

Seleção genômica no melhoramento animal

Dr. Fernando Baldi
UNESP-FCAV

1

Seleção Genômica

- ✓ A seleção genômica é realizada com base nos valores genômicos (GEBV) preditos a partir de marcadores de alta densidade espalhados em todo o genoma (Meuwissen et al., 2001).
- ✓ Continuamos dentro da genética quantitativa
 - Herança poligênica
 - Estudo a nível de populações
 - Variações contínuas e efeitos do meio

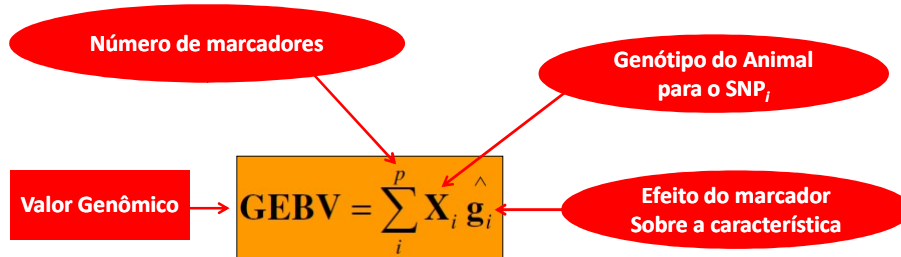
2

Princípio da Seleção genômica

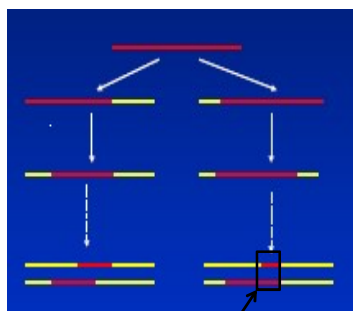


Seleção genômica

Predizer os valores genéticos genômicos (GEBV) como soma de efeitos de todos os marcadores



Seleção genômica



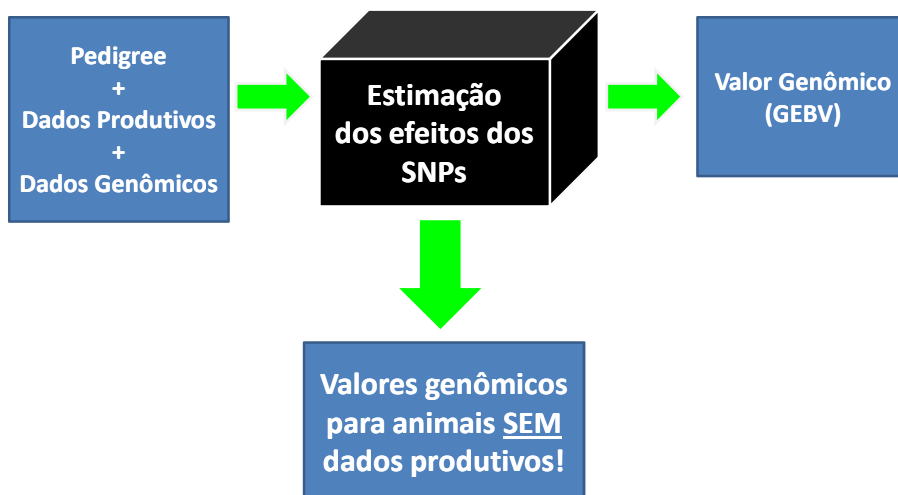
Segmentos de cromossomos, que provem de um antepassado em comum sem intervenção da recombinação (carregam alelos ou haplótipos idênticos).

No caso de existir um QTL dentro do segmento de cromossomo, dois indivíduos parentes vão compartilhar alelos do QTL idênticos

1 2 Q 1 1

5

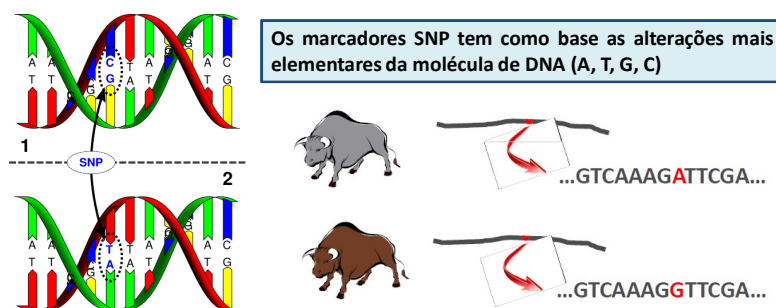
Por que é diferente?



6

Informação Genômica

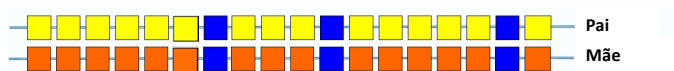
O Sequenciamento do genoma tem permitido a descoberta de milhares de polimorfismos de base única (SNPs):



RARAMENTE UM ÚNICO SNP É RESPONSÁVEL POR VARIAÇÃO EM CARACTERÍSTICAS DE INTERESSE ECONÔMICO.

Os “genes” determinam as DEPs

Parte de 1 par de cromossomos



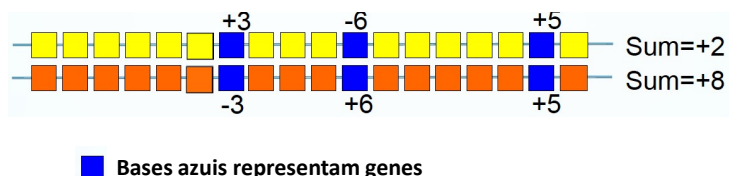
- Bases azuis representam genes
- Bases amarelas são herdadas do pai
- Bases vermelhas são herdadas da mãe

Bovinos têm 30 pares de cromossomos

Cada membro do par é herdado do pai e outro da mãe

Cada cromossomo possui próximo de 3.000 milhões de pares de bases (A, G, T, C)

DEP é soma da metade dos efeitos dos marcadores



A DEP é metade da somatória de todos os efeitos dos marcadores
 (metade porque a progênie recebe metade dos cromossomos de cada pai)

Seleção genômica

$$MVP = \sum_i^p \text{SNP}_i * b_i$$



MVP ou DGV: Valor Molecular Predito

Conhecendo o **efeito dos marcadores** é possível obter o **valor molecular genômico** inclusive antes do indivíduo chegar à idade reprodutiva ou expressar características de interesse.

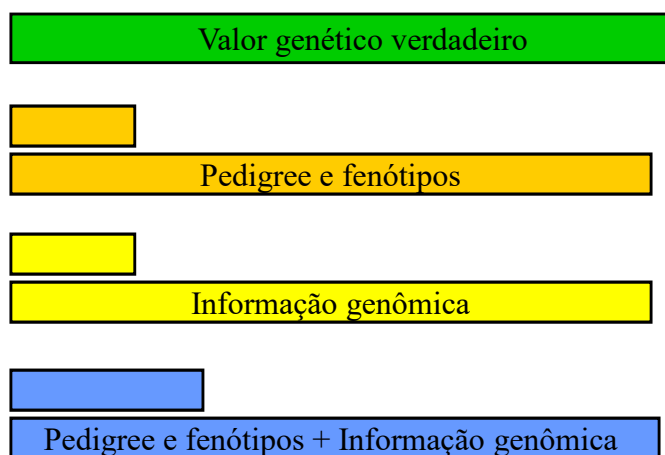
Combinação do MVP com informação do pedigree:

$$GEBV = w_1 MVP + w_2 DEP_{\text{Pais}}$$

GEBV: Valor genômico estimado ou DEP genômica

10

Múltiplas fontes de informação



Uma moeda comum

- Uma ÚNICA estimativa de valor de genético com base em todas as informações disponíveis:
 - Inf. genômica
 - Pedigree
 - Fenótipos
- ÚNICA medida de acurácia

Maior acurácia no início da vida

Cenário 1

		Pedigree + Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	
	SIM		

Cenário 2

		Pedigree + Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	
	SIM		

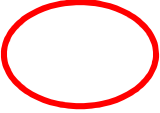
Uma moeda comum



Uma moeda comum

		Pedigree + Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	DEP
	SIM		

Cenário 3

		Pedigree + Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	DEP
	SIM		

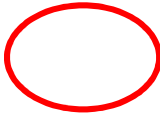
Uma moeda comum



Uma moeda comum

		Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	DEP
	SIM	MVP	

Cenário 4

		Pedigree + Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	DEP
	SIM	MVP	

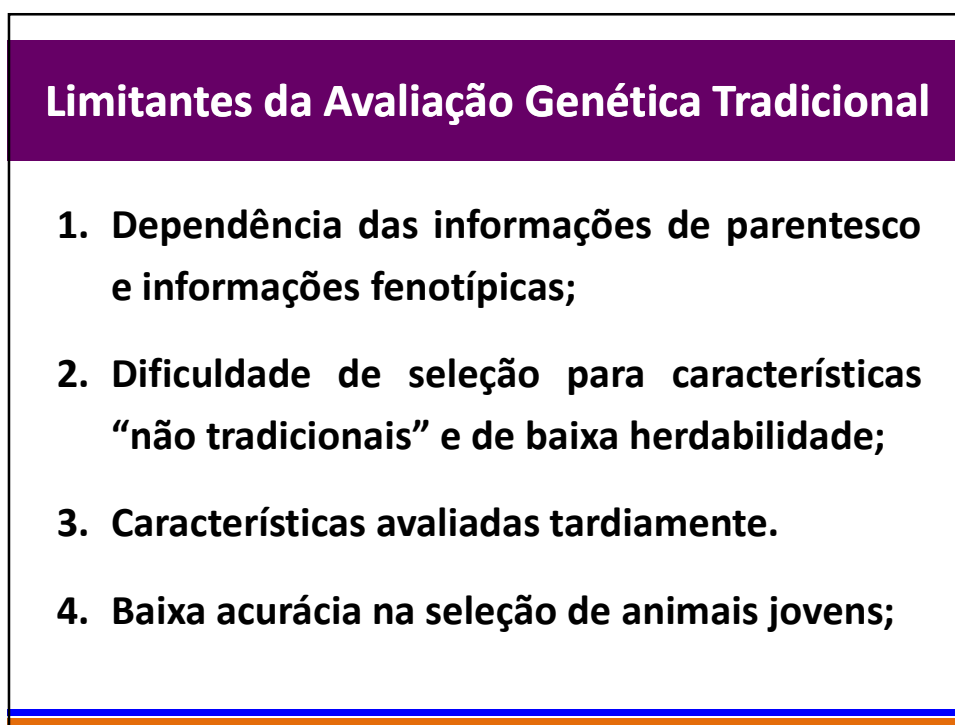
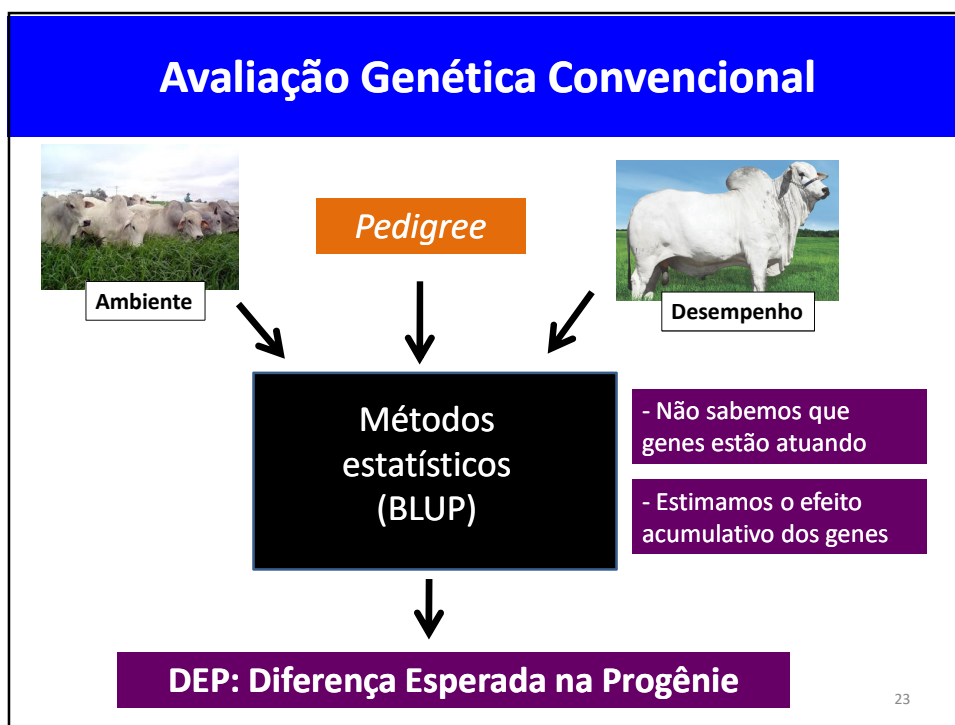
Uma moeda comum



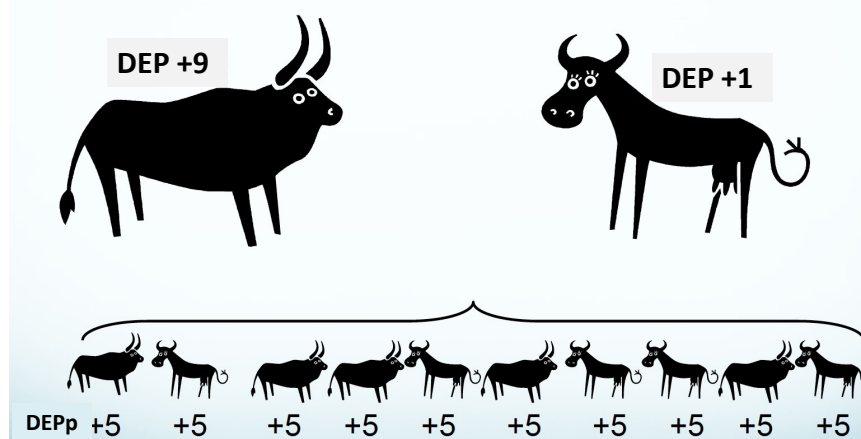
Uma moeda comum para seleção

		Pedigree + Fenótipo	
		NÃO	SIM
Informação genômica	NÃO	----	DEP
	SIM	MVP	DEP genômica

Mesmas unidades. Diferentes medidas de acurácia.

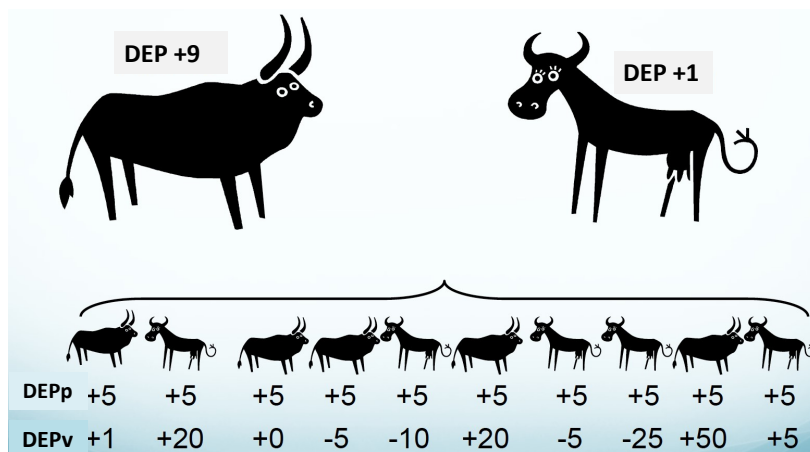


Animais Jovens: DEP é a média dos pais



DEPp = DEP de pedigree

Animais Jovens: com mais informações



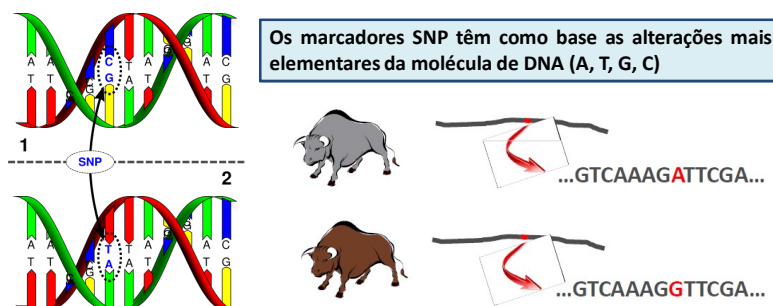
Precisamos dos fenótipos para identificar os animais que têm uma DEP maior à média dos pais (segregação Mendeliana)

Limitações da Avaliação Genética Convencional

- As DEPs são preditas a partir das informações do **fenótipo** e **pedigree** (na era pre-genômica)
- Com **suficiente informação** é possível chegar próximo da DEP "verdadeira" para cada característica.
- Na era da genômica temos os genótipos!!

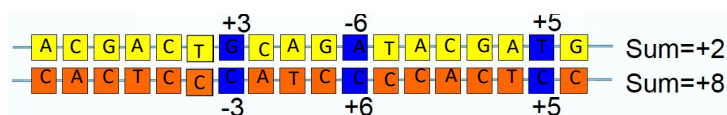
Informação Genômica

O sequenciamento do genoma tem permitido a descoberta de milhares de polimorfismos de base única (SNPs):



Difícilmente um único SNP é responsável pela variação em características de interesse econômico

DEP é soma da metade dos efeitos dos genes



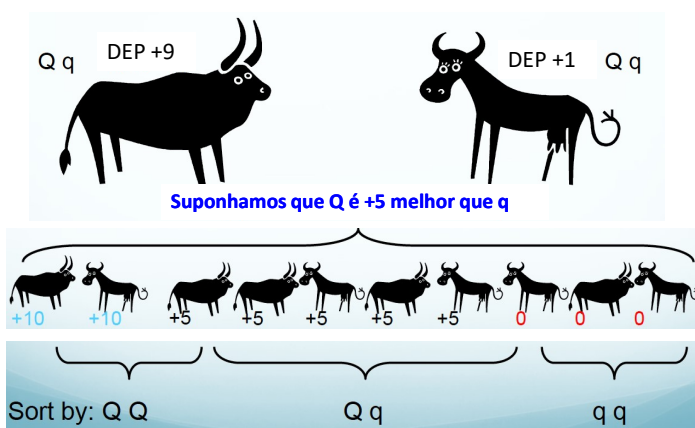
■ Bases azuis representam genes

MVP ou DGV: Valor Molecular Predit

$$MVP = \sum_i^p SNP_i * b_i$$

Conhecendo o efeito dos marcadores, é possível obter o **valor molecular genômico**, inclusive antes do indivíduo chegar à idade reprodutiva ou expressar características de interesse.

DEP da progênie jovem: média dos pais + **genômica**



Melhoramento Animal para uma Pecuária Eficiente

Predição do Ganho Genético (ΔG) **com genômica:**

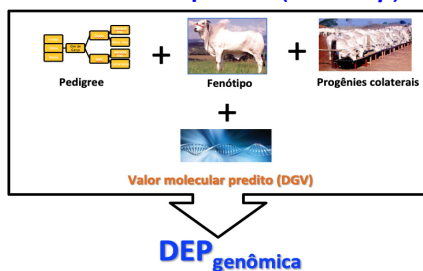
$$\Delta G_a = \frac{\left(\text{Acurácia de predição} \right) * \left(\text{Intensidade de seleção} \right) * \left(\text{Desvio padrão genético aditivo} \right)}{\text{Intervalo de gerações}}$$

31

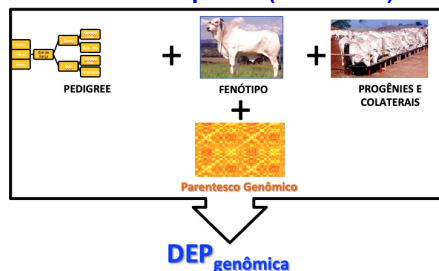
Abordagens em Avaliações Genômicas

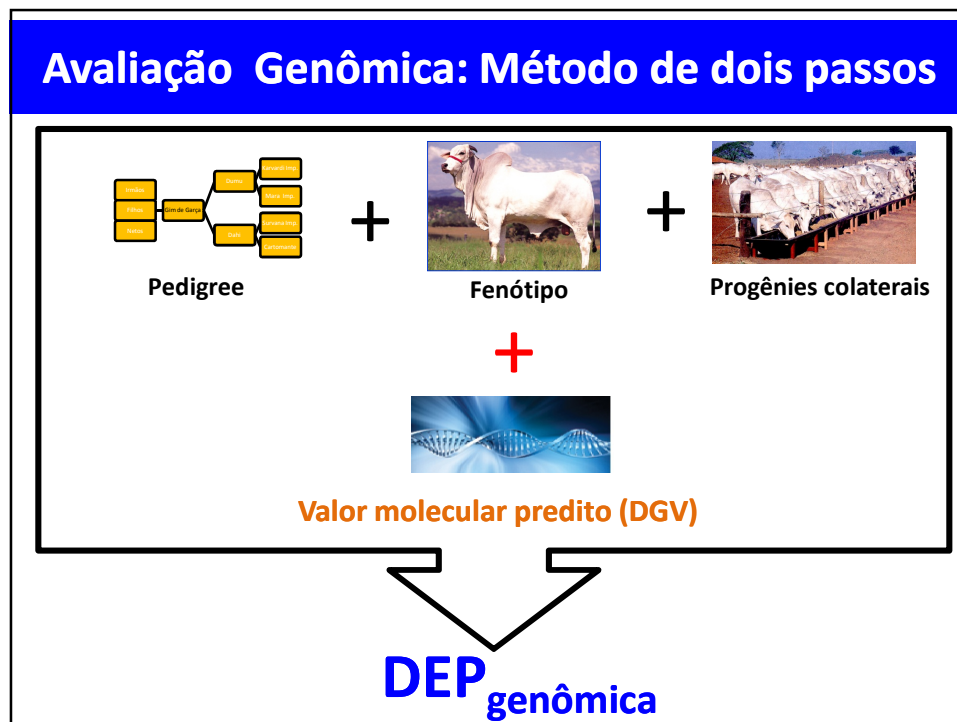
Inclusão da informação genômica:

Método de dois passos (Two-step):

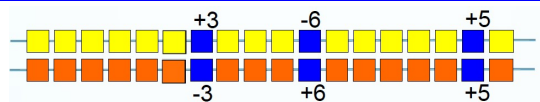


Método de um passo (ssGBLUP):





Seleção genômica: cálculo do DGV



$$DGV = \sum_i^p SNP_i * b_i$$



MVP ou DGV: Valor Molecular Predito

Conhecendo o efeito dos marcadores é possível obter o **valor molecular genômico**, antes do indivíduo chegar à idade reprodutiva ou expressar características de interesse.

