

IX WORKSHOP DE FÉRIAS EM
GENÉTICA E MELHORAMENTO
DE PLANTAS – 2015

**Biotecnologia e
Suas Aplicações
na Genética e
Melhoramento
de Plantas**



Mapeamento Genético

Cristiane Taniguti
Fernando Correr
Letícia Lara
Marianella Quezada



SUMÁRIO

1 ASPECTOS TEÓRICOS

2 POPULAÇÃO DE MAPEAMENTO

3 MARCADORES GENÉTICOS

4 CONSTRUÇÃO DE MAPAS GENÉTICOS

- ANÁLISE DE LIGAÇÃO
- GRUPOS DE LIGAÇÃO
- ORDENAÇÃO DOS MARCADORES
- FUNÇÕES DE MAPA

5 CONSIDERAÇÕES FINAIS

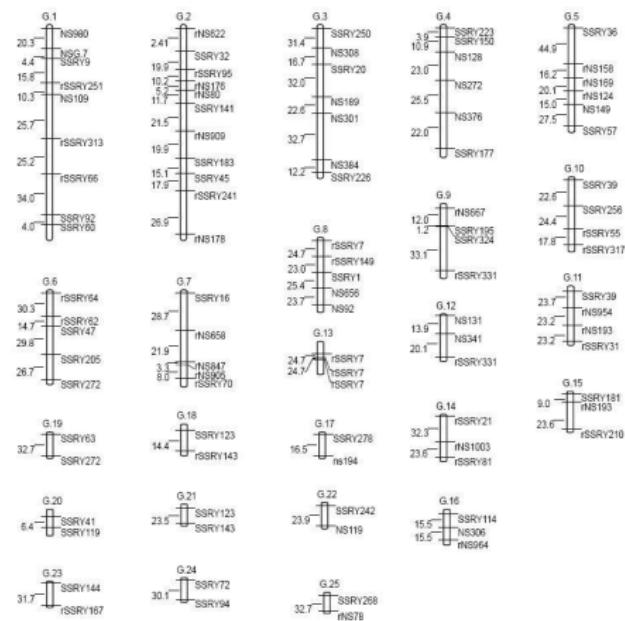
Aspectos teóricos

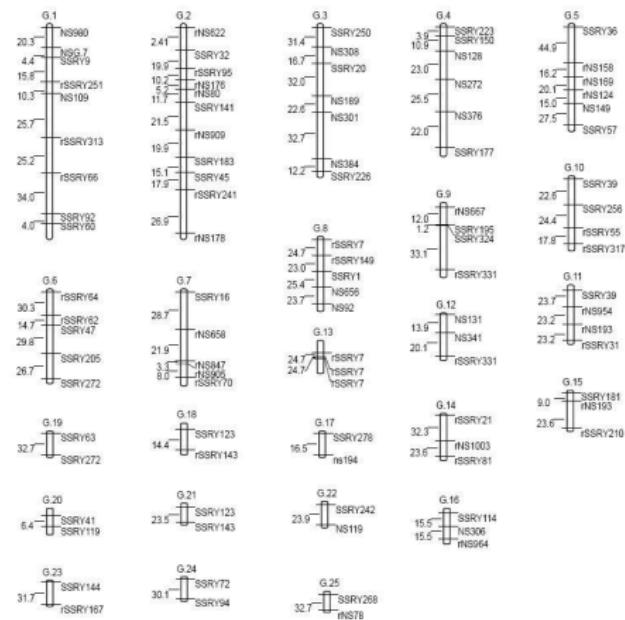
Mapas genéticos

Mapa genético ou mapa cromossômico ou mapa de ligação é uma representação gráfica das distâncias entre genes e de suas posições relativas em um cromossomo

O mapa genético estabelece as distâncias entre os genes. As mesmas estão diretamente relacionadas com a **frequência de recombinação** ou **fração de recombinação** entre eles



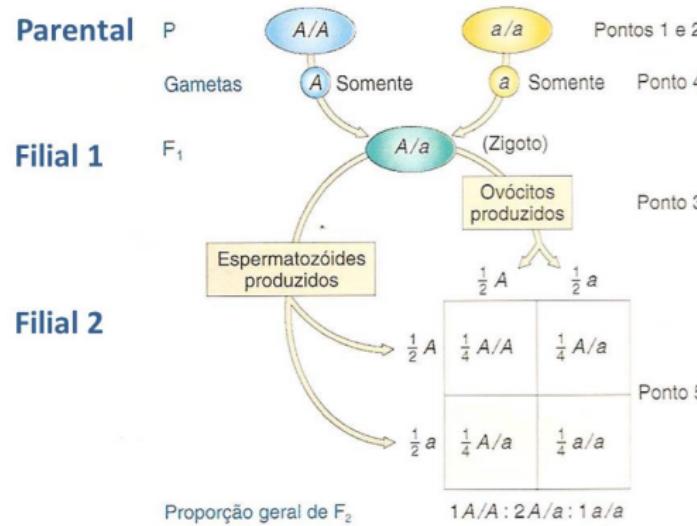




Número de grupos de ligação = Número de cromossomos da espécie

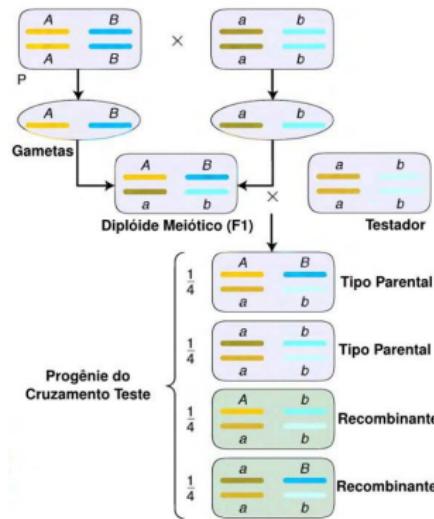
Primeira Lei de Mendel

Os dois membros de um par de genes segregam um do outro para os dois gametas; assim metade dos gametas leva um membro do par e a outra metade dos gametas leva o outro membro do par



Segunda Lei de Mendel

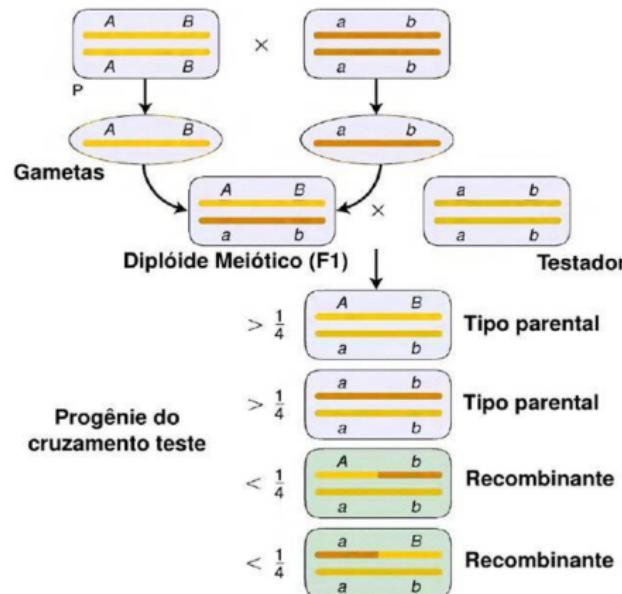
Pares de locos em cromossomos diferentes se distribuem independentemente na meiose



- Para genes não ligados a **fração de recombinação = 0,5 (50%)**

Genes ligados

Recombinação por *crossing over*



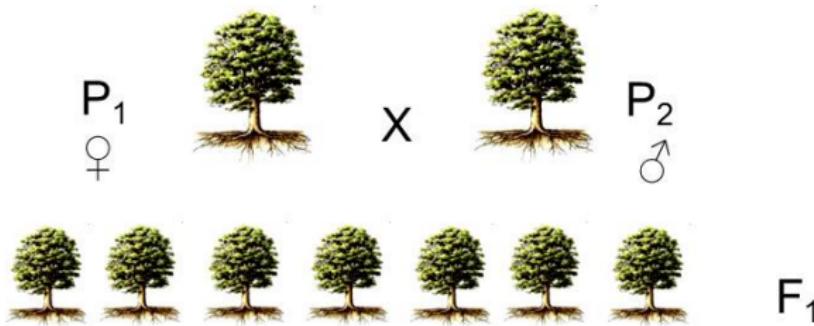
	Cromossomos meióticos	Produtos meióticos	
Meioses com um crossing entre os genes			Parental Recombinante Recombinante Parental

