UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE AGRONOMIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

**SELEÇÃO GENÔMICA EM BOVINOS DE CORTE**

MARIO LUIZ PICCOLI

Médico Veterinário/UFPel

Mestre em Zootecnia/UFRGS

Tese apresentada como um dos requisitos à obtenção do Grau de Doutor em Zootecnia – Área de Concentração: Produção Animal.

Porto Alegre (RS), Brasil

Março de 2015

MARIO LUIZ PICCOLI
Médico Veterinário e
Mestre em Zootecnia

<div align="center">

**TESE**

Submetida como parte dos requisitos
para obtenção do Grau de
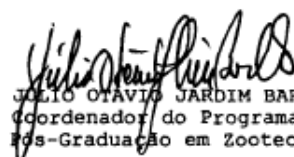
**DOUTOR EM ZOOTECNIA**

Programa de Pós-Graduação em Zootecnia
Faculdade de Agronomia
Universidade Federal do Rio Grande do Sul
Porto Alegre (RS), Brasil

</div>

Aprovado em: 25.03.2015
Pela Banca Examinadora
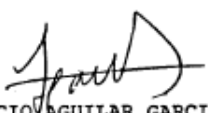
Homologado em: 20.05.2015
Por

JOSÉ BRACCINI NETO
PPG Zootecnia/UFRGS
Orientador

JÚLIO OTÁVIO JARDIM BARCELLOS
Coordenador do Programa de
Pós-Graduação em Zootecnia

JAIME ARAÚJO COBUCI
PPG ZOOTECNIA-UFRGS

FERNANDO SEBASTIÁN BALDI REY
UNESP - Jaboticabal/SP

IGNACIO AGUILAR GARCIA
INIA LAS BRUJAS - Uruguay

PEDRO ALBERTO SELBACH
Diretor da Faculdade de Agronomia

*"Ao Moacir (Tchile) e ao José que resolveram partir quando eu estava longe,*
*deixando muita saudade em meu coração;*
*ao Pedro e ao Bernardo, todos os dias próximos a mim,*
*preenchendo meu coração de amor e alegria"*

**Dedico**

# AGRADECIMENTOS

Ao meu orientador Dr. José Braccini Neto meu sincero agradecimento por seus ensinamentos e confiança a mim depositada durante o meu estudo no PPG de Zootecnia;

Ao meu co-orientador, em Portugal, Dr. Luís Telo da Gama por sua contribuição ao trabalho apresentado no capítulo II, através de sua experiência, e discussões ocorridas sempre nas quarta-feiras à tarde em seu gabinete;

Ao meu co-orientador, no Brasil, Dr. Fernando Flores Cardoso por sua contribuição ao trabalho apresentado nos capítulos III a VI, através de sua experiência que auxilaram no desenvolvimento e na interpretação dos resultados;

Ao meu co-orientador, no Canadá, Dr. Flávio Schramm Schenkel por sua contribuição ao trabalho apresentado nos capítulos III a VI, através de sua experiência e ensinamentos durante os cursos realizados no meu ano de estudo no Canadá;

Ao Dr. Mehdi Sargolzaei (University of Guelph) pela cedência dos programas FImpute, GEBV, entre outros, utilizados neste trabalho e por todo o seu auxílio, sempre de maneira muito eficiente e gentil;

Ao Dr. Ignacio Aguilar (INIA Uruguay) pelo auxílio sempre de maneira muito cordial, na utilização dos programas BLUPF90;

Ao PPG em Zootecnia da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul, ao Centre for Genetic Improvement of Livestock (CGIL) do Depto of Animal and Poultry Science da University of Guelph e a Faculdade de Medicina Veterinária da Universidade de Lisboa, Portugal, pela estrutura disponibilizada para a realização deste estudo;

À Conexão Delta G – Hereford e Braford, Embrapa Pecuária Sul, Programa Paint e à Associação Nacional de Criadores pela cedência dos dados;

Aos colegas e amigos da GenSys, Fernanda, Fries (in memoriam), Haroldo, Léia, Roberto, Severo, Vanerlei e Vânia, pela confiança e pela disposição em acumular mais trabalho na minha ausência durante o curso de doutorado;

Aos meus colegas e amigos do PPG em Zootecnia em Porto Alegre, do CGIL em Guelph e da Faculdade de Veterinária em Lisboa;

A Cláudia Bertoli e ao Luiz Brito, colegas e amigos, sempre dispostos a colaborar;

À CAPES, que me concedeu a bolsa de estudos no Brasil e no exterior;

Aos meus pais José e Lilia (in memoriam), ao meu irmão Moacir (in memoriam), às minhas irmãs Marise, Marlene, Marly e Maricir, sempre presentes na minha vida;

À minha esposa Márcia, sempre parceira e aos meus filhos Pedro e Bernardo, também, sempre parceiros.

*"Em geral quando termino um livro encontro-me numa confusão de sentimentos, um misto de alegria, alívio e vaga tristeza. Relendo a obra mais tarde, quase sempre penso 'Não era bem isto o que queria dizer'."*

*(O escritor diante do espelho)*

*Érico Veríssimo*

# [1]SELEÇÃO GENÔMICA EM BOVINOS DE CORTE

Autor: Mario Luiz Piccoli
Orientador: José Braccini
Co-orientadores: Luís Telo da Gama (Portugal), Fernando Flores Cardoso (Brasil); Flávio Schramm Schenkel (Canadá).

**RESUMO**
Objetivou-se com este trabalho: (1) avaliar parâmetros de diversidade genética e de estrutura populacional com base nas raças Angus, Devon, Hereford e Shorthorn. Taxa de endogamia (∆F), tamanho efetivo (Ne), grau de parentesco, entre outros parâmetros, foram estimados para fornecer subsídios aos programas de melhoramento. Os parâmetros indicaram adequada diversidade genética, com Ne variando entre 128 no Devon e 303 no Shorthorn e ∆F variando entre 1,50 no Angus e 3,92 no Devon; (2) avaliar estratégias de imputação de genótipos utilizando dados de Braford e Hereford através de painéis de baixa densidade (3K, 6K, 8K, 15K e 20K) para os painéis de 50K e 777K. Para o painel de 777K, também foram utilizados na imputação os painéis de 50K, 90iK e 90tK. Os resultados indicaram que, com exceção do painel de 3K, todos os demais painéis de baixa densidade poderiam ser utilizados como base visando à imputação para o painel de 50K e também que os painéis de média densidade (50K, 90iK e 90tK), poderiam ser utilizados como base na imputação para o painel de 777K. Esses painéis mostraram-se eficientes e possuem, em geral, custos compatíveis com a atividade pecuária; (3) avaliar a acurácia de predição dos valores genômicos utilizando alguns painéis de baixa densidade (8K e 15K) imputados para o painel de 50K, relacionando os resultados com o uso do painel original de 50K. A acurácia do valor genômico direto (DGV) e do valor genético genômico (GEBV) com o valor genético (EBV) utilizando painéis imputados ou não, indicaram que não houveram diferenças em acurácia e as perdas em acurácia por utilizar os painéis imputados ficaram entre -0,0002 e -0,0021 dependendo do painel, do cenário e da característica analisada; (4) usar marcadores moleculares na seleção genômica testando dois métodos BLUP (procedimento de passo único ou de multi passo) com dados simulados de bovinos de corte. Os resultados demonstraram, com base nos parâmetros estudados, igualdade de resultados entre os dois procedimentos; (5) avaliar a viabilidade do uso da seleção genômica usando dados de campo de animais Braford e Hereford, testando os dois métodos BLUP (passo único e multi passo). Os resultados com base nos parâmetros estudados, mostraram que as acurácias de predição do DGV e do GEBV foram iguais nos dois procedimentos, porém no método multi passo as predições genômicas foram menos viesadas.

**Palavras-chave:** diversidade genética, consanguinidade, imputação, SNP, seleção genômica, acurácia, single-step, two-steps, Braford, Hereford

---

[1] Tese de doutorado em Zootecnia- Produção Animal, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil. (166p.). Março, 2015.

# [2] GENOMIC SELECTION IN BEEF CATTLE

Author: Mario Luiz Piccoli
Supervisor: José Braccini
Co-supervisors: Luís Telo da Gama (Portugal); Fernando Flores Cardoso (Brazil); Flávio Schramm Schenkel (Canada).

**ABSTRACT**
The aim of this work were: (1) to evaluate parameters of genetic diversity and population structure based on Angus, Devon, Hereford and Shorthorn breeds. Inbreeding rate ($\Delta F$), effective size (Ne), relatedness, among other parameters, were estimated to provide subsidies for breeding programs. The parameters indicated a good genetic diversity, with Ne ranging from 128 in Devon to 303 in Shorthorn and $\Delta F$ ranging from 1.50 in Angus to 3.92 in Devon; (2) to evaluate strategies of genotype imputation with Braford and Hereford beef data using low density panels (3K, 6K, 8K, 15K and 20K) for 50K and 777K panels. For imputation to the 777K panel were also used the 50K, 90iK and 90tK panels. The results indicated that, except for the 3K panel, all other low density panels could be used of imputation the 50K panel and also the medium density panels (50K, 90iK and 90tK), could be used of imputation the 777K panel. These panels have been efficient and have, in general, compatible costs of the beef cattle operation; (3) to evaluate the accuracy of genomic prediction using some low density panels (8K and 15K) imputed to the 50K panel, relating the results with 50K original panel. The accuracy of direct genomic value (DGV) and genomic estimated breeding value (GEBV) with estimated breeding value (EBV) using imputed or not panels, indicated that there were no differences in accuracy and the losses in accuracy by using the imputed panels ranged from -0.0002 to -0.0021 depending on the panel, the scenario and the trait; (4) to use molecular markers in genomic selection testing two BLUP methods (single and two steps) with simulated beef cattle data. The results showed, based on the parameters studied, equality of results between the two methods; (5) to evaluate the viability of using the genomic selection using Braford and Hereford beef cattle and testing the two BLUP methods (single and two steps). The results, based on the parameters studied, showed that DGV and GEBV accuracies were similar in both methods, but the genomic predictions were less biased with then two-step method.

**Keywords:** genetic diversity, inbreeding, imputation, SNP, genomic seletion, accuracy, single-step, two-steps, Braford, Hereford

# SUMÁRIO

# LISTA DE TABELAS

# LISTA DE FIGURAS

# LISTA DE ABREVIATURAS E SÍMBOLOS

| | |
|---|---|
| BLUP | Melhor preditor linear não viciado |
| cM | centiMorgan |
| EBV | Valor genético estimado |
| DGV | Valor genômico direto |
| F e $\Delta$F | Coeficiente de endogamia |
| GEBV | Valor genético genômico |
| IBD | Idêntico por descendência |
| QTL | Locus de característica quantitativa |
| SNP | Polimorfismo de base única |
| Ne | Tamanho efetivo da população |

**CAPÍTULO I**

# INTRODUÇÃO

Os marcadores moleculares SNP (*Single Nucleotide Polymorphism*) vem sendo utilizados no que hoje se denomina de seleção genômica. O princípio desta tecnologia baseia-se no fato de que estes marcadores SNP podem capturar o efeito de todos os genes responsáveis por características de importância econômica. Para isto, estes marcadores devem estar espalhados e igualmente espaçados por todo o genoma do animal, próximos aos QTL (*Quantitative Trait Loci*) gerando uma associação entre o marcador e o QTL. Esta associação entre marcador e QTL é medida através do que se denomina de desequilíbrio de ligação, e vai depender do número de marcadores SNP utilizados, do tamanho efetivo da população (Ne) e da herdabilidade da característica.

O sequenciamento completo do bovino (The Bovine Genome Sequencing and Analysis Consortium, 2009) possibilitou a descoberta de um grande número de SNP (~3 milhões) e com isso o desenvolvimento de novas tecnologias de genotipagem com alto grau de repetibilidade, análises de alta performance e de baixo custo por marcador. Painéis de baixa até alta densidade já foram desenvolvidos e estão sendo utilizados visando à aplicação da seleção genômica em bovinos. Recentemente as empresas responsáveis pelo desenvolvimento destas tecnologias de genotipagem tem possibilitado o desenvolvimento de plataformas customizadas, ou seja, o desenvolvimento de painéis específicos de genotipagem para serem aplicados à determinada raça ou programa de melhoramento genético.

Se por um lado o avanço da tecnologia de genotipagem tem propiciado diversos estudos visando à aplicação da seleção genômica em bovinos de leite e de corte, por outro lado estas novas tecnologias para o produtor têm esbarrado no alto custo. Uma alternativa ao alto custo dos painéis de média a alta densidade, que tem mostrado os melhores resultados na aplicação da seleção genômica, é a utilização de painéis de baixa densidade e por consequência de mais baixo custo e a utilização da técnica de imputação. Esta técnica é utilizada para aumentar o número de marcadores uma vez que a densidade destes marcadores afeta a acurácia da seleção genômica.

A metodologia de modelos mistos baseada principalmente no modelo animal é utilizada no mundo todo para obter-se a melhor predição não viesada (BLUP) dos valores genéticos dos animais nos diferentes caracteres de produção utilizados na seleção. Os primeiros resultados de pesquisa apresentados em seleção genômica tem utilizado um procedimento desenvolvido em mais de uma etapa para a obtenção dos valores genéticos dos animais. Neste processo, os valores genéticos (EBV) são obtidas através de análises convencionais utilizando modelos mistos (com base no fenótipo e no pedigree) enquanto que os valores genômicos diretos (DGV) são obtidos com base em equações de predição geradas a partir da informação dos marcadores em análise de associação com os fenótipos, e após, ambas são combinadas por meio de diferentes índices gerando os valores genéticos genômicos (GEBV). Mais recentemente os pesquisadores tem estudado o procedimento em um único passo onde as informações dos SNP são combinadas com o fenótipo e o pedigree para a estimação simultânea dos valores genéticos. Vantagens e desvantagens têm sido apontadas para ambos

os procedimentos.

Em bovinos de leite os estudos em seleção genômica tem mostrado um ganho em confiabilidade (acurácia elevada ao quadrado) ao redor de 100% para os animais jovens quando a DGV foi estimada com base em aproximadamente 50.000 marcadores, e uma resposta a seleção duas vezes maior, principalmente pela diminuição do intervalo entre gerações (Schaeffer, 2006). Com isso os custos para identificar touros superiores podem ser reduzidos em até 92%. Com base nestes resultados vários países já adotam a seleção genômica em seus programas de melhoramento de bovinos de leite.

Pesquisas conduzidas com bovinos de corte têm mostrado resultados inferiores aos dos bovinos de leite. A estrutura genética em bovinos de corte envolve um número maior de raças que resulta em diferenças importantes em parâmetros como tamanho efetivo da população e extensão do desequilíbrio de ligação, além do custo da genotipagem para o produtor.

Este trabalho tem por objetivos: a) estimar parâmetros de diversidade genética e estrutura populacional das raças Angus, Devon, Hereford e Shorthorn; b) a acurácia de imputação de genótipos e também a utilização dos genótipos imputados na seleção genômica com dados de animais Braford, Hereford e Nelore; c) comparar as metodologias de passo único e multi passo na avaliação genômica em populações simuladas e de bovinos Braford e Hereford.

# REVISÃO BIBLIOGRÁFICA

### Diversidade genética

Diversidade genética é definida como sendo a variabilidade genética existente entre indivíduos de uma raça ou entre populações de uma espécie (Brown, 1983), influenciada por fatores tais como a mutação, a recombinação, a seleção e a deriva genética (Falconer & MacKay, 1996).

Análise com base na informação de pedigree pode ser usada para monitorar a evolução da diversidade de uma determinada raça e de como se comporta a estrutura genética desta população no decorrer do tempo (Falconer & MacKay, 1996). MacCluer et al. (1986), Lacy (1989) e Boichard et al. (1997) tem proposto diferentes métodos para análise de variabilidade genética em raças bovinas. Entre os métodos mais utilizados está o de estimação de medidas de distância genética (Nei, 1973) e a estatística F de Wright (1931).

Estudos de diversidade genética com bovinos, em processo de seleção, (Cleveland et al., 2005; Mc Parland et al., 2007; Márquez et al., 2010, Santana et al., 2012 e Piccoli et al., 2014a) são extremamente importantes e tem por objetivo obter parâmetros para o direcionamento ou redirecionamento da seleção dos caracteres de interesse econômico e também para o controle da perda de diversidade genética.

A metodologia de modelos mistos baseada principalmente no modelo animal é utilizada no mundo todo para predição dos valores genéticos dos animais dos diferentes caracteres de produção utilizados na seleção. Esta metodologia apresenta propriedades estatísticas BLUP (Henderson, 1975) gerando resultados satisfatórios de ganhos genéticos a curto e médio prazo. Entretanto, conduz a um aumento da taxa de endogamia e a perdas de variabilidade genética, por utilizar todas as relações de parentesco existentes na população e com isso os animais mais aparentados têm maior probabilidade de serem selecionados.

A endogamia, que significa o acasalamento de indivíduos que são relacionados por ascendência (Falconer & MacKay, 1996), se não controlada, faz com que os programas de melhoramento genético possam não apresentar ganho genético ou mesmo ganho pouco expressivo. Portanto, a endogamia é um importante parâmetro que deve ser constantemente monitorado e controlado em programa de melhoramento genético (Sørensen et al., 2005).

O coeficiente de endogamia (F), que é a probabilidade de dois genes em qualquer loco serem idênticos por ascendência, é a ferramenta utilizada para monitorar a endogamia. O coeficiente de endogamia foi apresentado por (Wright, 1922) e o cálculo é baseado no número de gerações até o antepassado comum. Valores de no máximo 1% por geração tem sido a recomendação da FAO (1998) para não haver perdas de diversidade genética.

Em populações com pedigree incompleto o cálculo do coeficiente de endogamia pelo método de Wright (1922) é subestimado e o cálculo do tamanho efetivo da população (Falconer & MacKay, 1996) é superestimado (Boichard et al., 1997). Com o objetivo de ajustar estes cálculos, VanRaden, (1992) apresentou um algoritmo onde a endogamia dos animais cujos pais são desconhecidos é igual à média de seus contemporâneos que têm os pais conhecidos. Estudo recente em bovinos de corte (Brito et al. 2013) compararam os dois métodos de cálculo do coeficiente de endogamia

utilizando informação de pedigree com mais de 50% dos acasalamentos sendo realizados com reprodutores múltiplos (neste caso específico o pai não é conhecido) e mostraram que a endogamia calculada da forma regular apresentou o valor F=0,25 sendo fortemente subestimada em relação ao método proposto por VanRaden, (1992) cujo valor foi de F=1,73. Em outra situação, Piccoli et al. (2014a) utilizando o pedigree de raças de origem Britânica, com 5% a 10% dos registros com falta de informação de um ou ambos pais, mostraram que o cálculo regular do coeficiente de endogamia foi pouco afetado (menos de 2%).

O tamanho efetivo de uma população pode ser entendido como o número de indivíduos que se reproduzem e conseguem deixar descendentes, transmitindo assim seus genes nas gerações recentes mantendo a diversidade genética. Este conceito foi introduzido por Wright (1931, 1938) como sendo o tamanho de uma população idealizada (infinitamente grande, sem ação da mutação, migração e seleção) que daria origem à mudança na frequência gênica ou na taxa de endogamia observada na população real.

Populações com tamanho efetivo pequeno perdem diversidade genética em consequência da deriva genética, onde os alelos presentes numa geração podem, aleatoriamente, se tornar mais ou menos frequentes, ou até mesmo extinguir-se em gerações subsequentes, sendo esta a causa principal de perda de variação genética à longo prazo, ameaçando assim a adaptabilidade das populações (Falconer & MacKay, 1996).

Um tamanho efetivo da população mínimo de 50 animais tem sido recomendado como um limite adequado para manter a diversidade genética em programas de conservação (Meuwissen & Woolliams, 1994; FAO, 2013) e em programas de seleção (Goddard e Smith, 1990), porém limites mais elevados têm sido recomendados, em alguns casos em raças bovinas, onde a inseminação artificial é comum (Leroy et al., 2013).

**Desequilíbrio e consistência de fase de ligação**

Desequilibrio de ligação ocorre quando dois genes estão suficientemente próximos no genoma que a recombinação durante a meiose entre eles é rara, e os segmentos do cromossomo são conservados de uma geração para a outra. O desequilíbrio de ligação descreve uma associação não aleatória de dois loci no mesmo cromossomo (Ardlie et al., 2002) e também é uma medida estatística da associação entre alelos de diferentes locos.

O desequilíbrio de ligação resulta do processo de seleção, migração, mutação e deriva genética, ou ainda, pode ser gerado por meio de cruzamentos entre linhagens ou raças diferentes (Lander & Schork, 1994; Zhao et al., 2005). Tanto a mutação quanto a seleção têm pequeno efeito no desequilíbrio de ligação. A mutação por ser um evento de frequência muito baixa e geralmente ter ocorrido muitas gerações atrás e a seleção por tem o seu efeito localizado ao redor de genes específicos e, portanto, tendo um efeito relativamente pequeno sobre a quantidade de desequilíbrio de ligação médio ao longo do genoma. Entretanto, a taxa de recombinação que ocorre entre os alelos de uma geração para outra é o principal fator de diminuição dos valores de desequilíbrio de ligação.

A extensão do desequilíbrio de ligação é uma função do tamanho efetivo da população. Em animais domésticos onde o tamanho efetivo pode ser

inferior a 100 animais (Riquet et al., 1999), níveis consideráveis de desequilibio de ligação podem se estender até mais do 5-10 cM (Hayes et al., 2003; Sargolzaei et al., 2008; De Roos et al., 2008). Desequilíbrio de ligação em gado de leite é, em geral, mais alto do que em gado de corte devido, principalmente, ao menor tamanho efetivo de população nas raças leiteiras (Goddard et al., 2006).

As principais formas de quantificar o desequilíbrio de ligação são através do cálculo do $D$ e do $r^2$ (Hill & Robertson, 1968). O cálculo do $D$ é definido como sendo: $D=f(A_1B_1)*f(A_2B_2)-f(A_1B_2)*f(A_2B_1)$, onde $f(A_1B_1)$ é a frequência dos haplótipos $A_1B_1$ na população, assim como os demais haplótipos. Esta medida de desequilíbrio de ligação é muito dependente da frequência dos alelos individuais. O cálculo do $r^2$ é definido como sendo $r^2=D^2/f(A_1)*f(A_2)*f(B_1)*f(B_2)$, onde $f(A_1)$ é a frequência do alelo $A_1$ na população, assim como os demais alelos. Esta medida representa a correlação entre dois loci e foi demonstrado ser mais adequada por ser menos sensível à frequência alélica e ao tamanho da amostra. (Zhao et al., 2007; Bohmanova et al., 2010).

A fase do desequilíbrio de ligação ou fase gamética é avaliada entre dois cromossomos homólogos. Se há consistência de fase entre os alelos, a fase é de acoplamento (sinal +) e se não há consistência de fase entre os alelos, a fase é de repulsão (sinal -). Sendo a fase do desequilíbrio de ligação consistente ao longo do genoma entre duas ou mais raças ou entre duas ou mais sub-populações da mesma raça, significa que existe a probabilidade de um marcador (ou QTL) ser comum em ambas populações estudadas e a seleção genômica poderia ser aplicada utilizando-se as informações conjuntas dos marcadores de ambas as raças (Lu et al., 2012). Calcula-se a fase do desequilíbrio de ligação como sendo $sinal\ D\sqrt{r^2}$ e usa-se a correlação de Pearson para confrontar a fase entre os marcadores presentes em ambas populações. Esta correlação vai representar o parentesco genético entre as populações (De Roos et al., 2008).

### Imputação de genótipos

A imputação de genótipos consiste na predição de marcadores SNP não genotipados a partir de uma população onde todos os marcadores SNP foram genotipados. Com o uso desta técnica é possível genotipar animais com painéis de baixa densidade e predizer os genótipos do painel de alta densidade (Druet et al., 2010; Zhang & Druet, 2010).

A imputação pode ser realizada com base na informação da população (Browning & Browning, 2007, 2009), com base na informação do pedigree (Hickey et al., 2011; Sargolzaei et al., 2014) ou mesmo uma combinação de ambas ( VanRaden et al., 2011; Sargolzaei et al., 2014). A imputação com base no pedigree utiliza as regras de ligação e segregação mendeliana para predizer os genótipos, sendo mais acurada para os indivíduos que possuem parentes genotipados, enquanto que a imputação com base na população utiliza o desequilíbrio de ligação entre os marcadores SNP observados na população utilizada como referência.

O método de imputação baseado na população assume que os indivíduos não apresentam relacionamento apesar de ser possível identificar relação de parentesco por haplótipos compartilhados (Browning & Browning,

2009). Neste método, pequenos segmentos do cromossomo carregam alelos ou haplótipos idênticos por descendência e, portanto, estas regiões estão conservadas, significando que dois indivíduos aparentados irão compartilhar os mesmos alelos. Quanto mais próximo o parentesco, os indivíduos irão compartilhar segmentos cromossômicos mais longos pois não haverá quebra por recombinação dos haplótipos idênticos por descendência (IBD). Por outro lado, quanto mais distante o parentesco, os haplótipos dos indivíduos serão mais curtos pois ao longo das gerações os segmentos IBD são perdidos, principalmente por recombinação. O método que combina a informação da população com a informação do pedigree é o método que vem sendo mais utilizado em animais domésticos, justamente por se ter disponível a informação do pedigree. Neste método, (Kong et al., 2008; Hickey et al., 2011; Sargolzaei et al., 2014) os segmentos IBD longos são identificados considerando o grau de parentesco entre os animais. A informação do pedigree representa importante fator para identificação da fase dos haplótipos e imputação (Kong et al., 2008). Os genótipos não observados de um indivíduo podem ser inferidos comparando haplótipos IBD herdados com haplótipos presentes em outro indivíduo proveniente da mesma família (Li et al. 2009).

Vários softwares, tais como, *BEAGLE*, *FastPHASE*, *MACH*, *IMPUTE2*, *FImpute*, *AlphaImpute* e *findhap* têm sido desenvolvidos para imputar com maior eficiência e precisão.

A imputação é dependente do tamanho e da distância genética da população de referência, da densidade dos marcadores e da frequência alélica (Zhang & Druet, 2010; Druet et al., 2010). A imputação, também, vai depender de qual método for utilizado (Zhang & Druet, 2010; Sargolzaei et al., 2014).

A eficiência da técnica é medida pela taxa de concordância e pelo quadrado da correlação alélica ($R^2$ alélico). A taxa de concordância corresponde a proporção de genótipos imputados corretamente enquanto o $R^2$ alélico é determinado pela quadrado da correlação entre a contagem de alelos (alelo de efeito menor) imputados e a contagem de alelos do genótipo original (Browning & Browning, 2009) e é uma medida que não depende da frequência alélica do marcador.

O objetivo em utilizar está técnica remete ao fato de que a densidade de marcadores no painel afeta a precisão da seleção genômica (Hayes et al., 2009; Brito et al., 2011). Em grandes populações a genotipagem tem um impacto muito forte no custo, principalmente com painéis de alta densidade. A alternativa é genotipar com painéis de mais baixa densidade e por consequência de menor custo, porém utilizando a imputação para inferir os genótipos para um painel mais denso, usufruindo assim de todo o potencial da seleção genômica e este menor custo da genotipagem podendo ser suportada pelos criadores (Sargolzaei et al., 2010; Dassonneville et al., 2012; Piccoli et al., 2014b).

A evolução da tecnologia de genotipagem resultou em muitos animais de diferentes raças sendo genotipados com uma variedade de painéis com diferentes densidades de marcadores SNP. Para a eficácia da seleção genômica todos os animais devem ter genótipos de densidade equivalente, portanto, a imputação também elimina a necessidade de refazer a genotipagem de animais importantes no processo da seleção genômica quando estes tenham sido genotipados com painéis diferentes da maioria dos demais

animais.

Os softwares que usam em seu processo de imputação a informação do pedigree têm possibilitado inferir genótipos de animais não genotipados com alta acurácia quando estes animais possuem ascendentes ou descendentes genotipados. Este procedimento tem sido utilizado principalmente para se obter os genótipos de animais que não se dispõem de material biológico, tais como alguns importantes touros, e também para se obter os genótipos de vacas até então não genotipadas com intuito de agregar informação porém não gerando custo.

### Seleção genômica

Com o desenvolvimento dos marcadores moleculares criou-se uma expectativa de que as informações destes marcadores em associação com características de interesse econômico pudessem ser utilizadas na seleção de animais com a perspectiva de um aumento nos ganhos por seleção. Com isso vários estudos foram realizados (Casas et al., 2000; Schenkel et al., 2006) buscando a associação dos marcadores com os QTL. Muitos QTL foram detectados e mapeados, porém não foram aplicados de forma prática em programas de melhoramento genético (Bernardo, 2008), principalmente porque a resposta aos efeitos destes marcadores somente se verificava dentro de cada família avaliada e ao fato de serem feitas apenas a detecção de um pequeno número de QTL de grande efeito, os quais, devido à natureza poligênica das características de interesse econômico, não explicavam suficientemente toda a variação genética (Dekkers, 2004).

As bases da seleção genômica apresentada por Meuwissen et al. (2001) teve como princípio, o uso de painéis densos de marcadores distribuídos ao longo de todo o genoma para estimar o valor genético dos animais baseado no desequilíbrio de ligação entre marcadores e genes. Na seleção genômica não se necessita conhecer quais marcadores estão próximos a QTL pois todos os marcadores e QTL são considerados nas análises. Também não são necessários testes para genes específicos pois o objetivo é o melhoramento genético para características quantitativas, ou seja, características controladas por muitos pares de genes.

Na seleção genômica os efeitos dos marcadores são estimados em uma população de animais (população de treinamento) que possuem informação de genotipagem e que também tenham fenótipos coletados. Os efeitos estimados dos marcadores serão utilizados na predição genômica de outros animais (população de predição) que pertençam à mesma população tal que o desequilíbrio de ligação entre marcadores e QTL persista da subpopulação de treinamento para a subpopulação de predição. A soma dos efeitos estimados de todos os marcadores vai compor o que se denomina de valor genômico direto (DGV).

O uso da informação dos marcadores moleculares poderá levar a maiores ganhos genéticos em períodos mais curtos de tempo. Alguns resultados de uso da seleção genômica demonstraram maiores ganhos genéticos em programas de melhoramento através da incorporação das predições genômicas ao mérito genético do animal em comparação com as avaliações genéticas tradicionais (Schenkel et al., 2009; Hayes et al., 2009). Estes ganhos estariam associados com menor intervalo entre gerações, com o

aumento na intensidade e acurácia da seleção (Meuwissen et al., 2001; Aguilar et al., 2010). Economicamente, o uso dos marcadores moleculares, resultando em predições genética mais acuradas e mais cedo na vida dos animais permitiriam uma economia substancial na condução de testes de progênie (Schaeffer, 2006; Hayes et al., 2009).

A acurácia da predição do DGV estaria ligada: 1) ao nível de desequilíbrio de ligação entre os marcadores e os QTL; 2) ao número de animais na população de treinamento com fenótipo e genótipo que serão utilizados para estimação dos efeitos dos marcadores; 3) a herdabilidade da característica ou a acurácia da DEP, se esta for utilizado como resposta na equação de predição dos efeitos dos marcadores e 4) a distribuição dos efeitos dos QTL (Hayes et al., 2009).

Na seleção genômica se tem o interesse na utilização dos marcadores estimados em uma determinada população serem utilizados na predição genômica de animais não relacionados a esta população, seja por questões econômicas ou por questões técnicas. Neste caso, é esperado que a acurácia das predições genômicas podem ser menores, tanto quanto forem geneticamente mais distantes as populações ou por apresentarem diferente fase de ligação. Neste aspecto, a tendência é de que as populações utilizadas para estimação dos efeitos dos SNP sejam formadas de animais de diferentes populações e raças. Alguns estudos foram realizados com animais de diferentes raças em gado de leite (Hayes et al., 2009a; Erbe et al., 2012), em gado de corte (Weber et al., 2012; Kachman et al., 2013) e em ovinos e caprinos leiteiros (Moghaddar et al., 2014; Carillier et al., 2013). Os valores das acurácias encontradas pelos diversos autores quando na análise estavam envolvidas mais de uma população ou raça, em geral, foram inferiores na comparação dentro da mesma população ou raça.

Em casos em que a população de predição não têm parentesco com os animais da população de estimação dos efeitos dos marcadores, mesmo sendo da mesma raça, as predições dos valores genéticos genômicos requerem uma maior densidade de marcadores e maior tamanho da população treinamento (Meuwissen, 2009). A densidade dos marcadores, igualmente distribuídos pelo genoma, irá aumentar a probabilidade de que cada QTL esteja em alto desequilíbrio de ligação com pelo menos um marcador (Calus et al., 2008; Goddard, 2009). O tamanho da população de treinamento está diretamente relacionado com a herdabilidade da característica. Herdabilidades menores requerem maior tamanho da população para estimação dos efeitos dos marcadores e vice-versa. Hayes et al. (2009b) citam que para se obter uma acurácia de 0.50 e 0.70 para uma característica de herdabilidade de 0.20 é necessário aproximadamente 5.000 e 18.000 animais com genótipos e fenótipos na população de treinamento, respectivamente, e para uma herdabilidade de 0.50 os números seriam de 2.000 e 5.000 animais, respectivamente.

# HIPÓTESES E OBJETIVOS

### Hipóteses

a) a seleção aplicada em programas de melhoramento genético animal afeta os parâmetros de diversidade genética e de estrutura populacional das populações de bovinos de corte;

b) a imputação de genótipos para os painéis de 50K e 777K a partir de painéis de baixa e média densidade permite a predição de SNP não genotipados de forma acurada, possibilitando seu uso na seleção genômica;

c) o uso de genótipos imputados na seleção genômica produzem acurácias semelhantes na compração com o uso de genótipos não imputados;

d) as metodologias de passo único e multi passo produzem acurácias semelhantes nas predições genômicas.

### Objetivos

a) estimar parâmetros de diversidade genética e estrutura populacional dos rebanhos Angus, Devon, Hereford e Shorthorn registrados na Associação Nacional de Criadores – Head Book Collares;

b) estimar a acurácia da imputação de genótipos utilizando dados de animais Braford e Hereford da Conexão Delta G, através de painéis de baixa densidade, (3K, 6K, 8K, 15K e 20K) para o painel de 50K e dos painéis de baixa e média densidade (3K, 6K, 8K, 15K, 20K, 50K, 90iK e 90tK) para o painel de 777K;

c) estimar a acurácia da seleção genômica com base nos painéis imputados para 50K e 777K comparando os resultados com os painéis originais de 50K e 777K;

d) comparar as metodologias de um passo único e multi passos na avaliação genômica em populações simuladas e de bovinos Braford e Hereford.

**<sup>1</sup>CAPÍTULO II**

*Pedigree analyses of British cattle breeds in Brazil*

**Origins and genetic diversity of British cattle breeds in Brazil assessed by pedigree analyses[1]**

M.L. Piccoli[*][§], J. Braccini Neto [*], F.V. Brito[§], L.T. Campos [¶], C.D. Bértoli[*], G.S. Campos[*], J.A. Cobuci [*], C.M. MacManus [*], J.O.J. Barcellos [*], L.T. Gama [#][2]

[*] Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul, RS.

[§] GenSys Consultores Associados SS, Porto Alegre, RS.

[¶] Associação Nacional de Criadores - Herd Book Collares, Pelotas, RS.

[#] CIISA - Faculdade de Medicina Veterinária, Universidade de Lisboa, Portugal.

[2] Corresponding author: ltgama@fmv.utl.pt

Phone: +351-213652836

Fax: +351-213652810

**Abstract**

Pedigree information available for Angus (ANG), Devon (DEV), Hereford (HER) and Shorthorn (SHO) cattle in Brazil was analyzed to appraise the genetic diversity and population structure of these breeds. Pedigree records collected from the beginning of the 20th century until 2010 were used in the analyses. Over time, the number of herdbook registrations declined in HER after a peak in the 1970's, remained low in DEV and SHO and increased steadily in ANG since the 1990s, such that it the latter is now the leading British cattle breed in Brazil. The average number of offspring registered per sire ranged from about 12 (SHO) to 20 (DEV) and the mean generation interval ranged from about 6.0 (HER and SHO) to 6.4 (ANG) years. In the reference population (calves born in 2009 and 2010, plus those born in 2008 for SHO) the mean equivalent number of generations known ranged from about 7 (SHO) to 9 (HER). In the four breeds studied, nearly all animals born over the last few years are inbred, even though the mean level of inbreeding in the reference population is below 4% in all breeds. The rate of inbreeding per generation, computed from the individual increase in inbreeding, ranged from about 0.2% (ANG) to 0.5% (DEV), with a corresponding effective population size of 245 and 92, respectively, which is above the recommended minimum critical threshold. The number of founders/ancestors contributing with 50% of the reference population gene pool was 211/26 for ANG, 41/14 for DEV, 164/25 for HER and 79/10 for SHO, with effective number of founders/ancestors/founder genomes of 470/68/36, 89/33/16, 289/59/30 and 200/28/18 for ANG, DEV, HER and SHO, respectively. The genetic contribution of different countries to the gene pool of each breed indicated that, throughout the period studied, DEV genes originated predominantly from the UK, while for the other breeds there was a changing pattern over time. Until the 1970's Argentina was the major supplier of ANG, while HER and SHO genes were mostly from Uruguay, but since then the USA took the leading role as supplier of ANG, HER and SHO genes to Brazil. Our results reveal a mild increase in inbreeding in all breeds studied, with effective population size estimates indicating that reasonable levels of genetic diversity have been maintained in all 4 breeds. Continuous monitoring of inbreeding trends and of parameters derived from probability of gene origin should be ensured, to warrant the long term maintenance of genetic diversity.

**Introduction**

Brazil is one the leading beef producing countries in the world (Anualpec, 2011), and it is estimated that currently over 80% of its cattle are of the Zebu type (Mariante et al., 2003). Nevertheless, British breeds of cattle have been used in crossbreeding programs in Brazil since the beginning of the 20th century. Their popularity has changed over the years, with an early interest in Shorthorn, followed by an increased importance of Hereford, and in recent years the Angus has become more prominent. Given the extreme diversity of environmental constraints in different regions in Brazil (Hermuche et al., 2013), purebred British cattle are only able to cope with the milder climate found in Southern Brazil, but British bulls, in particular Angus and Hereford, are extensively used in crossbreeding programs throughout the country, either by artificial insemination or in natural matings with Zebu-type females.

Purebred cattle of the Angus, Devon, Hereford and Shorthorn breeds registered in Brazil trace their origins back to imported animals, and pedigree records have been kept for

these breeds in Brazil since the late 19th century. Over the years, Brazil continued to import animals from these breeds, in an attempt to prevent the undesirable consequences of inbreeding and to capture some of the genetic progress made abroad.

The analyses of pedigree information for a breed can be used to monitor the evolution of its genetic diversity and population structure over time (Falconer and MacKay, 1996), and parameters based on the probability of genetic origin from different founders (James, 1972; Lacy, 1989), and ancestors (Boichard et al., 1997) provide additional information, which can be used to examine population changes over a short period of time.

In our study, we analyzed all pedigree information available for the Angus, Devon, Hereford and Shorthorn breeds in Brazil, to assess how genetic diversity has evolved in each breed and identify the major factors contributing to its change over time.

## 2. Materials and Methods

### *Animal Welfare*
Animal welfare and use committee approval was not necessary for this study because data were obtained from preexisting databases.

### *Data*
Pedigree information was obtained for purebred animals of the Angus (ANG), Devon (DEV), Hereford (HER) and Shorthorn (SHO) breeds with records tracing back to founder animals imported to Brazil, which were considered to be of "Pure Origin" (PO). Data were provided by the National Breeder's Association (Associação Nacional de Criadores - Herd Book Collares) which is responsible for keeping pedigree records for PO animals of European descent in Brazil.

For imported animals, the information on sire and dam plus any known ancestors up to the third generation were entered in the data base, and pedigrees registered from then on, making the information for Brazilian-born calves progressively more complete. The first years of recorded information were 1901, 1909, 1905 and 1897 for the ANG, DEV, HER and SHO, respectively, and records collected up to 2010 were considered in all breeds. For each animal the information available included the sire, dam, sex, date of birth, herd and country of origin. Records were edited and validated regarding the consistency of information on pedigree, sex, date of birth and repeated records.

### *Pedigree analysis*
The degree of pedigree completeness was assessed by calculating an equivalent number of complete generations known per animal ($n_i$), as in Carolino and Gama (2008). Briefly, $n_i$ was obtained as $n_i = \dfrac{n_s + n_d}{2} + 1$, where $n_s$ and $n_d$ correspond to the number of generations known for the sire and dam, and assume a value of -1 if the corresponding parent is unknown. For founder animals, a value of 0 was attributed to $n_i$.

Generation intervals were calculated for the 4 selection paths, i.e., sires of sires, sires of dams, dams of sires, and dams of dams, and these were then averaged in a pooled generation interval.

The mean additive genetic relationship between pairs of animals and the individual coefficient of inbreeding ($F_i$) were obtained from the numerator relationship matrix (Van Vleck, 1993). An inbreeding level of zero was assigned to individuals with one or both parents unknown. Based on the individual $F_i$, the rate of inbreeding per year was

estimated by the regression coefficient of $F_i$ on year of birth, obtained with the GLM procedure of SAS v.9.3. (SAS Inst. Inc., Cary, NC).

An estimate of the rate of inbreeding per generation was obtained in the reference population from the change in average inbreeding in the period between 2002 and 2009, as this corresponds to approximately 1 generation interval. In this case, the approximate rate of inbreeding per generation ($\Delta F_g$) was calculated as:

$$\Delta F_g = \frac{\overline{F}_{2009} - \overline{F}_{2002}}{1 - \overline{F}_{2002}}$$

where $\overline{F}_{2009}$ and $\overline{F}_{2002}$ are the mean coefficients of inbreeding for animals born in 2009 and 2002, respectively.

Another estimate of the rate of inbreeding per generation was obtained from the individual rate of inbreeding ($\delta F_i$), which was computed as in Gutiérrez et al. (2009):

$$\delta F_i = 1 - \sqrt[n_i - 1]{1 - F_i}$$

where $F_i$ is the coefficient of inbreeding of an individual and $n_i$ is its equivalent number of complete generations known. The $\delta F_i$ for animals in the reference population were averaged by breed to obtain a mean rate of inbreeding per generation ($\Delta F_i$).

The two estimates of the rate of inbreeding per generation were used to estimate the effective population size ($N_e$), which was computed as in Falconer and Mackay (1996):

$$N_e = \frac{1}{2\,\Delta F}$$

where $\Delta F$ corresponds to either $\Delta F_g$ or $\Delta F_i$.

For the calculation of genetic contributions from founders, ancestors and countries, a reference population intended to represent the current gene pool of a breed is often assumed. This reference population usually covers a short period of time, to avoid the possibility of including both parents and their offspring, as this could bias the results (Boichard, 1997). In our analyses, the reference population corresponded to the group of calves born in the years 2009 and 2010 for the ANG, DEV and HER breeds. For the SHO breed, given the small number of animals registered in the last few years (162 calves born in 2009 and 2010), the reference population was expanded to also include calves born in 2008, but still no parent-offspring pairs were represented. The genetic contributions to the reference population of founder animals and ancestors were computed, as described by James (1972) and Boichard et al. (1997). From these contributions, the effective number of founders ($f_e$) and ancestors ($f_a$) was calculated (Boichard et al., 1997), and the effective number of founder genomes ($f_g$) was obtained as described by MacCluer et al. (1986) and Lacy (1989). For the purpose of these calculations, founders were considered both the individuals in the pedigree with no parents known, and the unknown parents of an animal with only one parent known ("phantom founders"; Gutiérrez and Goyache, 2005). In these analyses, $f_e$ and $f_a$ corresponded to the number of founders and ancestors that would be expected to generate the same level of genetic diversity in the reference population, if they all had the same contribution (Boichard et al., 1997). The $f_g$, on the other hand, accounts for the effects of unequal founder contributions, bottlenecks and genetic drift, and thus corresponds to the number of equally contributing founders that would result in the same genetic diversity in the reference population, taking into account the random loss of founder alleles (Lacy, 1989). The ratios $f_e/f_a$ and $f_e/f_g$ were calculated for each breed, a higher value indicating a stronger bottleneck in the pedigrees.

In addition to the computation of probabilities of gene origin in the reference population as defined above, the effective numbers of founders and ancestors were calculated in a series of reference populations defined as groups of calves born in subsequent 10 years intervals, to assess how genetic diversity has evolved in British cattle breeds after they were introduced in Brazil at the beginning of the 20th century.

In demographic analyses, the genetic contributions of founder herds are often computed (Gutiérrez and Goyache, 2005; Carolino and Gama, 2008). In our case, more than the specific contributions of individual herds, the interest was to assess the genetic contribution that individuals from different countries have provided to the genetic pool of the breeds analyzed. Therefore, the country of origin of imported animals was considered in a manner similar to that usually used for founder herds. In these analyses, each founder was associated with its corresponding country of origin, and the genetic contribution of a given "founder country" was obtained by summing the genetic contributions to the reference population of the founders originating from that country. To assess how genetic contributions of different countries to the genetic pool of each breed have changed over time, reference populations were defined in intervals of 10 years, as outlined above for the contributions of founders and ancestors.

All demographic and statistical analyses were carried out within-breed, using the ENDOG V.4.8 software (Gutiérrez and Goyache, 2005). This software computes individual inbreeding coefficients based on the algorithm proposed by Meuwissen and Luo (1992), assigning a null coefficient of inbreeding to animals that do not have both parents known. In our data set the HER herdbook had missing pedigree information in a period of 3 years and some imported animals had limited pedigree information registered. In these conditions, the assumption of a null inbreeding coefficient for animals with unknown parents could underestimate the coefficient of inbreeding, especially for animals born in more recent years. Van Raden (1992) proposed an alternative algorithm to compute inbreeding that incorporates genetic groups for animals with missing parents, assigning them the mean inbreeding coefficient of the corresponding generation. To investigate the influence that incomplete pedigrees in our data set may have had on the computation of inbreeding, the PEDIG software (Boichard, 2002) was used to compute inbreeding with the Van Raden algorithm, forming groups according to year of birth of calves with unknown parents.

**Results**

Pedigree records used in the analyses included information on 175,179, 61,295, 245,942 and 37,751 calves registered for ANG, DEV, HER and SHO, respectively (Table 1). Even though the largest number of registered calves in the whole period studied was found in HER breed, the evolution of the various herdbook registrations differed considerably over time for the 4 breeds analyzed (Figure 1). In the last few years, a noteworthy increment of registered calves has been observed in ANG, followed by DEV and HER with similar numbers, while SHO has progressively reduced its census to an extremely small number of registered new-born calves.

