influenciar em diferentes graus. Na genética humana, por exemplo, fenótipos de características comportamentais, tais como ansiedade e depressão, por exemplo, podem apresentar aproximadamente uma distribuição normal; e quando sob estudo, devem ser considerados sob um modelo complexo ou poligênico (SHALOM & DARVASI, 2002). Nessas condições, a arquitetura genética da característica deve ser convenientemente estudada sob um modelo quantitativo; e os fatores genéticos analisados são denominados “*quantitative trait loci*”, ou simplesmente QTL; termo que foi criado por GELDERMANN (1975) que se baseia no princípio da existência de locos de maior importância relacionados com a expressão de determinada característica métrica.

Os fundamentos teóricos do mapeamento de QTL foram entendidos a partir do trabalho de THODAY (1961), o qual sugeriu que se um oligogene (gene controlador de um caráter de herança complexa) estivesse ligado a um monogene (gene controlador de um caráter de herança simples), os efeitos fenotípicos do primeiro poderiam ser indiretamente estudados com base nos efeitos de um gene vizinho. O termo QTL diz respeito a uma região de DNA associada a um particular fenótipo de interesse. No entanto, mesmo uma pequena região, ou um pequeno fragmento de cromossomo, pode conter dezenas ou centenas de genes; fatores estes que são os verdadeiros responsáveis pela característica expressa. Nas páginas a seguir, sempre que se for referir ao(s) gene(s) presente em uma específica região genômica, será utilizado o termo gene(s) QTL.

O processo de detecção de QTL, bem como o seu mapeamento, tem como objetivo identificar os fatores genéticos que controlam a característica, por meio da identificação de segmentos ou regiões cromossômicas específicas e, em uma segunda etapa, os genes QTL ou os elementos regulatórios que influenciam a expressão da característica fenotípica. Isto ainda representa um dos grandes desafios da genética neste século (LUO et al., 2002; PRICE, 2006), mas tem sido possível com o desenvolvimento de novas tecnologias analíticas. Nos últimos anos, os pesquisadores têm conseguido, com sucesso, detectar e mapear genes QTL que afetam características complexas em humanos, organismos modelos, e espécies agrícolas e florestais (TANKSLEY, 2004; MISSIAGGIA et al., 2005; WERNER et al., 2005, ROSADO et al., 2010).

É de grande importância notar que os vários estágios da análise de QTL constituem um procedimento multidisciplinar, no qual diversos métodos e tecnologias são requeridos nos diferentes estágios do processo. O primeiro estágio constitui o processo de detecção do QTL, no qual se avalia a hipótese de existência de ligação entre o marcador genético, ou um conjunto de marcadores, e o loco causal da característica (SHALOM & DARVASI, 2002). No segundo estágio tem-se o mapeamento, que constitui o processo de posicionar o QTL em um cromossomo ou mapa genético. A menos que o tamanho da progênie avaliada seja grande o suficiente, este estágio proporciona uma informação limitada sobre a posição do QTL e, conseqüentemente, dos genes relativos a este (DAVARSI, 1998). A medida que se diminui o comprimento do intervalo em que se quer posicionar o QTL, aumenta-se o número de indivíduos requeridos para detectar pelo menos um evento de recombinação na região, assim como o número de marcadores genéticos necessários para detectá-lo (MACKAY et al., 2009).

Uma vez que os QTLs tenham sido identificados e mapeados, a próxima etapa constitui identificar os genes, por meio do estudo e avaliação dos genes candidatos. Entretanto, na maioria das situações, o intervalo do mapa genético no qual o QTL se posiciona abrange algumas unidades, quando não, dezenas de centiMorgans, situação na qual dezenas ou até centenas de genes são abrangidos. Em tais situações, um “mapeamento fino” deve ser procedido, o que envolve o uso de uma segunda população de mapeamento, maior que a utilizada inicialmente (MACKAY et al., 2009), e estratégias

específicas. Cerca de milhares de indivíduos são genotipados para marcadores genéticos (atualmente SNP's, "single nucleotide polymorphism" ou polimorfismos de um único nucleotídeo) que se posicionam na região do QTL. Aqueles que apresentam eventos de recombinação na região do QTL são fenotipados para a característica, permitindo posicionamento mais acurado do QTL, usualmente, frações de centiMorgan ($< 1,0$ cM), distância que abrange usualmente poucas dezenas de genes, (~1–50 genes) – (PRICE, 2006). Ainda, além de requerer gastos consideráveis no que concerne ao tamanho populacional e o processo de genotipagem, o processo de mapeamento fino apresenta graves limitações. Genes QTL de pequeno efeito limitam a habilidade do pesquisador em acessar, com acurácia, diferenças fenotípicas entre as variantes alélicas. A maioria destes não é rastreável usando o mapeamento fino (PRICE, 2006), o que acarreta que tal estratégia sejam melhor utilizável apenas na identificação de genes com moderado a grande efeito sobre a expressão da característica fenotípica. Chega-se que as diversas estratégias, bem como metodologias apresentadas na literatura científica, são mais complementares do que concorrentes, no sentido de se estudar e entender as conseqüências da variação genética, com o intuito de melhor manipulá-la.

No contexto do melhoramento genético, o leitor poderia questionar sobre a validade desta estratégia, sendo ela aplicável "apenas" a genes de efeito considerável. Deve-se ponderar ainda os tipos de populações de mapeamento necessárias ao processo, bem como o seu tamanho amostral. Ainda assim, justifica-se tal esforço, uma vez que tais genes, uma vez detectados e mapeados, poderiam ser utilizados em processos de transgenia com o intuito de tornar melhores as variedades comerciais, uma vez que esforço considerável seria requerido nessas situações quando utilizadas as estratégias de melhoramento convencional.

Exemplos de genes QTL que apresentam moderado e grande efeito são abundantes na literatura. Gene QTL *fw2.2*, que controla o tamanho e peso do fruto, em tomate, um dos primeiros QTLs de efeito maior detectado, sendo responsável por explicar 30-47% da variação fenotípica – detectado em população de 264 indivíduos derivados de retrocruzamento (TANKSLEY, 2004; ALPERT et al. apud PRICE, 2006). Gene QTL *ovate*, que controla o formato do

fruto, em tomate, responsável por explicar 47-67% da variação fenotípica (TANKSLEY, 2004; KU et al. apud PRICE, 2006). Gene QTL *cbf3*, cujo produto gênico, um fator de transcrição regulado por temperatura, atua na característica tolerância à geada, explicando 40-48% da variação na característica (VAGUJFALVI et al., 2003). Gene QTL de efeito maior, *gcp-6B1*, que controla conteúdo de proteína no grão de trigo, explicando 66% da variação fenotípica. QTL FT1, mapeado junto ao marcador *satt365* e 0,1 cM do gene efetor *FT1*, responsável por explicar 66% da variação fenotípica na característica tempo de florescimento em soja (YAMANAKA, et al., 2005). Diversos outros exemplos são encontrados para espécies tais como *Eucalyptus sp* (MISSIAGGIA et al., 2005; ROSADO et al., 2010), *Brassica oleracea* (PELEMAN et al., 2005), e o modelo *Arabidopsis thaliana* (EL-DIN EL-ASSAL et al., 2001; WERNER et al., 2005).

2.2. ANÁLISE DE LIGAÇÃO

De acordo com FERREIRA (2004), a análise de ligação constitui uma de duas estratégias complementares – sendo a outra análise proporcionada pelos métodos da genética de associação – utilizadas no processo de detecção de mapeamento de genes. De forma geral, uma análise de ligação é conduzida com o intuito de “associar” um loco que controla uma determinada característica a uma determinada região do genoma, região à qual este loco se apresenta ligado, mas que tipicamente é grande o suficiente para conter centenas de genes.

Dois métodos conjuntos de métodos são bastante utilizados na análise de ligação; aqueles baseados na regressão de Haseman-Eston, e os baseados em componentes de variância (WANG & ELSTON, 2007). Apesar das vantagens e desvantagens inerentes a cada método, ambos são bastante utilizados na genética humana, sendo amplamente referenciados na literatura. SHEN et al. (2007) fizeram uso do método de análise de ligação com o intuito de associar variações em marcadores SNP e microssatélites a características de doença coronariana e infarto do miocárdio. GRAY-MCGUIRE, et al. (2010) utilizaram este tipo de metodologia para auxiliar na associação de um

haplótipo, fator de risco familiar, na predisposição em um tipo específico de câncer colorretal que fora posicionado no cromossomo 9q22. STEIN & ELSTON (2009) discorrem sobre o uso de tais metodologias, focando a busca de genes que afetam doenças humanas. Diversos outros exemplos são citados (LARKIN et al. 2008; SCHELLING et al., 2008; SHEN et al., 2007; STEIN et al., 2007). Tais metodologias, no entanto, têm sido pouco aplicadas no estudo e identificação de genes em espécies vegetais, sejam oligogenes, ou QTLs de efeito maior.

O método baseado em componentes de variância pode ser aplicável a qualquer tipo de *pedigree* e apresenta substancialmente maior poder estatístico quando a característica aproxima-se da distribuição Gausiana. Essa dependência, no entanto, faz com o método seja falho, caso os pressupostos de normalidade sejam violados, situações freqüentes quando o tamanho amostral é reduzido, ou quando se trabalha com amostras selecionadas (FEINGOLD, 2002). Já os métodos baseados na regressão Haseman-Elston (oHE), por sua vez, compreendem metodologias simples e robustas, sendo portanto grandemente utilizadas na prática. No entanto, uma das limitações deste método é a perda de poder quando comparado com o método de componentes de variância, e a não utilização de toda informação disponível nos valores fenotípicos da característica estudada (WANG & ELSTON, 2007).

O método oHE realiza regressão entre uma medida relativa a avaliação fenotípica de dois membros (um par de irmãos, por exemplo) e uma medida de similaridade genética, a qual é usualmente descrita como o número de alelos IBD, "*identical by descent*" ou idênticos por descendência, compartilhados pelos membros, em um específico loco gênico. A lógica, subjacente ao método, é que quando o marcador estudado não está ligado ao loco que afeta a característica, o número de alelos IBD compartilhado pelos membros não depende dos valores fenotípicos associados aos mesmos. Por outro lado, quando o marcador estudado está ligado ao loco causal, espera-se uma correlação entre o número de alelos IBD compartilhados e a similaridade fenotípica. Os métodos baseados em regressão avaliam essa correlação por meio do coeficiente de regressão. Sob este ponto de vista, é crítico para o poder, e conseqüentemente, a validade do método de análise de ligação, a

maneira de calcular a medida de similaridade fenotípica, e o número de alelos IBD compartilhados (WANG & ELSTON, 2007).

Também sob este ponto de vista, é crucial o conceito de IBD. Classicamente, este constitui um coeficiente de relacionamento genético, definido como o número de alelos em comum de um dado loco em um particular par de cromossomos que dois indivíduos aparentados herdaram de um ancestral comum (AMOS et al. 1990; POWELL et al. 2010). Os vários métodos IBD derivados dependem deste conceito, embora os primeiros métodos de análise de ligação tenham sido baseados em alelos marcadores idênticos por estado, IBS ou "*identical by state*", ou seja, a similaridade entre dois indivíduos era baseada somente em sua similaridade alélica com respeito ao loco avaliado.

Nos métodos baseados em regressão, primeiramente as probabilidades IBD são calculadas e em seguida o modelo de regressão é construído e aplicado. Obviamente, o método de estimação, o que reflete em o quão boa são as estimativas das probabilidades IBD, vai influenciar o poder da análise de ligação (WANG & ELSTON, 2007). A estimação do IBD compartilhado por dois membros de um pedigree é baseado na observação da informação genotípica associada ao marcador genético.

Os primeiros métodos baseados em IBD utilizavam somente irmãos completos. Caso os genitores também fossem genotipados para um marcador codominante, ou se os genótipos destes pudessem ser deduzidos, seria possível contabilizar o número de alelos que um determinado par de irmãos compartilham. Tais métodos ainda eliminavam aqueles pares os quais não era possível, com certeza, contabilizar o número de alelos IBD; o que conduzia a viés estatístico sempre que os marcadores não eram completamente informativos. HASEMAN & ELSTON (1972) propuseram estimar a probabilidade IBD para o loco, utilizando a informação disponível do marcador genético, e dessa forma, a proporção de alelos IBD compartilhados pelos indivíduos. Para um determinado par de indivíduos, a proporção de alelos IBD, denotada por π pode assumir os valores: 0, $\frac{1}{2}$ ou 1,0; sendo determinado por $\hat{\pi}_j = \frac{1}{2}f_{j1} + f_{j2}$; em que f_{j1} e f_{j2} , são as probabilidades estimadas de o j-ésimo par de indivíduos compartilhem um ou dois alelos IBD.

Vários algoritmos, para estimação de IBD ou de probabilidades IBD, dados os mais diversos tipos de relacionamento genético, estão disponíveis na literatura (KRUGLYAK et al., 1995; LANDER & KRUGLYAK, 1995; SOBEL & LANGE, 1996; THOMPSON & HEATH, 1998).

A informatividade de um marcador para estimação do IBD compartilhado entre indivíduos está intimamente relacionada com o grau de polimorfismo deste. Para marcadores altamente polimórficos, os indivíduos fundadores de *pedigree* apresentam alelos únicos, e portanto, o número de alelos IBD compartilhado pelos indivíduos na progênie pode ser inequivocadamente determinada. No entanto, o número de alelos para um dado marcador, na prática, é limitado – isso vai depender da natureza do marcador genético utilizado – o que vai resultar que o número de alelos IBD nem sempre pode ser especificado com certeza (HASEMAN & ELSTON, 1972; WANG & ELSTON, 2007).

De acordo com CORDELL (2004), a melhor maneira de minimizar a ambigüidade na determinação da origem dos alelos transmitidos pelos indivíduos genitores ou ancestrais comuns, é utilizar um maior número de marcadores em um mapa genético densamente saturado. Com a recente disponibilização de marcadores SNP, “*single nucleotide polymorphism*” ou polimorfismo de nucleotídeo único, para as mais diversas espécies, bem como tecnologias de genotipagem automatizada, há grandes expectativas de tais métodos possam ser utilizados na busca de genes efetores para as mais diversas características.

2.2.1. REGRESSÃO DE HASEMAN-ELSTON (oHE)

A regressão de Haseman-Elston (oHE) constitui um método baseado em análise de ligação, que é comumente utilizado para mapear genes que afetam doenças complexas, devido ao fato de não requerer especificações detalhadas sobre o modo de herança da característica sob estudo. Um forma de demonstrar a existência de um loco com controle genético sobre uma característica quantitativa seria determinar uma relação de ligação entre este loco e um específico loco marcador. Este constitui o princípio do método de

oHE, um procedimento de regressão linear simples, na qual a variável dependente, o quadrado da diferença entre pares de indivíduos aparentados, é regressado em relação à proporção de alelos IBD compartilhado entre os pares de indivíduos.

Considera-se X_{1j} e X_{2j} os valores observados para o primeiro e o segundo irmãos, respectivamente, do j -ésimo par; e assume-se o seguinte modelo geral: $X_{1j} = \mu + g_{1j} + \varepsilon_{1j}$ e $X_{2j} = \mu + g_{2j} + \varepsilon_{2j}$; em que, μ corresponde a um efeito de média geral; g_{ij} , ao efeito genético associado ao indivíduo i do j -ésimo par de irmãos; e ε_{ij} , ao efeito ambiental associado ao indivíduo i do j -ésimo par de irmãos.

Seria dessa forma, conveniente assumir que somente um loco g_{ij} é responsável por determinar a característica. Supõe-se, dessa forma, que somente dois alelos, B e b , com freqüências p e q , estejam envolvidos no fenótipo. Assim, os valores genotípicos podem ser definidos como:

$$g_{ij} = a \text{ para o indivíduo } BB;$$

$$g_{ij} = d \text{ para o indivíduo } Bb;$$

$$g_{ij} = -a \text{ para o indivíduo } bb.$$

Chega-se que a variância genética (σ_g^2) é composta por um componente aditivo (σ_a^2) e um componente de dominância (σ_d^2), que em uma população sob acasalamento ao acaso, são dados por:

$$\sigma_a^2 = 2pq[a - d(p - q)]^2; \text{ e}$$

$$\sigma_d^2 = (2pqd)^2.$$

O método consiste de uma análise de ligação utilizando pares de irmãos. Adota-se o quadrado da diferença da característica fenotípica do par de irmãos como variável dependente em uma análise de regressão. A variável independente é constituída pela quantidade de alelos IBD (identical by descendent; idênticos por descendência) compartilhado pelo par de irmãos. Desta forma, a equação de regressão pode ser expressa como a seguir: $Y_j = \beta_0 + \beta_1\pi_j + \varepsilon_j$; em que Y_j é estabelecido pelo quadrado da diferença fenotípica entre os indivíduos do j -ésimo par de irmãos, e π_j corresponde ao

valor de alelos idênticos por descendência (IBD) do par de irmãos considerados. β_0 e β_1 são os parâmetros do modelo. O valor de Y_j é, dessa forma, dado por $Y_j = (X_{1j} - X_{2j})^2$; sendo que X_{1j} e X_{2j} , correspondem aos valores observados da característica para o primeiro e o segundo irmão do j -ésimo par. Chega-se, portanto, que: $Y_j = (X_{1j} - X_{2j})^2 = (g_{1j} - g_{2j} + \varepsilon_j)^2$; o que reflete o fato de que a variável dependente corresponde a uma diferença genética entre os pares de irmão; e que deve ser regredido em relação aos alelos IBD dos referidos pares de irmãos para o loco marcador sob análise. Com relação ao erro, tem-se que $\varepsilon_j = \varepsilon_{1j} - \varepsilon_{2j}$; e por conveniência, denota-se $E(\varepsilon_j^2)$ por σ_ε^2 ; sendo esta função da variância ambiental.