- **Fração de recombinação** significativamente menor que 50% mostra que os genes estão ligados
- A fração de recombinação varia dependendo dos genes que estão sendo estudados
- Assim estas variações representam as distâncias reais entre os genes nos cromossomos

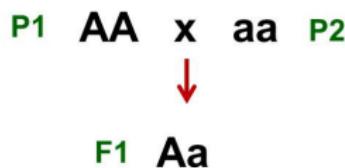
População de mapeamento

População de mapeamento

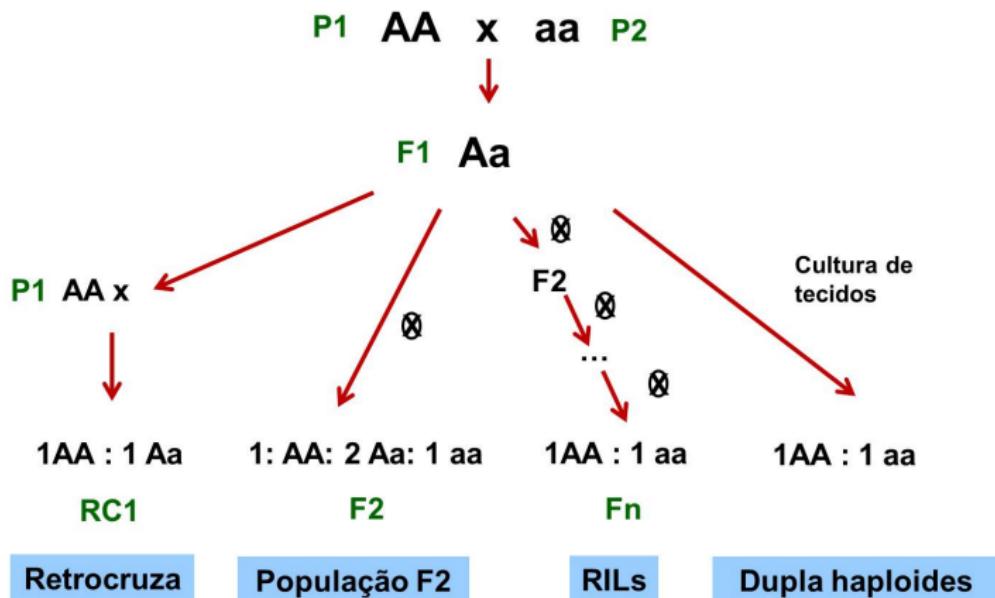
- **População segregante** (polimórfica para alguma variante de interesse)
- Genótipos parentais contrastantes em pelo menos uma característica
- Tamanho de 100 a 250 indivíduos



Genitores homozigotos (linhagens)



Genitores homozigotos (linhagens)



Genitores **não** endogâmicos

- Até 4 alelos segregando para cada loco

P1 A1A2 x A3A4 P2



F1 1 A1A3 : 1 A1A4 : 1 A2A3 : 1 A2A4

- Podem ocorrer diferentes variantes alélicas (Ex: $A_1 = A_3$)
- Posso esperar padrões de segregação: 3:1, 1:2:1 e 2:1
- A mistura de segregação dificulta as estimativas da fração de recombinação

Marcadores genéticos

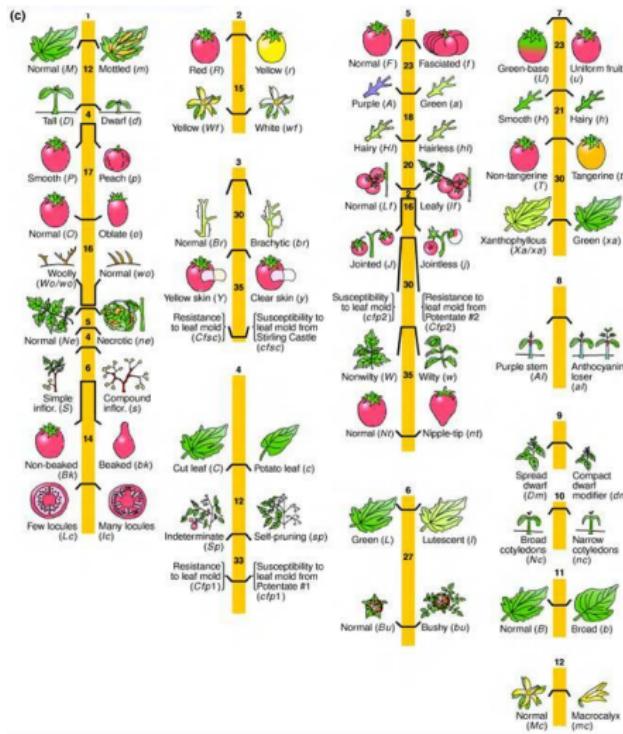
Marcadores genéticos

- Característica que é capaz de detectar diferenças entre dois ou mais indivíduos ou organismos
- Nos estudos de mapeamentos os marcadores servem para identificar um local ou uma região de um cromossomo
- Como principais características deve ter:
 - Alto nível de polimorfismo
 - Estabilidade em diferentes ambientes
 - Detectar grande número de locos não ligados
 - Apresentar herança mendeliana

Marcadores Morfológicos

- Fenótipo de fácil identificação
 - Ex: alelo de coloração em milho ligado ao alelo responsável da macho esterilidade
- Limitação: número reduzido de marcadores (baixa cobertura do genoma)

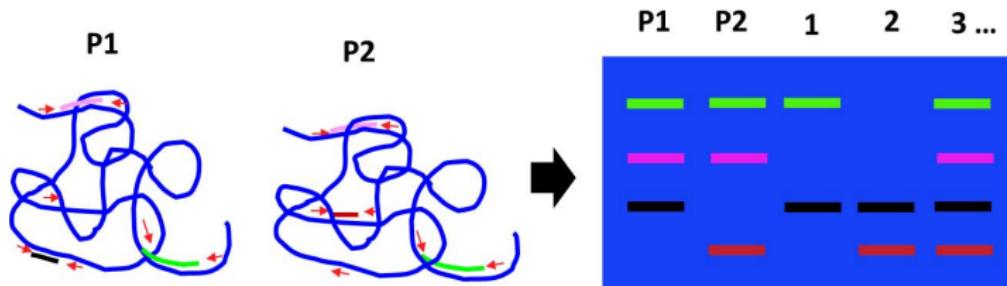
Mapa genético com marcadores morfológicos



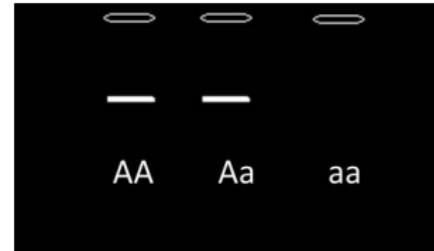
Marcadores Moleculares

- Baseados em polimorfismos na própria sequência de DNA
- Vantagem: grande variabilidade entre indivíduos de mesma espécie
- Podemos obter uma alta cobertura do genoma
- Permitem marcar grande parte dos genes de interesse para uma espécie
- No mapeamento genético as principais vantagens são:
 - Codominante
 - Multialelico
 - Hipervariável (polimórfico em varias espécies)

Marcadores RAPD (*Random Amplified Polymorphic DNA*)

