

Estimativa de frequência de recombinação no mapeamento genético de famílias de irmãos completos

Leonardo Lopes Bhering⁽¹⁾, Cosme Damião Cruz⁽¹⁾ e Pedro Ivo Vieira Good God⁽¹⁾

⁽¹⁾Universidade Federal de Viçosa, Departamento de Biologia Geral, CEP 36571-000 Viçosa, MG. E-mail: leobhering@yahoo.com.br, cdcruz@ufv.br, pivggod@yahoo.com.br

Resumo – O objetivo deste trabalho foi obter as estimativas de frequência de recombinação, por meio de simulação, para diferentes situações do mapeamento genético de famílias de irmãos completos. Foi simulado um genoma constituído de três grupos de ligação, em que cada um apresentava 11 marcas moleculares multialélicas, codominantes, distribuídas a cada 10 cM. A partir desse genoma, foram simulados dois cenários: um com genitores completamente informativos e outro com genitores formados aleatoriamente. Após a obtenção de todas as estimativas das frequências de recombinação, concluiu-se que, para locos completamente informativos, pode-se calcular a frequência de recombinação entre pares de locos, a partir da frequência gamética de cada genitor ou a partir da frequência genotípica da progênie. Para locos parcialmente informativos, a obtenção da frequência de recombinação a partir da frequência genotípica conjunta é mais apropriada.

Termos para indexação: genômica, máxima verossimilhança, populações exogâmicas.

Estimation of recombination frequency in genetic mapping of full-sib families

Abstract – The objective of this work was to obtain the estimates of recombination frequency, through simulation, for different situations in genetic mapping involving full sibling families. A genome with three linkage groups simulated, each one with 11 molecular marks which were multi-allelic, codominants, with saturation of 10 cM. Starting from this genome, two situations were simulated: one with informative genitors, and other with genitors formed randomly. Once all the recombination frequencies were estimated, it was found out that in completely informative loci, the recombination frequency can be calculated among equal pairs of loci starting from the gametic frequency of each genitor, or starting from the genotypic frequency of the progeny. For partially informative loci, the recombination frequency starting from the united genotypic frequency is more appropriate.

Index terms: genomics, exogamic populations, maxim likelihood.

Introdução

A construção de mapas genéticos de populações exogâmicas apresenta determinadas complicações que não são encontradas em progênies derivadas de linhagens endogâmicas, como populações F₂, RILs (recombinant inbred lines), duplo-haplóides, retrocruzamentos, entre outros. Em populações segregantes, derivadas de linhagens endogâmicas, todos os locos segregam para apenas dois alelos. Além disso, a fase de ligação do duplo heterozigoto pode ser claramente determinada, com base na análise da segregação dos gametas recombinantes da população. Ainda, na descendência de um cruzamento entre dois indivíduos não idênticos de uma população exogâmica, o número de alelos que segregam, por loco marcador, poderá variar em até quatro para uma espécie diplóide.

Além disso, a fase de ligação normalmente é desconhecida.

Em determinadas espécies vegetais, não é possível se obterem populações segregantes derivadas de linhagens endogâmicas, em razão da auto-incompatibilidade, depressão endogâmica ou do longo período juvenil. Assim, com tais espécies, é preciso empregar delineamentos genéticos de populações exogâmicas, como famílias de meios-irmãos e famílias de irmãos completos (FIC).

Na construção de um mapa genético, o primeiro passo é a análise de ligação entre pares de locos, para se verificar a existência de ligação entre eles. O teste apropriado para essa situação é o qui-quadrado. Entretanto, tal teste é apenas qualitativo e, quando se detecta evidência de ligação, não é possível obter-se a percentagem de recombinação entre os pares de

marcadores. O método mais utilizado para a estimação da percentagem de recombinação é o método da máxima verossimilhança (MV). Após a estimação, é necessário definir a frequência máxima de recombinação e o LOD mínimo, para inferir se dois locos estão ligados. O objetivo é estabelecer critérios para a formação dos grupos de ligação. Depois de definidos os grupos de ligação, a última abordagem é estimar as frequências de recombinação, com base em informações multilocos entre pares de marcas, que considere todos os marcadores ligados em um grupo de ligação simultaneamente, o que resulta em análise única para cada grupo de ligação (Lynch & Walsh, 1998).

