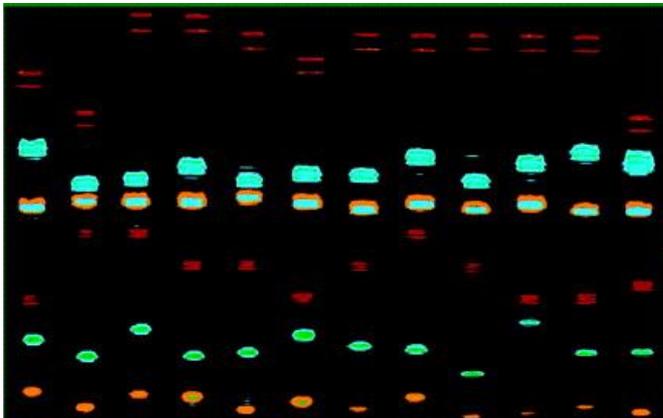


MARCADORES MOLECULARES: DO MELHORAMENTO A CONSERVAÇÃO

Aula 12

LGN232 – Genética Molecular



Maria Carolina Quecine
Departamento de Genética
mquecine@usp.br

SUMÁRIO

- Revisão as ferramentas moleculares;
- Conceito de marcador molecular;
- Marcadores morfológico;
- Marcadores moleculares;
- Microssatélites;
- SNPs;
- Aplicações dos marcadores moleculares;
- Estudo dirigido.

Aplicações de Marcadores no Melhoramento

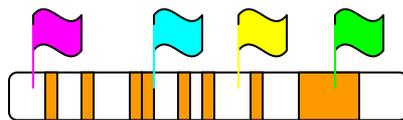
- **Mapeamento genômico**
 - Seleção Assistida por Marcadores - MAS
 - Seleção Genômica - GWAS
 - Clonagem/Identificação de genes
- **Diversidade genética e filogenia**
 - Conservação e caracterização de germoplasma
 - Identificação de acessos
 - Seleção de genitores para cruzamentos e populações
 - Filogenia e Evolução

Genética e Melhoramento

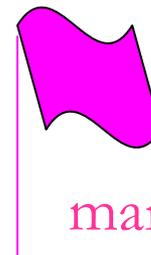
- Seleção inconsciente e “arte”
 - da invenção da Agricultura até século XIX
- 1900s - Descoberta dos princípios genéticos
- 1920-50 - Melhoramento genético científico
 - genética quantitativa e biometria
 - (fenótipo é previsor ruim do valor genético!)*
- 1970-80s em diante - Utilização de marcadores genéticos moleculares

Genética e Melhoramento

- Sucesso no melhoramento depende da capacidade de **distinguir fatores genéticos herdáveis dos ambientais**
- Marcadores genéticos são unidades herdáveis simples
- Marcadores genéticos quando associados a características de interesse aumentam a eficiência de seleção



cromossomo



marcadores

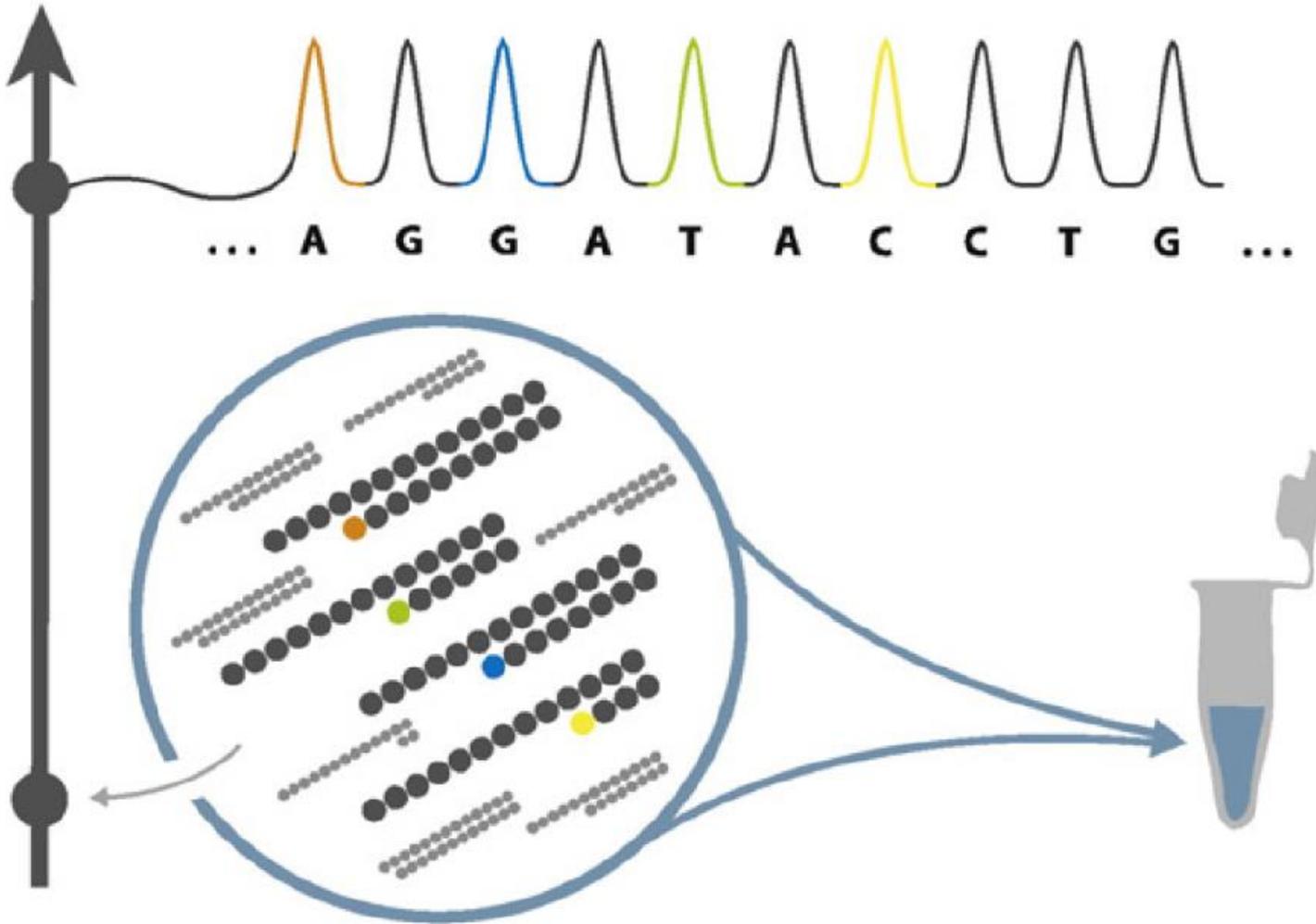
Mas o que é um Marcador Genético?

O ideal seria sequenciar todos os genótipos, mas.....

Se usa Marcador Genético ou Molecular como proxy!

- Variação ou polimorfismo que, numa população segregante, se comportam de acordo com as leis Mendelianas!
- Baseiam-se na existência de variabilidade genética e na possibilidade de sua detecção
- **Tipos:**
 - Morfológicos
 - Bioquímicos (isoenzimas)
 - Moleculares – DNA (microssatélites, SNPs)

SEQUENCIAMENTO DE DNA



Histórico de Marcadores

1. Karl Sax (1923): propôs método para localização de QTLs

ligação entre genes de característica qualitativa (cor de semente) e quantitativa (peso de semente)

Problema: ausência de mutações múltiplas em estoque de elite, baixa viabilidade

2. Hunter & Marker (1957): marcas bioquímicas - proteínas

desenvolveram isoenzimas em gel de amido

3. Hubby & Lewontin (1966)

demonstraram que 30% de loci de isoenzimas exibiam polimorfismo em populações selvagens de *Drosophila*

