

ANÁLISE DE LIGAÇÃO ENTRE MARCADORES DOMINANTES E CO-DOMINANTES EM FAMÍLIAS DE IRMÃOS COMPLETOS DE ESPÉCIES ALÓGAMAS

Alexandre Alonso Alves¹, Leonardo Lopes Bhering², Cosme Damião Cruz³ e Acelino Couto Alfenas⁴

Resumo

Tendo em vista que ferramentas genômicas como microarranjos de DNA têm levado ao desenvolvimento de novos procedimentos de genotipagem, é provável que no futuro mapas de alta densidade sejam construídos a partir de marcadores dominantes e co-dominantes. Recentemente, uma abordagem estritamente genética foi descrita para a estimação da frequência de recombinação (r) entre marcadores moleculares em famílias de irmãos completos de espécies alógamas. Entretanto, uma configuração em particular envolvendo marcadores dominantes que segregam na razão de 3:1, e marcadores co-dominantes não foi considerada. Neste trabalho, nove estimadores para cálculo de r entre marcadores dominantes e co-dominantes são apresentados, de modo que todas as combinações entre marcadores moleculares (dois a quatro alelos), incluindo segregação em um ou ambos os genitores, dominância e todas as fases de ligação em famílias de irmãos completos são contempladas.

Introdução

Recentemente, em um trabalho extensivo de mapeamento genético em famílias de irmãos completos, Bhering *et al.*, (2008) usando uma abordagem estritamente genética obtiveram estimadores para a frequência de recombinação entre marcadores moleculares diferentes daqueles apresentados anteriormente por Malliepaard *et al.*, (1997). Com base neste trabalho, um módulo de mapeamento de populações exogâmicas foi implementado no software GQMOL (Cruz, 2009), amplamente utilizado no Brasil para o mapeamento genético e análise de QTLs (Cruz *et al.*, 2009). Entretanto, uma configuração em particular não foi tratada nesse trabalho, tendo em vista que a estimação da distância entre marcadores dominantes que segregam na razão de 3:1 e marcadores co-dominantes não foi considerada. Poderia se argumentar que esta não é necessária, uma vez que a maioria dos marcadores moleculares atualmente em uso para mapeamento genético em culturas agrícolas ou espécies florestais são co-dominantes como, por exemplo, os microssatélites (SSRs). Todavia, considerando-se que no futuro mapas de alta densidade sejam construídos a partir de marcadores dominantes e co-dominantes tais estimadores são necessários. Como extensão do trabalho de Bhering *et al.*, (2008), no presente trabalho nove estimadores da frequência de recombinação entre marcadores dominantes e marcadores co-dominantes são apresentados, de modo que todas as combinações entre marcadores moleculares em famílias de irmãos completos são contempladas, incluindo segregação em um ou ambos os genitores, dominância e todas as possíveis fases de ligação. Os estimadores foram desenvolvidos com base nas frequências genotípicas esperadas para cada classe. Essas frequências foram utilizadas para construir funções de máxima verossimilhança para cada possível configuração de marcadores. Com base em suas propriedades, e pelo fato de estarem implementados em um software livre (GQMOL), esses estimadores devem ser úteis para grupos de pesquisa, cujo interesse é o mapeamento e/ou a utilização de marcadores moleculares para auxiliar a seleção de características monogênicas objeto de melhoramento em espécies alógamas.

¹ Primeiro Autor é Aluno de Doutorado do Programa de Pós-graduação em Genética e Melhoramento, Universidade Federal de Viçosa, Viçosa, MG, CEP 36570-000. E-mail: alexandre.alves@ufv.br

² Segundo Autor é Pesquisador da Embrapa Agroenergia, Brasília, DF, CEP 70770-901. Email: leonardo.bhering@embrapa.br

³ Terceiro Autor é Professor Titular do Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, CEP 36570-000. E-mail: cdcruz@ufv.br

⁴ Quarto Autor é Professor Titular do Departamento de Fitopatologia, Universidade Federal de Viçosa, Viçosa, MG, CEP 36570-000. E-mail: aalfenas@ufv.br