A1

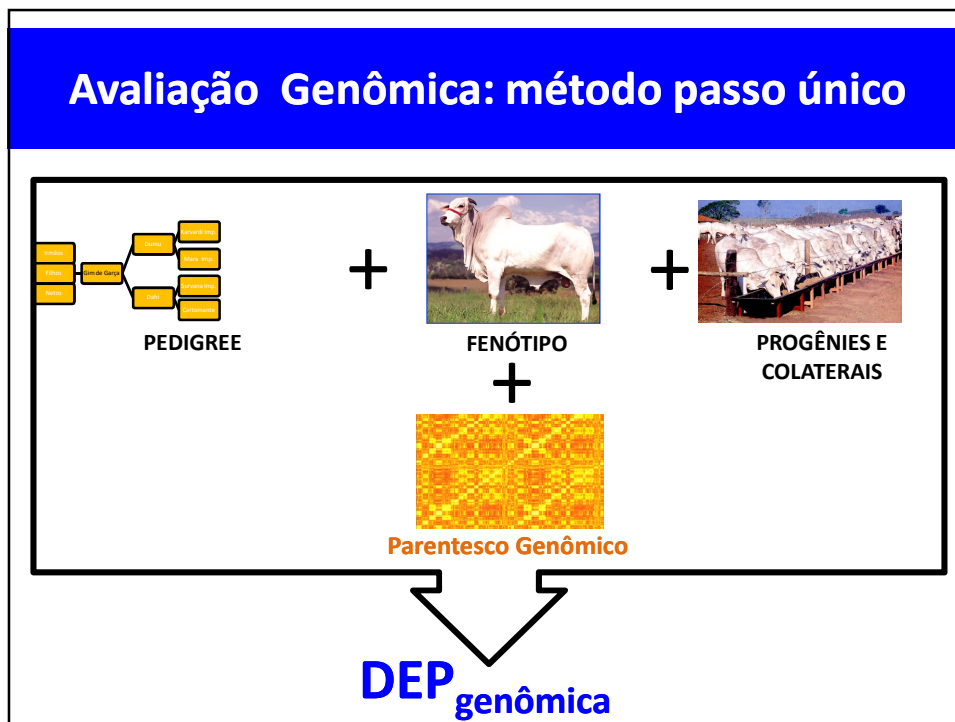
Combinação do DGV com informação do pedigree:

$$GEBV = w_1 DGV + w_2 DEP_{Pais}$$

GEBV: Valor genômico estimado ou DEP genômica

Diapositivo 34

A1 ISTO ESTE CONCEITO JÁ ESTÁ INSERIDO EM OUTRO SLIDE (13)
User; 25/10/2018



ssGBLUP: pedigree + informações genômicas

- A inversa da matriz de parentesco H^{-1} é baseada no pedigree e na informação genômica:

Inversa da matriz de parentesco (pedigree)

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

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

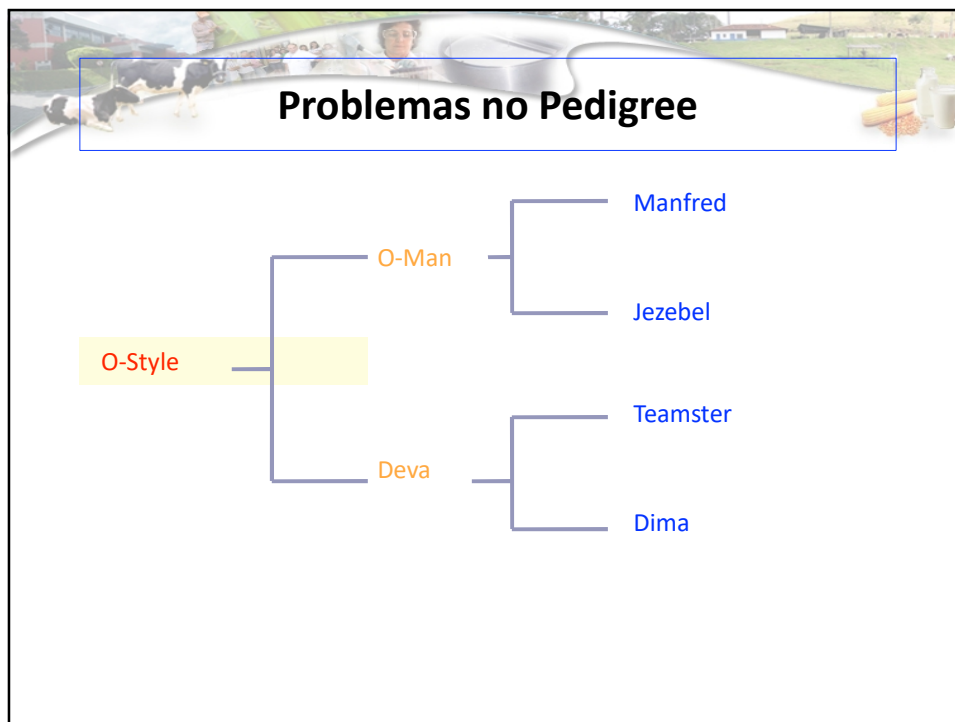
→

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

Matriz de parentesco genômica

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

Fonte: Aguilar. Journal of Dairy Science Vol. 93 No. 2, 2010



Matriz de parentesco esperada (avôs e avós não aparentados)

	AvôP	AvóP	AvôM	AvóM	Pai	Mãe	Touro
Manfred	1.0	.0	.0	.0	.5	.0	.25
Jezebel	.0	1.0	.0	.0	.5	.0	.25
Teamster	.0	.0	1.0	.0	.0	.5	.25
Dima	.0	.0	.0	1.0	.0	.5	.25
O-Man	.5	.5	.0	.0	1.0	.0	.5
Deva	.0	.0	.5	.5	.0	1.0	.5
O-Style	.25	.25	.25	.25	.5	.5	1.0

Workshop de Melhoramento Genético Animal

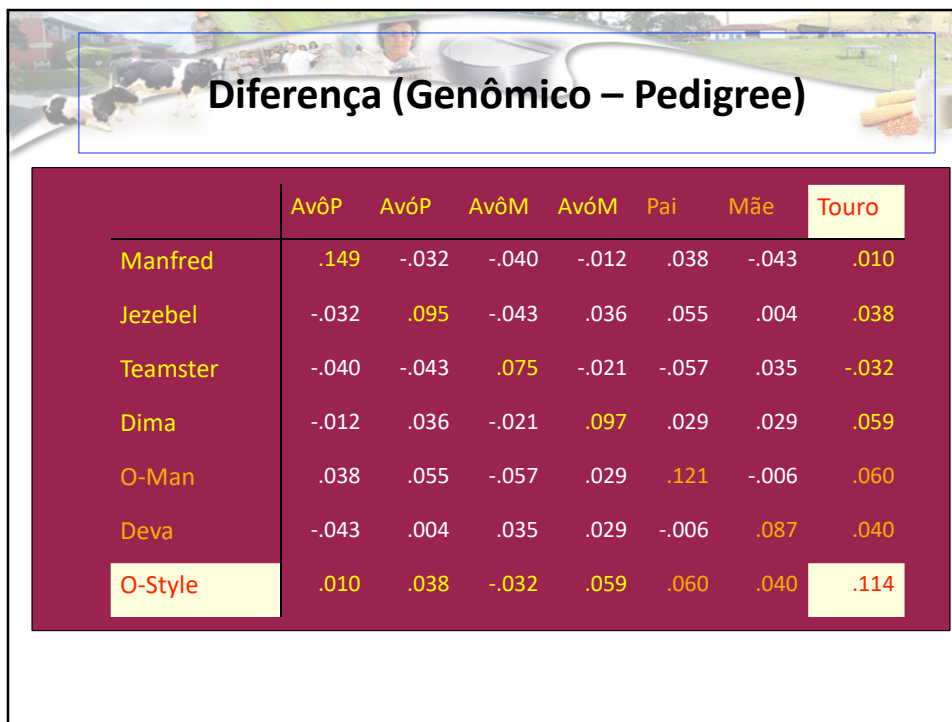
Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Matriz de parentesco esperada (considerando parentesco entre avôs e avós)

	AvôP	AvóP	AvôM	AvóM	Pai	Mãe	Touro
Manfred	1.053	.090	.090	.105	.571	.098	.334
Jezebel	.090	1.037	.051	.099	.563	.075	.319
Teamster	.090	.051	1.035	.120	.071	.578	.324
Dima	.105	.099	.120	1.042	.102	.581	.342
O-Man	.571	.563	.071	.102	1.045	.086	.566
Deva	.098	.075	.578	.581	.086	1.060	.573
O-Style	.334	.319	.324	.342	.566	.573	1.043

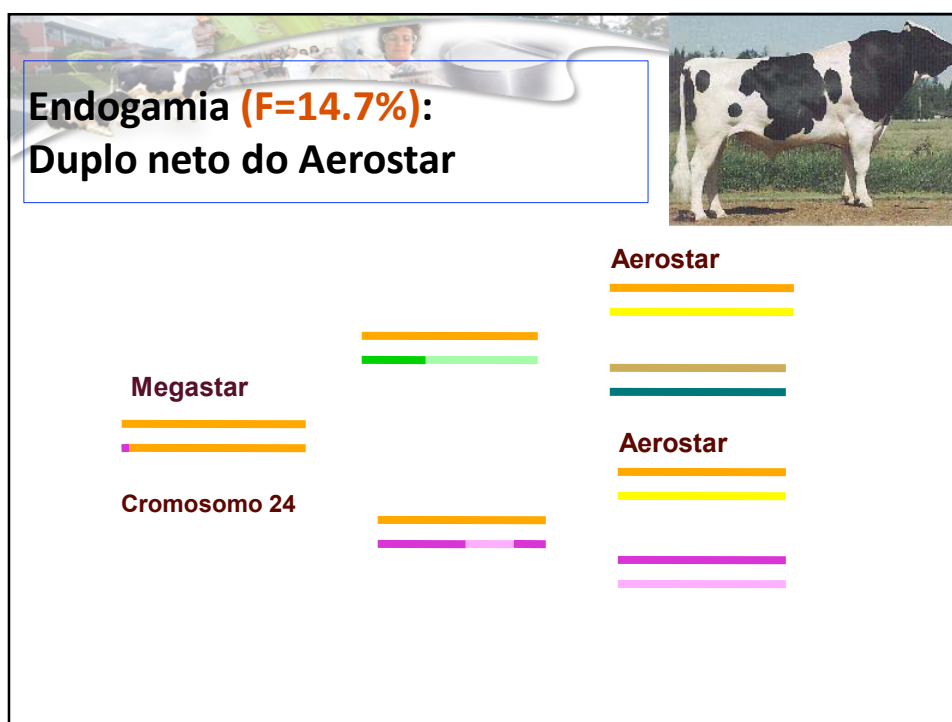
Matriz de parentesco genômico

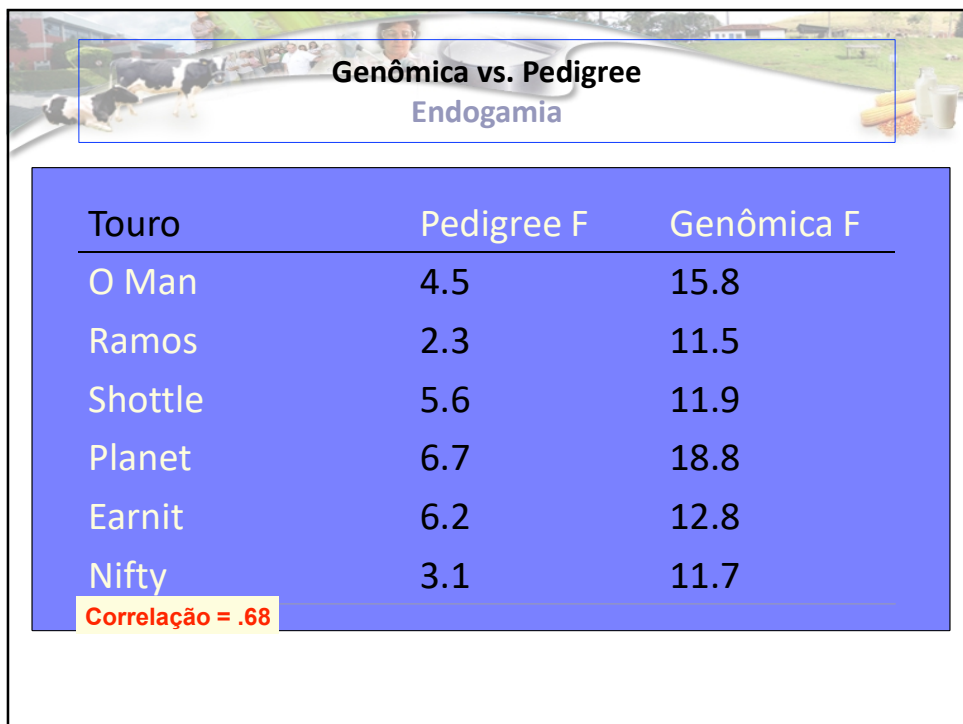
	AvôP	AvóP	AvôM	AvóM	Pai	Mãe	Touro
Manfred	1.201	.058	.050	.093	.609	.054	.344
Jezebel	.058	1.131	.008	.135	.618	.079	.357
Teamster	.050	.008	1.110	.100	.014	.613	.292
Dima	.093	.135	.100	1.139	.131	.610	.401
O-Man	.609	.618	.014	.131	1.166	.080	.626
Deva	.054	.079	.613	.610	.080	1.148	.613
O-Style	.344	.357	.292	.401	.626	.613	1.157



Diferença (Genômico – Pedigree)

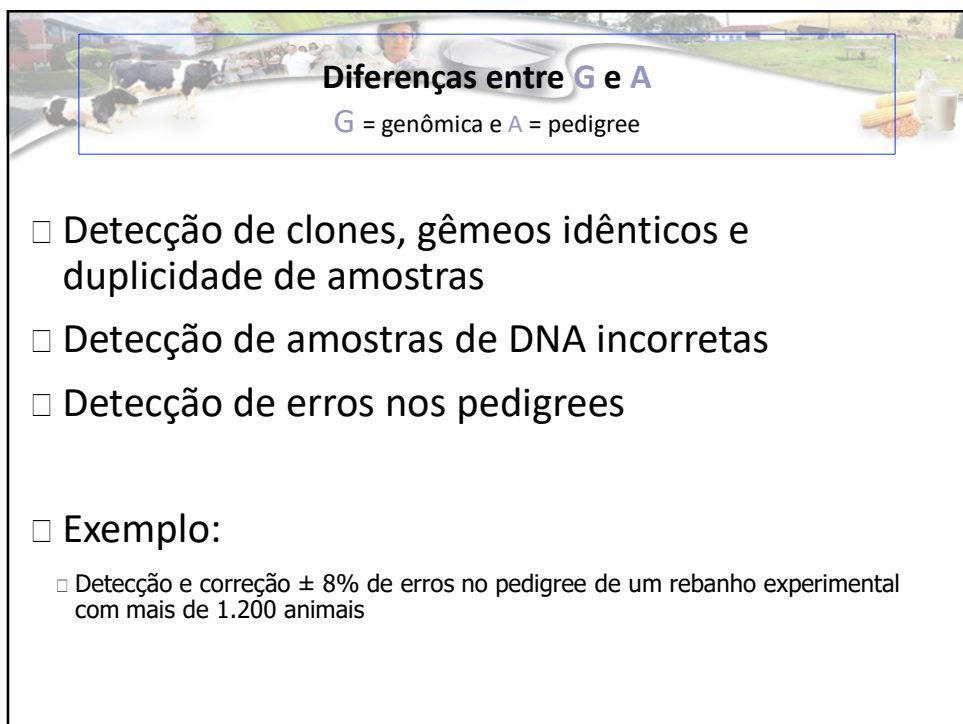
	AvôP	AvóP	AvôM	AvóM	Pai	Mãe	Touro
Manfred	.149	-.032	-.040	-.012	.038	-.043	.010
Jezebel	-.032	.095	-.043	.036	.055	.004	.038
Teamster	-.040	-.043	.075	-.021	-.057	.035	-.032
Dima	-.012	.036	-.021	.097	.029	.029	.059
O-Man	.038	.055	-.057	.029	.121	-.006	.060
Deva	-.043	.004	.035	.029	-.006	.087	.040
O-Style	.010	.038	-.032	.059	.060	.040	.114





Touro	Pedigree F	Genômica F
O Man	4.5	15.8
Ramos	2.3	11.5
Shottle	5.6	11.9
Planet	6.7	18.8
Earnit	6.2	12.8
Nifty	3.1	11.7

Correlação = .68

- 
- Diferenças entre G e A**
G = genômica e A = pedigree
- ☐ Detecção de clones, gêmeos idênticos e duplicidade de amostras
 - ☐ Detecção de amostras de DNA incorretas
 - ☐ Detecção de erros nos pedigrees
 - ☐ Exemplo:
 - ☐ Detecção e correção $\pm 8\%$ de erros no pedigree de um rebanho experimental com mais de 1.200 animais

Método do passo único

Combina todas as fontes de informação de um indivíduo em único passo (*single step*):

The diagram illustrates the single-step method equation: $GEBV = w_1PA + w_2YD + w_3PC + w_4DGV - w_5PP$. Callouts identify the components:
- PA : Contribuição média dos pais
- YD : Desempenho do próprio animal
- PC : Contribuição da progênie do animal
- DGV : Valor genômico direto
- PP : Contribuição do pedigree dos animais genotipados

$$GEBV = w_1PA + w_2YD + w_3PC + w_4DGV - w_5PP,$$

Fonte: Aguiar. Journal of Dairy Science Vol. 93 No. 2, 2010

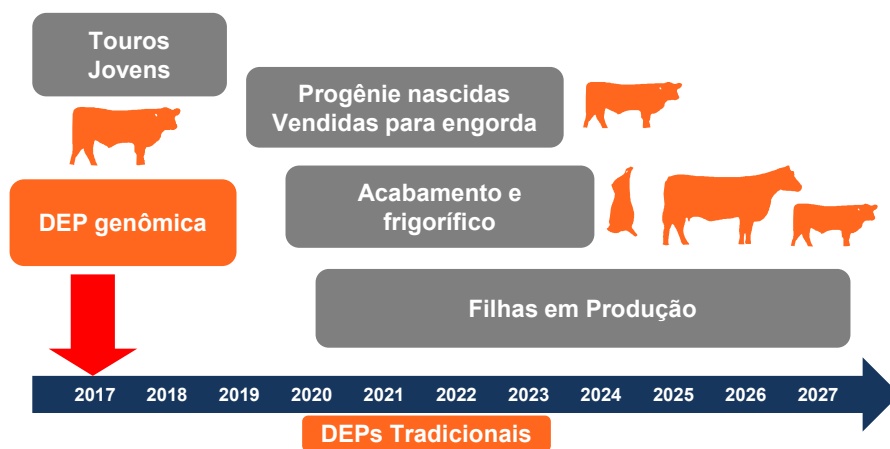
Vantagens da informação genômica para o melhoramento genético de bovinos de corte

- Aumentar a acurácia na seleção de animais jovens;
- Características que dependem de teste de progênie ou difícil mensuração;
- Características de baixa herdabilidade;
- Melhorar a estrutura do pedigree e controle da endogamia;
- Identificar QTL para características poligênicas e monogênicas;
- Variações estruturais e adaptação ambiental.

Avaliação de animais jovens e características de difícil mensuração

Seleção genômica

Podemos acelerar o ganho genético se identificarmos os animais superiores mais cedo na vida do animal?