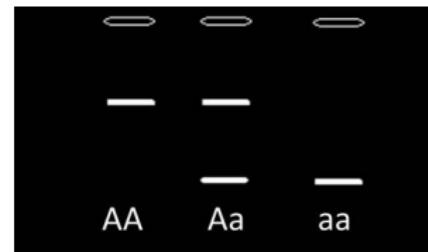


■ Marcadores Dominantes

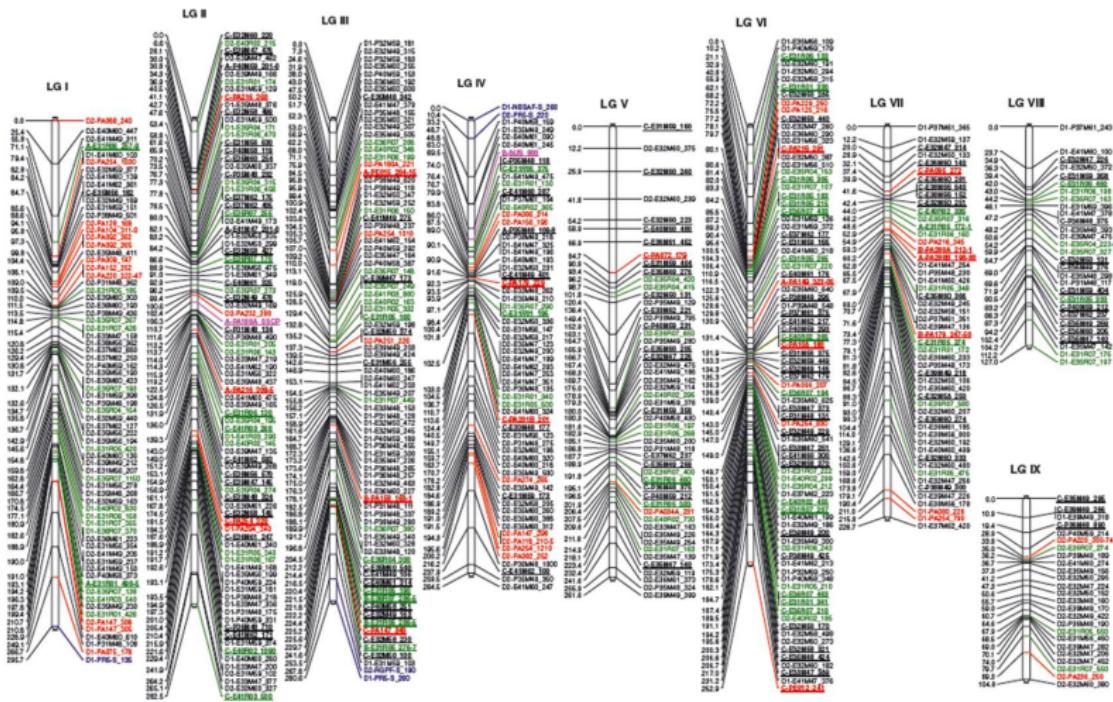


Marcadores SSR (*Single Sequence Repeat*)

■ Marcadores Codominantes



Mapa genético com marcadores moleculares



Marcadores SNP (*Single Nucleotide Polymorphism*)

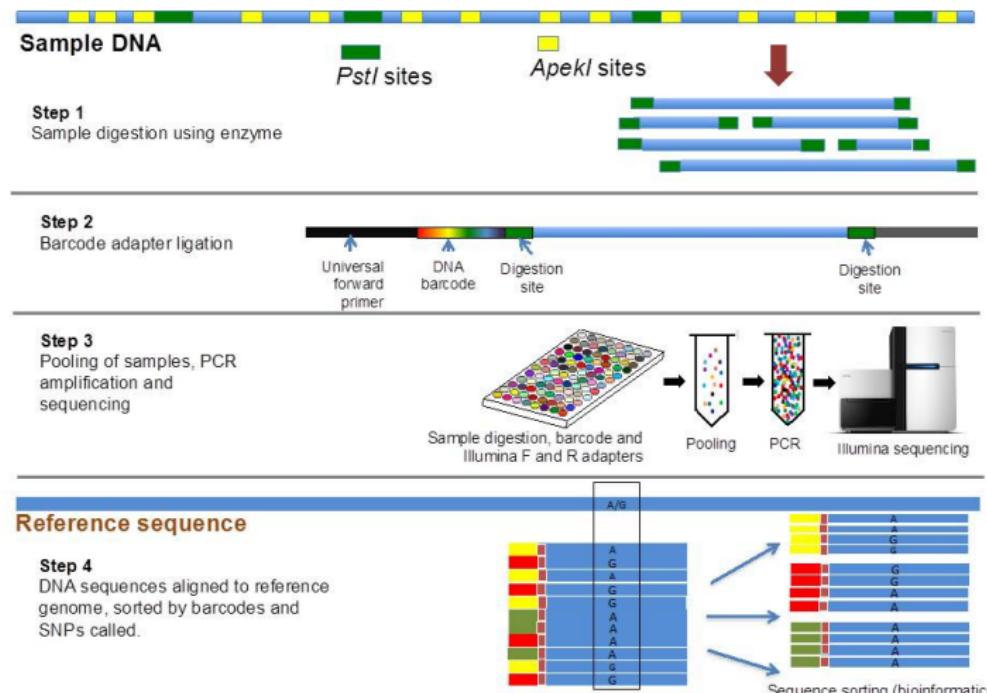
- Pequena mudança ou variação na sequência de DNA
- Marcadores de preferência pela sua grande abundância e pelo desenvolvimento de tecnologias de genotipagem em larga escala

SNP

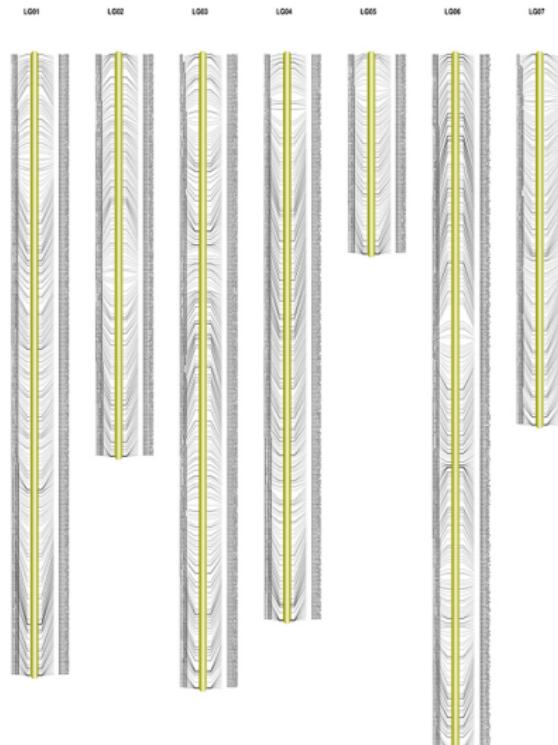
↓

Tree 1	A C G T G T C G G T C T T A
	A C G T G T C A G T C T T A
Tree 2	A C G T G T C G G T C T T A
	A C G T G T C G G T C T T A
Tree 3	A C G T G T C A G T C T T A
	A C G T G T C A G T C T T A

Marcadores SNP em grande escala (GBS, RADseq, DArT,...)



Mapa genético com marcadores SNPs



Codificação dos Marcadores Moleculares (Wu *et al.*, 2002)

		Parent			Offspring			
	Cross type	Cross	Observed band	Remark	Observed bands	Segregation	No. phenotypes	
A	1	$ab \times cd$	$ab \times cd$	Asymmetry	ac, ad, bc, bd	1:1:1:1	4	
	2	$ab \times ac$	$ab \times ac$	Asymmetry	a, ac, ba, bc	1:1:1:1	4	
	3	$ab \times co$	$ab \times c$	Asymmetry	ac, a, bc, b	1:1:1:1	4	
	4	$ao \times bo$	$a \times b$	Asymmetry	ab, a, b, o	1:1:1:1	4	
B	B ₁	5	$ab \times ao$	$ab \times a$	Asymmetry	$ab, 2a, b$	1:2:1	3
	B ₂	6	$ao \times ab$	$a \times ab$	Asymmetry	$ab, 2a, b$	1:2:1	3
	B ₃	7	$ab \times ab$	$ab \times ab$	Symmetry	$a, 2ab, b$	1:2:1	3
C		8	$ao \times ao$	$a \times a$	Symmetry	$3a, o$	3:1	2
D	D ₁	9	$ab \times cc$	$ab \times c$	Asymmetry	ac, bc	1:1	2
	D ₁	10	$ab \times aa$	$ab \times a$	Asymmetry	a, ab	1:1	2
	D ₁	11	$ab \times oo$	$ab \times o$	Asymmetry	a, b	1:1	2
	D ₁	12	$bo \times aa$	$b \times a$	Asymmetry	ab, a	1:1	2
	D ₁	13	$ao \times oo$	$a \times o$	Asymmetry	a, o	1:1	2
D ₂	D ₂	14	$cc \times ab$	$c \times ab$	Asymmetry	ac, bc	1:1	2
	D ₂	15	$aa \times ab$	$a \times ab$	Asymmetry	a, ab	1:1	2
	D ₂	16	$oo \times ab$	$o \times ab$	Asymmetry	a, b	1:1	2
	D ₂	17	$aa \times bo$	$a \times b$	Asymmetry	ab, a	1:1	2
	D ₂	18	$oo \times ao$	$o \times a$	Asymmetry	a, o	1:1	2