Material e Métodos

Considere que os genótipos de dois indivíduos de uma população alógama para um marcador em particular seja A_1A_2 e A_3A_4 , respectivamente. Se esses dois indivíduos forem cruzados para formar uma família de irmãos-completos o padrão de segregação esperado é: $1A_1A_3$: $1A_1A_4$: $1A_2A_3$: $1A_2A_4$. Considere agora que os genótipos desses mesmos dois indivíduos sejam B_1B_2 e B_3B_4 , para outro marcador. Se esses dois indivíduos forem cruzados para formar uma família de irmãos-completos o padrão de segregação esperado; é: $1B_1B_3$: $1B_1B_4$: $1B_2B_3$: $1B_2B_4$. Considerando os gametas produzidos por esses dois indivíduos 16 classes genotípicas são esperadas na progênie. Considerando-se ainda que $B_1 = B_3 = B$ e $B_2 = B_4 = b$, e que BB e Bb são indistinguíveis, o que torna B um marcador tipicamente dominante, a estimativa da frequência de recombinação entre esses dois marcadores pode ser obtida através da aplicação do método da máxima verossimilhança. Ao se maximizar o logaritmo natural da função de verossimilhança, o estimador para a frequência de recombinação entre os dois marcadores pode ser obtido. No entanto, diferentes tipos de cruzamentos podem ocorrer em uma família de irmãos-completos (Haseman & Elston, 1972). Portanto, para desenvolver fórmulas gerais para os estimadores da frequência de recombinação entre marcadores dominantes que segregam na razão de 3:1 e marcadores co-dominantes em famílias de irmãos-completos tem-se que de considerar todos os diferentes padrões de segregação e fases de ligação do marcador co-dominante. Enquanto o genótipo para o marcador dominante é sempre Bb (para ambos os genitores), considerando os diferentes tipos de cruzamentos previstos por Haseman & Elston, (1972), os genótipos para o marcador co-dominante pode ser: 2 alelos - $A_1A_1 \times A_1A_2$, $A_1A_2 \times A_2A_2$, $A_1A_2 \times A_1A_2$; 3 alelos - $A_1A_1 \times A_2A_3$, $A_1A_2 \times A_3A_3$, $A_1A_2 \times A_1A_3$, $A_1A_2 \times A_2A_3$ e para 4 alelos - $A_1A_2 \times A_3A_4$. Esses alelos podem ainda estar ligados em acoplamento em um dos genitores e indefinido no segundo; em repulsão pode em um dos genitores e indefinido no segundo, ou alelos podem estar ligados em acoplamento ou repulsão em ambos os genitores (Maliepaard *et al.*, 1997).

Resultados e Discussão

Considere que os genótipos de dois indivíduos de uma população alógama para um marcador em particular seja A_1A_2 e A_3A_4 , respectivamente. Se esses dois indivíduos forem cruzados para formar uma família de irmãos-completos o padrão de segregação esperado é: $1A_1A_3$: $1A_1A_4$: $1A_2A_3$: $1A_2A_4$. Considerando-se que o genótipo destes mesmos dois indivíduos sejam Bb para um marcador dominante ligado ao marcador A . Uma função de verossimilhança pode ser usada para calcular a frequência de recombinação entre esses dois marcadores se as frequências esperadas para cada classe genotípica forem conhecidas. Para tanto, considere os haplótipos para os marcadores no primeiro genitor em fase de acoplamento e suas frequências como: $f(A_1B) = f(A_2b) = (1-r)/2 = P$ e $f(A_1b) = f(A_2B) = r/2 = R$. E para o segundo genitor como: $f(A_3B) = f(A_4b) = (1-r)/2 = P$ e $f(A_3b) = f(A_4B) = r/2 = R$. Considerando os gametas produzidos por ambos os indivíduos, a estimação da frequência de recombinação entre esses dois genitores pode ser realizada com base nas frequências esperadas para cada classe genotípicas considerando que BB e Bb são indistinguíveis. Aplicando o método da máxima verossimilhança, a função de verossimilhança pode ser expressa por:

$$L(r, ni) = \prod_{i=1}^n p_i^{ni}, \text{ que é}$$

$$L(r; ni) = \lambda (1/4 - R^2)^{na} (R^2)^{nb} (1/4 - PR)^{nc} (PR)^{nd} (1/4 - PR)^{ne} (PR)^{nf} (1/4 - P^2)^{ng} (P^2)^{nh},$$

