e-ISSN 2764-2488

Documentos

197

Campinas, SP / Setembro, 2025

Utilização do software brblup para avaliação genética

Estudos de caso com as raças Canchim e Brangus

Embrapa
Agricultura Digital

Empresa Brasileira de Pesquisa Agropecuária
Embrapa Agricultura Digital
Ministério da Agricultura e Pecuária

e-ISSN 2764-2488

Documentos 197

Setembro, 2025

Utilização do software brblup para avaliação genética

Estudos de caso com as raças Canchim e Brangus

Roberto Hiroshi Higa
Ivan Carvalho Filho
Marcos Jun-Iti Yokoo

Embrapa Agricultura Digital
Campinas, SP
2025

Embrapa Agricultura Digital
Av. Dr. André Tosello, 209 - Cidade Universitária
Campinas, SP, 13083-886
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações
Presidente
Júlio Cesar Dalla Mora Esquerdo
Secretário-executivo
Sonia Ternes
Membros
Adauto Luiz Mancini, Alan Massaru Nakai, Carla Cristiane Osawa, Geraldo Magela de Almeida Cançado, Graziella Galinari, Joice Machado Bariani, Juliana Erika de Carvalho Teixeira Yassitepe, Luiz Manoel Silva Cunha, Magda Cruciol, Paula Regina Kuser Falcão

Revisão de texto
Graziella Galinari

Projeto gráfico
Leandro Sousa Fazio

Diagramação
Giulia Mizuno

Foto da capa
Danilo de Paula Moreira
Keke Barcellos

Publicação digital: PDF

Todos os direitos reservados
A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)
Embrapa Agricultura Digital

Higa, Roberto Hiroshi

Utilização do software brblup para avaliação genética: estudos de caso com as raças Canchim e Brangus / Roberto Hiroshi Higa, Ivan Carvalho Filho, Marcos Jun-Iti Yokoo. – Campinas : Embrapa Agricultura Digital, 2025
PDF (24 p.) : il. color. - (Documentos / Embrapa Agricultura Digital, ISSN 2764-2488 ; 197).

1. Software. 2. Software brblup. 3. Avaliação genética. I. Carvalho Filho, II. Yokoo, M. J. I. III. Título. IV. Embrapa Agricultura Digital. V. Série.

CDD (21. ed.) 005.1068

Carla Cristiane Osawa (CRB-8/10421)

© 2025 Embrapa

Autores

Roberto Hiroshi Higa
Engenheiro Eletricista, doutor em Engenharia Elétrica, pesquisador da Embrapa Agricultura Digital, Campinas, SP

Ivan Carvalho Filho
Zootecnista, doutor em Genética e Melhoramento Animal, geneticista da Geneplus Consultoria, Campo Grande, MS

Marcos Jun-Iti Yokoo
Zootecnista, doutor em Genética e Melhoramento Animal, pesquisador da Embrapa Pecuária Sudeste, São Carlos, SP

Apresentação

Programas de melhoramento genético animal têm por objetivo promover o aumento de alelos desejáveis em uma população, identificando os indivíduos superiores para características de interesse da cadeia produtiva, particularmente aquelas de importância econômica. Atualmente, são poucos os softwares de avaliação genética utilizando a genômica que apoiam esse objetivo, servindo como ferramentas na pesquisa e desenvolvimento em avaliação genética e melhoramento animal. O software brblup, desenvolvido pela Embrapa Agricultura Digital, está inserido nesse contexto, porém, utiliza uma plataforma de desenvolvimento (Python) diferente daquela utilizada pelos poucos softwares já estabelecidos nesta área (Fortran) e com o qual seu público-alvo está habituado. Diante desse contexto, esse documento apresenta uma demonstração da capacidade do brblup ao realizar avaliações genéticas de populações de animais vinculados a programas de melhoramento genético reais, utilizando como estudo de caso as populações de animais vinculadas aos programas de melhoramento Geneplus/Canchim e Brangus+.