Construção de Mapas Genéticos

- 1** Analise de ligação
- 2** Formação de Grupos de Ligação
- 3** Ordenação
- 4** Funções de mapa

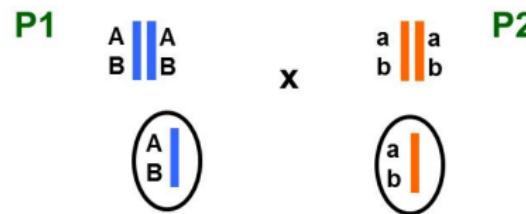
Analise de ligação

- Estimação de **Fração de recombinação** (\hat{r})
- Estimativa realizada para cada par de marcadores (Teste de dois pontos)
- Diferentes populações de mapeamento

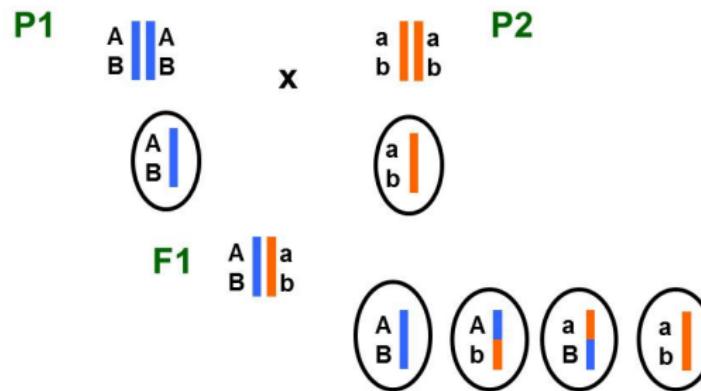
Fração de recombinação

\hat{r} = Probabilidade de ocorrer um evento de recombinação entre os locos

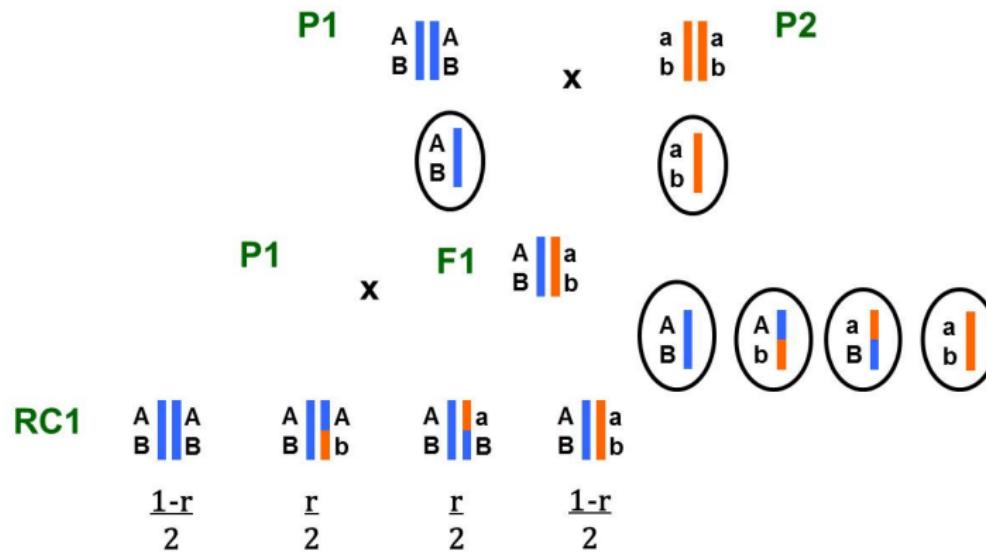
População de retrocruzamento



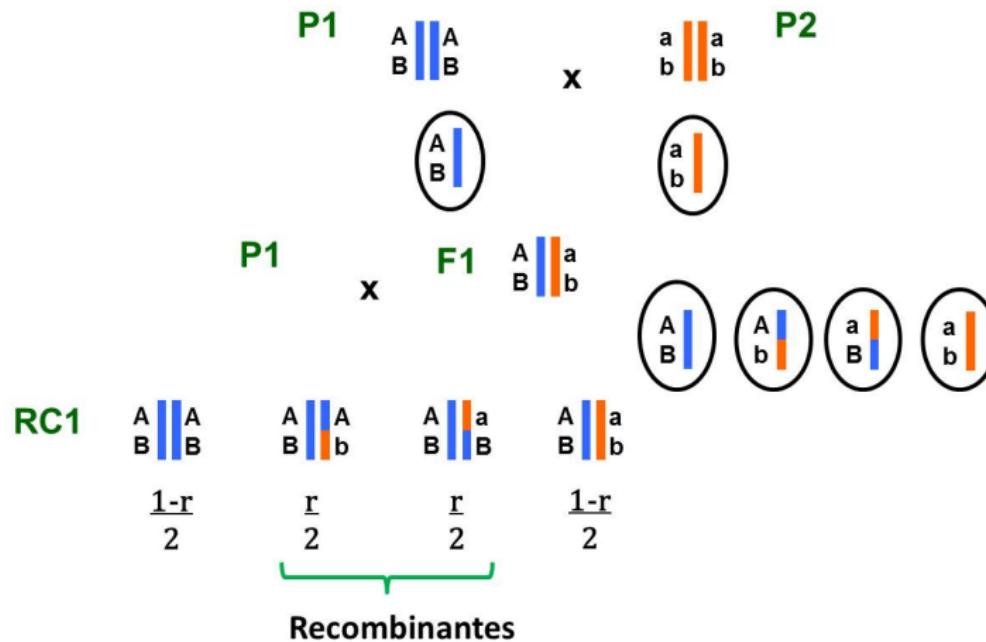
População de retrocruzamento



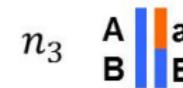
População de retrocruzamento



População de retrocruzamento



- Dados de contagem das 4 classes: n_1 , n_2 , n_3 e n_4



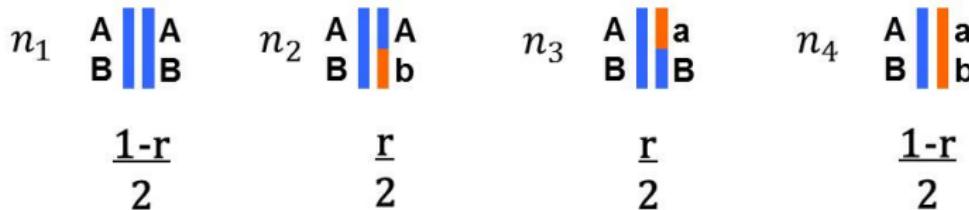
$$\frac{1-r}{2}$$

$$\frac{r}{2}$$

$$\frac{r}{2}$$

$$\frac{1-r}{2}$$

- Dados de contagem das 4 classes: n_1, n_2, n_3 e n_4



Função de Verossimilhança na população RC_1

$$L(r) \propto \left(\frac{1-r}{2}\right)^{n_1} \cdot \left(\frac{r}{2}\right)^{n_2} \cdot \left(\frac{r}{2}\right)^{n_3} \cdot \left(\frac{1-r}{2}\right)^{n_4}$$

$$l(r) = (n_1 + n_4) \log\left(\frac{1-r}{2}\right) + (n_2 + n_3) \log\left(\frac{r}{2}\right)$$

Estimador de Máxima Verossimilhança para \hat{r}

$$\hat{r} = \frac{n_2 + n_3}{n_2 + n_3 + n_1 + n_4} = \frac{\text{recombinantes}}{\text{recombinantes} + \text{não recombinantes}}$$

Fração de recombinação

\hat{r} = Probabilidade de ocorrer um evento de recombinação entre os locos

Exemplo

Conjunto de dados:

Ind	M1	M2	Ind	M1	M2	
1	AA	BB	1	1	1	$n_1 = 58$
2	AA	Bb	2	1	0	$n_2 = 3$
3	Aa	BB	3	0	1	$n_3 = 4$
4	Aa	Bb	4	0	0	$n_4 = 38$
...	

