



Construção de Mapas Genéticos, Mapeamento de QTLs e Aplicações no Melhoramento de Plantas

*Departamento de Genética
Escola Superior de Agricultura "Luiz de Queiroz"
Universidade de São Paulo*

Cristiane Hayumi Taniguti

Orientador: Antonio Augusto Franco Garcia

















- ▶ Introdução
- ▶ Construção de Mapas Genéticos
- ▶ Princípios de Mapeamento de QTLs
- ▶ Comparação com Mapeamento Associativo (GWAS)
- ▶ Recomendações
- ▶ Referências

Mendel

1865



- ▶ Segregação independente
- ▶ Proporções 3:1, 1:2:1

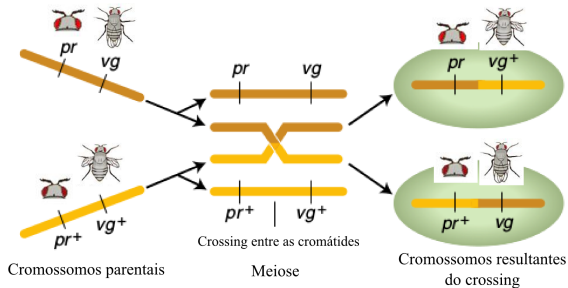
	<i>Forma da vagem</i>	<i>Cor da vagem</i>	<i>Forma da semente</i>	<i>Cor da semente</i>	<i>Cor da casca</i>	<i>Posição das flores</i>	<i>Altura da planta</i>
<i>Recessivo</i>	 Comprimida	 Amarela	 Rugosa	 Verde	 Branca	 Terminal	 Baixa
<i>Dominante</i>	 Inflada	 Verde	 Lisa	 Amarela	 Alta	 Auxilar	 Alta



Mapa de Morgan e Sturtevant

1913

- ▶ Dois locos ligados
- ▶ Distância proporcional ao número de recombinações





Marcadores Genéticos

▶ Marcadores morfológicos

- ▶ Raros
- ▶ Sofrem forte influência ambiental

▶ Marcadores moleculares

- ▶ Abundantes
- ▶ Sofrem pouca influência ambiental
- ▶ Exigem esforço em bancada

▶ Marcadores moleculares providos de tecnologias de sequenciamento

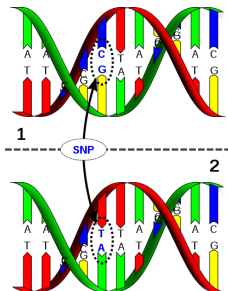
- ▶ Mais Abundantes ainda
- ▶ Sofrem pouca influência ambiental
- ▶ Facilidade de obtenção
- ▶ Exigem esforço computacional



Marcadores Genéticos

Single-Nucleotide Polymorphism

- ▶ Vantagens
 - ▶ Abundância
 - ▶ Custo
 - ▶ Automatização
- ▶ Desvantagens
 - ▶ Baixa informatividade (F1 segregante, poliplóides)
 - ▶ Erros na chamada de SNPs



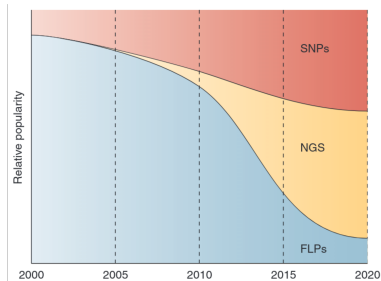
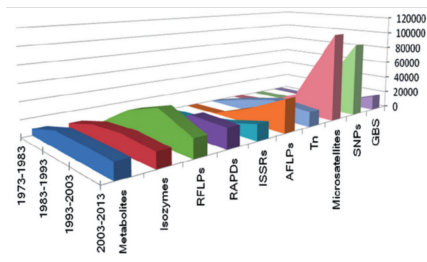
Marcadores Genéticos

Frequência de uso



[Nadeem et al., 2017]

DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing

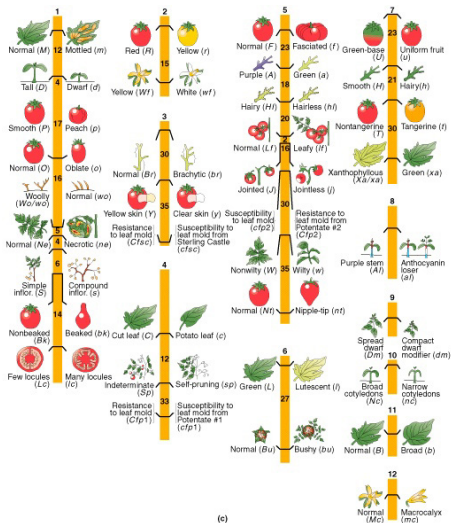


[Grover and Sharma, 2014] [Seeb et al., 2011]

Um mapa de 1952



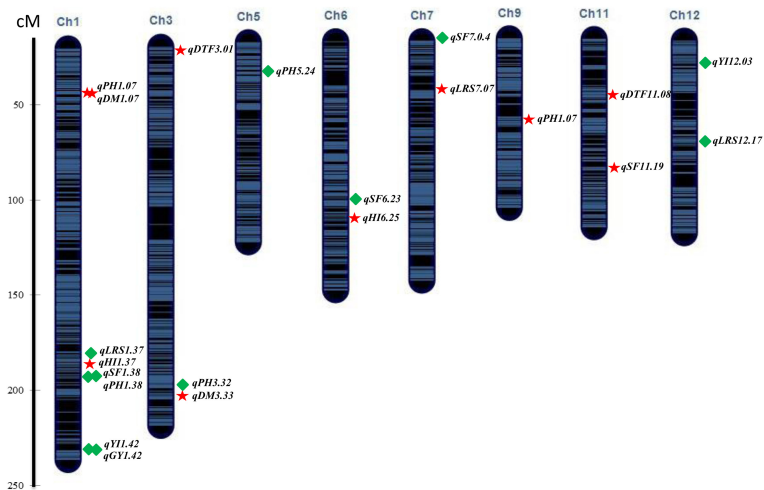
► Marcadores Morfológicos



"The Tomato", C. M. Rick Copyright 1978 de Scientific American, Inc

Mapas agora

Arroz

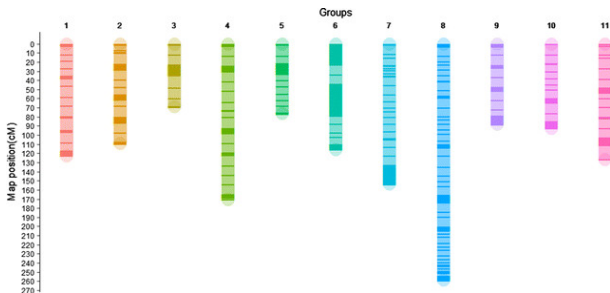


4748 SNPs e 181 RILs [Bhattarai and Subudhi, 2018]



Mapas agora

Jatropha curcas purgueira, jatropa ou pinhão-manso



3422 SNPs e 153 indivíduos F1 segregante [Xia et al., 2018]



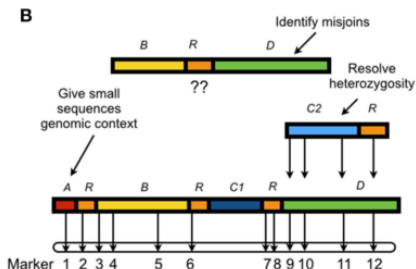
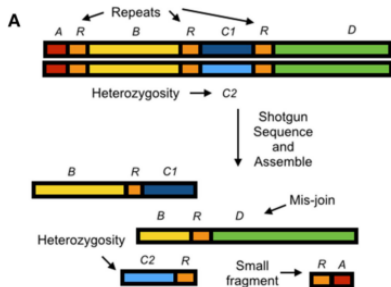
Aplicações

