

# O USO DE HAPLÓTIPOS EM ASSOCIAÇÃO GENÔMICA: RODANDO ANÁLISES

*Fabíeli Loíse Braga Feitosa*

*Doutora em Genética e Melhoramento Animal*

*Pós-doutoranda em Zootecnia na UFBA*



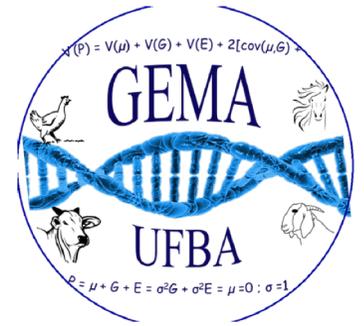
CURSO: APLICABILIDADE DE SOFTWARES EM ANÁLISES GENÔMICAS



ESCOLA DE MEDICINA  
VETERINÁRIA E ZOOTECNIA



# GRUPO DE ESTUDOS DE GENÉTICA E MELHORAMENTO ANIMAL - GEMA



## DOCENTES:

- Prof. Dr. Gregório Miguel Ferreira de Camargo
- Prof. Dr. Luís Fernando Batista Pinto
- Prof. Dr. Raphael Bernal Costa
- Profa. Dra. Thereza Cristina Borio dos Santos Calmon de Bittencourt

## PÓS-DOCTORANDAS:

- Dra. Iara Del Pilar Solar Diaz
- Dra. Valdecy Aparecida Rocha da Cruz
- Dra. Fabieli Loise Braga Feitosa

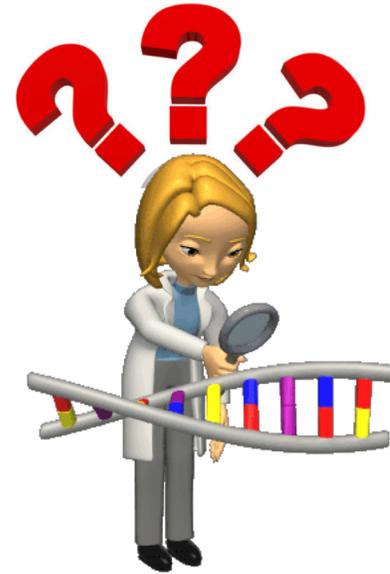
**DOUTORANDOS: 8**

**MESTRANDOS: 5**

**INICIAÇÃO CIENTÍFICA: 5 (bolsistas) e 10 (voluntários)**

# INTRODUÇÃO

O QUE SÃO HAPLÓTIPOS?



A combinação de alelos em *loci* adjacentes, que fazem parte do mesmo cromossomo e são transmitidos de forma conjunta, ele pode ser formado por um ou vários alelos, ou até mesmo pelo cromossomo inteiro.

-A C T T T G C T C -  
-A C T T A G C T T -  
-A A T T T G C T C -

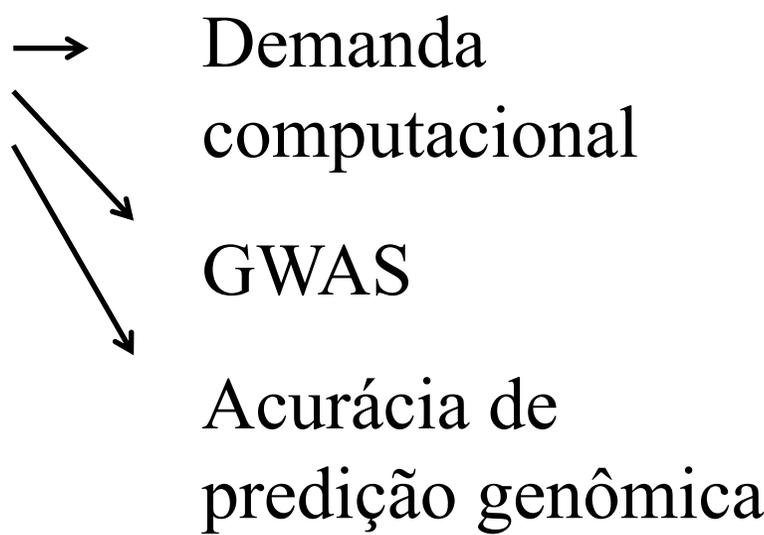
↑    ↑    ↑  
SNP<sub>1</sub> SNP<sub>2</sub> SNP<sub>3</sub>



