

**ADÉSIO FERREIRA**

**MAPEAMENTO GENÉTICO UTILIZANDO MARCADORES  
MOLECULARES COM DISTORÇÃO DE SEGREGAÇÃO GAMÉTICA E  
GENOTÍPICA**

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  
2006**

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

T

F383m  
2006

Ferreira, Adésio, 1971-  
Mapeamento genético utilizando marcadores moleculares  
com distorção de segregação gamética e genotípica /  
Adésio Ferreira. – Viçosa : UFV, 2006.  
xii, 132f. : il. ; 29cm.

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

Referências bibliográficas: f. 114-132.

1. Genética. 2. Melhoramento genético. 3. Mapeamento  
cromossômico. 4. Biologia molecular. 5. Marcadores  
genéticos. 6. Mendel, Lei de. 7. Variação (Biologia). I. Universidade  
Federal de Viçosa. II. Título.

CDD 22.ed. 576.5

ADÉSIO FERREIRA

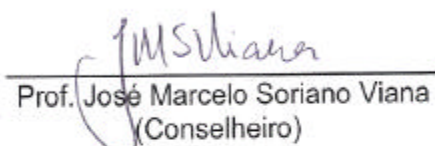
MAPEAMENTO GENÉTICO UTILIZANDO MARCADORES  
MOLECULARES COM DISTORÇÃO DE SEGREGAÇÃO GAMÉTICA E  
GENOTÍPICA

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 abril de 2006



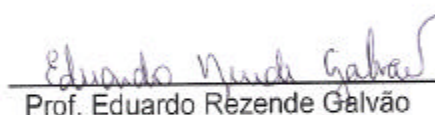
Prof. Paulo Roberto Cecon  
(Conselheiro)



Prof. José Marcelo Soriano Viana  
(Conselheiro)



Prof. Sebastião Martins Filho



Prof. Eduardo Rezende Galvão



Prof. Cosme Damião Cruz  
(Orientador)

A Deus.

A minha esposa Marcia

A minha filha Glória Maria

Aos meus pais, Sebastião e  
Glória Marlene.

Aos meus irmãos, Adenílson,  
Denize e Deizimar

Aos meus sobrinhos e  
afilhado

A minha tia Augusta

Ao meu primo Ciro e  
minhas primas Solange, Sônia e Luíza

À minha amiga Janete

A toda minha família

Aos meus queridos, que  
não estão mais comigo nesta vida:

Avô Joaquim, tio  
Ciro, tia Alzira, tio Roberto e prima  
Rita.

## **AGRADECIMENTO**

A Deus, pela infinita bondade, constante proteção e por mais esta graça concedida.

À Universidade Federal de Viçosa, pela oportunidade de realizar o Curso de Doutorado em Genética e Melhoramento.

Ao CNPq pela concessão da bolsa de estudo.

Ao professor Cosme Damião Cruz, pela orientação, ensinamentos, amizade, compreensão, respeito, dedicação e oportunidade.

Ao professor Paulo Roberto Cecon, pela amizade, estímulo, ensinamentos e apoio.

Ao professor José Marcelo Soriano Viana, pelos ensinamentos, respeito e amizade.

Ao professor Sebastião Martins Filho, pela amizade, ensinamentos, apoio e estímulo, desde a graduação.

Ao Professor Eduardo Rezende Galvão, pelas críticas e sugestões.

Ao professor Edvaldo Fialho dos Reis, pela confiança, incentivo e amizade.

A todos professores que tive em minha vida. Pois, cada um desfruta parcela de contribuição na concretização deste meu sonho.

Aos funcionários da genética, Dona Rita, Conceição e Sr. Paulo, pelo carinho e amizade.

Aos colegas de curso pela amizade e convívio.

Aos colegas de laboratório: Luciano, Edmar, Tatiana, Dudu, Luís, Fábio e Ricardo pela amizade e convívio.

A minha nova família Maria Eunice, Miguel, Adílson, César, Jacira, Aline, Eloísa.

Aos meus irmãos e familiares, pelo carinho e compreensão.

Aos meus queridos pais Sebastião e Glória Marlene pelo amor, ensinamentos, incentivos e apoio.

À minha esposa Marcia Flores da Silva Ferreira e minha filha Glória Maria Flores Ferreira que são meus amores, fontes de minha felicidade.

A todos que contribuíram, direta e indiretamente, para a realização deste trabalho.

## **BIOGRAFIA**

Adésio Ferreira, filho de Sebastião Dias Ferreira e de Glória Marlene Ferreira, nasceu em Mimoso do Sul, Espírito Santo, em 12 de outubro de 1971.

Formou-se como Técnico em Agropecuária, pela Escola Agrotécnica Federal de Alegre, em dezembro de 1990.

Em 29 de julho de 1992, foi contratado pela Empresa Florestas Rio Doce S/A, do grupo Companhia Vale do Rio Doce S/A, na área de reflorestamento e florestamento, no setor de manutenção, como Gerente de Reflorestamento.

Em agosto de 1997, iniciou o curso de graduação em Engenharia Agrônoma, pela Universidade Federal do Espírito Santo, onde em abril de 2002, graduou-se como Engenheiro Agrônomo.

Em 04 de abril de 2002, iniciou curso de Mestrado em Genética e Melhoramento, onde em 29 de julho de 2003 defendeu a tese de mestrado.

Em agosto de 2003 iniciou o curso de Doutorado em Genética e Melhoramento, na Universidade Federal de Viçosa, submetendo-se aos exames finais de defesa de tese no dia 03 de abril de 2006.

## CONTEÚDO

	Página
<b>RESUMO</b> .....	ix
<b>ABSTRACT</b> .....	xi
<b>1. INTRODUÇÃO</b> .....	1
<b>2. REVISÃO DE LITERATURA</b> .....	4
2.1. <i>Ocorrência e Utilização de Marcadores com DS no Mapeamento Genético</i> .....	4
2.2. <i>Causas de Distorção de Segregação</i> .....	6
2.3. <i>Ocorrência da Distorção de Segregação em Espécies Vegetais</i> .....	8
2.4. <i>Segregação Distorcida em Populações de Mapeamento</i> .....	10
2.5. <i>Efeito da DS sobre Regiões Cromossômicas</i> .....	12
2.6. <i>Distorção de Segregação em Cruzamentos Inter e Intraespecíficos</i> .....	14
2.7. <i>Distorção de Segregação em Cruzamentos Divergentes</i> .....	15
2.8. <i>Tipos de Marcadores</i> .....	16
<b>3. MATERIAL E MÉTODOS</b> .....	18
3.1. <i>Fundamentação Teórica</i> .....	18
3.1.1. <i>Testes de Segregação e Detecção de Segregação Distorcida</i> .....	18
3.1.2. <i>Tamanho de Populações para Detecção de Segregação Ambígua</i> ..	20
3.1.3. <i>Taxa de Segregação Distorcida</i> .....	20
3.1.4. <i>Porcentagem de Recombinação</i> .....	21
3.2. <i>Mapeamento Incluindo Marcadores com Segregação de Distorção</i> .	21
3.2.1. <i>Estratégia de Mapeamento</i> .....	21
3.2.2. <i>Simulação de Populações com Distorção de Segregação</i> .....	22
3.2.2.1. <i>Simulação do Genoma</i> .....	22



3.2.2.2. <i>Simulação de Genitores</i> .....	22
3.2.2.3. <i>Tamanho de População</i> .....	23
3.2.2.4. <i>Procedimento de Simulação dos Indivíduos da População</i> .....	23
3.2.3. <i>Mapeamento</i> .....	24
3.2.3.1. <i>Análise de Segregação de Locus Individuais</i> .....	24
3.2.3.2. <i>Análise de Pares de Marcas – Estimação da Percentagem de Recombinação</i> .....	24
3.2.3.3. <i>Determinação dos Grupos de Ligação</i> .....	25
3.2.3.4. <i>Ordenamento das Marcas no Grupo de Ligação</i> .....	25
3.2.4. <i>Comparação de Genomas</i> .....	25
3.2.4.1. <i>Número de Grupos de Ligação Obtidos no Mapeamento</i> .....	26
3.2.4.2. <i>Número de Marcas Obtidas por GL no Mapeamento</i> .....	26
3.2.4.3. <i>Tamanho do Grupo de Ligação</i> .....	26
3.2.4.4. <i>Distância Média Entre Marcadores Adjacentes</i> .....	26
3.2.4.5. <i>Variância das Distâncias Entre Marcas Adjacentes</i> .....	26
3.2.4.6. <i>Estresse</i> .....	26
<b>4. RESULTADO E DISCUSSÃO</b> .....	28
4.1. <i>Segregação Ambígua</i> .....	28
4.1.1. <i>População <math>F_2</math> com Marcadores Dominante</i> .....	28
4.1.2. <i>População de Retrocruzamento</i> .....	32
4.2. <i>Estimadores das Taxas de Distorção de Segregação</i> .....	35
4.2.1. <i>Taxa de Distorção Esperada</i> .....	36
4.2.1.1. <i>Taxa de Distorção Gamética em População <math>F_2</math>, com Marcadores Dominantes</i> .....	36
4.2.1.2. <i>Taxa de Distorção Genotípica em População <math>F_2</math>, com Marcadores Dominantes</i> .....	38
4.2.1.3. <i>Taxa de Distorção Gamética em População Retrocruzamento</i> .....	40
4.2.1.4. <i>Taxa de Distorção Genotípica em População Retrocruzamento</i> ...	41
4.2.2. <i>Taxa de Distorção Observada</i> .....	42
4.2.2.1. <i>População <math>F_2</math>, com Marcadores Dominantes</i> .....	43
4.2.2.2. <i>População de Retrocruzamento</i> .....	44
4.3. <i>Segregação de Pares de Locus</i> .....	45
4.3.1. <i>Distorção Gamética</i> .....	45

4.3.1.1. População $F_2$ , com Marcadores Dominantes em Fase de Aproximação .....	45
4.3.1.2. População de Retrocruzamento, $F_1$ em Aproximação .....	54
4.3.1.3. População $F_2$ com Marcadores Codominantes, Derivada de $F_1$ em Aproximação .....	56
4.3.1.4. População $F_2$ , com Marcadores Codominantes, Derivada de $F_1$ em Repulsão .....	58
4.3.1.5. População $F_2$ , com Marcadores Dominantes, Derivada de $F_1$ em Repulsão .....	60
4.3.1.6. População de Retrocruzamento, $F_1$ em Repulsão .....	62
4.3.1.7. População de Duplo-Haplóide, Derivada de $F_1$ em Aproximação ..	64
4.3.1.8. População de Duplo-Haplóide, Derivada de $F_1$ em Repulsão .....	65
4.3.1.9. População de RIL, Derivada de $F_1$ em Aproximação .....	66
4.3.1.10. População de RIL, Derivada de $F_1$ em Repulsão .....	68
4.3.2. Distorção Genotípica .....	69
4.4. Mapeamento Incluindo Marcadores com Distorção de Segregação .....	75
4.4.1. Distorção Gamética .....	75
4.4.1.1. Análise de Segregação de Locos Individuais .....	76
4.4.1.2. Mapeamento Genético .....	80
4.4.1.2.1. Número de Grupos de Ligação .....	81
4.4.1.2.2. Número de Marcas Obtidas por GL no Mapeamento .....	81
4.4.1.2.3. Tamanhos de Grupos de Ligação .....	82
4.4.1.2.4. Distância Média Entre Marcadores Adjacentes .....	83
4.4.1.2.5. Variância das Distâncias Entre Marcadores Adjacentes .....	84
4.4.1.2.6. Estresse .....	85
4.4.1.2.7. Considerações Gerais .....	87
4.4.2. Distorção Genotípica .....	104
4.4.2.1. Mapeamento Genético.....	105
4.4.2.2. Considerações Gerais.....	110
<b>5. RESUMO E CONCLUSÕES .....</b>	<b>112</b>
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>114</b>

## RESUMO

FERREIRA, Adésio, D.S., Universidade Federal de Viçosa, abril de 2006.  
**Mapeamento genético utilizando marcadores moleculares com distorção de segregação gamética e genotípica.** Orientador: Cosme Damião Cruz. Conselheiros: Paulo Roberto Cecon e José Marcelo Soriano Viana.

O desenvolvimento de mapas genéticos fundamentados em marcadores de DNA tem propiciado consideráveis avanços à genômica de plantas e animais. A construção de um mapa genético é feita com dados oriundos de populações segregantes, usualmente, retrocruzamentos, geração  $F_2$ , RILs, duplo-haplóides, dentre outras. Para cada marcador utilizado, espera-se padrão mendeliano de segregação, típico para cada população, que possibilita prever relações genotípicas e estimar recombinações a partir de pressupostos dos modelos probabilísticos. A ausência do padrão de segregação mendeliano é chamada de distorção de segregação (DS), que pode afetar testes estatísticos usados para detectar a ligação e gerar falsos positivos em um mapa genético. Diante dos problemas que a DS pode provocar em relação à construção de mapas de ligação genética, foram abordados, neste trabalho, temas relativos à detecção, aos efeitos e às implicações da DS no estabelecimento de mapas genéticos acurados. Verificou-se que a DS genotípica é mais prejudicial que a DS gamética. Valores de estimativas de distância entre marcas, mesmo com DS gamética, são viesados, ao negligenciar a distorção, principalmente com níveis de razão de segregação mais distantes que a segregação esperada. Diante destas constatações, foram desenvolvidas funções de verossimilhança que levam em consideração a DS gamética para as populações de retrocruzamentos, geração  $F_2$ , RIL e duplo-haplóides.

Adicionalmente, foram propostas duas estratégias para a utilização destas funções de verossimilhança. Na primeira estratégia, é pressuposto serem conhecidas as marcas que apresentam distorção de segregação e também a taxa esperada, de forma que o valor da taxa de DS pode ser assumido como paramétrico. Na segunda estratégia, o valor da taxa de DS é estimado a partir dos dados observados, relativos a todos os marcadores estudados. Este valor é, então, utilizado para o cálculo do valor da distância entre os marcadores  $i$  e  $i'$ . Como, para o marcador  $i$ , a taxa de distorção será dada por  $\hat{s}_i$  e o marcador  $i'$  apresenta taxa de distorção  $\hat{s}_{i'}$ , são obtidas duas estimativas da medida da porcentagem de recombinação entre  $i$  e  $i'$ , sendo o valor da distância a estimativa média. Observou-se, com base em análise de dados simulados de uma população  $F_2$  com marcadores dominantes, que as duas estratégias são mais eficientes para a construção de mapas genéticos acurados, do que aquela que negligencia a distorção de segregação. Constatou-se, ainda, que a segunda estratégia pode ser utilizada na construção de mapa genético acurado, mesmo sem a presença nítida de DS e dispensa a utilização do teste de qui-quadrado, proporcionando mapas mais acurados. Os efeitos sobre o mapeamento genético, em razão de distorção genotípica, também foram enfatizados neste trabalho.

## ABSTRACT

FERREIRA, Adésio, D.S., Universidade Federal de Viçosa, April 2006.  
**Genetic mapping by molecular markers with gametic and genotypic segregation distortion.** Adviser: Cosme Damiano Cruz. Committee members: Paulo Roberto Cecon and José Marcelo Soriano Viana.

The development of the genetic maps based on DNA markers has been providing remarkable progresses to the genomics of plants and animals. The construction of a genetic map is performed, by using data from the segregant populations, that usually are backcrossing, F2 generation, RILs, double-haploids and others. A mendelian segregation pattern that is typical to each population and makes possible to foresee genotypic relationships, as well as to estimate recombinations from the presuppositions within the probabilistic models is expected for each marker. The absence of the mendelian segregation pattern is so-called segregation distortion (DS), that may affect the statistical tests used to detect the linkage as well as to generate positive falses in a genetic map. Taking into account the problems the DS may cause in relation to the construction of the genetic linkage maps, this study is concerning to the themes on the detection, effects and implications of DS in the establishment of accurate genetic maps. It was found that genotypic DS is more prejudicial than gametic DS. The estimative values of the distances among marks even with gametic DS are biased when neglecting the distortion, mainly with more distant levels of segregation ratio than the expected segregation. Due to those verifications, some likelihood functions taking into account the gametic DS for the backcrossing populations, F2 generation, RIL and double-haploid were developed. In

addition, two strategies were proposed for using these likelihood functions. In the first strategy, it is presupposed that the marks presenting segregation distortion and also the expected rate are known, so the value of the DS rate could be assumed as a parametric one. In the second strategy, the value of the DS rate is estimated from the observed data concerning to the markers under study. Then, this value is used for calculating the value of the distance between the markers  $i$  and  $i'$ . As for the marker  $i$  the distortion rate will be given by  $\hat{s}_i$  and the marker  $i'$  inhibits a distortion rate  $\hat{s}_{i'}$ , two estimates of the recombination percentage measure between  $i$  and  $i'$  are obtained, whereas the distance value is the mean estimate.

## 1. INTRODUÇÃO

As bases da hereditariedade foram formuladas por Mendel, e a tão conhecida primeira lei, ou Lei da Segregação, postula sobre a transmissão previsível dos alelos dos pais para a progênie, ressaltando que cada membro de um par de alelos tem igual oportunidade de ser transmitido para a próxima geração; também postula sobre a formação previsível de genótipos em consequência dessa transmissão dos alelos.

A primeira lei de Mendel tem sido aplicada nos estudos de diversos caracteres, em uma gama extensiva de espécies. Outras áreas da ciência, como a genética de populações, quantitativa e evolucionária assumem em seus princípios esta clássica definição de segregação mendeliana para os caracteres, desde que De Vries, Von Tschermak e Correns redescobriram os trabalhos de Mendel em 1909 (Säll & Nilsson, 1994; Lorieux et al., 1995a). Zamir & Tadmor (1986) relatam que, embora as leis de Mendel terem sido baseadas numa variação intra-específica, elas têm sido geralmente aplicadas também em variações interespecíficas.

Entretanto, logo após a redescoberta dos trabalhos de Mendel, Correns (1902) reportou segregação monogênica aberrante, denominada de Distorção de Segregação, ou simplesmente DS, em que a segregação gamética não ocorreria como previsto por Mendel. Desde então, numerosos casos têm sido relatados e descritos pormenorizadamente, particularmente em cruzamentos interespecíficos (Grant, 1975). Estes desvios da proporção mendeliana esperada nos indivíduos, em dada classe genotípica, em populações segregantes, é definida como distorção da razão de segregação

(DS) (Sandler & Novitski, 1957; Sandler & Golic, 1985; Lyttle, 1991). Ela resulta na falha ou violação de pressupostos da teoria genética convencional e suas análises (Lu et al., 2002).

Distorção de segregação tem sido relatada em uma extensiva gama de organismos, incluindo plantas, nas quais espécies ou raças híbridas exibem preferencialmente disfunção de gametas (Xu et al., 1997). Pode ocorrer no pólen (Cameron & Moav, 1957; Loegering & Sears, 1963; No 1982; Finch et al., 1984; Tsujimoto & Tsunewaki, 1985), no megásporo (Scoles & Kibirge-Sebunya, 1983), ou em ambos (Rick, 1966). A base genética da distorção de segregação pode ser o aborto de gametas macho ou fêmeas e também a fertilização seletiva de uma estrutura gamética com configuração genotípica particular. Pode envolver todos cromossomos e pode ser causada por processos de seleção no estágio gamético ou zigótico (Gadish & Zamir, 1986; Zamir & Tadmor 1986).

Em análise genômica, em especial relacionada ao estudo de mapa de ligação genético, são confeccionadas populações especiais, denominadas populações de mapeamento, em que se espera a ocorrência de segregação mendeliana clássica dos marcadores analisados, de forma que a segregação esperada dos marcadores seja conhecida. A presença de marcadores com DS podem ser um inconveniente, devido à construção de mapas de ligação genética, ser realizada a partir de frequências de recombinação entre locos em populações experimentais ou populações naturais com pedigrees conhecidos.

De acordo com Säll & Nilsson (1994) e Lorieux et al. (1995a), as estimativas de recombinação entre pares de locos, em que a DS ocorre em pelo menos um deles, são viesadas. Assim, seriam obtidos mapas poucos acurados, o que é desvantajoso devido ao interesse crescente em mapas mais acurados e de alta resolução (Hackett & Broadfoot, 2003). Deve-se ter mente que estes mapas são ferramentas essenciais para aplicações práticas, tais como seleção assistida por marcadores e clonagem de genes. Para o caso de clonagem de genes com base em mapas, por exemplo, a região cromossômica onde um gene é localizado é primeiramente determinada com um mapa de ligação com altíssimo grau de saturação, e a



ordenação dos marcadores neste tipo de mapa e a localização deles, podem ser usadas para o screening de livrarias genômicas (Frisch et al., 2004).

Diante dos problemas que a DS pode provocar em relação à construção de mapas de ligação genética, métodos que a consideram poderiam ser úteis na análise e no entendimento do processo, na construção de mapas genéticos acurados e utilização em processos como mapeamento de QTLs, seleção assistida e clonagem de genes. Negligenciar a ocorrência de DS não é prática recomendável, tendo em vista a grande necessidade de se obter informações confiáveis para tomada de decisão em estudos genéticos teóricos e aplicados ao melhoramento.

Diante do exposto, este trabalho foi conduzido com objetivo de abordar a detecção, os efeitos e as implicações da DS em estudos genômicos, principalmente relacionados com o estabelecimento de mapas genéticos acurados. Também tem-se por objetivo propor metodologia para o mapeamento genômico, admitindo a existência de marcadores moleculares com DS em razão de segregação gamética diferencial.

## 2. REVISÃO DE LITERATURA

### 2.1. *Ocorrência e Utilização de Marcadores com DS no Mapeamento Genético*

Na construção de mapas genéticos, feito com dados oriundos de populações segregantes, espera-se um padrão mendeliano típico para cada tipo de população, que é pressuposto com base na segregação mendeliana. Para verificar se um loco está segregando como esperado, procede-se uma comparação do número de indivíduos observados em cada classe com o esperado de acordo com a razão de segregação. Utiliza-se para este fim o teste de qui-quadrado ( $\chi^2$ ), que leva em consideração os desvios ocorridos entre os valores esperados e observados, o qual é sensível ao tamanho da amostra. A estatística qui-quadrado é dada por:

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

em que:

$\chi^2$  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 ( $i= 1, 2, \dots, n$ ).

Quando ocorre a presença de marcadores que fogem da segregação mendeliana clássica, diz-se ter ocorrido a DS. Segundo Liu (1998), os locos que apresentam distorção mendeliana afetam os testes estatísticos usados para detectar a ligação, podendo gerar falsos positivos. De acordo com Oliveira et al. (2004), a presença de marcadores com DS localizados entre

marcadores com segregação mendeliana esperada, acarreta acréscimo na estimativa da distância entre os marcadores.

Quanto ao que deve ser feito em relação aos marcadores que apresentam DS é discutível. De acordo com Kao et al. (1999), é recomendado, preferencialmente, o descarte dos locos que apresentam distorções da segregação mendeliana, para não comprometer a qualidade do mapa. No entanto, há trabalhos de mapeamento genético em que se utilizou marcadores com distorção na razão de segregação, e outros apenas indicando-os no grupo de ligação com o nível de probabilidade utilizado no teste de segregação individual, sem saber os efeitos da utilização de tais marcadores (CostaeSilva et al., 2004).

De acordo com Oliveira et al. (2004) e Jarrell et al. (1992), deve-se ter cautela no uso de locos que apresentam DS, pois apesar deles poderem mostrar informações genéticas importantes, também podem alterar as distâncias e ordens lineares de outros marcadores no grupo de ligação.

Com o objetivo de estudar as conseqüências da utilização de locos com distorção na razão de segregação, para a estimação da freqüência de recombinação e para determinação da ligação gênica, CostaeSilva et al. (2004) utilizaram uma população de retrocruzamento do F<sub>1</sub> em fase de aproximação. Foram obtidas estimativas de LOD e de freqüência de recombinação para vários tamanhos de população, em que se considerou dois genes com freqüência de recombinação variando de 0,01 a 0,50, variando também os graus de distorção, de 0,01 ao valor de  $r$ , para cada situação. A eficácia em detectar a ligação com marcadores com segregação distorcida foi maior quanto maior o tamanho da população utilizada e quanto mais saturado o mapa.

São raros os casos em que conjuntos de dados com a presença de marcadores com DS são utilizados para o mapeamento (Harushima et al., 1996; Kuang et al., 1998). Dois exemplos são os trabalhos de Yu & Guo (2003) e Crouzilalat et al. (1996), que incluíram na análise de mapeamento os marcadores que apresentaram DS para uma possível identificação de regiões de distorção. Em *Citrus* geralmente observa-se a construção de mapas de ligação com base em marcadores moleculares considera a

presença de marcadores com DS (Cai et al., 1994; Luro et al., 1994; Kijas et al., 1997; Cristofani et al., 1999).

Oliveira et al. (2004) sugerem a utilização de marcadores que apresentam DS, mas somente aqueles que conferem pequeno grau de alteração das distâncias no mapa, quando acrescidos. E, ainda, que marcadores com DS presentes entre marcadores com segregação mendeliana esperada poderiam ser eliminados, ou então, estimar a sua distância correta por um artifício adequado. Em relação àqueles marcadores que se encontram no final dos grupos de ligação ou os que formam grupos de ligação, os autores relatam que é essencial a remoção destes para estudos genéticos e mapeamento.

Normalmente, após a construção de mapas de ligação vem a análise de QTLs. Assim, se frações de recombinação, ou pior, a ordem dos marcadores forem inferidas incorretamente, as pressuposições básicas da análise de QTLs irão ser violadas e os resultados poderão ser imprecisos (Vogl & Xu, 2000).

Diante dos fatos, verifica-se a necessidade de estudos dos diferentes tipos e intensidade de DS, considerando diferentes tipos de população, no sentido de esclarecer a questão da utilização ou não de marcadores com DS. Deve-se ter em mente que a obtenção de marcadores demanda mão-de-obra especializada, tempo e custo relativamente alto. Portanto, investimentos para o uso adequado das informações é fundamental para promover conhecimento e utilidade do investimento até então aplicado na pesquisa.

## 2.2. *Causas de Distorção de Segregação*

Em plantas, a distorção na razão de segregação de marcadores pode ser atribuída a uma série de causas genéticas e fisiológicas (Grant, 1975), podendo ser manifestada pela transmissão diferencial de alelos tanto de linhagens germinativas macho, como de fêmea (Longley, 1945; Rhoades & Dempsey, 1966; Lavery & James, 1987; Yanagihara et al., 1995), incluindo competição de tubo polínico (Mangelsdorf & Jones, 1926; Levin & Berude, 1972; Liedl & Anderson, 1993); pólen letais (Rick, 1996), fertilização preferencial (Schwemmler, 1968; Gadish & Zamir, 1986) e eliminação seletiva

de zigotos (Rick, 1963; Gadish & Zamir, 1986; Zamir & Tadmor, 1986); locos gênicos que apresentam seleção natural próximos a genes que levam a menor viabilidade de gametas, como observado em arroz por He et al. (2001); ou sendo resultado de seleção pós-zigótica, anterior à primeira avaliação genotípica.

Freqüentemente, a distorção de segregação parece aumentar pela seleção gametofítica em macho, pela influência seletiva do gineceu, incluindo incompatibilidade genética, efeitos ambientais e habilidades diferenciais competitivas do pólen geneticamente variado (Xu et al., 1997), e a presença de classes de genótipos homozigotos ou heterozigotos em excesso é forte indicação de DS por meio de fatores gametofíticos.

As distorções de segregação mendeliana podem ser atribuídas à seleção que ocorre em uma ou mais das seguintes fases do ciclo de vida da planta: 1) esporogênese - da meiose à formação dos gametófitos maduros macho e fêmea; 2) função do esporo – começando na polinização e terminando com a cariogamia; 3) desenvolvimento da semente – do zigoto à semente; 4) germinação da semente e crescimento da planta até que a população seja mensurada (Grant, 1975; Willson & Burley, 1983; Zamir, 1983).

Uma outra possível causa da DS é a duplicação de marcadores. A distinção da DS causada entre duplicação de marcadores e seleção zigótica pode ser verificada por meio de alguns fatos descritos a seguir. Quando a DS tem causa na duplicação de marcadores, geralmente ela se manifesta somente em um ou poucos locos específicos (algumas vezes denominados de locos), de forma que locos vizinho, pertencentes ao mesmo grupo de ligação, permanecem com o padrão de segregação mendeliano esperado. Em contraste, quando a DS ocorre em razão de seleção zigótica ou gamética, a distorção não se manifesta apenas para o loco que é afetado por seleção, mas também para os que estão ligados a ele (Frisch et al., 2004). Assim, conseqüentemente, se distorção de segregação é detectada somente em um loco, a chance é alta que seja devido à duplicação de marcadores. Entretanto, nos casos em que a DS é detectada em muitos locos ligados, isto poderá ser interpretado como uma indicação de seleção zigótica (Frisch et al., 2004).

Xu et al. (1997) relatam que a DS em plantas pode depender da presença de fatores específicos, como já relacionados, mas, também, pode ser condicionada por combinações destes fatores.

Em determinadas situações a viabilidade de seleção após a fertilização pode ser mais importante que seleção gamética, por exemplo, em acasalamentos consangüíneos em que há possibilidade de redução da taxa de sobrevivência de homozigotos, quando comparado com heterozigotos (Charlesworth & Charlesworth, 1987). Uma distinção precisa entre DS antes e após a fertilização pode ser feita. Um marcador com DS atuando antes da fertilização pode apenas alterar proporções gaméticas, e as proporções genótípicas serão alteradas apenas se, indiretamente as proporções gaméticas restringirem as combinações realizáveis das proporções genótípicas. Por outro lado, a DS após a fertilização pode alterar diretamente as proporções genótípicas (Vogl & Xu, 2000) e, inclusive, romper a estrutura genética da população, tornando-a inapropriada para fins de mapeamento.

### 2.3. *Ocorrência da Distorção de Segregação em Espécies Vegetais*

A constatação de DS tem sido observada em inúmeras espécies. Um dos primeiros trabalhos que reportam DS foi em milho (Mangelsdorf & Jones, 1926). Subseqüentemente, distorção em milho foi citada por Burnham (1936), Rhoades (1942), Longley (1945), Helentjaris et al. (1986), Wendel et al. (1987) e Gardiner et al. (1993).

Exemplos de taxas de DS têm sido reportados em diversas culturas vegetais, e com variáveis porcentagens, como arroz (*Oryza sativa* L.) (Nakagahra, 1972; McCouch et al., 1988; Xu et al., 1997), de 36% *Oryza* (Virk et al., 1998), 33% em *Pumus* (Foolad et al., 1995), 23% em *Helianthus* (Quillet et al., 1995), 8,4% em *Lens sp.* (Eujayl et al. 1998), 1,4% em *Hevea spp.* (Lespinasse et al., 2000), 51% a 80% em *Lycopersicum*, no cruzamento de uma espécie selvagem com a cultivada *L. esculentum* (Paterson et al., 1988, 1991), cevada (*Hordeum vulgare* L.) (Graner et al., 1991; Devaux et al., 1995), sorgo (*Sorghum bicolor* L.) (Pereira et al., 1994), alfafa (*Medicago sativa* L.) (Echt et al., 1994) e Café (*Coffea sp.*) (Ky et al., 2000).

Em cada espécie as causas e a intensidade de DS podem variar. A seguir são sumarizados alguns exemplos em diferentes culturas.

Em milho, Lu et al. (2002) verificaram que os fatores genéticos mais comumente associados com a DS são gametofíticos, e identificaram 34% de marcadores distorcidos, os quais foram desigualmente distribuídos nos dez cromossomos do milho.

Em sugi (*Cryptomeria japonica*) tem sido considerado que fatores de natureza biológica, como a existência de genes letais, seja uma das razões da segregação distorcida (Ohba, 1979; Kawasaki & Ohba, 1984; Tsumura et al., 1989). Porém, estes autores ressaltam que é complicado afirmar que a distorção em todas as marcas envolvidas seja devido a uma única causa biológica.

Ky et al. (2000), estudando *Coffea spp*, em 62 retrocruzamentos derivados de cruzamento interespecífico, constataram 30% de DS. Apesar da segregação mendeliana esperada ser de 1:1, os autores obtiveram razão de segregação de 3:1 e 1:3, em favor de um parental ou outro. Os autores relatam que as distorções são, na maioria das vezes, devidas à seleção de gametas, zigoto ou pós-zigoto, mas que estes motivos não são suficientes para explicar a distribuição simétrica e polimodal de 3:1 e 1:3, como eles obtiveram, pois este processo tenderia às razões de 5:3 e 3:5.

Citando Leblon & Rossignol (1973), Nicolas & Rossignol (1983) e Nag et al. (1989) relatam que a conversão gênica pode ter sido a causadora de muitas razões de segregação distorcida. Mas ratificam, baseados em Nag et al. (1989), que a melhor explicação para os seus resultados de DS seria a ocorrência de alta taxa de conversão gênica e baixa taxa de segregação pós-meiótica, reflexo do alto nível de formação de heteroduplex e eficiente reparo dos pareamentos errados.

A conversão gênica é a situação na qual os produtos de uma meiose de um indivíduo AA'são 3A e 1A' ou 1A e 3A', e não 2A e 2A' como é usualmente esperado. Assim, pode-se ter a impressão que um gene A foi convertido em A' (ou vice versa). A conversão gênica envolve também um erro raro no reparo do DNA, que ocorre enquanto a recombinação está ocorrendo na prófase meiótica. Um filamento duplo quebra e um bivalente é alargado para eliminar um alelo do filamento complementar.

Em uma população  $F_2$  de soja, Yang et al. (2000) verificaram 37% de distorção da razão segregação para locos polimórficos de AFLP. Os autores atribuíram a DS a fatores causados por um fator parcial letal, agindo no final das gerações, tais como expressão de genes ligados a genes letais ou no desenvolvimento de estágios zigóticos ou gaméticos, ou pela seleção parcial gamética em ambos os genitores, macho e fêmea.

A DS em arroz é citada freqüentemente, especialmente entre cruzamentos de subespécies selvagens (Nakagahra, 1972; Nakagahra et al., 1972; Maekawa et al., 1981; Maekawa & Kita, 1985; Lin et al., 1992; Xu & Shen, 1992; Lin & Ikehashi, 1993), e também quando se utiliza marcadores de isoenzimas (Ishikawa et al., 1987; Ranjhan et al., 1988; Wu et al., 1988; Guiderdoni et al., 1989) e com a utilização de RFLP (McCough et al., 1995; Huang et al., 1994; Causse et al., 1994; Xu et al., 1995; Harushima et al., 1996; Xiao et al., 1996; Yamagishi et al., 1996). Um exemplo o trabalho desenvolvido por Xu et al. (1997), que estudaram o controle genético da distorção da razão de segregação, usando marcadores isoenzimáticos e morfológicos, 15 locos de genes gametofíticos, distribuídos em 11 regiões cromossômicas. Também se constatou a presença de 20 locos de genes de esterilidade em sete regiões cromossômicas, além de em outros nove cromossomos terem sido identificados genes de esterilidade e gametofíticos. Em arroz, sugere-se muito que a distorção de segregação em um loco marcador possa ser causada pela ligação entre um marcador e um gene gametofítico (ga) conferindo baixa habilidade de polinização, também referido como eliminador de gameta ou pólen letal causando aborto dos gametas (Iwata et al., 1964; Nakagahra, 1972 e 1986; Sano et al., 1979; Sano, 1983 e 1990).

