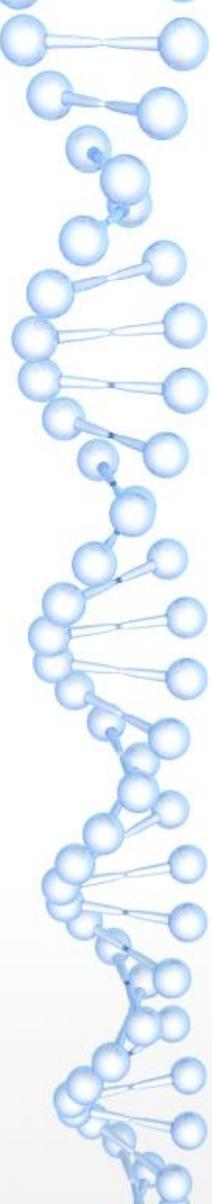


Uso da Visualização Computacional em agrupamentos genômicos

Discente: Diógenes Lodi Pinto
Docente: Profa. Rosane Minghin
Disciplina: SCC 0252/0652
09/2019



Objetivos

- Apresentar usos diferenciados de técnicas de visualização computacional na predição de ganho genético em populações de animais cruzados

Grupo de pesquisa



- Bioinformática e Melhoramento Animal
 - VNP - FMVZ - USP/Pirassununga

Seleção genômica

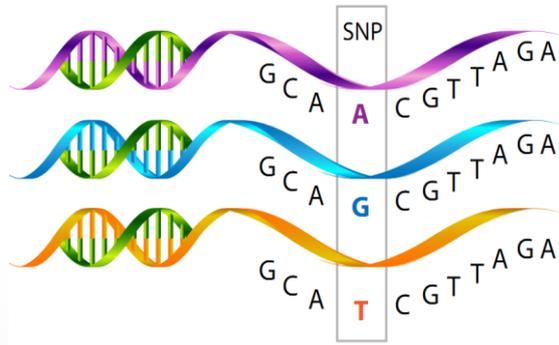
A seleção genômica consiste no uso de marcadores moleculares distribuídos densamente ao longo de todo o genoma para predizer o valor genético dos animais para características poligênicas (MEUWISSEN et al., 2001)



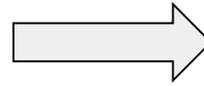
Fonte: <https://www.beefpoint.com.br/usda-publica-guia-de-genomica-animal/>

SNPs (Single Nucleotide Polymorphism)

Os marcadores SNP tem como base as alterações mais elementares da molécula de DNA, ou seja, mutações em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina).



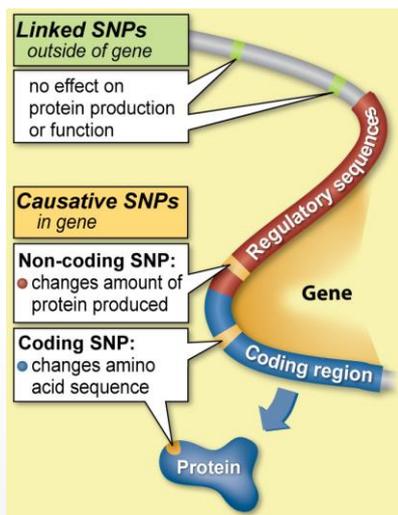
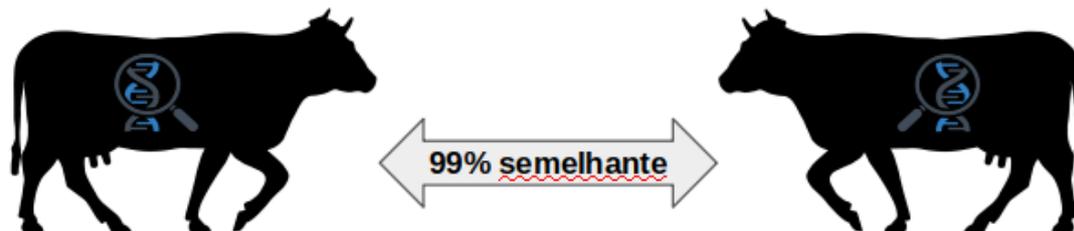
Fonte: <http://cahartmanfiction.com/understanding-genetics-snps/>