MARCADORES MORFOLÓGICOS

- Monomorfismo



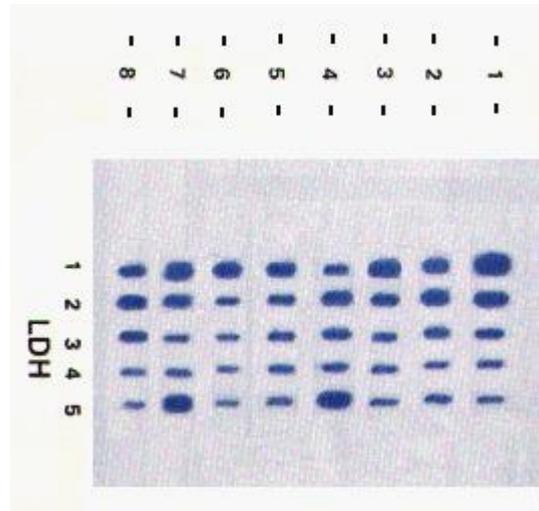
- Polimorfismo



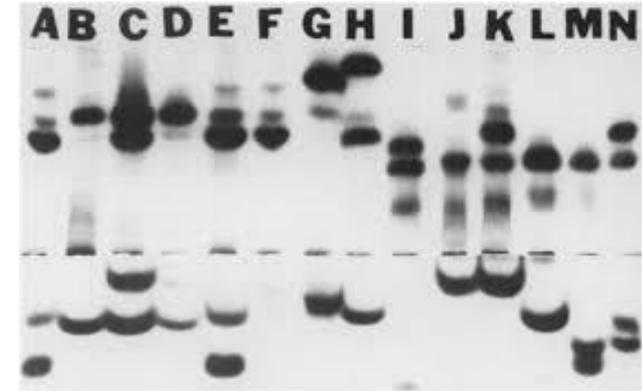
MARCADORES BIOQUÍMICOS

- Variabilidade observada como resultado da tradução de RNA em proteínas

- Monomorfismo

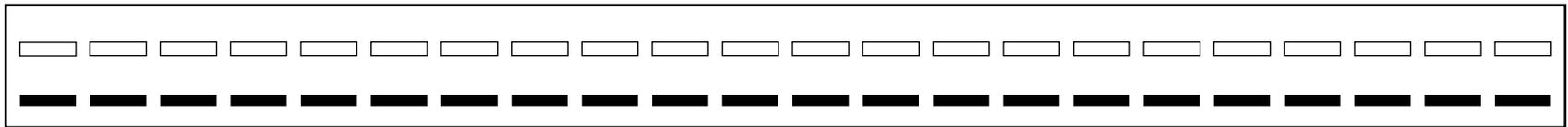


- Polimorfismo

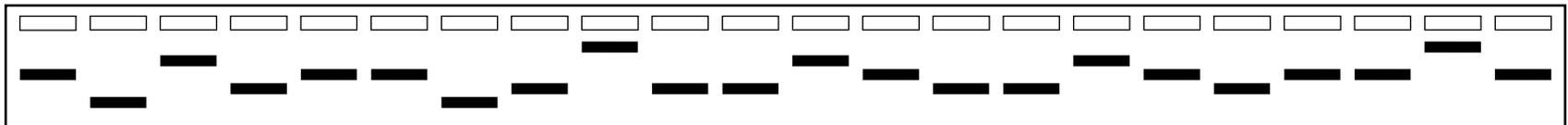


MARCADORES MOLECULARES

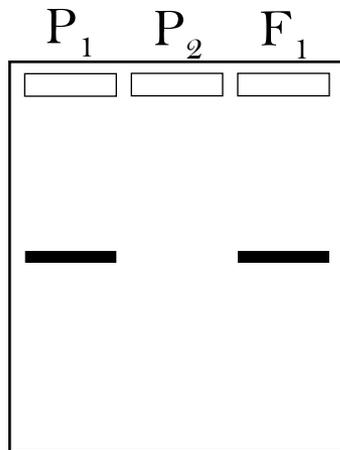
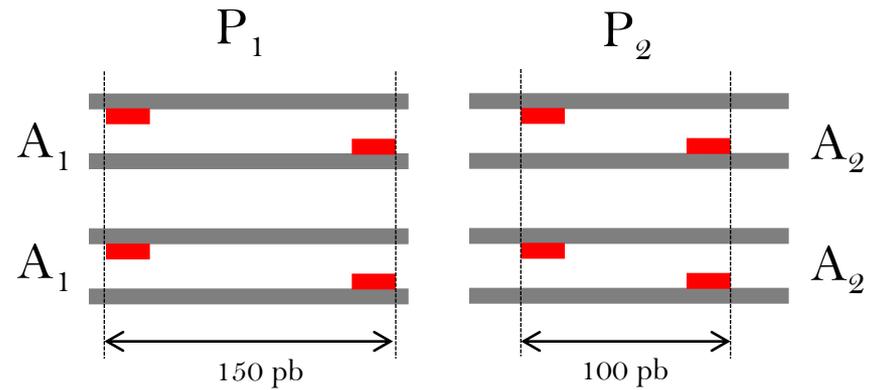
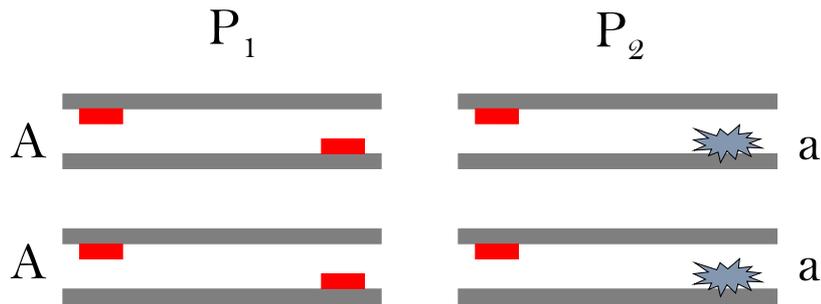
- Variabilidade surge por mutação, que é a base para identificação de marcadores
- Monomorfismo



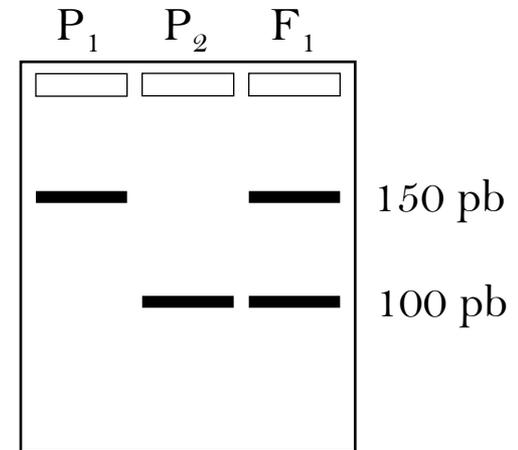
- Polimorfismo



MARCADORES DOMINANTES E CODOMINANTES?



Dominante



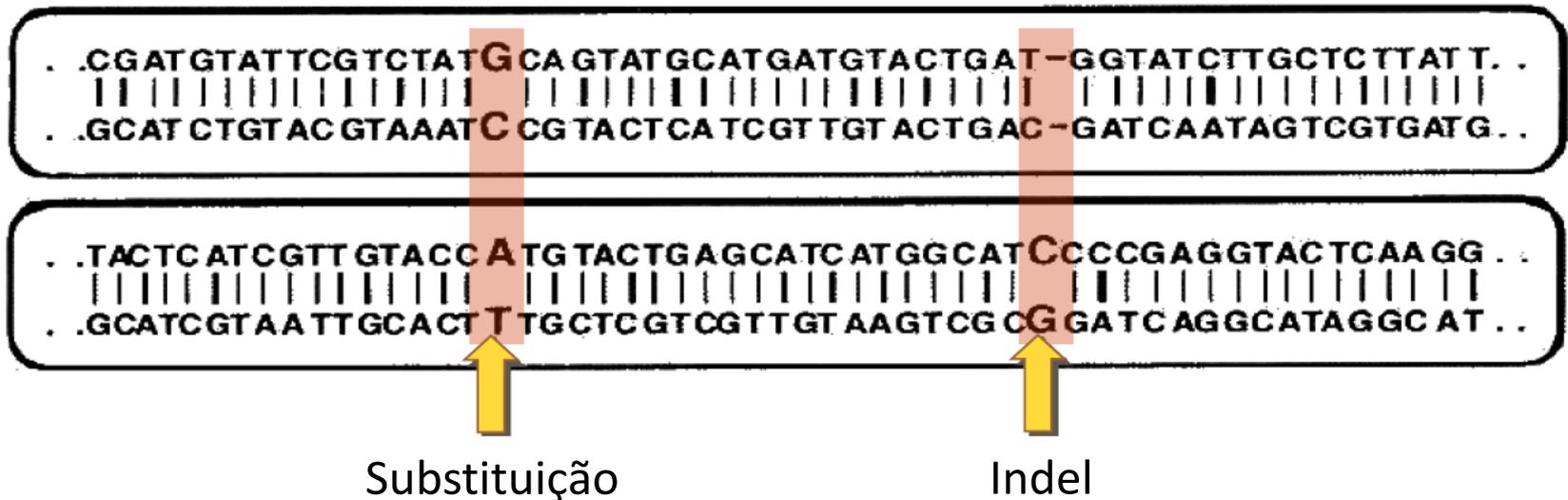
Codominante

MARCADORES MOLECULARES

Origem do polimorfismo

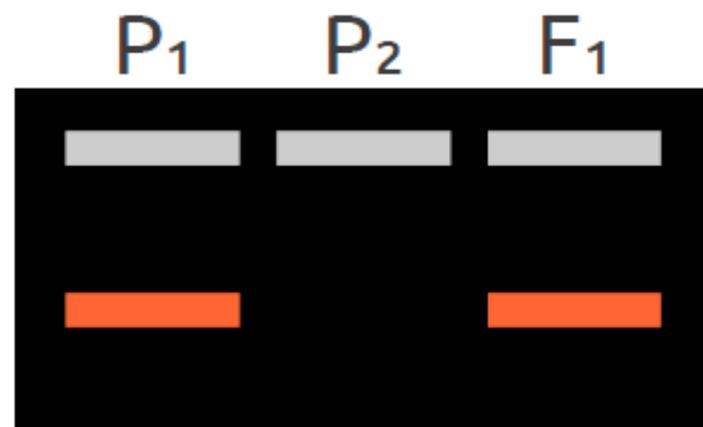
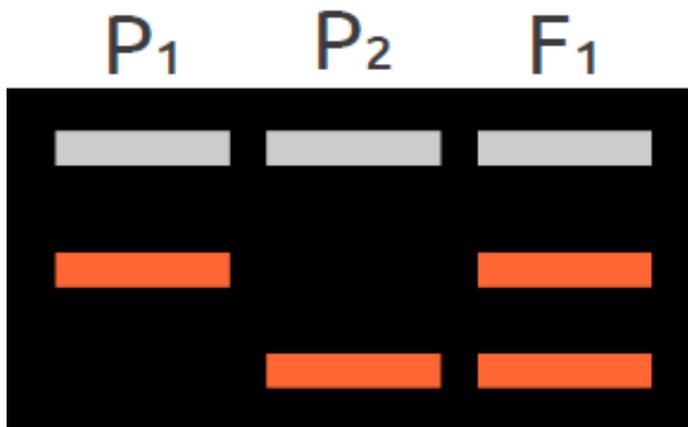
Polimorfismos de tamanho: originados por indels (inserções ou deleções)