Cronograma de atividades de teste de progênie

Exemplo da raça Holstein no Canadá

Tempo (meses)	Atividade
0	Vacas de elite escolhidas e acasaladas
9	Bezerros filhos de vacas de elite
+ de 5 anos para obter dados fenotípicos da progênie do reprodutor	
57	Primeiras estimativas de valores genéticos para os touros jovens pelo <i>test-day model</i>
64	Filhas completam a primeira lactação, seleção de touros

Schaeffer (2006)

49

Impacto da seleção genômica no gado de leite

$$\Delta G = \frac{iR\sigma_g}{L}$$

Estratégia de seleção	SEM SELEÇÃO GENÔMICA					
	% Seleção	i	r	Intervalo, L	i x r	Mudança no desvio padrão genético por ano
Pai de touros	5	2,06	0,99	6,5	2,04	0,31
Pai de vacas	20	1,40	0,75	6	1,05	0,18
Mae de touros	2	2,42	0,60	5	1,45	0,29
Mae de vacas	85	0,27	0,50	4,25	0,14	0,03
Total				21,75	4,68	0,21
Estratégia de seleção	COM SELEÇÃO GENÔMICA					
	% Seleção	i	r	Intervalo, L	i x r	Mudança no desvio padrão genéticos por ano
Pai de touros	5	2,06	0,75	1,75	1,54	0,88
Pai de vacas	20	1,40	0,75	1,75	1,05	0,60
Mae de touros	2	2,42	0,75	2	1,82	0,91
Mae de vacas	85	0,27	0,50	4,25	0,14	0,03
Total				9,75	4,55	0,47

50
Adaptado de Schaeffer, 2006

Avaliação genômica utilizando o método do ssGBLUP

Resultados de pesquisa com dados da ANCP:

Acurácia para P455 obtida pelo BLUP e ssGBLUP para diferentes grupos de animais

Categoria	BLUP	ssGBLUP	%
Todos os animais: 60.325	0,34	0,37	10
Touros: 1363	0,46	0,50	9
Animais genotipados: 3809	0,26	0,53	107
Genotipados com fenótipo: 1973	0,33	0,57	72
Genotipados sem fenótipo: 1836	0,17	0,48	180
Animais jovens: 13.529	0,16	0,24	44
Animais jovens sem genótipo: 12.014	0,16	0,21	25
Animais Jovens com genótipo: 1515	0,16	0,47	189

51

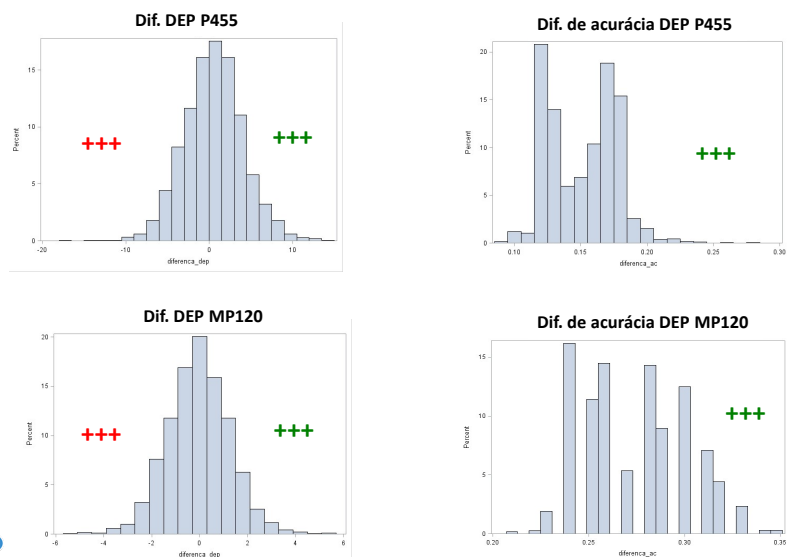
Número de progênes equivalentes para a DEP genômica (Clarified v3.1) para animais nascidos na safra 2016 e 2017

Característica	N°	h ²	DEP	Media %	DEPg_v3.1	Nº Progên. Equi.
IPP	2.891	0,11	0,16	44	0,23	6
3P	2.815	0,24	0,17	135	0,40	11
STAY	2.908	0,12	0,11	155	0,28	15
DPG	2.916	0,38	0,23	57	0,36	3
PAC	2.857	0,20	0,12	200	0,36	14
MP120	2.792	0,13	0,14	193	0,41	25
MP240	2.804	0,12	0,13	146	0,32	17
P120	2.923	0,28	0,23	61	0,37	5
P210	2.927	0,3	0,23	65	0,38	5
P365	2.924	0,36	0,27	52	0,41	4
P455	2.925	0,38	0,27	56	0,42	4
PE455	2.925	0,41	0,21	81	0,38	4
PE365	2.924	0,45	0,2	80	0,36	3
ED	2.931	0,66	0,18	106	0,37	3
ES	2.931	0,61	0,18	83	0,33	2
MD	2.931	0,66	0,18	83	0,33	2
MS	2.931	0,66	0,16	131	0,37	3
PD	2.931	0,61	0,17	118	0,37	3
PS	2.931	0,63	0,16	100	0,32	2
DALT	2.929	0,25	0,10	100	0,20	4
AOL	2.926	0,35	0,19	89	0,36	5
ACAB	2.926	0,35	0,19	68	0,32	3

*Animais no conjunto de validação



Distribuição da diferença entre acurácia e DEP genômica e DEP tradicional de animais jovens



Resposta em acurácia para CAR e IMS em animais jovens

Animais sem filhos na base de dados

Variável	Desvio	Std Dev	Min	Max
ims_ssgblup	0.196		0.0100	0.5000
ims_blup	0.148	+32%	0.0100	0.3400
car_ssgblup	0.185		0.0100	0.4800
car_blup	0.135	+37%	0.0100	0.3000

Animais com fenótipo e sem genótipo

Variável	Media	Desvio	Min	Max
ims_ssgblup	0.350		0.160	0.500
ims_blup	0.296	+18%	0.110	0.340
car_ssgblup	0.316		0.150	0.480
car_blup	0.259	+22%	0.100	0.300

Animais com fenótipo e genótipos

Variável	Media	Desvio	Min	Max
ims_ssgblup	0.438		0.320	0.500
ims_blup	0.292	+50%	0.240	0.330
car_ssgblup	0.415		0.280	0.480
car_blup	0.255	+63%	0.200	0.300

Animais sem fenótipo e com genótipo

Variável	Media	Desvio	Min	Max
ims_ssgblup	0.370		0.200	0.440
ims_blup	0.165	+124%	0.100	0.220
car_ssgblup	0.362		0.190	0.430
car_blup	0.154	+134%	0.090	0.210



Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population¹

R. M. O. Silva,^{*2} B. O. Fragomeni,[†] D. A. L. Lourenco,[†] A. F. B. Magalhães,^{*} N. Irano,^{*} R. Carvalheiro,^{*} R. C. Canesin,[‡] M. E. Z. Mercadante,[‡] A. A. Boligon,[§] F. S. Baldi,^{*} I. Misztal,[†] and L. G. Albuquerque^{*}

^{*}Faculdade de Ciências Agrárias de Veterinárias, UNESP – Univ Estadual Paulista, Department of Animal Science, Jaboticabal, São Paulo, Brazil, 14884-900; [†]University of Georgia, Department of Animal and Dairy Science, Athens 30602-2771; [‡]Centro APTA Bovinos de Corte, Animal Science Institute, Sertãozinho, São Paulo, Brazil, 13460-000; and [§]Department of Animal Science, Federal University of Pelotas, Pelotas, RS, Brazil, CEP 96160-000

ABSTRACT

Animal feeding is the most important economic component of beef production systems. Selection for feed efficiency has not been effective mainly due to difficult and high costs to obtain the phenotypes. The application of genomic selection using SNP can decrease the cost of animal evaluation as well as the generation interval. The objective of this study was to compare methods for genomic evaluation of feed efficiency traits using different cross-validation layouts in an experimental beef cattle population genotyped for a high-density SNP panel (BovineHD BeadChip assay 700k, Illumina Inc., San Diego, CA). After

Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population¹

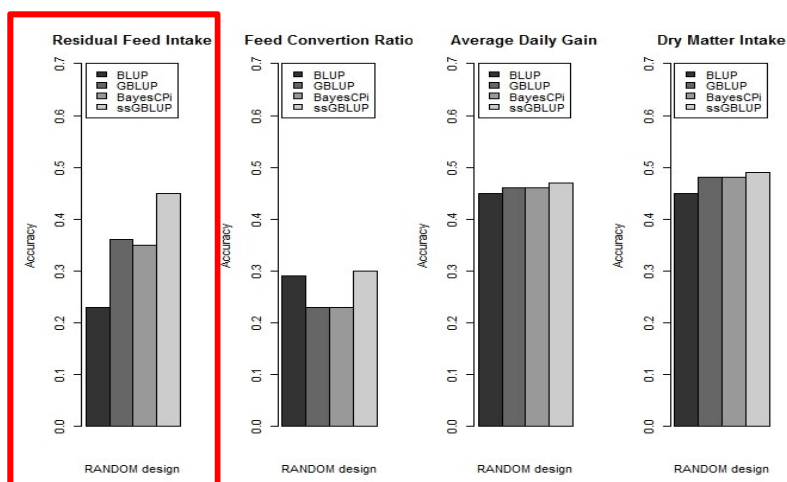


Figure 4 - Average accuracies for each studied trait using the RANDOM cross-validation design obtained using BLUP, genomic BLUP, BayesCπ and single step genomic BLUP.

Fonte: Silva et al., 2016 J. Anim. Sci. 2016.94:1–11 doi:10.2527/jas2016-0401



Contents lists available at ScienceDirect

Meat Science

journal homepage: www.elsevier.com/locate/meatsci





Genomic prediction for beef fatty acid profile in Nellore cattle

Hermenegildo Lucas Justino Chiaia ^{a,*}, Elisa Peripoli ^a, Rafael Medeiros de Oliveira Silva ^a, Carolyn Aboujaoude ^a, Fabiele Loise Braga Feitosa ^a, Marcos Vinicius Antunes de Lemos ^a, Mariana Piatto Berton ^a, Bianca Ferreira Olivieri ^a, Rafael Espigolan ^a, Rafael Lara Tonussi ^a, Daniel Gustavo Mansan Gordo ^a, Tiago Bresolin ^a, Ana Fabrícia Braga Magalhães ^a, Gerardo Alves Fernandes Júnior ^a, Lúcia Galvão de Albuquerque ^a, Henrique Nunes de Oliveira ^a, Joyce de Jesus Mangini Furlan ^b, Adrielle Mathias Ferrinho ^b, Lenise Freitas Mueller ^c, Humberto Tonhati ^a, Angélica Simone Cravo Pereira ^b, Fernando Baldi ^a

A B S T R A C T

The objective of this study was to compare SNP-BLUP, BayesCrr, BayesC and Bayesian Lasso methodologies to predict the direct genomic value for saturated, monounsaturated, and polyunsaturated fatty acid profile, omega 3 and 6 in the *Longissimus thoracis* muscle of Nellore cattle finished in feedlot. A total of 963 Nellore bulls with phenotype for fatty acid profiles, were genotyped using the Illumina BovineHD BeadChip (Illumina, San Diego, CA) with 777,962 SNP. The predictive ability was evaluated using cross validation. To compare the methodologies, the correlation between DGV and pseudo-phenotypes was calculated. The accuracy varied from -0.40 to 0.62. Our results indicate that none of the methods excelled in terms of accuracy, however, the SNP-BLUP method allows obtaining less biased genomic evaluations, thereby; this method is more feasible when taking into account the analyses' operating cost. Despite the lowest bias observed for EBV, the adjusted phenotype is the preferred pseudophenotype considering the genomic prediction accuracies regarding the context of the present study.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Meat Science 128 (2017) 60–67

Table 4
 Accuracies of genomic predictions measured by Pearson's correlation between pseudophenotype and direct genomic breeding values ($r_{(Y|DGV)}$), and standard error (SE) for beef polyunsaturated fatty acids of Nellore cattle based on different methods.

Fatty acid (g/100 g)	Type	$r_{(Y DGV)} \pm (SE)$		Bayesian LASSO		BayesC		BayesCrr	
		SNP-BLUP							
C18:2 n-6	Yc	0.25	(0.10)	0.23	(0.06)	0.22	(0.06)	0.12	(0.11)
	EBV	0.21	(0.02)	0.13	(0.06)	0.11	(0.06)	0.11	(0.07)
C18:2 cis-9 trans-11	Yc	0.13	(0.06)	-0.11	(0.13)	-0.04	(0.09)	-0.07	(0.10)
	EBV	0.21	(0.02)	0.08	(0.04)	0.10	(0.01)	0.18	(0.03)
C18:3 n-3	Yc	-0.01	(0.09)	0.00	(0.05)	0.06	(0.07)	-0.03	(0.04)
	EBV	0.14	(0.01)	0.07	(0.05)	0.06	(0.03)	0.07	(0.05)
C18:3 n-6	Yc	-0.02	(0.04)	-0.05	(0.06)	0.10	(0.03)	0.09	(0.01)
	EBV	0.01	(0.01)	0.00	(0.05)	0.08	(0.03)	0.05	(0.04)
C20:3 n-3	Yc	-0.06	(0.12)	0.13	(0.06)	0.12	(0.04)	0.19	(0.08)
	EBV	0.12	(0.03)	0.06	(0.05)	0.09	(0.03)	0.10	(0.03)
C22:6 n-3	Yc	0.10	(0.08)	0.18	(0.11)	0.15	(0.13)	0.18	(0.12)
	EBV	0.36	(0.18)	0.36	(0.19)	0.33	(0.22)	0.31	(0.21)
Sum of n-3	Yc	0.12	(0.02)	0.24	(0.11)	0.11	(0.12)	0.25	(0.10)
	EBV	0.07	(0.06)	0.07	(0.06)	0.05	(0.07)	0.12	(0.05)
Sum of n-6	Yc	0.22	(0.09)	0.26	(0.07)	0.34	(0.07)	0.24	(0.08)
	EBV	0.22	(0.03)	0.14	(0.08)	0.22	(0.03)	0.15	(0.07)
n-6/n-3	Yc	-0.07	(0.12)	-0.02	(0.06)	-0.12	(0.15)	0.19	(0.06)
	EBV	0.11	(0.05)	-0.02	(0.01)	0.05	(0.05)	0.07	(0.05)
Sum of PUFA	Yc	0.50	(0.13)	0.45	(0.12)	0.56	(0.07)	0.48	(0.10)
	EBV	0.24	(0.03)	0.15	(0.08)	0.16	(0.07)	0.15	(0.09)
PUFA/SFA	Yc	0.28	(0.15)	0.17	(0.13)	0.21	(0.12)	0.13	(0.17)
	EBV	0.32	(0.14)	0.32	(0.15)	0.33	(0.17)	0.32	(0.15)

^a For the Yc as response variable, $r_{(Y|DGV)}$ was divided by the square root of heritability of the trait; Yc phenotype adjusted for fixed effects, EBV estimated breeding value, SFA saturated fatty acid, n-3 and n-6 total of omega 3 and 6 fatty acids respectively, n-6/n-3 ratio of omega 6 to omega 3 fatty acids, PUFA Polyunsaturated fatty acid, PUFA/SFA ratio of polyunsaturated to saturated fatty acid.

Table 2
 Accuracies of genomic predictions measured by Pearson's correlation between pseudophenotype and direct genomic breeding values ($r_{(Y|DGV)}$), and standard error (SE) for beef saturated fatty acids of Nellore cattle based on different methods.

Fatty acid (g/100 g)	Type	$r_{(Y DGV)} \pm (SE)$		Bayesian LASSO		BayesC		BayesCrr	
		SNP-BLUP							
C14:0	Yc	0.30	(0.03)	0.19	(0.07)	0.15	(0.08)	0.15	(0.10)
	EBV	0.20	(0.02)	0.16	(0.06)	0.10	(0.07)	0.16	(0.08)
C16:0	Yc	0.25	(0.06)	0.13	(0.06)	0.15	(0.03)	0.19	(0.10)
	EBV	0.14	(0.03)	0.07	(0.04)	0.03	(0.05)	0.09	(0.04)
C18:0	Yc	0.16	(0.14)	0.28	(0.03)	0.13	(0.13)	0.19	(0.15)
	EBV	0.05	(0.03)	0.07	(0.03)	-0.01	(0.04)	0.03	(0.04)
Sum of SFA	Yc	0.20	(0.19)	0.12	(0.07)	0.17	(0.11)	0.24	(0.12)
	EBV	0.17	(0.04)	0.04	(0.05)	0.10	(0.02)	0.09	(0.03)

^a For the Yc as response variable, $r_{(Y|DGV)}$ was divided by the square root of heritability of the trait; Yc phenotype adjusted for fixed effects, EBV estimated breeding value, SFA saturated fatty acid.

4. Conclusions

Beef FA profile is a very important study area due to its implication in human health. Genomic information can assist in improving FA profile in Zebu animals, since the use of genomic information yielded genomic values for FA profile with accuracies ranging from low to moderate. None of the methods excelled in terms of accuracy, however, the SNP-BLUP method allows obtaining less biased genomic evaluations, thereby; this method is more feasible when taking account the computational cost. Despite the lowest bias observed for EBV, the adjusted phenotype is the preferred pseudophenotype considering the genomic prediction accuracies regarding the context of the present study.

Journal of Applied Genetics
<https://doi.org/10.1007/s13353-018-0470-5>

ANIMAL GENETICS • ORIGINAL PAPER

CrossMark

Genomic prediction ability for beef fatty acid profile in Nelore cattle using different pseudo-phenotypes

Hermenegildo Lucas Justino Chiaia¹ • Elisa Peripolli¹ • Rafael Medeiros de Oliveira Silva¹ • Fabiele Loise Braga Feitosa¹ • Marcos Vinicius Antunes de Lemos¹ • Mariana Piatto Berton¹ • Bianca Ferreira Olivieri¹ • Rafael Espigolan¹ • Rafael Lara Tonussi¹ • Daniel Gustavo Mansan Gordo¹ • Lucia Galvão de Albuquerque¹ • Henrique Nunes de Oliveira¹ • Adrielle Mathias Ferrinho² • Lenise Freitas Mueller³ • Sabrina Kluska¹ • Humberto Tonhati¹ • Angélica Simone Cravo Pereira² • Ignacio Aguilar⁴ • Fernando Baldi¹

Abstract

The aim of the present study was to compare the predictive ability of SNP-BLUP model using different pseudo-phenotypes such as phenotype adjusted for fixed effects, estimated breeding value, and genomic estimated breeding value, using simulated and real data for beef FA profile of Nelore cattle finished in feedlot. A pedigree with phenotypes and genotypes of 10,000 animals were simulated, considering 50% of multiple sires in the pedigree. Regarding to phenotypes, two traits were simulated, one with high heritability (0.58), another with low heritability (0.13). Ten replicates were performed for each trait and results were

Fatty acid (g/100 g)	Pseudo-phenotype	BIF accuracy	Pedigree SNP-BLUP				
			BLUP prediction ability				
			$r_{(Y,LDGV)}$	$\pm$ (SE)	$\hat{h}_{(Y,LDGV)}$	$\pm$ (SE)	
Sum of SFA	Yc		0.30 $\pm$ 0.04	0.20	(0.19)	0.07	(0.02)
	EBV	0.06		0.17	(0.04)	0.17	(0.04)
	GEBV	0.08		0.54	(0.04)	0.54	(0.04)
Sum of MUFA	Yc		0.11 $\pm$ 0.04	0.12	(0.03)	0.05	(0.01)
	EBV	0.01		0.07	(0.02)	0.08	(0.02)
	GEBV	0.02		0.34	(0.03)	0.34	(0.03)
Sum of n-3	Yc		0.44 $\pm$ 0.03	0.12	(0.02)	0.04	(0.01)
	EBV	0.06		0.07	(0.06)	0.10	(0.05)
	GEBV	0.07		0.47	(0.02)	0.47	(0.02)
Sum of n-6	Yc		0.31 $\pm$ 0.05	0.32	(0.09)	0.15	(0.04)
	EBV	0.12		0.22	(0.03)	0.22	(0.03)
	GEBV	0.14		0.47	(0.04)	0.47	(0.04)
Sum of PUFA	Yc		0.25 $\pm$ 0.19	0.38	(0.11)	0.14	(0.04)
	EBV	0.04		0.24	(0.28)	0.24	(0.03)
	GEBV	0.05		0.45	(0.03)	0.16	(0.08)
PUFA/SFA	Yc		0.41 $\pm$ 0.19	0.28	(0.15)	0.07	(0.04)
	EBV	0.05		0.32	(0.14)	0.32	(0.14)
	GEBV	0.06		0.45	(0.10)	0.45	(0.10)

Phenotype adjusted for fixed effects (Yc), estimated breeding value using pedigree (EBV), genomic estimated breeding value using pedigree and genomic information (GEBV); correlation between pseudo-phenotype and direct genomic breeding values ($r_{(Y,LDGV)}$), regression coefficient of the pseudo-phenotype on direct genomic breeding values ($\hat{h}_{(Y,LDGV)}$) and standard error (SE). Beef Improvement Federation (BIF). Best linear unbiased prediction (BLUP)

Conclusions

Genomic information can assist in improving beef fatty acid profile in Zebu cattle, since the use of genomic information yielded genomic values for fatty acid profile with accuracies ranging from low to moderate. Considering both simulated and real data, the ssGBLUP model is an appropriate alternative to obtain more reliable and less biased GEBVs as pseudo-phenotype in situations of missing pedigree, due to high proportion of multiple sires, being more adequate than EBV and Yc to predict direct genomic value for beef fatty acid profile.

Situações de incerteza de paternidade

O próximo desafio: Parentesco genômico

- Podemos utilizar a informação genômica para “recuperar” parentesco;
- Potencial para a avaliação de rebanhos comerciais e situações com paternidade incerta;
- A partir do parentesco genômico é possível quantificar a endogamia genômica.

62

Melhoramento genético: estamos mesmo fazendo melhoramento?

Nº Matrizes (2011)	74.487.416
Quantidade de touros (1/25)	2.979.497

Necessidade de touros de reposição: ~3 milhões x 20% de reposição = ~600.000/ano

Quantos touros são avaliados geneticamente?

~17.000 CEIP + ~18.000 PO superiores = 35.000 <6% de nossas necessidades.

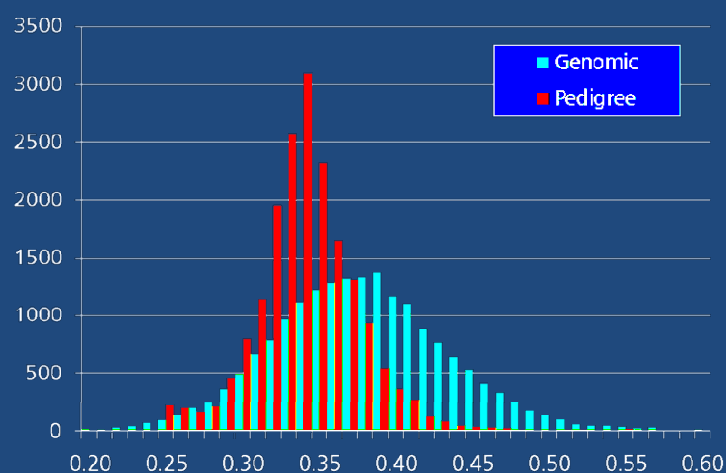
94% dos touros de reposição do Brasil → geneticamente desconhecidos.

Baruselli, SIIRA, Londrina, 08/2014

**Como a informação genômica pode contribuir para
reverter este cenário?**

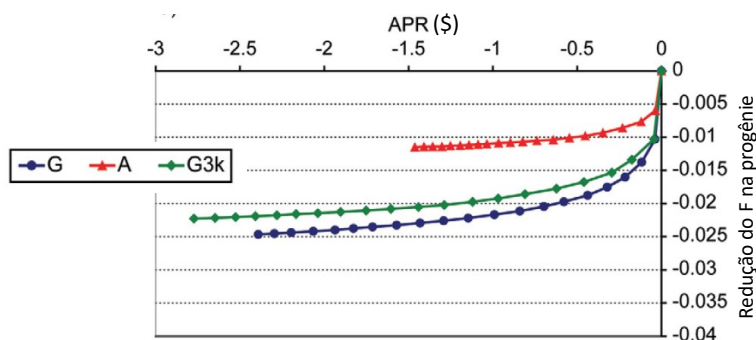
63

Comparação do parentesco tradicional e genômico (Parentesco neto-avô materno)



USDA-AIPL information

Estratégias para minimizar a endogamia na progênie maximizando o ganho genético



Efeito da endogamia por pedigree (A) e endogamia genômica (G, G3k) sobre a redução do ganho genético para APR (Australian Profit Index \$) na progênie

Fonte: Pryce et al., 2012.

65



RESEARCH ARTICLE

Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data

Rafael Lara Tonussi¹, Rafael Medeiros de Oliveira Silva¹, Ana Fabricia Braga Magalhães¹, Rafael Espigolan¹, Elisa Peripolli¹, Bianca Ferreira Olivieri¹, Fabieli Loise Braga Feitosa¹, Marcos Vinicius Antunes Lemos¹, Mariana Piatto Berton¹, Hermenegildo Lucas Justino Chialia¹, Angelica Simone Cravo Pereira², Raysildo Barbosa Lôbo³, Luiz Antônio Framartino Bezerra⁴, Cláudio de Ulhoa Magnabosco⁵, Daniela Andressa Lino Lourenço⁶, Ignácio Aguiar⁷, Fernando Baldi^{1*}

Abstract

The objective of this study was to investigate the application of BLUP and single step genomic BLUP (ssGBLUP) models in different scenarios of paternity uncertainty with different strategies of scaling the G matrix to match the A22 matrix, using simulated data for beef cattle. Genotypes, pedigree, and phenotypes for age at first calving (AFC) and weight at 550 days (W550) were simulated using heritabilities based on real data (0.12 for AFC and 0.34 for W550). Paternity uncertainty scenarios using 0, 25, 50, 75, and 100% of multiple sires (MS) were studied. The simulated genome had a total length of 2,333 cM, containing 735,293 biallelic markers and 7,000 QTLs randomly distributed over the 29 BTA. It was assumed that QTLs explained 100% of the genetic variance. For QTL, the amount of alleles per loci randomly ranged from two to four. The BLUP model that considers phenotypic and

Workshop de Melhoramento Genético Animal
 Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
 Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Table 3. Breeding value accuracy estimates and standard errors for AFC and W550 using BLUP and ssGBLUP models with different proportions of multiple sires.