Stanley Robson de Medeiros Oliveira
Chefe-Geral da Embrapa Agricultura Digital

Sumário

Introdução 9

Material e método 11

Resultados 15

Comentários finais 21

Referências 23

Introdução

Softwares de avaliação genética apoiam os programas de melhoramento em suas rotinas de avaliação, também servindo como ferramentas na pesquisa e desenvolvimento em genética e melhoramento. Sua funcionalidade principal consiste em utilizar dados de animais com genotipagem, fenotipagem e a genealogia completa da população de animais em avaliação para, com base em um modelo genético-estatístico pré-definido, calcular a DEPG[1] (diferença esperada na progênie aprimorada pela genômica) para os animais dessa população. A maioria dos softwares estabelecidos no mercado contento esta funcionalidade, tais como o Blupf90 (Misztal et al., 2014), o DMU (Madsen et al., 2014), o MiX99 (Vuori et al., 2006) e o MixBlup (Vandenplas et al., 2022), baseiam-se em linguagens tradicionais na comunidade de computação científica, principalmente o Fortran.

O software brblup (Higa, 2025), desenvolvido pela Embrapa Agricultura Digital, é escrito em linguagem Python[2] e integra um pacote denominado brbv, que disponibiliza um conjunto de interfaces por linha-de-comando de softwares para suporte às atividades de avaliação genética de animais. Essa escolha da linguagem de programação baseou-se no fato de Python ser orientada a objetos, moderna, simples e flexível, o que facilita bastante o seu desenvolvimento, bem como futuras evoluções. Outro fator importante que subsidiou essa escolha foi a existência de um robusto ecossistema de ferramentas associadas, tais como Jupyter[3] para análise iterativa, NumPy[4] e SciPy[5]

---

(1) A DEP refere-se à diferença esperada na progênie para descrever o valor genético dos animais. No presente documento, será utilizado a EBV (valor genético estimado) que refere-se a candidatos para seleção. Mantidas inalteradas todas as condições, DEP = EBV/2.

(2) Disponível em: https://www.python.org/.

(3) Disponível em: https://jupyter.org/.

(4) Disponível em: https://numpy.org/.

(5) Disponível em: https://scipy.org/.

para análise numérica e Matplotlib[6] para visualização, entre outros. Tudo isso, além do fato das licenças de uso dessas ferramentas serem variantes de licenças do tipo “open source” e da existência de uma comunidade de suporte bastante ativa, têm feito com que Python figure hoje entre as principais ferramentas na comunidade de computação científica.

Especificamente em relação a softwares para avaliação genética, até onde se tem conhecimento, o brblup é o primeiro software para avaliação genética de animais que baseia-se em Python e suas ferramentas associadas. Em termos de plataforma, ele difere muito daquela (Fortran) com a qual seu público-alvo, formado por melhoristas e pesquisadores vinculados a programas nacionais de melhoramento genético animal, estão habituados.

Assim, diante desse contexto, cabe uma demonstração da capacidade do brblup em realizar avaliações genéticas de populações de animais vinculados a programas de melhoramento genético reais. Este documento apresenta o resultado de avaliações genética das populações de animais vinculadas aos programas de melhoramento Geneplus/Canchim e Brangus+, comparando-os com as avaliações (EBVs, acurácias) oficiais dos respectivos programas.

(6) Disponível em: https://matplotlib.org/.

Material e método

O conjunto de dados utilizado para avaliação genética dos animais da raça Brangus é oriundo da avaliação genética oficial da raça realizada junto à Associação Brasileira de Brangus (ABB) no dia 10 de julho de 2024, sendo composto pelos seguintes arquivos, após conversão para o formato utilizado pelo brblup:

- Observações de 40.347 animais, incluindo a fenotipagem (peso ao desmame e peso ao nascer) dos animais e correspondentes efeitos fixos e covariáveis.
- Informações de genealogia, formando um pedigree com 109.780 animais.
- Dados de genotipagem de 8.848 animais, sendo que aproximadamente 65% possuem fenótipo.
- Mapeamento entre os IDs dos animais e seus correspondentes genótipos (75.972 SNPs).

O seguinte modelo animal bi-característico – Peso ao nascimento (PN) e Peso à desmama (PD) – de avaliação genética foi utilizado:

$$Característica = ani + mat + pe + gc + cov1 + cov2 + cov3 + cov4 + \\ + idv + cov5 + cov6$$

Onde:

- ***Característica: Peso ao nascimento (PN) ou Peso à desmama (PN).***
- ani: efeito direto do animal.
- mat: efeito materno.
- pe: efeito ambiental permanente materno.
- gc: grupo de contemporâneos.