The HER was the predominant British breed in Brazil up until the 1980s, but a steady decline was observed in herdbook registrations after the mid-1960s, when registrations reached a peak of about 6,500 calves per year (Figure 1), while in 2009 the breed accounted for about 2,000 calves recorded. A sharp drop in HER registrations was observed in years 1971-1974, corresponding to a partial loss of herdbook records. When the herdbook returned to its normal activity, a steady decline in registrations was observed, confirming the continuous drop in the census of HER in Brazil. The ANG

breed showed an opposite trend, with a sharp increase in registrations in the 1990s, reaching more than 7,000 calves registered in 2008. As a consequence, ANG is currently the largest British cattle breed in Brazil. Concerning the SHO, the number of annually registered calves never exceeded 1,000, but the census increased up to 1960, and declined thereafter reaching about 100 calves registered in 2009. The DEV has shown a steady but moderate increase in number of registrations since the early 1960s, with about 1,500 calves registered per year after 2000.

Overall, the number of sires registered in Brazil in the period studied averaged about 11,000 in ANG and HER, and nearly 3,000 in DEV and SHO (Table 1). The average number of offspring per sire ranged from 12 to 20, depending on the breed considered, with the lowest values in SHO and the highest in DEV (Table 1). The variability in number of offspring/sire was great, especially in ANG. In all breeds, the majority of the bulls sire less than 50 calves, especially in SHO. On the other hand, nearly 25% of the ANG bulls had more than 300 offspring registered, while the percentage of bulls in this category was about 7.5% for DEV and 9.5% for HER, but only 1.5% for the SHO (results not shown). The average number of offspring per dam ranged from about 2.3 calves in SHO to 3.1 in DEV (Table 1).

Pedigree completeness for the breeds analyzed was assessed by evaluating the proportion of ancestors known in different generations (results not shown). In the whole population, the DEV had the most complete pedigree information (84% of the calves had great-grandparents known), followed by ANG (73%) but pedigree information was less complete for HER (40%) and SHO (32%). More complete pedigrees were seen when only the reference population was considered, such that nearly all calves had parents known in all four breeds, and at least 97% of the calves had registered grandparents (results not shown).

When the percentage of known ancestors per generation is considered for animals in the reference population the lowest values of pedigree depth were observed in SHO, which looking back 10 generations reported only 11% of known ancestors in the reference population (results not shown). At the same number of generations back, the other breeds reported nearly 40% of the ancestors known. The mean number of complete generation equivalents known per breed was lowest for SHO and HER, with a mean of about 2.7 and 3.3, respectively, but was similar for the other two breeds, where animals had, on average, more than 5.7 complete generations known (Table 1). Over time, the number of generation equivalents increased steadily, with a distinct pattern depending on the breed (Figure 2). As previously mentioned, pedigree records for HER were lost for calves born in years 1971-1974, which resulted in a nearly null number of generations known for animals born in this period. The situation improved after this loss, such that the HER had the highest mean number of generation equivalents known for animals born in 2010. For the ANG and DEV there was a gradual increase in pedigree information with year of birth, with a mean of nearly 8.5 generations known for animals born in 2010. The SHO had a more moderate increase in pedigree information over the years, showing about 1.5 generations less than the other breeds for calves born in 2010.

Overall, the mean generation interval (Table 1) was nearly 6.4 years in ANG, and was shorter by about 0.4 years in HER and SHO, and by 0.14 years in DEV. For the different selection paths, generally the dam-daughter path was the longest, except in DEV, where the dam-son path was longest. In most cases, differences among the various selection paths were not large, except for the SHO where son paths had generation intervals shorter than daughter paths by about 0.7 years (results not shown).

Matings among full- and half-sibs were rare in all breeds, while the percentage of registered calves resulting from parent-offspring matings ranged from 0.9% in ANG to 2.4% in SHO (Table 1).

The percentage of inbred animals in the whole population was about 20% in HER and SHO, 60% in ANG and 74% in DEV, with an average inbreeding in inbred animals of about 4.9, 8.7, 2.3 and 3.8%, respectively (Table 1). Although the percentage of inbred animals has increased over time in all breeds, this increase occurred mostly over the last 30 years and nearly all animals born over the last few years are inbred in all the breeds studied (results not shown).

For the whole population, the mean inbreeding coefficient ranged from 1.16% in HER to 2.82% in DEV (Table 1), while the average relatedness was about 2.4% in DEV but below 1% in all other breeds. The average inbreeding of registered calves has shown a very mild increase over time (Figure 3), with an annual rate of inbreeding which ranged from about 0.003%/year in ANG to about 0.054%/year in DEV (Table 1). These moderate trends led to an average inbreeding for calves born in 2010 of about 1.5, 1.8, 3.1 and 4.0% for the ANG, SHO, HER and DEV breeds, respectively.

To investigate the impact of missing pedigree information, an alternative estimate of the individual coefficient of inbreeding was obtained with the Van Raden algorithm as implemented by the PEDIG software (Boichard, 2002), grouping calves with unknown parents by year of birth. The breed means for inbreeding obtained with this procedure (results not shown) were higher by a margin not exceeding 2% of the inbreeding means obtained when a null coefficient of inbreeding is assumed for animals with unknown parents. Given the negligibility of the differences, the original procedure to estimate inbreeding was followed in further analyses.

The number of founders and ancestors represented in each population differed considerably among the 4 breeds studied (Table 1), in part reflecting the differences in number of existing animals in each pedigree data set. However, the DEV had the lowest number of founders and ancestors even though the SHO had the lowest number of registered animals.

The number of ancestors supplying 50% of the gene pool to the whole population was 34 for DEV, 76 for ANG, 142 for SHO and 443 for HER (Table 1). For each breed, the number of founders contributing with 50% of the gene pool was about 2 to 4 times that computed for the number of ancestors, but the ranking of breeds followed a similar pattern. These results indicate that HER has the broadest representation of founders and ancestors and that, in spite of its very small census, the SHO has been able to maintain a certain level of genetic diversity. Conversely, genetic erosion is probably taking place in both ANG and DEV breeds. The reference population was represented by about 12,700 animals in ANG, 2,400 in DEV, 3,800 in HER and less than 300 in SHO (Table 2). In this reference population, the average inbreeding per breed ranged from 1.50% in ANG to 3.92% in DEV, while the percentage of inbred animals was nearly 85% in SHO and above 92% for the other breeds. The average relatedness by breed was below 1% in HER and SHO, about 1.3% in ANG and 3.4% in DEV. When the distribution of animals by different levels of inbreeding was considered (results not shown), the majority of the calves in the general population in all breeds showed a level of inbreeding below 6.25%, such that nearly 6% of the ANG and HER calves, 9% of the SHO and 15% of the DEV had a coefficient of inbreeding above this level.

The $\Delta F_i$ in the reference population ranged from about 0.20% in ANG to 0.54% in DEV, and was close to 0.40% in the HER and SHO (Table 2). Comparatively, The $\Delta F_g$ was very similar for the ANG, but somewhat smaller for the other 3 breeds. The $N_e$

estimated from $\Delta F_i$ ranged from 92 in DEV to 245 in ANG, with intermediate values in HER ($N_e$=129) and SHO ($N_e$=118). In comparison, when $N_e$ computed from $\Delta F_g$, the estimate was very close to the above value in ANG, but it was about 30% larger in DEV and HER, and about one-half in SHO (Table 2).

Of the total number of founders that contributed to the gene pool of British cattle breeds in Brazil (Table 1), about 47% are still represented in the reference populations for ANG and DEV breeds, but only 16% for HER and SHO (Table 2). The representativeness of ancestors is much smaller, with about 22% of all ancestors still represented in the ANG reference population, 31% in DEV, and about 5% in HER and SHO. The cumulative contribution of founders to the gene pool of each breed (Figure 4a) shows a distinct pattern among breeds: HER and ANG have a mild increase in founder contributions, indicating that there is not a predominant group of animals with a major influence on the breed; DEV has an opposite pattern, with the first 10 founders accounting for about 30% of the gene pool; SHO shows an intermediate trend. As a consequence of the different pattern of genetic contributions, the number of founders contributing with 50% of the gene pool was 41 for DEV, 79 for SHO, 164 for HER and 211 for ANG (Table 2).

When compared to founders, fewer ancestors contributed to most of the genetic variation, e.g., the 3 major ancestors contributed nearly 30% of the gene pool in DEV and SHO, and 18% in HER and ANG, while the number of ancestors contributing with 50% of the gene pool was 10 for SHO, 14 for DEV, 25 for HER and 26 for ANG (Table 2, Figure 4b).

The $f_a$ reflects the uneven contribution of ancestors to the genetic pool, and in this case the estimates were about 30 for DEV and SHO, and 60 for ANG and HER (Table 2). The ratio $f_e/f_a$ was highest in ANG and SHO, indicating a stronger bottleneck in the pedigree, and lowest in DEV. Average values for $f_g$ were about 17 in DEV and SHO, and 32 in HER and ANG indicating a lower retention of genetic diversity in the first two breeds. Over the years, the effective number of ancestors has shown a mild decline in DEV, SHO and ANG, while the effective number of founders was relatively stable in DEV and SHO and showed some increase in ANG over the last 20 years (Figure 5). A different pattern was observed in HER, where both the effective number of founders and ancestors has declined steadily since the late 1970s.

The analyses of contributions of different countries of origin of founder animals (Figure 6) showed that, throughout the period studied, there is a clear predominance of DEV genes originating from the UK, with very minor contributions from other countries. For the other breeds the situation has been quite different, with important changes over time in the ranking of country contributions. In ANG, Argentina was the major supplier until the 1970s, while HER and SHO genes to Brazil were mostly from Uruguay. In these three breeds, genes from the USA became predominant from the 1970s up to the present.

## Discussion

The first cattle arrived in Brazil in the 16th century, brought by Portuguese and Spanish settlers, and gave origin to what are currently known as Creole, local or naturalized breeds (Primo, 2004). In the late 18th century, British cattle started spreading throughout the Americas (Martinez et al., 2012), but their expansion in Brazil was only meaningful in the late 19th century, and was mostly limited to the Southern part of the country, where local pasture conditions and subtropical climate favored their adaptation (Cardellino, 2000; Elias, 2006). The development of British cattle in Brazil was largely

based on the Angus, Devon, Hereford and Shorthorn breeds, which have characteristics desired by producers, such as medium-frame body size and early maturing, both in terms of reproduction, growth and finishing (Smith et al., 1976).

The relative importance of British breeds in Brazil has evolved over the years, and their census numbers reflected those changes. In 2012, the number of herds registering animals in each herdbook in Brazil was 512 for Angus, 57 for Devon, 83 for Hereford and 7 for Shorthorn with, respectively, 9,029, 1,867, 1,726 and 93 purebred breeding animals enrolled in each herdbook in that year (ANC, 2012). Nowadays, herdbook information includes records of both imported animals, with their respective country, and of those born in Brazil. The analysis of population structure based on pedigree information has been applied to the genetic characterization of different cattle breeds in Brazil (Faria et al., 2010; Reis Filho et al., 2010; Oliveira et al., 2012; Santana et al., 2012) as well as in several other countries (see review by Carolino and Gama, (2008), and is now a common tool in managing genetic diversity in livestock species (FAO, 2013).

Concerning the breeds of the present study, pedigree structure has been analyzed in other countries for some of them, e.g., the Hereford and Red Angus were studied in the United States by Cleveland et al. (2005) and Márquez et al. (2010), respectively, while McParland et al. (2007) studied the Hereford and Aberdeen Angus in Ireland. Nevertheless, to our knowledge, demographic information is very limited for the DEV and SHO breeds, but they show reduced numbers worldwide, even though they were among the most important cattle breeds at the beginning of the 20th century (Brassley, 2000).

In our work, we analyzed pedigree information for the Angus, Devon, Hereford and Shorthorn breeds in Brazil, as they show a very distinct history in this country, reflecting the evolution of market trends over the years and the consequent decisions taken by farmers. The SHO had some popularity in the 1950's but in general it has been a breed of minor expression in Brazil, while the DEV was able to occupy a niche market, with a slow increase in the number of registered animals since the 1960's. The situation has been quite different for the HER and ANG, which have played a major role in crossbreeding programs in Brazil and have thus expanded their census in the country, even though their relative importance was not the same throughout the period analyzed. The HER was by far the most popular British breed up until the mid-1980s, but from then on its census dropped steadily, while the ANG gained popularity, becoming the major British breed in Brazil over the last decade.

As a result of the accurate recording of genealogical information from the first imports, animals in the reference population had almost 9 complete generation equivalents known, with the exception of the SHO which only had 7, possibly because the widespread use of crossbreeding caused a reduction in pedigree recording. In any case, the pedigree is deep enough to provide a comprehensive picture of the evolution of genetic diversity in the major British cattle breeds raised in Brazil.

For the breeds considered here, herdbooks are open to the access of registered animals imported to Brazil, in contrast to other reports considering herdbooks which were essentially closed (Gutiérrez et al., 2003; Cleveland et al., 2005; McParland et al., 2007; Carolino and Gama, 2008). Hence, the integration of foreign animals in the herdbook over the years in our study is expected to increase genetic variability and diversify the origins represented. This was the case with most breeds, with the exception of DEV, which originated almost exclusively from the UK. For the other breeds, until the mid-1970s the ANG was essentially of Argentinian origin while the HER and SHO were

mostly from Uruguay. In the late 1970s, the source of imported animals changed, with the USA becoming the major supplier of ANG, HER and SHO genetics. This switch probably reflects the change from the import of live animals to be used in natural matings in the early years, which would be mostly from nearby countries, to the widespread use of artificial insemination in the mid-1970s and, to a lesser extent, embryo transfer that facilitated the direct import of semen and embryos from the USA.

The genetic contributions of the major founders and ancestors to the reference population showed two distinct patterns, i.e., the DEV and SHO with a more pronounced contribution of a few major ancestors when compared with ANG and HER. As a consequence of the unbalanced representation of ancestors, the estimated $f_a$ was nearly 30 for the DEV and SHO breeds, which is in line with estimates reported for endangered beef breeds such as the Grauvieh (Sölkner et al., 1998) and Sayaguesa (Gutiérrez et al., 2003), and even for some dairy breeds where high selection intensity is applied (Boichard et al., 1997; Sørensen et al., 2005). For the ANG and HER breeds in our study, the $f_a$ was about 60, which is almost twice the estimates published for the same breeds in Ireland (McParland et al., 2007) but is lower than the estimated $f_a$ for Limousine and Charolais cattle in different European countries (Bouquet et al., 2011). Still, the estimated $f_a$ was fairly high in all the breeds included in our study.

Statistics such as the ratios between founders, ancestors and founder genomes revealed the occurrence of pedigree bottlenecks in all breeds, which are confirmed by a ratio $f_e/f_a$ >1 (Sørensen et al., 2005). This occurred in spite of the fact that herdbooks remained open to the access of imported animals, and is a consequence of the small population size of DEV and SHO, as well as of the extensive use of some popular sires in ANG and HER. Nearly all animals in the reference population had some degree of inbreeding in all four breeds, probably as a consequence of the strong pedigree depth, implying that at some point in the pedigree the ancestors might be related. Nevertheless, the overall level of inbreeding was moderate in all breeds, with means in the reference population below 4% in all breeds. These moderate levels of inbreeding are confirmed by the small number of animals in the reference population with levels of inbreeding exceeding 6.25%, and by the reduced incidence of matings among close relatives in all breeds. The means for inbreeding in ANG and HER in our study are slightly higher than those reported for the same breeds in Ireland (McParland et al., 2007), and for the Red Angus in the United States (Marquez et al., 2010), but much lower than the estimate of about 10% in American Hereford (Cleveland et al., 2005). For the DEV, our estimate was close to the mean inbreeding of nearly 4.4% reported for the American Milking Devon by Splan and Sponenberg (2003), while for SHO our results were lower than the estimate of 5.5% reported for the American dairy Shorthorn (AIPL, 2013). In Brazil, the low levels of inbreeding found in ANG and, to a lesser extent, in HER, can be justified by some continuous influx of imported animals of diverse origins. The DEV was the breed with the highest level of inbreeding and mean relatedness in the reference population, and it had the highest rate of inbreeding and the lowest $f_e$ among the breeds studied. This could be a result of the narrow recruitment of DEV overseas, mostly from the United Kingdom, where it is considered to be an endangered breed (DEFRA, 2013). It is, therefore, reasonable to consider the DEV as the British breed undergoing the highest level of genetic erosion in Brazil, even though it is not the breed with the smallest census. One unexpected result in our study is the low mean inbreeding in SHO, as this is a breed with very small census, both in Brazil and in countries supplying SHO germplasm. However, Shorthorn Associations in North America have allowed the registration of animals resulting from crossbreeding with, e.g., Red Holstein or Swedish

Red (ASA, 2013; CMSS, 2013), which should maintain inbreeding at low levels in Brazil when semen from these animals is imported. As Brazilian SHO has been kept mostly free from the influence of other breeds, it could be an interesting source population for other countries where SHO is now experiencing genetic erosion.

The negative consequences of inbreeding in cattle are well known (Burrow, 1993) and minimizing the rate of inbreeding is often the major objective in conservation of within-breed genetic diversity (Hill, 2000; Caballero and Toro, 2002). In this perspective, a maximum rate of inbreeding of 1% per generation, corresponding to a minimum effective population size of 50, has been recommended as an adequate target to maintain genetic diversity in conservation (Meuwissen and Woolliams, 1994; FAO, 2013) and selection programs (Goddard and Smith, 1990), but higher thresholds have been in some cases recommended, as in cattle breeds where artificial insemination is common (Leroy et al., 2013). Nevertheless, in all breeds included in our study the rate of inbreeding was below the critical threshold of 1% (FAO, 1998), and consequently $N_e$ was always above 50, indicating that a possible loss of genetic diversity is not presently a matter of concern in the breeds studied. Still, regardless of the method used for computation, the rate of inbreeding was higher in DEV, with estimates ranging from about 0.39 to 0.54% per generation. The $N_e$, computed from the rate of inbreeding estimated according to different methods, ranged between 92 and 128 for the DEV, and the estimates were always above 100 for the remaining breeds, tending to be higher in ANG. Recently, Leroy et al. (2013) used different approaches to estimate $N_e$ in various dog, sheep, cattle and horse breeds, and concluded that the consistency of estimates depended on the species considered and the specific genetic structure of the population analyzed, especially when methods based on the evolution of coancestry or inbreeding were compared. In our case, $N_e$ was estimated from the rate of inbreeding per generation, either $\Delta F_g$ or $\Delta F_i$, and the comparison is not straightforward, because the former only takes into account the mean level of inbreeding in two points in time, while the latter considers the evolution of inbreeding throughout the period analyzed, and should thus be more reliable.

In conclusion, our estimates point to a mild increase in inbreeding in the 4 British breeds studied, with estimates of effective population size indicating that genetic diversity is being maintained at a reasonable level in all breeds, above the recommended minimal critical threshold (FAO, 1998). Nevertheless, the genetic structure of the breeds analyzed is somewhat different. The ANG had a major increase in its census in the 1990s and is now the leading British breed in Brazil, with a continuous infusion of imported germplasm. This has led to low levels of inbreeding, in spite of some bottlenecks in pedigrees due to the heavy use of a few selected sires. The HER was the major British breed in Brazil until the 1980s, when ANG gained the leadership. Notwithstanding its lower census, the HER still has a high effective population size and a good representation of founders and ancestors, thus showing indications of good management of the population gene pool. The DEV has, for many years, maintained a small census, which has shown a very mild increase since the 1960s. Among the British breeds studied, DEV is the one with the highest rate of inbreeding, and the low effective number of founders and ancestors suggests that imported germplasm, which is exclusively of British origin, should be diversified, to avoid bottlenecks in the population. The SHO is a breed with a rather small population size in Brazil, but inbreeding has been kept under control, possibly due to the common practice of importing crossbred animals. Overall, results of this study indicate that, so far, the continuous flow of imported genes has contributed to the maintenance of genetic

variability in all breeds studied, but continuous monitoring of inbreeding trends and of parameters derived from probability of gene origin are of primary importance to ensure breed conservation and a long term maintenance of genetic diversity.

**Literature Cited**

AIPL. 2013. Animal Improvement Programs Laboratory. Inbreeding coefficients for milking Shorthorn cows. http://aipl.arsusda.gov/dynamic/inbrd/current/MSt.html. (Accessed 06 August 2013.)

ANC. 2012. Associação Nacional de Criadores - Herd Book Colares. Relatório estatístico anual 2012. http://www.herdbook.org.br/conteudo/relatorios/Relatorio_de_Atividades_do_Ano _de_2012.pdf (Accessed 06 December 2013.)

ANUALPEC. 2011. Anuário da pecuária brasileira. 2011. FNP Prol Editora Gráfica. São Paulo, Brazil.

ASA. 2013. American Shorthorn Association - Shorthorn breed information. http://www.shorthorn.org/breedinfo/about/history_about_breedinfo.html. (Accessed 06 August 2013.)

Boichard, D. 2002. Pedig: a Fortran package for pedigree analysis suited for large populations. Proc 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. 33: 755-756.

Boichard, D., L. Maignel, and E. Verrier 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol. 29:5–23.

Bouquet, A., E. Venot, D. Laloë, F. Forabosco, A. Fogh, T. Pabiou, K. Moore, J. -Å. Eriksson, G. Renand, and F. Phocas. 2011. Genetic structure of the European Charolais and Limousin cattle metapopulations using pedigree analyses. J. Anim. Sci. 89:1719–1730.

Brassley, P. 2000. Livestock breeds. In The Agrarian History of England and Wales. E. J. T. Collins. ed. Cambridge University Press. Cambridge, U.K.

Burrow, H. M. 1993. The effects of inbreeding in beef cattle. Anim. Breed. Abstr. 61:737–751.

Caballero, A., and M. A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3:289–299.

Cardelino, R. A. 2000. Animal genetic resources in southern Brazil. Arch. Zootec. 49:327-331.

Carolino, N., and L. T. Gama. 2008. Indicators of genetic erosion in an endangered population: the Alentejana cattle breed in Portugal. J. Anim. Sci. 86:47–56.

Cleveland, M. A., H. D. Blackburn, R. M. Enns, and D. J. Garrick. 2005. Changes in inbreeding of U.S. Herefords during the twentieth century. J. Anim. Sci. 83:992–1001.

CMSS. 2013. Canadian Milking Shorthorn Society - The breed. http://www.cmss.on.ca/breed.php. (Accessed 06 August 2013.)

DEFRA. 2013. Department for Environment, Food and Rural Affairs. UK country report on farm animal genetic resources 2012. https://www.gov.uk/government/uploads/ system/uploads/attachment_data/file/191781/pb13938-fangr-country-report-2012.pdf. (Accessed 06 August 2013.)

Elias, A. C. 2006. O centenário do Herd-Book Collares, 1th ed. Futura.rs, Porto Alegre, Brasil.

Falconer, D. S., and T. F. C. MacKay. 1996. Introduction to Quantitative Genetics, 4th ed. Longmans Green, Harlow, Essex, U.K.

FAO. 1998. Secondary guidelines for development of national farm animal genetic resources management plans: management of small populations at risk. FAO, Rome, Italy.

FAO. 2013. In vivo conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14. FAO, Rome, Italy.

Faria, L. C., S. A. Queiroz, P. A. Vozzi, R. B. Lôbo, C. U. Magnabosco, and J. A. Oliveira. 2010. Variabilidade genética da raça Brahman no Brasil detectada por meio de análise de pedigree. Pesq. Agrop. Bras. 45:1133–1140.

Goddard, M. G., and C. Smith. 1990. Optimum number of bull sires in dairy cattle breeding. J. Dairy Sci. 73:1113–1122.

Gutiérrez, J. P., J. Altarriba, C. Díaz, R. Quintanilla, J. Cañón, and J. Piedrafita. 2003. Pedigree analysis of eight Spanish beef cattle breeds. Genet. Sel. Evol. 35:43–63.

Gutiérrez, J. P., and F. Goyache. 2005. A note on ENDOG: a computer program for analysing pedigree information. J. Anim. Breed. Genet. 122:172–176.

Gutiérrez, J. P., I. Cervantes, and F. Goyache. 2009. Improving the estimation of realized effective population sizes in farm animals. J. Anim. Breed. Genet. 126:327–332.

Hermuche, P., R. F. Guimarães, O. A. Carvalho, R. A. T. Gomes, S. R. Paiva, and C. M. McManus. 2013. Environmental factors that affect sheep production in Brazil. Appl. Geog. 44, 172–181.

Hill, W. G. 2000. Maintenance of quantitative genetic variation in animal breeding programmes. Livest. Prod. Sci. 63:99–109.

James, J. W. 1972. Computation of genetic contributions from pedigrees. Theor. Appl. Genet. 42:272–273.

Lacy, R. C. 1989. Analysis of founder representations in pedigrees: Founder equivalents and founder genome equivalents. Zoo Biol. 8:111–123.

Leroy, G., T. Mary-Huard, E. Verrier, S. Danvy, E. Charvolin, and C. Danchin-Burge. 2013. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and house. Genet. Sel. Evol. 45:1-10.

MacCluer, J. W., J. L. VandeBerg, B. Read, and O. A. Ryder. 1986. Pedigree analysis by computer simulation. Zoo Biol. 5:147–160.

Mariante, A. S., C. McManus, and J. F. Mendonça. 2003. Country report on the state of animal genetic resources Brasil. Documents Embrapa Genetic Resources and Biotechnology. No.99. Brasília, Brazil.

Márquez, G. C., S. E. Speidel, R. M. Enns, and D. J. Garrick, D. J. 2010. Genetic diversity and population structure of American Red Angus cattle. J. Anim. Sci. 88:59–68.

Martínez, A. M., L. T. Gama, J. Cañón, C. Ginja, J. V. Delgado, S. Dunner, V. Landi, I. Martín-Burriel, M. C. T. Penedo, C. Rodellar, J. L. Vega-Pla, A. Acosta, L. A. Alvarez, E. Camacho, O. Cortés, J. R. Marques, R. Martínez, R. D. Martínez, L. Melucci, G. Martínez-Velázquez, J. E. Muñoz, A. Postiglioni, J. Quiroz, P. Sponenberg, O. Uffo, A. Villalobos, D. Zambrano, and P. Zaragoza. 2012. Genetic footprints of Iberian cattle in America 500 years after the arrival of Columbus. PLoS ONE, 7:e49066.

Mc Parland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. J. Anim. Sci. 85:322–331.

Meuwissen, T. H. E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24:305-313.

Meuwissen, T. H. E., and J. A. Woolliams. 1994. Effective sizes of livestock populations to prevent a decline in fitness. Theor. Appl. Genet. 89:1019–1026.

Oliveira, A. P., C. H. M. Malhado, P. L. S. Carneiro, R. M. Filho, É. S. Silveira, L. A. Souza, L. M. S. Muniz, and D. M. M. R. Azevêdo. 2012. Pedigree analysis on the population of Gir cattle in northeast Brazil. R. Bras. Zootec. 41:1153–1157.

Primo, A. T. 2004. América: conquista e colonização - A fantástica história dos conquistadores ibéricos e seus animais na era dos descobrimentos. ed. Movimentos. Porto Alegre, Brazil.

Reis Filho, J. C., P. S. Lopes, R. S. Verneque, R. A. Torres, R. L. Teodoro, and P. L. S. Carneiro. 2010. Population structure of Brazilian Gyr dairy cattle. R. Bras. Zoot. 39:2640–2645.

Santana, M. L., P. S. Oliveira, J. P. Eler, J. P. Gutiérrez, and J. B. S. Ferraz. 2012. Pedigree analysis and inbreeding depression on growth traits in Brazilian Marchigiana and Bonsmara breeds. J. Anim. Sci. 90:99–108.

Smith, G. M., H. A. Fitzhugh Jr, L. V. Cundiff, T. C. Cartwright, and K. E. Gregory. 1976. A genetic analysis of maturing patterns in straightbred and crossbred Hereford, Angus and Shorthorn cattle. J. Anim. Sci. 43:389–395.

Sölkner, J., L. Filipcic, and N. Hampshire. 1998. Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. Anim. Sci. 67:249–256.

Sørensen, A. C., M. K. Sørensen, and P. Berg. 2005. Inbreeding in Danish dairy cattle breeds. J. Dairy Sci. 88:1865–1872.

Splan, R. K., and D. P. Sponenberg. 2003. Characterization and conservation of the American Milking Devon. Anim. Genet. Res. Inform. 34:11–16.

Van Raden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J. Dairy Sci. 75:3136-3144.

Van Vleck, L. D. 1993. Selection index and introduction to mixed model methods. CRC Press, Boca Raton, FL.

Table 1. Summary statistics from pedigree analyses of the whole population for the Angus, Devon, Hereford and Shorthorn breeds.

| Parameter | Angus | Devon | Hereford | Shorthorn |
|---|---|---|---|---|
| Total number of animals | 175,179 | 61,295 | 245,942 | 37,751 |
| Number of calves | 106,144 | 39,190 | 144,770 | 20,996 |
| Number of sires | 10,876 | 2,972 | 12,247 | 2,708 |
| Number of dams | 58,159 | 19,133 | 88,925 | 14,047 |
| Average number of offspring per bull | 15.19±62.20 | 19.97±42.26 | 17.81±48.22 | 12.12±26.86 |
| Average number of offspring per dam | 2.84±2.75 | 3.10±2.22 | 2.45±1.75 | 2.33±1.71 |
| Mean equivalent number of complete generations | 5.70 | 5.83 | 3.27 | 2.67 |
| Mean generation interval | 6.42±3.34 | 6.28±3.27 | 6.04±2.70 | 6.03±7.14 |
| Matings between sibs (%) | 0.03 | 0.08 | 0.01 | 0.09 |
| Parent-offspring matings (%) | 0.92 | 1.61 | 1.59 | 2.39 |
| Average inbreeding coefficient (%) | 1.38±3.46 | 2.82±4.81 | 1.16±3.93 | 1.60±5.16 |
| Inbred animals (%) | 60.68 | 73.58 | 23.90 | 18.55 |
| Average inbreeding coefficient for inbred animals (%) | 2.27±4.20 | 3.83±5.25 | 4.86±6.82 | 8.65±9.10 |
| Maximum inbreeding coefficient (%) | 39.93 | 42.94 | 43.75 | 47.66 |
| Animals with inbreeding coefficient above 6.25%(%) | 6.13 | 14.63 | 6.01 | 9.14 |
| Average relatedness (%) | 0.89±0.59 | 2.42±1.57 | 0.25±0.34 | 0.27±0.23 |
| Rate of inbreeding/year (%) | 0.0033±0.0005 | 0.0538±0.0013 | 0.0254±0.0005 | 0.0165±0.0015 |
| Number of founders | 10,194 | 1,977 | 27,867 | 4,980 |
| Number of ancestors | 8,616 | 1,751 | 25,434 | 4,268 |
| Number of founders explaining 50% of the gene pool | 330 | 74 | 1,089 | 319 |
| Number of ancestors explaining 50% of the gene pool | 76 | 34 | 443 | 142 |

Table 2. Summary statistics from pedigree analyses of the reference populations[a] for the Angus, Devon, Hereford and Shorthorn breeds.

| Parameter | Angus | Devon | Hereford | Shorthorn |
|---|---|---|---|---|
| Number of animals in the reference population[a] | 12,703 | 2,375 | 3,833 | 287 |
| Mean equivalent number of complete generations | 8.43 | 8.49 | 8.98 | 7.00 |
| Average inbreeding coefficient % | 1.50±2.52 | 3.92±4.69 | 2.99±4.01 | 2.37±3.64 |
| Average inbreeding coefficient for inbred animals (%) | 1.57±2.56 | 4.22±4.74 | 3.06±4.03 | 2.81±3.81 |
| Maximum inbreeding coefficient (%) | 37.58 | 42.94 | 30.98 | 25.02 |
| Inbred animals (%) | 96.52 | 92.84 | 98.02 | 84.67 |
| Average relatedness (%) | 1.29±0.39 | 3.38±0.99 | 0.68±0.18 | 0.41±0.13 |
| Rate of inbreeding/generation ($\Delta F_g$)[b] (%) | 0.214 | 0.390 | 0.270 | 0.165 |
| Rate of inbreeding/generation ($\Delta F_i$)[c] (%) | 0.204 | 0.542 | 0.388 | 0.424 |
| Effective population size from $\Delta F_g$ | 234 | 128 | 185 | 303 |
| Effective population size from $\Delta F_i$ | 245 | 92 | 129 | 118 |
| Number of founders | 4,903 | 913 | 4,372 | 780 |
| Number of ancestors | 1,888 | 546 | 1,374 | 156 |
| Number of founders explaining 50% of the gene pool | 211 | 41 | 164 | 79 |
| Number of ancestors explaining 50% of the gene pool | 26 | 14 | 25 | 10 |
| Effective number of founders ($f_e$) | 470 | 89 | 289 | 200 |
| Effective number of ancestors ($f_a$) | 68 | 33 | 59 | 28 |
| $f_e/f_a$ ratio | 6.91 | 2.70 | 4.90 | 7.14 |
| Founder genome equivalents ($f_g$) | 36 | 16 | 30 | 18 |
| $f_e/f_g$ ratio | 13.06 | 5.56 | 9.63 | 11.11 |

[a] Reference population defined as the group of calves born in 2009 and 2010 for the Angus, Devon and Hereford breeds, and calves born in 2008, 2009 and 2010 for the Shorthorn breed.
[b] Obtained from the evolution of inbreeding in the period between 2002 and 2009.
[c] Obtained from the individual increase in inbreeding, as in Gutiérrez et al. (2009).
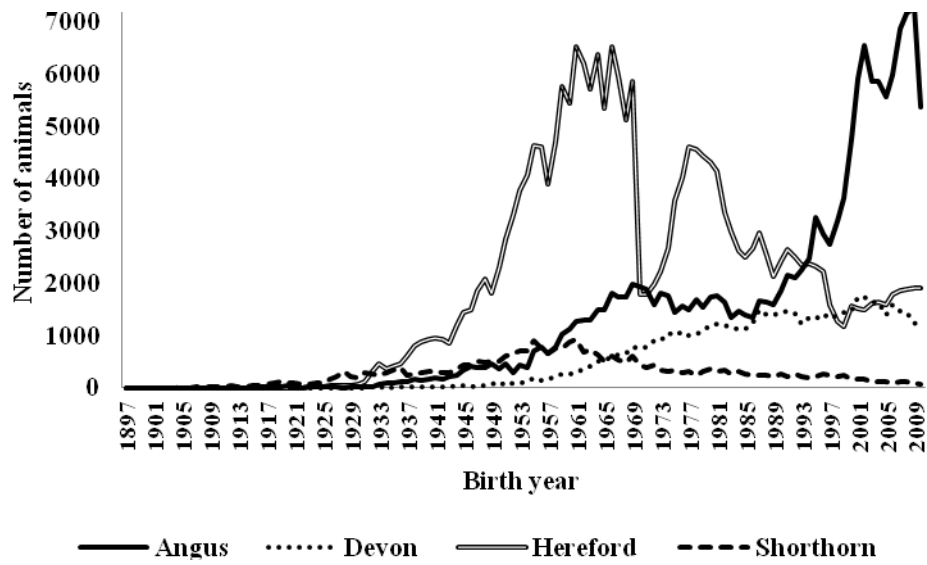
**Figure 1.** Number of calves of the Angus, Devon, Hereford and Shorthorn breeds registered in the herdbook, by year of birth.
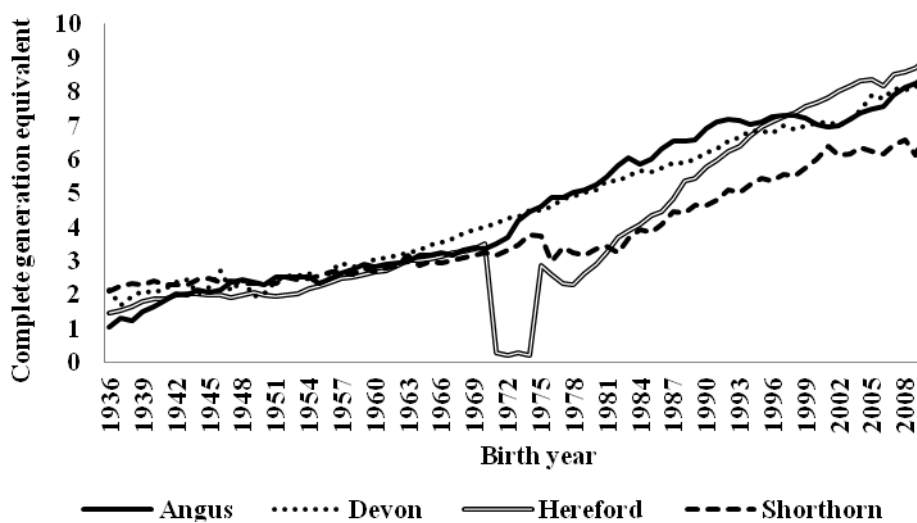
**Figure 2.** Level of pedigree completeness for Angus, Devon, Hereford and Shorthorn calves, by year of birth.
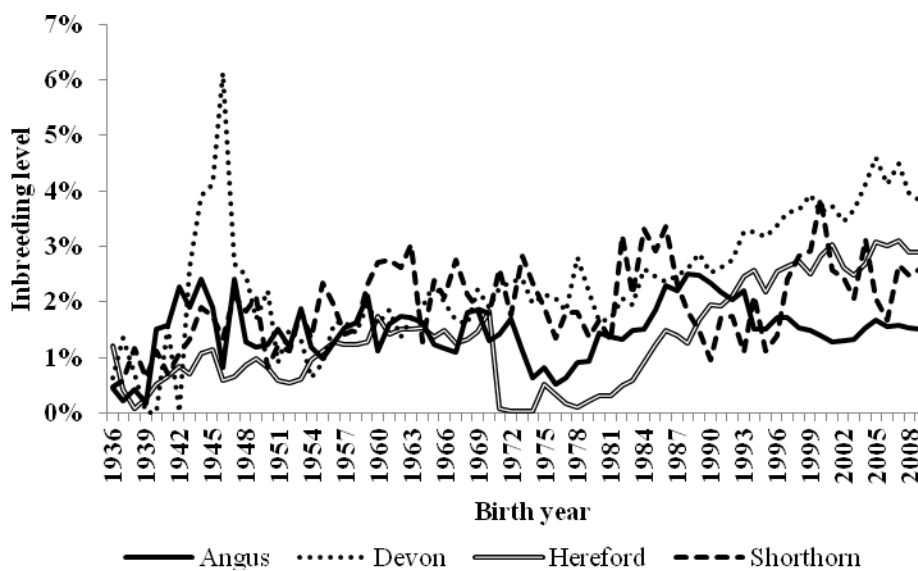


**Figure 3.** Average inbreeding by year of birth for the Angus, Devon, Hereford and Shorthorn breeds.
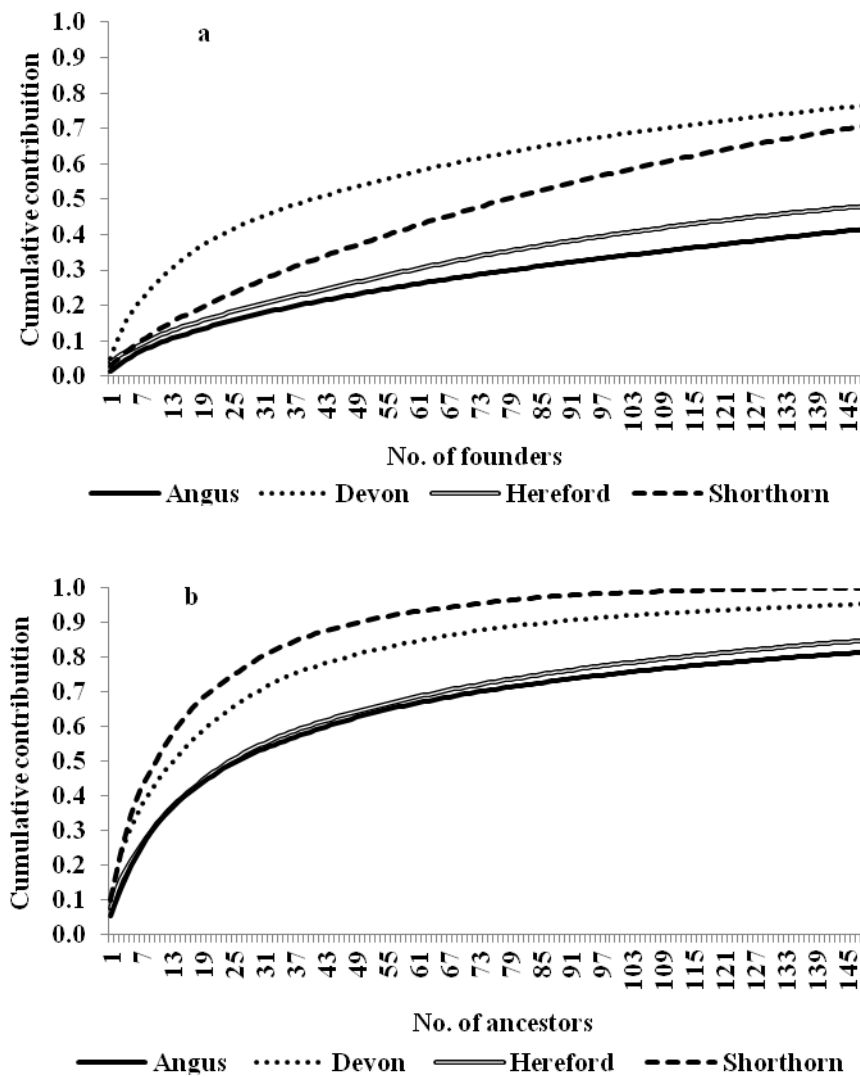
**Figure 4.** Cumulative genetic contribution to the reference population of the 150 most influential a) founders and b) ancestors, for the Angus, Devon, Hereford and Shorthorn breeds.
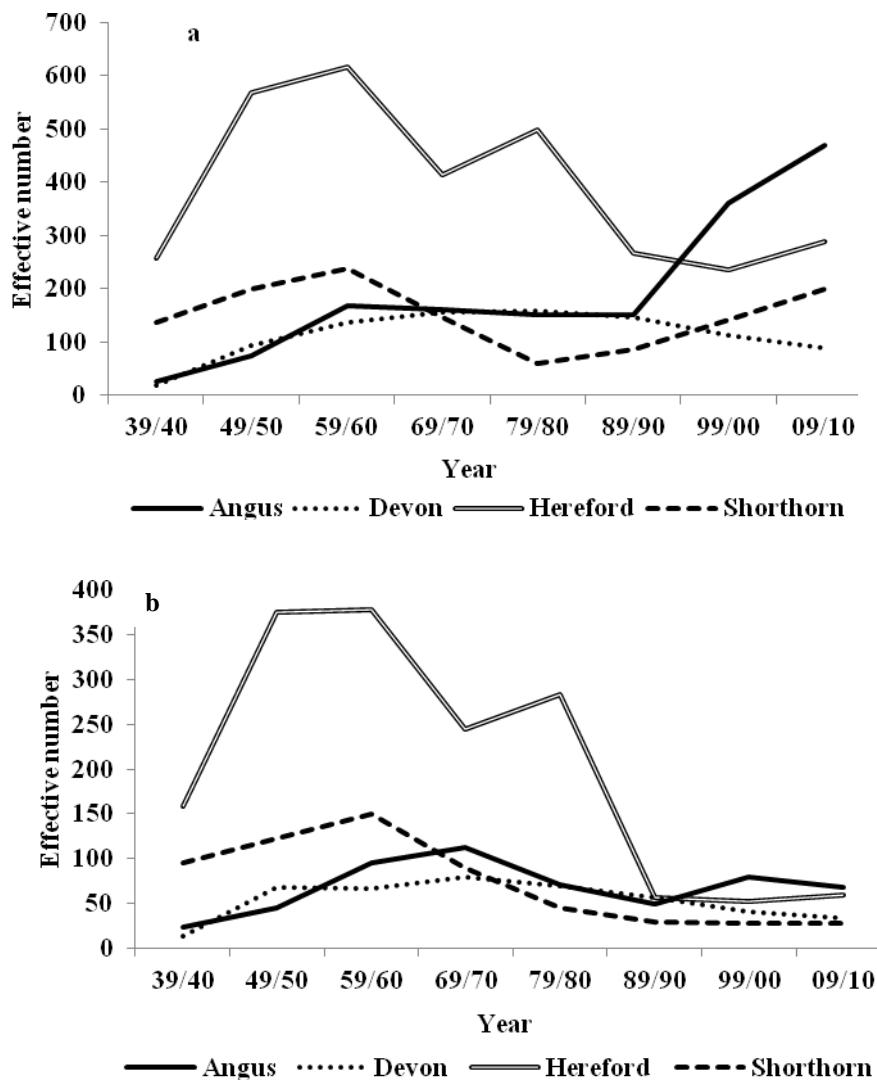
**Figure 5.** Effective number of founders (a) and ancestors (b) by year of birth for the Angus, Devon, Hereford and Shorthorn breeds, shown in 10 year-intervals.
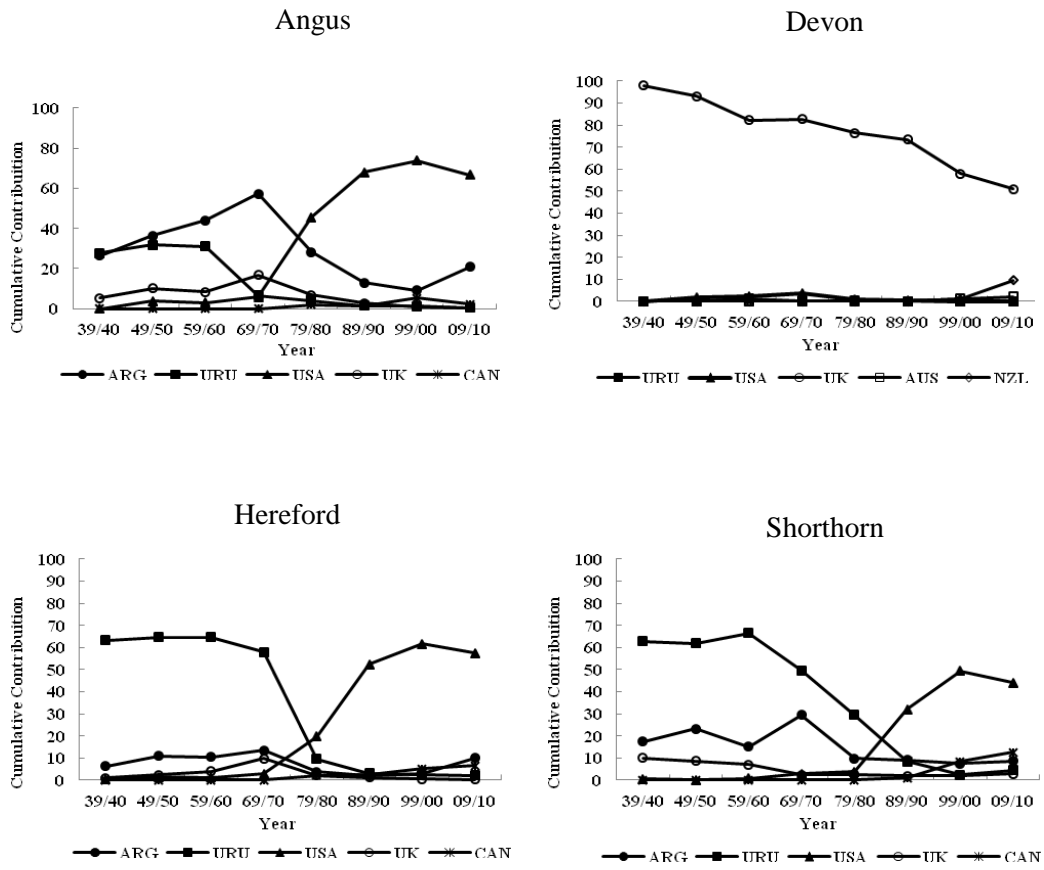
**Figure 6.** Genetic contribution of different countries of origin (ARG: Argentina; AUS: Australia; CAN: Canada; NZL: New Zealand; UK: United Kingdom; URU: Uruguay; USA: United States of America) to the genetic pool of the Angus, Devon, Hereford and Shorthorn breeds, by year of birth.