Diferentes configurações de marcadores podem segregar em famílias de irmãos completos, originadas a partir do cruzamento entre parentais derivados de uma população exogâmica. De acordo com Haseman & Elston (1972), para o caso geral de sistemas multialélicos com quatro ou mais alelos haverá, basicamente, três categorias e sete tipos distintos de acasalamentos ou diferentes tipos de pares de irmãos que caracterizam a herança de marcas individuais. Assim, considerando-se um loco A com alelos i, j, k e l, têm-se os seguintes tipos de cruzamentos: entre genitores homocigotos: I- $A_iA_i-A_iA_i$, II- $A_iA_i-A_jA_j$; entre genitores homocigotos e heterocigotos: III- $A_iA_i-A_jA_j$; IV- $A_iA_i-A_jA_k$; entre genitores heterocigotos: V- $A_iA_j-A_iA_j$, VI- $A_iA_j-A_iA_k$, VII- $A_iA_j-A_kA_l$.

Na análise de segregação, deve-se considerar que somente acasalamentos que envolvam pelo menos um dos genitores heterocigoto sejam informativos, para fins de mapeamento. Assim, apenas cruzamentos dos tipos III, IV (cruzamentos entre genitores homocigotos e heterocigotos) e V, VI, e VII (cruzamentos entre genitores heterocigotos) são informativos. Além disso, deve ser observada a existência de indivíduos informativos na prole, que são aqueles nos quais é possível identificar-se a origem de seus alelos em relação aos parentais. De forma resumida, podem-se classificar os seguintes tipos de acasalamentos, quanto ao grau de informação da progênie (Lynch & Walsh, 1998): famílias derivadas de cruzamentos completamente informativos; famílias derivadas de retrocruzamentos; e famílias derivadas de intercruzamentos. A partir desses cruzamentos, diferentes conteúdos de informação podem ser obtidos (Liu, 1998).

Um fator complicador na análise de ligação, em cruzamentos que envolvem genitores exogâmicos, é a determinação da fase de ligação que, em geral, não é

conhecida a priori. Em populações segregantes derivadas de linhagens endogâmicas, a fase de ligação é a mesma nos parentais, ao passo que, em FIC, as seguintes combinações de fase de ligação podem ser encontradas: acoplamento no primeiro parental e indefinido no segundo, ou vice-versa; repulsão no primeiro parental e indefinido no segundo, ou vice-versa; acoplamento em ambos os parentais; repulsão em ambos os parentais; acoplamento no primeiro parental e repulsão no segundo ou vice-versa (Maliepaard et al., 1997).

O objetivo deste trabalho foi obter as estimativas de frequência de recombinação, por meio de simulação, para diferentes situações no mapeamento genético de família de irmãos completos.

Material e Métodos

Um genoma hipotético foi simulado com três grupos de ligação (GL). Os grupos de ligação possuíam tamanho de 100 cM, com 11 marcas moleculares multialélicas e codominantes, equidistantes a cada 10 cM. Foram gerados, por meio de simulação, dois cenários para se exemplificar o mapeamento genético em famílias de irmãos completos, descritos a seguir.

Cenário 1: os genitores foram gerados de maneira a serem completamente informativos para todos os locos e, portanto, $A_iA_j \times A_kA_l$. Neste caso, os pais possuem marcadores diferentes e heterocigotos. Assim todas as progênies são informativas em distinção aos alelos alternativos, provenientes de ambos os pais, de forma que os alelos parentais podem ser distintos e ambos os pais podem ser examinados, pela comparação de valores da característica nas progênies $A_i \times A_j$ e $A_k \times A_l$. A partir dos cruzamentos foi obtida uma família de irmãos completos com 200 indivíduos.