Polimorfismos de base: causados por substituição de bases nitrogenadas



MARCADORES MOLECULARES

- Herança
 - Dominantes: não se identificam heterozigotos
 - Codominantes: identificam-se heterozigotos



ALGUNS EXEMPLOS DE MARCADORES MOLECULARES

MICROSSATÉLITES ou SSR

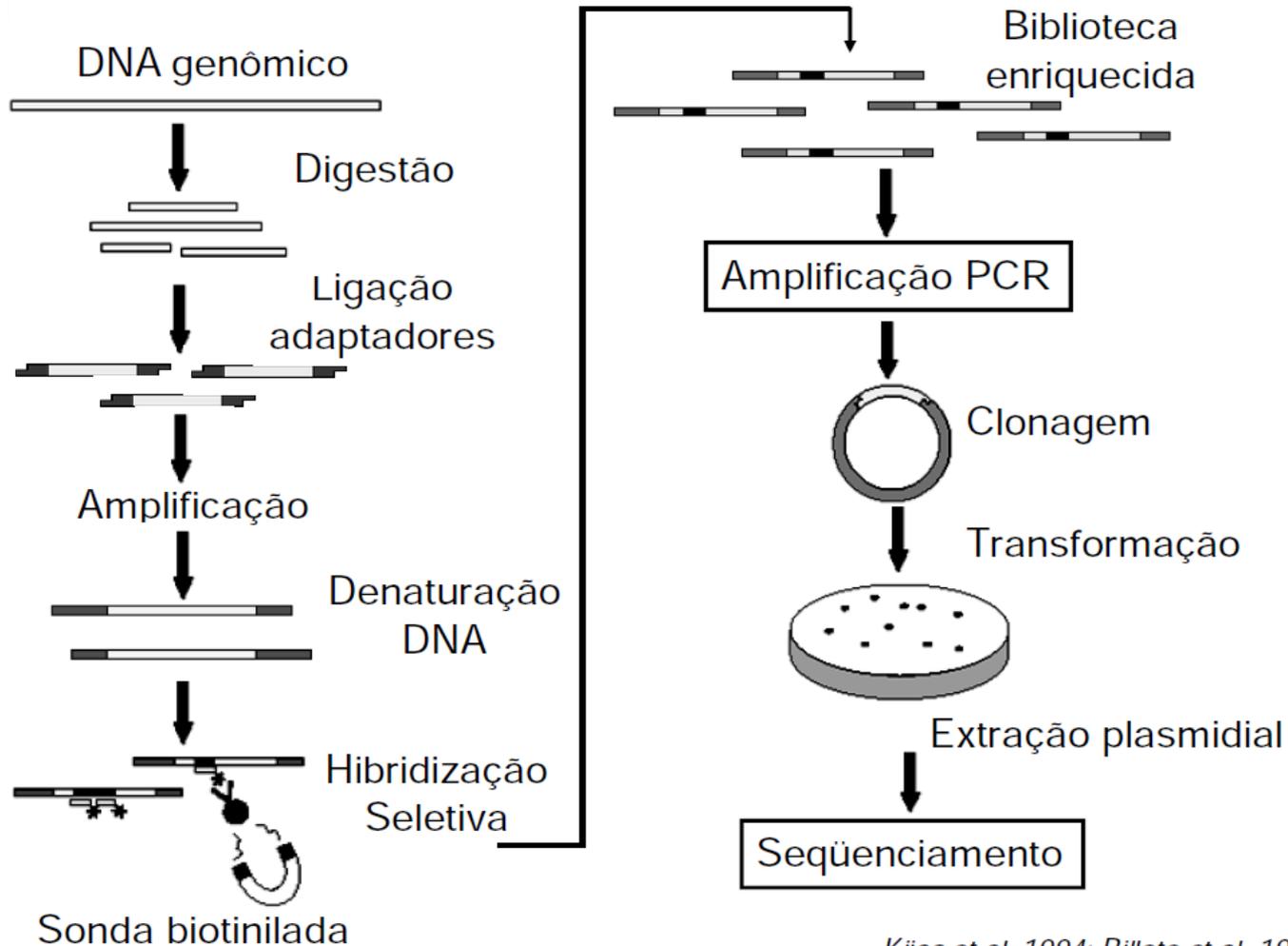
Simple Sequence Repeats

Sequências simples repetidas

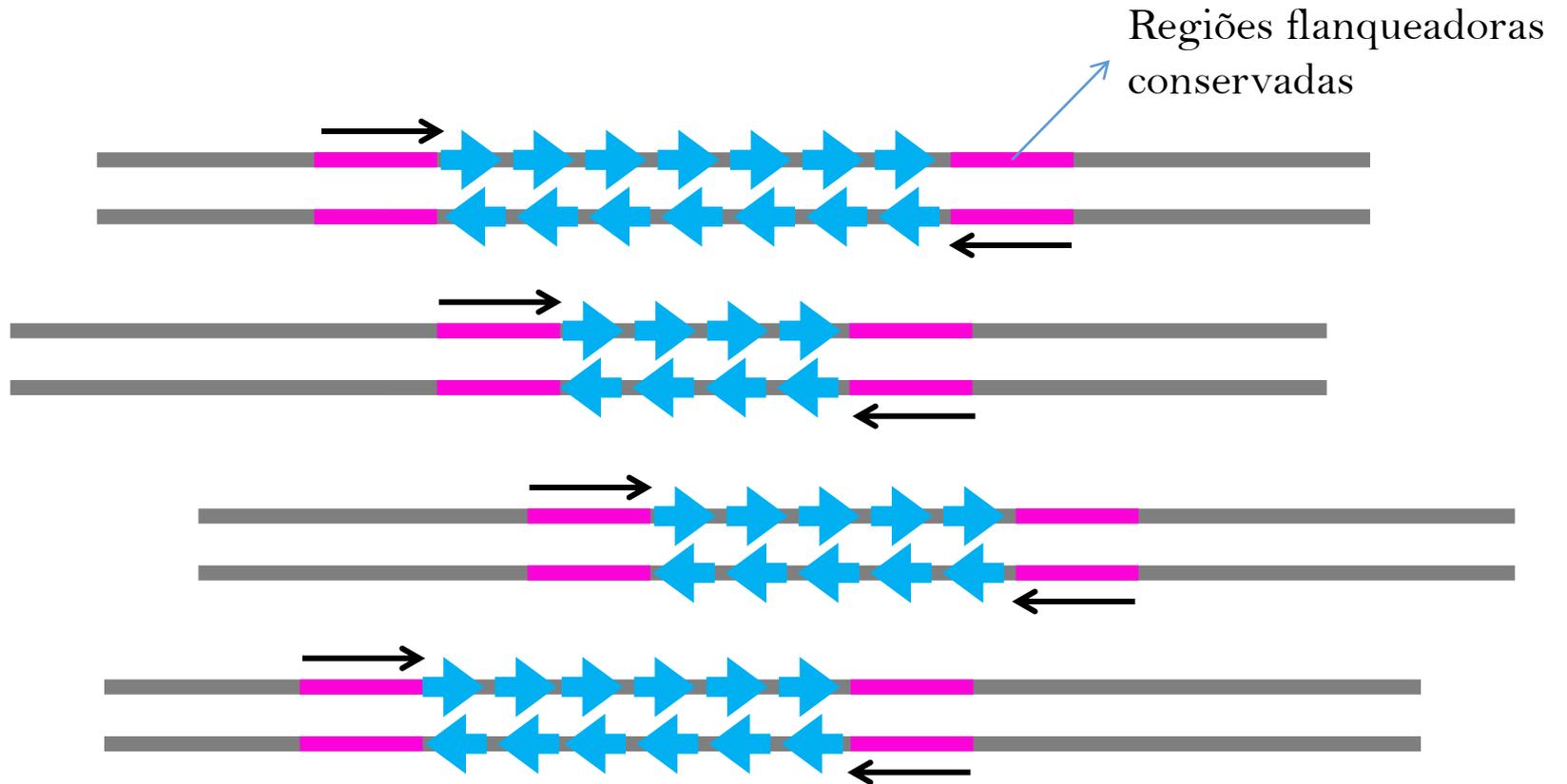
- Pequenas sequências de DNA (1 a 6 nucleotídeos) repetidos em tandem
- **Codominante**
 - Indivíduos heterozigotos são detectados
- Multialélicos e amplamente distribuído no genoma
- **Base genética:**
 - Amplificação de locos específicos de sequências repetitivas
 - Baseados em PCR com primers específicos