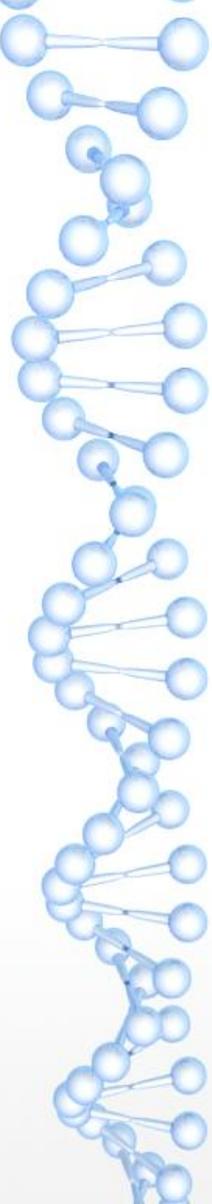
Fonte: https://www.segredosdesalao.com.br/noticia/conheca-os-8-tipos-de-cabelos-existent-no-brasil-e-descubra-qual-e-o-seu_a1973/1

SNPs (Single Nucleotide Polymorphism)

Comparando o genoma de dois animais



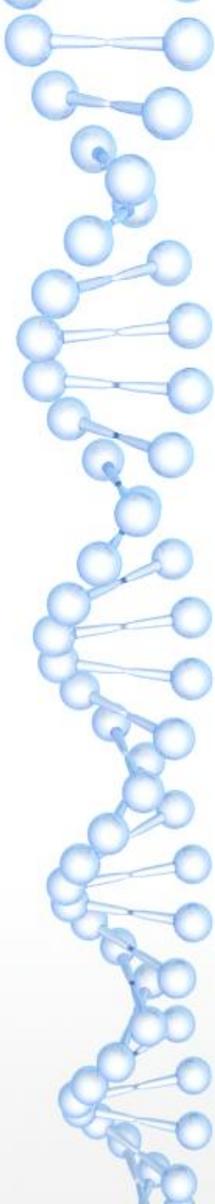
- 99% dos nucleotídeos são os mesmos em todo o gado
- 1% de diferença é baseada praticamente em SNPs.
- Aproximadamente há 50 milhões de SNPs no genoma bovino identificados



Chips de genotipagem

- Os chips de genotipagem de SNPs de alta densidade mudaram radicalmente várias questões práticas e técnicas em relação à execução de estudos de mapeamento. Atualmente, a genotipagem de uma amostra com o chip da Illumina para bovinos (Bovine SNP50®) tem um custo da ordem de US\$30 por amostra.