Exemplo

Conjunto de dados:

Ind	M1	M2	Ind	M1	M2	
1	AA	BB	1	1	1	$n_1 = 58$
2	AA	Bb	2	1	0	$n_2 = 3$
3	Aa	BB	3	0	1	$n_3 = 4$
4	Aa	Bb	4	0	0	$n_4 = 38$
...	

Fração de recombinação

$$\hat{r} = \frac{3 + 4}{3 + 4 + 58 + 38} = \frac{7}{103} = 0,0679 \ (6,79\%)$$

- Os marcadores M1 e M2 estão ligados

LOD score

- Teste de razão de verossimilhança
- Comparar valores de verossimilhança considerando diferentes valores dos parâmetros (\hat{r} = fração de recombinação)
- Sabemos que para marcadores não ligados: $r = 0,5$

$$LOD = \log_{10} \frac{L(\hat{r})}{L(r = 0,5)}$$

Valores significantes de $LOD > 3$

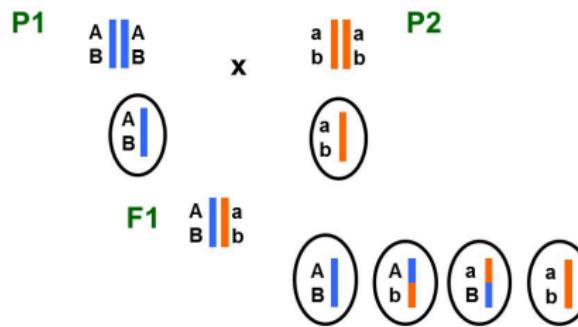
LOD score

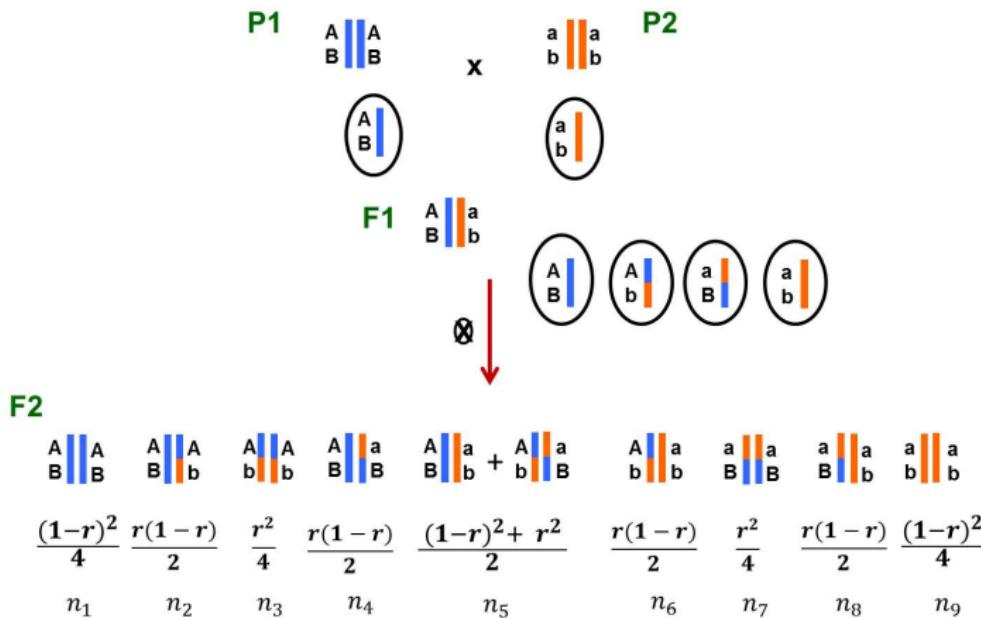
- Teste de razão de verossimilhança
- Comparar valores de verossimilhança considerando diferentes valores dos parâmetros (\hat{r} = fração de recombinação)
- Sabemos que para marcadores não ligados: $r = 0,5$

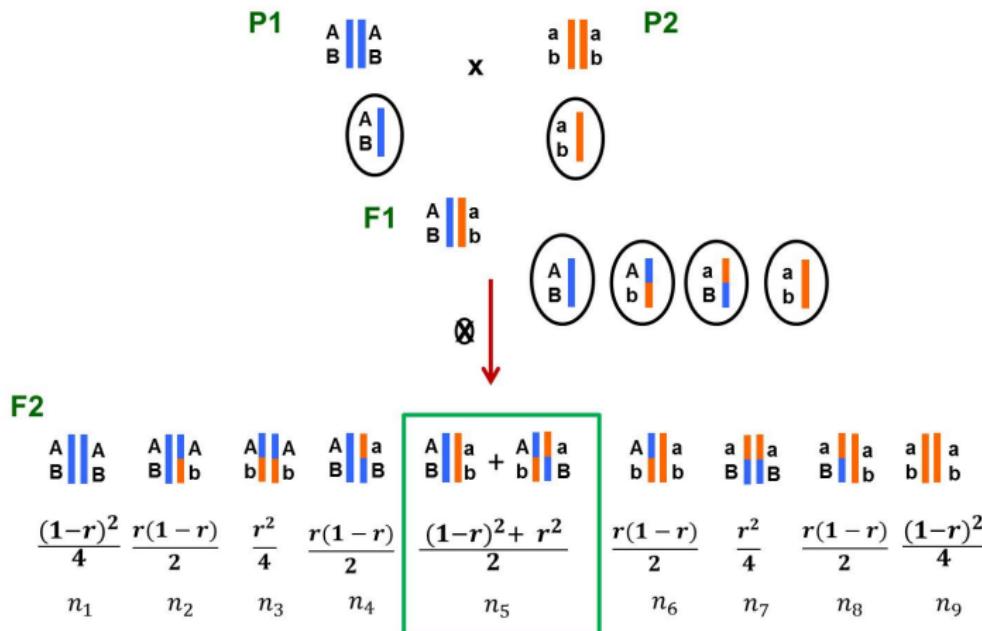
$$LOD = \log_{10} \frac{L(\hat{r})}{L(r = 0,5)}$$

Valores significantes de $LOD > 3$

Para os marcadores M1-M2: $LOD = 19,89$

População F_2 

População F_2 

População F_2 

Função de Verossimilhança na população F_2

$$L(r) \propto \left[\frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[\frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[\frac{r^2}{4} \right]^{n_3+n_7}$$
$$\times \left[\frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

$$l(r) = (n_1 + n_9) \log \left[\frac{(1-r)^2}{4} \right] + (n_2 + n_4 + n_6 + n_8) \log \left[\frac{r(1-r)}{2} \right]$$
$$+ (n_3 + n_7) \log \left[\frac{r^2}{4} \right] + (n_5) \log \left[\frac{(1-r)^2}{2} + \frac{r^2}{2} \right]$$

Função de Verossimilhança na população F_2

$$L(r) \propto \left[\frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[\frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[\frac{r^2}{4} \right]^{n_3+n_7}$$

$$\times \left[\frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

$$l(r) = (n_1 + n_9) \log \left[\frac{(1-r)^2}{4} \right] + (n_2 + n_4 + n_6 + n_8) \log \left[\frac{r(1-r)}{2} \right]$$

$$+ (n_3 + n_7) \log \left[\frac{r^2}{4} \right] + (n_5) \log \left[\frac{(1-r)^2}{2} + \frac{r^2}{2} \right]$$