#### 2.4. *Segregação Distorcida em Populações de Mapeamento*

Para o mapeamento genético de plantas, diversos tipos de populações podem ser utilizadas, cada uma com características próprias, apresentando vantagens e desvantagens que devem ser levadas em consideração pelo pesquisador no momento de selecionar populações para fins de estudos genéticos (Schuster & Cruz, 2004).



Estudos comprovam que a distorção de segregação apresenta-se de maneira diferenciada nas populações. Um relato é apresentado de Xu et al. (1997), que compararam os resultados de ocorrência de distorção da razão de segregação de marcadores em 53 populações de diferentes tipos ( $F_2$ , RC, RIL e DH) em diferentes espécies, e demonstraram que populações de RIL apresentaram maior frequência de distorção de marcadores ( $39,4 \pm 2,5\%$ ). As frequências médias de distorção em DH e RC foram quase sempre idênticas (DH:  $29,4 \pm 3,5\%$ ; RC:  $28,6 \pm 2,8\%$ ). A população  $F_2$  foi a que apresentou menor frequência de distorção de marcadores ( $19,3 \pm 11,2\%$ ).

Lu et al. (2002), trabalhando com quatro populações de milho, analisaram populações de RIL e  $F_2$  e observaram também menor ocorrência de DS na população  $F_2$ .

Veldboon & Lee (1994), analisando famílias  $F_3$  e RILs em milho, encontraram marcadores com 9% de DS nas famílias  $F_3$  e com 29% entre as RILs, concordando com os resultados encontrados Xu et al. (1997), que também verificaram que populações de RIL apresentam as mais altas frequências de marcadores distorcidos. Todos estes resultados indicam que DS são acumuladas com gerações adicionais de meioses, fato inerente à obtenção de RILs. Isto é, existe uma positiva relação entre o número de gerações de meioses e a frequência de DS. Deve-se ressaltar que, sendo a população de RILs obtida via SSD, resultante de sucessivas gerações de autofecundação, concentra o efeito cumulativo de fatores genéticos e ambientais e a interação destes em múltiplas gerações. Assim, é difícil distinguir causas genéticas e ambientais de frequências de alelos distorcidos neste tipo de população (Xu et al., 1997).

Entre todas as 53 populações exemplificadas no trabalho de Xu et al. (1997), uma população de RIL em tomate (Paran et al., 1995) teve a mais alta taxa de DS, 73%. E um cruzamento intraespecífico em população  $F_2$  de *Cuphea lanceolata* (Webb et al., 1992) apresentou a menor taxa de DS, igual a 5,4%. Um exemplo extremo relatado na literatura é o encontrado em arroz, numa população intraespecífica de RIL (Wang et al., 1994) em que se observou a presença de 98,8% de locos marcadores apresentando DS.

Em populações de DH, derivadas de cultura de anteras (gametófitos machos), a distorção de segregação pode ser atribuída à viabilidade diferencial e letalidade do pólen ou à regeneração seletiva em cultura in vitro, que claramente não sofre influência do gineceu ou habilidade diferencial competitiva do pólen, estando associada aos efeitos ambientais (Xu et al., 1997).

A regeneração seletiva de genótipos tem sido freqüentemente relatada na literatura, como no estudo desenvolvido por Graner et al. (1991) que, trabalhando com população de DH de cevada, encontraram 44% de marcadores apresentando DS, causada principalmente por prevalência de alelos do pai que respondem melhor em cultura in vitro.

Um exemplo de DS em DH, causada por genes letais, é relatado em *Coffea canephora*, em que a principal causa do pronunciado grau de DS é a seleção zigótica, devida à expressão de genes letais e sub-letais nos homocigotos (Lashermes et al., 2001).

Deve-se observar, ainda, que razões de segregação 3:1 ou 15:1 em DH, podem ser causadas por co-migração de bandas representando dois ou três locos respectivamente. Conseqüentemente, algumas supostas distorções podem ser consideradas como sendo originárias da superimposição de locos provocadas por duplicações gênicas (Virk et al., 1998).

Estudos com populações de retrocruzamento também revelam DS, principalmente em retrocruzamentos derivados de cruzamentos interespecíficos. O trabalho apresentado por Ky et al. (2000) é um exemplo deste fenômeno.

## 2.5. *Efeito da DS sobre Regiões Cromossômicas*

Em estudos de mapeamento em que há ocorrência de DS em certos marcadores tem sido constatado que se um gene que manifesta DS está segregando em população, então marcadores ligados a ele tenderão a exibir razões de distorção em grau diferenciado (Zamir & Tadmor, 1986). Se muitas populações estão segregando para o mesmo fator gamético, ou outro fator desconhecido que causa DS, então estas populações exibirão DS para a mesma região cromossômica. Assim análises com marcadores

moleculares em diversas populações é uma ferramenta útil para encontrar regiões comuns com DS e para a identificação futura de genes desconhecidos que causam DS nestas regiões.

Na literatura existem diversos trabalhos, em diferentes espécies, que localizaram regiões cromossômicas associadas a marcadores que apresentavam DS. Um destes é o de Wendel et al. (1987), em milho, que observaram 65% de locos, utilizando marcadores aloenzimáticos, com DS significativa numa população  $F_2$ . Gardiner et al. (1993), trabalhando com a mesma cultura, detectaram regiões cromossômicas associadas com DS nos cromossomos 1, 2, 3 e 5. Em outro trabalho com milho (Lu et al., 2002), verificaram 34% de marcadores com DS, distribuídos ao longo dos 10 cromossomos, com exceção, apenas para o cromossomo 3. Os autores relataram que estes resultados são fortes indicativos que fatores genéticos de DS existem para a maioria dos cromossomos e eles determinam a direção de menor distorção no cromossomo.

Em tomate, Paterson et al. (1988) encontraram 68% dos marcadores, de 21 regiões distintas com DS em um retrocruzamento interespecífico.

Xu et al. (1997) caracterizaram 17 regiões cromossômicas associadas com marcadores que apresentavam DS, em seis populações segregantes de arroz, uma região no cromossomo 1, duas no cromossomo 2, duas no cromossomo 3, uma no cromossomo 4, uma no cromossomo 5, duas no cromossomo 6, duas no cromossomo 7, uma no cromossomo 8, uma no cromossomo 9, uma no cromossomo 10, uma no cromossomo 11 e duas no cromossomo 12.

Yamada et al. (2001), trabalhando com uma população  $F_2$  de soja, encontraram algumas regiões envolvidas com DS.

A implicação de regiões cromossômicas associadas à DS na estimativa de distância de mapa depende do número e tipo de fatores genéticos que causam a DS. De acordo com Lu et al. (2002) e Lorieux et al. (1995a, b), a estimação da frequência de recombinação permanece inalterada se a distorção de segregação é devida somente a um loco em uma região de distorção de segregação. Tal fato parece ser verdadeiro em alguns tipos de populações, tais como RC, RILs e DH, nas situações em que a distorção ocorre por segregação gamética diferencial. Porém, segundo

Lorieux et al. (1995a, b), quando a DS está relacionada com a presença de mais de um fator gametofítico, a estimativa da frequência de recombinação e, conseqüentemente, a distância entre marcadores em um mapa de ligação será viesada. Lu et al. (2002) relatam que na maioria das vezes um fator gametofítico está presente em uma região de distorção de segregação, e que métodos de mapeamento por intervalo, para o mapeamento de locos que causam distorção de segregação, podem ser úteis para solucionar problemas de regiões com distorção de segregação.

#### 2.6. *Distorção de segregação em Cruzamentos Inter e Intraespecíficos*

A transferência de genes de espécies selvagens para outras cultivadas, por meio de cruzamentos amplos, é uma das estratégias de melhoramento. Trocas gênicas são possíveis devido ao processo de recombinação meiótica, que permite a informação de um cromossomo parental ser combinada dentro de uma nova entidade genética e, que posteriormente, são repassados para as próximas gerações. Entretanto, problemas inerentes a cruzamentos interespecíficos, tais como instabilidade híbrida, infertilidade, segregação não mendeliana e baixos níveis de crossing-over inter-genômico, podem ser importantes limitações (Stebbins, 1950, em Herrera et al., 2002). Em relação à DS, tem sido freqüentemente relatada em cruzamentos interespecíficos para todos os tipos de marcadores (morfológicos, izoenzimas, RFLP ou AFLP) (Ky et al., 2000).

Em cruzamentos intra-específicos, muitos marcadores obedecem a lei de transmissão de fatores genéticos postulada por Mendel. Nos cruzamentos interespecíficos, muitos destes locos são heterozigotos, conduzindo a reprodução desbalanceada (Zamir & Tadmor, 1986). Neste tipo de cruzamento (interespecífico) a alta taxa de formação de heteroduplex e o pareamento errôneo do DNA são esperados durante o pareamento de cromossomos homólogos, e obviamente a conversão gênica parece ser característica de cruzamentos interespecíficos, não dependendo do tipo de marcador (Lu et al., 2002).

Zamir & Tadmor (1986) estudaram a proporção de genes com DS em cruzamentos intra e interespecíficos (*Lens*, *Capsicum* e *Lycopersicum*). A proporção de desvio da razão da segregação monogênica encontrada nos

gêneros *Lens*, *Capsicum* e *Lycopersicum* foi significativamente alta (61/114 genes, 54%) nas progênes de híbridos interespecífico, em relação às progênes de híbridos intraespecífico (7/52 genes, 13%).

A DS em híbridos é comumente observada e potencialmente representa níveis de isolamento reprodutivo devido aos arranjos cromossômicos ou interações gênicas (Rieseberg et al., 1995). Em híbridos pode-se esperar a presença de supressão de recombinação, quando comparados com a espécie pura, devido à diferenciação dos cromossomos homólogos das espécies parentais (Jackson, 1985; Tenhoopen et al., 1996; Chetelat et al., 2000). Distorção de segregação em híbridos é comumente atribuída a fatores como incompatibilidade pólen-pistilo (Lord & Russell, 2002); competição gamética (Snow, 1984; Carney et al., 1996; Lu et al., 2002); interação epistática negativa entre alelos (Li et al., 1997; Fishman et al., 2001); presença de locos com segregação distorcida, que resulta na destruição alternada de gametas (processo meiótico: Hartl, 1974; Sano, 1990); ou seleção positiva para introgressão de alelos (Jiang et al., 2000; Burke & Arnold, 2001). Os híbridos interespecíficos são freqüentemente mais caracterizados por distorção de segregação (Zamir & Tadmor, 1986; Ky et al., 2000),

A investigação do padrão de segregação mapeamento de marcadores moleculares entre híbridos, pode identificar locos que apresentam ações como barreiras de reprodutividade, mesmo que eles não contribuam para a clara diferenciação fenotípica e taxa parental. Por exemplo, marcadores que exibem DS em populações híbridas poderiam estar ligados a genes que causam letalidade híbrida ou esterilidade, ou competição gametofítica e ao mesmo tempo não serem importantes para diferenciação fenotípica e taxa parental (Harushima et al., 2001).

## 2.7. *Distorção de segregação em Cruzamentos Divergentes*

Tem sido previamente sugerido que DS poderia ser acrescida com o aumento do nível de divergência entre pais (Zamir & Tadmor, 1986). Um menor nível de distorção em cruzamentos interespecíficos é freqüentemente correlacionado com a proximidade genética entre as plantas.

O grau de transmissão da razão de distorção é possível de ser positivamente correlacionado com nível de divergência (Palopoli & Wu, 1996; Jenczewski et al., 1997; Whitkus, 1998; Harushima et al., 2001; Taylor & Ingvarsson, 2003). Estudos empíricos têm demonstrado evidência com marcadores pouco distorcidos em cruzamentos intraespecífico, em relação a cruzamentos interespecíficos em plantas (Zamir & Tadmor, 1986; Causse et al., 1994; Jenczewski et al., 1997), sugerindo correlação positiva entre grau de transmissão da razão de distorção e o nível de divergência genômica.

## 2.8. Tipos de Marcadores

A distorção de segregação pode ser detectada em quase todo tipo de marcadores genéticos, incluindo marcadores morfológicos, izoenzimáticos e de DNA (Foltz, 1986; Zamir & Tadmor, 1986; Abe & Tsuda, 1987; Wendel et al., 1987; Konishi et al. 1990) em Xu et al. (1997). Porém, tem sido verificado que ao se utilizar marcadores AFLP para o estabelecimento de mapa de ligação, a segregação de marcadores com distorção é freqüentemente mais elevada. Segundo Ky et al. (2000), marcadores AFLP exibem co-migração e superimposição de produtos de amplificação não-alélica, possibilitando a ocorrência de pseudodistorção da razão (3:1). A situação simétrica entre as razões 1:3 e 3:1 constituem o principal argumento em favor da verdadeira distorção de marcadores. De acordo com Nikaido et al. (1999), a utilização de AFLP pode ser a maior causa da detecção de DS. Os autores consideram ser provável que o uso de mais nucleotídeos seletivos poderiam reduzir a DS. Mas é impraticável estender o número de nucleotídeos seletivos indefinidamente.

Na construção de um mapa de ligação de *Cryptomeria japonica*, 50% dos marcadores apresentaram desvios na segregação esperada e por isto não foram utilizados nas análises de ligação. Os autores atribuíram esse resultado à sobreposição de fragmentos de AFLP, de tamanho bastante semelhantes, o que provavelmente dificultou a sua correta identificação (Nikaido et al., 1999).

Lu et al. (1998) relatam que 15% de marcadores AFLP desviaram da razão de segregação esperada (3:1), em família de pêsego. Paillard et al. (1996) verificaram que 12% dos marcadores AFLP em um mapa de ligação

de população de duplo-haplóide, em *Coffea canephora*, apresentaram DS. Em soja, ocorreu em 25% de marcadores AFLP (Prabhu & Gresshoff, 1994).

Nikaido et al. (1999), trabalhando com sugi (*Cryptomeria japonica*) e utilizando AFLP no estabelecimento de mapa de ligação, verificaram que a DS foi elevada, atingindo taxas acima de 40%. A proporção de DS que tem sido geralmente revelada em mapeamento com AFLP em outras espécies tem sido menores que este valor (Eujayl et al., 1998; Menendez et al., 1997; Marques et al., 1998; Travis et al., 1998; Lu et al., 1998).

### 3. MATERIAL E MÉTODOS

Neste trabalho será investigado o uso de marcadores moleculares com segregação distorcida em mapeamentos genéticos, com uma abordagem teórica sobre detecção, taxa de segregação, funções de verossimilhança para fins de estimação de porcentagem de recombinação, ordenamento e mapeamento. Além dos aspectos teóricos serão considerados os aspectos aplicados, no qual será apresentada proposta de metodologia de mapeamento de marcadores, com e sem distorção na segregação, e sua utilização no mapeamento de uma população  $F_2$  a partir de dados simulados.

#### 3.1. *Fundamentação Teórica*

Serão considerados os vários aspectos pertinentes à análise de locos individuais, envolvendo a detecção de marcadores com segregação distorcida, teste estatístico, tamanhos de amostras para distinção entre segregações ambíguas. Também será abordado a estimação da porcentagem de recombinação entre locos, o efeito da distorção sobre marcadores em regiões cromossômicas vizinhas e seu mapeamento.

##### 3.1.1. *Testes de Segregação e Detecção de Segregação Distorcida*

A primeira etapa do mapeamento genético utilizando marcadores moleculares é a certificação de que eles segregam de acordo com o esperado, para a manipulação de mapeamento empregada. Para a



verificação da razão de segregação esperada dos locos individuais, procede-se uma comparação do número de indivíduos observados em cada classe com o esperado de acordo com o tipo de população estudada, por meio através do teste de qui-quadrado ( $\chi^2$ ). Este teste é bastante útil e eficiente, para este fim, pois além de levar em consideração os desvios ocorridos entre os valores esperados e observados, também é sensível ao tamanho da amostra (Cruz & CostaeSilva, 2006).

A estatística qui-quadrado é dada por:

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

em que  $\chi^2$  é valor de qui-quadrado calculado; e

$\text{Obs}_i$  e  $\text{Esp}_i$ , são os valores observado e esperado, para a i-ésima classe fenotípica ( $i= 1, 2, \dots, n$ ), respectivamente.

Para comprovação da segregação esperada geralmente adota-se um determinado nível de significância para teste individual e conjunto e conclui-se que o marcador segrega como esperado se o nível de significância estimado superar o mínimo estabelecido, geralmente 1% ou 5% de probabilidade. Assim, têm-se as segregações esperadas para os vários tipos de populações normalmente empregadas em estudos de mapeamento genético:

População	Hipótese de Segregação ( $H_0$ )
Duplo-Haplóide	$H_0:1:1$
Retrocruzamento	$H_0:1:1$
RIL	$H_0:1:1$
$F_2$ Dominante	$H_0:3:1$
$F_2$ Codominante	$H_0:1:2:1$

O teste de qui-quadrado geralmente tem sido utilizado como estatística conclusiva a respeito da ocorrência de segregação esperada mendeliana ou, em caso de sua rejeição, conclusivo a respeito da evidência de ocorrência de distorção de segregação. Entretanto alguns questionamentos podem ser formulados e serão considerados neste trabalho. O teste é eficiente em rejeitar a segregação mendeliana nos casos

em que há distorção de segregação, em populações de tamanho finito? O teste é eficiente em distinguir entre duas razões de segregação em populações de tamanho finito?

### *3.1.2. Tamanho de Populações para Detecção de Segregação Ambígua*

Uma pressuposição comum em alguns estudos de mapeamento com marcadores com distorção de segregação é admitir a existência de uma segregação alternativa, dentro de uma expectativa fundamentada nos princípios da genética. Assim, pode-se admitir, por exemplo, que em uma população  $F_2$  a segregação ocorre numa razão 2:1 ao invés de 3:1, em razão de possível letalidade genotípica. Esta situação é específica, pois considera-se que tanto a segregação mendeliana esperada (3:1), quanto a distorcida (2:1), sejam conhecidas. Entretanto, questiona-se a eficácia da estatística qui-quadrado em distinguir uma da outra razão de segregação.

Certamente que o poder de discriminação do teste é diretamente proporcional ao tamanho da amostra considerada. Assim, a detecção do número mínimo de indivíduos necessário para a distinção de duas razões de segregações, com determinado nível de significância, em determinada população, e a demonstração de segregação ambígua deve ser considerada. Neste estudo foram consideradas varias razões de segregação ocorrendo em populações  $F_2$ , com marcadores dominantes, e retrocruzamento com diferentes tamanhos populacionais.

Todas as análises de segregação e obtenção de estimativas de tamanhos ótimos para distinção entre segregações foram efetuadas por meio do programa estatístico GQMOL/UFV (Cruz, 2006).

### *3.1.3. Taxa de Segregação Distorcida*

Em muitas situações a distorção de segregação ocorre em razão da probabilidade diferencial de transmissão de gametas ou em razão de viabilidade diferencial dos diferentes genótipos. Assim, neste trabalho considerou-se uma situação mais geral de distorção de segregação, originária em razão de valores seletivos diferentes de gametas (s) ou de genótipos (g). Expressões para estimação destes valores foram consideradas neste trabalho. Dessa forma, admite-se ser dispensável um

teste qualitativo da taxa de segregação, mendeliana ou distorcida, em substituição a uma medida deste grau de distorção. Quando os valores de  $s$  (para seleção gamética) ou de  $g$  (para seleção genotípica) forem nulos admite-se que a segregação ocorre como esperado pelos princípios mendelianos. Valores diferentes de zero implicam em distorções em taxas diversas.

#### *3.1.4. Porcentagem de Recombinação*

Foram apresentados os valores esperados para as classes genotípicas estabelecidas por dois locos gênicos nas populações de retrocruzamento, RIL, duplo-haplóide,  $F_2$  dominante e  $F_2$  codominante em diferentes tipos de populações. A partir das frequências esperadas, sob hipótese de independência e ligação gênica, foram estabelecidas funções de verossimilhança que possibilitaram estimar a porcentagem de recombinação entre os marcadores, na presença de distorção de segregação. Neste estudo foram consideradas as populações de retrocruzamento, RIL, duplo-haplóide,  $F_2$  dominante e  $F_2$  codominante.

### *3.2. Mapeamento Incluindo Marcadores com Segregação de Distorção*

#### *3.2.1. Estratégia de Mapeamento*

O processo de mapeamento consiste em, inicialmente, estimar as porcentagens de recombinação entre pares de marcadores moleculares, geralmente usando princípios de verossimilhança. Neste contexto, as frequências esperadas das diferentes classes genotípicas, incluindo marcadores com DS, foram de grande utilidade nesta etapa. Posteriormente, as marcas moleculares são agrupadas considerando valores referenciais de LOD mínimo e porcentagem de recombinação máxima para que duas marcas sejam consideradas ligadas.

Neste trabalho foram consideradas duas táticas, perfazendo três estratégias, para obtenção da porcentagem de recombinação entre duas marcas moleculares. A primeira tática desconsidera a DS, perfazendo apenas uma estratégia nesta tática, a E1. A segunda tática considera a DS e possui duas estratégias. A primeira, estratégia E2, pressupõe conhecer

quais são as marcas que apresentam distorção de segregação e também a taxa esperada, de forma que o valor de  $s$  pode ser assumido como paramétrico. Uma matriz de distâncias entre pares de marcadores é confeccionada e utilizada no mapeamento. Na segunda estratégia, a E3, considera-se o valor estimado de distorção observado em todos os marcadores estudados, e este valor é utilizado para o cálculo do valor da distância entre os marcadores  $i$  e  $i'$ . Como para o marcador  $i$  a taxa de distorção será dada por  $\hat{s}_i$  e o marcador  $i'$  apresenta taxa de distorção  $\hat{s}_{i'}$ , são obtidas duas estimativas da medida da porcentagem de recombinação entre  $i$  e  $i'$ , sendo o valor da distância a estimativa média.

### 3.2.2. *Simulação de Populações com Distorção de Segregação*

Foram gerados dados de populações  $F_2$ , considerando o estudo de marcadores dominantes, admitindo-se diferentes tamanhos de populações e níveis de saturação do genoma.

Para a geração dos dados das populações de  $F_2$  dominante foi utilizado o módulo de simulação do programa computacional GQMOL/UFV (CRUZ, 2006), o qual permite gerar informações sobre genomas, genótipos de genitores, indivíduos de diferentes tipos de populações e a presença de marcadores que exibem distorção de segregação. Detalhes sobre o processo de simulação são descritos a seguir.

#### 3.2.2.1. *Simulação do Genoma*

O comprimento total do genoma foi de 300 cM, genoma com nível de saturação de 10 cM entre marcas moleculares. O genoma foi composto por 3 grupos de ligação, 100 cM em cada grupo e, portanto, com 11 marcas moleculares por grupo de ligação perfazendo 33 marcas moleculares no total.

#### 3.2.2.2. *Simulação de Genitores*

Na análise de ligação gênica foi simulado apenas um tipo de arranjo de genes na geração  $F_1$ , qual seja, um genitor homocigoto dominante e o outro homocigoto recessivo. Situação esta que produziu uma geração  $F_1$  com todos os locos em aproximação.

### 3.2.2.3. *Tamanho de População*

Para cada uma das situações estudadas foram geradas populações com 200, 400 e 1000 indivíduos.

### 3.2.2.4. *Procedimento de Simulação dos Indivíduos da População*

Para cada indivíduo foram produzidos os dados genotípicos de marcador, de acordo com a informação do genoma. A estratégia de simulação básica foi caminhar ao longo dos cromossomos e ir realizando permutas nos intervalos entre marcas adjacentes, de acordo com as distâncias dos marcadores.

O processo de simulação seguiu os seguintes passos: i) a partir do genoma simulado foram construídos os genótipos parentais homocigotos e contrastantes para os marcadores, conforme descrito no item anterior, de tal forma que a geração  $F_1$  apresentava-se em aproximação para todos os pares de marcadores; ii) a partir do genótipo da geração  $F_1$  foram gerados os gametas para a formação dos indivíduos das populações de  $F_2$  dominante. A produção de gametas foi realizada simulando-se o pareamento dos homólogos e realizando-se permutas ao longo dos cromossomos, nas regiões delimitadas por dois marcadores adjacentes. A probabilidade de ocorrência de recombinação numa região entre marcadores adjacentes foi dada de acordo com a distância destes marcadores no genoma simulado, no caso 10 cM. Após a decisão de ocorrência ou não da recombinação nesta região, passou-se para a próxima região, delimitada pelo segundo marcador e o terceiro. O procedimento continuou até que todas as regiões entre marcadores adjacentes no cromossomo foram alcançadas. Para formação de cada indivíduo nas populações foram simulados 1000 (mil) gametas, sendo sorteado apenas um gameta para formação de cada indivíduo, sendo que, o genótipo do indivíduo foi obtido pela duplicação do gameta. Na formação dos gametas levou-se em consideração a existência de fator seletivo que proporcionasse valores desiguais de probabilidade de um indivíduo receber determinado gameta, de forma que estaria caracterizada a distorção de segregação na descendência. Admitiu-se que em cada grupo

de ligação havia um marcador sob seleção, e que estes ocupavam posição intermediária dentro do grupo, ou seja, foi estabelecido que os marcadores na quinta posição apresentavam distorção de segregação.

### 3.2.3. *Mapeamento*

Após a geração dos dados das populações  $F_2$ , com análise de marcadores dominante, seguiram-se as demais etapas do processo de mapeamento levando-se em consideração a distorção e a negligenciando, como descrito a seguir:

#### 3.2.3.1. *Análise de Segregação de Locos Individuais*

Foram aplicados testes de qui-quadrado para verificação da razão de segregação em cada marca, para todas as populações geradas, sendo que a proporção esperada para as populações simuladas era de 3:1. Este teste serviu para ratificar a distorção de segregação nos marcadores em que tal condição foi imposta e também para avaliar o efeito que este marcador exercia sobre os demais em regiões cromossômicas vizinhas.

#### 3.2.3.2. *Análise de Pares de Marcas – Estimação da Percentagem de Recombinação*

Após a aplicação dos testes de segregação, seguiu-se a etapa da estimação da percentagem de recombinação entre pares de marcas. Para esta estimação foi utilizado o método da máxima verossimilhança. Nesta etapa foram utilizadas todas as marcas, incluindo aquelas que não segregaram conforme o esperado para a população.

Para esta estimação pelo método de máxima verossimilhança foram utilizadas duas funções de verossimilhança. A primeira foi a normalmente conhecida para a população, a qual não leva em consideração a distorção de segregação presente (Liu, 1998; Schuster & Cruz, 2004). E a segunda uma função desenvolvida neste trabalho, que considera a presença de marcas com distorção de segregação.

Assim, foram obtidos para cada população simulada três mapas genéticos. Um ignorando a existência de distorção de segregação e outros

dois admitindo a existência de distorção, porém usando duas estratégias já citadas anteriormente.

#### 3.2.3.3. *Determinação dos Grupos de Ligação*

O passo seguinte à estimação das frequências de recombinação entre pares de marcas é transformação destas distâncias por alguma das funções de mapeamento e o estabelecimento dos grupos de ligação. Na formação do grupo 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, então o loco A está ligado ao loco C, independente da frequência de recombinação estimada entre A e C e, portanto, A, B e C pertencem ao mesmo grupo de ligação. Os critérios a serem utilizados no agrupamento são a frequência máxima de recombinação ( $r_{\max}$ ) e o LOD mínimo ( $\text{LOD}_{\min}$ ), para inferir que dois locos estão ligados. Neste trabalho foram utilizados os valores de 30% e 3, respectivamente, para  $r_{\max}$  e  $\text{LOD}_{\min}$ .

#### 3.2.3.4. *Ordenamento das Marcas no Grupo de Ligação*

Quando se tem um grupo de ligação com apenas duas marcas, apenas uma ordem é possível, uma vez que a orientação da ordem pode ser ignorada. Porém, para três marcadores são possíveis 3 ordens, e para  $n$  marcadores, são possíveis  $n!/2$  ordens. Portanto, após a formação dos grupos de ligação, é necessário à determinação da melhor ordem das marcas nos grupos. O método utilizado foi o da SARF (Sum of Adjacent Recombination Fractions).

Realizados todos os passos descritos até então, os grupos de ligação para as populações simuladas estarão formados.

#### 3.2.4. *Comparação de Genomas*

Para as comparações foram avaliadas: 1) número de grupos de ligação obtidos no mapeamento; 2) número de marcas obtidas por grupo de ligação no mapeamento; 3) tamanho de grupos de ligação; 4) distância média entre marcadores adjacentes; 5) variância das distâncias entre marcas adjacentes e; 6) estresse.

#### 3.2.4.1. Número de Grupos de Ligação Obtidos no Mapeamento

Para todos os genomas analisados foi feita uma contagem do número de grupos de ligação obtidos do mapeamento das populações simuladas.

#### 3.2.4.2. Número de Marcas Obtidas por GL no Mapeamento

Para todos os genomas analisados foi feita uma contagem do número de marcas em todos os grupos de ligação obtidos do mapeamento das populações simuladas

#### 3.2.4.3. Tamanho do Grupo de Ligação

Dado pelo somatório das distâncias entre marcas adjacentes no grupo de ligação analisado, como segue:

$$L = \sum_{k=1}^{m-1} d_k$$

em que L é o tamanho do grupo de ligação,  $d_k$  é a distância entre marcas adjacentes  $m_k$  e  $m_{k+1}$  no grupo de ligação analisado ( $k= 1, \dots, m-1$ ), e m é o número de marcadores no grupo de ligação analisado.

#### 3.2.4.4. Distância Média entre Marcadores Adjacentes

É a razão do tamanho do grupo de ligação pelo número de intervalos entre marcas adjacentes no grupo de ligação, como segue:

$$\bar{d} = \frac{L}{m-1}$$

#### 3.2.4.5. Variância das Distâncias entre Marcas Adjacentes

Esta medida foi útil uma vez que o genoma original apresentava marcas equidistantes tendo-se variância nula. É dada por:

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{m-1} (d_k - \bar{d})^2}{l-1}, \text{ l em que é o número de intervalos dado por } m-1.$$

#### 3.2.4.6. Estresse

O coeficiente de estresse (S) é utilizado como medida de adequação das distâncias estimadas em representar as verdadeiras distâncias



estabelecidas no genoma real. Foi estabelecido com o propósito similar ao apresentado Cruz e Carneiro (2003) em estudos de divergência genética. É dado por:

$$S = 100 \cdot \sqrt{\frac{\sum_{k=1}^{m-1} (d_{ok} - d_k)^2}{\sum_{k=1}^{m-1} d_{ok}^2}}$$

em que  $d_{ok}$  é a distância entre marcas adjacentes  $m_k$  e  $m_{k+1}$  no genoma original (real) e  $d_k$  é a distância entre marcas adjacentes  $m_k$  e  $m_{k+1}$  no grupo de ligação analisado ( $k= 1, \dots, m-1$ ).

Para melhor entendimento, o estresse pode ser dado em função do desvio médio das distâncias entre marcas adjacentes no grupo de ligação e aquelas no genoma original, como segue abaixo e representado na fórmula.

$$S = \left( \frac{d}{d_o} \right) \cdot 100$$

em que  $d$  é o desvio médio das distâncias por intervalo no grupo de ligação analisado, em relação ao genoma original com saturação de 10cM e  $d_o$  é a distância entre marcas adjacentes no genoma original (neste caso 10cM).

## 4. RESULTADO E DISCUSSÃO

### 4.1. Segregação Ambígua

#### 4.1.1. População $F_2$ com Marcadores Dominante

Uma das dificuldades encontradas em análises genômicas com finalidade de se estabelecer o mapeamento, envolvendo marcas com segregação distorcida, é a própria constatação da distorção e, obviamente, da razão de segregação distorcida. Esta dificuldade é observada, por exemplo, se for considerada uma população  $F_2$  em que se consideram marcas dominantes com segregação esperada 3:1. Caso ocorra distorção de segregação, numa razão 2:1, é possível que tal razão possa não ser percebida ou detectada nos testes estatísticos. Para tal, deve-se ter um tamanho mínimo da amostra para possibilitar a distinção de duas razões de segregações numa determinada população, com determinado nível de significância.

Assim, a Tabela 1 ilustra o número mínimo de indivíduos em certas situações em que há duas possíveis razões de segregação. Verifica-se, portanto que, para a distinção entre as razões de segregação 2:1 e 3:1 seriam necessários cerca de 454 indivíduos. Tal fato é de grande importância uma vez que em muitos estudos de mapeamento o tamanho da população tem sido consideravelmente reduzido. São exemplos, entre muitos na literatura, os trabalhos de:

- Torres et al. (1993), trabalhando com *Vicia faba* com 20 indivíduos;
- Helentjaris et al. (1986) trabalhando com soja com 50 indivíduos;

- Xu et al. (1994) trabalhando com *Shorgum bicolor* com 50 indivíduos;
- Bonierbale et al. (1998) trabalhando com *Solanum spp.* com 65 indivíduos;
- Nodari et al. (1993) trabalhando com *Phaseolus vulgaris* com 75 indivíduos;
- Vilanova et al. (2003) trabalhando com *Prunus armeniaca* L. com 76 indivíduos;
- Schwarz-Sommer et al. (2003), trabalhando com *Antirrhinum* com 92 indivíduos.

Tabela 1 – Número mínimo de indivíduos necessários para distinção de razões de segregação, considerando 5% de probabilidade

Segregação	Segregação Alternativa	Tamanho da População
3:1	1:1	53,52
3:1	2:1	454,65
3:1	4:1	1065,37
3:1	5:1	358,83
3:1	7:1	143,27

Com o propósito de ressaltar a importância do número de indivíduos de uma população na caracterização da distorção e evidenciar a ambigüidade de segregação, foi considerada uma população F<sub>2</sub> com marcas dominantes, em que a segregação esperada é de 3:1. Com um tamanho de população de 455 indivíduos poder-se-ia obter o resultado ilustrado na Tabela 2. Com nítida situação de ambigüidade de segregação.

Tabela 2 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em duas populações F<sub>2</sub> dominante com 455 indivíduos

	Obs	Esp 3:1	Esp 2:1
A-	323	341,25	303,33
aa	132	113,75	151,67
$\chi^2$		3,90	3,82
$\alpha$ (%)		4,817	5,005

Na situação apresentada, as duas hipóteses apresentam valor de qui-quadrado próximo, tornando difícil à tomada de decisão quanto à razão mais apropriada. Um único indivíduo classificado incorretamente seria capaz de alterar o valor de probabilidade em torno do nível crítico, neste exemplo referenciado como 5%, alterando a conclusão sobre a hipótese de segregação mais apropriada. Situações como esta, certamente podem ter ocorrido em muitos estudos mapeamento, passando despercebido, uma vez que tal assunto não tem merecido a devida importância.

Uma situação ainda mais drástica, a ser considerada em muitos estudos de mapeamento genético, é aquela em que há possibilidade de se cometer erro Tipo II, em que segregações falsas são admitidas como verdadeiras. Isto pode ocorrer em razão do baixo poder de detecção do teste estatístico quando aplicados em amostras de tamanho reduzido. Um exemplo bastante evidente, para maiores reflexões, é apresentado, considerando novamente uma população  $F_2$  com marcas dominantes (Tabela 3), em que uma marca segrega nitidamente na proporção 2:1, numa população representada por apenas 30 indivíduos.