Haplotype 1  
Haplotype 2  
Haplotype 3

C    T    C  
C    A    T  
A    T    C

↑    ↑    ↑  
SNP<sub>1</sub> SNP<sub>2</sub> SNP<sub>3</sub>

- Haplótipos:
  - $\uparrow$  LD  $\rightarrow$  identificação efeito aditivo maior;
  - Várias mutações  $\rightarrow$  região cromossômica  $\rightarrow$  pequenos efeitos
  - Número de blocos e alelos  $\rightarrow$  
    - Demanda computacional
    - GWAS
    - Acurácia de predição genômica

- Uso de informações de haplótipos:
  - Seleção Genômica
  - Associação Genômica
- Maiores acurácias de predição;
- Variações em características economicamente importantes;
- Efeitos dos haplótipos sobre o fenótipo;
- Predição dos valores genéticos

- Métodos de análises/reconstrução dos haplótipos:
  - LD;
  - Tamanho fixo (Mb);
  - Número fixo de SNPs.

- Etapas para análise utilizando blocos de haplótipos:
  - Imputação dos missing;
  - *Phasing* (“faseamento”);
  - Reconstrução dos blocos de haplótipos;
  - Construção da matriz de pseudoSNPs.

- Imputação dos missing genotypes:
  - FImpute;
  - BEAGLE.
- *Phasing* (“Faseamento”):
  - FImpute;
  - BEAGLE;
  - PLINK.

- Reconstrução dos blocos de haplótipos:
  - HaploView.
- Construção da matriz de pseudoSNPs:
  - Unix;
  - R software.

- Arquivo genótipos:
  - Controle de qualidade (BLUPf90 family):
    - ☒ SNPs posição genômica desconhecida;
    - ☒ marcadores monomórficos;
    - ☒ SNPs que não estão em Equilíbrio de Hardy-Weinberg, com probabilidade muito baixa ( $p\text{-valor} < 10^{-6}$ );
    - ☒ MAF (frequência de alelos menores)  $< 0,05$ ;
    - ☒ taxa de atribuição dos genótipos (Call Freq)  $< 90\%$ ;
    - ☒ excesso de genótipos heterozigotos (Het Excess).

- FImpute (missing genotypes and phasing):

- Input Genotype File Example:

ANIMAL	CHIP	GENÓTIPO
		1801565 1 102111111111111020212111102222222220010202022021221012
		1801214 1 01122200202020220020202000200000021011211022020212102
		1802409 1 111111111111111021101110011221112220001111111010222012
		1274961 1 002222000200220020202110011221112110101111121112111101
		1801235 1 111111111111111021101110011221112220001111111011222012
		1757562 1 101212111110111010212111102222222220021211022021212102
		1802258 1 00222200020022002020211001122111211110111112111111101
		1802276 1 002222000200220020202020002222222221000202022022222002
		1801947 1 00221211202010210021211110211111120010202022021221012
		1802404 1 00222200202020220020202000200000022010202022022222002
		1277110 1 101222000200220020202110011221112110101111121112111101
		1802418 1 510111111111111021101110011221112220001111111011222012
		1801237 1 101222000200220020202020002222222220000202022022222002
		1801105 1 002222000200220020212111102222222220000202022022222002
		1801582 1 00222200202020220020202000200000022101111111010222012

## - Mapa File Example:

SNP_ID	CHR	Posição
BovineHD0100046367	1	89725 3
BovineHD0100000033	1	111645 4
BovineHD0100000037	1	127885 5
Hapmap43437-BTA-101873	1	135098 6
BovineHD0100000039	1	146011 7
BovineHD0100000040	1	147231 8
BovineHD0100000042	1	149772 9
BovineHD0100000043	1	151060 10

- Parameters card:

```
title="Imputacao";  
genotype_file="Gclean3";  
snp_info_file="mapa.txt";  
output_folder="outputimp4";  
parentage_test /find_identical;  
exclude_snp="exclude";  
save_hap_lib;  
njob=5;
```

## - Genotypes\_imp example:

ID	Chip Calls...
1295341	1 302444333433443020232203
1291603	1 002222002020202200202020
1802246	1 002222000200220020242434
1801240	1 443303222024002044333203
1801203	1 302222002020202200202020
1801935	1 002222003430243340202020

Haplotypes codes:

If  $i=0$ : s1: 1    s2: 1 ( $A_1A_1$ )

If  $i=2$ : s1: 2    s2: 2 ( $A_2A_2$ )

If  $i=3$ : s1: 1    s2: 2 ( $A_1A_2$ )

If  $i=4$ : s1: 2    s2: 1 ( $A_2A_1$ )

## - File example:

```
Animal → 1295341  
1 1 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2 → Strand 1  
1 1 2 1 2 1 2 2 2 2 1 2 1 1 1 2 1 1 2 → Strand 2  
1291603  
1 1 2 1 2 2 1 2 1 2 2 1 2 2 2 1 1 2 2  
1 1 2 1 2 1 2 2 1 2 2 1 2 2 2 1 1 1 2
```

## - File input example for HaploView:

```
GENOTY 1291603 1 1 2 1 2 2 1 2 1 2 2 1 2 2  
GENOTY 1291603 1 1 2 1 2 1 2 2 1 2 2 1 2 2  
GENOTY 1802246 1 1 1 2 2 1 2 2 2 2 1 2 1 1  
GENOTY 1802246 1 1 1 2 2 1 2 2 2 2 1 2 1 1  
GENOTY 1801240 1 1 2 1 2 1 2 2 1 2 2 1 2 1  
GENOTY 1801240 1 1 2 1 2 2 1 2 1 2 2 1 2 2
```

- HaploView:
  - Método Intervalo de Confiança baseado no LD (Gabriel et al., 2002):
    - intervalo de confiança  $\geq 95\%$ ;
    - $0,98 \leq D' > 0,7$ ;
    - “strong LD”;
    - blocos não sobrepostos.

## - Haplotype output file example:

```
BLOCK 1.  MARKERS: 1 2
11 (0.929)      |0.646  0.281  0.000|
22 (0.071)      |0.016  0.000  0.053|
Multiallelic Dprime: 0.769
BLOCK 2.  MARKERS: 5 6 7 8
2122 (0.662)    |0.413  0.128  0.095  0.022|
2212 (0.281)    |0.017  0.256  0.005  0.000|
1121 (0.054)    |0.000  0.002  0.052  0.000|
Multiallelic Dprime: 0.737
BLOCK 3.  MARKERS: 9 10 11 12
2212 (0.431)    |0.416  0.016  0.000|
1221 (0.387)    |0.000  0.142  0.243|
1121 (0.155)    |0.017  0.135  0.002|
1222 (0.022)    |0.006  0.016  0.000|
Multiallelic Dprime: 0.828
```

- File with the alleles of the 2 strands of each haplotype, in which the lines are the 2 strands of each animal (1 and 2 - animal 1; 3 and 4 - animal 2; ...) and the columns are the haplotype blocks.

```
11 2212 1221 121 22 11 121121 21 122
11 2122 2212 112 22 22 112212 22 211
11 2212 1221 221 22 22 112212 22 211
11 2122 1221 221 22 22 121121 11 112
11 2122 2212 112 22 22 212212 22 112
11 2122 2212 112 11 11 121121 11 112
11 2122 1221 121 11 21 112212 22 211
11 2212 1221 221 22 22 112212 22 211
```

- File of the incidence matrix, where the lines is the identification of the animal and the columns is the haplotype block and the allele type.

```
hap.1.11 hap.1.12 hap.1.22 hap.2.1121 hap
1295341 2 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0
1291603 2 0 0 0 0 0 0 1 1 0 0 0 2 0 0 0 0
1802246 2 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0
1801240 2 0 0 0 0 0 0 1 1 0 0 0 2 0 0 0 0
1801203 2 0 0 0 0 0 0 1 1 1 0 0 1 0 0 0 0
1801935 2 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0
1801593 2 0 0 0 0 0 0 1 1 1 0 0 1 0 0 0 0
1801565 2 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0
1801214 2 0 0 0 0 0 0 1 1 0 0 0 2 0 0 0 0
```

# CONSIDERAÇÕES FINAIS

- Escolher a melhor metodologia para os seus dados;
- Definir os parâmetros de controle de qualidade dos dados e análise;
- Definir a metodologia de associação e seleção genômica;
- Análise funcional.

**OBRIGADA PELA ATENÇÃO!!!**



**DÚVIDAS?**

[bfeitosa@hotmail.com](mailto:bfeitosa@hotmail.com)