

CAIO CÉSIO SALGADO

**ABORDAGEM SOBRE ESTUDO DE HERANÇA DE CARACTERES PELA  
GENÉTICA CLÁSSICA E MOLECULAR**

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

S164a  
2012

Salgado, Caio César, 1980-  
Abordagem sobre estudo de herança de caracteres pela  
genética clássica e molecular / Caio César Salgado. – Viçosa,  
MG, 2012.  
viii, 90f. : il. ; 29cm.

Orientador: Cosme Damião Cruz  
Tese (doutorado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. Genética molecular - Testes. 2. Genética quantitativa -  
Testes. 3. População biológica - Métodos de simulação.  
4. Genômica. I. Universidade Federal de Viçosa. II. Título.

CDD 22. ed. 572.8

CAIO CÉSIO SALGADO

**ABORDAGEM SOBRE ESTUDO DE HERANÇA DE CARACTERES PELA  
GENÉTICA CLÁSSICA E MOLECULAR**

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: 03 de agosto de 2012.

---

Paulo Roberto Cecon

---

Eveline Teixeira Caixeta

---

Leonardo Lopes Bhering

---

Marciane da Silva Oliveira

---

Cosme Damião Cruz

**“Devemos fazer tudo como se tudo dependesse de nós; e esperar tudo  
como se tudo dependesse de Deus”**

**Santo Inácio de Loyola**

## AGRADECIMENTOS

À Deus e seus anjos, por sempre se fazerem presentes me iluminando todos os dias.

Aos meus pais, Antonia Bressani Salgado e Fernando Amir Salgado e meu irmão Rider Walber Salgado, pelo apoio e compreensão em todos os momentos.

A minha tia Neomisa, pela amizade, boas conversas e incentivo no caminho do bem.

Ao tio Tenir, que mesmo nos encontrando pouco, sempre presente em minha vida.

À Lívia Meirelles, companheira de todas as horas, pelo amor, paciência, carinho e acreditar no meu potencial.

Ao professor Cosme Damião Cruz, pela amizade, paciência, confiança e incentivo, um exemplo a ser seguido.

À Universidade Federal de Viçosa, pela oportunidade de realização deste curso.

Aos amigos do laboratório de Bioinformática, Marciane, Gislayne, Moysés, Leonardo, Dário, Gilson, Isis, Felipe, Rafael, Danielle, Eliel, Jaqueline e Tales pelo convívio agradável durante a realização deste curso.

Aos amigos e alunos de iniciação científica Vinícius, Lívia, Luiza, Rafael Mauri e Letícia.

Aos grandes amigos Lucas e Flávia, que mesmo distantes, estão sempre presentes.

Aos amigos de república Felipe, Rafael, Dalcionei e Fábio pela convivência e companheirismo.

Ao Gustavo, pela grande amizade, companheirismo, paciência e apoio aos meus ideais.

Aos amigos do curso de pós-graduação pela parceria e afinidade: Camila, Ana Maria, André, Lívia Tomé, Éder, Kátia.

Aos amigos de outras datas que me apoiaram e me incentivaram nesta empreitada: Lisete, Soraya, e Zuy.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudo.

Aos professores de graduação e de pós-graduação, pela atenção, pela disponibilidade e pelos ensinamentos.

Às secretárias do curso de pós-graduação em genética e melhoramento, Edna e Rita, pelo apoio, dedicação, atenção e amizade.

A mãe Joaquina, aonde quer que esteja, pelo carinho, amor e amizade que sempre teve comigo.

Meus sinceros agradecimentos!!!

**Pode ser que um dia nos afastemos...  
Mas, se formos amigos de verdade,  
A amizade nos reaproximará.**

Albert Einstein

## SUMÁRIO

RESUMO .....	vii
ABSTRACT .....	viii
INTRODUÇÃO GERAL .....	1
REVISÃO DE LITERATURA .....	4
Utilização de marcadores moleculares no melhoramento de plantas.....	13
Simulação .....	15
Hipóteses Genéticas .....	18
Método do mapeamento por intervalo simples .....	22
Aplicativos utilizados para simulação.....	23
Referências .....	25
<b>CAPÍTULO 1 .....</b>	<b>32</b>
<b>ABORDAGEM SOBRE ESTUDO DE HERANÇA DE CARATERES PELA GENÉTICA CLÁSSICA E MOLECULAR PARA PADRÃO BINÁRIO DE EXPRESSÃO FENOTÍPICA .....</b>	<b>32</b>
Introdução.....	34
Material de Métodos.....	38
Resultados e discussão.....	43
Conclusões.....	51
Referências .....	52
<b>CAPÍTULO II .....</b>	<b>60</b>
<b>ABORDAGEM SOBRE ESTUDO DE HERANÇA DE CARATERES PELA GENÉTICA CLÁSSICA E MOLECULAR PARA PADRÃO MULTICATEGÓRICO DE EXPRESSÃO FENOTÍPICA .....</b>	<b>60</b>
Introdução.....	62
Material de Métodos.....	65
Resultados e Discussão.....	71
Conclusões.....	80
Referências .....	82

## RESUMO

SALGADO, Caio César, D.Sc. Universidade Federal de Viçosa, agosto de 2012. **Abordagem sobre estudo de herança de caracteres pela genética clássica e molecular.** Orientador: Cosme Damião Cruz. Coorientadores: Eveline Teixeira Caixeta e Leonardo Lopes Bhering.

O trabalho teve por objetivo caracterizar hipóteses genéticas de maior similaridade utilizadas para prever padrão de herança de características oligogênicas. Foi utilizado o delineamento genético clássico, envolvendo população  $F_2$  derivada de  $F_1$  proveniente de cruzamento entre pais homocigotos contrastantes, de característica controlada por um, dois e três genes independentes e ligados. A este estudo foi também considerada a agregação de informações moleculares de indivíduos  $F_2$  possibilitando o emprego de análises genômicas para fins de complementação de informações e elucidação de testes de hipóteses no caso de existir segregação ambígua em hipóteses concorrentes. Foram simulados genomas parentais e populações  $F_2$  de 100, 200, 400 e 600 indivíduos, fenotipados e genotipados em relação a cinco grupos de ligação contendo 11 marcadores co-dominantes por grupo de ligação, e saturação de 10 cM entre marcadores. Às populações simuladas foram acrescentadas características fenotípicas simuladas representando um padrão de herança binário de presença e ausência (0 e 1) e um padrão multicategórico representado por notas que variavam de zero a cinco. Testes de genética clássica, baseados em qui-quadrado, foram incapazes de identificar corretamente o padrão de herança de características controladas por genes ligados e proporcionaram grande percentual de erro, principalmente em tamanhos populacionais mais reduzidos, em teste com hipóteses com segregações concorrentes próximas tais como 3:1 e 13:3, 9:7 e 37:27 e 1:2:1 e 3:9:4, para características fenotípicas binárias e multicategóricas. As análises genômicas foram realizadas pelas metodologias de intervalo simples aplicadas aos estudos de detecção de QTL (quantitative trait loci) demonstrando seu potencial na identificação dos genes, posicionamento e quantificação de seus efeitos sobre a expressão da característica.

## ABSTRACT

SALGADO, Caio César, D.Sc. Universidade Federal de Viçosa, August, 2012. **Approach to study the inheritance of characters by classical genetics and molecular.** Adviser: Cosme Damião Cruz. Co-Advisers: Eveline Teixeira Caixeta and Leonardo Lopes Bhering

The study aimed to characterize genetic hypothesis of greater similarity used to predict the pattern of inheritance of characteristics oligogênicas. We used classical genetic design involving F1, F2 population derived from crosses between homozygous parents contrasting trait controlled by one, two and three independent genes and linked. In this study, we also considered the aggregation of molecular information from the F2 allowing the use of genomic analysis for completeness of information and clarification of hypothesis testing where there is segregation ambiguous competing hypotheses. Parental genomes were simulated and F2 populations 100, 200, 400 and 600 individuals, phenotyped and genotyped for five linking groups containing 11 co-dominant marker by linking group, and overrun of 10 cm between markers. Simulated populations were added to the phenotypic characteristics simulated representing a binary inheritance pattern of presence and absence (0 and 1) and multicategorico represented by a standard scale ranging from zero to five. Test of classical genetics, based on chi-square, were unable to correctly identify the pattern of inheritance of traits controlled by genes linked and provided a large percentage of error, especially in smaller population sizes in test hypotheses with close competitors such as segregation 3:1 and 13:3, 9:7 and 37:27, and 1:2:1 and 3:9:4 for binary and phenotypic characteristics multicategoric. The genomic analyzes were performed by simple interval methods applied to studies of detection of QTL (quantitative trait loci) showing its potential in identifying genes, positioning and quantification of their effects on the expression of the trait.

## INTRODUÇÃO GERAL

Uma definição amplamente conhecida de melhoramento de plantas foi a de Fehr (1987) que o definiu como a arte e a ciência de modificar as plantas a fim de torná-las mais úteis ao homem. Durante o século XX, a população do planeta passou de 1,5 milhão de habitantes para 6,2 bilhões. Esse espetacular incremento exerceu forte pressão sobre a necessidade de aumentar a produtividade das principais espécies cultivadas. Efetivamente este fato ocorreu e foi uma das maiores contribuições da ciência para a sociedade. Projeções da ONU mostram que a população mundial deve alcançar 8,2 bilhões de pessoas em 2030 e 9,1 bilhões de pessoas em 2050. Esse crescimento populacional deve ser concentrado principalmente nos países menos desenvolvidos da Ásia, África e América Latina. Esses dados mostram um grande desafio que as nações, principalmente as menos desenvolvidas, terão de enfrentar nos próximos anos: aumentar a produção de alimentos para que não haja fome no mundo.

Devido ao melhoramento, seja ele animal ou vegetal, os níveis de produtividade atingido nos mais diversos campos da agricultura são surpreendentes. Para que o melhoramento de plantas seja cada vez mais eficiente, as diferenças a serem detectadas exigem que os experimentos sejam cada vez mais precisos e os conhecimentos pontuais, gerados em diversos setores das ciências biológicas, se tornam cada vez mais importantes para geração de tecnologia de análise de dados permitindo maior acurácia nos programas de melhoramento e melhoria efetiva da produtividade.

O que se observa é que modernas ferramentas de análises genéticas não tem sido aplicadas efetivamente ou contemplada a possibilidade de seu uso durante o

estabelecimento e a condução de experimentos. Como a genômica tem uma abordagem ampla na genética, o conhecimento dos processos e as descobertas científicas, que culminaram em uma ferramenta que permite maior compreensão sobre as bases químicas da herança, são de fundamental importância para o entendimento das diversas aplicações da genética quantitativa na era da genômica e suas implicações para o melhoramento de plantas.

Na determinação da herança, um método utilizado para caracteres qualitativos se baseia nas leis mendelianas, onde se avalia as proporções fenotípicas em  $F_2$  e retrocruzamentos. Um dos pontos mais importantes para o estudo das características da distribuição discreta é o teste de hipótese genética que permite, com uma margem de erro, concluir a respeito do padrão predominante da segregação: se governada por um, dois ou mais genes e o tipo de interação predominante (LIU, 1997).

A consideração inapropriada de uma hipótese genética pode resultar em problemas de interpretação. O equívoco nas conclusões em relação à natureza da herança da característica muitas vezes devido ao tamanho da amostra sendo comum encontrar na literatura populações avaliadas com tamanho amostral não suficiente para dar respaldo a veracidade do tipo de segregação, ao genótipo de genitores, direcionamento de cruzamentos e posicionamento dos genes nos mapas de ligação apresenta potencial para causar prejuízos na continuidade da pesquisa.

A análise de experimentos clássicos como teste de segregação aliado a análise genômica poderia solucionar definitivamente o problema de erro tipo II para características oligogênicas, trazendo maior confiabilidade sobre o número de genes que controlam a herança da característica principalmente no que se refere a resistência de plantas a pragas e doenças.

## **Objetivo Geral**

O objetivo deste trabalho é caracterizar as hipóteses genéticas de maior similaridade utilizadas para prever o padrão de herança de características oligogênicas em experimentos envolvendo o delineamento de gerações P<sub>1</sub>, P<sub>2</sub>, F<sub>1</sub> e F<sub>2</sub> e fazer uso efetivo da genômica no melhoramento de plantas

## **Objetivos específicos**

-Avaliar erros de rejeição de hipóteses verdadeira, em particular com hipóteses concorrentes próximas tais como 3:1 e 13:3 ou 1:2:1 e 3:9:4, por exemplo, serão analisados e discutidos.

-Avaliar a contribuição da análise genômica, com a agregação de informações da genotipagem da F<sub>2</sub>, na elucidação do padrão de segregação além de prover informações adicionais sobre posicionamento e quantificação do efeito de locos individuais sobre a expressão da característica.

## REVISÃO DE LITERATURA

Logo após as redescobertas dos trabalhos de Mendel, no final do século XIX, foram realizados inúmeras pesquisas visando explicar a herança de vários caracteres nas mais diversas espécies de plantas e animais. Em 1902, W. Bateson, E.R. Saunders e R.C. Punnet, demonstraram em ervilha-doce, que as segregações dos caracteres cor da flor e formato do pólen não ocorriam de forma independente (Lander e Weinberg, 2000). T.H.Morgan e colaboradores trabalhando com *Drosophila melanogaster* ao analisarem o padrão de herança de um gene mutante ligado ao sexo, para cor dos olhos observaram que alguns caracteres não segregavam de acordo com a segunda lei de Mendel e forneceram a primeira evidência de que os genes estão localizados em posições definidas nos cromossomos e que podem ser manipulados e avaliados experimentalmente. Morgan sugeriu que alguns genes estariam situados no mesmo cromossomo e que durante a meiose, ocasionalmente, ocorreriam, entre os homólogos, trocas de segmentos denominados crossing-over ou permuta.

Em 1913, A.H.Sturtevant, interpretando dados oriundos da segregação de genes ligados, sugeriu o uso da porcentagem de recombinantes como indicador quantitativo da distância linear entre dois genes na construção de mapas genéticos. Os mapas mostravam que a posição dos genes correspondiam à sua ordem linear nos cromossomos. Assim, o conceito de localização dos genes em uma ordem linear passou a ser incorporado à teoria cromossômica da herança (Griffiths et al, 1998).

Grandes debates começaram a surgir se o modelo genético de Gregor Mendel poderia ser aplicado às características quantitativas. Entre os primeiros pesquisadores a fornecer evidência afirmativa para tal acontecimento destacava-se estava Edward East (1916), que trabalhava com populações de tabaco de flor longa (*Nicotiana longiflora*). A característica que ele estudava era o comprimento da corola, a parte da flor formada

pela pétalas. No tabaco de flor longa, a corola tem a forma de um tubo, o que levou East a 2 predições; Primeiro, a menos que se faça o cruzamento e a mensuração de milhares de plantas, a amplitude de variação que veremos nas  $F_2$ s não se estenderá até os genótipos dos parentais originais e a segunda predição de East era de que dentro de poucas gerações de cruzamento seletivo para corolas curtas ou longas, seremos capazes de recuperar os fenótipos dos pais.

Quando esse pesquisador cruzou genitores com flores curtas e longas, foram produzidas  $F_1$ s com flores médias. Quando deixou que as  $F_1$ s se autofecundassem, foram produzidas  $F_2$ s que mostraram maior variação fenotípica do que as  $F_1$ s. Contudo, como East examinou somente 454 plantas da  $F_2$ , não milhares de plantas, não encontrou  $F_2$ s com fenótipos próximos aos extremos da geração parental. Finalmente, partindo de plantas da  $F_2$ , East cruzou-as seletivamente para corolas curtas e corolas longas. Na época em que chegou a geração  $F_5$ , a maioria das plantas de suas linhagens selecionadas tinha comprimentos de corola dentro das amplitudes dos genitores originais. Os dados de East confirmam sua predição. Seu experimento, além de outros semelhantes como o de Yule e Nilsson-Ehle, estabeleceu que as características quantitativas são determinadas pela influência conjunta de alelos mendelianos de muitos locos.

A recuperação de caracteres quantitativos exatamente igual ao da geração parental nem sempre acontecia, foi notado que existia uma variação fenotípica até mesmo entre plantas geneticamente idênticas. A razão disso é que cada planta mesmo no jardim experimental de East, estava exposta a um ambiente exclusivo. Wilhelm Johanssen, mostrou que o fenótipo de um caráter quantitativo era devido à ação conjunta do genótipo mais o efeito do ambiente. Foi também este pesquisador que atribuiu o termo gene a fatores mendelianos.

A reconciliação entre a genética mendeliana com a descrição biométrica da variação contínua em populações reais foi obtida por vários autores em muitos estágios, mas neste contexto, um artigo publicado em 1918, de R.A.Fisher, foi particularmente importante. Fisher demonstrou que os resultados alcançados pelos biometristas, poderiam ser explicados pelos princípios mendelianos e neste momento, surgia a genética quantitativa propriamente dita.

A publicação de Fisher fazia referencia à covariância e à correlação genética entre indivíduos aparentados. A variância genética foi decomposta em variância aditiva, devida aos efeitos médios dos genes, a variância de dominância, devida aos efeitos das interações intralélicas e a variância epistática. Após este trabalho, surgiram muitos outros, como o de Wright (1927), Haldane (1932), Cockermam (1954, 1956), Kempthorne (1954). O advento da genética quantitativa possibilitou avaliar as vantagens e as deficiências dos métodos de melhoramento já conhecidos, bem como propor alternativas de melhorar os métodos de melhoramento aplicados a plantas autógama e alógamas já existentes e criar novos métodos. Vários trabalhos mostram o progresso genético obtido em várias espécies como em Wilcox, 2001 ; Bernzonsky & Lafever, 1993, grande parte deste progresso foi devido à utilização da heterose e ciclos contínuos seletivos submetidos as espécies sob seleção.

Paralelamente a estes trabalhos a natureza química do DNA começava a ser elucidada e as bases para a genômica começavam a ser construídas. Griffith, por volta de 1928 mostrava que devia existir “princípio transformante” que atuava nas bactérias S transformando em R virulentas. Avery, MacLeod e McCarty em 1944 demonstravam que a substancia responsável pela transformação destas bactérias era o DNA. Em 1953 Watson e Crick propuseram a estrutura molecular do DNA, com base nos trabalhos de

Chargaff sobre composição das bases do DNA de diferentes espécies e, também, os resultados dos estudos de difração de raios X feitos por Franklin e Wilkins.

Dentre as principais dificuldades encontradas para a confecção dos primeiros mapas genéticos, pode-se destacar a limitação em se obter marcadores genéticos consistentes e adequados para a análise de ligação. Embora os marcadores morfológicos tenham possibilitado o desenvolvimento de mapas genéticos no início do século passado, tais marcadores não apresentam qualidades esperadas para este tipo de abordagem, apresentando baixo nível de polimorfismo, pouca estabilidade ambiental e número limitado de locos disponíveis para estudos de mapeamento (Ferreira e Grattapaglia, 1995).

Mapas genéticos se tornaram rapidamente uma poderosa ferramenta para os geneticistas uma vez que são de suma importância para o programa de melhoramento. O estudo de genomas inteiros, com a utilização de marcadores genéticos na análise de caracteres quantitativos, teve início a partir do século XX, quando foram estabelecidos os primeiros mapas genéticos intra-específicos em *Drosophila melanogaster* e em *Phaseolus vulgaris* (SAX, 1923).

Zimogramas começaram ser utilizados para representação da variabilidade alélica. Hunter e Market, em 1957, adaptaram à eletroforese métodos histoquímicos de coloração, para localizar as zonas de atividade enzimática diretamente no meio suporte. Linn e Arber(1968) descobriram as enzimas de restrição que cortam o DNA em pontos específicos fornecendo as respaldo para, em 1973, Stanley, Cohen e Brown isolarem genes e introduzi-los em outro organismo via *Agrobacterium*. A manipulação do material genético começou efetivamente em 1973, na Universidade de Stanford, nos EUA, quando o Dr. Paul Berg conseguiu unir sequência de DNA de *Escherichia coli* à

do vírus Simian papilona. Esse feito pode ser considerado o marco zero da denominada ‘era genômica’ ( Ramalho e Lambert 2004).

Começava, efetivamente, a corrida para decifrar o código genético da vida, o livro que abriria as portas aos segredos mais íntimos sobre nossa espécie. A partir da década de 80, marcadores genéticos baseados na análise direta de seqüências polimórficas de DNA foram sendo implementados e assim, o mapeamento tornou-se ilimitados a todas a espécies. Diversos tipos de marcadores polimorficos de DNA foram desenvolvidos como RAPD<sup>1</sup>, AFLP<sup>2</sup>, RFLP<sup>3</sup>, SSR<sup>4</sup> e SNP<sup>5</sup>. A utilização destes marcadores na construção de mapas genéticos difere apenas em alguns aspectos: número de locos que pode ser detectado, graus de polimorfismo entre e dentro de acessos e características de dominância (Maliepaard et al., 1997). Aspectos como custos, tempo necessário para a realização das avaliações, e dificuldades práticas inerentes à execução de cada técnica variam de acordo com a técnica empregada, ou seja de acordo com o conteúdo informativo da técnica empregada: marcadores co-dominantes (RFLPs e SSRs) ou dominantes (RAPSs e AFLPs) (Coelho 2000).

Fragmentos mapeados de grupos de ligação podem também servir como sondas para trabalhos com hibridização *in situ* em metáfases mitóticas, contribuindo para a identificação grupos de ligação e cromossomo, é possível ainda associar marcas mapeadas a genes, simplesmente pelo sequenciamento dos fragmenos mapeados, os quais servem como ponto de partida para a comparação com seqüências depositadas nos bancos, facilitando a identificação e clonagem destes genes. Bancos de dados genômicos foram construídos a partir de sequenciamento de nucleotídeos. Atualmente inúmeros projetos genomas estão em andamento em todo o mundo gerando um enorme

---

<sup>1</sup> Random Amplified Polymorphic DNA

<sup>2</sup> Amplified Fragment Length Polymorphic

<sup>3</sup> Restriction Fragment Length Polymorphic

<sup>4</sup> Simple Sequence Repeat

<sup>5</sup> Single Nucleotide Polymorphisms

volume de seqüências de DNA, as quais estão depositadas em bancos de dados públicos como o National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>) ou privados. Um grande número de mapas genéticos prontos esta disponível na internet no SOL Genomics Network (SGN) <http://sgn.cornell.edu> e <http://www.ncbi.nlm.nih.gov/mapview> ou em sites especializados em determinadas culturas como tomate em <http://tomatomap.net> (Van Deynze et al. 2006). Estes sites, além de apresentarem os mapas prontos, apresentam uma série de informações úteis como o número e freqüência dos mapas mais utilizados para determinadas culturas, os grupos de ligação e seus relativos tamanhos, estatísticas sobre o mapa, nome da sonda e acesso no GenBank, dentre outras informações.

Dentre as principais utilizações do mapa genético destacam-se a localização e o mapeamento de QTL (Quantitative Traits Loci), localização de genes ligados a doenças, estresse abiótico e mapeamento comparativo de genomas de diferentes espécies e também estudos de sintenia e clonagem genes com base em mapas genéticos que passaram a representar complementos importantes a serem considerados em programas de melhoramento das mais variadas espécies vegetais (Lee, 1995).