► Mapas densos e auxílio na montagem de genomas



[Fierst, 2015]

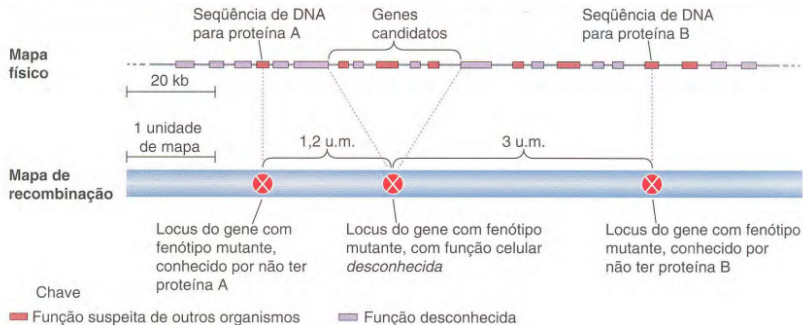
Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools





Aplicações

- Complementariedade com mapas físicos para estudo de funções gênicas



[Griffiths et al., 2004]

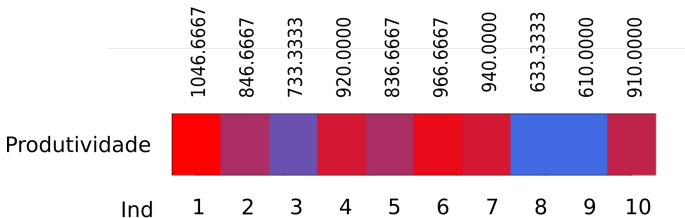


Aplicações

A ideia de associação genótipo x fenótipo

Conjunto de dados exemplo: Produtividade de Feijão em Lavras

► 10 indivíduos

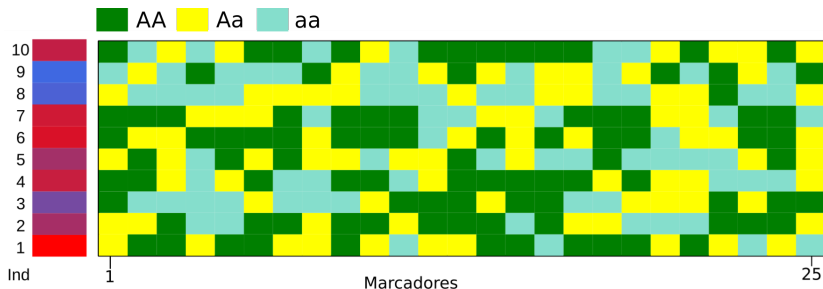




Aplicações

A ideia de associação genótipo x fenótipo

- ▶ 25 marcadores
- ▶ 15 marcadores ligados à característica
- ▶ 10 marcadores não ligados à característica