onde: PP é $(1-r)^2/4$, PR é $r(1-r)/4$, RR é $r^2/4$, n_a número total de indivíduos com o genótipo $A_1A_3B_{-}$, n_b número total de indivíduos com o genótipo A_1A_3bb , n_c número total de indivíduos com o genótipo $A_1A_4B_{-}$, n_d número total de indivíduos com o genótipo A_1A_4bb , n_e número total de indivíduos com o genótipo $A_2A_3B_{-}$, n_f número total de indivíduos com o genótipo A_2A_3rr , n_g número total de indivíduos

com o genótipo $A_2A_4B_-$, n_h número total de indivíduos com o genótipo A_2A_4bb e N número total de indivíduos.

Ao se maximizar o logaritmo natural da função de verossimilhança como se segue

$$\frac{\partial \ln(L)}{\partial r} = \sum_{i=a}^h \frac{n_i}{p_i} \frac{\partial p_i}{\partial r} = 0$$

o estimador para a frequência de recombinação entre os dois marcadores é obtido (Tabela 1). Entretanto, os alelos de cada genitor podem estar ligados não em acoplamento, mas em repulsão e existe também possibilidade que os alelos de um genitor estejam ligados em acoplamento enquanto os alelos do outro genitor estejam ligados em repulsão e *vice versa*. Os estimadores da frequência de recombinação entre marcadores dominantes e co-dominantes, para todos os tipos de cruzamentos, segregações e fases de ligação em famílias de irmãos completos de espécies alógamas, obtidos segundo o procedimento acima descrito são apresentados na Tabela 1.

Conclusões

A partir de uma perspectiva genética, toda a gama de situações envolvendo marcadores moleculares em famílias de irmãos completos de espécies alógamas, isto é diferentes combinações entre marcadores moleculares com dois a quatro alelos (sem epistasia), incluindo segregação em um ou ambos os genitores, dominância e todas as possíveis fases de ligação em cruzamentos com espécies alógamas (famílias de irmãos-completos), está contemplada com a adição nove novos estimadores àqueles apresentados por Bhering *et al.*, (2008). Com base nas propriedades desses estimadores, e pelo fato destes estarem implementados em um software livre (GQMOL), estes devem ser excepcionalmente úteis para grupos de pesquisa, cujo interesse é o mapeamento e/ou a utilização de marcadores moleculares para auxiliar a seleção de características monogênicas em espécies alógamas.

Agradecimentos

Este trabalho contou com apoio do CNPq, através da concessão de uma bolsa de doutorado para o primeiro autor e de bolsas de produtividade em pesquisa para o terceiro e quarto autores, e da CAPES e FAPEMIG através da concessão de financiamento de pesquisa.

Referências

- BHERING, L.L.; CRUZ, C.D.; GOD, P.I.V.G. Estimativa de frequência de recombinação no mapeamento genético de famílias de irmãos completos. *Pesquisa Agropecuária Brasileira* n. 43, p. 363-369. 2008.
- CRUZ, C.D. GQMOL – Software de Análise em Genética Quantitativa e Molecular. Viçosa, 2009. Disponível para download em: <http://www.ufv.br/dbg/gqmol/gqmol.htm>. Acesso em 21 maio.2009.
- CRUZ, C.D.; GOD, P.I.V.G.; BHERING, L.L. Mapeamento de QTLs em Populações Exogâmicas. In: BORÉM, A.; CAIXETA, E.T. *Marcadores Moleculares*. 2^{ed}. Viçosa: Editora Folha de Viçosa, 2009. p. 443-482.
- HASEMAN, J.K.; ELSTON, R.C. The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* n. 2, p. 3-19. 1972.
- MALIEPAARD, C.; JANSSEN, J.; VAN OOIJEN, J.W. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genetics Research* n. 70, p. 237-250. 1997.

Tabela 1: Expressões para o cálculo da frequência de recombinação entre marcadores dominantes e co-dominantes em famílias de irmãos completos de espécies alógamas (diferentes tipos de cruzamentos – Cruz. e fases de ligação - FL são consideradas).