A comparação das estruturas genômicas de diferentes espécies (synteny mapping), do ponto de vista de homologias dos genes, conservação das distâncias e ordem de ligação nos cromossomos contribui para o entendimento sobre a evolução dos genomas (Tanksley et al., 1988; Kianian e Quiros, 1992; Ahn e Tanksley, 1993). O mapeamento comparativo é também uma estratégia para obtenção de mapa único de referência para a maioria das espécies vegetais cultivadas (Moore et al., 1995; Sewell et al., 1999), pelo menos ao nível de famílias taxonômicas.

Levaram-se apenas 100 anos da redescobertas dos trabalhos de Mendel ao completo seqüenciamento de uma planta superior (*Arabidopsis*). Análises preliminares

de sequências de *Arabidopsis* detectaram cerca de 25500 genes, quase o dobro do número de genes detectados em *Drosophila* (13600). Aproximadamente 45% de todos os genes detectados em *Arabidopsis* estão presentes em quatro ou mais cópias, de modo que destes 25500 genes podem ser classificados em aproximadamente 11600 tipos de proteínas. *Arabidopsis* oferece a chave para estudo do genoma de plantas superiores, devido ao tamanho de seu genoma ser relativamente pequeno em relação as demais plantas superiores cultivadas pelo homem. Embora não haja nenhuma razão para pensar que o conjunto básico de tipos de proteínas de um indivíduo seja maior que 12000 devido a extensa história evolucionária de poliploidização e duplicação segmental na maioria das plantas superiores (Wendel, 2000) sugere que mais de 26000 genes de *Arabidopsis* podem estar presentes na maioria das nossa plantas cultivadas.

Pensava-se que, quando a corrida para decifrar o código genético da vida chegasse ao fim, iríamos conhecer os segredos relativos a expressão da vida, porém este relato ainda reservava muitas surpresa aos cientistas, muito do que se acreditava fossem as regras da evolução e da manutenção da vida se esvaneceu a medida que aquelas pequenas letras eram traduzidas, afinal a “língua falada por nosso genoma” não era tão simplista quanto imaginávamos. Dogmas como cada gene seria responsável por uma proteína foram abandonados, pois os cientistas descobriram que a linguagem genética é muito mais sofisticada e menos linear do que se acreditava. Trechos inteiros do DNA, que inicialmente foram descartados como “lixo” evolucionário, estão sendo reciclados, pois tem funções e importância inimaginadas e ainda não totalmente descritas.

É claro que a informação genômica tem grandes impactos na genética quantitativa, devido ao grande volume de informações alcançado na era das “omicas”, essa informação já tem modificado bastante a maneira como se processa a informação

em genética quantitativa e ainda não se sabe quão profunda irão ser as mudanças nesta área do conhecimento em paralelo as tecnologias recentes.

Criou-se a expectativa de que as informações genóticas dos marcadores moleculares, uma vez correlacionados com características fenotípicas de interesse, pudessem ser amplamente utilizadas na obtenção e seleção de indivíduos com maior valor genético. Esta técnica ficou conhecida como seleção assistida por marcadores moleculares (MAS – Marker Assisted Selection).

Com a perspectiva de um aumento nos ganhos de seleção e redução nos ciclos de melhoramento via seleção assistida por marcadores, muitas pesquisas foram feitas e QTLs foram detectados e mapeados nas mais variadas culturas (Frary, et. al. 2000; Yano et. al. 2000; Takahashi et. al. 2001; El-Din El-Assal et. al. 2001; Liu et. al, 2002). No entanto, grande parte destes QTLs detectados e mapeados em cada espécie, não foram aplicados de forma prática nos seus programas de melhoramento (Bernardo et al. 2008)

As principais causas deste insucesso foram a necessidade do estabelecimento de associações entre os marcadores e os QTLs para cada família avaliada e o fato de serem feitas apenas a detecção de um pequeno número de QTLs de grande efeito, os quais, devido a natureza poligênica e a alta influência ambiental dos caracteres quantitativos, não explicam suficientemente toda a variação genética (Dekkers, 2004). Além disso, pode se destacar, também, que a seleção baseada em marcadores moleculares somente é superior em relação a seleção fenotípica quando esta é aplicada em uma família com tamanho superior a 500 indivíduos (Resende, 2007).

Os avanços de tecnologias de genotipagem em larga escala, a descoberta de novos marcadores como os SNPs e a automação do processo de genotipagem de marcadores (Jenkins and Gibson, 2002) proporcionaram a redução do preço por data

point e permitiram que um grande número de marcadores fosse usado para várias culturas.

Uma vez gerado um grande número de marcadores espalhados por todo o genoma de um indivíduo, alguns destes marcadores estarão muito perto do QTL e em desequilíbrio de ligação (LD) com este (Hastbacka et. al., 1994). O conceito de desequilíbrio de ligação refere-se a associação não aleatória entre dois genes ou entre um QTL e um loco marcador. Quando as frequências alélicas e genotípicas de um ou mais locos autossômicos são constantes de uma geração para a outra e as frequências genotípicas são determinadas pelas frequências alélicas, diz-se que este loco se encontra em equilíbrio de ligação. Com a ligação gênica, dois genes ligados apresentam uma associação que não se dá ao acaso e estão em desequilíbrio de ligação. Com os eventos de recombinação, a cada nova geração, os locos tendem a uma situação de equilíbrio, e o tamanho de um dado segmento cromossômico que contém dois locos quaisquer e que não sofreu recombinação diminui, o que conseqüentemente reduz o LD. Por essa razão, o uso eficiente de marcadores moleculares para auxiliar o melhoramento genético requer grande número desses, para que, mesmo em uma população que já passou por várias gerações e sucessivas recombinações históricas, exista um marcador tão próximo do QTL que apresente uma associação com este que não tem razão aleatória.

Uma vez que um marcador se encontra em LD com o QTL, alguns alelos destes marcadores, estarão correlacionados com efeitos positivos dos QTLs em todas as famílias e podem ser utilizados sem que seja necessário o estabelecimento da fase de ligação em cada família (Meuwissen et.al., 2001). Além disso, características quantitativas são controladas por muitos genes e para que se tenha grande parte da variação genética explicada pelos marcadores moleculares é preciso obter um marcador em desequilíbrio de ligação com cada loco controlador da característica quantitativa.

Teoricamente, todas as possíveis interações entre QTLs podem ser medidas por meio da análise de variância ou estatísticas correlatas. Na prática, Tanksley (1993) comenta que existem vários problemas com esta aproximação: 1- as amostras das populações segregantes são normalmente pequenas para estimativa da interação dos vários locos. Para medir a interação entre dois locos, todos os genótipos possíveis para estes locos devem estar presente na população com frequência suficiente para realizar as comparações estatísticas; 2- o número potencial de interações multilocos é muito grande, requerendo muitos testes estatísticos; 3- o efeito das interações poderia ser subestimado devido à recombinação entre o QTL e o marcador molecular, nos quais as análises foram baseadas. Este problema pode ser minimizado pelo uso de alta densidade de marcadores moleculares fortemente ligados aos QTLs, facilitando sua análise e detecção.

### **Utilização de marcadores moleculares no melhoramento de plantas**

Marcadores moleculares podem ser utilizados com diferentes finalidades em estudos de genética: proteção de cultivares (Brondani et al. 1998), confirmação de cruzamentos controlados e identidade clonal (Rocha et al. 2002), seleção de genitores (Dias et al. 2001), estudo de diversidade genética (Muro Abad et al. 2005a; Muro Abad et al. 2005b; Dias et al. 2005), mapeamento genético (Grattapaglia & Sederoff, 1994) e detecção de QTLs (Verhaegen et al. 1997).

A detecção e caracterização dos QTL dependem, principalmente, do mecanismo de herança da característica, da população avaliada, do número de indivíduos genotipados e do tipo de marcador molecular utilizado (Dekkers & Hospital, 2002; Mackay, 2001; Schuster & Cruz, 2004). O efeito do QTL na expressão da característica é associado à precisão experimental, à herdabilidade da característica avaliada e aos

efeitos aditivos e de dominância (Liu, 1998; Lynch & Walsh, 1998; Schuster & Cruz, 2004). Além destas três características, também destacam-se o efeito do QTL na expressão da característica, a precisão experimental e a herdabilidade.

Portanto, utilizando-se informação indireta dos marcadores, é possível construir um mapa de QTL. Entende-se por mapeamento de QTL não apenas sua localização física no genoma, mediante alguma escala apropriada, mas também a quantificação e caracterização de seus efeitos, como interações alélicas (grau de dominância) de cada um, e mesmo interações espistáticas ou pleiotrópicas (Rocha, 2004; Machado, 2004). Estudos mais detalhados possibilitam, ainda, a detecção e a decomposição da interação genótipos x ambientes em nível de cada QTL (Silva, 2001; Bento, 2006).

Brondani (2000) avaliou 11 características de interesse agrônomo no arroz a partir de 96 famílias  $RC_2F_2$  em experimentos conduzidos em dois locais, utilizando marcadores microssatélites. Foram identificados QTLs para todas as 11 características mensuradas, além de locos com agrupamentos de QTLs para caracteres ligados à produtividade nos cromossomos 1, 2, 3, 4, 5, 7 e 11.

Faleiro (2000) avaliou oito características quantitativas no feijoeiro a partir de 154 linhas endogâmicas recombinantes (RIL's) utilizando 42 marcadores SCAR. Os marcadores associados a cada característica, em conjunto, explicaram 39,79% do número de dias até o florescimento, 40, 14% do número de dias até a maturação, 28,99% do número médio de vagens por planta, 36,07% do peso de 100 sementes, 14,03 do número médio de sementes por planta, 23,62% do número médio de sementes por vagem, 19, 20% da produção média por planta e 17,13% da produção média por vagem.

Soares (2000), utilizando marcadores microssatélites., identificou dois QTL's associados a teor de proteína nos grupos de ligação C1 e C2, para as famílias cultivadas

em Cascavel, que explicaram respectivamente, 11,13% e 12,19% da variação do teor de proteínas.

A seleção fenotípica das espécies domesticadas tem sido praticada consciente ou inconscientemente por milênios, porém os avanços da genética molecular prometem aumentar a eficiência das técnicas de melhoramento até então utilizadas. Por isso, existem muitas razões para que a genética molecular não substitua os atuais processos de melhoramento, mas deve estar integrada a esses para que se possa obter o máximo de ganho, de forma econômica e racional (Lande e Thompson, 1990). O objetivo da utilização de marcadores moleculares é incrementar o ganho genético por unidade de tempo (Badu et al.2004; Muro Abad et al. 2000). Neste sentido, a construção de mapas de ligação utilizando marcadores moleculares é a estratégia que vem sendo utilizada para a detecção de locos relacionados com a expressão de características quantitativas. Em trabalhos atuais é necessário que seja apresentada uma abordagem diferente da genômica na qual ela se presta como ferramenta efetiva para identificação de padrões de segregação corroborando com a agregação de conhecimento e de maior acurácia nos processos iniciais dos programas de melhoramento.

### **Simulação de dados para análise genômica**

Direta ou indiretamente a simulação tem contribuído para o avanço tanto na genômica quanto para as demais áreas do melhoramento genético. A simulação consiste em construir um sistema que imite o funcionamento de uma realidade, com a finalidade de averiguar o que aconteceria no sistema real se alterações de interesse fossem efetuadas em seu funcionamento (Dachs, 1998).

Ao se fazer estudo baseado em simulação, deve-se considerar a abstração que consiste em substituir o objeto real de interesse por outro modelo semelhante. O modelo deve ser suficientemente simples para ser operacionalizado e interpretado

adequadamente, mas seu desempenho deve ser comparável com o modelo real e, se a defasagem for grande, ele deve ser eliminado ou refinado (Cruz, 2001).

Na utilização de uma técnica de simulação o pesquisador deve precaver-se contra erros, seja estes devidos a problemas como levantamentos amostrais, escolha inadequada das distribuições de probabilidades nos eventos de natureza aleatória, simplificação inadequada da realidade e erros de implantação do sistema simulado. Para a garantia de sua eficiência pode-se lançar mão de processos de validação. Essa validação consiste em fazer o sistema simulado operar nas condições do sistema real e verificar através de testes de hipóteses e outras análises estatísticas ou através de comparação com situações reais já analisadas, se os resultados observados na simulação condizem com os observados no sistema real (Cruz, 2001).

Uma das dificuldades atualmente enfrentadas pelos pesquisadores é a identificação da posição de QTLs e do tamanho do seu efeito. Através da simulação pode-se inferir e testar, sob determinadas pressuposições, o que aconteceria sob determinadas condições, como um maior distanciamento entre marcas moleculares, qual seria o efeito da presença de um ou mais QTLs próximos a um QTL flanqueado por marcadores, o que aconteceria caso existissem QTLs externamente a um intervalo limitado por marcadores, porém, sem a existência de QTL entre eles, e outras situações que se mostrassem de interesse a serem analisadas. Martínez (1992), no trabalho em que procurou estimar a localização e o tamanho dos efeitos de QTLs, usando marcadores flanqueadores, gerou um conjunto de dados simulados que exemplificaram o problema a ser pesquisado. Os dados foram analisados pelo método de mapeamento por intervalo e pelo modelo de regressão.

Darvasi et al. (1993) realizaram estudo de simulação, usando populações de retrocruzamento, para determinar o efeito do espaçamento de marcadores, efeito do

gene e o tamanho da população no poder da ligação entre o marcador e o QTL, no erro-padrão dos estimadores de máxima verossimilhança e do efeito do QTL sobre a localização no mapa de ligação. Neste estudo constatou-se que o poder de detecção do QTL foi virtualmente o mesmo, usando marcadores espaçados de 10 cM ou um número infinito de marcadores, com a ocorrência de pequeno decréscimo quando os marcadores foram espaçados de 20 ou até mesmo 50 cM. A vantagem de utilizar intervalo de mapeamento em vez de análise com apenas um marcador foi sem importância. O poder de resolução de uma ligação marcador-QTL foi definido como sendo de 95% do intervalo de confiança, para um mapa de localização de QTL que seria obtido através da utilização de infinitos marcadores. Descobriu-se que, reduzindo o espaçamento entre marcadores para menos do que o poder de resolução, não houve ganho apreciável estreitando o intervalo de confiança. É interessante observar que, nesse trabalho de simulação, foram geradas 1000 repetições para cada parâmetro de combinação.

Ferreira (1995) avaliou a eficiência do mapeamento de QTLs e da seleção assistida por marcadores. Para isso foi gerada uma população  $F_2$  com 400 plantas. Considerou-se o caráter sendo controlado por genes distribuídos em todos os 10 cromossomos, mapeando-se os QTLs em todos eles e considerando uma distribuição aleatória. Os valores genotípicos e fenotípicos foram gerados considerando ausência e presença de dominância e ausência de epistasia e quatro níveis de herdabilidade (0,750; 0,500; 0,250 e 0,125). Foi constatado que, com o aumento do número de genes controlando a característica quantitativa, houve um aumento de probabilidade de se detectar significativamente marcadores ligados a QTLs. O tamanho populacional de 400 plantas  $F_2$  foi adequado, pois foram obtidos, mesmo com baixos níveis de herdabilidade ( $h^2 = 0,125$ ), marcadores ligados aos QTLs com ligação menor que 10% de recombinação.

Uma clara vantagem do uso da simulação é o número de amostras que podem ser geradas. Como exemplo pode-se citar o trabalho de Martinez e Curnow (1992), que teve o objetivo de ilustrar como é possível aparecer um “QTL fantasma” quando é usada a análise de mapeamento por intervalo de Lander e Botstein (1989) ou o mapeamento por regressão usando apenas marcadores flanqueadores.

O que se verifica é que profissionais da área de informática têm pouco conhecimento dos problemas da área de genética, que apresentam certa complexidade por tratar de fenômenos biológicos e envolver princípios e, de certa forma, complexas distribuições probabilísticas na sua análise. Por outro lado, são raros os geneticistas com conhecimento e aptidão para atuarem na área da informática (Cruz, 2001). A integração entre essas áreas será fundamental para tornar a simulação computacional muito mais eficaz e utilizada entre melhoristas.

### **Hipóteses Genéticas**

Na determinação da herança, um método utilizado para caracteres qualitativos se baseia nas leis mendelianas, onde se avalia as proporções fenotípicas em  $F_2$  e retrocruzamentos. No entanto, para caracteres quantitativos fica impossível o estudo de herança por meio da segregação fenotípica, pois estes exibem os valores fenotípicos nas populações segregantes em distribuição contínua.

Algumas etapas devem ser seguidas para o estudo da herança monogênica ou oligogênica. Deve-se escolher genitores contrastantes para o caráter em questão, realizar os cruzamentos e obter as gerações  $F_1$  e  $F_2$ . Pode-se também obter os retrocruzamentos (RC), ou seja, o cruzamento da geração  $F_1$  com ambos ou apenas um dos genitores. Posteriormente, avaliam-se os genitores e todas as gerações de cruzamentos obtidas e verifica-se a segregação fenotípica, determinando o controle genético dos caracteres

com base nas leis mendelianas. Sendo os genitores puros, não haverá segregação na geração F<sub>1</sub>, apenas nas demais.

Um dos pontos mais importantes para o estudo das características da distribuição discreta é o teste de hipótese genética que permite, com uma margem de erro, concluir a respeito do padrão predominante da segregação: se governada por um, dois ou mais genes e o tipo de interação predominante (Liu, 1997). O teste de qui-quadrado tem se mostrado bastante prático e eficiente para o teste de hipóteses de padrões de segregação, uma vez que considera os desvios ocorridos entre valores previstos e observados e o número de observações avaliado (Schuster & Cruz, 2004).

Testes de qui-quadrado ( $\chi^2$ ) para verificar a razão de segregação de cada marca em todas as populações de mapeamento, com suas particularidades é dado por:

$$\chi^2 = \sum_{i=1}^n \left[ \frac{(\text{Obs}_i - \text{Esp}_i)^2}{\text{Esp}_i} \right]$$

onde:

- $\chi^2$  é valor de qui-quadrado calculado;
- $\text{Obs}_i$  e  $\text{Esp}_i$ , são os valores observado e esperado, para a i-ésima classe fenotípica (i= 1, 2,...,n), respectivamente.

A hipótese (H<sub>0</sub>) de segregação específica para cada loco é testada a um nível  $\alpha$  de probabilidade (usualmente  $\alpha=5\%$ ) que representa o erro tipo 1. Se o valor de probabilidade calculado foi inferior ao pré-estabelecido, a hipótese H<sub>0</sub> é rejeitada, ou seja, há a evidência de que a segregação não ocorre de acordo com o esperado.

Porém o teste de qui-quadrado é sensível ao tamanho da amostra, um número reduzido de informações pode resultar na impossibilidade de discriminação de hipóteses genéticas. Liu (1997) apresenta um estimador que permite a comparação de hipóteses

genéticas em relação ao número de eventos que deveriam ser analisados para discriminá-las considerando a distribuição de  $\chi^2$  e um nível crítico associado. Como mostrado por Liu (1997), a obtenção considera a resolução do sistema de equações apresentado abaixo, em que  $x$  e  $y$  representam padrões de segregação diferentes, e  $a$  e  $b$ , as frequências de suas classes, como mostrado abaixo:

$$(a - xb)^2 / nx$$

$$(a - yb)^2 / ny$$

O desenvolvimento deste sistema de equações permite a obtenção do estimador do número de indivíduos necessários para discriminar duas hipóteses genéticas:

$$n \geq \left[ \frac{1 + \sqrt{xy}}{\sqrt{x} + \sqrt{y}} \right]^2 x^2 \mathbf{1, \alpha}$$

Diferentemente da abordagem métrica, que, em relação a uma série de pressuposições, suporta desvios em relação ao não atendimento, a consideração inapropriada de uma hipótese genética pode resultar em problemas de interpretação. O equívoco nas conclusões em relação à natureza da herança da característica muitas vezes devido ao tamanho da amostra sendo comum encontrar na literatura populações avaliadas com tamanho amostral não suficiente para dar respaldo a veracidade do tipo de segregação, ao genótipo de genitores, direcionamento de cruzamentos e posicionamento dos genes nos mapas de ligação apresenta potencial para causar prejuízos na continuidade da pesquisa.

Visando avaliar controle genético da reação do feijoeiro ao *Phaeoisariopsis griseola* (fungo causador da mancha angular), Vieira (2004) cruzou duas linhagens

contrastantes quanto a característica, sendo elas a linhagem ESAL 693 (resistente) e a cultivar Rosinha (suscetível) e obteve as gerações F<sub>1</sub>, F<sub>2</sub> e progênes F<sub>2:3</sub>. Os dados da reação à mancha angular obtidos em F<sub>2</sub> e F<sub>2:3</sub> foram utilizados no estudo do controle genético do caráter, sendo as proporções fenotípicas observadas testadas pelo teste  $\chi^2$ . O resultado da F<sub>2</sub> se ajustou perfeitamente às proporções esperadas de ¾ resistentes e ¼ suscetível, sugerindo herança monogênica, sendo o alelo dominante responsável pela resistência e o recessivo pela suscetibilidade. A segregação 1:2:1 observada entre as famílias F<sub>2:3</sub> confirmou o resultado da F<sub>2</sub>. Desta forma, o controle genético da reação ao *P. griseola* é monogênico, sendo o alelo dominante da linhagem ESAL 693 responsável pela resistência (Tabela 1).

**Tabela 1.** Segregação para reação à mancha angular no feijoeiro nas gerações F<sub>2</sub> e F<sub>2:3</sub>.

Fenótipo	Nº de plantas F <sub>2</sub>		Fenótipo	Famílias F <sub>2:3</sub>	
	FO	FE		FO	FE
Resistente	111	110,25	Resistente	22	29,75
			Segregante	60	59,60
Suscetível	36	36,75	Suscetível	37	29,75

Estudando o controle genético do escurecimento precoce de grãos de feijão tipo carioca, por meio da segregação obtida em F<sub>2</sub>, Silva (2007) constatou que segregação ajustou-se a proporção de 3 escuros:1 claro,  $\chi^2 = 4,58$ , (P ≤ 0,032), indicando que o caráter deve ser controlado por um gene com dominância do alelo que condiciona grãos escuros precocemente (Tabela 2). A avaliação do escurecimento foi feita por meio de uma escala de notas variando de 1 a 5, sendo 1 a cor de fundo de grão muito clara, 2 mediamente claro, 3 claro, 4 mediamente escuro, 5 muito escuro. As notas foram atribuídas por dois avaliadores isoladamente. O processo foi realizado aos 30, 60 e 90 dias, após a colheita.

**Tabela 2.** Freqüências observadas e esperadas, considerando-se a segregação de 3 escuros: 1claro, na geração F<sub>2</sub>. Foram considerados claros os grãos que receberam, na média dos avaliadores, nota inferior a 1,5.