# Accuracy of genome-wide imputation in Braford and Hereford beef cattle

**Mario L Piccoli[1,2,3], José Braccini[1,5], Fernando F Cardoso[4,5], Medhi Sargolzaei[3,6], Steven G Larmer[3], Flávio S Schenkel[3]§**

[1] Universidade Federal do Rio Grande do Sul, Departamento de Zootecnia, Porto Alegre, Brazil.

[2] GenSys Consultores Associados S/S, Porto Alegre, Brazil.

[3] Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada.

[4] Embrapa Southern Region Animal Husbandry, Bagé, Brazil.

[5] National Council for Scientific and Technological Development, Brasília, Brazil

[6] The Semex Alliance, Guelph, Canada.

§ Corresponding author: Flávio S. Schenkel (schenkel@uoguelph.ca)

Email addresses:

      MLP: mariopiccoli@gensys.com.br

      JB: jose.braccini@ufrgs.br

      FFC: fernando.cardoso@embrapa.br

      MS: msargol@uoguelph.ca

      SL: slarmer@uoguelph.ca

      FSS: schenkel@uoguelph.ca

## Abstract

**Background:** Strategies for imputing genotypes from the Illumina-Bovine3K, Illumina-BovineLD (6K), BeefLD-GGP (8K), a non-commercial-15K and IndicusLD-GGP (20K) to either Illumina-BovineSNP50 (50K) or to Illumina-BovineHD (777K) SNP panel, as well as for imputing from 50K, GGP-IndicusHD (90iK) and GGP-BeefHD (90tK) to 777K were investigated. Imputation of low density (<50K) genotypes to 777K was carried out in either one or two steps. Imputation of ungenotyped parents (n=37 sires) with four or more offspring to the 50K panel was also assessed. There were 2,946 Braford, 664 Hereford and 88 Nellore animals, from which 71, 59 and 88 were genotyped with the 777K panel, while all others had 50K genotypes. The reference population was comprised of 2,735 animals and 175 bulls for 50K and 777K, respectively. The low density panels were simulated by masking genotypes in the 50K or 777K panel for animals born in 2011. Analyses were performed using both Beagle and FImpute software. Genotype imputation accuracy was measured by concordance rate and allelic $R^2$ between true and imputed genotypes.
**Results:** The average concordance rate using FImpute was 0.943 and 0.921 averaged across all simulated low density panels to 50K or to 777K, respectively, in comparison with 0.927 and 0.895 using Beagle. The allelic $R^2$ was 0.912 and 0.866 for imputation to 50K or to 777K using FImpute, respectively, and 0.890 and 0.826 using Beagle. One and two steps imputation to 777K produced averaged concordance rates of 0.806 and 0.892 and allelic $R^2$ of 0.674 and 0.819, respectively. Imputation of low density panels to 50K, with the exception of 3K, had overall concordance rates greater than 0.940 and allelic $R^2$ greater than 0.919. Ungenotyped animals were imputed to 50K panel with an average concordance rate of 0.950 by FImpute.
**Conclusion:** FImpute accuracy outperformed Beagle on both imputation to 50K and to 777K. Two-step outperformed one-step imputation for imputing to 777K. Ungenotyped animals that have four or more offspring can have their 50K genotypes accurately inferred using FImpute. All low density panels, except the 3K, can be used to impute to the 50K using FImpute or Beagle with high concordance rate and allelic $R^2$.

**Keywords:** Braford, Imputation accuracy, Low density panel, Hereford, High density panel.

## Background

Traditional animal breeding methods utilized phenotypic data and relationships among individuals to make informed mating decision to improve traits of economic significance. Recent advances in DNA technology, led to the full sequencing of several species, including cattle [1] and to the development of new genomic technologies. SNP genotyping is now possible at a cost reasonable for producers. This includes the Illumina BovineHD (Illumina Inc., San Diego, USA), that makes it possible to genotype 777,962 SNPs in a single chip. The first panel of medium density for bovine was the Parallel 10K SNP released in 2006 by the Parallel Company. In 2007, the Illumina Inc., San Diego, USA developed the Illumina BovineSNP50 panel with 54,609 SNPs and in 2011 it released the Illumina BovineHD panel with 777,962 SNPs. These new genotyping technologies have stimulated the development of new research areas, including techniques to infer SNPs on high density genotype panels for animals that have been genotyped at a lower density.

Procedures for imputation of genotypes, a technique that refers to prediction of ungenotyped SNP genotypes, have been the subject of recent studies in some species, such as, dairy cattle [2], [3], beef cattle [4], [5], horse [6] and pig [7]. Software programs have been developed to more efficiently and accurately impute high density genotypes [8], [9], [10], [11], [12]. Density of markers genotyped affects genomic selection accuracy [13], [14], [15], and to reduce the cost of genotyping large populations, less dense, less expensive panels can be used and imputation can infer a more dense genotype, enabling broader uptake of genotyping technology by cattle producers [16], [17]. The evolution of genotyping technology has resulted in many animals of different breeds being genotyped with a variety of SNP panels. For effective genomic selection, all animals should have genotypes of equivalent density. It has been shown that there is a need to evaluate different panels for imputation to higher density panels. Imputation also eliminates the need for re-genotyping of key animals, reducing costs of genomic selection and association analysis.

The Brazilian cattle industry plays a significant role in the national economy. Brazil has a herd of more than 211 million cattle of which 80% is zebu cattle [18]. Hereford and Braford breeds, together with Angus and Brangus account for 50% of the approximate 8 million doses of beef cattle semen commercialized in Brazil in 2013 [19]. Much of this semen, as well as most live bulls sold are mated to Zebu females with the primary objective of improving carcass quality [20].

The main objective of this research was to assess accuracy of imputation from lower density SNP panels to genotypes from the Illumina BovineSNP50 and the Illumina BovineHD panels (Illumina Inc., San Diego, USA) in Brazilian Braford and Hereford cattle.

## Methods

### Animal welfare

Animal welfare and use committee approval was not necessary for this study because data were obtained from existing databases.

### Data

Data was from the Conexão Delta G's genetic improvement program - Hereford and Braford (Zebu x Hereford) cattle (Conexão Delta G, Dom Pedrito/RS, Brazil), containing approximately 520,000 animals from 97 farms located in the South, Southeast, Midwest and Northeast regions of Brazil. A total of 683 Hereford and 2,997 Braford animals from these farms were genotyped. Of the genotyped animals, there were 624 Hereford and 2,926 Braford animals genotyped with the Illumina BovineSNP50 panel, and 59 Hereford and 71 Braford animals genotyped with the Illumina BovineHD panel from 17 farms located in the South of Brazil. Data also included 88 Nellore bulls from the Paint Program (Lagoa da Serra, Sertãozinho/SP, Brazil) genotyped with the Illumina BovineHD panel.

### Data editing

For imputation to the 50K SNP panel, animals genotyped with 777K SNP genotypes had SNPs not contained on the 50K SNP panel removed. This resulted in a population of 3,768 animals genotyped for 49,345 SNPs. Sites were filtered for GenCall score (>=0.15) [21], [22], Call Rate (>=0.90) [21], [22] and Hardy-Weinberg Equilibrium (P>=$10^{-6}$) [23], [24]. Only autosomes were considered [3], [4]. The

individual sample quality control considered GenCall Score (>=0.15) [21], [22], Call Rate (>=0.90) [21], [22], heterozygosity deviation [21] (limit of ± 3 SD), repeated sampling and paternity errors [22]. After quality control, 3,698 animals and 43,248 SNP were used for further analysis.

For imputation to the 777K SNP panel, only the animals genotyped with the 777K SNP panel could be used as reference. The SNP quality control was the same as for the imputation to the 50K SNP panel (SNP in the 50K panel that were not in common with the 777K were also removed from 50K). After the quality control, 218 bulls (Hereford=59, Braford=71, Nellore=88) and 587,620 SNPs remained.

Table 1 shows the numbers of genotyped animals after data editing as well as the pedigree structure of the genotyped animals.

**Reference and imputation populations**

For imputation to the 50K SNP panel, the dataset was split into two populations. The imputation population was comprised of all animals born in 2011. The remainder of the population was assigned to the reference population for imputation. This division resulted in 2,735 animals in the reference population when Nellore animals were included and 2,647 when Nellore animals were not included. A total of 963 animals were sorted into the imputation population.

Hereford and Braford animals in the reference population included 129 sires born before 2008 and 2,518 animals born between 2008 and 2010. From these 2,518 animals, 3.8% had at least one genotyped offspring.

For animals in the imputation population, the 3K, 6K, 8K, 15K and 20K low density SNP panels were created by masking the non-overlapping SNP between the 50K SNP panel and each of these SNP panels. The imputation population included 33 animals with two parents genotyped and 308 animals with one parent genotyped. Moreover, 52% of the imputation animals were offspring of multiple sire matings.

The data set for imputation to the 777K SNP panel contained 71, 59 and 88 Braford, Hereford and Nellore animals, respectively. The strategy used to test the imputation was to create three different data sets randomly alternating animals in the reference population and in the imputation population, always keeping the Nellore animals in reference population as the objective was to test the imputation accuracy of Braford and Hereford cattle. Each reference population was composed by 175 animals (88 Nellore plus 87 Hereford and Braford animals) and each imputation population had 43 Hereford and Braford animals. For animals in the imputation population the 3K, 6K, 8K, 15K, 20K, 50K, 90iK and 90tK SNP panels were created by masking non-overlapping SNP from 777K SNP panel.

All panels, but one, were commercial panels: Illumina Bovine3K (3K), Illumina BovineLD (6K), Illumina BovineSNP50 (50K) and Illumina BovineHD (777K) panels (Illumina Inc., San Diego, USA), Beef LD GGP (8K), Indicus LD GGP (20K), GGP Taurus HD (90tK) and GGP Indicus HD (90iK) panels (Gene Seek Inc., Lincoln, USA) (Table 2).

All the SNPs from 8K SNP panel were part of the customized 15K SNP panel. The remaining SNPs (7K) were selected from the 50K SNP panel using high minor allele frequency, low linkage disequilibrium, and location (approximately evenly spaced between two SNPs in the 8K SNP panel) as selection criteria. The best possible threshold values to meet the three criteria were a minor allele frequency greater than 0.23 and a linkage disequilibrium, as measured by $r^2$, less than 0.088.

**Imputation scenarios**

For imputation to the 50K SNP panel, four different scenarios were explored as follows: including Nellore genotypes in the reference population and either including pedigree information (NE-P) or not including pedigree information (NE-NP); not including Nellore genotypes in the reference population and either including pedigree information (NNE-P) or not including pedigree information (NNE-NP).

For imputation to the 777K SNP panel, a third set of Hereford and Braford bulls were imputed in four different scenarios: including Nellore genotypes and pedigree information in the reference population (NE-P) or including Nellore genotypes and not including pedigree information in the reference population (NE-NP). Each of these two scenarios was carried out in one or two steps. Two-step imputation was carried out only for panels with density less than 50K SNP. Two-step imputation involved: 1) in the first step, the animals genotyped with 3K, 6K, 8K, 15K and 20K SNP panels were imputed to the 50K SNP panel using in the reference population all the animals genotyped with the 50K SNP panel; 2) in the second step, all the animals imputed to the 50K SNP panel were then imputed to the 777K SNP panel using as reference two-thirds of the Hereford and Braford and all Nellore bulls genotyped with the 777K SNP panel. One-step imputation was performed by imputing from the simulated low density panels directly to the 777K SNP panel.

Imputation accuracy of above scenarios was assessed by concordance rate (CR), which corresponds to the proportion of genotypes correctly imputed, and by allelic $R^2$, which corresponds to the square of the correlation between the number of minor alleles in the imputed genotype and the number of minor alleles in the original genotype [25].

There were twenty imputation scenarios from low density panels to the 50K SNP panel. Thirty-four scenarios were examined for imputation from low and medium density panels to 777K SNP panel and thirty scenarios were used to assess differences in imputation accuracy in one or two steps (Table 3).

**Imputation methods**

Imputation was carried out by FImpute v.2.2 [11] and Beagle v.3.3 [8]. Beagle was used in scenarios that did not include pedigree information and ungenotyped animals. FImpute was used in all scenarios.

Imputation methods can be based on linkage disequilibrium information between markers in the population, but also can use the inheritance information within family. Beagle software is based on linkage disequilibrium between markers in the population and uses a Hidden Markov model [26] for inferring haplotype phase and filling in genotypes. Beagle also exploits family information indirectly by searching for long haplotypes. Contrary to Beagle, FImpute software uses a deterministic algorithm and makes use of both family and population information directly. Family information is taken into account only when pedigree information is available. The population imputation in FImpute is based on an overlapping sliding window method [11] in which information from close relatives (long haplotype match) is first utilized and information from more distant relatives is subsequently used by shortening the window size. The algorithm assumes that all animals are related to each other to some degree ranging from very close to very distant relationships.

**Comparison between scenarios**

Analysis of variance was carried out using the GLM procedure in SAS version 9.2 (SAS Inst. Inc., Cary, NC) to compare the average CR and allelic $R^2$ of each

scenario. An arcsine square root [27] transformation was applied to CR and allelic $R^2$ to normalize the residuals.

**Results**

Of the 3,698 animals genotyped with the 50K SNP panel, ~24% had sire and/or dam genotyped and ~65% had at least one parent unknown in the pedigree. With respect to the animals genotyped with the 777K SNP panel, ~15% had sire and/or dam genotyped and ~35% had at least one parent unknown. Table 1 shows pedigree structure for each breed.

Table 4 provides the computing run time for each imputation scenario. Using FImpute, the run-time ranged between 2 and 48 minutes for different scenarios, while Beagle took between 25 and 2,280 minutes for the same scenarios. Table 5 provides the means and standard deviations of CR and allelic $R^2$ for imputation to 50K and 777K SNP panels.

**Imputation of the low density panels to the 50K SNP panel**

There were significant differences ($P<0.05$) in CR and allelic $R^2$ between the two algorithms and between pairs of simulated low density panels, as well as a significant algorithm by panel interaction ($P<0.05$). However, there were no significant differences ($P>0.05$) in CR and allelic $R^2$ between scenarios (Table 6).

The non-commercial 15K SNP panel resulted in the highest imputation accuracy of the low density panels with an overall CR of 0.973 and allelic $R^2$ of 0.962, 0.109 and 0.175 points higher than the 3K SNP panel, respectively (Table 5). The use of Nellore genotypes or use of pedigrees in FImpute did not improve CR or allelic $R^2$ when imputing to the 50K SNP panel (Table 6). The average CR and allelic $R^2$ for the four scenarios were 0.940 and 0.905, respectively. Using FImpute resulted in an overall average CR of 0.943 and allelic $R^2$ of 0.912 while for Beagle the same average features were 0.927 and 0.890, respectively (Table 5). The algorithm by panel interaction, showed larger differences in CR and allelic $R^2$ between FImpute and Beagle for sparser panels (0.021 in CR and 0.031 in allelic $R^2$ for the 3K SNP panel) when compared to denser panels (0.012 in CR and 0.016 in allelic $R^2$ for the 15K SNP panel), with FImpute being consistently more accurate. Imputation accuracy for 8K and 20K SNP panels were not significantly different using Beagle ($P>0.05$) with respect to CR and allelic $R^2$ (Table 6). The highest CR ($>0.977$) and allelic $R^2$ ($>0.967$) were obtained using the 15K SNP panel and FImpute.

An important measurement of imputation success is the number of animals imputed with modest accuracy (assumed <0.950 CR here).Using the 15K SNP panel resulted in 93% and 83% of the animals being imputed with a CR above 0.950 (average of all scenarios) for FImpute and Beagle, respectively, while using the 3K SNP panel as the low density panel resulted in only 6.3% and 0.8% of animals above this accuracy threshold using FImpute and Beagle, respectively. The results for the other panels ranged between 62% and 70% using FImpute and between 40% and 48% using Beagle (Figure 1).

The CR (average of all scenarios) for the 3K SNP panel, from either FImpute or Beagle, were lower than all other panels with CR values over all BTAs at or below 0.900. All other panels produced CR above 0.930 for all chromosomes. Imputation accuracy was found to be relative to chromosome length with the highest CRs obtained for BTA1 while the lowest CRs were obtained for BTA28 in all scenarios and both algorithms, however little difference was seen across the genome (Figure 2).

The average CR for imputation from the alternative low density panels (3K, 6K, 8K, 15K and 20K) to the 50K SNP panel was calculated for three different classes of minor allele frequency (MAF) (<0.01, 0.01-0.05, and >0.05). For the MAF class <0.01 the average CR was close to 1.00 for all panel densities. For SNPs with MAF 0.01-0.05 and >0.05 the average CRs ranged similarly from 0.84 to 0.97, depending on the panel density (Figure 3).

**Imputation of the ungenotyped animals to the 50K SNP panel**

FImpute allows for accurate imputation of 50K genotypes for ungenotyped animals that have four or more offspring [11]. Thirty-seven animals that had four or more offspring were imputed and showed an average CR of 0.950 and with 99.86% of the SNPs imputed. When average CR were examined based on the number of offspring, accuracies of 0.924, 0.941, 0.972, 0.961 and 0.990 were found for bulls with 4-9, 10-19, 20-29, 30-39 and over 40 offspring, respectively. There were 11, 11, 9, 3 and 3 bulls in each of those progeny size classes, respectively. The lowest CR (0.900) corresponded to two Hereford animals with five offspring each, while the highest CR (above 0.980) was for six Braford animals with more than twenty offspring each.

**Imputation of the low density panels to the 777K SNP panel**

There were significant differences (P<0.05) in CR and allelic $R^2$ between algorithms, panels and scenarios when imputing to 777K SNP panel. The algorithm by panel interaction was also significant (P<0.05) (Table 7).

Using FImpute resulted in an overall average CR of 0.921 and allelic $R^2$ of 0.866, while Beagle yielded an average CR of 0.895 and allelic $R^2$ of 0.826 (Table 5). The 6K, 8K and 20K SNP panels did not significantly differ (P>0.05) in their average CR and allelic $R^2$ (Table 7). The highest CR and allelic $R^2$ were obtained with the 90tK SNP panel (CR=0.955; allelic $R^2$=0.925) and the lowest CR and allelic $R^2$ with the 3K SNP panel (CR=0.838; allelic $R^2$=0.728). For the other panels, CR was between 0.898 and 0.952 and allelic $R^2$ was between 0.829 and 0.919 (Table 4). The use of the pedigree information (NE-P) slightly decreased the CR and allelic $R^2$ for imputation to the 777K SNP panel (P<0.05) (Table 6). The interaction algorithm by panel, showed larger differences in CR and allelic $R^2$ between FImpute and Beagle for sparse panels (0.028 in CR and 0.044 in allelic $R^2$ for the 3K SNP panel) when compared to denser panels (0.016 in CR and 0.024 in allelic $R^2$ for the 90tK SNP panel), with FImpute resulting in consistently higher accuracy of imputation.

The distributions of animals in high classes of CR varied between FImpute and Beagle. For FImpute, the proportion of animals imputed above a CR of 0.95 ranged from 12.8% for the 3K SNP panel to 73.6% for the 90iK SNP panel. For the other panels, the proportion of animals was between 20% and 48% (Figure 4a). For Beagle, with the exception of the 90iK SNP panel (39.5%) and the 90tK SNP panel (53.5%), the proportion of animals imputed above a CR of 0.95 was around 3% (Figure 4b).

Imputation accuracy per chromosome using Beagle was only greater than 0.900 when 50K or more dense panels were used (Figure 5b), while the same was observed using FImpute for all panels denser than 6K (Figure 5a). Per chromosome accuracies followed the results from 50K, where the highest accuracy was observed on BTA1, and the lowest on BTA28.

Imputation to the 777K SNP panel performed in two steps was statistically superior (P<0.05) to imputation in a one-step both when measured by CR and allelic $R^2$, and this difference was observed for all scenarios (Table 8). The interaction between

number of steps and algorithm showed larger difference between CR and allelic $R^2$ from one and two steps imputation when Beagle was used (0.107 in CR and 0.181 in allelic $R^2$). The interaction between number of steps and low density panel showed that the difference between CR and allelic $R^2$ from one to two steps imputation was larger for sparse panels (0.178 in CR and 0.298 in allelic $R^2$ for the 3K SNP panel) when compared to denser panels (0.020 in CR and 0.034 in allelic $R^2$ for the 20K SNP panel.

The relative increase in CR for the two-step imputation with respect to the one-step imputation was 27%, 12%, 11%, 5% and 2% for 3K, 6K, 8K, 15K and 20K SNP panels, respectively, and the relative increase in allelic $R^2$ was 69%, 21% 22% 9% and 4% for 3K, 6K, 8K, 15K and 20K SNP panels, respectively.

The average CR for imputation from the alternative low density panels (3K, 6K, 8K, 15K, 20K, 50K, 90iK and 90tK) to the 777K SNP panel was calculated for three different classes of MAF (<0.01, 0.01-0.05, and >0.05). For the MAF class <0.01 the average CR was close to 0.99 for all panel densities, for MAF class 0.01-0.05 and >0.05 the average CRs ranged from 0.84 to 0.97 and from 0.65 to 0.96, respectively, depending on the panel density (Figure 6).

**Discussion**

**Imputation of the low density panels to the 50K SNP panel**

There was no significant difference when imputation was performed using Nellore genotypes in the reference population and when the imputation was based on either family and population imputation or population imputation only. This means including pedigree information did not improve the CR and allelic $R^2$ and is not required for accurate imputation. When Nellore genotypes were included in the reference population, it was expected that it would increase CR and allelic $R^2$ because imputation population was mostly formed by Braford animals that have in their breed composition from 15% to 75% of zebu breeds, including the Nellore breed. This implies that the haplotypes present in the Braford animals available in the reference population are able to account for almost all of the haplotypes in the population. Ventura et al. [5] also did not find differences in imputation accuracies when the reference population included Angus plus multiple breeds or Charolais plus multiple breeds to impute crossbreds in Canada. Berry et al. [28], studying seven dairy and beef breeds in Ireland, concluded that reference populations formed by multiple breeds did not significantly increase the accuracy of the imputation of purebreds.

Including pedigree information did not increase CR or allelic $R^2$. This could be expected due to the weak structure of the pedigree within the set of genotyped animals and in the whole pedigree file. Similar results were found by Carvalheiro et al. [21] when working with Nellore in Brazil with similar pedigree structure. However, Ma et al. [29] found increases in CR between 1% and 2% using Beagle and FImpute in Nordic Red cattle in Sweden when including genotypes of the bull-sires of the imputation population into the reference population. It would not, however, require pedigree information to detect these relationships in either algorithm.

The interaction between algorithm and panel was significant and yielded greater differences in CR and allelic $R^2$ between FImpute and Beagle for low density panels, showing a greater advantage to using FImpute when a sparser low density panel is used. Carvalheiro et al. [21], working with Nellore in Brazil, also reported that FImpute outperformed Beagle for different low density panels and that there was a trend of greater differences between algorithms as low density panel density decreased.

The CR and allelic $R^2$ values from FImpute in all analyses were consistently higher than those from Beagle, showing that the overlapping windows approach used by FImpute better infer missing genotypes than Hidden Markov models used by Beagle. Similar results were obtained by Carvalheiro et al. [21] in Nellore in Brazil and Larmer et al. [30], who worked on imputation from 6K and 50K SNP panels to 777K SNP panel in dairy cattle in Canada.

The 20K SNP panel was mainly developed for imputation to the 777K SNP panel and it has only 7,320 common SNPs with the 50K SNP panel. No difference between the 8K and 20K SNP panel was found using Beagle algorithm as they had similar number and average distance between the SNPs present on the 50K SNP panel. A few studies have tested the accuracy of imputation using different densities of markers and denser low density panels have consistently led to higher imputation accuracy in several beef cattle breeds, observed in Wang et al. [31] in Angus, Dassonneville et al. [17] in Blonde d'Aquitaine , Huang et al. [32] in Hereford and Chud [33] in Canchim cattle. The customized 15K SNP panel created in this study showed higher CR and allelic $R^2$ when compared to the other low density panels, including the 20K SNP panel. The reason for that may because of a higher density of markers in low linkage disequilibrium with adjacent SNPs and medium to high minor allele frequency in the population, allowing a better haplotype reconstruction. The superiority of the customized 15K SNP panel in relation to the commercial panels, however, might be expected because it was created based on criteria specific for this population. Carvalheiro et al. [21], working with Nellore cattle in Brazil, also developed a 15K SNP panel for imputation to the 777K SNP panel. They found slightly better results when compared to imputation from the 50K SNP panel. One possible disadvantage of customized panels is the cost will likely be higher in comparison to already available commercial panels of similar density.

The highest accuracies were obtained for all low density panels when examining BTA1, whereas the worst results were obtained for BTA28. Sun et al. [34], working with Angus genotypes in the United States, reported that genotype imputation was more difficult in the initial and end regions of the chromosomes. Therefore, the shorter are the chromosomes, which is the case of BTA28 (46 Mb), the lower the overall chromosome accuracy, as the poorly imputed distal regions comprise a greater proportion of the overall chromosome. Similar results were found by Berry & Kearney [35] in Irish Holstein cattle, when imputing from the 3K to the 50K SNP panel. Moreover, Pausch et al. [24], working with Fleckvieh in Germany and imputing from 50K to 777K SNP panel, and Wang et al. [31], working with Angus in the United States and imputing to the 50K SNP panel from various low density panels, found higher and lower accuracies for BTA1 and BTA28, respectively, when compared to the average accuracy of imputation for all chromosomes.

**Imputation of low density panels to the 777K SNP panel**

On average, the imputation population had seven animals with one of the parents genotyped and the reference population had twenty-four animals with one of the parents also genotyped. The inclusion of pedigree information did not result in an increase in CR and allelic $R^2$. Carvalheiro et al. [21] studying, among other factors, the effect of using or not the pedigree information in Nellore, also did not observe significant difference in CR when imputing from 15K and 50K to the 777K SNP panel using FImpute.

The two-step imputation procedure consistently out-performed imputation in one-step. This result confirmed that more SNPs contained on the low density panel, results in greater accuracy of imputation [31], [17], [7], [33]. Similar results were found by Larmer et al. [30] in Canadian Holstein cattle, when imputing in two steps from 6K to the 50K and from 50K to the 777K SNP panel. The interaction between algorithm and one or two steps was significant and showed greater difference in CR and allelic $R^2$ between one and two steps methods when using Beagle. The percentage of animals with CR above 0.95, in general, was higher for higher density panels, as expected. However, the 15K SNP panel showed higher percentage than the 20K SNP panel, most likely due to the criteria that were used for developing the 15K SNP panel. Moreover, it may be also due to the fact that the 20K SNP panel was developed mainly for genotype imputation in *Bos Taurus Indicus* cattle.

The results by chromosome followed the same pattern found for imputation to the 50K SNP panel, with longer chromosomes having greater imputation accuracies [34], [33].

**Imputation of the ungenotyped animals to the 50K SNP panel**

Genotype imputation for ungenotyped animals is now a lower cost alternative that can be used to increase the training population towards the implementation of genomic selection. Important ungenotyped ancestors that may have no available biological material to perform genotyping can also be accurately imputed using genotyped progeny information. Also, groups of cows that were ungenotyped due to the costs can have their genotypes inferred [36], [37], [38]. Different software, such as AlphaImpute [39], FindHap [12], PedImpute [23] and FImpute [11] are able to infer genotypes ungenotyped animals with high CR using different approaches, such as imputation based on: genotyped parents; sire and maternal grandsire, dam and paternal grand dam, sire only, dam only, and offspring. However, the accuracy of each approach is different [37], [36], [38].

Ungenotyped animals in this study were imputed using FImpute, using offspring. FImpute requires at least 4 offspring be available for imputation of ungenotyped individuals (default parameter). Preliminary results obtained by Sargolzaei et al. [11] and Berry et al. [38] using FImpute clearly showed an inability to impute the genotype of sires when a paternal halfsib family size of three or less was used. However, the results indicated that the greater the number of genotyped offspring, the higher were the CR values.

These results were similar to the ones reported by Berry et al. [38] studying seven dairy and beef breeds in Ireland with five offspring per ungenotyped individual and Bouwman et al. [36] studying dairy cattle in Netherlands with four offspring per ungenotyped animal. The average value found in this study was compatible to what is considered an accurate imputation from low density, i.e. average CR above 0.950 and having a very low missing rate.

**Conclusions**

All low density panels, except the 3K SNP panel, can be used to impute to the 50K SNP panel with average concordance rates higher than 0.940. The customized 15K SNP panel yielded the highest percentage of animals with concordance rate above 0.950 of all the low density panels studied.

The 50K, 90iK and 90tK SNP panels can be used to impute to the 777K SNP panel with average concordance rates higher than 0.940. A two-step imputation is

recommended for lower density panels, making use of all available intermediate density panel genotypes.

FImpute outperformed Beagle in all scenarios for imputation to both the 50K and to the 777K SNP panels both in terms of accuracy and computing time required.

Ungenotyped animals that have four or more offspring and do not have available biological material to carry out genotyping may have their 50K SNP panel genotype inferred with an average concordance rate of 0.950 in the Hereford/Braford population analyzed.

**Abbreviations**
BTA: Bos taurus autosomal chromosome; CR: Concordance rate; DNA: Deoxyribonucleic acid, GLM: General linear models; GGP: GeneSeek genomic profiler; K: Kbytes; MAF: Minor allele frequency; Mb: Mega base pairs; NC: North Carolina; SAS: Statistical analysis sistem; SD: Standard deviation; SNP: Single nucleotide polymorphism; USA: United States of America.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
MLP participated in the design of the study, carried out the analyses, was involved in the discussions, prepared and drafted the manuscript. JB participated in the design of the study, was involved in the discussions and helped to draft the manuscript. FFC was involved in the field experimental design and data collection, in the discussions and helped to draft the manuscript. MS developed the FImpute software, was involved in the discussions, and helped to draft the manuscript. SGL helped to draft the manuscript. FSS participated in the design of the study, was involved in the discussions and helped to draft the manuscript. All authors read and approved the final manuscript.

**References**
1. The Bovine Genome Sequencing and Analysis Consortium., Elsik CG, Tellam RL, Worley KC: **The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution**. *Science* 2009, **324**:522–528.

2. Zhang Z, Druet T: **Marker imputation with low-density marker panels in Dutch Holstein cattle**. *J Dairy Sci* 2010, **93**:5487–5494.

3. Druet T, Schrooten C, de Roos APW: **Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle.** *J Dairy Sci* 2010, **93**:5443–5454.

4. Hozé C, Fouilloux M-N, Venot E, Guillaume F, Dassonneville R, Fritz S, Ducrocq V, Phocas F, Boichard D, Croiseau P: **High-density marker imputation accuracy in sixteen French cattle breeds.** *Genet Sel Evol* 2013, **45**:33.

5. Ventura RV, Lu D, Schenkel FS, Wang Z, Li C, Miller SP: **Impact of reference population on accuracy of imputation from 6K to 50K SNP chips in purebred and crossbreed beef cattle**. *J Anim Sci* 2014, **92**:1433–1444.

6. Corbin LJ, Kranis A, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams J a: **The utility of low-density genotyping for imputation in the Thoroughbred horse.** *Genet Sel Evol* 2014, **46**:9.

7. Huang Y, Hickey JM, Cleveland MA, Maltecca C: **Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost.** *Genet Sel Evol* 2012, **44**:25.

8. Browning SR, Browning BL: **Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.** *Am J Hum Genet* 2007, **81**:1084–1097.

9. Druet T, Georges M: **A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping.** *Genetics* 2010, **184**:789–798.

10. Howie B, Marchini J, Stephens M: **Genotype imputation with thousands of genomes.** *G3 (Bethesda)* 2011, **1**:457–469.

11. Sargolzaei M, Chesnais JP, Schenkel FS: **A new approach for efficient genotype imputation using information from relatives.** *BMC Genomics* 2014, **15**:478.

12. VanRaden PM, O'Connell JR, Wiggans GR, Weigel KA: **Genomic evaluations with many more genotypes.** *Genet Sel Evol* 2011, **43**:10.

13. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: **Invited review: Genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**:433–443.

14. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score.** *J Dairy Sci* 2010, **93**:743–752.

15. Brito F V, Neto JB, Sargolzaei M, Cobuci JA, Schenkel FS: **Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle.** *BMC Genet* 2011, **12**:80.

16. Sargolzaei M, Schenkel FS, Chesnais J: **Impact of amount of dam genotypic information on family-based imputation accuracy**. In *Dairy Cattle Breed Genet Comm Meet 2010*. Guelph/ON Canada; 2010.

17. Dassonneville R, Fritz S, Ducrocq V, Boichard D: **Short communication: Imputation performances of 3 low-density marker panels in beef and dairy cattle.** *J Dairy Sci* 2012, **95**:4136–4140.

18. IBGE: *Produção Da Pecuária Municipal 2012* (In Portuguese). Brasília, Brazil; 2012:71.

19. **Índex ASBIA - importação, exportação e comercialização de sêmen 2011** (In portuguese) [http://www.asbia.org.br/novo/upload/mercado/relatorio2011.pdf]

20. Fries LA: **Cruzamentos em Gado de Corte** (In portuguese). In *4o Simpósio sobre Pecuária Corte, 8-10 Outubro 1996*. Piracicaba/SP, Brasil; 1996.

21. Carvalheiro R, Boison SA, Neves HHR, Sargolzaei M, Schenkel FS, Utsunomiya YT, O'Brien AMP, Sölkner J, McEwan JC, Tassell CP Van, Sonstegard TS, Garcia JF: **Accuracy of genotype imputation in Nelore cattle**. *Genet Sel Evol* 2014, **46**:69.

22. Dassonneville R, Brøndum RF, Druet T, Fritz S, Guillaume F, Guldbrandtsen B, Lund MS, Ducrocq V, Su G: **Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations.** *J Dairy Sci* 2011, **94**:3679–3686.

23. Nicolazzi EL, Biffani S, Jansen G: **Short communication: imputing genotypes using PedImpute fast algorithm combining pedigree and population information.** *J Dairy Sci* 2013, **96**:2649–2653.

24. Pausch H, Aigner B, Emmerling R, Edel C, Götz K-U, Fries R: **Imputation of high-density genotypes in the Fleckvieh cattle population.** *Genet Sel Evol* 2013, **45**:3.

25. Browning BL, Browning SR: **A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals.** *Am J Hum Genet* 2009, **84**:210–223.

26. Rabiner LR, Juang BH: **An introduction to hidden Markov models.** *IEEE/ASSP* 1986.

27. Cochran WG, Snedecor GW: *Statistical Methods*. 8th edition. Ames: Iowa State University Press; 1989.

28. Berry DP, McClure MC, Mullen MP: **Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes**. *J Anim Breed Genet* 2013:1–8.

29. Ma P, Brøndum RF, Zhang Q, Lund MS, Su G: **Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle.** *J Dairy Sci* 2013, **96**:4666–4677.

30. Larmer SG, Sargolzaei M, Schenkel FS: **Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds.** *J Dairy Sci* 2014, **97**:1–14.

31. Wang H, Woodward B, Bauck S, Rekaya R: **Imputation of missing SNP genotypes using low density panels**. *Livest Sci* 2012, **146**:80–83.

32. Huang Y, Maltecca C: **Effects of reduced panel, reference origin, and genetic relationship on imputation of genotypes in Hereford cattle**. *J Anim Sci* 2012, **90**:4203–4208.

33. Chud TCS: *Metodologias E Estratégias de Imputação de Marcadores Genéticos Em Bovinos Da Raça Canchim* (in Portuguese). Joboticabal/SP - Brazil; 2014.

34. Sun C, Wu X-L, Weigel KA, Rosa GJM, Bauck S, Woodward BW, Schnabel RD, Taylor JF, Gianola D: **An ensemble-based approach to imputation of moderate-density genotypes for genomic selection with application to Angus cattle.** *Genet Res (Camb)* 2012, **94**:133–150.

35. Berry DP, Kearney JF: **Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection.** *Animal* 2011, **5**:1162–1169.

36. Bouwman AC, Hickey JM, Calus MPL, Veerkamp RF: **Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle.** *Genet Sel Evol* 2014, **46**:6.

37. Pimentel ECG, Wensch-Dorendorf M, König S, Swalve HH: **Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture.** *Genet Sel Evol* 2013, **45**:12.

38. Berry DP, McParland S, Kearney JF, Sargolzaei M, Mullen MP: **Imputation of ungenotyped parental genotypes in dairy and beef cattle from progeny genotypes.** *Animal* 2014, **8**:895–903.

39. Hickey JM, Kinghorn BP, Tier B, van der Werf JHJ, Cleveland MA: **A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation.** *Genet Sel Evol* 2012, **44**:9.

**Table 1 Summary statistics of genotyped animals and pedigree structure of the 50K and the 777K SNP panels.**

| Parameter | Braford | Hereford | Nellore |
|---|---|---|---|
| *Imputation to the 50K SNP panel* | | | |
| Total of genotyped animals | 2,946 | 664 | 88 |
| Sires | 39 | 29 | 6 |
| Dams | 76 | 21 | 0 |
| Offspring | 2,831 | 614 | 82 |
| Offspring with sire and/or dam genotyped (%) | 22.81 | 32.68 | 12.50 |
| Average number of offspring per sire | 15.28±17.38 | 6.76±6.46 | 1.83±0.90 |
| Smallest and largest number of offspring per sire | 1-76 | 1-26 | 1-3 |
| Average number of offspring per dam | 1.00±0.00 | 1.00±0.00 | 1.00±0.00 |
| Offspring with sire and/or dam unknown (%) | 69.86 | 48.04 | 18.18 |
| | | | |
| *Imputation to the 777K SNP panel* | | | |
| Total of genotyped animals | 71 | 59 | 88 |
| Sires | 8 | 3 | 5 |
| Dams | 0 | 0 | 0 |
| Offspring | 63 | 56 | 83 |
| Offspring with sire and/or dam genotyped (%) | 25.35 | 8.47 | 10.23 |
| Average number of offspring per sire | 2.25±1.09 | 1.67±0.94 | 1.80±0.98 |
| Smallest and largest number of offspring per sire | 1-4 | 1-3 | 1-3 |
| Average number of offspring per dam | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 |
| Offspring with sire and/or dam unknown (%) | 53.52 | 38.98 | 18.18 |

**Table 2 Number of SNPs on each simulated panel before and after quality control for imputation to 50K or 777K SNP panels[1].**

| Commercial name | Label | Number of SNPs | Number of SNPs in the imputation to 50K | Number of SNPs in the imputation to 777K |
|---|---|---|---|---|
| Illumina Bovine3K | 3K | 2,900 | 2,321 | 2,359 |
| Illumina BovineLD | 6K | 6,909 | 6,205 | 6,216 |
| Beef LD GeneSeek Genomic Profiler | 8K | 8,762 | 7,033 | 7,478 |
| 15K panel[2] | 15K | 14,195 | 12,304 | 12,345 |
| Indicus LD GeneSeek Genomic Profiler | 20K | 19,721 | 7,320 | 16,047 |
| Illumina BovineSNP50 | 50K | 54,609 | 43,247 | 43,247 |
| GeneSeek Genomic Profiler Indicus HD | 90iK | 74,085 | - | 55,819 |
| GeneSeek Genomic Profiler Beef HD | 90tK | 76,992 | - | 61,445 |
| Illumina BovineHD | 777K | 787,799 | - | 587,620 |

[1] The SNP quality control included GenCall score (>=0.15), Call Rate (>=0.90), Hardy-Weinberg Equilibrium (P>=$10^{-6}$), removal of non-autosomal chromosomes and SNPs not in common with reference panel;

[2] Non commercial panel. The 15K panel was created based on the Beef LD GeneSeek Genomic Profiler (8K) panel by expanding it with SNPs selected based on minor allele frequency greater than 0.23, linkage disequilibrium less than 0.088 and preferably located evenly spaced between two SNPs in the 8K SNP panel.

**Table 3 Imputation scenarios used in the study.**

| Imputation | | Software | Pedigree information | Nellore genotypes | Method |
|---|---|---|---|---|---|
| From | To | | | | |
| 3K, 6K, 8K, 15K, 20K | 50K | Fimpute | Yes | Yes | One-step |
| | | | | No | |
| | | | No | Yes | |
| | | | | No | |
| | | Beagle | No | Yes | |
| | | | | No | |
| 3K, 6K, 8K, 15K, 20K | 777K | Fimpute | Yes | Yes | One-step |
| | | | No | | |
| | | Beagle | No | | Two-step |
| 50K, 90iK, 90tK | 777K | Fimpute | Yes | Yes | One-step |
| | | | No | | |
| | | Beagle | No | | |

**Table 4 Overall computing run time in minutes for the different imputation scenarios[1,2].**

| Panel | FImpute | | | | Beagle | |
|---|---|---|---|---|---|---|
| | NE-P | NNE-P | NE-NP | NNE-NP | NE-NP | NNE-NP |
| *Imputation to the 50K SNP panel[3]* | | | | | | |
| 3K | 2 | 6 | 41 | 39 | 2280 | 2131 |
| 6K | 3 | 7 | 46 | 45 | 828 | 772 |
| 8K | 3 | 7 | 45 | 45 | 808 | 656 |
| 15K | 3 | 9 | 48 | 48 | 328 | 317 |
| 20K | 3 | 7 | 37 | 42 | 708 | 622 |
| *Imputation to the 777K SNP panel[4,5]* | | | | | | |
| 3K | 16 (17,24) | - | 4 (5,8) | - | 64 (224,41) | - |
| 6K | 17 (23,24) | - | 4 (19,21) | - | 49 (238,33) | - |
| 8K | 17 (23,24) | - | 3 (20,23) | - | 45 (177,34) | - |
| 15K | 15 (24,23) | - | 8 (20,23) | - | 40 (127,42) | - |
| 20K | 17 (23,23) | - | 9 (20,23) | - | 44 (161,42) | - |
| 50K | 3 | - | 11 | - | 29 | - |
| 90iK | 17 | - | 11 | - | 25 | - |
| 90tK | 17 | - | 10 | - | 33 | - |

[1] Run time based on 10 parallel jobs with computer with 4*6-core processors (Intel Xeon X5690 @ 3.47GHz) and 128 Gigabytes of memory in OS x86-64 GNU/Linux;

[2] Scenarios for imputation. (NE-P) - using Nellore genotypes in the reference population and considering pedigree information; (NNE-P) - not using Nellore genotypes in the reference population and considering pedigree information; (NE-NP) - using Nellore genotypes in the reference population and not using pedigree information; (NNE-NP) - not using Nellore genotypes in the reference population and not using pedigree information;

[3] 2,735 or 2,647 (not using Nellore genotypes) animals in the reference population and 963 animals in the imputation population;

[4] Values outside the brackets refer to the one-step imputation. The reference and imputation population were formed by 175 and 43 animals, respectively;

[5] Values inside the brackets refer to the two-step imputation. The reference population were formed by 3,567 in the imputation from low density panel to the 50K SNP panel and 175 animals in the imputation from the 50K SNP panel to the 777K SNP panel. The imputation population was formed by 43 animals.

**Table 5 Mean and standard deviation (SD) of concordance rate and allelic $R^2$ calculated for different algorithms, panel densities and scenarios for both imputation to 50K and 777K SNP panels.**

| | No. | CR | | Allelic $R^2$ | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| | | | | | |
| *Imputation to the 50K SNP panel* | | | | | |
| *Algorithm* | | | | | |
| Beagle | 10 | 0.927 | 0.042 | 0.890 | 0.067 |
| Fimpute | 20 | 0.943 | 0.038 | 0.912 | 0.061 |
| *Panel* | | | | | |
| 3K | 6 | 0.864 | 0.011 | 0.787 | 0.016 |
| 6K | 6 | 0.946 | 0.008 | 0.919 | 0.011 |
| 8K | 6 | 0.952 | 0.008 | 0.927 | 0.011 |
| 15K | 6 | 0.973 | 0.006 | 0.962 | 0.008 |
| 20K | 6 | 0.953 | 0.008 | 0.929 | 0.011 |
| *Scenario* | | | | | |
| NE-P | 5 | 0.943 | 0.041 | 0.913 | 0.065 |
| NE-NP | 10 | 0.935 | 0.041 | 0.901 | 0.066 |
| NNE-P | 5 | 0.943 | 0.042 | 0.912 | 0.067 |
| NNE-NP | 10 | 0.935 | 0.042 | 0.901 | 0.066 |
| | | | | | |
| *Imputation to the 777K SNP panel [1]* | | | | | |
| *Algorithm* | | | | | |
| Beagle | 8 | 0.895 | 0.040 | 0.826 | 0.066 |
| Fimpute | 16 | 0.921 | 0.035 | 0.866 | 0.059 |
| *Panel* | | | | | |
| 3K | 3 | 0.838 | 0.017 | 0.728 | 0.025 |
| 6K | 3 | 0.898 | 0.016 | 0.829 | 0.025 |
| 8K | 3 | 0.902 | 0.017 | 0.836 | 0.026 |
| 15K | 3 | 0.918 | 0.017 | 0.863 | 0.027 |
| 20K | 3 | 0.903 | 0.017 | 0.837 | 0.026 |
| 50K | 3 | 0.930 | 0.016 | 0.882 | 0.025 |
| 90iK | 3 | 0.952 | 0.010 | 0.919 | 0.016 |
| 90tK | 3 | 0.955 | 0.009 | 0.925 | 0.014 |
| *Scenario* | | | | | |
| NE-P | 8 | 0.9199 | 0.037 | 0.865 | 0.062 |
| NE-NP | 16 | 0.9082 | 0.039 | 0.846 | 0.065 |
| *Step* | | | | | |
| One-step | 15 | 0.8064 | 0.884 | 0.674 | 0.147 |
| Two-step | 15 | 0.8920 | 0.032 | 0.819 | 0.053 |

[1] Means and standard deviation for the two-step analysis.