O valor de Y_j pode, ainda, assumir sete valores, dependendo dos genótipos dos pares de irmãos (Tabela 1). Por exemplo, para um par de irmãos com os genótipos $BB - Bb$, tem-se que $X_{1j} = \mu + a + \varepsilon_{1j}$, $X_{2j} = \mu + d + \varepsilon_{2j}$; e portanto:

$$Y_j = (X_{1j} - X_{2j})^2 = (a + \varepsilon_{1j} - d - \varepsilon_{2j})^2 = (a - d + e_j)^2$$

TABELA 1. Distribuição condicional de Y_j . HASEMAN & ELSTON (1972).

		Probabilidade condicional		
		$\pi_j = 0$	$\pi_j = 1/2$	$\pi_j = 1$
$BB - BB$	} e_j^2 {	p^4	p^3	p^2
$bb - bb$		q^4	q^3	q^2
$Bb - Bb$		$(2pq)^2$	pq	$2pq$
$BB - Bb$	$(a - d + e_j)^2$	$2p^3q$	$2p^2q$	0
$Bb - BB$	$(-a + d + e_j)^2$	$2p^3q$	$2p^2q$	0
$Bb - bb$	$(a + d + e_j)^2$	$2pq^3$	$2pq^2$	0
$bb - Bb$	$(-a - d + e_j)^2$	$2pq^3$	$2pq^2$	0
$BB - bb$	$(2a + e_j)^2$	$2p^2q^2$	0	0
$bb - BB$	$(-2a + e_j)^2$	$2p^2q^2$	0	0

Com o intuito de obter a esperança de Y_j condicional à proporção de alelos IBD, é necessário conhecer a distribuição de Y_j condicional a esta proporção. Um particular par de irmãos deve apresentar zero, um ou dois genes IBD, com probabilidades iguais a 0, $\frac{1}{2}$ ou 1. O valor de π_j é dado por: $\hat{\pi}_j = \frac{1}{2}f_{j1} + f_{j2}$; em que f_{j1} é a probabilidade de que o j-ésimo par de irmão tenha 1 alelo IBD e f_{j2} a probabilidade de que o j-ésimo par de irmão tenha 2 alelos IBD.

Quando os genótipos dos indivíduos e o dos parentais são conhecidos, $\hat{\pi}_j$ pode ser facilmente calculado para cada par de irmãos, considerando tipo de cruzamento dos quais eles são provenientes. Em uma situação genérica com múltiplos alelos, existem sete tipos de cruzamentos considerando os genótipos parentais e, de forma similar, sete combinações de genótipos considerando os pares de irmãos. Todas as possibilidades, assim como as respectivas probabilidades e os apropriados valores de $\hat{\pi}_j$ estão apresentados na Tabela 2. Os números dentro dos parênteses correspondem ao número de combinações, dado o tipo de cruzamento, passíveis de serem gerados.

O conceito de identidade por descendência (IBD, Identity by descent) é de grande importância, pois quantifica a semelhança genética entre o par de indivíduos. Um par de indivíduos aparentados possui um alelo IBD se esse alelo tem origem comum em um ancestral. Se um determinado marcador e uma característica estiverem ligados geneticamente, a proporção de alelos IBD compartilhados excede o valor esperado. Sob a hipótese de ligação, irmãos que compartilham proporção maior de alelos IBD terão fenótipos mais parecidos, enquanto que sob a hipótese nula (sem ligação) os fenótipos dos pares serão independentes do compartilhamento de alelos, no gene marcador (FEITOSA & KRIEGER, 2002). Ainda, segundo HASEMAN & ELSTON (1972), o valor de π_j nem sempre é conhecido com certeza, uma vez que para alguns pares de irmãos – dependendo da configuração dos genótipos e da quantidade de alelos que segregam no loco marcador – tal valor é ambíguo.

TABELA 2. IBD estimado ($\hat{\pi}_j$) dos pares de irmãos quando os genótipos parentais são conhecidos. HASEMAN & ELSTON (1972).

Cruzamento	Par de irmãos	Prob.	f_{j0}	f_{j1}	f_{j2}	π_j
I: $AiAi \times AiAi$	I: $AiAi - AiAi$	p_i^4	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
II: $AiAi \times AjAj$	V: $AiAj - AiAj$	$2p_i^2 p_j^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
III: $AiAi \times AiAj$	I: $AiAi - AiAi$	$p_i^3 p_j$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
	III: $AiAi - AiAj$	$2p_i^3 p_j$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
	V: $AiAj - AiAj$	$p_i^3 p_j$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
IV: $AiAj \times AjAk$	V: (2)	$p_i^2 p_j p_k$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
	VI: $AiAj - AiAk$	$2p_i^2 p_j p_k$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
V: $AiAj \times AiAj$	I: (2)	$p_i^2 p_j^2 / 4$	0	0	1	1
	II: $AiAi - AjAj$	$p_i^2 p_j^2 / 2$	1	0	0	0
	III: (2)	$p_i^2 p_j^2$	0	1	0	$\frac{1}{2}$
	V: $AiAj - AiAj$	$p_i^2 p_j^2$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$
VI: $AiAj \times AiAk$	I: $AiAi - AiAi$	$p_i^2 p_j p_k / 2$	0	0	1	1
	III: (2)	$p_i^2 p_j p_k$	0	1	0	$\frac{1}{2}$
	IV: $AiAi - AjAk$	$p_i^2 p_j p_k$	1	0	0	0
	V: (3)	$p_i^2 p_j p_k / 2$	0	0	1	1
	VI: $AiAj - AiAk$	$p_i^2 p_j p_k$	1	0	0	0
	VI: $AiAj - AjAk$	$\left. \begin{array}{l} \\ \end{array} \right\} p_i^2 p_j p_k$	0	1	0	$\frac{1}{2}$
	VI: $AiAk - AjAk$		0	1	0	$\frac{1}{2}$
VII: $AiAj \times AkAl$	V: (4)	$p_i p_j p_k p_l / 2$	0	0	1	1
	VI: (4)	$p_i p_j p_k p_l$	0	1	0	$\frac{1}{2}$
	VII: (4)	$p_i p_j p_k p_l$	1	0	0	0

Em geral, a frequência de combinações informativas, para os pares de irmãos, cresce substancialmente com o aumento do número de alelos no marcador sob análise. Com dois alelos por loco, poucas combinações são informativas. Entretanto, para locos marcadores em que pelo menos cinco alelos podem ser observados, a grande maioria de combinações são completamente informativas. Mesmo quando os marcadores não se mostram

informativos ou em situações nas quais os genótipos dos parentais são desconhecidos, o valor de π_j pode ser estimado (Tabela 3).

TABELA 3. IBD estimado ($\hat{\pi}_j$) dos pares de irmãos, quando os genótipos parentais são desconhecidos – caso especial de ausência de dominância. HASEMAN & ELSTON (1972).

Par de irmãos	Prob.	f_{j0}	f_{j1}	f_{j2}	π_j
I: $AiAi - AiAi$	$p_i^2(1 + p_i)^2 / 4$	$\frac{p_i^2}{(1 + p_i)^2}$	$\frac{2p_i}{(1 + p_i)^2}$	$\frac{1}{(1 + p_i)^2}$	$\frac{1}{(1 + p_i)}$
II: $AiAi - AjAj$	$p_i^2 p_j^2 / 2$	1	0	0	0
III: $AiAi - AiAj$	$p_i^2 p_j(1 + p_i)$	$\frac{1}{1 + p_i}$	$\frac{1}{1 + p_i}$	0	$\frac{1}{2(1 + p_i)}$
IV: $AiAi - AjAk$	$p_i^2 p_j p_k$	1	0	0	0
V: $AiAj - AiAj$	$\frac{1}{2} p_i p_j \phi$	$\frac{2p_i p_j}{\phi}$	$\frac{(p_i + p_j)}{\phi}$	$1/\phi$	$\frac{(2 + p_i + p_j)}{2\phi}$
VI: $AiAj - AiAk$	$p_i p_j p_k(1 + 2p_i)$	$\frac{2p_i}{1 + 2p_i}$	$\frac{1}{1 + 2p_i}$	0	$\frac{1}{2(1 + 2p_i)}$
VII: $AiAj \times AkAl$	$2p_i p_j p_k p_l$	1	0	0	0

$\phi = (1 + p_i + p_j + 2p_i p_j)$

Os valores da Tabela 1 podem ser utilizados para calcular a esperança de Y_j condicional aos valores de π_j . Dessa forma, tem-se:

$$\begin{aligned}
 E(Y_j / \pi_j = 1) &= E\{e_j^2 [p^2 + 2pq + q^2]\} = E(e_j^2) = \sigma_e^2; \\
 E(Y_j / \pi_j = 1/2) &= E\{e_j^2 [p^3 + pq + q^3] + [(a - d + e_j)^2 + (-a + d + e_j)^2] p^2 q + \\
 &\quad + [a + d + e_j)^2 + (-a - d + e_j)^2] pq^2\} \\
 &= \sigma_e^2 + (a^2 + d^2)[2p^2 q + 2pq^2] + 4ad[pq^2 - p^2 q] \\
 &= \sigma_e^2 + 2pq[a^2 + d^2 - 2ad(p - q)] \\
 &= \sigma_e^2 + 2pq[a - (p - q)d]^2 + 2pqd^2[1 - (p - q)^2] \\
 &= \sigma_e^2 + \sigma_a^2 + 2pqd^2[4pq] \\
 &= \sigma_e^2 + \sigma_a^2 + 2(2pqd)^2 \\
 &= \sigma_e^2 + \sigma_a^2 + 2\sigma_d^2; \text{ e de forma similar;}
 \end{aligned}$$

$$E(Y_j / \pi_j = 0) = \sigma_e^2 + 2\sigma_a^2 + 2\sigma_d^2.$$

Mais uma vez, ao se considerar o caso especial de ausência de dominância ($d = 0$), pode-se escrever $E(Y_j / \pi_j) = (\sigma_e^2 + 2\sigma_g^2) - 2\sigma_g^2\pi_j$, para $\pi_j = 0, 1/2$ ou 1 . Isto pode ainda ser escrito por meio do modelo de regressão: $E(Y_j / \pi_j) = \alpha + \beta\pi_j$.

Nessas condições, se o loco controlador da característica quantitativa – loco QTL – se posiciona sobre o loco marcador, então $\hat{\alpha}$ passa ser um estimador de $-2\sigma_g^2 + \sigma_e^2$, enquanto que $-\hat{\beta}/2$ passa a ser um estimador de quadrados mínimos não viesado de σ_g^2 . HASEMAN & ELSTON (1972) mostram ainda que mesmo quando existe efeito de dominância, esta relação permanece válida.

Ainda, se o QTL se posiciona a uma distância do loco marcador, com uma frequência de recombinação θ , então β torna-se: $\beta = -2(1-2\theta)^2\sigma_g^2$. Ao se considerar mais uma vez o caso especial de ausência de dominância e completa informação sobre os genitores, temos que $\beta = -2(1-2\theta)^2\sigma_a^2$, em que σ_a^2 corresponde à variância genética aditiva.

A hipótese de nulidade a ser testada é a de não existência de ligação genética entre o loco marcador e QTL, podendo ser testada por meio de um teste t unilateral, realizado sobre o coeficiente da regressão, assumindo que os resíduos normalmente distribuídos. WAN et al. (1997) propuseram a execução de um procedimento permutação, por meio da manutenção dos valores de π_j na ordem original, e permutando aleatoriamente os valores de Y_j . Para um “grande” número de pares de irmãos, os autores demonstraram que a variância de β da regressão permutada e a variância estimada por quadrados mínimos sob hipótese de nulidade, se equivalem. Dessa forma, um teste t convencional se aproxima muito do resultado proporcionado pelo teste de permutação. Alternativamente, o coeficiente de regressão pode ser avaliado por meio da estatística t , elevado ao quadrado, testados por meio de valores de χ^2 , assumindo uma distribuição normal bivariada (FULKER & CHERNY, 1996).

Segundo os autores, para elevados tamanhos amostrais, a distribuição de t^2 corresponde à distribuição de χ^2 com um grau de liberdade.

É importante notar, ainda, que o procedimento só permite realizar a detecção da ligação entre o marcador e o loco QTL, não sendo possível estimar θ , meio da análise de regressão. O valor de σ_a^2 fica confundido com a fração de recombinação (θ) e, dessa forma, a ligação fatorial pode ser detectada, mas não estimada. Dessa forma a análise deve ser realizada pressupondo a completa ligação entre o marcador e o loco da característica quantitativa, embora, uma vez detectada a significância da estimativa de β , não podemos afirmar que se trata de um loco QTL de grande efeito, longe do marcador, ou de um QTL de pequeno efeito próximo à marca.

A regressão de Haseman- Elston constitui uma metodologia robusta de regressão utilizando pares de irmãos, com o propósito de detectar ligação e estimar a variância do QTL. Apesar deste método de pares de irmãos ser adequado quanto a diversos tipos de distribuição de dados e ser independente do modelo genético do QTL, ele apresenta o problema – assim como qualquer metodologia de marca simples – de não discernir entre o efeito do QTL e a sua distância com marcador avaliado. Tal fato implica que apenas a ligação entre o QTL e o marcador é detectável, não sendo possível estimar se é devido a um QTL de grande efeito localizado longe do marcador, ou devido a um QTL de pequeno efeito, próximo ao loco marcador (MARTINEZ & VUKASINOVIC, 2000).

Talvez a maior limitação deste método se dê ao fato de ele não ser aplicado a qualquer tipo de dados; a qualquer pedigree; e apresentar um menor poder estatístico quando comparado com os métodos baseados em componentes de variância. Com o intuito de superar essas limitações, diversos pesquisadores têm estendido este método a diversas situações. Os métodos derivados permitem com que a informação fenotípica seja melhor aproveitada, e que a metodologia possa ser estendida a quaisquer tipos de indivíduos, independente do parentesco. Será exposto a seguir algumas das modificações do método oHE, apresentadas na literatura.

2.2.2. MODIFICAÇÕES DO MÉTODO ORIGINAL

O método oHE, assim como suas várias extensões, tem sido largamente utilizados em estudos de ligação em características complexas. Devido ao fato de o método original ignorar a informação disponível na soma da característica para dado par de irmãos, ELSTON et al. (2000), desenvolveram uma extensão da metodologia original, a qual eles denominaram Haseman-Elston revisado, revisited Haseman-Elston (rHE), e que apresenta um apreciável aumento de poder estatístico ao método original (oHE), sendo similar, em termos de poder, ao método da máxima verossimilhança. Em tal método, a variável dependente é assumida como o produto cruzado dos valores fenotípicos associados ao j-ésimo par de irmãos, corrigido pela média populacional: $Y_j = (x_{1j} - \mu)(x_{2j} - \mu)$.

A idéia básica dos vários métodos baseados, assim como o oHE, é detectar a associação entre a similaridade fenotípica de dois indivíduos aparentados, e a quantidade de alelos IBD compartilhados. Devido ao fato de a similaridade fenotípica também poder ser resultante de efeitos poligênicos ou de efeitos ambientais comuns, do que simplesmente os alelos compartilhados para dado loco, essa covariância proporcionada por esses casos deveria ser levada em conta na análise de regressão. Em teoria, isto poderia ser alcançado simplesmente modificando a metodologia rHE. Além da estimativa dos elementos de β ser, em média, 50% maior em magnitude do que o modelo do método original, os autores demonstram que neste método, a esperança de Y_j corresponde à covariância entre os pares de irmãos, que desta forma, pode ser modelada diretamente (ELSTON et al., 2000). Na prática, um valor médio associado aos valores fenotípicos deve ser estimado; e diferentes estimativas poderiam ser utilizadas. Identificando uma perda de poder, a qual pode ser resultante do uso da média populacional, no modelo que considera o produto cruzado corrigido, quando da existência correlações em específicos pares de indivíduos, ou quando efeitos ambientais ou poligênicos estão presentes, WANG et al. (2001) propuseram utilizar um valor médio específico associado ao par de irmãos (smHE – sibship sample mean Haseman-Elston). Estes autores demonstraram que esta metodologia detém um ganho significativo de

poder estatístico em tais situações, ou quando o número de pares de irmãos é maior. Para o caso de uma única família, este método equivale ao método original.

Diversas outras alterações no método original foram propostos, com o intuito de modelar e aproveitar a covariância existente entre os pares de irmãos, bem como, a covariância existente entre os elementos de Y_j , como aqueles, por exemplo, que compartilham um indivíduo do par. Tais proposições não serão avaliadas nem discutidas neste trabalho.

3. MATERIAL E MÉTODOS

3.1. SIMULAÇÃO

Foi simulado um genoma hipotético utilizando o módulo de simulação do aplicativo computacional GQMOL (CRUZ, 2008). Foram simulados genomas parentais e amostras populacionais de famílias de irmãos completos.

Foram tomadas como referência duas espécies diplóides, a primeira com $2n = 2x = 20$ (Cenário 1), composto por dez (10) grupos de ligação; e a segunda, $2n = 2x = 8$ (Cenário 2), composto por quatro (04) grupos de ligação. Cada grupo de ligação possui tamanho de 100 cM, sobre o qual se apresentam dispostos 21 marcadores codominantes equidistantes.

3.1.1 CENÁRIOS SIMULADOS

O Cenário 1 foi estabelecido com o intuito de permitir avaliar a eficiência do processo de detecção de QTL para locos gênicos cujos efeitos variam de 1 a 25% da característica. Foram considerados nove (09) QTLs independentes, com efeitos gênicos de respectivamente 1, 3, 5, 7, 10, 13, 16, 20, 25%, posicionados, cada qual, no meio (50 cM das extremidades) dos respectivos grupos de ligação simulados.