F <sub>2</sub> 60 dias	FO	FE
Escuro	160	147
Claro	36	49
Total	196	196

$$\chi^2 = 4,58, (P \leq 0,032)$$

### Método do mapeamento por intervalo simples

O método de mapeamento por intervalo apresentado por Lander & Botstein (1989) é considerado mais preciso que o método da marca simples, pois se baseia na identificação de QTLs por meio da análise dos intervalos entre marcadores adjacentes ao longo de todos os grupos de ligação. Os testes do mapeamento por intervalo são feitos pela incorporação da informação dos marcadores que flanqueiam determinado intervalo no modelo de regressão. A título de exemplificação, a seguir é apresentado o modelo de mapeamento por intervalo da população de retrocruzamento, segundo Schuster & Cruz(2004):

$$Y_j = \gamma + gx_j^* + \varepsilon_j$$

Em que:

$Y_j$  é o valor mensurado para a característica no i-ésimo indivíduo;

$\gamma = u + d_j$ : Valor genético do efeito u+d da característica na população;

$g = a + d$ : Efeito confundido aditivo e devido à dominância do loco em estudo;

$x^*j$  Variável condicionadora cujos valores são dependentes dos genótipos dos

marcadores que flanqueiam o QTL do indivíduo  $j$ ;  $e$

$\varepsilon_j$  erro aleatório na  $j$ -ésima observação em que  $\varepsilon_j \sim N(0, \sigma^2)$

De posse das probabilidades condicionais de ocorrência do genótipo marcador e do genótipo QTL, as mesmas estimativas também podem ser obtidas pelo método da máxima verossimilhança. Neste caso específico, a consistência entre as estimativas faz destas duas estratégias equivalentes para a obtenção das estimativas, sendo que a facilidade aritmética de obtenção das estimativas, utilizando o modelo de regressão, é um atrativo deste método (Schuster & Cruz, 2004). As estimativas de efeito e a posição dos QTLs obtidas pelo mapeamento por intervalo simples são mais precisas e acuradas do que aquelas obtidas pelo método de marcas simples. No entanto, o mapeamento por intervalo simples também possui limitações. A obtenção de estimativas viesadas de  $r$  causadas pela ocorrência de dois QTLs em um mesmo grupo de ligação é uma limitação já bem caracterizada desta metodologia. O QTL de efeito e posicionamento viesado, denominado QTL fantasma, é um artefato da técnica geralmente encontrado entre dois QTLs adjacentes em um mesmo grupo de ligação.

### **Aplicativos utilizados para simulação**

Diversos softwares são encontrados com a finalidade de realizar simulação de dados. Aqui será feito um breve relato sobre os dois diretamente relacionados com estudos genéticos.

O programa GENES, amplamente utilizado em análises de modelos aplicados ao melhoramento de plantas e animais, é um software destinado à análise e processamento de dados por meio de diferentes modelos biométricos, contando com procedimentos uni e multivariados, enfatizando estimação de parâmetros genéticos. Também estão disponíveis procedimentos para análise de dados binários, geralmente obtidos de estudos moleculares, permitindo a análise e interpretação de fenômenos

particulares desta área. Possui ainda o modo “simulação”, onde o usuário poderá avaliar tamanhos de amostras, número de famílias, plantas, repetições, em diferentes estudos.

O programa GENES está disponível para download, gratuitamente, sem nenhuma restrição de uso ou de divulgação, no endereço <http://www.ufv.br/dbg/genes/genes.htm> ou <http://www.ufv.br/dbg/biodata.htm>.

O programa GQMOL (Cruz, 2012), foi desenvolvido com o propósito de analisar dados obtidos de estudos moleculares. O programa pode ser usado análise de segregação de locos individuais, estimação de porcentagem de recombinação, agrupamento de marcas moleculares e mapeamento, incluindo estudos de QTL com populações controladas, populações exogâmicas, simulação e análise de imagem. Além disso, conta com um módulo de ensino, apresentando vários procedimentos para entendimento de princípios estatísticos e genéticos envolvidos na análise genômica. Possui também o módulo “simulação”, onde diferentes genomas podem ser simulados, levando em conta diferentes tamanhos amostrais, tipos de populações, variáveis quantitativas entre outras variáveis.

O programa GQMOL está disponível para download, gratuitamente, sem nenhuma restrição de uso ou de divulgação, no endereço <http://www.ufv.br/dbg/gqmol/gqmol.htm> ou <http://www.ufv.br/dbg/biodata.htm>.

Os dois programas tem sido de grande utilidade na área de melhoramento genético uma vez que conta com recursos para análise e processamento de dados fundamentados em diferentes metodologias biométricas. Também tem sido indispensável em muitos estudos por contar com módulo de simulação em que é possível estabelecer diferentes genomas, níveis de saturação, tipo de marcadores, grau de polimorfismo, tipo de população segregante e características quantitativas.

O uso destes aplicativos em estudos genômicos tem sido rotineiro, proporcionando valiosas contribuições na experimentação científica.

## Referências

AHN, S.; TANKSLEY, S.D. Comparative linkage maps of the rice and mize genomes. **Proceedings of the National Academic of Sciences of the United States of America**, Washigton, v.90, n.17, p.7980-7984, 1993.

BADU R., SUDHA K.N., PRASANNA B.M., GUPTA H.S. Integrating marker assisted selection in crop breeding – Prospects and challenges. **Current Science**, v.87, n.5, p.607-619, 2004.

BENTO, D.A.V. **Mapeamento de QTLs para produção de grãos e seus componentes em uma população de milho tropical**. 2006. 133p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.

BERNARDO R (2008). Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. **Crop Science**, 48:1649-1664.

BERZONSKY, W .A.; LAFEVER, H. N. Progress in Ohio soft red winter wheat breeding: grain yield and agronomic traits of cultivars released from 1871 to 1987. **Crop Science**, Madison, v. 33, n. 6, p. 1382, 1993.

BRONDANI, C. **Desenvolvimento de marcadores microssatélites, construção de mapa genético interespecífico de *Oryza glumaepatula* x *oryza sativa* e análise de QTLs para caracteres de importância agrônômica**. Brasília: Universidade de Brasília, 2000. 226p. (Tese – Doutorado em Genética e Melhoramento de Plantas).

BRONDANI, R.P.V., BRONDANI C., TARCHINI R., GRATTAPAGLIA D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla* . **Theoretical and Applied Genetics**, v.97, p.816-827, 1998.

COELHO, A.S.G. Considerações gerais sobre a análise de QTL's. In: PINHEIRO, J.B.; CARNEIRO, I.F. (Eds.). **Análise de QTL no melhoramento de plantas**. Goiânia: FUNAPE, 2000. p.1-36.

CRUZ, C.D. A informática no melhoramento genético. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos genéticos e melhoramento – plantas**. Rondonópolis: Fundação MT, 2001, p.1085-1118.

DACHS, N. **Estatística computacional**. ed. São Paulo: Livros Técnicos e Científicos Editora. S.A. 236p. 1998.

DARVASI, A.; WEINREB, A.; MINKE, V.; WELLER, J.I.; SOLLER, M. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. **Genetics**, v. 134, p.943-951, July, 1993.

DEKKERS J.C.M., HOSPITAL F. The use of molecular genetics in the improvement of agricultural populations. **Nature reviews**, v.3, p.22-32, 2001.

DEKKERS JCM Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. **Journal of Animal Science** 82:313-328. (2004).

DIAS L.A.S., ROCHA, R.B., MURO ABAD J.I., SALOMÃO T.F., ALFENAS A.C. Cocoa cultivar distinctiveness and hybrid prediction using RAPD markers. **Ingenic Newsletter**, v.6, p.19-22, 2001.

DIAS, L.A.S., ROCHA, R. B., PICOLI, E.A.T Distinctness of cacao cultivars using yield component data and RAPD markers. **Crop breeding and Applied Biotechnology**, v.5, p.47 - 54, 2005.

EL-DIN EL-ASSAL S, ALONSO-BLANCO C, PEETERS AJ, RAZ V, KOORNNEEF M: A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2, **Nat Genet**, 29:435-440. (2001).

FALEIRO, F.G. **Melhoramento e mapeamento genético do feijoeiro-comum: análise de características quantitativas, morfológicas, moleculares e de resistência a doenças.** Viçosa: UFV, 2000. 1776p. (Tese – Doutorado em Genética e Melhoramento de Plantas).

FERREIRA, M.E., GRATTAPAGLIA, D. **Introdução ao uso de marcadores RAPD e RFLP em análise genética.** Brasília: EMBRAPA-CENARGEN. 220p. 1995

FERREIRA, M.E., GRATTAPAGLIA, D.(1995). **Introdução ao uso de marcadores RAPD e RFLP em análise genética.** Brasília: EMBRAPA-CENARGEN. 220p.

FRARY A, NESBITT TC, GRANDILLO S, KNAAP E, CONG B, LIU J, MELLER J, ELBER R, ALPERT KB, TANKSLEY SD: fw2,2: a quantitative trait locus key to the evolution of tomato fruit size. **Science**, 289:85-88. (2000)

GRATTAPAGLIA, D., SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, v.137, p.1121-1137, 1994.

GRIFFITHS, A.J.F.; MILLER, J.H.; SUZUKI, D.T.; LEWONTIN, R.C.; GELBART, W.M. **Introdução à genética.** Rio de Janeiro: Guanabara, 1998.856p.

KIANIAN, S.F.; QUIROS, C.F.; Geration of a Brassica oleracea composite RFLP map: linkage arrangements among various populations and evolutionary implications. **Theoretical and Applied Genetics**, New York, v.84, n.5/6, p.544-554, 1992.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in improvement of quantitative traits. **Genetics**, v.124, n.3, p.743-756, 1990.

LANDER, E.S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Bethesda, v.121, n.1, 185-199, 1989. (erratum in 1994 *Genetics*, v.136, n.2, 705).

LANDER, E.S.; WEINBERG, R.A. Genomics: journey to the century of biology. **Science**, Washington, v.287, n.5459, p.1777-1782,2000.

LEE, M. **DNA markers and plant breeding programs**. Advances in Agronomy, San Diego, v.287, n.5459, p.1777-1782, 1995.

LIN, S.; ARBER, W. Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. **Proceeding of the National Academy of sciences** 59:1300-1306. (1968)

LIU B.H., **Statistical Genomics: linkage mapping and QTL analysis**. Boca raton, Florida, USA:CRC Press, 1997, 610p.

LIU J, VAN ECK J, CONG B, TANKSLEY SD. A new class of regulatory,genes underlying the cause of pear-shaped tomato fruit. **Proc Natl Acad Sci USA** 99:13302-13306. (2002)

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, USA: Sinauer Associates, Inc.Ed., 1998, 978p.

MACHADO, C.F. **Repetibilidade, correlações fenotípicas e mapeamento de QTLs em populações segregantes de café arábica**. 2004. 188p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.

MACKAY T.F.C., Genetic architecture of quantitative traits. **Annual Review of Genetics**, v.35, p.303-339, 2001.

MALIEPAARD, C.; JANSEN, J.; VAN OOIJEN, J.W. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetics Research**, v.70, p.237-250, 1997.

MOORE, G.; DEVOS, K.M.; WANG, Z.; GALE, M.D. Cereal Genome Evolution: Grasses, line up and from a circle. **Current Biology**, London, v.5, n.7, p.737-749, 1995.

MURO ABAD, J.I., ROCHA, R.B., CRUZ, C.D., ARAÚJO, E. F. Crosses recommendation method for obtaining Eucalyptus spp. hybrids assisted by molecular markers. **Crop Breeding and Applied Biotechnology**, v.5, p.71- 73, 2005b.

MURO ABAD, J.I., ROCHA, R.B., CRUZ, C.D., ARAÚJO, E.F. Obtainment of Eucalyptus spp. hybrids aided by molecular markers - SSR analysis. **Scientia Forestalis**, v.67, p.53 - 63, 2005a.

RAMALHO, M.A.P., LAMBERT, E.S. Biometria e o melhoramento de plantas na era da genômica. **Revista Brasileira de Milho e Sorgo** v3, p.221-242, 2004.

RESENDE, MDV. Matemática e estatística na análise de experimentos e no melhoramento genético. Colombo: Embrapa Florestas. 561p. (2007).

ROCHA, R.B. **Mapeamento de QTL para características de qualidade da madeira e crescimento em híbridos (*Eucalyptus grandis* x *Eucalyptus urophylla*)**. 2004. 77p. Dissertação (Mestrado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.

SAX, K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. **Genetics**. v.8, p.552-560. (1923).

SCHUSTER I., CRUZ C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos**. Viçosa , MG. Editora UFV, 2004. 585p.

SEWELL, M.M.; SHERMAN, B.K.; NEALE, D.B. A consensus map for Loblolly Pine (*Pinus taeda* L.): I. Construction and integration of individual linkage maps from two outbreed three-generation pedigrees. *Genetics*, Baltimore, v.151, n.1, p.321-330, 1999.

SILVA, G. S. **Controle genético do escurecimento precoce de grãos de feijão tipo carioca**. 2007. 52p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.

SILVA, H.D. **Aspectos biométricos da detecção de QTL's ("quantitative trait loci") em espécies cultivadas**. 2001. 166p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

SOARES, T.C.B. **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Viçosa: UFV, 2000. 58p.. (Tese – Doutorado em Bioquímica).

TAKAHASHI Y, SHOMURA A, SASAKI T, YANO M: Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the a subunit of protein kinase CK2, **Proc Natl Acad Sci USA**, 98:7922-7927. (2001)

TANKSLEY, S.D. Mapping polygenes. **Annual Review of Genetics**, Palo Alto, v.27, p.205- 233, 1993.

TANKSLEY, S.D.; BERNATZKY, R.; LAPITAN, N.L.; PRINCE, J.P. Conservation of gene repertoire but not gene order in pepper and tomato. **Proceedings of the National Academic of Sciences of the United States of America**, Washington, v.85, n.17, p. 6519-6423, 1988.

VAN DEYNZE, A.; VAN DER KNAAP, E.; FRANCIS, D. Development and application o fan informative seto f anchored markers for tomato breeding. In: **Plant and Animal Genome XVI conf**, San Diego, CA, USA. P.188. (2006)

VERHAEGEN, D., PLOMION, C., GION, J.-M.;POITEL, M., COSTA, P., KREME, A. Quantitative trait dissection analysis in Eucalyptus using RAPD markers: I. Detection of QTL in Interespecific hybrid progeny, stability of QTL, expression across different ages. . **Theoretical and Applied Genetics**. v. 95, p. 597-608, 1997.

VIEIRA, F. C. **Controle genético da reação do feijoeiro ao *Phaeoisariopsis . griseola* e seleção de famílias baseada em caracteres agronômicos**. 2004. 31p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.

WENDEL, J. G. Genome evolution in polyploids. **Plant Molecular Biology** 42: 225-249. (2000)

WILCOX, J. R. Sixty years of improvement in publicly developed elite soybean lines. **Crop Science**, Madison, v. 49, p. 1711-1716, 2001.

YANO M, KATAYOSE Y, ASHIKARI M, YAMANOUCHI U, MONNA L, FUSE T, BABA T, YAMAMOTO K, UMEHARA Y, NAGAMURA Y, SASAKI T: Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. **Plant Cell**, 12:2473-2484. (2000)

## **CAPÍTULO 1**

### **ABORDAGEM SOBRE ESTUDO DE HERANÇA DE CARATERES PELA GENÉTICA CLÁSSICA E MOLECULAR PARA PADRÃO BINÁRIO DE EXPRESSÃO FENOTÍPICA**

Resumo – O trabalho teve por objetivo caracterizar hipóteses genéticas de maior similaridade utilizadas para prever padrão de herança em características oligogênicas. Foi utilizado o delineamento genético clássico, envolvendo população  $F_2$  derivada de  $F_1$  proveniente de cruzamento entre pais homocigotos contrastantes, de característica controlada por um, dois e três genes independentes e ligados. A este estudo foi também considerada a agregação de informações moleculares de indivíduos  $F_2$  possibilitando o emprego de análises genômicas para fins de complementação de informações e elucidação de testes de hipóteses no caso de existir segregação ambígua em hipóteses concorrentes. Foram simulados genomas parentais e populações  $F_2$  de 100, 200, 400 e 600 indivíduos, fenotipados (0 e 1) e genotipados em relação a cinco grupos de ligação contendo 11 marcadores co-dominantes por grupo de ligação, e saturação de 10 cM entre marcadores. Testes de genética clássica, baseados em qui-quadrado, foram incapazes de identificar corretamente o padrão de herança de características controladas por genes ligados e proporcionaram grande percentual de erro, principalmente em tamanhos populacionais mais reduzidos, em teste com hipóteses com segregações concorrentes próximas tais como 3:1 e 13:3, 9:7 e 37:27 e 1:2:1 e 3:9:4. As análises genômicas foram realizadas pelas metodologias de intervalo simples aplicadas aos estudos de detecção de QTL (quantitative trait loci) demonstrando seu potencial na identificação dos genes, posicionamento e quantificação de seus efeitos sobre a expressão da característica.

Termos para indexação: simulação; genômica; ambigüidade de segregação; identificação de QTL; dados binários.

## Introdução

Entre as principais necessidades de esclarecimento dentro das ciências biológicas citam-se os mecanismos e os processos da hereditariedade e da variação. A genética clássica descreve uma maneira apropriada para a realização de tais estudos, que são as avaliações e cruzamentos de genitores contrastantes para o caráter em questão, e gerações  $F_1$  e  $F_2$ . Posteriormente, avaliam-se os genitores e todas as gerações de cruzamentos obtidas e verifica-se, por meio da análise da distribuição fenotípica de caracteres qualitativos ou através das médias e variâncias de caracteres quantitativos. Esses estudos proporcionam informações fundamentais e complementares para o estudo e entendimento da diversidade e da variabilidade genética.

O desenvolvimento de ferramentas para o estudo da herança e da variação das características hereditárias associados a técnicas moleculares associados ao DNA tem permitido o mapeamento e a identificação de genes para quaisquer características. O conhecimento adquirido por estes processos tem permitido aos geneticistas a manipulação mais eficaz da hereditariedade com abordagem ampla englobando características monogênicas, oligogênicas ou poligênicas. Nesse sentido, o emprego dos delineamentos genéticos convencionais agregado a informações moleculares devem ser investigados com vistas a quantificar o ganho de informação sem onerar a pesquisa.

No estudo da herança de característica qualitativa, por meio da genética clássica, é de interesse do pesquisador conhecer o número de genes e alelos envolvidos no controle e as suas interações para a expressão da característica. Testes de hipóteses para o reconhecimento do padrão de herança de características monogênicas e oligogênicas, determinadas por genes independentes, foram introduzidos na genética pelos trabalhos originais de Mendel e até hoje utilizados em pesquisa com organismos diploides.

A pormenorização de estudos permitindo a detecção, posição relativa e a quantificação dos efeitos individuais dos genes nos cromossomos, assim com a sua relação com outros genes só foi possível a partir da disponibilidade de mapas genéticos obtidos pela análise de um número relativamente grande de marcadores abrangendo todo o genoma da espécie. Dessa forma, as análises genômicas passaram a ter também papel fundamental na disponibilização de ferramentas em vários campos da pesquisa genética, uma vez que permitem a visualização, mesmo que de forma relativa, da organização dos genes nos cromossomos.

A genética se associa ao melhoramento vegetal por meio da experimentação e da modelagem que envolvem a mensuração sistemática e criteriosa de um caráter em campo, a aplicação e a associação dos valores obtidos com os parâmetros e os conceitos básicos de genética. O tratamento estatístico destas associações e informações é, por fim, estabelecido pela biometria. A agregação de técnicas modernas de análise genômica aos procedimentos clássicos de seleção de genótipos e classificação de padrões hereditários constitui grande desafio contemporâneo para pesquisadores.

É indiscutível a importância do estudo de caracteres monogênicos para o melhoramento de plantas. Tais características são governadas por um gene, ou pelo menos um gene de efeito maior. As características oligogênicas possuem padrão de herança mendeliano governado por dois, três ou mais genes de interação epistática, podendo seu padrão de expressão fenotípico ser alocado em classes como, por exemplo, os casos em que se atribui notas a severidade ao ataque de doenças. Alguns exemplos de características importantes para o melhoramento vegetal e de provável herança oligogênica são: espessura de casca e acúmulo de óleos essenciais em eucalipto (Zobel & Jett, 1995), resistência ao *Ceratocystis* spp. em eucalipto (Alfenas et al. 2004), resistência à mancha angular e à ferrugem e à antracnose no feijoeiro comum (Vale &

Zambolim; 1997), resistência à ferrugem em soja (Vijayalakshmi et al. 2005; Junghans et al. 2003; Wilcox et al. 1996).

A caracterização do padrão da herança é um dos pontos mais importantes na manipulação dos componentes hereditários das características, uma vez que a determinação incorreta do padrão de herança pode gerar resultados equivocados nos ensaios de seleção. O teste de hipóteses genéticas de segregação permite, com uma margem de erro, concluir a respeito do padrão predominante de herança: se governado por um, dois ou mais genes e o tipo de interação predominante entre genes e alelos (Schuster & Cruz 2004; Lynch & Walsh, 1998). Porém podem ocorrer casos de segregação ambígua que é a impossibilidade de discriminação entre duas ou mais hipóteses genéticas. A não rejeição de uma hipótese genética que não representa o verdadeiro mecanismo de herança de uma característica pode resultar em erros grosseiros e conclusões equivocadas.

Na literatura não é raro encontrar falta de concordância de relatos sobre o número de genes responsáveis pelo controle de certas características. Kukhang et al. (1993) identificou padrão de segregação 3:1 para resistência a *H. vastatrix* em híbridos do Timor (*Coffea Arabica*) enquanto (Pereira 1995) encontrou um padrão 63:1 e (Capucho et al 2009) encontraram um padrão de segregação de 2 genes dominantes e independentes.

De maneira geral, estudo de herança tem sido realizados por meio da mensuração da expressão fenotípica de indivíduos em delineamentos que envolvem as gerações genitoras ( $P_1$  e  $P_2$ ), geralmente homocigotas e contrastantes, a  $F_1$  e a geração segregante  $F_2$ . Para características supostamente de herança monogênica e oligogênica o estudo genético é feito por meio da identificação do padrão de segregação na  $F_2$  e, se necessário, a ratificação da hipótese de segregação em gerações derivadas por

retrocruzamento. Para estudo de caracteres quantitativo este delineamento também é apropriado, recorrendo-se às gerações sem variabilidade genética ( $P_1$ ,  $P_2$  e  $F_1$ ) para quantificar a influência da variação ambiental. Retrocruzamento também tem sido indicado para estudar os componentes aditivos e não-aditivos da variação genotípica. Associar a este delineamento experimental as informações da genotipagem da geração  $F_2$  certamente contribuirá para a elucidação de questões fundamentais da herança do caráter e minimizará erros e equívocos em análises e interpretações de uma questão tão relevante para o melhoramento genético.

O que se observa é que modernas ferramentas de análises genéticas não tem sido aplicadas efetivamente ou contemplada a possibilidade de seu uso durante o estabelecimento e a condução de experimentos. A análise de experimentos clássicos como teste de segregação aliado a análise genômica poderia solucionar definitivamente o problema de erro tipo II para características oligogênicas, trazendo maior confiabilidade sobre o número de genes que controlam a herança da característica principalmente no que se refere a resistência de plantas a pragas e doenças.

O objetivo deste trabalho é o de caracterizar as hipóteses genéticas de maior similaridade utilizadas para prever o padrão de herança de características oligogênicas, de interesse agrônomo, em experimentos envolvendo o delineamento de gerações  $P_1$ ,  $P_2$ ,  $F_1$  e  $F_2$  e prover informações adicionais sobre posicionamento e quantificação do efeito de locos individuais sobre a expressão da característica. Erros de rejeição de hipóteses verdadeira, em particular com hipóteses concorrentes próximas tais como 3:1 e 13:3 ou 1:2:1 e 3:9:4, por exemplo, serão analisados e discutidos. Também é avaliada a contribuição da análise genômica, com a agregação de informações da genotipagem da  $F_2$ , na elucidação do padrão de segregação.