**Table 6 Analysis of variance performed on the average concordance rate and allelic $R^2$ of the animals in the imputation population from each scenario for imputation from low density panels to the 50K SNP panel[1,2].**

| Concordance rate | | | Allelic $R^2$ | | |
|---|---|---|---|---|---|
| Source | Mean | Scheffé test[3] | Source | Mean | Scheffé test[3] |
| *Algorithm[4] (P-value < 0.0001)* | | | *Algorithm[4] (P-value < 0.0001)* | | |
| FImpute | 1.340 | a | FImpute | 1.283 | a |
| Beagle | 1.306 | b | Beagle | 1.244 | b |
| *Panel[5] (P-value < 0.0001)* | | | *Panel[5] (P-value < 0.0001)* | | |
| 15K | 1.402 | a | 15K | 1.368 | a |
| 20K | 1.347 | b | 20K | 1.295 | b |
| 8K | 1.345 | c | 8K | 1.292 | c |
| 6K | 1.332 | d | 6K | 1.276 | d |
| 3K | 1.189 | e | 3K | 1.085 | e |
| *Scenario[6] (P-value 0.0147)* | | | *Scenario[6] (P-value 0.0277)* | | |
| NE-P | 1.323 | a | NE-P | 1.264 | a |
| NNE-P | 1.323 | a | NE-NP | 1.263 | a |
| NE-NP | 1.323 | a | NNE-P | 1.264 | a |
| NNE-NP | 1.322 | a | NNE-NP | 1.262 | a |
| *Algorithm*Panel (P-value < 0.0001)* | | | *Algorithm*Panel (P-value 0.0265)* | | |
| FImpute - 15K | 1.420 | a | FImpute -15K | 1.388 | a |
| Beagle - 15K | 1.384 | b | Beagle - 15K | 1.347 | b |
| FImpute - 20K | 1.365 | c | FImpute - 20K | 1.316 | c |
| FImpute - 8K | 1.362 | d | FImpute - 8K | 1.312 | d |
| FImpute - 6K | 1.349 | e | FImpute - 6K | 1.295 | e |
| Beagle - 20K | 1.330 | f | Beagle - 20K | 1.275 | f |
| Beagle - 8K | 1.328 | f | Beagle - 8K | 1.272 | f |
| Beagle - 6K | 1.316 | g | Beagle - 6K | 1.257 | g |
| FImpute - 3K | 1.204 | h | FImpute - 3K | 1.104 | h |
| Beagle - 3K | 1.174 | i | Beagle - 3K | 1.067 | i |

[1] Concordance rate and allelic $R^2$ were arcsine square root transformed for the analyses;

[2] Interactions between Algorithm*Scenario and Panel*Scenario were not statistically significant (P>0.05);

[3] Different letters within a group means that there is a statistical difference between two means (P<0.05);

[3] Algorithm used was either FImpute v.2.2 [11] or Beagle v.3.3 [8];

[5] 3K, 6K, 8K, 15K and 20K are low-density panels;

[6] Scenarios for imputation to the 50K SNP panel. (NE-P) - using Nellore genotypes in the reference population and considering pedigree information; (NNE-P) - not using Nellore genotypes in the reference population and considering pedigree information; (NE-NP) - using Nellore genotypes in the reference population and not using pedigree information; (NNE-NP) - not using Nellore genotypes in the reference population and not using pedigree information.

**Table 7 Analysis of variance performed on the average concordance rate and allelic $R^2$ of the animals in the imputation population from each scenario for imputation from low density panels to the 777K SNP panel[1,2,3].**

| Concordance rate | | | Allelic $R^2$ | | |
|---|---|---|---|---|---|
| Source | Mean | Scheffé test[4] | Source | Mean | Scheffé test[4] |
| *Algorithm[5] (P-value < 0.0001)* | | | *Algorithm[5] (P-value < 0.0001)* | | |
| FImpute | 1.291 | a | FImpute | 1.203 | a |
| Beagle | 1.244 | b | Beagle | 1.145 | b |
| *Panel[6] (P-value < 0.0001)* | | | *Panel[6] (P-value < 0.0001)* | | |
| 90tK | 1.351 | a | 90tK | 1.286 | a |
| 90iK | 1.343 | b | 90iK | 1.275 | b |
| 50K | 1.295 | c | 50K | 1.210 | c |
| 15K | 1.273 | d | 15K | 1.181 | d |
| 20K | 1.247 | e | 20K | 1.146 | e |
| 8K | 1.245 | e | 8K | 1.144 | e |
| 6K | 1.239 | e | 6K | 1.135 | e |
| 3K | 1.150 | f | 3K | 1.013 | f |
| *Scenario[7] (P-value  0.0258)* | | | *Scenario (P-value  0.0346)* | | |
| NE-NP | 1.269 | a | NE-NP | 1.175 | a |
| NE-P | 1.267 | b | NE-P | 1.172 | b |
| *Algorithm*panel (P-value =0.0052)* | | | *Algorithm*panel (P-value =0.0107)* | | |
| FImpute - 90tK | 1.370 | a | FImpute - 90tK | 1.309 | a |
| FImpute - 90iK | 1.364 | a | FImpute - 90iK | 1.301 | a |
| Beagle - 90tK | 1.331 | b | Beagle - 90tK | 1.262 | b |
| FImpute - 50K | 1.322 | b | Beagle - 90iK | 1.249 | b |
| Beagle - 90iK | 1.322 | b | FImpute - 50K | 1.244 | b |
| FImpute - 15K | 1.300 | c | FImpute - 15K | 1.215 | c |
| FImpute - 20K | 1.271 | d | Beagle - 50K | 1.176 | d |
| FImpute - 8K | 1.269 | d | FImpute - 20K | 1.176 | d |
| Beagle - 50K | 1.269 | d | FImpute - 8K | 1.174 | d |
| FImpute - 6K | 1.262 | d | FImpute - 6K | 1.165 | d |
| Beagle - 15K | 1.245 | e | Beagle - 15K | 1.146 | e |
| Beagle - 20K | 1.222 | f | Beagle - 20K | 1.115 | f |
| Beagle - 8K | 1.221 | f | Beagle - 8K | 1.114 | f |
| Beagle - 6K | 1.215 | f | Beagle - 6K | 1.106 | f |
| FImpute - 3K | 1.169 | g | FImpute - 3K | 1.039 | g |
| Beagle - 3K | 1.130 | h | Beagle - 3K | 0.988 | h |

[1] Concordance rate and allelic $R^2$ were arcsine square root transformed for the analyses;

[2] Interaction effects between Algorithm*Scenario and Panel*Scenario were not statistically significant ($P>0.05$);

[3] 3K, 6K, 8K, 15K and 20K are low-density panels were imputed in two steps (firstly they were imputed to the 50K and then to the 777K SNP panel);

[4] Different letters within a group means that there is a statistical difference between two means ($P<0.05$);

[5] Algorithm used was either FImpute v.2.2 [11] or Beagle v.3.3 [8];

[6] 3K, 6K, 8K, 15K, 20K, 50K, 90iK and 90tK are low-density panels;

[7] Scenarios for imputation to the 777K SNP panel. (NE-P) - using Nellore genotypes in the reference population and considering pedigree information; (NE-NP) - using Nellore genotypes in the reference population and not using pedigree information.

**Table 8 Analysis of variance performed on the average concordance rate and allelic $R^2$ of the animals in the imputation population from each scenario for imputation to the 777K SNP panel by one or two steps[12].**

| Concordance rate | | | Allelic $R^2$ | | |
|---|---|---|---|---|---|
| Source | Mean | Scheffé test[3] | Source | Mean | Scheffé test[3] |
| *Step[4] (P-value < 0.0001)* | | | *Step[4] (P-value < 0.0001)* | | |
| Two-step | 1.231 | a | Two-step | 1.125 | a |
| One-step | 1.110 | b | One-step | 0.997 | b |
| *Algorithm[5] (P-valeu < 0.0001)* | | | *Algorithm[4] (P-valeu 0.0001)* | | |
| FImpute | 1.202 | a | FImpute | 1.080 | a |
| Beagle | 1.140 | b | Beagle | 0.997 | b |
| *Panel[6] (P-value < 0.0001)* | | | *Panel[6] (P-value < 0.0001)* | | |
| 15K | 1.236 | a | 15K | 1.130 | a |
| 20K | 1.229 | b | 20K | 1.120 | a |
| 8K | 1.180 | c | 8K | 1.052 | b |
| 6K | 1.167 | d | 6K | 1.034 | c |
| 3K | 1.042 | e | 3K | 0.855 | d |
| *Scenario[7] (P-value 0.7638)* | | | *Scenario[7] (P-value 0.9983)* | | |
| NE-NP | 1.171 | a | NE-NP | 1.038 | a |
| NE-P | 1.170 | a | NE-P | 1.038 | a |
| *Step*Algorithm (P-value < 0.0001)* | | | *Step*Algorithm (P-value < 0.0001)* | | |
| Two-step - FImpute | 1.254 | a | Two-step - FImpute | 1.154 | a |
| Two-step - Beagle | 1.208 | b | Two-step - Beagle | 1.095 | b |
| One-step - FImpute | 1.149 | c | One-step - FImpute | 1.006 | c |
| One-step - Beagle | 1.072 | d | One-step - Beagle | 0.898 | d |
| *Step*Panel (P-value < 0.0001)* | | | *Step*Panel (P-value < 0.0001)* | | |
| Two-step - 15K | 1.274 | a | Two-step - 15K | 1.183 | a |
| Two-step - 20K | 1.247 | b | Two-step - 20K | 1.147 | b |
| Two-step - 8K | 1.246 | b | Two-step - 8K | 1.145 | b |
| Two-step - 6K | 1.239 | b | Two-step - 6K | 1.136 | b |
| One-step - 20K | 1.210 | c | One-step - 20K | 1.094 | c |
| One-step - 15K | 1.198 | c | One-step - 15K | 1.078 | c |
| Two-step - 3K | 1.149 | d | Two-step - 3K | 1.013 | d |
| One-step - 8K | 1.114 | e | One-step - 8K | 0.960 | e |
| One-step - 6K | 1.094 | f | One-step - 6K | 0.932 | e |
| One-step - 3K | 0.936 | g | One-step - 3K | 0.696 | f |

[1] Concordance rate and allelic $R^2$ were arcsine square root transformed for the analyses;

[2] Interaction effects between step*scenario, algorithm*panel, algorithm*scenario and panel*scenario were not statistically significant (P>0.05);

[3] Different letters within a group means that there is a statistical difference between two means (P<0.05);

[4] One-step is the imputation from the low-density panels to the 777K SNP panel and two-step is the imputation from low-density panels to 50K SNP panel and after the imputation from 50K SNP panel to 777K SNP panel;

[5] Algorithm used was either FImpute v.2.2 [11] or Beagle v.3.3 [8];

[6] 3K, 6K, 8K, 15K, and 20K are low-density panels;

[7] Scenarios for imputation to the 777K SNP panel. (NE-P) - using Nellore genotypes in the reference population and considering pedigree information; (NE-NP) - using Nellore genotypes in the reference population and not using pedigree information.

a)



b)
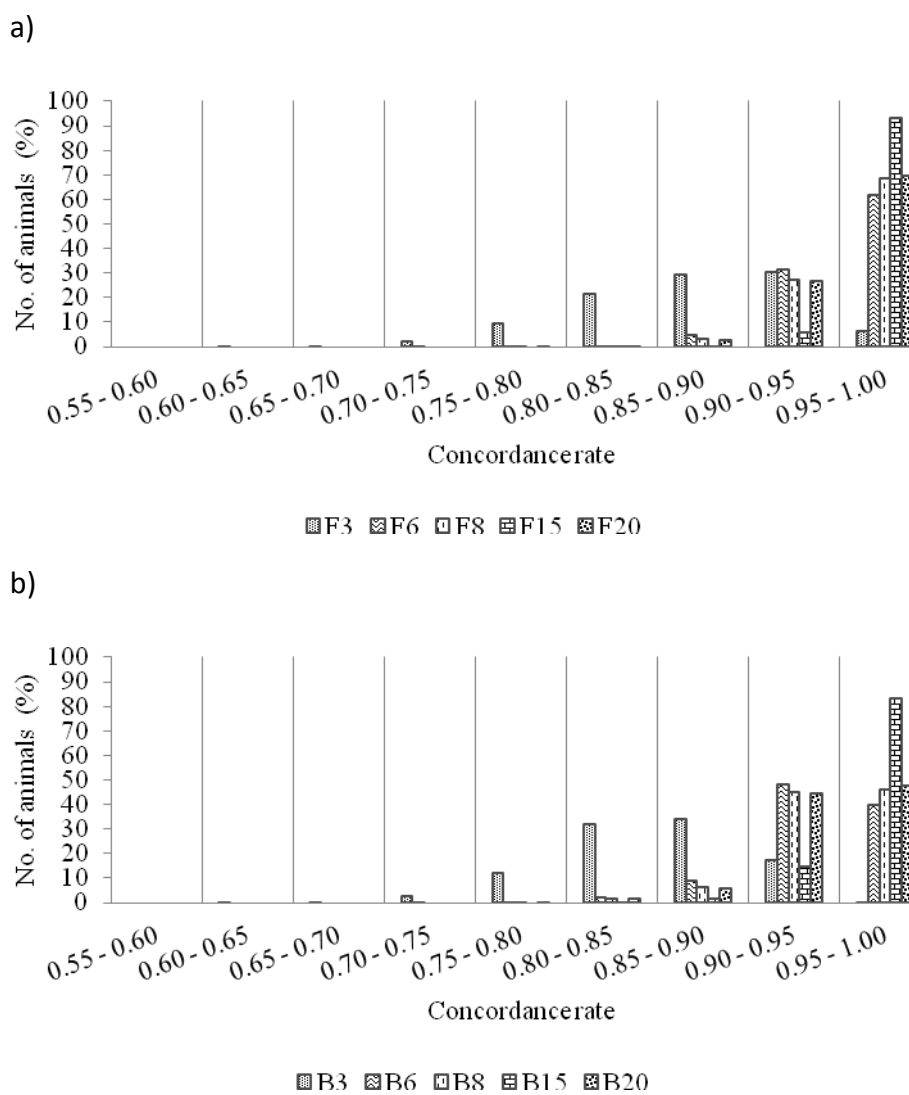


**Figure 1 Concordance rate of imputation to the 50K panel in different concordance rate bins.**
Average over scenarios of imputation from alternative low density panels (3K, 6K, 8K, 15K and 20K) to the 50K SNP panel. a) using FImpute; b) using Beagle.

a)



b)
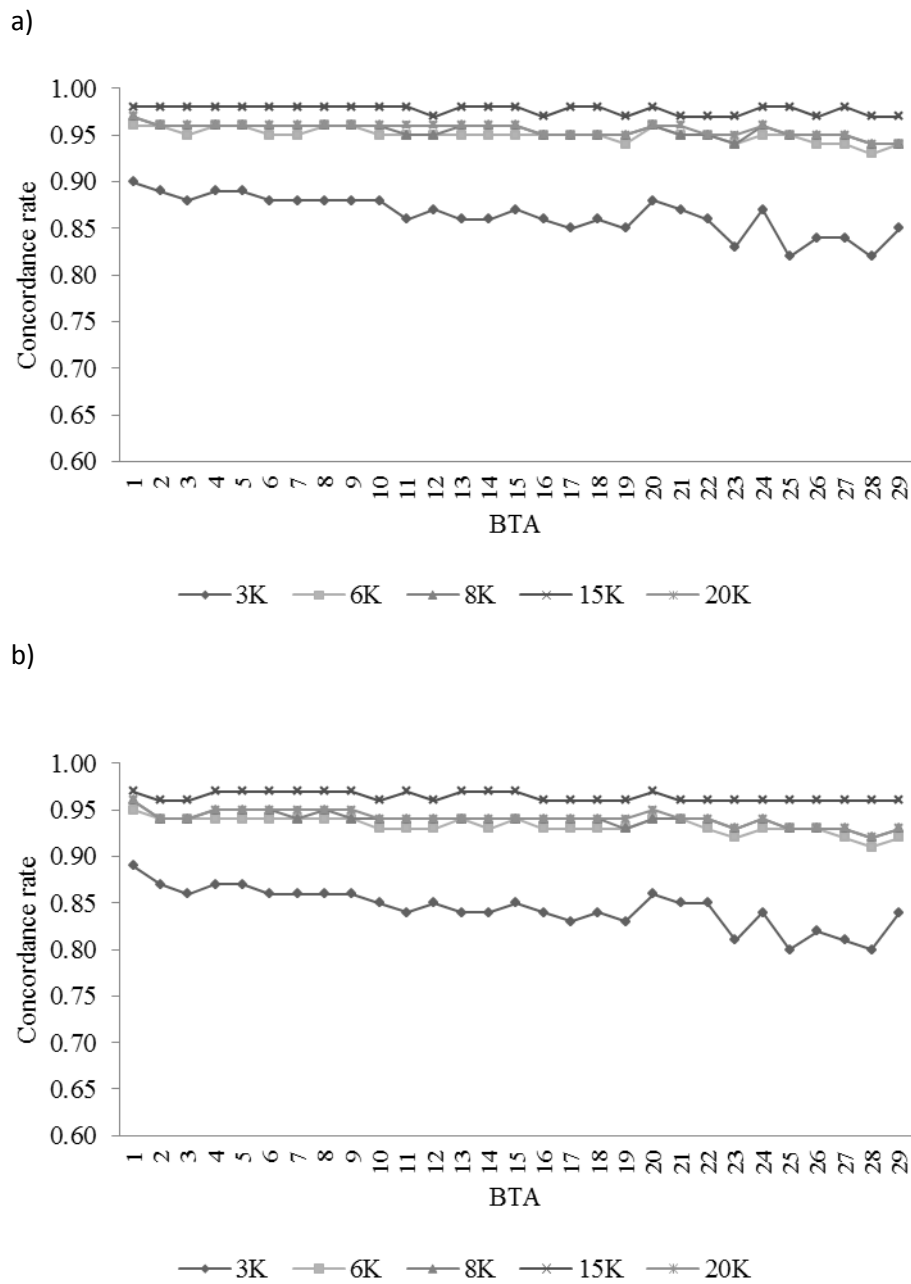


**Figure 2 Concordance rate of imputation to the 50K panel for all BTAs and scenarios.**
a) using FImpute; b) using Beagle.
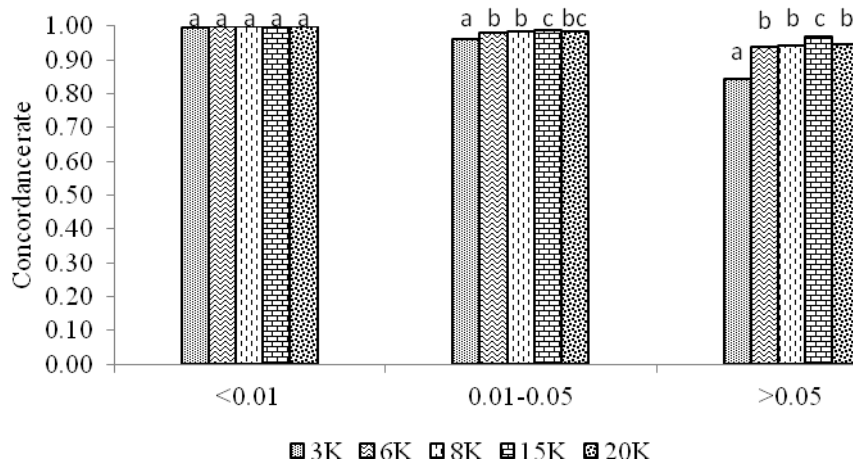
**Figure 3 Concordance rate of imputation by MAF classes.**
Average over scenarios of imputation from alternative low density panels (3K, 6K, 8K, 15K and 20K) to the 50K SNP panel. Within a group of colums, two different letters means a statistical difference (P<0.05).

a)
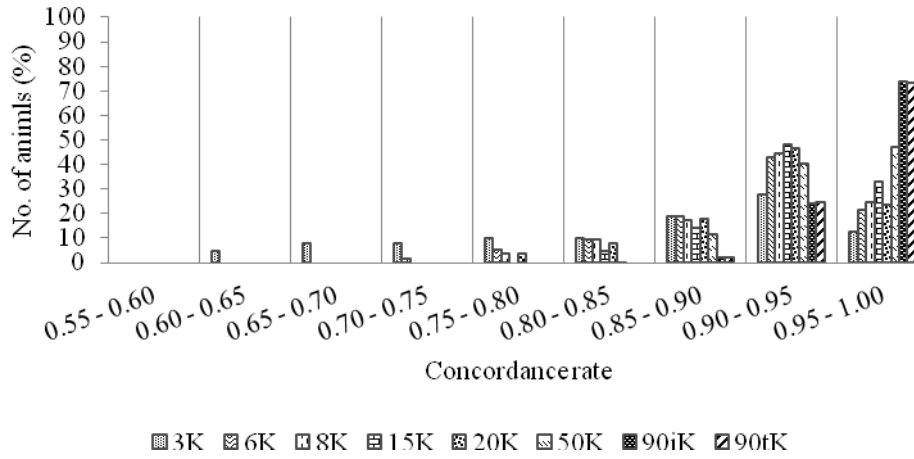


b)
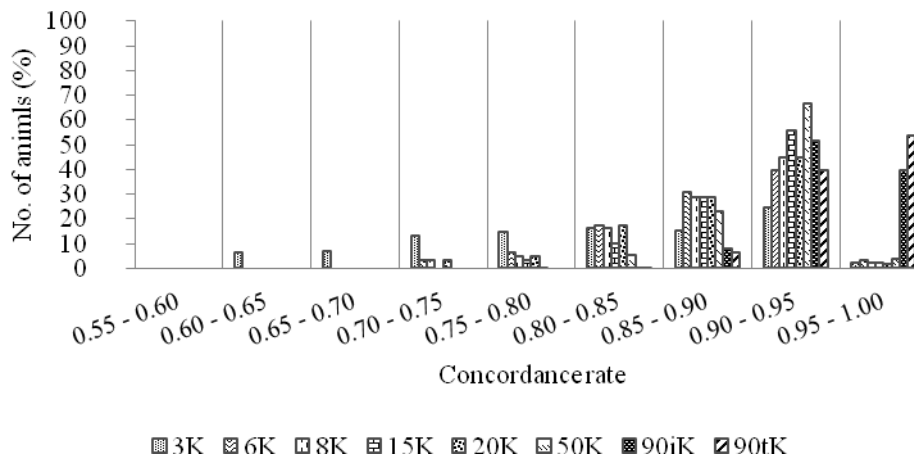


**Figure 4 Concordance rate of imputation to the 777K panel in different concordance rate bins.**
Average over scenarios of imputation from alternative low density panels (3K, 6K, 8K, 15K, 20K, 50K, 90iK and 90tK) to the 777K SNP panel. a) using FImpute; b) using Beagle.

a)



b)
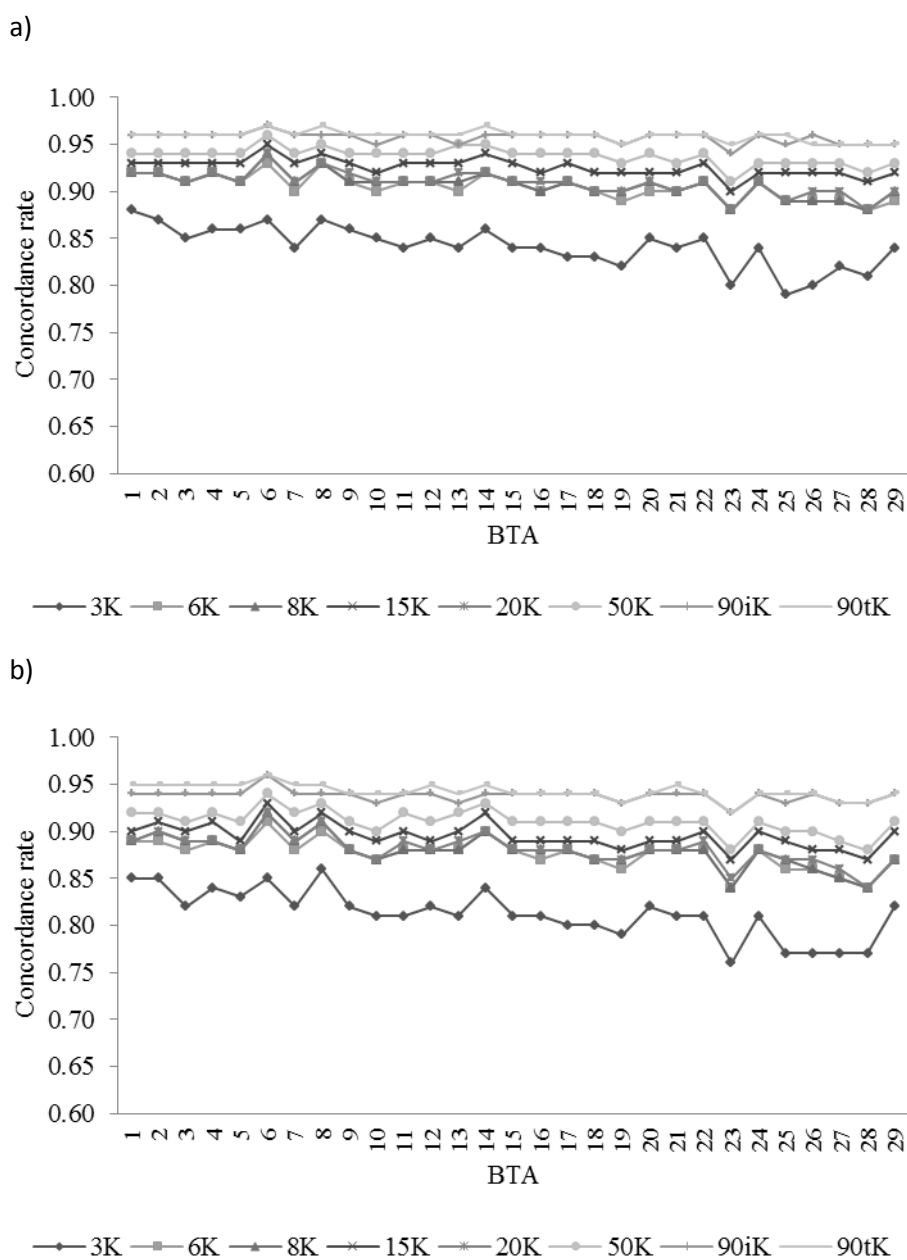


**Figure 5 Concordance rate of imputation to the 777K panel for all BTAs and scenarios.**
a) using FImpute; b) using Beagle.

a)



b)
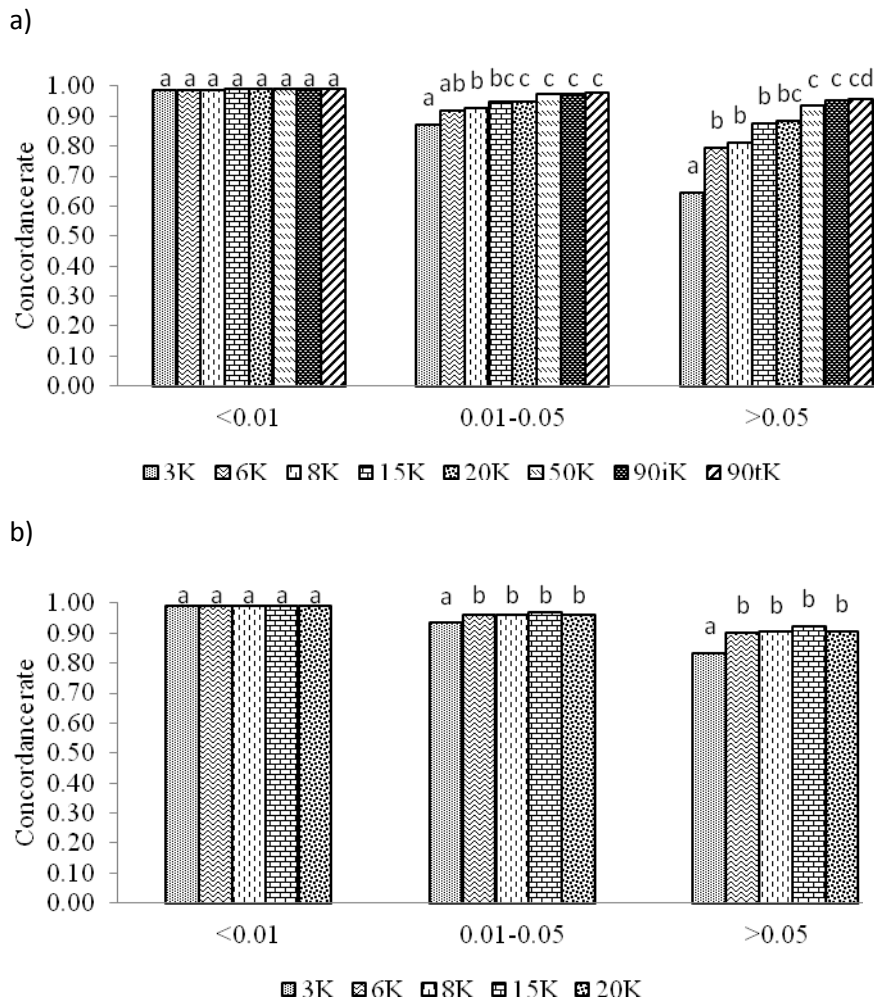


**Figure 6 Concordance rate of imputation by MAF classes.**
a) Average over scenarios of imputation from alternative low density panels (3K, 6K, 8K, 15K and 20K, 50K, 90iK and 90tK) to the 777K SNP panel; b) Average over scenarios of imputation from alternative low density panels (3K, 6K, 8K, 15K, 20K) to the 777K SNP panel in two-step imputation. Within a group of colums, two different letters means a statistical difference (P<0.05).

[7] **CAPÍTULO IV**

---

[7] Artigo redigido de acordo com as normas do periódico *BMC Genetics*

# Accuracy of genomic prediction in Braford and Hereford beef cattle from true and imputed genotypes

**Mario L Piccoli[123]\*, José Braccini[1], Fernando F Cardoso[45], Medhi Sargolzaei[36], Flávio S Schenkel[3].**

[1] Universidade Federal do Rio Grande do Sul, Departamento de Zootecnia, Porto Alegre, Brazil.

[2] GenSys Consultores Associados S/S, Porto Alegre, Brazil.

[3] Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada.

[4] Embrapa Southern Region Animal Husbandry, Bagé, Brazil.

[5] National Council for Scientific and Technological Development, Brasília, Brazil

[6] The Semex Alliance, Guelph, Canada.

\* Corresponding author: Mario L Piccoli (mariopiccoli@gensys.com.br)

Email addresses:

     MLP: mariopiccoli@gensys.com.br

     JB: jose.braccini@ufrgs.br

     FFC: fernando.cardoso@embrapa.br

     MS: msargol@uoguelph.ca

     FSS: schenkel@uoguelph.ca

**Abstract**

**Background:** Alternate scenarios composed by different percentages of animals with imputed genotypes and different sizes of the training population were used to investigate the effect of the use of imputed genotypes in the accuracy of genomic selection in twenty economic important traits in Brazilian Braford and Hereford beef cattle breeds. The training population was formed by animals born before 2011 and the validation population by animals born in 2011. Deregressed EBVs were used as pseudo phenotypes in a GBLUP model using two different mimicked panels derived from the 50K panel, the 8K panel and the 15K panel, which were subsequently imputed to the 50K panel. In addition, a 777K was also used in the analyses, which was imputed from the 50K panel.
**Results:** DGV validated accuracy (Pearson's correlation) in the prediction population for the twenty traits ranged from 0.38 to 0.40 in the different scenarios. The average losses in GEBV expected accuracy (accuracy obtained from the inverse of the mixed model equations) relative to the real 50K genotypes ranged from -0.0007 to -0.0012 and from -0.0002 to -0.0005 when using the 50K panel imputed from the 8K and 15K panel, respectively. When using the imputed 777K panel the average losses in GEBV expected accuracy was -0.0021. The average gain in EBV expected accuracy by including molecular markers when compared to simple BLUP was between 0.02 and 0.03 across scenarios and traits.
**Conclusion:** The percentage of animals with genotypes imputed in the training population did not influence the DGV validated accuracy, but the size of the training population, in general, influenced DGV validated accuracy. The losses in GEBV expected accuracy were lower when using the 50K panel imputed from the 15K panel than form the 8K panel. The increase in GEBV expected accuracy in the prediction population by adding information from molecular markers was small compared to simple BLUP.

**Keywords:** Direct genomic value, Genomic selection, Imputation, Beef cattle, Accuracy.

**Background**

Animal breeding for important economic traits has been practiced over the years based on phenotypes and relationships among individuals. Recent advances in DNA (Deoxyribonucleic Acid) analysis led to the complete sequencing of several species including cattle [1]. From these advances, new technologies emerged and currently dense panels for genotyping SNPs (Single Nucleotide Polymorphisms) are available, such as Illumina BeadChip BovineHD (Illumina Inc., San Diego, USA) that enables genotyping 777K SNPs from a sample in a single panel. This new genotyping technology has triggered the worldwide development of new research, with substantial allocation of human and financial resources, in order to use the information generated from SNP genotypes in animal breeding.

Incorporation of genomic information in livestock breeding programs is expected to result in substantially higher genetic gain in shorter period of time [2], [3], [4]. These gains would be associated with decreased generation intervals, increased accuracy of selection and incorporation of new traits of economic importance [2], [5], [6], [7].

Particularly, in dairy cattle, this new technology has provided substantial financial savings in progeny tests [8], [9]. Accuracy of direct genomic value (DGV), calculated from the estimated effects of molecular markers, depends on many factor such as the level of linkage disequilibrium between markers and quantitative trait loci (QTL), the number of animals in the training population, the heritability and the distribution of QTL effects [9]. The genomic selection has changed considerably the dairy cattle breeding, especially young bull testing, where some countries today have partially or completely eliminated the traditional progeny test. The success of genomic selection in dairy cattle, mainly in the Holstein breed, is associated with a large number of genotyped animals, a small effective population size, key sires are used in all countries with important milk production and there is cooperation between countries with respect to the use of genotypes [9]. The picture for beef cattle is different, because the effective population size is larger, there are several important breeds with few key sires used across countries and also the cooperation between countries is minimal [6].

Genotyping cost in large commercial herds is the major financial constrain for implementing genomic selection. Farmers would likely use in their commercial breeding programs lower density panels, which are more affordable. These low density panels would be imputed to a medium or high density panel [10], [11] and would be used to predict the genomic values for the animals [12], [13].

The aim of this study was to investigate the accuracy of genomic predictions using real 50K SNP genotypes, as well as using different percentages of imputed SNP genotypes and different sizes of training population in Braford and Hereford cattle.

## Methods

### Animal welfare

Animal welfare and use committee approval was not necessary for this study because data were obtained from preexisting databases.

### Genotype and phenotype data

Data was obtained from the Conexão Delta G's genetic improvement program - Hereford and Braford (Zebu x Hereford) cattle (Conexão Delta G, Dom Pedrito/RS, Brazil). The dataset contained approximately 520,000 animals from 97 farms located in the South, Southeast, Midwest and Northeast regions of Brazil. Out of these animals there were 683 Hereford and 2,997 Braford animals genotyped born from 2008 to 2011 plus 130 sires. There were 624 Hereford and 2,926 Braford animals genotyped with the Illumina BovineSNP50 panel and 59 Hereford and 71 Braford siresgenotyped with the Illumina BovineHD panel from 17 farms located in the South of Brazil. Data also included 88 Nellore bulls from the Paint Program (Lagoa da Serra, Sertãozinho/SP, Brazil) genotyped with the Illumina BovineHD panel.

### Genotype data editing

For imputation to the 50K SNP panel, animals genotyped with 777K SNP genotypes had SNPs not contained on the 50K SNP panel removed. The missing genotypes (= 0.46%) in the 50K SNP panel were previously imputed. Sites were filtered for GenCall score ($>=0.15$), Call Rate ($>=0.90$) and Hardy-Weinberg Equilibrium ($P>=10^{-6}$). Only autosomes were considered. The individual sample quality control considered GenCall Score ($>=0.15$), Call Rate ($>=0.90$), heterozygosity deviation (limit of $\pm$ 3 SD), repeated sampling and paternity errors. The SNP and sample quality control,

for imputation to the 777K SNP panel, were the same as for the imputation to the 50K SNP panel. The 8K and 15K SNP panels were used for imputation to the 50K SNP panel and the 50K SNP panel was used for imputation to the 777K SNP panel [11] using FImpute v.2.2 [14].

For estimation the markers effect were used the same quality controls used in the imputation plus Minor Allele Frequency (>=0.05).

**Training and prediction populations**

The dataset was split into two groups for the analysis. The training group included all animals born before 2011 and the prediction group included all the animals born in 2011 and some bulls born before 2011 that did not have phenotypic information and had no offspring until 2010. Training and prediction groups varied in size for each trait (Table 1).

**Scenarios**

Three groups of scenarios per trait were defined. The first two groups of scenarios were created from 8K and 15K SNP panels imputed to the 50K SNP panel. The third group/scenario was created from the 50K SNP panel imputed to the 777K SNP panel. The first group of scenarios (SCE1) was created with different percentages of animals with imputed genotypes and unequal training population sizes. The second group of scenarios (SCE2) was also created with different percentages of animals with imputed genotypes, but with equal training population size. The third group/scenario (SCE3) was created with only one percentage of animals with imputed genotypes and only one training population size. The definitions of the scenarios are presented in the Table 2.

**Traditional genetic evaluation**

The package used to obtain the trait estimated breeding values (EBV) was written in Fortran language, developed by GenSys (GenSys Consultores Associados, Porto Alegre, Brazil) and considers the degree of connectedness among contemporaneous groups in multi trait animal models with robust estimation procedures in relation to the heterogeneity of variance of contemporary groups and the residual size for each observation [15]. Two EBVs sets were generated. The first set was estimated using all available information to date while the second set was estimated when information for all animals born after 2010 was discarded. These two sets of EBVs were then used for validation and training, respectively.

**Deregressed EBV**

The second EBV set of the training population was deregressed and used as pseudo phenotypes to estimate markers effects. The approach of VanRaden & Wiggans [16] was used to calculate deregressed EBVs using EBVs and reliabilities of genotyped animals and their sires and dams. Deregressed EBVs were calculated for animals of the training population with EBV reliability greater than the average ($r^2$=0.09) and that satisfied the condition:

$$abs\left( \frac{(EBV - dEBV)}{dpEBV} \right) \leq 10 dpEBV \, ,$$

where, EBV is the estimated breeding value, dEBV is a deregressed EBV and dpEBV is the standard deviation of the EBVs.

**Prediction of DGV and GEBV**

Direct genomic values (DGV) were estimated using GBLUP method [5] for all the twenty traits (Table 2), using either 50K or 777K SNP panels and deregressed EBVs in the GEBV package [17]. The following linear model was assumed:

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Zg} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of deregressed EBV for the trait, $\mu$ is the overall mean, $\mathbf{1_n}$ is a vector of ones, $\mathbf{Z}$ is the design matrix that relates deregressed EBVs to animals, $\mathbf{g}$ is the vector of DGV to be predicted, and $\mathbf{e}$ is the vector of residual effects. It was assumed that $\mathbf{g} \sim N (0, \mathbf{G^*}\sigma^2 g)$ where $\sigma^2 g$ is the additive genetics variance and $\mathbf{G^*}$ is a combined relationship matrix (80% genomic relationship and 20% pedigree-based relationship) , and $\mathbf{e} \sim N (0, \mathbf{R}\sigma^2 e)$ where $\sigma^2 e$ is the residual variance and $\mathbf{R}$ is a diagonal matrix whose elements account for the differences in reliabilities of the deregressed EBVs.

The genomic estimated breeding values (GEBV) were estimated using the blending procedure outlined by Hayes et al. [9] and described below:

$$\mathbf{GEBV} = \frac{\mathbf{r_{DGV}^2 * DGV + r_{EBV}^2 * EBV}}{\mathbf{r_{DGV}^2 + r_{EBV}^2}} \text{ where,}$$

$\mathbf{r_{DGV}^2}$ and $\mathbf{r_{EBV}^2}$ are the reliability of DGV and EBV, respectively.

**Comparison criteria between scenarios**

The accuracies of genomic predictions were calculated in two different ways and were used to express the results of this study.

(1) Pearson's correlation between DGVs and EBVs in the prediction population was used as measure of accuracy in each scenario in this study and was termed as *validated accuracy.*

(2) Accuracy obtained from the mixed model equation in the prediction population was used to quantify the losses in GEBV accuracy by the use of imputed panel compared to the original panel and also was used to quantify the gain in EBV accuracy by the use of molecular marker information in the EBV estimation, was termed as *expected accuracy.*

Validated accuracy and losses in GEBV expected accuracy were used in the analysis of variance carried out by ANOVA procedure of SAS version 9.2 (SAS Inst. Inc., Cary, NC).

The results are presented based on the traits that make up the selection index used by Conexão Delta G's genetic improvement program and another group of traits that are not included in the selection index, but are used for independent culling selection. The selection index puts the following weights on the traits: 25% for weight gain from birth to weaning (WGBW), 25% for weight gain from weaning to yearling (WGWY), 4% for conformation score at weaning (CW), 4% for conformation score at yearling (CY), 8% for precocity score at weaning (PW), 8% for precocity score at yearling (PY), 8% for muscularity score at weaning (MW), 8% for muscularity score at yearling (MY), 5% for scrotal circumference adjusted for age at yearling (SCa), 5% for scrotal circumference adjusted for age and weight at yearling (SCaw).

## Results

### Traits in the selection index

Validated accuracy for the ten traits in the selection index were between 0.29 to 0.31 across all the scenarios in SCE1, SCE2 and SCE3 (Table 3 and 4).

In the SCE1scenarios, where the size of the training population was increased by an increased percentage of imputed animals, the size of training population had significant impact on the validated accuracy, generally in favor of the scenario with the highest number of training animals. The comparison of the 8K and 15K SNP panels imputed to the 50K SNP panel to the true 50K SNP panel showed significant differences (P<0.05) in 45% and 40% of the cases, respectively (Table 5). These differences were generally in favor of the true 50K SNP panel. The differences between the 8K and 15K SNP panels were not significant (P>0.05) in 55% of the cases (Table 5).

In the SCE2 scenarios, where the training population size was held constant but the percentage of imputed animals varied, there were no significant difference among the alternate percentage of imputed animals (P>0.05) for all traits. The comparison of the 8K and 15K SNP panels to the true 50K SNP panel showed significant differences (P<0.05) in 45% and 60% of the cases, respectively (Table 6). These differences were generally in favor of the true 50K SNP panel. The differences between the 8K and 15K SNP panels were not significant (P>0.05) in 55% of the cases (Table 6).

Average EBV expected accuracy in the training and prediction population, for the ten traits in the selection index, were 0.64 and 0.63, respectively. Average GEBV expected accuracy was 0.66 for the scenario with all animals and with 60% imputed genotypes (SCE1-60% and SCE2-60%) and 0.65 for the SCE3 scenario. The increase in average GEBV expected accuracy in the prediction population by adding the information of the markers was 0.03. The average DGV expected accuracy across traits was 0.40 (Table 7).

Losses in GEBV expected accuracy were measured within each level of the scenario in relation to the same level of the scenario using only the real genotypes. All losses were statistically different from real 50K SNP panel (P<0.05) and were higher when using the 8K SNP panel in relation to the 15K SNP panel. For the 8K and 15K SNP panel the average losses in GEBV expected accuracy were between -0.004 and -0.0011 and -0.0002 and -0.0011 across scenarios in SCE1 and SCE2, respectively (Tables 8 and 9). In the SCE3 scenario using the 777K SNP panel imputed from the 50K SNP panel the average loss in GEBV expected accuracy was -0.0021 (Table 9).

### Traits not in the selection index

Validated accuracy for the ten traits not included in the selection index ranged between 0.47 and 0.50 across scenarios in SCE1, SCE2 and SCE3. However, for the traits related to fitness (NW, NY, HW, HY, TR and OP), the average validated accuracy varied from 0.63 to 0.65 (Tables 3 and 4).

As observed for the traits included in the selection index, the validated accuracy showed no significant difference (P>0.05) in SCE2, regardless of which panel the genotypes were imputed from. Regarding to the panels, there were no significant differences in 60% of the cases (P>0.05) between the 8K and 15K SNP panels and the true 50K SNP panel (Table 6).

When evaluating the size and percentage of animals with imputed genotypes in the training population in SCE1 scenarios, the validated accuracy showed that 86% of

the comparisons were statistically different (P<0.05). In general, the comparisons were in favor of scenarios with larger training population. Regarding the panel used, 60% of the comparisons of the 8K and 15K SNP panel to the true 50K panel were statistically significant (P<0.05) and in favor of the true 50K SNP panel (Table 5).

Average EBV expected accuracy in the training and prediction population, of these ten traits, were 0.64 and 0.63, respectively. These results were equal to those attained for traits in the selection index. Average GEBV expected accuracy was 0.65 for the scenarios with all animals and with 60% imputed genotypes and for the SCE3 scenario. The increase in GEBV expected accuracy by adding marker information was about 0.02 in all scenarios. The average DGV expected accuracy, was 0.40, which was also equal to that for traits in the selection index (Table 7).

Losses in GEBV expected accuracy were statistically different from real 50K SNP panel (P<0.05) for this group of traits which have higher values than the traits in the selection index. Using the 8K and 15K SNP panels imputed to the 50K SNP panel the average losses in GEBV expected accuracy were between -0.0004 and -0.0013 and -0.0003 to -0.0013, respectively, across SCE1 and SCE2 scenarios (Table 8 and 9). In the SCE3 scenario that used the 777K SNP panel imputed from the 50K SNP panel, the loss in average GEBV expected accuracy was -0.0021. The same value was attained for the selection index traits (Table 9).

## Discussion

### Traits in the selection index

Conexão Delta G's genetic improvement program - Hereford and Braford (Nellore x Hereford) started in the 1970s. During the first years, animals were selected using a selection index that included weight gain, scrotal circumference and conformation score traits [18]. In 1975, precocity, muscling and body size scores [19] were incorporated into the selection index. In the 90s, body size score was excluded from the selection index. The selection index showed some variations over the years in the weighting of the traits, but in general the selection index put 50% for weight gain, 40% for conformation, precocity and muscling score and 10% for scrotal circumference.

The correlation between DGV and EBV (validated accuracy) has been used to represent the accuracy of DGV [4], [20], [21]. The validated accuracy for traits in the selection index showed lower values than those reported by Saatchi et al. [20] working with Limousin and Simmental breeds in the United States and Saatchi et al. [22] and Boddhireddy et al. [21] both working with Angus in the United States. While Neves et al. [23], working with Nellore in Brazil and with the same set of traits in the selection index, also found larger validated accuracy than those found in this study, except for WGBW and CW. These lower values of validated accuracy in this study are also related to the lowest values of EBV expected accuracy in the training population (r=0.64). In dairy cattle, the validated accuracy are much higher in comparison with beef cattle [24], [13], [25]. Those higher values are associated with more reliable EBV values for training group, higher genomic relationship between training and prediction groups, and higher extent of linkage disequilibrium [26]. Validated accuracy were greater at higher trait heritability values (e.g., post weaning traits). These results were similar to those found by Brito et al. [4], working with simulated data in beef cattle, by Akanno et al. [27], working with simulated data in pigs, and by Khatkar et al. [28], working with dairy cattle in Australia.