		Percentage of multiple sires					
		0%	25%	50%	75%	100%	
AFC	ALL	BLUP	0.79±0.006	0.76±0.005	0.73±0.005	0.70±0.005	0.68±0.004
		ssGBLUP	0.80±0.006	0.78±0.005	0.76±0.005	0.73±0.005	0.69±0.004
	BULL	BLUP	0.87±0.005	0.86±0.005	0.84±0.006	0.83±0.006	0.85±0.005
		ssGBLUP	0.87±0.005	0.85±0.005	0.81±0.004	0.79±0.004	0.85±0.005
	GEN	BLUP	0.46±0.007	0.20±0.005	0.16±0.004	0.16±0.005	0.30±0.006
		ssGBLUP	0.55±0.009	0.57±0.009	0.56±0.009	0.56±0.009	0.56±0.010
	FEM	BLUP	0.81±0.005	0.78±0.005	0.76±0.005	0.74±0.005	0.72±0.004
		ssGBLUP	0.81±0.005	0.80±0.005	0.78±0.005	0.76±0.005	0.73±0.004
	YOUNG	BLUP	0.41±0.007	0.14±0.005	0.08±0.005	0.06±0.007	0.05±0.008
		ssGBLUP	0.50±0.009	0.50±0.008	0.46±0.008	0.42±0.007	0.31±0.007
W550	ALL	BLUP	0.92±0.001	0.89±0.001	0.86±0.002	0.84±0.002	0.82±0.002
		ssGBLUP	0.92±0.001	0.92±0.001	0.90±0.001	0.88±0.001	0.84±0.001
	BULL	BLUP	0.97±0.001	0.96±0.001	0.95±0.001	0.93±0.001	0.97±0.001
		ssGBLUP	0.97±0.001	0.97±0.001	0.95±0.001	0.94±0.001	0.97±0.001
	GEN	BLUP	0.60±0.005	0.29±0.003	0.25±0.003	0.28±0.003	0.41±0.003
		ssGBLUP	0.72±0.006	0.70±0.005	0.68±0.004	0.67±0.004	0.66±0.004
	FEM	BLUP	0.93±0.001	0.91±0.001	0.89±0.001	0.88±0.001	0.87±0.002
		ssGBLUP	0.93±0.001	0.92±0.001	0.91±0.001	0.89±0.001	0.87±0.002
	YOUNG	BLUP	0.50±0.004	0.16±0.003	0.09±0.005	0.06±0.005	0.07±0.006
		ssGBLUP	0.65±0.006	0.60±0.005	0.53±0.003	0.46±0.003	0.36±0.003

AFC = age at first calving; W550 = weight at 550 days; ALL = all animals; BULL = bulls; GEN = genotyped animals; FEM = females; YOUNG = young males; BLUP = best linear unbiased prediction; ssGBLUP = single step genomic BLUP

Conclusions

Despite the ssGBLUP procedure was not developed to deal with paternity uncertainty situations, the ssGBLUP model is an appropriate alternative to obtaining more reliable and less biased breeding values in situations of missing pedigree, especially for young animals with few or no phenotypic records.

It is important to scale the G matrix to be compatible with the numerator relationship matrix for genotyped animals, even in situations where the latter is less informative due to the presence of missing pedigree. For accurate and unbiased genomic predictions with the ssGBLUP model, it is necessary to assure that the G matrix is compatible with the A_{22} matrix even in situations with paternity uncertainty.

Accuracy of genomic breeding values and predictive ability for growth and reproductive traits in a Nellore cattle population with multiple sire mating system¹

Rafael Lara Tonussi², Rafael Medeiros de Oliveira Silva^{2,8}, Ana Fabricia Braga Magalhães², Rafael Espigolan², Elisa Peripolli², Bianca Ferreira Olivieri², Fabieli Loise Braga Feitosa², Marcos Vinicius Antunes Lemos², Mariana Piatto Berton², Angelica Simone Cravo Pereira³, Raysildo Barbosa Lôbo⁴, Ignácio Aguilar⁵, Fernando A. Di Croce⁶, Jason B. Osterstock⁶, Sue Denise⁶, Daniela Andressa Lino Lourenco⁷ and Fernando Baldi²

Table 3. Accuracies and standard errors of (G)EBV for age at first calving using BLUP and ssGBLUP with different proportions of multiple sires (0, 25, 50, 75, and 100%).

Group		Percentage of multiple sires				
		0%	25%±SE	50%±SE	75%±SE	100%
PHE	BLUP	0.14	0.14±0.0003	0.14±0.0004	0.14±0.0002	0.14
	ssGBLUP	0.17	0.17±0.0002	0.17±0.0003	0.16±0.0002	0.16
SIR	BLUP	0.17	0.17±0.0007	0.17±0.0005	0.16±0.0002	0.16
	ssGBLUP	0.21	0.21±0.0002	0.20±0.0003	0.19±0.0002	0.18
GEN	BLUP	0.10	0.08±0.0008	0.07±0.0002	0.05±0.0001	0.04
	ssGBLUP	0.21	0.19±0.0008	0.17±0.0009	0.14±0.0007	0.11
YNG	BLUP	0.09	0.09±0.0007	0.09±0.0002	0.08±0.0001	0.08
	ssGBLUP	0.12	0.12±0.0001	0.11±0.0002	0.11±0.0001	0.10
YNGEN	BLUP	0.09	0.07±0.008	0.06±0.004	0.04±0.003	0.03
	ssGBLUP	0.25	0.18±0.009	0.16±0.001	0.13±0.002	0.11

BLUP = best linear unbiased predictor; ssGBLUP = single step genomic BLUP; SE = standard error; PHE = all animals in the population; SIR = only bulls with ten or more progenies; GEN = genotyped animals and, YNG = male and female young animals without phenotypes; YNGEN = genotyped young animals

Fonte: Tonussi et al. (2017). Submitted to Animal Journal.

68

Accuracy of genomic breeding values and predictive ability for growth and reproductive traits in a Nellore cattle population with multiple sire mating system¹

Rafael Lara Tonussi², Rafael Medeiros de Oliveira Silva^{7,8}, Ana Fabricia Braga Magalhães², Rafael Espigolan², Elisa Peripolli², Bianca Ferreira Olivieri², Fabieli Loise Braga Feitosa², Marcos Vinicius Antunes Lemos², Mariana Piatto Berton², Angelica Simone Cravo Pereira³, Raysildo Barbosa Lôbo⁴, Ignácio Aguilar⁵, Fernando A. Di Croce⁶, Jason B. Osterstock⁶, Sue Denise⁶, Daniela Andressa Lino Lourenco⁷ and Fernando Baldi²

Table 6. Pearson correlation between DGV_y and EBV or Y_{adj} and standard errors for different proportions of multiple sires

Trait	Benchmark	Percentage of multiple sires				
		0%	25%±SE	50%±SE	75%±SE	100%
AFC	Y _{adj}	0.25	0.25±0.002	0.24±0.003	0.23±0.002	0.23
	EBV	0.68	0.67±0.002	0.64±0.007	0.58±0.007	0.47
W450	Y _{adj}	0.12	0.11±0.002	0.10±0.002	0.09±0.002	0.08
	EBV	0.56	0.54±0.004	0.49±0.009	0.41±0.012	0.32

DGV_y=direct genetic value for young genotyped animals; Y_{adj}=adjusted phenotype for fixed effects; EBV=estimated breeding value; SE=standard error; AFC=age of first calving; W450= weight at 450 days

Fonte: Tonussi et al. (2017). Submitted to Animal Journal.

Avaliação genômica de rebanhos comerciais utilizando diferentes estratégias

Duas abordagens ou estratégias para avaliação genômica:

- 1) **ssGBLUP_comercial**: Animais “cara limpa” com painel ZL5
- 2) **ssGBLUP_comercial+registrado**: Animais “cara limpa”+ animais registrados

Característica	Estratégia	N_Dados	N_Genótipos	Acurácia	% Aumento
IPP	ssGBLUP_comercial ¹	53.866	2.017	0,28	46
	ssGBLUP_comercial+registrado ²	456.951	2.017(12.206)	0,41	
PAC	ssGBLUP_comercial	18.759	2.017	0,27	48
	ssGBLUP_comercial+registrado ²	181.871	2.017(12.206)	0,40	
MP120	ssGBLUP_comercial	143.614	2.017	0,28	46
	ssGBLUP_comercial+registrado ²	1.085.945	2.017(12.206)	0,41	
DP120	ssGBLUP_comercial	143.614	2.017	0,31	42
	ssGBLUP_comercial+registrado ²	1.085.945	2.017(12.206)	0,44	
P455	ssGBLUP_comercial	124.169	2.017	0,31	45
	ssGBLUP_comercial+registrado ²	840.145	2.017(12.206)	0,45	

¹272534 animais no pedigree

²1.757.011 animais no pedigree



Projeto Fapesp #2016/22751-6

Identificação de QTL para características poligênicas e monogênicas

 PLOS ONE

RESEARCH ARTICLE

Genomic Regions Associated with Feed Efficiency Indicator Traits in an Experimental Nellore Cattle Population

Bianca Ferreira Olivieri^{1*}, Maria Eugênia Zerloti Mercadante², Joseline Noely dos Santos Gonçalves Cyrillo², Renata Helena Branco², Sarah Figueiredo Martins Bonilha², Lucila Galvão de Albuquerque¹, Rafael Medeiros de Oliveira Silva¹, Fernando Baldi¹

Abstract

The objective of this study was to identify genomic regions and metabolic pathways associated with dry matter intake, average daily gain, feed efficiency and residual feed intake in an experimental Nellore cattle population. The high-density SNP chip (Illumina High-Density Bovine BeadChip, 777K) was used to genotype the animals. The SNP markers effects and their variances were estimated using the single-step genome wide association method. The (co)variance components were estimated by Bayesian inference. The chromosome segments that are responsible for more than 1.0% of additive genetic variance were selected to explore and determine possible quantitative trait loci. The bovine genome Map Viewer was used to identify genes. In total, 51 genomic regions were identified for all analyzed traits. The heritability estimated for feed efficiency was low magnitude (0.13±0.06). For average daily gain, dry matter intake and residual feed intake, heritability was moderate to high (0.43±0.05; 0.47±0.05, 0.18±0.05, respectively). A total of 8, 17, 14 and 12 windows that are responsible for more than 1% of the additive genetic variance for dry matter intake, average daily gain, feed efficiency and residual feed intake, respectively, were identified. Candidate genes *GOLIM4*, *RFX6*, *CACNG7*, *CACNG6*, *CAPN8*, *CAPN2*, *AKT2*, *GPRC6A*, and *GPR45* were associated with feed efficiency traits. It was expected that the response to selection would be higher for residual feed intake than for feed efficiency. Genomic regions harboring possible QTL for feed efficiency indicator traits were identified. Candidate genes identified are involved in energy use, metabolism protein, ion transport, transmembrane transport, the olfactory system, the immune system, secretion and cellular activity. The identification of these regions and their respective candidate genes should contribute to the formation of a genetic basis in Nellore cattle for feed efficiency indicator traits, and these results would support the selection for these traits.

Received: April 20, 2016
Accepted: September 23, 2016
Published: October 19, 2016
Copyright: © 2016 Olivieri et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Data Availability Statement: All relevant data are within the paper and its Supporting Information files.
Funding: This work was supported by Sao Paulo Research Foundation (FAPESP, grant #2009/18118-5) for financial support, and Coordination for the Improvement of Higher Education Personnel (CAPES) for providing grants to BFO.
Competing Interests: The authors have declared that no competing interests exist.

Lemos et al. BMC Genomics (2016) 17:213
 DOI 10.1186/s12864-016-2511-y

BMC Genomics

RESEARCH ARTICLE

Open Access



Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure

Marcos V. A. Lemos^{1*}, Hermenegildo Lucas Justino Chiala¹, Mariana Piatto Berton¹, Fabieli L. B. Feitosa¹,
 Carolyn Aboujaoud², Gregório M. F. Camargo¹, Angélica S. C. Pereira¹, Lucia G. Albuquerque¹, Adrielle M. Ferrinho²,
 Lenise F. Mueller², Monica R. Mazzali², Joyce J. M. Furlan², Roberto Carneiro², Daniel M. Gordo¹, Rafael Tonussi¹,
 Rafael Espigolan¹, Rafael Medeiros de Oliveira Silva¹, Henrique Nunes de Oliveira¹, Susan Duckett³,
 Ignacio Aguilar⁴ and Fernando Baldi¹

Abstract

Background: Saturated fatty acids can be detrimental to human health and have received considerable attention in recent years. Several studies using taurine breeds showed the existence of genetic variability and thus the possibility of genetic improvement of the fatty acid profile in beef. This study identified the regions of the genome associated with saturated, mono- and polyunsaturated fatty acids, and n-6 to n-3 ratios in the *Longissimus thoracis* of Nellore finished in feedlot, using the single-step method.

Results: The results showed that 115 windows explain more than 1 % of the additive genetic variance for the 22 studied fatty acids. Thirty-one genomic regions that explain more than 1 % of the additive genetic variance were observed for total saturated fatty acids, C12:0, C14:0, C16:0 and C18:0. Nineteen genomic regions, distributed in sixteen different chromosomes accounted for more than 1 % of the additive genetic variance for the monounsaturated fatty acids, such as the sum of monounsaturated fatty acids, C14:1 cis-9, C18:1 trans-11, C18:1 cis-9, and C18:1 trans-9. Forty genomic regions explained more than 1 % of the additive genetic variance for the polyunsaturated fatty acids group, which are related to the total polyunsaturated fatty acids, C20:4 n-6, C18:2 cis-9 cis-12 n-6, C18:3 n-3, C18:3 n-6, C22:6 n-3 and C20:3 n-6 cis-8 cis-11 cis-14. Twenty-one genomic regions accounted for more than 1 % of the genetic variance for the group of omega-3, omega-6 and the n-6:n-3 ratio.

Conclusions: The identification of such regions and the respective candidate genes, such as *ELCVL5*, *ESSRG*, *PCYT1A* and genes of the ABC group (*ABC5*, *ABC6* and *ABC10*), should contribute to form a genetic basis of the fatty acid profile of Nellore (*Bos indicus*) beef, contributing to better selection of the traits associated with improving human health.

Keywords: *Bos indicus*, Fatty acid composition, Heritability, Genetic markers, Mapping, ssGWAS

Journal of Applied Genetics
 https://doi.org/10.1007/s13353-018-0436-7

ANIMAL GENETICS - ORIGINAL PAPER



Association study between copy number variation and beef fatty acid profile of Nellore cattle

Marcos Vinicius Antunes de Lemos¹ · Elisa Peripolli¹ · Mariana Piatto Berton¹ · Fabiele Loise Braga Feitosa¹ ·
 Bianca Ferreira Olivieri¹ · Nedenia Bonvino Stafuzza¹ · Rafael Lara Tonussi¹ · Sabrina Kluska¹ ·
 Hermenegildo Lucas Justino Chiala¹ · Lenise Mueller² · Adrielli Mathias Ferrinho² · Angelica Simone Cravo Prereira² ·
 Henrique Nunes de Oliveira¹ · Lucia Galvão de Albuquerque¹ · Fernando Baldi¹

Received: 3 October 2017 / Revised: 23 January 2018 / Accepted: 26 February 2018
 © Institute of Plant Genetics, Polish Academy of Sciences, Poznan 2018

Abstract

The aim of this study was to analyze the association between the copy number variation regions (CNVRs) and fatty acid profile phenotypes for saturated (SFA), monounsaturated (MUFA), polyunsaturated (PUFA), ω 6 and ω 3 fatty acids, PUFA/SFA and ω 6/ ω 3 ratios, as well as for their sums, in Nellore cattle (*Bos primigenius indicus*). A total of 963 males were finished in feedlot and slaughtered with approximately 2 years of age. Animals were genotyped with the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA). The copy number variation (CNV) detection was performed using the PennCNV algorithm. Log R ratio (LRR) and allele B frequency (BAF) were used to estimate the CNVs. The association analyses were done using the CNVRuler software and applying a logistic regression model. The phenotype was adjusted using a linear model considering the fixed effects of contemporary group and the animal age at slaughter. The fatty acid profile was analyzed on samples of *longissimus thoracis* muscle using gas chromatography with a 100-m capillary column. For the association analysis, the adjusted phenotypic values were considered for the traits, while the data was adjusted for the effects of the farm and year of birth, management groups at birth, weaning, and superannuation. A total of 186 CNVRs were significant for SFA (43), MUFA (42), PUFA (66), and omega fatty acid (35) groups, totaling 278 known genes. On the basis of the results, several genes were associated with several fatty acids of different saturations. Olfactory receptor genes were associated with C12:0, C14:0, and C18:0 fatty acids. The *SAMD8* and *BSCL2* genes, both related to lipid metabolic process, were associated with C12:0. The *RAPGEF6* gene was found to be associated with C18:2 cis-9 cis-12 n-6, and its function is related to regulation of GTPase activity. Among the results, we highlighted the olfactory receptor activity (GO:0004984), G-protein-coupled receptor activity (GO:0004930), potassium:proton antiporter activity (GO:0015386), sodium:proton antiporter activity (GO:0015385), and odorant-binding (GO:0005549) molecular functions. A large number of genes associated with fatty acid profile within the CNVRs were identified in this study. These findings must contribute to better elucidate the genetic mechanism underlying the fatty acid profile of intramuscular fat in Nellore cattle.



RESEARCH ARTICLE

Genome-wide association study provides insights into genes related with horn development in Nelore beef cattle

Nedenia Bonvino Stafuzza^{1,2*}, Rafael Medeiros de Oliveira Silva^{2,3*}, Elisa Peripolli⁴, Luiz Antônio Framartino Bezerra⁵, Raysildo Barbosa Lôbo⁶, Cláudio de Ulhoa Magnabosco⁷, Fernando A. Di Croce⁸, Jason B. Osterstock⁸, Danísio Prado Munari¹, Daniela A. Lino Lourenço², Fernando Baldi⁴

Received: February 11, 2018
Accepted: August 13, 2018
Published: August 30, 2018

Copyright: © 2018 Stafuzza et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The causal mutation for polledness in Nelore (*Bos taurus indicus*) breed seems to have appeared first in Brazil in 1957. The expression of the polled trait is known to be ruled by a few groups of alleles in taurine breeds; however, the genetic basis of this trait in indicine cattle is still unclear. The aim of this study was to identify genomic regions associated with the hornless trait in a commercial Nelore population. A total of 107,294 animals had phenotypes recorded and 2,238 were genotyped/imputed for 777k SNP. The weighted single-step approach for genome-wide association study (WssGWAS) was used to estimate the SNP effects and variances accounted for by 1 Mb sliding SNP windows. A centromeric region of chromosome 1 with 3.11 Mb size (BTA1: 878,631–3,987,104 bp) was found to be associated with hornless in the studied population. A total of 28 protein-coding genes are mapped in this region, including the taurine *Polled* locus and the *IFNAR1*, *IFNAR2*, *IFNGR2*, *KRTAP11-1*, *MIS18A*, *OLIG1*, *OLIG2*, and *SOD1* genes, which expression can be related to the horn formation as described in literature. The functional enrichment analysis by DAVID tool revealed cytokine-cytokine receptor interaction, JAK-STAT signaling, natural killer cell mediated cytotoxicity, and osteoclast differentiation pathways as significant ($P < 0.05$). In addition, a runs of homozygosity (ROH) analysis identified a ROH island in polled animals with 2.47 Mb inside the region identified by WssGWAS. Polledness in Nelore cattle is associated with one region in the genome with 3.1 Mb size in chromosome 1. Several genes are harbored in this region, and they may act together in the determination of the polled/horned phenotype. Fine mapping the locus responsible for polled trait in Nelore breed

Genomic study for beef tenderness in a polled Nelore cattle population

Silva RMO¹; Castro LM², Peripolli E¹, Lopes FB²; Pereira ASC³; Baldi F¹, Rosa GJM⁴, Regitano LA⁶, Sainz RD⁶, Magnabosco CU^{*2}

Abstract: The aim of this study was to evaluate the impact of the use of genomic information in the traditional genetic evaluation for beef tenderness (Warner-Bratzler Shear Force - WBSF) and identify genomic regions associated with these phenotypes in a Nelore cattle population. The analyzed dataset was performed by Guaporé Agropecuária S/A – Marca OB,

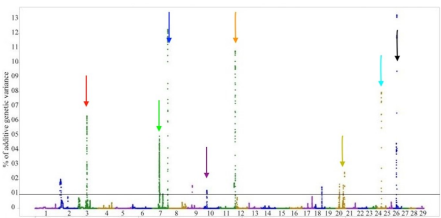


Figure 1. Manhattan plot of additive genetic variance explained (axis y) by 1 Mb size windows of chromosomes for beef tenderness.

Table 4. Prediction ability of ridge regression method using all markers available (376k) and preselected markers (5k).