Chips de genotipagem



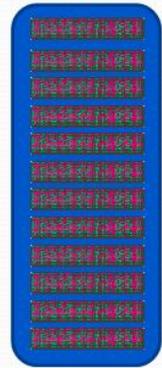
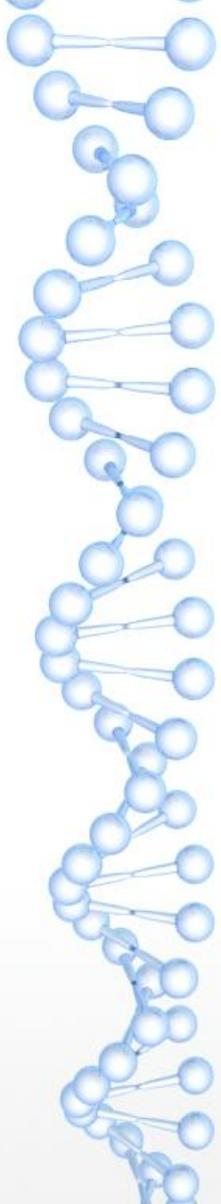
```
genotype_100.txt
~/Documents/HACKA-BIOMA

Open [icon] Save [icon]

ID      Genotypes (0 = a1,a1; 2 = a2,a2; 3 = a1,a2; 4 = a2,a1; 5 = missing; The first allele is
paternal and the second allele is maternal) ...
4201
4040433302244000033020002420424032200223200200020202030202203033424202020000222020002020244030340230
4202
4343200002242003300343032440242332203240320323302020200020223033324420232043303222020034323224300020200
4203
43202000022420030303200342404223322034402000204020202033402223033344440402323303242343002320442330430340
4204
43432000344423030003433222022233220324302002002020233040423030042440402323303242343002320244033043430
4205
0202240003302043200202422440242332202303430020402020233040423030042440402323303242343002320244030233020
4206
43432000033020232004022222042200220433043332340234340304042303334444040232333242340034023222000200020
4207
02022400033020432002024224402420323400323032320202020334044330332222020204000344432000202022330200020
4208
02342233333040404002024222022233220320303340402343403040443000022220202020000222020002020244030200320
4209
43432000022420032002024224402423322023002303232020200020223033324420232043303222020034323224300430340
4210
43432000022420033003430324402423322032430200200020202330404203033222023204333022232000202022033340200
4211
323224303330404040020442224322200220230343002040202023304042303004244040232000344432000202022333200020
4212
2020200002222000000020002220222002200220020020202000202200002222020202000222020002020222000020200
4213
0202240003302043200202422440242332202300230020402020233040420033344440202043330222320002343222000430340
```

```
diogenes@diogenes-Inspiron-5480: ~/Documents/ICMC_SCC5836
File Edit View Search Terminal Help
GAS GAS_14180 0 0 0 U G G G A A A C C 0 0 G A G G G G A G G G A A A A A G G G
GAS GAS_14183 0 0 0 U G G G G A A C C 0 0 A G G G G G A G G G A A A A A G A G
GAS GAS_14185 0 0 0 U A A G A A A C C A G G G G G G A G G A A A A A A G G G
GAS GAS_14186 0 0 0 U G A G G A A C C 0 0 G A G G G G A A G G G A A A A G G G
GAS GAS_14187 0 0 0 U G G A A A A C C 0 0 A A G G G G A A G G A A A A A G G G
GAS GAS_9048 0 0 0 U G A G A A C C 0 0 G A A G G G A G G G A A A A A G A G G
GAS GAS_9049 0 0 0 U G A G A A A C C 0 0 G A G G G G A G G G A A A A A G A G G
GAS GAS_9050 0 0 0 U A A G G A A A C 0 0 A A G G G G A A G G G A A A A A G A G G
GAS GAS_9051 0 0 0 U G G G A A A C C 0 0 A A G G G G A A G G A A A A A G G G G
C C A A G G G G G G A A G G A A A A A A G G G G
C C A A G G G G G G A A G G A A A A A A G G G G
C C G G G A G G G G A G G G A A A A A A G G G G
C C 0 0 G A A G G A A A G G A A A A A A G G G G
A C 0 0 G A A G G A A A G G A A A A A A G G G G
C C 0 0 A A G G G G A A G G G G A A A A G G G G
C C 0 0 G A G G G G A G G G G A A A A G G A G G
C C 0 0 A A G G G G A A G G A A A A A A G G G G
A C G G G G A G G A A G G A A A A A A G G A G G
C C 0 0 G A G G G G A A G G G A A A A A G G G G
C C A G G G G G A A G G G G A A A A A G A G G G
0 0 A C 0 0 G A A G 0 0 A A G G G A A A A G
```