Tabela 3 - Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em duas populações  $F_2$  dominante com 30 indivíduos

	Obs	Esp 3:1	Esp 2:1
A-	20	22,5	20
aa	10	7,5	10
$\chi^2$		1,11	0,00
$\alpha$ (%)		29,184	100,0

Na Tabela 3, pode ser visualizada uma situação bastante rotineira em estudos de mapeamento genético, em que o pesquisador avalia a segregação mendeliana adotando um nível de significância pré-estabelecido. Assim, pode ser constatado que, considerando um nível mínimo de significância igual a 5%, a hipótese de segregação 3:1 não seria rejeitada, apesar de ser, obviamente, falsa. Neste caso, o tamanho da amostra de 30 indivíduos foi insuficiente para promover níveis de probabilidade

significativos. Um exemplo prático é o trabalho de Eujayl et al. (1997) no qual estudaram população  $F_2$  de *Lens culinaris* Medik, em que não se rejeita a segregação de 3:1 devido ao número de indivíduos reduzido na análise por meio do teste de qui-quadrado para A- e aa, com 27 e 5 indivíduos, respectivamente e para 25 e 13 indivíduos. Ficando evidente que estes resultados podem ser devidos ao tamanho reduzido da amostra.

Na Tabela 4, fica evidenciado de forma mais geral o problema de não se rejeitar uma hipótese falsa considerando diferentes tamanhos de amostras variando de 30 a 120 indivíduos. Deve ser realçado que estes tamanhos muitas vezes estão dentro da faixa de muitos estudos realizados e descritos na literatura.

Tabela 4 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em quatro populações  $F_2$  dominante com evidente segregação observada de 2:1, porém com hipótese de segregação 3:1

	Obs	Obs	Obs	Obs
A-	20	40	60	80
aa	10	20	30	40
$\chi^2$ ( $H_0:3:1$ )	1,11	2,22	3,33	4,44
$\alpha$ (%)	29,18	13,60	6,79	3,50

Verifica-se, na Tabela 4, que para a situação considerada, que apenas com 120 indivíduos, com valores observados de 80 e 40 indivíduos para A- e aa, respectivamente, foi possível rejeitar a hipótese de segregação 3:1. O trabalho de Jenczewski et al. (1997) é um bom exemplo para observar na prática resultados próximos. Estes autores, trabalhando com *Medicago* (Leguminosa) rejeitam ou não a segregação de 3:1 por meio de pequena variação no número de indivíduos. Como ilustração, quando A- e aa ocorrem com 81 e 40 indivíduos respectivamente rejeita-se a segregação 3:1, porém quando A- e aa se apresentam, com respectivamente, 82 e 38 não se rejeita a segregação 3:1. Assim, verifica-se na prática que mesmo com 120 indivíduos uma pequena variação de indivíduos nas classes pode modificar

a conclusão da hipótese  $H_0$ , mesmo com forte evidencia de uma segregação, no caso de 2:1.

Esta situação de segregação 2:1 em populações  $F_2$  com marcas dominantes, já foi relatada na literatura em diversos trabalhos (Eujayl et al., 1997; Matsushita et al., 2003; Zamir & Tadmor, 1986; Lu et al., 2002; Jenczewski et al., 1997).

#### 4.1.2. *População de Retrocruzamento*

As situações anteriormente apresentadas relatam de fato as dificuldades na definição da própria distorção em população  $F_2$ , com marcadores dominante. Uma situação também a ser considerada, porém mais amena, é aquela encontrada em análises de populações derivadas de retrocruzamento em que a razão de segregação esperada é de 1:1. Para uma abordagem sobre o tema e oferecer dados para reflexão, admitiu-se possível encontrar segregação do tipo 2:1, 3:1 e/ou outras, entre marcas moleculares em confronto com a segregação mendeliana esperada de 1:1.

A Tabela 5 apresenta o número mínimo de indivíduos em certas situações em retrocruzamento para distinção em relação a uma hipótese de 1:1. Constata-se que o número necessário para distinção entre segregação ambígua é comparativamente menos elevado do que o verificado numa população  $F_2$ , demonstrando-se, que neste aspecto, uma população de retrocruzamento deve possibilitar identificar mais acertadamente a verdadeira razão de segregação

Tabela 5 – Número mínimo de indivíduos necessários para distinção de razões de segregação, considerando 5% de probabilidade

Segregação	Segregação Alternativa	Tamanho da População
1:1	2:1	130,44
1:1	3:1	53,48
1:1	4:1	34,56
1:1	5:1	26,32
1:1	7:1	18,83

Para melhor entendimento sobre a ambigüidade, foi considerado a análise de uma população de retrocruzamento com 130 indivíduos, proporcionado um particular resultado, descrito na Tabela 6. Evidencia-se também a ambigüidade de segregação, no qual as duas hipóteses apresentam valor de qui-quadrado próximo, tornando difícil a tomada de decisão quanto à razão mais apropriada. Um único indivíduo classificado incorretamente seria capaz de alterar o valor de probabilidade em torno do nível crítico, alterando a conclusão sobre a hipótese de segregação mais apropriada. Tal fato é de grande importância uma vez que também em retrocruzamentos usados em muitos estudos de mapeamento, o tamanho da população tem sido consideravelmente reduzido. São exemplos, entre muitos na literatura, os trabalhos de:

- Gebhardt et al. (1989), trabalhando com *Solanum tuberosum* com 67 indivíduos;
- Gebhardt et al. (1989), trabalhando com *Solanum spp.* com 100 indivíduos;
- Vallejos et a. (1992), trabalhando com *Phaseolus vulgaris* com 68 indivíduos;
- Echt et al. (1994), trabalhando com *Medicago sativa* com 86 indivíduos;
- Fjeellstrom and Parfitt (1994), trabalhando com *Juglans spp.* com 63 indivíduos;
- Ky et al. (2000), trabalhando com *Coffea sp.* com 62 indivíduos.

Tabela 6 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em duas razões de segregações numa população de retrocruzamentos, com 130 indivíduos

	Obs	Esp 1:1	Esp 2:1
Aa	76	65	86,7
aa	54	65	43,3
$\chi^2$		3,72	3,94
$\alpha$ (%)		5,34	4,72

Em população de retrocruzamento também é possível encontrar situações que levam o pesquisador a cometer erro Tipo II, em que segregações falsas são admitidas como verdadeiras, em consequência do baixo poder de detecção do teste estatístico, em razão do tamanho reduzido da amostra utilizado. Novamente, para fins de melhores fundamentações sobre a questão levantada, é apresentado, como ilustração, uma marca segregando nitidamente na proporção 2:1, numa população representada por apenas 30 indivíduos (Tabela 7). Verifica-se, pelo resultado encontrado ao realizar o teste de qui-quadrado, que o pesquisador seria levado a não rejeitar a hipótese de segregação 1:1, ao nível de 5% de probabilidade, apesar de, nitidamente, a segregação ocorrer numa taxa 2:1.

Tabela 7 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em relação a duas hipóteses de segregação em população de retrocruzamento, com 30 indivíduos

	Obs	Esp 1:1	Esp 2:1
Aa	20	15	20
aa	10	15	10
$\chi^2$		3,33	0,00
$\alpha$ (%)		6,79	100,0

Pode ser verificado, considerando um nível mínimo de significância igual a 5%, que a hipótese de segregação 1:1 não seria rejeitada, apesar de falsa. Tal situação pode já ter ocorrido em muitos estudos e passado despercebido. Neste caso, o tamanho da amostra de 30 indivíduos foi insuficiente para promover níveis de probabilidade significativos. Na Tabela 8, fica evidenciado, de forma mais geral, as situações, em população de retrocruzamento, em que o problema de não se rejeitar uma hipótese falsa se verifica para diferentes tamanhos de populações.

Na Tabela 8, que ilustra, de forma mais geral, o problema de não rejeitar uma hipótese falsa, também fica evidente que o problema de detecção de segregação distorcida, em retrocruzamento, é mais ameno que em população  $F_2$ , com marcas dominantes. Com 36 indivíduos, com valores



observados de 24 e 12 indivíduos para Aa e aa, respectivamente, já foi possível rejeitar a hipótese de segregação 1:1.

Esta situação de segregação 2:1 em populações de retrocruzamento, já foi relatada na literatura, como pode ser constatado nos trabalhos de Herrera, et al. (2002) e Cheng et al. (1998).

Tabela 8 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em quatro populações de retrocruzamento com evidente segregação observada de 2:1

	Obs	Obs	Obs	obs
Aa	20	22	24	26
aa	10	11	12	13
$\chi^2$	3,33	3,67	4,00	4,33
$\alpha$ (%)	6,79	5,55	4,55	3,74

#### 4.2. *Estimadores das Taxas de Distorção de Segregação*

A distorção gamética é aquela em que a sua ocorrência resulta numa proporção gamética diferente da esperada mendeliana. Porém as frequências esperadas das classes genotípicas são compatíveis com os fundamentos biológicos relacionados à segregação de genes, distribuição de cromátides, viabilidade gamética e zigótica, dentre outras. Neste caso, a estrutura da população estudada não é perdida, podendo ser predita se conhecida a taxa de distorção da segregação dos gametas. Deve ser ressaltado que a existência de um marcador com DS atuando antes da fertilização pode apenas alterar proporções gaméticas, e as proporções genotípicas serão alteradas apenas se, indiretamente as proporções gaméticas restringirem as combinações realizáveis das proporções genotípicas (Vogl & Xu, 2000).

A distorção genotípica é aquela que incide sobre o genótipo já estabelecido (após a fertilização), que desestrutura a população, em termos do estabelecimento da frequência genotípica do loco sobre distorção e dos demais a ele ligados. Em determinadas situações a viabilidade por seleção após a fertilização pode ser mais importante que aquela que se manifesta

com a seleção gamética, por exemplo, em acasalamentos consangüíneos em que se reduz a sobrevivência de homozigotos comparado com heterozigotos (Charlesworth & Charles-worth, 1987). A DS após a fertilização geralmente afetará e alterará diretamente as proporções genótípicas (Vogl & Xu, 2000).

Assim, principalmente em distorção gamética, existem muitos casos em que a razão de segregação pode não ser nítida em termos de valores inteiros. Ou seja, numa população que deveria segregar 3:1 há possibilidade de várias taxas de segregação e não apenas 1:1, 2:1, 4:1 etc. Desta forma, torna-se fundamental trabalhar não com a razão de segregação, mas com a taxa de distorção, que pode ser predita, a partir da concepção de um fenômeno biológico, ou estimada a partir dos dados experimentais.

Para cada tipo de população estudada tem-se uma taxa de distorção esperada relativa a determinado loco, e o seu conhecimento e entendimento é importante para o entendimento dos processos de distorção de segregação e estabelecimento de valores esperados de segregação conjunta. Tais conceitos serão abordados a seguir.

#### 4.2.1. *Taxa de Distorção Esperada*

##### 4.2.1.1. *Taxa de Distorção Gamética em População F<sub>2</sub>, com Marcadores Dominantes*

No estabelecimento da taxa de distorção gamética, espera-se que os valores de probabilidade de ocorrência de gametas A e a, de um heterozigoto, seja diferencial. Admite-se que o valor da taxa possa ser representado por  $s$ , e que um heterozigoto Aa seria capaz de produzir os seguintes gametas:

$$f(A) = \frac{1}{2} + s$$

$$f(a) = \frac{1}{2} - s$$

Com as freqüências dos gametas tem-se o seguinte quadro de estabelecimento de genótipos a partir de cruzamento entre heterozigotos:

Gametas	A ( $\frac{1}{2} + s$ )	a ( $\frac{1}{2} - s$ )
A ( $\frac{1}{2} + s$ )	AA ( $\frac{1}{2} + s$ ) ( $\frac{1}{2} + s$ )	Aa ( $\frac{1}{2} + s$ ) ( $\frac{1}{2} - s$ )
a ( $\frac{1}{2} - s$ )	Aa ( $\frac{1}{2} + s$ ) ( $\frac{1}{2} - s$ )	aa ( $\frac{1}{2} - s$ ) ( $\frac{1}{2} - s$ )

Assim a freqüência fenotípica esperada na descendência será:

$$f(A-) = [(\frac{1}{2} + s)]^2 + 2(\frac{1}{4} - s^2)$$

e

$$f(aa) = (\frac{1}{2} - s)^2$$

Portanto, numa população  $F_2$ , derivada do cruzamento entre heterozigoto espera-se a constituição genotípica (ou fenotípica) dada a seguir:

Classes	Esperado (s=0)	Esperado (s 0)
A-	$\frac{3}{4}$	$\frac{3}{4} + s - s^2$
aa	$\frac{1}{4}$	$\frac{1}{4} - s + s^2$

Se há razões biológicas para se acreditar numa razão de segregação distorcida, por exemplo, 2:1, ou genericamente  $a_1: a_2$  pode-se assumir, genericamente, que:

$$f(A-) = \frac{3}{4} + s - s^2 = \frac{a_1}{a_1 + a_2}$$

e

$$f(aa) = \frac{1}{4} - s + s^2 = \frac{a_2}{a_1 + a_2}$$

dessa forma, o valor da taxa de distorção gamética poderia ser estabelecida da seguinte maneira:

$$s = \frac{1}{2} - \sqrt{\frac{a_2}{a_1 + a_2}}$$

Assim, se for novamente tomado como ilustração uma razão de segregação de 2:1, o valor representativo da taxa de distorção gamética esperada, associada a esta razão, será igual a -0,077350269 (valor de s).

Na Tabela 9 encontram-se valores da taxa de distorção gamética associada a algumas razões de segregação, possíveis de serem encontradas em várias situações de análise genômica.

Tabela 9 – Valores de distorção gamética em diversas razões de segregação

Razão de segregação	Taxa de distorção gamética
1:1	-0,20710
2:1	-0,0735
3:1	0,0
4:1	0,05279
5:1	0,09175
7:1	0,14644
9:1	0,18377
10:1	0,19848

#### 4.2.1.2. Taxa de Distorção Genotípica em População $F_2$ , com Marcadores Dominante

Certamente que também se pode inferir que a distorção na razão de segregação constatada pode ser calculada admitindo que ela seja conseqüência de uma viabilidade diferencial dos genótipos que surgem após o processo de fertilização. Neste caso, deve-se estabelecer uma taxa de distorção genotípica que surge em razão da segregação distorcida ter ocorrido por morte zigótica. Neste estudo, será admitido que a taxa de distorção genotípica seja representada por  $g$ , de forma que seria esperado na população  $F_2$  as seguintes classes fenotípicas:

Classes	Esperado ( $g=0$ )	Esperado ( $g \neq 0$ )
A-	$\frac{3}{4}$	$\frac{3}{4} + g$
aa	$\frac{1}{4}$	$\frac{1}{4} - g$

Novamente, se há razões biológicas para se acreditar numa razão de segregação distorcida, por exemplo, de 2:1, ou genericamente  $a_1: a_2$  pode-se definir:

$$f(A-) = \frac{3}{4} + g = \frac{a_1}{a_1 + a_2}$$

e

$$f(aa) = \frac{1}{4} - g = \frac{a_2}{a_1 + a_2}$$

dessa forma, o valor da taxa de distorção genotípica poderia ser estabelecida por meio do seguinte estimador:

$$g = \frac{a_1}{a_1 + a_2} - \frac{3}{4} = \frac{1}{4} - \frac{a_2}{a_1 + a_2}$$

Assim, se for novamente considerada que a razão de segregação esperada seja 2:1 o valor da taxa de distorção genotípica estimado será de - 0,08333. Na Tabela 10 encontram-se valores da taxa de distorção genotípica associados a várias razões de segregação.

Tabela 10 – Valores de distorção genotípica associados a diversas razões de segregação

Razão de segregação	Taxa de distorção gamética
1:1	- 0,25
2:1	-0,08333
3:1	0,0
4:1	0,05
5:1	0,08333
7:1	0,125
9:1	0,15
10;1	0,15909

As duas taxas, gamética e genotípicas estão relacionadas por meio da seguinte expressão:

$$s = \frac{1 - \sqrt{1 - 4g}}{2}$$

#### 4.2.1.3. Taxa de Distorção Gamética em População Retrocruzamento

A taxa de segregação distorcida poderá ser estimada em diferentes populações, como se pretende evidenciar neste estudo. Assim, se agora for admitido que ocorreu uma segregação diferente do esperado pela lei mendeliana, a taxa, representada por  $s$ , pode ser estimada considerando, novamente, que um heterozigoto  $Aa$  produzirá os seguintes gametas:

$$f(A) = \frac{1}{2} + s$$

e

$$f(a) = \frac{1}{2} - s$$

E, o  $F_1$  quando retrocruzado com o genitor recessivo  $aa$  proporcionará as seguintes classes genótípicas na descendência:

Gametas	$a$ (1)
$A$ ( $\frac{1}{2} + s$ )	$Aa$ ( $\frac{1}{2} + s$ ) (1) = ( $\frac{1}{2} + s$ )
$a$ ( $\frac{1}{2} - s$ )	$aa$ ( $\frac{1}{2} - s$ ) (1) = ( $\frac{1}{2} - s$ )

Desta forma a população de retrocruzamento, com e sem distorção de segregação, é constituída pelos genótipos e as freqüências genótípicas especificadas a seguir:

Classes	Esperado ( $s=0$ )	Esperado ( $s \neq 0$ )
$Aa$	$\frac{1}{2}$	$\frac{1}{2} + s$
$aa$	$\frac{1}{2}$	$\frac{1}{2} - s$

Pode-se observar que, neste tipo de população, as freqüências genótípicas são dadas exatamente pela freqüências gaméticas. Assim, se há razões biológicas para se acreditar numa razão de segregação distorcida, por exemplo, de 2:1, ou genericamente  $a_1 : a_2$  pode-se estabelecer a igualdade:

$$f(Aa) = \frac{1}{2} + s = \frac{a_1}{a_1 + a_2}$$

e

$$f(aa) = \frac{1}{2} - s = \frac{a_2}{a_1 + a_2}$$

dessa forma, o valor da taxa de distorção gamética poderia ser estimada por meio do estimador:

$$s = \frac{1}{2} - \frac{a_2}{a_1 + a_2} \quad \text{ou} \quad s = \frac{a_1}{a_1 + a_2} - \frac{1}{2}$$

Assim, se for novamente considerada a razão de segregação de 2:1 o valor da taxa de distorção gamética esperada será de 0,16667. Na Tabela 11 encontram-se valores da taxa de distorção gamética, numa população de retrocruzamentos, associada a várias situações de razão de segregação com e sem distorção.

Tabela 11 – Valores de distorção gamética em diversas razões de segregação

Razão de segregação	Taxa de distorção gamética
1:3	-0,25
1:1	0,00
2:1	0,16667
3:1	0,25
4:1	0,30
5:1	0,33333
7:1	0,375
9:1	0,4
10:1	0,4090

#### 4.2.1.4. Taxa de Distorção Genotípica em População Retrocruzamento

Em relação à taxa de distorção genotípica, em que se admite que a razão de segregação distorcida tenha ocorrido em razão de morte zigótica, também representada por  $g$ , de forma que seria esperado, na população de retrocruzamento, as seguintes classes genotípicas:

Classes	Esperado (g=0)	Esperado (g ≠ 0)
Aa	½	½ + g
aa	½	½ - g

Novamente, e se há razões biológicas para se acreditar numa razão de segregação distorcida, genericamente dada por  $a_1 : a_2$ , então se tem:

$$f(A-) = \frac{1}{2} + g = \frac{a_1}{a_1 + a_2}$$

e

$$f(aa) = \frac{1}{2} - g = \frac{a_2}{a_1 + a_2}$$

Dessa forma, o valor da taxa de distorção genotípica poderia ser estabelecida da seguinte maneira:

$$g = \frac{1}{2} - \frac{a_2}{a_1 + a_2}$$

ou

$$g = \frac{a_1}{a_1 + a_2} - \frac{1}{2}$$

A fórmula obtida para obtenção de  $g$  é idêntica à obtida para o cálculo da taxa de distorção gamética  $s$ . Assim, se for novamente considerada que a razão de segregação esperada seja 2:1 o valor da taxa de distorção genotípica  $g$  esperada será também igual a 0,16667, como  $s$ . Assim, os valores da taxa de distorção genotípica em várias situações de razão de segregação será idêntica à taxa de distorção gamética, demonstrada na Tabela 11.

#### 4.2.2. Taxa de Distorção Observada

Nesta situação não há inferência sobre a causa da distorção, apenas uma constatação de que a segregação mendeliana não ocorre e, conseqüentemente, é possível por meio dos valores observados estimar a taxa de distorção.



#### 4.2.2.1. População $F_2$ , com Marcadores Dominante

Para a estimação da taxa de distorção será considerado uma população  $F_2$ , em estudo de mapeamento com marcadores dominantes. A taxa de distorção poderá ser estimada a partir dos próprios dados experimentais bastando considerar que as proporções teóricas  $a_1:a_2$  possam ser substituídas pelas proporções observadas  $o_1:o_2$  em que  $o_1$  representa a número observado de indivíduos A- e  $o_2$  o número observado de indivíduos aa. Assim, pode-se obter de:

$$s = \frac{1}{2} - \sqrt{\frac{a_2}{a_1 + a_2}}$$

com a substituição apropriada, a expressão:

$$\hat{s} = \frac{1}{2} - \sqrt{\frac{o_2}{o_1 + o_2}}$$

para distorção de segregação gamética

e de:

$$g = \frac{a_1}{a_1 + a_2} - \frac{3}{4} \quad \text{ou} \quad \frac{1}{4} - \frac{a_2}{a_1 + a_2}$$

com a substituição apropriada, a expressão:

$$\hat{g} = \frac{o_1}{o_1 + o_2} - \frac{3}{4} \quad \text{ou} \quad \frac{1}{4} - \frac{o_2}{o_1 + o_2}$$

para distorção de segregação genotípica.

Assim, supondo uma distorção de segregação, gamética ou genotípica, de 2:1 tem-se os valores estimados:

- distorção gamética

$$\hat{s} = \frac{1}{2} - \sqrt{\frac{o_2}{o_1 + o_2}} = \frac{1}{2} - \sqrt{\frac{1}{2+1}} = -0,077350269$$

- distorção genotípica

$$\hat{g} = \frac{1}{4} - \frac{o_2}{o_1 + o_2} = \frac{1}{4} - \frac{1}{2+1} = -0,08333$$

Apesar de evidente, é importante lembrar o exemplo dado anteriormente para  $F_2$  com marcas dominantes (Tabela 4), em que o teste de qui-quadrado não rejeitava, em algumas situações, a hipótese de

segregação 3:1, mas a taxa de distorção gamética é, de fato, própria da segregação 2:1 (Tabela 12).

Tabela 12 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em quatro populações  $F_2$  dominante com evidente segregação observada de 2:1

	Obs	Obs	Obs	Obs
A-	20	40	60	80
aa	10	20	30	40
$\chi^2$	1,11	2,22	3,33	4,44
$\alpha$ (%)	29,18	13,60	6,79	3,50
s(ou $\hat{s}$ )	-0,07735	-0,07735	-0,07735	-0,07735

#### 4.2.2.2. População de Retrocruzamento

Em relação à população de retrocruzamento, a taxa de distorção também poderá ser estimada a partir dos próprios dados experimentais bastando considerar que as proporções teóricas  $a_1: a_2$  possam ser substituídas pelas proporções observadas  $o_1: o_2$ , em que  $o_1$  representa o número observada de indivíduos Aa e  $o_2$  o número observada de indivíduos aa. Assim, pode-se obter de:

$$s = g = \frac{1}{2} - \frac{a_2}{a_1 + a_2}$$

com a substituição apropriada, tem-se o estimador

$$\hat{s} = \hat{g} = \frac{1}{2} - \frac{o_2}{o_1 + o_2}$$

para as distorções de segregação gamética ( $\hat{s}$ ) e genotípica ( $\hat{g}$ ).

Supondo uma distorção de segregação gamética e genotípica de 2:1 tem-se:

$$\hat{s} = \hat{g} = \frac{1}{2} - \frac{o_2}{o_1 + o_2} = \frac{1}{2} - \frac{1}{2+1} = 0,16667$$

Na mesma situação apresentado para  $F_2$  dominante, também em retrocruzamento é importante lembrar o exemplo dado anteriormente (Tabela 8), em que o teste de qui-quadrado não rejeitava, em algumas

situações, a hipótese de segregação 1:1, mas com a taxa de distorção gamética sendo de fato, própria da segregação 2:1 (Tabela 13).

Tabela 13 – Valores de  $\chi^2$  e nível de significância ( $\alpha$ ) em quatro populações de retrocruzamento com evidente segregação observada de 2:1

	Obs	Obs	Obs	Obs
Aa	20	22	24	26
aa	10	11	12	13
$\chi^2$	3,33	3,67	4,00	4,33
$\alpha$ (%)	6,79	5,55	4,55	3,74
s(ou $\hat{s}$ )	0,1667	0,1667	0,1667	0,1667

#### 4.3. Segregação de Pares de Locos

Outro problema a ser considerado no mapeamento genético envolvendo locos com segregação distorcida é a estimação da porcentagem de recombinação entre dois genes (ou marcas moleculares) em que pelo menos uma delas apresenta distorção de segregação. O que pode ser devido, por exemplo, a distorção de segregação gamética ou genotípica.

##### 4.3.1. Distorção Gamética

##### 4.3.1.1. População $F_2$ , com Marcadores Dominantes em Fase de Aproximação

Considerando uma população  $F_2$  com duas marcas dominantes (A/a e B/b), em que a marca A/a apresenta segregação distorcida, por exemplo, segrega na proporção, 2:1, e a marca B/b segrega na proporção 3:1. Admitindo que estas marcas encontram-se em fase de aproximação no  $F_1$  duplo-heterozigoto, com a configuração AB//ab, e que a distorção é explicada por variação na probabilidade gamética do loco A/a, pode-se admitir os seguintes valores de probabilidade conjunta:

$$P(AB) = P(A)P(B/A) = \left(\frac{1}{2} + s\right)(1-r) = \frac{(1+2s)(1-r)}{2}$$

$$P(Ab) = P(A)P(b/A) = \left(\frac{1}{2} + s\right)r = \frac{(1+2s)r}{2}$$

$$P(aB) = P(a)P(B/a) = \left(\frac{1}{2} - s\right)r = \frac{(1-2s)r}{2}$$

$$P(ab) = P(a)P(b/a) = \left(\frac{1}{2} - s\right)(1-r) = \frac{(1-2s)(1-r)}{2}$$

Dessa forma, tem-se na Tabela 14 os valores comparativos de freqüências gaméticas esperadas com e sem distorção. Constatase que, no modelo apresentado, há um pressuposto interessante de que se o loco A/a apresenta distorção gamética, o loco B/b ligado a ele certamente apresentará distorção pela ação indireta da seleção sobre o loco em ligação fatorial.

Tabela 14 – Comparação de valores de freqüências gaméticas esperadas de um F<sub>1</sub> em aproximação, sem e com distorção de segregação gamética

Gameta	Tipo	Freqüência (s=0)	Freqüência (s 0)
AB	Paternal	$\frac{1-r}{2}$	p(1-r)
Ab	Recombinante	$\frac{r}{2}$	pr
aB	Recombinante	$\frac{r}{2}$	qr
ab	Paternal	$\frac{1-r}{2}$	q(1-r)

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$ , sendo  $p+q = 1$

Assim se a distorção gamética, direta e indireta, pode ser observada por meio das seguintes considerações:

- a segregação esperada para o loco A/a é dada por:

$$f(A) = p(1-r) + pr = p = \frac{1}{2} + s$$

e

$$f(a) = qr + q(1-r) = qr+q-qr = q = \frac{1}{2} - s$$

A segregação esperada para o loco B/b também será distorcida, não pelo fator seletivo que atua diretamente sobre este loco, mas única e exclusivamente pela ação indireta em razão da existência da ligação fatorial com o loco A/a sob distorção. Assim, tem-se:

$$f(B) = p(1-r) + qr = \frac{1}{2} + s - 2rs$$

e

$$f(b) = pr + q(1-r) = \frac{1}{2} - s + 2rs$$

Se o loco B/b está completamente ligado ao loco A/a, então  $r = 0$  e se verifica:

$$f(B) = f(A) = \frac{1}{2} + s$$

e

$$f(b) = f(a) = \frac{1}{2} - s$$

Se o loco B/b tem distribuição independente de A/a, então  $r = 1/2$  e se verifica:

$$f(B) = \frac{1}{2}$$

e

$$f(b) = \frac{1}{2}$$

Este pressuposto do modelo é coerente com fatos observados na literatura, uma vez que diversos trabalhos relatam que genes causadores de distorção manifestam distorção de segregação em marcas ligadas. Dentre eles pode-se citar os trabalhos de Zamir & Tadmor (1986), Pereira et al. (1993), Xu et al. (1997) e Vogl & Xu (2000). Entretanto, em alguns outros trabalhos há relatos que a distorção incide unicamente em um loco, não

afetando aqueles em regiões adjacentes. Assim, a não observação do pressuposto na teoria apresentada, poderia ocorrer se a DS tivesse causa em erros de genotipagem (Vogl & Xu, 2000) ou na duplicação da informação genética (Frisch et al., 2004), fatores que afetam apenas o marcador específico.

Assim, com as freqüências esperadas dos gametas pode-se estabelecer as freqüências esperadas das classes genóticas com a presença de distorção de segregação gamética conforme o Tabela 15. Estas freqüência serão indispensáveis no estabelecimento das funções de verossimilhança, possibilitando obter estimativas mais acuradas da porcentagem de recombinação entre as marcas estudadas.

Tabela 15 – Freqüências genóticas esperadas com e sem distorção gamética em população  $F_2$  dominante em aproximação

Genótipo	Freqüência esperada com distorção $s \neq 0$	Freqüência esperada sem distorção ( $s=0$ )
A-B-	$1 - q^2 - [pr + q(1-r)]^2 + q^2(1-r)^2$	$\frac{1}{2} + \frac{1}{4}(1-r)^2$
A-bb	$[pr + q(1-r)]^2 - q^2(1-r)^2$	$\frac{1}{4} - \frac{(1-r)^2}{4}$
aaB-	$q^2 - q^2(1-r)^2$	$\frac{1}{4} - \frac{(1-r)^2}{4}$
aabb	$q^2(1-r)^2$	$\frac{(1-r)^2}{4}$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$  sendo  $p+q = 1$

Apesar de ocorrer diferenças teóricas entre as freqüências fenotípicas quando se consideram duas marcas moleculares, em que uma apresenta distorção de segregação, não severa, constata-se que, na prática, o valor estimado da freqüência de recombinação é bem próximo quando esta distorção é negligenciada. Neste contexto, será considerado como ilustração duas populações  $F_2$  com 100 e 10000 indivíduos, em que o loco A/a apresenta segregação 2:1 e o loco B/b segregação distorcida em razão da

ligação fatorial, e a porcentagem de recombinação é de 10 centimorgans (cM), conforme Tabela 16.

Para as populações, o cálculo do valor da porcentagem de recombinação considerando a DS, deve ser estimado a partir da função de verossimilhança:

$$L(r; n_i) = \lambda [p_1]^{n_1} [p_2]^{n_2} [p_3]^{n_3} [p_4]^{n_4}$$

em que:

$$p_1 = 1 - q^2 + [pr + q(1-r)]^2 + q^2(1-r)^2$$

$$p_2 = [pr + q(1-r)]^2 - q^2(1-r)^2$$

$$p_3 = q^2 - q^2(1-r)^2$$

$$p_4 = q^2(1-r)^2$$

pode-se adotar as medidas auxiliares:

$$p = (1+2s) / 2$$

$$q = (1-2s) / 2$$

$$p+q=1$$

e os totais observados:

$$n_1 = n_{A-B-}; n_2 = n_{A-bb}; n_3 = n_{aaB-}; n_4 = n_{aabb};$$

$$\text{definindo, ainda, } \lambda = \frac{N!}{n_1!n_2!n_3!n_4!} \text{ sendo } N = \sum_{i=1}^4 n_i$$

Tabela 16 – Valores de classes fenotípicas com segregação de 2:1 em duas populações F<sub>2</sub> dominante, com 100 e 10000 indivíduos

Classes	Observado	Observado
A-B-	62	6210
A-bb	5	457
aaB-	6	633
aabb	27	2700
Total	100	10000
A-	67	6667
aa	33	3333
B-	68	6842
bb	32	3157

O LOD score (Likelihood of Odds), é logaritmo da razão entre a probabilidade de dois locos estarem ligados e a probabilidade de não estarem ligados. O entendimento destas questões é fundamental, pois elas determinam a confiabilidade do mapa e sua aplicação nos programas de melhoramento. Assim, além da freqüência de recombinação, o LOD score é utilizado também como critério para inferir sobre a presença ou ausência de ligação. Este critério é um teste de razão de verossimilhança que utiliza o logaritmo de base 10 ( $\log_{10}$ ) da razão entre a hipótese de ligação com um determinado valor de freqüência de ligação diferente de 0,5 e a hipótese de ausência de ligação, ou seja, freqüência de recombinação igual a 0,5. A expressão de LOD score é apresentada a seguir:

$$\text{LOD} = \log_{10} \left[ \frac{L(r/n_i)}{L(r = 0,5/n_i)} \right]$$

Para as populações de  $F_2$  com marcadores dominantes a expressão de cálculo do valor de LOD score para qualquer estimativa de freqüência de recombinação ( $r$ ) e tamanho de população é dada por:

$$\text{LOD} = \log_{10} \left\{ \frac{\lambda \left[ \frac{1}{2} + \frac{1}{4}(1-r)^2 \right]^{n_1} \left[ \frac{1}{4} - \frac{(1-r)^2}{4} \right]^{n_2} \left[ \frac{1}{4} - \frac{(1-r)^2}{4} \right]^{n_3} \left[ \frac{(1-r)^2}{4} \right]^{n_4}}{\lambda \left( \frac{1}{4} \right)^N} \right\}$$

O LOD score é considerado um teste de significância usado para testar a hipótese de ligação de dois locos gênicos. O resultado deste teste pode ser interpretado da seguinte forma: quando  $L(r/n_i) > L(r = 0,5/n_i)$ , o valor de LOD score é positivo. Conclui-se que os locos estão ligados geralmente quando o valor de LOD score é maior do que 3. O valor de LOD score igual a 3 implica que há uma probabilidade 1.000 vezes maior da presença de ligação do que da ausência de ligação entre os locos gênicos estudados.

Assim para a população com 100 indivíduos o valor da porcentagem de recombinação e do LOD seria dado por

$$r = 10 \text{ cM} \quad \text{e} \quad \text{LOD} = 13,28$$

Para a população de 10000 indivíduos o valor da porcentagem de recombinação e do LOD seriam:



$$r = 10,0 \text{ cMorgans e LOD} = 1328,46$$

Entretanto, muitas vezes, por desconhecimento ou falta de recursos computacionais, certos pesquisadores tem negligenciado a existência de distorção de segregação e estimado a freqüência de recombinação a partir da função de verossimilhança dada por;

$$L(r; n_i) = \lambda \left[ \frac{1}{2} + \frac{1}{4}(1-r)^2 \right]^{n_1} \left[ \frac{1}{4} - \frac{1}{4}(1-r)^2 \right]^{n_2+n_3} \left[ \frac{1}{4}(1-r)^2 \right]^{n_4}$$

em que:

$r$  = freqüência de recombinação.