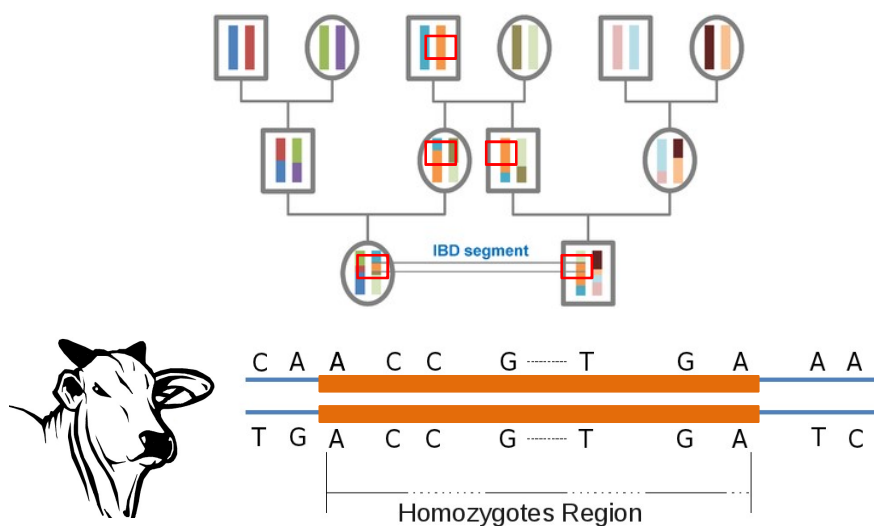
Prediction Ability	Folder*				Mean
	F1	F2	F3	F4	
RR-BLUP_376k (EBV x DGV)	0.38	0.35	0.35	0.38	0.37±0.10
RR-BLUP_376k (Yhat x DGV)	0.08	0.06	0.05	0.09	0.07±0.11
RR-BLUP_5k (EBV x DGV)	0.45	0.45	0.45	0.46	0.45±0.09
RR-BLUP_5k (Yhat x DGV)	0.26	0.27	0.26	0.28	0.27±0.09

EBV= Estimated Breeding Value obtained using the complete data set; DGV= Direct genomic value;

Fonte: Silva et al. (2017). Submitted to Plos One Journal.

Importância das variações estruturais

Segmentos em homozigose: ROH



Peripolli et al. *BMC Genomics* (2018) 19:680
<https://doi.org/10.1186/s12864-018-5060-8>

RESEARCH ARTICLE

Open Access

BMC Genomics

Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits

Elisa Peripolli^{1*}, Julia Metzger², Marcos Vinicius Antunes de Lemos¹, Nedenia Bonvino Stafuzza³, Sabrina Kluska¹, Bianca Ferreira Olivieri¹, Fabiell Louie Braga Feltosa¹, Mariana Piatto Berton¹, Fernando Brito Lopes¹, Danisio Prado Munari², Raysido Barbosa Lôbo⁴, Cláudio de Ulhoa Magnabosco⁵, Fernando Di Croce⁶, Jason Osterstock⁶, Sue Denise⁶, Angélica Simone Cravo Pereira⁷ and Fernando Baldi¹

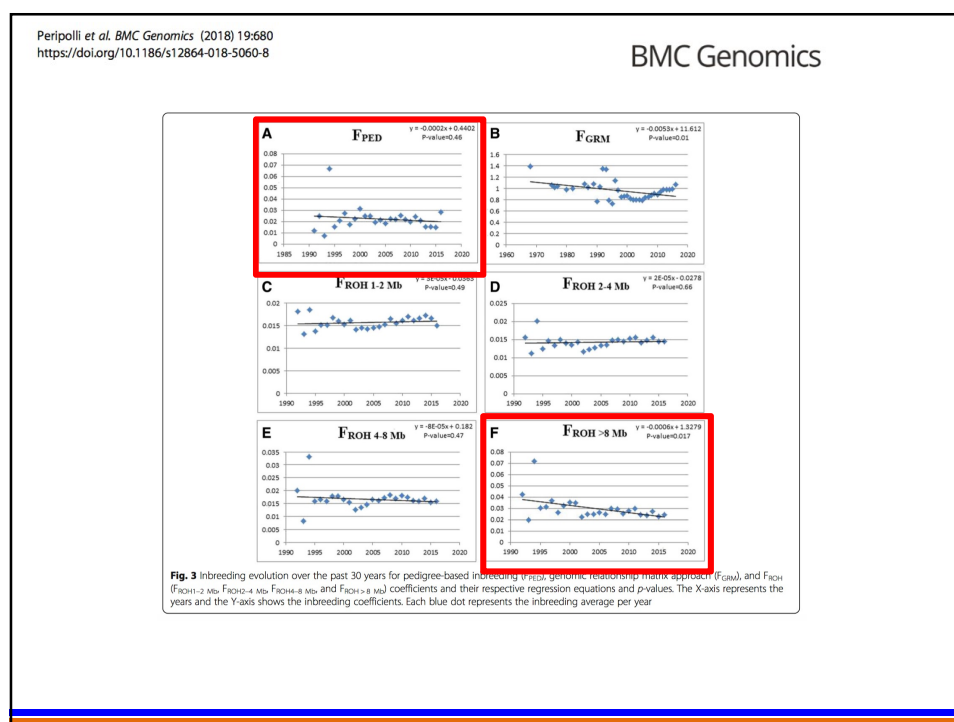
Abstract

Background: The aim of this study was to assess genome-wide autozygosity in a Nellore cattle population and to characterize ROH patterns and autozygosity islands that may have occurred due to selection within its lineages. It attempts also to compare estimates of inbreeding calculated from ROH (F_{ROH}), genomic relationship matrix (F_{GRM}), and pedigree-based coefficient (F_{PED}).

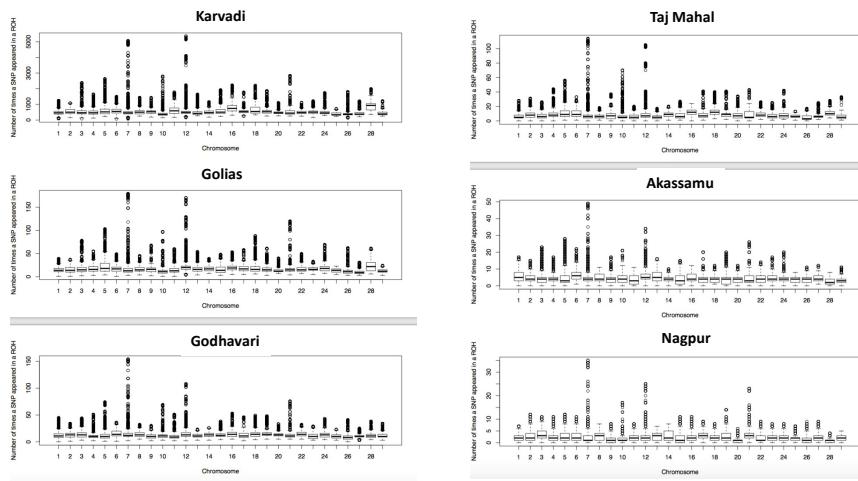
Results: The average number of ROH per animal was 55.15 ± 13.01 with an average size of 3.24 Mb. The Nellore genome is composed mostly by a high number of shorter segments accounting for 78% of all ROH, although the proportion of the genome covered by them was relatively small. The genome autozygosity proportion indicates moderate to high inbreeding levels for classical standards, with an average value of 7.15% (178.70 Mb). The average of F_{PED} and F_{ROH} and their correlations (-0.05 to 0.26) were low. Estimates of correlation between F_{GRM} - F_{PED} was zero, while the correlation (-0.01 to -0.07) between F_{GRM} - F_{ROH} decreased as a function of ROH length, except for $F_{ROH} > 8$ Mb (-0.03). Overall, inbreeding coefficients were not high for the genotyped animals. Autozygosity islands were evident across the genome ($n = 62$) and their genomic location did not largely differ within lineages. Enriched terms ($p < 0.01$) associated with defense response to bacteria (GO:0042742), immune complex reaction (GO:0045647), pregnancy-associated glycoproteins genes (GO:0030163), and organism growth (GO:0040014) were described within the autozygotic islands.

Conclusions: Low F_{PED} - F_{ROH} correlation estimates indicate that F_{PED} is not the most suitable method for capturing ancient inbreeding when the pedigree does not extend back many generations and F_{ROH} should be used instead. Enriched terms ($p < 0.01$) suggest a strong selection for immune response. Non-overlapping islands within the lineages greatly explain the mechanism underlying selection for functionally important traits in Nellore cattle.

Keywords: *Bos indicus*, Indicine, Genomic inbreeding, Gene ontology



Endogamia nas principais linhagens na raça Nelore



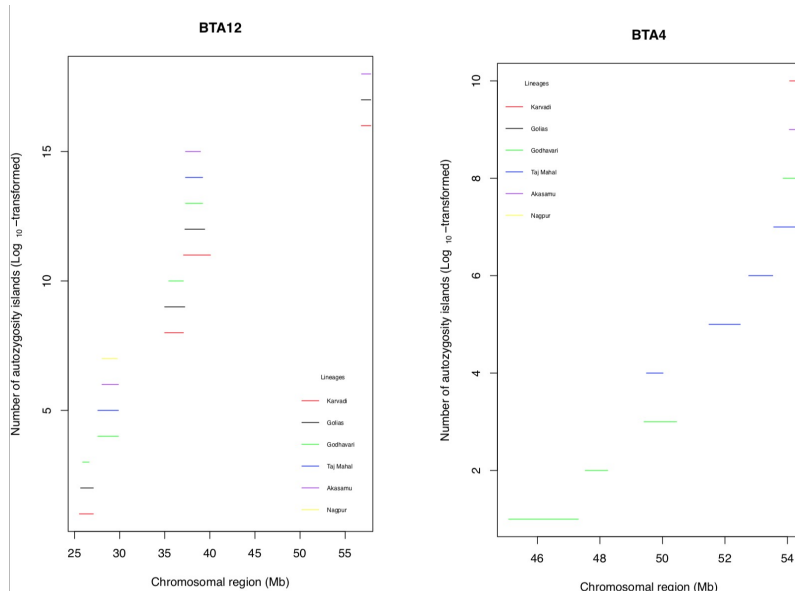
Peripolli et al. (2017).



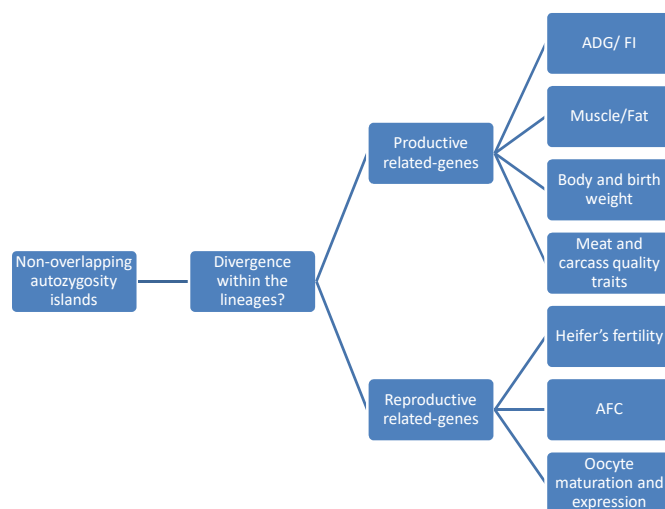
81

Peripolli et al. *BMC Genomics* (2018) 19:680
<https://doi.org/10.1186/s12864-018-5060-8>

BMC Genomics



Ilhas de endogamia em linhagens de Nelore: Evidências de seleção para características funcionalmente importantes e adaptação ambiental



83

Ilhas de endogamia em linhagens de Nelore: Evidências de seleção para características funcionalmente importantes e adaptação ambiental

Table 2. Gene Ontology (GO) terms and KEGG pathways annotation analysis enriched ($P < 0.01$) based on autozygosity islands set of genes

Terms	Genes	P-value
GO Biological Process		
(GO:0042742) Defense response to bacteria	14	7.07E-5
(GO:0030163) Protein catabolic process	9	6.33E-4
(GO:0070200) Establishment of protein localization to telomere	4	1.70E-3
(GO:0040014) Regulation of multicellular organism growth	6	2.68E-3
(GO:0045647) Negative regulation of erythrocyte differentiation	4	4.46E-3
(GO:0030901) Midbrain development	6	4.84E-3
GO Molecular Function		
(GO:0008289) Lipid binding	13	2.07E-4
(GO:0004190) Aspartic-type endopeptidase activity	9	3.24E-4
GO Cellular Component		
(GO:0005776) Autophagosome	8	3.07E-3
(GO:0005634) Nucleus	155	6.11E-3
(GO:0005815) Microtubule organizing center	10	8.36E-3
(GO:0005730) Nucleolus	41	8.50E-3
KEGG pathway		
(bta01100) Metabolic pathways	72	4.21E-4

act to protect the cell/organism

immune and inflammatory response

Immune complex reaction

84

Workshop de Melhoramento Genético Animal
 Projeto ALT-BiotechRepGen - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
 Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019



Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



 CrossMark

Copy number variation regions in Nellore cattle: Evidences of environment adaptation

Marcos Vinicius Antunes de Lemos^{a,*}, Mariana Piatto Berton^{a,1},
 Gregório Miguel Ferreira de Camargo^{b,1}, Elisa Peripolli^{a,1}, Rafael Medeiros de Oliveira Silva^{a,1},
 Bianca Ferreira Olivieri^{a,1}, Aline S.M. Cesar^{c,1}, Angélica Simone Cravo Pereira^{a,1},
 Lucia Galvão de Albuquerque^{a,1}, Henrique Nunes de Oliveira^{a,1}, Humberto Tonhati^{a,1},
 Fernando Baldi^{a,1}

^a Department of Animal Science, State University of São Paulo, FCAV – UNESP, Via de acesso Prof. Paulo Donato Castellane, Jaboticabal, São Paulo, Brazil
^b Universidade Federal da Bahia (UFBA), Departamento de Zootecnia, Salvador, Bahia, Brazil
^c Department of Animal Science, University of São Paulo, Piracicaba, São Paulo, Brazil
^d University of São Paulo, Faculdade de Medicina Veterinária e Zootecnia, Department of Nutrition and Animal Production, Pirassununga, SP, Brazil

ARTICLE INFO

Keywords:
 CNVR
 Natural selection
 Structural variation
 Nellore
 Subspeciation

ABSTRACT

The aim of the present study was to analyze the distribution of CNV regions (CNVRs) as well as to address hypothesis about the natural/artificial selection process in Nellore cattle. A total of 399,361 CNVs were identified, using the PennCNV algorithm, in 3794 Nellore cattle (*Bos taurus indicus*) genotyped with the Bovine HD BeadChip array. The default quality control was applied and 2902 samples and 195,873 CNVs remained. The medium CNV length size was 54,744 bp with a maximum of 870,000 bp and a minimum of 3000 bp. The CNVRs were generated by overlapping the 195,873 identified CNVs using the CNVRuler program. There was a higher incidence of CNVRs on BTA19 (24.26%), BTA23 (18.68%), and BTA25 (18.05%). The chromosomes that showed a lower incidence of CNVR were BTA29 (1.63%), BTA13 (9.72%), and BTA8 (9.72%). According to the type, 38.5%, 28.5% and 33.0% of the CNVRs were characterized as insertion, deletion and mixed (insertion and deletion in the same region), respectively. The 3905 CNVR estimated in the present study covered approximately 13.05% of the cattle genome (UMD3.1, 2,649,685,063 bp) and overlapped with 5495 genes. These genes have functions described as involved in biological processes that might be related to the environmental adaptation of the subspecies to tropical areas, such as regulation of vasodilatation, immune system response, hair follicle morphogenesis, among others. This study confirms the existence of large structural variations in the Nellore cattle genome and contributes to understanding the differences between cattle subspecies. Besides, it can also work as a guideline for future studies in which structural variations are present.

Cardoso et al. Genet Sel Evol (2018) 50:22
<https://doi.org/10.1186/s12711-018-0381-2>



Genetics
Selection
Evolution

RESEARCH ARTICLE

Open Access

 CrossMark

Genome-wide scan reveals population stratification and footprints of recent selection in Nellore cattle

Dierces F. Cardoso^{1*}, Lucia Galvão de Albuquerque^{1,2}, Christian Reimer³, Saber Qanbari³, Malena Erbe^{3,4},
 André V. do Nascimento¹, Guilherme C. Venturini¹, Daiane C. Becker Scalez¹, Fernando Baldi^{1,2},
 Gregório M. Ferreira de Camargo¹, Maria E. Zerlotti Mercadante^{1,5}, Joslaine N. do Santos Gonçalves Cyrillo⁵,
 Henner Simianer³ and Humberto Tonhati^{1,2}

Abstract

Background: This study aimed at (1) assessing the genomic stratification of experimental lines of Nellore cattle that have experienced different selection regimes for growth traits, and (2) identifying genomic regions that have undergone recent selection. We used a sample of 763 animals genotyped with the Illumina BovineHD BeadChip, among which 674 animals originated from two lines that are maintained under directional selection for increased yearling body weight and 89 animals from a control line that is maintained under stabilizing selection.

Results: Multidimensional analysis of the genomic dissimilarity matrix and admixture analysis revealed a substantial level of population stratification between the directional selection lines and the stabilizing selection control line. Two of the three tests used to detect selection signatures (F_{ST} , XP-EHH and iHS) revealed six candidate regions with indications of selection, which strongly indicates truly positive signals. The set of identified candidate genes included several genes with roles that are functionally related to growth metabolism, such as *COL14A1*, *CPT1C*, *CRH*, *TBC1D1*, and *XKR4*.

Conclusions: The current study identified genetic stratification that resulted from almost four decades of divergent selection in an experimental Nellore population, and highlighted autosomal genomic regions that present patterns of recent selection. Our findings provide a basis for a better understanding of the metabolic mechanism that underlies the growth traits, which are modified by selection for yearling body weight.

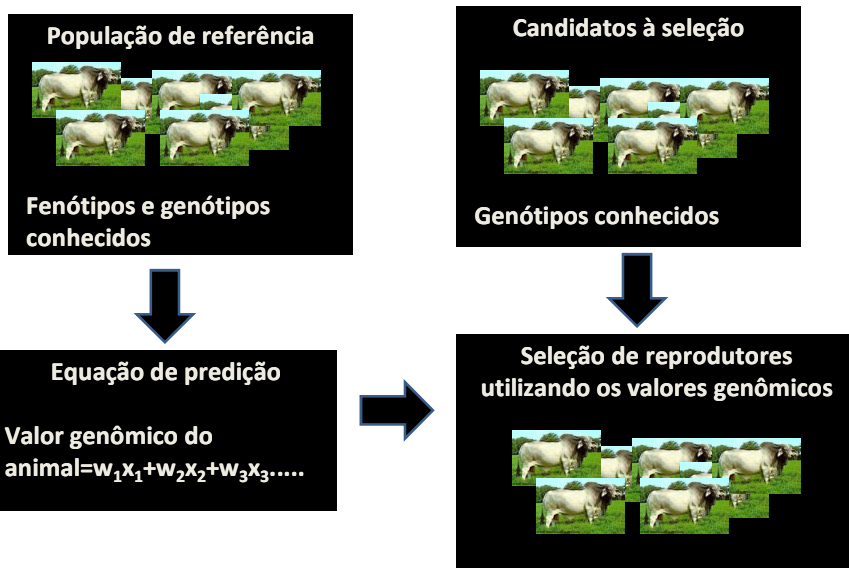
Seleção genômica

Implementação da seleção genômica conceitualmente procede em duas etapas:

1. Os efeitos dos marcadores são estimados em uma população de referência
 - Com genótipos e fenótipos
 - Equação de predição ($GEBV = b_1x_1 + b_2x_2 + \dots + b_nx_n$)
2. Predição da DEP genômica ($GEBVs$) para os animais candidatos à seleção (sem informação fenotípica)

87

Passos para implementar a seleção genômica



88

Uma questão importante é quais animais incluir na população de referência?

- Usar touros provados
- Teoricamente: a população de referência ideal deve incluir toda a gama de fenótipos e genótipos, para permitir a previsões confiáveis
- População de referência deve incluir todas as linhas dentro de raças (varias raças?)
- Importância da densidade de marcadores

89

Metodologias para seleção genômica

- Diversas metodologias têm sido propostas para a incorporação de dados de marcadores densos em modelos de regressão.
 - Métodos que explicitamente regridem registros fenotípicos sobre os marcadores através da função de regressão: $u_i = g(x_i, \beta)$
 - Métodos onde os valores genéticos são uma função dos indivíduos e utiliza a informação dos marcadores para construir uma matriz de (co)variância entre os indivíduos

90

O modelo em GS

O modelo para prever os GEBVs, considerando-se apenas os efeitos aditivos é descrito como:

$$y_i = \text{Efeitosfixos} + \text{animal}_i + \Sigma(\text{SNP}_{ijk}) + e_i$$

onde y_i pode ser um fenótipo (DEP; DEP desregredida; o media de produção de leite das filhas); Efeitos fixos e o conjuntos de efeitos fixos; animal_i e o efeito poligênico do animal; SNP_{ijk} e a somatória de todos os efeitos de SNP ($k= 1, 2$)

91

Metodologias para seleção genômica

- Várias abordagens têm sido propostas para estimar os efeitos dos marcadores (b_n)
- Uma das principais diferenças entre estas abordagens é a suposição sobre os efeitos e variâncias dos marcadores
 - Suposição mais simples é que a variância dos efeitos dos marcadores é a mesma para todos os segmentos de cromossomos

92

Metodologias para seleção genômica

Metodologias propostas para estimar os efeitos de cada marcador associado a características de importância econômica:

- Método dos quadrados mínimos
- Regressão de cumeira ou “ridge regression”
- “*Best Linear Unbiased Predictor*” – BLUP
- Inferência Bayesiana
 - BayesA, BayesB, BayesC, Lasso

93

Seleção genômica com BLUP

- BLUP = Melhor Preditor Linear não Viesado “*Best Linear Unbiased Prediction*”
- Os efeitos dos marcadores são estimados todos ao mesmo tempo

- Modelo:

$$\mathbf{y} = \mu \mathbf{1}_n + \sum_{i=1}^p \mathbf{X}_i \mathbf{g}_i + \mathbf{e}$$

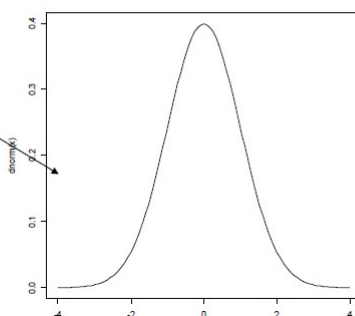
- Em BLUP assumimos a mesma variância para os efeitos dos marcadores (SNPs ou haplótipos)
- $E(\mathbf{g}) \approx N(0, \sigma_g^2)$, onde $\mathbf{g} = [g_1, g_2, g_3, \dots, g_n]$

94

Seleção genômica com BLUP

$$a_i \sim N(0, \sigma_a^2)$$

Poucos <<grandes>> efeitos



95

Seleção genômica com BLUP

- Podemos estimar os efeitos dos marcadores ou segmentos como:

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

$$\lambda = \sigma_e^2 / \sigma_a^2$$

- $$\mathbf{GEBV} = \sum_i^p \mathbf{X}_i \hat{\mathbf{g}}_i$$

96

Seleção genômica com BLUP

- As estimativas de variâncias dos marcadores com maiores variâncias tendem a ser superestimadas com BLUP
 - Diminuição na precisão da seleção genômica
- Estimativas de GEBV mais acuradas podem ser obtidas por métodos que permitem diferente variância para os efeitos dos marcadores (ou segmentos cromossômicos)

97

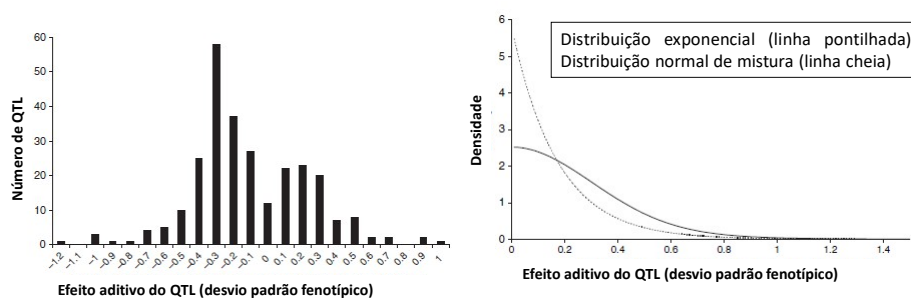
Inferência Bayesiana e seleção genômica

- BLUP assume igual variância para todos os efeitos dos marcadores
- Não considera o conhecimento prévio sobre a distribuição dos efeitos de QTL no modelo
- Com abordagem bayesiana: existem alguns segmentos do cromossomo de grande efeito (conhecimento *a priori*), e outros segmentos com efeitos moderados a pequenos ou sem efeito

98

Distribuição dos efeitos dos QTL

Distribuição dos efeitos dos QTL para 34 características a de qualidade da carne e de carcaça. Dados de mapeamento de três cruzamentos em suínos F2 .