MICROSSATÉLITES

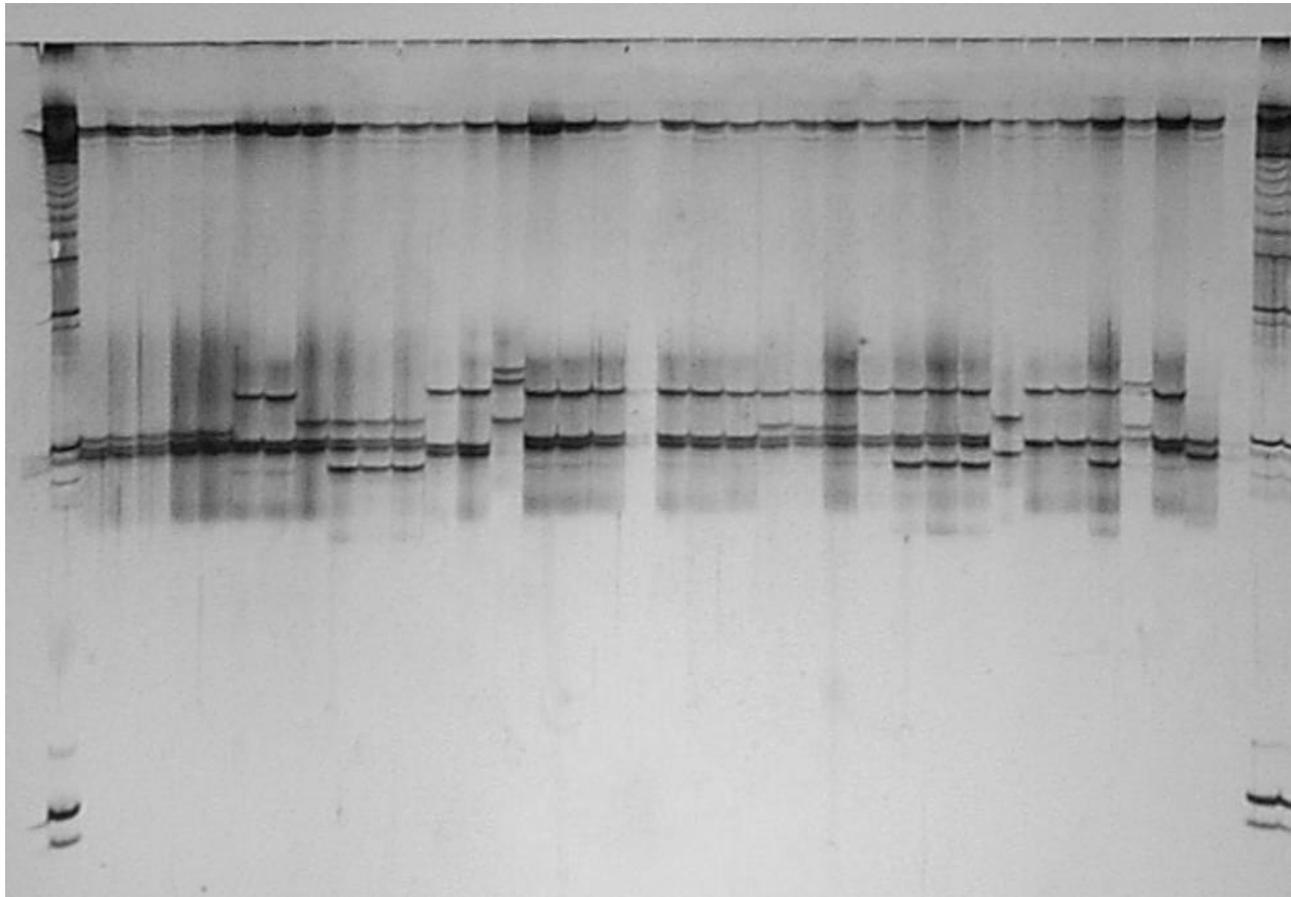
Biblioteca genômica enriquecida com microssatélites



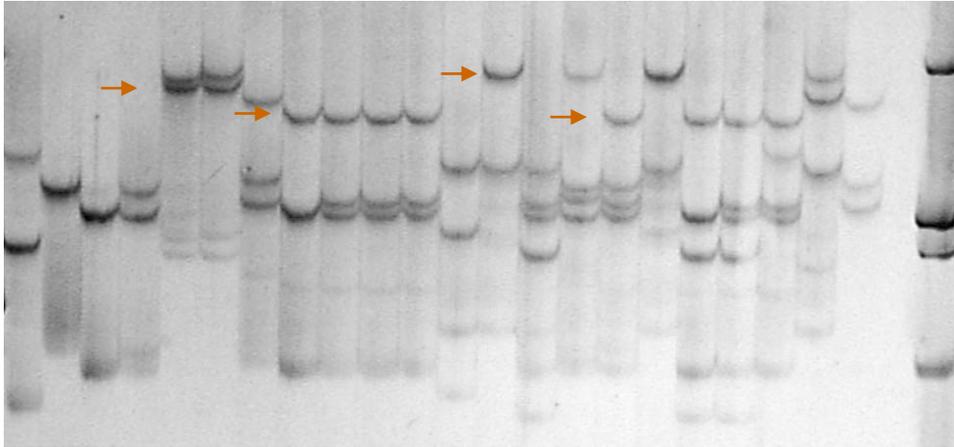
MICROSSATÉLITES



Microsatélites (SSR)

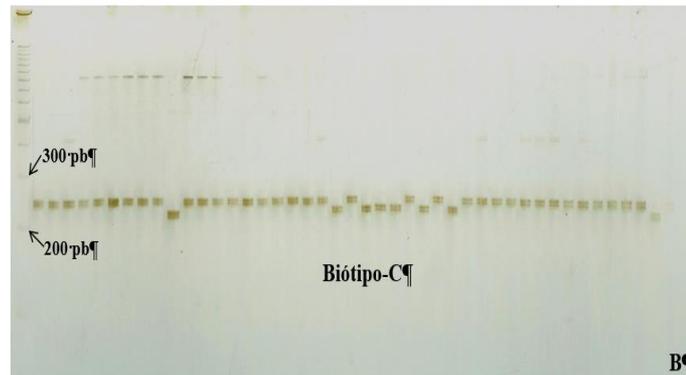
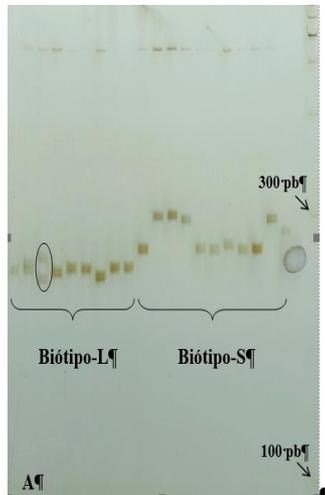


Microssatélites (SSR)



Seta: alelos genoma
B

Musa balbisiana



MICROSSATÉLITES

Vantagens

- ❑ Baseiam-se em PCR
- ❑ Altamente reprodutíveis
- ❑ Codominantes e multialélicos

Desvantagens

- ❑ Há necessidade de informações de sequências oriundas de bibliotecas genômicas ou de cDNA
- ❑ Elevado custo inicial

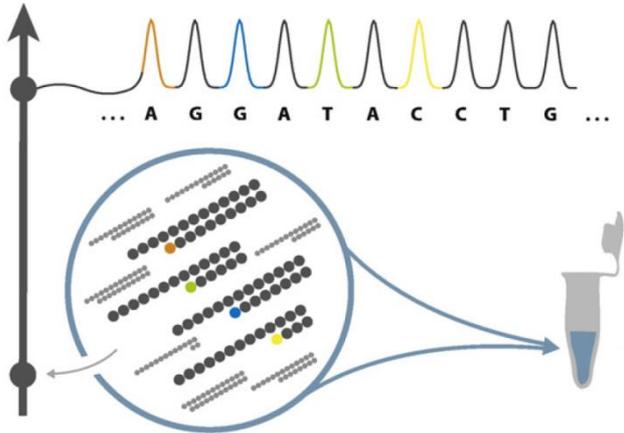
SNP

Single Nucleotide Polymorphism Polimorfismo de base única

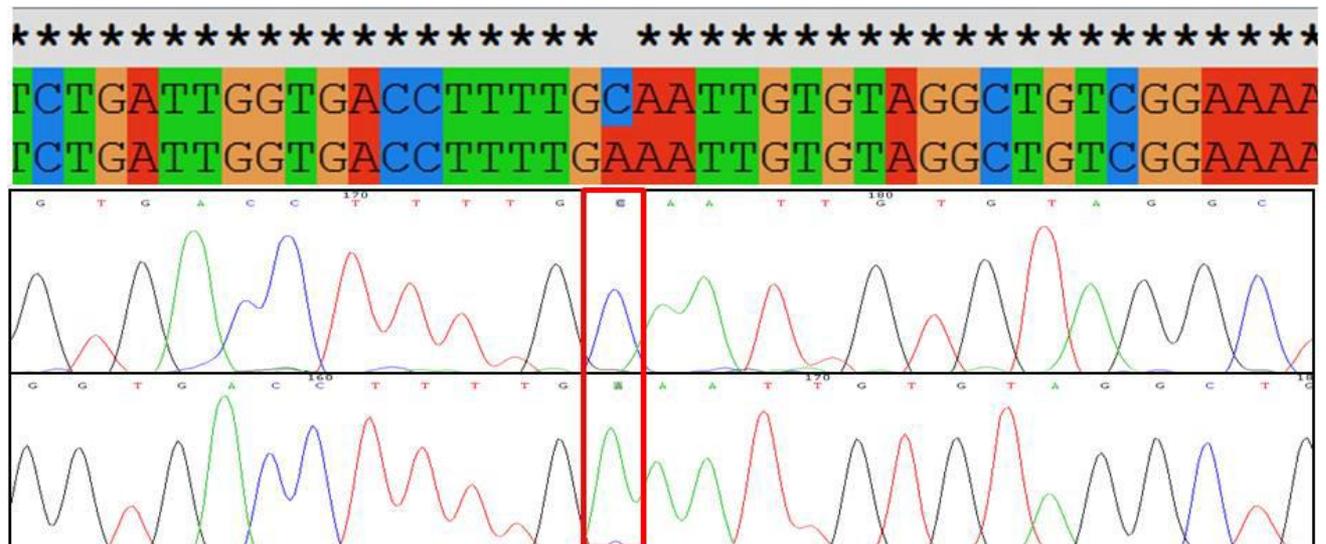
- Polimorfismo resultante da alteração de uma única base
- Abundantes no genoma
- **Codominante e bialélico**
- Há diversas maneiras de se avaliar a variação:
 - Alinhamento e comparação de sequências
 - Métodos baseados em géis
 - Métodos baseados em chips
 - Sequenciamento de 2ª geração, etc.