Apesar de o cálculo ser errôneo o valor obtido é próximo daquele preconizado como ideal. Assim, para a população de 100 indivíduos é estimado:

$$r = 10,3 \text{ cM e LOD} = 13,28$$

e, para a população de 10000 indivíduos é estimado:

$$r = 10,3 \text{ cM e LOD} = 1328,05$$

Assim, valores muito próximos são encontrados mesmo utilizando um estimador inapropriado indicando que a manutenção de mapas distorcidas em certos estudos genômicos pode ser tolerada principalmente quando a proporção de distorção não é tão discrepante da razão de segregação esperada na população. Mas se for considerado a existência de distorção de segregação mais drástica, por exemplo, de 7:1, os valores poderiam ser mais discrepantes. Para maiores reflexões pode ser considerado os dados da Tabela 17. Para a população com 100 indivíduos o valor da porcentagem de recombinação e do LOD considerando a DS na função de verossimilhança seriam:

$$r = 10,2 \text{ cM e LOD} = 8,9479$$

e, para a população de 10000 indivíduos é estimado:

$$r = 10 \text{ cM e LOD} = 910, 52495$$

E a obtenção da estimativa da porcentagem de recombinação e do LOD negligenciando a DS na função de verossimilhança, para a população com 100 indivíduos seriam:

$$r = 8,9 \text{ cM e LOD} = 8,96621$$

e, para a população de 10000 indivíduos é estimado:

$r = 8,7 \text{ cM}$  e  $\text{LOD} = 908,99187$

Tabela 17 – Valores de classes fenotípicas com segregação de 7:1 em duas populações  $F_2$  dominante, com 100 e 10000 indivíduos

Classes	Observado	Observado
A-B-	83	6210
A-bb	5	457
aaB-	2	633
aabb	10	2700
Total	100	10000
A-	88	6667
aa	12	3333
B-	85	6843
bb	15	3157

Analisando os resultados verifica-se que a desconsideração da DS mais pronunciada, nas funções de verossimilhança pode resultar em um viés considerável na distância entre marcadores, mesmo com 10000 indivíduos, que é um tamanho de população consideravelmente elevado. Deve ficar registrado que este nível de distorção em torno de 7:1 em população  $F_2$  dominante, não é raro. Na literatura há relatos de valores superiores a este nível de distorção, de segregação (Zamir & Tadmor, 1986; Eujayl et al., 1997; Jenczewski et al., 1997; Zhang et al., 2006).

A Tabela 18 apresenta a estimativa de frequência de recombinação em uma população de 10000 indivíduos para diferentes níveis de DS e valores estimados da porcentagem de recombinação admitindo ou não a taxa de distorção. O viés obtido nos valores foi de 0,3 a 2,0 cMorgans.

Na Tabela 19 são apresentadas as estimativas de frequência de recombinação em uma população de 100 indivíduos, para diferentes níveis de DS. O tamanho da população afeta o valor de LOD, podendo influenciar o agrupamento e ordenamento de marcas em grupos de ligação, uma vez que a utilização de um valor mínimo de LOD, como critério de definição da

ligação fatorial, é utilizado em todos os estudos de estabelecimento de mapas de ligação.

Tabela 18 – Estimativas de frequência de recombinação (FR) e LODs em uma população F<sub>2</sub> dominante com 10000 indivíduos, com diferentes DS, considerando a porcentagem de recombinação verdadeira de 10 cMorgans

Nível de DS	FR		FR	
	sem considerar s	LOD	considerando s	LOD
1:1	10,5	1720,43	10	1722,63
2:1	10,3	1328,05	10	1328,46
3:1	10	1146,63	10	1146,63
4:1	9,6	1043,73	10	1043,95
5:1	9,3	981,04	10	981,67
7:1	8,7	908,99	10	910,52
9;1	8,2	873,17	10	875,51
10:1	8,0	861,80	10	864,38

Tabela 19 – Estimativas de frequência de recombinação (FR) e LODs em uma população F<sub>2</sub> dominante com 100 indivíduos, com diferentes DS

Nível de DS	FR		FR	
	sem considerar s	LOD	considerando s	LOD
1:1	10,5	17,20	10	17,22
2:1	10,3	13,28	10	13,28
3:1	10	11,47	10	11,47
4:1	9,6	10,44	10	10,44
5:1	9,3	9,81	10	9,81
7:1	8,7	9,09	10	9,10
9;1	8,2	8,73	10	8,76
10:1	8,0	8,62	10	8,64

Os valores de frequência de recombinação observados nas duas tabelas apresentadas anteriormente evidenciam o viés mais pronunciado à

medida que o nível de segregação se distancia da freqüência esperada, demonstrando que a negligenciação da DS em nível mais severo proporciona um resultado prejudicial na estimativa da distância entre marcas ou do LOD, independente do tamanho da população.

#### 4.3.1.2. População de Retrocruzamento, $F_1$ em Aproximação

Considere uma população de retrocruzamento e duas marcas dominantes (A/a e B/b), em que a marca A/a apresenta segregação distorcida e a marca B/b segrega na proporção 1:1. Admitindo que estas marcas encontram-se em fase de aproximação no  $F_1$  duplo-heterozigoto, com a configuração AB//ab, e que a distorção é explicada por variação na probabilidade gamética do loco A/a, deve-se tomar os mesmos valores de probabilidade conjunta gamética, apresentados anteriormente. Desta forma, tem-se na Tabela 20, os valores de freqüências gaméticas esperadas com e sem distorção na  $F_1$  em aproximação.

Tabela 20 – Comparação de valores de freqüências gaméticas esperadas sem e com distorção

Gameta	Tipo	Freqüência (s=0)	Freqüência (s 0)
AB	Paternal	$\frac{1-r}{2}$	$p(1-r)$
Ab	Recombinante	$\frac{r}{2}$	$pr$
aB	Recombinante	$\frac{r}{2}$	$qr$
ab	Paternal	$\frac{1-r}{2}$	$q(1-r)$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$ , sendo  $p+q = 1$

Em retrocruzamento, da mesma forma que demonstrado para  $F_2$  com marcadores dominantes, se o loco A/a apresenta distorção gamética, o loco B/b ligado a ele certamente apresentará distorção pela ação indireta da seleção sobre o loco em ligação fatorial.

Assim, com a obtenção das freqüências esperadas dos gametas da  $F_1$  com distorção de segregação e da esperada do pai recorrente (recessivo) que é unitária, não influenciando, foi obtida a freqüência esperada de cada classe genotípica com a presença de distorção de segregação gamética Tabela 21.

Tabela 21 – Freqüências genotípicas esperadas com e sem distorção gamética em população de retrocruzamento,  $F_1$  em aproximação

Genótipo	Freqüência esperada com distorção ( $s \neq 0$ )	Freqüência esperada sem distorção ( $s=0$ )
AaBb	$p(1-r)$	$\frac{1}{2}(1-r)$
Aabb	$pr$	$\frac{1}{2}r$
aaBb	$qr$	$\frac{1}{2}r$
aabb	$q(1-r)$	$\frac{1}{2}(1-r)$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$  sendo  $p+q = 1$

Com a dedução das freqüências esperadas, considerando a DS, o cálculo da porcentagem de recombinação é obtido a partir da seguinte função de verossimilhança:

$$L(r; n_i) = \lambda [p(1-r)]^{n_1} [pr]^{n_2} [qr]^{n_3} [q(1-r)]^{n_4}$$

em que:

$$p = (1+2s) / 2$$

$$q = (1-2s) / 2$$

$$p+q = 1$$

sendo

$$n_1 = n_{AaBb}; n_2 = n_{Aabb}; n_3 = n_{aaBb}; n_4 = n_{aabb}$$

$$\text{tendo } \lambda = \frac{N!}{n_1!n_2!n_3!n_4!}; e$$

$$\text{sendo } N = \sum_{i=1}^4 n_i$$

Apesar de existir valores diferenciados das freqüências genotípica na população de retrocruzamento, o valor da estimativa de verossimilhança da porcentagem de recombinação, considerando ou não a taxa de distorção, será exatamente o mesmo. Evidencia-se que em população de retrocruzamento a distância entre duas marcas não é alterada se a distorção de segregação for gamética. As duas funções de verossimilhança, a que considera DS (deduzida) e a que não considerando a DS (tradicional), fornecerão estimativas de distância entre as marcas não viesadas.

#### 4.3.1.3. População $F_2$ , Marcadores Codominantes, Derivada de $F_1$ em Aproximação

Considerando uma população  $F_2$  codominante descendente do cruzamento entre dois pais heterozigotos com duas marcas codominantes (A/a e B/b) e admitindo que estas marcas encontram-se em fase de aproximação no  $F_1$  duplo-heterozigoto, com a configuração AB//ab e que a distorção é explicada por variação na probabilidade gamética do loco A/a, pode-se admitir os seguintes valores de probabilidade conjunta, já apresentados anteriormente. Desta forma, tem-se na Tabela 22, os valores de freqüências gaméticas esperadas com e sem distorção na  $F_1$  em aproximação.

Tabela 22 – Comparação de valores de freqüências gaméticas esperadas sem e com distorção

Gameta	Tipo	Freqüência (s=0)	Freqüência (s 0)
AB	Paternal	$\frac{1-r}{2}$	$p(1-r)$
Ab	Recombinante	$\frac{r}{2}$	$pr$
aB	Recombinante	$\frac{r}{2}$	$qr$
ab	Paternal	$\frac{1-r}{2}$	$q(1-r)$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$ , sendo  $p+q = 1$

Neste tipo de população também ocorre o fato de que se o loco A/a apresenta distorção gamética, o loco B/b ligado a ele certamente apresentará distorção pela ação indireta da seleção sobre o loco em ligação fatorial. Com o estabelecimento das freqüências esperadas dos gametas, considerando a existência de distorção de segregação em um dos locos, foi possível a obtenção da freqüência esperada de cada classe genotípica com a presença de distorção de segregação gamética conforme apresentado na Tabela 23.

Tabela 23 – Freqüências genotípicas esperadas com e sem distorção gamética em população F<sub>2</sub>, marcadores codominantes, obtida de F<sub>1</sub> em aproximação

Genótipo	Freqüência esperada com distorção (s ≠ 0)	Freqüência esperada sem distorção (s=0)
(1) AABB	$p^2(1-r)^2$	$\frac{1}{4}(1-r)^2$
(2) AABb	$2p^2r(1-r)$	$\frac{1}{2}r(1-r)$
(3) AAbb	$p^2r^2$	$\frac{1}{4}r^2$
(4) AaBB	$2pqr(1-r)$	$\frac{1}{2}r(1-r)$
(5) AaBb	$2pq(1-r)^2 + 2pqr^2$	$\frac{1}{2}(1-r)^2 + \frac{1}{2}r^2$
(6) Aabb	$2pqr(1-r)$	$\frac{1}{2}r(1-r)$
(7) aaBB	$q^2r^2$	$\frac{1}{4}r^2$
(8) aaBb	$2q^2r(1-r)$	$\frac{1}{2}r(1-r)$
(9) aabb	$q^2(1-r)^2$	$\frac{1}{4}(1-r)^2$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$  sendo  $p+q = 1$

A função de verossimilhança, considerando a DS, para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r; n_i) = \lambda [p^2(1-r)^2]^{n_1} [2p^2r(1-r)]^{n_2} [p^2r^2]^{n_3} [2pqr(1-r)]^{n_4+n_6} [2pq(1-r)^2 + 2pqr^2]^{n_5} \\ [q^2r^2]^{n_7} [2q^2r(1-r)]^{n_8} [q^2(1-r)^2]^{n_9}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q = 1;$$

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!n_5!n_6!n_7!n_8!n_9!};$$

$$N = \sum_{i=1}^9 n_i.$$

#### 4.3.1.4. População $F_2$ , Marcadores Codominantes, Derivada de $F_1$ em Repulsão

Considerando uma população  $F_2$  codominante descendente do cruzamento entre dois pais heterozigotos com duas marcas dominantes (A/a e B/b) e admitindo que estas marcas encontram-se em fase de repulsão no  $F_1$  duplo-heterozigoto, com a configuração Ab//aB e que a distorção é explicada por variação na probabilidade gamética do loco A/a, pode-se admitir os valores de probabilidade conjunta já apresentados anteriormente. Dessa forma tem-se a Tabela 24, com os valores de freqüências gaméticas esperadas com e sem distorção na  $F_1$  em repulsão.

Tabela 24 – Comparação de valores de freqüências gaméticas esperadas sem e com distorção

Gameta	Tipo	Freqüência (s=0)	Freqüência (s=0)
Ab	Paternal	$\frac{1-r}{2}$	$p(1-r)$
AB	Recombinante	$\frac{r}{2}$	$pr$
ab	Recombinante	$\frac{r}{2}$	$qr$
aB	Paternal	$\frac{1-r}{2}$	$q(1-r)$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$ , sendo  $p+q = 1$



Também em  $F_2$  codominante em repulsão ocorre o evento de que se o loco A/a apresenta distorção gamética, o loco B/b ligado a ele certamente apresentará distorção pela ação indireta da seleção sobre o loco em ligação fatorial. Embora quando em repulsão tem-se:

$$f(b) = p(1-r) + qr = \frac{1}{2} + s - 2rs$$

e

$$f(B) = pr + q(1-r) = \frac{1}{2} - s + 2rs$$

Da mesma forma das populações anteriormente relatadas, a obtenção das freqüências esperadas dos gametas considerando DS capacitou a obtenção da freqüência esperada de cada classe genotípica com a presença de distorção de segregação gamética, conforme a Tabela 25.

Tabela 25 – Freqüências genotípicas esperadas com e sem distorção gamética em população  $F_2$ , com marcadores codominantes, derivada de  $F_1$  em repulsão

Genótipo	Freqüência esperada com distorção $s$	Freqüência esperada sem distorção ( $s=0$ )
(1) AABB	$p^2r^2$	$\frac{1}{4}r^2$
(2) AABb	$2p^2r(1-r)$	$\frac{1}{2}r(1-r)$
(3) AAbb	$p^2(1-r)^2$	$\frac{1}{4}(1-r)^2$
(4) AaBB	$2pqr(1-r)$	$\frac{1}{2}r(1-r)$
(5) AaBb	$2pq(1-r)^2 + 2pqr^2$	$\frac{1}{2}(1-r)^2 + \frac{1}{2}r^2$
(6) Aabb	$2pqr(1-r)$	$\frac{1}{2}r(1-r)$
(7) aaBB	$q^2(1-r)^2$	$\frac{1}{4}(1-r)^2$
(8) aaBb	$2q^2r(1-r)$	$\frac{1}{2}r(1-r)$
(9) aabb	$q^2r^2$	$\frac{1}{4}r^2$

A função de verossimilhança, considerando a DS, para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r; n_i) = \lambda L(r; n_i) = \lambda [p^2 r^2]^{n_1} [2p^2 r(1-r)]^{n_2} [p^2(1-r)^2]^{n_3} [2pqr(1-r)]^{n_4+n_6} [2pq(1-r)^2 + 2pqr^2]^{n_5} \\ [q^2(1-r)^2]^{n_7} [2q^2 r(1-r)]^{n_8} [q^2 r^2]^{n_9}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q=1;$$

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!n_5!n_6!n_7!n_8!n_9!};$$

$$N = \sum_{i=1}^9 n_i .$$

#### 4.3.1.5. População $F_2$ , Marcadores Dominantes, Derivada de $F_1$ em Repulsão

Considerando uma população  $F_2$  dominante descendente do cruzamento entre dois pais heterozigotos com duas marcas dominantes (A/a e B/b) e admitindo que estas marcas encontram-se em fase de repulsão no  $F_1$  duplo-heterozigoto, com a configuração Ab//aB, e que a distorção é explicada por variação na probabilidade gamética do loco A/a, pode-se admitir os seguintes valores de probabilidade conjunta:

$$P(Ab) = P(A)P(b/A) = \left(\frac{1}{2} + s\right)(1-r) = \frac{(1+2s)(1-r)}{2}$$

$$P(AB) = P(A)P(B/A) = \left(\frac{1}{2} + s\right)r = \frac{(1+2s)r}{2}$$

$$P(ab) = P(a)P(b/a) = \left(\frac{1}{2} - s\right)r = \frac{(1-2s)r}{2}$$

$$P(aB) = P(a)P(B/a) = \left(\frac{1}{2} - s\right)(1-r) = \frac{(1-2s)(1-r)}{2}$$

Desta forma tem-se a Tabela 26 com os valores de freqüências gaméticas esperadas com e sem distorção na  $F_1$  em repulsão.

Tabela 26 – Comparação de valores de freqüências gaméticas esperadas sem e com distorção

Gameta	Tipo	Freqüência (s=0)	Freqüência (s=0)
Ab	Paternal	$\frac{1-r}{2}$	$p(1-r)$
AB	Recombinante	$\frac{r}{2}$	$pr$
ab	Recombinante	$\frac{r}{2}$	$qr$
aB	Paternal	$\frac{1-r}{2}$	$q(1-r)$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$ , sendo  $p+q = 1$

Também em  $F_2$  dominante, derivada de  $F_1$  em repulsão, ocorre o evento de que se o loco A/a apresenta DS gamética, o loco B/b ligado a ele apresentará DS pela ação indireta da seleção sobre o loco em ligação fatorial. A configuração é idêntica à de  $F_2$  codominante em repulsão que é:

$$f(b) = p(1-r) + qr = \frac{1}{2} + s - 2rs \quad \text{e} \quad f(B) = pr + q(1-r) = \frac{1}{2} - s + 2rs$$

Obtidas as freqüências esperadas dos gametas, obteve-se a freqüência esperada das classes genotípicas com DS gamética, Tabela 27.

Tabela 27 – Freqüências genotípicas esperadas com e sem distorção gamética em população  $F_2$  com marcadores dominantes, derivada de  $F_1$  em repulsão

Genótipo	Freqüência esperada com distorção ( $s \neq 0$ )	Freqüência esperada sem distorção ( $s=0$ )
(1) A_B_	$1 - q^2 - \{[p(1-r)] + (qr)\}^2 + q^2r^2$	$\frac{1}{2} + \frac{1}{4}r^2$
(2) A_bb	$\{[p(1-r)] + qr\}^2 - q^2r^2$	$\frac{1}{4} - \frac{1}{4}r^2$
(3) aaB_	$q^2 - q^2r^2$	$\frac{1}{4} - \frac{1}{4}r^2$
(4) aabb	$q^2r^2$	$\frac{1}{4}r^2$

A função de verossimilhança, considerando a DS, para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r; n_i) = \lambda \left[ 1 - q^2 - \{[p(1-r)] + (qr)\}^2 + q^2 r^2 \right]^{n_1} \left\{ [p(1-r)] + qr \right\}^2 - q^2 r^2 \right]^{n_2} \\ \left[ q^2 - q^2 r^2 \right]^{n_3} \left[ q^2 r^2 \right]^{n_4}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q = 1;$$

$$\lambda = \frac{N!}{n_1! n_2! n_3! n_4!};$$

$$N = n_1 + n_2 + n_3 + n_4.$$

#### 4.3.1.6. População de Retrocruzamento em Repulsão

Em relação a uma população de retrocruzamento, resultante do cruzamento do  $F_1$  descendente do cruzamento entre dois pais heterozigotos, com duas marcas dominantes (A/a e B/b), admitindo que estas marcas encontram-se em fase de repulsão no  $F_1$  duplo-heterozigoto, com a configuração Ab//aB, e que a distorção é explicada por variação na probabilidade gamética do loco A/a, pode-se admitir os seguintes valores de probabilidade conjunta:

$$P(Ab) = P(A)P(b/A) = \left( \frac{1}{2} + s \right) (1-r) = \frac{(1+2s)(1-r)}{2}$$

$$P(AB) = P(A)P(B/A) = \left( \frac{1}{2} + s \right) r = \frac{(1+2s)r}{2}$$

$$P(ab) = P(a)P(b/a) = \left( \frac{1}{2} - s \right) r = \frac{(1-2s)r}{2}$$

$$P(aB) = P(a)P(B/a) = \left( \frac{1}{2} - s \right) (1-r) = \frac{(1-2s)(1-r)}{2}$$

Desta forma tem-se a Tabela 28, comparativa de valores de freqüências gaméticas esperadas com e sem distorção na  $F_1$  em repulsão.

Tabela 28 – Comparação de valores de freqüências gaméticas esperadas sem e com distorção

Gameta	Tipo	Freqüência (s=0)	Freqüência (s=0)
Ab	Paternal	$\frac{1-r}{2}$	$p(1-r)$
AB	Recombinante	$\frac{r}{2}$	$pr$
ab	Recombinante	$\frac{r}{2}$	$qr$
aB	Paternal	$\frac{1-r}{2}$	$q(1-r)$

em que:  $p = (1+2s) / 2$  e  $q = (1-2s) / 2$ , sendo  $p+q = 1$

Conforme demonstrado, se o loco A/a apresenta DS gamética, o loco B/b ligado certamente apresentará DS pela ação indireta da seleção sobre o loco em ligação fatorial. Assim, com a obtenção das freqüências esperadas dos gametas da F<sub>1</sub> com DS e da esperada do pai recorrente (recessivo), que é unitária (não influencia), obteve-se a freqüência esperada de cada classe genotípica com a presença de DS gamética, conforme a Tabela 29.

Tabela 29 – Freqüências genotípicas esperadas com e sem DS gamética em população de retrocruzamento, derivada de F<sub>1</sub> em repulsão

Genótipo	Freqüência esperada com distorção $\underline{s}$	Freqüência esperada sem distorção (s=0)
(1) AaBb	$pr$	$\frac{1}{2}(1-r)$
(2) Aabb	$p(1-r)$	$\frac{1}{2}r$
(3) aaBb	$q(1-r)$	$\frac{1}{2}r$
(4) aabb	$qr$	$\frac{1}{2}(1-r)$

A função de verossimilhança, considerando a DS, para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r; n_i) = \lambda [pr]^{n_1} [p(1-r)]^{n_2} [q(1-r)]^{n_3} [qr]^{n_4}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q = 1;$$

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!};$$

$$N = n_1 + n_2 + n_3 + n_4.$$

O estimador da porcentagem de recombinação entre os pares de locos considerados não se altera com a consideração de que uma das marcas apresenta segregação distorcida.

#### 4.3.1.7. *População de Duplo-Haplóide, Derivada de F<sub>1</sub> em Aproximação*

As populações de duplo-haplóides são obtidas pela duplicação do número de cromossomos de plantas haplóides adquiridas por cultura de anteras ou embriões haplóides imaturos. Todos os indivíduos são homozigotos e representam a variabilidade genética encontrada no indivíduo parental que produziu os grãos de pólen ou o embrião haplóide. Tem vantagem de ser uma população que pode ser perpetuada. Permite o cultivo em vários locais, possibilitando a obtenção de estimativa mais precisa dos parâmetros dos caracteres quantitativos a serem mapeados. Além disso, é possível mapear locos considerando a interação genótipo x ambiente nessas características. Mas, por serem compostas apenas de indivíduos homozigotos, apenas dois genótipos são possíveis em cada loco (AA e aa), razão pela qual não é possível estimar parâmetros associados ao grau de dominância dos locos mapeados. Outra desvantagem é que requer o emprego de técnicas de cultura de tecidos, que não estão disponíveis para algumas espécies (Schuster & Cruz, 2004).

As frequências esperadas de cada classe genotípica numa população de duplo-haplóides, derivada da F<sub>1</sub> em fase de aproximação, é apresentada na Tabela 30.

Tabela 30 – Freqüências genotípicas esperadas com e sem distorção gamética em população duplo-haplóide, derivada de F<sub>1</sub> em aproximação

Genótipo	Freqüência esperada Com distorção (s ≠ 0)	Freqüência esperada sem distorção (s=0)
(1) AABB	p(1 - r)	$\frac{1}{2}(1 - r)$
(2) AAbb	pr	$\frac{1}{2}r$
(3) aaBB	qr	$\frac{1}{2}r$
(4) aabb	q(1 - r)	$\frac{1}{2}(1 - r)$

A função de verossimilhança considerando a DS para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r; n_i) = \lambda [p(1 - r)]^{n_1} [pr]^{n_2} [qr]^{n_3} [q(1 - r)]^{n_4}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q = 1;$$

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!};$$

$$N = n_1 + n_2 + n_3 + n_4 .$$

A distorção não altera o estimador de máxima verossimilhança entre os pares de marcas estudados.

#### 4.3.1.8. População de Duplo-Haplóide, Derivada de F<sub>1</sub> em Repulsão

As freqüências esperadas de cada classe genotípica numa população de duplo-haplóides, derivada da F<sub>1</sub> em fase de repulsão, são apresentadas na Tabela 31.

Tabela 31 – Freqüências genótípicas esperadas com e sem distorção gamética em população duplo-haplóide, derivada de F<sub>1</sub> em repulsão

Genótipo	Freqüência esperada com distorção (s ≠ 0)	Freqüência esperada sem distorção (s=0)
(1) AAbb	p(1-r)	$\frac{1}{2}(1-r)$
(2) AABB	pr	$\frac{1}{2}r$
(3) aabb	qr	$\frac{1}{2}r$
(4) aaBB	q(1-r)	$\frac{1}{2}(1-r)$

A função de verossimilhança, considerando a DS, para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r; n_i) = \lambda [p(1-r)]^{n_1} [pr]^{n_2} [qr]^{n_3} [q(1-r)]^{n_4}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q=1;$$

e

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!};$$

$$N = n_1 + n_2 + n_3 + n_4 .$$

A distorção não altera o estimador de máxima verossimilhança entre os pares de marcas estudados.

#### 4.3.1.9. População de RIL, Derivada de F<sub>1</sub> em Aproximação

As populações de RILs são derivadas por sucessivas autofecundações a partir de uma população F<sub>1</sub>, pelo método da descendência de uma única semente (SSD) ou por cruzamentos entre irmãos, até atingirem um elevado grau de homozigose, de maneira que ao final do processo todos os locos gênicos terão uma segregação de 1:1



(AA:aa). Na formação das RILs por meio de autofecundações, cada planta  $F_2$  gera uma planta  $F_3$ , que por sua vez gera uma planta  $F_4$ , e assim por diante. Quando a população atinge a geração  $F_6$  ou  $F_7$ , abrem-se linhas, que são as RILs. Nelas, cada planta  $F_2$  é representada por uma linha endogâmica homozigota. Assim, toda a variabilidade existente na população  $F_2$  original estará representada pelas RILs, desde que um tamanho de população adequado seja utilizado.

As desvantagens da utilização de RILs são o tempo requerido para a obtenção da população, de 6 a 8 gerações de autofecundações e a impossibilidade da estimação de efeitos devido à dominância, uma vez que, igualmente em população de duplo-haplóide, apresentam apenas dois genótipos na população segregante. Entretanto, por serem compostas apenas por indivíduos homozigotos, as RILs também podem ser perpetuadas e a disponibilidade de grande quantidade de sementes de cada RIL permite o cultivo em vários locais, possibilitando a obtenção de estimativas mais precisas dos parâmetros dos caracteres quantitativos a serem mapeados. Além disso, é possível mapear locos associados à interação genótipo x ambiente destas características. A necessidade de muitos ciclos de meiose para atingir a homozigose resulta em oportunidades adicionais para a recombinação.

A frequência de recombinação nas RILs ( $r'$ ) aproxima-se do valor  $r' = 2r/(1+2r)$  para linhas autofecundadas e  $r' = 4r/(1+6r)$  para linhas formadas a partir de cruzamentos entre irmãos, sendo  $r$  a frequência de recombinação na geração  $F_2$ , que expressa a distância entre os locos (Burr e Burr, 1991). Contudo, cada RIL tem dois cromossomos homólogos idênticos, enquanto que indivíduos  $F_2$  têm dois homólogos distintos, os quais contribuem com diferentes informações.

As frequências esperadas de cada classe genotípica numa população de RILs, derivada da  $F_1$  em fase de aproximação, é apresentada na Tabela 32.

Tabela 32 – Freqüências genótípicas esperadas com e sem distorção gamética em população de RIL em aproximação

Genótipo	Freqüência esperada com distorção ( $s \neq 0$ )	Freqüência esperada sem distorção ( $s=0$ )
(1) AABB	$p(1 - r')$	$\frac{1}{2}(1 - r')$
(2) AAbb	$pr'$	$\frac{1}{2}r'$
(3) aaBB	$qr'$	$\frac{1}{2}r'$
(4) aabb	$q(1 - r')$	$\frac{1}{2}(1 - r')$

A função de verossimilhança, considerando a DS, para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r'; n_i) = \lambda [p(1 - r')]^{n_1} [pr']^{n_2} [qr']^{n_3} [q(1 - r')]^{n_4}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q = 1;$$

e

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!};$$

$$N = n_1 + n_2 + n_3 + n_4.$$

A distorção não altera o estimador de máxima verossimilhança entre os pares de marcas estudados.

#### 4.3.1.10. População de RIL, Derivada de $F_1$ em Repulsão

As freqüências esperadas de cada classe genotípica numa população de RILs, derivada da  $F_1$  em fase de repulsão, é apresentada na Tabela 33.

Tabela 33 – Freqüências genótípicas esperadas com e sem distorção gamética em população de RIL em repulsão

Genótipo	Freqüência esperada com distorção ( $s \neq 0$ )	Freqüência esperada sem distorção ( $s=0$ )
(1) AAbb	$p(1 - r')$	$\frac{1}{2}(1 - r')$
(2) AABB	$pr'$	$\frac{1}{2}r'$
(3) aabb	$qr'$	$\frac{1}{2}r'$
(4) aaBB	$q(1 - r')$	$\frac{1}{2}(1 - r')$

A função de verossimilhança considerando a DS para o cálculo da porcentagem de recombinação é a seguinte:

$$L(r'; n_i) = \lambda [p(1 - r')]^{n_1} [pr']^{n_2} [qr']^{n_3} [q(1 - r')]^{n_4}$$

em que:

$$p = (1+2s) / 2;$$

$$q = (1-2s) / 2;$$

$$p+q = 1;$$

e

$$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!};$$

$$N = n_1 + n_2 + n_3 + n_4.$$

A distorção não altera o estimador de máxima verossimilhança entre os pares de marcas estudados.

#### 4.3.2. Distorção Genotípica

Estudando o processo de distorção de segregação genotípica, considerou-se que para este tipo de DS a probabilidade de ocorrência de genótipos associados aos loco A/a ou B/b não é possível de ser predita. Esse tipo de DS geralmente causa quebra da estrutura da população e as freqüências genótípicas são alteradas de forma não sistemática.

Para fins de abordagem sobre as conseqüências da alteração na freqüência genotípica em razão da viabilidade diferencial de um genótipo, foi admitido, inicialmente, a disponibilidade de uma população  $F_2$  dominante com 1000 indivíduos, em que as marcas  $A/a$  e  $B/b$  segregam como esperado, 3:1. Entretanto, é possível que um fator qualquer possa a vir ocorrer proporcionando a morte, pós-fertilização, de genótipo(s) associado a um loco. Admitindo que a marca  $A/a$  apresentará distorção de segregação, não será possível prever a probabilidade conjunta dos locos, devido a morte poder ter ocorrido em qualquer genótipo com a presença de  $A$ . Assim, por exemplo, as classe do genótipos  $A-B-$  e/ou  $A-bb$  serão afetadas de forma imprevisível. Para melhor compreensão pode-se exemplificar, supondo que tenha ocorrido, dentre os 1000 indivíduos que constituíam originalmente a população  $F_2$ , a morte de metade deles (500 indivíduos) resultando segregação distorcida do loco  $A/a$ , deixando de ser o esperado 3:1 e passando a ser de 1:1, conforme descrito na Tabela 34.

Tabela 34 – População  $F_2$  dominante com segregação genotípica de 1:1 no loco  $A/a$  por mortalidade genotípica

Tam. Pop.	Seg.	$A-B-$	$A-bb$	$aaB-$	$aabb$
1000	3:1	562,50	187,50	187,50	62,50
500	1:1	62,00	188,00	188,00	62,00
500	1:1	250,00	0,00	188,00	62,00

Verifica-se nas duas últimas linhas, dois casos hipotéticos. No primeiro a morte de genótipos  $A-$  ocorreu apenas entre os indivíduos  $B-$  e no segundo caso, a morte de  $A-$  ocorreu em todos os indivíduos  $bb$  e em alguns portadores de  $bb$ . Certamente que inúmeras outras situações poderiam ter ocorrido. A idéia que se tem é a de que é impossível prever as freqüências esperadas das classes genotípicas com a presença de distorção de segregação genotípica, pois é possível até mesmo não serem observados indivíduos em uma dada classe genotípica, resultando, como relatado anteriormente, a quebra da estrutura genética preditiva da população de mapeamento.

Com intenção de melhor compreensão dos efeitos causados por DS genotípica na estimação de frequência de recombinação entre duas marcas, estudaram-se as consequências em populações de  $F_2$  dominante, com presença de DS genotípica, utilizando as funções de verossimilhança normal e a deduzida para DS gamética em diversas razões de segregação com tamanhos iniciais de 1000, 500, 200 e 100 indivíduos nas populações com frequências de recombinação de 50cM, 20cM e 10cM (Tabelas 35, 36 e 37). A DS de 1:1 e 2:1 considerou-se a morte de indivíduo(s) em dada(s) classe(s) genotípica(s) e a DS de 4:1 foi devido ao acréscimo de indivíduos em dada(s) classe(s) genotípica(s).

Observando os resultados da Tabela 35, verificou-se na situação considerada referente à DS genotípica, que pode ocorrer falsa ligação se considerado o mínimo valor de 40 cMorgans, fato ratificado nas populações de 1000 e 500 com segregação observada de 2:1, em que ocorreu a morte de todos os indivíduos de uma dada classe genotípica que no exemplo foi a classe A-bb.

Estes fatos evidenciam falha grave na construção de mapas genéticos. Deve-se ter em mente que a utilização do valor 40 cMorgans como valor mínimo de frequência de recombinação em mapeamentos genéticos é comumente observado (Faleiro et al., 2003).

Examinando os resultados obtidos nas Tabelas 36 e 37 é demonstrado que as estimativas das frequências de recombinação são extremamente viesadas, podendo até mesmo apresentar falsa independência entre marcas, como observada em algumas populações com segregação de 4:1.

Portanto os resultados das Tabelas 35, 36 e 37, comprovam que mesmo com a função de verossimilhança deduzida a estimação da frequência de recombinação com a utilização de marcadores com DS genotípica promoverá mapas não acurados, com possibilidade, por exemplo, de ordenamento incorreto das marcas, acréscimo ou decréscimo do tamanho do genoma, entre outros fatores. Em diversos trabalhos é relatado que DS de segregação causa viés na frequência de recombinação em mapas genéticos (Lorieux et al., 1995ab; Liu, 1998; Vogl e Xu, 2000 e Eujayl et al., 1997).