Scrotal circumference has the highest heritability in the group in the selection index, but showed the lowest validated accuracy. This was probably related to the smaller number of animals in the training population (n=708).

In the SCE2 scenarios, where the size of the training population was constant, the validated accuracy were no statistical differences and in the SCE1 scenarios, where the size and the percentage of animals with imputed genotypes in the training population varied together, the validated accuracy were statistical differences, showed that the size of the training population was more important than the percentage of animals with imputed genotypes. Berry and Kearney [12], Khatkar et al. [28], Dassonneville et al. [29], Segelke et al. [30] and Mulder et al. [13], studying the effect of the presence of imputed genotypes in the training population in dairy cattle, showed that small losses in reliability were observed using imputed genotypes to predict the effect of the markers. The highest GEBV expected accuracy were observed with the traits with higher heritabilities (post weaning traits, exception WGWY) but with slightly higher gains in GEBV expected accuracy for the pre weaning traits (lower heritability) in accordance with the results showed by Brito et al. [4].

The DGV expected accuracy for the different traits and SNP panels were about 0.40 being slightly higher for the pre weaning traits in comparison to the post weaning traits. DGV expected accuracy were close to the parent average expected accuracy for the traits in the selection index (average of ten traits was 0.55), showing that the selection based on the DGV when the parent average is not known, can be used with some loss in accuracy. The lower expected accuracies obtained for GEBV and DGV in SCE3 scenario are probably related to the fact that only 212 animals in the training population had true genotypes. Brito et al. [4], working with beef cattle simulated data, found an increase of 0.09 in the DGV accuracy by using a 777K SNP panel instead of a 50K SNP panel when the training population was formed by 480 sires of highly accurate EBVs.

Despite of all the losses in GEBV expected accuracy using 8K and 15K SNP panels being statistically significant when compared to the real 50K panel, they were lower with the 15K SNP panel. These results are associated with the highest concordance rate in the 50K SNP panel imputed from the 15K SNP panel [11]. This same behavior was found by Segelke et al. [30] when they analyzed the losses in reliability from imputed panels of two different densities of SNPs in dairy cattle in Germany, while Sargolzaei et al. [31], working with dairy cattle in Canada with one density of SNP panel in the imputed panel, also showed losses in reliability with values around -0.02. When the percentage of animals with imputed genotypes in the training population in the SCE2 scenarios increased the losses also increased. The same behavior was observed in the SCE1 scenarios. In the SCE3 scenario, losses were much higher and related to higher percentage imputation error, since, in this scenario, only 212 animals had true genotypes [11].

**Traits not in the selection index**

The independent culling level was carried out systematically since the beginning of the Conexão Delta G's genetic improvement program for BW, BA and OP traits, particularly in Hereford, and NW, NY, HW and HY in Braford. The SW and SY traits were part of the selection index between the 1970s and 1990s while the selection of TR has been performed with greater emphasis on young bulls in the last decade.

The traits related to fitness (NW, NY, HW, HY, TR and OP) had values of validated accuracy higher in relation to the other traits, including those from the

selection index. These higher values are probably associated with greater genetic variability due to a milder selection. The results found by Akanno et al. [27], working with simulated data in swine strengthens this theory because they found much higher accuracy for the indigenous population (smaller selection pressure) in comparison with the exotic population (high selection pressure). However, Neves et al. [23] studying Nellore in Brazil found validated accuracy lower than those attained for the NW and NY traits. This may be due to a strong selection performed in Nellore breed.

The validated accuracy for the BW and the BA in this study were lower than other traits studied. This was probably related to the strong selection which is carried out in Hereford breed. Saatchi et al. [22], [20] and [32] working with Angus, Limousin, Simmental and Hereford breeds, found higher validated accuracy for these two traits compared to those reported in this study. Validated accuracy for BW in the SCE2 scenario were not influenced by either the panel or the percentage of imputed animals in the training population. Different results were found in the SCE1 scenario, where both the panel and the number of animals in the training population influenced the validated accuracy of this trait. Hayes et al. [9] showed that the values of accuracy varied according to the size of the training population and Brito et al. [4] working with simulated data from beef cattle, showed that the size of the training population has a major effect on the accuracies. These results were also observed for the fitness traits (NW, NY, HW, HY, TR and OP).

For the SW and the SY in the SCE2 scenario, the percentage of animals with imputed genotypes was not significant and in the SCE1 scenario it was significant when varying the size and the percentage of animals with imputed genotypes in the training population. This showed that the effect could due to the population size. The behavior of these two traits was similar to these traits' scores in the selection index, most likely because these traits were used together in the selection index for many years.

The GEBV and DGV expected accuracy in this group of traits were similar to the expected accuracy of the traits in the selection index. Despite the fact that the DGV of TR and OP traits have presented a high correlation with EBV, the DGV expected accuracy for all scenarios was much lower than the accuracy of the parents average to the same level of heritability. In general, the DGV expected accuracy for all traits across levels of each scenario were lower than the accuracy of the parents average reported by Brito et al. [4] which was 0.44 to 0.58 for traits with heritability between 0.10 to 0.40.

The losses in GEBV expected accuracy of each scenario were always analyzed relative to the scenario where only real genotypes were used. For this group of traits in the different scenarios, the losses in GEBV expected accuracy were, on average, higher compared to the group of traits in the selection index. However, the losses in GEBV expected accuracy were higher when using to the 50K SNP imputed panel from the 8K SNP panel, which were similar to the traits in the selection index and all differences were statistically different from real 50K SNP panel. The higher the percentage of animals with imputed genotypes, the higher the losses in GEBV expected accuracy, regardless of the scenario. The same behavior was observed in the SCE3 scenario, in other words, losses in GEBV expected accuracy were higher due to the higher error rate in the imputation [11].

## Conclusion

The percentage of animals with imputed genotypes in the training population did not significantly influence the validated accuracy (Pearson's correlation), but the size of the training population influenced these validated accuracy in the prediction population.

A small gain in EBV expected accuracy (accuracy obtained from the inverse of the mixed model equations) was found when including molecular marker information in the EBV estimation. The losses in GEBV expected accuracy due to imputation of genotypes were lower when using the 50K SNP panel imputed from the 15K SNP panel instead of imputation from the 8K SNP panel.

**Abbreviations**
SNP: Single nucleotide polymorphism; DNA: Deoxyribonucleic acid, QTL: Quantitative trait loci; K: Kbytes; SD: Standard deviation; DGV: Direct genomic value; GEBV: Genomic estimated breeding value; EBV: Estimated breeding value; dEBV: Deregressed estimated breeding value; BLUP: Best linear unbiased prediction; GBLUP: Genomic best linear unbiased prediction; SAS: Statistical analysis sistem; ANOVA: analysis of variance; NC: North Carolina; USA: United States of America; SCE1: First scenario; SCE2: Second scenario; SCE3: Third scenario; WGBW: Weight gain from birth to weaning (kg); WGWY: Weight gain from weaning to yearling (kg); CW: Conformation score at weaning (scores 1-5); CY: Conformation score at yearling (scores 1-5); PW: Precocity score at weaning (scores 1-5); PY: Precocity score at yearling (scores 1-5); MW: Muscularity score at weaning (scores 1-5); MY: Muscularity score at yearling (scores 1-5); SCa: Scrotal circumference adjusted for age at yearling (cm); SCaw: Scrotal circumference adjusted for age and weight at yearling (cm); BW: Birth weight (kg); BA: Birth assistance score (scores 1-5); SW: Size score at weaning (scores 1-5); SY: Size score at yearling (scores 1-5); NW: Prepuce (navel) score at weaning (scores 1-5); NY: Prepuce (navel) score at yearling (scores 1-5); HW: Hair length score at weaning (scores 1-3); HY: Hair length score at yearling (scores 1-3); TR: Ticks resistance (ticks unit); OP: Ocular pigmentation score (scores 1-3).

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
MLP participated in the design of the study, carried out the analyses, was involved in the discussions, and prepared and drafted the manuscript. JB participated in the design of the study, was involved in the discussions and helped to draft the manuscript. FFC was involved in the discussions and helped to draft the manuscript. MS developed the GEBV software, was involved in the discussions, and helped to draft the manuscript. FSS participated in the design of the study, was involved in the discussions and helped to draft the manuscript. All authors read and approved the final manuscript.

**References**

1. The Bovine Genome Sequencing and Analysis Consortium., Elsik CG, Tellam RL, Worley KC: **The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution**. *Science* 2009, **324**:522–528.

2. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819–1829.

3. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score.** *J Dairy Sci* 2010, **93**:743–752.

4. Brito F V, Neto JB, Sargolzaei M, Cobuci JA, Schenkel FS: **Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle.** *BMC Genet* 2011, **12**:80.

5. VanRaden PM: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91**:4414–4423.

6. Miller S: **Genetic improvement of beef cattle through opportunities in genomics**. *Rev Bras Zootec* 2010, **2010 supl.**:247–255.

7. Goddard M, Hayes B, Meuwissen T: **Genomic selection in farm animal species-lessons learnt and future perspectives**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany; 2010.

8. Schaeffer LR: **Strategy for applying genome-wide selection in dairy cattle.** *J Anim Breed Genet* 2006, **123**:218–223.

9. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: **Invited review: Genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**:433–443.

10. Ventura RV, Lu D, Schenkel FS, Wang Z, Li C, Miller SP: **Impact of reference population on accuracy of imputation from 6K to 50K SNP chips in purebred and crossbreed beef cattle**. *J Anim Sci* 2014, **92**:1433–1444.

11. Piccoli ML, Braccini J, Cardoso FF, Sargolzaei M, Schenkel FS: **Accuracy of genome-wide imputation in Braford and Hereford beef cattle**. *BMC Genet* 2014, 15: 1275.

12. Berry DP, Kearney JF: **Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection.** *Animal* 2011, **5**:1162–1169.

13. Mulder HA, Calus MPL, Druet T, Schrooten C: **Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle.** *J Dairy Sci* 2012, **95**:876–889.

14. Sargolzaei M, Chesnais JP, Schenkel FS: **A new approach for efficient genotype imputation using information from relatives.** *BMC Genomics* 2014, **15**:478.

15. Carvalheiro R, Fries LA, Schenkel FS, Albuquerque LG: **Effects of Heterogeneity of Residual Variance among Contemporary Groups on Genetic Evaluation of Beef Cattle**. *R Bras Zootec* 2002, **31**:1680–1688.

16. VanRaden PM, Wiggans GR: **Derivation, calculation, and use of national animal model information.** *J Dairy Sci* 1991, **74**:2737–2746.

17. Sargolzaei M, Schenkel FS, VanRaden PM: **gebv: Genomic breeding value estimator for livestock**. In *Tech Rep to Dairy Cattle Breed Genet Comm Univ Guelph*. Guelph/ON - Canada; 2009.

18. BIF: *Guidelines For Uniform Beef Improvement Programs*. 2th ed. Athens, GA, USA; 1972.

19. Long RA: *El Sistema de Evaluación de Ankony Y Su Aplicación En La Mejora Del Ganado*. 1975:1–13.

20. Saatchi M, Schnabel RD, Rolf MM, Taylor JF, Garrick DJ: **Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle.** *Genet Sel Evol* 2012, **44**:38.

21. Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, S. D: **Genomic predictions in Angus cattle : Comparisons of sample size , response variables , and clustering methods for cross-validation**. *J Anim Sci* 2014, **92**:485–497.

22. Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel RD, Garrick DJ, Taylor JF: **Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation.** *Genet Sel Evol* 2011, **43**:40.

23. Neves HHR, Carvalheiro R, O'Brien AMP, Utsunomiya YT, do Carmo AS, Schenkel FS, Sölkner J, McEwan JC, Van Tassell CP, Cole JB, da Silva MVGB, Queiroz S a, Sonstegard TS, Garcia JF: **Accuracy of genomic predictions in Bos indicus (Nellore) cattle.** *Genet Sel Evol* 2014, **46**:17.

24. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited review: reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92**:16–24.

25. Su G, Brøndum RF, Ma P, Guldbrandtsen B, Aamand GP, Lund MS: **Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations.** *J Dairy Sci* 2012, **95**:4657–4665.

26. Habier D, Fernando RL, Dekkers JCM: **The impact of genetic relationship information on genome-assisted breeding values.** *Genetics* 2007, **177**:2389–2397.

27. Akanno EC, Schenkel FS, Sargolzaei M, Friendship RM, Robinson J a B: **Persistency of accuracy of genomic breeding values for different simulated pig breeding programs in developing countries.** *J Anim Breed Genet* 2014:1–12.

28. Khatkar MS, Moser G, Hayes BJ, Raadsma HW: **Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle.** *BMC Genomics* 2012, **13**:538.

29. Dassonneville R, Brøndum RF, Druet T, Fritz S, Guillaume F, Guldbrandtsen B, Lund MS, Ducrocq V, Su G: **Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations.** *J Dairy Sci* 2011, **94**:3679–3686.

30. Segelke D, Chen J, Liu Z, Reinhardt F, Thaller G, Reents R: **Reliability of genomic prediction for German Holsteins using imputed genotypes from low-density chips.** *J Dairy Sci* 2012, **95**:5403–5411.

31. Sargolzaei M, Schenkel FS, Chesnais J: **Comparison between the use of true and imputed genotypes for predicting the GPA of young bulls Data : Strategy :** *Rep Present Dairy Cattle Breed Genet Comm Meet* 2010:1–8.

32. Saatchi M, Ward J, Garrick D: **Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations**. *J Anim Sci* 2013, **91**:1538–1551.

33. Madsen P, Jensen J, Labouriau R, Christensen OF, Sahana G: **DMU - A Package for Analyzing Multivariate Mixed Models in quantitative Genetics and Genomics**. In *Proc 8th World Congr Genet Appl to Livest Prod 13-18 August 2006*. Belo Horizonte/MG - Brazil; 2006.

**Table 1 Number of phenotypes, EBVs and genotypes for each economic trait in the training and prediction population after data editing**

| Economic traits | Abbr. | $h^2$ | Phenotypes | Genotypes | Training population | | | | Prediction population | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Animal | Sire | Dam | Total | Animal | Sire | Dam | Total |
| *Traits in the selection index* | | | | | | | | | | | | |
| Weight gain from birth to weaning (kg) | WGBW | 0.25 | 354,255 | 3,305 | 2,231 | 91 | 3 | 2,325 | 944 | 9 | 27 | 980 |
| Weight gain from weaning to yearling (kg) | WGWY | 0.31 | 164,140 | 2,539 | 1,520 | 91 | 3 | 1,614 | 907 | 6 | 12 | 925 |
| Conformation score at weaning (scores 1-5) | CW | 0.25 | 348,020 | 2,970 | 1,896 | 91 | 3 | 1,990 | 944 | 9 | 27 | 980 |
| Conformation score at yearling (scores 1-5) | CY | 0.32 | 171,406 | 2,768 | 1,717 | 91 | 3 | 1,811 | 939 | 6 | 12 | 957 |
| Precocity score at weaning (scores 1-5) | PW | 0.25 | 330,312 | 2,961 | 1,887 | 91 | 3 | 1,981 | 944 | 9 | 27 | 980 |
| Precocity score at yearling (scores 1-5) | PY | 0.32 | 160,261 | 2,768 | 1,717 | 91 | 3 | 1,811 | 939 | 6 | 12 | 957 |
| Muscularity score at weaning (scores 1-5) | MW | 0.25 | 330,059 | 2,968 | 1,894 | 91 | 3 | 1,988 | 944 | 9 | 27 | 980 |
| Muscularity score at yearling (scores 1-5) | MY | 0.32 | 159,706 | 2,768 | 1,717 | 91 | 3 | 1,811 | 939 | 6 | 12 | 957 |
| Scrotal circumference a (cm) | SCa[1] | 0.43 | 46,823 | 1,581 | 623 | 85 | 0 | 708 | 865 | 5 | 3 | 873 |
| Scrotal circumference aw (cm) | SCaw[1] | 0.43 | 46,823 | 1,581 | 623 | 85 | 0 | 708 | 865 | 5 | 3 | 873 |
| *Traits not in the selection index* | | | | | | | | | | | | |
| Birth weight (kg) | BW | 0.33 | 221,038 | 3,434 | 2,401 | 88 | 3 | 2,492 | 905 | 10 | 27 | 942 |
| Birth assistance score (scores 1-5) | BA | 0.10 | 26,058 | 1,581 | 1,123 | 44 | 0 | 1,167 | 395 | 3 | 16 | 414 |
| Size score at weaning (scores 1-5) | SW | 0.25 | 140,681 | 2,911 | 1,848 | 81 | 3 | 1,932 | 944 | 8 | 27 | 979 |
| Size score at yearling (scores 1-5) | SY | 0.41 | 84,261 | 2,737 | 1,694 | 84 | 3 | 1,781 | 939 | 5 | 12 | 956 |
| Prepuce (navel) score at weaning (scores 1-5) | NW | 0.46 | 265,802 | 3,343 | 2,291 | 89 | 2 | 2,382 | 927 | 7 | 27 | 961 |
| Prepuce (navel) score at yearling (scores 1-5) | NY | 0.41 | 122,409 | 2,825 | 1,793 | 89 | 2 | 1,884 | 924 | 5 | 12 | 941 |
| Hair length score at weaning (scores 1-3) | HW | 0.23 | 110,163 | 1,579 | 612 | 86 | 2 | 700 | 846 | 6 | 27 | 879 |
| Hair length score at yearling (scores 1-3) | HY | 0.31 | 73,621 | 2,480 | 1,561 | 88 | 2 | 1,651 | 813 | 4 | 12 | 829 |
| Ticks resistance (ticks unit) | TR | 0.19 | 56,978 | 1,640 | 594 | 60 | 0 | 654 | 948 | 3 | 3 | 954 |
| Ocular pigmentation score (scores 1-3) | OP | 0.20 | 139,082 | 1,635 | 587 | 90 | 1 | 678 | 921 | 9 | 27 | 957 |

[1] SCa is the scrotal circumference adjusted for age at yearling and SCaw is the scrotal circumference adjusted for age and weight at yearling;

[2] The heritability estimates were obtained prior to this study using the DMU package [33].

**Table 2 Number of animals with true and imputed genotypes in each scenario for the weight gain from birth to weaning (WGBW) trait[1]**

| Scenarios SCE1 – 50K SNP panel | | | | Scenarios SCE2 – 50K SNP panel | | | | Scenario SCE3 – 777K SNP panel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Genotypes | | | | Genotypes | | | | Genotypes | | |
| Level[2] | True | Imputed | Total | Level[2] | True | Imputed | Total | Level[2] | True | Imputed | Total |
| 0 | 2,325 | 0 | 2.325 | 0 | 2.325 | 0 | 2.325 | 90 | 212 | 2,113 | 2,325 |
| 10 | 930 | 103 | 1,033 | 10 | 2,092 | 233 | 2,325 | | | | |
| 20 | 930 | 233 | 1,163 | 20 | 1,860 | 465 | 2,325 | | | | |
| 30 | 930 | 399 | 1,329 | 30 | 1,627 | 698 | 2,325 | | | | |
| 40 | 930 | 620 | 1,550 | 40 | 1,395 | 930 | 2,325 | | | | |
| 50 | 930 | 930 | 1,860 | 50 | 1,162 | 1,163 | 2,325 | | | | |
| 60 | 930 | 1,395 | 2,325 | 60 | 930 | 1,395 | 2,325 | | | | |

[1] The same criteria were used to define the scenarios of other traits, but the number of genotypes varies for each trait;

[2] Percentage of animals with imputed genotypes.

**Table 3 DGV validated accuracy in the prediction population for the SCE1 scenarios[12]**

| Traits[3] | 8K[4] | | | | | | 15K[4] | | | | | | 50K[4] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 10 | 20 | 30 | 40 | 50 | 60 | 10 | 20 | 30 | 40 | 50 | 60 |
| *Traits in the selection index* | | | | | | | | | | | | | | | | | | |
| WGBW | 0.32 | 0.33 | 0.34 | 0.34 | 0.34 | 0.32 | 0.32 | 0.33 | 0.34 | 0.33 | 0.33 | 0.32 | 0.31 | 0.33 | 0.34 | 0.33 | 0.34 | 0.32 |
| WGWY | 0.33 | 0.33 | 0.33 | 0.34 | 0.36 | 0.35 | 0.33 | 0.33 | 0.33 | 0.33 | 0.36 | 0.35 | 0.33 | 0.33 | 0.33 | 0.33 | 0.36 | 0.35 |
| CW | 0.24 | 0.27 | 0.29 | 0.30 | 0.31 | 0.29 | 0.25 | 0.28 | 0.29 | 0.30 | 0.31 | 0.28 | 0.25 | 0.28 | 0.29 | 0.30 | 0.31 | 0.29 |
| CY | 0.29 | 0.29 | 0.29 | 0.28 | 0.29 | 0.26 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.26 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.26 |
| PW | 0.28 | 0.29 | 0.30 | 0.31 | 0.33 | 0.31 | 0.27 | 0.29 | 0.29 | 0.31 | 0.33 | 0.31 | 0.27 | 0.29 | 0.30 | 0.31 | 0.33 | 0.31 |
| PY | 0.32 | 0.32 | 0.31 | 0.32 | 0.35 | 0.33 | 0.32 | 0.32 | 0.31 | 0.32 | 0.35 | 0.34 | 0.32 | 0.32 | 0.31 | 0.33 | 0.35 | 0.34 |
| MW | 0.32 | 0.33 | 0.34 | 0.34 | 0.37 | 0.35 | 0.32 | 0.33 | 0.34 | 0.34 | 0.37 | 0.35 | 0.31 | 0.33 | 0.34 | 0.34 | 0.37 | 0.35 |
| MY | 0.35 | 0.35 | 0.33 | 0.34 | 0.37 | 0.34 | 0.35 | 0.35 | 0.33 | 0.35 | 0.38 | 0.35 | 0.35 | 0.35 | 0.33 | 0.34 | 0.38 | 0.35 |
| SCa | 0.21 | 0.21 | 0.20 | 0.20 | 0.17 | 0.18 | 0.22 | 0.23 | 0.21 | 0.21 | 0.17 | 0.18 | 0.22 | 0.22 | 0.21 | 0.20 | 0.18 | 0.18 |
| SCaw | 0.24 | 0.23 | 0.22 | 0.21 | 0.17 | 0.20 | 0.25 | 0.24 | 0.22 | 0.22 | 0.18 | 0.19 | 0.25 | 0.25 | 0.23 | 0.22 | 0.18 | 0.20 |
| **Average** | **0.29** | **0.30** | **0.30** | **0.30** | **0.31** | **0.29** | **0.29** | **0.30** | **0.30** | **0.30** | **0.31** | **0.29** | **0.29** | **0.30** | **0.30** | **0.30** | **0.31** | **0.30** |
| *Traits not in the selection index* | | | | | | | | | | | | | | | | | | |
| BW | 0.21 | 0.21 | 0.22 | 0.21 | 0.21 | 0.22 | 0.21 | 0.21 | 0.23 | 0.21 | 0.21 | 0.21 | 0.20 | 0.21 | 0.22 | 0.20 | 0.20 | 0.21 |
| BA | 0.12 | 0.14 | 0.12 | 0.13 | 0.14 | 0.14 | 0.14 | 0.15 | 0.14 | 0.14 | 0.15 | 0.15 | 0.13 | 0.14 | 0.13 | 0.13 | 0.14 | 0.14 |
| SW | 0.28 | 0.30 | 0.31 | 0.32 | 0.31 | 0.33 | 0.28 | 0.30 | 0.31 | 0.31 | 0.31 | 0.33 | 0.27 | 0.30 | 0.31 | 0.31 | 0.31 | 0.33 |
| SY | 0.29 | 0.30 | 0.30 | 0.30 | 0.33 | 0.36 | 0.30 | 0.31 | 0.31 | 0.30 | 0.33 | 0.35 | 0.30 | 0.31 | 0.31 | 0.30 | 0.33 | 0.36 |
| NW | 0.47 | 0.49 | 0.49 | 0.50 | 0.51 | 0.53 | 0.47 | 0.48 | 0.49 | 0.50 | 0.51 | 0.53 | 0.47 | 0.49 | 0.49 | 0.50 | 0.52 | 0.54 |
| NY | 0.46 | 0.48 | 0.50 | 0.50 | 0.52 | 0.54 | 0.46 | 0.48 | 0.50 | 0.50 | 0.51 | 0.54 | 0.46 | 0.48 | 0.50 | 0.50 | 0.52 | 0.54 |
| HW | 0.73 | 0.72 | 0.72 | 0.71 | 0.71 | 0.71 | 0.73 | 0.73 | 0.72 | 0.71 | 0.71 | 0.71 | 0.73 | 0.73 | 0.72 | 0.71 | 0.72 | 0.71 |
| HY | 0.79 | 0.79 | 0.78 | 0.78 | 0.81 | 0.81 | 0.79 | 0.79 | 0.78 | 0.79 | 0.81 | 0.81 | 0.80 | 0.79 | 0.79 | 0.79 | 0.81 | 0.81 |
| TR | 0.65 | 0.64 | 0.63 | 0.62 | 0.62 | 0.62 | 0.65 | 0.64 | 0.63 | 0.62 | 0.62 | 0.62 | 0.65 | 0.64 | 0.63 | 0.62 | 0.62 | 0.62 |
| OP | 0.70 | 0.69 | 0.69 | 0.67 | 0.67 | 0.66 | 0.70 | 0.69 | 0.68 | 0.67 | 0.66 | 0.66 | 0.70 | 0.69 | 0.69 | 0.68 | 0.67 | 0.66 |
| **Average** | **0.47** | **0.48** | **0.48** | **0.47** | **0.48** | **0.49** | **0.47** | **0.48** | **0.48** | **0.48** | **0.48** | **0.49** | **0.47** | **0.48** | **0.48** | **0.47** | **0.48** | **0.49** |
| **Overall mean** | **0.38** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.38** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.38** | **0.39** | **0.39** | **0.39** | **0.40** | **0.39** |

[1] DGV validated accuracy means Pearson's correlation between DGVs and EBVs in the prediction population;

[2] SCE1 scenario that the number of animals and the percentage of animals with imputed genotypes in the training population varied;

[3] WGBW: Weight gain from birth to weaning (kg); WGWY: Weight gain from weaning to yearling (kg); CW: Conformation score at weaning (scores 1-5); CY: Conformation score at yearling (scores 1-5); PW: Precocity score at weaning (scores 1-5); PY: Precocity score at yearling (scores 1-5); MW: Muscularity score at weaning (scores 1-5); MY: Muscularity score at yearling (scores 1-5); SCa: Scrotal circumference adjusted for age at yearling (cm); SCaw: Scrotal circumference adjusted for age and weight at yearling (cm); BW: Birth weight (kg); BA: Birth assistance score (scores 1-5); SW: Size score at weaning (scores 1-5); SY: Size score at yearling (scores 1-5); NW: Prepuce (navel) score at weaning (scores 1-5); NY: Prepuce (navel) score at yearling (scores 1-5); HW: Hair length score at weaning (scores 1-3); HY: Hair length score at yearling (scores 1-3); TR: Ticks resistance (ticks unit); OP: Ocular pigmentation score (scores 1-3);

[4] 8K: means that the base panel is the 8K SNP panel imputed to the 50K SNP panel; 15K: means that the base panel is the 15K SNP panel imputed to the 50K SNP panel; 50K: means the true 50K SNP panel.

**Table 4 DGV validated accuracy in the prediction population for the SCE2 and SCE3 scenarios[12]**

| Traits[3] | 8K[4] | | | | | | 15K[4] | | | | | | 50K[4] | 777K[45] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 10 | 20 | 30 | 40 | 50 | 60 | | |
| *Traits in the selection index* | | | | | | | | | | | | | | |
| WGBW | 0.33 | 0.33 | 0.32 | 0.33 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.34 |
| WGWY | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.36 |
| CW | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.28 | 0.28 | 0.28 | 0.28 | 0.29 | 0.29 | 0.29 | 0.32 |
| CY | 0.27 | 0.26 | 0.26 | 0.26 | 0.25 | 0.26 | 0.26 | 0.26 | 0.26 | 0.26 | 0.26 | 0.26 | 0.26 | 0.29 |
| PW | 0.31 | 0.31 | 0.31 | 0.31 | 0.31 | 0.32 | 0.31 | 0.30 | 0.30 | 0.31 | 0.31 | 0.31 | 0.31 | 0.32 |
| PY | 0.34 | 0.34 | 0.34 | 0.33 | 0.33 | 0.34 | 0.34 | 0.34 | 0.34 | 0.34 | 0.33 | 0.34 | 0.34 | 0.34 |
| MW | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.37 |
| MY | 0.35 | 0.35 | 0.35 | 0.34 | 0.34 | 0.35 | 0.35 | 0.35 | 0.34 | 0.34 | 0.34 | 0.35 | 0.35 | 0.36 |
| SCa | 0.18 | 0.19 | 0.18 | 0.19 | 0.19 | 0.18 | 0.18 | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 | 0.18 | 0.17 |
| SCaw | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.19 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
| **Average** | **0.30** | **0.30** | **0.30** | **0.30** | **0.29** | **0.30** | **0.29** | **0.29** | **0.29** | **0.29** | **0.29** | **0.30** | **0.30** | **0.31** |
| *Traits not in the selection index* | | | | | | | | | | | | | | |
| BW | 0.22 | 0.21 | 0.21 | 0.22 | 0.21 | 0.21 | 0.21 | 0.21 | 0.21 | 0.21 | 0.21 | 0.21 | 0.21 | 0.21 |
| BA | 0.13 | 0.14 | 0.14 | 0.14 | 0.14 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.16 | 0.14 | 0.11 |
| SW | 0.33 | 0.34 | 0.34 | 0.34 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.36 |
| SY | 0.36 | 0.36 | 0.36 | 0.35 | 0.35 | 0.35 | 0.36 | 0.36 | 0.36 | 0.35 | 0.36 | 0.36 | 0.36 | 0.37 |
| NW | 0.54 | 0.54 | 0.53 | 0.53 | 0.53 | 0.53 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.56 |
| NY | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.55 | 0.55 | 0.54 | 0.54 | 0.57 |
| HW | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 |
| HY | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 | 0.81 | 0.80 | 0.81 | 0.81 | 0.81 | 0.80 |
| TR | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 | 0.62 |
| OP | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.64 |
| **Average** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.49** | **0.50** |
| **Overall mean** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.39** | **0.40** |

[1] DGV validated accuracy means Pearson's correlation between DGVs and EBVs in the prediction population;

[2] SCE2 scenario that the percentage of animals with imputed genotypes in the training population varied and SCE3 scenario was created with only one percentage of animals with imputed genotypes and only one training population size;

[3] WGBW: Weight gain from birth to weaning (kg); WGWY: Weight gain from weaning to yearling (kg); CW: Conformation score at weaning (scores 1-5); CY: Conformation score at yearling (scores 1-5); PW: Precocity score at weaning (scores 1-5); PY: Precocity score at yearling (scores 1-5); MW: Muscularity score at weaning (scores 1-5); MY: Muscularity score at yearling (scores 1-5); SCa: Scrotal circumference adjusted for age at yearling (cm); SCaw: Scrotal circumference adjusted for age and weight at yearling (cm); BW: Birth weight (kg); BA: Birth assistance score (scores 1-5); SW: Size score at weaning (scores 1-5); SY: Size score at yearling (scores 1-5); NW: Prepuce (navel) score at weaning (scores 1-5); NY: Prepuce (navel) score at yearling (scores 1-5); HW: Hair length score at weaning (scores 1-3); HY: Hair length score at yearling (scores 1-3); TR: Ticks resistance (ticks unit); OP: Ocular pigmentation score (scores 1-3);

[4] 8K: means that the base panel is the 8K SNP panel imputed to the 50K SNP panel; 15K: means that the base panel is the 15K SNP panel imputed to the 50K SNP panel; 50K: means the true 50K SNP panel; 777K: means that the base panel is the 50K SNP panel imputed to the 777K SNP panel;

[5] When the 777K SNP panel was used, all animals had their genotypes imputed from the 50K SNP panel to the 777K SNP panel, except for 212 animals in the training population (SCE3).

**Table 5 Results of analysis of variance of the DGV validated accuracy for the SCE1 scenarios[123]**

| Economic traits | Abbr. | Panel[4] | | | Scenario[5] | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 8-15 | 8-50 | 15-50 | 10-60 | 20-60 | 30-60 | 40-60 | 50-60 |
| *Traits in the selection index* | | | | | | | | | |
| Weight gain from birth to weaning (kg) | WGBW | * (8) | ns | ns | * (60) | * (20) | * (30) | * (40) | * (50) |
| Weight gain from weaning to yearling (kg) | WGWY | ns | ns | ns | * (60) | * (60) | * (60) | * (60) | ns |
| Conformation score at weaning (scores 1-5) | CW | ns | * (50) | ns | * (60) | * (60) | Ns | * (40) | * (50) |
| Conformation score at yearling (scores 1-5) | CY | * (15) | * (50) | ns | * (10) | * (20) | * (30) | * (40) | * (50) |
| Precocity score at weaning (scores 1-5) | PW | ns | ns | ns | * (60) | * (60) | * (60) | ns | * (50) |
| Precocity score at yearling (scores 1-5) | PY | ns | ns | ns | * (60) | * (60) | * (60) | * (60) | * (50) |
| Muscularity score at weaning (scores 1-5) | MW | ns | ns | * (15) | * (60) | * (60) | * (60) | * (60) | * (50) |
| Muscularity score at yearling (scores 1-5) | MY | ns | ns | ns | ns | ns | * (60) | ns | * (50) |
| Scrotal circumference a (cm) | SCa[6] | * (15) | ns | ns | * (10) | * (20) | * (30) | * (40) | ns |
| Scrotal circumference aw (cm) | SCaw[6] | * (15) | * (50) | * (50) | * (10) | * (20) | * (30) | * (40) | * (60) |
| *Traits not in the selection index* | | | | | | | | | |
| Birth weight (kg) | BW | ns | * (8) | * (15) | * (60) | ns | * (30) | * (60) | * (60) |
| Birth assistance score (scores 1-5) | BA | * (15) | ns | * (15) | * (60) | ns | * (60) | * (60) | Ns |
| Size score at weaning (scores 1-5) | SW | * (8) | * (8) | ns | * (60) | * (60) | * (60) | * (60) | * (60) |
| Size score at yearling (scores 1-5) | SY | ns | ns | ns | * (60) | * (60) | * (60) | * (60) | * (60) |
| Prepuce (navel) score at weaning (scores 1-5) | NW | ns | ns | ns | * (60) | * (60) | * (60) | * (60) | * (60) |
| Prepuce (navel) score at yearling (scores 1-5) | NY | ns | ns | ns | * (60) | * (60) | * (60) | * (60) | * (60) |
| Hair length score at weaning (scores 1-3) | HW | * (15) | * (50) | * (50) | * (10) | * (20) | * (30) | ns | Ns |
| Hair length score at yearling (scores 1-3) | HY | * (15) | * (50) | * (50) | * (60) | * (60) | * (60) | * (60) | Ns |
| Ticks resistance (ticks unit) | TR | ns | * (50) | * (50) | * (10) | * (20) | * (30) | ns | * (60) |
| Ocular pigmentation score (scores 1-3) | OP | * (15) | * (50) | * (50) | * (10) | * (20) | * (30) | * (40) | * (50) |

[1] DGV validated accuracy means Pearson's correlation between DGVs and EBVs in the prediction population;

[2] SCE1 scenario that the number of animals and the percentage of animals with imputed genotypes in the training population varied;

[3] "*" means that there was a significant difference (P <0.05). The value between brackets indicates which panel/scenario had higher estimated accuracy. "ns" means that there was no significant difference (P> 0.05);

[4] 8,15 and 50 means the 8K, 15K and 50K SNP panel. 8-15, 8-50 and 15-50 are the contrast between the two panels;

[5] 10,20,30,40,50 and 60 means the percentage of imputed genotypes. 10-60, 20-60, 30-60, 40-60, 50-60 are the contrasts between the two percentages of the imputed genotypes;

[6] SCa is the scrotal circumference adjusted for age at yearling and SCaw is the scrotal circumference adjusted for age and weight at yearling.

**Table 6 Results of analysis of variance of the DGV validated accuracy for the SCE2 scenarios[123]**

| Economic traits | Abbr. | Panel[4] | | | Scenario[5] | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 8-15 | 8-50 | 15-50 | 10-60 | 20-60 | 30-60 | 40-60 | 50-60 |
| *Traits in the selection index* | | | | | | | | | |
| Weight gain from birth to weaning (kg) | WGBW | ns | ns | ns | ns | ns | Ns | ns | ns |
| Weight gain from weaning to yearling (kg) | WGWY | ns | * (50) | * (50) | ns | ns | Ns | ns | ns |
| Conformation score at weaning (scores 1-5) | CW | * (8) | ns | * (50) | ns | ns | Ns | ns | ns |
| Conformation score at yearling (scores 1-5) | CY | ns | ns | * (50) | ns | ns | Ns | ns | ns |
| Precocity score at weaning (scores 1-5) | PW | ns | * (50) | * (50) | ns | ns | Ns | ns | ns |
| Precocity score at yearling (scores 1-5) | PY | ns | * (50) | * (50) | ns | ns | Ns | ns | ns |
| Muscularity score at weaning (scores 1-5) | MW | ns | ns | * (50) | ns | ns | Ns | ns | ns |
| Muscularity score at yearling (scores 1-5) | MY | * (8) | ns | * (50) | ns | ns | Ns | ns | ns |
| Scrotal circumference a (cm) | SCa[6] | ns | * (8) | * (15) | ns | ns | Ns | ns | ns |
| Scrotal circumference aw (cm) | SCaw[6] | * (15) | * (50) | Ns | ns | ns | Ns | ns | ns |
| *Traits not in the selection index* | | | | | | | | | |
| Birth weight (kg) | BW | ns | ns | Ns | ns | ns | ns | ns | ns |
| Birth assistance score (scores 1-5) | BA | * (15) | ns | * (15) | ns | ns | ns | ns | ns |
| Size score at weaning (scores 1-5) | SW | * (8) | * (8) | Ns | ns | ns | ns | ns | ns |
| Size score at yearling (scores 1-5) | SY | ns | ns | Ns | ns | ns | ns | ns | ns |
| Prepuce (navel) score at weaning (scores 1-5) | NW | * (15) | ns | ns | ns | ns | ns | ns | ns |
| Prepuce (navel) score at yearling (scores 1-5) | NY | * (15) | * (50) | ns | ns | ns | ns | ns | ns |
| Hair length score at weaning (scores 1-3) | HW | * (15) | * (50) | * (50) | ns | ns | ns | ns | ns |
| Hair length score at yearling (scores 1-3) | HY | ns | ns | ns | ns | ns | ns | ns | ns |
| Ticks resistance (ticks unit) | TR | * (8) | * (8) | * (50) | ns | ns | ns | ns | ns |
| Ocular pigmentation score (scores 1-3) | OP | ns | ns | * (50) | ns | ns | ns | ns | ns |

[1] DGV validated accuracy means Pearson's correlation between DGVs and EBVs in the prediction population;

[2] SCE2 scenario that the percentage of animals with imputed genotypes in the training population varied;

[3] "*" means that there was a significant difference (P <0.05). The value between brackets indicates which panel/scenario had higher estimated accuracy. "ns" means that there was no significant difference (P> 0.05);

[4] 8,15 and 50 means the 8K, 15K and 50K SNP panel. 8-15, 8-50 and 15-50 are the contrasts between the two panels;

[5] 10,20,30,40,50 and 60 means the percentage of imputed genotypes. 10-60, 20-60, 30-60, 40-60, 50-60 is the contrast between the two percentages of the imputed genotypes;

[6] SCa is the scrotal circumference adjusted for age at yearling and SCaw is the scrotal circumference adjusted for age and weight at yearling.

**Table 7 EBV expected accuracy in the training and prediction population and GEBV and DGV expected accuracy in the prediction population in the scenario with the largest training population[1]**

| Economic trait | Abbr. | EBV training | EBV prediction | 8K[2] GEBV | DGV | 15K[2] GEBV | DGV | 50K[2] GEBV | DGV | 777K[2] GEBV | DGV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Traits in the selection index* | | | | | | | | | | | |
| Weight gain from birth to weaning (kg) | WGBW | 0.64 | 0.64 | 0.67 | 0.42 | 0.67 | 0.42 | 0.67 | 0.43 | 0.66 | 0.42 |
| Weight gain from weaning to yearling (kg) | WGWY | 0.62 | 0.60 | 0.63 | 0.39 | 0.63 | 0.40 | 0.63 | 0.40 | 0.63 | 0.39 |
| Conformation score at weaning (scores 1-5) | CW | 0.60 | 0.61 | 0.64 | 0.40 | 0.64 | 0.40 | 0.64 | 0.40 | 0.63 | 0.39 |
| Conformation score at yearling (scores 1-5) | CY | 0.62 | 0.61 | 0.64 | 0.40 | 0.64 | 0.40 | 0.65 | 0.41 | 0.64 | 0.40 |
| Precocity score at weaning (scores 1-5) | PW | 0.60 | 0.61 | 0.64 | 0.40 | 0.64 | 0.40 | 0.64 | 0.40 | 0.63 | 0.39 |
| Precocity score at yearling (scores 1-5) | PY | 0.62 | 0.61 | 0.64 | 0.40 | 0.64 | 0.40 | 0.65 | 0.41 | 0.64 | 0.40 |
| Muscularity score at weaning (scores 1-5) | MW | 0.60 | 0.61 | 0.64 | 0.40 | 0.64 | 0.40 | 0.64 | 0.40 | 0.63 | 0.39 |
| Muscularity score at yearling (scores 1-5) | MY | 0.62 | 0.61 | 0.64 | 0.40 | 0.64 | 0.40 | 0.65 | 0.41 | 0.64 | 0.40 |
| Scrotal circumference a (cm) | SCa[3] | 0.74 | 0.70 | 0.72 | 0.39 | 0.72 | 0.39 | 0.72 | 0.39 | 0.72 | 0.39 |
| Scrotal circumference aw (cm) | SCaw[3] | 0.73 | 0.70 | 0.71 | 0.38 | 0.71 | 0.39 | 0.71 | 0.39 | 0.71 | 0.38 |
| **Average** | | **0.64** | **0.63** | **0.66** | **0.40** | **0.66** | **0.40** | **0.66** | **0.40** | **0.65** | **0.40** |
| *Traits not in the selection index* | | | | | | | | | | | |
| Birth weight (kg) | BW | 0.66 | 0.65 | 0.68 | 0.44 | 0.68 | 0.44 | 0.68 | 0.44 | 0.68 | 0.43 |
| Birth assistance score (scores 1-5) | BA | 0.73 | 0.73 | 0.75 | 0.43 | 0.75 | 0.43 | 0.75 | 0.43 | 0.74 | 0.42 |
| Size score at weaning (scores 1-5) | SW | 0.60 | 0.61 | 0.63 | 0.40 | 0.64 | 0.40 | 0.64 | 0.40 | 0.63 | 0.39 |
| Size score at yearling (scores 1-5) | SY | 0.69 | 0.68 | 0.71 | 0.43 | 0.71 | 0.44 | 0.71 | 0.44 | 0.71 | 0.43 |
| Prepuce (navel) score at weaning (scores 1-5) | NW | 0.72 | 0.71 | 0.74 | 0.47 | 0.74 | 0.47 | 0.74 | 0.47 | 0.74 | 0.46 |
| Prepuce (navel) score at yearling (scores 1-5) | NY | 0.68 | 0.68 | 0.70 | 0.43 | 0.70 | 0.44 | 0.70 | 0.44 | 0.70 | 0.43 |
| Hair length score at weaning (scores 1-3) | HW | 0.58 | 0.55 | 0.58 | 0.35 | 0.58 | 0.35 | 0.58 | 0.35 | 0.58 | 0.35 |
| Hair length score at yearling (scores 1-3) | HY | 0.61 | 0.60 | 0.63 | 0.40 | 0.63 | 0.40 | 0.63 | 0.40 | 0.63 | 0.39 |
| Ticks resistance (ticks unit) | TR | 0.52 | 0.51 | 0.53 | 0.31 | 0.53 | 0.31 | 0.53 | 0.31 | 0.53 | 0.31 |
| Ocular pigmentation score (scores 1-3) | OP | 0.57 | 0.54 | 0.57 | 0.35 | 0.57 | 0.35 | 0.57 | 0.35 | 0.57 | 0.35 |
| **Average** | | **0.64** | **0.63** | **0.65** | **0.40** | **0.65** | **0.40** | **0.65** | **0.40** | **0.65** | **0.40** |
| **Overall mean** | | **0.64** | **0.63** | **0.65** | **0.40** | **0.65** | **0.40** | **0.66** | **0.40** | **0.65** | **0.39** |

[1] EBV, DGV and GEBV expected accuracy means that accuracy were obtained from the mixed model equation;

[2] 8K: means that the base panel is the 8K SNP panel imputed to the 50K SNP panel; 15K: means that the base panel is the 15K SNP panel imputed to the 50K SNP panel; 50K: means that the true 50K SNP panel; 777K: means that the base panel is the 50K SNP panel imputed to the 777K SNP panel;

[3] SCa is the scrotal circumference adjusted for age at yearling and SCaw is the scrotal circumference adjusted for age and weight at yearling.