Cenário 2: os genitores foram gerados de forma aleatória, com quatro alelos em igual frequência (25% para cada), de modo que o conteúdo de informação fosse variável para cada loco, uma vez que todas as sete diferentes classes relatadas anteriormente podem ocorrer na população. Assim, nem todas as progênies foram informativas; e houve diferentes situações em relação à informatividade que deverão ser analisadas. A partir dos genitores foi obtida uma família de irmãos completos com 200 indivíduos.

Resultados e Discussão

No cenário 1, em que foram utilizados genitores completamente informativos do tipo $A_iA_j \times A_kA_l$, a

segregação esperada a ser testada é a seguinte: 1 A_iA_k:1 A_iA_j:1 A_jA_k:1 A_jA_i, ou seja, 1:1:1:1, de modo que a combinação entre todos os locos será do tipo (1:1:1:1) x (1:1:1:1). Dessa forma, a combinação de dois locos completamente informativos deve produzir uma progênie totalmente informativa e, é possível, sem dificuldade adicional, determinar-se a frequência de recombinação entre os dois locos.

No caso de os locos serem completamente informativos, podem ser obtidas três estimativas da porcentagem de recombinação. Nas duas primeiras situações, podem ser obtidas estimativas da frequência de recombinação, com base nas frequências gaméticas marginais de cada genitor. Alternativamente, pode-se utilizar a informação da frequência genotípica conjunta da progênie e é possível, para este tipo de combinação, a reconstrução completa dos haplótipos parentais na progênie.

Pode-se escrever, de forma generalizada, o seguinte cruzamento:

$$P_1: A_1A_2B_1B_2 \times P_2: A_3A_4B_3B_4.$$

Considerando-se os haplótipos para os marcadores no parental 1, em fase de acoplamento, os gametas produzidos e suas frequências serão do tipo:

$$f(A_1B_1) = f(A_2B_2) = (1 - r)/2 = P$$

$$f(A_1B_2) = f(A_2B_1) = r/2 = R.$$

Da mesma forma, para o segundo parental, tem-se:

$$f(A_3B_3) = f(A_4B_4) = (1 - r)/2 = P$$

$$f(A_3B_4) = f(A_4B_3) = r/2 = R.$$

Combinando-se os gametas dos dois parentais, teremos 16 classes genotípicas, conforme exemplificado na Tabela 1.

Assim, o cálculo da porcentagem de recombinação pode ser feito de três maneiras descritas a seguir.

Com base nas frequências gaméticas marginais do genitor 1, de modo que o genitor 1 estará em fase de acoplamento, se for observado que (n₁ + n₄) > (n₂ + n₃); nesta situação, a porcentagem de recombinação será calculada por:

$$r_{G1CIS} = (n_2 + n_3)/N.$$

Caso seja observado que (n₁ + n₄) < (n₂ + n₃), o genitor 1 estará em fase de repulsão, de modo que:

$$r_{G1TRANS} = (n_1 + n_4)/N.$$

Com base nas frequências gaméticas marginais do genitor 2, de modo que o genitor 2 estará em fase de acoplamento, se for observado que (n₁ + n₄) > (n₂ + n₃); nesta situação, a porcentagem de recombinação será calculada por: r_{G2CIS} = (n₂ + n₃)/N.

Caso seja observado que (n₁ + n₄) < (n₂ + n₃), o genitor 2 estará em fase de repulsão, de modo que: r_{G2TRANS} = (n₁ + n₄)/N.

E ainda, o cálculo da frequência de recombinação pode ser feito com base na frequência genotípica conjunta da progênie, por meio do método da máxima verossimilhança; de modo que, tendo-se estabelecido adequadamente a fase de ligação, pode-se estimar a porcentagem de recombinação por meio do método da máxima verossimilhança, com base nas frequências genotípicas da progênie. Uma alternativa é a obtenção da frequência de recombinação, para as quatro combinações possíveis na fase de ligação, e assim, tem-se que a fase de ligação real é aquela onde se obtém o maior valor de LOD.