Chips de genotipagem



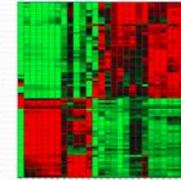
Illumina
BeadChip

+

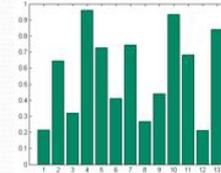


BeadXpress
reader

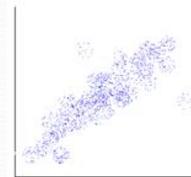
=



Heat map

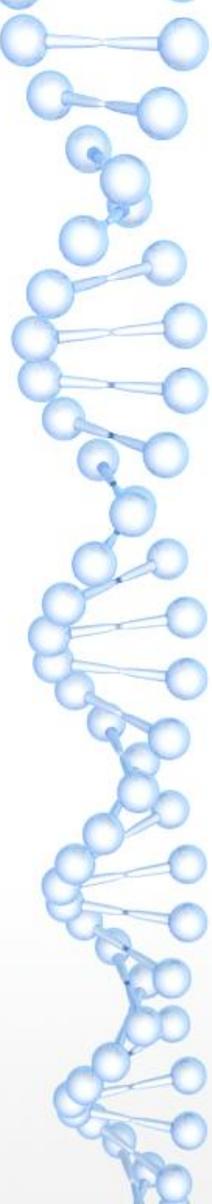


Bar plots



Scatter
plots

- https://en.wikipedia.org/wiki/Illumina_Methylation_Assay#/media/File:Dataoutput.png

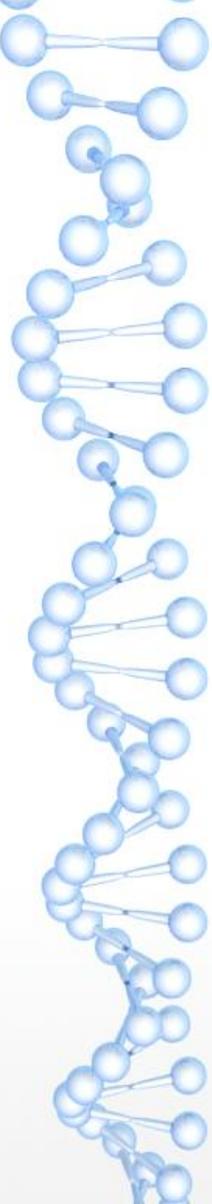


Artigo Seleccionado

Genomic clustering helps to improve prediction in a multibreed population¹

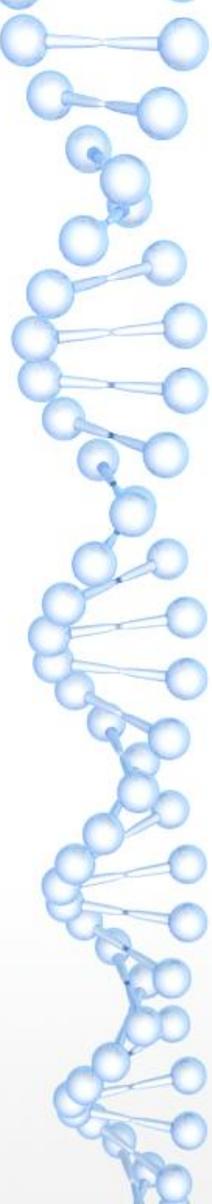
R. Ventura,^{*†2} S. Larmer,^{*} F. S. Schenkel,^{*} S. P. Miller,^{*§} and Peter Sullivan[‡]

^{*}Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada, N1G2W1; [†]Beef Improvement Opportunities, Elora, ON, Canada, N0B1S0; [‡]Canadian Dairy Network, Guelph, ON, Canada, N1K1E5; and [§]AgResearch, Invermay Agricultural Centre, Mosgiel, 9010, New Zealand