## **Material de Métodos**

### *Material Genético*

Neste trabalho foram utilizadas, para fins de estudo de herança, populações simuladas  $F_1$ , derivada de pais homocigotos contrastantes, e  $F_2$ , e informações fenotípicas e moleculares desta geração segregante. Foram consideradas populações de tamanho 100, 200, 400 e 600 indivíduos.

### *Informações moleculares da $F_2$*

As informações fenotípicas e moleculares das populações  $F_2$  foram obtidas de diferentes tamanhos populacionais simulados. Considerou que esta geração foi derivada de parentais homocigotos ( $P_1$  e  $P_2$ ) contrastantes com  $P_1$  dominante e  $P_2$  recessivo, pelo módulo de simulação do programa para análise de dados moleculares e quantitativos – GQMOL (Cruz, 2012).

As informações moleculares foram simuladas considerando genoma com cinco grupos de ligação e saturação de 11 marcas moleculares codominantes, espaçadas a 10 cM, portanto, cada grupo de ligação apresentou o tamanho de 100 cM.

### *Informações fenotípicas da $F_2$*

Para cada população genotipada simulada também foram consideradas 10 características simuladas com um padrão de expressão controladas por 1, 2 ou 3 genes com segregação variável. Para as características governadas por um gene, considerou-se as segregações 3:1 e 1:2:1; para as características governadas por dois genes, considerou-se as segregações 9:7, 13:3, 15:1, 9:6:1, 12:3:1 e 9:3:4 e, para as características governadas por três genes considerou-se a segregação 27:37 e 63:1.

Os genes controladores destas características foram posicionados de forma a se ter resultados de padrão de segregação típico de genes independentes e ligados. Os

genes independentes foram localizados nas posições correspondentes aos marcadores 5 (para características controlada por um gene) ; 5 e 20 (para características controlada por dois genes); e 5, 20 e 50 (para características controlada por três genes). Para as características controladas por genes ligados a posição dos genes foi aquela correspondente aos marcadores 15 e 16 ou 13, 14 e 15, de modo que o padrão de herança foi determinado utilizando-se os dados genotípicos relativos a cada marcador responsável pela característica em estudo, nos diferentes conjuntos populacionais e codificando-os de acordo com o padrão de herança desejado.

Assim, caracterizando o processo de simulação para segregação do tipo 9:7, com possibilidade de extrapolação para os demais casos, os dados relativos aos marcadores 5 e 20 codificados com 0, 1 e 2 foram transformados em padrão fenotípico recebendo código 1 na combinação 2-2, 2-1, 1-2 e 1-1 e código 0 nas demais combinações.

#### *Análises estatísticas moleculares*

Após a simulação dos dados moleculares da população  $F_2$ , foi realizada a análise de segregação de locos individuais. Foi aplicado o teste de qui-quadrado, a 5% de probabilidade, para constatação da razão de segregação 1:2:1 em cada marcador para todas as populações.

Para estes mesmos conjuntos de dados foram construídos mapas genéticos. Para construção dos grupos de ligação, utilizou-se a propriedade transitiva, ou seja, se o loco A está ligado ao loco B, e o loco B está ligado ao loco C, logo o loco A está ligado ao loco C, independente da frequência de recombinação de A e C e, portanto, A, B e C pertencem ao mesmo grupo de ligação. Os critérios utilizados no agrupamento foram a frequência máxima de recombinação ( $r_{max}$ ) e o LOD mínimo ( $LOD_{min}$ ), para inferir se dois locos estão ligados. Foram utilizados os valores 30% e 3, respectivamente, para

$r_{\max}$  e  $\text{LOD}_{\min}$ . As marcas mais próximas em relação às marcas já consideradas que atenderem aos dois critérios adotados serão incorporadas ao grupo de ligação. Assim, o processo continua investigando a porcentagem de recombinação e o LOD entre marcas e possíveis vizinhos a serem incorporados às extremidades do grupo de ligação em construção.

Para o ordenamento das marcas no grupo de ligação foi utilizado o método da soma das frações de recombinação adjacente (SARF – Sum of Adjacent Recombination Fraction) para os mapas construídos sem análise multiloco. Neste processo, a melhor ordem é aquela que apresenta menor soma de recombinações adjacentes. Considera-se a ordem original estabelecida pelo processo de agrupamento e aplica-se o algoritmo RCD (Rapid Chain Delineation), que consiste em realizar permutas entre dois marcadores vizinhos ou distantes envolvendo três ou quatro marcadores. A ordem é alterada se, após a permuta, a soma das distâncias adjacentes for reduzida. Após todas as permutas conclui-se que a melhor ordem é aquela de menor soma de distâncias adjacentes (Schuster & Cruz, 2008).

#### *Estudo do padrão de herança com base em informações fenotípicas*

O estudo da herança de cada característica, obtida por simulação considerando a ação de genes com posição previamente conhecida e sem efeito ambiental, foi realizado da forma convencional em estudo de Genética Mendeliana. Nesse sentido formula-se uma determinada hipótese de segregação e aplica-se um teste apropriado, geralmente o teste de qui-quadrado. Assim, por exemplo, a não rejeição da segregação 9:7 será indicativo de que a característica é governada por dois genes independentes e com interação epistática. Entretanto, é possível obter resultado falso (positivos ou negativos)

em razão do tamanho inadequado da população, da metodologia empregada e o tipo de informação disponível, que neste caso é apenas fenotípica.

Para comprovar a influência do tamanho populacional sobre o resultado do teste de segregação foram realizados testes de qui-quadrado para verificar a ocorrência de ambigüidade de aceitação de hipóteses (positiva para mais de um tipo de segregação) em cada população dos conjuntos relacionados ao tamanho populacional para todas as hipóteses de segregação que mostravam mesmo número de classes, em um total de 8320 testes.

Para características controladas por genes ligados não há uma proporção prévia a ser adotada pelo desconhecimento da porcentagem de recombinação entre estes genes. Assim, para dois genes  $A/a$  e  $B/b$  ligados em que  $A-B-$  determina um padrão fenotípico e  $(A-bb, aaB-$  e  $aabb)$  determina o outro padrão fenotípico a segregação 9:7 é considerada um “falsa hipótese de segregação”. Nesse caso os genes estavam localizados em mesmo cromossomo, e , portanto, a segregação não era independente.

A adequação da metodologia clássica para estudo de herança, apenas com base em informação fenotípica, foi avaliada pela quantidade de erros e acertos da não rejeição da hipótese de segregação verdadeira ou não rejeição simultânea de duas hipóteses pela ambigüidade da razão observada de segregação.

Para melhor elucidação do processo adotado neste estudo, considerou, por exemplo, uma característica sabidamente controlada por um gene e segregação simulada de 3:1. De posse dos dados fenotípicos foi avaliada a segregação 3:1 e todas as demais proporções adequadas tais como: 9:7, 13:3, 15:1, 27:37 e 63:1. Quando se esperava um padrão de segregação com três classes, como no caso de 1:2:1, foram também realizados os testes de qui-quadrado em relação 9:6:1, 12:3:1 e 9:3:4. Assim, algumas situações foram encontradas e quantificadas: i. Apenas a hipótese verdadeira não foi

rejeitada; ii. A hipótese verdadeira foi rejeitada e outra falsa não rejeitada; iii. Tanto a hipótese verdadeira quanto a falsa não foram rejeitas com nível de significância favorecendo, ou não, a hipótese verdadeira.

#### *Estudo do padrão de herança com base em informações fenotípicas e moleculares*

Considerou-se que, para os mesmos conjuntos de dados em que o padrão de herança da característica fora estudado por meio de técnicas convencionais da Genética Clássica, também havia disponibilidade de mapas genéticos de forma que o estudo poderia ser aprimorado ou elucidado por meio de análises genômica demonstrando e a utilidade desta para melhorar acurácia das informações nos programas de melhoramento.

Os conjuntos de dados provenientes de diferentes tamanhos populacionais envolvendo característica com diferentes controles gênicos foram submetidas aos procedimentos similares ao de identificação de QTLs pelo método do intervalo simples. Os resultados favoráveis esperados por esta metodologia seria o de identificar apropriadamente o número de genes controladores de cada característica, o posicionamento destes genes, principalmente quando a característica era controlada por genes ligados e a genética clássica mostra-se ineficiente neste estudo e identificar o padrão epistático de herança, ou seja, ser capaz, de no caso de dois genes, caracterizar apropriadamente as segregações 9:7, 15:1, 13: 3, 9:3:4 e 9:6:1.

Os resultados não completamente satisfatórios por esta metodologia consistiam em: não identificação; identificação do número de genes correto porém na posição próxima a esperada; identificação do genes na posição certa com ocorrência de gene fantasma ou “extrapolado” e identificação do número e posição incorretos.

Dados relativos a valores de cada genes expressos por **m**, **a** e **d** foram analisados, em que **u** representa a média fenotípica dos genótipos homozigotos, ou o ponto médio, **a** mede o afastamento dos homozigotos em relação ao ponto médio, efeito aditivo e **d** mede o afastamento do heterozigoto em relação ao ponto médio. As relações entre **a** e **d** permite conhecer o tipo de interação alélica presente, em que **d/a** mede o grau de dominância (gd) de um determinado loco. Os valores de **m**, **a** e **d** foram anotados para cada indivíduo de todos os 8 conjuntos populacionais. Foram utilizados valores médios dos 20 indivíduos para cada padrão de segregação analisado com intuito de corroborar com as informações proporcionadas pela genômica e de caracterizar o tipo de segregação fenotípica. Todas as análises foram realizadas utilizando o aplicativo computacional GQMOL (Cruz, 2012), disponível em <http://dl.dropbox.com/u/34614950/gqmolexe.zip> <http://dl.dropbox.com/u/34614950/dadosgq.zip>.

## **Resultados e discussão**

Estão apresentados na Tabela 1 os resultados de rejeição, ou não, de hipótese de segregação para 10 características hipotéticas cujo controle gênico era previamente conhecido em famílias  $F_2$ . Sabe-se que, entre as populações tradicionalmente avaliadas no melhoramento de plantas, as  $F_{2s}$  derivadas de  $F_1$  (Harushima et al 1998) e as populações de retrocruzamento (Weiguo et al 2002) são intensivamente utilizadas para o mapeamento genético e estudo de padrões de herança fenotípica por sua praticidade e facilidade de obtenção (Schuster & Cruz 2004). Pode-se observar que, para as características controladas por um ou dois genes, a porcentagem de réplicas de simulação em que não houve rejeição da hipótese verdadeira ( $Pc1$ ) foi relativamente alta. Observa-se também que o aumento do tamanho da população não proporcionou grandes mudanças na não rejeição de  $H_0$ , ou seja a característica estudada estava

segregando conforme o esperado. Em alguns testes de segregação foi observado nível de significância inferior a 5%, porém a porcentagem de ocorrência de tais resultados não ultrapassou 10% das réplicas dentro de alguns conjuntos populacionais podendo ser explicado pela própria aleatoriedade do processo de simulação. Exceção ocorreu para a característica controlada por três genes e segregação esperada de 27:37 em que o valor de  $Pc1$  (porcentagem de não rejeição de hipótese verdadeira) diminuiu consideravelmente com o aumento do tamanho população indicando que a dificuldade do teste de qui-quadrado em proporcionar classificação correta para este tipo de segregação. O teste de qui-quadrado é bastante sensível ao tamanho da amostra e a não rejeição de hipóteses verdadeiras em tamanhos populacionais elevados só ocorrem quando o desvio, entre observado e esperado, for relativamente pequeno. Assim, como ilustração, pode-se considerar experimento com ocorrência de dois eventos com frequência paramétrica  $p=q=0,5$ . Entretanto, em um experimento 10 observações pode-se encontrar resultado iguais a  $p=0,7$  e  $q=0,3$  e o teste qui-quadrado estará associado a um nível de significância estimado de 20,59% indicando a não rejeição da hipótese. Em outro experimento, com 1000 observações, pode-se encontrar  $p = 0,55$  e  $q = 0,45$ , bem mais próximos do valor paramétrico e o teste de qui-quadrado estará associado a um nível de significância 0,1565% indicando a rejeição da hipótese.

Diferentemente da abordagem métrica que, em relação a uma série de pressuposições, suporta desvios em relação ao não atendimento, a consideração inapropriada de uma hipótese genética pode resultar em problemas de interpretação. O equívoco nas conclusões em relação a natureza da característica, ao genótipo dos genitores e ao direcionamento de cruzamentos possui grande potencial para causar prejuízos a qualidade da pesquisa.

Liu (1997) descreve metodologia que permite obter um tamanho populacional ótimo para realizar a comparação de hipóteses genéticas de forma a discriminá-las apropriadamente considerando a distribuição de qui-quadrado e um nível crítico associado. Assim, hipótese de segregação 3:1 e 13:3 demandariam 666 indivíduos para inferir sobre a hipótese mais apropriada, adotando 5% como nível crítico de probabilidade. Usando o mesmo método, a distinção entre hipóteses alternativas de 9:7 e 27:37 demandariam 23.383 indivíduos.

Na Tabela 1 verificou-se que os testes de hipótese comumente adotado em estudos de genética clássica não se mostra eficiente quando se observa os valores de qui-quadrado associados à hipóteses alternativas, pois em todos os tamanhos populacionais houve problemas de ambigüidade de segregação. Para estabelecer o tamanho apropriado da população considerou-se a informação na literatura em que amostra de tamanho reduzido, inferiores a 50 indivíduos são inadequadas para avaliações fenotípicas e proporciona baixa resolução em estudo de mapeamento genético tornando difícil a detecção de genes (ou QTLs) de pequeno efeito (Young, 1994). Atualmente, para estudos moleculares, o tamanho das populações utilizadas tem sido maiores. São encontrados trabalhos que estudam a influência do tamanho populacional na recuperação dos grupos de ligação, tanto para cruzamento controlado quanto para populações exogâmicas. Observou-se que, em populações de 200 indivíduos, recupera-se as informações genômicas, dependendo da saturação (Bhering & Cruz, 2008; Yan et al., 2009). Assim, considerou ser apropriado neste estudo utilizar amostras de tamanho 100, 200, 400 e 600 indivíduos.

Para o tamanho populacional de 100 e 200 indivíduos, sete e cinco proporções maiores que 50% ( $P_c2$ ) foram observados, respectivamente. Somente para as proporções 9:6:1 e 12:3:1 não houveram contradições em nenhum tamanho

populacional. O problema é ainda mais crítico quando o experimento indica uma hipótese alternativa falsa como mais apropriada para descrever os dados, que neste estudo foi quantificado por  $Pc3$ . No tamanho populacional de 100, os valores de  $Pc3$  foram bem expressivos alcançando valores de 50% e 90% para as segregações 9:7 e 27:37, respectivamente. Este problema foi minimizado com o aumento do tamanho populacional, percebendo redução do problema da ambigüidade de segregação, de forma que, no tamanho populacional de 600 indivíduos, apenas as segregações 9:7; 9:3:4 e 27:37 apresentaram valores expressivos de  $Pc2$ , porém a frequência  $Pc3$  foi menor que a de  $P1c$  para a segregação 9:7.

A verdadeira segregação da característica controlada por dois genes com segregação 9:7 não foi encontrada por completo com o aumento populacional com 45% dos valores de  $Pc3$ . Somente para a proporção 27:37 que houve o inverso, ou seja, com o aumento da população houve um maior índice de aceitação da hipótese concorrente 9:7 (25%) em detrimento da verdadeira, um aumento na ocorrência do erro tipo 1.

A Tabela 1 também traz informações sobre identificação de QTL no intuito de verificar a quantidade de informação adicional gerada pela genômica em comparação aos métodos clássicos de estudo de características oligogênicas. Foi considerado que em estudos de detecção de genes (ou de QTL), primeiramente, precisa-se saber qual o objetivo da pesquisa, que pode ser somente a detecção ou o seu posicionamento. Quando o objetivo for apenas a detecção, a metodologia da marca simples é eficiente. Corroborando tais afirmações, encontra-se trabalhos que utiliza metodologia da marca simples (Santos et al., 2007; Mignon et al., 2009) entretanto o método do intervalo simples tem sido mais apropriado por proporcionar informações sobre o posicionamento e os efeitos dos genes sobre a característica (McClosky et al., 2011; Borovsky & Paran, 2011).

Observou-se que, de maneira geral, com o aumento do tamanho populacional houve maior eficiência na metodologia de identificação dos genes controladores da característica, chegando a 100% na identificação dos genes para o tamanho populacional de 600 indivíduos. Para o tamanho populacional de 100 indivíduos foi verificado que o emprego da análise genômica para elucidar questões sobre o número e tipo de ação de genes sobre o controle da característica foi ineficiente apenas para a segregação 63:1. Além disso, 10%, 47,5% e 30% dos genes não foram encontrados para as proporções 13:3; 12:3:1 e 27:37, respectivamente e que os demais foram posicionados no local esperado ou próximo. Para os tamanhos populacionais de 200 e 400 indivíduos foram encontrados genes falsos (genes fantasmas) além dos esperados, fato este resolvido também com o tamanho populacional.

Na Tabela 3 são apresentados os valores teóricos de **m**, **a** e **d** calculados com base nos valores fenotípicos de cada classe genotípica para o gene controlador da característica. Assim, pode-se argumentar que a análise genômica seria eficiente em detectar o número de genes controladores da característica, mas que omitiria informação sobre o tipo de ação gênica, em especial das epistasias. Na genética clássica, ao se testar hipótese de segregação 9:7 já há indicação de se tratar de dois genes independentes e epistasia recessiva dupla. Para a análise genômica, o tipo de epistasia pode ser reconhecido a partir da magnitude dos efeitos médios, ou individuais, dos locos denotados por **m**, **a** e **d**. Quando se considerou genes ligados, os valores de **m**, **a** e **d** foram calculados admitindo que a posição dos genes eram conhecidas e estavam localizados a 10 cM de distância (valor de  $r=0,1$ ), caracterizando então os valores de **u**, **a** e **d** médios para as características estudadas representativas dos 10 tipos de segregação. Assim, a comparação dos valores desta tabela com as estimativas médias obtidas para **u**, **a** e **d** para os 2 cenários (caracteres controlados por genes independentes

ou ligados) aumentará a informação da análise genômica pelo fato de possibilitar caracterizar também o tipo de segregação que está ocorrendo com a característica em estudo.

Na Tabela 4 encontram-se descritos os valores médios de **m**, **a** e **d** obtidos para genes independentes controladores das características com padrão de segregação variado. Estes resultados, quando comparados com os da Tabela 3, revelam grande coincidência de magnitude, ou pouca variação em relação a classe estudada com forte indicativo de que os valores médios sejam, de fato, informativos do padrão de segregação epistáticos. Para as segregações 3:1 e 1:2:1 os valores são todos coincidentes com os estimados e descritos na Tabela 3, com exceção para o tamanho populacional de 400 indivíduos em que o valor médio de **a**, para  $Y_{1:2:1}$ , foi de 0,87. Para as características  $Y_{13:3}$  e  $Y_{15:1}$  foram encontrados valores discrepantes para **a**, no tamanho populacional de 400 indivíduos. No entanto, para o tamanho populacional de 600 indivíduos, esta variação é menor.

A característica  $Y_{27:37}$  foi a que apresentou maiores variações em relação aos valores esperados de **m**, **a** e **d**, mas com o aumento da população os valores médios se aproximaram do valor esperado.

No cenário em que se consideraram genes ligados, tendo-se 2 ou 3 genes controlando a característica, alocados a distância de 10 e 20 cM, pode-se observar que, com o aumento do tamanho da população, não houveram grandes mudanças na não rejeição de  $H_0$  expresso por  $Pc1$  (Tabela 2). Em quase a totalidade dos casos, o teste de qui-quadrado falhou, na identificação da suposta segregação prevista, como esperado. O surpreendente foram os resultados da hipótese de segregação para as características  $Y_{12:3:1}$  e  $Y_{27:37}$ . A primeira, que com tamanho populacional de 100 indivíduos, estava segregando conforme o falso esperado em 100% dos casos, mas com o aumento do

tamanho da população o valor de  $Pc1$ , obtido por meio do teste de qui-quadrado, diminuiu para 65% dos casos positivos. No caso da segunda característica ( $Y_{27:37}$ ) obteve-se  $Pc1$  de 75%, com tamanho populacional 100, reduzindo para 30%, para tamanho populacional de 200, e atingindo zero para os tamanhos populacionais de 400 e 600 indivíduos.

Para o tamanho populacional de 100 indivíduos (Tabela 2) foi ainda constatado três testes falsos positivos para a característica  $Y_{63:1}$  ( $Pc1 = 15\%$ ), dois para  $Y_{9:7}$  e apenas um para  $Y_{9:6:1}$ . Para o tamanho populacional de 200 indivíduos, verificou-se apenas um teste de segregação positivo para a característica  $Y_{9:7}$ .

A hipótese genética é definida como o mecanismo de herança mais provável de uma característica com base na avaliação de dados disponíveis. Assim, como ressaltado anteriormente, quando se considera genes ligados tem-se na realidade “falsas hipóteses de segregação”, ou seja, as características  $Y_{9:7}$  seqüencialmente até  $Y_{63:1}$  não tem proporções de segregação esperada. Neste caso os genes não estão segregando independente e o valor da porcentagem de recombinação é desconhecido. Assim, este fato justifica o resultado encontrado para a grande maioria dos testes de segregação com valores de  $Pc1$  abaixo de 5% e grande quantidade de ambigüidade de segregação.

Em todos os tamanhos populacionais houve ambigüidade de segregação, para o conjunto de características governadas por genes ligados. Observa-se que houveram casos positivos de ambigüidade de segregação para todas as hipóteses testadas. A exceção foi a característica  $Y_{12:3:1}$ , e que, em relação às hipóteses alternativas, foram encontrados seis padrões concorrentes com valores maiores que 85% de número de testes de segregação falso positivo (erro tipo I). Para os demais conjuntos populacionais não foram encontrados mais de uma hipótese concorrente. Como ocorreram elevados índices de erro tipo I para as hipóteses testadas (conhecidas) e também elevados índices

de erro tipo II para as hipóteses concorrentes, conseqüentemente a porcentagem do número de testes concorrentes positivos é elevada e quando comparados a hipótese supostamente conhecida com a concorrente em termos de maior probabilidade.

A eficácia da análise genômica nos testes de identificação do número de genes controladores das características foi comprovada para genes ligados (Tabela 2). Estas análises foram eficientes na quantificação e na identificação dos genes responsáveis pela característica. Pode-se observar que, de maneira geral, com o aumento do tamanho populacional houve melhor eficiência da metodologia de identificação de genes, chegando a 100% na identificação dos mesmos para o tamanho populacional de 600 indivíduos. Para o tamanho populacional de 100 indivíduos foi verificado grande quantidade de testes de identificação de genes negativos, com destaque para as características  $Y_{15:1}$ ;  $Y_{27:37}$  e  $Y_{63:1}$ . Este problema foi resolvido com o aumento do tamanho populacional, com 200 indivíduos foi constatado apenas 2,5% (Pg2) de testes negativos para  $Y_{9:7}$  e 12,5 para  $Y_{13:3}$ . Tais resultados são coerentes com alguns obtidos em análise genômica de detecção de QTLs que ressaltam que o tamanho populacional de 200 indivíduos é apropriado para este fim. São encontrados pesquisas visando à detecção de QTLs utilizando populações de até 200 indivíduos (Couto et al., 2010; Borovsky e Paran, 2011; Lu et al., 2011) e acima de 200 indivíduos (Ding et al., 2011; Sousa et al., 2011).