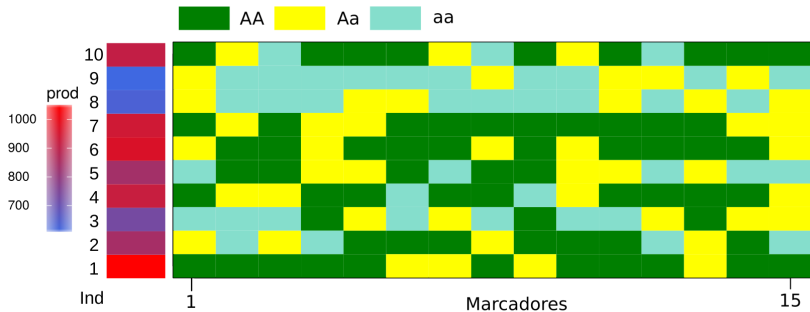




Aplicações

A ideia de associação genótipo x fenótipo

- ▶ 25 marcadores
- ▶ 15 marcadores ligados à característica

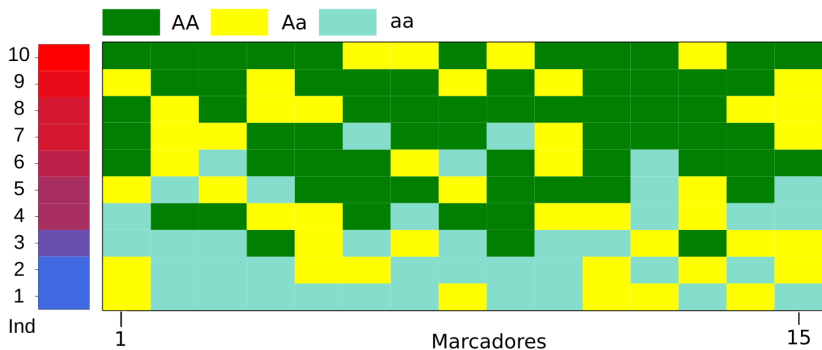




Aplicações

A ideia de associação genótipo x fenótipo

- ▶ 25 marcadores
- ▶ 15 marcadores ligados à característica

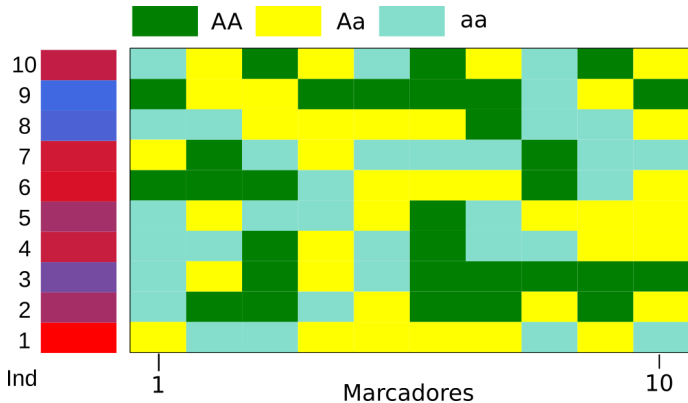




Aplicações

A ideia de associação genótipo x fenótipo

- ▶ 25 marcadores
- ▶ 10 marcadores não ligados à característica

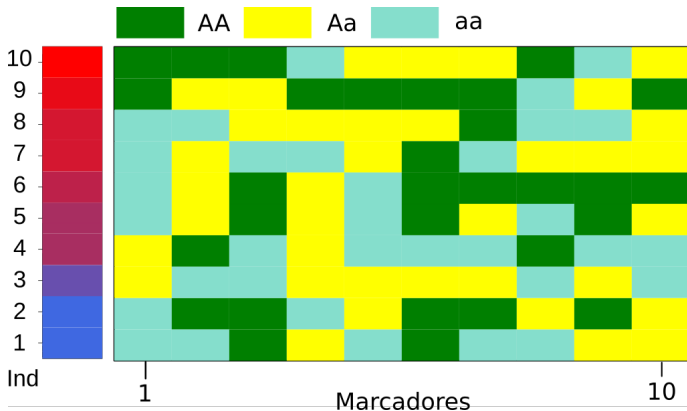




Aplicações

A ideia de associação genótipo x fenótipo

- ▶ 25 marcadores
- ▶ 10 marcadores não ligados à característica



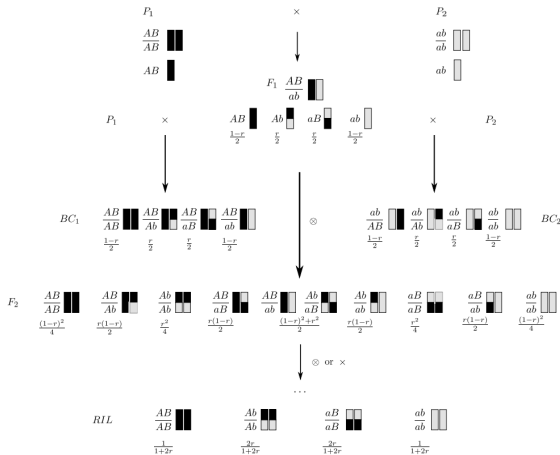
Construção de Mapas Genéticos



Construção de Mapas Genéticos

Populações de mapeamento

► Provienda de linhagens

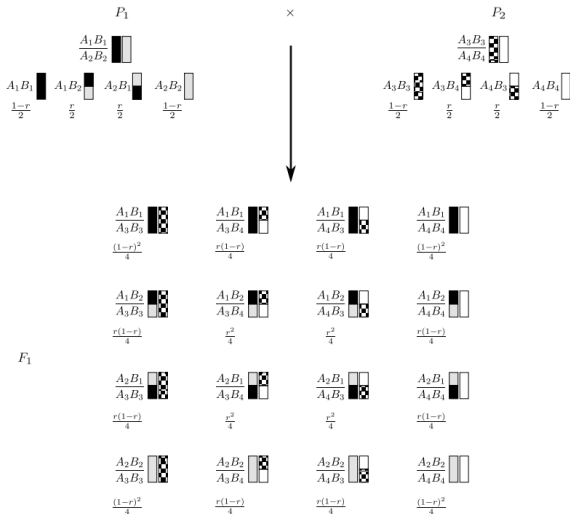




Construção de Mapas Genéticos

Populações de mapeamento

► F1 segregante





Construção de Mapas Genéticos

Obtenção de marcadores

- ▶ **Microssatélites**
 - ▶ Poliplóides e F1 segregante

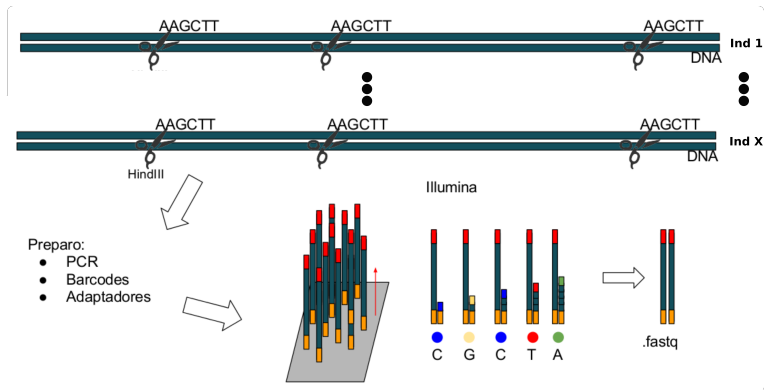
- ▶ **SNPs**
 - ▶ Restriction site-associated DNA sequencing (RAD)
 - ▶ Sequenom
 - ▶ Amplified-fragment single nucleotide polymorphism and methylation (AFSM)
 - ▶ Genotyping-by-sequencing (GBS)
 - ▶ DArT Seq e outras tecnologias arrays



Construção de Mapas Genéticos

Genotipagem por Sequenciamento

► Metodologia



[Elshire et al., 2011]



Construção de Mapas Genéticos

Chamada de SNPs

- ▶ Principais softwares
 - ▶ TASSEL, STACKS, GATK e freebayes
- ▶ Vantagem do Genoma de Referência
- ▶ Erros
 - ▶ Profundidade x Plexagem
 - ▶ Qualidade do Sequenciamento
 - ▶ Viés da PCR
 - ▶ Alinhamento



Construção de Mapas Genéticos

Filtragem dos marcadores

► Padrão de segregação

P1		P2
Aa	x	AA
Aa	AA	aa

P1		P2
aa	x	AA
Aa	AA	aa

- Dados perdidos
- Profundidade
- MAF



Construção de Mapas Genéticos

Estimação da fração de recombinação

- ▶ Estimador de máxima verossimilhança para o caso mais simples:

População de retrocruzamento

$$r = \frac{n_r}{n}$$

n_r : Número de indivíduos com genótipo recombinante

n : Número total de indivíduos na progênie

- ▶ Estimadores para todas os possíveis cruzamentos, inclusive para espécies outcrossing:



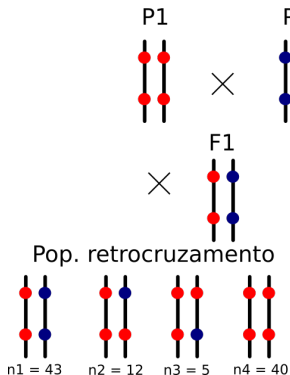
[Maliepaard et al., 1997]

Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications.



Construção de Mapas Genéticos

Estimação da fração de recombinação



$$r_{AB} = \frac{12+5}{100} = 0.17$$
$$r_{AC} = 0.22$$
$$r_{BC} = 0.3$$

Qual a ordem?
B-A-C



Construção de Mapas Genéticos

Agrupamento

Critérios para determinar se marcadores pertencem ao mesmo grupo

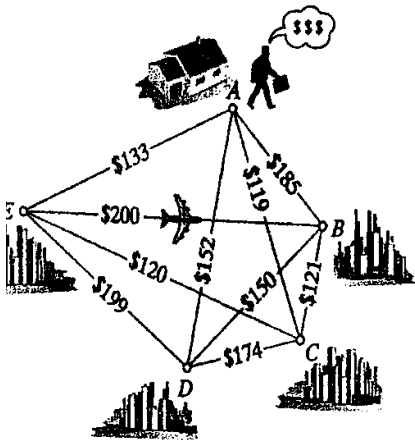
- ▶ Estimativa de dois pontos de r (inferior à 0.5)
- ▶ p-valor relativo à H_0
- ▶ LOD Score



Construção de Mapas Genéticos

Algoritmos de ordenamentos

Problema do caixeiro viajante





Construção de Mapas Genéticos

Algoritmos de ordenamento

Testar todas as ordens possíveis?