- cov1: efeito linear da idade do animal na data da medida (para PD).
- cov2: efeito quadrático da idade do animal na data da medida (para PD).
- cov3: efeito da proporção zebuína individual.
- cov4: efeito da proporção zebuína materna.
- idv: efeito fixo de classe de idade da vaca (mãe)
- cov5: efeito da heterozigose individual.
- cov6: efeito da heterozigose materna.

As seguintes matrizes de covariância completam o modelo.

Matriz de covariância residual	$\begin{pmatrix} 389{,}2 & 9{,}028 \\ 8{,}0871 & 363{,}1378 \end{pmatrix}$
Matriz de covariância genética	$\begin{pmatrix} 73{,}02 & 9{,}216 & 0{,}0 & 0{,}0 \\ 9{,}216 & 7{,}852 & 0{,}0 & 0{,}0 \\ 0{,}0 & 0{,}0 & 35{,}07 & 2{,}822 \\ 0{,}0 & 0{,}0 & 2{,}822 & 0{,}8427 \end{pmatrix}$
Matriz de covariância para o efeito ambiental permanente materno	$\begin{pmatrix} 89{,}26 & 3{,}387 \\ 3{,}387 & 0{,}4658 \end{pmatrix}$

Para a raça Canchim, o conjunto de dados utilizado para avaliação dos animais é oriundo da avaliação genética oficial da raça realizada em maio de 2024, sendo composto pelos seguintes arquivos, após conversão para o formato utilizado pelo brblup:

- Observações de 232.187 animais, incluindo a fenotipagem (peso à desmama e peso ao sobreano) dos animais e correspondentes efeitos fixos e covariáveis.
- Informações de pedigree com 286.810 animais.
- Dados de genotipagem de 2.251 animais.
- Mapeamento entre os IDs dos animais e seus correspondentes genótipos (197.511 SNPs).

O seguinte modelo animal bi-característico (PN e PD) de avaliação genética foi utilizado:

$$\boxed{\textit{Característica} = ani + mat + pe + gc + idv + cov\ 1 + cov\ 2 + cov\ 3}$$

Onde:

- Característica: Peso ao nascimento (PN) ou Peso à desmama (PD).
- ani: efeito direto do animal.
- mat: efeito materno.
- pe: efeito ambiental permanente materno.
- gc: grupo de contemporâneos.
- idv: efeito fixo de classe de idade da vaca x sexo do bezerro.
- cov1: efeito linear da consanguinidade.
- cov2: efeito linear da idade do animal na data da medida (para PD).
- cov3: efeito quadrático da idade do animal na data da medida (para PD).

As seguintes matrizes de covariância completam o modelo.

Matriz de covariância residual:	$\begin{pmatrix} 8{,}0473 & 8{,}0871 \\ 8{,}0871 & 363{,}1378 \end{pmatrix}$
Matriz de covariância genética:	$\begin{pmatrix} 3{,}9249 & 12{,}6348 & 0{,}0 & 0{,}0 \\ 12{,}6348 & 137{,}2880 & 0{,}0 & 0{,}0 \\ 0{,}0 & 0{,}0 & 0{,}8742 & 3{,}8331 \\ 0{,}0 & 0{,}0 & 3{,}8331 & 82{,}3728 \end{pmatrix}$
Matriz de covariância para o efeito ambiental permanente materno:	$\begin{pmatrix} 0{,}7707 & 2{,}6150 \\ 2{,}6150 & 82{,}3728 \end{pmatrix}$

Em ambos os casos, o brblup foi utilizado para obtenção dos valores genéticos (EBV) e erro quadrático (s.e), para o posterior cálculo

das acurácias, a partir de métodos direto (modo direto) e iterativo (modo iterativo) e valores padrão para as demais opções: critério de parada para o algoritmo PCG (gradiente conjugado pré-condicionado) no modo iterativo igual a 10-8 e critério de parada para o aproximação do s.e igual a 10-3 em ambos os casos. Tanto para a obtenção do EBV quanto do s.e foram medidos o número de iterações, quando pertinente, o tempo de processamento e a quantidade de memória utilizada. Uma vez finalizado o processamento, os resultados foram comparados com aqueles oficiais de cada programa de melhoramento, utilizando as seguintes medidas: máxima diferença absoluta e correlações de Pearson e Spearman.