Tabela 35 – Estimativas de freqüências de recombinação em populações de F<sub>2</sub> dominante em diversas razões de segregação originada por DS genotípica ocasionada por morte de genótipo(s), em que inicialmente a distância era de 50 cMorgans

TAM.POP	SEG.	A-B-	A-bb	aaB-	aabb	H <sub>0</sub> : s=0		H <sub>0</sub> : s 0	
						r <sup>17</sup>	LOD	r <sup>17</sup>	LOD
1000	3:1	562,50	187,50	187,50	62,50	50,00	0,00	50,00	0,00
	1:1	62	188	188	62	61,00	0,00	61,00	0,00
	1:1	250	0	188	62	41,30	2,10	44,30	0,69
	2:1	312	188	188	62	56,19	0,00	56,19	0,00
	2:1	500	0	188	62	32,90	8,99	33,60	7,05
	2:1	483	62	188	62	41,00	2,65	41,70	1,95
	4:1	812	188	188	62	43,90	1,56	43,80	1,72
	4:1	562	438	188	62	63,30	0,00	63,30	0,00
	4:1	687	313	188	62	55,46	0,00	55,46	0,00
500	3:1	281,25	93,75	93,75	31,25	50,00	0,00	50,00	0,00
	1:1	31	94	94	31	61,28	0,00	61,00	0,00
	1:1	125	0	94	31	41,30	1,05	44,30	0,35
	2:1	156	94	94	31	56,19	0,00	56,19	0,00
	2:1	250	0	94	31	32,90	4,50	33,6	3,52
	4:1	406	94	94	31	43,90	0,73	43,80	0,86
	4:1	281	219	94	31	63,30	0,00	63,30	0,00
	4:1	344	156	94	31	55,38	0,00	55,46	0,00
200	3:1	112,50	37,50	37,50	12,50	50,00	0,00	50,00	0,00
	1:1	12	38	38	12	62,02	0,00	62,00	0,00
	1:1	38	12	38	12	50,61	0,00	50,61	0,00
	2:1	62	38	38	12	56,99	0,00	56,99	0,00
	2:1	100	0	38	12	33,50	1,65	34,20	1,26
	4:1	162	38	38	12	44,70	0,22	44,50	0,27
	4:1	112	88	38	12	63,93	0,00	63,93	0,00
	4:1	138	62	38	12	55,80	0,00	55,80	0,00
100	3:1	56,25	18,75	18,75	6,25	50,00	0,00	50,00	0,00
	1:1	6,00	19	19	6	62,02	0,00	61,00	0,00
	1:1	25,00	0	19	6	42,10	0,17	44,30	0,05
	2:1	31,00	19	19	6	56,99	0,00	56,19	0,00
	2:1	50,00	0	19	6	33,50	0,82	33,60	0,63
	4:1	81,00	19	19	6	44,70	0,11	43,80	0,13
	4:1	56,00	44	19	6	63,93	0,00	63,30	0,00
	4:1	69,00	31	19	6	55,80	0,00	55,46	0,00

<sup>17</sup> valores acima de 50% devem ser considerados como estimativa de 50% de recombinação

Tabela 36 – Estimativas de freqüências de recombinação em populações de F<sub>2</sub> dominante em diversas razões de segregação originada por DS genotípica ocasionada por morte de genótipo(s), em que inicialmente a distância era de 20 cMorgans

TAM.POP	SEG.	A-B-	A-bb	aaB-	aabb	H <sub>0</sub> : s=0		H <sub>0</sub> : s 0	
						r <sup>1/</sup>	LOD	r <sup>1/</sup>	LOD
1000	3:1	660	89	89	160	20,00	53,82	20,00	53,82
	1:1	160	90	90	160	28,10	23,88	26,80	26,88
	1:1	250	0	90	160	15,40	55,60	14,90	53,53
	2:1	410	90	90	160	23,60	37,53	23,00	38,09
	2:1	500	0	90	160	12,30	77,12	12,00	75,70
	4:1	910	90	90	160	17,20	72,08	17,70	72,43
	4:1	660	340	90	160	38,20	9,65	39,30	8,14
	4:1	785	215	90	160	28,40	30,44	29,20	29,42
500	3:1	330	45	45	80	20,00	26,88	20,00	26,91
	1:1	80	45	45	80	28,10	11,94	26,80	13,44
	1:1	125	0	45	80	15,40	27,80	14,90	26,77
	2:1	205	45	45	80	23,60	18,76	23,00	19,05
	2:1	250	0	45	80	12,30	38,56	12,00	37,85
	4:1	455	45	45	80	17,20	36,04	17,70	36,22
	4:1	330	170	45	80	38,20	4,82	39,30	4,07
	4:1	392	108	45	80	28,50	15,10	29,20	14,71
200	3:1	132	17	17,98	32,00	20,00	10,76	20,00	10,76
	1:1	32	18	18	32	28,10	4,78	26,80	5,38
	1:1	50	0	18	32	15,40	11,12	14,90	10,71
	2:1	82	18	18	32	23,60	7,51	23,00	7,62
	2:1	100	0	18	32	12,30	15,42	12,00	15,14
	4:1	182	18	18	32	17,20	14,42	17,70	14,49
	4:1	132	68	18	32	38,20	1,93	39,30	1,63
	4:1	157	43	18	32	28,40	6,09	29,20	5,88
100	3:1	66	9	9	16	20,00	5,38	20,00	5,38
	1:1	16	9	9	16	28,10	2,39	26,80	2,69
	1:1	25	0	9	16	15,40	5,56	14,90	5,35
	2:1	41	9	9	16	23,60	3,75	23,00	3,81
	2:1	50	0	9	16	12,30	7,71	12,00	7,57
	4:1	91	9	9	16	17,20	7,21	17,70	7,24
	4:1	66	34	9	16	38,20	0,96	39,30	0,81
	4:1	78	22	9	16	28,90	2,93	29,20	2,94

<sup>1/</sup> valores acima de 50% devem ser considerados como estimativa de 50% de recombinação

Tabela 37 – Estimativas de freqüências de recombinação em populações de F<sub>2</sub> dominante em diversas razões de segregação originada por DS genotípica ocasionada por morte de genótipo(s), em que inicialmente a distância era de 10 cMorgans

TAM.POP	SEG.	A-B-	A-bb	aaB-	aabb	H <sub>0</sub> : s=0		H <sub>0</sub> : s 0	
						r <sup>1/</sup>	LOD	r <sup>1/</sup>	LOD
1000	3:1	702,50	47,40	47,40	202,50	10,00	114,66	10,00	114,66
	1:1	202	48	48	202	14,70	68,38	14,00	70,33
	1:1	250	0	48	202	7,70	99,48	7,30	98,29
	2:1	452	48	48	202	12,00	90,23	11,70	90,56
	2:1	500	0	48	202	6,10	125,80	5,90	125,02
	4:1	952	48	48	202	8,70	138,22	9,00	138,42
	4:1	702	298	48	202	29,50	32,01	30,40	29,62
500	3:1	351,25	23,70	23,70	101,25	10,00	57,33	10,00	57,33
	1:1	101	24	24	101	14,7	34,19	14,00	33,84
	1:1	125	0	24	101	7,7	49,74	7,30	49,92
	2:1	250	0	24	101	6,1	62,90	5,90	63,20
	2:1	226	24	24	101	12	45,12	11,70	44,96
	4:1	476	24	24	101	8,7	69,11	9,00	69,21
	4:1	351	149	24	101	29,5	16,00	30,40	14,81
200	3:1	140,50	9,48	9,48	40,50	10,00	22,94	10,00	22,93
	1:1	41	9	9	41	13,80	14,55	14,00	13,54
	1:1	50	0	9	41	7,20	20,65	7,30	19,97
	2:1	91	9	9	41	11,30	19,04	5,90	25,28
	2:1	100	0	9	41	5,70	25,98	11,70	17,98
	4:1	191	9	9	41	8,10	28,82	9,00	27,68
	4:1	141	59	9	41	29,00	6,82	30,40	5,92
100	3:1	70,25	4,74	4,74	20,25	10,00	11,47	10,00	11,47
	1:1	20	5	5	20	15,40	6,56	14,00	6,77
	1:1	25	0	5	20	8,00	9,70	7,30	9,98
	2:1	45	5	5	20	12,60	8,70	5,90	12,64
	2:1	50	0	5	20	6,40	12,31	11,70	8,99
	4:1	95	5	5	20	9,10	13,44	9,00	13,84
	4:1	70	30	5	20	30,00	3,06	30,40	2,96
4:1	82	18	5	20	20,40	6,53	20,10	6,85	

<sup>1/</sup> valores acima de 50% devem ser considerados como estimativa de 50% de recombinação



#### 4.4. Mapeamento Incluindo Marcadores com Segregação de Distorção

##### 4.4.1. Distorção Gamética

Neste tópico é abordado o emprego de metodologia para o estabelecimento de mapas genéticos, considerando a existência de marcadores com distorção de segregação, causada por distorção gamética. Para isso, foram gerados, por meio do módulo de simulação do programa computacional GQMOL/UFV (CRUZ, 2006), dados de populações  $F_2$ , considerando o estudo de marcadores dominantes. O genoma original foi constituído por três (3) grupos de ligação (GL), com 11 marcas/GL eqüidistantes (10cM) e comprimento de 100cM/GL. Os tamanhos das populações estudadas foram de 200, 400 e 1000 indivíduos e o nível de distorção de segregação gamética foi de 1:1, 2:1, 7:1 e 10:1, presentes nos marcadores posicionados na quinta posição de cada grupo de ligação.

A estimação da porcentagem de recombinação foi efetuada pelo método da máxima verossimilhança, utilizando duas funções: a normalmente conhecida para a população, a qual não leva em consideração a distorção de segregação presente (Liu, 1998; e Schuster & Cruz, 2004), considerada como estratégia 1 (E1); e a função desenvolvida neste trabalho (item 3.1.1.), que considera a presença de marcas com distorção de segregação, para a qual o mapeamento foi realizado considerando duas estratégias, para obtenção da porcentagem de recombinação. A primeira admite que se conhecem quais são as marcas que apresentam distorção de segregação e também a taxa esperada, de forma que o valor de  $s$  pode ser assumido como paramétrico e assim faz-se a indicação destas marcas e suas respectivas taxa de distorção, considerada como estratégia 2 (E2). A segunda estima a taxa de DS e o valor estimado de distorção observado em todos os marcadores que foram utilizados para o cálculo do valor da distância entre os marcadores  $i$  e  $i'$ . Como, para o marcador  $i$ , a taxa de distorção será dada por  $\hat{s}_i$  e o marcador  $i'$  apresenta taxa de distorção  $\hat{s}_{i'}$ , foram obtidas duas estimativas da medida da porcentagem de recombinação entre  $i$  e  $i'$ , sendo o valor da distância a estimativa média. Então, tem-se a estratégia 3 (E3).

Desta maneira foram obtidos, para cada população simulada, três mapas genéticos, sendo que apenas um deles é ignorado a existência de distorção de segregação e, nos outros dois, é admitindo a existência de distorção.

#### 4.4.1.1. Análise de Segregação de Locos Individuais

Previamente à realização do mapeamento genético para todas as populações segregante geradas, aplicou-se testes de qui-quadrado ( $\chi^2$ ) para verificar se, de fato, a razão de segregação de locos individuais ocorreria como esperado de acordo com segregações mendelianas. Assim, a hipótese  $H_0:3:1$  do loco A-:aa foi avaliada, a 5% de probabilidade (erro Tipo I).

Pode-se exemplificar a aplicação do teste de qui-quadrado para a marca D15 de uma população  $F_2$  dominante de segregação esperada de 3:1, constituída de 1000 indivíduos em que se simulou, para esta marca distorção, frequência esperada de 7:1. Obteve-se número observado de 875 indivíduos, para A- e, e de 125 indivíduos, para aa. Assim tem-se o teste conforme descrito na Tabela 38.

Tabela 38 – Teste de segregação de qui-quadrado em uma população  $F_2$  dominante

Fenótipos	Razão Fenotípica	Número Observado	Número Esperado	$(O - E)^2 / E$
A-	3	875	750	20,83
aa	1	125	250	62,50

GL= 1  
 Probabilidade: 0,00%

$$\chi^2_{\text{calc}} = \sum_{i=1}^n \left[ \frac{(\text{Obs}_i - \text{Esp}_i)^2}{\text{Esp}_i} \right] = \left[ \frac{(875 - 750)^2}{750} \right] + \left[ \frac{(125 - 250)^2}{250} \right] = 83,33$$

Verificou-se a significância do teste qui-quadrado ao testar a hipótese de segregação 3:1, do loco D15. Tal fato ocorreu em razão da distorção de

segregação conforme a condição imposta na simulação da população de 7:1.

Na Tabela 39 e 40 são apresentadas as probabilidades dos testes de segregação em relação à hipótese  $H_0$ : 3:1, para todas as marcas na população de tamanho 1000 e ainda as frequências esperada e observada de indivíduos e as razões de segregação observada de todas as marcas.

Verifica-se, pela Tabela 39, que praticamente todas as marcas adjacentes a D15, D25 e D35 apresentaram DS, devido à ligação fatorial com as marcas citadas, que possuíam DS estabelecidas na simulação das populações. Com exceção da marca D211, todas as demais foram significativas, a 5% de probabilidade pelo teste de qui-quadrado, concluindo sobre a rejeição da segregação esperada de 3:1, ( $H_0$ ) para população  $F_2$  dominante. Este fato é biologicamente possível e esperado principalmente com alta taxa de DS, como de 7:1, em questão. Xu et al. (1997) relatam que a DS nas marcas adjacentes pode ocorrer com a cobertura de todo o GL ou em grande extensão deste. E ainda, os autores destacam esta ocorrência em seu trabalho, que em dois grupos praticamente todas as marcas apresentaram DS (GL: 3 e 9) e, em outros dois grupos, a maioria das marcas, em determinada região dos cromossomos (GL: 8 e 11).

Pode-se observar (Tabela 40) que em uma população de 1000 indivíduos, em que as marcas D15, D25 e D35 apresentam DS gamética de 4:1 (A-:aa), imposta na simulação da população, que marcas adjacentes a estas podem não manifestar DS por ligação fatorial. Observou-se o referido na marca D16 adjacente à marca D15 com DS gamética (Tabela 40). A ocorrência do fenômeno é esperado devido a razão de segregação das marcas com DS gamética (4:1). Pois esse nível de razão de segregação das marcas com DS gamética afeta a segregação das marcas adjacentes. Mas, este o efeito, pode não ser grande o suficiente para que o teste de segregação de qui-quadrado rejeite a hipótese esperada de 3:1 (A-:aa).

Os fatores observados nas populações exemplificadas exibem um indicativo da boa qualidade dos dados gerados no processo de simulação, proporcionando adequabilidade e confiabilidade das populações estudadas, com concordância a fenômenos biológicos observados e relatados no meio científico.

Tabela 39 – Segregação dos marcadores em população F<sub>2</sub> com 1000 indivíduos, com as marcas D15, D25 e D35 apresentando distorção de segregação de 7:1

Marcador	A-	aa	Qui-quadrado	Segregação A-/aa	Probabilidade(%)	
D11	778	222	4,18	3,50:1	4,0872	*
D12	797	203	11,78	3,93:1	0,0598	*
D13	836	164	39,45	5,10:1	0,0	*
D14	855	145	58,80	5,90:1	0,0	*
<u>D15</u>	875	125	83,33	7:1	0,0	*
D16	849	151	52,27	5,62:1	0,0	*
D17	845	155	48,13	5,45:1	0,0	*
D18	826	174	30,81	4,75:1	0,0	*
D19	810	190	19,20	4,26:1	0,0012	*
D110	807	193	17,33	4,18:1	0,0031	*
D111	791	209	8,97	3,78:1	0,2752	*
D21	791	209	8,96	3,78:1	0,2752	*
D22	802	198	14,42	4,05:1	0,0146	*
D23	829	171	33,29	4,85:1	0,0	*
D24	839	161	42,24	5,21:1	0,0	*
<u>D25</u>	863	137	68,10	6,30:1	0,0	*
D26	830	170	34,13	4,88:1	0,0	*
D27	823	177	28,42	4,65:1	0,0	*
D28	810	190	19,20	4,26:1	0,0012	*
D29	797	203	11,78	3,93:1	0,0598	*
D210	795	205	10,80	3,88:1	0,1015	*
D211	774	226	3,07	3,42:1	7,9651	ns
D31	784	216	6,16	3,63:1	1,3028	*
D32	820	180	26,13	4,56:1	0,0	*
D33	838	162	41,30	5,17:1	0,0	*
D34	852	148	55,49	5,76:1	0,0	*
<u>D35</u>	868	132	74,26	6,58:1	0,0	*
D36	845	155	48,13	5,45:1	0,0	*
D37	835	165	38,53	5,06:1	0,0	*
D38	812	188	20,50	4,32:1	0,0006	*
D39	803	197	14,98	4,08:1	0,0109	*
D310	795	205	10,80	3,88:1	0,1015	*
D311	782	218	5,46	3,59:1	1,9442	*

\* valor abaixo do nível crítico testado a 5%, para segregação esperada de 3:1; ns não significativo.

Tabela 40 – Segregação dos marcadores em população F<sub>2</sub> com 1000 indivíduos com as marcas D15, D25 e D35 apresentando distorção de segregação de 4:1

Marcador	A-	aa	Qui-quadrado	Segregação A-/aa	Probabilidade(%)
D11	746	254	0,08	2,94	77,02 ns
D12	773	227	2,82	3,41	9,30 ns
D13	771	229	2,35	3,37	12,51 ns
D14	780	220	4,80	3,55	2,85 *
<u>D15</u>	787	213	7,30	3,69	0,69 *
D16	770	230	2,13	3,35	14,41 ns
D17	758	242	0,34	3,13	55,91 ns
D18	756	244	0,19	3,10	66,13 ns
D19	767	233	1,54	3,29	21,44 ns
D110	766	234	1,36	3,27	24,26 ns
D111	777	223	3,89	3,48	4,86 *
D21	751	249	0,005	3,02	94,18 ns
D22	758	242	0,34	3,13	55,91 ns
D23	781	219	5,12	3,57	2,36 *
D24	789	211	8,11	3,74	0,44 *
<u>D25</u>	784	216	6,16	3,63	1,30 *
D26	781	219	5,12	3,57	2,36 *
D27	777	223	3,89	3,48	4,86 *
D28	767	233	1,54	3,29	21,44 ns
D29	767	233	1,54	3,29	21,44 ns
D210	765	235	1,20	3,26	27,33 ns
D211	773	227	2,82	3,41	9,30 ns
D31	740	260	0,53	2,85	46,52 ns
D32	761	239	0,64	3,18	42,18 ns
D33	768	232	1,73	3,31	18,87 ns
D34	789	211	8,11	3,74	0,44 *
<u>D35</u>	798	202	12,29	3,95	0,05 *
D36	796	204	11,28	3,90	0,08 *
D37	782	218	5,46	3,59	1,94 *
D38	791	209	8,96	3,78	0,28 *
D39	788	212	7,70	3,72	0,55 *
D310	765	235	1,20	3,26	27,33 ns
D311	747	253	0,048	2,95	82,66 ns

\* valor abaixo do nível crítico testado a 5%, para segregação esperada de 3:1; ns não significativo.

Zamir & Tadmor(1986) afirmam que se um gene que causa DS está segregando em uma população, marcadores ligados a ele apresentaram razões de distorção. Vogl & Xu (2000) e Pereira et al. (1993) relatam que este fator pode ser causado devido a regiões específicas que apresentam a DS.

Xu et al. (1997) descrevem, em várias populações  $F_2$  de arroz, a existência de regiões associadas com DS, envolvendo fatores gaméticos, por exemplo. Yamada et al. (2001) relatam que resultados semelhantes aos encontrados por Xu et al. (1997), são descritos em outros trabalhos e que também em seu estudo com duas variedades de soja (*Glycine max.*) o fenômeno foi observado.

Por meio da análise dos dados da Tabela 39, demonstra-se também uma característica importante que é a presença de um gradiente de razão de distorção, em que as marcas mais próximas da marca sob seleção, neste estudo com DS devido a fator gamético, apresentam altas taxas de DS e à medida que as marcas se distanciam, diminuem os valores das taxas de DS. Pereira et al. (1993) observou em seu trabalho um gradiente de distorção com essas mesmas características observadas. Xu et al. (1997) também verificaram que marcadores apresentavam progressiva redução de DS à medida que se distanciavam dos marcadores com razão de segregação mais severa. Assim mais uma vez é demonstrado e evidenciado a adequabilidade do modelo proposto neste estudo, por meio de exemplos biológicos práticos.

#### 4.4.1.2. Mapeamento Genético

Após a aplicação dos testes de segregação estimou-se a porcentagem de recombinação entre pares de marcas pelo método da máxima verossimilhança, adotando as três estratégias citadas anteriormente, gerando, assim, três mapas genéticos, em cada população, com determinado valor de DS, para os quais foi utilizada a distância em Morgan e a propriedade transitiva para a determinação dos grupos de ligação. Os critérios utilizados no agrupamento foram  $r_{\max} = 30\text{cM}$  e LOD mínimo de 3, para inferir a ligação entre dois locos. E para o ordenamento se utilizou o método da SARF. Com intuito de se estudar os efeitos de marcadores com DS, todas as marcas foram utilizadas no mapeamento.

Os mapas genéticos assim construídos (Figuras 1 - 12), foram comparados com o genoma a partir do qual cada população foi gerada utilizando para a comparação os seguintes critérios: 1) número de grupos de ligação obtidos no mapeamento; 2) número de marcas obtidas por grupo de

ligação no mapeamento; 3) tamanho de grupos de ligação; 4) distância média entre marcadores adjacentes; 5) variância das distâncias entre marcas adjacentes e; 6) estresse (Tabelas 41, 42 e 43).

#### *4.4.1.2.1. Número de Grupos de Ligação*

O número de grupos de ligação esperado no processo de mapeamento das populações simuladas era igual a três (3), independente do tamanho de população e nível de distorção de segregação. Em todas as populações observou-se (Figuras 1 a 12) que o número de grupos de ligação foi exatamente de três não apresentando nenhuma exceção. Assim evidenciou-se que a DS gamética não afeta o número de grupos de ligação, mesmo com a utilização da estratégia que não considera a DS em sua função de verossimilhança. Demonstra-se para este critério, por meio desses resultados (Figuras 1 a 12), que a utilização de marcadores que apresentam DS gamética na construção de mapas de ligação é aconselhável e o descarte desses para construção do mapa seria inadequado.

O tamanho das populações utilizadas, de 200, 400 e 1000 indivíduos, são eficientes para se verificar o efeito da DS gamética na formação de GL. Pois de acordo com Ferreira et al. (2006), para populações com nível de saturação de 10cM, a partir do número mínimo de 200 indivíduos a formação de grupo de ligação não é afetada devido ao seu tamanho.

#### *4.4.1.2.2. Número de Marcas Obtidas por GL no Mapeamento*

O número esperado de marcas por grupo de ligação no mapeamento das populações simuladas era de 11 para todas as populações simuladas a partir dos genomas simulados.

Para esta característica foi observado em todas as populações (Figuras 1 a 12) que o número de marcas por grupo de ligação foi exatamente igual aos genomas simulados 11 marcas, sem nenhuma exceção, evidenciando, também, como observado para o número de grupos de ligação formado, que a DS gamética não afeta o número de marcas por grupo de ligação. Também não ocorreu nenhuma inversão de marcas, independente do tamanho e nível de distorção de segregação, e

independente das estratégias utilizadas considerarem ou não a DS em suas funções de verossimilhança. Demonstra-se também para este critério (Figuras 1 a 12), que a utilização de marcadores que apresentam DS gamética na construção de mapas de ligação é aconselhável e o descarte desses para construção do mapa seria inadequado.

#### 4.4.1.2.3. *Tamanho de Grupos de Ligação*

O tamanho de grupo de ligação foi obtido pela média aritmética dos tamanhos dos três grupos de ligação para cada população. O tamanho de cada grupo de ligação esperado após o mapeamento das populações segregantes era de 100 cM, pois, este era o tamanho de cada grupo de ligação para todos os genomas utilizados para geração das populações segregantes.

Analisando-se o efeito do tamanho de população dentro do nível de distorção de segregação (Tabelas 41, 42 e 43), verifica-se a maior proximidade do tamanho médio dos grupos ao valor esperado de 100 cM à medida que a DS gamética é mais próxima da segregação esperada de 3:1. No trabalho o nível mais próximo é de 2:1. Assim à medida que o nível de DS se distancia da frequência esperada têm-se maior viés do tamanho do grupo de ligação. Esse efeito é evidenciado em todos os tamanhos de populações quando se utiliza a função de verossimilhança que não considera a DS. Mas quando são utilizadas as funções de verossimilhança que consideram a DS este efeito não é propriamente uma regra. Por exemplo, na Tabela 37 o viés para o tamanho de população com taxa de DS de 1:1 é maior que o viés observado com a taxa de DS 7:1 com as duas estratégias que consideram a DS.

Esta característica pode ser devido às funções de verossimilhança que consideram a DS, que pode-se dizer que são mais sensíveis em relação ao número de indivíduos em cada classe genotípica. Deve-se ter em mente que na função de verossimilhança que não considera a DS, as classes recombinantes são somadas e para o cálculo das distâncias não consideram o número de indivíduos em cada classe de recombinantes. Já em relação às funções de verossimilhança que consideram a DS, leva-se em consideração o número de indivíduos observados em cada classe recombinante.



Em relação ao tamanho de população, as médias de cada estratégia (média com todas as razões de DS), nos diferentes tamanhos, demonstra que o viés nas três estratégias diminui quando o tamanho de população aumenta, com exceção observada apenas para a estratégia E2, em que na população com 400 indivíduos com 99,75 cM foi mais próximo ao esperado, de 100 cM, do que para a população com 1000 indivíduos com 101,13 cM. Uma explicação para aproximação do tamanho médio dos grupos de ligação ao valor esperado de 100 cM, à medida que se aumenta o tamanho da população segregante, é dada pela melhoria das estimativas das frequências de recombinação.

A média geral, em relação as três estratégias utilizadas em todas as populações, foi de: 96,08 cM para estratégia E1, 96,99 cM para estratégia E2 e de 97,75 cM para E3, demonstrando a disponibilidade de melhores estimativa com a utilização da estratégia E3, seguida da E2.

#### 4.4.1.2.4. *Distância Média entre Marcadores Adjacentes*

A distancia média de marcas adjacentes dentro de cada população foi obtida pela estimativa de duas médias, como descrito a seguir. Primeiramente, obteve-se uma média aritmética das distâncias entre marcas adjacentes dentro de cada grupo de ligação. Posteriormente, obteve-se uma segunda média aritmética, entre as médias aritméticas primeiramente obtidas em cada população.

Os valores de distância média entre marcas adjacentes, para os vários tamanhos de população segregantes e DS gamética utilizadas no mapeamento genético, estão apresentados nas Tabelas 41, 42 e 43. A distância média entre marcas adjacentes esperada era de 10 cM, para todas as populações segregantes geradas.

Analisando-se o efeito da distância média adjacente dentro de cada nível de distorção de segregação (Tabelas 41, 42 e 43), as mesmas conclusões observadas para tamanho de população podem ser referenciadas para distância média entre marcas adjacentes, com maior proximidade de distância média entre marcas adjacente ao valor esperado de 10 cM à medida que a DS gamética é mais próxima da segregação esperada de 3:1. Assim à medida que o nível de DS se distancia da

freqüência esperada, têm-se um maior viés da distância média entre marcas adjacentes. Ressalvando-se as mesmas considerações observadas para tamanho de grupo de ligação.

Em relação ao tamanho de população, observa-se também que o viés das distâncias médias entre marcas adjacentes tem as mesmas características observadas para tamanho médio de grupos de ligação.

A distância média geral, em relação às três estratégias utilizadas em todas as populações, foi de 9,60 cM para estratégia E1, 9,70 cM para estratégia E2 e de 9,77 cM para E3, demonstrando a disponibilidade de melhores estimativa com a utilização da estratégia E3, seguida da E2.

#### 4.4.1.2.5. *Variância das Distâncias entre Marcadores Adjacentes*

A partir das distâncias de marcas adjacentes obtidas nos grupos de ligação, foi estimada uma variância amostral. Os valores apresentados nas Tabelas 41, 42 e 43, para cada população, são as médias aritméticas das variâncias obtidas nos três grupos de ligação, nas populações simuladas.

As estimativas esperadas de variância na análise dos mapas obtidos das populações segregantes simuladas eram de zero para todas as populações, uma vez que nos genomas simulados, para a geração das populações segregantes, os marcadores estavam distribuídos de forma equidistante dentro dos três grupos de ligação.

Quanto menores forem os valores de variância, mais equidistantemente estão distribuídos os marcadores dentro do grupo de ligação. Portanto, a obtenção de valores de variância pequenos e distâncias médias entre marcas adjacentes próximos aos valores esperados nas populações, indicam boa recuperação do genoma com o mapeamento das populações segregantes.

Analisando-se o efeito do tamanho de população (Tabelas 41, 42 e 43), verifica-se uma redução da variância média das distâncias entre marcas adjacentes à medida que se aumenta o tamanho da população segregante, em todas as três estratégias de mapeamento. Por exemplo, em relação à estratégia E1 verificou-se os valores de 3,82; 2,02 e; 0,99, para as populações de 200, 400 e 1000 indivíduos respectivamente.

Pode-se observar as médias das três estratégias de mapeamento nas populações com 200, 400 e 1000 indivíduos nas Tabelas 41, 42 e 43. A estratégia E1 é a que apresentou as maiores estimativas de variâncias em todos os diferentes tamanhos de população. E nas populações com 200 e 400 indivíduos também os maiores viés das distâncias médias entre marcas adjacentes em relação ao valor esperado de 10 cM, indicando que esta estratégia foi menos eficiente que as demais na recuperação do genoma simulado.

#### 4.4.1.2.6. Estresse

O valor de estresse expressa o grau de concordância dos valores de distância entre cada par de marcas adjacentes no grupo de ligação obtido, no mapeamento das populações simuladas, com as distâncias dos respectivos pares de marcas no genoma utilizado para simulação das populações.

De acordo com o apresentado no Material e Métodos (item 3.2.4.6.), a expressão que estima o estresse é:

$$S = 100. \sqrt{\frac{\sum_{k=1}^{m-1} (d_{ok} - d_k)^2}{\sum_{k=1}^{m-1} d_{ok}^2}}$$

em que  $d_{ok}$  é a distância entre marcas adjacentes  $m_k$  e  $m_{k+1}$  no genoma original (real) e  $d_k$  é a distância entre marcas adjacentes  $m_k$  e  $m_{k+1}$  no grupo de ligação analisado ( $k= 1, \dots, m-1$ ).

Para melhor entendimento, o estresse pode ser dado em função do desvio médio das distâncias entre marcas adjacentes no grupo de ligação e aquelas no genoma original, como segue abaixo e representado a seguir:

$$S = \left( \frac{d}{d_o} \right) \cdot 100$$

em que  $d$  é o desvio médio das distâncias por intervalo no grupo de ligação analisado em relação ao genoma original com saturação de 10cM e  $d_o$  é a distância entre marcas adjacentes no genoma original (neste caso 10cM).

O valor de estresse possui significados diferentes quando se comparam genomas com níveis de saturação distintos. Por exemplo, para

uma mesma estimativa de estresse igual a 20%, o desvio médio referente a este estresse pode assumir valores de 1, 2 ou 4 cM, se estiver associado ao genoma com nível de saturação de 5, 10 ou 20 cM, respectivamente.

Os valores de estresse médio para cada população, em relação às estratégias englobando todas as razões de segregação são apresentados nas Tabelas 41, 42 e 43. Observa-se que a estratégia E2 apresenta a menor estimativa de estresses com população de 200 e 400 indivíduos e a estratégia E3 é a com menor estimativa em população constituída com 1000 indivíduos. Os valores médios de todos os tamanhos de população e razões de segregação em relação às estratégias são: 75,83%; 70,86% e; 73,23% para as estratégias E1, E2 e E3 respectivamente. Estas estimativas expressam um valor de desvio médio entre marcas adjacentes de 7,69cM, 7,18cM e 7,42cM, respectivamente. Demonstra-se assim por meio das estimativas de estresse que as estratégias que consideram a DS (E2 e E3) são relativamente mais eficientes que a estratégia E1 que não considera a função de verossimilhança. Se for efetuada a diferença entre o desvio médio observado para a estratégia E1, em relação aos desvios observados para as estratégias E2 e E3, ter-se-á:

$$E1 - E2 = 7,69 - 7,18 = 0,51 \text{ cM}$$

e

$$E1 - E3 = 7,69 - 7,42 = 0,27 \text{ cM.}$$

Então, como trata-se de desvios, percebe-se que, por exemplo, a estratégia E2 pode apresentar um valor distância mais próximo ao valor da distância verdadeira acima de 1cM ( $0,51\text{cM} + 0,51\text{cM} = 1,02\text{cM}$ ) em relação à estratégia E1. Enquanto que para estratégia E3 valor de 0,54cM.

Os resultados encontrados ratificam que as estratégias que consideram a DS gamética em suas funções de verossimilhança são extremamente úteis na obtenção de mapas genéticos mais acurados, particularmente nos mapeamentos finos que objetivam a clonagem posicional de genes.

Na literatura tem sido detectados QTLs em intervalos de 1 a 10 cM entre marcadores (Tanksley et al., 1995). Este intervalo pode corresponder entre centenas e milhares de quilo pares de base (kpb) na estrutura física do cromossomo. Intervalos dessa magnitude geralmente abrigam dezenas ou

centenas de genes. Em *Arabidopsis*, uma distância de mapa de 1cM corresponde em média a uma distância física de 250 kpb (Grattapaglia e Ferreira, 2006). Para *Eucalyptus*, 1cM de recombinação corresponde em média a 500 kpb (Grattapaglia e Bradshaw, 1994).

Nos últimos anos, vários genes responsáveis por QTL foram clonados em plantas, mais comumente em arroz, tomate, *Arabidopsis* e trigo, pela estratégia de clonagem posicional. Um exemplo é o mapeamento de alta resolução do QTL fw2.2, que controla peso de fruto em tomate, baseado na utilização de 51 linhagens quase isogênicas para a construção de mapa de alta resolução, bem como para a avaliação fenotípica, permitindo localizar o QTL em um intervalo de 0,13 cM, correspondendo a uma distância física de cerca de 150 kpb (Alpert e Tanksley, 1996).

#### 4.4.1.2.7. Considerações Gerais

O presente trabalho demonstra um fator importante, desejável e adequado, no estudo e entendimento de DS considera mais que um único marcador com DS no genoma. Vogl & Xu (2000) relatam exemplos, no passado que consideravam apenas um único marcador, como o trabalho de Sorensen (1967).

Eujayl et al. (1997) relatam que, em populações  $F_2$ , marcadores com DS causam viés nas estimativas de distâncias entre marcas na construção de mapas genéticos, corroborando com os resultados obtidos neste trabalho.

Oliveira et al. (2004) estudaram os efeitos da inclusão de marcadores com DS em *Citrus*, em  $F_1$ , e observaram que marcadores com distorção de segregação entre marcas em um dado GL geralmente afetavam as distâncias genéticas. O número de grupos de ligação foram afetados com a exclusão de marcadores com DS. Pode-se inferir, com os resultados encontrados no presente trabalho, que a observação da alteração do número de GL é esperado se for realizada a exclusão dos marcadores com distorção de segregação, como observado por Oliveira et al. (2004). Pois ainda neste estudo verificou-se que marcadores com DS podem formar um grupo de ligação inteiro, ou marcas sem DS podem estar presentes em um dado grupo de ligação, devido estarem ligadas a marcadores com DS, em que a exclusão das marcas com DS farão com que as marcas sem DS não

apresentem-se ligadas, em nenhum grupo ou algumas formarem outro grupo de ligação.