99

J. Bennewitz & T.H.E. Meuwissen. The distribution of QTL additive and dominance effects in porcine F2 crosses. J. Anim. Breed. Genet. 127 (2010) 171–179

Distribuições *a priori* para os efeitos do marcador

Várias distribuições tem sido propostas:

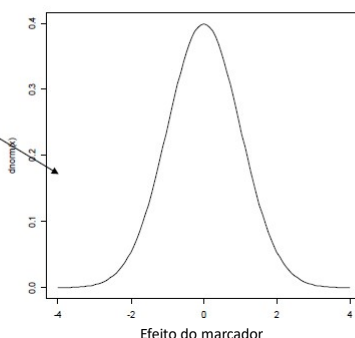
- Normal (Meuwissen et al., Genetics 2001; Van Raden JDS 2008)
- Mixture of normal e BayesC(Pi) (Van Raden JDS 2008; Habier et al., 2011)
- BayesA, BayesB (Meuwissen et al. 2001)
- Lasso (Yi, N. and S. XU, 2008 Genetics 179: 1045; De los Campos, submitted; Park and Casella, Journal of the American Statistical Association)

100

Distribuição Normal

$$a_i \sim N(0, \sigma_a^2)$$

Poucos <<grandes>> efeitos



101

Inferência Bayesiana e seleção genômica

- Método Bayesiano: é possível assumir uma distribuição *a priori* para os efeitos e variâncias dos marcadores
 - muitos têm um efeito pequeno e poucos têm um grande efeito sobre a característica
- $V(\mathbf{g}_i) = \sigma_{gi}^2$ devem ser estimados
- Trabalho de Meuwissen et al. (2001)

102

Inferência Bayesiana e seleção genômica

Copyright © 2001 by the Genetics Society of America

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes[†] and M. E. Goddard^{†,‡}

^{*}Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, [†]Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and [‡]Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
 Accepted for publication January 17, 2001

ABSTRACT

Recent advances in molecular genetic techniques will make dense marker maps available and genotyping many individuals for these markers feasible. Here we attempted to estimate the effects of ~50,000 marker haplotypes simultaneously from a limited number of phenotypic records. A genome of 1000 cM was simulated with a marker spacing of 1 cM. The markers surrounding every 1-cM region were combined into marker haplotypes. Due to finite population size ($N_e = 100$), the marker haplotypes were in linkage disequilibrium with the QTL located between the markers. Using least squares, all haplotype effects could not be estimated simultaneously. When only the biggest effects were included, they were overestimated and the accuracy of predicting genetic values of the offspring of the recorded animals was only 0.32. Best linear unbiased prediction of haplotype effects assumed equal variances associated to each 1-cM chromosomal segment, which yielded an accuracy of 0.73, although this assumption was far from true. Bayesian methods that assumed a prior distribution of the variance associated with each chromosome segment increased this accuracy to 0.85, even when the prior was not correct. It was concluded that selection on genetic values predicted from markers could substantially increase the rate of genetic gain in animals and plants, especially if combined with reproductive techniques to shorten the generation interval.

103

Inferência Bayesiana e seleção genômica

O método modelados os dados em dois níveis. O primeiro é a nível dos dados:

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}_i \mathbf{g}_i + \mathbf{e}$$

μ a distribuição *a priori* foi uniforme e não informativa

A dist. *a priori* para σ_e^2 é $\chi^2(-2, 0)$

A distribuição *a priori* para os efeitos de haplótipos ou marcadores para o segmento i é $\mathbf{g}_i \approx N(0, \sigma_{gi}^2)$

O segundo nível do modelo está nas variâncias dos efeitos dos marcadores:

A distribuição *a priori* foi uma distribuição qui-quadrado invertida:

$$Prior(\sigma_{gi}^2) \sim \chi^{-2}(v, S)$$

$v=4.012$ and $S=0.002$ Meuwissen et al. (2001)

Onde S é um parâmetro de escalar e v o número de graus de liberdade

A distribuição *a posteriori* da σ_{gi}^2 combina informações da distribuição *a priori* e informações dos dados:

$$P(\sigma_{gi}^2 | \mathbf{g}_i) \text{ é uma distribuição qui-quadrado com um escalar } S + \mathbf{g}_i' \mathbf{g}_i \text{ e } v + n_i \text{ graus de liberdade}$$

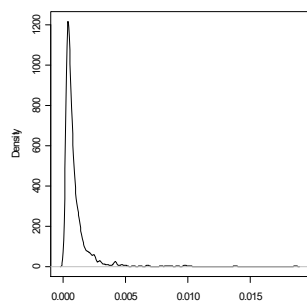
$$P(\sigma_{gi}^2 | \mathbf{g}_i) = \chi^{-2}(v + n_i, S + \mathbf{g}_i' \mathbf{g}_i) \text{ onde } n_i \text{ é o número de efeitos haplótipo no segmento } i$$

104

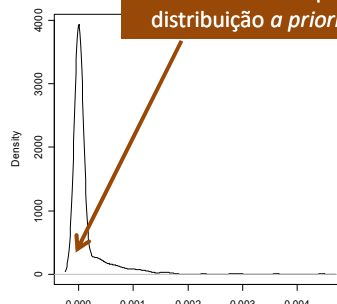
Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Diferença entre a distribuição *a priori* de variâncias dos efeitos dos marcadores utilizando o método de BayesB e o método BayesA



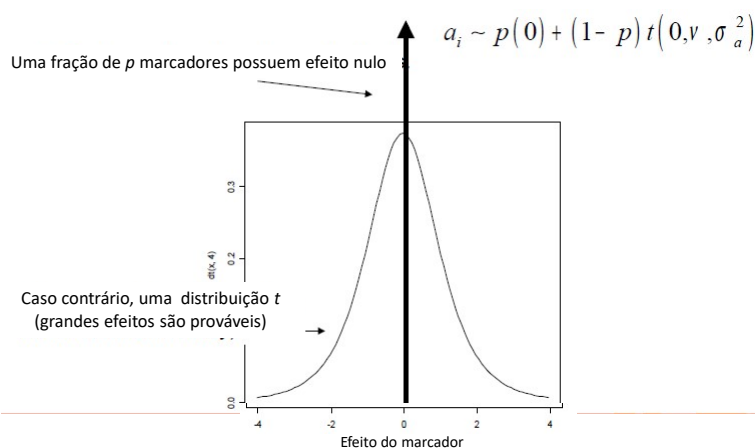
Bayes A



Bayes B

105
Meuwissen et al. (2001)

BayesB: efeito do marcador



106

Distribuição de mistura ou BayesCpi

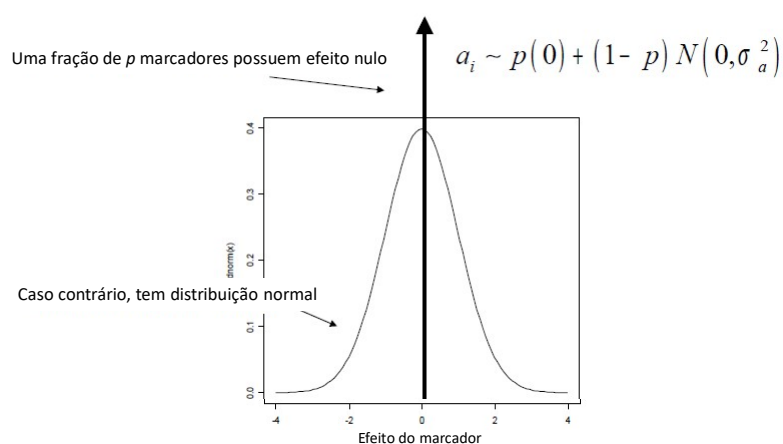
- É razoável assumir que a maioria dos loci não estão em LD com marcadores.
- BayesB: componentes de variância com valores de zero.
- Alternativa: variável indicadora (δ) informando se o marcador tem efeito (1) ou não (0)
- Modelo:

$$y_i = \text{other effects} + \sum_{j=1}^n (z_{ij} a_j \delta_j) + e_i$$

Onde $\delta_i = (0, 1)$. A distribuição de $\delta = (\delta_1, \delta_2, \dots, \delta_n)$ é binomial com probabilidade π .

107

Distribuição de mistura ou BayesCpi



108

Resumo das especificações de diferentes distribuições *a priori*

Method	Prior Distribution
Bayes BLUP	$\beta_i \sigma^2 \sim N(0, \sigma^2)$ $\sigma^2 \sim \chi^{-2}(r, s)$
Bayes A	$\beta_i \sigma_i^2 \sim N(0, \sigma_i^2)$ $\sigma_i^2 \sim \chi^{-2}(r, s)$
Bayes A/B (Hybrid)	$\beta_i \sigma_i^2 \sim N(0, \sigma_i^2)$ $\sigma_i^2 = 0$ with probability $1 - \pi$ $\sigma_i^2 \sim \chi^{-2}(r, s)$ with probability π
Bayes C	$\beta_i \gamma_i, \sigma_i^2 \sim (1 - \gamma_i)N(0, \sigma_i^2 / 100) + \gamma_i N(0, \sigma_i^2)$ $\sigma_i^2 \sim \chi^{-2}(r, s)$ $\gamma_i \sim \text{bernoulli}(\pi)$ $1 - p(\gamma_i = 0) = p(\gamma_i = 1) = \pi$

β_i is the effect for the i^{th} SNP and γ_i is the indicator variable for the i^{th} SNP.

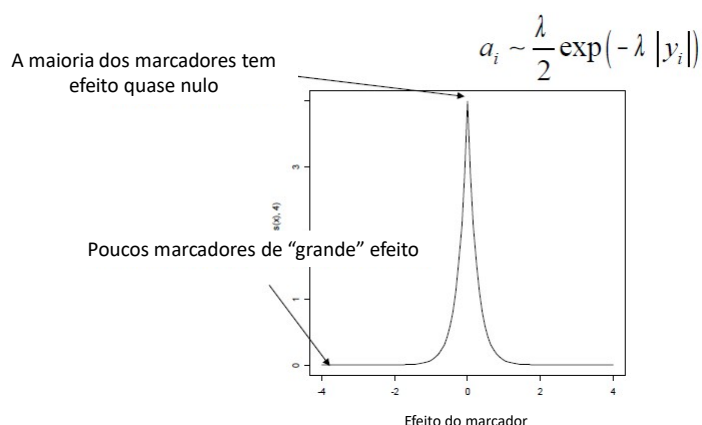
109
Verbyla et al., 2010

Bayesian LASSO

- O Método de Lasso (*Bayesian least absolute selection and shrinkage operator*)
 - Método de seleção e penalização de variáveis
- Método de um passo para estimar os efeitos dos SNPs mais importantes, reduzindo para zero as estimativas dos SNPs menos importantes
- Assume que a variância dos efeitos dos SNPs seguem uma distribuição dupla exponencial (alguns com grandes efeitos, muitos com efeitos insignificantes)

110

Bayesian LASSO



111

Bayesian LASSO

- Este método minimiza a soma de quadrados dos resíduos limitando a soma dos valores absolutos dos coeficientes de regressão
- Esse tipo especial de restrição permite que alguns coeficientes de regressão sejam exatamente zero
- Com a utilização do modelo LASSO o grau de restrição nos coeficientes de regressão (efeitos dos SNP's) é estimado a partir de informações do conjunto de dados, sem necessidade de pré-determinação do grau de restrição nos parâmetros.

112

Bayesian LASSO 2 Var

$$m | \lambda \sim \prod_i^n \frac{\lambda}{2} \exp(-\lambda |a_i|); \quad \varepsilon | \sigma_\varepsilon^2 \sim \text{MVN}(0, I\sigma_\varepsilon^2)$$

$$\text{Var}(m) = \frac{2}{\lambda^2},$$

onde MNV representa uma distribuição normal multivariada e λ é o parâmetro de “sharpness” ou parâmetro de regularização.

Usando uma formulação em termos de um modelo hierárquico incluindo um τ_{2i} componente de variância adicional associado com cada locus, temos:

$$p(m | \tau) \sim N(0, D); \quad \text{diag}(D) = (\tau_1^2 \dots \tau_n^2)$$

$$p(\tau | \lambda) = \prod_i \left(\frac{\lambda^2}{2} \right) \exp\left(-\frac{\lambda^2 \tau_i^2}{2} \right)$$

Portanto, $\text{Var}(m_i) = \sigma_{mi}^2 = \tau_{i2}^2$.

A distribuição *a priori* da variância residual foi χ^2 com 4 graus de liberdade.

113

Bayesian LASSO 2 Var

- Assume SNP effects have a different « variance »
 $y = Xb + Za + e$

$$p(a | \tau) \sim N(0, D\sigma_a^2)$$

$$p(e) \sim N(0, \sigma_e^2)$$

- This is more similar to BLUP_SNP, BayesA, BayesC, etc etc.
- In fact σ_a^2 is confounded with $1/\lambda$ and can be set to 1
 - equivalent to the Tibshirani's original Lasso
 - Therefore, the Bayesian hierarchical formulation of Tibshirani's Lasso is:

114

Bayesian LASSO 2 Var

$$y = Xb + Za + e$$

$$p(a|\tau) \sim N(0, D) \quad \Rightarrow \quad D = \begin{bmatrix} \tau_1^2 & 0 & 0 & 0 \\ 0 & \tau_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tau_n^2 \end{bmatrix} \quad \Rightarrow \quad p(\tau|\lambda) = \prod_i \frac{\lambda^2}{2} e^{-\lambda^2 \tau_i^2}$$

$$p(e) \sim N(0, I\sigma_e^2)$$

τ^2 are « variances » of
SNP effects

115

Bayesian Lasso 2 Var vs. BayesA

$$y = Xb + Za + e$$

$$p(a|\tau) \sim N(0, D) \quad \Rightarrow \quad D = \begin{bmatrix} \tau_1^2 & 0 & 0 & 0 \\ 0 & \tau_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tau_n^2 \end{bmatrix} \quad \tau^2 \text{ are « variances » of SNP effects } (\sigma_{ai}^2 \text{ in BayesA})$$

$$p(e) \sim N(0, I\sigma_e^2)$$

BL

Distribution of the variances

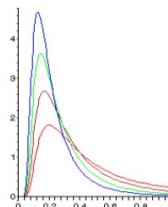
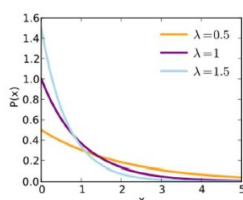
BayesA

$$p(\tau|\lambda) = \prod_i \frac{\lambda^2}{2} e^{-\lambda^2 \tau_i^2}$$

Exponential

Inverted chi-squared

$$p(\tau|\lambda) \sim \prod_i \chi_v^{-2} S_a^{-2}$$



19

116

Variação genética

- É possível estimar a variância genética na população utilizando os SNPs

– Assumindo H-W e LE (Gianola et al., 2009)

- BL2Var: $\sigma_u^2 = 2 \sum_i p_i (1 - p_i) \text{Var}(a) = 2 \sum_i p_i (1 - p_i) \frac{2}{\lambda^2}$
- BL1Var: $\sigma_u^2 = 2 \sum_i p_i (1 - p_i) \sigma_a^2 \frac{2}{\lambda^2}$
- BLUP_SNP: $\sigma_u^2 = 2 \sum_i p_i (1 - p_i) \sigma_a^2$
- BayesA: $\sigma_u^2 = 2 \sum_i p_i (1 - p_i) \hat{\sigma}_{ai}^2$
- BayesCpi: $\sigma_u^2 = 2 \sum_i p_i (1 - p_i) \sigma_a^2 (1 - \pi)$

117

Comparação da acurácia dos métodos (Meuwissen et al. 2001)

Comparação dos valores genômicos estimados versus valores genômicos verdadeiros em progênies sem registros fenotípicos (Meuwissen et al. (2001). Segmentos cromossômicos foram estimados em uma população de 2000 animais.

	$r_{\text{TBV,EBV}} + \text{SE}$	$b_{\text{TBV,EBV}} + \text{SE}$
MQ	0.318 ± 0.018	0.285 ± 0.024
BLUP	0.732 ± 0.030	0.896 ± 0.045
BayesA	0.798	0.827
BayesB	0.848 ± 0.012	0.946 ± 0.018

$r_{\text{TBV,EBV}}$: correlação entre os valores genéticos estimados e verdadeiros. $r_{\text{TBV,EBV}}$: regressão Do valor genômico estimado sobre o valor genômico estimado.

118

Comparação da acurácia dos métodos

- O método dos mínimos quadrados superestimou o efeito dos haplótipos
- Maior precisão da abordagem bayesiana
 - Muitos segmentos com efeitos próximos de zero, em BayesA, ou zero no BayesB
 - Melhor modelagem da “arquitetura genética”
- As acurácias foram elevadas, semelhantes daquelas obtidas no teste de progênie

119

Comparação dos métodos utilizados para a predição dos valores genômicos

Dados:

- **População de referência**
 - 1.098 touros da raça Holandesa progênie testada <2003
- **Validação**
 - 400 touros da raça Holandesa progênie testada > 2003
- **Fenótipos**
 - valores genéticos para produção de proteína, gordura, leite, % de proteína, % de gordura
- **Todos os touros foram genotipados para 39.048 marcadores**
- **Avaliar os métodos por:**
 - $r(\text{GEBV}, \text{ABV})$ onde ABV= Valor genético aditivo (Austrália)

120

Hayes, B. 2009. "Genomic selection in the era of the \$1000 genome sequence" In: Symposium SGLPGE Statistical Genetics of Livestock for the Post-Genomic Era.

Comparação dos métodos utilizados para a predição dos valores genômicos

Resultados $r(\text{GEBV}, \text{ABV})$:

	Proteína	Gordura	Leite	Prot%	Gord%	Média
SNP_BLUP	0.59	0.48	0.64	0.65	0.65	0.60
GBLUP	0.60	0.48	0.65	0.64	0.65	0.60
BayesA	0.55	0.48	0.61	0.64	0.71	0.60
Bayes C	0.57	0.51	0.63	0.67	0.74	0.62
Bayesian Lasso	0.56	0.47	0.62	0.62	0.64	0.58

• Interação entre o método e a característica

121

Hayes, B. 2009. "Genomic selection in the era of the \$1000 genome sequence" In: Symposium SGLPGE Statistical Genetics of Livestock for the Post-Genomic Era.

Qual é a melhor?