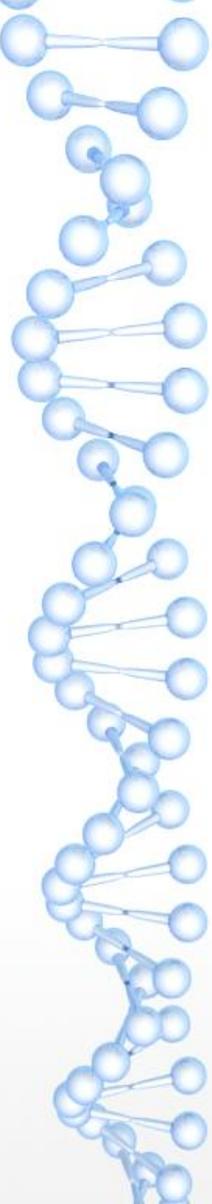
Proposta do método

- O método consiste em agrupar animais puros ou cruzados (baseado exclusivamente nos genótipos) previamente à etapa de predição genômica.
- A proposta é investigar a similaridade genética de diferentes agrupamentos (clustering) de animais utilizando vários métodos de relacionamento genômico entre indivíduos, de tal forma a se evitar grupos distintos em um mesmo processo preditivo.



Algoritmos hierárquicos aglomerativos

- O algoritmo escolhido foi o average-link que é computado como sendo a distância média entre os elementos de cada subcluster.
- A acurácia da seleção genômica foi avaliada utilizando 1.500 touros com 54.609 marcadores cada.



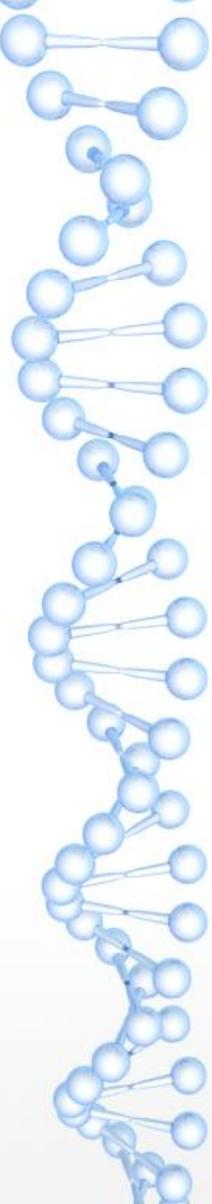
Fenótipos registrados

- Peso no nascimento
 - Ganho de peso no desmame
 - Ganho de peso pós desmame
 - Ganho de peso em um ano
-
- Foi usada a metodologia BLUP

Raças



1. **Black Angus** (<https://i.ytimg.com/vi/gZCycSuQEZI/hqdefault.jpg>), 2. **Red Angus** (<https://i.pinimg.com/originals/de/80/42/de8042288b12e0cd9a7c388981bff696.jpg>), 3. **Simmental** (<https://i.ytimg.com/vi/HS4FjoNlaws/hqdefault.jpg>), 4. **Limousin** (<http://fermesaintmichel.be/wp-content/themes/aripop/img/32-5.jpg>), 5. **Hereford** (<https://i.pinimg.com/originals/96/8b/41/968b4114078713b73dbdd1b2dd4c8648.jpg>), 6. **Charolais** (<https://i.ytimg.com/vi/7DqTYMw8H6w/hqdefault.jpg>),



Software Admixture

- Composição racial (via SNPs) de cada animal foi obtida para fins de entendimento biológico de cada relação animal-cluster