- Estimador de Máxima Verossimilhança difícil de resolver analiticamente

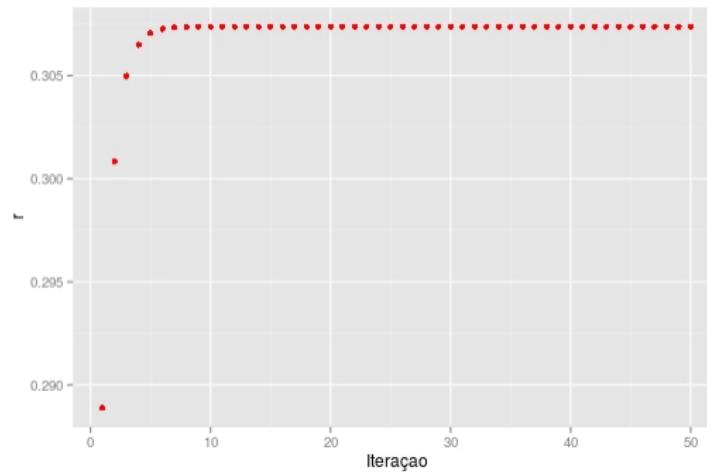
Algoritmo Expectation-Maximization (Dempster *et al.*, 1977)

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Empregado quando temos uma variável latente (não observável)
 - Se busca um modelo apropriado para a distribuição da variável latente
 - Calcula-se sua esperança (empregando um valor inicial para os parâmetros)
 - São estimados novamente os parâmetros pelo método de máxima verossimilhança
 - O processo é repetido até convergência
- Muito usado em análises de mapeamento genético

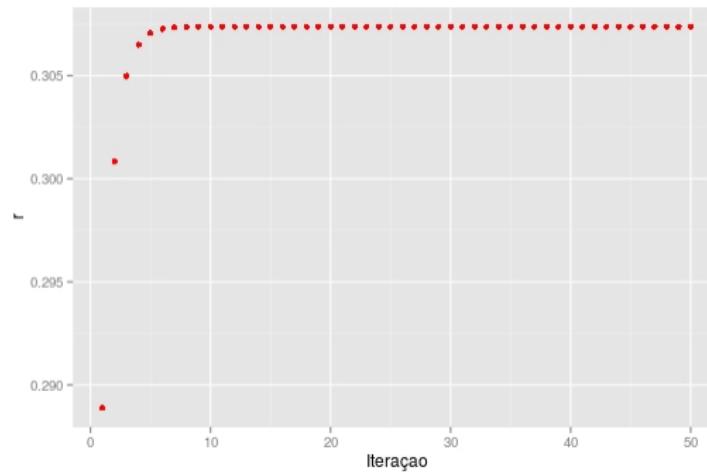
Exemplo

Conjunto de dados:

Classe	M1	M2	Classe	M1	M2	
n_1	AA	BB	n_1	2	2	$n_1 = 23$
n_2	AA	Bb	n_2	2	1	$n_2 = 18$
n_3	AA	bb	n_3	2	0	$n_3 = 2$
n_4	Aa	BB	n_4	1	2	$n_4 = 21$
n_5	Aa	Bb	n_5	1	1	$n_5 = 49$
n_5	Aa	Bb	n_5	1	1	$n_5 = \dots$
n_6	Aa	bb	n_6	1	0	$n_6 = 16$
n_7	aa	BB	n_7	0	2	$n_7 = 4$
n_8	aa	Bb	n_8	0	1	$n_8 = 22$
n_9	aa	bb	n_9	0	0	$n_9 = 16$



- Fração de recombração entre M1-M2: $\hat{r} = 0,307$ (30,7%)



- Fração de recombração entre M1-M2: $\hat{r} = 0,307$ (30,7%)
- Os marcadores M1 e M2 estão ligados

Grupos de ligação

Grupos de ligação

- Biologicamente: grupos de genes/locos no mesmo cromossomo
- Estatisticamente: grupos de locos que segregam conjuntamente

Para dois locos (dois marcadores)

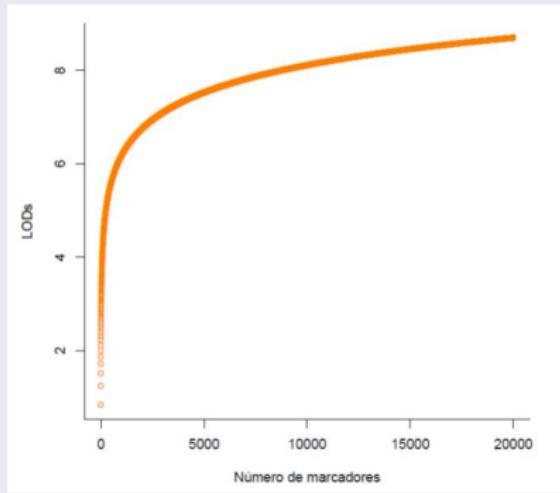
- Valor de fração de recombinação entre marcadores (r)
- Hipóteses H_0 : marcadores **não** estão no mesmo grupo de ligação ($r = 0,5$)
- Uso de valor de LOD score

Critérios para formação de grupos de ligação

- **Máxima fração de recombinação** para declarar ligação
- **Mínimo valor de LOD** para declarar ligação
- Estos são definidos pelo pesquisador, a partir dos dados (espécie, delineamento, marcadores moleculares, etc)
- Em geral são empregados:
 - Valores de fração de recombinação $r = 0,5$
 - Valores de LOD score = 3

Regra prática valor de LOD score

- Uso de regra prática para estabelecer o valor de LOD score
- Considera o número de marcadores (número de múltiplos testes)
- Emprega a correção de Bonferroni



Exemplo

Conjunto de dados:

r	M1	M2	M3	M4
M1	-	0,12	0,17	0,50
M2		-	0,27	0,50
M3			-	0,50
M4				-

LOD	M1	M2	M3	M4
M1	-	13,80	8,34	0,01
M2		-	4,65	0,10
M3			-	0,87
M4				-

Exemplo

Conjunto de dados:

r	M1	M2	M3	M4
M1	-	0,12	0,17	0,50
M2		-	0,27	0,50
M3			-	0,50
M4				-

LOD	M1	M2	M3	M4
M1	-	13,80	8,34	0,01
M2		-	4,65	0,10
M3			-	0,87
M4				-

Os marcadores M1-M2-M3 formam um grupo de ligação

Exemplo

Conjunto de dados:

r	M1	M2	M3	M4
M1	-	0,12	0,17	0,50
M2		-	0,27	0,50
M3			-	0,50
M4				-

LOD	M1	M2	M3	M4
M1	-	13,80	8,34	0,01
M2		-	4,65	0,10
M3			-	0,87
M4				-

Os marcadores M1-M2-M3 formam um grupo de ligação

O marcador M4 **não** forma parte do grupo de ligação

Ordenação dos marcadores

Ordenação dos marcadores



Qual a ordem dos locos?

$$\hat{r} (M1-M2) = 0,12$$

$$\hat{r} (M1-M3) = 0,17$$

$$\hat{r} (M2-M3) = 0,27$$

Ordenação dos marcadores



Qual a ordem dos locos?

$$\hat{r} (M1-M2) = 0,12$$

$$\hat{r} (M1-M3) = 0,17$$

$$\hat{r} (M2-M3) = 0,27$$

Ordem correta: M2-M1-M3



Qual a ordem dos locos?

$$\hat{r} (M1-M2) = 0,32$$

$$\hat{r} (M1-M3) = 0,28$$

$$\hat{r} (M1-M4) = 0,07$$

$$\hat{r} (M1-M5) = 0,02$$

$$\hat{r} (M2-M3) = 0,60$$

$$\hat{r} (M2-M4) = 0,25$$

$$\hat{r} (M2-M5) = 0,30$$

$$\hat{r} (M3-M4) = 0,35$$

$$\hat{r} (M3-M5) = 0,30$$

$$\hat{r} (M4-M5) = 0,05$$

Qual a ordem dos locos?

$$\hat{r} (M1-M2) = 0,32$$

$$\hat{r} (M1-M3) = 0,28$$

$$\hat{r} (M1-M4) = 0,07$$

$$\hat{r} (M1-M5) = 0,02$$

$$\hat{r} (M2-M3) = 0,60$$

$$\hat{r} (M2-M4) = 0,25$$

$$\hat{r} (M2-M5) = 0,30$$

$$\hat{r} (M3-M4) = 0,35$$

$$\hat{r} (M3-M5) = 0,30$$

$$\hat{r} (M4-M5) = 0,05$$

- Difícil de estabelecer a ordem correta quando temos muitos (milhares) de marcadores

Comparação de todas as ordens possíveis

- Com o aumento do número de marcadores torna-se inviável comparar todas as ordens
- O número de ordens aumenta exponencialmente

Locos	Ordens	Tempo
5	60 ($5!/2$)	0 seg
10	1.814.400 ($10!/2$)	0,0162 seg
15	653.837.184.000	2,54 horas
20	$1,2165 \times 10^{18}$	732,4 anos
25	$7,755 \times 10^{24}$	5.898.373.012,27 anos

Algoritmos para ordenar marcadores

- Comando TRY (Lander, 1987)
 - Ordena exaustivamente um sub-grupo de marcadores e depois testa a posição dos marcadores remanescentes, um a um
- Comando ORDER
 - Realiza o procedimento anterior de forma automática
- Seriation (Buetow & Chakravarti, 1987)
- Rapid Chain Delination (Doerge, 1996)
- Record (Van Os *et al.*, 2005)
- ...