Locos	Ordens	Tempo
5	$5!/2$	Insignificante
10	1.814.400	0,0062 seg
15	653837.184.000	2,5427 h
20	$1,21 \times 10^{18}$	732,4 anos
25	$7,75 \times 10^{24}$	5.898.373.012,27 anos

Utilização de Heurística



Construção de Mapas Genéticos

Algoritmos de ordenamento

- ▶ Comando TRY (MAPMAKER/EXP)
 - ▶ Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- ▶ Comando ORDER (MAPMAKER/EXP)
 - ▶ Automatização do TRY
- ▶ Rapid Chain Delineation (RCD)
- ▶ Multidimensional scaling (MDS)
- ▶ RIPPLE



Construção de Mapas Genéticos

Estimativa Multiponto das Distâncias

Solução para os problemas

- ▶ Dados perdidos
- ▶ Erros de genótipos
- ▶ Marcadores com baixa informatividade
- ▶ Integração de mapas em F1 segregante

Uso de abordagem com Modelos Gráficos Probabilísticos



Construção de Mapas Genéticos

Funções de Mapeamento

- ▶ Morgan
 - ▶ Proporcional à fração de recombinação
 - ▶ Distâncias não podem ser somadas

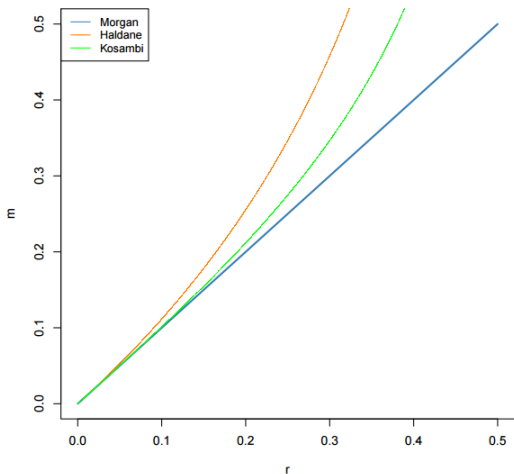
- ▶ Haldane
 - ▶ Considera ausência de interferência e distribuição de Poisson
 - ▶ Distâncias podem agora ser somadas

- ▶ Kosambi
 - ▶ Baseada em observações empíricas
 - ▶ Considera interferência
 - ▶ Geralmente a mais adequada



Construção de Mapas Genéticos

Funções de Mapeamento





Construção de Mapas Genéticos

Software OneMap



[Margarido et al., 2007]

OneMap: software for genetic mapping in outcrossing species.

- ▶ Populações diplóides: retrocruzamento, F2, RILs e F1 segregante
- ▶ Teste de Segregação
- ▶ Agrupamento
- ▶ Teste de dois pontos
- ▶ Diversos algoritmos de ordenamento



Construção de Mapas Genéticos

Software OneMap

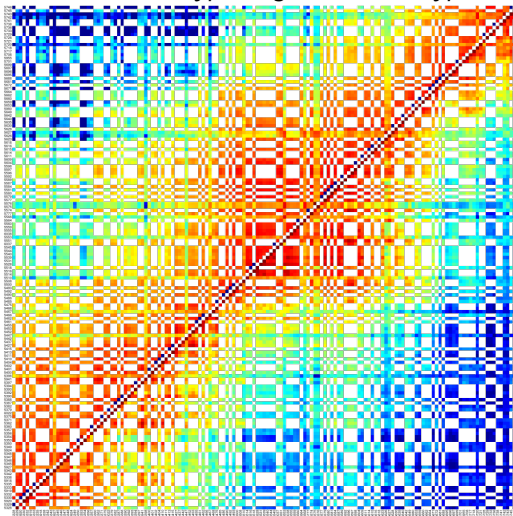
- ▶ Estimativa Multiponto
- ▶ Pondera confiabilidade dos marcadores conforme profundidade (Futuro)
- ▶ Combina SNPs em haplótipo para maior informatividade (Futuro)
- ▶ Possibilidade de utilizar mapas anteriores ou informações de genoma
- ▶ Tutoriais didáticos
- ▶ Gráficos elegantes



Construção de Mapas Genéticos

Software OneMap

Cromossomo 11 de *Eucalyptus grandis* x *Eucalyptus urophylla*





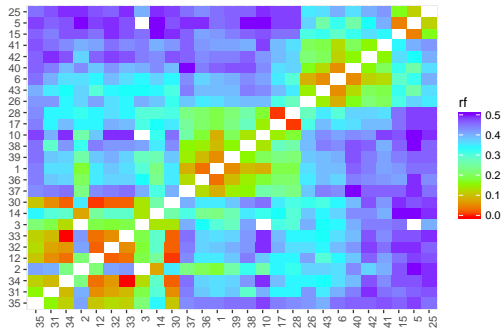
Construção de Mapas Genéticos

Software OneMap

Como saber se está bom?

- ▶ Verossimilhança: -1823.505
- ▶ Distância entre marcadores: 229.78 cM
- ▶ Padrão de cores no gráfico

Conjunto de dados exemplo do OneMap



Esse está ruim!



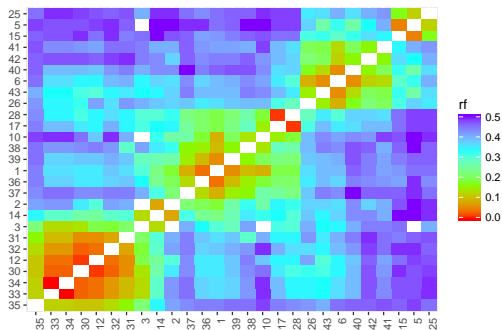
Construção de Mapas Genéticos

Software OneMap

Como saber se está bom?

- ▶ Verossimilhança: -1574.924
- ▶ Distância entre marcadores: 120.49 cM
- ▶ Padrão de cores no gráfico

Conjunto de dados exemplo do OneMap



Esse está bom!

Mapeamento de QTLs



Princípios de Mapeamento de QTLs

- ▶ O que são QTLs?
 - ▶ Quantitative Trait Loci: locos que contribuem para a variação de um caráter quantitativo
 - ▶ Mapeamento: identificação de tais locos em cruzamentos experimentais

Estudos da arquitetura genética

- ▶ Número de QTLs
- ▶ Localização dos mesmo no mapa/genoma
- ▶ Seus efeitos
- ▶ Interações entre si (epistasia)
- ▶ Interações QTLxE



Princípios de Mapeamento de QTLs

Planejamento do experimento

- ▶ Lidar com variações que não são foco do estudo
 - ▶ Diferenças de irrigação, composição do solo, fertilização, captação de luz, replantio, etc.
- ▶ Obedecer princípios básicos de experimentação:
 - ▶ Aleatorização
 - ▶ Repetições
 - ▶ Bordaduras (experimentos em campo)
 - ▶ Escolha de um delineamento que melhor controle variações do experimento específico



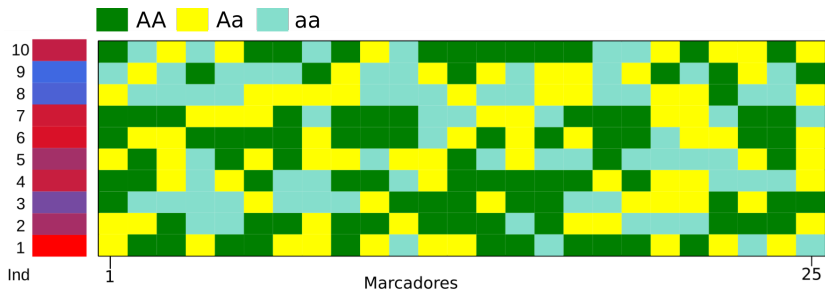
Princípios de Mapeamento de QTLs

Planejamento do experimento

- ▶ Utilização de modelo estatístico apropriado
 - ▶ Modelo fixo
 - ▶ Modelo aleatório
 - ▶ Modelo misto
- ▶ Uso dos BLUEs ou BLUPs nos modelos de QTL