1. Não existe uma boa teoria de genética quantitativa para escolher um dos modelos
2. Distribuições variadas (BayesA, BayesB, exponencial dupla) e têm parâmetros complicados
- Qual *a priori* definir?
3. O métodos mais adequado é diferente dependendo da base de dados (características)
4. A distribuição normal tem um bom comportamento

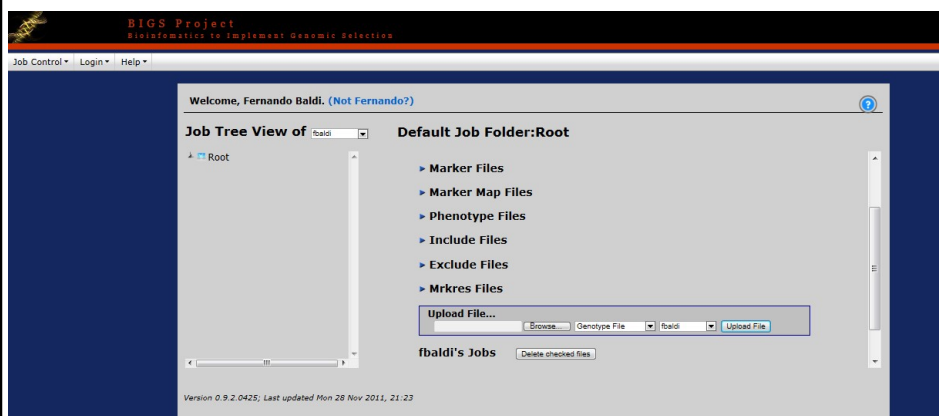
122

Softwares para seleção genômica

- **Gensel** (Fernando and Garrick 2008)
 - Gen2bin
 - Markov Chain Monte Carlo (MCMC)
 - BayesA, BayesB, BayesC, BayesCPI, BayesCCSub
 - Tipo de análise: Bayes, StepWise, Predict, generateData ou LD
 - Disponível: <http://big.ansci.iastate.edu/bigsgui/>

123

Gensel



124

Workshop de Melhoramento Genético Animal

Projeto ALT-BiotechRepGen - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Gensel

Job Name: Full_10-Aug08_041027
Input File: /
Marker File: Upload Local File (Browse...)
Marker Map File: No Map File (Browse...)
Phenotype File: Upload Local File (Browse...)
Include Type: None
Analysis Type: Bayes
Bayes Type: Bayes A
Chain Length: 4000
Burnin: 1000
Genotypic Variance: 1
Residual Variance: 1
DF for Residual Variance: 10
DF for Marker Variance: 4
Is Categorical: 1
Output Frequency: 100
Random Number Seed: 1234
Output MCMC Samples: /
Prior Distributions: /
Estimate Scale Factor: No
Output Model Sequence: /
Gensel version: gensel
Run on cluster: /
Comment:
Submit Job Cancel Save Save As
Version 0.9.2.0425, Last updated Mon 28 Nov 2011, 21:23

Softwares para seleção genômica

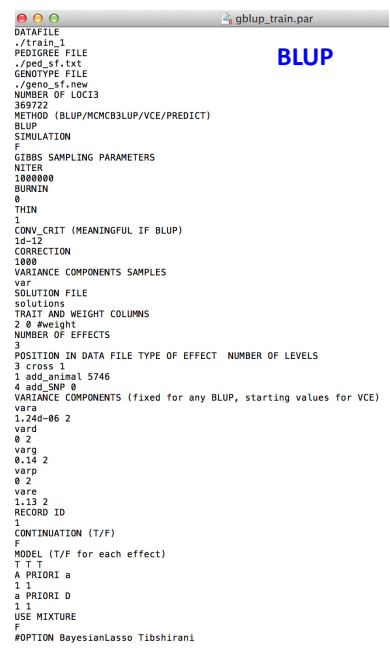
- Gebv (Sargolzaei et al., 2010)
 - Inclui G-BLUP e R-BLUP para a seleção genômica
 - Testes de paternidade
 - Imputa genótipos (*family-based haplotype reconstruction algorithm*)
 - Opção de predizer (novos genótipos)
 - Limpeza de SNPs por MAF ou baixa qualidade de genotipagem

126

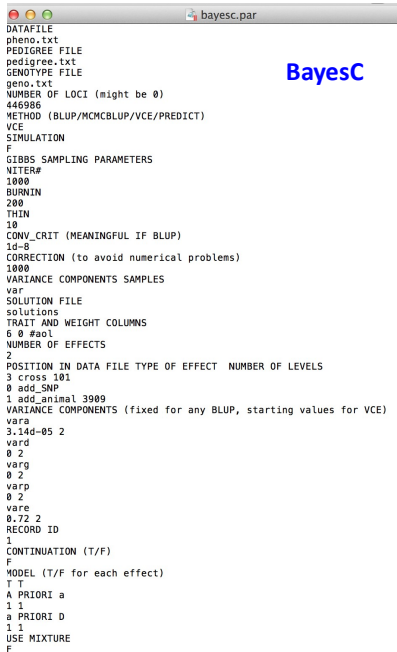
Softwares para seleção genômica

- GS3 (Legarra et al. 2011)
 - Estima os efeitos fixos e aleatórios, os valores genéticos e efeitos SNP para seleção genômica
 - Modelo linear e não linear (*threshold model*)
 - Distribuição de mista (BayesCPi), BayesC, normal (GBLUP), ou dupla exponencial (Bayesian Lasso)
 - Incluir efeitos poligênicos e ambientais, e também a inclusão de variâncias heterogêneas
 - Opções: VCE e PREDICT
 - Disponível: <http://snp.toulouse.inra.fr/~alegarra/>

127



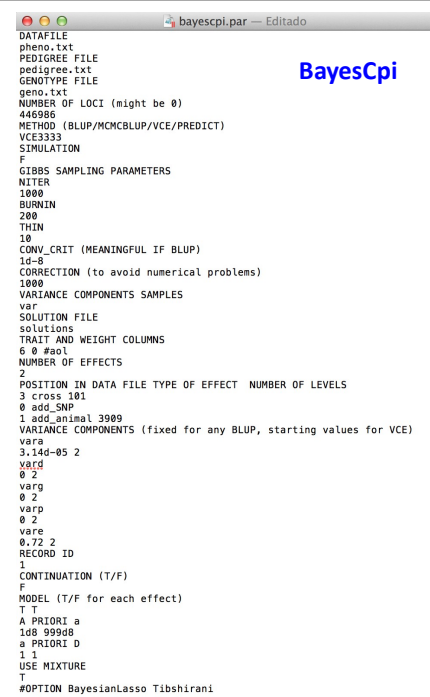
BLUP



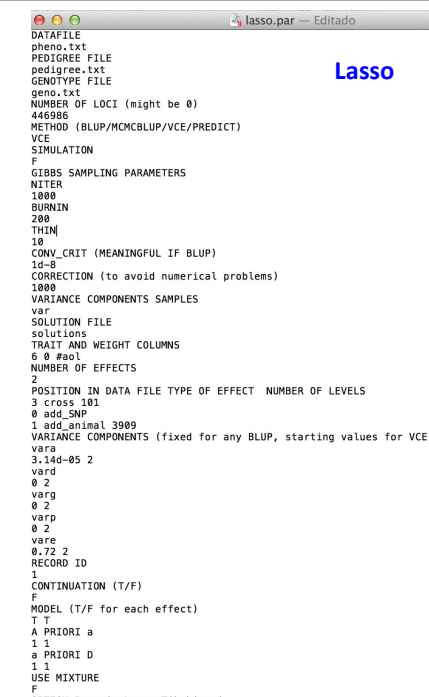
BayesC

128

Workshop de Melhoramento Genético Animal
 Projeto ALT-BiotechRepGen - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
 Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019



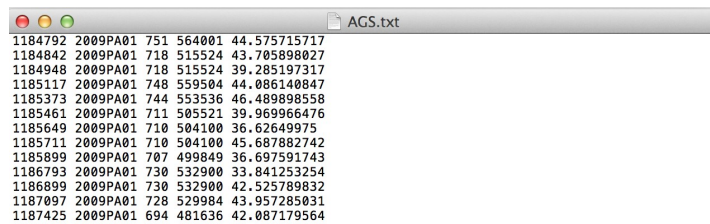
BayesCpi



Lasso

GS3

- Arquivo de Dados



ID	sire	dam	value
1184792	2009PA01	751	564001
1184842	2009PA01	718	515524
1184948	2009PA01	718	515524
1185117	2009PA01	748	559504
1185373	2009PA01	744	553536
1185461	2009PA01	711	505521
1185649	2009PA01	710	504100
1185711	2009PA01	710	504100
1185899	2009PA01	707	499849
1186793	2009PA01	730	532900
1186899	2009PA01	730	532900
1187097	2009PA01	728	529984
1187425	2009PA01	694	481636

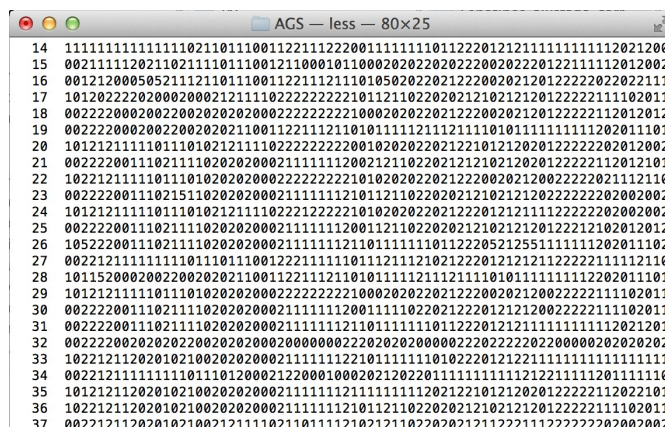
- Arquivo de pedigree



ID	sire	dam	sex
1271595	1341625	205705	1
1295341	1218561	115287	1
1757701	1761700	185220	1
1291603	1248757	285526	1
1202718	1325094	83300	1
1758117	1342407	948893	1
1801058	1363927	997515	1
1802246	0	922736	1
1802089	0	998524	1
1802517	0	1007096	1
1802563	0	923546	1
1801633	0	923528	1
1802501	1363915	1007878	1
1802880	0	1001300	1
1801240	0	921988	1
1801161	0	907574	1

GS3

- Arquivo de Genotipos



131

Abordagem semi-paramétrica

- Desvios em relação ao modelo linear
- Supostos em relação ao modelo
- Influencia das distribuições *a priori*
- Melhores resultados com efeitos não aditivos no modelo
- Maior habilidade preditiva

132

Abordagem semi-paramétrica

A informação dos marcadores é utilizada para criar uma estrutura de (co)variâncias entre os valores genômicos:

$$u_i, \text{Cov}(u_i, u_j) = K(x_i, x_j) \sigma_\alpha^2$$

onde $K(.,.)$ é alguma função positiva-definida e σ_α^2 é um parâmetro a ser estimado a partir dos dados.

133

Uma abordagem completamente diferente

- Utilização de técnicas semi-paramétricas
 - Reproducing Kernel Hilbert Spaces
 - Machine Learning
 - ...etc
 - Gianola & Gonzalez-Recio & De los Campos...

134

Seleção genômica

Fatores que afetam a precisão da seleção genômica $r(\text{GEBV}, \text{TBV})$:

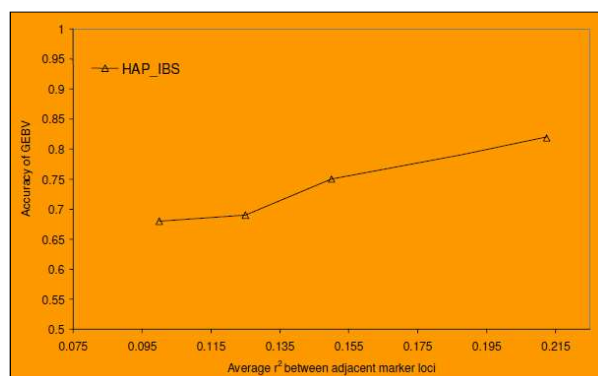
- Desequilíbrio de ligação entre QTL e os marcadores
- Tipo de Marcadores: únicos, haplótipos ou IBD
- Número de Marcadores
- Número de observações utilizadas para estimar os efeitos dos marcadores

135

Seleção genômica

Fatores que afetam a precisão da seleção genômica:

- Desequilíbrio de ligação entre QTL e marcadores:



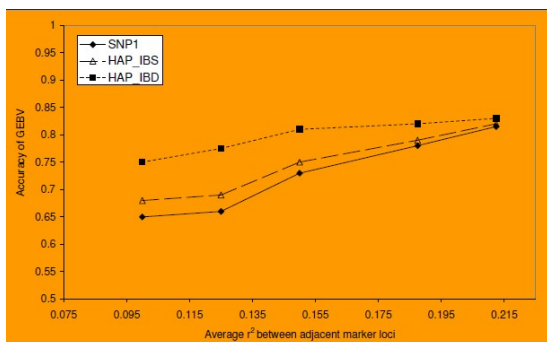
136

Calus et al. (2008)

Seleção genômica

Comparando a acurácia da seleção genômica com:

- Abordagem de genes IBD
- Haplótipos
- Marcadores únicos (SNPs)

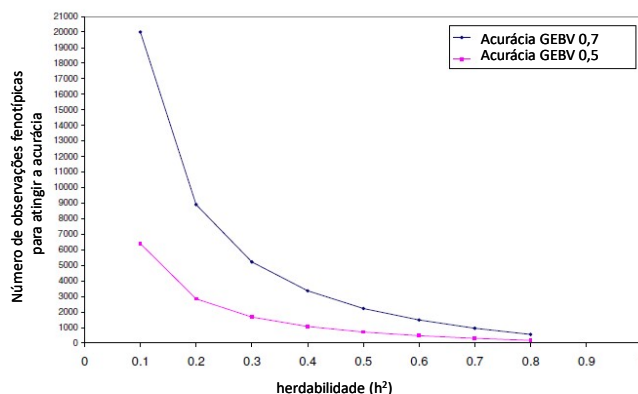


Acurácia dos valores genômicos estimados para os animais sem informações fenotípicas com três modelos diferentes de seleção genômica. Calus et al. (2008)

137

Seleção genômica

Número de observações utilizadas para estimar os efeitos dos marcadores:



138

Seleção genômica

Número de observações utilizadas para estimar os efeitos dos marcadores:

- Os efeitos dos marcadores são estimados numa população de referência
- Tamanho da população de referência

	No. Observações		
	500	1000	2200
Mínimos quadrados	0.124	0.204	0.318
Melhor preditor linear não viesado (BLUP)	0.579	0.659	0.732
BayesB	0.708	0.787	0.848

Meuwissen *et al.* (2001)¹³⁹

Impacto do tamanho da população de referência e número de marcadores

VanRaden *et al. Genetics Selection Evolution* 2011, 43:10
<http://www.gsejournal.org/content/43/1/10>

Page 9 of 11

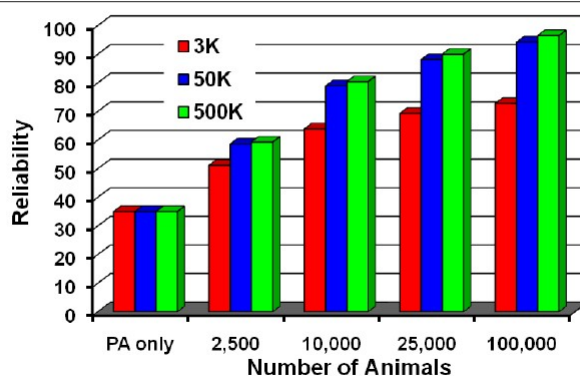


Figure 4 Expected reliabilities by number of bulls in reference population using 3,000, 50,000, or 500,000 SNP.

Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Confiabilidade de modelos HD linear, HD e 50k não linear, de touros da raça Holandesa utilizando *single-trait*

Trait	HD linear	HD non-linear	50K non-linear
Milk	64.0	65.2	65.5
Fat	66.9	68.7	68.5
Protein	59.8	60.1	60.6
Productive life	77.3	78.6	77.0
Somatic cell score	65.3	65.7	65.2
Dtr pregnancy rate	66.9	67.3	66.6
Final score	54.9	55.0	55.2
Stature	65.5	66.2	64.4
Sire calving ease	29.6	31.5	33.8
Mean (28 traits)	60.3	61.1	60.7

141
Van Randen et al. Presented at Interbull annual meeting, Cork, Ireland, May 29, 2012

Seleção genômica entre raças

142

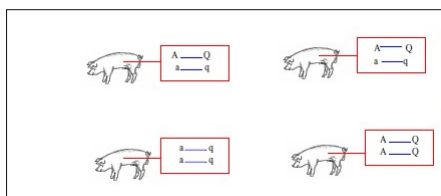
Seleção genômica entre raças

- Seleção genômica depende da fase de LD entre marcadores e QTL
- Mesma fase na população dos candidatos à seleção e na população de referência.
- Se as populações são muito diferentes, a fase entre o marcador e QTL provavelmente não seja a mesma
 - especialmente se a distância entre marcadores e QTL é relativamente grande (densidade do painel de marcadores?).

143

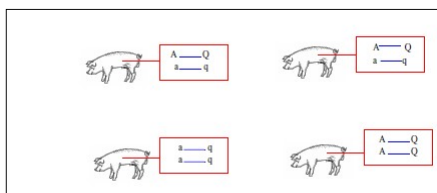
A mesma raça ou população próxima

Predizer g_i na população de referência

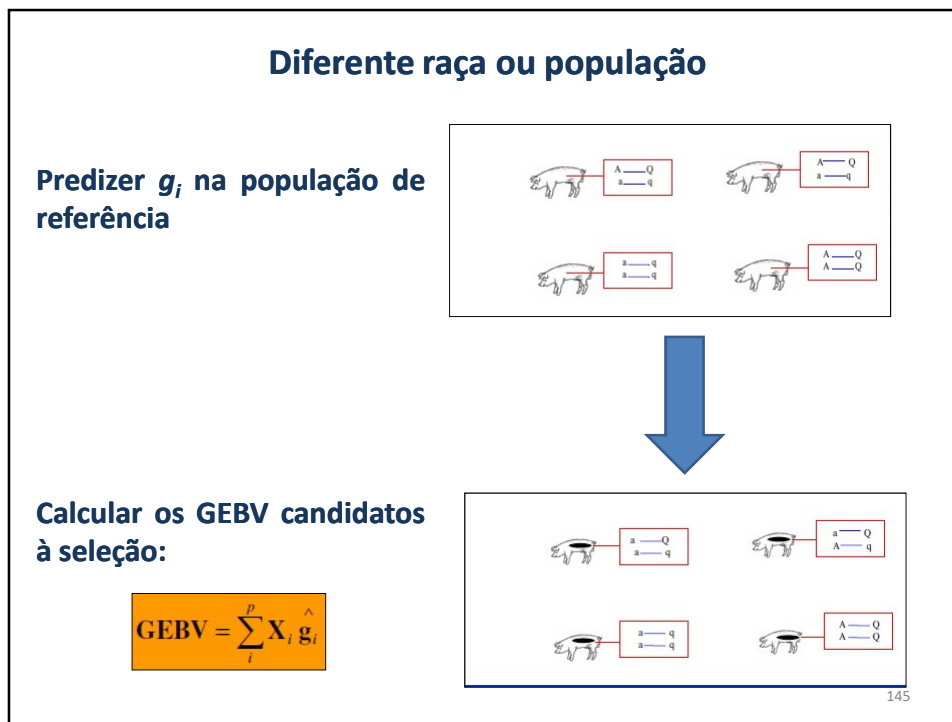


Calcular os GEBV candidatos à seleção:

$$\text{GEBV} = \sum_i^p X_i \hat{g}_i$$



144



Seleção genômica entre raças

- Correlação entre duas populações, $\text{corr}(r_1, r_2)$, para avaliar a persistência (fase) do LD nas populações
 - Se o mesmo sinal em diferentes raças (1 e 2), o alelo marcador na raça 1 é o mesmo alelo associado com alelo favorável do QTL na raça 2
- Se os efeitos dos marcadores são estimados na população 1, e a acurácia do GEBVs na população 1 é x_1 ,
- GEBVs dos animais na população 2 pode ser estimados a partir das informações da população 1 com uma acurácia $x_2 = x_1 \text{corr}^*(r_1, r_2)$

146

Seleção genômica entre raças

- Utilizar uma população de referência de várias raças, de modo que, potencialmente, todas as variantes genéticas estejam segregando dentro e entre raças
- Maior benefício para raças com menor número de indivíduos na população de referência, sempre que a fase de ligação seja a mesma

147

Research

Open Access

Accuracy of genomic breeding values in multi-breed dairy cattle populations

Ben J Hayes^{*1}, Phillip J Bowman¹, Amanda C Chamberlain¹, Klara Verbyla² and Mike E Goddard^{1,2}

Genetics Selection Evolution 2009, **41**:51 doi:10.1186/1297-9686-41-51

- EPD des-regredidos para produção de leite, proteína e gordura
- 54.001 SNP
- População de referência (progênie<2004): Holstein (n=781) Jersey (n=287)
- População de validação (progênie>2004): Holstein (n = 400) Jersey (n = 77)
- Acurácia GEBV: $r(\text{GEBV}, \text{ABV})$ (correlação)
 - ABV (Australian Breeding Value): valor genético tradicional para os touros na população de validação

148

Acurácia dos GEBV obtidos com os modelos BAYESA e BAYES_SSVS							
Reference Set	Validation set	Method	Trait				
			Protein	Fat	Milk	Prot%	Fat%
Holstein only	Holstein	BAYESA	0.47	0.44	0.59	0.59	0.71
		BAYES_SSVS	0.47	0.44	0.59	0.58	0.70
	Jersey	BAYESA	0.24	0.35	0.37	0.33	0.63
		BAYES_SSVS	0.27	0.31	0.23	0.29	0.42
Jersey only	Holstein	BAYESA	0.01	0.02	-0.02	0.05	0.17
		BAYES_SSVS	0.03	0.04	0.01	0.02	0.11
	Jersey	BAYESA	0.43	0.37	0.59	0.51	0.67
		BAYES_SSVS	0.43	0.37	0.59	0.51	0.65
Holstein and Jersey	Holstein	BAYESA	0.47	0.44	0.55	0.54	0.69
		BAYES_SSVS	0.46	0.45	0.55	0.54	0.70
	Jersey	BAYESA	0.47	0.51	0.58	0.67	0.82
		BAYES_SSVS	0.47	0.51	0.58	0.67	0.82

Genetics Selection Evolution 2009, 41:51 doi:10.1186/1297-9686-41-51

✓ População de referência de uma raça para prever o GEBV da outra raça resultou em baixas acurácias (próxima de zero).

✓ As acurácias dos GEBV aumentaram quando utilizou uma população de referência com as duas raças.

✓ Para raças com uma população de referência pequena, combinação com outras raças para formar uma população de referência multirracial é uma possibilidade.

Workshop de Melhoramento Genético Animal

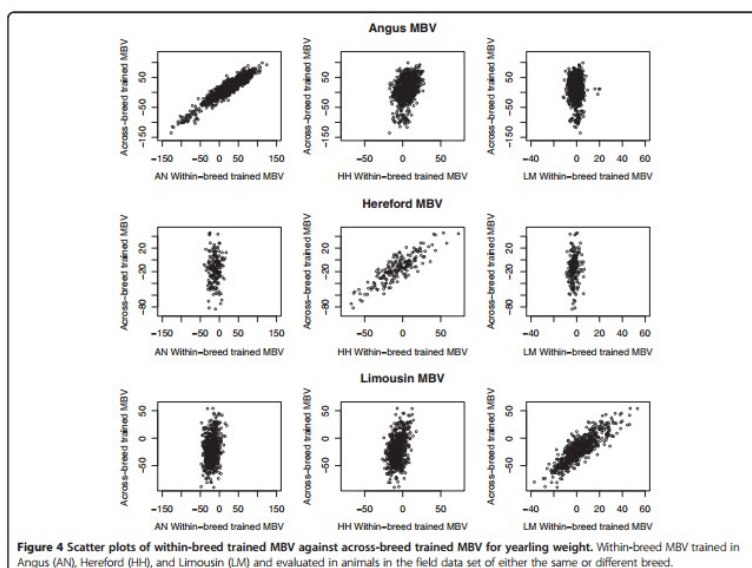
Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Estimated genetic correlations and standard errors for within-breed trained MBV for Angus, Hereford, and Limousin

Breed	Weaning weight MBV			Yearling weight MBV		
	Angus	Hereford	Limousin	Angus	Hereford	Limousin
Angus	0.36±0.07	0.14±0.08	-0.06±0.08	0.51±0.07	0.25±0.08	-0.12±0.09
Red Angus	0.16±0.16	0.09±0.16	0.25±0.16	0.08±0.18	-0.11±0.17	0.16±0.18
Charolais	-0.17±0.19	-0.06±0.19	0.35±0.19	0.09±0.18	-0.26±0.17	0.64±0.11
Gelbvieh	0.12±0.14	0.31±0.13	-0.13±0.14	0.10±0.16	0.27±0.16	0.16±0.17
Hereford	0.04±0.21	0.42±0.18	0.27±0.21	0.05±0.22	0.06±0.22	0.23±0.22
Limousin	0.02±0.09	0.23±0.09	0.40±0.08	0.06±0.09	0.17±0.09	0.28±0.08
Simmental	-0.14±0.13	0.10±0.14	0.01±0.14	-0.11±0.17	0.06±0.18	-0.36±0.16

¹Animals in the pedigree of the field data evaluation population bulls were excluded from training; genetic correlations and their standard errors are in bold characters when the MBV was evaluated in the breed in which it was trained.