A função de verossimilhança pode ser descrita conforme a seguinte equação:

$$L(r; n_i) = [N! / (n_{11}! \dots n_{44}!)] \times (P^2)^{n_{11} + n_{14} + n_{41} + n_{44}} \times (R^2)^{n_{22} + n_{23} + n_{32} + n_{33}} \times (PR)^{n_{12} + n_{13} + n_{21} + n_{24} + n_{31} + n_{34} + n_{42} + n_{43}}.$$

Maximizando-se o logaritmo natural da função de verossimilhança, obtém-se:

$$r = S_R/n + 0,5S_{PR}/n = (r_{G1} + r_{G2})/2,$$

em que S_R é soma do número de indivíduos originados de duas cromátides recombinantes, cuja frequência esperada é R²; S_{PR} é a soma do número de indivíduos originados de uma cromátide recombinante e outra paternal, cuja frequência esperada é PR.

De forma geral, para os locos em fase de aproximação nos dois pais, tem-se:

$$L(r; n_i) = \lambda(P^2)^{n_1}(PR)^{n_2}(R^2)^{n_3}, \text{ em que: } \lambda = N! / (n_1!n_2! \dots n_n!),$$

$$L(r; n_i) = \lambda[0,25(1-r)^2]^{n_1}[0,25r(1-r)]^{n_2}(0,25r^2)^{n_3}. \text{ Ao se}$$

Tabela 1. Frequências gaméticas dos genitores e frequências genotípicas da progênie, para cruzamento entre dois locos completamente informativos (1:1:1:1) e (1:1:1:1).

Gametas		A ₃ B ₃ P	A ₃ B ₄ R	A ₄ B ₃ R	A ₄ B ₄ P	Frequência gamética	Total de indivíduos
A ₁ B ₁	P	P ² ₁₃₋₁₃	PR ₁₃₋₁₄	PR ₁₄₋₁₃	P ² ₁₄₋₁₄	(1-r)/2	n ₁
A ₁ B ₂	R	PR ₁₃₋₂₃	R ² ₁₃₋₂₄	R ² ₁₄₋₂₃	PR ₁₄₋₂₄	r/2	n ₂
A ₂ B ₁	R	PR ₂₃₋₁₃	R ² ₂₃₋₁₄	R ² ₂₄₋₁₃	PR ₂₄₋₁₄	r/2	n ₃
A ₂ B ₂	P	P ² ₂₃₋₂₃	PR ₂₃₋₂₄	PR ₂₄₋₂₃	P ² ₂₄₋₂₄	(1-r)/2	n ₄
Frequência gamética		(1-r)/2	r/2	r/2	(1-r)/2		
Total		n ₁	n ₂	n ₃	n ₄		

desenvolver esta expressão, obtém-se: $r = (2n_3 + n_2)/2N$, em que n_3 é a soma do número de indivíduos originados de duas cromátides recombinantes, cuja frequência esperada é R^2 ; n_2 é a soma do número de indivíduos originados de uma cromátide recombinante e outra paternal, cuja frequência esperada é PR ; e N é o número total de indivíduos da população.

Para se testar se um par de marcadores está ligado, se $r < 0,5$, o LOD score pode ser usado como teste estatístico. O LOD score é o logaritmo de base 10, da razão entre a estimativa de máxima verossimilhança determinada ($r = \hat{r}$) e a estimativa de verossimilhança sob a hipótese de nulidade (considerando-se $r = 0,5$, ausência de ligação). Um LOD score de 3 é comumente utilizado e significa que a estimativa da percentagem de recombinação é mil vezes mais provável do que a hipótese de nulidade. Esta alta restringência é necessária, pois são realizados múltiplos testes para os diferentes pares de marcadores (Maliépaard et al., 1997).

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

No cenário 2, como foram utilizados marcadores que não são completamente informativos, as marcas analisadas segregam em diferentes proporções, de modo que são encontrados locos com segregação completamente informativa (1:1:1:1), informativa (1:1) ou parcialmente informativa (1:2:1).