Cruz.	FL	Estimadores
$A_1A_1 \times A_1A_2$	A^*	$r^3 \cdot (N) - r^2 \cdot (2 \cdot b + 3 \cdot c + d) - r \cdot (a + b - 2 \cdot (c - d)) + 2 \cdot b = 0$
$A_1A_1 \times A_2A_3$		
$A_1A_2 \times A_2A_2$	R	$r^3 \cdot (N) - r^2 \cdot (3 \cdot a + b + 2 \cdot d) + r \cdot (2 \cdot a - 2 \cdot b - c - d) + 2 \cdot d = 0$
$A_1A_2 \times A_3A_3$		
	A	$2 \cdot r^7 \cdot (N) - r \cdot (2 \cdot a + 2 \cdot b + c + d + 4 \cdot f) - 2 \cdot r^6 \cdot (4 \cdot a + 5 \cdot b + 6 \cdot c + 5 \cdot d + 4 \cdot e + f) + r^5 \cdot (14 \cdot a + 16 \cdot b + 10 \cdot c + 11 \cdot d + 2 \cdot (5 \cdot e + 3 \cdot f)) - r^4 \cdot (14 \cdot a + 6 \cdot b - 8 \cdot c + 3 \cdot d + 2 \cdot (e + 2 \cdot f)) + r^3 \cdot (4 \cdot a - 10 \cdot b - 9 \cdot c - 2 \cdot (5 \cdot d + 4 \cdot e + f)) + r^2 \cdot (14 \cdot b + 2 \cdot c + 9 \cdot d + 4 \cdot (2 \cdot e + f)) - 2 \cdot (2 \cdot b + d + e) = 0$
	AR	$(2 \cdot r - 1) \cdot (2 \cdot r^4 \cdot (N) - 4 \cdot r^3 \cdot (a + b + c + d + e + f) + r^2 \cdot (3 \cdot a + 5 \cdot b + 3 \cdot c + 4 \cdot d + 3 \cdot e + 5 \cdot f) - r \cdot (a + 3 \cdot b + c + 2 \cdot d + e + 3 \cdot f) + b + f) = 0$
$A_1A_2 \times A_1A_2$	R	$2 \cdot r^7 \cdot (N) - r^6 \cdot (4 \cdot b + c + d + 2 \cdot (e + f)) - 2 \cdot r^5 \cdot (4 \cdot a + b + 6 \cdot c + 5 \cdot d + 4 \cdot e + 5 \cdot f) + r^4 \cdot (10 \cdot a + 6 \cdot b + 10 \cdot c + 11 \cdot d + 2 \cdot (7 \cdot e + 8 \cdot f)) - r^3 \cdot (2 \cdot a + 4 \cdot b - 8 \cdot c + 3 \cdot d + 2 \cdot (7 \cdot e + 3 \cdot f)) - r^2 \cdot (8 \cdot a + 2 \cdot b + 9 \cdot c + 2 \cdot (5 \cdot d - 2 \cdot e + 5 \cdot f)) + r \cdot (8 \cdot a + 4 \cdot b + 2 \cdot c + 9 \cdot d + 14 \cdot f) - 2 \cdot (a + d + 2 \cdot f) = 0$
	A	$2 \cdot r^7 \cdot (N) - r^6 \cdot (2 \cdot a + 2 \cdot b + c + d + e + f + 4 \cdot h) - 2 \cdot r^5 \cdot (4 \cdot a + 5 \cdot b + 6 \cdot c + 5 \cdot d + 6 \cdot e + 5 \cdot f + 4 \cdot g + h) + r^4 \cdot (14 \cdot a + 16 \cdot b + 10 \cdot c + 11 \cdot d + 10 \cdot e + 11 \cdot f + 2 \cdot (5 \cdot g + 3 \cdot h)) - r^3 \cdot (14 \cdot a + 6 \cdot b - 8 \cdot c + 3 \cdot d - 8 \cdot e + 3 \cdot f + 2 \cdot (g + 2 \cdot h)) + r^2 \cdot (4 \cdot a - 10 \cdot b - 9 \cdot c - 10 \cdot d - 9 \cdot e - 2 \cdot (5 \cdot f + 4 \cdot g + h)) + r \cdot (14 \cdot b + 2 \cdot c + 9 \cdot d + 2 \cdot e + 9 \cdot f + 4 \cdot (2 \cdot g + h)) - 2 \cdot (2 \cdot b + d + f + g) = 0$