Nos grupos de características e populações de tamanho 400 e 600 foram encontrados resultados com genes extras (ou genes fantasmas), chegando a 37,5% dos testes para a característica  $Y_{9:7}$  com presença de um gene extra e 33,3% para a  $Y_{63:1}$ , podendo inferir que este fato deva-se ao espaçamento entre as marcas ou a metodologia aplicada. Problemas similares são encontrados em estudos de QTL, cuja elucidação da

ocorrência de QTL fantasma tem sido feito por meio do uso do método do intervalo composto.

Os valores médios de **m**, **a** e **d** para os conjuntos de caracteres controlados por genes ligados estão apresentados na Tabela 4. Esse dados quando comparados aos valores referenciais da Tabela 3, permite constatar grande coincidência ou valores próximos com pouca variação em relação a classe estudada e com forte indicativo de que os valores médios destes sejam indicativo da proporção de segregação. Assim, com base no valor médio de **m**, **a** e **d** pode-se inferir sobre o tipo de controle genético da característica. Conclusões desta natureza não seriam possíveis de serem obtidas por meio da análise genética clássica devido a impossibilidade do estabelecimento de uma hipótese de segregação que contemplasse informações sobre o número de genes ligados, distância entre estes e o padrão epistático. Observa-se que, com o aumento do tamanho populacional, os valores médios de **m**, **a** e **d** para cada classe de característica se aproximaram do valor paramétrico (Tabela 4).

## **Conclusões**

- Os testes de segregação empregados em genética clássica foram eficientes na identificação do padrão de herança de característica quando se tratava de genes independentes, porém problemas de ambigüidade de segregação foram encontrados em tamanhos populacionais de 100 a 600 indivíduos.

- Testes clássicos de segregação quando aplicados em características controladas por genes ligados são inconclusivos e proporcionam grandes equívocos em análise e interpretação.

- Os testes de identificação de genes em análises genômica, a partir de informações fenotípicas e moleculares da F<sub>2</sub>, se mostraram eficientes tanto para estudo de herança de características controlada por genes independentes quanto ligados.

- Delineamento genético para fins de estudo de herança envolvendo informações fenotípicas e moleculares das gerações segregantes devem ser recomendados por permitir o emprego de análise de genética clássica e genômica gerando informações complementares e reduzindo equívocos nos casos de ocorrência de segregações genéticas ambíguas.

## Referências

ALFENAS A.C., ZAUZA, E.A.V., MAFIA R.G., ASSIS T.F. **Clonagem e doenças do Eucalipto**. Viçosa, MG. Editora UFV, 2004. 442p.

BHERING, L. L.; CRUZ, C. D. Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. **Pesquisa agropecuária brasileira**, v.43, p.379-385, 2008.

BOROVSKY, Y.; PARAN, I. Characterization of *fs* 10.1, a major QTL controlling fruit in elongation in *Capsicum*. **Theor Appl Genet**, 2011. doi: 10.1007/s00122-011-1615-7

CAPUCHO, A. S.; CAIXETA, E. T.; MACIEL-ZAMBOLIM, E.; ZAMBOLIM, L. Herança da resistência do Híbrido de Timor UFV443-03 à ferrugem-do-cafeeiro. **Pesq. Agropec. Bras.**, Brasília, v. 44, n. 3, p. 276-282, 2009.

COUTO, K. R.; SANTOS, J. B.; RAMALHO, M. A. P.; SILVA, G. S. Identificação de marcadores microssatélites relacionados ao escurecimento do grão de feijão. **Pesquisa Agropecuária Brasileira**, v.45, n.11, p.1268-1274, 2010.

CRUZ, C. D. **Programa para análise de dados moleculares e quantitativos – GQMOL - Versão 2012.1**. Viçosa: UFV, 2012.

DING, J. Q.; MA, J. L.; ZHANG, C. R.; DONG, H. F.; XI, Z. Y.; XIA, Z. L.; WU, J. Y. QTL mapping for test weight by using F<sub>2:3</sub> population in maize. **Journal of Genetics**, v.90, n.1, p.75-80, 2011

HARUSHIMA Y., YANO M., SHOMURA A., SATO M., SHIMANO T., KUBOKI, Y., YAMAMOTO T., LIN S.Y., BALTAZAR A.A., PARCO A., KAJIYA H., HUANG N., YAMAMOTO K., NAGAMURA Y., KURATA N., KHUSH G.S., SASAKI T.A. high-density rice genetic linkage map with 2275 markers using a single F<sub>2</sub> population. **Genetics**, v.148, p.479-494, 1998.

JUNGHANS D. T., ALFENAS A. C., BROMMONSHENKEL S. H., ODA S., MELLO E. J., GRATTAPAGLIA D. Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**. v.108, p.175-180, 2003.

KUKHANG, T. D.; MAWARDI, S.; ESKES, A. B. Studies on the inheritance of the SH3 resistance factor to coffee leaf rust. In: **Colloque Scientifique International sur Le Café**, ASIC, 1993, Montpellier, França. **Resumos...** Montpellier, 1993. p. 776-778

LIU B.H., **Statistical Genomics: linkage mapping and QTL analysis**. Boca raton, Florida, USA:CRC Press, 1997. 610p.

LU, B.; XIE, K.; YANG, C.; ZHANG, L.; WU, T.; LI, L.; LIU, X.; JIANG, L.; WAN, J. Efficient QTL detection for heading date in bechcross inbred line and F<sub>2</sub> population derived from the same rice cross. **African Journal of Agriculture Research**, v.6, n.10, p.2372-2378, 2011.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, USA: Sinauer Associates, Inc.Ed., 1998, 978p.

MCCLOSKEY, B.; MA, X.; TANKSLEY, S.D. Quantifying the Relative Contribution of the Heterozygous Class to QTL Detect Power. **Statistical Application in Genetics and Molecular Biology**, v.10, n.1, 2011. doi: 10.2202/1544-6116.1622

MIGNON, G.L.; PITEL, F.; GILBERT, H.; BIHAN-DURVAL, E.L.; VIGNOLES, F.; DEMEURE, O.; LAGARRIGUE, S.; SIMON, J.; COGBURN, L.A.; AGGREY, S.E.; DOUAIRE, M.; ROY, P.L. A comprehensive analysis of QTL for abdominal fat and breast muscle weights on chicken chromosome 5 using a multivariate approach. **Animal Genetics**, v.40, p.157-164, 2009.

PEREIRA, A. A. **Herança da resistência a *Hemileia vastatrix* Berk. et Br. em cafeeiros derivados do Híbrido de Timor**. 1995. 66 f. Tese (Doutorado em Fitopatologia) – Universidade Federal de Viçosa, Viçosa, MG, 1995.

SANTOS, R. M. F.; LOPES, U. V.; BAHIA, R. C.; MACHADO, R. C. R.; AHNERT, D.; CORREA, R. X. Marcadores microssatélites relacionados com a resistência à vassoura-de-bruxa do cacaueteiro. **Pesquisa agropecuária brasileiro**, v.42, n.8, p.1137-1142, 2007.

SCHUSTER I., CRUZ C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos**. Viçosa , MG. Editora UFV, 2004. 585p.

SOUSA, K. R. S.; GUIMARÃES, S. E. F.; SILVA FILHO, M. I.; LOPES, M. S.; PINTO, A. P. G.; VERANDO, L. L.; BRACCINI NETO, J.; LOPES, P. S. Mapeamento de locos de características quantitativas nos cromossomos 5, 7 e 8 de suínos. **Revista Brasileira de Zootecnia**, v.40, n.1, p.115-123, 2011.

VALE, F. X. R.; ZAMBOLI, L. (ED) **Controle de doenças de plantas – Grandes culturas**. v.2, Viçosa, MG, UFV, Depto. de Fitopatologia; Brasília, Ministério da Agricultura, 1997, 1065p.

VIJAYALAKSHMI S., YADAV K., KUSHWAHA C., SARODE S.B., SRIVASTAVA C.P., CHAND R., SINGH B.D. Identification of RAPD markers linked to the rust (*Uromyces fabae*) resistance gene in pea (*Pisum sativum*). **Euphytica**, v.144, p.265-274, 2005.

WEIGUO L., MING Y.Z., ENRIQUE A.P., CALVIN F.K. Highly efficient doubledhaploid production in wheat (*Triticum aestivum* L.) via induced microspore embryogenesis. **Crop Science**, v.42, p.682-692, 2002.

WILCOX P.L., AMERSON H.V., KUHLMAN E.G., LIU B., O'MALLEY D.M., SEDEROFF R.R. Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. **Proceedings of the National Academy of Sciences**, v.93, p.3859-3864, 1996.

YAN, A.; ZHI-QIU, H.; ZAI-XIANG, T.; XUE-FENG, W.; CHEN-WU, X. A General Method for QTL Mapping in Multiple Related populations Derived from Multiple Parents. **Rice Science**, v. 16, n. 1, 2009.

YOUNG, N. D. **Constructing a plant genetic linkage map with DNA markers**. In: PHILIPS, P. L.; VASIL, I. K. DNA-Based Markers in plants. Dordrecht, The Netherlands: Kluwer Academic Publisher, p.39-57, 1994.

ZOBEL, B., JETT J. B. **Genetics of wood production**. Berlin: Springer-Verlag, 1995. 337p.

Tabela 1- Compilação do padrão de herança por métodos de genética clássica e por análise genômica para caracteres controlados por um, dois ou três genes independentes.

Caráter	Genética Clássica				Genômica			
	Pc1	Ha	Pc2	Pc3	Pg1.	Pg2	Pg3	Pg4
N=100								
Y <sub>(3:1)</sub>	100	13:3	50	25	50	-	-	50
Y <sub>(1:2:1)</sub>	100	9:3:4	90	40	40	-	-	60
Y <sub>(9:7)</sub>	90	27:37	95	50	22,5	-	-	77,5
Y <sub>(13:3)</sub>	100	3:1	80	25	10	10	-	80
Y <sub>(15:1)</sub>	100	63:1	15	15	12,5	-	-	87,5
Y <sub>(9:6:1)</sub>	100	-	-	-	17,5	-	-	82,5
Y <sub>(12:3:1)</sub>	100	-	-	-	7,5	47,5	2,5	42,5
Y <sub>(9:3:4)</sub>	100	1:2:1	70	20	22,5	-	-	77,5
Y <sub>(27:37)</sub>	100	9:7	100	90	31,7	30	-	38
Y <sub>(63:1)</sub>	95	15:1	55	5	-	100	-	-
N=200								
Y <sub>(3:1)</sub>	100	13:3	60	20	90	-	-	10
Y <sub>(1:2:1)</sub>	100	9:3:4	70	40	90	-	-	10
Y <sub>(9:7)</sub>	100	27:37	95	40	60	-	40	-
Y <sub>(13:3)</sub>	100	3:1	45	-	45	12,5	12,5	30
Y <sub>(15:1)</sub>	100	-	-	-	32,5	10	57,5	-
Y <sub>(9:6:1)</sub>	100	-	-	-	60	-	40	-
Y <sub>(12:3:1)</sub>	100	-	-	-	40	45	15	-
Y <sub>(9:3:4)</sub>	100	1:2:1	60	15	57,5	12,5	25	5
Y <sub>(27:37)</sub>	90	9:7	100	-	41,7	33,3	21,7	3,3
Y <sub>(63:1)</sub>	100	-	-	-	15	61,7	-	23,3
N=400								
Y <sub>(3:1)</sub>	95	13:3	15	5	100	-	-	-
Y <sub>(1:2:1)</sub>	100	9:3:4	40	15	85	-	15	-
Y <sub>(9:7)</sub>	100	27:37	80	25	87,5	-	-	12,5
Y <sub>(13:3)</sub>	95	3:1	15	5	70	-	-	30
Y <sub>(15:1)</sub>	100	-	-	-	37,5	-	2,5	60
Y <sub>(9:6:1)</sub>	100	-	-	-	82,5	-	2,5	15
Y <sub>(12:3:1)</sub>	95	-	-	-	52,5	12,5	-	35
Y <sub>(9:3:4)</sub>	100	1:2:1	25	5	72,5	-	2,5	25
Y <sub>(27:37)</sub>	25	9:7	70	65	71,7	3,3	-	25
Y <sub>(63:1)</sub>	100	-	-	-	38,3	5	1,7	55
N=600								
Y <sub>(3:1)</sub>	100	-	-	-	100	-	-	-
Y <sub>(1:2:1)</sub>	95	-	-	-	100	-	-	-
Y <sub>(9:7)</sub>	100	27:37	95	45	100	-	-	-
Y <sub>(13:3)</sub>	100	-	-	-	90	-	-	10
Y <sub>(15:1)</sub>	95	-	-	-	72,5	-	-	27,5
Y <sub>(9:6:1)</sub>	95	-	-	-	90	-	-	10
Y <sub>(12:3:1)</sub>	100	-	-	-	85	-	-	15
Y <sub>(9:3:4)</sub>	100	1:2:1	10	-	95	-	-	5
Y <sub>(27:37)</sub>	10	9:7	25	25	100	-	-	-
Y <sub>(63:1)</sub>	90	-	-	-	100	-	-	-

Pc1: percentagem de teste qui-quadrado que não rejeitou a hipótese verdadeira a 5% de probabilidade; Ha: hipótese alternativa ou concorrente; Pc2: percentagem de teste qui-quadrado que não rejeitou a hipótese concorrente a 5% de probabilidade; Pc3: percentagem de teste qui-quadrado que não rejeitou ambas hipóteses, mas com a concorrente associada a maior nível de significância; Pg1: percentagem de análise genômica em que os genes foram detectados na quantidade e na posição exata; Pg2: percentagem de análise em que os genes não foram detectados; Pg3: percentagem de análise que detectou genes extras (fantasma) além do esperado; Pg4: percentagem de análise em que os genes foram detectados em quantidade correta mas em posição aproximada. Pg1+Pg2+Pg4=100.

Tabela 2- Compilação do padrão de herança por métodos de genética clássica e por análise genômica para caracteres controlados por um, dois ou três genes ligados

Caráter	Genética Clássica				Genômica			
	Pc1	Ha	Pc2	Pc3	Pg1.	Pg2	Pg3	Pg4
N=100								
Y <sub>(9:7)</sub>	10	13:3-15:1- 27:37	85-15- 25	85-15- 25	27,5	27,5	45	-
Y <sub>(13:3)</sub>	0	15:1 - 63:1	100 35	100 - 35	30	65	5	-
Y <sub>(15:1)</sub>	0	3:1 - 13:3	95 - 95	95 - 95	60	17,5	22,5	-
Y <sub>(9:6:1)</sub>	5	12:3:1 - 9:3:4	85 -10	85 - 10	-	50	50	-
Y <sub>(12:3:1)</sub>	100	9:6:1	5	-	22,5	30	47,5	-
Y <sub>(9:3:4)</sub>	0	9:6:1 - 12:3:1	25 - 75	25 - 75	17,5	40	42,5	-
Y <sub>(27:37)</sub>	75	3:1 - 9:7	45 - 50	45 - 0	41,7	37,6	26,7	-
Y <sub>(63:1)</sub>	15	3:1 - 13:3	45 - 100	45 - 100	33,3	6,7	60	-
N=200								
Y <sub>(9:7)</sub>	5	27:37	65	60	2,5	2,5	-	95,0
Y <sub>(13:3)</sub>	0	3:1	95	95	17,5	12,5	-	70
Y <sub>(15:1)</sub>	0	13:3	100	100	2,5	-	-	97,5
Y <sub>(9:6:1)</sub>	0	9:3:4	50	45	-	-	-	100
Y <sub>(12:3:1)</sub>	95	-	-	-	7,5	-	-	92,5
Y <sub>(9:3:4)</sub>	0	1:2:1	25	25	15	-	-	85
Y <sub>(27:37)</sub>	30	9:7	20	-	10	-	-	90
Y <sub>(63:1)</sub>	0	15:1	95	45	1,7	-	-	98,3
N=400								
Y <sub>(9:7)</sub>	0	3:1	50	50	-	-	10	90
Y <sub>(13:3)</sub>	0	15:1	75	75	72,5	-	2,5	25,0
Y <sub>(15:1)</sub>	0	3:1 - 13:3	20 - 95	20 - 95	-	-	10	90
Y <sub>(9:6:1)</sub>	0	12:3:1	20	20	-	-	17,5	82,5
Y <sub>(12:3:1)</sub>	90	-	-	-	2,5	-	12,5	85
Y <sub>(9:3:4)</sub>	0	12:3:1	30	30	10	-	25	65
Y <sub>(27:37)</sub>	0	-	-	-	8,3	-	16,7	75,0
Y <sub>(63:1)</sub>	0	13:3	70	70	1,7	-	-	98,3
N=600								
Y <sub>(9:7)</sub>	0	3:1	25	25	37,5	-	37,5	25
Y <sub>(13:3)</sub>	0	15:1	65	65	47,5	-	25	27,5
Y <sub>(15:1)</sub>	0	3:1 - 13:3	20 - 85	20 - 85	30	-	25	45,0
Y <sub>(9:6:1)</sub>	0	12:3:1	15	15	2,5	-	-	97,5
Y <sub>(12:3:1)</sub>	65	-	-	-	50	-	10	40,0
Y <sub>(9:3:4)</sub>	0	12:3:1	5	5	50	-	7,5	42,5
Y <sub>(27:37)</sub>	0	-	-	-	20	-	33,33	46,67
Y <sub>(63:1)</sub>	0	13:3	65	65	1,67	-	20	78,33

Pc1: porcentagem de teste qui-quadrado que não rejeitou a hipótese verdadeira a 5% de probabilidade; Ha: hipótese alternativa ou concorrente; Pc2: porcentagem de teste qui-quadrado que não rejeitou a hipótese concorrente a 5% de probabilidade; Pc3: porcentagem de teste qui-quadrado que não rejeitou ambas hipóteses, mas com a concorrente associada a maior nível de significância; Pg1: porcentagem de análise genômica em que os genes foram detectados na quantidade e na posição exata; Pg2: porcentagem de análise em que os genes não foram detectado; Pg3: porcentagem de análise que detectou genes extras (fantasma) além do esperado; Pg4: porcentagem de análise em que os genes foram detectados em quantidade correta mas em posição aproximada. Pg1+Pg2+Pg4=100.

Tabela 3: Valores médios esperados de **m**, **a** e **d** determinantes da expressão fenotípica de características com diferentes padrões de segregação

Caráter	Genes independentes			Genes ligados (r =0,10cM)		
	m	a	d	m	a	d
$Y_{(3:1)}$	0,5	0,5	0,5	-	-	-
$Y_{(1:2:1)}$	1	1	0	-	-	-
$Y_{(9:7)}$	0,375	0,375	0,375	0,495	0,495	0,415
$Y_{(13:3)}$	0,75	0,125	0,125	0,95	0,045	0,005
$Y_{(15:1)}$	0,875	0,125	0,125	0,595	0,405	0,405
$Y_{(9:6:1)}$	0,625	0,25	0,25	0,545	0,45	0,41
$Y_{(12:3:1)}$	0,75	0,1875	0,1875	0,57	0,4275	0,4075
$Y_{(9:3:4)}$	0,5	0,3125	0,3125	0,52	0,4725	0,4125
$Y_{(27:37)}$	0,25	0,25	0,25	0,48	0,45	0,34
$Y_{(63:1)}$	0,95	0,041	0,041	0,66	0,31	0,31

Tabela 4- Compilação do valores médios dos efeitos **m**, **a** e **d** relativos a 20 réplicas de cada conjunto populacional simulado

Caráter	Genes independentes			Genes ligados		
	m	a	d	m	a	d
N=100						
Y <sub>(3:1)</sub>	0.50	0.50	0.51	-	-	-
Y <sub>(1:2:1)</sub>	1.00	1.01	0.00	-	-	-
Y <sub>(9:7)</sub>	0.37	0.39	0.39	0.47	0.54	0.48
Y <sub>(13:3)</sub>	0.74	0.13	0.14	0.86	0.14	0.14
Y <sub>(15:1)</sub>	0.87	0.13	0.14	0.56	0.43	0.46
Y <sub>(9:6:1)</sub>	0,62	0,25	0,26	0,49	0,5	0,51
Y <sub>(12:3:1)</sub>	0,75	0,19	0,19	0,52	0,47	0,49
Y <sub>(9:3:4)</sub>	0,49	0,32	0,32	0,48	0,51	0,46
Y <sub>(27:37)</sub>	0.39	0.18	0.25	0.47	0.50	0.40
Y <sub>(63:1)</sub>	0.97	0.03	0.03	0.62	0.38	0.41
N=200						
Y <sub>(3:1)</sub>	0.50	0.50	0.50	-	-	-
Y <sub>(1:2:1)</sub>	1.00	1.00	0.00	-	-	-
Y <sub>(9:7)</sub>	0.36	0.38	0.40	0.47	0.52	0.45
Y <sub>(13:3)</sub>	0.74	0.13	0.14	0.94	0.13	0.10
Y <sub>(15:1)</sub>	0.87	0.13	0.14	0.57	0.44	0.67
Y <sub>(9:6:1)</sub>	0,62	0,24	0,25	0,49	0,5	0,47
Y <sub>(12:3:1)</sub>	0,75	0,18	0,19	0,51	0,47	0,5
Y <sub>(9:3:4)</sub>	0.49	0.31	0.33	0.48	0,52	0.49
Y <sub>(27:37)</sub>	0.36	0.24	0.31	0.47	0.52	0.40
Y <sub>(63:1)</sub>	0.97	0.03	0.04	0.62	0.37	0.40
N=400						
Y <sub>(3:1)</sub>	0.50	0.50	0.50	-	-	-
Y <sub>(1:2:1)</sub>	1.00	0.87	0.00	-	-	-
Y <sub>(9:7)</sub>	0.37	0.38	0.38	0.48	0.52	0.45
Y <sub>(13:3)</sub>	0.75	0.33	0.13	1.39	0.06	0.00
Y <sub>(15:1)</sub>	0.87	0.38	0.14	0.58	0.42	0.44
Y <sub>(9:6:1)</sub>	0,62	0,37	0,26	0,49	0,50	0,51
Y <sub>(12:3:1)</sub>	0,74	0,38	0,20	0,51	0,47	0,49
Y <sub>(9:3:4)</sub>	0,72	0,40	0,32	0,48	0,52	0,49
Y <sub>(27:37)</sub>	0.80	0.45	0.30	0.46	0.52	0.42
Y <sub>(63:1)</sub>	0.97	0.31	0.21	0.64	0.36	0.39
N=600						
Y <sub>(3:1)</sub>	0.50	0.50	0.50	-	-	-
Y <sub>(1:2:1)</sub>	1.00	1.00	0.00	-	-	-
Y <sub>(9:7)</sub>	0.37	0.38	0.39	0.50	0.49	0.41
Y <sub>(13:3)</sub>	0.75	0.12	0.13	0.94	0.05	0.02
Y <sub>(15:1)</sub>	0.87	0.13	0.14	0.60	0.40	0.40
Y <sub>(9:6:1)</sub>	0,62	0,25	0,26	0,50	0,49	0,41
Y <sub>(12:3:1)</sub>	0,75	0,19	0,19	0,55	0,44	0,43
Y <sub>(9:3:4)</sub>	0,5	0,33	0,32	0,50	0,46	0,44
Y <sub>(27:37)</sub>	0.36	0.25	0.31	0.49	0.48	0.36
Y <sub>(63:1)</sub>	0.96	0.04	0.04	0.67	0.32	0.33

**CAPÍTULO II**  
**ABORDAGEM SOBRE ESTUDO DE HERANÇA DE CARATERES PELA**  
**GENÉTICA CLÁSSICA E MOLECULAR PARA PADRÃO**  
**MULTICATEGÓRICO DE EXPRESSÃO FENOTÍPICA**

Resumo- Características oligogênicas de distribuição discreta e expressão governada por poucos genes de maior efeito têm se mostrado importantes na condução dos programas de melhoramento. O trabalho teve por objetivo caracterizar hipóteses genéticas de maior similaridade utilizadas para prever padrão de herança de características oligogênicas quando se utilizou um padrão multicategórico de dados fenotípicos. Foi utilizado o delineamento genético clássico, envolvendo população  $F_2$  derivada de  $F_1$  proveniente de cruzamento entre pais homocigotos contrastantes, de característica controlada por um, dois e três genes independentes e ligados. A este estudo foi também considerada a agregação de informações moleculares de indivíduos  $F_2$  possibilitando o emprego de análises genômicas para fins de complementação de informações e elucidação de testes de hipóteses no caso de existir segregação ambígua em hipóteses concorrentes. Foram simulados genomas parentais e populações  $F_2$  de 100, 200, 400 e 600 indivíduos, fenotipados e genotipados em relação a cinco grupos de ligação contendo 11 marcadores co-dominantes por grupo de ligação, e saturação de 10cM entre marcadores. Testes de genética clássica, baseados em qui-quadrado, foram incapazes de identificar corretamente o padrão de herança de características controladas por genes ligados e proporcionaram grande percentual de erro, principalmente em tamanhos populacionais mais reduzidos, em teste com hipóteses com segregações concorrentes próximas tais como 3:1 e 13:3, 9:7 e 37:27 e 1:2:1 e 3:9:4. As análises genômicas foram realizadas pelas metodologias de intervalo simples aplicadas aos estudos de detecção de QTL (quantitative trait loci) demonstrando seu potencial na identificação dos genes, posicionamento e quantificação de seus efeitos sobre a expressão da característica.