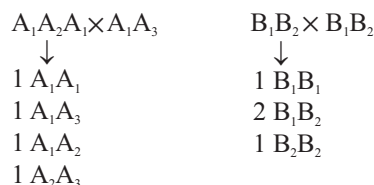
Como discutido anteriormente, na análise de pares de marcas que envolvem combinações do tipo (1:1:1:1)(1:1:1:1), os haplótipos que geraram a progênie podem ser reconstruídos completamente, e é possível diferenciar os gametas de origem parental e aqueles oriundos de recombinação, por meio da produção de uma progênie totalmente informativa. Entretanto, para os dados do cenário 2, nem todas as combinações serão totalmente informativas. Diferentes combinações podem surgir ente locos completamente informativos, informativos e parcialmente informativos.

De posse da população simulada, para exemplificar o cálculo da percentagem de recombinação foram obtidos nove diferentes situações (Tabela 2).

Para facilitar o entendimento das expressões de máxima verossimilhança, utilizadas ao longo desta simulação, foram reunidas, na Tabela 3, todas as classes de probabilidades esperadas e as suas estimativas. Os valores ali reunidos foram utilizados, posteriormente, nas funções de verossimilhança.

Para exemplificar a obtenção das estimativas, demonstra-se, a seguir, o cruzamento do caso (1:1:1:1)(1:2:1).

Cruzamento Tipo: $A_1A_2B_1B_2 \times A_1A_3B_1B_2$. Deste cruzamento, tem-se:



Este caso aborda a segregação entre um loco completamente informativo e um loco parcialmente informativo. Verifica-se que, para o loco B ($B_1B_2 \times B_1B_2$), não é possível distinguir os genótipos do tipo B_1B_2 , quanto à origem dos alelos dos parentais. Assim, os genótipos $A_1A_1-B_1B_2$, $A_1A_2-B_1B_2$, $A_1A_3-B_1B_2$ e $A_2A_3-B_1B_2$ estão confundidos em mais de uma classe genotípica, e não é possível distingui-la (Tabela 4). O fato de se encontrarem genótipos confundidos em mais de uma classe pode reduzir a precisão na estimativa da distância entre os dois locos, uma vez que menor número de classes será utilizado, para a obtenção das estimativas da percentagem de recombinação. Além disso, o valor de r deixa de ser a média entre os valores marginais da tabela.

De posse das classes e número de indivíduos, observados em cada uma dessas classes, pode-se obter a expressão de máxima verossimilhança, de forma a se desenvolver uma fórmula geral para o cálculo da percentagem de recombinação, na situação que envolve cruzamento do tipo (1:1:1:1)(1:2:1). Assim, tem-se:

Tabela 2. Tipos de segregação encontrados na população simulada, marcadores pertencentes a cada um desses tipos e as classes de probabilidades esperadas em cada uma das segregações.

Segregação	Classes
(1:1:1:1) (1:1:1:1)	$P^2; PR; R^2$
(1:1:1:1) (1:2:1)	$P^2; PR; R^2; 2PR; P^2+R^2$
(1:2:1)(1:1:1:1)	$P^2; PR; R^2; 2PR; P^2+R^2$
(1:2:1)(1:2:1)	$P^2; R^2; 2PR; 2P^2+2R^2$
(1:2:1)(1:1)	$P^2 + PR; R^2 + PR; P^2 + R^2 + 2PR$
(1:1) (1:2:1)	$P^2 + PR; R^2 + PR; P^2 + R^2 + 2PR$
(1:1) (1:1)	$2P^2 + 2PR; 2R^2 + 2PR$
(1:1) (1:1:1:1)	$P^2 + PR; R^2 + PR$
(1:1:1:1) (1:1)	$P^2 + PR; R^2 + PR$

$$L(r;n_i) = \lambda(P^2)^{n_1}(PR)^{n_2}(R^2)^{n_3}(2PR)^{n_4}(P^2 + R^2)^{n_5},$$

em que $\lambda = N!/(n_1!n_2!\dots n_n!)$,
 $L(r;n_i) = \lambda[0,25(1-r)^2]^{n_1}[0,25r(1-r)]^{n_2}(0,25r^2)^{n_3}$
 $[0,5r(1-r)]^{n_4}[0,25(2r^2 - 2r + 1)]^{n_5}$.