O esquema do grupo de ligação de cada genitor está representado na Figura 1. Neste esquema, demonstra-se que o genitor 1 sempre possui o genótipo 12 (A_1A_2), enquanto que o genitor 2 possui o genótipo do 34 (A_3A_4), de maneira que a descendência produzida por este cruzamento seja completamente informativa – configuração de cruzamento número VII ($A_iA_j \times A_kA_l$) segundo HASEMAN & ELSTON (1972). A saturação média de 5 cM por grupo de ligação simulado também pode ser observada.

O Cenário 2 foi estabelecido com o intuito de permitir avaliar a eficiência do processo de detecção de QTL para locos gênicos com efeitos superiores a 25%. Foram considerados três (03) QTLs independentes, com efeitos gênicos de 27, 30 e 43%, posicionados, cada qual, no meio (50 cM das extremidades) dos respectivos grupos de ligação simulados. De forma semelhante ao cenário 1, do grupo de ligação de cada genitor está representado na Figura 1.

Federal de Viçosa e disponível em <http://www.ufv.br/dbg>; segundo o seguinte modelo: $Y_i = \mu + \sum_{t=1}^k a_t x_{t(i)} + e_i$; em que Y_i corresponde ao valor fenotípico do i-ésimo indivíduo; μ corresponde a média geral; k corresponde ao número de QTLs controlando o carácter; a_t corresponde ao efeito aditivo do t-ésimo QTL; $x_{t(i)}$ corresponde a uma variável codificadora do genótipo no t-ésimo QTL; e e_i , ao erro aleatório.

3.2. DETECÇÃO DE QTL

A associação entre marcadores individuais e a expressão da característica foram avaliadas por meio das metodologias, baseadas em marca única, de regressão proposta por Haseman-Elston (oHE – original HE), e o método revisited Haseman-Elston (rHE). Foi ainda utilizado o método original (oHE) considerando o delineamento EDAC (Extreme Discordant and Concordant strategy).

Por se tratar de um procedimento de teste múltiplos foi utilizada proteção de Bonferroni no estabelecimento de nível de significância para teste individual a partir de um nível de significância global de 10% ($\alpha_T = 0,10$). Os níveis individuais foram, portanto, de $2,38 \times 10^{-4}$ e $5,95 \times 10^{-4}$ para os cenários 1 e 2 respectivamente.

As metodologias possibilitaram testar a hipótese de herança oligogênica e/ou quantitativa, simulada por meio do aplicativo computacional GQMOL (CRUZ, 2008) desenvolvido na Universidade Federal de Viçosa e disponível em <http://www.ufv.br/dbg>.

3.2.1. REGRESSÃO DE HASEMAN-ELSTON (oHE)

Esta estratégia, proposta por HASEMAN & ELSTON (1972) baseia-se em uma regressão na qual a variável dependente (Y_j) é estabelecida pelo quadrado da diferença fenotípica entre os indivíduos do j-ésimo par de irmãos e

a variável independente que é dada pelo valor de alelos idênticos por descendência (IBD) do par de irmãos considerados.

O modelo adotado foi: $Y_j = \beta_0 + \beta_1\pi_j + \varepsilon_j$, em que;

$Y_j = (X_{1j} - X_{2j})^2$, sendo X_{1j} e X_{2j} os valores observados da característica para o primeiro e o segundo irmão do j-ésimo par;

$\pi_j = \frac{1}{2}f_{j1} + f_{j2}$, sendo f_{j1} a probabilidade de que o j-ésimo par de irmão tenha 1 alelo IBD e f_{j2} a probabilidade de que o j-ésimo par de irmão tenha 2 alelos IBD;

β_0 e β_1 são os parâmetros do modelo.

Como X_{1j} e X_{2j} são os valores observados da característica para o primeiro e o segundo irmão, respectivamente, no j-ésimo par, foi assumido o seguinte modelo geral:

$$X_{1j} = \mu + g_{1j} + e_{1j}$$

e

$$X_{2j} = \mu + g_{2j} + e_{2j}$$

em que:

μ : média geral;

g_{ij} : efeito genético associado ao indivíduo i do j-ésimo par de irmãos;

e_{ij} : efeito ambiental sobre o indivíduo i do j-ésimo par de irmãos.

Assumindo que existe ligação completa entre o marcador e o QTL, que $\sigma_D^2 = 0$, e reunindo as informações das diferenças quadráticas (Y_j) e as frequências esperadas para cada par de irmãos condicionais à π_j , pode-se calcular a esperança $E(Y_j | \pi_j)$, a qual é dada por:

$$E(Y_j | \pi_j) = (\sigma_e^2 + 2\sigma_a^2) - 2\sigma_a^2\pi_j$$

e que pode ser reescrito pelo modelo:

$$E(Y_j | \pi_j) = \beta_0 + \beta_1\pi_j, \text{ em que:}$$

$$\beta_0 = \sigma_e^2 + 2\sigma_a^2$$

$$\beta_1 = -2\sigma_a^2$$

Conhecendo-se π_j é possível ajustar uma equação linear por regressão a partir do modelo $E(Y_j | \pi_j) = \beta_0 + \beta_1 \pi_j$ em que $-\frac{1}{2}\beta_1$ é um estimador não viesado de σ_a^2 .

Quando o QTL e o marcador não estiverem completamente ligados, a proporção esperada de IBD do QTL deve ser estimada indiretamente pelo IBD do marcador, cuja informação é disponibilizada nas análises moleculares. Desta forma, ao se substituir π_j por $\hat{\pi}_j$ têm-se o seguinte modelo:

$$E(Y_j / \hat{\pi}_j) = \beta_0 + \beta_1 \hat{\pi}_j$$

Considerando-se o caso especial de ausência de dominância e completa informação sobre os genitores, tem-se que: $\beta_1 = -2(1-2r)^2 \sigma_a^2$; em que r é a porcentagem de recombinação entre os locos do marcador e o loco controlador da característica avaliada.

Em nível da família, a herdabilidade do QTL pode ser estimada por meio da expressão: $\hat{h}_{ad}^2 = -\frac{\hat{\beta}_1}{4\hat{\sigma}_F^2}$; em que $\hat{\sigma}_F^2$ é a estimativa da variância fenotípica da característica.

3.2.2. Revisited HASEMAN-ELSTON (rHE) – (ELSTON et al., 2000)

Foi adotado o modelo: $Y_j = \beta_0 + \beta_1 \pi_j + \varepsilon_j$, em que; $Y_j = (X_{1j} - \mu)(X_{2j} - \mu)$, sendo X_{1j} e X_{2j} os valores observados da característica para o primeiro e o segundo irmão do j-ésimo par; e μ a média amostral do par de irmãos considerados. Ainda, $\pi_j = \frac{1}{2}f_{j1} + f_{j2}$, sendo f_{j1} a probabilidade de que o j-ésimo par de irmãos tenha 1 alelo IBD e f_{j2} a probabilidade de que o j-ésimo par de irmão tenha 2 alelos IBD; β_0 e β_1 são os parâmetros do modelo.

3.2.3. EXTREME DISCORDANT AND CONCORDANT (EDAC) STRATEGY.

A seleção ótima de amostras a serem avaliadas atenta para solucionar um dos grandes problemas relativos ao processo de detecção de QTL, que é a baixo poder de detecção (PURCELL et al., 2001) associado a QTLs de pequeno efeito. Tal abordagem foi utilizada quando os custos associados ao processo de genotipagem eram suficientemente elevados, o que inviabilizava os custos do processo de detecção, fato que atualmente não constitui um grande limitante.

Tal estratégia constitui um tipo de delineamento para análise de pares de irmãos, e pode ter sido utilizada com o intuito de proporcionar um acréscimo no poder de detecção por meio da seleção dos indivíduos mais informativos.

Estudados, desta forma, as implicações do uso do delineamento EDAC, estratégia na qual são selecionados para análise os indivíduos discordantes e concordantes extremos; no qual são selecionados e avaliados conjuntamente os indivíduos cujos valores fenotípicos apresentam-se nos extremos da distribuição.

O método oHE foi procedido sobre os cenários simulados, considerando:

- 30% dos indivíduos; EDAC T15B50 - 15% maiores e 15% menores valores fenotípicos;
- 40% dos indivíduos; EDAC T20B20 - 20% maiores e 20% menores valores fenotípicos.

Foram consideradas as populações de 50, 75, 100 e 200 indivíduos para análise.

3.3. INTERPRETAÇÃO DOS RESULTADOS

A análise das populações simuladas foi realizada por meio do software livre R para Estatística Computacional (R Development Core Team, 2010), versão 2.12.1, disponível em <http://www.r-project.org>. Foram desenvolvidos

scripts – disponíveis em ANEXOS – para execução das metodologias acima descritas.

Ao avaliar os resultados das análises de ligação proporcionadas pelo método oHe e extensões, deve-se considerar dois importantes fatores. O primeiro diz respeito ao p-valor das estimativas dos coeficientes de regressão; o qual é equivalente à quantidade de evidência de ligação genética. O p-valor associado à estimativa é derivado de uma distribuição t de Student, e representa a probabilidade da hipótese de nulidade $\beta_1 = 0$, que considera a ausência de ligação entre o marcador e o loco QTL. Ainda, a magnitude do coeficiente β é capaz de indicar a extensão com a o marcador está ligado ao QTL que afeta a característica, uma vez que tal parâmetro é função da variância genética; variância devido a efeitos aditivos e de dominância do QTL.

A hipótese de nulidade a ser testada é a de não existência de ligação genética entre o loco marcador e QTL, podendo ser realizada por meio de um teste t unilateral, realizado sobre o coeficiente da regressão, assumindo resíduos normalmente distribuídos. Ainda, o poder estatístico do teste de inferência, representado por $(1-\beta)$. Trata-se da probabilidade de se rejeitar corretamente a hipótese de nulidade quando esta deve ser rejeitada, ou seja, está relacionada com a capacidade do teste estatístico em detectar diferenças. (FERREIRA, 2004). Dessa forma, o poder do teste está definido como a probabilidade do erro Tipo II. Para análise, foi utilizada a função *power.t.test* do pacote *Stats* do software livre R para Estatística Computacional (R Development Core Team, 2010), avaliando os marcadores completamente ligados aos QTLs.

4. RESULTADOS E DISCUSSÃO

4.1. AVALIAÇÃO DO PODER ESTATÍSTICO DOS MÉTODOS

O poder estatístico ou potência de um teste de inferência diz respeito à probabilidade de se rejeitar corretamente a hipótese de nulidade quando esta deve ser rejeitada – evitando-se um erro tipo II –; estando desta forma relacionado com a capacidade do teste estatístico em detectar diferenças. Dessa forma, é o poder estatístico que demonstra a probabilidade de encontrar a significância quando esta realmente existe (FERREIRA, 2004).

É geralmente aceito que a potência do teste deve ser de 80% ou superior, ou seja, uma probabilidade de 80% ou mais de se encontrar uma diferença estatisticamente significativa quando esta realmente existe. Ainda assim, tal parâmetro não é utilizado como referência de controle do teste, uma vez que este deve ser exercido pelo valor de alfa, o nível de significância. A explicação disto é simples, dado que cometer um erro Tipo I é muito mais grave do que cometer um erro Tipo II na análise. Ao se cometer um erro Tipo I, se aceita uma hipótese falsa, sendo a pesquisa erroneamente finalizada. No entanto, quando se comete um erro Tipo II, se rejeita uma hipótese verdadeira, mas a pesquisa continua para buscar uma efetiva hipótese verdadeira (FERREIRA, 2004).

Ainda assim, deter o conhecimento a respeito do potência do teste estatístico é de grande importância para o planejamento do tamanho da amostral e para interpretação dos resultados de análises estatísticas inferenciais.

Com o intuito de avaliar o poder de detecção de QTL do método oHE e metodologias derivadas, e ainda, investigar a extensão da influência de fatores tais como: tamanho populacional, efeito de QTL, herdabilidade da característica e genotipagem seletiva; foram simuladas populações derivadas de genitores completamente informativos, nos quais QTLs com diferentes ações gênicas atuam na determinação da característica. É citado na literatura que vários fatores são capazes de afetar o poder do teste inferência: o tipo de população, o tamanho amostral (número de indivíduos e/ou famílias), o tipo e a distribuição da característica fenotípica, erros de genotipagem, a má especificação de

relacionamento genético, dentre outros (PURCELL et al., 2001; FEINGOLD, 2002; FERREIRA, 2004). Ainda, com relação ao controle genético da característica sob estudo, são capazes de afetar o poder estatístico do teste: o número de QTLs que afetam a característica, o efeito dos QTLs, a interação entre QTLs, epistáticos ou não, e a interação QTL e ambiente, dentre outros (FEINGOLD, 2002).

Nas Tabelas 4 a 7 estão apresentadas as estimativas de poder estatístico associado aos marcadores completamente ligados aos QTLs, para os métodos oHE, rHE, oHE com seleção: EDAC T15B15 e EDAC T20T20. Estão apresentadas as estimativas para os diferentes tamanhos populacionais, e os diferentes valores de herdabilidade. As áreas marcadas de cinza nas tabelas correspondem às estimativas superiores a 80%.

Verifica-se, por meio dos dados apresentados na Tabela 4, que a metodologia oHE apresenta dificuldade em detectar QTLs de pequeno ($< 10\%$) e de moderado efeito ($10 - 25\%$) para características de baixa herdabilidade (característica simulada de herdabilidade de 35%). A exceção ocorre quando o tamanho da população é suficientemente grande (≥ 200 indivíduos), o que permite que sejam detectados genes/QTLs com ação gênica superior a 20% . Oligogenes com efeitos iguais ou superiores a 30% podem ser detectados em populações de tamanho superior a 75 indivíduos. Populações com tais tamanhos amostrais podem ser, portanto, úteis em estudos preliminares para identificação de locos de características tais como: floração precoce, resistência a patógenos e/ou condições ambientais; que muitas vezes são determinadas por genes QTL de grande efeito.

Ao se considerar valores de herdabilidade de 50% e de 65% , passam a ser identificáveis QTLs de ação gênica superior a 10% (quando avaliada população de 200 indivíduos), e superior a 25% em população de 100 indivíduos. A utilização de populações menores não condiciona ganhos reais na detecção de genes. O mesmo se aplica à característica de herdabilidade simulada de 80% . Nesta última condição, genes QTL de ação gênica superior a 13% passam a ser detectáveis em populações de 100 indivíduos.

Os método oHE mostrou-se incapaz de detectar QTLs de pequeno efeito (inferior a 10%) ou mesmo genes/QTLs de efeito moderado com ação gênica inferior a 20% em características de baixa herdabilidade. Isto constitui

um limitante ao estudo de características quantitativas, situação para a qual o métodos de regressão de Haseman-Elston foi inicialmente proposto. Verifica-se, no entanto, que a metodologias se presta bem ao estudo de características oligogênicas, caracterizadas por apresentarem alta herdabilidade, e poucos genes QTL atuando no controle genético da característica. Mesmo nestas condições, recomenda-se sua aplicação a populações de tamanho considerável (iguais ou superiores a 200 indivíduos).

O método rHE (Tabela 5) foi o que apresentou as maiores estimativas de poder estatístico. Na característica de herdabilidade simulada de 35%, o método permitiu a detecção de locos QTLs de efeito superiores a 13% em população de 200 indivíduos. Para a característica de herdabilidade simulada de 50% ou superior, o método foi capaz de detectar locos QTLs de efeito superior a 7%, em população de 200 indivíduos. Para características oligogênicas, o método rHE detém as mesmas características do oHE, permitindo, a detecção de genes/QTLs em populações de tamanho 25 e 50 indivíduos. Para herdabilidade de 50% ou superior, genes QTL de efeito $\geq 27\%$ são detectáveis em populações de 50 indivíduos; e para herdabilidade de 80% ou superior, genes QTL de efeito $\geq 30\%$ são detectáveis.

Os resultados para as metodologias EDAC T15B15 e EDAC T20B20 estão apresentados nas Tabelas 7 e 8. A seleção de amostras foi originalmente proposta com o intuito de resolver dois problemas práticos: o custo de genotipagem, que nos primeiros cinco anos do século XXI era demasiado alto, e o baixo poder de detecção de QTLs, principalmente os de pequeno efeito (PURCELL et al. 2001). De acordo com PURCELL (2001), utilizando um nível de significância para definir indivíduos fenotipicamente extremos, três tipos de pares de irmãos podem ser selecionados com o intuito de fornecer maior informatividade com respeito à ligação gênica: indivíduos concordantes com alto valor fenotípico, indivíduos concordantes com baixo valor fenotípico, e pares discordantes. Uma estratégia em particular, a EDAC – *extreme discordant and concordant strategy* – propõe selecionar pares de indivíduos mais discordantes e concordantes, sendo que o nível de significância para a partes selecionadas pode ser ajustada de acordo com o conhecimento do pesquisador sobre o modelo genético da característica. Caso haja o conhecimento, por exemplo, de que um alelo raro, cuja frequência se deseja

umentar na população, o nível de significância pode ser ajustado de forma a selecionar mais indivíduos concordantes com alto valor, e que possuam tal alelo.