Voltando ao nosso exemplo intuitivo:

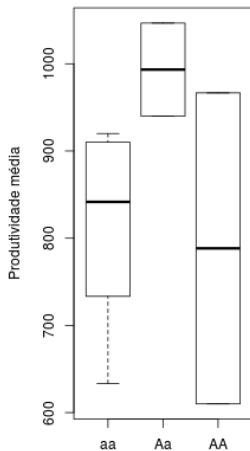
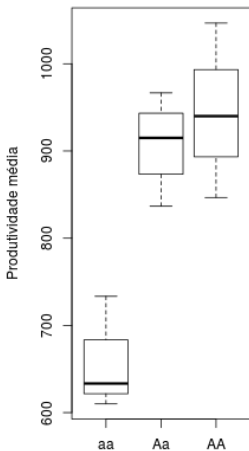




Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

- ▶ Testar se há diferenças fenotípicas entre indivíduos de diferentes classes genotípicas

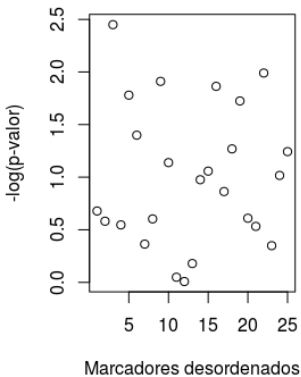
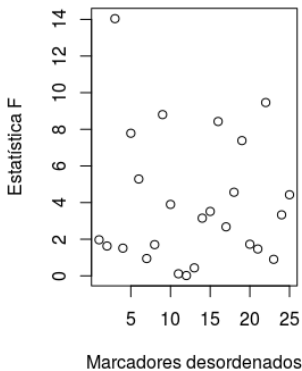




Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

- ▶ Estatística F e p-valor

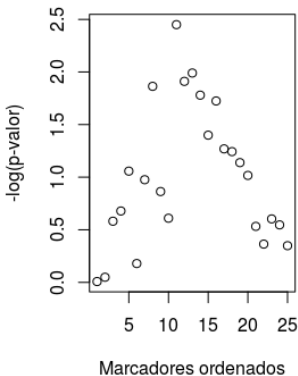
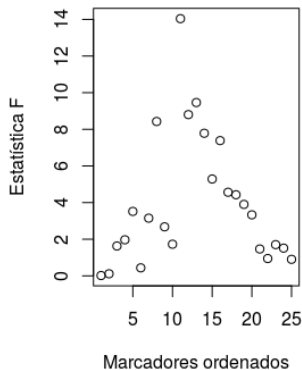




Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

► Estatística F e p-valor





Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

- ▶ Regressão Linear simples - caso de retrocruzamento

$$y_j = \mu + \beta x_j + \varepsilon_j$$

$j = 1, 2, 3, \dots, n$

y_j = valor fenotípico do indivíduo j

μ = intercepto

x_j =

$$\left\{ \begin{array}{l} 1, \text{ se o indivíduo } j \text{ tem genótipo AA} \\ 0, \text{ se o indivíduo } j \text{ tem genótipo Aa} \end{array} \right\}$$

β = coeficiente de regressão linear (efeitos genéticos)

$\varepsilon_j \sim N(0, \sigma^2)$



Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

- Regressão Linear múltipla - caso de uma população F_2

$$y_j = \mu + \beta_1 x_{1j} + \beta_2 x_{2j} + \varepsilon_j$$

$j = 1, 2, 3, \dots, n$

y_j = valor fenotípico do indivíduo j

μ = intercepto

$x_{1j} =$

$$\left\{ \begin{array}{l} 1, \text{ se o indivíduo } j \text{ tem genótipo AA} \\ 0, \text{ se o indivíduo } j \text{ tem genótipo Aa} \\ -1, \text{ se o indivíduo } j \text{ tem genótipo aa} \end{array} \right\}$$

$x_{2j} =$

$$\left\{ \begin{array}{l} -\frac{1}{2}, \text{ se o indivíduo } j \text{ tem genótipo AA} \\ 1, \text{ se o indivíduo } j \text{ tem genótipo Aa} \\ -\frac{1}{2}, \text{ se o indivíduo } j \text{ tem genótipo aa} \end{array} \right\}$$

β_1 e β_2 = coeficientes de regressão linear (efeitos genéticos aditivo e de dominância)

$\varepsilon_j \sim N(0, \sigma^2)$



Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

- ▶ Hipótese $H_0 : \beta = 0$ e $H_1 : \beta \neq 0$

Com um pouco de conhecimento em Modelos Lineares

- ▶ Obtenção dos efeitos genéticos por quadrados mínimos ou máxima verossimilhança
- ▶ Análise de variância, estatística F, p-valor, LOD score



Princípios de Mapeamento de QTLs

Mapeamento por Marca Individual

Vantagens

- ▶ Fácil implementação e entedimento
- ▶ Não requer softwares específicos
- ▶ Não precisa que o mapa tenha sido estimado
- ▶ Pode ser aplicada para marcadores não ligados
- ▶ Base para métodos como GWAS e GWS

Desvantagens

- ▶ Não é possível saber se há mais de um QTL ligado ao marcador
- ▶ Não há estimativa da posição do QTL
- ▶ Considera um único QTL, logo, não permite estudos de epistasia
- ▶ O efeito do QTL é subestimado



Princípios de Mapeamento de QTLs

Mapeamento por Intervalos



[Lander and Botstein, 1989]

Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps

- ▶ Abordagem inovadora e precursora de outros métodos de análise (CIM, MIM, etc).



Princípios de Mapeamento de QTLs

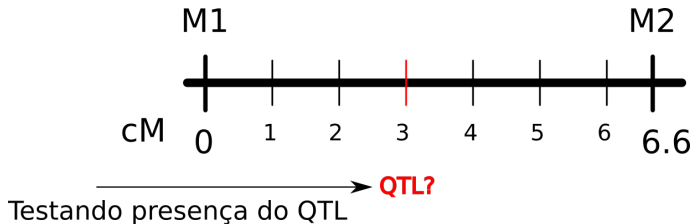
Mapeamento por Intervalos

- ▶ Percorre sistematicamente o genoma em busca de QTLs
- ▶ Extensão da análise de marcador individual, permitindo análise nos intervalos
- ▶ Usa a informação dos marcadores flanqueadores aos intervalos (Na modernidade: multiponto)



Princípios de Mapeamento de QTLs

Mapeamento por Intervalos



- Ideia central: considerar os QTLs como dados perdidos ou variáveis latentes (modelo de misturas)



Princípios de Mapeamento de QTLs

Mapeamento por Intervalos

► Modelo para retrocruzamento

$$y_j = \mu + \beta^* x_j^* + \varepsilon_j$$

$$j = 1, 2, 3, \dots, n$$

y_j = valor fenotípico do indivíduo j

μ = intercepto

$$x_j^* =$$

$$\left\{ \begin{array}{l} 1, \text{ se o genótipo do QTL do indivíduo } j \text{ é } QQ \\ 0, \text{ se o genótipo do QTL do indivíduo } j \text{ é } Qq \end{array} \right\}$$