Termos para indexação: simulação; genômica; ambigüidade de segregação; identificação de QTL; dados multicategóricos.

## **Introdução**

Algumas etapas devem ser seguidas para o estudo da herança monogênica ou oligogênica. Deve-se escolher genitores contrastantes para o caráter em questão, realizar os cruzamentos e obter as gerações  $F_1$  e  $F_2$ . Pode-se também obter os retrocruzamentos (RC), ou seja, o cruzamento da geração  $F_1$  com ambos ou apenas um dos genitores. Posteriormente, avaliam-se os genitores e todas as gerações de cruzamentos obtidas e verifica-se, por meio da análise da distribuição fenotípica de caracteres qualitativos ou por meio das médias e variâncias de caracteres quantitativos, o controle genético com base nas leis mendelianas. Sendo os genitores puros, também não haverá segregação na geração  $F_1$ , mas apenas nas demais ( $F_2$  e retrocruzamentos). Deve-se testar a hipótese de segregação com o teste estatístico de qui-quadrado ( $\chi^2$ ), o qual verifica se os desvios apresentados entre o número observado e o esperado, de cada classe fenotípica, são significativos em determinado nível de probabilidade (normalmente, 5%) (Ramalho et al., 2004).

Características oligogênicas, de distribuição discreta e expressão governada por poucos genes de maior efeito, têm se mostrado importantes na condução dos programas de melhoramento de várias culturas (Rao & Li, 2000). Para o melhoramento vegetal, a resistência das plantas à ação de patógenos é provavelmente a característica mais importante de natureza oligogênica (Alzate-Marin et al., 2005; Barbosa-Prestes et al., 2008; Gagliardi & Camargo, 2009). Tais características são governadas por poucos genes, ou pelo menos de alguns de efeitos maiores, sendo importantes para várias culturas. As características oligogênicas possuem padrão de herança mendeliano governado por dois, três ou mais genes de interação epistática, podendo seu padrão de expressão fenotípico alocado em classes a exemplo quando se atribui notas a severidade ao ataque de doenças.

As variáveis multicatóricas são bastante utilizadas na caracterização de cultivares. Essas variáveis estão relacionadas com particularidades morfológicas e estruturais da planta, além de características que conferem qualidade ao produto comercializado e quantidade de infecção por doenças em plantas. Essas características sofrem pouca influência do ambiente por serem de herança simples, além disso são relativamente fáceis e rápidas de se mensurar (Cruz & Carneiro, 2006).

A coleta de dados multicatóricos tem sido de maior interesse pelos pesquisadores pela vantagem, em relação à coleta de dados moleculares e quantitativos, de serem facilmente observados e requererem menos tempo e mão de obra na sua realização (Sudré et al., 2006). Entretanto, os limites estabelecidos entre uma e outra classe podem não ser tão facilmente identificados e, assim, as informações coletadas poderão ser carregadas de certa imprecisão tendo em vista a subjetividade de classificação realizada pelo avaliador. Outro aspecto a ser considerado é que não se pode negar que a avaliação, por se consistir também na emissão de um juízo de valor, está necessariamente permeada pela visão de quem avalia. Portanto, a explicitação de critérios e de suas conseqüências contidos no processo de avaliação, inclusive na interpretação dos resultados, são os caminhos indicados para deixar visível a especificidade buscada na abordagem qualitativa e a sua validade.

A idéia de que apenas o que pode ser expresso em números é permeado com a objetividade exigida para dar veracidade à avaliação vigora no meio científico, porém a análise e conclusões obtidas no processo de avaliação adotando-se a abordagem quantitativa também não estão isentas de imprecisões e subjetividades. O mais importante é ser rigoroso, criterioso e atento na execução da avaliação e deixar sempre claro quais os elementos que foram adotados e que permitiram as conclusões obtidas.

O uso de descritores multicategóricos, com mais de duas classes por variável, tem sido utilizado como ferramenta para estruturação da variabilidade genética. Técnicas de análises multivariadas, como por exemplo a análise de agrupamento, tem sido empregadas para a quantificação da divergência fenotípica em várias espécies, como por exemplo hortaliças (Barelli et al., 2007; Bertan et al., 2006 ; Pereira et al., 2003; Sudré et al., 2006) e na caracterização fenotípica de padrões por infecção a doenças e posterior identificação de QTL (Kamei et al., 2010; Palomares-Rius., 2011; Sharma et al., 2011) considerando dados de características multicategóricas.

No estudo da herança de característica qualitativa, por meio da genética clássica, é de interesse do pesquisador conhecer o número de genes e alelos envolvidos no controle e as suas interações para a expressão da característica. Entretanto, a pormenorização destes estudos permitindo a detecção, posição relativa e a quantificação dos efeitos individuais desses genes nos cromossomos, assim com a sua relação com outros genes só foi possível a partir da disponibilidade de mapas genéticos obtidos pela análise de um número relativamente grande de marcadores abrangendo todo o genoma da espécie. Assim, as análises genômicas passaram a ter também papel fundamental na disponibilização ferramental útil em vários campos da pesquisa genética, uma vez que permitem a visualização, mesmo que de forma relativa, da organização dos genes nos cromossomos.

Modernas ferramentas de análises genéticas não tem sido aplicadas efetivamente ou contemplada a possibilidade de seu uso durante o estabelecimento e a condução de experimentos. A análise de experimentos clássicos como teste de segregação aliado a análise genômica poderia solucionar definitivamente o problema de erro tipo II para características oligogênicas, trazendo maior confiabilidade sobre o número de genes que

controlam a herança da característica principalmente no que se refere a resistência de plantas a pragas e doenças.

O objetivo deste trabalho é o de caracterizar as hipóteses genéticas de maior similaridade utilizadas para prever o padrão de herança de características oligogênicas em experimentos envolvendo o delineamento de gerações  $P_1$ ,  $P_2$ ,  $F_1$  e  $F_2$  por simulação de dados genotípicos e fenotípicos para características multicategóricas. Erros de rejeição de hipóteses verdadeira, em particular com hipóteses concorrentes próximas tais como 3:1 e 13:3 ou 27:37 e 9:7, por exemplo, serão analisados e discutidos. Também é avaliada a contribuição da análise genômica, com a agregação de informações da genotipagem da  $F_2$ , na elucidação do padrão de segregação além de prover informações adicionais sobre posicionamento e quantificação do efeito de locos individuais sobre a expressão da característica.

## **Material de Métodos**

### *Material Genético*

Neste trabalho foram utilizadas, para fins de estudo de herança, populações das gerações  $F_1$ , derivada de pais homocigotos contrastantes, e  $F_2$ , e informações fenotípicas e moleculares desta geração segregante.

### *Informações moleculares da $F_2$*

As informações fenotípicas e moleculares das populações  $F_2$  foram obtidas de diferentes tamanhos populacionais 100, 200, 400 e 600. Considerou que esta geração foi derivada a partir de genomas parentais homocigotos contrastantes com  $P_1$  dominante e  $P_2$  recessivo, pelo módulo de simulação do programa para análise de dados moleculares e quantitativos – QQMOL (Cruz, 2012).

As informações moleculares foram geradas de genoma com cinco grupos de ligação e saturação de 11 marcas moleculares codominantes, espaçadas a 10 cM, portanto, cada grupo de ligação apresentou o tamanho de 100 cM.

### *Informações fenotípicas da F<sub>2</sub>*

Para cada população genotipada simulada também foram consideradas 10 características fenotípicas simuladas com um padrão de expressão que seriam controladas por 1, 2 ou 3 genes com segregação mendeliana variável. Para as características governadas por um gene, considerou-se as segregações 3:1 e 1:2:1; para as características governadas por dois genes, considerou-se as segregações 9:7, 13:3, 15:1, 9:6:1, 12:3:1 e 9:3:4 e, para as características governadas por três genes considerou-se a segregação 27:37 e 63:1. As características fenotípicas foram inseridas em um padrão de notas que variavam de zero a cinco.

Os genes controladores destas características foram posicionados de forma a se ter resultados de padrão de segregação típico de genes independentes e ligados. Os genes independentes foram localizados nas posições correspondentes aos marcadores 5 (para características controlada por um gene) ; 5 e 20 (para características controlada por dois genes) e 5, 20 e 50 (para características controlada por três genes). Para as características controladas por genes ligados a posição dos genes foi aquela correspondente aos marcadores 15 e 16 ou 13, 14 e 15 de modo que, o padrão de herança foi determinado selecionando-se os dados genotípicos relativos a cada marcador responsável pela característica em estudo nos diferentes conjuntos populacionais e codificando-os de acordo com o padrão de herança desejado.

Assim, caracterizando o processo de simulação para segregação do tipo 13:3, com possibilidade de extrapolação para os demais casos com 2 e 3 genes, os dados

genotípicos relativos aos marcadores 5 e 20 codificados com 0, 1 e 2 foram transformados em padrão fenotípico recebendo um determinado valor na combinação 2-2, 2-1, 1-2 e 1-1 e outro valor nas demais combinações. Com base nestes valores foi determinada a variância genotípica e, uma vez especificado um determinado valor de herdabilidade, foi possível prever a magnitude da variância ambiental e seus efeitos que ocorriam aleatoriamente sobre os genótipos determinando o valor fenotípico final, agrupado nas classes de 0 a 5.

#### *Análises estatísticas moleculares*

Após a simulação dos dados moleculares da população  $F_2$ , foi realizada a análise de segregação de locos individuais. Foi aplicado o teste de qui-quadrado, a 5% de probabilidade, para constatação da razão de segregação 1:2:1 em cada marcador para todas as populações  $F_2$ .

Para estes mesmos conjuntos de dados foram construídos mapas genéticos. Para construção dos grupos de ligação, utilizou-se a propriedade transitiva, ou seja, se o loco A está ligado ao loco B, e o loco B está ligado ao loco C, logo o loco A está ligado ao loco C, independente da frequência de recombinação de A e C e, portanto, A, B e C pertencem ao mesmo grupo de ligação. Os critérios utilizados no agrupamento foram a frequência máxima de recombinação ( $r_{\max}$ ) e o LOD mínimo ( $LOD_{\min}$ ), para inferir se dois locos estão ligados. Foram utilizados os valores 30% e 3, respectivamente, para  $r_{\max}$  e  $LOD_{\min}$ . As marcas mais próximas em relação às marcas já consideradas que atenderem aos dois critérios adotados serão incorporadas ao grupo de ligação. Assim, o processo continua investigando a porcentagem de recombinação e o LOD entre marcas e possíveis vizinhos a serem incorporados às extremidades do grupo de ligação em construção.

Para o ordenamento das marcas no grupo de ligação foi utilizado o método da soma das frações de recombinação adjacente (SARF – Sum of Adjacent Recombination Fraction) para os mapas construídos sem análise multiloco. Neste processo, a melhor ordem é aquela que apresenta menor soma de recombinações adjacentes. Considera-se a ordem original estabelecida pelo processo de agrupamento e aplica-se o algoritmo RCD (Rapid Chain Delineation), que consiste em realizar permutas entre dois marcadores vizinhos ou distantes envolvendo três ou quatro marcadores. A ordem é alterada se, após a permuta, a soma das distâncias adjacentes for reduzida. Após todas as permutas conclui-se que a melhor ordem é aquela de menor soma de distâncias adjacentes (Schuster & Cruz, 2008).

#### *Estudo do padrão de herança com base em informações fenotípicas*

O estudo da herança de cada característica, cujos dados foram obtidos por simulação considerando a ação de genes com posição previamente conhecida e efeito ambiental que conduzisse a herdabilidade de 80%, foi realizado da forma convencional em estudo de Genética Mendeliana em que se formula uma determinada hipótese de segregação e aplica-se um teste apropriado, geralmente o teste de qui-quadrado. Como as características eram multicategóricas foi necessário agrupar as classes em duas ou três categorias de acordo com a hipótese genética testada. Quando a hipótese era relativa a duas classes agrupou-se os indivíduos com notas 0, 1 e 2 numa classe e 3, 4 e 5 em outra classe. E, quando a hipótese era relativa a três classes foram agrupados os indivíduos com notas 0 e 1, 2 e 3 e 4 e 5.

A não rejeição de determinada hipótese de segregação, como por exemplo 13:3 ou 9:3:4, é indicativo de que a característica é governada por dois genes independentes e com interação epistática. Entretanto, é possível obter resultado falso (positivos ou

negativos) em razão do tamanho inadequado da população, da metodologia empregada e o tipo de informação disponível, que neste caso é apenas fenotípica.

Para comprovar a influência do tamanho populacional sobre o resultado do teste de segregação foram realizados testes de qui-quadrado para verificar a ocorrência de ambigüidade de aceitação de hipóteses (positiva para mais de um tipo de segregação) em cada população dos conjuntos relacionados ao tamanho populacional para todas as hipóteses de segregação que mostravam mesmo número de classes, em um total de 8320 testes.

Salienta-se que para características controladas por genes ligados não há uma proporção prévia a ser adotada pelo desconhecimento da porcentagem de recombinação entre estes genes. Assim, para dois genes A/a e B/b ligados em que A-B- determina um padrão fenotípico e (A-bb, aaB- e aabb) determina o outro padrão fenotípico a segregação 13:3 é considerada um “falsa hipótese de segregação” pois os genes estavam localizados em mesmo cromossomo, ou seja, a segregação não era independente.

A adequação da metodologia clássica para estudo de herança, apenas com base em informação fenotípica, foi avaliada pela quantidade de erros e acertos da não rejeição da hipótese de segregação verdadeira ou não rejeição simultânea de duas hipóteses pela ambigüidade da razão observada de segregação.

Assim, para melhor elucidação do processo adotado neste estudo, considerou, por exemplo, uma característica sabidamente controlada por um gene e segregação simulada de 3:1. De posse dos dados fenotípicos foi avaliada a segregação 3:1 e todas as demais proporções adequadas tais como: 9:7, 13:3, 15:1, 27:37 e 63:1. Quando se esperava um padrão de segregação com três classes, como no caso de 1:2:1, foram também realizados os testes de qui-quadrado em relação 9:6:1, 12:3:1 e 9:3:4. Assim, algumas situações foram encontradas e quantificadas: i. Apenas a hipótese verdadeira

não foi rejeitada; ii. A hipótese verdadeira foi rejeitada e outra falsa não rejeitada; iii. Tanto a hipótese verdadeira quanto a falsa não foram rejeitas com nível de significância favorecendo, ou não, a hipótese verdadeira.

#### *Estudo do padrão de herança com base em informações fenotípicas e moleculares*

Neste estudo considerou que, para os mesmos conjuntos de dados em que o padrão de herança da característica fora estudado por meio de técnicas convencionais da Genética Clássica, também havia disponibilidade de mapas genéticos de forma que o estudo poderia ser aprimorado ou elucidado por meio de análises genômica demonstrando a utilidade desta para melhorar a acurácia das informações nos programas de melhoramento.

Foi considerado que em estudos de detecção de genes (ou de QTL), primeiramente, precisa-se saber qual o objetivo da pesquisa, que pode ser somente a detecção ou o seu posicionamento. Quando o objetivo for apenas a detecção, a metodologia da marca simples é eficiente.

Os conjuntos de dados provenientes de diferentes tamanhos populacionais envolvendo característica com diferentes controles gênicos foram submetidas aos procedimentos similares ao de identificação de QTLs pelo método do intervalo simples. Os resultados favoráveis esperados por esta metodologia seria o de identificar apropriadamente o número de genes controladores de cada característica, o posicionamento destes genes, principalmente quando a característica era controlada por genes ligados e no caso de quando a genética clássica mostra-se ineficiente neste estudo, identificar o padrão epistático de herança, ou seja, ser capaz, de no caso de dois genes, caracterizar apropriadamente as segregações 9:7, 15:1, 13:3, 9:3:4 e 9:6:1 ou 63:1 e 27:37 para 3 genes.

Os resultados não completamente satisfatórios por esta metodologia consistiam em: não identificação; identificação do número de genes correto porém na posição próxima a esperada; identificação do genes na posição certa com ocorrência de gene fantasma ou “extrapolado” e identificação do número e posição incorretos.

Dados relativos a valores de cada genes expressos por **m**, **a** e **d** foram analisados, em que **m** representa a média fenotípica dos genótipos homozigotos, ou o ponto médio, **a** mede o afastamento dos homozigotos em relação ao ponto médio, efeito aditivo e **d** mede o afastamento do heterozigoto em relação ao ponto médio. As relações entre **a** e **d** permite conhecer o tipo de interação alélica presente, em que **d/a** mede o grau de dominância (gd) de um determinado loco. Os valores de **m**, **a** e **d** foram anotados para cada indivíduo de todos os 8 conjuntos populacionais e foram utilizados valores médios dos 20 indivíduos para cada padrão de segregação analisado com intuito de corroborar com as informações proporcionadas pela genômica e de caracterizar o tipo de segregação fenotípica. Todas as análises foram realizadas utilizando o aplicativo computacional GQMOL (Cruz, 2012), disponível em <http://dl.dropbox.com/u/34614950/gqmolexe.zip>  
<http://dl.dropbox.com/u/34614950/dadosgq.zip>.

## **Resultados e Discussão**

Sabe-se que, entre as populações tradicionalmente avaliadas no melhoramento de plantas, as F<sub>2</sub>s derivadas de F<sub>1</sub> (Harushima et al 1998) e as populações de retrocruzamento (Weiguo et al 2002) são intensivamente utilizadas para o mapeamento genético e estudo de padrões de herança fenotípica por sua praticidade e facilidade de obtenção (Schuster & Cruz 2004). Para estabelecer o tamanho apropriado da população

considerou-se a informação na literatura em que amostra de tamanho reduzido, inferiores a 50 indivíduos são inadequadas para avaliações fenotípicas e proporciona baixa resolução em estudo de mapeamento genético tornando difícil a detecção de genes (ou QTLs) de pequeno efeito (Young, 1994). Atualmente, para estudos moleculares, o tamanho das populações utilizadas tem sido maiores. São encontrados trabalhos que estudam a influência do tamanho populacional na recuperação dos grupos de ligação, tanto para cruzamento controlado quanto para populações exogâmicas, sendo verificado que em populações de 100 indivíduos recupera as informações genômicas, dependendo da saturação (Bhering & Cruz, 2008; Yan et al., 2009).

Na Tabela 1 estão apresentados os resultados de rejeição, ou não, de hipótese de segregação para 10 características cujo controle gênico era previamente conhecido. Pode-se observar que, para as características controladas por um ou dois genes, a porcentagem de réplicas de simulação em que não houve rejeição da hipótese verdadeira ( $P_c1$ ) foi relativamente alta, mas o efeito ambiental, a existência de vários padrões de classes e a subjetividade no agrupamento destas classes para aplicação do teste de hipótese contribuíram para a redução da frequência de não rejeição da hipótese verdadeira. Também observa-se que aumento do tamanho da população proporcionou mudanças na não rejeição de  $H_0$ , ou seja na maioria das características estudadas com o aumento do tamanho populacional houve um pequeno decréscimo na aceitação de  $H_0$  o que pode ser explicado pela própria aleatoriedade do processo de simulação, pela herdabilidade imputada aos dados e até mesmo pelo número de classes. Salienta-se as características controladas por três e dois genes e segregação esperada de 27:37 e 9:3:4 em que o valor de  $P_{1c}$  (porcentagem de não rejeição de hipótese verdadeira) diminuiu consideravelmente com o aumento do tamanho população indicando que a dificuldade do teste de qui-quadrado em proporcionar classificação correta para estes tipos de

segregação. Pelos resultados encontrados o problema enfrentado por melhoristas em questão de ambigüidade atribuído a tamanhos inadequados de amostra é fortemente ampliado quando se acrescenta variações não genéticas e grupos de classes estabelecidas de forma subjetiva.

Os padrões de segregação 3:1 e 13:3 apresentaram 100% de aceitação tanto para a população de 200 quanto de 600, o aumento da percentagem de Pc1 com o aumento da população, como era esperado, somente ocorreu com a segregação 63:1.

Um ponto importante a ser discutido é sensibilidade do tamanho da amostra do teste de qui-quadrado e a não rejeição de hipóteses verdadeiras em tamanhos populacionais elevados só ocorrem quando o desvio, entre observado e esperado, for relativamente pequeno. Assim, como ilustração, pode-se considerar experimento com ocorrência de três eventos com frequência paramétrica  $p=q=r=0,33$ , ou seja em uma expectativa de 1:1:1. Entretanto, em um experimento 10 observações pode-se encontrar resultado iguais a  $p=0,5$ ,  $q=0,3$  e  $r=0,2$  e o teste qui-quadrado estará associado a um nível de significância estimado de 49,65% indicando a não rejeição da hipótese. Em outro experimento, com 1000 observações, pode-se encontrar  $p = 0,4$ ,  $q = 0,3$  e  $r=0,3$ , aparentemente mais apropriado, o teste de qui-quadrado estará associado a um nível de significância 0,001% indicando a rejeição da hipótese.

Diferentemente da abordagem métrica, que, em relação a uma série de pressuposições, suporta desvios em relação ao não atendimento, a consideração inapropriada de uma hipótese genética pode resultar em problemas de interpretação e o equívoco nas conclusões em relação a natureza da característica, ao genótipo dos genitores e ao direcionamento de cruzamentos com grande potencial para causar prejuízos a qualidade da pesquisa.