Inicialmente, as avaliações foram realizadas utilizando o seguinte equipamento:

- Desktop HP com 12 processadores modelo Intel(R) Xeon(R) E-2236 CPU @ 3.40GHz e 32 GB de memória RAM com sistema operacional Ubuntu 20.04.6 LTS.

Especificamente ao realizar a avaliação genética da população de Brangus utilizando o método direto[7], a quantidade de memória RAM disponível neste equipamento foi insuficiente para o processamento dos dados de genótipos. Dessa forma, essa avaliação foi realizada utilizando o seguinte equipamento:

- Servidor disponibilizado em nuvem, com 16 processadores modelo Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz e 128 GB de memória RAM com sistema operacional Ubuntu 22.04.4 LTS.

(7) Para o método iterativo, a avaliação foi realizada utilizando o Desktop HP.

Resultados

A Tabela 1 apresenta os resultados obtidos para os dados da raça Canchim. Nela, observa-se que utilizando o modo direto o tempo total de execução foi de 31,836 minutos, enquanto que utilizando o modo iterativo foi de 26,396 minutos, sendo que em ambos os casos o cálculo das EBVs consumiu a maior parte do tempo de processamento, 70% no modo direto e 51% no modo iterativo. Em relação ao consumo de memória, a quantidade utilizada no modo direto foi 34% maior que no modo iterativo (2.870 GB vs 3.752 GB)[8].

Tabela 1. Consumo de recursos computacionais pelo processo de avaliação e comparação com resultados oficiais do programa de melhoramento Geneplus/Canchim.

	EBV		S.E	
	Direto	**Iterativo**	**Direto**	**Iterativo**
Número de iterações	-	676	557	519
Tempo (minutos)	22,463	13,479	9,373	12,917
Memória RAM (GB)	3,852	2,870	3,852	2,870
Máxima diferença	1,115	1,219	10,863	9,501
Correlação (Pearson)	1,000	1,00	0,986	0,991
Correlação (Spearman)	1,000	0,991	0,941	0,969

Em relação à comparação entre resultados obtidos utilizando o br-blup e os resultados da avaliação oficial, observa-se que independente do modo de execução tanto as EBVs quanto os s.e apresentam fortes

(8) Nas Tabelas 1 e 2, a memória utilizada refere-se a todo o processo envolvendo o cálculo das EBVs e do s.e. Por isso, os valores nas respectivas colunas Direto e Iterativo são iguais.

correlações de Pearson e de Spearman, com máximas diferenças observadas de 1,115 e 10,863 para as EBVs e s.e obtidas no modo direto e de 1,219 e 9,501 para as EBVs e s.e obtidas no modo iterativo. A ligeira superioridade dos resultados obtidos no modo direto pode ser melhor observada na Figura 1, em que são apresentados os histogramas das diferenças em relação aos resultados oficiais. As Figuras 1A a 1D sugerem que os valores médios tendem a zero, indicando a normalidade das estimativas. Adicionalmente, as Figuras 1C e 1D sugerem que valores médios mais próximos de zero e, portanto, um resultado mais preciso pode ser obtido utilizando um número maior de iterações. A Figura 1D também sugere que a estimativa do s.e obtida por meio do brblup é ligeiramente superestimada.