β^* = coeficiente de regressão linear (efeito do possível QTL)

$$\varepsilon_j \sim N(0, \sigma^2)$$



Princípios de Mapeamento de QTLs

Mapeamento por Intervalos

- ▶ O modelo considera o genótipo do QTL e não dos marcadores
- ▶ Genótipo dos QTLs são uma variável não observada, ou latente. Isso caracteriza um modelo de misturas.
- ▶ Os marcadores flancueadores de um intervalo são utilizados para calcular a probabilidade do genótipo do QTL numa dada posição θ do intervalo

$$p_{kj} = P(x_j^* = k | M_i M_{i+1}, \theta)$$
$$k = 0, 1$$



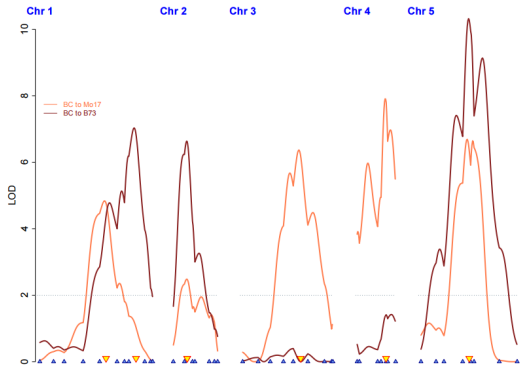
Princípios de Mapeamento de QTLs

Mapeamento por Intervalos



[Lander and Botstein, 1989]

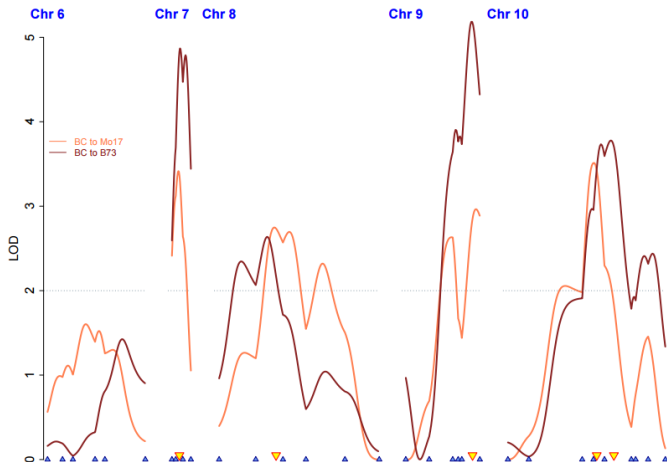
Identification of Genetic Factors Contributing to Heterosis in a Hybrid From Two Elite Maize Inbred Lines Using Molecular Markers





Princípios de Mapeamento de QTLs

Mapeamento por Intervalos





Princípios de Mapeamento de QTLs

Mapeamento por Intervalos

Cromos.	Marcador	LOD	R^2	Efeito
RC com <i>B73</i>				
1	<i>NPI255</i>	6,91	15,1	10,40
2	<i>NPIB1</i>	6,63	13,3	9,72
5	<i>Amp3</i>	9,73	18,0	11,30
7	<i>NPI216</i>	4,44	8,8	7,98
9	<i>NPI427</i>	4,80	10,3	8,70
10	<i>NPI264</i>	3,16	6,2	6,52
RC com <i>Mo17</i>				
1	<i>NPI429</i>	4,78	9,5	9,50
3	<i>NPI212</i>	6,53	14,4	12,38
4	<i>NPI444</i>	8,01	13,9	11,34
5	<i>Amp3</i>	6,86	12,9	13,72
7	<i>NPI216</i>	3,31	6,4	8,02
8	<i>BLN1.45</i>	2,73	5,8	7,68
9	<i>NPI427</i>	2,97	5,6	7,52
10	<i>Glu1</i>	3,56	6,5	7,06



Princípios de Mapeamento de QTLs

Mapeamento por Intervalos

Vantagens

- ▶ Possível fazer inferência sobre a posição do QTL
- ▶ Maior poder estatístico que as análises feitas individualmente
- ▶ Observação
 - ▶ No caso de mapas muito saturados, o IM é equivalente à análise individual dos marcadores



Princípios de Mapeamento de QTLs

Mapeamento por Intervalos

Desvantagens

- ▶ Não exclui influência de QTLs de fora da região avaliada
- ▶ Influência de outros QTLs podem gerar falsos positivos
- ▶ Não se usa toda a informação do genoma para a avaliação
- ▶ Não permite estudos de epistasia de forma apropriada



Princípios de Mapeamento de QTLs

Nível de Significância

Como determinar um nível de significância (threshold)?

Pode ser considerada distribuição χ^2 sob H_0 , mas existem complicações decorrente do grande número de testes feitos, já que percorre todo genoma

▶ Métodos analíticos

- ▶ Suposições gerais (tamanho do genoma, saturação do mapa, número de indivíduos, tipo de marcador, etc)



Princípios de Mapeamento de QTLs

Nível de Significância

Uma boa ideia: Permutações



[Doerge and Churchill, 1996]

Empirical Threshold Values for Quantitative Trait Mapping.



[Doerge, 1996]

Permutation Tests for Multiple Loci Affecting a Quantitative Character.



Princípios de Mapeamento de QTLs

Nível de Significância

Vantagens

- ▶ Threshold particular para cada conjunto de dados
- ▶ Se suposições dos métodos analíticos estiverem corretas, gera resultados semelhantes
- ▶ Método mais aceito para publicações

Desvantagem

- ▶ Tendência de ser conservativo
- ▶ Difícil de implementar para modelos muito complexos
- ▶ Elevado tempo de processamento (em algumas situações)



Princípios de Mapeamento de QTLs

Nível de Significância

Passos

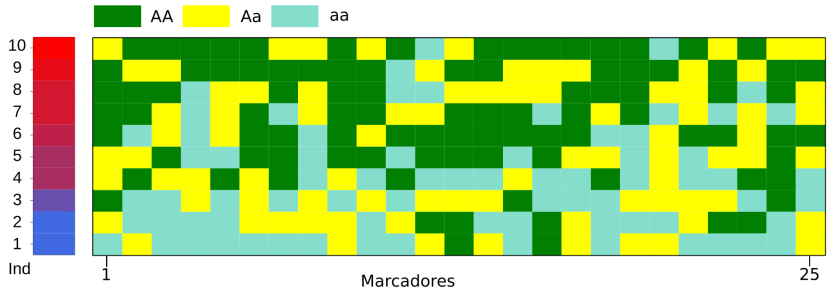
- ▶ Permutar o genótipo em relação ao fenótipo (torna os valores de associação aleatórios)
- ▶ Fazer o teste de associação (IM e outros)
- ▶ Armazenar o maior valor de LOD
- ▶ Repetir (normalmente 1000x) até que todos os valores armazenados formem um distribuição empírica
- ▶ Limiar é obtido pelo percentil da distribuição

Princípios de Mapeamento de QTLs

Nível de Significância



- ▶ Permutar o genótipo em relação ao fenótipo (torna os valores de associação aleatórios)

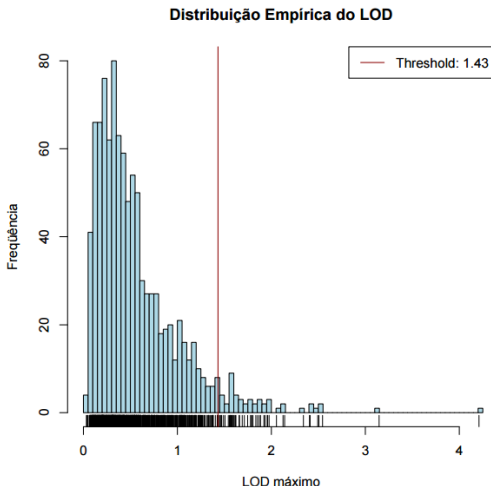




Princípios de Mapeamento de QTLs

Nível de Significância

- ▶ Limiar é obtido pelo percentil da distribuição





Princípios de Mapeamento de QTLs

Mapeamento por Intervalo Composto



[Zeng, 1993]

Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci.