Ao se desenvolver a expressão, obtém-se:
 $r^3[-4N] + r^2[2(2N + n_2 + 2n_3 + n_4 + n_5)] + r[-2(N + n_2 + 2n_3 + n_4)] + (n_2 + 2n_3 + n_4) = 0$

Ao se obter a raiz do polinômio, temos o valor de r. Vários são os programas que calculam a raiz do polinômio de terceiro grau; uma forma fácil de se conseguir esta estimativa é fornecida pelo Método de Tartaglia (ou Método de Cardano) para a obtenção das raízes de uma equação de terceiro grau.

Tabela 3. Classes de probabilidades e respectivas estimativas, utilizadas na função de verossimilhança⁽¹⁾.

Probabilidade	Estimativas
P^2	$\left(\frac{1-r}{2}\right) \times \left(\frac{1-r}{2}\right) = \left(\frac{1-r}{2}\right)^2 = \frac{(1-r)^2}{4}$
R^2	$\left(\frac{r}{2}\right) \times \left(\frac{r}{2}\right) = \frac{r^2}{4}$
PR	$\left(\frac{1-r}{2}\right) \times \left(\frac{r}{2}\right) = \frac{r(1-r)}{4}$
2PR	$2 \times \frac{r(1-r)}{4} = \frac{r(1-r)}{2}$
$P^2 + R^2$	$\frac{(1-r)^2}{4} + \frac{r^2}{4} = \frac{2r^2 - 2r + 1}{4}$
$2P^2 + 2R^2$	$2 \times \frac{2r^2 - 2r + 1}{4} = \frac{2r^2 - 2r + 1}{2}$
$P^2 + PR$	$\frac{(1-r)^2}{4} + \frac{r(1-r)}{4} = \frac{1-r}{4}$
$R^2 + PR$	$\left(\frac{r^2}{4}\right) + \left(\frac{r(1-r)}{4}\right) = \frac{r}{4}$
$P^2 + R^2 + 2PR$	$\left(\frac{1-r}{4}\right) + \left(\frac{r}{4}\right) = \frac{1}{4}$
$2P^2 + 2PR$	$2 \times \left(\frac{1-r}{4}\right) = \frac{1-r}{2}$
$2R^2 + 2PR$	$2 \times \left(\frac{r}{4}\right) = \frac{r}{2}$

⁽¹⁾ $P + R = 1/2$; $R = r/2$; $P = (1 - r)/2$.

Tabela 4. Frequências gaméticas dos genitores e as frequências genotípicas das progênie, para cruzamento do tipo $A_1A_2B_1B_2 \times A_1A_3B_1B_2$ (1:1:1:1)(1:2:1).

Gametas		A_1B_1 P	A_1B_2 R	A_3B_1 R	A_3B_2 P	Frequência gamética
A_1B_1	P	P^2_{11-11}	$PR_{11-12}^{(1)}$	PR_{13-11}	$P^2_{13-12}^{(1)}$	$(1-r)/2$
A_1B_2	R	$PR_{11-12}^{(1)}$	R^2_{11-22}	$R^2_{13-12}^{(1)}$	PR_{13-22}	$r/2$
A_2B_1	R	PR_{12-11}	$R^2_{12-12}^{(1)}$	R^2_{23-11}	$PR_{23-12}^{(1)}$	$r/2$
A_2B_2	P	$P^2_{12-12}^{(1)}$	PR_{12-22}	$PR_{23-12}^{(1)}$	P^2_{23-22}	$(1-r)/2$
Frequência gamética		$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$	

⁽¹⁾Classes não distinguíveis.

De maneira similar à apresentada anteriormente, podem-se obter as expressões para os valores da frequência de recombinação, por meio do método da máxima verossimilhança. Na Tabela 5, estão representadas as expressões deduzidas para cada um dos casos relacionados anteriormente.

Um importante aspecto a se destacar é a acurácia das estimativas obtidas para a distância entre os marcadores. Embora o LOD score significativo indique a ligação entre um par de marcas, isto não implica que a estimativa da frequência de recombinação seja acurada. Assim, no processo de mapeamento, o interesse não está apenas em se detectar a ligação entre pares de marcas, mas também em se obterem estimativas acuradas, necessárias para determinar a ordem e as distâncias dos marcadores, como foram discutidas no presente estudo.

