

## Seleção genômica no melhoramento animal

Dr. Fernando Baldi  
UNESP-FCAV  
Dr. Marcos Vinicius Barbosa da Silva  
EMBRAPA GADO DE LEITE

1

### Avaliação genômica em um único passo: *Single step* (Misztal et al., 2009)

Avaliações genômicas são realizadas atualmente em um processo de várias etapas. Uma avaliação típica requer:

1. Avaliação tradicional com um modelo animal
2. Obtenção das pseudo-observações, tais como DEP desregredidas (Y)
3. Estimativa de efeitos genômicos para animais genotipados ( $Y=b_1x_1+ b_2x_2+ \dots + b_nx_n$ )
4. Combinação das informações genômicas com avaliação tradicional

2

## Avaliação genômica em um único passo (Misztal et al., 2009)

Esses passos são dependentes de vários parâmetros e pressupostos:

1. Para estimar os efeitos genômicos temos várias opções  
BayesA, BayesB ..... Kernel Regression
2. Efeitos dos marcadores SNP podem ser estimados com pressupostos diferentes quanto à distribuição prévia dos efeitos  
-Difícil de comprovar alguns supostos
3. Regressões estimados são tendenciosos por genotipagem seletiva
4. Pseudo-observações são mal definidas ou de baixa qualidade (por exemplo, para animais com número reduzido de progênie)

3



J. Dairy Sci. 93:743–752  
doi:10.3168/jds.2009-2730  
© American Dairy Science Association®, 2010.

### **Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score<sup>1</sup>**

I. Aguilar,\*†‡ I. Misztal,\* D. L. Johnson,‡ A. Legarra,§ S. Tsuruta,\* and T. J. Lawlor#

\*Animal and Dairy Science Department, University of Georgia, Athens 30602  
†Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay  
‡Livestock Improvement Corp., Private Bag 3016, Hamilton 3240, New Zealand  
§INRA, UR631 SAGA, BP 52627, 32326 Castanet-Tolosan, France  
#Holstein Association USA Inc., Brattleboro, VT 05302-0808

#### ABSTRACT

The first national single-step, full-information (phenotype, pedigree, and marker genotype) genetic evaluation was developed for final score of US Holsteins. Data included final scores recorded from 1955 to 2009 for 6,232,548 Holsteins cows. BovineSNP50 (Illumina, San Diego, CA) genotypes from the Cooperative Dairy DNA Repository (Beltsville, MD) were available for 6,508 bulls. Three analyses used a repeatability animal model as currently used for the national US evaluation. The first 2 analyses used final scores recorded up to 2004. The first analysis used only a pedigree-based relationship matrix. The second analysis used a rela-

be increased up to 0.92 by scaling differences between the genomic and pedigree-based relationship matrices with little loss in accuracy of prediction. One complete evaluation took about 2 h of computing time and 2.7 gigabytes of memory. Computing times for single-step analyses were slightly longer (2%) than for pedigree-based analysis. A national single-step genetic evaluation with the pedigree relationship matrix augmented with genomic information provided genomic predictions with accuracy and bias comparable to multiple-step procedures and could account for any population or data structure. Advantages of single-step evaluations should increase in the future when animals are pre-selected on genotypes.

4

## Avaliação genômica em um único passo (Misztal et al., 2009)

- Única etapa de avaliação em que a matriz de parentesco (pedigree) é aumentada pela contribuição da matriz genômica de parentesco (Misztal et al., 2009)
  - Muitos animais não têm genótipos, incluí-los na matriz de parentesco genômica
- Procedimento de um passo fornece uma estrutura unificada, e elimina várias hipóteses e pressuposições sobre os parâmetros

5

## Matriz de parentesco com base no pedigree + informações genômicas

A inversa da matriz de parentesco  $H^{-1}$  é baseada no pedigree e na informação genômica (Aguilar et al. 2010)

Inversa da matriz de parentesco (pedigree)