Sobre os dados simulados, foi procedida a estratégia de seleção EDAC, baseada na seleção dos valores fenotípicos, realização do método de regressão oHE, e análise do poder de detecção dos QTLs. Na Tabela 6 e na Tabela 7 estão apresentados os resultados obtidos quando se procede as estratégias EDAC T15B15 – na qual se selecionam os 15% indivíduos com maiores valores (*top* 15%) e os 15% menores valores (*bottom* 15%) – e EDAC T20B20 – *top* 20% *bottom* 20% – respectivamente. Quando comparado com o método original, em características de menores herdabilidades, foi constatada pequena perda de poder estatístico do método T15B15, sem grandes comprometimentos dos efeitos gênicos detectáveis. Por exemplo, para a característica de herdabilidade de 35%, e população de 200 indivíduos, o poder de detecção para genes QTL com 20 e 25% de ação gênica apresentou uma redução < 1% e 4,19% para o método EDAC T15B15. Tal fato à primeira vista inesperado, justifica-se devido ao fator baixa herdabilidade, o que conduz ao fenótipo deixar de ser um bom preditor do genótipo. Mesmo assim, neste caso específico, a estratégia de seleção permitiria uma redução de 70% no número de indivíduos genotipados em ambos os casos. No entanto, para grandes valores de herdabilidade, os métodos EDAC T15B15 e T20B20 se equiparam ao método sem seleção, e por vezes superam. O método EDAC T20B20, por exemplo, para população de 100 indivíduos e herdabilidade de 80%, o poder de detecção de genes/QTLs com ação gênica de 20 e de 25% apresenta aumento de 1,10% e 3,19% sobre o método original; com a vantagem de proporcionar redução de indivíduos genotipados em 60%.

TABELA 4. Poder do método oHE em detectar QTLs de diversos efeitos, sob diferentes tamanhos populacionais. A área marcada corresponde às condições nas quais o poder de detecção excede 80%. (A) herdabilidade de 35%; (B) herdabilidade de 50%; (C) herdabilidade de 65%; (D) herdabilidade de 80%.

(A)						(B)					
Efeito do QTL	25	50	75	100	200	Efeito do QTL	25	50	75	100	200
1	0,04	0,06	0,60	0,06	0,06	1	0,10	0,10	0,12	0,12	0,10
3	0,04	0,08	0,06	0,22	0,08	3	0,08	0,10	0,10	0,12	0,22
5	0,06	0,10	0,14	0,16	0,10	5	0,14	0,20	0,20	0,10	0,36
7	0,06	0,14	0,24	0,18	0,32	7	0,12	0,16	0,34	0,26	0,70
10	0,16	0,18	0,24	0,20	0,52	10	0,14	0,22	0,28	0,40	0,80
13	0,14	0,24	0,30	0,32	0,74	13	0,18	0,20	0,44	0,44	0,86
16	0,16	0,36	0,28	0,60	0,78	16	0,18	0,40	0,50	0,54	0,94
20	0,16	0,30	0,50	0,58	0,84	20	0,32	0,40	0,82	0,58	0,90
25	0,18	0,40	0,66	0,58	1,00	25	0,30	0,52	0,80	0,90	1,00
27	0,20	0,36	0,72	0,76	1,00	27	0,34	0,56	0,70	0,86	1,00
30	0,24	0,44	0,86	0,86	1,00	30	0,40	0,76	0,90	0,94	1,00
43	0,30	0,66	0,90	0,88	1,00	43	0,56	0,84	1,00	1,00	1,00

(C)						(D)					
Efeito do QTL	25	50	75	100	200	Efeito do QTL	25	50	75	100	200
1	0,08	0,06	0,10	0,12	0,06	1	0,10	0,10	0,10	0,08	0,12
3	0,10	0,08	0,10	0,14	0,22	3	0,06	0,12	0,12	0,12	0,34
5	0,14	0,14	0,14	0,26	0,52	5	0,12	0,10	0,24	0,30	0,62
7	0,20	0,20	0,34	0,62	0,78	7	0,16	0,20	0,20	0,30	0,76
10	0,26	0,24	0,50	0,58	0,84	10	0,30	0,26	0,34	0,52	0,82
13	0,22	0,32	0,65	0,56	0,90	13	0,30	0,36	0,60	0,82	0,94
16	0,20	0,56	0,55	0,72	0,94	16	0,30	0,50	0,56	0,92	0,98
20	0,32	0,46	0,76	0,86	1,00	20	0,36	0,62	0,66	0,90	1,00
25	0,44	0,50	0,78	0,96	1,00	25	0,34	0,74	0,86	0,94	1,00
27	0,42	0,52	0,78	0,96	1,00	27	0,60	0,70	0,90	0,96	1,00
30	0,48	0,58	0,94	0,96	1,00	30	0,68	0,76	1,00	1,00	1,00
43	0,66	0,90	1,00	1,00	1,00	43	0,70	0,90	1,00	1,00	1,00

TABELA 5. Poder do método rHE em detectar QTLs de diversos efeitos, sob diferentes tamanhos populacionais. A área marcada corresponde às condições nas quais o poder de detecção excede 80%. (A) herdabilidade de 35%; (B) herdabilidade de 50%; (C) herdabilidade de 65%; (D) herdabilidade de 80%.

(A)						(B)					
Efeito do QTL	25	50	75	100	200	Efeito do QTL	25	50	75	100	200
1	0,06	0,06	0,10	0,14	0,10	1	0,08	0,14	0,06	0,06	0,06
3	0,08	0,10	0,14	0,24	0,28	3	0,06	0,10	0,14	0,18	0,26
5	0,06	0,10	0,26	0,26	0,22	5	0,10	0,30	0,16	0,34	0,50
7	0,06	0,16	0,34	0,24	0,46	7	0,10	0,20	0,20	0,32	0,84
10	0,10	0,20	0,26	0,40	0,64	10	0,12	0,30	0,16	0,54	0,88
13	0,20	0,30	0,40	0,34	0,82	13	0,18	0,30	0,50	0,60	0,96
16	0,20	0,46	0,34	0,66	0,82	16	0,26	0,40	0,56	0,74	0,96
20	0,22	0,30	0,48	0,60	0,90	20	0,44	0,40	0,84	0,76	0,98
25	0,20	0,54	0,62	0,76	0,94	25	0,30	0,66	0,90	0,96	1,00
27	0,24	0,48	0,66	0,76	0,96	27	0,34	0,72	0,90	1,00	1,00
30	0,24	0,50	0,80	0,90	1,00	30	0,40	0,80	0,94	1,00	1,00
43	0,50	0,70	1,00	1,00	1,00	43	0,62	0,96	1,00	1,00	1,00

(C)						(D)					
Efeito do QTL	25	50	75	100	200	Efeito do QTL	25	50	75	100	200
1	0,10	0,08	0,06	0,10	0,10	1	0,10	0,16	0,20	0,18	0,16
3	0,10	0,12	0,20	0,32	0,26	3	0,08	0,10	0,20	0,20	0,60
5	0,14	0,26	0,38	0,34	0,62	5	0,14	0,20	0,36	0,46	0,60
7	0,20	0,26	0,46	0,76	0,84	7	0,14	0,24	0,40	0,50	0,84
10	0,32	0,30	0,46	0,72	0,90	10	0,30	0,30	0,50	0,74	0,92
13	0,34	0,46	0,70	0,70	0,96	13	0,30	0,50	0,76	0,84	0,96
16	0,36	0,60	0,80	0,88	0,96	16	0,36	0,66	0,78	0,94	0,98
20	0,42	0,50	0,82	0,96	1,00	20	0,50	0,76	0,78	0,98	1,00
25	0,50	0,70	0,96	1,00	1,00	25	0,40	0,80	0,94	1,00	1,00
27	0,48	0,80	0,94	0,96	1,00	27	0,70	0,80	0,90	0,96	1,00
30	0,58	0,84	0,96	1,00	1,00	30	0,82	0,86	0,96	1,00	1,00
43	0,70	1,00	1,00	1,00	1,00	43	0,94	1,00	1,00	1,00	1,00

TABELA 6. Poder do método oHE-T15B15 em detectar QTLs de diversos efeitos, sob diferentes tamanhos populacionais. Os valores entre parênteses indicam o número de indivíduos selecionados para realização da análise. A área marcada corresponde às condições nas quais o poder de detecção excede 80%. (A) herdabilidade de 35%; (B) herdabilidade de 50%; (C) herdabilidade de 65%; (D) herdabilidade de 80%.

(A)					(B)				
Efeito do QTL	50 (15)	75 (23)	100 (30)	200 (60)	Efeito do QTL	50 (15)	75 (23)	100 (30)	200 (60)
1	0,08	0,07	0,06	0,12	1	0,05	0,05	0,09	0,11
3	0,02	0,09	0,10	0,26	3	0,09	0,09	0,09	0,22
5	0,10	0,14	0,18	0,24	5	0,14	0,13	0,21	0,39
7	0,12	0,13	0,15	0,32	7	0,15	0,22	0,30	0,54
10	0,12	0,21	0,29	0,70	10	0,20	0,26	0,39	0,67
13	0,22	0,28	0,31	0,64	13	0,21	0,39	0,44	0,82
16	0,18	0,27	0,43	0,76	16	0,34	0,48	0,52	0,90
20	0,28	0,41	0,53	0,84	20	0,41	0,64	0,70	0,95
25	0,30	0,49	0,63	0,96	25	0,49	0,73	0,82	0,99
27	0,54	0,57	0,66	0,95	27	0,51	0,67	0,85	0,99
30	0,40	0,57	0,74	0,94	30	0,59	0,75	0,90	1,00
43	0,48	0,73	0,88	1,00	43	0,79	0,89	0,99	1,00

(C)					(D)				
Efeito do QTL	50 (15)	75 (23)	100 (30)	200 (60)	Efeito do QTL	50 (15)	75 (23)	100 (30)	200 (60)
1	0,07	0,07	0,08	0,14	1	0,05	0,08	0,11	0,16
3	0,11	0,14	0,17	0,34	3	0,11	0,13	0,15	0,34
5	0,19	0,19	0,29	0,52	5	0,18	0,24	0,31	0,48
7	0,20	0,29	0,39	0,68	7	0,18	0,29	0,34	0,68
10	0,29	0,37	0,54	0,86	10	0,27	0,49	0,55	0,85
13	0,34	0,49	0,61	0,91	13	0,40	0,53	0,65	0,93
16	0,46	0,61	0,71	0,96	16	0,44	0,59	0,79	0,97
20	0,47	0,70	0,80	1,00	20	0,56	0,72	0,87	1,00
25	0,62	0,78	0,90	1,00	25	0,72	0,85	0,94	1,00
27	0,63	0,81	0,93	1,00	27	0,71	0,94	0,99	1,00
30	0,70	0,91	0,95	1,00	30	0,82	0,93	0,97	1,00
43	0,87	0,99	1,00	1,00	43	0,92	0,98	1,00	1,00

TABELA 7. Poder do método oHE-T20B20 em detectar QTLs de diversos efeitos, sob diferentes tamanhos populacionais. Os valores entre parênteses indicam o número de indivíduos selecionados para realização da análise. A área marcada corresponde às condições nas quais o poder de detecção excede 80%. (A) herdabilidade de 35%; (B) herdabilidade de 50%; (C) herdabilidade de 65%; (D) herdabilidade de 80%.

(A)

Efeito do QTL	50 (20)	75 (30)	100 (40)	200 (80)
1	0,06	0,05	0,05	0,08
3	0,04	0,13	0,10	0,21
5	0,16	0,17	0,16	0,32
7	0,16	0,14	0,18	0,45
10	0,08	0,21	0,28	0,66
13	0,24	0,32	0,30	0,62
16	0,18	0,33	0,48	0,83
20	0,30	0,46	0,58	0,88
25	0,52	0,54	0,66	0,98
27	0,53	0,58	0,69	0,96
30	0,57	0,64	0,77	0,97
43	0,62	0,74	0,89	1,00

(B)

Efeito do QTL	50 (20)	75 (30)	100 (40)	200 (80)
1	0,06	0,07	0,07	0,14
3	0,12	0,10	0,11	0,26
5	0,13	0,14	0,23	0,39
7	0,14	0,28	0,30	0,58
10	0,22	0,29	0,41	0,74
13	0,22	0,40	0,50	0,83
16	0,36	0,53	0,54	0,92
20	0,39	0,66	0,77	0,96
25	0,56	0,78	0,85	0,99
27	0,54	0,69	0,88	1,00
30	0,61	0,80	0,92	1,00
43	0,83	0,93	0,99	1,00

(C)

Efeito do QTL	50 (20)	75 (30)	100 (40)	200 (80)
1	0,07	0,08	0,07	0,15
3	0,11	0,19	0,14	0,31
5	0,16	0,25	0,29	0,57
7	0,24	0,29	0,40	0,70
10	0,27	0,40	0,56	0,91
13	0,33	0,56	0,66	0,93
16	0,48	0,66	0,77	0,98
20	0,54	0,74	0,84	1,00
25	0,64	0,82	0,94	1,00
27	0,67	0,83	0,96	1,00
30	0,74	0,92	0,98	1,00
43	0,90	0,99	1,00	1,00

(D)

Efeito do QTL	50 (20)	75 (30)	100 (40)	200 (80)
1	0,07	0,07	0,13	0,20
3	0,10	0,13	0,18	0,36
5	0,14	0,25	0,30	0,51
7	0,19	0,32	0,36	0,71
10	0,25	0,53	0,61	0,88
13	0,39	0,56	0,74	0,96
16	0,48	0,65	0,79	1,00
20	0,59	0,77	0,91	1,00
25	0,69	0,86	0,97	1,00
27	0,72	0,95	0,99	1,00
30	0,82	0,94	0,98	1,00
43	0,96	1,00	1,00	1,00

Assim como relatado na literatura (FERREIRA, 2004; FEINGOLD, 2002), a análise dos dados permitiu a constatação de que quanto maior o efeito do loco QTL, maior a probabilidade de detectá-lo. Ainda, para QTLs de mesma ação gênica, quanto maior o tamanho amostral, maior a probabilidade de detecção. Ao considerar QTLs de efeito pequeno efeito (inferior a 10%, por exemplo), são necessárias populações maiores para que ele possa ser detectado. Para exemplificar tal fato, na Figura 2, está apresentado o poder estatístico associado ao método rHE (método de maior poder dentre os avaliados) para característica simulada de herdabilidade 80%.

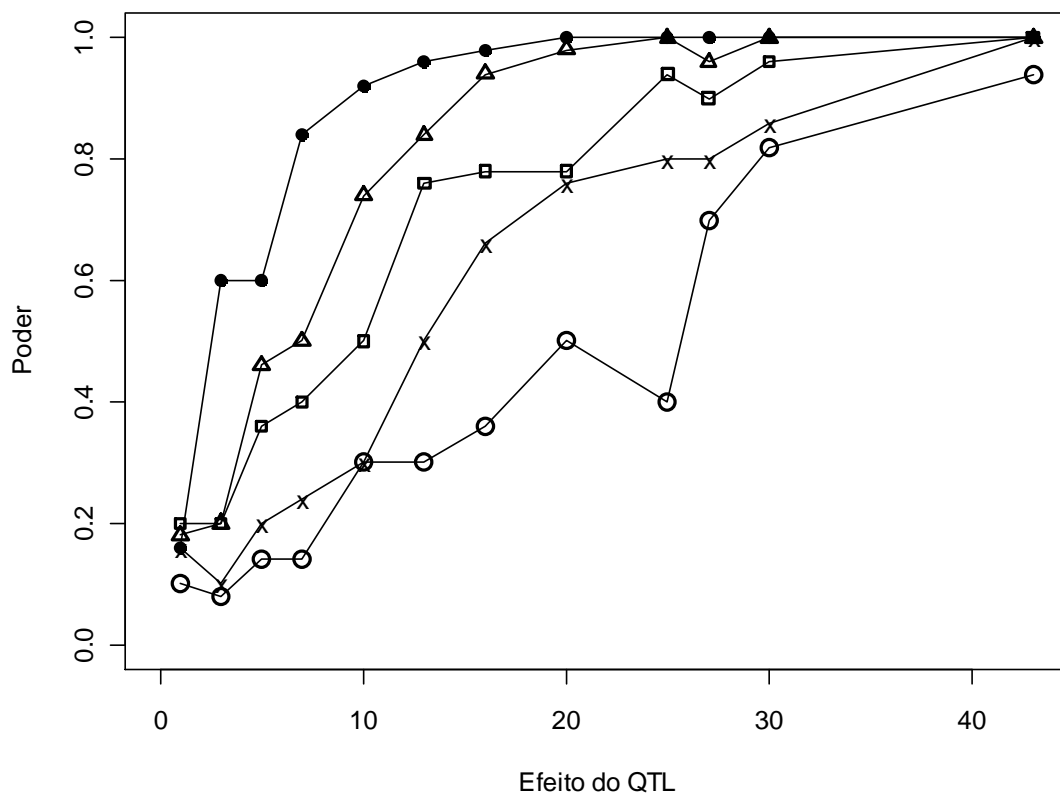


Figura 2. Poder estimado do teste estatístico relativo aos diferentes tamanhos populacionais avaliados. Característica simulada de herdabilidade de 80%. Resultados referentes à metodologia rHE. (●) 200 indivíduos. (△) 100 indivíduos; (□) 75 indivíduos. (x) 50 indivíduos. (○) 25 indivíduos.

Pode-se constatar que somente populações de 200 indivíduos permitem, com confiabilidade, detectar locos QTLs que contribuem com 7–10% da característica fenotípica; populações de 100 indivíduos permitem a detecção de locos com efeito superior a 13%; populações de 75 indivíduos, efeito superior a 20%; e populações de 50 indivíduos, efeito gênico superior a 25%. Nas situações avaliadas, populações de 25 indivíduos se mostraram poucos úteis para detecção de locos quantitativos ou oligogênicos.

Deve-se considerar ainda que os cenários avaliados constitui situação idealizada à identificação de locos QTL. Na prática, as dezenas de genes (ou locos QTLs) participam das rotas bioquímicas que culminam na expressão da característica quantitativa e/ou oligogênica, apresentam os mais variados modos de ação gênica, o que torna o processo de identificação de genes (QTLs) ainda mais difícil.