**Table 8 Losses in GEBV expected accuracy using the 8K and 15K SNP panel imputed to the 50K SNP panel compared to the real 50K SNP panel in the SCE1 scenarios[12]**

| Traits[3] | 8K[4] | | | | | | 15K[4] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 10 | 20 | 30 | 40 | 50 | 60 |
| *Traits in the selection index* | | | | | | | | | | | | |
| WGBW | -0.0008 | -0.0009 | -0.0005 | -0.0009 | -0.0009 | -0.0010 | -0.0003 | -0.0003 | 0.0001 | -0.0003 | -0.0004 | -0.0004 |
| WGWY | -0.0010 | -0.0011 | -0.0011 | -0.0011 | -0.0012 | -0.0008 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | 0.0000 |
| CW | -0.0010 | -0.0011 | -0.0012 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 |
| CY | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 |
| PW | -0.0010 | -0.0011 | -0.0012 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 |
| PY | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 |
| MW | -0.0010 | -0.0011 | -0.0012 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 |
| MY | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0004 |
| SCa | -0.0008 | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0012 | -0.0003 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0005 |
| SCaw | -0.0008 | -0.0009 | -0.0010 | -0.0011 | -0.0011 | -0.0012 | -0.0003 | -0.0004 | -0.0004 | -0.0005 | -0.0005 | -0.0005 |
| **Average** | **-0.0009** | **-0.0010** | **-0.0010** | **-0.0011** | **-0.0011** | **-0.0011** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0004** |
| *Traits not in the selection index* | | | | | | | | | | | | |
| BW | -0.0007 | -0.0008 | -0.0008 | -0.0008 | -0.0009 | -0.0009 | -0.0003 | -0.0003 | -0.0003 | -0.0003 | -0.0003 | -0.0003 |
| BA | -0.0004 | -0.0004 | -0.0005 | -0.0006 | -0.0006 | -0.0007 | -0.0002 | -0.0002 | -0.0002 | -0.0002 | -0.0003 | -0.0003 |
| SW | -0.0011 | -0.0011 | -0.0012 | -0.0011 | -0.0011 | -0.0012 | -0.0004 | -0.0004 | -0.0005 | -0.0004 | -0.0004 | -0.0004 |
| SY | -0.0007 | -0.0008 | -0.0009 | -0.0009 | -0.0010 | -0.0010 | -0.0003 | -0.0003 | -0.0003 | -0.0004 | -0.0004 | -0.0004 |
| NW | -0.0005 | -0.0006 | -0.0006 | -0.0007 | -0.0007 | -0.0008 | -0.0002 | -0.0002 | -0.0002 | -0.0003 | -0.0003 | -0.0003 |
| NY | -0.0007 | -0.0007 | -0.0008 | -0.0008 | -0.0009 | -0.0010 | -0.0003 | -0.0003 | -0.0003 | -0.0003 | -0.0004 | -0.0004 |
| HW | -0.0014 | -0.0015 | -0.0016 | -0.0017 | -0.0018 | -0.0018 | -0.0006 | -0.0006 | -0.0007 | -0.0007 | -0.0007 | -0.0008 |
| HY | -0.0009 | -0.0010 | -0.0011 | -0.0011 | -0.0012 | -0.0013 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0005 |
| TR | -0.0016 | -0.0017 | -0.0018 | -0.0019 | -0.0020 | -0.0020 | -0.0007 | -0.0007 | -0.0007 | -0.0008 | -0.0008 | -0.0008 |
| OP | -0.0015 | -0.0016 | -0.0017 | -0.0018 | -0.0019 | -0.0019 | -0.0006 | -0.0007 | -0.0007 | -0.0008 | -0.0008 | -0.0008 |
| **Average** | **-0.0010** | **-0.0010** | **-0.0011** | **-0.0012** | **-0.0012** | **-0.0013** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0005** | **-0.0005** | **-0.0005** |
| **Overall average** | **-0.0009** | **-0.0010** | **-0.0011** | **-0.0011** | **-0.0012** | **-0.0012** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0004** |

[1] GEBV expected accuracy means that accuracy were obtained from the mixed model equation in the prediction population;

[2] SCE1 scenario that the number of animals and the percentage of animals with imputed genotypes in the training population varied;

[3] WGBW: Weight gain from birth to weaning (kg); WGWY: Weight gain from weaning to yearling (kg); CW: Conformation score at weaning (scores 1-5); CY: Conformation score at yearling (scores 1-5); PW: Precocity score at weaning (scores 1-5); PY: Precocity score at yearling (scores 1-5); MW: Muscularity score at weaning (scores 1-5); MY: Muscularity score at yearling (scores 1-5); SCa: Scrotal circumference adjusted for age at yearling (cm); SCaw: Scrotal circumference adjusted for age and weight at yearling (cm); BW: Birth weight (kg); BA: Birth assistance score (scores 1-5); SW: Size score at weaning (scores 1-5); SY: Size score at yearling (scores 1-5); NW: Prepuce (navel) score at weaning (scores 1-5); NY: Prepuce (navel) score at yearling (scores 1-5); HW: Hair length score at weaning (scores 1-3); HY: Hair length score at yearling (scores 1-3); TR: Ticks resistance (ticks unit); OP: Ocular pigmentation score (scores 1-3);

[4] 8K: means that the base panel is the 8K SNP panel imputed to the 50K SNP panel; 15K: means that the base panel is the 15K SNP panel imputed to the 50K SNP panel and 10, 20, 30, 40, 50 and 60 means the percentage of animals with imputed genotypes.

**Table 9 Losses in GEBV expected accuracy using the 8K and 15K SNP panel imputed to the 50K SNP panel in the SCE2 scenarios and the 777K SNP panel imputed from the 50K SNP panel in the SCE3 scenario compared to the real 50K SNP panel[12]**

| Traits[3] | 8K[4] | | | | | | 15K[4] | | | | | | 777K[4] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 10 | 20 | 30 | 40 | 50 | 60 | |
| WGBW | -0.0001 | -0.0003 | -0.0004 | -0.0004 | -0.0006 | -0.0006 | 0.0002 | 0.0001 | 0.0001 | 0.0001 | 0.0000 | 0.0000 | -0.0020 |
| WGWY | -0.0002 | -0.0004 | -0.0005 | -0.0006 | -0.0008 | -0.0009 | 0.0002 | 0.0001 | 0.0001 | 0.0001 | 0.0000 | -0.0001 | -0.0020 |
| CW | -0.0008 | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0012 | -0.0003 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0022 |
| CY | -0.0007 | -0.0009 | -0.0009 | -0.0010 | -0.0012 | -0.0012 | -0.0002 | -0.0003 | -0.0004 | -0.0004 | -0.0005 | -0.0005 | -0.0026 |
| PW | -0.0008 | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0012 | -0.0003 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0023 |
| PY | -0.0007 | -0.0009 | -0.0009 | -0.0010 | -0.0012 | -0.0012 | -0.0002 | -0.0003 | -0.0004 | -0.0004 | -0.0005 | -0.0005 | -0.0026 |
| MW | -0.0008 | -0.0009 | -0.0010 | -0.0010 | -0.0011 | -0.0012 | -0.0003 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0023 |
| MY | -0.0007 | -0.0009 | -0.0009 | -0.0010 | -0.0012 | -0.0012 | -0.0002 | -0.0003 | -0.0004 | -0.0004 | -0.0005 | -0.0005 | -0.0026 |
| SCa | -0.0005 | -0.0006 | -0.0007 | -0.0008 | -0.0009 | -0.0011 | -0.0002 | -0.0002 | -0.0003 | -0.0003 | -0.0004 | -0.0005 | -0.0013 |
| SCaw | -0.0006 | -0.0006 | -0.0007 | -0.0008 | -0.0009 | -0.0011 | -0.0002 | -0.0002 | -0.0003 | -0.0003 | -0.0004 | -0.0005 | -0.0013 |
| **Average** | **-0.0006** | **-0.0007** | **-0.0008** | **-0.0009** | **-0.0010** | **-0.0011** | **-0.0002** | **-0.0002** | **-0.0003** | **-0.0003** | **-0.0003** | **-0.0004** | **-0.0021** |
| BW | -0.0005 | -0.0007 | -0.0007 | -0.0008 | -0.0009 | -0.0010 | -0.0002 | -0.0003 | -0.0003 | -0.0003 | -0.0004 | -0.0004 | -0.0021 |
| BA | -0.0004 | -0.0003 | -0.0004 | -0.0004 | -0.0005 | -0.0006 | -0.0002 | -0.0001 | -0.0002 | -0.0002 | -0.0002 | -0.0003 | -0.0020 |
| SW | -0.0008 | -0.0010 | -0.0010 | -0.0010 | -0.0012 | -0.0012 | -0.0003 | -0.0004 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0022 |
| SY | -0.0005 | -0.0006 | -0.0007 | -0.0008 | -0.0009 | -0.0011 | -0.0002 | -0.0002 | -0.0003 | -0.0003 | -0.0004 | -0.0005 | -0.0024 |
| NW | -0.0004 | -0.0006 | -0.0006 | -0.0007 | -0.0008 | -0.0008 | -0.0002 | -0.0002 | -0.0002 | -0.0003 | -0.0003 | -0.0003 | -0.0024 |
| NY | -0.0005 | -0.0007 | -0.0007 | -0.0008 | -0.0010 | -0.0010 | -0.0002 | -0.0003 | -0.0003 | -0.0003 | -0.0004 | -0.0004 | -0.0025 |
| HW | -0.0015 | -0.0015 | -0.0016 | -0.0016 | -0.0018 | -0.0020 | -0.0006 | -0.0006 | -0.0007 | -0.0007 | -0.0008 | -0.0009 | -0.0014 |
| HY | -0.0007 | -0.0009 | -0.0010 | -0.0010 | -0.0012 | -0.0013 | -0.0003 | -0.0004 | -0.0004 | -0.0004 | -0.0005 | -0.0005 | -0.0027 |
| TR | -0.0016 | -0.0016 | -0.0017 | -0.0017 | -0.0018 | -0.0020 | -0.0006 | -0.0007 | -0.0007 | -0.0007 | -0.0008 | -0.0009 | -0.0016 |
| OP | -0.0017 | -0.0017 | -0.0019 | -0.0019 | -0.0020 | -0.0022 | -0.0007 | -0.0007 | -0.0008 | -0.0008 | -0.0008 | -0.0010 | -0.0011 |
| **Average** | **-0.0009** | **-0.0010** | **-0.0010** | **-0.0011** | **-0.0012** | **-0.0013** | **-0.0003** | **-0.0004** | **-0.0004** | **-0.0004** | **-0.0005** | **-0.0006** | **-0.0021** |
| **Overall average** | **-0.0007** | **-0.0009** | **-0.0009** | **-0.0010** | **-0.0011** | **-0.0012** | **-0.0002** | **-0.0003** | **-0.0003** | **-0.0004** | **-0.0004** | **-0.0005** | **-0.0021** |

[1] GEBV expected accuracy means that accuracy were obtained from the mixed model equation in the prediction population;

[2] SCE2 scenario that the percentage of animals with imputed genotypes in the training population varied and SCE3 scenario was created with only one percentage of animals with imputed genotypes and only one training population size;

[3] WGBW: Weight gain from birth to weaning (kg); WGWY: Weight gain from weaning to yearling (kg); CW: Conformation score at weaning (scores 1-5); CY: Conformation score at yearling (scores 1-5); PW: Precocity score at weaning (scores 1-5); PY: Precocity score at yearling (scores 1-5); MW: Muscularity score at weaning (scores 1-5); MY: Muscularity score at yearling (scores 1-5); SCa: Scrotal circumference adjusted for age at yearling (cm); SCaw: Scrotal circumference adjusted for age and weight at yearling (cm); BW: Birth weight (kg); BA: Birth assistance score (scores 1-5); SW: Size score at weaning (scores 1-5); SY: Size score at yearling (scores 1-5); NW: Prepuce (navel) score at weaning (scores 1-5); NY: Prepuce (navel) score at yearling (scores 1-5); HW: Hair length score at weaning (scores 1-3); HY: Hair length score at yearling (scores 1-3); TR: Ticks resistance (ticks unit); OP: Ocular pigmentation score (scores 1-3);

[4] 8K: means that the base panel is the 8K SNP panel imputed to the 50K SNP panel; 15K: means that the base panel is the 15K SNP panel imputed to the 50K SNP panel; 777K: means that the base panel is the 50K SNP panel imputed to the 777K SNP panel and 10, 20, 30, 40, 50 and 60 means the percentage of animals with imputed genotypes.

[8]**CAPÍTULO V**

---

[8] Artigo redigido de acordo com as normas do periódico *BMC Genetics*

# Genomic evaluation using single and two steps procedures in the presence of selection in simulated data of beef cattle

**Mario L Piccoli[1,2,3]\*, José Braccini[16], Fernanda V Brito[2], Fernando F Cardoso[5,6], Jaime A Cobuci[1], Luiz F Brito[3], Medhi Sargolzaei[3,4], Flávio S Schenkel[3]**

[1] Universidade Federal do Rio Grande do Sul, Departamento de Zootecnia, Porto Alegre, Brazil.

[2] GenSys Consultores Associados S/S, Porto Alegre, Brazil.

[3] Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada.

[4] The Semex Alliance, Guelph, Canada.

[5] Embrapa Southern Region Animal Husbandry, Bagé, Brazil.

[6] National Council for Scientific and Technological Development, Brasília, Brazil

\* Corresponding author: Mario L Piccoli (mariopiccoli@gensys.com.br)

Email addresses:

MLP: mariopiccoli@gensys.com.br

JB: jose.braccini@ufrgs.br

FVB: fernandabrito@gensys.com.br

FFC: fernando.cardoso@embrapa.br

JAC: jaime.cobuci@ufrgs.br

LFB: lbrito@uoguelph.ca

MS: msargol@uoguelph.ca

FSS: schenkel@uoguelph.ca

## Abstract

**Background:** Some studies have demonstrated greater genetic gains in breeding programs through incorporation of genomic predictions of genetic merit compared to traditional genetic evaluations. Single and two steps procedures were developed to analyze pedigree and phenotype jointly with molecular markers. A single trait with heritability of 0.25 and with 40K SNPs (single nucleotide polymorphism) and 750 QTLs (quantitative trait loci) across the 29 *Bos taurus* autosomes was simulated based on forward-in-time process, using QMSim. The training population was composed by 1,920 sires under selection from generation three to eight with more than 50 offspring each one. The prediction population was composed by 3,060 individuals randomly selected from the $10^{th}$ generation. Direct genomic value (DGV) and genomic estimated breeding value (GEBV) were estimated in the training population based on GEBV package (two-steps) and BLUPF90 package (single-step). The aim of this study was to compare the single and two steps procedures using beef cattle simulated data.

**Results:** There were no statistical differences ($P>0.05$) in DGVs accuracies between single and two steps procedures. There were no statistical differences ($P>0.05$) in GEBVs accuracies between single and two steps procedures. GEBV accuracy by Hayes method in two-steps procedure was lower ($P<0.05$) when used PP2 prediction population (including phenotypes of genotyped animals). DGVs regression coefficients ranged from 0.837 to 0.954 and from 0.957 to 1.106 using single and two steps procedure, respectively, and GEBVs regression coefficients ranged from 0.937 to 1.032 and from 0.998 to 1.348 using single and two steps procedures, respectively. The maximums gains in GEBVs accuracy, using default polygenic effect in the genomic relationship matrix, were 0.364 and 0.341 by single and two steps, respectively, compared to the EBVs parent's average (EBVpa). In general, when used different levels of polygenic effect in the genomic relationship matrix, there were no statistical differences ($P>0.05$) in DGVs and GEBVs accuracies by single and two steps procedures. Using DGV or GEBV to select 10% of males and 50% of females, there were between 36% and 49% and between 68% and 75% of coincidence regarding the use of TBV in males and females, respectively.

**Conclusion:** DGVs and GEBVs predicted by single and two steps produced the same level of accuracies, except for the GEBVs by Hayes blend method in the two-steps procedure. DGVs regression coefficients were equal to 1.00 ($P>0.05$) when using two-steps procedure with dEBV as pseudo phenotype and were different to 1.00 ($P<0.05$) in single-step e two-steps with EBV as pseudo phenotype. GEBVs regression coefficients were equal to 1.00 ($P>0.05$) using PP1 prediction population (single-step) and were equal to 1.00 ($P>0.05$) using PP2 prediction population (single-step and two-steps with VanRaden blend method). The polygenic effect in the genomic relationship matrix did not affect the DGV and GEBV accuracies in single and two steps procedures. Genetic evaluation using pedigree, phenotypes and genotypes information resulted in gains of accuracy greater than 100% compared to the EBVpa. There were no difference between the selected animals (10% males and 50% females) using single and two steps**.**

## Background

The incorporation of single nucleotide polymorphism (SNP) markers in genetic evaluations in order to obtain more accurate predictions and in earlier stages of the

animal production have been more usual in breeding programs, mainly in dairy cattle [1], [2], [3], [4]. SNP panels from low to high density have been used in genomic predictions in various species and specifically for cattle it is possible to genotype animals from 3K to 777K SNP panels. The use of molecular markers information will lead to greater genetic gains in shorter periods of time. Some results in genomic selection based on the estimated effects of SNP markers in general, have demonstrated greater genetic gains in breeding programs through incorporation of genomic predictions of genetic merit compared to traditional genetic evaluations. These gains would be associated with shorter generation intervals, increasing in selection intensity and accuracy of selection [5], [6]. Economically, the use of molecular markers information resulting in genetics predictions more accurate and earlier in the life of animals would allow substantial savings in conducting breeds progeny testing [7], [1].

Single and two steps procedures have been developed to analyze the pedigree and the phenotype together with the information of SNP markers [8], [9]. The two-steps means that estimated breeding values (EBVs) are obtained by conventional analyzes (based on pedigree and phenotype) and direct genomic value (DGV) are obtained based on prediction equations (through phenotypes and genotypes information) and after both are combined by different indexes generating the genomic estimated breeding values (GEBVs). The single-step means that the information of the SNPs are combined with the phenotype and pedigree for the simultaneous estimation of breeding values.

The first tests to combine the use of genomics data with the EBVs were based on two-steps procedure, where DGVs and EBVs were combined by different indexes weighted by the accuracy and heritability. The use of the two sources of information, DGVs and EBVs, is important because if the effect of the quantitative trait loci (QTL) is not captured by a SNP, this may be captured by the polygenic breeding value [10], [2], [1]. The disadvantages of this method according to Legarra et al. [11] and Misztal et al. [9] have been associated with the estimation process in more than one step and the need for pre-estimated parameters. However, it indicated the advantage that there was no change in the statistical model in routine evaluations.

The advantages of single-step in relation to the two-steps according to Vitezica et al. [12] and Christensen and Lund [13] were: a) simplicity, in other words, no need for the multiple steps and therefore, less errors can be occured during the estimation process; b) slightly higher computational time compared to the traditional estimation of the EBVs but lower when compared to the two-steps procedure and, c) generalization to other models and species. While Aguilar et al. [6] mentioned that the single-step procedure provided a unified structure, eliminating various assumptions and allowed to calculate more accurately genomic evaluations. However, the single-step procedure that uses the traditional relationship matrix enlarged with the relationship markers information have computational difficulty in obtaining this matrix and its inverse [6], [13].

Previous results presented by Vitezica et al. [12], Aguilar et al. [6], Garrick [14] and Chen et al. [15] showed advantages and disadvantages when using both procedures. More recently, Legarra et al. [16] have related that in dairy sheep [17], dairy goats [18], pigs and chickens [19], in general, the results have shown an advantage when using the single-step compared to the two-steps procedure. In dairy cattle, the genomic evaluations have been conducted through the two-steps procedure [1], [2], [3], [4] and recently Koivula et al. [20], Harris et al. [21], Pribyl et al. [22] have been evaluating the single-step procedure with equal or better results in comparison to the two-steps procedure.

In beef cattle there were not found studies relating the application of single-step procedure in genomic evaluation [16]. Furthermore, beef data sets are more complex due to more missing information in the pedigree, smaller sib ships, and the presence of maternal effects [16].

The aim of this study was to compare the single and two steps BLUP procedures using simulated data of beef cattle.

## Methods

### Simulation

Studies with simulated data can be efficient when there is a need to compare different methodologies due to the possibility to simulate the true breeding values (TBVs). In the simulations it was used parameters based on the study populations in order to mimic real data. The simulated data used in this study mimicked the extent of linkage disequilibrium in beef cattle and was part of the simulated data used by Brito et al. [23]. More detailed information about the dataset can be found in the original paper of the mentioned authors.

### Population structure

The populations were simulated based on forward-in-time process, using the QMSim software [24] with 40K SNPs markers and, 750 QTLs across the 29 *Bos taurus* autosomes (BTA). Firstly, 1,000 generations with a constant size of 1,000 animals were simulated, followed by 1,020 generations with a gradual decrease in population size from 1,000 to 200 in order to create initial linkage disequilibrium and to establish mutation-drift equilibrium in historical generations. The number of males and females remained constant and the mating system was based on random union of gametes, randomly sampled. In the second step, an expansion of the population was created by initially randomly selecting 100 founder males and 100 founder females from the last generation of the historical population. In the third step, in order to enlarge the population, eight generations were simulated with five offspring per dam. The mating was based on the random union of gametes and no selection. In the fourth step, the two recent generation sets were simulated from the last generation by selecting 640 males and 32,000 females, one male to 50 females. The parameters used in the recent generations mimicked more closely to a real production system with one progeny per dam per year, 50% of male progeny, selection for high values of EBV and culling for low values of EBV with a replacement rate of 60% for sires and 20% for dams. Sires and dams were randomly mated.

The whole process was used to generate ten different populations in order to obtain ten replications.

### Genome

The simulated genome consisted of 29 pairs of autosomes with length identical to the real bovine genome based on Btau_3.1 assembling [25] totaling 2,333 cM. The SNP markers were evenly distributed such that it would generate one density of segregating bi-allelic loci with minor allele frequency (MAF)>0.1. The markers were neutral in their effect on the trait. A number of QTL was simulated to generate 750 segregating loci with two, three or four alleles and MAF>0.1, whose positions were randomly distributed. Additive allelic effects were randomly sampled from gamma distribution with shape parameter equal to 0.4. The rate of missing marker genotypes

was 0.01 and the rate of marker genotyping error was 0.005. A recurrent mutation rate of $10^{-5}$ for both markers and QTLs was considered to establish mutation-drift equilibrium in historical generations. The same mutation rate was also applied in all subsequent generations after the historical ones.

**Simulated trait and genetic values**

A single trait with heritability of 0.25 and phenotypic variance of 1.0 was simulated. The EBVs were predicted by BLUPF90 package [26] for an individual animal model, considering the true additive genetic variance. The rate of missing sire and dam information was 5%. The TBV of an individual was equal to the sum of the QTL additive effects. The phenotypes were generated by adding random residuals to the TBV.

**Training and prediction population for using in the two-steps procedure**

The training population was composed of 1,920 sires under selection from generation three to eight and each on had 50 or more offspring. The prediction population was composed by 3,060 individuals randomly chosen to the $10^{th}$ generation with parents born until the $8^{th}$ generation.

Three EBVs sets were generated using the BLUPF90 package [26]. The first two sets were formed by the prediction population with all animals born in the $10^{th}$ generation not including the phenotypes of genotyped animals (PP1) and including the phenotypes of genotyped animals (PP2), while the third set was formed by the training population with all animals born until $8^{th}$ generation (TP).

The EBVs and deregressed EBVs (dEBV) of the training population were used as pseudo phenotypes to estimate markers effects. The approach of VanRaden and Wiggans [27] was used to calculate dEBVs free of parent average effects from the EBVs and reliabilities of genotyped animals and their sires and dams.

The DGVs were estimated in the training population based on the GEBV package [8] in GBLUP model considering 5%, 10%, 15% and 20% (default) for the polygenic effect in the genomic relationship matrix [10]. It can be described as:

$$y = 1_n \mu + Z_g + e,$$

where **y** is the vector of EBVs or dEBVs for the trait, **μ** is the overall mean, **1$_n$** is a vector of ones, **Z** is the design matrix that relates records to breeding values, **g** is the vector of DGV to be predicted, and **e** is the vector of residual effects. It was assumed that $g \sim N(0, G^* \sigma^2 g)$ where $\sigma^2 g$ is the additive genetic variance and **G\*** is a combined relationship matrix, and $e \sim N(0, R\sigma^2 e)$ where $\sigma^2 e$ is the residual variance and **R** is a diagonal matrix whose elements account for the differences in reliabilities of the EBVs or dEBVs.

In the two-steps procedure were used two ways to combine DGVs with EBVs into genomic estimated breeding values (GEBVs).

The first way used the approach of Hayes et al. [1] where:

$$GEBV = \frac{r^2_{DGV} * DGV + r^2_{EBV} * EBV}{r^2_{DGV} + r^2_{EBV}} \text{ , where,}$$

$r^2_{DGV}$ and $r^2_{EBV}$ are the reliabilities of DGV and EBV, respectively.

The second way used the approach of VanRaden et al. [2] where:

$$GEBV = b_1 * DGV + b_2 * EBV_1 + b_3 * EBV \text{ , where,}$$

EBV$_1$ was predicted for the subset of genotyped animals using traditional relationships and their dEBV, excluding data from ungenotyped animals, and b$_1$, b$_2$ and b$_3$ are weights based on reliabilities of DGV, EBV$_1$ and EBV.

**Population for using in the single-step procedure**

The pedigree information was used until the 10$^{th}$ generation. The phenotype information was used until the 8$^{th}$ generation in one analysis (PP1) and in another was added phenotypes of genotyped animals of the prediction population (PP2). The genotypes included in the analysis contained 1,920 sires with more than 50 offspring (training population) plus 3,060 genotyped animals of the prediction population.

The DGVs and GEBVs were estimated based on the BLUPF90 package [26] in GBLUP model considering the inclusion of different weights to create $[\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}]$ ($\mathbf{A}_{22}^{-1}$ is the polygenic effect in the genomic relationship matrix with 5% (default), 10%, 15% and 20%). It can be described as:

$$\mathbf{y} = \mathbf{1}_n \boldsymbol{\mu} + \mathbf{Z}_g + \mathbf{e},$$

where $\mathbf{y}$ is the vector of phenotypes for the trait, $\boldsymbol{\mu}$ is the overall mean, $\mathbf{1_n}$ is a vector of ones, $\mathbf{Z}$ is the design matrix that relates records to breeding values, $\mathbf{g}$ is the vector of animals to be predicted, and $\mathbf{e}$ is the vector of residual effects. It was assumed that $\mathbf{g} \sim \mathrm{N}$ (0, $\mathbf{H}\sigma^2 g$) where $\sigma^2 g$ is the additive genetic variance and $\mathbf{H}$ is a combined relationship matrix, and $\mathbf{e} \sim \mathrm{N}$ (0, $\mathbf{R}\sigma^2 e$) where $\sigma^2 e$ is the residual variance and $\mathbf{R}$ is a diagonal matrix whose elements account for the differences in reliabilities of the observations in $\mathbf{y}$.

In the single-step procedure GEBVs were obtained by directly combining phenotypic, genomic and pedigree information [9], [6] where the traditional relationship matrix ($\mathbf{A}$) was replaced by a matrix that includes the genomic information ($\mathbf{H}$).

**Comparison between single and two steps procedures**

The average accuracy of ten repetitions, measured by Pearson's correlation between DGV and GEBV with TBV in the prediction population, were used as response in the analysis of variance carried out with the ANOVA procedure of SAS version 9.2 (SAS Inst. Inc., Cary, NC).

**Results**

There were no significant differences (P>0.05) in DGVs accuracies by single-step and two-steps (using EBVs or dEBVs) procedures within the same level of polygenic effect in the genomic relationship matrix considered. The levels of polygenic effect in the genomic relationship matrix were not significant (P>0.05) to the DGV accuracy, except to the DGV by single-step procedure in the levels 15% and 20%. When single and two steps procedures considering default polygenic effect in the genomic relationship matrix were used, the DGVs accuracies were 0.584 and 0.564, respectively (Table 1).

There were no significant differences (P>0.05) in GEBVs accuracies by single and two steps procedures within the same level of polygenic effect in the genomic relationship matrix using PP1 and PP2 prediction population. However, there were significant differences (P<0.05) in GEBVs accuracies between VanRaden and Hayes blending methods in two-steps procedure using PP1 and PP2 prediction population. The levels of polygenic effect in the genomic relationship matrix were not significant (P>0.05) to the GEBVs accuracy, except to the GEBVs by Hayes blend method in two-

steps procedure in the levels 15% and 20% using PP1 prediction population (Tables 2 and 3). When single and two steps procedures considering default polygenic effect in the genomic relationship matrix were used, the GEBVs accuracies with PP1 prediction population were 0.589 and from 0.542 to 0.604, respectively (Table 2) and with PP2 prediction population were 0.699 and from 0.639 to 0.676, respectively (Table 3).

The EBV parent's average accuracy (EBVpa) was 0.335 and the EBV accuracy (EBVphe) was 0.534 showing an increase of 0.199 (59.4%) in accuracy by phenotypes addition. When were added genotypes in the single-step procedure, the ssGEBVpa and ssGEBVphe accuracies (using default polygenic effect in the genomic relationship matrix) were 0.589 and 0.699, respectively. It showed an increase of 0.254 (75.8%) and 0.055 (10.3%), and 0.364 (108.7%) and 0.165 (30.9%) compared to the EBVpa and the EBVphe accuracies, respectively (Table 4). Adding genotypes in two-steps procedure with VanRaden blending method and default polygenic effect in the genomic relationship matrix, the tsGEBVv_pa and tsGEBVv_phe accuracies were 0.604 and 0.676, respectively, showing an increase of 0.269 (80.3%) and 0.07 (13.1%), and 0.341 (101.8%) and 0.142 (26.6%) in comparison to the EBVpa and the EBVphe, respectively. Using the Hayes blending method there were decreases of 0.06 and 0.04 with VanRaden blending method, (Table 5).

Slope coefficient for the regression was expected to be close to 1.00, which would indicate that DGV or GEBV predictions were not inflated or deflated (Tables 6, 7 and 8). Analysis with two-steps procedure using dEBV as pseudo phenotypes in the SNPs estimation presented the slope of the regression on DGV close to 1.00 (P>0.05), however using the single-step procedure the slope of the regression on DGVs were different to 1.00 (P>0.05), showed that DGVs prediction were inflated. Slope coefficient for the regression of GEBV using two-steps procedure, the results not showed deflation or inflation with PP2 predicted population and VanRaden blend method. However with Hayes blend method, the results showed deflation with PP1 and PP2 prediction population. Using single-step, the results not showed deflation or inflation with PP1 and PP2 predicted population.

Assuming selection of 10% in males and 50% in females the exchange of animal percentage were analysed when changed the selection criteria of TBV to DGV and GEBV estimated by single and two steps. Tables 9, 10 and 11 showed the percentage of coincidence between DGV or GEBV (PP1 and PP2 prediction population) with TBV were statistically equal (P>0.05) when used single and two steps procedure. Using DGV, the percentage of coincidence was ~40% in males and ~70% in females, while that using GEBV with PP1 prediction population was ~42% in males and ~70% in females and with PP2 prediction population was ~48% in males and ~73% in females.

## Discussion

### Using EBV or dEBV in two-steps procedure

Different pseudo phenotypes of the animal, such as DYD, dEBV and EBV have been used for estimating the effect of molecular markers. Some studies have used the daughter yield deviations (DYD) [17], [28], [29] and other studies have used dEBV or EBV [23], [30], [31]. These different pseudo phenotypes have been used since TBV of animals are unknown and the goal is a better approximation of the TBV. The average reliability for the data set was 85% (sires in training population had between 50 and 250 offspring) and with this level of reliability, the results showed that there is no need to deregress the EBVs because there were no statistical differences in accuracy between

the estimation of the markers effects (DGVs) using dEBV or EBV. Neves et al. [31] studied simulated data of beef cattle, did not found differences in accuracies when the pseudo phenotypes was the dEBV, the EBV or the DYD and Guo et al. [28] also using simulated data did not found differences in accuracies when the pseudo phenotypes was the EBVs or DYDs.

The scenario in which the whole training population showed high reliability generally does not occur in production systems of commercial beef cattle. Boddhireddy et al. [30] studied Angus cattle in USA showed that the use of the dEBV compared to the EBV doubled the accuracy values. Furthermore, Ostersen et al. [29] studying pigs in Denmark showed that the use of dEBV in relation to the EBV produced from 18% to 39% higher reliabilities.

**Polygenic effect in the genomic relationship matrix**

Genomic selection using medium to high density marker panels does not cover the whole genome yet [32] and therefore part of the markers effects will be explained by polygenic effect. In this sense was analyzed the accuracy based on different levels of polygenic effect in the genomic relationship matrix, pondered 5%, 10%, 15% and 20%. Regarding to the two-steps procedure, the accuracy were always higher when the weight was 95% for the markers effect, but the differences were not statistically significant. In the single-step procedure the values were also higher when using 95% of the markers effect, but the differences were statistically significant when using a lot of polygenic effect. Onogi et al. [33] studied carcass traits in black Japanese cattle using single-step procedure with different weightings to the $[\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}]$ (20%<=$\mathbf{G}^{-1}$<=100%) and obtained better results in terms of accuracy when the fraction of $\mathbf{G}^{-1}$ was larger. However, Neves et al. [31] studied fifteen traits in Nellore in Brazil and showed higher accuracies when two-steps procedure considered 20% of polygenic effects in the genomic relationship matrix instead of not considering it. The same trend was observed in the study of Gao et al. [34] with sixteen traits in Nordic Holstein population. Calus and Veerkamp [35] studied simulated data including traits of different heritabilities in GBLUP model with the aim of evaluate the effect of including or not the polygenic effect analyzed by the accuracy of prediction. The authors concluded that the inclusion of polygenic effect in the model increased the accuracy of DGV. Liu et al. [36] studied German Holstein cattle and showed that adjusting for the polygenic effect reduced GEBV bias and concluded that weighting for polygenic effect seems to differ between the traits.

**Accuracy**

One of the main reasons to use the information of molecular markers in the prediction of breeding values is to be able to carry out the prediction to very young ages, even before obtaining phenotypes. The DGVs accuracies presented in this study (0.584 for single-step and 0.564 for two-steps) showed that the use of genetic values obtained only including the markers produced gains in accuracies of 74% and 68% compared to the EBVs parent's average for single and two steps procedures (polygenic effect default), respectively. It means that when selecting animals based only in the markers effects the accuracy increased from 0.229 to 0.249 for the trait with heritability equal to 0.25 showing that the selection using the molecular markers is better than selection by the EBV parent's average These accuracies were higher than the accuracies reported by Neves et al. [37] for traits with similar heritability in Nellore data in Brazil and lower than those reported by Boddhireddy et al. [30] using Angus cattle data in

USA. DGV accuracy, in this study, by single and two steps procedures were statistically equal, agreement with the results by Vitezica et al. [12] studying simulated data in the presence of selection. However, Koivula et al. [20] studying Nordic Red breed, and Baloche et al. [17] studying Lacaune dairy sheep breed in France, showed reliability gains for the single-step compared to the two-steps procedure. Pribyl et al. [22], studying Holstein cattle breed from Czech Republic, concluded that single-step procedure should not cause a big increase in accuracy in comparison to the traditional EBVs.

The GEBV from single and two steps (VanRaden blending method) procedures were statistically equal and showed higher accuracies in relation to the Hayes blending method in two-steps. Cardoso et al. [38] studied tick resistance in Hereford and Braford cattle in Brazil and found differences in GEBV accuracy between the two blending methods in two-steps procedure, however favorable to VanRaden method and the GEBV accuracy by single-step was superior in comparison to the two-steps procedure. In the same direction, Su et al. [4] studied Nordic Red cattle found a difference in GEBV accuracy between single and two steps (VanRaden blending method) in GBLUP model.

In beef cattle raised on pasture, which is common in tropical countries, the first culling of animals is done at weaning. Until this time the calves still suckling and therefore there is no need to culling animals. Thus it is possible to obtain various phenotypes for using in the EBV prediction such as weight gain between birth and weaning. Therefore, in this situation it will be possible to use information from parents and phenotypes to predict the animals EBVs. If an animal was genotyped, this information would also be used to predict the EBVs. In this study this situation was analysed. Firstly, there were gains in accuracy of 59% using only the phenotypes jointly with the pedigree information. The accuracy increased from 0.335 to 0.534. Based on the results of this study it was evident that the use the DGVs for selection of animals have produced equal or superior accuracy in comparison to the EBVs based on pedigree and phenotype information.

When carrying out the genotyping of animals, it is recommended to collect all possible phenotypes in order to re-estimate the markers effects and to expand the database of genotyped animals. In this case, when the information used for animal selection was the pedigree, phenotype and genotype, the increase in accuracy compared to the EBVs parent's average was up to 100%. These results were in agreement with the results obtained by VanRaden et al. [2] and Schenkel et al. [39] when studying North American Holstein using two-steps procedure, where it was observed gains greater than 100% in reliability in the category of young animals.

**Scale of DGV and GEBV**

The scale of genomic predictions should be a matter of concern, especially to determine whether DGV and GEBV can be compared to traditional EBV from routine evaluations in breeding programs. If the regression coefficient was smaller than 1.0 it would be indicating that there was overestimation of genomic predictions and if the regression coefficient was larger than 1.0 it would indicates that there was underestimation of genomic predictions. Vitezica et al. [41] have discussed that criterion under the aspect of selection. If the parents of the next generation come from only genotyped selection candidates, they share a common mean for belonging to the same generation, then the bias would not be a concern. However, if for different candidates there are different amounts of information such as progeny test males and

newborn animals and in presence of bias (genetic gains over or under estimated), thus, newborns could be considered better than they really are. Regarding to the DGV, the regression coefficients of this study showed that the two-step procedure generated estimates not biased when using dEBV as pseudo phenotypes in the SNPs estimation and regarding to the GEBV, in general, the regression coefficients showed that using single and two steps procedure generated estimates not biased, except when using Hayes blend method in two-steps procedure. The regression coefficients obtained in this study were similar to the regression coefficients obtained by Su et al. [25], studying Nordic Red breed with single and two steps procedure and by Gao et al. [40], studying Nordic Holstein breed with single and two steps.

**Selection of 10% in males and 50% in females**

Breeding programs use genetic predictions to select animals will remain in herds in order to produce the next crop of calves. Therefore, it was simulated a breeding programs with 10% selection of males (exchange 100% of sires for each year) and 50% selection of females (empty cows discarded). The results showed that the percentage of coincidences in selected males (~40%) and females (~70%) were equal using DGV or GEBV estimated by single and two steps procedure. These results reinforces the results obtained in terms of accuracy, showing that use of genomic predictions from single and two steps procedures generated almost the same list of selected males and females.

**Conclusions**

DGVs and GEBVs predicted by single and two steps produced the same level of accuracies, except for the GEBVs by Hayes blend method in the two-steps procedure. DGVs regression coefficients were equal to 1.00 (P>0.05) when using two-steps procedure with dEBV as pseudo phenotype and were different to 1.00 (P<0.05) in single-step e two-steps with EBV as pseudo phenotype. GEBVs regression coefficients were equal to 1.00 (P>0.05) using PP1 prediction population (single-step) and were equal to 1.00 (P>0.05) using PP2 prediction population (single-step and two-steps with VanRaden blend method). The polygenic effect in the genomic relationship matrix did not affect the DGV and GEBV accuracies in single and two steps procedures. Genetic evaluation using pedigree, phenotypes and genotypes information resulted in gains of accuracy greater than 100% compared to the EBVpa. There were no difference between the selected animals (10% males and 50% females) using single and two steps**.**

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
MLP participated in the design of the study, carried out the analyses, was involved in the discussions, prepared and drafted the manuscript. JB participated in the design of the study, was involved in the discussions and helped to draft the manuscript. FVB carried out the simulated data used in this study, was involved in the discussions and helped to draft the manuscript. FFC participated in the design of the study, was involved in the discussions and helped to draft the manuscript. JAC was involved in the discussions and helped to draft the manuscript. LFB was involved in the discussions and helped to draft the manuscript.MS developed the QMsim and GEBV software, was involved in the discussions, and helped to draft the manuscript. FSS participated in the design of the study, was involved in the discussions and helped to draft the manuscript.

All authors read and approved the final manuscript.

**References**

1. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: **Invited review: Genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**:433–443.

2. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited review: reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92**:16–24.

3. Harris BL, Johnson DL: **Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation.** *J Dairy Sci* 2010, **93**:1243–1252.

4. Su G, Madsen P, Nielsen US, Mäntysaari EA, Aamand GP, Christensen OF, Lund MS: **Genomic prediction for Nordic Red Cattle using one-step and selection index blending.** *J Dairy Sci* 2012, **95**:909–917.

5. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819–1829.

6. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score.** *J Dairy Sci* 2010, **93**:743–752.

7. Schaeffer LR: **Strategy for applying genome-wide selection in dairy cattle.** *J Anim Breed Genet* 2006, **123**:218–223.

8. Sargolzaei M, Schenkel FS, VanRaden PM: **gebv: Genomic breeding value estimator for livestock**. In *Tech Rep to Dairy Cattle Breed Genet Comm Univ Guelph*. Guelph/ON - Canada, 2009.

9. Misztal I, Legarra A, Aguilar I: **Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information.** *J Dairy Sci* 2009, **92**:4648–4655.

10. VanRaden PM: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91**:4414–4423.

11. Legarra A, Aguilar I, Misztal I: **A relationship matrix including full pedigree and genomic information.** *J Dairy Sci* 2009, **92**:4656–4663.

12. Vitezica ZG, Aguilar I, Legarra A: **One-Step vs . Multi-Step Methods For Genomic Prediction In Presence Of Selection**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany, 2010.

13. Christensen OF, Lund MS: **Genomic prediction when some animals are not genotyped.** *Genet Sel Evol* 2010, **42**:2.

14. Garrick DJ: **The Nature , Scope And Impact Of Some Whole- Genome Analyses In Beef Cattle**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany, 2010.

15. Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE: **Genome Wide Marker Assisted Selection in Chicken : Making the Most of All Data , Pedigree , Phenotypic , and Genomic in a Simple One Step Procedure**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany, 2010.

16. Legarra A, Christensen OF, Aguilar I, Misztal I: **Single Step, a general approach for genomic selection**. *Livest Sci* 2014, **166**:54–65.

17. Baloche G, Legarra A, Sallé G, Larroque H, Astruc J-M, Robert-Granié C, Barillet F: **Assessment of accuracy of genomic prediction for French Lacaune dairy sheep.** *J Dairy Sci* 2014, **97**:1107–1116.

18. Carillier C, Larroque H, Robert-Granié C: **Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population**. *Genet Sel Evol* 2014, **46**:67.

19. Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE, Aggrey SE, Wing T, Muir WM: **Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens.** *J Anim Sci* 2011, **89**:23–28.

20. Koivula M, Strandén I, Su G, Mäntysaari EA: **Different methods to calculate genomic predictions--comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP).** *J Dairy Sci* 2012, **95**:4065–4073.

21. Harris BL, Winkelman AM, Johnson DL: **Impact of Including a Large Number of Female Genotypes on Genomic Selection**. In *Interbull Bull No 47 23-25 August 2013*. Nantes, France, 2013:23–27.

22. Přibyl J, Madsen P, Bauer J, Přibylová J, Simečková M, Vostrý L, Zavadilová L: **Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production.** *J Dairy Sci* 2013, **96**:1865–1873.

23. Brito F V, Neto JB, Sargolzaei M, Cobuci JA, Schenkel FS: **Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle.** *BMC Genet* 2011, **12**:80.

24. Sargolzaei M, Schenkel FS: **QMSim: a large-scale genome simulator for livestock.** *Bioinformatics* 2009, **25**:680–1.

25. Snelling WM, Chiu R, Schein JE, Hobbs M, Abbey CA, Adelson DL, Aerts J, Bennett GL, Bosdet IE, Boussaha M, Brauning R, Caetano AR, Costa MM, Crawford AM, Dalrymple BP, Eggen A, Everts-van der Wind A, Floriot S, Gautier M, Gill CA, Green RD, Holt R, Jann O, Jones SJ, Kappes SM, Keele JW, de Jong PJ, Larkin DM, Lewin HA, McEwan JC, et al.: **A physical map of the bovine genome.** *Genome Biol* 2007, **8**:R165.

26. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH: **BLUPF90 AND RELATED PROGRAMS (BGF90)**. In *7th World Congr Genet Appl to Livest Prod 19-23 August 2002*. Montpellier, France, 2002.

27. VanRaden PM, Wiggans GR: **Derivation, calculation, and use of national animal model information.** *J Dairy Sci* 1991, **74**:2737–2746.

28. Guo G, Lund MS, Zhang Y, Su G: **Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables.** *J Anim Breed Genet* 2010, **127**:423–432.

29. Ostersen T, Christensen OF, Henryon M, Nielsen B, Su G, Madsen P: **Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs.** *Genet Sel Evol* 2011, **43**:38.

30. Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, S. D: **Genomic predictions in Angus cattle : Comparisons of sample size , response variables , and clustering methods for cross-validation**. *J Anim Sci* 2014, **92**:485–497.

31. Neves HHR, Carvalheiro R, Queiroz SA: **Adequacy of Different Pseudo-Phenotypes for Model Training and Validation of Genomic Predictions in Beef Cattle**. In *Proc 10th World Congr Genet Appl to Livest Prod 17-22 August 2014*. Vancouver, Canada, 2014.

32. De Roos APW, Schrooten C, Mullaart E, Calus MPL, Veerkamp RF: **Breeding value estimation for fat percentage using dense markers on Bos taurus autosome 14.** *J Dairy Sci* 2007, **90**:4821–4829.

33. Onogi A., Ogino A, Komatsu T, Shoji N, Simizu K, Kurogi K, Yasumori T, Togashi K, Iwata H: **Genomic prediction in Japanese Black cattle: Application of a single-step approach to beef cattle**. *J Anim Sci* 2014, **92**:1931–1938.

34. Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G: **Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population.** *Genet Sel Evol* 2012, **44**:8.

35. Calus M, Veerkamp R: **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**. *J Anim Breed Genet* 2007, **124**:362–368.

36. Liu Z, Seefried FR, Reinhardt F, Rensing S, Thaller G, Reents R: **Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction.** *Genet Sel Evol* 2011, **43**:19.

37. Neves HHR, Carvalheiro R, O'Brien AMP, Utsunomiya YT, do Carmo AS, Schenkel FS, Sölkner J, McEwan JC, Van Tassell CP, Cole JB, da Silva MVGB, Queiroz SA, Sonstegard TS, Garcia JF: **Accuracy of genomic predictions in Bos indicus (Nellore) cattle.** *Genet Sel Evol* 2014, **46**:17.

38. Cardoso FF, Sollero BP, Gomes CCG, Oliveira MM, Roso VM, Higa RH, Yokoo MJ, Caetano AR, Aguilar I: **Genomic selection for tick resistance**. *J Anim Sci* 2014, **submitted**.

39. Schenkel FS, Sargolzaei M, Kistemaker G, Jansen GB, Sullivan P, Van Doormaal BJ, VanRaden PM, Wiggans GR: **Reliability of Genomic Evaluation of Holstein Cattle in Canada**. In *Interbull Int Work Genomic Inf Genet Eval 26-29 january 2009*. Uppsala, Sweden, 2009.

**Table 1 DGVs accuracies (Acc) and standard deviations (SD)[123]**

| DGV | $_G95\text{-}_A5$ | | | $_G90\text{-}_A10$ | | | $_G85\text{-}_A15$ | | | $_G80\text{-}_A20$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc | SD | | Acc | SD | | Acc | SD | | Acc | SD | |
| ssDGV | 0.584 | 0.018 | a,a | 0.571 | 0.018 | a,a | 0.554 | 0.018 | b,a | 0.535 | 0.020 | b,a |
| tsDGVdebv | 0.577 | 0.018 | a,a | 0.573 | 0.018 | a,a | 0.569 | 0.019 | a,a | 0.564 | 0.019 | a,a |
| tsDGVebv | 0.576 | 0.019 | a,a | 0.573 | 0.019 | a,a | 0.569 | 0.019 | a,a | 0.563 | 0.019 | a,a |

[1] ssDGV predicted by single-step procedure and tsDGVdebv and tsDGVebv predict by two-steps procedure using dEBV or EBV;

[2] $_G95\text{-}_A5$ means to 5% polygenic effect, $_G90\text{-}_A10$ means to 10% polygenic effect, $_G85\text{-}_A15$ means to 15% polygenic effect and $_G80\text{-}_A20$ means to 20% polygenic effect;

[3] Different letters indicate significant differences (P < 0.05) by Scheffes's test. The first letter indicates differences within rows, while the second letter indicates differences within columns.