Algoritmos para ordenar marcadores

- Comando TRY (Lander, 1987)
 - Ordena exaustivamente um sub-grupo de marcadores e depois testa a posição dos marcadores remanescentes, um a um
- Comando ORDER
 - Realiza o procedimento anterior de forma automática
- Seriation (Buetow & Chakravarti, 1987)
- Rapid Chain Delination (Doerge, 1996)
- Record (Van Os *et al.*, 2005)
- ...

Ripple

- Verifica possíveis inversões entre marcadores próximos

Rapid Chain Delination

A partir da matriz com as frações de recombinação:

- 1 Inicie o grupo de ligação com o par de marcadores com o menor \hat{r}
- 2 O grupo de ligação é estendido com a adição de marcadores remanescentes nos terminais do grupo de ligação. Os marcadores com os menores \hat{r} são adicionados um a um
- 3 Repetir o passo 2 até que nenhum marcador possa ser adicionado à cadeia
- 4 Encerrar o processo quando não restarem marcadores remanescentes ou quando os restantes não estiverem significativamente ligados a nenhum outro marcador

Exemplo

\hat{r}	M1	M2	M3	M4
M1	—	0,09	0,19	0,17
M2		—	0,26	0,22
M3			—	0,32
M4				—

Algoritmo RCD

Ordem dos marcadores

Exemplo

\hat{r}	M1	M2	M3	M4
M1	—	0, 09	0, 19	0, 17
M2		—	0, 26	0, 22
M3			—	0, 32
M4				—

Exemplo

\hat{r}	M1	M2	M3	M4
M1	—	0,09	0,19	0,17
M2		—	0,26	0,22
M3			—	0,32
M4				—

Algoritmo RCD

Ordem dos marcadores

M1-M2

Exemplo

\hat{r}	M1	M2	M3	M4
M1	—	0,09	0,19	0,17
M2		—	0,26	0,22
M3			—	0,32
M4				—

Exemplo

\hat{r}	M1	M2	M3	M4
M1	—	0,09	0,19	0,17
M2		—	0,26	0,22
M3			—	0,32
M4				—

Algoritmo RCD

Ordem dos marcadores

M4-M1-M2

Exemplo

\hat{r}	M1	M2	M3	M4
M1	—	0,09	0,19	0,17
M2		—	0,26	0,22
M3			—	0,32
M4				—

Exemplo

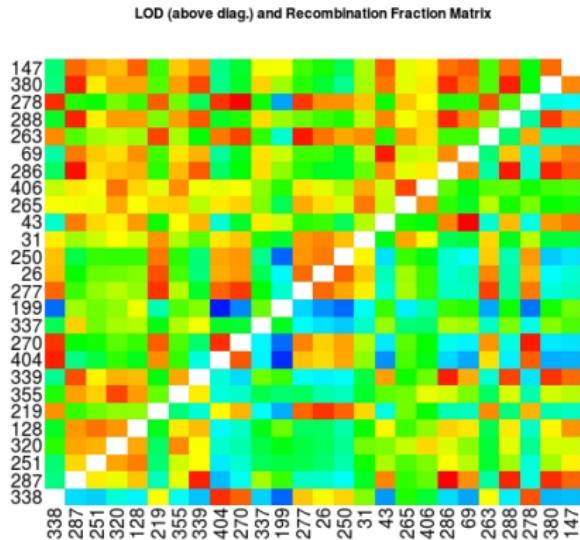
\hat{r}	M1	M2	M3	M4
M1	—	0,09	0,19	0,17
M2		—	0,26	0,22
M3			—	0,32
M4				—

Algoritmo RCD

Ordem dos marcadores

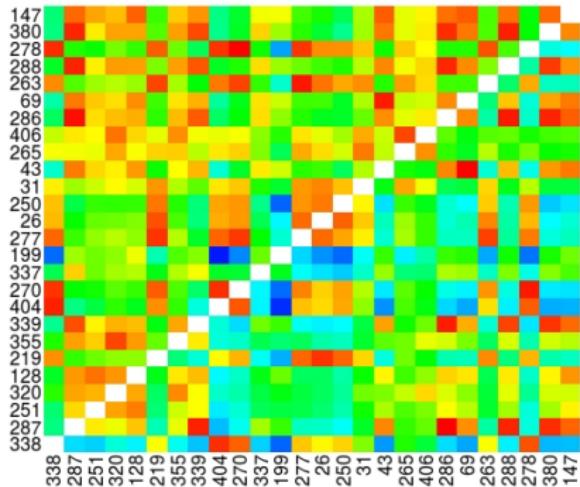
M4-M1-M2-M3

Uso de *Heatmap* para avaliar a ordem dos marcadores

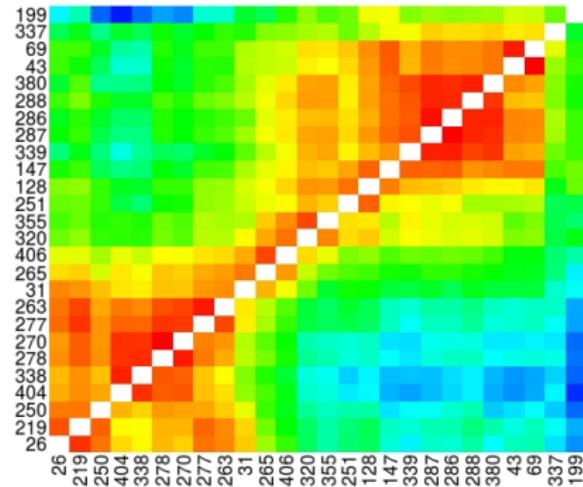


Uso de *Heatmap* para avaliar a ordem dos marcadores

LOD (above diag.) and Recombination Fraction Matrix



LOD (above diag.) and Recombination Fraction Matrix



Funções de mapa

Funções de mapa

- Frações de recombinação não podem ser somadas
- Uso de Funções de mapeamento que são aditivas
- Unidade de mapa (distancia): Morgan ou cM

Morgan

- Fração de recombinação(r) = unidade de mapa $u.m$

Morgan

- Fração de recombinação(r) = unidade de mapa $u.m$

Haldane

$$d_{ij} = -\frac{1}{2} \log(1 - 2r_{ij})$$

Morgan

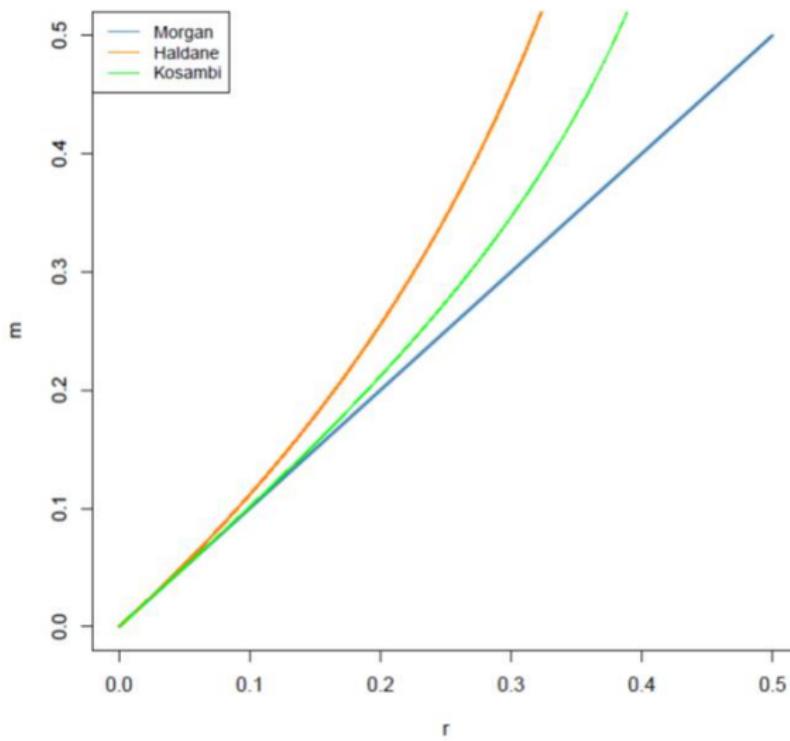
- Fração de recombinação(r) = unidade de mapa $u.m$