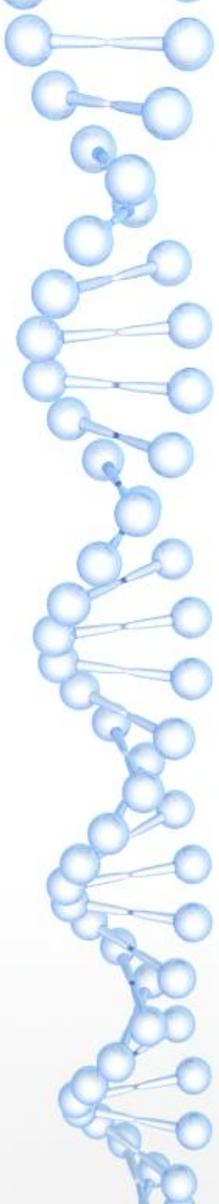
Table 1. Average breed composition as estimated by ancestral components in Admixture software (Alexander et al., 2009)

| Feature | Breed ¹ | | | | | | |
|----------------------------------|--------------------|-------|-------|-------|-------|-------|-------|
| | BLK | RED | SIM | LIM | HER | CHA | OTH |
| Number of purebreds ² | 160 | 23 | 191 | 60 | 23 | 107 | 4 |
| Average breed composition | 0.320 | 0.196 | 0.244 | 0.065 | 0.040 | 0.047 | 0.087 |

¹Breeds were determined using pedigree information on available purebred animals. BLK = Black Angus; RED = Red Angus; SIM = Simmental; LIM = Limousin; HER = Hereford; CHA = Charolais; OTH = ancestry component not identified to contain any purebred animals.

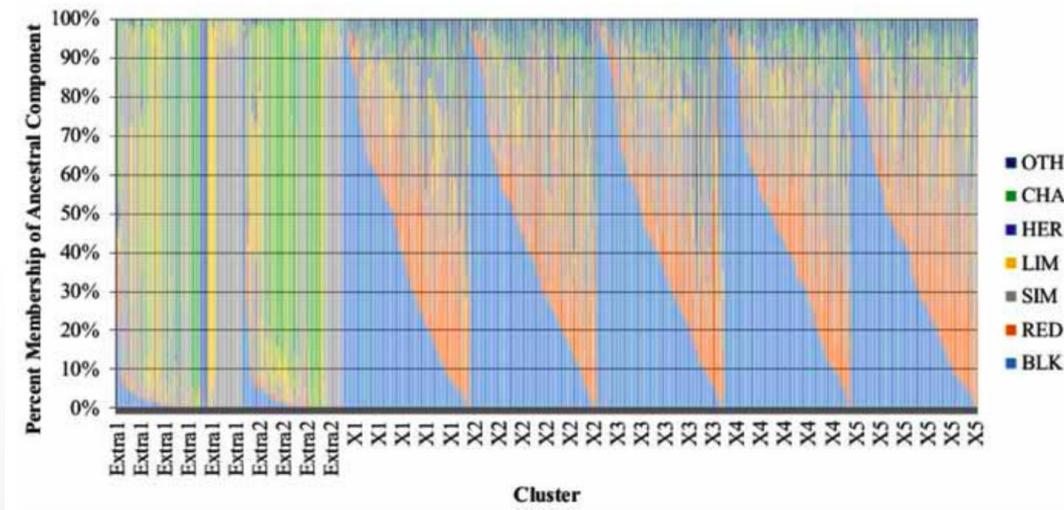
²Purebreds were defined as having greater than 75% of ancestry component in Admixture.