- ▶ Promove idenpendência do QTL que esta sendo avaliado em relação aos outros contidos no genoma
- ▶ Inclusão de covariáveis no modelo



Princípios de Mapeamento de QTLs

Mapeamento por Intervalo Composto

$$y_j = \mu + \beta^* x_j^* + \sum_k \beta_k x_{jk} + \varepsilon_j$$

$j = 1, 2, \dots, n$

y_j = valor fenotípico do indivíduo j

μ = intercepto

x_j^* =

$$\left\{ \begin{array}{l} 1, \text{ se o genótipo do QTL é QQ} \\ 0, \text{ se o genótipo do QTL é Qq} \end{array} \right\}$$

β^* = efeito do possível QTL

x_{jk} = cofatores (marcas selecionadas para controlar variação)

β_k = efeito do k -ésimo cofator

$\varepsilon_j \sim N(0, \sigma^2)$



Princípios de Mapeamento de QTLs

Mapeamento por Intervalo Composto

Seleção das covariáveis

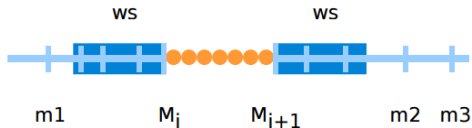
- ▶ Poucos marcadores: não há redução da variância residual
- ▶ Muitos marcadores: redução do poder da análise
- ▶ Deve ser avaliado para cada situação:
 - ▶ Análise de regressão múltipla com procedimento stepwise



Princípios de Mapeamento de QTLs

Mapeamento por Intervalo Composto

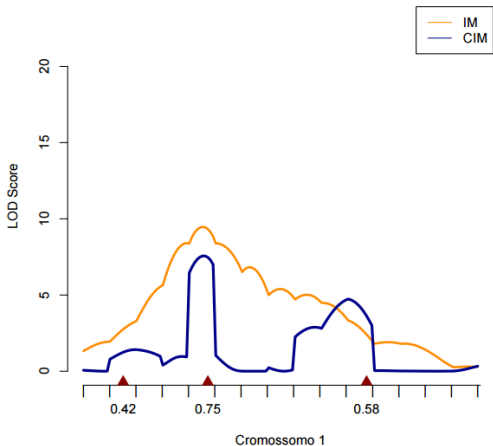
- ▶ Utiliza-se modelo stepwise para definir marcadores significativos
- ▶ Estabelece-se um tamanho para a janela adjacente ao intervalo sendo avaliado (normalmente 10 ou 15 cM)
- ▶ Marcadores selecionados que estão fora da janela são considerados cofatores
- ▶ Usa-se modelo específico (alterando os cofatores) para cada intervalo avaliado





Princípios de Mapeamento de QTLs

Mapeamento por Intervalo Composto



[Zeng, 1994]



Princípios de Mapeamento de QTLs

Mapeamento por Intervalo Composto

Metodologia expandida para F1 segregante:



[Gazaffi et al., 2014]

A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny.



Princípios de Mapeamento de QTLs

Mapeamento por Múltiplos Intervalos

Modelos CIM realizam mapeamento de um único QTL de cada vez, buscando controlar variações de QTLs de fora do intervalo.

A ideia

Já que o objetivo é controlar a variação causada por outros QTLs, por que não incluí-los diretamente no modelo? [Kao and Zeng, 1997], [Kao et al., 1999], [Zeng et al., 1999]

Exemplo: Modelo com três QTLs:

$$y_j = \mu + \beta_1^* x_{1j}^* + \beta_2^* x_{2j}^* + \beta_3^* x_{3j}^* + \varepsilon_j$$



Princípios de Mapeamento de QTLs

Mapeamento por Múltiplos Intervalos

- ▶ Modelo considera múltiplos QTLs
- ▶ Conjunto de procedimentos para mapeamento: seleção e comparação de modelos

Vantagens

- ▶ Combina alta precisão e poder estatístico
- ▶ Permite a inclusão de epistasia (interação entre locos)
- ▶ Permite o estudo da arquitetura genética dos caracteres quantitativo
- ▶ Permite estimar o número, posição, efeitos e interação entre QTLs
- ▶ Permite estimar o breeding value: seleção assistida



Princípios de Mapeamento de QTLs

Mapeamento por Múltiplos Intervalos

Uma estratégia para realizar o MIM no R/QTL

- ▶ Fazer uma regressão múltipla para selecionar marcas significativas
- ▶ Ajustar modelo CIM para cada intervalo
- ▶ Fazer um teste preliminar de interação entre as marcas
- ▶ Definir um modelo com m QTLs e t interações
- ▶ Testar cada parâmetro desse pré-modelo no contexto do MIM; eliminar os efeitos não significativos

Assim seleciona-se um pré-modelo



Princípios de Mapeamento de QTLs

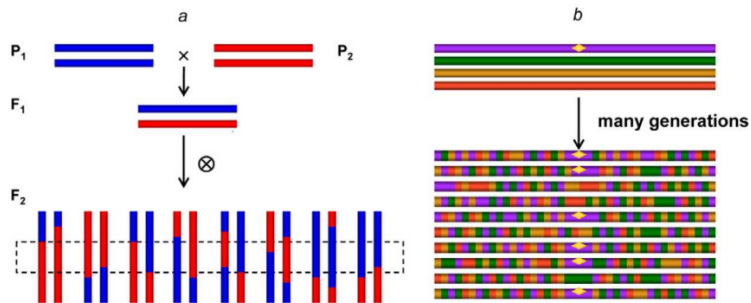
Mapeamento por Múltiplos Intervalos

- ▶ Começar com um modelo com m QTLs e t efeitos epistáticos
- ▶ Procurar ao longo do genoma um $(m + 1)$ -ésimo QTL
- ▶ Procurar um $(t + 1)$ -ésimo efeito epistático. Repetir o processo até não encontrar mais efeitos significativos
- ▶ Reavaliar os efeitos de todos os QTLs no modelo. Manter os significativos (ou com epistasia significativa)
- ▶ Otimizar as estimativas das posições dos QTLs do modelo. Normalmente, isso é feito para cada QTL individualmente, condicionado à presença dos demais QTLs no modelo
- ▶ Retornar ao passo 2 e repetir o processo até não haver mais variações (efeitos e posições otimizados)