Kachman et al. Genetics Selection Evolution 2013, 45:30



Kachman et al. Genetics Selection Evolution 2013, 45:30

Utilização das informações genômicas para determinar as relações de parentesco entre os animais

153

Matriz de parentesco (A)

- Matriz que descreve a relação aditiva entre os indivíduos na população:
 - descreve as covariâncias entre os indivíduos devido à variação genética aditiva
 - covariância entre os valores genéticos aditivos (u) de indivíduos X e Y ($\text{Cov}(u_x, u_y) = a_{xy} \sigma_u^2$)
- **Coeficiente de parentesco:** probabilidade de que um gene extraído aleatoriamente de X é idêntico por descendência com um gene extraído aleatoriamente de Y
- **Coeficiente de consanguinidade:** a probabilidade de que os dois alelos no Z, descendente de X e Y são idênticos por descendência

154

Matriz de parentesco genômica

- **Endogamia e parentesco:**
 - São definidas com relação a uma população-base (geralmente fundadores)
 - Onde uma relação arbitrária entre os indivíduos é definida (normalmente 0).
- A matriz de parentesco (**A**) pedigree é baseada em uma relação média que assume vários genes não ligados (desvios dos que existem na realidade).

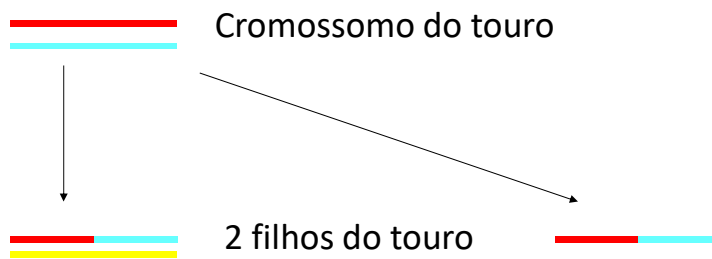
155

Matriz de parentesco genômica

- Matriz de parentesco genômica (**G**): utilizar os marcadores moleculares para determinar as relações de parentesco
 - Deduzir as relações de parentesco «verdadeiras»
 - Rastrear os alelos que cada indivíduo recebeu realmente!
- Amostragem mendeliana: durante o processo de formação das gametas resulta em variação na proporção de genes em comum entre dois indivíduos (irmãos inteiros)

156

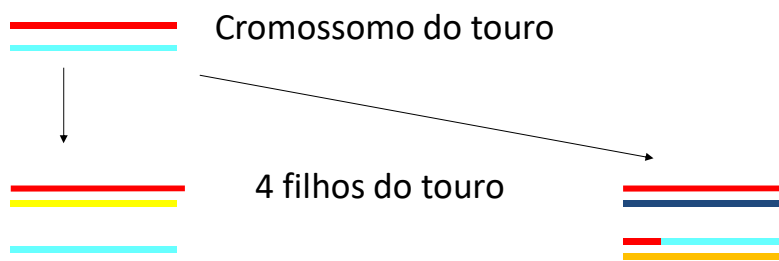
Exemplo



No modelo infinitesimal (assume-se), cada filho recebe exatamente metade dos genes do pai (mesmo grupo de genes)

157

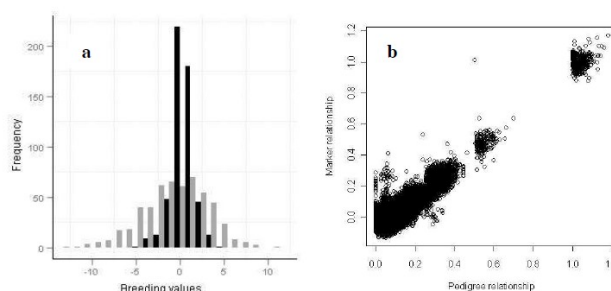
Exemplo



Na realidade, dois filhos são idênticos (recebem o mesmo grupo de genes do pai), e os outros dois são muito diferentes

158

Aplicação das informações genômicas para estimar as relações de parentesco



Histogram of direct genomic values (DGV) and estimated breeding values (EBV), (■) represents the EBV predicted by the polygenic model and (■) represents the DGV predicted by the model including the SNP information
Figure 1b: Comparison of coefficients of the polygenic model (pedigree relationship) and coefficients of the model including SNP information (marker relationship)

159
Verbyla et al. (2010) WCGALP

Matriz de parentesco genômica

- **Em muitas populações de animais a matriz de parentesco pode não estar disponível, ou ser incompleta ou conter erros**
 - Reprodutores múltiplos
- **Construção da matriz de parentesco aditivo utilizando as informações dos marcadores**
 - Animais fundadores podem ser relacionados entre si (compartem genes)
- **Combinações entre matriz de parentesco tradicional e genômica**
 - Não todos os animais são genotipados

160

Matriz de parentesco genômica

- Os marcadores (SNP) devem capturar informações de parentesco não contidas no pedigree tradicional (valores probabilísticos)
- Diminuir o viés e aumentar a acurácia das estimativas dos parâmetros genéticos
- Diminuir ΔF

161

Seleção Genômica: modelos

- Modelo de efeitos do marcador(segmento):

$$y = Xb + \sum_{i=1}^m Z_i g_i + e \quad (1)$$

- Modelo de efeito genético total:

$$y = Xb + Zu + e \quad (2)$$

Pressuposição sobre a arquitetura genética da característica de interesse	Modelo 1	Modelo 2
Modelo Infinitesimal	RR-BLUP	G-BLUP
Modelo Não Infinitesimal	BayesA, B, C,...	TA-BLUP

162

Matriz de parentesco no MME

$$y = Xb + Zu + e$$

$$Vu = \begin{cases} A \sigma_u^2 & \rightarrow \text{BLUP tradicional} & \leftarrow \text{Pedigree} \\ G \sigma_u^2 & \rightarrow \text{G-BLUP} & \leftarrow \text{Marcadores} \\ T_a \sigma_u^2 & \rightarrow \text{TA-BLUP} & \leftarrow \begin{matrix} \text{Marcadores} \\ + \\ \text{Ponderados} \end{matrix} \end{cases}$$

A: Matriz de parentesco (coeficientes de parentesco)

G: Matriz de parentesco (marcadores)

TA: Matriz de parentesco específica para a característica (marcadores)

163

Três vias para obter G

- **M** é a matriz ($n \times m$) que especifica quais alelos do marcador (m) cada indivíduo (n) possui.
- Os elementos de **M** são -1, 0, 1
- **MM'** a diagonal conta número de locos homozigóticos para cada indivíduo, e fora da diagonal mede o número de alelos compartilhados por indivíduos parentes
- Seja a frequência (p_i) do segundo alelo do locus i
- **P** especifica a frequência como a diferença de 0,5 multiplicado por 2. $P = 2(p_i - 0,5)$
- **Z = M - P** (define os valores médios dos efeitos do alelo para 0)
- As frequências alélicas em **P** devem ser da população-base não selecionada; e não as frequências que ocorrem após a seleção ou a endogamia

VanRaden¹⁶⁴, 2008

Requisitos para Cálculo matriz G

Exemplo

M - matriz de marcadores

- dimensões n (número de indivíduos) por m (número de loci)
- elementos para definir a homozigoto, heterozigoto e homozigoto (-1, 0, 1)

MM' - Diagonal - conta o número de locos homozigotos para cada indivíduo

- Fora da diagonal – mede o número de alelos compartilhados por parentes

M'M - Diagonal - conta o número de indivíduos homozigotos para cada locus

- Fora da diagonal - medir o número de vezes que os alelos em loci diferentes foram herdados pelo mesmo indivíduo

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \end{bmatrix}$$

$$\mathbf{MM}' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 3 \end{bmatrix}$$

$$\mathbf{M}'\mathbf{M} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -2 \\ 1 & 1 & -1 \\ -2 & -1 & 2 \end{bmatrix}$$

165

$$\mathbf{P} = 2(p_i - 0.5)$$

$$p_1 = 0.833, \quad p_2 = 0.667, \quad p_3 = 0.167$$

$$\mathbf{P} = \begin{bmatrix} 0.666 & 0.334 & -0.666 \\ 0.666 & 0.334 & -0.666 \\ 0.666 & 0.334 & -0.666 \end{bmatrix}$$

$$\mathbf{Z} = \mathbf{M} - \mathbf{P} = \begin{bmatrix} 0.334 & -0.334 & -0.334 \\ -0.666 & -0.334 & 0.666 \\ 0.334 & 0.666 & -0.334 \end{bmatrix}$$

166

Três vias para obter **G**

1. $\mathbf{G} = \frac{\mathbf{ZZ}'}{2 \sum p_i(1-p_i)}$

variação média

G escala análoga **A** (Wright, 1922)

$F_{ij} = G_{ij} - 1$
2. $\mathbf{G} = \mathbf{ZDZ}'$ onde **D** é diagonal com $D_{ii} = \frac{1}{m[2p_i(1-p_i)]}$

Cada marcador é ponderado pela sua variância, ou seja, **D** é uma matriz com diferentes pesos para cada SNP. Esta fórmula foi proposta para estudos de genética humana.
3. $\mathbf{MM}' = g_0 \mathbf{11}' + g_1 \mathbf{A} + \mathbf{E}$, Para a obtenção de **G** não necessita de frequências alélicas. Ajusta para homozigose média regredindo **MM'** em **A** para obter **G**.
 onde g_0 é a interceptação e g_1 é o coef. de regressão.

$$\mathbf{G} = \frac{\mathbf{MM}' - g_0(\mathbf{11}')}{g_1} \quad \longrightarrow \quad \begin{bmatrix} n^2 & \sum_k A_{jk} \\ \sum_k A_{jk} & \sum_k \sum_k A_{jk}^2 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \end{bmatrix} = \begin{bmatrix} \sum_k (MM')_{jk} \\ \sum_k \sum_k (MM')_{jk} A_{jk} \end{bmatrix}$$

167

Três vias para obter **G**

- Matriz **G** é positiva semi-definida com os 2 primeiros métodos, mas pode ser positiva não definida se o número de locos são limitados ou se dois indivíduos têm genótipos idênticos
- **G** é singular quando $m < n$
- Para evitar problemas potenciais **G** pode ser ponderada:

$$\mathbf{G} = 0.99 \frac{\mathbf{ZZ}'}{2 \sum p_i(1-p_i)} + 0.01 \mathbf{A}$$

168

Propriedades de G

- Em H-W, equilíbrio de ligação:
 - Média de Diag (G) = 1 $G = ZZ' / 2 \sum_{all\ SNPs} p_i (1 - p_i)$
 - Média fora da diagonal (G) = 0
 - Valor genético médio de indivíduos genotipados = 0
 - Isto corresponde à definição da população de base
- Com F endogamia média
 - Média de Diag (G) = 1 + F

	AA	Aa	aa
freq	$q^2 + pqF$	$2pq(1-F)$	$p^2 + pqF$

169

GBLUP

- GBLUP é um «BLUP» construído com G assim definido:

– A para G:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}W \\ W'R^{-1}X & W'R^{-1}W + G\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ W'R^{-1}y \end{bmatrix}$$

$$G = \frac{ZZ'}{2 \sum p_i (1 - p_i)}$$

- Como em BLUP, podemos incluir os animais com o genótipo, mas sem fenótipo

170

GBLUP

- GBLUP dá resultados idênticos ao BLUP_SNP se é usada a mesma variância:

$$\sigma_u^2 = 2 \sum_{\text{all SNPs}} p_i (1 - p_i) \sigma_a^2$$

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2 \sum p_i (1 - p_i)}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{I}\sigma_a^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

$\hat{\mathbf{g}}_{\text{de GBLUP}} = \mathbf{Z} \hat{\mathbf{a}}_{\text{de BLUP_SNP}}$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{I} \\ \mathbf{I}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{I}'\mathbf{R}^{-1}\mathbf{I} + \mathbf{G}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{I}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

171

GBLUP

- Podemos passar de GBLUP para BLUP_SNP:

$$\hat{\mathbf{g}} = \mathbf{Z}\hat{\mathbf{a}}$$

GEBV a partir de SNP

Efeitos a partir de SNP GEBV de (Henderson, 1973; Strandén and Garrick, 2009)

$$\hat{\mathbf{a}} = \mathbf{I}\sigma_a^2 \mathbf{Z}'\mathbf{G}^{-1}\sigma_u^{-2} \hat{\mathbf{g}}$$

172

GBLUP

- **Vantagens:**
 - Se ajusta muito bem em softwares já existentes com BLUP
 - ... e na teoria existente (REML, características múltiplas ... *Single Step*)
 - Fornece medidas de acurácia a partir da inversa do LHS
 - Utiliza todos os animais
- **Desvantagens:**
 - Não pode ajustar genes principais (utilização de pesos na construção de G)
 - Dificuldades no cálculo de G e de inversão

173

A variância genética utilizando a matriz G ou A é a mesma?

- Se temos infinita quantidade de dados e executamos um G-BLUP, obtemos a mesma herdabilidade que aplicando A-BLUP?
- Em **A**, a base genética é uma população em equilíbrio H-W e possui as mesmas frequências alélicas como aqueles usadas para calcular **G**
 - Se usamos animais selecionados para genotipar, possivelmente esta pressuposição falsa.
 - Utilizar frequências alélicas observadas ou $p_i=0.5$
 - Apenas importante se queremos misturar as fontes de informação (pedigree+genômica)

174

A variância genética utilizando a matriz G ou A é a mesma?

- **É necessário utilizar a mesma população base:**
 - Mesma restrição sobre os valores genéticos (valor genético médio da base = 0)
 - Mesma variação genética
- **Estimar as frequências alélicas na população base**
 - Difícil ou impossível de obter
 - Várias alternativas

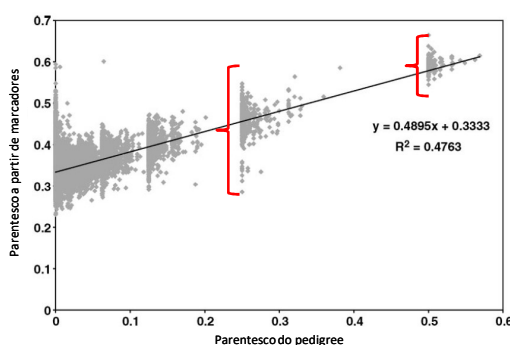
175

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Technical note: Prediction of breeding values using marker-derived relationship matrices

B. J. Hayes and M. E. Goddard
 J. Anim. Sci. 2008. 96:2089–2091



Coefficientes de parentesco aditivo estimados a partir das informações de marcadores vs. coeficientes de parentesco aditivo obtidos a partir do pedigree em animais Angus

Conjunto de 9323 SNP distribuídos aleatoriamente por todo o genoma em 379 bovinos da raça Angus

Animais selecionados para alta e baixo consumo alimentar residual

Um modelo aditivo foi ajustado com a espessura característica da gordura no P8 (mm):

$$\text{fatp8}_{ij} = \mu + \text{grupo contemporâneo}_i + a_j + e_{ij}$$

μ é média, a_j é o efeito poligênico aditivo do animal j , Sendo que $a_j \sim N(0, A, \sigma_a^2)$, A a matriz de parentesco aditiva (obtida a partir dos marcadores (GNRM) ou do pedigree (NRM))

h^2 P8 (NRM): 0,38
 h^2 P8 (GRM): 0,57

176

Aplicação de G

- Seleção Genômica (GBLUP)
- Estimativa de parâmetros genômicos (GREML)
 - Em populações sem informações de pedigree
 - Magnitude da variação explicada pelos marcadores
- Melhorar o modelo de análise de associação de marcadores
 - Estrutura da população

177

Comparison of predicted accuracy using genomic relationship matrices and true accuracy from correlation between true BV and EBV

h^2	Predicted accuracy		True accuracy
	G_m	$\hat{G}$	
0.2	0.52	0.42	0.43
0.5	0.68	0.60	0.59
0.9	0.87	0.70	0.73

¹Number of training records is $T = 500$

$$G_m = XX' / M \rightarrow M = \sum 2p_j(1 - p_j)$$

$$\hat{G} = A + b(G_m - A) \rightarrow \text{where } b \text{ is the regression of elements of } G - A \text{ on elements of } G_m - A$$

Goddard et al., J. Anim. Breed. Genet. 128 (2011) 409–421 ¹⁷⁸

Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Em definitiva, existem várias situações em que as relações de parentesco podem ser derivadas das informações do parentesco.

- Para características que não podem ser medidas nos candidatos à seleção
- Reprodutores múltiplos
- Estudos de associação genômica (marcadores)
- Melhorar a matriz A

179

Frequência com que devem ser re-estimados os efeitos dos marcadores

180

Frequência com que devem ser re-estimados os efeitos dos marcadores

- Se os marcadores utilizados foram realmente as mutações subjacentes que causam o efeito de QTL, as estimativas dos efeitos dos marcadores seriam “perfeitas”.
- GEBVs para todas as gerações posteriores podem ser estimados a partir dos marcadores
- Na prática, haverá marcadores com níveis baixos a moderados de r^2 com as mutações que causam o efeito de QTL.

181

Frequência com que devem ser re-estimados os efeitos dos marcadores

- A recombinação entre os marcadores e o QTL irá reduzir a acurácia da GEBV
- Os marcadores não capturam toda variância do QTL
- Cada quanto tempo os efeitos do marcadores devem ser re-estimados?

182

Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

Frequência com que devem ser re-estimados os efeitos dos marcadores

Correlação entre os valores genéticos estimados (EBV) e verdadeiro (TBV) nas gerações 1003-1008, onde os valores genéticos estimados foram obtidos a partir das estimativas do marcador nas gerações 1001 e 1002 (BayesB).

Geração	$r_{TBV;EBV}$
1003	0.848
1004	0.804
1005	0.768
1006	0.758
1007	0.734
1008	0.718

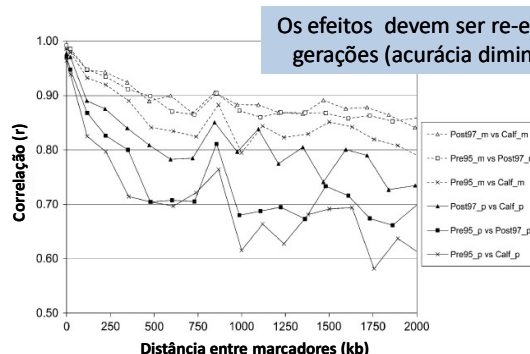
As gerações 1004-1008 foram obtida da mesma forma como a partir da 1003 e as gerações parental.

183
Meuwissen et al. (2001)

Linkage Disequilibrium and Persistence of Phase in Holstein-Friesian, Jersey and Angus Cattle

A. P. W. de Roos,^{*,1} B. J. Hayes,[†] R. J. Spelman[‡] and M. E. Goddard^{†,§}

^{*}CRV, 6800 AL Arnhem, The Netherlands, [†]Animal Genetics and Genomics, Primary Industries Research Victoria, Attwood 3049, Australia, [‡]Livestock Improvement Corporation, Hamilton, New Zealand and [§]Faculty of Land and Food Resources, University of Melbourne, Parkville 3052, Australia



Correlação (r) entre os grupos de animais Holstein-Friesian, em função da distância entre marcadores, para touros nascidos antes de 1995 (Pre95), touros nascidos após 1997 (Post97), e bezerros nascidos em 2006 (Calf)

184

Seleção genômica com baixo LD entre os marcadores

185

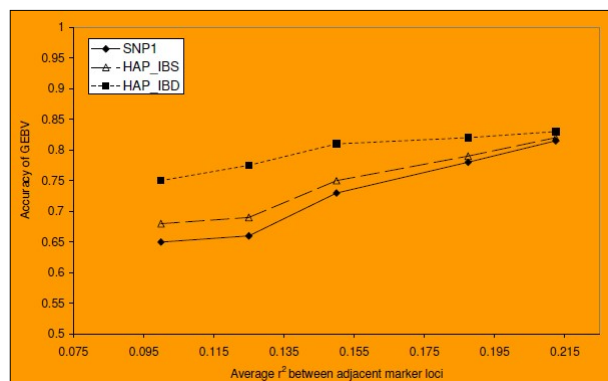
Seleção genômica com marcadores de densidade baixa

- Densidade de marcadores no genoma suficiente para garantir $r^2 \geq 0,2$.
- Não captura toda a variância genética com os marcadores
- Duas estratégias
 - Explorar a ligação (dentro de famílias), bem como o desequilíbrio de ligação, utilizando a abordagem IBD

186

Workshop de Melhoramento Genético Animal

Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019



Calus et al. (2007)¹⁸⁷

Seleção genômica com marcadores de densidade baixa

- Densidade de marcadores no genoma suficiente para garantir $r^2 \geq 0,2$.
- Não captura toda a variância genética com os marcadores
- Duas estratégias
 - Explorar a ligação, bem como o desequilíbrio de ligação, utilizando a abordagem IBD
 - Incluir o efeito poligênico para capturar as variações genéticas não capturadas pelos marcadores (explorar pedigree)

$$\text{GEBV} = \hat{u} + \sum_i^p \mathbf{X}_i \hat{\mathbf{g}}_i$$

188

Workshop de Melhoramento Genético Animal

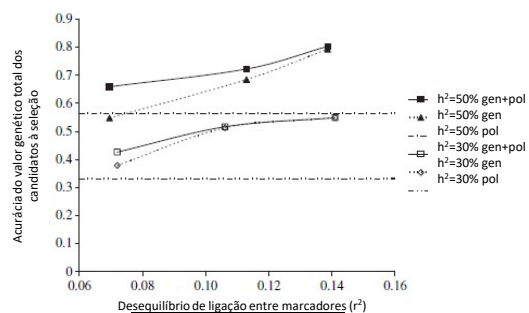
Projeto ALT-Biotech^{RepGen} - Recursos Genéticos Animais e Biotecnologias: projeção para o futuro
Estação Zootécnica Nacional – Fonte Boa, 17 de Dezembro de 2019

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM

M.P.L. Calus & R.F. Veerkamp



Acurácias dos valores genéticos dos candidatos à seleção para características com herdabilidades de 50 e 10%, com estimativa de valor genômico incluindo (gen+pol) ou excluindo efeitos poligênicos (gen), ou utilizando estimativa do valor genético, considerando apenas o efeito poligênico (pol).

189

M.P.L. Calus & R.F. Veerkamp. J. Anim. Breed. Genet. 124 (2007) 362–368