Heat map plot

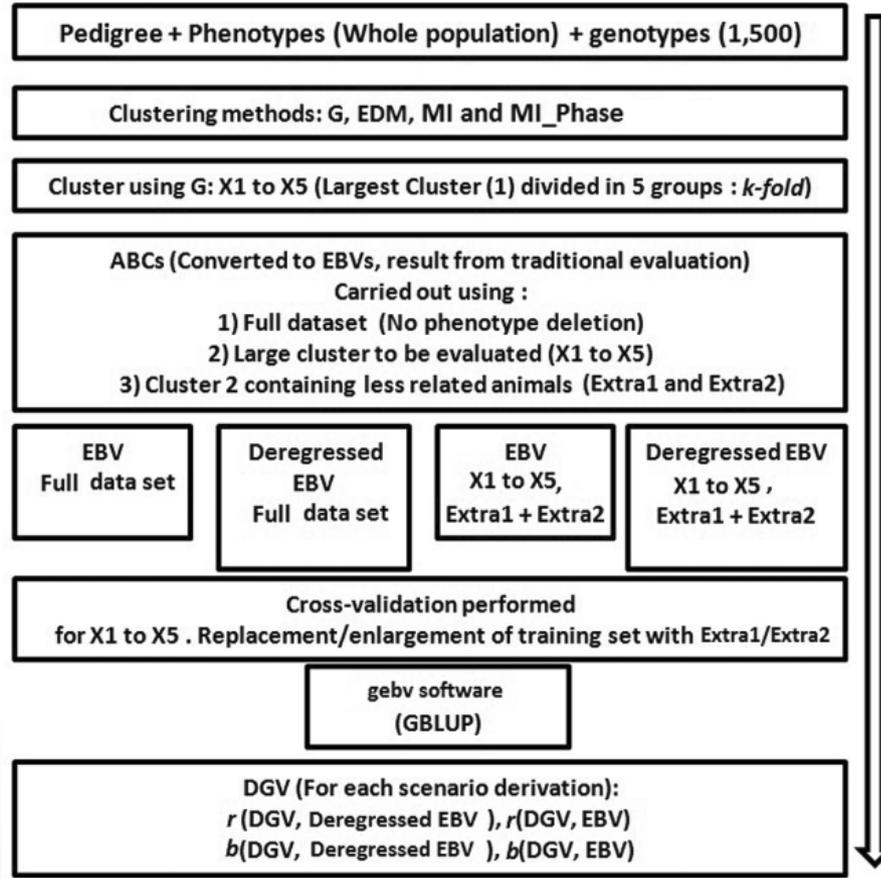


Composição de raças

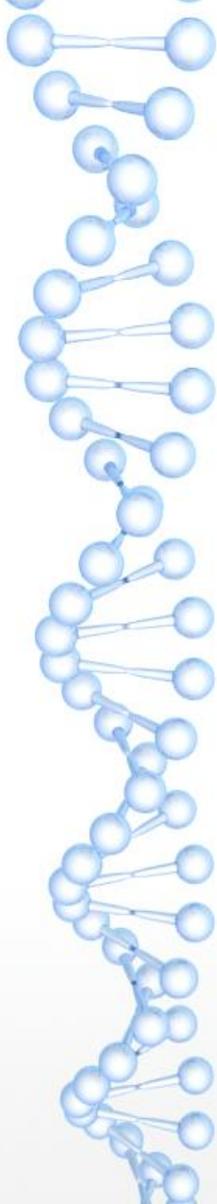
- Por um método supervisionado, foram identificados 7 grupos genéticos (CHA, HER, etc), representados pelas diferentes cores. Observa-se a ocorrência de animais puros (única cor na linha vertical), assim como cruzados



Fluxograma



Método k-fold



adding 2 subsets
of Cluster 2 in the
k-fold validation
process

k-fold

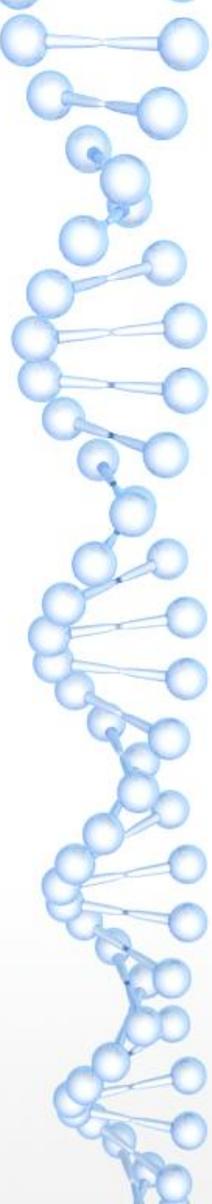
| | CLUSTER 1 | | | | | CLUSTER 2 | |
|---|-----------|----|----|----|----|-----------|--------|
| | X1 | X2 | X3 | X4 | X5 | Extra1 | Extra2 |
| Validation of group X5 as part of the <i>k-fold</i> strategy | T | T | T | T | V | | |
| | | T | T | T | V | T | |
| | | | T | T | V | T | T |
| | T | | T | T | V | T | T |
| | T | | | T | V | T | T |
| | T | T | | T | V | T | T |
| Validation of group X4 as part of the <i>k-fold</i> strategy | T | T | T | V | T | | |
| | | T | T | V | T | T | T |
| | T | | T | V | T | T | T |
| | T | | | V | T | T | T |
| | T | T | | V | T | T | T |
| | T | T | T | V | | T | T |