Haldane

$$d_{ij} = -\frac{1}{2} \log(1 - 2r_{ij})$$

Kosambi

$$d_{ij} = \frac{1}{4} \log \frac{1 + 2r_{ij}}{1 - 2r_{ij}}$$



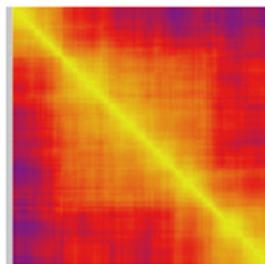
Considerações finais

Considerações finais

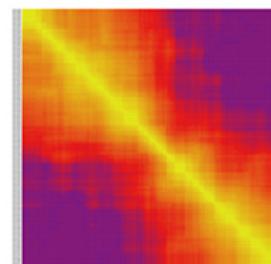
- **Mapa genético:** diagrama no qual são representados os genes com suas respectivas posições no cromossomo
- Baseados na **fração de recombinação** entre os genes (crossing-over)
- Permite estabelecer a organização física do genoma de uma espécie
- Contém informação sobre regiões específicas do genoma
- **Centimorgan (cM):** unidade de medida utilizada para descrever a distância entre os genes

Mapas genéticos de *nova geração*

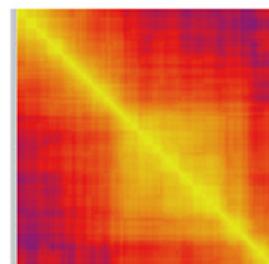
Chr1



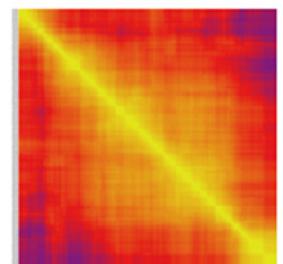
Chr2



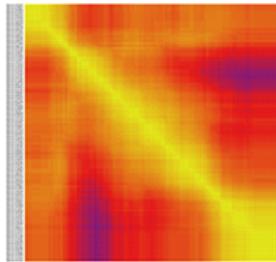
Chr3



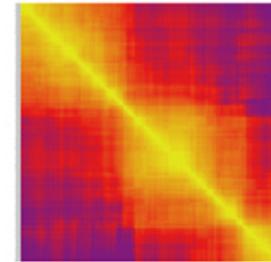
Chr4



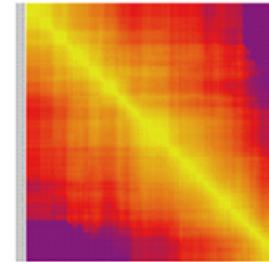
Chr5



Chr6



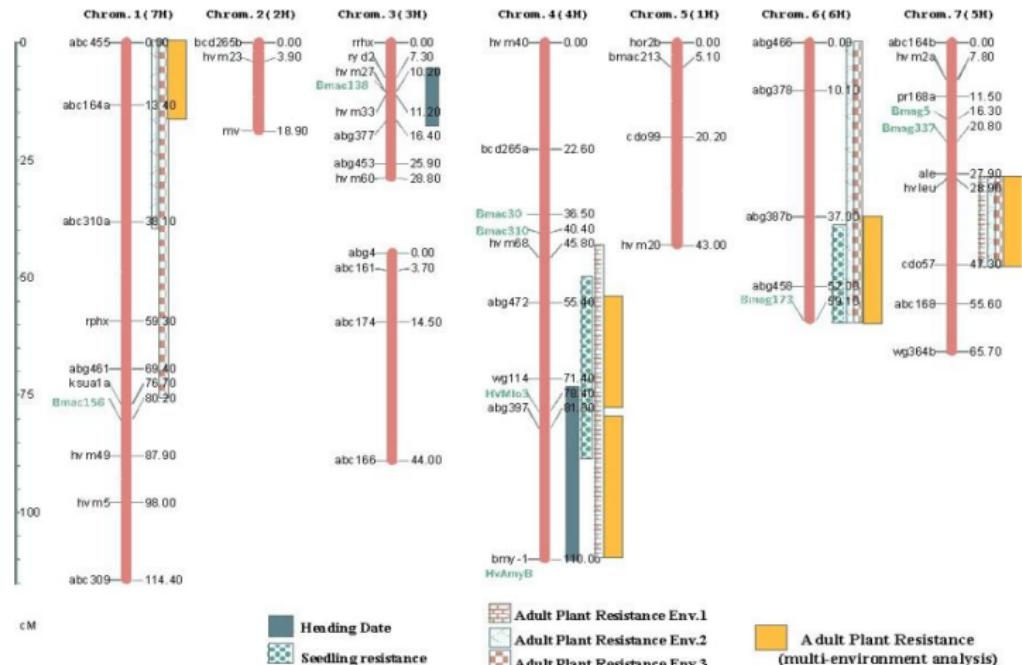
Chr7



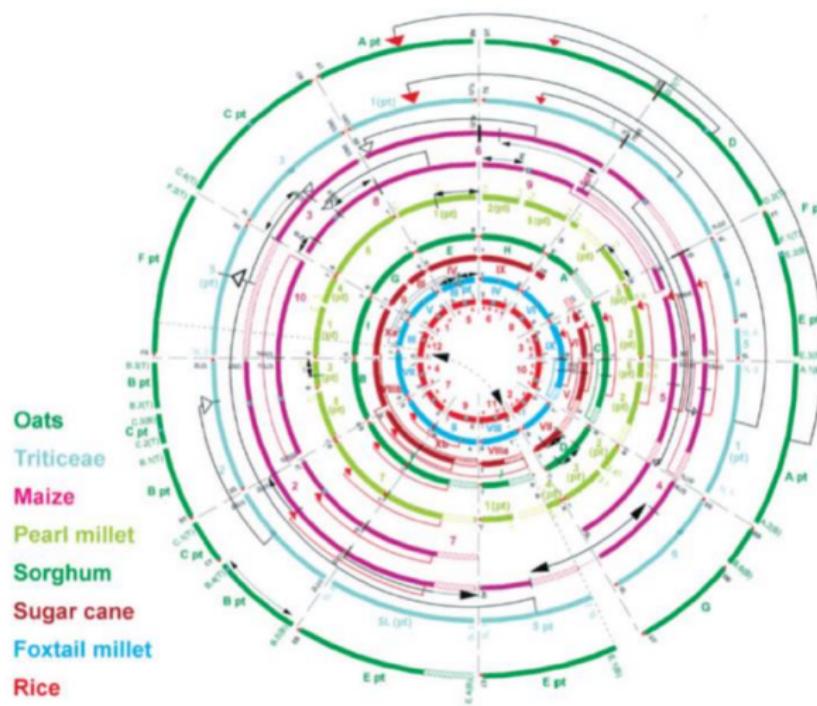
Aplicações dos mapas genéticos

- Identificação de QTL (*Quantitative trait loci*)- Regiões do genoma associadas a caracteres fenotípicos
- Estudo de funções de novos genes
- Analises de mapeamento comparativo e sintenia entre especies relacionadas
- Seleção assistida por marcadores em programas de melhoramento
- Plataforma para a montagem do genoma

Identificação de QTLs



Mapeamento comparativo



Obrigado