SNP

Sequenciamento



Alinhamento de
sequências

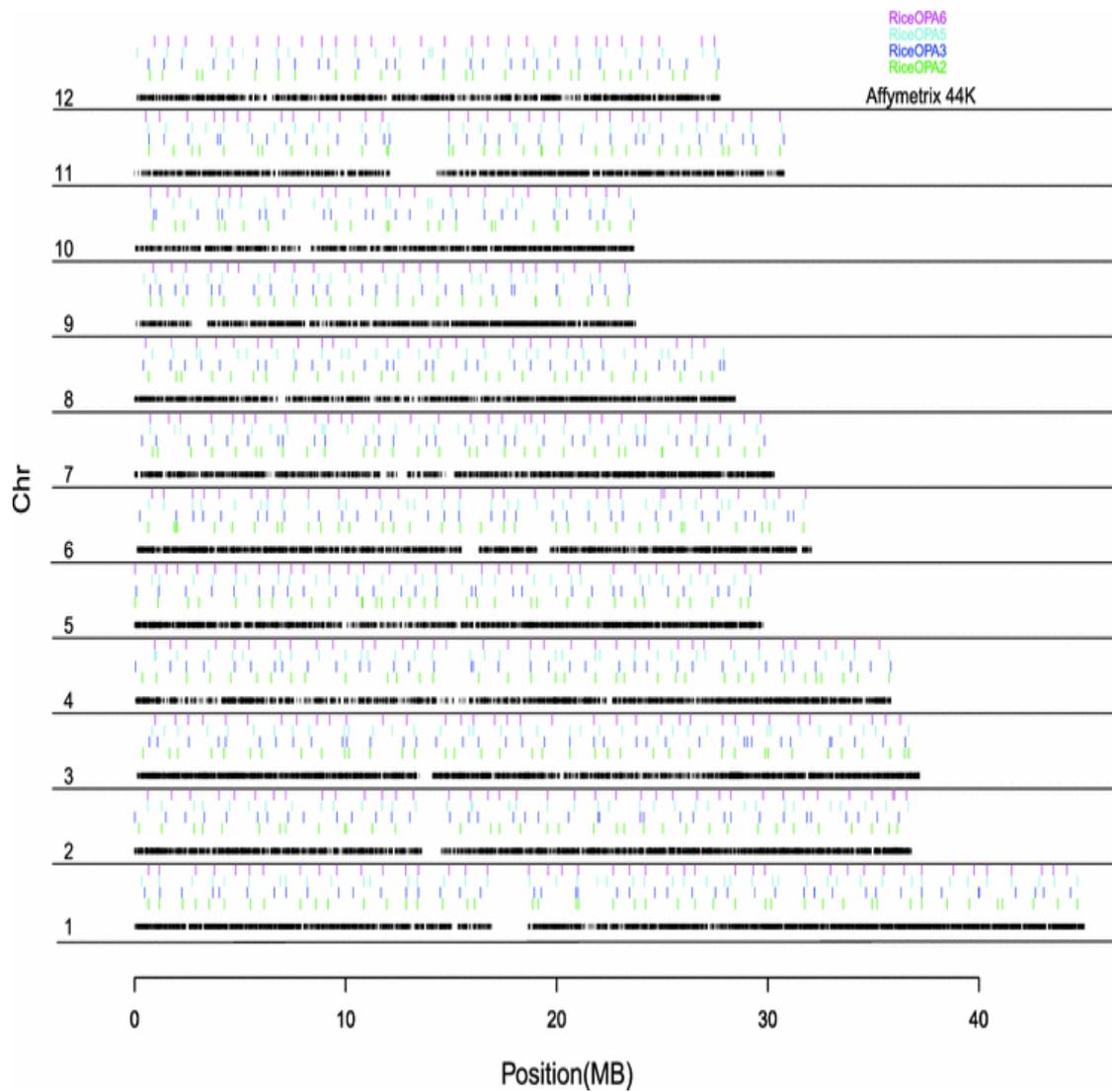


SNP

Exemplos de plataforma de genotipagem por SNPs

| Genotyping Platform | Technology | SNP x sample combinations | Capital investment | Cost per sample | Advantages |
|------------------------------|--------------------------|--|--|------------------|-----------------------------------|
| Illumina Infinium iSelect HD | Fixed array | 3,072 – 700K SNPs x 24 samples | High (iScan) | Moderate to high | Highly multiplexed |
| Affymetrix Axiom | Fixed array | 50K SNPs x 384 samples; 650K SNPs x 96 samples | High (GeneTitan) | Moderate to high | Highly multiplexed |
| Douglas Array Tape | Flexible, PCR-based | 1 SNP/sample x 76,800 reactions/reel | Very High (Nexar, Soellex, Araya) | Very low | Ultra high-throughput |
| Fluidigm Dynamic Arrays | Flexible, PCR-based | 96 SNPs x 96 samples; 24 SNPs x 192 samples | Moderate (IFC Controller, FC1, EP1) | Low | High-throughput |
| RE-based GBS | Genotyping by sequencing | ~10K-100K SNPs x 96 or 384 samples | Low to moderate (NGS outsourced or in-house) | Low to moderate | Lots of data relative to the cost |
| Amplicon sequencing | Genotyping by sequencing | Variable (e.g. 20-500 SNPs x 48-384 samples) | Low to moderate (NGS outsourced or in-house) | Low to moderate | Multiple targeted loci at once |

SNP em arroz



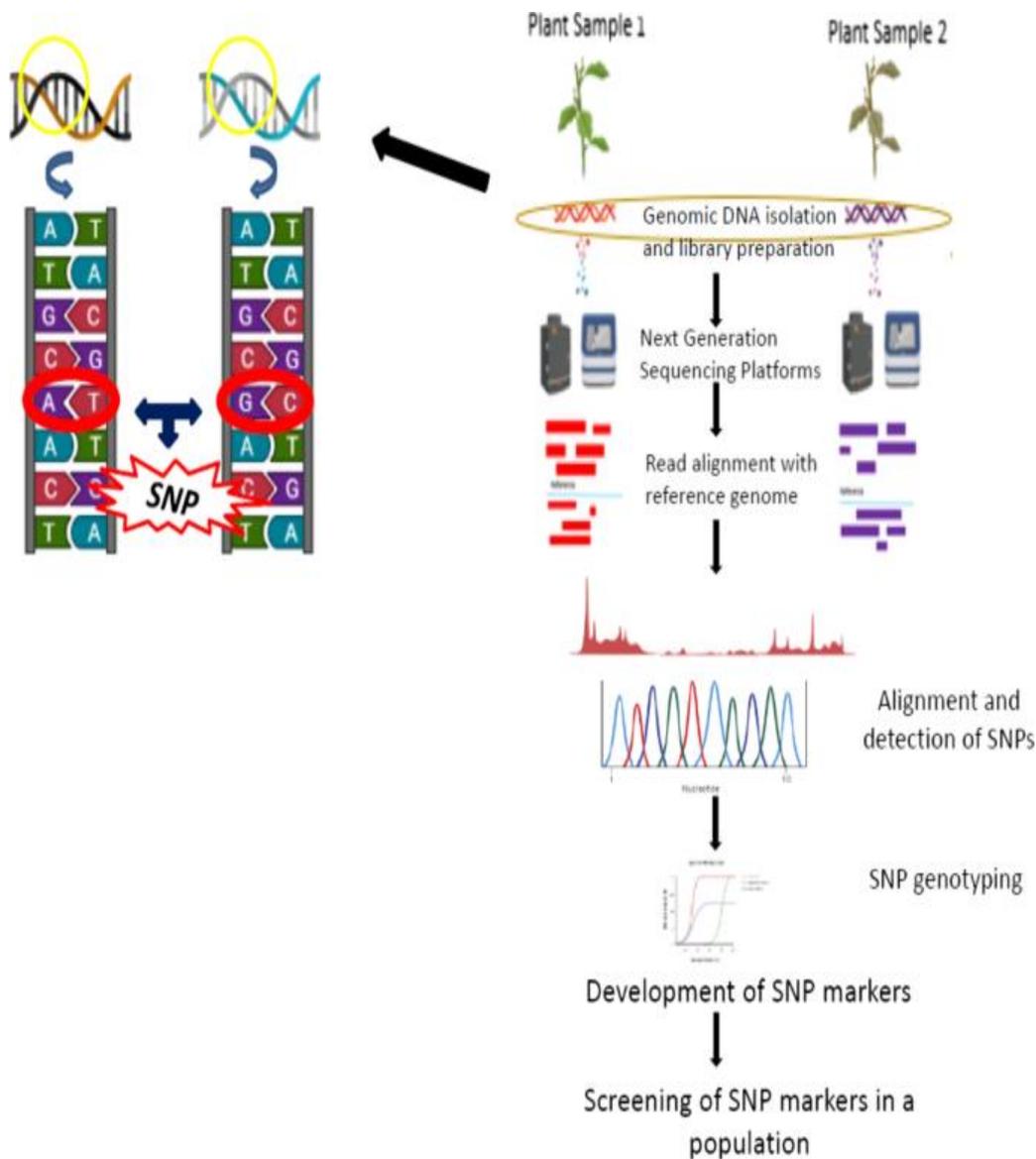
SNP

Target
↓

| Subgroup | id5000197 | id5000200 | id5000204 | id5000205 | id5000025 | id5000209 | id5000217 | id5000223 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| IND | A | T | G | T | T | A | A | C |
| IND | A | T | G | T | T | A | A | G |
| IND | A | T | G | N | T | A | N | C |
| IND | A | T | G | T | T | A | A | C |
| IND | A | T | G | T | T | A | A | C |
| IND | A | C | G | T | T | A | A | G |
| IND | A | T | G | T | T | A | A | G |
| IND | A | T | G | T | T | A | A | C |
| IND | A | T | G | T | T | A | N | C |
| IND | A | T | G | T | T | A | A | G |
| IND | A | T | G | T | T | A | A | G |
| IND | A | T | G | T | T | A | A | C |
| AUS | A | C | G | T | T | A | A | G |
| AUS | A | C | A | C | T | A | C | C |
| AUS | C | C | G | T | C | G | C | C |
| AUS | A | C | G | T | T | A | A | G |
| AUS | C | C | G | T | C | G | C | C |
| AROMATIC | A | C | G | T | T | A | A | G |
| AROMATIC | A | C | G | T | T | A | A | G |
| AROMATIC | A | C | G | T | T | A | A | G |
| AROMATIC | A | C | G | T | T | A | A | G |
| AROMATIC | A | C | G | T | T | A | A | G |
| AROMATIC | A | C | G | T | T | A | A | G |
| TEJ | A | C | A | C | T | A | C | C |
| TEJ | A | C | A | C | T | A | C | C |
| TEJ | A | C | A | C | T | A | C | C |
| TEJ | A | C | A | C | T | A | C | C |
| TEJ | A | C | A | C | T | A | C | C |
| TEJ | A | C | A | C | T | A | C | C |
| TEJ | A | C | A | C | T | A | C | C |