Diante de todos os resultados observados neste estudo, pode-se inferir que as estratégias E2 e E3, são mais eficientes que a estratégia E1. E com ligeira superioridade para a estratégia E2, a qual se faz a indicação das marcas que apresentam DS gamética, devido a fatores biológicos e que considera as marcas adjacentes a estas, com DS devido à ligação fatorial e as demais como se não apresentassem DS. O que não é verdade, comprovado com exemplos da literatura e por meio das funções de verossimilhança deduzidas neste trabalho. Deve-se, neste contexto, ser considerado um exemplo, em que uma marca apresenta DS e que primeiramente será estimada a distância entre esta marca e outra sem considerar a DS e a seguir estima-se a distância considerando a distorção de segregação, como a seguir:

Uma população com 400 indivíduos com um distância exata de 10,9 cM com a seguinte configuração genotípica observada;

A-B- = 330 indivíduos

A-bb = 20 indivíduos

aaB- = 10 indivíduos

aabb = 40 indivíduos

Ter-se-á a estimativa de distância sem considerar a DS de 9,5 cM. E a estimativa de distância considerando a DS na função de verossimilhança de 10,9 cM.

Assim é evidenciado que os resultados encontrados pela estratégia E3 no presente trabalho é mais confiável que as demais estratégias, embora tenha apresentado, às vezes, resultados menos adequados. Isto pode ser entendido e comprovado pela observação das funções de verossimilhança. Pois quando não se considera a DS em  $F_2$  dominante, as classes genotípicas recombinantes são somadas para o cálculo da distância. O que não ocorre com a função de verossimilhança com DS, que considera o número de indivíduos observado em cada classe genotípica. Fato que não é adequado, pois com DS os valores esperados nessas classes genotípicas recombinantes são diferentes. Desta forma, a estratégia E3, que leva este fato em consideração, é sensível ao desvio do número de indivíduos para

cada classe genotípica recombinante e/ou paternas. Com a utilização da estratégia E3, quando o software estima a distorção de segregação observada para todas as marcas, as estimativas de distâncias tendem a ser mais precisas.

Diante do exposto, a estratégia E3 pode ser usada rotineiramente com mais confiabilidade na construção de mapas genéticos com DS. E também mesmo sem a presença nítida de DS em populações. Uma vez que marcas podem apresentar razão de segregação, por exemplo, de 3,5:1 e será estimada a distância considerando, por exemplo, 3:1 devido o teste de qui-quadrado não rejeitar a hipótese de 3:1.

Uma questão também interessante para utilização da estratégia E3 é que esta também dispensa o uso do teste de qui-quadrado para a verificação da segregação das marcas.

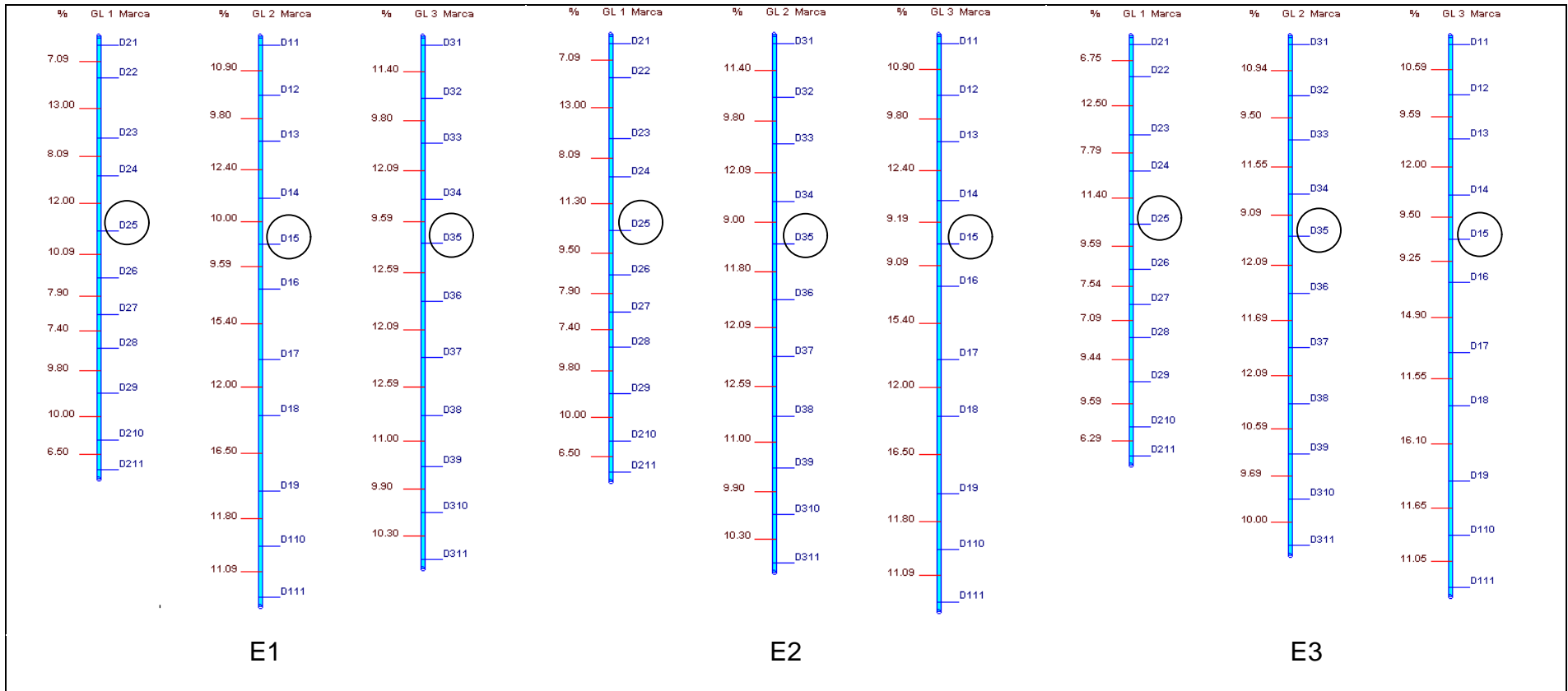


Figura 1 – População F<sub>2</sub> dominante com distorção de segregação de 1:1 nos locos D15, D25 e D35, com tamanho de população de 200 indivíduos.



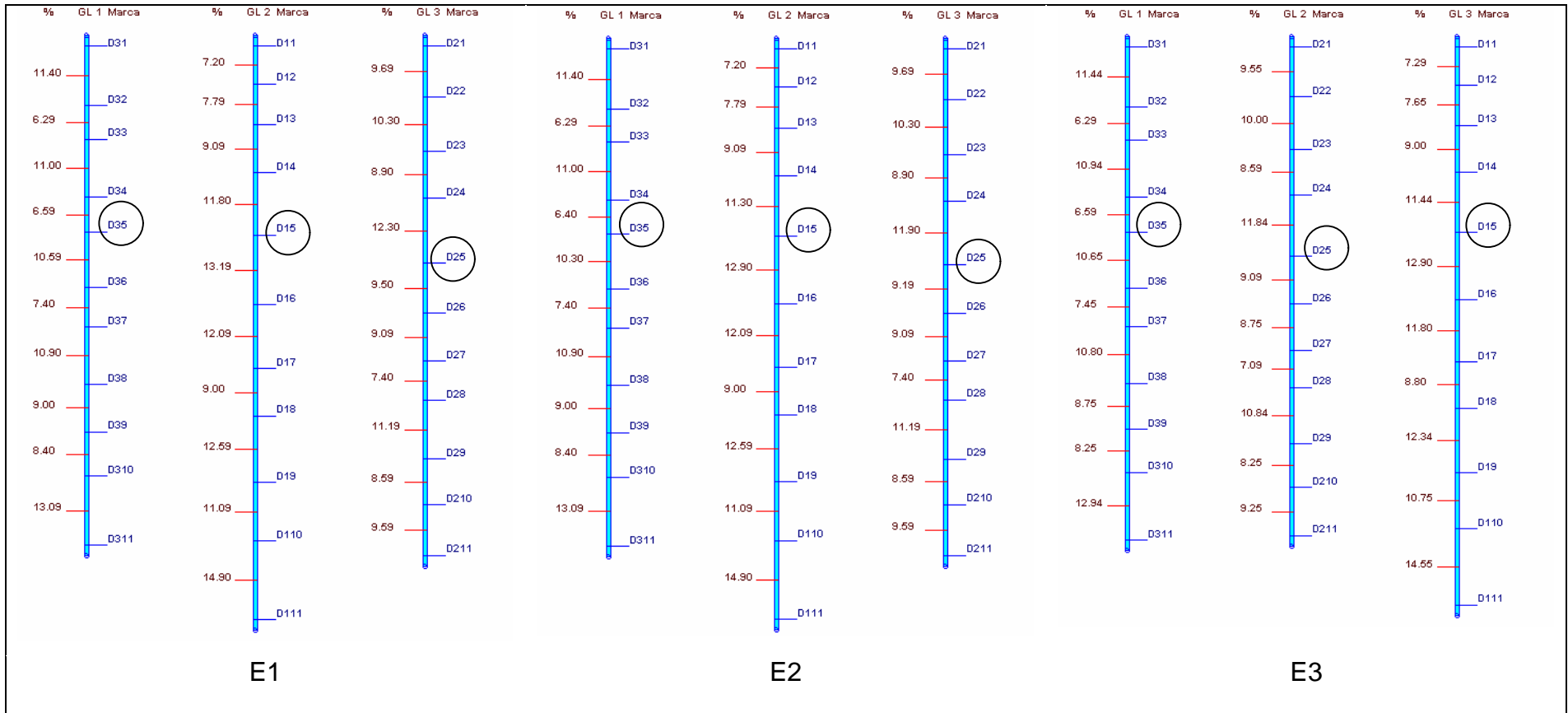


Figura 2 – População  $F_2$  dominante com distorção de segregação de 2:1 nos locos D15, D25 e D35, com tamanho de população de 200 indivíduos

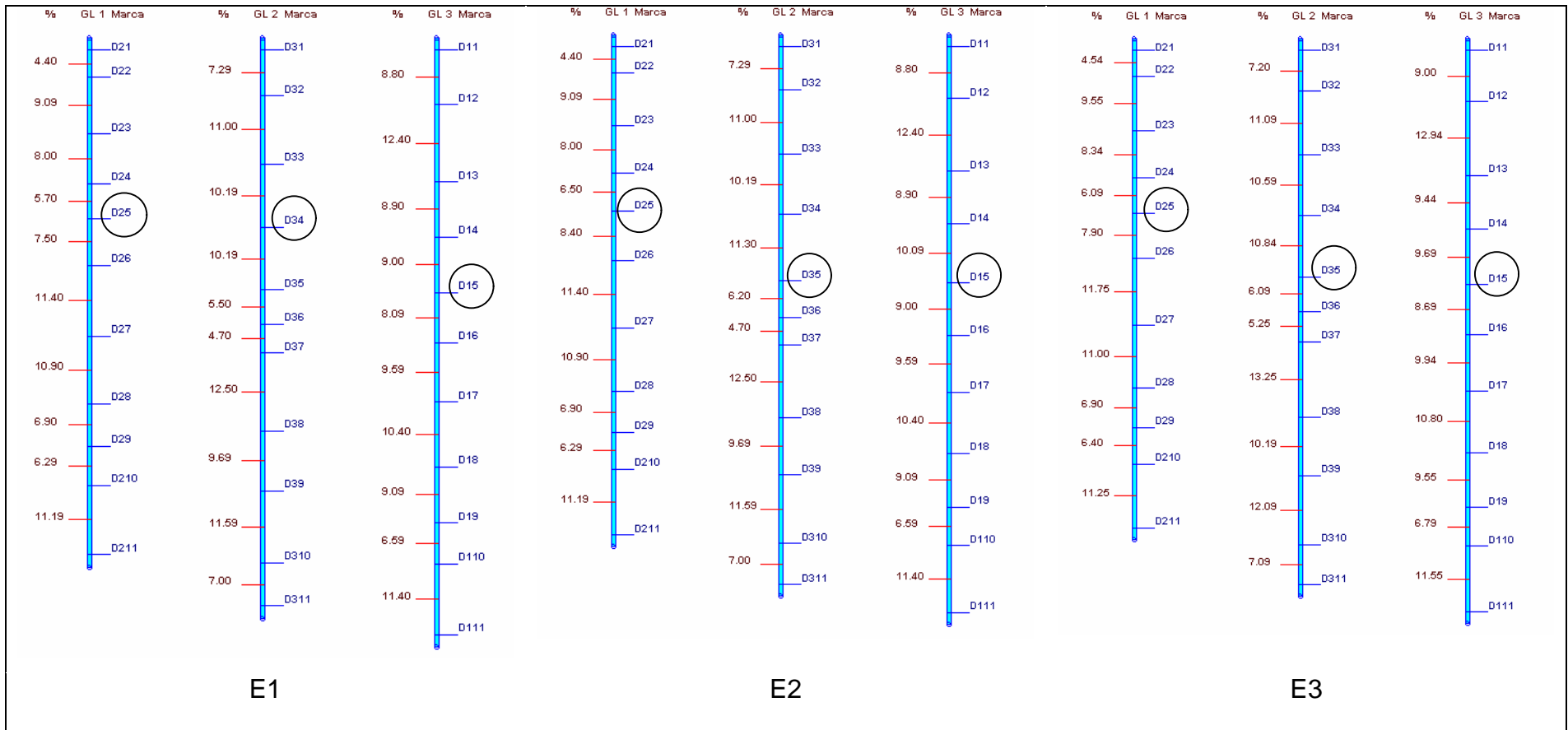


Figura 3 – População F<sub>2</sub> dominante com distorção de segregação de 7:1 nos locos D15, D25 e D35, com tamanho de população de 200 indivíduos

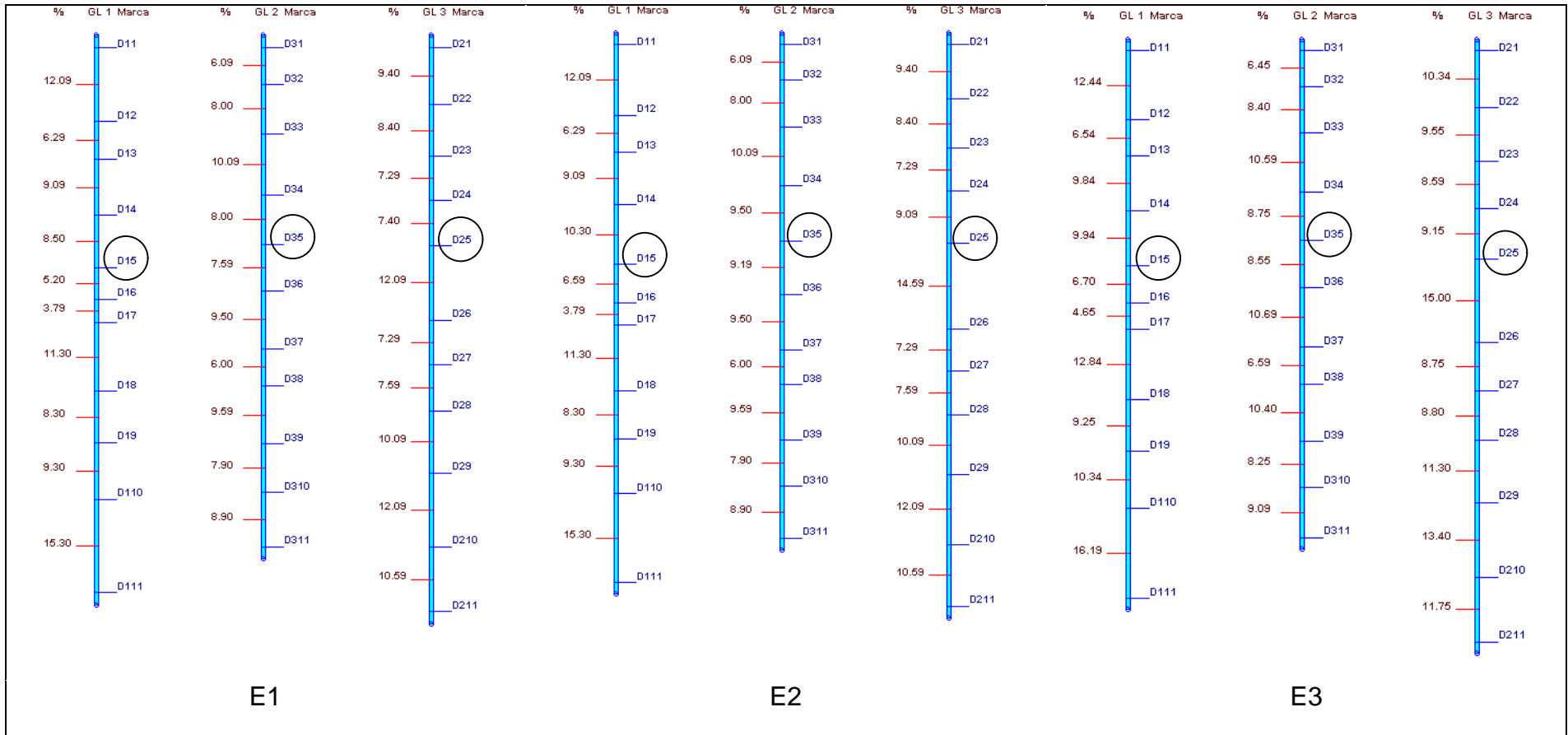


Figura 4 – População F<sub>2</sub> dominante com distorção de segregação de 10:1 nos locos D15, D25 e D35, com tamanho de população de 200 indivíduos

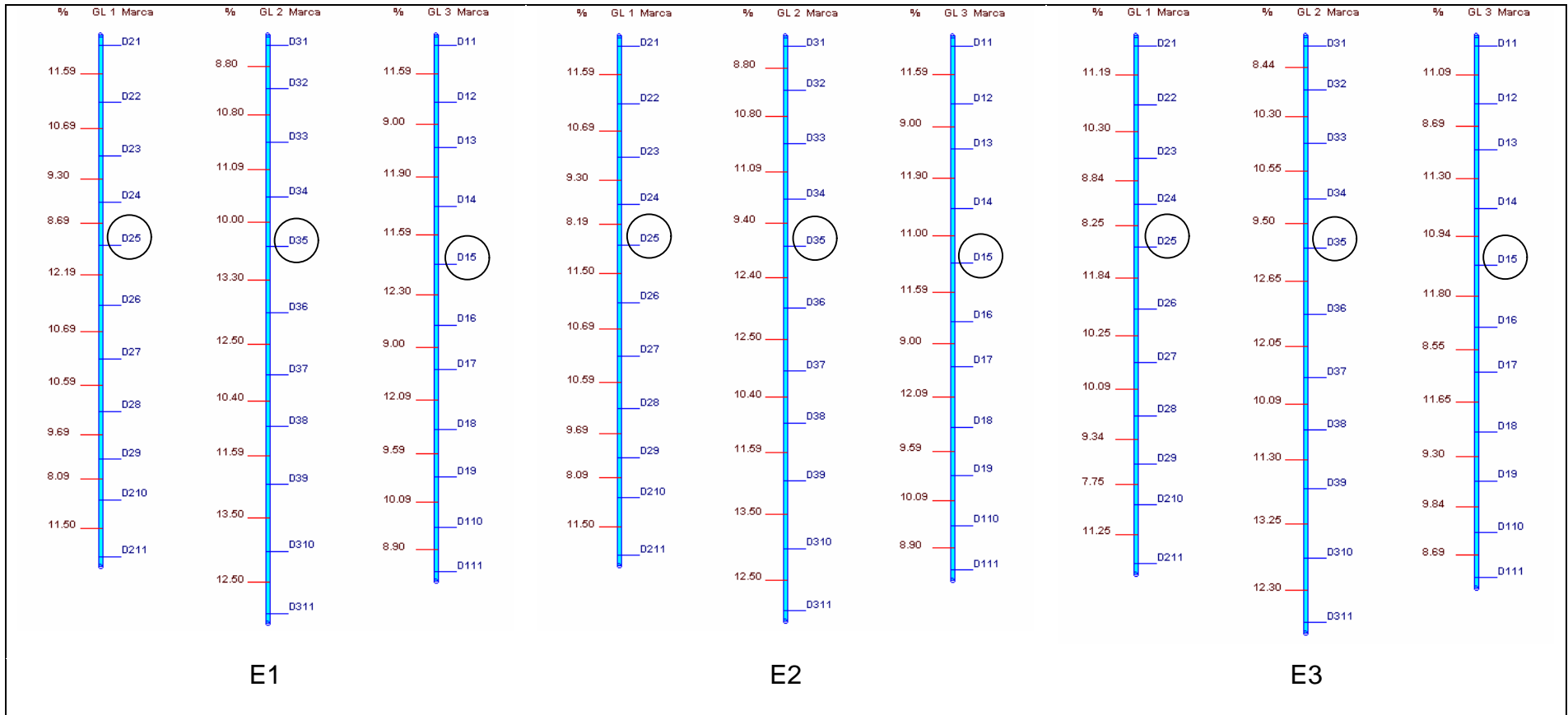


Figura 5 – População  $F_2$  dominante com distorção de segregação de 1:1 nos locos D15, D25 e D35, com tamanho de população de 400 indivíduos.

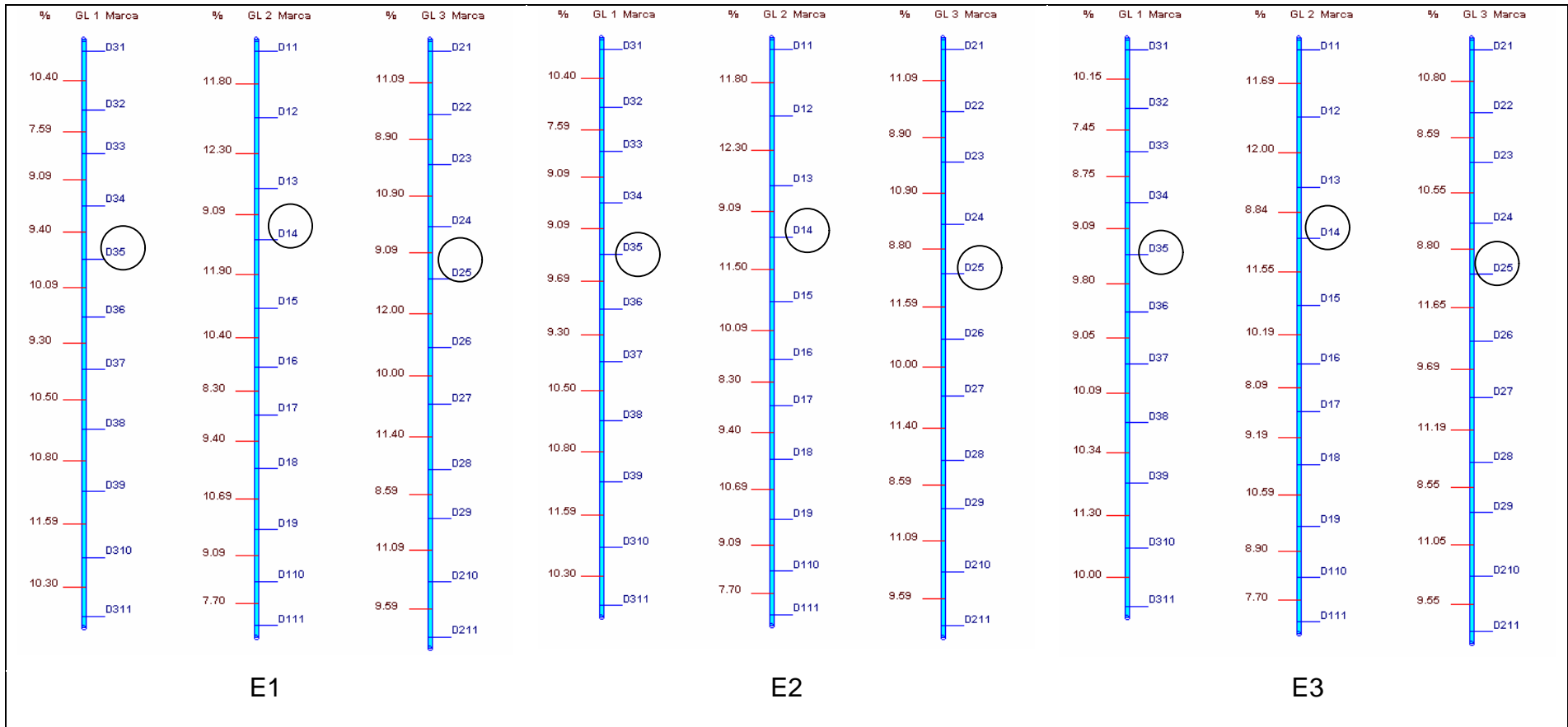


Figura 6 – População F<sub>2</sub> dominante com distorção de segregação de 2:1 nos locos D15, D25 e D35, com tamanho de população de 400 indivíduos.

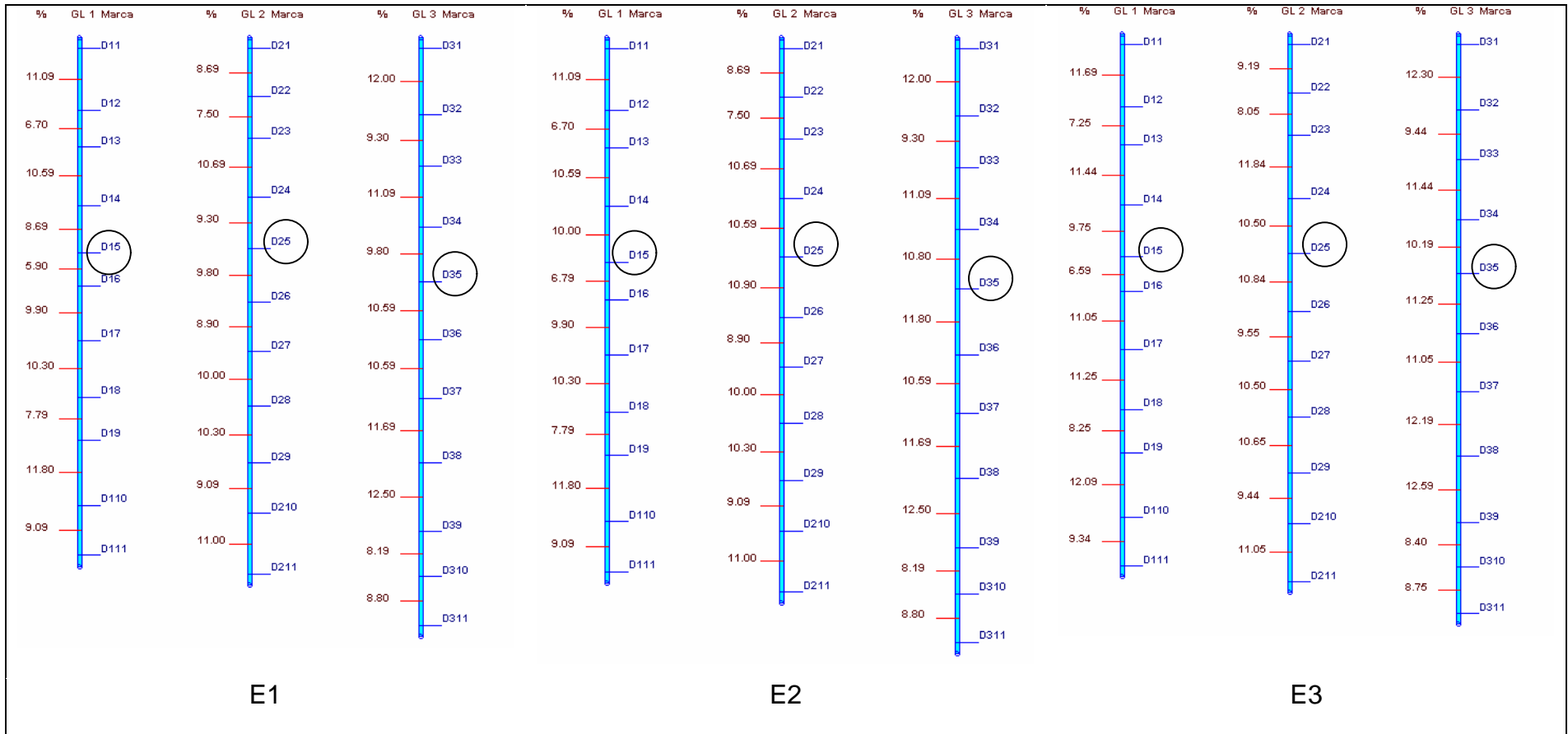


Figura 7 – População F<sub>2</sub> dominante com distorção de segregação de 7:1 nos locos D15, D25 e D35, com tamanho de população de 400 indivíduos.

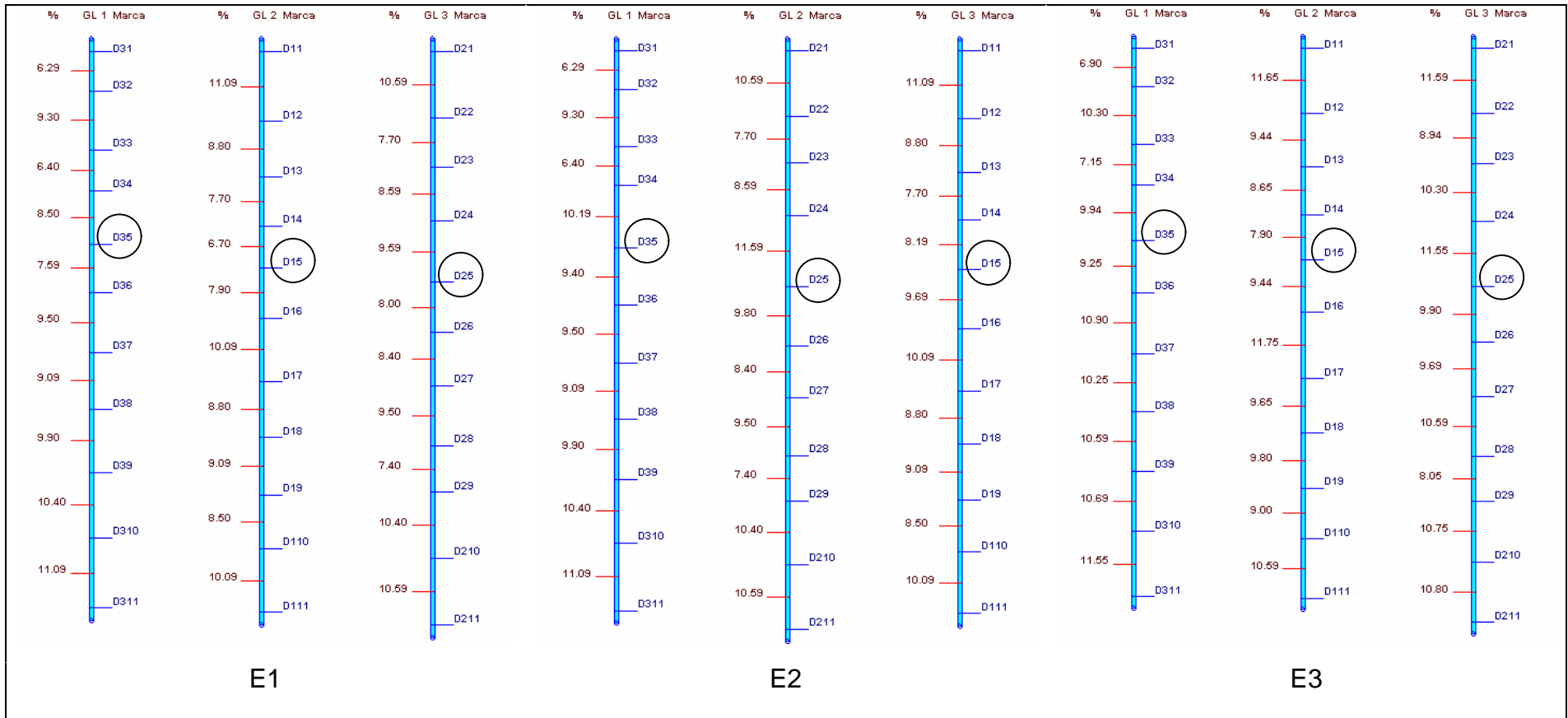


Figura 8 – População F<sub>2</sub> dominante com distorção de segregação de 10:1 nos locos D15, D25 e D35, com tamanho de população de 400 indivíduos.

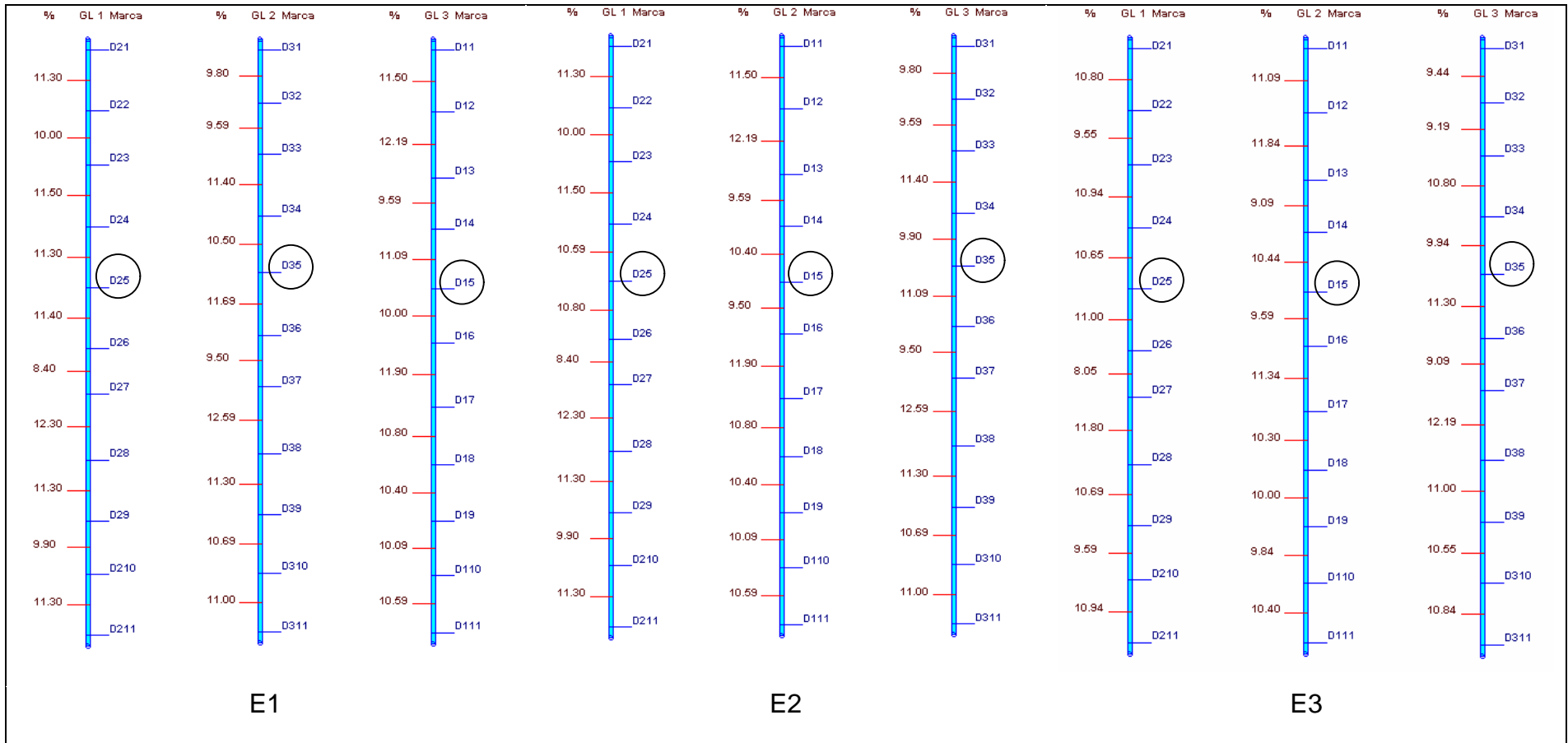


Figura 9 – População F<sub>2</sub> dominante com distorção de segregação de 1:1 nos locos D15, D25 e D35, com tamanho de população de 1000 indivíduos.