4.2. ESTIMATIVA DE PARÂMETROS GENÉTICOS

É sabido que fatores como herdabilidade afetam grandemente a detecção de locos que afetam a característica fenotípica. Genes de características de alta herdabilidade, tais como monogenes ou oligogenes, são mais facilmente detectáveis, pois sua influência na variação fenotípica é maior.

Foi estimada a contribuição de cada loco QTL para herdabilidade da característica. Nas Tabelas 8 a 23 estão apresentadas as estimativas de herdabilidades para os marcadores completamente ligados aos locos QTLs, para os métodos oHE, rHE, EDAC T15B15 e EDAC T20B20, considerando os tamanhos populacionais e herdabilidades previamente simulados. Os resultados corroboram o fato de que quanto maior o efeito do loco, maior a contribuição deste para a variação fenotípica, e portanto, maior a estimativa de herdabilidade associada ao loco. A fração da herdabilidade associada a cada

loco foi estimada por meio da equação:
$$h^2 = \frac{\hat{\beta}_1}{2\hat{\sigma}_F^2}.$$

A metodologia oHE (Tabela 9 a 12) freqüentemente apresentou estimativas de herdabilidade superestimadas, e em muitos casos, superiores à unidade. Para o tamanho populacional de 25 indivíduos, particularmente, tal

fato, mostrou-se mais pronunciado. Os métodos EDAC T20B20 e T15B15 (Tabela 16 – 23), por constituírem essencialmente do método oHE no qual uma fração dos indivíduos foram selecionados, apresentaram o mesmo problema de superestimação.

TABELA 8. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método oHE; herdabilidade simulada de 35%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1000	0,1175	0,0531	0,0453	0,0282	0,0236	0,0272	0,0182	0,0111	0,0107
3	0,1072	0,0654	0,0578	0,0441	0,0352	0,0186	0,0404	0,0257	0,0277	0,0144
5	0,0697	0,0697	0,0611	0,0632	0,0440	0,0417	0,0293	0,0223	0,0215	0,0213
7	0,0898	0,0836	0,0384	0,0415	0,0672	0,0611	0,0524	0,0520	0,0246	0,0216
10	0,1222	0,0897	0,0603	0,0479	0,0522	0,0437	0,0519	0,0360	0,0419	0,0284
13	0,1304	0,0886	0,0697	0,0562	0,0663	0,0417	0,0498	0,0456	0,0501	0,0264
16	0,1264	0,0992	0,0890	0,1101	0,0609	0,0349	0,0636	0,0312	0,0479	0,0290
20	0,1231	0,0901	0,0947	0,0964	0,0631	0,0461	0,0701	0,0526	0,0725	0,0371
25	0,1795	0,1036	0,1338	0,0941	0,1072	0,0890	0,0838	0,0716	0,0920	0,0338
TOTAL	1,0482		0,6580		0,5242		0,4684		0,3893	
27	0,1153	0,0784	0,1206	0,0848	0,0854	0,0603	0,1005	0,0613	0,0786	0,0393
30	0,1475	0,1301	0,0945	0,0596	0,1434	0,0737	0,1265	0,0724	0,1128	0,0324
43	0,1781	0,1219	0,1247	0,0616	0,1323	0,0580	0,1698	0,0899	0,1411	0,0476
TOTAL	0,4408		0,3398		0,3611		0,3968		0,3325	

TABELA 9. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método oHE; herdabilidade simulada de 50%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0632	0,0734	0,0584	0,0456	0,0389	0,0544	0,0224	0,0158	0,0133	0,0087
3	0,0804	0,0661	0,0457	0,0491	0,0518	0,0589	0,0246	0,0188	0,0214	0,0199
5	0,1306	0,1078	0,0860	0,0789	0,0517	0,0417	0,0352	0,0249	0,0292	0,0307
7	0,0894	0,0419	0,0696	0,0500	0,0727	0,0488	0,0434	0,0317	0,0429	0,0176
10	0,1115	0,0825	0,0752	0,0665	0,0597	0,0579	0,0530	0,0357	0,0447	0,0261
13	0,1348	0,1183	0,0694	0,0526	0,0605	0,0439	0,0647	0,0395	0,0630	0,0369
16	0,1200	0,0919	0,1122	0,0763	0,0839	0,0527	0,0794	0,0432	0,0764	0,0342
20	0,2094	0,1574	0,1160	0,0851	0,1287	0,0722	0,0998	0,0774	0,0840	0,0452
25	0,1732	0,1237	0,1616	0,1124	0,1224	0,0619	0,1349	0,0806	0,1167	0,0396
TOTAL	1,1125		0,7940		0,6704		0,5574		0,4915	
27	0,1977	0,1393	0,1374	0,0744	0,1339	0,0766	0,1405	0,0675	0,1236	0,0447
30	0,1743	0,1591	0,1706	0,0928	0,1483	0,0734	0,1528	0,0613	0,1487	0,0467
43	0,3126	0,1962	0,2572	0,1301	0,1899	0,0836	0,2105	0,0799	0,2270	0,0474
TOTAL	0,6846		0,5651		0,4721		0,5038		0,4994	

TABELA 10. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método oHE; herdabilidade simulada de 65%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0990	0,0570	0,0557	0,0954	0,0401	0,0385	0,0317	0,0241	0,0155	0,0074
3	0,1232	0,1101	0,0382	0,0396	0,0519	0,0349	0,0450	0,0378	0,0236	0,0220
5	0,1218	0,1270	0,0734	0,0738	0,0561	0,0313	0,0335	0,0265	0,0353	0,0307
7	0,1842	0,1490	0,0773	0,0574	0,0727	0,0469	0,0744	0,0342	0,0500	0,0257
10	0,2147	0,1192	0,0938	0,0726	0,0814	0,0561	0,0951	0,0625	0,0806	0,0308
13	0,1687	0,1307	0,1016	0,0865	0,1199	0,0849	0,0869	0,0632	0,0786	0,0404
16	0,1541	0,1011	0,1561	0,0817	0,0849	0,0561	0,1230	0,0857	0,1107	0,0537
20	0,1911	0,1294	0,1393	0,1076	0,1545	0,0973	0,1306	0,0612	0,1331	0,0425
25	0,2204	0,1330	0,1425	0,0920	0,1663	0,0812	0,1558	0,0579	0,1561	0,0541
TOTAL	1,4772		0,8778		0,8276		0,7758		0,6834	
27	0,1418	0,0930	0,2041	0,1178	0,1557	0,1015	0,1947	0,0796	0,1826	0,0475
30	0,2105	0,1100	0,1918	0,0980	0,2024	0,0811	0,1812	0,0744	0,1924	0,0545
43	0,3399	0,2119	0,2267	0,1039	0,2892	0,0967	0,2970	0,0833	0,2535	0,0578
TOTAL	0,6922		0,6227		0,6473		0,6730		0,6285	

TABELA 11. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método oHE; herdabilidade simulada de 80%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1352	0,0770	0,0583	0,0455	0,0559	0,0691	0,0248	0,0228	0,0161	0,0105
3	0,0754	0,0440	0,0526	0,0427	0,0500	0,0361	0,0299	0,0275	0,0296	0,0211
5	0,1169	0,0983	0,0635	0,0556	0,0581	0,0270	0,0509	0,0394	0,0395	0,0233
7	0,1015	0,0839	0,1091	0,0791	0,0519	0,0546	0,0461	0,0339	0,0463	0,0255
10	0,1715	0,1549	0,0974	0,0720	0,1046	0,1018	0,0518	0,0297	0,0684	0,0413
13	0,1275	0,0825	0,0944	0,0682	0,1095	0,0586	0,1019	0,0598	0,0823	0,0371
16	0,1811	0,1403	0,1217	0,0850	0,0974	0,0649	0,1344	0,0558	0,1223	0,0417
20	0,1823	0,1245	0,1688	0,0969	0,1166	0,0747	0,1277	0,0656	0,1543	0,0322
25	0,2294	0,1795	0,1742	0,1005	0,1575	0,0880	0,1568	0,0714	0,1601	0,0426
TOTAL	1,3207		0,9400		0,8016		0,7242		0,7189	
27	0,3009	0,1316	0,2014	0,1196	0,1971	0,1134	0,2060	0,0705	0,1809	0,0580
30	0,2798	0,2017	0,1831	0,0698	0,2027	0,0582	0,1770	0,1116	0,2344	0,0675
43	0,3049	0,1162	0,3308	0,1528	0,3410	0,0756	0,2736	0,0810	0,3187	0,0593
TOTAL	0,8856		0,7154		0,7407		0,6566		0,7340	

TABELA 12. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método rHE; herdabilidade simulada de 35%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0350	0,0371	0,0164	0,0145	0,0090	0,0081	0,0086	0,0075	0,0042	0,0042
3	0,0278	0,0234	0,0166	0,0195	0,0114	0,0087	0,0114	0,0097	0,0078	0,0069
5	0,0212	0,0122	0,0190	0,0192	0,0152	0,0154	0,0112	0,0087	0,0090	0,0093
7	0,0245	0,0214	0,0166	0,0152	0,0220	0,0273	0,0182	0,0214	0,0117	0,0103
10	0,0391	0,0388	0,0184	0,0195	0,0197	0,0213	0,0191	0,0188	0,0162	0,0136
13	0,0403	0,0376	0,0304	0,0245	0,0262	0,0210	0,0189	0,0187	0,0227	0,0140
16	0,0339	0,0391	0,0463	0,0441	0,0220	0,0159	0,0247	0,0148	0,0228	0,0142
20	0,0495	0,0414	0,0375	0,0361	0,0326	0,0176	0,0307	0,0252	0,0368	0,0184
25	0,0407	0,0318	0,0504	0,0417	0,0464	0,0410	0,0383	0,0343	0,0465	0,0164
TOTAL	0,3119		0,2514		0,2042		0,1809		0,1775	
27	0,0369	0,0203	0,0475	0,0409	0,0401	0,0267	0,04585	0,03215	0,03825	0,02055
30	0,0369	0,0288	0,0410	0,0295	0,0673	0,0398	0,0588	0,0347	0,05575	0,01455
43	0,0773	0,0592	0,0581	0,0363	0,0701	0,0241	0,08595	0,04475	0,0706	0,02425
TOTAL	0,1511		0,1466		0,1775		0,1906		0,1646	

TABELA 13. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método rHE; herdabilidade simulada de 50%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0320	0,0288	0,0181	0,0154	0,0125	0,0148	0,0063	0,0064	0,0039	0,0029
3	0,0269	0,0201	0,0145	0,0132	0,0179	0,0229	0,0058	0,0048	0,0085	0,0081
5	0,0383	0,0268	0,0323	0,0347	0,0162	0,0152	0,0162	0,0117	0,0127	0,0114
7	0,0282	0,0184	0,0207	0,0194	0,0227	0,0226	0,0124	0,0131	0,0178	0,0073
10	0,0264	0,0221	0,0271	0,0179	0,0190	0,0243	0,0235	0,0178	0,0211	0,0131
13	0,0433	0,0498	0,0232	0,0235	0,0275	0,0241	0,0261	0,0201	0,0327	0,0183
16	0,0427	0,0397	0,0378	0,0400	0,0349	0,0228	0,0332	0,0213	0,0375	0,0157
20	0,0725	0,0645	0,0474	0,0428	0,0613	0,0379	0,0480	0,0339	0,0412	0,0221
25	0,0634	0,0605	0,0693	0,0518	0,0594	0,0322	0,0659	0,0377	0,0582	0,0188
TOTAL	0,3736		0,2901		0,2712		0,2373		0,2334	
27	0,0676	0,0610	0,0598	0,0354	0,0640	0,0373	0,0681	0,0315	0,0602	0,0219
30	0,0541	0,0474	0,0734	0,0472	0,0738	0,0390	0,0736	0,0296	0,0733	0,0240
43	0,1399	0,0851	0,1245	0,0612	0,0923	0,0383	0,1014	0,0389	0,1125	0,0226
TOTAL	0,2615		0,2576		0,2300		0,2431		0,2459	

TABELA 14. Estimativas dos valores médios herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método rHE; herdabilidade simulada de 65%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0289	0,0198	0,0182	0,0214	0,0102	0,0093	0,0095	0,0089	0,0046	0,0036
3	0,0362	0,0377	0,0124	0,0072	0,0173	0,0165	0,0169	0,0176	0,0081	0,0103
5	0,0323	0,0354	0,0251	0,0288	0,0183	0,0128	0,0159	0,0131	0,0176	0,0146
7	0,0536	0,0652	0,0355	0,0288	0,0301	0,0273	0,0317	0,0186	0,0218	0,0139
10	0,0588	0,0530	0,0357	0,0346	0,0307	0,0278	0,0409	0,0319	0,0400	0,0150
13	0,0606	0,0550	0,0437	0,0362	0,0510	0,0391	0,0430	0,0321	0,0383	0,0194
16	0,0582	0,0390	0,0596	0,0434	0,0418	0,0238	0,0613	0,0433	0,0546	0,0277
20	0,0590	0,0502	0,0556	0,0490	0,0704	0,0489	0,0639	0,0288	0,0643	0,0207
25	0,0871	0,0657	0,0686	0,0445	0,0774	0,0342	0,0769	0,0267	0,0780	0,0264
TOTAL	0,4745		0,3543		0,3468		0,3598		0,3272	
27	0,0599	0,0451	0,0908	0,0556	0,0773	0,0434	0,0973	0,0373	0,0905	0,0234
30	0,0924	0,0513	0,0866	0,0499	0,0926	0,0386	0,0868	0,0363	0,0948	0,0279
43	0,1563	0,1038	0,1106	0,0503	0,1431	0,0460	0,1457	0,0380	0,1281	0,0281
TOTAL	0,3085		0,2879		0,3129		0,3297		0,3134	

TABELA 15. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método rHE; herdabilidade simulada de 80%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0291	0,0220	0,0189	0,0187	0,0176	0,0256	0,0061	0,0125	0,0017	0,0064
3	0,0230	0,0194	0,0192	0,0171	0,0164	0,0125	0,0067	0,0156	0,0107	0,0107
5	0,0390	0,0338	0,0222	0,0212	0,0200	0,0145	0,0198	0,0191	0,0153	0,0120
7	0,0361	0,0312	0,0333	0,0399	0,0242	0,0262	0,0194	0,0198	0,0219	0,0125
10	0,0541	0,0506	0,0381	0,0359	0,0461	0,0484	0,0288	0,0146	0,0339	0,0183
13	0,0463	0,0363	0,0436	0,0288	0,0503	0,0313	0,0483	0,0292	0,0401	0,0180
16	0,0655	0,0652	0,0521	0,0371	0,0508	0,0324	0,0654	0,0291	0,0616	0,0196
20	0,0723	0,0571	0,0807	0,0473	0,0549	0,0356	0,0649	0,0294	0,0765	0,0148
25	0,0788	0,0681	0,0793	0,0427	0,0812	0,0412	0,0772	0,0303	0,0788	0,0201
TOTAL	0,4440		0,3871		0,3613		0,3364		0,3402	
27	0,1289	0,0772	0,0894	0,0567	0,0969	0,0513	0,1034	0,0340	0,0909	0,0281
30	0,1290	0,0933	0,0733	0,0405	0,0972	0,0309	0,0857	0,0513	0,1173	0,0338
43	0,1473	0,0596	0,1634	0,0706	0,1615	0,0364	0,1341	0,0393	0,1577	0,0295
TOTAL	0,4052		0,3260		0,3556		0,3232		0,3658	

TABELA 16. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 35%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1136	0,1213	0,0873	0,0707	0,0833	0,0654	0,0498	0,0331
3	0,1176	0,0820	0,0831	0,0577	0,0938	0,0817	0,0623	0,0512
5	0,1355	0,1287	0,1345	0,1322	0,0922	0,0806	0,0505	0,0619
7	0,1396	0,1234	0,1381	0,1198	0,1899	0,1760	0,0840	0,0943
10	0,1174	0,1299	0,1414	0,1119	0,1289	0,0852	0,1410	0,0649
13	0,2148	0,1469	0,1251	0,0901	0,1410	0,1158	0,1393	0,1110
16	0,2539	0,2057	0,1045	0,0843	0,1637	0,1322	0,1343	0,0814
20	0,2920	0,1572	0,1696	0,0832	0,1447	0,1247	0,1827	0,1079
25	0,3855	0,2299	0,2256	0,1756	0,2047	0,1710	0,2452	0,1034
TOTAL	1,7699		1,2092		1,2421		1,0890	
27	0,3021	0,1455	0,2487	0,1422	0,2558	0,1797	0,2155	0,1259
30	0,2543	0,1877	0,2976	0,2211	0,2695	0,1768	0,2947	0,0988
43	0,2906	0,1622	0,3431	0,1051	0,4117	0,1982	0,3565	0,1251
TOTAL	0,8470		0,8895		0,9370		0,8667	

TABELA 17. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 50%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1297	0,0811	0,0716	0,0510	0,0579	0,0501	0,0400	0,0304
3	0,0938	0,0850	0,0862	0,0916	0,0597	0,0543	0,0578	0,0486
5	0,2280	0,2248	0,1112	0,0976	0,0898	0,0603	0,0827	0,0851
7	0,2043	0,1946	0,1371	0,1387	0,1382	0,1119	0,0963	0,0583
10	0,1897	0,1291	0,1182	0,1315	0,1206	0,0943	0,1040	0,0600
13	0,1749	0,1681	0,1627	0,1046	0,1465	0,1267	0,1675	0,0947
16	0,2410	0,1922	0,2219	0,1479	0,2069	0,1294	0,1645	0,0833
20	0,3443	0,2556	0,2612	0,1944	0,2561	0,2239	0,2019	0,1188
25	0,4049	0,2510	0,3075	0,2034	0,3457	0,2203	0,2910	0,1018
TOTAL	2,0105		1,4775		1,4214		1,2057	
27	0,3800	0,1973	0,3229	0,1816	0,3826	0,1624	0,2951	0,1070
30	0,3791	0,2584	0,3163	0,1448	0,3372	0,1515	0,3489	0,1477
43	0,5317	0,2472	0,3949	0,1622	0,4975	0,1997	0,5475	0,1140
TOTAL	1,2907		1,0340		1,2173		1,1915	