Example of patterns of informative SNPs within and between subgroups. A subset of the rice 44K SNP data from Zhao *et al.* 2011 is shown for representative accessions from four subgroups: *indica* (IND), *aus* (AUS), *aromatic* (Aromatic), and *temperate japonica* (TEJ). Eight SNP loci are shown flanking a gene target, with three SNPs outlined: id5000200 is an example of a SNP mostly monomorphic within subgroups, but polymorphic between *indica* the others; id5000025 is an example of a SNP monomorphic within all groups except for two *aus* accessions; and id5000223 is segregating within *indica* and *aus*, and polymorphic between *aromatic* and *temperate japonica*. In practice, the minor allele frequencies (MAF) within and between subgroups will be used as criteria during the SNP selection process.

SNP – Genotype by Sequencing



SNP

Vantagens:

- ❑ Abundante no genoma
- ❑ Cobertura em alta densidade do genoma
- ❑ Gera muita informação em pouco tempo
- ❑ Marcador codominante (bialélico)

Desvantagens:

- ❑ Elevado custo total (mas por dado, é barato)

COMPARAÇÃO DOS MARCADORES

| | SSR | SNP |
|-------------------------|-----------------------------|-----------------|
| Base genética | PCR com primers específicos | Sequencia-mento |
| Tipo de herança | Codominante | Codominante |
| Número de locos | Único | Único |
| Número de alelos | Vários | Dois |

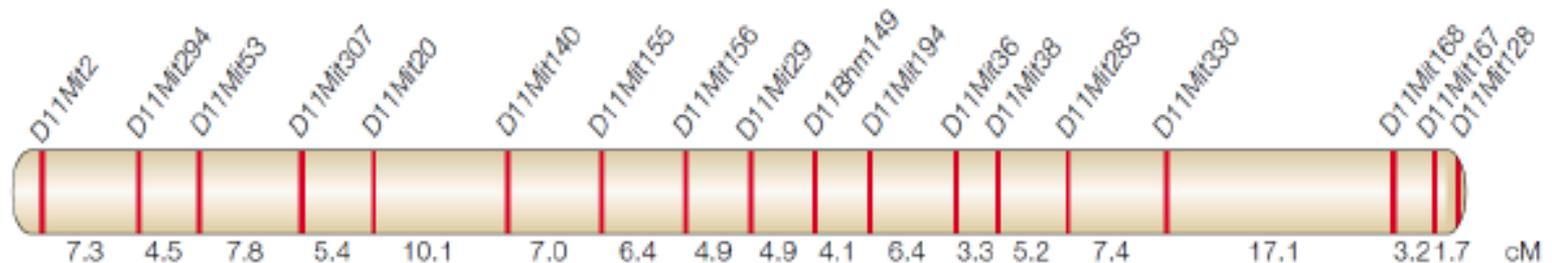
USO DOS MARCADORES NO MELHORAMENTO DE PLANTAS

CONSTRUÇÃO DE MAPAS GENÉTICOS

Mapa genético: representação de ordem e distância entre marcadores genéticos

Ligação genética

- A herança conjunta de diferentes locos dá-se por conexão física

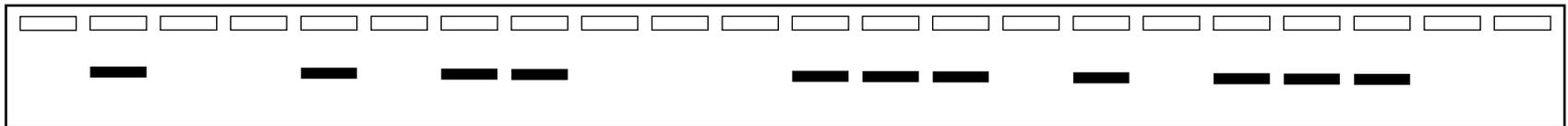
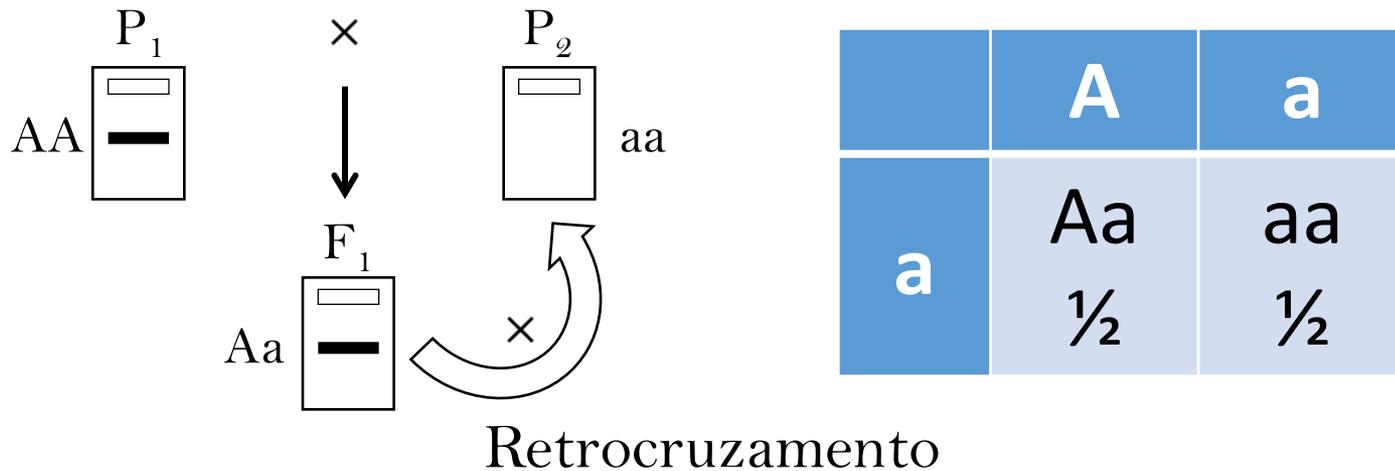


Mus musculus L.

Doerge (2002) *Nature Reviews* 3:43-52 adaptado de:
Butterfield et al. (1999) *The Journal of Immunology* 162:3096-3102

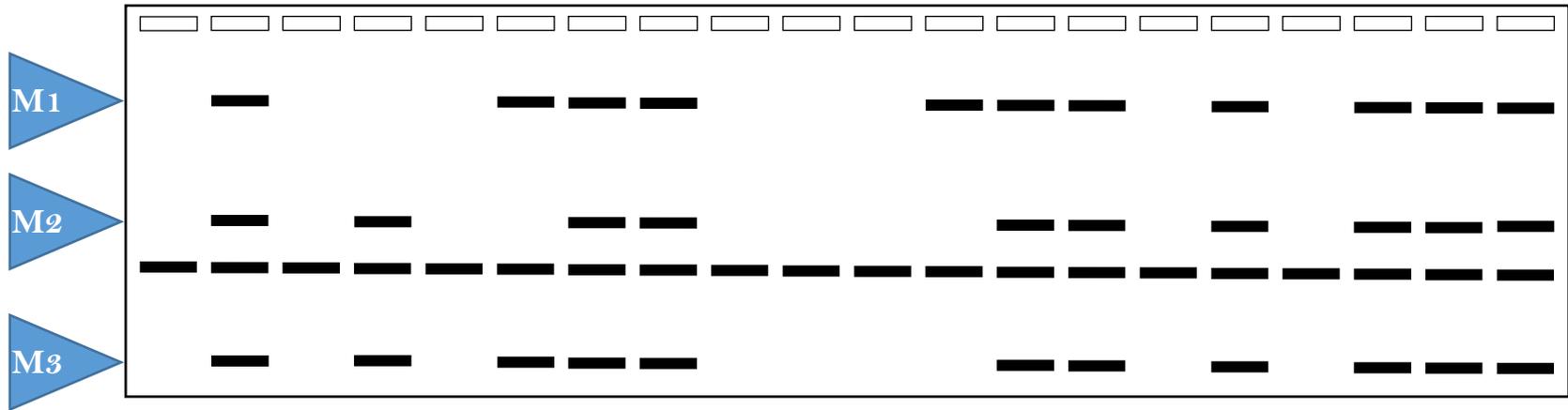
CONSTRUÇÃO DE MAPAS GENÉTICOS

Consiste na avaliação de uma população segregante (retrocruzamentos, F_2 , por exemplo) por meio de vários locos marcadores