Liu (1997) descreve metodologia que permite obter um tamanho populacional ótimo para realizar a comparação de hipóteses genéticas de forma a discriminá-las apropriadamente considerando a distribuição de qui-quadrado e um nível crítico associado. Assim, adotando o método de Liu, hipóteses de segregação 3:1 e 13:3 demandariam 666 indivíduos para inferir sobre a hipótese mais apropriada, adotando 5% como nível crítico de probabilidade. Usando o mesmo método, a distinção entre hipóteses alternativas de 9:7 e 27:37 demandariam 23.383 indivíduos.

Na Tabela 1 verifica-se ainda que os testes de hipótese comumente adotado em estudos de genética clássica não se mostra eficiente quando se observa os valores de qui-quadrado associados à hipóteses alternativas em quase todos os casos de segregação para tamanho populacional de 100 indivíduos, com exceção para as características com padrão de segregação 1:2:1 e 9:3:4.

O problema é ainda mais crítico quando o experimento indica uma hipótese alternativa falsa como mais apropriada para descrever os dados, que neste estudo foi quantificado por  $Pc_3$ . No tamanho populacional de 100, os valores de  $Pc_3$  foram expressivos alcançando valores de 45% e 90% para as segregações 13:3 e 27:37, respectivamente. Observa-se, porém, que com o aumento populacional a ambigüidade de segregação deixa de existir para a maioria das segregações testadas com exceção para as segregações 9:7, 12:3:1 e 27:37 em que para o tamanho populacional de 600 indivíduos casos positivos foram encontrados, porém com pequenos índices de  $Pc_3$  e sempre menores que  $Pc_1$ , este fato somente não ocorreu para a característica 27:37 em que  $Pc_3 > Pc_1$  mas com indicativo de que com um tamanho populacional maior este problema poderia ser resolvido. Vale novamente ressaltar que a ocorrência de resultados desfavoráveis, associados a alto  $Pc_3$  e baixo  $Pc_1$ , foram mais pronunciados em razão dos efeitos ambientais que incidem sobre os valores fenotípicos das características e a

subjetividade na transformação dos valores multicategóricos em classes arbitrários para estudos da genética mendeliana.

A Tabela 1 também traz informações sobre identificação de QTL no intuito de verificar a quantidade de informação adicional gerada pela genômica em comparação aos métodos clássicos de estudo de características oligogênicas. Foi utilizada a metodologia do intervalo simples pois apesar de se ter relatos do uso da metodologia de marca simples em estudos desta natureza (Santos et al., 2007; Mignon et al., 2009), verifica-se que o método do intervalo simples tem sido mais apropriado por proporcionar informações sobre o posicionamento e os efeitos dos genes sobre a característica (McClosky et al., 2011; Borovsky & Paran, 2011). Pode-se observar que, de maneira geral, com o aumento do tamanho populacional há maior eficiência na metodologia de identificação dos genes controladores da característica, chegando a 100% na identificação dos genes para o tamanho populacional de 600 indivíduos. Para o estudo genômico, a utilização de classes multicategóricas favoreceu a identificação dos locos controladores das características tornando o seu uso mais eficiente mesmo sob certa influência do ambiente. Assim, para o tamanho populacional de 100 indivíduos foi verificado que o emprego da análise genômica para elucidar questões sobre o número e tipo de ação de genes sobre o controle da característica foi ineficiente com mais de 25% de Pg2 para as segregações 13:3; 12:3:1 e 27:37 com destaque para a segregação 63:1 com 90% de Pg2, problema que foi resolvido com o aumento do tamanho populacional. Para os tamanhos populacionais de 100, 200 e 400 indivíduos foram encontrados genes falsos (genes fantasmas) além dos esperados para praticamente todos os padrões de segregação testados, fato este resolvido também com o tamanho populacional.

Na tabela 3 são apresentados os valores teóricos de **m**, **a** e **d** calculados com base nos valores fenotípicos de cada classe genotípica de cada gene controlador da

característica. Assim, pode-se argumentar que a análise genômica seria eficiente em detectar o número de genes controladores da característica, mas que omitiria informação sobre o tipo de ação gênica, em especial das epistasias. Na genética clássica, ao se testar hipótese de segregação 9:7 já há indicação de se tratar de dois genes independentes e epistasia recessiva dupla. Para a análise genômica, o tipo de epistasia pode ser reconhecido a partir da magnitude dos efeitos médios, ou individuais, dos locos denotados por **m**, **a** e **d**. Quando se considerou genes ligados, os valores de **m**, **a** e **d** foram calculados admitindo que a posição dos genes eram conhecidas e estavam localizados a 10 cM de distância (valor de  $r=0,1$ ), caracterizando então os valores de **m**, **a** e **d** médios para as características estudadas representativas dos 10 tipos de segregação. Assim, a comparação dos valores desta tabela com as estimativas médias obtidas para **m**, **a** e **d** para os 2 cenários (caracteres controlados por genes independentes ou ligados) aumentará a informação da análise genômica pelo fato de possibilitar caracterizar também o tipo de segregação que esta ocorrendo com a característica em estudo.

Na tabela 4 encontram-se descritos os valores de **m**, **a** e **d** obtidos para genes independentes controladores das características com padrão de segregação variado. Estes resultados, quando comparados com os da Tabela 3, revela grande coincidência de magnitude, ou pouca variação em relação a classe estudada com forte indicativo de que os valores médios sejam, de fato, informativos do padrão de segregação epistáticos. Para as segregações 3:1 e 1:2:1 os valores são todos praticamente coincidentes com os estimados e descritos na Tabela 3. Para as características  $Y_{13:3}$  e  $Y_{15:1}$  foram encontrados valores pouco discrepantes para **a** e **d**, no tamanho populacional de 400 indivíduos. No entanto, para o tamanho populacional de 600 indivíduos, esta variação não é tão abrupta.

A característica  $Y_{63:1}$  foi a que apresentou maiores variações em relação aos valores esperados de **m**, **a** e **d**, mas com o aumento da população os valores médios se aproximaram do valor esperado. Salienta-se também que para a característica  $Y_{27:37}$  para gene marcador situado na posição 20 cM nos tamanhos populacionais de 100 e 200 não foram identificados nesta posição apresentando valores nulos para **m**, **a** e **d**.

No cenário em que se consideraram genes ligados, tendo-se 2 ou 3 genes controlando a característica, alocados a distância de 10 e 20 cM, pode-se observar que, com o aumento do tamanho da população, não houveram grandes mudanças na não rejeição de  $H_0$  expresso por  $Pc1$  (Tabela 2). Em quase a totalidade dos casos, o teste de qui-quadrado falhou, como era de se esperar, na identificação da suposta segregação prevista. O surpreendente foram os resultados da hipótese de segregação para as características  $Y_{9:3:4}$ ,  $Y_{27:37}$  e  $Y_{12:3:1}$ . A primeira, que com tamanho populacional de 100 indivíduos, estava segregando conforme o falso esperado em apenas 65% dos casos, mas com o aumento do tamanho da população o valor de  $Pc1$ , obtido por meio do teste de qui-quadrado, decaiu para 50% dos casos positivos. No caso da segunda característica ( $Y_{27:37}$ ) ocorreu resultado de  $Pc1$  de 75%, com tamanho populacional 100, reduzindo para 50%, para tamanho populacional de 200, e atingindo 15% para os tamanhos populacionais de 400 e 600 indivíduos.

Para o tamanho populacional de 100 indivíduos (Tabela 2) foi ainda constatado de dois a quatro testes falsos positivos em todas características avaliadas e em 5 dos casos de segregação estudados os valores de  $Pc3$  foram maiores que  $Pc1$  com destaque para a característica  $Y_{63:1}$   $Pc1 = 0\%$  e valores para  $Pc3$  que chegaram a 50% para  $Y_{9:7}$  e 80%  $Y_{27:37}$ . Para o tamanho populacional de 200 indivíduos observa-se uma redução dos valores de  $Pc1$  para todas as características, valores elevados de  $Pc3$  foram encontrados

para  $Y_{15:1}$  e  $Y_{63:1}$  em que 13:3 foi aceito em 100% dos casos em detrimento da hipótese verdadeira.

Define-se hipótese genética como o mecanismo de herança mais provável de uma característica com base na avaliação de dados disponíveis. Assim, como ressaltado anteriormente, quando se considera genes ligados tem-se na realidade “falsas hipóteses de segregação”, ou seja, as características  $Y_{9:7}$  seqüencialmente até  $Y_{63:1}$  não tem proporções de segregação esperada previamente pelo fato dos genes não estarem segregando independente e o valor da porcentagem de recombinação ser desconhecido. Assim, este fato justifica o resultado encontrado para a grande maioria dos testes de segregação com valores de  $Pc1$  abaixo de 15% e grande quantidade de ambigüidade de segregação.

Em todos os tamanhos populacionais houve ambigüidade de segregação, para o conjunto de características governadas por genes ligados. Observa-se que houveram casos positivos de ambigüidade de segregação para todas as hipóteses testadas com exceção da característica  $Y_{9:3:4}$  a partir do tamanho populacional de 200 indivíduos, e que, em relação às hipóteses alternativas, foram encontrados 12 padrões concorrentes com valores maiores que 70% de número de testes de segregação falso positivo (erro tipo I). Para os tamanhos populacionais de 400 e 600 indivíduos apenas  $Y_{15:1}$  apresentou duas hipóteses concorrentes as demais apresentaram apenas uma. Como ocorreram elevados índices de erro tipo I para as hipóteses testadas (conhecidas) e também elevados índices de erro tipo II para as hipóteses concorrentes, conseqüentemente a porcentagem do número de testes concorrentes positivos é elevada e quando comparados a hipótese supostamente conhecida com a concorrente em termos de maior probabilidade.

Na Tabela 2 pode ser comprovada a eficácia da análise genômica nos testes de identificação do número de genes controladores das características. Estas análises foram eficientes na quantificação e na identificação dos genes responsáveis pela característica. Pode-se observar que, de maneira geral, com o aumento do tamanho populacional e a utilização das informações de classes multicategóricas houve melhor eficiência da metodologia de identificação de genes, chegando a 100% na identificação dos mesmos para o tamanho populacional de 600 indivíduos para quase a totalidade dos casos de segregação estudados. Para o tamanho populacional de 100 indivíduos foi verificado grande quantidade de testes de identificação de genes foram negativos, com destaque para as características  $Y_{15:1}$ ;  $Y_{27:37}$ ;  $Y_{12:3:1}$  e  $Y_{63:1}$ , porém este problema foi resolvido com o aumento do tamanho populacional, com 200 indivíduos foi constatado apenas 25% (Pg2) de testes negativos para  $Y_{13:3}$ , 37,5 para  $Y_{12:3:1}$  e 85% para  $Y_{63:1}$ . Tais resultados são coerentes com alguns obtidos em análise genômica de detecção de QTLs que ressaltam que o tamanho populacional de 200 indivíduos trazem boas informações na identificação de QTL. São encontrados pesquisas visando à detecção de QTLs utilizando populações de até 200 indivíduos (Couto et al., 2010; Borovsky e Paran, 2011; Lu et al., 2011) e acima de 200 indivíduos (Ding et al., 2011; Sousa et al., 2011).

Nos grupos de características e população de tamanho 400 indivíduos foram encontrados resultados com genes extras (ou genes fantasmas), chegando a 37,5% dos testes para a característica  $Y_{15:1}$  com presença de um gene extra e 30% para a  $Y_{63:1}$ , podendo inferir que este fato deva-se ao espaçamento entre as marcas ou a metodologia aplicada. Problemas similares são encontrados em estudos de QTL, cuja elucidação da ocorrência de QTL fantasma tem sido feito por meio do uso do método do intervalo composto.

Os valores médios de **m**, **a** e **d** para os conjuntos de caracteres controlados por genes ligados estão apresentados na Tabela 4 que, quando comparados aos valores referenciais da Tabela 3, permite constatar grande coincidência ou valores próximos com pouca variação em relação a classe estudada e com forte indicativo de que os valores destes sejam indicativo da proporção de segregação a única exceção foi para  $Y_{13:3}$  em que somente a partir do tamanho de 600 indivíduos que se observou a identificação da marca/gene situada na posição M16 . Assim, com base no valor médio de **m**, **a** e **d** pode-se inferir sobre o tipo de controle genético da característica. Conclusões desta natureza não seriam possíveis de serem obtidas por meio da análise genética clássica pela impossibilidade do estabelecimento de uma hipótese de segregação que contemplasse informações sobre o número de genes ligados, distância entre estes e o padrão epistático. Observa-se que, com o aumento do tamanho populacional, os valores médios de **m**, **a** e **d** para cada classe de característica se aproximaram do valor paramétrico (Tabela 4).

### **Conclusões**

- Os testes de segregação empregados em genética clássica foram eficientes na identificação do padrão de herança de característica quando se tratava de genes independentes, porém problemas de ambigüidade de segregação foram ampliados em razão do efeito ambiental sobre a característica e a subjetividade no agrupamento de classes necessário para testar as hipóteses genéticas formuladas.

- Testes clássicos de segregação quando aplicados em características controladas por genes ligados são inconclusivos e proporcionam grandes equívocos em análise e interpretação.

- Os testes de identificação de genes em análises genômica, a partir de informações fenotípicas e moleculares da F<sub>2</sub>, se mostraram eficientes tanto para estudo de herança de características controlada por genes independentes quanto ligados.

- O uso de informações fenotípicas multicategóricas e influenciadas pelo ambiente, reduziu a eficiência dos estudos, pela Genética Clássica, em razão da necessidade de se agrupar, subjetivamente, classes para teste de hipótese genética.

- O uso de informações fenotípicas multicategóricas aumentou a eficiência dos estudos do controle genético, pela Genômica, em razão da maior variabilidade dos valores fenotípicos empregados nos modelos de detecção de fatores genéticos.

- Delineamento genético para fins de estudo de herança envolvendo informações fenotípicas e moleculares das gerações segregantes devem ser recomendados por permitir o emprego de análise de genética clássica e genômica gerando informações complementares e removendo equívocos nos casos de ocorrência de segregações genéticas ambíguas.

## Referências

ALZATE-MARIN A.L. et al. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v.30, n.4, p.333-342, 2005.

BARBOSA-PRESTES M.M. et al. Controle genético da resistência parcial à ferrugem da folha em aveia (*Avena sativa* L.). **Ciência Rural**, v.38, n.2, p.308-314, 2008.

BARELLI,M.A.A.; SCHAWINSKI, E.C.; NEVES, L.G. Avaliação de acessos de *Manihot esculenta*, Crantz. Com uso de variáveis multicategóricas no município de Cáceres-MT. **Raízes e Amidos Tropicais**. Botucatu: CERAT/UNESP. V.3,2007.

BERTAN,I.;CARVALHO,F.I.F.C.;OLIVEIRA,A.C.;VIEIRA,E.A. Comparação de métodos de agrupamento na representação da distância morfológica entre genótipos de trigo. **R. Bras. Agrocência**, Pelotas, v. 12, n. 3, p. 279-286, julset, 2006.

BHERING, L. L.; CRUZ, C. D. Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. **Pesquisa agropecuária brasileira**, v.43, p.379-385, 2008.

BOROVSKY, Y.; PARAN, I. Characterization of *fs* 10.1, a major QTL controlling fruit in elongation in *Capsicum*. **Theor Appl Genet**, 2011. doi: 10.1007/s00122-011-1615-7

COIMBRA, R.R.; MIRANDA, G.V.; MOREIRA, G.R.; SILVA, D.J.H.; CRUZ, C.D.; CARNEIRO, P.C.S.; SOUZA, L.V.; GUIMARÃES, L.J.M.; MARCASSO, R.C.; CANIATO, F.F. Divergência genética de cultivares de milho baseada em descritores qualitativos. In: **Simpósio de recursos genéticos para a América Latina e Caribe, III SIRGEALC.**,2001,Londrina. *Anais...* Londrina, PR, 2001. p. 266-268.

COUTO, K. R.; SANTOS, J. B.; RAMALHO, M. A. P.; SILVA, G. S. Identificação de marcadores microssatélites relacionados ao escurecimento do grão de feijão. **Pesquisa Agropecuária Brasileira**, v.45, n.11, p.1268-1274, 2010.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético**, vol. 2. Viçosa: UFV, 2006. 585 p.

DING, J. Q.; MA, J. L.; ZHANG, C. R.; DONG, H. F.; XI, Z. Y.; XIA, Z. L.; WU, J. Y. QTL mapping for test weight by using  $F_{2:3}$  population in maize. **Journal of Genetics**, v.90, n.1, p.75-80, 2011

GAGLIARDI P.R.; CAMARGO L.E.A. Resistência de variedades comerciais de cana-de-açúcar ao agente causal do raquitismoda-soqueira. **Ciência Rural**, v.39, n.4, p.1211-1214, 2009.

HARUSHIMA Y., YANO M., SHOMURA A., SATO M., SHIMANO T., KUBOKI, Y., YAMAMOTO T., LIN S.Y., BALTAZAR A.A., PARCO A., KAJIYA H., HUANG N., YAMAMOTO K., NAGAMURA Y., KURATA N., KHUSH G.S., SASAKI T.A. high-density rice genetic linkage map with 2275 markers using a single  $F_2$  population. **Genetics**, v.148, p.479-494, 1998.

Kamei, A.; Tsuru, M.; Kubo, N.; Hayashi, T.; Wang, N.; Fujimura, T.; Hirai, M.; QTL mapping of clubroot resistance in radish (*Raphanus sativus* L.). **Theoretical and Applied Genetics**, v.120, p.1021–1027, 2010.

LU, B.; XIE, K.; YANG, C.; ZHANG, L.; WU, T.; LI, L.; LIU, X.; JIANG, L.; WAN, J. Efficient QTL detection for heading date in bechcross inbred line and  $F_2$  population derived from the same rice cross. **African Journal of Agriculture Research**, v.6, n.10, p.2372-2378, 2011.

MCCLOSKEY, B.; MA, X.; TANKSLEY, S.D. Quantifying the Relative Contribution of the Heterozygous Class to QTL Detect Power. **Statistical Application in Genetics and Molecular Biology**, v.10, n.1, 2011. doi: 10.2202/1544-6116.1622.

MIGNON, G.L.; PITEL, F.; GILBERT, H.; BIHAN-DURVAL, E.L.; VIGNOLES, F.; DEMEURE, O.; LAGARRIGUE, S.; SIMON, J.; COGBURN, L.A.; AGGREY, S.E.; DOUAIRE, M.; ROY, P.L. A comprehensive analysis of QTL for abdominal fat and breast muscle weights on chicken chromosome 5 using a multivariate approach. **Animal Genetics**, v.40, p.157-164, 2009.

PALOMARES-RIUS, F. J.; VIRUEL, M. A.; YUSTE-LISBONA, F. J.; LÓPEZ-SESÉ, A. I.; GÓMEZ-GUILLAMÓN, M. L.; Simple sequence repeat markers linked to QTL for resistance to Watermelon mosaic virus in melon. **Theoretical and Applied Genetics**, v.123, p.1207–1214, 2011.

PEREIRA, F.H.F.; PUIATTI, M.; MIRANDA, G.V.; SILVA, D.J.H.; FINGER, F.L. Divergência genética entre acessos de taro utilizando caracteres morfoqualitativos de inflorescência. **Horticultura Brasileira**, Brasília, v. 21, n. 3, p. 520-524, julho-setembro 2003.

SANTOS, R. M. F.; LOPES, U. V.; BAHIA, R. C.; MACHADO, R. C. R.; AHNERT, D.; CORREA, R. X. Marcadores microssatélites relacionados com a resistência à vassoura-de-bruxa do cacauzeiro. **Pesquisa agropecuária brasileiro**, v.42, n.8, p.1137-1142, 2007.

SCHUSTER I., CRUZ C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos**. Viçosa , MG. Editora UFV, 2004. 585p.

SHARMA, S.; SHARMA, S.; KOPISCH-OBUCH, F. J.; KEIL, T.; LAUBACH, E.; STEIN, N.; GRANER, ANDREAS.; JUNG, C.; QTL analysis of root-lesion nematode resistance in barley: 1. *Pratylenchus neglectus*. **Theoretical and Applied Genetics**, v.122, p.1321–1330, 2011.

SOUSA, K. R. S.; GUIMARÃES, S. E. F.; SILVA FILHO, M. I.; LOPES, M. S.; PINTO, A. P. G.; VERANDO, L. L.; BRACCINI NETO, J.; LOPES, P. S. Mapeamento de locos de características quantitativas nos cromossomos 5, 7 e 8 de suínos. **Revista Brasileira de Zootecnia**, v.40, n.1, p.115-123, 2011.

SUDRÉ C.P.; CRUZ C.D.; RODRIGUES R.; RIVA E.M.; AMARAL JÚNIOR A.T.; SILVA D.J.H.; PEREIRA T.N.S. Variáveis multicatóricas na determinação da divergência genética entre acessos de pimenta e pimentão. **Horticultura Brasileira**, 24: 88-93, 2006.

WEIGUO L., MING Y.Z., ENRIQUE A.P., CALVIN F.K. Highly efficient doubledhaploid production in wheat (*Triticum aestivum* L.) via induced microspore embryogenesis. **Crop Science**, v.42, p.682-692, 2002.

YOUNG, N. D. **Constructing a plant genetic linkage map with DNA markers**. In: PHILIPS, P. L.; VASIL, I. K. DNA-Based Markers in plants. Dordrecht, The Netherlands: Kluwer Academic Publisher, p.39-57, 1994.

Tabela 1- Compilação do padrão de herança por métodos de genética clássica e por análise genômica para caracteres controlados por um, dois ou três genes independentes.