TABELA 18. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 65%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1442	0,1199	0,0925	0,1482	0,0522	0,0660	0,0454	0,0380
3	0,0701	0,0830	0,1013	0,1072	0,1256	0,1491	0,0609	0,0603
5	0,1856	0,2441	0,0788	0,0652	0,1159	0,0990	0,0813	0,0638
7	0,1946	0,1554	0,1995	0,1248	0,1842	0,1085	0,1343	0,0804
10	0,2477	0,2516	0,1762	0,1384	0,2214	0,1245	0,2008	0,0921
13	0,2536	0,1813	0,2597	0,1857	0,2953	0,1610	0,1966	0,1056
16	0,2836	0,1902	0,2161	0,1334	0,3085	0,1904	0,2708	0,1399
20	0,3260	0,2443	0,3143	0,2184	0,2911	0,1514	0,3155	0,1160
25	0,3724	0,2134	0,4388	0,1982	0,4003	0,1587	0,3781	0,1293
TOTAL	2,0778		1,8771		1,9945		1,6837	
27	0,5050	0,2518	0,3467	0,1977	0,4364	0,1701	0,4407	0,1186
30	0,4566	0,2426	0,4643	0,2053	0,4933	0,2001	0,4735	0,1102
43	0,4782	0,2786	0,6544	0,1457	0,6561	0,1601	0,5958	0,1278
TOTAL	1,4398		1,4653		1,5858		1,5101	

TABELA 19. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 80%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1348	0,1296	0,1642	0,1991	0,0948	0,0896	0,0396	0,0235
3	0,1728	0,1295	0,1580	0,0860	0,0927	0,0760	0,0667	0,0650
5	0,2766	0,1915	0,1199	0,0986	0,1507	0,1059	0,0802	0,0587
7	0,2949	0,2576	0,1679	0,1496	0,1178	0,0807	0,1236	0,0722
10	0,2715	0,2409	0,2499	0,2240	0,1534	0,0907	0,1679	0,0891
13	0,2192	0,1423	0,2473	0,1513	0,2297	0,1163	0,2055	0,1004
16	0,2498	0,1625	0,2611	0,1790	0,3378	0,1381	0,2998	0,1255
20	0,4035	0,2030	0,2991	0,1940	0,3333	0,1574	0,3943	0,0925
25	0,4058	0,1991	0,3730	0,1877	0,3997	0,1311	0,3971	0,1252
TOTAL	2,4288		2,0405		1,9099		1,7748	
27	0,4486	0,2747	0,4391	0,1991	0,5012	0,1626	0,4641	0,1398
30	0,3884	0,1798	0,4682	0,1462	0,3921	0,1914	0,5500	0,1443
43	0,6941	0,2219	0,6853	0,1286	0,6566	0,1567	0,7103	0,1158
TOTAL	1,5310		1,5927		1,5498		1,7243	

TABELA 20. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 35%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0840	0,1018	0,0738	0,0649	0,0499	0,0384	0,0376	0,0271
3	0,0851	0,0730	0,0756	0,0717	0,0809	0,0571	0,0563	0,0436
5	0,1256	0,1193	0,1182	0,1000	0,0781	0,0600	0,0437	0,0526
7	0,1353	0,1131	0,1209	0,1087	0,1211	0,1170	0,0603	0,0541
10	0,1522	0,1198	0,1056	0,1059	0,1279	0,0837	0,1146	0,0475
13	0,1696	0,1053	0,1326	0,0700	0,0941	0,0814	0,1047	0,0763
16	0,2607	0,1967	0,0960	0,0668	0,1225	0,0824	0,1076	0,0581
20	0,2138	0,1591	0,1513	0,0697	0,1384	0,1017	0,1533	0,0786
25	0,2900	0,2031	0,1878	0,1468	0,1781	0,1602	0,1946	0,0849
TOTAL	1,5163		1,0617		0,9911		0,8727	
27	0,2392	0,1453	0,2013	0,1386	0,2161	0,1564	0,1768	0,0963
30	0,2680	0,1838	0,2610	0,1810	0,2389	0,1525	0,2499	0,0668
43	0,2222	0,1259	0,3024	0,1034	0,3487	0,1716	0,3075	0,1174
TOTAL	0,7294		0,7646		0,8036		0,7342	

TABELA 21. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 50%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1291	0,0688	0,0576	0,0627	0,0615	0,0357	0,0268	0,0213
3	0,1085	0,1085	0,0757	0,0820	0,0456	0,0275	0,0461	0,0454
5	0,1514	0,1687	0,1098	0,0867	0,0902	0,0685	0,0672	0,0642
7	0,1581	0,1685	0,1138	0,0991	0,1107	0,0832	0,0774	0,0471
10	0,1713	0,1365	0,1234	0,1217	0,1059	0,0804	0,0909	0,0586
13	0,1484	0,1180	0,1483	0,0876	0,1168	0,0921	0,1370	0,0621
16	0,1866	0,1658	0,1747	0,1318	0,1470	0,0920	0,1519	0,0702
20	0,2695	0,2097	0,2479	0,1615	0,2244	0,1618	0,1778	0,0983
25	0,3529	0,2222	0,2680	0,1470	0,2866	0,1794	0,2470	0,0810
TOTAL	1,6757		1,3192		1,1886		1,0221	
27	0,3148	0,1678	0,2708	0,1360	0,3164	0,1530	0,2553	0,1083
30	0,3113	0,2020	0,2885	0,1340	0,2914	0,1233	0,3017	0,1262
43	0,5053	0,2244	0,3725	0,1602	0,4280	0,1796	0,4722	0,0981
TOTAL	1,1314		0,9317		1,0357		1,0293	

TABELA 22. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 65%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,1078	0,1260	0,1028	0,0967	0,0473	0,0418	0,0371	0,0241
3	0,0861	0,0848	0,1304	0,0881	0,1027	0,0944	0,0517	0,0473
5	0,1494	0,1973	0,0889	0,0589	0,0945	0,0784	0,0671	0,0603
7	0,1873	0,1172	0,1652	0,1255	0,1493	0,0756	0,1021	0,0655
10	0,2020	0,1970	0,1643	0,1127	0,2063	0,1115	0,1623	0,0622
13	0,2016	0,1507	0,2283	0,1831	0,2037	0,1521	0,1628	0,0843
16	0,2608	0,1753	0,1898	0,1148	0,2589	0,1589	0,2331	0,1135
20	0,2805	0,2032	0,2985	0,1708	0,2605	0,1260	0,2834	0,1046
25	0,3159	0,2166	0,3537	0,1386	0,3499	0,1322	0,3350	0,1140
TOTAL	1,7913		1,7219		1,6730		1,4347	
27	0,4181	0,2618	0,3025	0,1656	0,4014	0,1533	0,3787	0,1062
30	0,3968	0,2105	0,4070	0,1817	0,4157	0,1758	0,3872	0,0950
43	0,4622	0,2337	0,5859	0,1773	0,5769	0,1413	0,5324	0,1080
TOTAL	1,2771		1,2955		1,3940		1,2983	

TABELA 23. Estimativas dos valores médios de herdabilidade explicados pelos QTLs de diferentes efeitos, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 80%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0583	0,0455	0,1049	0,1525	0,0732	0,0576	0,0319	0,0244
3	0,0526	0,0427	0,1301	0,0788	0,0716	0,0821	0,0516	0,0482
5	0,0635	0,0556	0,1031	0,0808	0,1207	0,0953	0,0730	0,0490
7	0,1091	0,0791	0,1215	0,1350	0,1044	0,0862	0,0975	0,0483
10	0,0974	0,0720	0,2155	0,2025	0,1105	0,0678	0,1431	0,0896
13	0,1000	0,0660	0,2140	0,1204	0,2068	0,0855	0,1620	0,0779
16	0,1217	0,0850	0,2073	0,1410	0,2828	0,1277	0,2641	0,1052
20	0,1688	0,0969	0,2735	0,1558	0,2608	0,1336	0,3355	0,0773
25	0,1742	0,1005	0,3372	0,1723	0,3411	0,1123	0,3432	0,0953
TOTAL	0,9456		1,7072		1,5719		1,5019	
27	0,3991	0,2715	0,3982	0,1703	0,4368	0,1408	0,3821	0,1032
30	0,3583	0,1800	0,4163	0,1205	0,3526	0,1947	0,4851	0,1306
43	0,6390	0,2182	0,6300	0,1106	0,5788	0,1346	0,6337	0,0991
TOTAL	1,3964		1,4445		1,3682		1,5009	

Foram avaliados ainda o parâmetro coeficiente de determinação (R^2) da análise de regressão, para os marcadores completamente ligados aos locos QTLs. Em modelos lineares, o R^2 mensura a proporção da variância total que é explicada pela variável explicativa. No entanto, apesar de o método de Haseman-Elston constituir-se de um modelo de regressão, a interpretação do coeficiente de determinação não pode ser realizada de maneira direta.

Consideremos, por exemplo, o uso e a interpretação do R^2 , para análise de marca simples em uma população F2, como descrita por SCHUSTER & CRUZ (2004). Seja uma população constituída por genótipos Marcadores M_1M_1 , M_1M_2 e M_2M_2 ao qual são atribuídos valores -1, 0 e 1, para ajuste do coeficiente de regressão associado a um modelo aditivo; e 0, 1 e 0, para ajuste do coeficiente associado a um modelo dominante.

Dessa forma, pode ser ajustado o modelo: $Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j$ ou o modelo detalhado $Y_j = \beta_0 + \beta_1 X_j + \beta_2 X_j + \varepsilon_j$; em que Y_j corresponde ao valor para característica; β_0 , o intercepto da regressão; β_1 , o coeficiente associado ao efeito aditivo do marcador; β_2 , o coeficiente associado ao efeito de dominância do marcador, X_j , a codificação associada ao genótipo como especificado anteriormente; e ε_j , o erro aleatório manifestado pela característica no j-ésimo indivíduo. Constatada significância para o modelo reduzido ou o modelo completo, o R^2 calculado do modelo corresponde à proporção da variância fenotípica explicada pelo modelo ajustado. Quanto mais próximo à unidade, melhor o marcador genético explica a variação fenotípica.

Para o método de Haseman-Elston, no entanto, o vetor de observações Y_j que é submetido à análise de regressão, constitui uma diferença entre dois indivíduos, elevada ao quadrado. O elemento da variável explicativa associada ao coeficiente de regressão constitui uma proporção de alelos IBD compartilhado pelos dois indivíduos considerados no respectivo elemento de Y_j . Diferente da análise regressão realizada para população F2, o coeficiente de determinação do método oHE e derivados não é interpretável.

Para os métodos oHE e rHE, verificou-se nas análises, que mesmo para os marcadores completamente ligados aos QTLs, os valores de R^2

médios (das análises cujos resultados foram significativos) são considerados baixos. Os resultados de R^2 médio estão apresentados nas Tabelas 29 a 40, nas quais são considerados os diferentes tamanhos populacionais e valores de herdabilidades.

Em geral, há um acréscimo no valor de R^2 a medida que aumenta a contribuição do loco QTL para a característica. Tal fato mais pronunciado em valores maiores de herdabilidade. Embora o R^2 neste tipo de análise não seja interpretável, WANG & ELSTON (2004) sugerem a utilização deste parâmetro para mensurar a eficiência comparativa entre os métodos. Por meio da avaliação dos valores médios de R^2 , chega-se que o método rHE é o mais eficiente, proporcionando estimativas de R^2 : 1,56; 1,58; 1,68 e 1,74 vezes maior para características simuladas de herdabilidades de 35, 50, 65 e 80%, respectivamente.

Ainda, o processo de seleção proporcionou um grande aumento nas estimativas de R^2 do método oHE com seleção, embora tais valores ainda sejam considerados baixos. Os resultados de R^2 médio para os métodos EDAC T15B15 e T20B20 estão apresentados nas Tabelas 24 a 39, nas quais são considerados os diferentes tamanhos populacionais e valores de herdabilidades. Comparados com o método oHE sem seleção, o método EDAC T15B15 proporcionou estimativas: 11,91; 12,09; 10,99 e 12,61 vezes maior para os valores de herdabilidade de 35, 50, 65 e 80%, respectivamente – tais estimativas excluem o tamanho populacional de 25 indivíduos. Para o método EDAC T20B20, foram encontradas estimativas: 8,31; 8,51; 7,72 e 6,86 vezes maior para os valores de herdabilidade de 35, 50, 65 e 80%, respectivamente – tais estimativas excluem o tamanho populacional de 25 indivíduos.

TABELA 24. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método oHE; herdabilidade simulada de 35%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0050	0,0118	0,0009	0,0014	0,0003	0,0004	0,0002	0,0002	0,0000	0,0001
3	0,0024	0,0018	0,0009	0,0012	0,0003	0,0004	0,0005	0,0005	0,0002	0,0002
5	0,0019	0,0027	0,0015	0,0028	0,0008	0,0015	0,0003	0,0004	0,0002	0,0004
7	0,0026	0,0038	0,0006	0,0012	0,0016	0,0033	0,0007	0,0015	0,0002	0,0004
10	0,0052	0,0061	0,0010	0,0016	0,0011	0,0017	0,0008	0,0011	0,0006	0,0006
13	0,0059	0,0084	0,0020	0,0028	0,0013	0,0015	0,0008	0,0016	0,0007	0,0008
16	0,0048	0,0080	0,0050	0,0092	0,0010	0,0012	0,0011	0,0010	0,0008	0,0008
20	0,0046	0,0063	0,0038	0,0066	0,0016	0,0017	0,0018	0,0032	0,0017	0,0015
25	0,0063	0,0084	0,0068	0,0085	0,0050	0,0080	0,0028	0,0058	0,0024	0,0018
27	0,0043	0,0052	0,0047	0,0061	0,0028	0,0031	0,0036	0,0037	0,0019	0,0019
30	0,0070	0,0123	0,0029	0,0030	0,0065	0,0058	0,0056	0,0068	0,0035	0,0020
43	0,0119	0,0196	0,0053	0,0052	0,0589	0,0210	0,0097	0,0088	0,0055	0,0034

TABELA 25. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método oHE; herdabilidade simulada de 50%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0026	0,0040	0,0007	0,0014	0,0005	0,0011	0,0002	0,0002	0,0001	0,0001
3	0,0025	0,0033	0,0008	0,0012	0,0008	0,0023	0,0002	0,0002	0,0002	0,0003
5	0,0054	0,0091	0,0029	0,0048	0,0008	0,0013	0,0004	0,0006	0,0004	0,0010
7	0,0031	0,0035	0,0013	0,0017	0,0016	0,0022	0,0004	0,0007	0,0005	0,0004
10	0,0043	0,0055	0,0022	0,0031	0,0013	0,0028	0,0009	0,0010	0,0007	0,0007
13	0,0061	0,0114	0,0014	0,0021	0,0015	0,0017	0,0011	0,0015	0,0013	0,0015
16	0,0045	0,0079	0,0039	0,0054	0,0024	0,0026	0,0018	0,0018	0,0018	0,0014
20	0,0147	0,0210	0,0046	0,0057	0,0057	0,0059	0,0041	0,0047	0,0023	0,0026
25	0,0101	0,0184	0,0098	0,0117	0,0048	0,0049	0,0062	0,0068	0,0039	0,0026
27	0,0140	0,0206	0,0052	0,0051	0,0065	0,0069	0,0061	0,0043	0,0046	0,0032
30	0,0118	0,0274	0,0084	0,0095	0,0075	0,0078	0,0070	0,0047	0,0063	0,0036
43	0,0398	0,0499	0,0219	0,0181	0,0113	0,0100	0,0135	0,0087	0,0141	0,0060

TABELA 26. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método oHE; herdabilidade simulada de 65%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0033	0,0031	0,0016	0,0041	0,0005	0,0010	0,0003	0,0004	0,0001	0,0001
3	0,0077	0,0133	0,0005	0,0011	0,0007	0,0010	0,0006	0,0010	0,0002	0,0003
5	0,0066	0,0109	0,0017	0,0030	0,0007	0,0009	0,0004	0,0006	0,0006	0,0009
7	0,0102	0,0195	0,0025	0,0035	0,0016	0,0017	0,0017	0,0016	0,0007	0,0009
10	0,0093	0,0146	0,0029	0,0047	0,0022	0,0023	0,0030	0,0040	0,0020	0,0016
13	0,0095	0,0173	0,0043	0,0075	0,0046	0,0067	0,0030	0,0039	0,0021	0,0019
16	0,0091	0,0099	0,0062	0,0069	0,0027	0,0033	0,0062	0,0092	0,0039	0,0034
20	0,0101	0,0132	0,0070	0,0102	0,0085	0,0119	0,0055	0,0043	0,0051	0,0028
25	0,0160	0,0236	0,0070	0,0076	0,0086	0,0083	0,0074	0,0057	0,0074	0,0066
27	0,0059	0,0065	0,0142	0,0150	0,0094	0,0120	0,0123	0,0094	0,0094	0,0046
30	0,0147	0,0125	0,0113	0,0129	0,0123	0,0089	0,0103	0,0079	0,0105	0,0065
43	0,0491	0,0593	0,0169	0,0151	0,0244	0,0174	0,0265	0,0133	0,0179	0,0081