CONSTRUÇÃO DE MAPAS GENÉTICOS

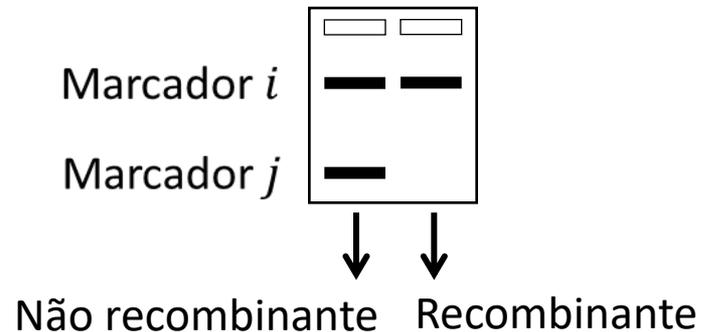
População de retrocruzamento genotipada com três locos AFLP



- Cálculo da fração de recombinação:

$$r_{ij} = \frac{n_{\text{recombinantes}}}{n}$$

Comparando dois locos polimórficos



CONSTRUÇÃO DE MAPAS GENÉTICOS

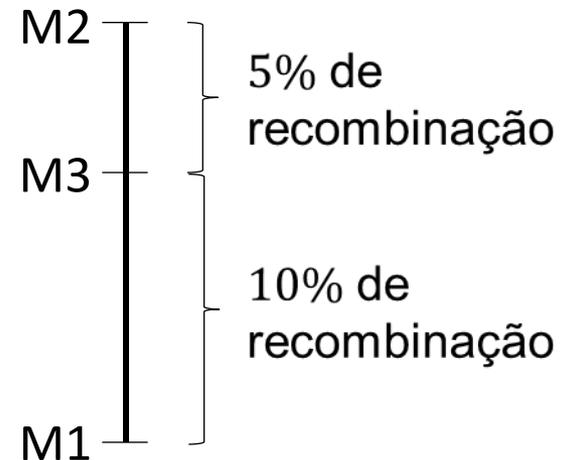
Leitura do gel

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| M1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| M2 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| M3 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |

$$r_{1,2} = \frac{3}{20} = 0,15$$

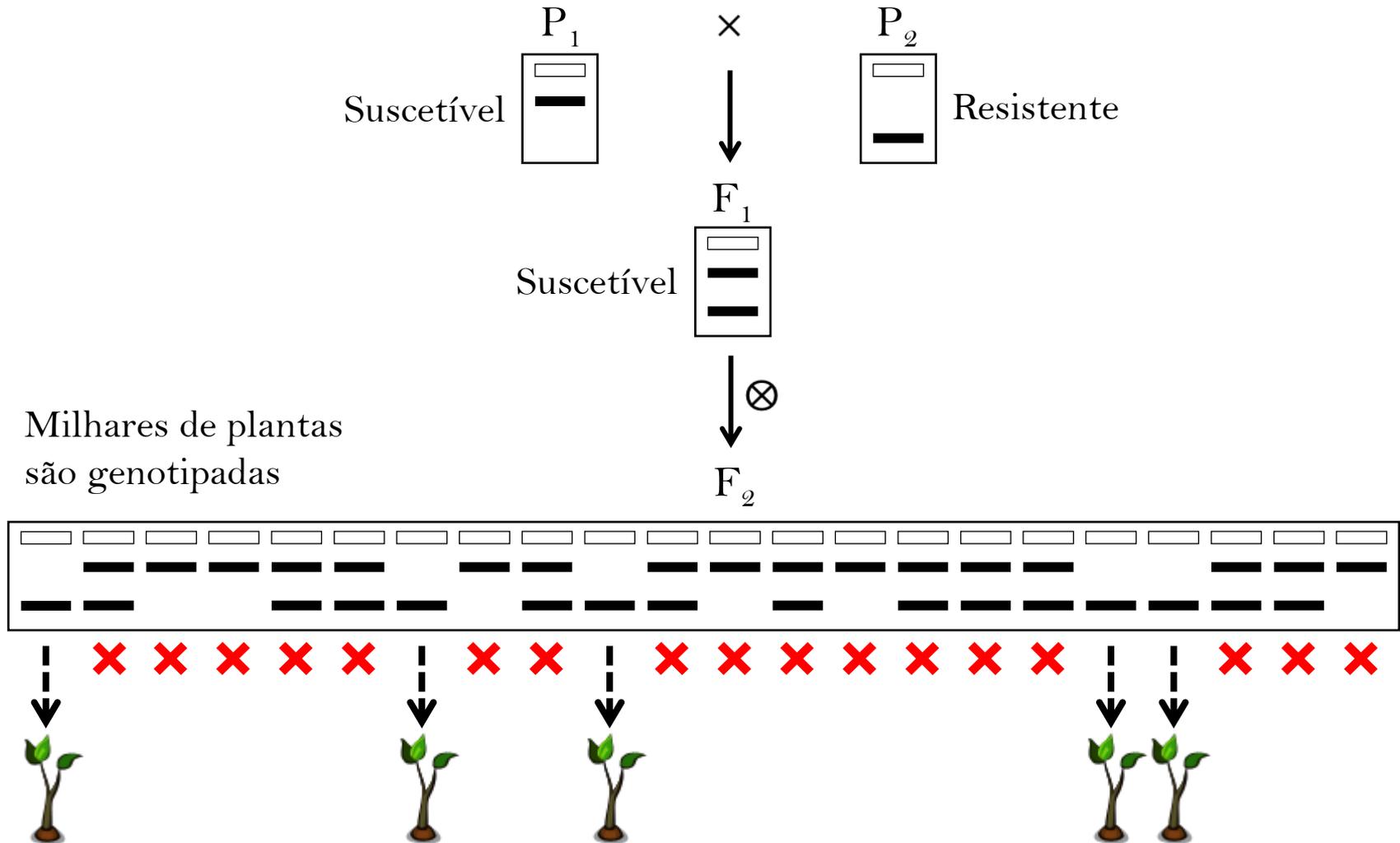
$$r_{1,3} = \frac{2}{20} = 0,10$$

$$r_{2,3} = \frac{1}{20} = 0,05$$



Mapa genético

SELEÇÃO ASSISTIDA POR MARCADORES



A seleção fenotípica é baseada em marcadores de DNA

QTL resistência a vassoura de bruxa do cacauero

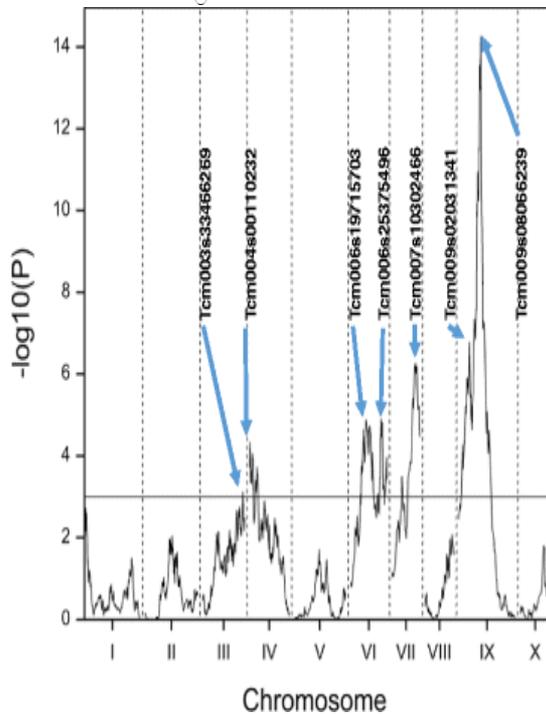
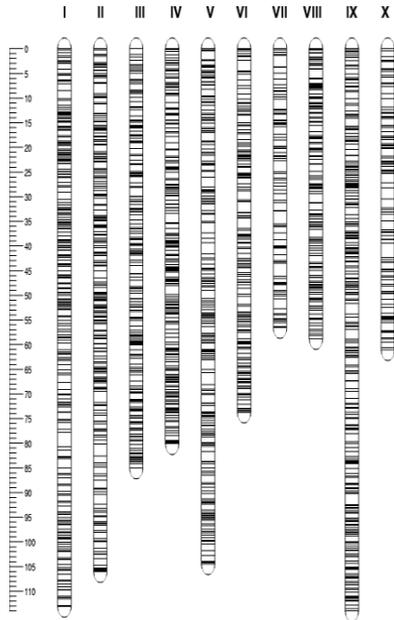


Table 4 Alleles for each of the parental haplotypes (T1, T2, C1 and C2) of 'TSH 1188' and 'CCN 51' for each QTL and representative SNP marker associated with WBD

| QTL | SNP marker | TSH 1188 | | CCN 51 | |
|--------|-----------------|-----------|-----------|-----------|-----------|
| | | Allele T1 | Allele T2 | Allele C1 | Allele C2 |
| QTL3.1 | Tcm003s33466269 | G | G | G | T |
| QTL4.1 | Tcm004s00110232 | T | C | C | C |
| QTL6.1 | Tcm006s19715703 | T | T | T | C |
| QTL6.2 | Tcm006s25375496 | G | T | G | G |
| QTL7.1 | Tcm007s10302466 | G | G | A | G |
| QTL9.1 | Tcm009s08066239 | A | G | A | A |
| QTL9.2 | Tcm009s02031341 | A | G | A | A |