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix} \longrightarrow H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

os animais são divididos em dois grupos, grupo 2 animais genotipados

Matriz de parentesco genômica

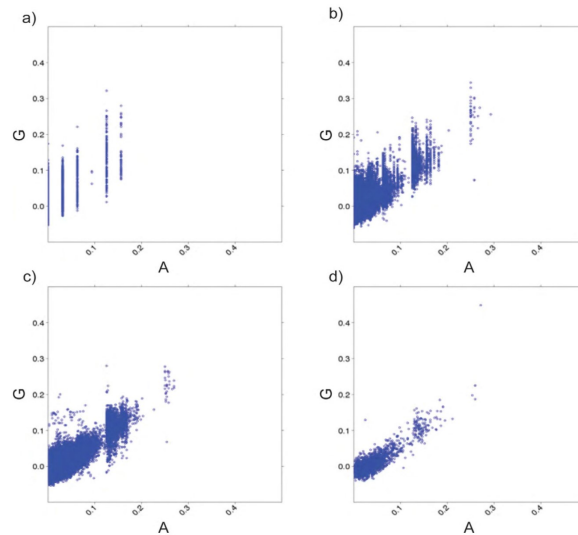
Matriz de parentesco ( $A^{-1}$ ) de animais genotipados

<sup>6</sup>  
Aguilar. Journal of Dairy Science Vol. 93 No. 2, 2010

## Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech<sup>RepGen</sup> - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro  
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

### Profundidade do pedigree (A)



Relação entre o parentesco tradicional (A) com o parentesco genômico utilizando 2(a), 4(b), 6(c) e 8 (d) gerações.

Fonte: Pryce et al., 2012.

7

### Propriedades da H

- Sempre semi-positiva definida (todos os seus valores próprios forem não negativos)
- Positiva definida & invertível, se G é invertível
- Na prática, se G é muito diferente (pedigree ou genótipos ruins) da matriz  $A_{22}$ , teremos problemas numéricos
- Se todos os animais são genotipados, ssBLUP é GBLUP

8

## **Workshop de Melhoramento Genético Animal**

Projeto ALT-Biotech<sup>RepGen</sup> - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro  
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

### **Introdução “curta” à família de programas BLUPf90**

#### **Programas BLUPf90**

- Desenvolvidos por Ignacy Misztal e colaboradores na Universidade de Georgia
- Conjunto de softwares aplicados ao melhoramento animal (modelos mistos)

## Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech<sup>RepGen</sup> - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro  
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

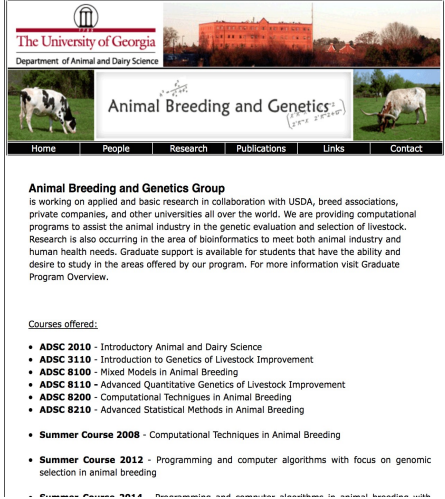
### Programas BLUPf90

- Conjunto de programas:
  - Utilizam a teoria dos Índices de seleção e modelos mistos
  - Programação simples utilizando Fortran90
  - Possuem uma programação geral que suporta diferentes modelos:
    - Modelos lineares e não lineares, com efeitos múltiplos correlacionados, modelo animal multicaracterística e modelos de dominância
- A filosofia geral dos programas está descrita em:
  - “Complex Model, More Data: Simple Programming?”. Misztal, I. 1999 InterBull.