	AR	$2 \cdot r^7 \cdot (N) - r^6 \cdot (a + b + 2 \cdot c + 2 \cdot d + 4 \cdot f + g + h) - 2 \cdot r^5 \cdot (6 \cdot a + 5 \cdot b + 4 \cdot c + 5 \cdot d + 4 \cdot e + f + 6 \cdot g + 5 \cdot h) + r^4 \cdot (10 \cdot a + 11 \cdot b + 14 \cdot c + 16 \cdot d + 10 \cdot e + 6 \cdot f + 10 \cdot g + 11 \cdot h) + r^3 \cdot (8 \cdot a - 3 \cdot b - 14 \cdot c - 6 \cdot d - 2 \cdot e - 4 \cdot f + 8 \cdot g - 3 \cdot h) - r^2 \cdot (9 \cdot a + 10 \cdot b - 4 \cdot c + 10 \cdot d + 8 \cdot e + 2 \cdot f + 9 \cdot g + 10 \cdot h) + r \cdot (2 \cdot a + 9 \cdot b + 14 \cdot d + 8 \cdot e + 4 \cdot f + 2 \cdot g + 9 \cdot h) - 2 \cdot (b + 2 \cdot d + e + h) = 0$
$A_1A_2 \times A_1A_3$	RA	$2 \cdot r^7 \cdot (N) - r^6 \cdot (a + b + 4 \cdot d + 2 \cdot e + 2 \cdot f + g + h) - 2 \cdot r^5 \cdot (6 \cdot a + 5 \cdot b + 4 \cdot c + d + 4 \cdot e + 5 \cdot f + 6 \cdot g + 5 \cdot h) + r^4 \cdot (10 \cdot a + 11 \cdot b + 10 \cdot c + 6 \cdot d + 14 \cdot e + 16 \cdot f + 10 \cdot g + 11 \cdot h) + r^3 \cdot (8 \cdot a - 3 \cdot b - 2 \cdot c - 4 \cdot d - 14 \cdot e - 6 \cdot f + 8 \cdot g - 3 \cdot h) - r^2 \cdot (9 \cdot a + 10 \cdot b + 8 \cdot c + 2 \cdot d - 4 \cdot e + 10 \cdot f + 9 \cdot g + 10 \cdot h) + r \cdot (2 \cdot a + 9 \cdot b + 8 \cdot c + 4 \cdot d + 14 \cdot f + 2 \cdot g + 9 \cdot h) - 2 \cdot (b + c + 2 \cdot f + h) = 0$
$A_1A_2 \times A_2A_3$		
$A_1A_2 \times A_3A_4$	R	$2 \cdot r^7 \cdot (N) - r^6 \cdot (4 \cdot b + c + d + e + f + 2 \cdot (g + h)) - 2 \cdot r^5 \cdot (4 \cdot a + b + 6 \cdot c + 5 \cdot d + 6 \cdot e + 5 \cdot f + 4 \cdot g + 5 \cdot h) + r^4 \cdot (10 \cdot a + 6 \cdot b + 10 \cdot c + 11 \cdot d + 10 \cdot e + 11 \cdot f + 2 \cdot (7 \cdot g + 8 \cdot h)) - r^3 \cdot (2 \cdot a + 4 \cdot b - 8 \cdot c + 3 \cdot d - 8 \cdot e + 3 \cdot f + 2 \cdot (7 \cdot g + 3 \cdot h)) - r^2 \cdot (8 \cdot a + 2 \cdot b + 9 \cdot c + 10 \cdot d + 9 \cdot e + 2 \cdot (5 \cdot f - 2 \cdot g + 5 \cdot h)) + r \cdot (8 \cdot a + 4 \cdot b + 2 \cdot c + 9 \cdot d + 2 \cdot e + 9 \cdot f + 14 \cdot h) - 2 \cdot (a + d + f + 2 \cdot h) = 0$

* A- acoplamento em ambos os genitores, R- repulsão em ambos os genitores, AR – acoplamento no primeiro genitor e repulsão no segundo genitor e RA repulsão no primeiro genitor e acoplamento no segundo genitor