TABELA 27. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método oHE; herdabilidade simulada de 80%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0043	0,0051	0,0010	0,0015	0,0013	0,0030	0,0002	0,0004	0,0001	0,0001
3	0,0018	0,0016	0,0009	0,0018	0,0007	0,0008	0,0003	0,0005	0,0003	0,0004
5	0,0046	0,0081	0,0010	0,0018	0,0008	0,0008	0,0009	0,0014	0,0005	0,0006
7	0,0041	0,0054	0,0024	0,0043	0,0014	0,0035	0,0008	0,0014	0,0007	0,0007
10	0,0118	0,0213	0,0028	0,0038	0,0049	0,0085	0,0010	0,0009	0,0016	0,0016
13	0,0055	0,0068	0,0033	0,0041	0,0038	0,0036	0,0038	0,0039	0,0021	0,0019
16	0,0120	0,0200	0,0063	0,0085	0,0034	0,0041	0,0057	0,0039	0,0044	0,0027
20	0,0123	0,0150	0,0096	0,0112	0,0051	0,0070	0,0058	0,0063	0,0065	0,0027
25	0,0155	0,0224	0,0109	0,0102	0,0086	0,0091	0,0079	0,0063	0,0073	0,0044
27	0,0315	0,0318	0,0142	0,0135	0,0146	0,0152	0,0127	0,0085	0,0093	0,0054
30	0,0382	0,0584	0,0092	0,0090	0,0120	0,0066	0,0113	0,0146	0,0159	0,0093
43	0,0331	0,0232	0,0363	0,0303	0,0325	0,0145	0,0212	0,0123	0,0278	0,0105

TABELA 28. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método rHE; herdabilidade simulada de 35%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0058	0,0149	0,0010	0,0022	0,0003	0,0005	0,0003	0,0004	0,0001	0,0001
3	0,0031	0,0064	0,0014	0,0032	0,0004	0,0005	0,0005	0,0007	0,0002	0,0003
5	0,0014	0,0013	0,0015	0,0032	0,0010	0,0018	0,0004	0,0006	0,0003	0,0006
7	0,0025	0,0042	0,0012	0,0021	0,0025	0,0062	0,0016	0,0032	0,0005	0,0009
10	0,0067	0,0137	0,0016	0,0041	0,0017	0,0030	0,0014	0,0025	0,0009	0,0011
13	0,0068	0,0106	0,0032	0,0044	0,0023	0,0030	0,0014	0,0026	0,0014	0,0014
16	0,0059	0,0121	0,0089	0,0172	0,0015	0,0018	0,0017	0,0017	0,0015	0,0015
20	0,0098	0,0144	0,0056	0,0093	0,0028	0,0027	0,0032	0,0054	0,0034	0,0030
25	0,0063	0,0084	0,0093	0,0111	0,0081	0,0136	0,0054	0,0105	0,0049	0,0033
27	0,0041	0,0045	0,0083	0,0116	0,0048	0,0052	0,0064	0,0073	0,0038	0,004
30	0,0050	0,0087	0,0055	0,0060	0,0129	0,0139	0,0096	0,0114	0,0067	0,0034
43	0,0229	0,0307	0,0099	0,0121	0,0114	0,0068	0,0195	0,0164	0,0113	0,0069

TABELA 29. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método rHE; herdabilidade simulada de 50%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0043	0,0084	0,0012	0,0019	0,0008	0,0023	0,0002	0,0004	-	0,0001
3	0,0026	0,0038	0,0008	0,0014	0,0017	0,0040	0,0001	0,0001	0,0003	0,0005
5	0,0049	0,0060	0,0047	0,0085	0,0010	0,0017	0,0008	0,0011	0,0006	0,0011
7	0,0026	0,0028	0,0017	0,0029	0,0021	0,0035	0,0007	0,0012	0,0008	0,0005
10	0,0027	0,0045	0,0022	0,0025	0,0020	0,0046	0,0018	0,0020	0,0012	0,0014
13	0,0101	0,0248	0,0023	0,0041	0,0027	0,0038	0,0022	0,0030	0,0028	0,0031
16	0,0077	0,0112	0,0064	0,0109	0,0036	0,0037	0,0032	0,0035	0,0033	0,0024
20	0,0212	0,0327	0,0086	0,0107	0,0108	0,0124	0,0070	0,0077	0,0044	0,0048
25	0,0169	0,0316	0,0157	0,0178	0,0095	0,0091	0,0118	0,0128	0,0076	0,0050
27	0,0196	0,0380	0,0104	0,0100	0,0116	0,0118	0,0116	0,0090	0,0083	0,0055
30	0,0115	0,0208	0,0164	0,0188	0,0146	0,0153	0,0131	0,0096	0,0121	0,0077
43	0,0623	0,0714	0,0410	0,0337	0,0207	0,0176	0,0243	0,0149	0,0268	0,0106

TABELA 30. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método rHE; herdabilidade simulada de 65%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0027	0,0035	0,0017	0,0054	0,0004	0,0008	0,0004	0,0006	0,0001	0,0001
3	0,0064	0,0154	0,0004	0,0004	0,0012	0,0020	0,0012	0,0022	0,0003	0,0007
5	0,0052	0,0132	0,0031	0,0066	0,0010	0,0011	0,0009	0,0011	0,0010	0,0015
7	0,0165	0,0340	0,0044	0,0065	0,0034	0,0048	0,0028	0,0030	0,0014	0,0016
10	0,0145	0,0241	0,0053	0,0087	0,0036	0,0047	0,0055	0,0082	0,0037	0,0027
13	0,0151	0,0238	0,0068	0,0101	0,0086	0,0109	0,0059	0,0071	0,0037	0,0032
16	0,0110	0,0137	0,0115	0,0124	0,0048	0,0051	0,0117	0,0160	0,0076	0,0074
20	0,0141	0,0193	0,0116	0,0173	0,0153	0,0201	0,0101	0,0073	0,0092	0,0054
25	0,0298	0,0443	0,0142	0,0156	0,0150	0,0113	0,0136	0,0092	0,0137	0,0103
27	0,0126	0,0145	0,0248	0,0246	0,0163	0,0178	0,0223	0,0151	0,0178	0,0085
30	0,0255	0,0233	0,0212	0,0223	0,0209	0,0150	0,0182	0,0139	0,0198	0,0124
43	0,0837	0,1017	0,0315	0,0257	0,0472	0,0302	0,0469	0,0216	0,0349	0,0152

TABELA 31. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método rHE; herdabilidade simulada de 80%.

Efeito do QTL	25		50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0030	0,0048	0,0015	0,0027	0,0020	0,0061	0,0004	0,0006	0,0001	0,0001
3	0,0022	0,0041	0,0014	0,0031	0,0009	0,0012	0,0006	0,0011	0,0004	0,0007
5	0,0059	0,0110	0,0020	0,0035	0,0013	0,0016	0,0015	0,0022	0,0008	0,0008
7	0,0052	0,0083	0,0058	0,0144	0,0026	0,0069	0,0016	0,0025	0,0013	0,0012
10	0,0124	0,0170	0,0057	0,0084	0,0091	0,0167	0,0021	0,0019	0,0030	0,0030
13	0,0078	0,0099	0,0057	0,0068	0,0073	0,0076	0,0065	0,0066	0,0039	0,0032
16	0,0189	0,0281	0,0086	0,0116	0,0075	0,0097	0,0106	0,0089	0,0085	0,0049
20	0,0195	0,0220	0,0185	0,0190	0,0088	0,0098	0,0107	0,0108	0,0124	0,0046
25	0,0263	0,0502	0,0173	0,0160	0,0172	0,0171	0,0142	0,0101	0,0135	0,0067
27	0,0526	0,0523	0,0238	0,0235	0,0251	0,0234	0,0245	0,0152	0,0184	0,0109
30	0,0561	0,0812	0,0148	0,0137	0,0218	0,0128	0,0207	0,0251	0,0302	0,0166
43	0,0567	0,0427	0,0687	0,0549	0,0571	0,0243	0,0406	0,0224	0,0524	0,0196

TABELA 32. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 35%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0096	0,0169	0,0049	0,0116	0,0026	0,0071	0,0006	0,0009
3	0,0062	0,0086	0,0056	0,0128	0,0035	0,0073	0,0020	0,0037
5	0,0166	0,0346	0,0090	0,0196	0,0059	0,0132	0,0022	0,0044
7	0,0152	0,0342	0,0073	0,0152	0,0073	0,0185	0,0043	0,0101
10	0,0099	0,0180	0,0129	0,0303	0,0095	0,0170	0,0100	0,0141
13	0,0204	0,0267	0,0144	0,0253	0,0096	0,0156	0,0124	0,0200
16	0,0240	0,0455	0,0166	0,0333	0,0182	0,0274	0,0127	0,0149
20	0,0322	0,0465	0,0239	0,0424	0,0263	0,0381	0,0193	0,0189
25	0,0649	0,0890	0,0309	0,0410	0,0320	0,0409	0,0298	0,0278
27	0,0650	0,0903	0,0401	0,0516	0,0339	0,0446	0,0282	0,0241
30	0,0546	0,0688	0,0439	0,0609	0,0451	0,0556	0,0335	0,0292
43	0,0779	0,1079	0,0615	0,0613	0,0773	0,0760	0,0673	0,0425

TABELA 33. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 50%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0098	0,0295	0,0042	0,0110	0,0031	0,0065	0,0013	0,0035
3	0,0121	0,0251	0,0053	0,0108	0,0041	0,0161	0,0021	0,0045
5	0,0158	0,0331	0,0079	0,0175	0,0075	0,0151	0,0042	0,0087
7	0,0171	0,0293	0,0132	0,0247	0,0099	0,0171	0,0072	0,0111
10	0,0229	0,0443	0,0142	0,0277	0,0141	0,0209	0,0099	0,0121
13	0,0218	0,0394	0,0256	0,0440	0,0165	0,0264	0,0169	0,0194
16	0,0374	0,0582	0,0387	0,0580	0,0232	0,0339	0,0217	0,0226
20	0,0595	0,0987	0,0432	0,0529	0,0436	0,0495	0,0335	0,0313
25	0,0781	0,1035	0,0643	0,0711	0,0584	0,0622	0,0442	0,0322
27	0,0654	0,0816	0,0538	0,0611	0,0690	0,0648	0,0532	0,0382
30	0,0839	0,1016	0,0694	0,0746	0,0667	0,0696	0,0661	0,0435
43	0,1525	0,1347	0,1230	0,1026	0,1444	0,0960	0,1380	0,0701

TABELA 34. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 65%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0093	0,0166	0,0051	0,0135	0,0030	0,0070	0,0015	0,0038
3	0,0121	0,0239	0,0096	0,0281	0,0063	0,0144	0,0026	0,0049
5	0,0177	0,0348	0,0122	0,0271	0,0117	0,0240	0,0059	0,0089
7	0,0224	0,0417	0,0144	0,0274	0,0144	0,0230	0,0109	0,0137
10	0,0376	0,0690	0,0232	0,0386	0,0231	0,0307	0,0178	0,0187
13	0,0418	0,0633	0,0312	0,0470	0,0253	0,0318	0,0246	0,0258
16	0,0549	0,0703	0,0469	0,0551	0,0438	0,0524	0,0350	0,0309
20	0,0746	0,0966	0,0602	0,0680	0,0541	0,0565	0,0495	0,0354
25	0,0850	0,0941	0,0783	0,0783	0,0829	0,0754	0,0747	0,0497
27	0,1014	0,1164	0,0872	0,0824	0,0977	0,0820	0,0937	0,0561
30	0,1234	0,1279	0,1177	0,0979	0,1214	0,0894	0,1113	0,0587
43	0,2175	0,1598	0,2007	0,1216	0,2043	0,1068	0,2028	0,0775

TABELA 35. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T15B15; herdabilidade simulada de 80%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0094	0,0219	0,0063	0,0194	0,0034	0,0076	0,0018	0,0044
3	0,0146	0,0295	0,0066	0,0141	0,0067	0,0150	0,0033	0,0066
5	0,0198	0,0362	0,0141	0,0337	0,0090	0,0140	0,0048	0,0073
7	0,0262	0,0522	0,0169	0,0333	0,0137	0,0245	0,0095	0,0124
10	0,0361	0,0682	0,0340	0,0559	0,0188	0,0277	0,0181	0,0194
13	0,0418	0,0612	0,0340	0,0417	0,0328	0,0397	0,0304	0,0267
16	0,0583	0,0757	0,0463	0,0574	0,0535	0,0568	0,0418	0,0326
20	0,0755	0,0927	0,0635	0,0679	0,0750	0,0701	0,0657	0,0432
25	0,1114	0,1132	0,0820	0,0773	0,0938	0,0852	0,0852	0,0529
27	0,1227	0,1314	0,1204	0,0984	0,1156	0,0834	0,1134	0,0629
30	0,1552	0,1378	0,1190	0,1033	0,1137	0,0830	0,1384	0,0671
43	0,2287	0,1561	0,2081	0,1331	0,2378	0,1207	0,2372	0,0887

TABELA 36. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 35%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0053	0,0093	0,0027	0,0065	0,0013	0,0032	0,0003	0,0007
3	0,0047	0,0066	0,0041	0,0083	0,0019	0,0038	0,0011	0,0021
5	0,0117	0,0219	0,0056	0,0112	0,0032	0,0073	0,0018	0,0033
7	0,0113	0,0249	0,0051	0,0108	0,0039	0,0089	0,0026	0,0043
10	0,0078	0,0189	0,0088	0,0188	0,0058	0,0103	0,0048	0,0066
13	0,0126	0,0177	0,0095	0,0154	0,0055	0,0087	0,0050	0,0089
16	0,0207	0,0443	0,0127	0,0252	0,0112	0,0173	0,0092	0,0106
20	0,0198	0,0324	0,0177	0,0290	0,0160	0,0214	0,0120	0,0124
25	0,0422	0,0541	0,0216	0,0283	0,0225	0,0311	0,0197	0,0171
27	0,0403	0,0553	0,0288	0,0377	0,0221	0,0259	0,0182	0,0146
30	0,0376	0,0527	0,0309	0,0403	0,0305	0,0376	0,0222	0,0190
43	0,0538	0,0743	0,0458	0,0474	0,0521	0,0489	0,0460	0,0300

TABELA 37. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 50%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0070	0,0198	0,0028	0,0056	0,0015	0,0028	0,0007	0,0018
3	0,0087	0,0174	0,0039	0,0079	0,0026	0,0067	0,0013	0,0024
5	0,0101	0,0225	0,0055	0,0121	0,0046	0,0092	0,0026	0,0054
7	0,0115	0,0218	0,0091	0,0161	0,0060	0,0106	0,0045	0,0069
10	0,0155	0,0266	0,0100	0,0188	0,0087	0,0146	0,0066	0,0081
13	0,0160	0,0307	0,0176	0,0294	0,0101	0,0163	0,0102	0,0106
16	0,0256	0,0389	0,0258	0,0374	0,0144	0,0231	0,0137	0,0139
20	0,0401	0,0654	0,0307	0,0353	0,0287	0,0332	0,0233	0,0208
25	0,0585	0,0721	0,0451	0,0488	0,0384	0,0386	0,0305	0,0223
27	0,0448	0,0549	0,0380	0,0418	0,0443	0,0421	0,0347	0,0258
30	0,0582	0,0694	0,0506	0,0544	0,0445	0,0445	0,0452	0,0297
43	0,1139	0,1023	0,0927	0,0754	0,0982	0,0650	0,0940	0,0480

TABELA 38 Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 65%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0060	0,0108	0,0031	0,0072	0,0016	0,0034	0,0008	0,0021
3	0,0089	0,0205	0,0072	0,0204	0,0035	0,0075	0,0015	0,0030
5	0,0114	0,0237	0,0081	0,0180	0,0068	0,0145	0,0036	0,0054
7	0,0159	0,0283	0,0096	0,0202	0,0092	0,0150	0,0065	0,0086
10	0,0253	0,0485	0,0155	0,0234	0,0155	0,0203	0,0115	0,0116
13	0,0285	0,0440	0,0228	0,0352	0,0164	0,0203	0,0149	0,0151
16	0,0397	0,0520	0,0328	0,0388	0,0284	0,0329	0,0236	0,0201
20	0,0550	0,0720	0,0446	0,0517	0,0366	0,0364	0,0349	0,0244
25	0,0639	0,0732	0,0555	0,0537	0,0558	0,0512	0,0514	0,0337
27	0,0774	0,0934	0,0631	0,0607	0,0653	0,0520	0,0630	0,0371
30	0,0917	0,0960	0,0884	0,0718	0,0808	0,0579	0,0743	0,0388
43	0,1624	0,1181	0,1509	0,0945	0,1474	0,0810	0,1448	0,0563

TABELA 39. Estimativas dos coeficientes de determinação (R^2) associados aos locos QTLs, dado o tamanho populacional. Método EDAC T20B20; herdabilidade simulada de 80%.