A acurácia das estimativas de máxima verossimilhança, para as distâncias entre pares de marcadores, pode ser obtida por meio de intervalos de confiança. Assim, é necessário determinar a variância das estimativas. Segundo Maliepaard et al. (1997), a variância dos estimadores de frequência de recombinação é constituída de dois componentes: o número de eventos de recombinação, criados pela amostragem dos gametas a partir da geração parental; e a habilidade ou impossibilidade de que esses eventos possam ser detectados, para certas combinações de configurações para dois locos. O primeiro componente é determinado pela própria frequência de recombinação e pelo tamanho amostral na progênie. O segundo componente é determinado pelo tipo de segregação dos locos envolvidos, de acordo com o conteúdo de informação e a fase de ligação entre os parentais. A variância de estimadores de máxima verossimilhança pode ser aproximadamente determinada pelo inverso do índice de informação de Fisher. Assim, funções de informação podem ser estabelecidas por meio do inverso da derivada segunda da função de verossimilhança. Tais funções de informação devem ser estabelecidas para diferentes configurações de acasalamento.

Tabela 5. Esquema resumido com todas as segregações analisadas, suas classes de probabilidades observadas, função de verossimilhança e expressão geral para obtenção rápida do valor da frequência de recombinação.

Cruzamento/ segregação	Classes	Função de verossimilhança	Expressão
(1:1:1:1) (1:1:1:1)	P ² ; PR; R ²	$L(r; n_i) = \lambda(P^2)^{n_1}(PR)^{n_2}(R^2)^{n_3}$	$r = \frac{2n_3 + n_2}{2N}$
(1:1:1:1) (1:2:1)	P ² ; PR; R ² ; 2PR; P ² + R ²	$L(r; n_i) = \lambda(P^2)^{n_1}(PR)^{n_2}(R^2)^{n_3}(PR + PR)^{n_4}(P^2 + R^2)^{n_5}$	$r^3[-4N] + r^2[2(2N + n_2 + 2n_3 + n_4 + n_5)] + r[-2(N + n_2 + 2n_3 + n_4)] + (n_2 + 2n_3 + n_4) = 0$
(1:2:1)(1:1:1:1)	P ² ; PR; R ² ; 2PR; P ² + R ²	$L(r; n_i) = \lambda(P^2)^{n_1}(PR)^{n_2}(R^2)^{n_3}(PR + PR)^{n_4}(P^2 + R^2)^{n_5}$	$r^3[-4N] + r^2[2(2N + n_2 + n_3 + n_4 + n_5)] + r[-2(N + n_2 + 2n_3 + n_4)] + (n_2 + 2n_3 + n_4) = 0$
(1:2:1)(1:2:1)	P ² ; R ² ; 2PR; 2P ² + 2R ²	$L(r; n_i) = \lambda(P^2)^{n_1}(2PR)^{n_2}(R^2)^{n_3}(R^2)^{n_4}(2P^2 + 2R^2)^{n_5}$	$r^3[-4N] + r^2[2(2N + n_2 + 2n_3 + n_4)] + r[-2(N + n_2 + 2n_3)] + (n_2 + 2n_3) = 0$
(1:2:1)(1:1)	P ² + PR; R ² + PR; P ² + R ² + 2PR	$L(r; n_i) = \lambda(P^2 + PR)^{n_1}(R^2)^{n_2}(R^2 + PR)^{n_3}(P^2 + R^2 + 2PR)^{n_4}$	$r = \frac{n_2}{n_1 + n_2}$
(1:1) (1:2:1)	P ² + PR; R ² + PR; P ² + R ² + 2PR	$L(r; n_i) = \lambda(P^2 + PR)^{n_1}(R^2)^{n_2}(R^2 + PR)^{n_3}(P^2 + R^2 + 2PR)^{n_4}$	$r = \frac{n_2}{n_1 + n_2}$
(1:1) (1:1)	2P ² + 2PR; 2R ² + 2PR	$L(r; n_i) = \lambda(2P^2 + PR)^{n_1}(R^2)^{n_2}(2R^2 + PR)^{n_3}$	$r = \frac{n_2}{N}$
(1:1) (1:1:1:1)	P ² + PR; R ² + PR	$L(r; n_i) = \lambda(P^2 + PR)^{n_1}(R^2 + PR)^{n_2}$	$r = \frac{n_2}{N}$
(1:1:1:1) (1:1)	P ² + PR; R ² + PR	$L(r; n_i) = \lambda(P^2 + PR)^{n_1}(R^2 + PR)^{n_2}$	$r = \frac{n_2}{N}$