QTLs x GWAS



Mapeamento Associativo



[Zhu et al., 2008]

- ▶ Feito com germoplasma (coleção de indivíduos, genitores, coleta, etc)
- ▶ Maior resolução



Mapeamento Associativo

Cruzamentos controlados x grupo de genótipos

Até agora:

- ▶ as frequências alélicas/genotípicas são conhecidas a priori

Exemplo: F2

- ▶ $p(A) = p(a) = 1/2$
- ▶ Se AB e ab são parentais então Ab e aB na progênie são recombinantes
- ▶ Frequencia na progênie é em função da fração de recombinação
- ▶ Estrutura populacional conhecida (todos pertencem a mesma)
- ▶ Parentesco é o mesmo



Mapeamento Associativo

Cruzamentos controlados x grupo de genótipos

Em GWAS existem outras situações:

- ▶ Necessário investigar se não há associação preferencial entre alelos (intra e inter loco)
- ▶ Desequilíbrio de ligação: pode ser quantificado usando alguma medida de associação entre os estados alélicos de pares de locos
- ▶ Pode haver desequilíbrio devido a outras causas que não a ligação genética!



Mapeamento Associativo

Desequilíbrio de ligação

Definição

DL é qualquer desvio das frequências alélicas em relação às frequências esperadas sob independência, indicando associação preferencial entre alelos de diferentes locos numa população (Lewontin & Kojima, 1960).

- ▶ DL e Ligação Física não são sinônimos
- ▶ Locos em DL podem não estar ligados
- ▶ Locos ligados podem ou não estar em DL (função do número de recombinações)



Mapeamento Associativo

Desequilíbrio de ligação

Fatores que afetam DL em grupo de genótipos:

- ▶ Recombinação (crossing-overs)
- ▶ Mutação
- ▶ Deriva Genética
- ▶ Seleção
- ▶ Migração
- ▶ Estrutura Populacional

Em mapeamento: DL é devido à ligação

Em GWAS: necessidade de explorar e considerar no modelo todas os fatores



Mapeamento Associativo

Desequilíbrio de ligação

Modelos Mistos

$$y = X\beta + S\alpha + Qv + Zu + e$$

y : fenótipos

$X\beta$: efeitos fixos do delineamento

α : efeito do SNP

v : efeito (fixo) da estrutura populacional

u : efeitos (aleatórios) dos poligenes do background genético

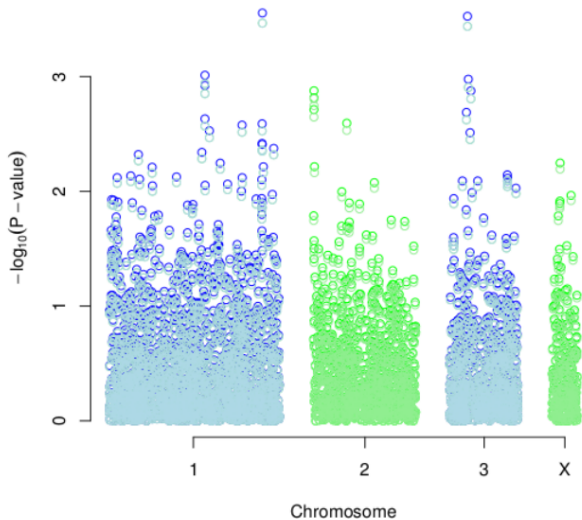
e : vetor de resíduos

$$V(u) = 2KV_g$$

K : kinship



Manhattan plot











Recomendações

- ▶ Biometria de Marcadores Genéticos - Programa de Genética e Melhoramento de Plantas (ESALQ)
- ▶ Modelos Lineares - Estatísticas e Experimentação Agronômica (ESALQ)
- ▶ Modelos Mistos - Estatísticas e Experimentação Agronômica (ESALQ)
- ▶ Tutoriais práticos
 - ▶ Curso de R
 - ▶ Construção de Mapas Genéticos a partir de linhagens
 - ▶ Construção de Mapas Genéticos em F1 segregante
 - ▶ Mapeamento de QTLs (IM e CIM)
 - ▶ Mapeamento de QTLs por Múltiplos Intervalos







-  Bhattarai, U. and Subudhi, P. K. (2018).
Genetic analysis of yield and agronomic traits under reproductive-stage drought stress in rice using a high-resolution linkage map.
Gene, 669(April):69--76.
-  Doerge, R. W. (1996).
Constructing genetic maps by rapid chain delineation.
Journal of Quantitative Trait Loci, 2:1--14.
-  Doerge, R. W. and Churchill, G. A. (1996).
Permutation Tests for Multiple Loci Affecting a Quantitative Character.
Genetics, 142:285--294.






-  Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011).
A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species.
PLOS ONE, 6(5):1--10.
-  Fierst, J. L. (2015).
Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools.
Frontiers in Genetics, 6(JUN):1--8.
-  Gazaffi, R., Margarido, G. R. a., Pastina, M. M., Mollinari, M., and Garcia, A. A. F. (2014).
A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny.
Tree Genetics & Genomes, 10(4):791--801.



-  Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart (2004).
An Introduction to Genetic Analysis.
W. H. Freeman, 8th edition.
-  Grover, A. and Sharma, P. C. (2014).
Development and use of molecular markers: past and present.
Critical reviews in biotechnology, 8551(00):1--13.
-  Kao, C. and Zeng, Z.-B. (1997).
General Formulas for Obtaining the MLEs and the Asymptotic Variance- Covariance Matrix in Mapping Quantitative Trait Loci When Using the EM Algorithm.
Biometrics, 53(2):653--665.
-  Kao, C., Zeng, Z.-B., and Teasdale, R. D. (1999).
Multiple Interval Mapping for Quantitative Trait Loci.
Genetics, 125(July):1203--1216.



-  Lander, E. S. and Botstein, D. (1989).
Mapping mendelian factors underlying quantitative traits using RFLP linkage maps.
Genetics, 121(January):185--199.
-  Maliepaard, C., Jansen, J., and Van Ooijen, J. W. (1997).
Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications.
Genetical Research, 70(3):237--250.
-  Margarido, G. R. A., Souza, A. P., and Garcia, A. A. F. (2007).
OneMap: software for genetic mapping in outcrossing species.
Hereditas, 144(3):78--9.



Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., and Baloch, F. S. (2017).

DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing.

Biotechnology & Biotechnological Equipment.







Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L. W. (2011).

Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms.

Molecular Ecology Resources, 11(SUPPL. 1):1--8.



-  Xia, Z., Zhang, S., Wen, M., Lu, C., Sun, Y., Zou, M., and Wang, W. (2018). Construction of an ultrahigh-density genetic linkage map for *Jatropha curcas* L. and identification of QTL for fruit yield. *Biotechnology for Biofuels*, 11(1):1--10.
-  Zeng, Z.-B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci USA*, 90(December):10972--10976.
-  Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics*, 1136(April):1457--1468.
-  Zeng, Z.-B., Kao, C., and Basten, C. J. (1999). Estimating the genetic architecture of quantitative traits. *Genetical Research*, 74:279--289.



Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008).
Status and Prospects of Association Mapping in Plants.
The Plant Genome, 1(1):5--20.