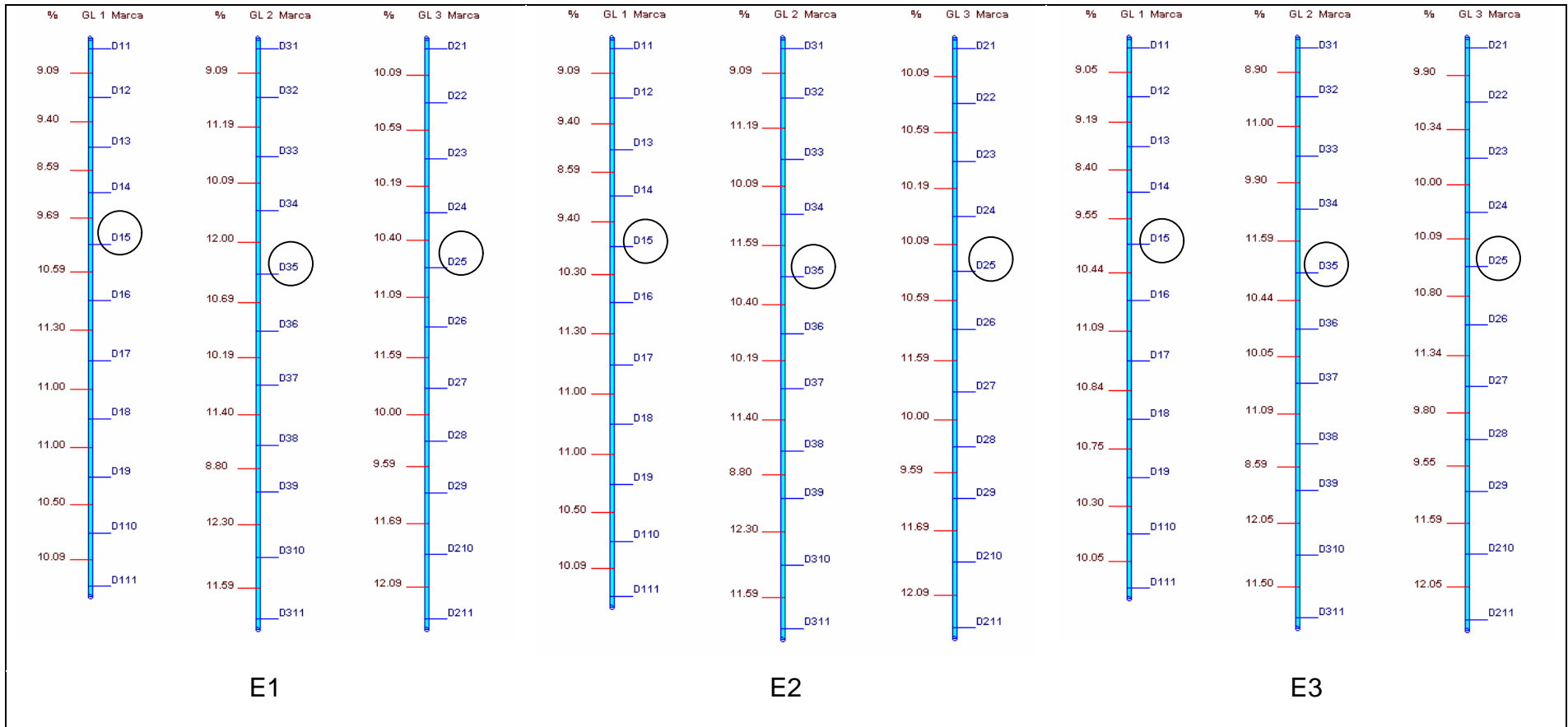


Figura 10 – População F<sub>2</sub> dominante com distorção de segregação de 2:1 nos locos D15, D25 e D35, com tamanho de população de 1000 indivíduos.

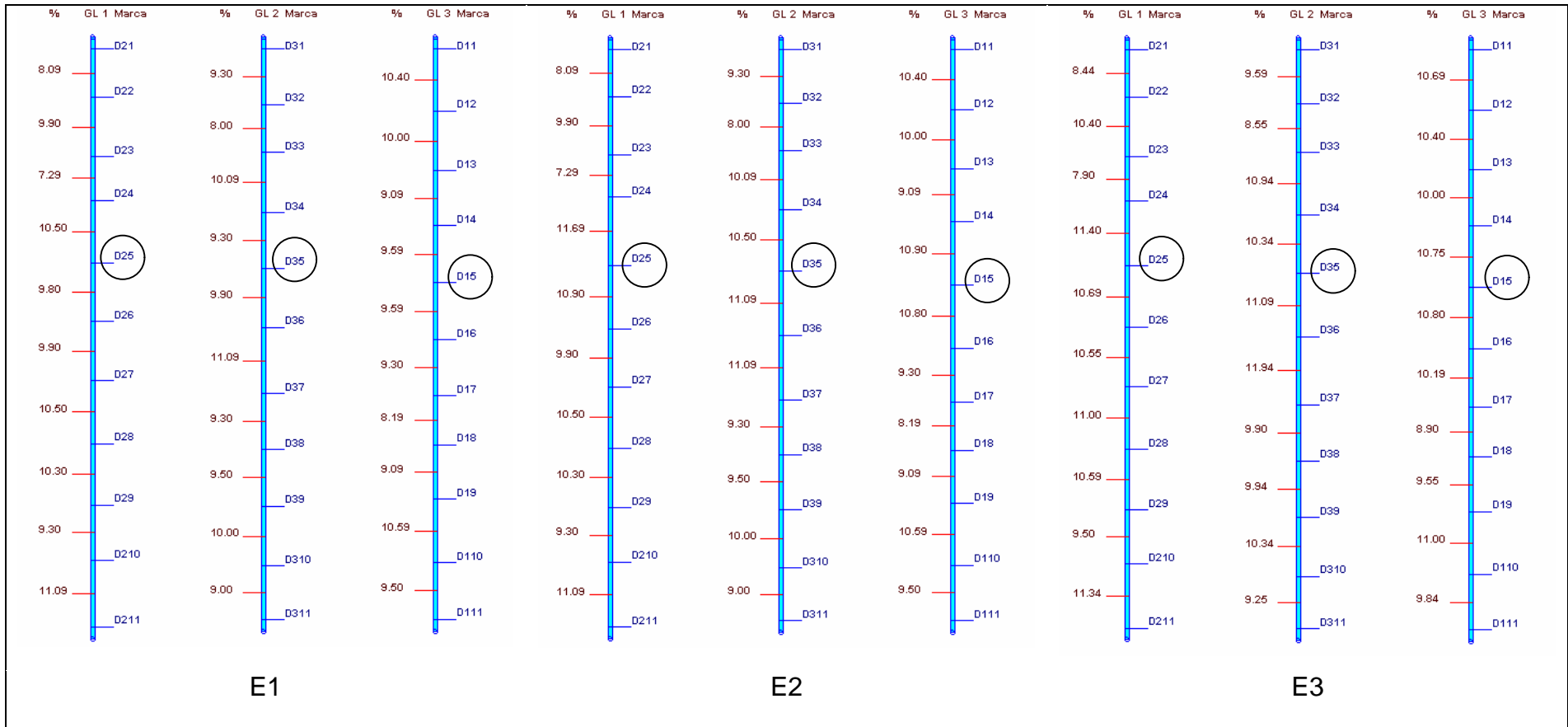


Figura 11 – População F<sub>2</sub> dominante com distorção de segregação de 7:1 nos locos D15, D25 e D35, com tamanho de população de 1000 indivíduos.

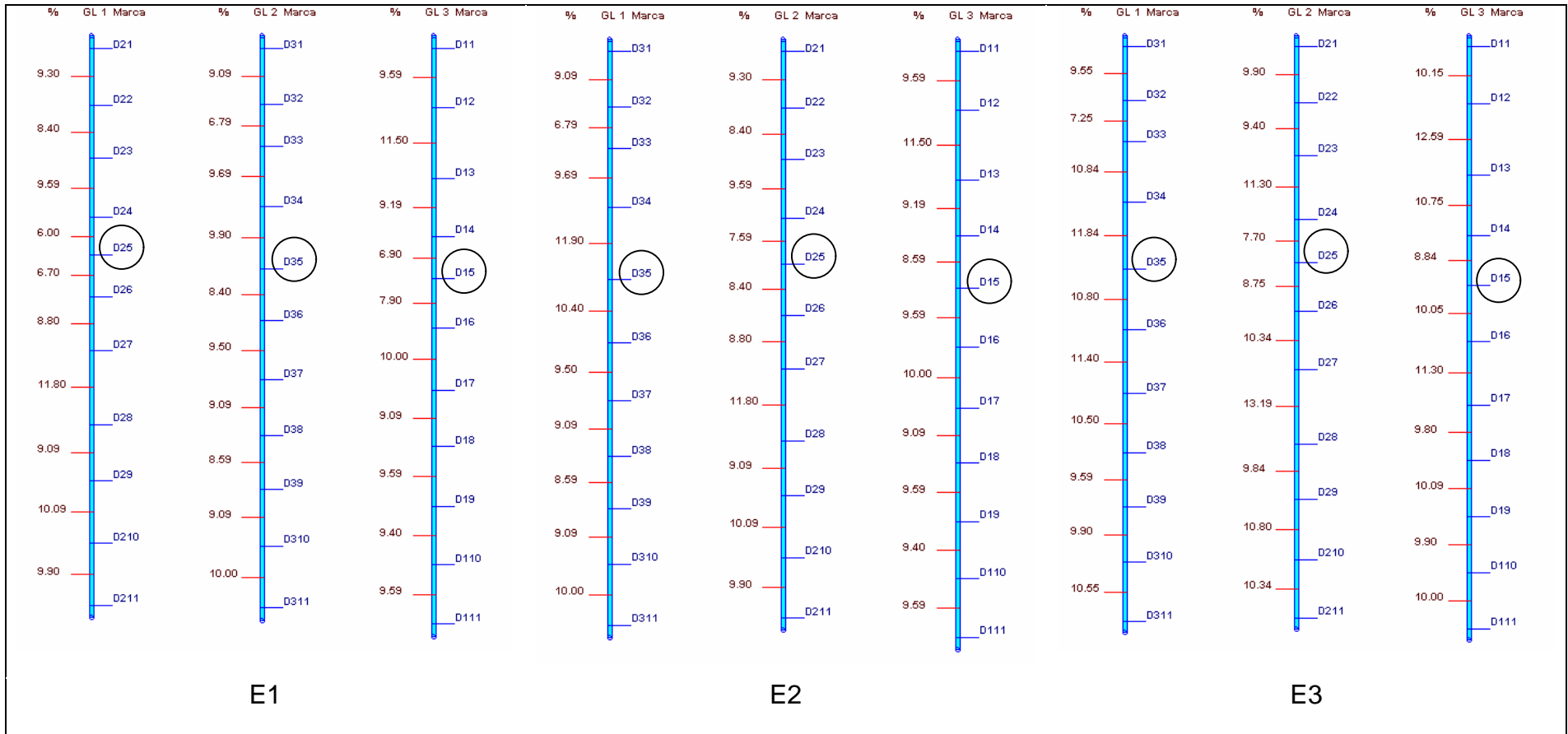


Figura 12 – População F<sub>2</sub> dominante com distorção de segregação de 10:1 nos locos D15, D25 e D35, com tamanho de população de 1000 indivíduos.

Tabela 41 – Valores de distância média, variância, tamanho médio de GL e estresse, em população  $F_2$  dominante, com 200 indivíduos em diferentes razões de segregação e estratégias de mapeamento

Estratégia	SEG.	TAM. GL	MED DIST	VARIÂNCIA	ESTRESSE
E1	1:1	107,60	10,76	3,82	71,33
E2	1:1	106,27	10,63	3,90	99,00
E3	1:1	103,85	10,39	3,55	99,00
E1	2:1	100,03	10,00	4,45	99,00
E2	2:1	99,37	9,94	4,32	99,00
E3	2:1	98,00	9,80	4,20	70,80
E1	7:1	88,47	8,85	5,27	99,00
E2	7:1	90,30	9,03	5,03	99,00
E3	7:1	91,98	9,20	5,43	99,00
E1	10:1	87,73	8,77	5,74	77,34
E2	10:1	91,23	9,12	6,08	76,65
E3	10:1	97,75	9,78	6,22	76,79
E1	MED	89,30	8,93	6,84	82,76
E2	MED	90,08	9,01	6,42	82,24
E3	MED	91,19	9,12	6,73	82,34

E1 - desconsidera a DS na função de verossimilhança; E2 - considera a DS e pressupõe conhecer as marcas com DS e a taxa esperada; E3 - considera a DS estimando o valor de distorção observado em todos os marcadores estudados.

Tabela 42 – Valores de distância média, variância, tamanho médio de GL e estresse em população  $F_2$  dominante com 400 indivíduos em diferentes razões de segregação e estratégias de mapeamento

Estratégia	SEG.	TAMANHO GL	MED DIST	VARIÂNCIA	ESTRESSE
E1	1:1	107,90	10,79	2,02	99,00
E2	1:1	106,57	10,66	1,88	99,00
E3	1:1	103,83	10,38	1,95	99,00
E1	2:1	100,83	10,08	1,72	99,00
E2	2:1	100,13	10,01	1,67	99,00
E3	2:1	98,43	9,84	1,62	99,00
E1	7:1	97,27	9,73	2,27	15,07
E2	7:1	99,53	9,95	2,20	14,83
E3	7:1	102,68	10,27	2,43	15,11
E1	10:1	89,23	8,92	1,93	99,00
E2	10:1	92,77	9,28	1,85	70,75
E3	10:1	99,22	9,92	1,73	99,00
E1	MED	98,81	9,88	1,99	78,02
E2	MED	99,75	9,98	1,90	70,89
E3	MED	101,04	10,10	1,93	78,03

E1 - desconsidera a DS na função de verossimilhança; E2 - considera a DS e pressupõe conhecer as marcas com DS e a taxa esperada; E3 - considera a DS estimando o valor de distorção observado em todos os marcadores estudados.

Tabela 43 – Valores de distância média, variância, tamanho médio de GL e estresse em população F<sub>2</sub> dominante com 1000 indivíduos em diferentes razões de segregação e estratégias de mapeamento

Estratégia	SEG.	TAMANHO GL	MED DIST	VARIÂNCIA	ESTRESSE
E1	1:1	108,33	10,83	0,99	99,00
E2	1:1	107,10	10,71	1,00	69,89
E3	1:1	104,15	10,42	0,94	69,49
E1	2:1	105,37	10,54	0,97	68,89
E2	2:1	104,67	10,47	0,95	68,89
E3	2:1	103,45	10,35	0,95	68,81
E1	7:1	95,87	9,59	0,83	99,00
E2	7:1	98,27	9,83	1,19	99,00
E3	7:1	101,98	10,20	0,94	99,00
E1	10:1	90,90	9,09	1,72	99,00
E2	10:1	94,47	9,45	1,21	70,35
E3	10:1	94,47	9,45	1,21	70,35
E1	MED	100,12	10,01	1,13	66,72
E2	MED	101,13	10,11	1,09	59,44
E3	MED	101,01	10,10	1,01	59,32

E1 - desconsidera a DS na função de verossimilhança; E2 - considera a DS e pressupõe conhecer as marcas com DS e a taxa esperada; E3 - considera a DS estimando o valor de distorção observado em todos os marcadores estudados.

#### 4.4.2. Distorção Genotípica

Neste tópico é abordado o emprego de metodologia para o estabelecimento de mapas genéticos, considerando a existência de marcadores distorção de segregação, causada por distorção genotípica.

#### 4.4.2.1. Mapeamento Genético

Gerou-se por meio do módulo de simulação do programa computacional GQMOL/UFV (CRUZ, 2006), dados de uma população  $F_2$ , com 1000 indivíduos, considerando o estudo de marcadores dominantes. O genoma original com três GL, e 10 marcas/GL equidistantes a 10cM. Após avaliou-se a hipótese de segregação mendeliana, nos dados iniciais ainda sem nenhum fator causador da DS genotípica Tabela 44, evidenciando-se a eficácia do processo de simulação em gerar informações das 30 marcas moleculares adequadas a segregação pressuposta numa população  $F_2$ .

Tabela 44 – Segregação dos marcadores em população  $F_2$  com 1000 indivíduos, e teste de comprovação da segregação 3:1

Marcadores	Observado (A-)	Observado (aa)	$\chi^2$	Probabilidade (%)	
x1	261	739	0,645	42,1786	ns
x2	253	747	0,048	82,6581	ns
x3	250	750	0,	100,0	ns
x4	235	765	1,2	27,3322	ns
x5	251	749	0,005	94,1783	ns
x6	256	744	0,192	66,1257	ns
x7	262	738	0,768	38,0837	ns
x8	250	750	0,0	100,0	ns
x9	259	741	0,432	51,1009	ns
x10	252	748	0,021	88,3875	ns
x11	264	736	1,045	30,6584	ns
x12	263	737	0,901	34,2424	ns
x13	262	738	0,768	38,0837	ns
x14	257	743	0,261	60,9206	ns
x15	260	740	0,533	46,5209	ns
x16	260	740	0,533	46,5209	ns
x17	256	744	0,192	66,1257	ns
x18	246	754	0,085	77,0196	ns
x19	257	743	0,261	60,9206	ns
x20	245	755	0,133	71,5001	ns
x21	249	751	0,005	94,1783	ns
x22	259	741	0,432	51,1009	ns
x23	262	738	0,768	38,0837	ns
x24	259	741	0,432	51,1009	ns
x25	236	764	1,045	30,6584	ns
x26	229	771	2,352	12,5122	ns
x27	224	776	3,605	5,7595	ns
x28	228	772	2,581	10,8131	ns
x29	243	757	0,261	60,9206	ns
x30	233	767	1,541	21,4419	ns

ns: não significativo a 5% de probabilidade

Após a aplicação dos testes de segregação, estimou-se a porcentagem de recombinação entre pares de marcas pelo método da máxima verossimilhança. Os critérios utilizados no agrupamento foram  $r_{\max} = 30\text{cM}$  e LOD mínimo de 3, para inferir a ligação entre dois locos. E para o ordenamento se utilizou o método da SARF. Assim, obteve-se o mapa genético conforme apresentado na Figura 13. Todos os grupos de ligações foram recuperados, tendo em vista o genoma originalmente simulado, e o ordenamento também foi estabelecido de forma correta.

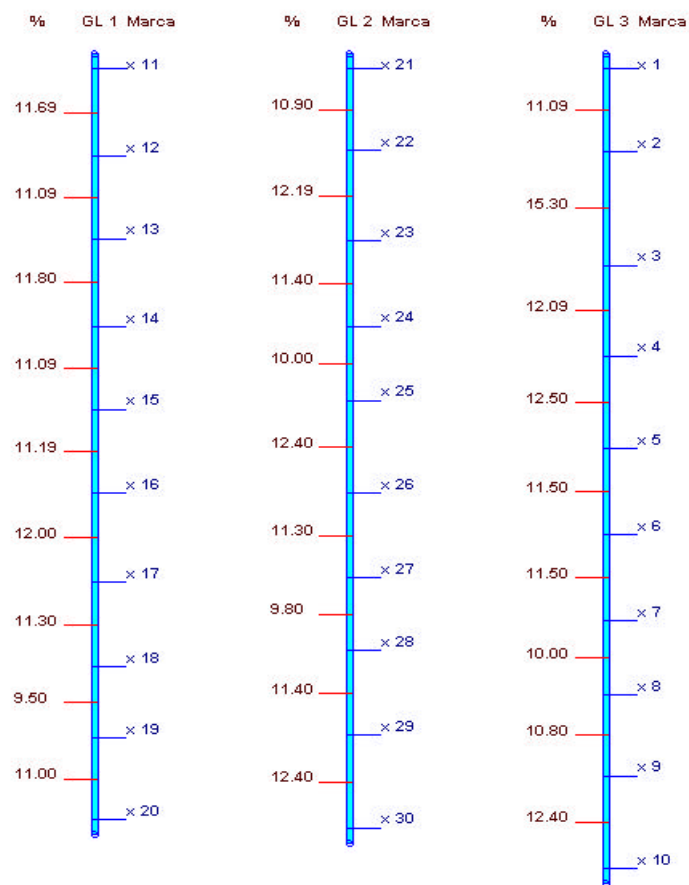


Figura 13 – População  $F_2$  dominante sem a presença de DS genotípica, com tamanho de população de 1000 indivíduos.

A partir do mesmo conjunto de dados, foi simulada a ocorrência de seleção ou inviabilidade genotípica de forma a pressupor que apenas algumas marcas apresentavam DS, sem afetar as marcas adjacentes. Neste estudo, considerou-se que apenas as marcas 5, 15 e 25 estavam sob seleção, de forma que suas segregações passaram a ser de 2:1 ao invés do esperado 3:1.



O novo conjunto de dados foi, novamente submetido ao teste de segregação de marca individuais, proporcionando os resultados apresentados na Tabela 45. O processo foi feito de forma que o tamanho da população foi mantido inalterado, mas as marcas especificadas tiveram, de fato, sua razão de segregação modificada, permanecendo todas as outras, inclusive, as adjacentes, com as mesmas razões de segregação. O teste de qui-quadrado foi capaz de detectar as alterações nas razões de segregação conforme visualizado na Tabela 45.

Tabela 45 – Segregação dos marcadores em população F<sub>2</sub> com 1000 indivíduos, e teste de comprovação da segregação 3:1 e 2:1

Marcas	Obs (A-)	Obs (aa)	Segregação 3:1			Segregação 2:1	
			$\chi^2$	Probabilidade (%)	$\chi^2$	Probabilidade (%)	
x1	261	739	0,645	42,1786 ns	23,545	0,0001 *	
x2	253	747	0,048	82,6581 ns	29,041	0,0 *	
x3	250	750	0,0	100,0 ns	31,25	0,0 *	
x4	235	765	1,2	27,3322 ns	43,513	0,0 *	
x5	334	666	37,632	0,0 *	0,002	96,4329 ns	
x6	256	744	0,192	66,1257 ns	26,912	0,0 *	
x7	262	738	0,768	38,0837 ns	22,898	0,0002 *	
x8	250	750	0,	100,0 ns	31,25	0,0 *	
x9	259	741	0,432	51,1009 ns	24,865	0,0001 *	
x10	252	748	0,021	88,3875 ns	29,768	0,0 *	
x11	264	736	1,045	30,6584 ns	21,632	0,0003 *	
x12	263	737	0,901	34,2424 ns	22,261	0,0002 *	
x13	262	738	0,768	38,0837 ns	22,898	0,0002 *	
x14	257	743	0,261	60,9206 ns	26,221	0,0 *	
x15	334	666	37,632	0,0 *	0,002	96,4329 ns	
x16	260	740	0,533	46,5209 ns	24,2	0,0001 *	
x17	256	744	0,192	66,1257 ns	26,912	0,0 *	
x18	246	754	0,085	77,0196 ns	34,322	0,0 *	
x19	257	743	0,261	60,9206 ns	26,221	0,0 *	
x20	245	755	0,133	71,5001 ns	35,113	0,0 *	
x21	249	751	0,005	94,1783 ns	32,005	0,0 *	
x22	259	741	0,432	51,1009 ns	24,865	0,0001 *	
x23	262	738	0,768	38,0837 ns	22,898	0,0002 *	
x24	259	741	0,432	51,1009 ns	24,865	0,0001 *	
x25	334	666	37,632	0,0 *	0,002	96,4329 ns	
x26	229	771	2,352	12,5122 ns	48,985	0,0 *	
x27	224	776	3,605	5,7595 ns	53,792	0,0 *	
x28	228	772	2,581	10,8131 ns	49,928	0,0 *	
x29	243	757	0,261	60,9206 ns	36,721	0,0 *	
x30	233	767	1,541	21,4419 ns	45,301	0,0 *	

\* valor abaixo do nível crítico testado a 5%, para segregação esperada de 3:1; ns não significativo.

Após a aplicação dos testes de segregação, estimou-se a porcentagem de recombinação entre pares de marcas pelo método da máxima verossimilhança, adotando as três estratégias citadas anteriormente, gerando, assim, três mapas genéticos, em cada população com determinado valor de DS, para os quais foi utilizada a distância em Morgan e a propriedade transitiva para a determinação dos grupos de ligação. Os critérios utilizados no agrupamento foram  $r_{\max} = 30\text{cM}$  e LOD mínimo de 3, para inferir a ligação entre dois locos. E para o ordenamento se utilizou o método da SARF. Com intuito de se estudar os efeitos de marcadores com DS, todas as marcas foram utilizadas no mapeamento.

Verifica-se através da observação da Figura 14 que a estratégia E1 foi incapaz de recuperar todo o grupo de ligação do genoma original, destacando o fato de que GL2 não incluiu a marca 25, considerada não ligada. Valores de distâncias em relação às marcas com segregação distorcidas também apresentaram considerável viés.

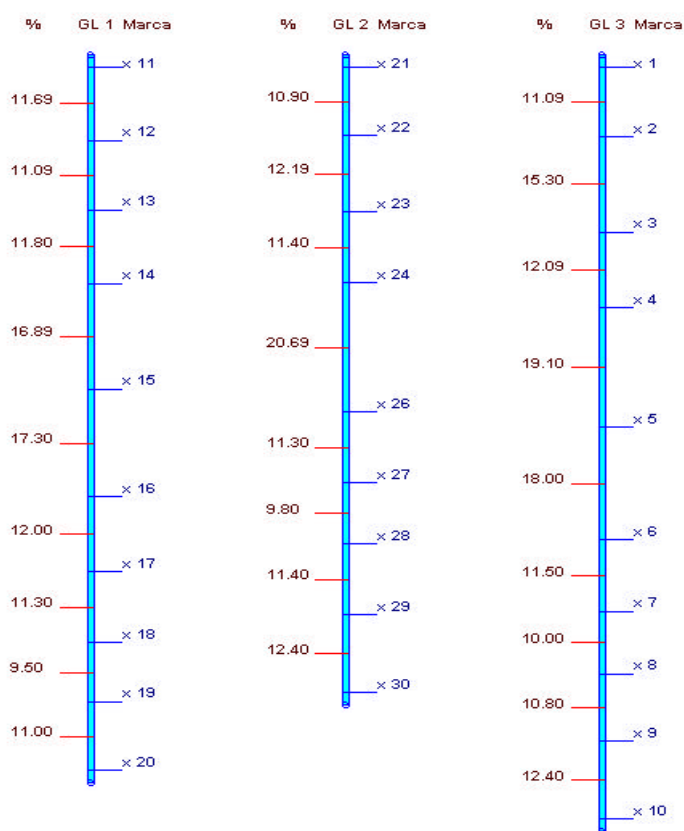


Figura 14 – População F<sub>2</sub> dominante com distorção de segregação de 2:1 nos locos X5, X15 e X25, com tamanho de população de 1000 indivíduos – Estratégia E1.

A estratégia E2 foi capaz de recuperar todos os grupos de ligação do genoma original, devendo, portanto ser preferida. Valores de distâncias em relação às marcas com segregação distorcidas também apresentaram considerável viés, apesar de ligeiramente inferiores aos obtidos com a estratégia E1.

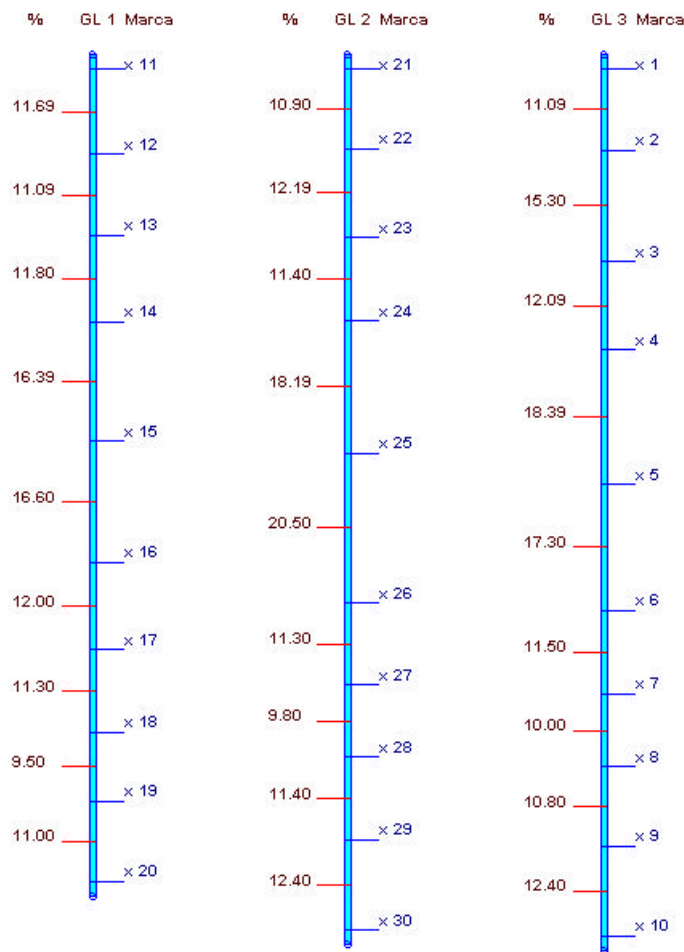


Figura 15 – População F<sub>2</sub> dominante com distorção de segregação de 2:1 nos locos X5, X15 e X25, com tamanho de população de 1000 indivíduos – Estratégia E2.

A metodologia E3 proporcionou resultado similar ao descrito para a metodologia E1 e, portanto, apresenta-se, neste exemplo, os mesmos problemas já descrito.

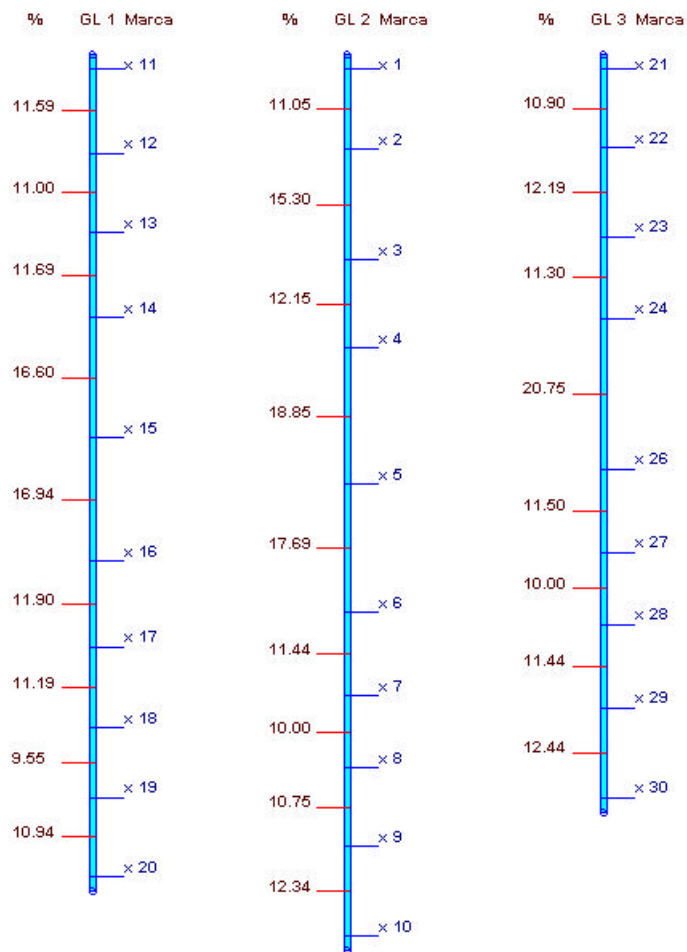


Figura 16 – População F<sub>2</sub> dominante com distorção de segregação de 2:1 nos locos X5, X15 e X25, com tamanho de população de 1000 indivíduos – Estratégia E3.

#### 4.4.2.2. Considerações Gerais

Deve-se ressaltar que este foi apenas um estudo de caso, e com uma razão de DS genotípica de 2:1, próxima ao esperado de 3:1, o que demonstra que o mapeamento com DS genotípica, mesmo com proximidade da razão esperada, pode causar problemas na construção de mapas de ligação. Por exemplo, até mesmo a incorreta independência de marcas no mapa genético como observado no GL3 em relação à marca x35. Em outras situações não apresentadas, mas que podem comprovadas pelos exemplos demonstrados no item 3.3.2. por meio das Tabelas 31, 32 e 33, que mapas genéticos em diferentes tamanhos de populações e níveis de DS genotípica,

reconstituídos com a utilização das três estratégias tiveram os mais diversos problemas em todas as estratégias e níveis de razão de segregação. Como a presença de grandes viés da distância, números de grupos de ligações diferentes do esperado, ordenamento incorreto das marcas, distância dentre marcas adjacentes e tamanho de grupo de ligação com valores muito discrepantes do esperado.

Evidencia-se que, para a DS genotípica mesmo com a utilização das estratégias E2 e E3, não se consegue geralmente a obtenção de mapas de ligação acurados, como ocorrido em DS gamética. A explicação deste fato, a qual já foi relatada anteriormente, é devido geralmente à ocorrência da quebra da estrutura do tipo de população.

## 5. RESUMO E CONCLUSÕES

Neste trabalho foram considerados os problemas proporcionados pela distorção de segregação no estabelecimento de mapas de ligação genética, sendo abordados aspectos teóricos relativos à detecção, aos efeitos e às implicações da distorção de segregação (DS), em relação ao estabelecimento de mapas genéticos acurados. Também foram abordados aspectos práticos relativos à construção de mapas genéticos, com inclusão de marcadores com distorção de segregação, tomando como referência dados simulados de uma população  $F_2$ .

A abordagem teórica permitiu concluir que:

- . O tamanho da população é fundamental para a detecção da distorção de segregação. Tamanho mínimo para detecção de segregação ambígua deve ser observado quando há suspeita de marcadores sob seleção.
- . A estimação da taxa de distorção de segregação foi apresentada como alternativa ao teste qualitativo, de qui-quadrado, para se avaliar a existência e o grau de distorção de segregação.
- . Distorções proporcionadas por inviabilidade gamética ou genotípica, que ocorre após fertilização, afetam de forma diferencial o mapeamento genético.
- . Os efeitos da distorção genotípica sobre o mapeamento genético são mais drásticos e erráticos.