### Programas BLUPf90

- Consiste de vários programas:
- Estimativa de componentes de variância
  - REMLF90, AIREMLF90, Thrgibbs1f90
  - Solver para equações dos modelos mistos
    - BLUPf90
    - BLUPf90IOD
  - Aproximação da acurácia
    - ACCF90 (grande banco de dados)
- Seleção genômica

**Workshop de Melhoramento Genético Animal**  
Projeto ALT-Biotech<sup>RepGen</sup> - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro  
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019



<http://nce.ads.uga.edu/wiki/doku.php>

## Programas BLUPf90

- Todos os programas são controlados pelo mesmo arquivo de parâmetros
- Opções extra podem ser usadas para funções específicas do programa (non-default)
- O entendimento do arquivo de parâmetros ajuda a resolver muitos problemas

## Arquivo de parâmetros do BLUPf90

```
# Each keyword can be preceded by comments, each starting with
#
DATAFILE
name of data file
NUMBER OF TRAITS
number of traits
NUMBER OF EFFECTS
number of effects
OBSERVATION(S)
position of observations in data file (one per trait)
WEIGHT(S)
position of weight(s) in data file (one per trait); blank if
all weights equal
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT
[EFFECT NESTED]
one line per effect:
  position of effect (one per trait)
  number of levels
  type of effect (one of cross for crossclassified or cov
for covariable)
  position of effect where nested (one per trait, optional)

RANDOM RESIDUAL VALUES
residual variance covariance matrix (full stored)
RANDOM_GROUP
number of one random effect or list of correlated effects
(order as in EFFECTS above); correlated effects must be
consecutive
RANDOM TYPE
type of random effect: one of diagonal, add animal, add sire,
add an upg, add an upginb, user_file,
user_file i or par_domin

FILE
relationship file; blank line if file unnecessary
(CO)VARIANCES
variance covariance matrix for given correlated effects and
traits (full stored)
```

Repetido para cada  
Efeito aleatório

## Arquivo de dados

- “free format”, ex: ao menos um espaço para separar as colunas
- TABs não são válidos para separar as colunas
- Apenas números: inteiros ou reais
- Separadores de decimais “.” não “,”
- “.” não é missing value
- Todos os efeitos devem ser renumerados a partir de 1 (RENUMF90)



## Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech<sup>RepGen</sup> - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro  
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

### PreGSF90

- PreGSF90 é um programa para processar informações genômicas para a família de programas BLUPF90
- Programa PreGSF90 implementa a seleção genômica seguindo a metodologia do único passo, tal como apresentado por Aguilar et al. JDS 2010
- Matriz A é substituída por uma matriz H, que combina as informações pedigree e genômico.
- Programa PreGSF90 poderia ser executado após RENUMf90.
- Também é executado automaticamente por programas aplicativos como BLUPF90, REMLF90, GIBBSxF90 ou BLUP90IOD quando seu arquivo de parâmetro contém OPTION SNP\_file filename.

17

### Programa PreGSf90

- Utiliza o modulo genômico
- Criação de matrizes genômica e relações de parentesco baseadas no pedigree
- Diferentes metodos para otimizar os calculos e utilizar processamento paralelo

18

## Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech<sup>RepGen</sup> - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro  
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

### Arquivos de entrada

- Os mesmos arquivos de parametros que os programas BLUPf90
  - OPTION SNP\_file XXXX
  - Indicar os comandos genomicos
- Arquivo de pedigree
- Informação dos marcadores (Arquivo de SNP)
- Arquivo de referência dos IDs
  - Ligação entre ID renumerados e originais

19

### Arquivo mapa (SNP map)

- Necessário para algumas análises de consistência
- Formato
  - Número do SNP
    - Índice com o número do SNP sorteado no mapa
  - Número do cromossomo
  - Posição
- Primeira linha corresponde à primeira coluna do SNP no arquivo de genótipos!!

20

## Opções para o arquivo de parâmetros BLUPf90

- PreGSf90
  - Adicionando opções (OPTIONS) ao arquivo de parâmetros
  - OPTION SNP\_file *marker.geno.clean*
  - Utiliza dois arquivos:
    - marker.geno.clean
    - Marker.geno.clean.XrefID

21