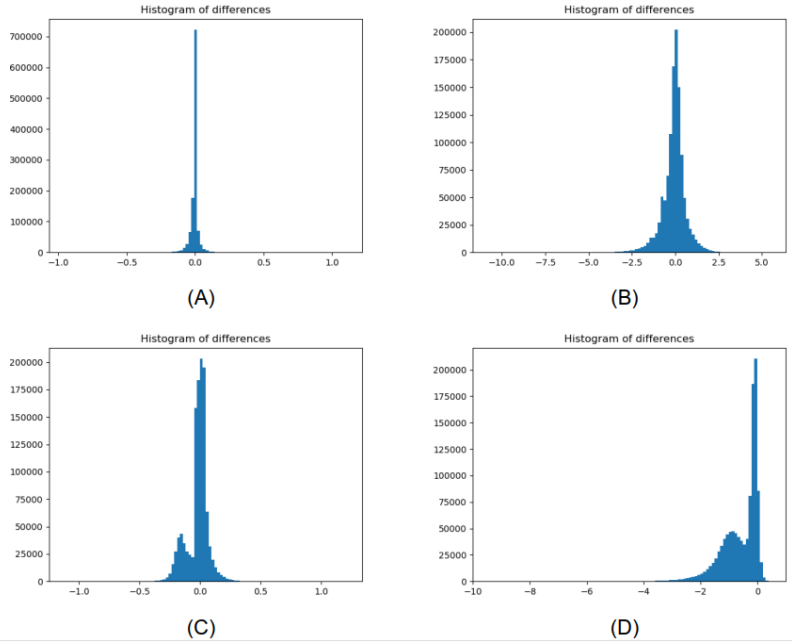

Figura 1. Histogramas das diferenças entre EBVs e s.e em relação aos resultados oficiais do programa Geneplus/Canchim. (A) EBVs no modo direto; (B) s.e no modo direto; (C) EBVs no modo iterativo; (D) s.e no modo iterativo.

A Tabela 2 apresenta os resultados obtidos para a raça Brangus. Nela, observa-se que utilizando o modo direto o tempo total de execução foi de 29,913 minutos, enquanto que utilizando o modo iterativo foi de 16,295 minutos, sendo que a maior parte do tempo de processamento no modo iterativo foi consumido pelo cálculo dos s.e (74%), enquanto que no modo direto o cálculo das EBVs consumiu a maior parte do tempo (55%). Em relação ao consumo de memória, a quantidade utilizada no modo direto foi nove vezes maior que no modo iterativo (33,781 GB vs 3,769 GB).

Tabela 2. Consumo de recursos computacionais pelo processo de avaliação e comparação com resultados oficiais do programa de melhoramento Brangus+.

	EBV		S.E	
	Direto	**Iterativo**	**Direto**	**Iterativo**
Número de iterações	-	694	500	500
Tempo (minutos)	16,343	4,281	13,570	12,014
Memória RAM2 (GB)	33,781	3,769	33,781	3,769
Máxima diferença	0,040	0,066	8,256	7,461
Correlação (Pearson)	1,000	1,000	0,991	0,989
Correlação (Spearman)	1,000	1,000	0,974	0,984

Em relação à comparação entre resultados obtidos utilizando o brblup e os resultados da avaliação oficial, da mesma forma que no caso dos dados da raça Canchim, foram observadas fortes correlações independente do modo de execução, sendo que no caso das EBVs todas as correlações são perfeitas positivas até a terceira casa decimal. As máximas diferenças observadas são de 0,040 e 8,256 para as EBVs e s.e obtidas no modo direto e de 0,066 e 7,461 para as EBVs e s.e obtidas no modo iterativo. Novamente, a ligeira superioridade dos resultados obtidos no modo direto pode ser melhor observada na Figura 2, em que são apresentados os histogramas

das diferenças em relação aos resultados oficiais. As Figuras 2A a 2D sugerem que os valores médios tendem a zero, indicando a normalidade das estimativas. Adicionalmente, as Figuras 2C e 2D sugerem que valores médios mais próximos de zero e, portanto, um resultado mais preciso pode ser obtido utilizando um número maior de iterações. A Figura 2D também sugere, neste caso, que a estimativa do s.e obtida por meio do brblup é ligeiramente subestimada.