No presente estudo, foram abordadas diferentes configurações para pares de marcas na análise de ligação em famílias de irmãos completos. Verificou-se que as diferentes configurações utilizadas variam em relação ao poder de detecção da ligação e na impossibilidade ou possibilidade de se estimar a fase de ligação para ambos os parentais. Assim, a acurácia do mapa genético depende da quantidade de informação, contida no grupo de dados utilizado para a construção do mapa. Obviamente, a quantidade de informação depende do delineamento experimental utilizado e da configuração dos locos em análise.

Conforme observado para o cenário 2, diferentes conteúdos de informação são obtidos para as estimativas de frequência de recombinação, de acordo com a configuração dos locos em análise. Assim, após os marcadores terem sido alocados nos grupos de ligação, informações conflitantes sobre a ordem dos marcadores nos grupos de ligação podem surgir, em razão das diferentes estimativas das frequências de recombinação, inerentes à configuração dos marcadores nos parentais.

Para o presente estudo, foi abordada a possibilidade de se estimarem as frequências de recombinação, com base na distribuição marginal dos gametas parentais para

cada genitor, de modo a se obter um mapa de ligação para cada parental, à semelhança de um esquema de “pseudo-testcross” (Grattapaglia & Sederoff, 1994). Alternativamente, pode-se utilizar a distribuição conjunta, observada na progênie para a construção de um mapa de ligação único. Verificou-se que as estimativas de percentagem de recombinação podem variar entre as diferentes abordagens, conforme o conteúdo de informação dos locos analisados. Quando se dispõe de locos completamente informativos, ambas as alternativas se igualam, isto é, a estimativa feita com base na média das frequências marginais se iguala com a estimativa baseada na frequência conjunta, como pode ser visto na análise do cenário 1; entretanto, quando são utilizados diferentes combinações de locos quanto ao seu conteúdo de informação, as estimativas nem sempre são iguais. Assim, a construção de mapas integrados pode variar, em relação aos mapas obtidos para cada parental.

Conclusões

1. Para locos completamente informativos, pode-se calcular a frequência de recombinação entre pares de locos, a partir da frequência gamética de cada genitor,

assim como a partir da frequência genotípica da progênie, que é a média aritmética da frequência de recombinação obtida com base na frequência gamética ou individual.

2. Para locos parcialmente informativos, o cálculo da frequência de recombinação, a partir das frequências gaméticas, pode não ser recomendável, pois parte da progênie não será informativa em relação à origem dos seus alelos; nesse caso, a obtenção da frequência de recombinação, a partir da frequência genotípica conjunta, é mais apropriada.

3. A construção de mapas de ligação para cada um dos genitores, para casos de existência de locos não completamente informativos, poderá proporcionar estimativas menos precisas da percentagem de recombinação, exclusão de muitas informações de progênies e necessidade posterior de emprego de técnicas de integração, uma vez que dois mapas serão gerados.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, por concessão de bolsa de estudos.

Referências

- GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. **Genetics**, v.137, p.1121-1137, 1994.
- HASEMAN, J.K.; ELSTON, R.C. The investigation of linkage between a quantitative trait and a marker locus. **Behavior Genetics**, v.2, p.3-19, 1972.
- LIU, B.H. **Statistical genomics: linkage, mapping and QTL analysis**. Boca Raton: CRC Press, 1998. 611p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980p.
- MALIEPAARD, C.; JANSEN, J.; VAN OOIJEN, J.W. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetics Research**, v.70, p.237-250, 1997.

Recebido em 22 de outubro de 2007 e aprovado em 18 de março de 2008