A abordagem prática, em que foram analisados dados simulados de uma população  $F_2$ , permitiu concluir que:

- . Mapas genéticos estabelecidos a partir de funções de verossimilhança que negligenciam a distorção de segregação de marcadores, proporcionam estimativas viesadas da distância entre marcas.
- . As funções de verossimilhanças propostas e as estratégias de mapeamento, apresentadas nestes trabalhos proporcionaram resultados satisfatórios e deverão ser consideradas em estudos futuros.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

ABE, J.; TSUDA, C. Genetic analysis for isozyme variation in the section vulgares genus *Beta*. **Jpn J Breed.** v.37, p.253-261. 1987.

ALPERT, K.B.; TANKSLEY, S.D. High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: A major fruit weight quantitative trait locus in tomato. **Proc Natl Acad Sci.** v.93, n.26, p.15503-15507. 1996.

BONIERBALE, M.W.; PLAISTED, R.L.; TANKSLEY, S.D. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in *Potato* and *Tomato*. **Genetics.** v.120, p.1095-1103. 1998.

BURKE, J.M.; ARNOLD, M.L. Genetics and the fitness of hybrids. **Annu Rev Genet.** v.35, p.31-52. 2001.

BURNHAM, C.R. Differential fertilization in the *Bt-Pr* linkage group of maize. **J Am Soc Agron.** v. 28, p.968-975. 1936.

BURR, B.; BURR, F. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. **Elsevier Science Publishers Reviews,** v.7, n.2, p. 55-60. 1991.

CAI, Q.; GUY, C.L.; MOORE, G.A. Extension of the linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP



mapping of cold-acclimation-responsive loci. **Theoretical and Applied Genetics**. v.89, p.606-614. 1994.

CAMERON, D.R.; MOAV, R. Inheritance in *Nicotiana tabacum*. XXVII. Pollen killer, an alien genetic locus inducing abortion of microspores not carrying it. **Genetics**. v.42, p.326-335. 1957.

CARNEY, S.E.; HODGES, S.A.; ARNOLD, M.L. Effects of differential pollen-tube growth on hybridization in the Louisiana irises. **Evolution**. v.50, p.1871-1878. 1996

CAUSSE, M.A.; FULTON, T.M.; CHO, Y.G.; AHN, S.N.; CHUNWONGSE, J; WU, K.; XIAO, J.; YU, Z.; RONALD, P.C.; HARRINGTON, S.E.; SECOND, G.; MCCOUCH, S.R.; TANKSLEY, S.D. Saturated molecular map of the rice genome based on an interspecific backcross population. **Genetics**. v.138, p.1251-1274. 1994.

CHARLESWORTH, B.; CHARLESWORTH, D. Inbreeding depression and its evolutionary consequences. **Annu Rev Ecol Syst** v.18, p.237-268. 1987.

CHENG, R.; KLEINHOF, A.; UKAI, Y. Method for mapping a partial lethal-factor locus on a molecular-marker linkage map of a backcross and doubled-haploid population. **Theoretical and Applied Genetics**. v.97, p.293-298. 1998.

CHETELAT, R.T.; MEGLIC, V.; CISNEROS, P. A genetic map of tomato based on BC1Lycopersicon esculentum × Solanum lycopersicoides reveals overall synteny but suppressed recombination between these homologous genomes. **Genetics**. v.154, p.857–867. 2000.

CORRENS, C. Scheinbare ausnahmen von der Mendel's hen spaltungsregel fur bastarde. **Ber. Dtsch. Bot. Ges.** v.20, p.159-172. 1902.

COSTAESILVA, L.; CRUZ, C.D.; PESSONI, L.A.; FERREIRA, A.; SILVA, M.F.; BARROS, E.G. Mapeamento genético: utilização de locos com distorção na razão de segregação. In: **50º Congresso Brasileiro de**

**Genética**. 2004. Resumos. Santa Catarina – S.C.: Sociedade Brasileira de Genética. 2004.

COSTAeSILVA, L. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs**. Viçosa, MG:UFV, 2005. 120p. (Dissertação Mestrado). Universidade Federal de Viçosa, 2005.

CRISTOFANI, M.; MACHADO, M.A.; GRATTAPAGLIA, D. Genetic linkage maps of *Citrus sunki* Hort. ex. Tan. and *Poncirus trifoliata* (L.) Raf. and mapping of citrus tristeza virus resistance gene. **Euphytica**, Wagenigem, v.109, n.1, p.25-32, 1999.

CROUZILLAT, D.; LERCETEAU, E.; PETIARD, V.; MORERA, J.; RODRIGUEZ, H.; WALKER, D.; PHILLIPS, W.; RONNING, C.; SCHNELL, R.; OSEI, J.; FRITZ, P. *Theobroma cacao* L.: A genetic linkage map and quantitative trait loci analysis. **Theoretical and Applied Genetics**. v.93, p.205-214, 1996.

CRUZ, C.D.; COSTAeSILVA, L. Mapeamento físico e clonagem posicional em plantas. In: Borém, A.; Caixeta, E.T. (Eds.) **Marcadores moleculares**. Viçosa, MG. Jard, 2006. p. 307-374

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Vol. 2, Viçosa: Imprensa Universitária, 2003. 279p.

DEVAUX, P.; KILIAN, A.; KLEINHOF, A. Comparative mapping of the barley genome with male and female recombination-derived, doubled haploid populations. **Mol Gen Genet**. v. 249. p.600-608. 1995.

ECHT, C.S.; KIDWELL, K.K.; KNAPP, S.J.; OSBORN, T.C.; MCCOY, T.J. Linkage mapping in diploid alfalfa (*Medicago sativa*). **Genome** v.37. p.61-71. 1994.

ENDO, T.R. Gametocidal chromosome of three *Aegilops* species in common wheat. **Can J Genet Cytol**. v.24, p.201-206. 1982.

EUJAYL, I.; BAUM, M.; ERSKINE, W.; PEHU, E.; MUEHLBAUER, F.J. The use of RAPD markers for lentil genetic mapping and the evaluation of distorted F2 segregation. **Euphytica**. v.96, p.405–412, 1997.

EUJAYL, I.; BAUM, M.; POWELL, W.; ERSKINE, W.; PEHU, E. A genetic linkage map of lentil (*Lens sp.*) based on RAPD and AFLP markers using recombinant inbred lines. **Theoretical and Applied Genetics**. v. 97, p.83–89. 1998.

FALEIRO, F.G.; RAGAGNIN, V.A.; SCHUSTER, I.; CORRÊA, R.X.; GOODGOD, P.I.; BROMMONSHENKEL, S.H.; MOREIRA, M.A.; BARROS, E.G. **Fitopatologia Brasileira**. v.28, n.1, p. 59-66, 2003.

FERREIRA, A.; SILVA, M.F. da; COSTAeSILVA, L. da; CRUZ, C.D. Genetic mapping: experimental population size and type effects. **Genetics and Molecular Biology**. v.29, n.1, p.187-192. 2006.

FINCH, R.A.; MILLER, T.E.; BENNETT, M.D. “Cukoo” *Aegilops* addition chromosome in wheat ensures its transmission by causing chromosome breaks in meiospores lacking it. **Chromosoma**. v.90, p.84-88. 1984.

FISHMAN, L.; KELLY, A.J.; MORGAN, E.; WILLIS, J.H. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. **Genetics**. v.159, p.1701-1716. 2001.

FJELLSTROM, R.G.; PARFITT, D.E. RFLP inheritance and linkage in walnut. **Theoretical and Applied Genetics**. v.89, p.665-670. 1994.

FOLTZ, D.W. Segregation and linkage studies of alloyme loci in pair crosses of the oyster *Crassostrea virginica*. **Biochem Genet**. v.24, p.941-956. 1986.

FOOLAD, M.R.; ARULSEKA, R. S.; BECERRA, V.; BLISS, F.A. A genetic map of *Prunus* based on an interspecific cross between peach and almond. **Theoretical and Applied Genetics**. v. 91, p.262–269. 1995.

FRISCH, M.; QUINT, M.; LÜBBERSTEDT, T.; MELCHINGER, A.E. Duplicate marker loci can result in incorrect locus orders on linkage maps. **Theoretical and Applied Genetics**. v.109, p.305-316. 2004.

GADISH, I.; ZAMIR, D. Differential zygotic abortion in an interspecific *Lycopersicum* cross. **Genome**. v.29, p.156–159. 1986.

GARDINER, J.M.; COE, E.H.; MELIA-HANCOCK, S.; HOISINGTON, D.A.; CHAO, S. Development of a core RFLP map in maize using an immortalized F2 population. **Genetics**. v.134, p.917-930. 1993.

GEBHARDT, C.; RITTER, E.; DEBENER, T.; SCHACHTSCHABEL, U.; WALKEMEIER, B.; UHRIG, H.; SALAMINI, F. RFLP analysis and linkage mapping in *Solanum tuberosum*. **Theoretical and Applied Genetics**. New York, v.78, p.65-75, 1989.

GQMOL. Viçosa: 11.03.2006. (<http://www.ufv.br/dbg/gqmol/gqmol.htm>).

GRANER, A.; JAHOOOR, A.; SCHONDELMAIER, J.; SIEDLER, H.; POLLEN, K.; FISCHBECK, G.; WENZEL, G.; HERRMANN, R.G. Construction of an RFLP map of barley. **Theoretical and Applied Genetics**. v.83, p.250-256. 1991.

GRANT, V. **Genetics of flowering plants**. Columbia University Press, New York, 1975.

GRATTAPAGLIA, D.; BRADSHAW JR, H.D. Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. **Can J For Res**. v.24, p.1074-1078. 1994.

GRATTAPLAGIA, D.; FERREIRA, M.E. Mapeamento físico e clonagem posicional em plantas. In: Borém, A.; Caixeta, E.T. (Eds.) **Marcadores moleculares**. Viçosa, MG. Jard, p. 231-272. 2006.

GUIDERDONI, E.; COURTOIS, B.; GLASZMANN, J.C. Use of isozyme markers to monitor recombination and assess gametic selection among anther culture derivatives of remote crosses of rice (*Oryza sativa* L.). In:

Mujeeb-Kazi A, Sitch LA (eds) Review of advances in plant biotechnology 1985-88: **2nd Int Symp Genet Manipulation Crops**. CIMMYT and IRRI, Mexico, D. F., Mexico and Manila, Philippines, p.43-55. 1989.

HACKETT, C.A. AND BROADFOOT, L.B. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. **Heredity**. v.90, p.33-38. 2003.

HARTL, D.L. Genetic dissection of segregation distortion. I. Suicide combinations of Sd genes. **Genetics**. v.76, p.477-486. 1974.

HARUSHIMA, Y.; MURATA, N.; YANO, M.; NAGAMURA, Y.; SASAKI, T.; MINOBE, Y.; NAKAGAHRA, M. Detection of segregation distortions in an *indica-japonica* rice cross using a high-resolution molecular map. **Theoretical and Applied Genetics**. v.92, p.145-150. 1996.

HARUSHIMA, Y.; NAKAGAHRA, M.; YANO, M.; SASAKI, T.; KURATA, N. A genome-wide survey of reproductive barriers in an intraspecific hybrid. **Genetics**. v.159, p.883-892. 2001.

HE, P.; Li, J. Z.; ZHENG, X. W.; SHEN, L. S.; LU, C. F.; CHEN, Y.; ZHU, L. H. Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. **Crop Science**. v.41, p.1240-1246. 2001.

HELENTJARIS, T.; SLOCUM, M.; WRIGHT, S.; SCHAEFER, A.; NIENHUIS, J. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. **Theoretical and Applied Genetics**. v.72, p.761-769. 1986.

HERRERA, J.C.; COMBES, M.C.; ANTHONY, F.; CHARRIER, A.; LASHERMES, P. Introgression into the allotetraploid coffee ( *Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the

tetraploid interspecific hybrid ( *C. arabicax C. canephora*). **Theoretical and Applied Genetics**. v.4, p.61-668. 2002.

HUANG, N.; MCCOUCH, S.; MEW, T.; PARCO, A.; GUIDERDONI, E. Development of an RFLP map from a doubled haploid population in rice. **Rice Genet News**. v.11, p.134-137. 1994.

ISHIKAWA, R.; KINOSHITA, T.; MORISHIMA, H. Trisomic analysis of genes for isozymes: location of *Cat-1*, *Acp-1* and *Pox-2* on chromosomes. **Rice Genet News**. v.4. p.75-76. 1987.

IWATA, T., NAGAMATSU, T., OMURA, T. Abnormal segregation of waxy and apiculus coloration by a gametophyte gene belonging to the first linkage group in rice. **Jpn J Breed**. v.14, p.33-39. 1964.

JACKSON, R.C. Genomic differentiation and its effect on gene flow. **Syst Bot**. v.10, p.391-404. 1985.

JARRELL, D.C.; ROOSE, M.L.; TRAUGH, S.N.; KUPPER, R.S. A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. **Theoretical and Applied Genetics**. v.84, p.49-56. 1992.

JENCZEWSKI, E.; GHERARDI, M.; BONNIN, I.; PROSPERI, J.M.; OLIVIERI, I.; HUGUET, T. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). **Theoretical and Applied Genetics**. v.94, p.682–691. 1997.

JIANG, C.X.; CHEE, P.W.; DRAYE, X.; MORRELL, P.L.; SMITH, C.W.; PATERSON, A.H. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). **Evolution**. v.54. p.798-814. 2000.

KAO, C.H.; ZENG, Z.B.; TEASDALE, R. Multiple interval mapping for quantitative trait loci. **Genetics**. v.152, p.1023-1216. 1999.

KAWASAKI, H.; OHBA, K. Detection of embryonic lethal genes in Wogon-Yoresugi with selfing of S1 families. **Transactions of the 95 th Meeting of the Japanese Forestry Society**. p.336-337. 1984.

KY, C.L.; BARRE, P.; LORIEUX, M.; TROUSLOT, P.; AKAFFOU, S.; LOUARN, J.; CHARRIER, A.; HAMON, S.; NOIROT, M. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetics**. v.101, p.669-676. 2000.

KIJAS, J. M. H.; THOMAS, M. R.; FOWLER, J. C. S.; ROOSE, M. L. Integration of trinucleotide microsatellites into a linkage map of *Citrus*. **Theoretical and Applied Genetics**. v. 94, p. 701-706. 1997.

KONISHI, T., ABE, K., MATSUURA, S., YANO, Y. Distorted segregation of the esterase isozyme genotypes in barley *Hordeum vulgare* L. **Jpn J. Genet.** v.65, p.411-416. 1990

KUANG, H.; RICHARDSON, T.E.; CARSON, S.D.; BONGARTEN, B.C. An allele responsible for seedling death in *Pinus radiata* D. Don. **Theoretical and Applied Genetics**. v.96, p.640-644. 1998.

KY, C-L., BARRE, P., LORIEUX, M., THOUSLOT, P., AKAFFOU, S., LOUARN, J., CHARRIER, A., HAMON, S., NOIROT, M. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetics**. v.101, p.669-676. 2000.

LASHERMES, P., COMBES, M.C., PRAKASH, N.S., TROUSLOT, P., LORIEUX, M., CHARRIER, A. Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome**. v.44, p.589-596. 2001.

LAVERY, P.; JAMES, S.H. Complex hybridity in *Isotoma petraea* V. Distorted segregation gametic lethal systems and population divergence. **Heredity**. v.58, p.401-408. 1987.

LEBLON, G.; ROSSIGNOL, J.L. Mechanism of gene conversion in *Ascobolus immersus*. The interaction of heteroalleles in the conversion process. **Mol Gen Genet.** v.122, p.165-182. 1973.

LESPINASSE, D.; RODIER-GOUD, M.; GRIVET, L.; LECONTE, A.; LEGNATE, H.; SEGUIN, M. A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. **Theoretical and Applied Genetics.** v.100, p.127-138. 2000.

LEVIN, D.A.; BERUBE, D.E. *Phlox* and *Colias*: the efficiency of a pollination system. **Evolution.** v.26, p.242-250. 1972.

LI, Z.; PINSON, S.R.M.; PATERSON, A.H.; PARK, W.D.; STANSEL, J.W. Genetics of hybrid sterility and hybrid breakdown in an inter subspecific rice (*Oryza sativa* L.) population. **Genetics.** v.145, p.1139-1148. 1997.

LIEDL, B.; ANDERSON, N.O. Reproductive barriers: identification, uses and circumvention. **Plant Breed Rev.** v.11, p.11-154. 1993.

LIN, S.Y.; IKEHASHI, H. A gamete abortion locus detected by segregation distortion of isozyme locus Est-9 in wide crosses of rice (*Oryza sativa* L.). **Euphytica.** v.67, p.35-40. 1993.

LIN, S.Y.; IKEHASHI, H.; YANAGIHARA, S.; KAWASHIMA, A. Segregation distortion via male gametes in hybrids between *Indica* and *Japonica* or wide-compatibility varieties in rice (*Oryza sativa* L). **Theoretical and Applied Genetics.** v.84, p.812-818. 1992.

LIU, H.B. **Statistical Genomics, linkage, mapping and QTL analysis.** CRC, Boca Raton, Florida, p.611. 1998.

LOEGERING, W.Q., SEARS, E.R. Distorted inheritance of stem-rust resistance of timstein wheat caused by a pollen-killing gene. **Can J Genet Cytol.** v.5, p.65-72. 1963.

LONGLEY, A.E. Abnormal segregation during meiosis in maize. **Genetics.** v.30, p.100-113. 1945.



LORD, E.M.; RUSSELL, S.D. The mechanisms of pollination and fertilization in plants. **Annu Rev Cell Dev Biol.** v.18, p.81-105. 2002.

LORIEUX, M.; GOFFINET, B.; PERRIER, X.; GONZÁLEZ DE LEÓN, D.; LANAUD, C. Maximum likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. **Theoretical and Applied Genetics.** v.90, p.73-80. 1995a.

LORIEUX, M.; PERRIER, X.; GOFFINET, B.; LANAUD, C.; GONZÁLEZ DE LEÓN, D. Maximum likelihood models for mapping genetic markers showing segregation distortion. 2. F<sub>2</sub>-populations **Theoretical and Applied Genetics.** v.90, p.81-89. 1995b.

LU, Z.X.; SOSINSKI, B.; REIGHARD, G.L.; BAIRD, W.V.; ABBOTT, A.G. Construction of a genetic linkage map and identification of AFLP markers for resistance to root-knot nematodes in peach rootstocks. **Genome.** v.41, p.199-207. 1998.

LU, H.; ROMERO-SEVERSON, J.; BERNARDO, R. Chromosomal regions associated with segregation distortion in maize. **Theoretical and Applied Genetics.** v.105, p.622-628. 2002.

LURO, F.; LORIEUX, M.; LAIGRET, F.; BOVÉ, J.M.; OLLITRAULT, P. Cartographie du génome des agrumes à l'aide des marqueurs moléculaires et distorsions de ségrégation. In: Berville and Tersac (eds) **Techniques et utilisations des marqueurs moléculaires**, INRA, Montpellier, France, v.72, p.69-82. 1994.

LYTTLE, T.W. Segregation distorters. **Annu Rev Genet.** v.25, p.511-557, 1991.

MAEKAWA, M.; KITA, F. New gametophyte genes located in the third linkage group (chromosome 3) of rice. **Jpn J Breed.** v.35, p. 25-31. 1985.

MAEKAWA, M.; KINOSHITA, T.; TAKAHASHI, M. Genetical studies on the rice plant. **LXXVI. A new gametophyte gene in the second linkage group of rice.** Jpn Fac Agric Hokkaido Univ. v.60, p.107-114. 1981.

MANGELSDORF, P.C.; JONES, D.F. The expression of Mendelian factors in the gametophyte of maize. **Genetics.** v.11, p.423-455. 1926.

MARQUES, C. M., J. A. ARAÚJO, J. G. FERREIRA, R. WHETTEN, and D. M. O'MALLEY *et al.*, AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. **Theoretical and Applied Genetics.** v.96, p.727-737. 1998.

MATSUSHITA, S.; ISEKI, T.; FUKUTA, Y.; ARAKI, E.; KOBAYASHI, S.; OSAKI, M.; YAMAGISHI, M. Characterization of segregation distortion on chromosome 3 induced in wide hybridization between indica and japonica type rice varieties **Euphytica.** v.134, p.27-32. 2003.

McCOUCH, S.R.; KOCHERT, G.; YU, Z.H.; WANG, Z.Y.; KHUSH, G.S.; GOFFMAN, W.R.; TANKSLEY, S.D. Molecular mapping of rice chromosomes. **Theoretical and Applied Genetics.** v.76, p.815-829. 1988.

MENÉNDEZ, C.M.; HALL, A.E.; GEPTS, P. A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between two inbred, domesticated lines. **Theoretical and Applied Genetics.** v.95, p.1210-1217, 1997.

NAG, D.K.; WHITE, M.A.; PETES, T.D. Palindromic sequences in heteroduplex DNA inhibit mismatch repair in yeast. **Nature.** v.340, p.318-320. 1989.

NAKAGAHRA, M.; OMURA, T.; IWATA, N. Gametophyte genes and their loci on the eleventh linkage group of cultivated rice. **Jpn J Breed.** v.22, p.305-312. 1972.

NAKAGAHRA, M. Genetic mechanism on the distorted segregation of marker gene belonging to the eleventh linkage group in cultivated rice. **Jpn J Breed.** v.22, p.232-238. 1972.

NAKAGAHRA, M. Geographic distribution of gametophyte genes in wide crosses of rice cultivars. In: **Rice Genetics**. IRRI, Manila, Philippines. p.73-82. 1986.

NICOLAS, A., ROSSIGNOL, J.L. Gene conversion: point-mutation heterozygosities lower heteroduplex formation. **EMBO J.** v.2, p.2265–2270. 1983.

NIKAIDO, A., YOSHIMARU, H., TSUMURA, Y., SUYAMA, Y., MURAI, M., NAGASAKA, K. Segregation distortion for AFLP markers in *Cryptomeria japonica*. **Genes Genet Systems**. v.74, p.55-59. 1999.

NODARI, R.O.; TSAI, S.M.; GILBERTSON, R.L.; GEPTS, P. Towards an integrated linkage map of common bean. 2. development of an RFLP-based linkage map. **Theoretical and Applied Genetics**. v.85, p.513-520. 1993.

OHBA, K. **Detection of embryonic lethal genes in sugi, *Cryptomeria japonica*** D. Don. Transactions of the 90<sup>th</sup> Meeting of the Japanese Forestry Society. p. 257-258. 1979.

OLIVEIRA, R.P. de; AGUILAR-VILDOSO, C.I.; CRISTOFANI, M.; MACHADO, M.A. Skewed RAPD markers in linkage maps of *Citrus*. **Genet. Mol. Biol.** v.27, n.3. 2004.

PAILLARD, M.; LASHERMES, P.; PÉTIARD, V. Construction of a linkage map in coffee. **Theoretical and Applied Genetics**. v.93, p.41-47. 1996.

PALOPOLI, M.F.; WU, C.I. Rapid evolution of a coadapted gene complex: Evidence from the Segregation Distorter (SD) system of meiotic drive in *Drosophila melanogaster*. **Genetics**. v.143, p.1675-88. 1996.

PARAN, I.; GOLDMAN, I.; TANKSLEY, S.D.; ZAMIR, D. Recombinant inbred lines for genetic mapping in tomato. **Theoretical and Applied Genetics**. v.90, p.542-548. 1995.

PATERSON, A.H.; LANDER, E.S.; HEWITT, J.D.; PETERSON, S.; LINCOLN, S.E.; TANKSLEY, S.D. Resolution of quantitative traits into

Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. **Nature**. v.335, p.721-726. 1988.

PATERSON, A.H.; DAMON, S.; HEWITT, J.D.; ZAMIR, D.; RABINOWITCH, H.D.; LINCOLN, S.E.; LANDER, E.S.; TANKSLEY, S.D. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. **Genetics**. v.127, p.181-197. 1991.

PEREIRA, M.G.; LEE, M.; BRAMEL-COX, P.; WODMAN, W.; DOEBLEY, J.; WHITKUS, R. Construction of an RFLP map in sorghum and comparative mapping in Maize. **Genome**. v.37, p. 236-243. 1993.

PEREIRA, M.G.; LEE, M.; BRAMEL-COX, P.; WOODMAN, W.; DOEBLEY, J.; WHITKUS, R. Construction of an RFLP map in sorghum and comparative mapping in maize. **Genome**. v.37, p.236-243. 1994.

PRABHU, R.R.; GRESSHOFF, P.M. Inheritance of polymorphic markers generated by DNA amplification fingerprinting and their use as genetic markers in soybean. **Plant Molecular Biology**. v.26, p.105-116. 1994.

QUILLET, M.C., MADJIDAN, N., GRIVEAU, Y., SERIEYS, H., TERSAC, M., LORIEUX, M., BERVILLÉ, A. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. **Theoretical and Applied Genetics**. v. 91, p.1195-1202. 1995.

RANJHAN, S.; GLASZMANN, J.C.; RAMIREZ, D.A.; KHUSH, G.S. Chromosomal localization of four isozyme loci by trisomic analysis in rice (*Oryza sativa* L.). **Theoretical and Applied Genetics**. v.75, p.541-545. 1988.

RHOADES, M.M.; DEMPSEY, E. The effect of abnormal chromosome 10 on preferential segregation and crossing over in maize. **Genetics**. v.53, p.989-1020. 1966.

RHOADES, M. M. Preferential segregation in maize. **Genetics**. v.27, p.395-407. 1942.

RICK, C.M. Differential zygotic lethality in a tomato species hybrid. **Genetics**. v.48, p.1497-1507. 1963.

RICK, C.M. Abortion of male and female gametes in the tomato determined by allelic interaction. **Genetics**. v.53, P.85-96. 1966.

RIESEBERG, L.; VAN FOSSEN, C.; DESROCHERS, A. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. **Nature**. v.375, p.313-316. 1995.

SÄLL, T.; NILSSON, N.O. The robustness of recombination frequency estimates in intercrosses with dominant markers. **Genetics**. v.137, p.589-596. 1994.

SANDLER, L.; GOLIC, K. Segregation distortion in *Drosophila*. **Trends Genet.** v.1, p.181-185. 1985.

SANDLER, L.; NOVITSKI, E. Meiotic drive as an evolutionary force. **Amer Naturalist**. v.41, p. 105-110. 1957.

SANO, Y.; CHU, Y.E.; OKA, H.I. Genetic studies of speciation in cultivated rice. 1. Genic analysis for the F<sub>1</sub> sterility between *Oryza sativa* L. and *O. glaberrima* steud. **Jpn J Genet.** v.54, p.121-132. 1979.

SANO, Y. A new gene controlling sterility in F<sub>1</sub> hybrids of two cultivated rice species. **J Hered.** v.74, p.435-439. 1983.

SANO, Y. The genic nature of gamete eliminator in rice. **Genetics**. v.125, p.183-191. 1990.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. v.1, Viçosa: Imprensa Universitária, 2004. 568p.

SCHWARZ-SOMMER, Z.; ANDRADE SILVA, E. de; BERNDTGEN, R.; LONNIG, W. E.; MULLER, A. *et al.* A linkage map of an F<sub>2</sub> hybrid population of *Antirrhinum majus* and *A. molle*. **Genetics**. v. 163, p.699-710. 2003.

SCHWEMMLE, J. Selective fertilization in *Oenothera*. **Adv Genet.** v.14, p.225-324. 1968.

SCOLES, G.J.; KIBIRGE-SEBUNYA, I.N. Preferential abortion of gametes in wheat induced by an *Agropyron* chromosome. **Can J Genet Cytol.** v.25, p.1-6. 1983.

SNOW, A.A. Mate choice in plants-tactics, mechanisms, and consequences. **Ecology**. v.65, p.1025–1026. 1984.

SORENSEN, F.C. Linkage between marker genes and embryonic lethal factors may cause disturbed segregation ratios. **Silvae Genet.** v.16, p.132-134.1967.

STEBBINS, G.L. **Variation and evolution in plants**. Colômbia University Press, New York. 1950.

TANKSLEY, S. D.; GANAL, M. W.; MARTIN, G. B. Chromosome landing: a paradigm for map-based cloning in plants with large genomes. *Trends in Genetics*. v.11, 63-68. 1995.

TAYLOR, D. R.; INGVARSSON, P. K. Common features of segregation distortion in plants and animals. **Genetica**. v.117, p.27-35. 2003.

TENHOOPEN, R.; ROBBINS, T.P.; FRANSZ, P.F.; MONTIJN, B.M.; OUD, O.; GERATS, A.G.M.; NANNINGA, N. Localization of T-DNA insertions in petunia by fluorescence in situ hybridization: physical evidence for suppression of recombination. **Plant Cell**. v.8, p.823-830. 1996.

TORRES, A.M.; WEEDEN, N.F.; MART, N. A. Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. **Theoretical and Applied Genetics**. v.85, p.937-945. 1993.

TRAVIS, S. E.; RITLAND, K. R.; WHITHAM, T.; KEIM, P. A genetic linkage map of Pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. **Theoretical and Applied Genetics**. v.97, p.871-880. 1998.

TSUJIMOTO, H.; TSUNEWAKI, K. Gametocidal genes in wheat and its relatives. II suppressor of the chromosome 3C gametocidal gene of *Aegilops triuncialis*. **Can J Genet Cytol**. v.27, p.178-185. 1985.

TSUMURA, Y.; UCHIDA, K.; OHBA, K. Genetic control of isozyme variation in needle tissues of *Cryptomeria japonica* **J. Hered**. v.80, p. 291-297. 1989.

VALLEJOS, C.E.; SAKIYAMA, N.S.; CHASE, C.D. A molecular marker based linkage map of *Phaseolus vulgaris* L. **Genetics**. v.131, p.733-740. 1992.

VELDBOOM, L.R.; LEE, M. Molecular-marker-facilitated studies of morphological traits in maize. II. Determination of QTLs for grain yield and yield components. **Theoretical and Applied Genetics**. v.89, p.451-458. 1994.

VILANOVA, S.; ROMERO, C.; ABBOTT, A.G.; LLÁCER, G.; BADENES, M.L. An apricot (*Prunus armeniaca* L.) F2 progeny linkage map based on SSR and AFLP markers, mapping plum pox virus resistance and self-incompatibility traits. **Theoretical and Applied Genetics**. v.107, p.239-247. 2003.

VIRK, P.S.; FORD-LLOYD, B.V.; NEWBURY, H.J. Mapping AFLP markers associated with subspecific differentiation of *Oryza sativa* (rice) and an investigation of segregation distortion. **Heredity**. v.81, p.613-620. 1998.

VOGL, C.; XU, S. Multipoint mapping of viability and segregation distorting loci using molecular markers. **Genetics**. v.155, p.1439-1447. 2000.

WANG, G.L.; MACKILL, D.J.; BONMAN, M.; MCCOUCH, S.R.; CHAMPOUX, M.C.; NELSON, R.J. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. **Genetics**. v.136, p.1421-1434. 1994.

WEBB, D.M.; KNAPP, S.J.; TAGLIANI, L.A. Restriction fragment length polymorphism and allozyme linkage map of *Cuphea lanceolata*. **Theoretical and Applied Genetics**. v.83, p.528-532. 1992.

WENDEL, J.F.; EDWARDS, M.D.; STUBER, C.W. Evidence for multilocus genetic control of preferential fertilization in maize. **Heredity**. v.58, p.297-302. 1987.

WHITKUS, R. Genetics of adaptive radiation in Hawaiian and Cook Island species of *Tetramolopium* (Asteraceae). II. Genetic linkage map and its implications for interspecific breeding barriers. **Genetics**. v.150, p.1209-1216. 1998.

WILLSON, M.F.; BURLEY, N. **Maite choice in plants: tactics, mechanisms and consequences**. Princeton University Press, Princeton, N.J. 1983.

WU, K.S.; GLASZMANN, J.C.; KHUSH, G.S. Chromosomal locations of ten isozyme loci in rice (*Oryza sativa* L.) through trisomic analysis. **Biochem Genet**. v.26, p.303-320. 1988.

XIAO, J.; LI, J.; YUAN, L.; TANKSLEY, S.D. Identification of QTLs affecting traits of agronomic importance in a recombination inbred population derived from a subspecific rice cross. **Theoretical and Applied Genetics**. v.92, p.230-244. 1996.

XU, Y.B.; SHEN, Z.T. Distorted segregation of waxy gene and its characterization in *indica-japonica* hybrids (in Chinese). **Chinese J Rice Sci**. v.6, p.89-92. 1992.

XU, G.W.; MAGILL, C.W.; SCHERTZ, K.F.; HART, G.E. A RFLP linkage map of *Sorghum bicolor* (L.) Moench. **Theoretical and Applied Genetics**. v.89, p.139-145. 1994.

XU, Y.; ZHU, L.; XIAO, J.; HUANG, N.; MCCOUCH, S.R. Chromosomal regions associated with segregation distortion of molecular markers in F<sub>2</sub>,



backcross, double haploids, and recombinant inbred populations in rice (*Oryza sativa* L.). **Mol Gen Genet.** v.253, p.535-545. 1997.

XU, Y.B.; SHEN, Z.T.; CHEN, Y.; ZHU, L.H. Distorted segregations of RFLP markers and their distribution on chromosomes in an *indica-japonica* F2 population of rice (*Oryza sativa* L.) (in Chinese). **Acta Bot Sin.** v.37, p.91-96. 1995.

YAMANAKA, N.; NINOMIYA, S.; HOSHI, M.; TSUBOKURA, Y.; YANO, M.; NAGAMURA, Y.; SASAKI, T.; HARADA, K. An Informative Linkage Map of Soybean Reveals QTLs for Flowering Time, Leaflet Morphology and Regions of Segregation Distortion. **DNA Research.** v.8, n.2, p.61-72. 2001.

YAMAGISHI, M.; YANO, M.; FUKUTA, Y.; FUKUI, K.; OTANI, M.; SHIMADA, T. Distorted segregation of RFLP markers in regenerated plants derived from anther culture of an F1 hybrid of rice. **Genes Genet Syst.** v.71, p.37-41. 1996.

YANAGIHARA, S.; McCOUCH, S.R.; ISHIKAWA, K.; OGI, Y.; MARUYAMA, K.; IKEHASHI, H. Molecular analysis of the inheritance of the S-5 locus, conferring wide compatibility in *Indica japonica* hybrids of rice (*O. sativa* L.). **Theoretical and Applied Genetics.** v.90, p.182-188. 1995.

YANG, W.; WEAVER, D.B.; NIELSEN, B.L.; QIU, J. A preliminary genetic linkage map of soybean using an intraspecific cross of two cultivars: 'peking' and 'lee'. **Soybean Genetics Newsletter.** v.27. 2000.

YU, Z.; GUO, X. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. **Biol. Bull.** v.204, p.327-338. 2003

ZAMIR, D.; TADMOR, Y. Unequal segregation of nuclear genes in plants. **Bot Gaz** v.147, p. 355–358. 1986.

ZAMIR, D. Pollen gene expression and selection: applications in plant breeding. In: S. D. Tanksley and T.J. Orton, eds. **Isozymes in plant genetics and breeding.** Elsevier, Amsterdam. p.313-330. 1983.

ZHANG, F.; XUE-QIN, W.; GUANG-TANG, P. QTL mapping of *Fusarium moniliforme* ear rot resistance in maize. 1. map construction with microsatellite and AFLP markers. **J. Appl. Genet.** v.47, n.1, p.9-15. 2006.