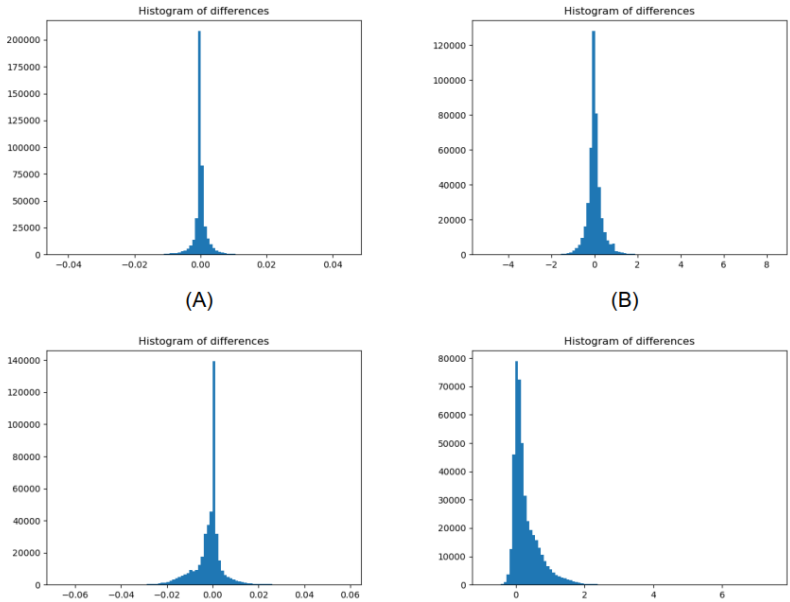

Figura 2. Histogramas das diferenças entre EBVs e s.e em relação aos resultados oficiais do programa Brangus+. (A) EBVs no modo direto; (B) s.e no modo direto; (C) EBVs no modo iterativo; (D) s.e no modo iterativo.Figura 2: Histogramas das diferenças entre EBVs e s.e em relação aos resultados oficiais do programa Brangus+. (A) EBVs no modo direto; (B) s.e no modo direto; (C) EBVs no modo iterativo; (D) s.e no modo iterativo.

A partir desses resultados, observa-se que, em geral, o modo direto tende a fornecer um resultado mais preciso que o iterativo, mas com uma maior demanda por recursos computacionais. Em particular,

observa-se que: 1) a quantidade de dados de genótipos tem grande influência na quantidade de memória RAM requerida pelo método direto; e que 2) o método iterativo tende a ser mais eficiente que o direto quando apenas a tarefa de obtenção das EBVs é realizada.

Finalmente, a partir desses resultados e supondo que cada programa realize dez avaliações sequencialmente para cobrir todas as características avaliadas, pode-se estimar que seriam necessários entre 4,399 horas e 5,370 horas no caso da raça Canchim e entre 3,033 horas e 7,803 horas no caso da raça Brangus, para realização de todas as avaliações.

Comentários finais

Este documento demonstrou a capacidade do brblup ao realizar avaliações genéticas de populações de animais vinculados a programas de melhoramento genético reais. Utilizando dados reais obtidos junto aos programas de melhoramento Geneplus/Canchim e Brangus+, foram realizadas avaliações genéticas de populações de animais vinculados a esses programas, comparando os resultados obtidos com os oficiais. Em ambos os casos, demonstrou-se que, a menos de diferenças que podem ser atribuídas à natureza numérica dos processos, os resultados obtidos são equivalentes. Em relação ao tempo necessário para realização dessa tarefa, obtenção das EBVs e s.e, estimou-se um tempo adequado para que ambos os programas processem todas as características avaliadas incluídas em suas avaliações.

Em termos de trabalhos futuros, uma vez que tanto o programa Geneplus/Canchim quanto o Brangus+ utilizam o software blupf90 para realização das avaliações, caberia uma comparação em termos de utilização de recursos computacionais. Outra questão interessante seria verificar o uso do brblup utilizando o algoritmo apy para avaliação de populações muito grandes de animais, contendo milhões de animais em seu pedigree.

Referências

HIGA, R. H. **Software BRBLUP**: manual do usuário. Campinas: Embrapa Agricultura Digital, 2025. 32 p. (Embrapa Agricultura Digital. Documentos, 196). Disponível em: https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1177861. Acesso em: 11 ago. 2025.

MADSEN, P.; JENSEN, J.; LABOURIAU, R.; CHRISTENSEN, O. F.; SAHANA, G. DMU - a package for analyzing multivariate mixed models in quantitative genetics and genomics. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10., 2014, Vancover. **Proceedings** [...]. Champaign: ASAS, 2014. 4 p.

MISZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA. Z. **Manual for BLUPF90 family of programs**. Athens, USA: University of Georgia, 2014. https://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all.pdf. Acesso em: 1 abr. 2025.

VANDENPLAS, J.; VEERKAMP, R. F.; CALUS, M. P. L.; LIDAUER, M. H.; STRANDÉN, I.; TASKINEN, M.; SCHRAUF, M.; TEN NAPEL, J. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 12., 2022, Rotterdam. **Proceedings** […]. Rotterdam: [s.n.], 2022. p. 1498-1501.

VUORI, K.; STRANDÉN, I.; LIDAUER, M.; MÄNTYSAARI, E. A. MiX99-effective solver for large and complex linear models. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Proceedings** […]. Minas Gerais: Instituto Prociência, 2006. p. 27-33.

CGPE 19285