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Validation of group X1 as part of the <i>k-fold</i> strategy | V | T | T | T | T | | |
| | V | | T | T | T | T | T |
| | V | | | T | T | T | T |
| | V | T | | T | T | T | T |
| | V | T | | | T | T | T |
| | V | T | T | | T | T | T |
| | V | T | T | T | | T | T |
| | V | T | T | T | | T | T |

Table 8. Regression coefficient (b) of genomic selection averaged across validation groups from k -fold validation when sets from the unrelated cluster (Extra2) were added in the validation process without replacing a set from the training population

| Trait used in gebv ¹ software | Validation group (compared with dEBV ²) | | | | | | Validation group (compared with EBV) | | | | | |
|---|---|---|-----------|--|-----------|---------------------------|--|--|-----------|---|-----------|---------------------------|
| | Within cluster 1 (k -fold) ³ | Adding Extra1 without replacing X1 to X5 ⁴ (#) | SE (#) | Adding Extra1 and Extra2 without replacing X1 to X5 ⁵ (#) | SE (#) | Traditional evaluation | Within cluster 1 (k -fold) ³ | Adding Extra1 without re- placing X1 to X5 (#) | SE (#) | Adding Extra1 and Extra2 without replacing X1 to X5 (#) | SE (#) | Traditional evaluation |
| Birth weight EBV | | | | N/A ⁶ | | | 1.240 | 1.320 | 0.008 | 1.178 | 0.009 | 1.02 |
| Birth weight dEBV | 1.148 | 1.150 | 0.009 | 1.141 | 0.010 | 1.05 | 1.023 | 1.016 | 0.006 | 1.001 | 0.007 | N/A |
| Weaning gain EBV | | | | N/A | | | 1.321 | 1.368 | 0.007 | 1.322 | 0.007 | 0.93 |
| Weaning gain dEBV | 1.086 | 1.092 | 0.027 | 1.113 | 0.020 | 0.91 | 0.865 | 0.921 | 0.009 | 0.927 | 0.011 | N/A |
| Postweaning gain EBV | | | | N/A | | | 1.655 | 1.711 | 0.012 | 1.733 | 0.011 | 1.05 |
| Postweaning gain dEBV | 1.849 | 1.881 | 0.045 | 1.749 | 0.059 | 1.11 | 1.514 | 1.489 | 0.011 | 1.421 | 0.014 | N/A |
| Yearling gain EBV | | | | N/A | | | 1.518 | 1.599 | 0.013 | 1.598 | 0.016 | 1.02 |
| Yearling gain dEBV | 1.650 | 1.820 | 0.036 | 1.690 | 0.016 | 1.12 | 1.254 | 1.303 | 0.012 | 1.309 | 0.015 | N/A |



Conclusões

- Foi demonstrado uma diminuição na acurácia da seleção genômica depois de incluir animais de cluster desconexos.
- a Predição genômica auxiliada por clusters (GPC) viabiliza a escolha da população de referência mais adequada para fins de melhor precisão durante o processo de predição genômica de animais mestiços e puros.