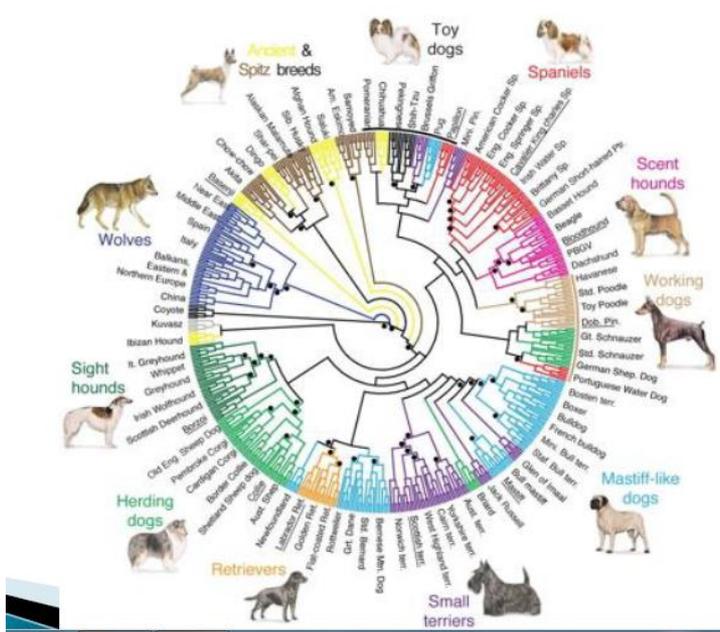
The alleles marked in bold are segregating in MP01 and are also the favorable alleles associated with WBD resistance

Table 3 Effects of selected QTL, values of the significance level $-\log_{10}(P)$ and % variance accounted for by including markers in the final QTL model

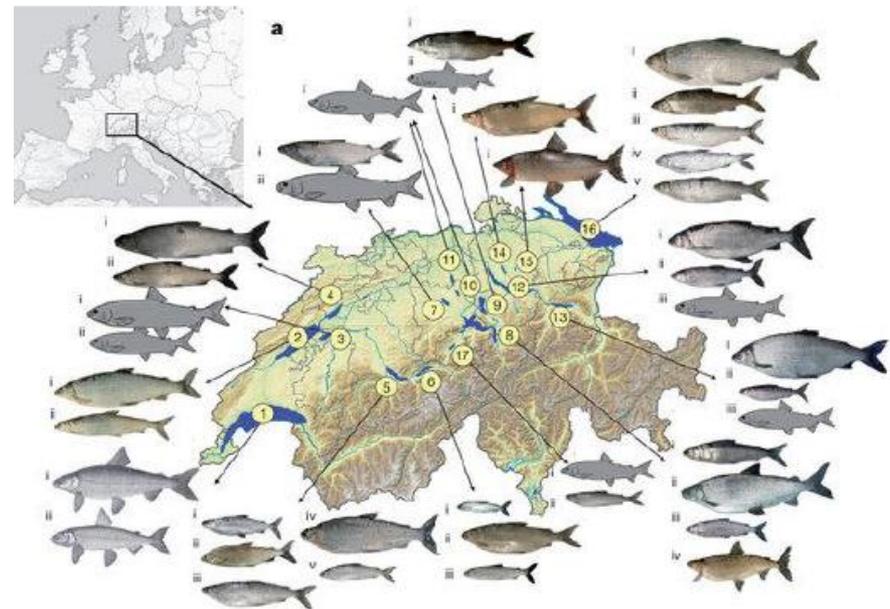
| Marker | Chromosome | Position (cM) | $-\log_{10}(P)$ | Trait | TSH-1188 | CCN-51 | TSH 1188 × CCN 51 | % variance |
|-----------------|------------|---------------|-----------------|-------|------------------------|------------------------|------------------------|------------|
| Tcm003s33466269 | III | 81.53 | 3.739 | VB | 0.069 (0.0483) | 0.086 (0.0480) | -0.025 (0.0479) | 0.6 |
| | | | | CB | 0.054 (0.0449) | 0.209 (0.0446) | 0.032 (0.0445) | 3.3 |
| Tcm004s00110232 | IV | 0.55 | 6.031 | VB | 0.177 (0.0483) | -0.111 (0.0487) | 0.015 (0.0486) | 2.9 |
| | | | | CB | 0.248 (0.0455) | -0.033 (0.0452) | 0.014 (0.0452) | 4.7 |
| Tcm006s19715703 | VI | 31.01 | 3.679 | VB | -0.107 (0.0520) | 0.190 (0.0551) | 0.118 (0.0496) | 5.2 |
| | | | | CB | -0.059 (0.0484) | 0.120 (0.0512) | 0.036 (0.0461) | 2.5 |
| Tcm006s25375496 | VI | 61.73 | 2.318 | VB | 0.015 (0.0551) | 0.015 (0.0551) | -0.097 (0.0491) | 3.1 |
| | | | | CB | -0.169 (0.0483) | 0.0206 (0.0512) | -0.111 (0.0456) | 5.4 |
| Tcm007s10302466 | VII | 47.55 | 8.306 | VB | -0.105 (0.0478) | -0.105 (0.0478) | -0.080 (0.0486) | 6.3 |
| | | | | CB | 0.146 (0.0452) | -0.100 (0.0444) | -0.139 (0.0452) | 3.4 |
| Tcm009s02031341 | IX | 13.73 | 3.557 | VB | -0.221 (0.0579) | -0.221 (0.0579) | -0.094 (0.0500) | 6.0 |
| | | | | CB | -0.065 (0.0498) | -0.203 (0.0538) | -0.024 (0.0465) | 1.7 |
| Tcm009s08066239 | IX | 44.36 | 11.24 | VB | 0.383 (0.0539) | 0.057 (0.0573) | -0.062 (0.0503) | 13.5 |
| | | | | CB | 0.239 (0.0501) | 0.110 (0.0533) | -0.049 (0.0467) | 6.2 |

| QTL and haplotypes | Total | Observed | | Expected | | P-value | % resistant |
|--------------------|------------------|-----------|----------|-----------|----------|-----------------|-------------|
| | | Resistant | Suscept. | Resistant | Suscept. | | |
| QTL9.1_T1 | 230 | 111 | 119 | 115 | 115 | 5.98E-01 | 48.26 |
| QTL9.1_T2 | 229 | 188 | 41 | 114.5 | 114.5 | 2.63E-22 | 82.10 |
| QTL9.1_C1 | 236 ^a | 159 | 77 | 118 | 118 | 9.41E-08 | 67.37 |
| QTL9.1_C2 | 218 ^a | 136 | 82 | 109 | 109 | 2.55E-04 | 62.39 |

USO DOS MARCADORES EM ESTUDOS DE DIVERSIDADE E CONSERVAÇÃO



Filogenia



Distribuição geográfica

POR QUE USAR MARCADORES?

Fornece informações a respeito de:

- ❑ Diversidade genética
- ❑ Endogamia e sistemas de cruzamento
- ❑ Fluxo gênico
- ❑ Paternidade
- ❑ Sexagem

Auxilia na definição de estratégias de manejo e conservação

Ajuda a organizar coleções de germoplasma

POR QUE CONSERVAR?

A variabilidade é a base da evolução

Espécies com baixa variabilidade tem seu potencial evolutivo reduzido

- Ex: uma praga que mata um indivíduo, numa população com baixa variabilidade, mata todos!

Base para o desenvolvimento de novas cultivares

- Alimentos mais ricos em nutrientes
- Variedade mais produtivas

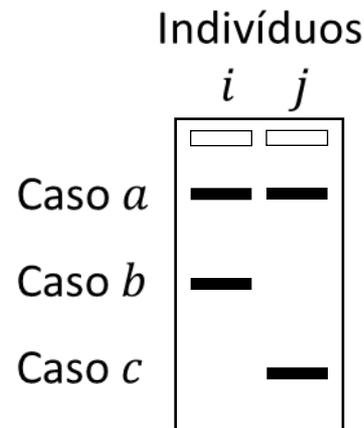
MARCADORES DOMINANTES

- Cálculo de similaridade genética: coeficiente de Jaccard

$$S_{ij} = \frac{a}{a + b + c}$$

Em que:

- $a = n^{\circ}$ de locos com alelos presentes nos indivíduos i e j
- $b = n^{\circ}$ de locos com alelos presentes em i e ausentes em j
- $c = n^{\circ}$ de locos com alelos ausentes em i e presentes em j



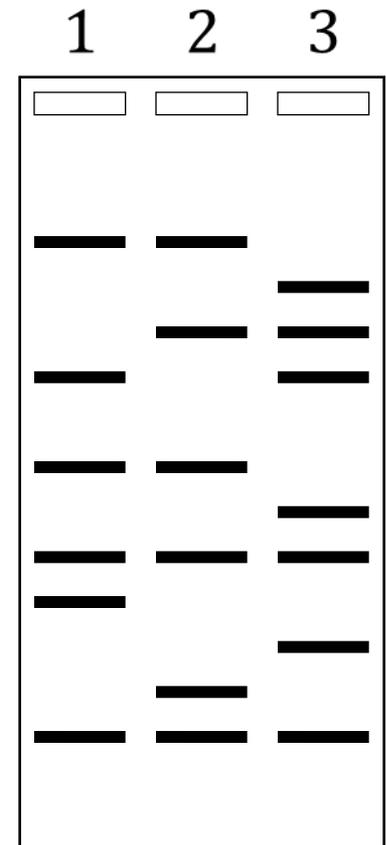
MARCADORES DOMINANTES

$$S_{12} = \frac{a}{a + b + c} \Rightarrow$$

$$S_{13} = \frac{a}{a + b + c} \Rightarrow$$

$$S_{23} = \frac{a}{a + b + c} \Rightarrow$$

- Conclusão: são mais similares os indivíduos e mais diferentes os indivíduos



ESTUDO DIRIGIDO

1. O que são marcadores moleculares?
2. Definir marcadores moleculares.
4. Que são marcadores dominantes e codominantes?
5. Quais os tipos de marcadores moleculares?
6. Qual a utilidade de marcadores moleculares no mapeamento genético?
7. Qual a utilidade de marcadores moleculares nos estudos de diversidade genética e conservação das espécies?

Leitura

Guimarães et al., 2009