Caráter	Genética Clássica				Genômica			
	Pc1	Ha	Pc2	Pc3	Pg1.	Pg2	Pg3	Pg4
N=100								
Y <sub>(9:7)</sub>	60	3:1 - 13:3 - 27:37	35 - 5 - 75	35 - 5 - 55	32,5	5	-	62,5
Y <sub>(13:3)</sub>	5	3:1 - 13:3 - 27:37	30 - 70 - 70	25 - 70 - 70	15	75	-	10
Y <sub>(15:1)</sub>	0	3:1 - 9:7 - 13:3 - 27:37	35 - 55 - 10 - 70	35 - 55 - 10 - 70	45	17,5	-	37,5
Y <sub>(9:6:1)</sub>	35	12:3:1 - 9:3:4	10 - 55	10 - 55	50	15	-	35
Y <sub>(12:3:1)</sub>	5	9:6:1 - 9:3:4	15 - 70	15 - 70	35	20	-	45
Y <sub>(9:3:4)</sub>	65	9:6:1 - 12:3:1	40 - 5	40 - 5	35	15	-	50
Y <sub>(27:37)</sub>	75	3:1 - 9:7 - 13:3	30 - 65 - 5	30 - 40 - 5	18,33	21,67	-	60
Y <sub>(63:1)</sub>	0	3:1 - 9:7 - 13:3 - 27:37	40 - 50 - 10 - 80	40 - 50 - 10 - 80	45	13,33	-	41,67
N=200								
Y <sub>(9:7)</sub>	15	3:1 - 27:37	30 - 25	30 - 20	100	-	-	-
Y <sub>(13:3)</sub>	10	15:1	80	80	35	62,5	-	2,5
Y <sub>(15:1)</sub>	-	3:1 - 13:3	85 - 100	85 - 100	100	-	-	-
Y <sub>(9:6:1)</sub>	-	9:3:4	55	55	100	-	-	-
Y <sub>(12:3:1)</sub>	15	9:3:4	65	65	100	-	-	-
Y <sub>(9:3:4)</sub>	50	-	-	-	100	-	-	-
Y <sub>(27:37)</sub>	50	3:1 - 9:7	25 - 35	25 - 0	100	-	-	-
Y <sub>(63:1)</sub>	-	3:1 - 13:3	20 - 100	20 - 100	100	-	-	-
N=400								
Y <sub>(9:7)</sub>	-	3:1	40	40	100	-	-	-
Y <sub>(13:3)</sub>	-	15:1	90	90	37,5	42,5	-	20
Y <sub>(15:1)</sub>	-	3:1 - 13:3	20-95	20-95	100	-	-	-
Y <sub>(9:6:1)</sub>	-	9:3:4	30	30	100	-	-	-
Y <sub>(12:3:1)</sub>	-	9:3:4	25	25	100	-	-	-
Y <sub>(9:3:4)</sub>	50	-	-	-	97,5	-	-	2,5
Y <sub>(27:37)</sub>	15	9:7	5	-	98,33	-	-	1,67
Y <sub>(63:1)</sub>	-	13:3	70	70	100	-	-	-
N=600								
Y <sub>(9:7)</sub>	-	3:1	10	10	100	-	-	-
Y <sub>(13:3)</sub>	-	15:1	80	80	87,5	12,5	-	-
Y <sub>(15:1)</sub>	-	3:1 - 13:3	20 - 85	20 - 85	100	-	-	-
Y <sub>(9:6:1)</sub>	-	9:3:4	50	50	100	-	-	-
Y <sub>(12:3:1)</sub>	10	9:3:4	5	5	100	-	-	-
Y <sub>(9:3:4)</sub>	30	-	-	-	100	-	-	-
Y <sub>(27:37)</sub>	15	3:1	5	5	100	-	-	-
Y <sub>(63:1)</sub>	-	13:3	65	65	100	-	-	-

Pc1: porcentagem de teste qui-quadrado que não rejeitou a hipótese verdadeira a 5% de probabilidade; Ha: hipótese alternativa ou concorrente; Pc2: porcentagem de teste qui-quadrado que não rejeitou a hipótese concorrente a 5% de probabilidade; Pc3: porcentagem de teste qui-quadrado que não rejeitou ambas hipóteses, mas com a concorrente associada a maior nível de significância; Pg1: porcentagem de análise genômica em que os genes foram detectados na quantidade e na posição exata; Pg2: porcentagem de análise em que os genes não foram detectados; Pg3: porcentagem de análise que detectou genes extras (fantasma) além do esperado; Pg4: porcentagem de análise em que os genes foram detectados em quantidade correta mas em posição aproximada. Pg1+Pg2+Pg4=100.

Tabela 2- Compilação do padrão de herança por métodos de genética clássica e por análise genômica para caracteres controlados por um, dois ou três genes ligados

Caráter	Genética Clássica				Genômica			
	Pc1	Ha	Pc2	Pc3	Pg1.	Pg2	Pg3	Pg4
N=100								
Y <sub>(3:1)</sub>	100	13:3 - 27:37	45 - 5	20 - 0	75	-	-	25
Y <sub>(1:2:1)</sub>	90	-	-	-	90	-	-	10
Y <sub>(9:7)</sub>	95	27:37	100	35	72,5	-	2,5	25
Y <sub>(13:3)</sub>	90	3:1	100	45	45	32,5	-	22,5
Y <sub>(15:1)</sub>	100	63:1	10	10	47,5	22,5	7,5	22,5
Y <sub>(9:6:1)</sub>	40	9:3:4	10	10	80	-	-	20
Y <sub>(12:3:1)</sub>	30	9:6:1 - 9:3:4	20 - 35	20 - 30	45	35	-	20
Y <sub>(9:3:4)</sub>	50	-	-	-	52,5	15	-	32,5
Y <sub>(27:37)</sub>	100	9:7	100	90	45	25	-	30
Y <sub>(63:1)</sub>	65	3:1-9:7-13:3-15:1-27:37	25-5-15-35-5	25-5-15-5-5	1,67	90	-	8,33
N=200								
Y <sub>(3:1)</sub>	100	13:3	45	15	100	-	-	-
Y <sub>(1:2:1)</sub>	75	-	-	-	100	-	-	-
Y <sub>(9:7)</sub>	95	27:37	90	35	97,5	-	-	2,5
Y <sub>(13:3)</sub>	100	3:1	45	10	72,5	25	2,5	-
Y <sub>(15:1)</sub>	95	-	-	-	75	5	5	15
Y <sub>(9:6:1)</sub>	40	-	-	-	97,5	-	-	2,5
Y <sub>(12:3:1)</sub>	30	9:6:1	25	25	52,5	37,5	-	10
Y <sub>(9:3:4)</sub>	5	-	-	-	90	-	7,5	2,5
Y <sub>(27:37)</sub>	75	9:7	100	95	63,33	35	-	1,67
Y <sub>(63:1)</sub>	95	15:1	5	-	6,67	85	8,33	-
N=400								
Y <sub>(3:1)</sub>	85	-	-	-	95	-	-	5
Y <sub>(1:2:1)</sub>	55	-	-	-	100	-	-	-
Y <sub>(9:7)</sub>	90	27:37	85	35	67,5	-	2,5	30
Y <sub>(13:3)</sub>	90	3:1	30	15	75	-	10	15
Y <sub>(15:1)</sub>	90	-	-	-	55	-	7,5	37,5
Y <sub>(9:6:1)</sub>	20	-	-	-	87,5	-	-	12,5
Y <sub>(12:3:1)</sub>	10	-	-	-	50	30	2,5	17,5
Y <sub>(9:3:4)</sub>	-	-	-	-	75	-	7,5	17,5
Y <sub>(27:37)</sub>	25	9:7	80	75	55	25	1,67	18,33
Y <sub>(63:1)</sub>	100	-	-	-	26,67	40	3,33	30
N=600								
Y <sub>(3:1)</sub>	100	-	-	-	100	-	-	-
Y <sub>(1:2:1)</sub>	45	-	-	-	100	-	-	-
Y <sub>(9:7)</sub>	80	27:37	75	35	100	-	-	-
Y <sub>(13:3)</sub>	100	-	-	-	100	-	-	-
Y <sub>(15:1)</sub>	95	-	-	-	100	-	-	-
Y <sub>(9:6:1)</sub>	25	-	-	-	100	-	-	-
Y <sub>(12:3:1)</sub>	25	9:6:1	10	10	92,5	7,5	-	-
Y <sub>(9:3:4)</sub>	-	-	-	-	100	-	-	-
Y <sub>(27:37)</sub>	15	9:7	45	45	95	5	-	-
Y <sub>(63:1)</sub>	90	-	-	-	93,33	6,67	-	-

Pc1: porcentagem de teste qui-quadrado que não rejeitou a hipótese verdadeira a 5% de probabilidade; Ha: hipótese alternativa ou concorrente; Pc2: porcentagem de teste qui-quadrado que não rejeitou a hipótese concorrente a 5% de probabilidade; Pc3: porcentagem de teste qui-quadrado que não rejeitou ambas hipóteses, mas com a concorrente associada a maior nível de significância; Pg1: porcentagem de análise genômica em que os genes foram detectados na quantidade e na posição exata; Pg2: porcentagem de análise em que os genes não foram detectado; Pg3: porcentagem de análise que detectou genes extras (fantasma) além do esperado; Pg4: porcentagem de análise em que os genes foram detectados em quantidade correta mas em posição aproximada. Pg1+Pg2+Pg4=100.

Tabela 3: Valores médios esperados de m, a e d determinantes da expressão fenotípica de características com diferentes padrões de segregação

Caráter	Genes independentes			Genes ligados ( $r = 0,10cM$ )			
	m	a	d	Caráter	m	a	d
Y <sub>(3:1)(5)</sub>	2,5434	1,4386	1,4224	Y <sub>(3:1)(5)</sub>	-	-	-
Y <sub>(1:2:1)(5)</sub>	2,3128	1,5894	-0,0242	Y <sub>(1:2:1)(5)</sub>	-	-	-
Y <sub>(9:7)(5)</sub>	1,9115	1,0025	1,0375	Y <sub>(9:7)(15)</sub>	2,5194	1,3343	1,0907
Y <sub>(9:7)(20)</sub>	1,9154	1,0656	1,047	Y <sub>(9:7)(16)</sub>	2,5117	1,3506	1,1269
Y <sub>(13:3)(5)</sub>	2,8374	1,0106	1	Y <sub>(13:3)(15)</sub>	3,5226	0,2892	0,2952
Y <sub>(13:3)(20)</sub>	3,4814	-0,3122	-0,2529	Y <sub>(13:3)(16)</sub>	3,7839	-0,0054	-0,2267
Y <sub>(15:1)(5)</sub>	4,1157	0,4623	0,4114	Y <sub>(15:1)(15)</sub>	2,9648	1,5181	1,4856
Y <sub>(15:1)(20)</sub>	4,1094	0,4191	0,4324	Y <sub>(15:1)(16)</sub>	2,9788	1,4962	1,4762
Y <sub>(9:6:1)(5)</sub>	3,0862	0,8718	0,8665	Y <sub>(9:6:1)(15)</sub>	2,7387	1,5959	1,4363
Y <sub>(9:6:1)(20)</sub>	3,1083	0,829	0,8386	Y <sub>(9:6:1)(16)</sub>	2,7509	1,5904	1,434
Y <sub>(12:3:1)(5)</sub>	2,8704	1,0316	1,0223	Y <sub>(12:3:1)(15)</sub>	2,5361	1,3993	1,3544
Y <sub>(12:3:1)(20)</sub>	3,3027	0,165	0,1871	Y <sub>(12:3:1)(16)</sub>	2,6797	1,2304	1,0833
Y <sub>(9:3:4)(5)</sub>	2,4212	1,3247	1,3366	Y <sub>(9:3:4)(15)</sub>	2,615	1,5808	1,385
Y <sub>(9:3:4)(20)</sub>	2,8033	0,5788	0,606	Y <sub>(9:3:4)(16)</sub>	2,7498	1,4039	1,1367
Y <sub>(27:37)(5)</sub>	1,7119	0,7861	1,2546	Y <sub>(27:37)(13)</sub>	2,6687	1,2778	0,9382
Y <sub>(27:37)(20)</sub>	2,1898	0,2607	0,3276	Y <sub>(27:37)(14)</sub>	2,7087	1,2105	0,8435
Y <sub>(27:37)(50)</sub>	1,9656	0,9948	0,8201	Y <sub>(27:37)(15)</sub>	2,6394	1,2675	1,0486
Y <sub>(63:1)(5)</sub>	4,6296	0,1224	0,1162	Y <sub>(63:1)(13)</sub>	3,2928	1,5317	1,5258
Y <sub>(63:1)(20)</sub>	4,627	0,1088	0,124	Y <sub>(63:1)(14)</sub>	3,2624	1,5587	1,5605
Y <sub>(63:1)(50)</sub>	4,6307	0,1143	0,1217	Y <sub>(63:1)(15)</sub>	3,3531	1,4662	1,4731

Tabela 4- Valores dos efeitos **m**, **a** e **d** relativos a 20 réplicas de cada conjunto populacional simulado

Genes independentes				Genes ligados			
Caráter	m	a	d	Caráter	m	a	d
N=100							
Y <sub>(3:1)</sub>	2,35	1,53	1,50	-	-	-	-
Y <sub>(1:2:1)</sub>	2,49	1,71	0,07	-	-	-	-
Y <sub>(9:7) (M5)</sub>	2,07	1,06	1,09	Y <sub>(9:7) (M15)</sub>	2,27	1,48	1,30
Y <sub>(9:7) (M20)</sub>	2,06	1,08	1,07	Y <sub>(9:7) (M16)</sub>	2,25	1,50	1,33
Y <sub>(13:3) (M5)</sub>	2,68	1,17	1,15	Y <sub>(13:3) (M15)</sub>	3,40	0,43	0,48
Y <sub>(13:3) (M20)</sub>	3,36	-0,31	-0,18	Y <sub>(13:3) (M16)</sub>	3,29	0,23	-0,13
Y <sub>(15:1) (M5)</sub>	3,61	0,57	0,58	Y <sub>(15:1) (M15)</sub>	2,63	1,70	1,71
Y <sub>(15:1) (M20)</sub>	3,58	0,60	0,60	Y <sub>(15:1) (M16)</sub>	2,51	1,65	1,64
Y <sub>(9:6:1) (M5)</sub>	2,63	0,98	0,97	Y <sub>(9:6:1) (M15)</sub>	2,54	1,64	1,56
Y <sub>(9:6:1) (M20)</sub>	2,59	0,95	1,09	Y <sub>(9:6:1) (M16)</sub>	2,56	1,61	1,59
Y <sub>(12:3:1) (M5)</sub>	2,88	1,21	1,25	Y <sub>(12:3:1) (M15)</sub>	2,47	1,69	1,74
Y <sub>(12:3:1) (M20)</sub>	3,25	0,51	0,66	Y <sub>(12:3:1) (M16)</sub>	2,47	1,63	1,57
Y <sub>(9:3:4) (M5)</sub>	2,24	1,37	1,38	Y <sub>(9:3:4) (M15)</sub>	2,29	1,53	1,40
Y <sub>(9:3:4) (M20)</sub>	2,55	0,63	0,73	Y <sub>(9:3:4) (M16)</sub>	2,41	1,55	1,31
Y <sub>(27:37) (M5)</sub>	1,93	0,82	1,33	Y <sub>(27:37) (M13)</sub>	2,24	1,45	1,21
Y <sub>(27:37) (M20)</sub>	2,44	0,66	0,34	Y <sub>(27:37) (M14)</sub>	2,23	1,44	1,24
Y <sub>(27:37) (M50)</sub>	2,10	1,00	1,01	Y <sub>(27:37) (M15)</sub>	2,21	1,39	1,20
Y <sub>(63:1) (M5)</sub>	0	0	0	Y <sub>(63:1) (M13)</sub>	3,08	1,70	1,78
Y <sub>(63:1) (M20)</sub>	4,36	0,41	0,26	Y <sub>(63:1) (M14)</sub>	3,07	1,69	1,79
Y <sub>(63:1) (M50)</sub>	4,40	0,33	0,44	Y <sub>(63:1) (M15)</sub>	3,10	1,68	1,76
N=200							
Y <sub>(3:1)</sub>	2,43	1,48	1,45	-	-	-	-
Y <sub>(1:2:1)</sub>	2,50	1,68	-0,01	-	-	-	-
Y <sub>(9:7) (M5)</sub>	2,04	0,99	1,02	Y <sub>(9:7) (M15)</sub>	2,31	1,34	1,14
Y <sub>(9:7) (M20)</sub>	2,07	1,01	0,98	Y <sub>(9:7) (M16)</sub>	2,31	1,33	1,11
Y <sub>(13:3) (M5)</sub>	2,71	1,13	1,20	Y <sub>(13:3) (M15)</sub>	3,40	0,37	0,37
Y <sub>(13:3) (M20)</sub>	3,52	-0,37	-0,47	Y <sub>(13:3) (M16)</sub>	-	-	-
Y <sub>(15:1) (M5)</sub>	3,77	0,34	0,42	Y <sub>(15:1) (M15)</sub>	2,80	1,52	1,54
Y <sub>(15:1) (M20)</sub>	3,78	0,45	0,46	Y <sub>(15:1) (M16)</sub>	2,76	1,56	1,57
Y <sub>(9:6:1) (M5)</sub>	2,65	0,87	0,94	Y <sub>(9:6:1) (M15)</sub>	2,57	1,55	1,40
Y <sub>(9:6:1) (M20)</sub>	2,69	0,89	0,90	Y <sub>(9:6:1) (M16)</sub>	2,56	1,55	1,39
Y <sub>(12:3:1) (M5)</sub>	2,94	1,11	1,08	Y <sub>(12:3:1) (M15)</sub>	2,55	1,58	1,60
Y <sub>(12:3:1) (M20)</sub>	3,29	0,35	0,19	Y <sub>(12:3:1) (M16)</sub>	2,68	1,41	1,29
Y <sub>(9:3:4) (M5)</sub>	2,25	1,22	1,31	Y <sub>(9:3:4) (M15)</sub>	2,33	1,46	1,36
Y <sub>(9:3:4) (M20)</sub>	2,64	0,58	0,57	Y <sub>(9:3:4) (M16)</sub>	2,46	1,32	1,08
Y <sub>(27:37) (M5)</sub>	1,94	0,76	1,32	Y <sub>(27:37) (M13)</sub>	2,36	1,30	0,99
Y <sub>(27:37) (M20)</sub>	0	0	0	Y <sub>(27:37) (M14)</sub>	2,42	1,26	0,89
Y <sub>(27:37) (M50)</sub>	2,19	0,99	0,79	Y <sub>(27:37) (M15)</sub>	2,37	1,30	0,99
Y <sub>(63:1) (M5)</sub>	4,40	0,22	0,29	Y <sub>(63:1) (M13)</sub>	3,14	1,63	1,60
Y <sub>(63:1) (M20)</sub>	4,51	0,24	0,26	Y <sub>(63:1) (M14)</sub>	3,14	1,63	1,61
Y <sub>(63:1) (M50)</sub>	4,27	0,207	0,31	Y <sub>(63:1) (M15)</sub>	3,17	1,60	1,58

...Continua

....Continuação

Genes independentes				Genes ligados			
Caráter	m	a	d	Caráter	m	a	d
N=400							
Y <sub>(3:1)</sub>	2,34	1,45	1,43	-	-	-	-
Y <sub>(1:2:1)</sub>	2,44	1,60	-0,04	-	-	-	-
Y <sub>(9:7) (M5)</sub>	2,12	0,97	0,99	Y <sub>(9:7) (M15)</sub>	2,38	1,32	1,08
Y <sub>(9:7) (M20)</sub>	2,13	0,99	0,98	Y <sub>(9:7) (M16)</sub>	2,37	1,32	1,11
Y <sub>(13:3) (M5)</sub>	2,79	1,19	1,15	Y <sub>(13:3) (M15)</sub>	3,44	0,31	0,32
Y <sub>(13:3) (M20)</sub>	3,54	-0,35	-0,42	Y <sub>(13:3) (M16)</sub>	3,87	-0,01	-0,39
Y <sub>(15:1) (M5)</sub>	3,75	0,50	0,49	Y <sub>(15:1) (M15)</sub>	2,77	1,52	1,52
Y <sub>(15:1) (M20)</sub>	3,77	0,47	0,47	Y <sub>(15:1) (M16)</sub>	2,79	1,51	1,48
Y <sub>(9:6:1) (M5)</sub>	2,70	0,87	0,90	Y <sub>(9:6:1) (M15)</sub>	2,52	1,55	1,38
Y <sub>(9:6:1) (M20)</sub>	2,72	0,89	0,87	Y <sub>(9:6:1) (M16)</sub>	2,52	1,54	1,40
Y <sub>(12:3:1) (M5)</sub>	3,00	1,12	1,12	Y <sub>(12:3:1) (M15)</sub>	2,52	1,57	1,58
Y <sub>(12:3:1) (M20)</sub>	3,34	0,29	0,38	Y <sub>(12:3:1) (M16)</sub>	2,70	1,37	1,25
Y <sub>(9:3:4) (M5)</sub>	2,20	1,23	1,22	Y <sub>(9:3:4) (M15)</sub>	2,37	1,47	1,34
Y <sub>(9:3:4) (M20)</sub>	2,56	0,54	0,53	Y <sub>(9:3:4) (M16)</sub>	2,50	1,30	1,10
Y <sub>(27:37) (M5)</sub>	1,84	0,75	1,28	Y <sub>(27:37) (M13)</sub>	2,28	1,31	1,05
Y <sub>(27:37) (M20)</sub>	2,33	0,38	0,30	Y <sub>(27:37) (M14)</sub>	2,36	1,29	0,89
Y <sub>(27:37) (M50)</sub>	2,02	0,90	0,89	Y <sub>(27:37) (M15)</sub>	2,29	1,30	1,06
Y <sub>(63:1) (M5)</sub>	4,60	0,21	0,19	Y <sub>(63:1) (M13)</sub>	3,28	1,51	1,50
Y <sub>(63:1) (M20)</sub>	4,57	0,18	0,20	Y <sub>(63:1) (M14)</sub>	3,26	1,52	1,52
Y <sub>(63:1) (M50)</sub>	4,57	0,20	0,21	Y <sub>(63:1) (M15)</sub>	3,30	1,49	1,48
N=600							
Y <sub>(3:1)</sub>	2,55	1,39	1,41	-	-	-	-
Y <sub>(1:2:1)</sub>	2,56	1,55	-0,01	-	-	-	-
Y <sub>(9:7) (M5)</sub>	2,06	0,98	0,99	Y <sub>(9:7) (M15)</sub>	2,43	1,30	1,07
Y <sub>(9:7) (M20)</sub>	2,07	0,98	1,006	Y <sub>(9:7) (M16)</sub>	2,44	1,30	1,07
Y <sub>(13:3) (M5)</sub>	2,82	1,07	1,07	Y <sub>(13:3) (M15)</sub>	3,49	0,30	0,29
Y <sub>(13:3) (M20)</sub>	3,51	-0,29	-0,29	Y <sub>(13:3) (M16)</sub>	3,87	-0,01	-0,29
Y <sub>(15:1) (M5)</sub>	3,83	0,52	0,51	Y <sub>(15:1) (M15)</sub>	2,74	1,48	1,48
Y <sub>(15:1) (M20)</sub>	3,84	0,51	0,51	Y <sub>(15:1) (M16)</sub>	2,76	1,45	1,46
Y <sub>(9:6:1) (M5)</sub>	2,78	0,86	0,90	Y <sub>(9:6:1) (M15)</sub>	2,55	1,47	1,35
Y <sub>(9:6:1) (M20)</sub>	2,79	0,86	0,88	Y <sub>(9:6:1) (M16)</sub>	2,57	1,47	1,34
Y <sub>(12:3:1) (M5)</sub>	3,03	1,10	1,07	Y <sub>(12:3:1) (M15)</sub>	2,61	1,58	1,56
Y <sub>(12:3:1) (M20)</sub>	3,49	0,24	0,24	Y <sub>(12:3:1) (M16)</sub>	2,78	1,39	1,23
Y <sub>(9:3:4) (M5)</sub>	2,26	1,27	1,26	Y <sub>(9:3:4) (M15)</sub>	2,46	1,44	1,31
Y <sub>(9:3:4) (M20)</sub>	2,63	0,54	0,56	Y <sub>(9:3:4) (M16)</sub>	2,60	1,30	1,05
Y <sub>(27:37) (M5)</sub>	1,91	0,76	1,19	Y <sub>(27:37) (M13)</sub>	2,47	1,23	0,95
Y <sub>(27:37) (M20)</sub>	2,44	0,26	0,30	Y <sub>(27:37) (M14)</sub>	2,52	1,20	0,85
Y <sub>(27:37) (M50)</sub>	2,12	0,88	0,86	Y <sub>(27:37) (M15)</sub>	2,46	1,23	0,98
Y <sub>(63:1) (M5)</sub>	4,57	0,15	0,17	Y <sub>(63:1) (M13)</sub>	3,28	1,54	1,55
Y <sub>(63:1) (M20)</sub>	4,58	0,17	0,16	Y <sub>(63:1) (M14)</sub>	3,30	1,52	1,53
Y <sub>(63:1) (M50)</sub>	4,61	0,15	0,13	Y <sub>(63:1) (M15)</sub>	3,29	1,52	1,53