Efeito do QTL	50		75		100		200	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0,0010	0,0017	0,0039	0,0125	0,0018	0,0039	0,0009	0,0020
3	0,0013	0,0026	0,0044	0,0095	0,0039	0,0084	0,0020	0,0034
5	0,0016	0,0026	0,0087	0,0180	0,0053	0,0086	0,0031	0,0049
7	0,0024	0,0054	0,0114	0,0225	0,0079	0,0127	0,0060	0,0081
10	0,0030	0,0050	0,0241	0,0409	0,0118	0,0164	0,0118	0,0113
13	0,0042	0,0064	0,0247	0,0320	0,0225	0,0265	0,0197	0,0179
16	0,0056	0,0078	0,0333	0,0418	0,0370	0,0408	0,0276	0,0215
20	0,0085	0,0107	0,0470	0,0479	0,0488	0,0458	0,0443	0,0276
25	0,0118	0,0132	0,0577	0,0587	0,0637	0,0598	0,0594	0,0370
27	0,0892	0,0965	0,0852	0,0671	0,0807	0,0593	0,0762	0,0403
30	0,1160	0,1039	0,0886	0,0750	0,0803	0,0603	0,0952	0,0497
43	0,1784	0,1272	0,1637	0,1053	0,1704	0,0906	0,1687	0,0662

4.3. IBD – IDENTICAL BY DESCENDENT

Na Tabela 40 está apresentado os valores de IBD médios, obtidos por meio da análise dos marcadores completamente ligados aos locos QTLs, considerando 200 populações simuladas. Tais valores de IBD foram submetidos à análise de regressão OHE como variável independente. Deve-se recordar que os valores de IBD esperados para pares de irmãos completos podem assumir três valores: 0,0; 0,5 ou 1,0; e portanto, a análise de regressão OHE, bem como os métodos derivados resume-se portanto à uma regressão de três pontos. Os valores médios de IBD para todos os locos apresentados na tabela se aproximaram se aproximaram 0,50; valor esperado para progênies de irmãos completos. Resultados semelhantes – não apresentados – foram encontrados para **todos** os outros locos, nas demais características simuladas.

TABELA 40. Tabela de valores médios de IBD referente aos marcadores completamente ligados aos locos QTLs em característica simulada de herdabilidade 35%.

Efeito do QTL	25	50	75	100	200
1	0.4980	0.5020	0.4985	0.5005	0.5000
3	0.5050	0.5010	0.5010	0.5035	0.5020
5	0.4915	0.5000	0.4970	0.5015	0.5000
7	0.5025	0.4975	0.5005	0.5040	0.5005
10	0.5025	0.4985	0.5015	0.5030	0.5005
13	0.5040	0.5005	0.4985	0.5000	0.5000
16	0.5045	0.5000	0.5005	0.5015	0.5005
20	0.4950	0.4965	0.5010	0.5015	0.5000
25	0.4980	0.5005	0.5045	0.5005	0.5000
MÉDIA	0.5001	0.4996	0.5003	0.5018	0.5004
27	0.4965	0.4965	0.5045	0.5020	0.5005
30	0.5060	0.5040	0.5040	0.5015	0.5000
43	0.4945	0.4975	0.5020	0.5010	0.5000
MÉDIA	0.4990	0.4993	0.5035	0.5015	0.5002

5. CONCLUSÕES

- Os métodos de detecção de QTL, baseados na regressão de Haseman-Elston e derivados, apesar de desenvolvidos para análise de pares de irmãos, necessitam de elevados tamanhos populacionais para detecção de genes/QTLs de efeitos moderados a elevados.
- A metodologia revisited Haseman-Elston (rHE), foi a que apresentou maior poder estatístico apresentando potencial para detecção de genes de efeito maior – em populações de tamanho superior a 100 indivíduos – para características fenotípicas de alta herdabilidade (superior a 35%).
- Os métodos de pares de irmãos avaliados mostraram-se pouco eficientes no estudo de características quantitativas, embora apresentem potencial para estudo de características oligogênicas
- Os métodos de pares de irmãos avaliados apresentaram potencialidade para estudo de características oligogênicas. Os métodos apresentaram, porém, pouco eficientes na detecção de genes de baixo efeito.

6. BIBLIOGRAFIA

- ALPERT, K.B. et al. (1995) fw-2.2: A major QTL controlling fruit weight is common to both red-fruited and green-fruited tomato species. *Theor. Appl. Genet.* 91, 994–1000.
- AMOS, C.I., DAWSON, D.V., ELSTON, R.C. (1990) The probabilistic determination of Identity-by-descent sharing. *Am J Hum Genet* 47:842–853.
- CORDELL, H.J. (2004) Bias toward the null hypothesis in model-free linkage analysis is highly dependent on the test statistic used. *Am J Hum Genet* 74:1294–1302.
- CRUZ, C.D. (2008) GQMOL: Programa para análise de dados moleculares e quantitativos. Universidade Federal de Viçosa, Viçosa, MG, Brasil. <http://www.ufv.br/dbg>
- DARVASI, A. (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genet.* 18(1), 19–24.
- EL-DIN EL-ASSAL, S. et al. (2001) A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. *Nat. Genet.* 29, 435–440
- ELSTON, R.C., BUXBAUM, S., JACOBS, J.B., OLSON, J.M. (2000) Haseman and Elston Revisited. *Genetic Epidemiology*, 19:1–17.
- FEINGOLD, E. (2002) Invited Editorial. Regression-Based Quantitative-Trait–Locus Mapping in the 21st Century. *Am. J. Hum. Genet.* 71:217–222.
- FEITOSA, M., KRIEGER, H. (2002) O futuro da epidemiologia genética de características complexas. *Ciência e Saúde Coletiva*, v. 7, p. 71-83, 2002.
- FERREIRA, M.A.R. (2004) Linkage Analysis: Principles and Methods for the Analysis of Human Quantitative Traits. *Twin Research*. V. 7 N. 5. 513–530.
- FULKER, D.W., CHERNY, S.S. (1996) An improved multipoint sib-pair analysis of quantitative traits. *Behav. Genet.* 26:527–32.
- GELDERMAN, H. (1975) Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor. Appl. Genet.* 46, 319–330.
- GRAY-MCGUIRE, C., et al. (2010) Confirmation of Linkage to and Localization of Familial Colon Cancer Risk Haplotype on Chromosome 9q22. *Cancer Res*; 70(13) July 1, 5409 - 5419. DOI:10.1158/0008-5472.CAN-10-0188

- HASEMAN, J.K., ELSTON, R.C. (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* 54:1092-1103.
- KRUGLYAK, L., DALY, M.J., LANDER, E.S. (1995) Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. *Am J Hum Genet* 56:519–527.
- KU, H.M. et al. (1999) The genetic basis of pear-shaped tomato fruit. *Theor. Appl. Genet.* 99, 844–850
- LANDER, E., KRUGLYAK, L. (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454.
- LARKIN, E.K., PATEL, S.R., ELSTON, R.C., GRAY-MCGUIRE, C., ZHU, X., REDLINE, S. (2008) Using Linkage Analysis to Identify Quantitative Trait Loci for Sleep Apnea in Relationship to Body Mass Index. *Ann Hum Genet.* 72(Pt 6): 762–773. doi:10.1111/j.1469-1809.2008.00472.x.
- LUO, Z.W. et al. (2002) Precision and high-resolution mapping of quantitative trait loci by use of recurrent selection, backcross or intercross schemes. *Genetics* 161, 915–929.
- MACKAY, T.F.C., STONE, E.A., AYROLES, J.F. (2009) The genetics of quantitative traits: Challenges and prospects. *Nature Genet.* V. 10, 565–577.
- MARTINEZ, M.L.; VUKASINOVIC, N. (2000) Algorithm to calculate proportion of genes identity by descent to map QTL in half-sibs families. *Rev. Bras. Zootec.* 29(2443-451).
- MISSIAGGIA, A.A., PIACEZZI, A.L., GRATTAPAGLIA, D. (2005) Genetic mapping of *Eef1*, a major effect QTL for early flowering in *Eucalyptus grandis*. V. 1, n. 2, 79-84, DOI: 10.1007/s11295-005-0011-3
- PELEMAN, J.D. et al. (2005) Quantitative trait locus (QTL) isogenic recombinant analysis: a method for high-resolution mapping of QTL within a single population. *Genetics* 171, 1341–1352.
- POWELL, J.E., VISSCHER, P.M., GODDARD, M.E. (2010) Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews | Genetics – Perspectives.* V.11, 800-805.
- PRICE, A.H. (2006) Believe it or not, QTLs are accurate! *TRENDS in Plant Science*, Vol.11 No.5., 213-216.

- PURCELL, S., CHERNY, S.S., HEWITT, J.K., SHAM, P.C. (2001) Optimal Sibship Selection for Genotyping in Quantitative Trait Locus Linkage Analysis. *Hum Hered*; 52:1–13.
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- ROSADO, T.B., TOMAZ, R.S., FERNANDES, L.S., ROSADO, A.M., ARAÚJO, E.F., CRUZ, C.D. Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic *Eucalytus* spp. Populations. *Crop Breeding and Applied Biotechnology*; v. 10; p. 321-328; 2010.
- SCHELLING, J.R., et al. (2008) Genome-Wide Scan for Estimated Glomerular Filtration Rate in Multi-Ethnic Diabetic Populations The Family Investigation of Nephropathy and Diabetes (FIND). *DIABETES*, VOL. 57, 235-243.
- SCHUSTER, I., CRUZ, C.D. (2004) Estatística genômica aplicada a populações derivadas de cruzamentos controlados. Viçosa: UFV, 585 p.
- SEMAGN, K.; BJØRNSTAD Å. and XU Y. The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology*, September 2010, vol. 13, no. 5. <http://dx.doi.org/10.2225/vol13-issue5-fulltext-14>
- SHALOM, A., DARVASI, A. (2002), “Experimental Designs for QTL Fine Mapping in Rodents”, in CAMP, N.J., COX, A. (orgs), *Quantitative Trait Loci: Methods and Protocols*. New Jersey: Humana Press, 199-223.
- SHEN, G.Q., et al. (2007) An LRP8 Variant Is Associated with Familial and Premature Coronary Artery Disease and Myocardial Infarction. *The American Journal of Human Genetics*. V. 81, 780 – 791.
- SOBEL, E., LANGE, K. (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 58:1323–1337.
- STEIN, C.M., ELSTON, R.C. (2009) Finding genes underlying human disease. *Clin Genet.*; 75(2): 101–106. doi:10.1111/j.1399-0004.2008.01083.x.
- STEIN, C.M., et al. (2007) Linkage and association analysis of candidate genes for TB and TNF α cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. *Hum Genet*. 121:663–673. DOI 10.1007/s00439-007-0357-8.

- TANKSLEY, S.D. (2004). The Genetic, Developmental, and Molecular Bases of Fruit Size and Shape Variation in Tomato. *The Plant Cell*, Vol. 16, S181–S189.
- THODAY, J.M. (1961) Location of Polygenes. *Nature* 191, 368 - 370 (22 July 1961); doi:10.1038/191368a0
- THOMPSON, E.A., HEATH, S.C. (1998) Estimation of conditional multilocus gene identity among relatives. IMS Lecture Note Series In Press.
- VAGUJFALVI, A. et al. (2003). The cold-regulated transcriptional activator *Cbf3* is linked to the frost-tolerance locus *Fr-A2* on wheat chromosome 5A. *Mol. Genet. Genomics* 269, 60–67.
- WAN, Y., COHEN, J.C., GUERRA, R. (1997) A permutation test for the robust sibpair method. *Ann Hum Genet.* 61:7987.
- WANG, D., LIN, S., CHENG, R., GAO, X., WRIGHT, F.A. (2001) Transformation of sib-pair values for the Haseman-Elston method. *Am J Hum Genet* 2001; 68:1238–1249.
- WANG, T., ELSTON, R.C. (2004) A modified Revisited Hanseman-Elston method to further Improve Power. *Human Heredity.* 57:109-116.
- WANG, T., ELSTON, R.C. (2007) “Regression-based Linkage Analysis Methods”. In DENG, H.W., SHEN, H., LIU, Y.J., HU, H. (orgs), CURRENT TOPICS IN HUMAN GENETICS - Studies in Complex Diseases. World Scientific Publishing Co. Pte. Ltd.
- WERNER, J.D. et al. (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2460–2465.
- YAMANAKA, N. et al. (2005) Fine mapping of the FT1 locus for soybean flowering time using a residual heterozygous line derived from a recombinant inbred line. *Theor. Appl. Genet.* 110, 634–639.

7. ANEXOS

ANEXO 1. Algoritmo para detecção de QTL por meio da metodologia de marca simples, Regressão de HASEMAN-ELSTON (1972), implementado para o ambiente do software livre R, segundo instruções descritas na seção Materiais e Métodos. Comentários estão precedidos pelo símbolo sharp (#).

```
# INSTRUÇÕES para entrada e leitura dos arquivos
prefixo <- "quant50-"           # dados dos arquivos
sufixo <- ".txt"                # dados dos arquivos
inicio <- 1                     # dados dos arquivos
iteracoes <- 200               # dados dos arquivos
marcadores <- 84                # número de marcadores
col <- 52                       # número de colunas dos arquivos

#####

resultado <- matrix(c(0), nrow=marcadores, ncol = 10)
colnames(resultado)<-c("Marcador","IBD","Beta_0","Beta_1","R2","F","p.value","H2","sigma2.a","n*")
w <- inicio
R2.print <- c(NULL)
H2.print <- c(NULL)
sigma.a.print <- c(NULL)
n.print <- c(NULL)

#####

for(w in 1:iteracoes){        # inicio do processo de avaliação das simulações
arquivo <- paste(prefixo,w,sufixo, sep="")
arquivo
dados <- read.table(arquivo, nrow = marcadores, header = FALSE)
quant <- read.table(arquivo, nrow = 1, header = FALSE, skip = marcadores)

#####

# CRIAÇÃO DO VETOR Y          # MÉTODO DE oHE

i <- 3  # primeiro dado a ser lido / terceira coluna
k <- i+1 # primeiro dado a ser comparado

vetory <- c(NULL)
for (i in 3:(col-1)){
  for (k in (i+1):col){
    y <- (quant[1,i]-quant[1,k])
    vetory <- c(vetory, y)
  }
  k <- i+1
}
rm(y)
vetory <- vetory*vetory
```

```

#####

# Variância Fenotípica
quant<- (quant[-c(1,2)])
media <- sum(quant)/(col-2)          # média
var.fen <- sum((quant-media)^2)/(col-3)

# ALGORITMO PARA COMPARAÇÃO E CRIAÇÃO DO IBD PARA VÁRIOS MARCADORES
i <- 3          # primeiro dado a ser lido / terceira coluna
k <- i+1      # dado a ser comparado
j <- 1        # linha

IBD <- c(NULL)
result <- c(NULL)

for (j in 1:marcadores){
  i <- 3
  k <- 4
  for (i in 3:(col-1)){
    k <- i+1
    for (k in (i+1):col){
# CRIAÇÃO DO IBD
      if (dados[j,1]==12 & dados[j,2]==34){      #Cruzamento tipo 7
        if(dados[j,i]==13 & dados[j,k]==13){IBD <- c(IBD, 0, 0,1,1)}
        else if(dados[j,i]==13 & dados[j,k]==14){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==13 & dados[j,k]==23){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==13 & dados[j,k]==24){IBD <- c(IBD,1,0,0,0)}
        else if(dados[j,i]==14 & dados[j,k]==13){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==14 & dados[j,k]==14){IBD <- c(IBD,0,0,1,1)}
        else if(dados[j,i]==14 & dados[j,k]==23){IBD <- c(IBD,1,0,0,0)}
        else if(dados[j,i]==14 & dados[j,k]==24){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==23 & dados[j,k]==13){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==23 & dados[j,k]==14){IBD <- c(IBD,1,0,0,0)}
        else if(dados[j,i]==23 & dados[j,k]==23){IBD <- c(IBD,0,0, 1,1)}
        else if(dados[j,i]==23 & dados[j,k]==24){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==24 & dados[j,k]==13){IBD <- c(IBD,1,0, 0,0)}
        else if(dados[j,i]==24 & dados[j,k]==14){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==24 & dados[j,k]==23){IBD <- c(IBD,0,1,0,0.5)}
        else if(dados[j,i]==24 & dados[j,k]==24){IBD <- c(IBD,0,0, 1,1)}
      }
    }
  }
}

i <- 3
k <- 4

IBD <- matrix(IBD, nrow = choose(col-2,2), ncol = 4, byrow=TRUE)
# CÁLCULO DAS ESTATÍSTICAS
result <- lm(formula= vetory ~ IBD[,4])
resultado[j,1] <- j          # Marcador
resultado[j,2] <- round(mean(IBD[,4]),2)      # IBD MÉDIO
resultado[j,3] <- round(result$coefficients[1],4)  # BETA 0
resultado[j,4] <- round(result$coefficients[2],4)  # BETA 1
result <- anova(lm(formula= vetory ~ IBD[,4]))      # ANOVA DA REGRESSÃO
resultado[j,5] <- round(result$Sum[1]/(result$Sum[1]+result$Sum[2]),4)  # R2
resultado[j,6] <- round(result$F[1],6)          # VALOR DE F
resultado[j,7] <- round(result$Pr[1],6)        # p.value
resultado[j,8] <- round(-resultado[j,4]/(2*var.fen),6)  # HERDABILIDADE

```



```

resultado[j,9] <- round(-resultado[j,4]/2,4)           # SIGMA2a

if(resultado[j,7]<alfa){resultado[j,10]<-resultado[j,10]+1} # valor de alfa a ser especificado.

#####
arquivo2 <- paste("resultado-",w,sufixo, sep="")
arquivo2
write.table(resultado, arquivo2)

IBD <- NULL
result <- NULL
}
rm(i,j,k)
#####

```

ANEXO 2. Algoritmo para criação do vetor de observações Y para metodologia Revisted Haseman e Elston (rHE), implementado para o ambiente do software livre R. Comentários estão precedidos pelo símbolo sharp (#).

```

# CRIAÇÃO DO VETOR Y           # MÉTODO DE rHE

i <- 3   # primeiro dado a ser lido / terceira coluna
k <- i+1
media <- sum(quant[-c(1,2)])/(col-2)
vetory <- c(NULL)
for (i in 3:(col-1)){
  for (k in (i+1):col){
    y <- (quant[1,i]-media)*(quant[1,k]-media)
    vetory <- c(vetory, y)
  }
  k <- i+1
}
rm(y)
vetory           # CRIA O VETOR DE DADOS FENOTÍPICOS DE ACORDO COM rHE

```