**Table 2 GEBVs accuracies (Acc) and standard deviations (SD) using PP1 prediction population[1234]**

| GEBV | $_G$95-$_A$5 | | $_G$90-$_A$10 | | $_G$85-$_A$15 | | $_G$80-$_A$20 | |
|------|------|------|------|------|------|------|------|------|
| | Acc | SD | Acc | SD | Acc | SD | Acc | SD |
| ssGEBV | 0.589 | 0.019 a,ab | 0.583 | 0.019 a,ab | 0.577 | 0.019 a,ab | 0.569 | 0.019 a,bc |
| tsGEBVv_debv | 0.612 | 0.016 a,a | 0.610 | 0.017 a,a | 0.608 | 0.017 a,a | 0.604 | 0.017 a,a |
| tsGEBVv_ebv | 0.611 | 0.017 a,a | 0.609 | 0.017 a,a | 0.605 | 0.017 a,a | 0.601 | 0.017 a,ab |
| tsGEBVh_debv | 0.570 | 0.020 ba,b | 0.562 | 0.020 ba,b | 0.554 | 0.021 ba,b | 0.545 | 0.021 b,c |
| tsGEBVh_ebv | 0.566 | 0.021 ba,b | 0.559 | 0.021 ba,b | 0.551 | 0.021 ba,b | 0.542 | 0.021 b,c |

[1] PP1 prediction population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals;

[2] ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv and tsGEBVh_ebv predicted by two-steps procedure using dEBV or EBV as pseudo phenotype, and VanRaden "v" or Hayes "h" blending method;

[3] $_G$95-$_A$5 means to 5% polygenic effect, $_G$90-$_A$10 means to 10% polygenic effect, $_G$85-$_A$15 means to 15% polygenic effect and $_G$80-$_A$20 means to 20% polygenic effect;

[4] Different letters indicate significant differences (P < 0.05) by Scheffes's test. The first letter indicates differences within rows, while the second letter indicates differences within columns.

**Table 3 GEBVs accuracies (Acc) and standard deviations (SD) using PP2 prediction population[1234]**

| GEBV | G95-A5 | | | G90-A10 | | | G85-A15 | | | G80-A20 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc | SD | | Acc | SD | | Acc | SD | | Acc | SD | |
| ssGEBV | 0.699 | 0.016 | a,a | 0.696 | 0.016 | a,a | 0.692 | 0.016 | a,a | 0.687 | 0.016 | a,a |
| tsGEBVv_debv | 0.685 | 0.016 | a,a | 0.683 | 0.016 | a,a | 0.680 | 0.016 | a,a | 0.676 | 0.016 | a,a |
| tsGEBVv_ebv | 0.684 | 0.016 | a,a | 0.681 | 0.016 | a,a | 0.678 | 0.016 | a,a | 0.674 | 0.016 | a,a |
| tsGEBVh_debv | 0.655 | 0.016 | a,b | 0.650 | 0.016 | a,b | 0.645 | 0.016 | a,b | 0.640 | 0.016 | a,b |
| tsGEBVh_ebv | 0.653 | 0.016 | a,b | 0.649 | 0.016 | a,b | 0.644 | 0.016 | a,b | 0.639 | 0.016 | a,b |

[1] PP2 prediction population formed with all animals born in the 10th generation and including the phenotypes of genotyped animals;

[2] ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv and tsGEBVh_ebv predicted by two-steps procedure using dEBV or EBV as pseudo phenotype, and VanRaden "v" or Hayes "h" blending methods;

[3] G95-A5 means to 5% polygenic effect, G90-A10 means to 10% polygenic effect, G85-A15 means to 15% polygenic effect and G80-A20 means to 20% polygenic effect;

[4] Different letters indicate significant differences ($P < 0.05$) by Scheffes's test. The first letter indicates differences within rows, while the second letter indicates differences within columns.

**Table 4 Accuracies (Acc) and standard deviations (SD) using single-step procedure[123]**

| Genetic merit | $_G$95-$_A$5 | | | $_G$90-$_A$10 | | | $_G$85-$_A$15 | | | $_G$80-$_A$20 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc | Sd | | Acc | Sd | | Acc | Sd | | Acc | Sd | |
| EBVpa | 0.335 | 0.019 | d | 0.335 | 0.019 | d | 0.335 | 0.019 | d | 0.335 | 0.019 | d |
| EBVphe | 0.534 | 0.017 | c | 0.534 | 0.017 | c | 0.534 | 0.017 | c | 0.534 | 0.017 | c |
| ssDGVpa | 0.584 | 0.018 | b | 0.571 | 0.018 | b | 0.554 | 0.018 | bc | 0.535 | 0.020 | c |
| ssGEBVpa | 0.589 | 0.019 | b | 0.583 | 0.019 | b | 0.577 | 0.019 | b | 0.569 | 0.019 | b |
| ssDGVphe | 0.695 | 0.016 | a | 0.686 | 0.015 | a | 0.674 | 0.014 | a | 0.661 | 0.014 | a |
| ssGEBVphe | 0.699 | 0.016 | a | 0.696 | 0.016 | a | 0.692 | 0.016 | a | 0.687 | 0.016 | a |

[1] EBVpa are EBVs parent's average, EBVphe are traditional EBVs, ssDGVpa and ssDGVphe are direct genomic value, ssGEBVpa and ssGEBVphe are genomic estimated breeding value. The "pa" means that the analyses were performed with prediction population formed with all animals born in the 10$^{th}$ generation and not including the phenotypes of genotyped animals (PP1) and "phe" including the phenotypes of genotyped animals (PP2);

[2] $_G$95-$_A$5 means to 5% polygenic effect, $_G$90-$_A$10 means to 10% polygenic effect, $_G$85-$_A$15 means to 15% polygenic effect and $_G$80-$_A$20 means to 20% polygenic effect;

[3] Different letters indicate significant differences ($P < 0.05$) by Scheffes's test. The letters indicate differences within columns.

**Table 5 Accuracies (Acc) and standard deviations (SD) using two-steps procedure[123]**

| Genetic merit | $_G95\text{-}_A5$ Acc | Sd | | $_G90\text{-}_A10$ Acc | Sd | | $_G85\text{-}_A15$ Acc | Sd | | $_G80\text{-}_A20$ Acc | Sd | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EBVpa | 0.335 | 0.019 | f | 0.335 | 0.019 | f | 0.335 | 0.019 | f | 0.335 | 0.019 | e |
| EBVphe | 0.534 | 0.017 | e | 0.534 | 0.017 | e | 0.534 | 0.017 | e | 0.534 | 0.017 | d |
| tsDGV | 0.577 | 0.020 | d | 0.573 | 0.018 | d | 0.569 | 0.019 | d | 0.564 | 0.019 | d |
| tsGEBVh_pa | 0.570 | 0.018 | d | 0.562 | 0.020 | de | 0.554 | 0.021 | de | 0.545 | 0.021 | d |
| tsGEBVv_pa | 0.612 | 0.016 | c | 0.610 | 0.017 | c | 0.608 | 0.017 | c | 0.604 | 0.017 | c |
| tsGEBVh_phe | 0.655 | 0.016 | b | 0.650 | 0.016 | b | 0.645 | 0.016 | b | 0.640 | 0.016 | b |
| tsGEBVv_phe | 0.685 | 0.016 | a | 0.683 | 0.016 | a | 0.680 | 0.016 | a | 0.676 | 0.016 | a |

[1] EBVpa are EBVs parent's average, EBVphe are traditional EBVs, tsDGV are direct genomic value, tsGEBVh_pa, tsGEBVv_pa, tsGEBVh_phe, tsGEBVv_phe are genomic estimated breeding value. The "pa" means that the analyses were performed with prediction population formed with all animals born in the $10^{th}$ generation and not including the phenotypes of genotyped animals (PP1) and "phe" including the phenotypes of genotyped animals (PP2). The "h" means that the Hayes blend method and the "v" means that the VanRaden blend method;

[2] $_G95\text{-}_A5$ means to 5% polygenic effect, $_G90\text{-}_A10$ means to 10% polygenic effect, $_G85\text{-}_A15$ means to 15% polygenic effect and $_G80\text{-}_A20$ means to 20% polygenic effect;

[3] Different letters indicate significant differences ($P < 0.05$) by Scheffes's test. The letters indicate differences within columns.

**Table 6 Regression coefficients (b1) and standard deviations (SD)[123]**

| DGV | G95-A5 | | | G90-A10 | | | G85-A15 | | | G80-A20 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | b1 | SD | | b1 | SD | | b1 | SD | | b1 | SD | |
| ssDGV | 0,944 | 0,033 | * | 0,919 | 0,036 | * | 0,882 | 0,040 | * | 0,837 | 0,045 | * |
| tsDGVdebv | 0,982 | 0,032 | ns | 1,005 | 0,034 | ns | 1,026 | 0,036 | ns | 1,045 | 0,038 | ns |
| tsDGVebv | 1,038 | 0,036 | ns | 1,063 | 0,038 | * | 1,085 | 0,041 | * | 1,106 | 0,043 | * |

[1] ssDGV predicted by single-step procedure and tsDGVdebv and tsDGVebv predict by two-steps procedure using dEBV or EBV;

[2] $_G$95-$_A$5 means to 5% polygenic effect, $_G$90-$_A$10 means to 10% polygenic effect, $_G$85-$_A$15 means to 15% polygenic effect and $_G$80-$_A$20 means to 20% polygenic effect;

[3] Within a group means b1(DGV,TBV), "ns" refers to statistically no different 1.00 ($P>0.05$) and "*" refers to statistically different 1.00 ($P<0.05$).

**Table 7 Regression coefficients (b1) and standard deviations (SD) using PP1 prediction population[1234]**

| GEBV | $_G$95-$_A$5 | | | $_G$90-$_A$10 | | | $_G$85-$_A$15 | | | $_G$80-$_A$20 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | b1 | SD | | b1 | SD | | b1 | SD | | b1 | SD | |
| ssGEBV | 0,967 | 0,032 | ns | 0,993 | 0,034 | ns | 1,014 | 0,036 | ns | 1,032 | 0,037 | ns |
| tsGEBVv_debv | 1,035 | 0,026 | ns | 1,058 | 0,027 | * | 1,079 | 0,028 | * | 1,098 | 0,029 | * |
| tsGEBVv_ebv | 1,076 | 0,027 | * | 1,099 | 0,028 | * | 1,119 | 0,029 | * | 1,136 | 0,030 | * |
| tsGEBVh_debv | 1,256 | 0,053 | * | 1,271 | 0,056 | * | 1,281 | 0,060 | * | 1,288 | 0,063 | * |
| tsGEBVh_ebv | 1,298 | 0,057 | * | 1,312 | 0,061 | * | 1,322 | 0,064 | * | 1,327 | 0,068 | * |

[1] PP1 prediction population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals;

[2] ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv and tsGEBVh_ebv predicted by two-steps procedure using dEBV or EBV as pseudo phenotype, and VanRaden "v" or Hayes "h" blending method;

[3] $_G$95-$_A$5 means to 5% polygenic effect, $_G$90-$_A$10 means to 10% polygenic effect, $_G$85-$_A$15 means to 15% polygenic effect and $_G$80-$_A$20 means to 20% polygenic effect;

[4] Within a group means b1(GEBV,TBV), "ns" refers to statistically no different 1.00 (P>0.05) and "*" refers to statistically different 1.00 (P<0.05).

**Table 8 Regression coefficients (b1) and standard deviations (SD) using PP2 prediction population[1234]**

| GEBV | G95-A5 | | | G90-A10 | | | G85-A15 | | | G80-A20 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | b1 | SD | | b1 | SD | | b1 | SD | | b1 | SD | |
| ssGEBV | 0,984 | 0,039 | ns | 0,995 | 0,040 | ns | 1,002 | 0,042 | ns | 1,007 | 0,043 | ns |
| tsGEBVv_debv | 1,006 | 0,025 | ns | 1,011 | 0,027 | ns | 1,013 | 0,029 | ns | 1,015 | 0,031 | ns |
| tsGEBVv_ebv | 1,039 | 0,027 | ns | 1,041 | 0,029 | ns | 1,041 | 0,031 | ns | 1,039 | 0,033 | ns |
| tsGEBVh_debv | 1,286 | 0,034 | * | 1,293 | 0,035 | * | 1,296 | 0,036 | * | 1,297 | 0,037 | * |
| tsGEBVh_ebv | 1,341 | 0,040 | * | 1,346 | 0,041 | * | 1,348 | 0,042 | * | 1,348 | 0,043 | * |

[1] PP1 prediction population formed with all animals born in the $10^{th}$ generation and not including the phenotypes of genotyped animals;

[2] ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv and tsGEBVh_ebv predicted by two-steps procedure using dEBV or EBV as pseudo phenotype, and VanRaden "v" or Hayes "h" blending method;

[3] G95-A5 means to 5% polygenic effect, G90-A10 means to 10% polygenic effect, G85-A15 means to 15% polygenic effect and G80-A20 means to 20% polygenic effect;

[4] Within a group means b1(GEBV,TBV), "ns" refers to statistically no different 1.00 (P>0.05) and "*" refers to statistically different 1.00 (P<0.05).

**Table 9 Percentage of coincidences in the selection of 10% males and 50% females[123]**

| DGV | $_{G}95-_{A}5$ | | $_{G}90-_{A}10$ | | $_{G}85-_{A}15$ | | $_{G}80-_{A}20$ | |
|---|---|---|---|---|---|---|---|---|
| | | | males | | | | | |
| ssDGV | 41 | a,a | 39 | a,a | 37 | a,a | 36 | a,a |
| tsDGVdebv | 40 | a,a | 40 | a,a | 39 | a,a | 39 | a,a |
| tsDGVebv | 40 | a,a | 40 | a,a | 39 | a,a | 39 | a,a |
| | | | females | | | | | |
| ssDGV | 70 | a,a | 70 | a,a | 69 | a,a | 69 | a,a |
| tsDGVdebv | 70 | a,a | 70 | a,a | 70 | a,a | 69 | a,a |
| tsDGVebv | 70 | a,a | 70 | a,a | 70 | a,a | 69 | a,a |

[1] ssDGV predicted by single-step procedure and tsDGVdebv and tsDGVebv predict by two-steps procedure using dEBV or EBV;

[2] $_{G}95-_{A}5$ means to 5% polygenic effect, $_{G}90-_{A}10$ means to 10% polygenic effect, $_{G}85-_{A}15$ means to 15% polygenic effect and $_{G}80-_{A}20$ means to 20% polygenic effect;

[3] Different letters indicate significant differences ($P < 0.05$) by Scheffes's test. The first letter indicates differences within rows, while the second letter indicates differences within columns.

**Table 10 Percentage of coincidences in the selection of 10% males and 50% females using PP1 prediction population[1234]**

| GEBV | $_G95-_A5$ | | $_G90-_A10$ | | $_G85-_A15$ | | $_G80-_A20$ | |
|---|---|---|---|---|---|---|---|---|
| Males | | | | | | | | |
| ssGEBV | 41 | a,a | 41 | a,a | 40 | a,a | 40 | a,a |
| tsGEBVv_debv | 42 | a,a | 42 | a,a | 42 | a,a | 42 | a,a |
| tsGEBVv_ebv | 42 | a,a | 42 | a,a | 42 | a,a | 42 | a,a |
| tsGEBVh_debv | 40 | a,a | 39 | a,a | 38 | a,a | 37 | a,a |
| tsGEBVh_ebv | 39 | a,a | 38 | a,a | 37 | a,a | 37 | a,a |
| Females | | | | | | | | |
| ssGEBV | 70 | a,a | 70 | a,a | 70 | a,a | 69 | a,a |
| tsGEBVv_debv | 71 | a,a | 71 | a,a | 70 | a,a | 70 | a,a |
| tsGEBVv_ebv | 70 | a,a | 70 | a,a | 70 | a,a | 70 | a,a |
| tsGEBVh_debv | 69 | a,a | 68 | a,a | 68 | a,a | 68 | a,a |
| tsGEBVh_ebv | 69 | a,a | 68 | a,a | 68 | a,a | 68 | a,a |

[1] PP1 prediction population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals;

[2] ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv and tsGEBVh_ebv predicted by two-steps procedure using dEBV or EBV as pseudo phenotype, and VanRaden "v" or Hayes "h" blending method;

[3] $_G95-_A5$ means to 5% polygenic effect, $_G90-_A10$ means to 10% polygenic effect, $_G85-_A15$ means to 15% polygenic effect and $_G80-_A20$ means to 20% polygenic effect;

[4] Different letters indicate significant differences (P < 0.05) by Scheffes's test. The first letter indicates differences within rows, while the second letter indicates differences within columns.

**Table 11 Percentage of coincidences in the selection of 10% males and 50% females using PP2 prediction population[1234]**

| Genetic merit | $_G95-_A5$ | | $_G90-_A10$ | | $_G85-_A15$ | | $_G80-_A20$ | |
|---|---|---|---|---|---|---|---|---|
| | | | males | | | | | |
| ssGEBV | 47 | a,a | 48 | a,a | 48 | a,a | 48 | a,a |
| tsGEBVv_debv | 49 | a,a | 49 | a,a | 49 | a,a | 48 | a,a |
| tsGEBVv_ebv | 48 | a,a | 48 | a,a | 48 | a,a | 47 | a,a |
| tsGEBVh_debv | 46 | a,a | 45 | a,a | 45 | a,a | 44 | a,a |
| tsGEBVh_ebv | 46 | a,a | 45 | a,a | 45 | a,a | 44 | a,a |
| | | | females | | | | | |
| ssGEBV | 75 | a,a | 75 | a,a | 74 | a,a | 74 | a,a |
| tsGEBVv_debv | 73 | a,a | 73 | a,a | 73 | a,a | 73 | a,a |
| tsGEBVv_ebv | 73 | a,a | 73 | a,a | 73 | a,a | 73 | a,a |
| tsGEBVh_debv | 72 | a,a | 72 | a,a | 72 | a,a | 72 | a,a |
| tsGEBVh_ebv | 72 | a,a | 72 | a,a | 71 | a,a | 71 | a,a |

[1] PP2 prediction population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals;

[2] ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv and tsGEBVh_ebv predicted by two-steps procedure using dEBV or EBV as pseudo phenotype, and VanRaden "v" or Hayes "h" blending method;

[3] $_G95-_A5$ means to 5% polygenic effect, $_G90-_A10$ means to 10% polygenic effect, $_G85-_A15$ means to 15% polygenic effect and $_G80-_A20$ means to 20% polygenic effect;

[4] Different letters indicate significant differences (P < 0.05) by Scheffes's test. The first letter indicates differences within rows, while the second letter indicates differences within columns.

[9]**CAPÍTULO VI**

# Genomic evaluation using single and two steps procedures in the presence of selection in Braford and Hereford beef cattle

**Mario L Piccoli[1,2,3]\*, José Braccini[1,5], Fernando F Cardoso[4,5], Vanerlei M Roso[2], Medhi Sargolzaei[3,6], Flávio S Schenkel[3]**

[1] Universidade Federal do Rio Grande do Sul, Departamento de Zootecnia, Porto Alegre, Brazil.
[2] GenSys Consultores Associados S/S, Porto Alegre, Brazil.
[3] Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada.
[4] Embrapa Southern Region Animal Husbandry, Bagé, Brazil.
[5] National Council for Scientific and Technological Development, Brasília, Brazil
[6] The Semex Alliance, Guelph, Canada.

\* Corresponding author: Mario L Piccoli (mariopiccoli@gensys.com.br)

Email addresses:

MLP: mariopiccoli@gensys.com.br

JB: jose.braccini@ufrgs.br

FFC: fernando.cardoso@embrapa.br

VMR: vanerleirosos@gensys.com.br

MS: msargol@uoguelph.ca

FSS: schenkel@uoguelph.ca

## Abstract

**Background:** Different strategies to the use of molecular markers data in genomic predictions have been proposed and greater genetic gains have been achieved in breeding programs. Best Linear Unbiased Prediction (BLUP) methods based on single and two steps procedures have been developed to analyze pedigree and phenotype data jointly with single nucleotide polymorphism (SNP) information. Data from Braford and Hereford breed animals born from 1975 to 2011, were used in order to assess genomic predictive ability for pre and post weaning weight gain, birth weight, scrotal circumference and conformation, precocity, muscularity, body size, prepuce (navel), hair length and ocular pigmentation score. Training populations were composed by animals born from 2008 to 2010 plus 130 sires and the prediction populations were composed by animals born in 2011. In total were genotyped 3,680 animals (2,997 Braford and 683 Hereford). Forward prediction schemes were adopted to predict the direct genomic value (DGV) and genomic estimated breeding value (GEBV) based on GEBV package (two-steps) and BLUPF90 package (single-step). The aim of this study was to compare the single and two steps procedures using Braford and Hereford beef cattle data.

**Results:** In general there were no statistical differences ($P>0.05$) in DGVs accuracies between single and two steps procedures. The empirical accuracies ranged from 0.08 to 0.42 when validating with deregressed estimated breeding value (dEBV) and from 0.16 to 0.66 when validating with estimated breeding value (EBV). There were no statistical differences ($P>0.05$) in GEBVs accuracies between single and two steps procedures. The empirical accuracies for single and two steps ranged from 0.10 to 0.41 and from 0.08 to 0.41 validating with dEBV and from 0.26 to 0.66 and from 0.27 to 0.70 validating with EBV, respectively. The regression coefficients for DGVs and GEBVs were closer to 1.00 when using two-steps procedure.

**Conclusion:** Both DGV and GEBV predicted by single and two steps procedures produce the same level of empirical accuracies, but with two-step procedure the genomic predictions were less biased.

**Keywords:** Genomic accuracy, Single-step, Two-step, Braford, Hereford.

## Background

Traditional animal breeding methods use phenotypic data and relationships among individuals to predict breeding value for using in mating decision to improve economically important traits. Recently, molecular markers information has been used jointly with phenotypic data and relationships among animals in order to obtain more accurate predictions in earlier stages of the animal breeding. Currently, single nucleotide polymorphism (SNPs) genotyping is viable at a reasonable cost for producers and has allowed to evaluate the contribution of its use in dairy cattle [1], [2], beef cattle [3], [4], pigs [5], [6], sheep [7], [8], goat [9], [10], poultry [11], [12] and horse [13] breeding programs.

Single and two steps procedures have been developed to analyze the pedigree and the phenotype jointly with the information of SNP markers [14], [15]. The two-step procedure means that estimated breeding value (EBV) were obtained by conventional analyzes (based on pedigree and phenotype data) and direct genomic value (DGV) were obtained based on prediction equations (through phenotypes and genotypes information) and after both were combined by different indexes generating the genomic

estimated breeding value (GEBV). The single-step means that the information of the SNPs was combined with the phenotype and pedigree for the simultaneous estimation of breeding values.

The first tests to combine the use of genomic data with the EBVs were based on two-step procedure. The use of the two sources of information, DGVs and EBVs, is important because if the effect of the quantitative trait loci (QTL) is not captured by a SNP, this may be captured by the polygenic breeding value [1], [16]. The disadvantage of this procedure according to Legarra et al. [17] and Misztal et al. [15] has been associated with the fact that the estimation process in more than one step and the need for pre-estimated parameters. However, advantage is that there was no change in the statistical model in routine genetic evaluations.

The single-step procedure does not need the multiple steps and therefore there is lower probability of errors during the estimation process and is easer to generalization to the other models and species [18], [19]. Aguilar et al. [20] mentioned that the single-step procedure provided a unified structure, eliminating various assumptions and allowed to calculate more accurately genomic evaluations. However, the single-step procedure that uses the traditional relationship matrix enlarged with the markers relationship information has computational difficulty in obtaining this matrix and its inverse [19], [20].

Previous results presented by Vitezica et al. [18], Aguilar et al. [20], Garrick [21] and Chen et al. [22] showed advantages and disadvantages when using both procedures. More recently, Legarra et al. [23] have related that in dairy sheep and goats, pigs and chickens, in general, the results have shown an advantage when using the single-step. In dairy cattle, genomic evaluations have been conducted using the two-step procedure [1], [2], [24], [25] and recently Koivula et al. [26], Harris et al. [27], Pribyl et al. [28] have been evaluating the single-step procedure with equal or better results in comparison to the two-step method.

In beef cattle it was not found studies comparing the application of single and two steps procedures in genomic evaluation. Furthermore, in general the beef cattle datasets are more complex because there are missing information in the pedigree file, smaller sib ships, and the presence of maternal effects [23]. Considering that, the aim of this study was to compare the single and two steps procedures using Braford and Hereford beef cattle data.

## Methods

### Animal welfare

Animal welfare and use committee approval was not necessary for this study because the data was obtained from existing databases.

### Data

Data was from the Conexão Delta G's genetic improvement program - Hereford and Braford (Zebu x Hereford) cattle (Conexão Delta G, Dom Pedrito/RS, Brazil), containing approximately 540,769 animals born between 1975 and 2011, from 97 farms located in the South, Southeast, Midwest and Northeast regions of Brazil. Of the 540,769 animals, 71% are Braford animals and 29% are Hereford animals, and 40% are males and 60% females. There were ~46% of animals with sire unknown (multiple sire mating), ~22% of animals with both parents unknown and ~33% of animals with both parents known (Table 1). A total of 624 Hereford and 2,926 Braford animals born from

2008 to 2011 plus 59 Hereford and 71 Braford sires were genotyped. Of the genotyped animals, there were 624 Hereford and 2,926 Braford animals genotyped with the Illumina BovineSNP50 panel, and 59 Hereford and 71 Braford animals genotyped with the Illumina BovineHD panel from 17 farms located in the South of Brazil.

Eighteen traits were analysed in this study including weight gain from birth to weaning in kilograms (WGBW), weight gain from weaning to yearling in kilograms (WGWY), conformation score at weaning (CW), conformation score at yearling (CY), precocity score at weaning (PW), precocity score at yearling (PY), muscularity score at weaning (MW), muscularity score at yearling (MY), scrotal circumference adjusted for age at yearling (SCa), scrotal circumference adjusted for age and weight at yearling (SCaw), birth weight (BW), size score at weaning (SW), size score at yearling (SY), prepuce (navel) score at weaning (NW), prepuce (navel) score at yearling (NY), hair length score at weaning (HW), hair length score at yearling (HY) and ocular pigmentation score (OP).

**Genotype data editing**

The SNP quality control included GenCall score (>=0.15), Call Rate (>=0.90), Hardy-Weinberg Equilibrium (P>=$10^{-6}$), Minor Allele Frequency (>=0.05) and only autosomal chromosome were considered, [1], [29]. The individual sample quality control considered GenCall Score (>=0.15), Call Rate (>=0.90), heterozygosity deviation (limit of $\pm$ 3 SD), repeated sampling and paternity errors [29]. After quality control remained 43,247 SNPs.

**Training and prediction population for use in the two-step procedure**

Two EBVs sets were generated using the BLUPF90 package [30] using animal model with maternal effect. The first one was formed by the prediction population with all genotyped animals while the second set was formed by the training population with all genotyped animals born until 2010 (Table 1).

The deregressed EBV (dEBV) and EBV of the training population were used as pseudo phenotypes to estimate markers effects. The approach of VanRaden and Wiggans [31] was used to calculate dEBVs using EBVs and reliabilities of genotyped animals and their sires and dams.

Direct genomic values (DGV) were estimated using GBLUP method [16] for all the eighteen traits (Table 2), using 43,247 SNPs and dEBVs or EBVs in the GEBV package [14]. The following linear model was assumed:

The DGVs were predicted in the training population based on the GEBV package [14] in GBLUP model [16]. It can be described as:

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Z_g} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of dEBV or EBV for the trait, $\mathbf{\mu}$ is the overall mean, $\mathbf{1_n}$ is a vector of ones, $\mathbf{Z}$ is the design matrix that relates dEBVs or EBVs to animals, $\mathbf{g}$ is the vector of DGV to be predicted, and $\mathbf{e}$ is the vector of residual effects. It was assumed that $\mathbf{g} \sim N$ (0, $\mathbf{G^*}\sigma^2 g$) where $\sigma^2 g$ is the additive genetics variance and $\mathbf{G^*}$ is a combined relationship matrix, and $\mathbf{e} \sim N$ (0, $\mathbf{R}\sigma^2 e$) where $\sigma^2 e$ is the residual variance and $\mathbf{R}$ is a diagonal matrix whose elements account for the differences in reliabilities of the dEBVs or EBVs.

The genomic estimated breeding values (GEBV) were estimated using the blending procedure outlined by Hayes et al. [9] and by VanRaden et al. [2] and described below:

The blending by Hayes et al. [1] where:

$$\text{GEBV} = \frac{r_{DGV}^2 * \text{DGV} + r_{EBV}^2 * \text{EBV}}{r_{DGV}^2 + r_{EBV}^2} \text{ , where,}$$

$r_{DGV}^2$ and $r_{EBV}^2$ are the reliabilities of DGV and EBV, respectively.

The blending by VanRaden et al. [2] where:

$\text{GEBV} = b_1 * \text{DGV} + b_2 * \text{EBV}_1 + b_3 * \text{EBV}$ , where,

$\text{EBV}_1$ was predicted for the subset of genotyped animals using traditional relationships and their dEBV or EBV, excluding data from ungenotyped animals, and $b_1$, $b_2$ and $b_3$ are weights based on reliabilities of DGV, $\text{EBV}_1$ and EBV.

## Population for use in single-step procedure

Two sets were generated using the BLUPF90 package [30] using animal model with maternal effect. The first one was formed by all animals born until 2010 and their respectively genotypes to estimate the DGVs. The second set included all animals, all genotypes and phenotypes of animals born until 2010 to estimate the DGVs and GEBVs (Table 1).

The DGVs and GEBVs were estimated based on the BLUPF90 package [30] using GBLUP model [16]. It can be described as:

$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Z_g} + \mathbf{e}$,

where $\mathbf{y}$ is the vector of dEBV or EBV for the trait, $\mu$ is the overall mean, $\mathbf{1_n}$ is a vector of ones, $\mathbf{Z}$ is the design matrix that relates dEBVs or EBVs to animals, $\mathbf{g}$ is the vector of DGV to be predicted, and $\mathbf{e}$ is the vector of residual effects. It was assumed that $\mathbf{g} \sim N$ (0, $\mathbf{H}\sigma^2 g$) where $\sigma^2 g$ is the additive genetic variance and $\mathbf{H}$ is a combined relationship matrix, and $\mathbf{e} \sim N$ (0, $\mathbf{R}\sigma^2 e$) where $\sigma^2 e$ is the residual variance and $\mathbf{R}$ is a diagonal matrix whose elements account for the differences in reliabilities of the dEBVs or EBVs.

The GEBVs were obtained by directly combining phenotypic, genomic and pedigree information [15], [20] where the traditional relationship matrix (**A**) was replaced by a matrix that includes the genomic information (**H**).

## Comparison between single and two steps procedures

The average accuracy, measured by Pearson's correlation, between DGV and GEBV with dEBV and EBV in the prediction population were used as response in the analysis of variance carried out in this study with the ANOVA procedure of SAS version 9.2 (SAS Inst. Inc., Cary, NC).

## Results

## Database structure

From 540,769 animals, only 2,734 animals showed inbreeding [32] (F=0.079) because in the pedigree structure had 68% of animals with unknown sire e/or dam parents with no known not allowing correctly calculate the number of inbred animals and the inbreeding of each animal and the average inbreeding. After the quality control, a total of 3,305 genotyped animals remained for the analysis. On average, there were 1,680 and 939 animals for the training and prediction population for the eighteen traits, respectively. On average, 68% and 59% of genotyped animals of training and prediction population had sire unknown (multiple sire mating), respectively. Phenotypes and genotypes used in training and prediction population for each of the eighteen traits are presented in Table 2.

**Empirical accuracy of DGV and GEBV**

The empirical accuracies of DGV are presented in the Table 3 and the empirical accuracies of GEBV are presented in the Table 4. There were no significant differences (P>0.05) between the means of DGVs accuracies when using two-step procedure with dEBV or EBV in SNPs training. When validations were made with dEBV the accuracies ranged from 0.08 to 0.40 and when validations were made with EBV the accuracies ranged from 0.16 to 0.51. There were no significant differences (P>0.05) between the mean of DGVs accuracies when using two different single-step approaches (1. SNPs effect were estimated with genotyped animals in the training population and the estimates were used in the prediction population (ssT); 2. SNPs effect were estimated with genotyped animals in the training e prediction population (ssTP)) and validations were made with dEBV and there were significant differences (P<0.05) between the mean of DGVs accuracies when validations were made with EBV. The accuracies ranged from 0.08 to 0.42 when validating with dEBV and ranged from 0.13 to 0.66 when validating with EBV. There were no significant differences (P>0.05) between single and two steps procedures while validating with dEBV or EBV. Although there were no statistical differences between the average of the single and two step, in general, the accuracy were always higher in single-step, with the exception of PW for validating with dEBV or EBV and BW for validating with dEBV (Table 3).

There were no significant differences (P>0.05) between the means of GEBVs accuracies when using VanRaden or Hayes blending methods in two-step procedure with dEBV or EBV in SNPs training. When validations were made with dEBV the accuracies ranged from 0.08 to 0.41 and when validations were made with EBV the accuracies ranged from 0.27 to 0.70 (Table 4). There were no significant differences (P>0.05) between single and two steps procedure when validating with dEBV or EBV. The accuracies of single-step ranged from 0.10 to 0.41 validating with dEBV and ranged from 0.26 to 0.66 validating with EBV. Although there were no statistical differences between the average of the single and two step, in general, the accuracy were higher in two-step, with the exception for eight traits (WGBW, WGWY, PY, MW, SW, SY, NY, OP) when validating it was made with dEBV (Table 4).

**Scale of DGV and GEBV**

Slope coefficient for the regression of dEBV or EBV on DGV or GEBV was expected to be close to 1.00, which would indicate that DGV or GEBV predictions were not inflated or deflated (Tables 3 and 4).

Analysis using two-step procedure, in general, presented the slope of the regression on DGV close to 1.00 (P>0.05) regardless if were used the dEBV or EBV in the SNPs estimation and validation with dEBV or EBV, excep when using dEBV in the SNPs estimation and EBV in the validation (b1=0.61). CW, PW and MW had the most inflated results. In the analysis using single-step procedure the average slope of the regression were statistically different from 1.00 (P<0.05), however CY, PY, MY, had the regression coefficient closer to 1.00 (Table 3).

Slope coefficient for the regression of dEBV or EBV on GEBV using two-step procedure, in general, the results not showed deflation or inflation. In other words, the average coefficient were statistically close to 1.00 (P>0.05), exception for tsDv (b1=0.71) and tsEh (b1=1.16) when using EBV in the SNPs validation (Table 4). In the single-step procedure the regression coefficient of the eighteen traits were statistically no different from 1.00 (P>0.05) when were used dEBV in the prediction population and statistically different form 1.00 (P<0.05) when were used dEBV in the prediction

population. In the regression on DGV traits, CY, PY, and MY had the regression coefficients closer to 1.00 (Table 4).

**Percentage of changes between the top 20% animals**

Assuming selection pressure of 20% in males the exchange of animal percentage were analysed when changed the selection criteria of EBV or $EBV_{PA}$ (parents average estimated breeding value) to DGV, $GEBV^2$ (blend with $EBV_{PA}$) or $GEBV^3$ (blend with EBV). Table 5 shows the percentage of mismatches when changing the classification criteria. In average for the eighteen analysed traits, 938 animals formed the prediction population with 175 animals classified as top 20%. The mismatch between models when the criteria changed in comparison to EBV were, in average for the eighteen traits, 52% to $EBV_{PA}$, greater than 50% to DGV and $GEBV^2$ and lower than 30% to $GEBV^3$. The mismatch, in comparison to the $EBV_{PA}$ was 48% to EBV, around 50% to DGV and $GEBV^3$ and lower than 40% to $GEBV^2$. These results indicated an important reordering of the animals.

**Discussion**

EBV and dEBV have been used for estimating and for predicting the effect of molecular markers, [33], [34], since true breeding values (TBV) of animals are unknown and the goal is a better approximation of the TBV. In simulation studies, the correlation between DGV and TBV has been used to represent the accuracy of DGV. In this sense, Piccoli et al. [35] and Neves et al. [34], working with simulated data of beef cattle, showed that there were no differences between the use of dEBV or EBV. These results were probably due to the fact that the training populations were formed by animals with high accuracy. The results of this study were similar to the results presented by Piccoli et al. [35] and Neves et al. [34]. There is no need to deregress the EBVs for training population because the results were very similar with original EBVs and there were no statistical differences in accuracy (average of eighteen traits) between the estimation of DGVs and GEBVs using the dEBV or EBV. However, the results of the DGV and GEBV estimated in the training population and when validated in the prediction population with dEBV or EBV, presented expressive accuracy gains (average for eighteen traits) in favor of EBV. It was observed probably due to double-counting since the EBV also has ancestral information and also to double shrinkage especially when the accuracies of EBVs are low. These latest results showed the same behavior to the results reported by Boddhireddy et al. [33] when studying Angus cattle in USA. On the other hand, Boddhireddy et al. [33] showed superiority of the estimates with EBV compared to dEBV in the training population. Ostersen et al. [5] studied pigs in Denmark and reported that the use of dEBV compared to EBV produced 18% to 39% higher reliabilities.

One of the main reasons to use information of molecular markers in the prediction of breeding values is to be able to carry out the prediction at very young ages, even before obtaining phenotypes. Studies with simulated data where the TBV is known, the correlation of DGV and GEBV with TBV have indicated the accuracy values. In studies using field data the TBV is not known and researchers have used for this purpose the correlation of DGV and GEBV with variable response that can be phenotype records, EBV, DYD or dEBV. Neves et al. [29] studied several traits in Nellore in Brazil using dEBV and divided the correlation between DGV with dEBV by the average accuracy of dEBV. Saatchi et al. [36] studied several traits in Angus in USA using dEBV and the accuracy was obtained by standardizing the estimated

covariance between DGV and dEBV using the genetic variance. Saatchi et al. [4] studied ten traits in Hereford from USA, Canada, Argentina and Uruguay using dEBV and the accuracy was obtained by simple correlation between DGV and dEBV and Boddhireddy et al. [33] studied seventeen traits in Angus in USA, also using dEBV and the accuracy was obtained by simple correlation. In this study were used the simple correlations of the DGV and GEBV with dEBV and EBV. In general the results of this study showed that empirical accuracies were lower than results reported in other studies with similar heritabilities. It is most probably due to the fact that the training populations were made up of animals with lower accuracies of EBVs (0.62 on average for eighteen traits and it ranged between 0.55 (HW) and 0.72 (NW)).

Bayesian methods such as bayesB, bayesC, bayesLasso and BLUP methods based on single and two steps procedures are often used to estimate the DGVs and GEBVs being one of the causes of variation in accuracy values [37], [38], [39], [29]. In this study were compared the BLUP method based on single and two steps procedure. Results of empirical accuracies of DGV in this study showed that ssTP outperformed the other methods when were used dEBVs or EBVs in prediction populations. However, in the average of eighteen analyzed traits ssTP did not differ statistically (P>0.05) from the other methods when using dEBV in the prediction population and also did not differ statistically (P>0.05) from the tsE (two-step with SNP estimation based on EBV) method when using EBV in the prediction population. The average empirical accuracy of ssTP was 0.23 using dEBV in the prediction population and it ranged from 0.10 (PW) to 0.42 (NW) and when using EBV in predicting population the average empirical accuracy was 0.44 and it ranged from 0.25 (SCa and SCaw) to 0.66 (HY). The results of empirical accuracies of GEBV in this study showed similar performance between blend method by VanRaden [16] and by Hayes [1]. The average of empirical accuracies of eighteen analyzed traits using dEBV in the prediction population ranged from 0.21 to 0.22 and using EBV in the prediction population ranged from 0.49 to 0.52, and from the single-step procedure were 0,23 and 0.46 using dEBV and EBV in the prediction population, respectively. The results obtained by Piccoli et al. [35] working with simulated data of beef cattle showed equality accuracies between the single and two steps procedures for DGV and GEBV, similar to this study, however with accuracies higher for the same level of heritability. Su et al. [25] studying fifteen traits of Nordic Red breed in Finland found an average empirical accuracy of 0.309, 0.322 and 0.318 for DGV and GEBV by single and two steps, respectively, higher than the average empirical accuracies obtained this study. In the study of by Su et al. [25] the average difference between GEBV for single and two steps was 0.04 in favor of single-step procedure, while in this study was in average 0.01 or 0.02 for single-step procedure when using dEBV and 0.03 or 0.06 in favor of two-step procedure when using EBV, but both were not statistically different. Gao et al. [40] studied sixteen traits in Nordic Holstein in Denmark and also found a superiority accuracy (0.02) in favor of GEBV by single-step procedure.

The scale of genomic predictions should be a matter of concern, especially to determine whether DGV and GEBV can be compared to traditional EBV from routine evaluations in breeding programs. If the regression coefficient was smaller than 1.0 it would be indicating that there was overestimation of genomic predictions and if the regression coefficient was larger than 1.0 it would indicates that there was underestimation of genomic predictions. Vitezica et al. [41] have discussed that criterion under the aspect of selection. If the parents of the next generation come from only genotyped selection candidates, they share a common mean for belonging to the

same generation, then the bias would not be a concern. However, if for different candidates there are different amounts of information such as progeny test males and newborn animals and in presence of bias (genetic gains over or under estimated), thus, newborns could be considered better than they really are.

Regarding to the DGV, the regression coefficients of this study showed that the two-step procedure generated estimates less biased and that the average for eighteen traits the results did not differ statistically from 1.0 (P>0.05) while that using the single-step procedure estimates were overestimated. Regarding the GEBV, the regression coefficients showed that the two-step procedure, generally, generated estimates underestimated while those estimates using the single-step procedure were overestimated. The results obtained by Koivula et al. [26] that studied Nordic Red breed in Finland overestimated genomic predictions in both single and two steps procedures with regression coefficients varying between 0.56 and 0.90. However, Su et al. [25] also studying Nordic Red breed showed regression coefficients average of the traits of 0.946 and 0.941 for single and two steps, respectively, and Gao et al. [40] studying Nordic Holstein breed obtained regression coefficients average of the traits of 0.958 and 0.960 for single and two steps, respectively, showing good genomic predictions.

Breeding programs conducted in Brazil and approved by the Ministry of Agriculture, Livestock and Food Supply (MAPA) to issue the Special Certification of Identification and Production (CEIP) can certify 20% of the best animals born each year based on breeding values, which is the case of Conexão Delta G's genetic improvement program - Hereford and Braford (Zebu x Hereford) cattle, whose provided the data for this study. The results in Table 5 show the percentage of animals that would no longer be certified with CEIP if the selection criterion based on $EBV_{PA}$ (selection at birth) or EBV (selection after collecting phenotype) would changed to DGV, $GEBV^2$ or $GEBV^3$ genomic predictions. The results show that the extent to which was added other sources of information to predict breeding values the classification by breeding values was being changed, which in fact is expected. The accuracy gains (average of eighteen traits) of $GEBV^2$ (average=0.47), EBV (average=0.62) and $GEBV^3$ (average=0.64) compared to $EBV_{PA}$ (average=0.37) were 0.10, 0.25 and 0.74, respectively. These results reinforce the importance of using molecular markers in the prediction of breeding values for young animals.

**Conclusions**

Both DGV and GEBV predicted by single and two steps procedures produce the same level of empirical accuracies, but with two-step procedure the genomic preditions were less biased.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
MLP participated in the design of the study, carried out the analyses, was involved in the discussions, prepared and drafted the manuscript. JB participated in the design of the study, was involved in the discussions and helped to draft the manuscript. FFC was involved in the discussions and helped to draft the manuscript. VMR helped to draft the manuscript. MS developed the GEBV software, was involved in the discussions, and helped to draft the manuscript. FSS participated in the design of the study, was involved in the discussions and helped to draft the manuscript. All authors read and approved the

final manuscript.

**References**
1. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: **Invited review: Genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**:433–443.

2. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited review: reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92**:16–24.

3. Weber K, Thallman R: **Accuracy of genomic breeding values in multi-breed beef cattle populations derived from deregressed breeding values and phenotypes**. *J Anim Sci* 2012:4177–4190.

4. Saatchi M, Ward J, Garrick D: **Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations**. *J Anim Sci* 2013, **91**:1538–1551.

5. Ostersen T, Christensen OF, Henryon M, Nielsen B, Su G, Madsen P: **Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs.** *Genet Sel Evol* 2011, **43**:38.

6. Wellmann R, Preuß S, Tholen E, Heinkel J, Wimmers K, Bennewitz J: **Genomic selection using low density marker panels with application to a sire line in pigs.** *Genet Sel Evol* 2013, **45**:28.

7. Baloche G, Legarra A, Sallé G, Larroque H, Astruc J-M, Robert-Granié C, Barillet F: **Assessment of accuracy of genomic prediction for French Lacaune dairy sheep.** *J Dairy Sci* 2014, **97**:1107–1116.

8. Moghaddar N, Swan AA, van der Werf JHJ: **Genomic prediction of weight and wool traits in a multi-breed sheep population**. *Anim Prod Sci* 2014, **54**:544.

9. Carillier C, Larroque H, Palhière I, Clément V, Rupp R, Robert-Granié C: **A first step toward genomic selection in the multi-breed French dairy goat population.** *J Dairy Sci* 2013, **96**:7294–305.

10. Carillier C, Larroque H, Robert-Granié C: **Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population**. *Genet Sel Evol* 2014, **46**:67.

11. Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE, Aggrey SE, Wing T, Muir WM: **Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens.** *J Anim Sci* 2011, **89**:23–28.

12. Simeone R, Misztal I, Aguilar I, Vitezica ZG: **Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure.** *J Anim Breed Genet* 2012, **129**:3–10.

13. Corbin LJ, Kranis A, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams J a: **The utility of low-density genotyping for imputation in the Thoroughbred horse.** *Genet Sel Evol* 2014, **46**:9.

14. Sargolzaei M, Schenkel FS, VanRaden PM: **gebv: Genomic breeding value estimator for livestock**. In *Tech Rep to Dairy Cattle Breed Genet Comm Univ Guelph*. Guelph/ON - Canada; 2009.

15. Misztal I, Legarra A, Aguilar I: **Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information.** *J Dairy Sci* 2009, **92**:4648–4655.

16. VanRaden PM: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91**:4414–4423.

17. Legarra A, Aguilar I, Misztal I: **A relationship matrix including full pedigree and genomic information.** *J Dairy Sci* 2009, **92**:4656–4663.

18. Vitezica ZG, Aguilar I, Legarra A: **One-Step vs . Multi-Step Methods For Genomic Prediction In Presence Of Selection**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany; 2010.

19. Christensen OF, Lund MS: **Genomic prediction when some animals are not genotyped.** *Genet Sel Evol* 2010, **42**:2.

20. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score.** *J Dairy Sci* 2010, **93**:743–752.

21. Garrick DJ: **The Nature , Scope And Impact Of Some Whole- Genome Analyses In Beef Cattle**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany; 2010.

22. Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE: **Genome Wide Marker Assisted Selection in Chicken : Making the Most of All Data , Pedigree , Phenotypic , and Genomic in a Simple One Step Procedure**. In *Proc 9th World Congr Genet Appl to Livest Prod 1-6 August 2010*. Leipzig - Germany; 2010.

23. Legarra A, Christensen OF, Aguilar I, Misztal I: **Single Step, a general approach for genomic selection**. *Livest Sci* 2014, **166**:54–65.

24. Harris BL, Johnson DL: **Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation.** *J Dairy Sci* 2010, **93**:1243–1252.

25. Su G, Madsen P, Nielsen US, Mäntysaari EA, Aamand GP, Christensen OF, Lund MS: **Genomic prediction for Nordic Red Cattle using one-step and selection index blending.** *J Dairy Sci* 2012, **95**:909–917.

26. Koivula M, Strandén I, Su G, Mäntysaari EA: **Different methods to calculate genomic predictions--comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP).** *J Dairy Sci* 2012, **95**:4065–4073.

27. Harris BL, Winkelman AM, Johnson DL: **Impact of Including a Large Number of Female Genotypes on Genomic Selection**. In *Interbull Bull No 47 23-25 August 2013*. Nantes, France; 2013:23–27.

28. Přibyl J, Madsen P, Bauer J, Přibylová J, Simečková M, Vostrý L, Zavadilová L: **Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production.** *J Dairy Sci* 2013, **96**:1865–1873.

29. Neves HHR, Carvalheiro R, O'Brien AMP, Utsunomiya YT, do Carmo AS, Schenkel FS, Sölkner J, McEwan JC, Van Tassell CP, Cole JB, da Silva MVGB, Queiroz SA, Sonstegard TS, Garcia JF: **Accuracy of genomic predictions in Bos indicus (Nellore) cattle.** *Genet Sel Evol* 2014, **46**:17.

30. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH: **BLUPF90 AND RELATED PROGRAMS (BGF90)**. In *7th World Congr Genet Appl to Livest Prod 19-23 August 2002*. Montpellier, France; 2002.

31. VanRaden PM, Wiggans GR: **Derivation, calculation, and use of national animal model information.** *J Dairy Sci* 1991, **74**:2737–2746.

32. Falconer DS, MacKay TFC: *Introduction to Quantitative Genetics*. 4th edition. Harlow, Essex, UK: Longmans Green; 1966.

33. Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, S. D: **Genomic predictions in Angus cattle : Comparisons of sample size , response variables , and clustering methods for cross-validation**. *J Anim Sci* 2014, **92**:485–497.

34. Neves HHR, Carvalheiro R, Queiroz SA: **Adequacy of Different Pseudo-Phenotypes for Model Training and Validation of Genomic Predictions in Beef Cattle**. In *Proc 10th World Congr Genet Appl to Livest Prod 17-22 August 2014*. Vancouver, Canada; 2014.

35. Piccoli ML, Braccini J, Brito F V, Cobuci JA, Cardoso FF, Brito L, Sargolzaei M, Schenkel FS: **Genomic evaluation using single and two steps procedures in presence of selection in simulated data of beef cattle**. *BMC Genet* 2015, **submitted**.

36. Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel RD, Garrick DJ, Taylor JF: **Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation.** *Genet Sel Evol* 2011, **43**:40.

37. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819–1829.

38. Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME: **Accuracy of genomic breeding values in multi-breed dairy cattle populations.** *Genet Sel Evol* 2009, **41**:51.

39. Habier D, Fernando RL, Dekkers JCM: **The impact of genetic relationship information on genome-assisted breeding values.** *Genetics* 2007, **177**:2389–2397.

40. Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G: **Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population.** *Genet Sel Evol* 2012, **44**:8.

41. Vitezica ZG, Aguilar I, Misztal I, Legarra a: **Bias in genomic predictions for populations under selection.** *Genet Res (Camb)* 2011, **93**:357–66.

42. Madsen P, Jensen J, Labouriau R, Christensen OF, Sahana G: **DMU - A Package for Analyzing Multivariate Mixed Models in quantitative Genetics and Genomics**. In *Proc 8th World Congr Genet Appl to Livest Prod 13-18 August 2006*. Belo Horizonte/MG - Brazil; 2006.

**Table 1 Summary statistics of pedigree structure.**

| Pedigree structure | Number | Percentage |
|---|---|---|
| Individuals in total | 540,769 | 100.00 |
| | | |
| Sires in total | 2,528 | 0.47 |
| -Progeny | 177,509 | 32.83 |
| Dams in total | 178,342 | 32.98 |
| -Progeny | 424,050 | 78.42 |
| | | |
| Individuals with progeny | 180,870 | 33.45 |
| Individuals with no progeny | 359,899 | 66.55 |
| | | |
| individuals with both known parents | 177,470 | 32.82 |
| individuals with both unknown parents | 116,680 | 21.58 |
| individuals with sire unknown | 246,580 | 45.60 |
| | | |
| Individuals Braford | 383,513 | 70.92 |
| Individuals Hereford | 157,256 | 29.08 |
| | | |
| Individuals males | 214,469 | 39.66 |
| Individuals females | 326,300 | 60.34 |

**Table 2 Number of phenotype records, genotyped animals and heritability for each economic trait.**

| Trait[1] | $h^2$ | Training population | | | | | | | Prediction population | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Phenotype | Genotype | | | | Unknown[3] | | Phenotype | Genotype | | | | Unknown[3] | |
| | | | Total | Animal | Sire | Dam | Sire | Dam | | Total | Animal | Sire | Dam | Sire | Dam |
| WGBW | 0.25 | 325,648 | 2,325 | 2,231 | 91 | 3 | 1,786 | 108 | 354,253 | 980 | 944 | 9 | 27 | 588 | 36 |
| WGWY | 0.31 | 153,993 | 1,614 | 1,520 | 91 | 3 | 1,257 | 108 | 164,140 | 925 | 907 | 6 | 12 | 541 | 17 |
| CW | 0.25 | 319,518 | 1,990 | 1,896 | 91 | 3 | 1,677 | 108 | 348,018 | 980 | 944 | 9 | 27 | 588 | 36 |
| CY | 0.32 | 161,033 | 1,811 | 1,717 | 91 | 3 | 1,416 | 108 | 171,403 | 957 | 939 | 6 | 12 | 569 | 17 |
| PW | 0.25 | 301,810 | 1,981 | 1,887 | 91 | 3 | 1,668 | 108 | 330,310 | 980 | 944 | 9 | 27 | 588 | 36 |
| PY | 0.32 | 149,890 | 1,811 | 1,717 | 91 | 3 | 1,416 | 108 | 160,260 | 957 | 939 | 6 | 12 | 569 | 17 |
| MW | 0.25 | 301,558 | 1,988 | 1,894 | 91 | 3 | 1,675 | 108 | 330,057 | 980 | 944 | 9 | 27 | 588 | 36 |
| MY | 0.32 | 149,336 | 1,811 | 1,717 | 91 | 3 | 1,416 | 108 | 159,706 | 957 | 939 | 6 | 12 | 569 | 17 |
| SCa | 0.43 | 44,086 | 708 | 623 | 85 | 0 | 430 | 96 | 46,823 | 873 | 865 | 5 | 3 | 498 | 7 |
| SCaw | 0.43 | 44,086 | 708 | 623 | 85 | 0 | 430 | 96 | 46,823 | 873 | 865 | 5 | 3 | 498 | 7 |
| BW | 0.33 | 197,472 | 2,492 | 2,401 | 88 | 3 | 1,839 | 102 | 221,038 | 942 | 905 | 10 | 27 | 559 | 37 |
| SW | 0.25 | 122,980 | 1,932 | 1,848 | 81 | 3 | 1,624 | 90 | 140,681 | 979 | 944 | 8 | 27 | 587 | 35 |
| SY | 0.41 | 78,036 | 1,781 | 1,694 | 84 | 3 | 1,297 | 94 | 84,259 | 956 | 939 | 5 | 12 | 568 | 16 |
| NW | 0.46 | 238,607 | 2,382 | 2,291 | 89 | 2 | 1,758 | 104 | 265,800 | 961 | 927 | 7 | 27 | 581 | 34 |
| NY | 0.41 | 112,597 | 1,884 | 1,793 | 89 | 2 | 1,398 | 104 | 122,409 | 941 | 924 | 5 | 12 | 563 | 17 |
| HW | 0.23 | 91,328 | 700 | 612 | 86 | 2 | 655 | 102 | 110,162 | 879 | 846 | 6 | 27 | 518 | 33 |
| HY | 0.31 | 65,085 | 1,651 | 1,561 | 88 | 2 | 1,320 | 102 | 73,621 | 829 | 813 | 4 | 12 | 477 | 16 |
| OP | 0.20 | 117,460 | 678 | 587 | 90 | 1 | 639 | 106 | 139,082 | 957 | 921 | 9 | 27 | 571 | 36 |
| Average | 0.32 | 165,251 | 1,680 | 1,590 | 89 | 2 | 1,317 | 103 | 181,603 | 939 | 914 | 7 | 19 | 557 | 25 |

[1] WGBW: weight gain from birth to weaning (Kg); WGWY: weight gain from weaning to yearling (Kg); CW: conformation score at weaning (scores 1-5); CY: conformation score at yearling (scores 1-5); PW: precocity score at weaning (scores 1-5); PY: precocity score at yearling (scores 1-5); MW: muscularity score at weaning (scores 1-5); MY: muscularity score at yearling (scores 1-5); SCa: scrotal circumference adjusted for age at yearling (cm); SCaw: scrotal circumference adjusted for age and weight at yearling (cm); BW: birth weight (Kg); SW: size score at weaning (scores 1-5); SY: size score at yearling (scores 1-5); NW: prepuce (navel) score at weaning (scores 1-5); NY: prepuce (navel) score at yearling (scores 1-5); HW: hair length score at weaning (scores 1-3); HY: hair length score at yearling (scores 1-3); OP: ocular pigmentation score (scores 1-3);

[2] The heritability estimates were obtained prior to this study by DMU package [42];

[3] Number of genotyped animals with unknown sire (generally multiple sire matings) and/or dam.

**Table 3 Accuracies of DGV and regression coefficients on DGV in the prediction population.**

| Trait[1] | Regarding dEBV of prediction population | | | | | | | | Regarding EBV of prediction population | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r(EBV/dEBV,DGV)^2$ | | | | $b1(EBV/dEBV,DGV)^3$ | | | | $r(EBV/dEBV,DGV)^2$ | | | | $b1(EBV/dEBV,DGV)^3$ | | | |
| | $tsD^4$ | $tsE^4$ | $ssT^4$ | $ssTP^4$ | $tsD^4$ | $tsE^4$ | $ssT^4$ | $ssTP^4$ | $tsD^4$ | $tsE^4$ | $ssT^4$ | $ssTP^4$ | $tsD^4$ | $tsE^4$ | $ssT^4$ | $ssTP^4$ |
| WGBW | 0.12 | 0.13 | 0.13 | 0.18 | 0.68 | 0.98 | 0.53 | 0.67 | 0.25 | 0.32 | 0.23 | 0.40 | 0.44 | 0.80 | 0.30 | 0.46 |
| WGWY | 0.08 | 0.09 | 0.09 | 0.13 | 0.61 | 0.85 | 0.46 | 0.57 | 0.27 | 0.30 | 0.25 | 0.38 | 0.58 | 0.80 | 0.39 | 0.51 |
| CW | 0.09 | 0.08 | 0.08 | 0.11 | 0.40 | 0.44 | 0.27 | 0.34 | 0.34 | 0.40 | 0.28 | 0.43 | 0.55 | 0.82 | 0.35 | 0.50 |
| CY | 0.18 | 0.22 | 0.19 | 0.26 | 0.81 | 1.36 | 0.69 | 0.81 | 0.33 | 0.42 | 0.32 | 0.49 | 0.67 | 1.16 | 0.50 | 0.67 |
| PW | 0.10 | 0.08 | 0.08 | 0.10 | 0.46 | 0.50 | 0.28 | 0.32 | 0.30 | 0.38 | 0.24 | 0.36 | 0.50 | 0.81 | 0.29 | 0.40 |
| PY | 0.25 | 0.26 | 0.25 | 0.28 | 1.07 | 1.54 | 0.84 | 0.86 | 0.39 | 0.46 | 0.36 | 0.50 | 0.73 | 1.17 | 0.53 | 0.66 |
| MW | 0.14 | 0.11 | 0.14 | 0.17 | 0.61 | 0.61 | 0.45 | 0.51 | 0.37 | 0.44 | 0.31 | 0.44 | 0.55 | 0.84 | 0.37 | 0.48 |
| MY | 0.24 | 0.27 | 0.25 | 0.29 | 0.94 | 1.37 | 0.77 | 0.86 | 0.40 | 0.47 | 0.38 | 0.53 | 0.71 | 1.11 | 0.54 | 0.70 |
| SCa | 0.10 | 0.11 | 0.09 | 0.15 | 0.52 | 0.81 | 0.35 | 0.52 | 0.17 | 0.20 | 0.14 | 0.26 | 0.44 | 0.73 | 0.27 | 0.47 |
| SCaw | 0.11 | 0.10 | 0.10 | 0.15 | 0.63 | 0.87 | 0.42 | 0.57 | 0.16 | 0.17 | 0.13 | 0.25 | 0.45 | 0.70 | 0.28 | 0.47 |
| BW | 0.15 | 0.15 | 0.14 | 0.14 | 0.82 | 1.23 | 0.57 | 0.54 | 0.21 | 0.24 | 0.19 | 0.25 | 0.46 | 0.81 | 0.30 | 0.40 |
| SW | 0.19 | 0.17 | 0.20 | 0.24 | 1.04 | 1.27 | 0.84 | 0.78 | 0.33 | 0.39 | 0.33 | 0.50 | 0.67 | 1.10 | 0.51 | 0.59 |
| SY | 0.25 | 0.29 | 0.27 | 0.33 | 1.03 | 1.74 | 0.82 | 0.85 | 0.34 | 0.42 | 0.35 | 0.51 | 0.70 | 1.27 | 0.54 | 0.66 |
| NW | 0.40 | 0.40 | 0.40 | 0.42 | 1.17 | 1.50 | 0.87 | 0.85 | 0.50 | 0.51 | 0.49 | 0.55 | 0.78 | 1.04 | 0.57 | 0.60 |
| NY | 0.37 | 0.33 | 0.37 | 0.40 | 1.45 | 1.74 | 1.14 | 1.18 | 0.46 | 0.45 | 0.45 | 0.51 | 0.88 | 1.15 | 0.67 | 0.74 |
| HW | 0.21 | 0.18 | 0.22 | 0.24 | 1.23 | 1.58 | 1.04 | 0.82 | 0.33 | 0.33 | 0.31 | 0.48 | 0.61 | 0.90 | 0.48 | 0.54 |
| HY | 0.17 | 0.17 | 0.18 | 0.20 | 0.69 | 1.07 | 0.50 | 0.54 | 0.28 | 0.33 | 0.27 | 0.42 | 0.45 | 0.78 | 0.29 | 0.42 |
| OP | 0.28 | 0.28 | 0.28 | 0.35 | 1.09 | 1.27 | 0.93 | 0.82 | 0.53 | 0.55 | 0.52 | 0.66 | 0.85 | 1.02 | 0.70 | 0.63 |
| Average | 0.19 | 0.19 | 0.19 | 0.23 | 0.85 | 1.15 | 0.65 | 0.69 | 0.33 | 0.38 | 0.31 | 0.44 | 0.61 | 0.95 | 0.44 | 0.55 |
| Anova[5] | a | a | a | a | ns | ns | * | * | b | ab | b | a | * | ns | * | * |

[1] WGBW: weight gain from birth to weaning (kg); WGWY: weight gain from weaning to yearling (kg); CW: conformation score at weaning (scores 1-5); CY: conformation score at yearling (scores 1-5); PW: precocity score at weaning (scores 1-5); PY: precocity score at yearling (scores 1-5); MW: muscularity score at weaning (scores 1-5); MY: muscularity score at yearling (scores 1-5); SCa: scrotal circumference adjusted for age (cm); SCaw: scrotal circumference adjusted for age and weight yearling (cm); BW: birth weight (kg); SW: size score at weaning (scores 1-5); SY: size score at yearling (scores 1-5); NW: prepuce (navel) score at weaning (scores 1-5); NY: prepuce (navel) score at yearling (scores 1-5); HW: hair length score at weaning (scores 1-3); HY: hair length score at yearling (scores 1-3); OP: ocular pigmentation score (scores 1-3);

[2] Accuracies measured as the Pearson's correlation r(dEBV,DGV) and r(EBV,DGV) of the animals in the prediction population;

[3] Inflation of genomic predictions measured by the slope of the regression b1(dEBV, DGV) and b1(EBV, DGV) in the prediction population;

[4] DGVs predictions wherein: (ts): DGVs predicted by two-step procedure; (ss): DGVs predicted by single-step procedure; (D): DGVs predicted by dEBV in the training population; (E): DGVs predicted by EBV in the training population;

[5] Different letters within a group means r(dEBV,DGV), r(EBV,DGV) that there is a statistical difference between two means (P<0.05). Within a group means b1(dEBV,DGV), b1(EBV,DGV), "ns" refers to statistically no different 1.00 (P>0.05) and "*" refers to statistically different 1.00 (P<0.05).

**Table 4 Accuracies of GEBV and regression coefficients on GEBV in the prediction population.**

| Trait[1] | Regarding dEBV of prediction population | | | | | | | | | | Regarding EBV of prediction population | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r(EBV/dEBV,GEBV)$[2] | | | | | $b1(EBV/dEBV,GEBV)$[3] | | | | | $r(EBV/dEBV,DGV)$[2] | | | | | $b1(EBV/dEBV,GEBV)$[3] | | | | |
| | tsDv[4] | tsDh[4] | tsEv[4] | tsEh[4] | ss[4] | tsDv[4] | tsDh[4] | tsEv[4] | tsEh[4] | ss[4] | tsDv[4] | tsDh[4] | tsEv[4] | tsEh[4] | ss[4] | tsDv[4] | tsDh[4] | tsEv[4] | tsEh[4] | ss[4] |
| WGBW | 0.27 | 0.23 | 0.28 | 0.23 | 0.18 | 1.19 | 1.42 | 1.42 | 1.57 | 0.67 | 0.61 | 0.56 | 0.67 | 0.61 | 0.41 | 0.83 | 1.07 | 1.06 | 1.33 | 0.48 |
| WGWY | 0.35 | 0.27 | 0.37 | 0.29 | 0.13 | 1.86 | 1.95 | 2.16 | 2.25 | 0.58 | 0.65 | 0.58 | 0.69 | 0.61 | 0.38 | 1.03 | 1.25 | 1.20 | 1.43 | 0.53 |
| CW | 0.23 | 0.20 | 0.23 | 0.19 | 0.11 | 0.74 | 0.82 | 0.79 | 0.83 | 0.35 | 0.71 | 0.69 | 0.76 | 0.72 | 0.47 | 0.84 | 1.08 | 0.98 | 1.20 | 0.54 |
| CY | 0.44 | 0.38 | 0.47 | 0.40 | 0.27 | 1.38 | 1.59 | 1.61 | 1.81 | 0.83 | 0.73 | 0.68 | 0.79 | 0.73 | 0.51 | 1.01 | 1.28 | 1.19 | 1.47 | 0.70 |
| PW | 0.23 | 0.19 | 0.22 | 0.18 | 0.10 | 0.82 | 0.93 | 0.88 | 0.89 | 0.33 | 0.66 | 0.64 | 0.72 | 0.68 | 0.39 | 0.83 | 1.08 | 1.00 | 1.21 | 0.43 |
| PY | 0.46 | 0.39 | 0.47 | 0.40 | 0.29 | 1.46 | 1.67 | 1.67 | 1.86 | 0.88 | 0.74 | 0.69 | 0.79 | 0.73 | 0.52 | 1.01 | 1.26 | 1.20 | 1.47 | 0.69 |
| MW | 0.26 | 0.22 | 0.24 | 0.19 | 0.16 | 0.88 | 0.95 | 0.89 | 0.87 | 0.51 | 0.69 | 0.67 | 0.74 | 0.70 | 0.46 | 0.83 | 1.03 | 0.98 | 1.15 | 0.51 |
| MY | 0.46 | 0.40 | 0.49 | 0.42 | 0.30 | 1.33 | 1.54 | 1.55 | 1.74 | 0.88 | 0.74 | 0.71 | 0.80 | 0.75 | 0.55 | 0.97 | 1.22 | 1.16 | 1.42 | 0.73 |
| SCa | 0.53 | 0.42 | 0.57 | 0.45 | 0.15 | 1.86 | 2.14 | 2.18 | 2.54 | 0.55 | 0.69 | 0.60 | 0.75 | 0.64 | 0.28 | 1.20 | 1.51 | 1.42 | 1.80 | 0.50 |
| SCaw | 0.57 | 0.44 | 0.61 | 0.46 | 0.15 | 2.17 | 2.51 | 2.58 | 3.01 | 0.60 | 0.70 | 0.59 | 0.76 | 0.63 | 0.26 | 1.31 | 1.65 | 1.57 | 2.00 | 0.50 |
| BW | 0.38 | 0.31 | 0.40 | 0.33 | 0.14 | 1.63 | 1.89 | 2.07 | 2.37 | 0.55 | 0.57 | 0.52 | 0.64 | 0.57 | 0.26 | 1.00 | 1.25 | 1.32 | 1.64 | 0.42 |
| SW | 0.37 | 0.33 | 0.36 | 0.31 | 0.24 | 1.25 | 1.58 | 1.34 | 1.60 | 0.78 | 0.76 | 0.73 | 0.80 | 0.76 | 0.52 | 0.96 | 1.28 | 1.10 | 1.43 | 0.62 |
| SY | 0.48 | 0.44 | 0.50 | 0.45 | 0.33 | 1.33 | 1.74 | 1.58 | 2.01 | 0.87 | 0.70 | 0.66 | 0.75 | 0.70 | 0.52 | 0.97 | 1.33 | 1.18 | 1.59 | 0.68 |
| NW | 0.53 | 0.50 | 0.54 | 0.49 | 0.41 | 1.27 | 1.66 | 1.47 | 1.87 | 0.87 | 0.69 | 0.66 | 0.72 | 0.67 | 0.55 | 0.89 | 1.19 | 1.06 | 1.39 | 0.62 |
| NY | 0.57 | 0.49 | 0.57 | 0.48 | 0.40 | 1.79 | 2.14 | 2.10 | 2.38 | 1.20 | 0.71 | 0.65 | 0.73 | 0.65 | 0.52 | 1.07 | 1.36 | 1.31 | 1.58 | 0.76 |
| HW | 0.38 | 0.32 | 0.41 | 0.34 | 0.21 | 1.71 | 2.07 | 1.92 | 2.35 | 0.84 | 0.67 | 0.62 | 0.74 | 0.67 | 0.43 | 0.96 | 1.25 | 1.14 | 1.50 | 0.55 |
| HY | 0.45 | 0.29 | 0.49 | 0.39 | 0.08 | 1.25 | 1.49 | 1.58 | 1.83 | 0.56 | 0.73 | 0.52 | 0.81 | 0.72 | 0.17 | 0.82 | 1.07 | 1.09 | 1.39 | 0.44 |
| OP | 0.47 | 0.42 | 0.48 | 0.43 | 0.35 | 1.24 | 1.55 | 1.34 | 1.65 | 0.83 | 0.81 | 0.79 | 0.83 | 0.80 | 0.66 | 0.87 | 1.18 | 0.95 | 1.27 | 0.64 |
| Average | 0.41 | 0.35 | 0.43 | 0.36 | 0.22 | 1.40 | 1.65 | 1.62 | 1.86 | 0.70 | 0.70 | 0.64 | 0.75 | 0.69 | 0.44 | 0.97 | 1.24 | 1.16 | 1.46 | 0.57 |
| Anova[5] | a | a | a | a | b | ns | * | * | * | ns | ab | b | a | ab | c | ns | * | ns | * | * |

[1] WGBW: weight gain from birth to weaning (kg); WGWY: weight gain from weaning to yearling (kg); CW: conformation score at weaning (scores 1-5); CY: conformation score at yearling (scores 1-5); PW: precocity score at weaning (scores 1-5); PY: precocity score at yearling (scores 1-5); MW: muscularity score at weaning (scores 1-5); MY: muscularity score at yearling (scores 1-5); SCa: scrotal circumference adjusted for age (cm); SCaw: scrotal circumference adjusted for age and weight yearling (cm); BW: birth weight (kg); SW: size score at weaning (scores 1-5); SY: size score at yearling (scores 1-5); NW: prepuce (navel) score at weaning (scores 1-5); NY: prepuce (navel) score at yearling (scores 1-5); HW: hair length score at weaning (scores 1-3); HY: hair length score at yearling (scores 1-3); OP: ocular pigmentation score (scores 1-3);

[2] Accuracies measured as the Pearson's correlation r(dEBV,GEBV) and r(EBV,GEBV) of the animals in the prediction population;

[3] Inflation of genomic predictions measured by the slope of the regression b1(dEBV, GEBV) and b1(EBV, GEBV) in the prediction population;

[4] GEBVs predictions wherein: (ts): GEBVs predicted by two-step procedure; (ss): GEBVs predicted by single-step procedure; (D): GEBVs predicted by dEBV in the training population; (E): GEBVs predicted by EBV in the training population; (v): GEBVs predicted by VanRaden blending method [16] in two-step procedure; (h): GEBVs predicted by Hayes blending method [1] in two-step procedure;

[5] Different letters within a group means r(dEBV,GEBV), r(EBV, GEBV) that there is a statistical difference between two means (P<0.05). Within a group means b1(dEBV,GEBV), b1(EBV,GEBV), "ns" refers to statistically no different 1.00 (P>0.05) and "*" refers to statistically different 1.00 (P<0.05).

**Table 5 Percentage of mismatch between 20% best animals in the prediction population.**

| Trait[1] | Regarding EBV | | | | | | | Regarding EBV$_{PA}$ | | | | | | |
| | EBV$_{PA}$ | DGV | | GEBV[2] | | GEBV[3] | | EBV | DGV | | GEBV[2] | | GEBV[3] | |
| | | tsD[4] | ssTP[4] | tsDv[5] | ssTP[5] | tsDv[5] | ssP[5] | | tsD[4] | ssTP[4] | tsDv[5] | ssTP[5] | tsDv[5] | ssP[5] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WG BW | 45 | 66 | 61 | 51 | 60 | 19 | 29 | 38 | 62 | 52 | 33 | 50 | 41 | 51 |
| WG WY | 42 | 72 | 60 | 44 | 58 | 18 | 23 | 39 | 54 | 35 | 18 | 31 | 41 | 41 |
| CW | 29 | 63 | 54 | 34 | 50 | 21 | 29 | 28 | 59 | 48 | 17 | 41 | 33 | 43 |
| CY | 35 | 60 | 46 | 31 | 43 | 11 | 24 | 31 | 55 | 40 | 17 | 33 | 34 | 43 |
| PW | 33 | 67 | 60 | 44 | 59 | 18 | 33 | 35 | 59 | 55 | 31 | 51 | 42 | 53 |
| PY | 39 | 60 | 48 | 40 | 46 | 17 | 26 | 28 | 53 | 38 | 22 | 36 | 37 | 40 |
| MW | 35 | 62 | 57 | 43 | 55 | 21 | 40 | 35 | 52 | 50 | 27 | 48 | 41 | 56 |
| MY | 28 | 62 | 49 | 34 | 48 | 17 | 28 | 26 | 53 | 43 | 24 | 40 | 32 | 43 |
| SCa | 43 | 75 | 62 | 43 | 62 | 14 | 25 | 31 | 64 | 52 | 20 | 49 | 33 | 41 |
| SCa w | 46 | 72 | 66 | 42 | 65 | 13 | 28 | 41 | 69 | 62 | 28 | 59 | 41 | 46 |
| BW | 53 | 73 | 70 | 59 | 70 | 17 | 30 | 44 | 71 | 58 | 39 | 57 | 47 | 52 |
| SW | 31 | 57 | 48 | 34 | 48 | 13 | 23 | 29 | 54 | 48 | 23 | 46 | 33 | 40 |
| SY | 38 | 61 | 49 | 41 | 49 | 14 | 22 | 32 | 57 | 37 | 28 | 36 | 39 | 44 |
| NW | 40 | 61 | 60 | 50 | 59 | 26 | 36 | 27 | 56 | 52 | 39 | 50 | 38 | 45 |
| NY | 39 | 59 | 57 | 44 | 57 | 14 | 25 | 23 | 57 | 50 | 37 | 49 | 25 | 33 |
| HW | 35 | 65 | 53 | 34 | 52 | 17 | 30 | 25 | 56 | 38 | 21 | 36 | 27 | 38 |
| HY | 34 | 66 | 63 | 42 | 60 | 24 | 39 | 36 | 59 | 50 | 25 | 46 | 36 | 41 |
| OP | 13 | 55 | 40 | 28 | 38 | 33 | 39 | 25 | 49 | 40 | 25 | 38 | 39 | 44 |
| Average | 37 | 64 | 56 | 41 | 54 | 18 | 29 | 32 | 58 | 47 | 26 | 44 | 37 | 44 |

[1] WGBW: weight gain from birth to weaning (kg); WGWY: weight gain from weaning to yearling (kg); CW: conformation score at weaning (scores 1-5); CY: conformation score at yearling (scores 1-5); PW: precocity score at weaning (scores 1-5); PY: precocity score at yearling (scores 1-5); MW: muscularity score at weaning (scores 1-5); MY: muscularity score at yearling (scores 1-5); SCa: scrotal circumference adjusted for age (cm); SCaw: scrotal circumference adjusted for age and weight yearling (cm); BW: birth weight (kg); SW: size score at weaning (scores 1-5); SY: size score at yearling (scores 1-5); NW: prepuce (navel) score at weaning (scores 1-5); NY: prepuce (navel) score at yearling (scores 1-5); HW: hair length score at weaning (scores 1-3); HY: hair length score at yearling (scores 1-3); OP: ocular pigmentation score (scores 1-3);

[2] $GEBV^2$ blend made with $EBV_{PA;}$

[3] $GEBV^3$ blend made with EBV;

[4] DGVs predictions wherein: (ts): DGVs predicted by two-step procedure; (ss): DGVs predicted by single-step procedure; (D): DGVs predicted by dEBV in the training population; (TP): DGVs predicted by all genotypes and phenotypes by training population;

[5] GEBVs predictions wherein: (ts): GEBVs predicted by two-step procedure; (ss): GEBVs predicted by single-step procedure; (D): GEBVs predicted by dEBV in the training population; (v): GEBVs predicted by VanRaden blending method [16] in two-step procedure; (TP): GEBVs predicted by single-step with $EBV_{PA}$; (P) GEBVs predicted by single-step with EBV.

**CAPÍTULO VII**

# CONSIDERAÇÕES FINAIS

Existe um grande potencial para o uso dos marcadores moleculares em programas de melhoramento genético animal em todo o mundo, principalmente impulsionado pela possibilidade de seleção dos animais a idades jovens (biópsia de embriões antes de serem implantados) acumulando maiores ganhos genéticos e diminuindo os custos dos testes de progênie (mais comum em bovinos de leite) e possibilitando a seleção de caracteres de importância econômica de difícil mensuração, tais como, consumo residual alimentar, resistência a doenças e a parasitas e características que necessitam o abate dos animais. Com base nestes aspectos foram estimados diversos parâmetros que darão suporte a estudos de viabilidade da aplicação da seleção genômica em rebanhos de bovinos de corte.

A análise de pedigree e estrutura populacional realizada nas raças Angus, Devon, Hereford e Shorthorn apontam para um ligeiro aumento da endogamia nas quatro raças britânicas estudadas, com estimativas de tamanho efetivo da população, indicando que a diversidade genética está sendo mantida em níveis razoáveis em todas as raças, acima do limiar crítico mínimo recomendado (FAO,1998), muito provavelmente devido ao fluxo continuo de importação de material genético (sêmen) de outros países.

Os parâmetros estudados com relação a imputação de genótipos de painéis de baixa densidade (3K, 6K, 8K, 15K e 20K) para o painel de 50K e dos painéis de baixa e média densidade (3K, 6K, 8K, 15K, 20K, 50K, 90iK e 90tK) para o painel de 777K indicaram que, com exceção do painel de 3K, todos os demais painéis poderiam ser utilizados como base visando a imputação para o painel de 50K e também que os painéis de 50K, 90iK e 90tK, poderiam ser utilizados como base na imputação para o painel de 777K, viabilizando o uso dos diferentes painéis na seleção genômica. O uso dos painéis imputados na seleção genômica (testes realizados com os painéis de 8K e 15K) em relação ao painel original de 50K, mostraram que não houveram diferenças em acurácia.

As análises mostraram que o uso dos marcadores moleculares na seleção genômica produziram ganhos em acurácia em relação à seleção sem o uso dos marcadores moleculares e que o procedimento de passo único ou de multi passo se equivalem na predição dos valores genéticos genômicos.

Um fato importante a ser considerado na implementação da seleção genômica é a genotipagem de animais (touros e vacas) largamente utilizados nos programas de melhoramento genético animal possibilitando criar uma população de treinamento que apresente acurácias elevadas, critério importante para obter melhores predições dos marcadores moleculares. Para este aspecto, é primordial que se crie um banco de material biológico, principalmente, com os animais mais importantes dos programas de melhoramento. Os custos da genotipagem, mesmo com painéis de baixa densidade, ainda podem ser elevados para muitos produtores. No entanto, espera-se uma diminuição destes custos com o passar do tempo, fato que se observa com outras tecnologias. O desenvolvimento de painéis com marcadores moleculares específicos para determinada raça ou programa de melhoramento genético, tem mostrado melhores resultados, tanto na imputação, visando a reconstrução de painéis mais densos, como nos

resultados da aplicação da seleção genômica medidos pelo valor das acurácias. No mundo globalizado de hoje, é imprescindível que se formem parcerias entre universidades, institutos de pesquisa, programas de melhoramento e entre países, com o intuito de partilhar os genótipos já coletados e direcionar as novas genotipagens, certamente promovendo maior benefício e qualidade nas pesquisas, que resultarão em maiores ganhos no agronegócio, que é o objetivo final.

O uso da seleção genômica em programas de melhoramento genético animal no Brasil, semelhantes ao programa da Conexão DeltaG, será importante principalmente: a) quando a seleção for praticada para caracteres de difícil mensuração e/ou de custo elevado, como por exemplo, resistência a ectoparasitas e consumo alimentar residual, b) para maximizar futuras produções pela utilização de touros e vacas de elevado valor genético e acurácia; c) para melhorar as informações de genealogia e por consequência as predições dos valores genéticos, visto que em média 50% dos animais são filhos de reprodutores múltiplos. A imputação de genótipos também terá um papel decisivo na seleção genômica dos bovinos de corte no Brasil. Esta técnica possibilitará: a) que os animais jovens sejam genotipados com painéis menos densos e, portanto, mais econômicos dentro do sistema de produção e após a reconstrução para painéis mais densos e mais apropriados para a seleção genômica; b) a reconstrução dos genótipos de animais que já possuem filhos genotipados, como por exemplo, animais importantes que não se dispunha de material biológico, bem como das vacas com várias crias.

Considerando os resultados desta pesquisa, seria importante a realização de estudos adicionais quanto ao desenvolvimento e uso de painéis específicos para raças e/ou programas de melhoramento genético; quanto a formação da população de treinamento, esta devendo ser constituida por mais animais e, principalmente, de maior acurácia e composta por diferentes raças; e quanto a imputação de genótipos de animais não genotipados e o uso destes na seleção genômica.

# REFERÊNCIAS BIBLIOGRÁFICAS

AGUILAR, I. et al. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Champaign, v. 93, n. 2, p. 743–752, 2010.

ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Nature Reviews. Genetics**, London, v. 3, n. 4, p. 299–309, 2002.

BERNARDO, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. **Crop Science**, Madison, v. 48, p. 1649–1664, 2008.

BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. S. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, London, v. 11, p. 421, 2010.

BOICHARD, D.; MAIGNEL, L.; VERRIER, E. The value of using probabilities of gene origin to measure genetic variability in a population. **Genetics Selection Evolution**, Paris, v. 29, n. 1, p. 5–23, 1997.

BRITO, F. V et al. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. **BMC Genetics**, London, v. 12, n. 1, p. 80, 2011.

BRITO, F. V. et al. In-depth pedigree analysis in a large Brazilian Nellore herd. **Genetics and Molecular Research**, Ribeirão Preto, v. 12, n. 4, p. 5758–5765, 2013.

BROWN W L. Genetic Diversity and Genetic Vulnerability-An Appraisal. **Economic Botany**, New York, v. 37, n. 1, p. 4–12, 1983.

BROWNING, B. L.; BROWNING, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. **American Journal of Human Genetics**, Amsterdam, v. 84, n. 2, p. 210–223, fev. 2009.

BROWNING, S. R.; BROWNING, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. **American Journal of Human Genetics**, Amsterdam, v. 81, n. 5, p. 1084–1097, nov. 2007.

CALUS, M. P. L. et al. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, Pittsburg, v. 178, n. 1, p. 553–561, 2008.

CARILLIER, C. et al. A first step toward genomic selection in the multi-breed French dairy goat population. **Journal of Dairy Science**, Champaign, v. 96, n. 11, p. 7294–7305, 2013.

CASAS, E. et al. Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. **Journal of Animal Science**, Champaign, v. 78, n. 3, p. 560–569, 2000.

CLEVELAND, M. A. et al. Changes in inbreeding of U . S . Herefords during the twentieth century. **Journal of Animal Science**, Champaign, v. 83, n. 5, p. 992–1001, 2005.

DASSONNEVILLE, R. et al. Short communication: Imputation performances of 3 low-density marker panels in beef and dairy cattle. **Journal of Dairy Science**, Champaign, v. 95, n. 7, p. 4136–4140, 2012.

DE ROOS, A. P. W. et al. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. **Genetics**, Pittsburg, v. 179, n. 3, p. 1503–1512, 2008.

DEKKERS, J. C. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, Champaign, v. 82 E-Suppl, 2004.

DRUET, T.; SCHROOTEN, C.; DE ROOS, A. P. W. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. **Journal of Dairy Science**, Champaign, v. 93, n. 11, p. 5443–5454, 2010.

ERBE, M. et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. **Journal of Dairy Science**, Champaign, v. 95, n. 7, p. 4114–4129, 2012.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Harlow: Pearson, 1996.

FAO. **Secondary guidelines for development of national farm animal genetic resources management plans: m**anagement of small populations at risk. Rome: FAO, 1998.

FAO. **In vivo conservation of animal genetic resources. Rome:** FAO, 2013. (FAO Animal Production and Health Guidelines. N. 14)

GODDARD, M. Genomic selection: prediction of accuracy and maximization of long term response. **Genetica**, Gif-sur-Yvette, v. 136, n. 2, p. 1–2, 2009.

GODDARD, M. E.; HAYES, B. J.; MCPARTLAN, H. Can the same genetic markers be used in multiple breeds? In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION. 8., 2006, Belo Horizonte. **Proceedings**... Belo Horizonte: Instituto Prociência, 2006. 1 CD-ROM.

HAYES, B. J. et al. Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. **Genome Research**, Cold Spring Harbor, v.13, n. 4, p. 635–643, 2003.

HAYES, B. J. et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. **Genetics Selection Evolution**, Paris, v. 41, p. 51, 2009a.

HAYES, B. J. et al. Invited review: Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, n. 2, p. 433–443, 2009b.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v. 31, n. 2, p. 423–447, 1975.

HICKEY, J. M. et al. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. **Genetics Selection Evolution**, Paris, v. 43, n. 1, p. 12, 2011.

HILL, W. C.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Stuttgart, v. 38, n. 6, p. 226–231, 1968.

KACHMAN, S. D. et al. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. **Genetics Selection Evolution**, Paris, v. 45, n. 1, p. 30, 2013.

KONG et al. Detection of sharing by descent , long range phasing and haplotype PubMed Commons. **Nature Genetics**, London, v. 40, n. 9, p. 1068–1075, 2008.

LACY, R. C. Analysis of founder representations in pedigrees: Founder equivalents and founder genome equivalents. **Zoo Biology**, Malden, v. 8, n. 2, p. 111–123, 1989.

LANDER, E. E. S.; SCHORK, N. J. N. Genetic dissection of complex traits. **Science**, Washington, v. 265, n. 5181, p. 2037–2048, 1994.

LEROY, G. et al. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. **Genetics Selection Evolution**, Paris, v. 45, n. 1, p. 1, 2013.

LU, D. et al. Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. **Frontiers in Genetics**, Lausanne, v. 3, n. 152, p. 1–10, 2012.

MACCLUER, J. W. et al. Pedigree analysis by computer simulation. **Zoo Biology**, Malden, v. 5, n. 2, p. 147–160, 1986.

MÁRQUEZ, G. C. et al. Genetic diversity and population structure of American Red Angus cattle. **Journal of Animal Science**, Champaign, v. 88, n. 1, p. 59–68, 2010.

MC PARLAND, S. et al. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. **Journal of Animal Science**, Champaign, v. 85, n. 2, p. 322–331, 2007.

MEUWISSEN, T. H. E. Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. **Genetics Selection Evolution**, Paris, v. 41, n. 45, p. 35, 2009.

MEUWISSEN, T. H. E.; WOOLLIAMS, J. A. Effective sizes of livestock populations to prevent a decline in fitness. **Theoretical and Applied Genetics**, Stuttgart, v. 89, p. 1019–1026, 1994.

MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Pittsburg, v. 157, n. 4, p. 1819–1829, 2001.

MOGHADDAR, N.; SWAN, A. A.; VAN DER WERF, J. H. J. Genomic prediction of weight and wool traits in a multi-breed sheep population. **Animal Production Science**, Clayton South, v. 54, n. 5, p. 544, 2014.

NEI, M. Analysis of gene diversity in subdivided populations. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 70, n. 12, p. 3321–3323, 1973.

PICCOLI, M. L. et al. Origins and genetic diversity of British cattle breeds in Brazil assessed by pedigree analyses. **Journal of Animal Science**, Champaign, v. 92, n. 5, p. 1920–1930, 2014a.

PICCOLI, M. L. et al. Accuracy of genome-wide imputation in Braford and Hereford beef cattle. **BMC genetics**, London, v. 15 n. 157, p. 1-15, 2014b.

RIQUET, J. et al. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 96, n. 16, p. 9252–9257, 1999.

SANTANA, M. L. et al. Pedigree analysis and inbreeding depression on growth traits in Brazilian Marchigiana and Bonsmara breeds. **Journal of Animal Science**, Champaign, v. 90, n. 1, p. 99–108, 2012.

SARGOLZAEI, M. et al. Extent of linkage disequilibrium in Holstein cattle in North America. **Journal of Dairy Science**, Champaign, v. 91, n. 5, p. 2106–2117, 2008.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC genomics**, London, v. 15, n. 1, p. 478, 2014.

SARGOLZAEI, M.; SCHENKEL, F. S.; CHESNAIS, J. Impact of amount of dam genotypic information on family-based imputation accuracy. In: DAIRY CATTLE BREEDING AND GENETICS COMMITTEE MEETING, 2010, Guelph. **Proceedings**...Guelph, Canada, 2010.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, Malden, v. 123, n. 4, p. 218–223, 2006.

SCHENKEL, F. S. et al. Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle. **Journal of Animal Science**, Champaign, v. 84, p. 291–299, 2006.

SCHENKEL, F. S. et al. Reliability of Genomic Evaluation of Holstein Cattle in Canada. In: INTERBULL INTERNATIONAL WORKSHOP, GENOMIC INFORMATION IN GENETIC EVALUATIONS, 2009, Uppsala, **Proceedings**... Uppsala, Sweden, 2009.

SØRENSEN, A. C.; SØRENSEN, M. K.; BERG, P. Inbreeding in Danish dairy cattle breeds. **Journal of Dairy Science**, Champaign, v. 88, n. 5, p. 1865–1872, 2005.

THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The genome sequence of taurine cattle: a window to ruminant biology and evolution **Science,** Washington, v.324, n. 5926, p.522-528, 2009.

VANRADEN, P. M. Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. **Journal of Dairy Science**, Champaign, v. 75, n. 11, p. 3136–3144, 1992.

VANRADEN, P. M. et al. Genomic evaluations with many more genotypes. **Genetics Selection Evolution**, Paris, v. 43, n. 1, p. 10, 2011.

WEBER, K.; THALLMAN, R. Accuracy of genomic breeding values in multi-breed beef cattle populations derived from deregressed breeding values and phenotypes. **Journal of Animal Science**, Champaign, v. 90, n. 12, p. 4177–4190, 2012.

WRIGHT, S. Coefficients of inbreeding and relationship. **The American Naturalist**, Chicago, v. 56, n. 645, p. 330–338, 1922.

WRIGHT, S. Evolution i n mendelian populations. **Genetics**, Pittsburg, v. 16, n. 2, p. 97–159, 1931.

LI, Y. et al. Genotype Imputation. **Annual Review of of Genomics and Human Genetics**, Palo Alto, v. 10, p. 387–406, 2009.

ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **Journal of Dairy Science**, Champaign, v. 93, n. 11, p. 5487–94, 2010.

ZHAO, H. et al. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. **Genetical Research**, MacKay, v. 86, n. 1, p. 77–87, 2005.

ZHAO, H.; NETTLETON, D.; DEKKERS, J. C. M. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. **Genetical Research**, MacKay, v. 89, n. 1, p. 1–6, 2007.

## VITA

Mario Luiz Piccoli, filho de Lilia Ribeiro Piccoli e José Piccoli, nasceu aos 05 dias do mês de maio de 1964 na cidade de São Marcos-RS.

Freqüentou da 1ª à 4ª série do ensino fundamental no Grupo Escolar Maranhão e da 5ª a 8ª série do ensino fundamental no Ginásio Estadual de São Marcos, ambos em São Marcos-RS. No ensino médio frequentou o Colégio Nossa Senhora do Carmo em Caxias do Sul-RS e realizou curso profissionalizante de Técnico em Contabilidade.

Ingressou no curso de Médicina Veterinária da Universidade Federal de Pelotas no ano de 1984. Cumpriu estágio curricular no Centro Rural Universitário de Treinamento e Ação Social - CRUTAC da Universidade Federal de Pelotas-RS e na Cooperativa Agropecuária Caxiense no setor de bovinos de leite em Caxias do Sul-RS. Realizou outros estágios durante o período acadêmico junto ao Serviço de Plantão do Hospital de Clínicas Veterinário da Universidade Federal de Pelotas-RS, na Suinocultura Eurotec Ltda em Caxias do Sul-RS, na Frangosul S/A nas sedes de Salvador do Sul-RS e de Caxias do Sul-RS, formou-se Médico Veterinário em julho de 1988.

Ingressou no curso de mestrado em Zootecnia da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul, em Porto Alegre-RS, área de produção animal, em março de 1989, sob a orientação do professor Luiz Alberto Fries, e obteve o grau de Mestre em Zootecnia em março de 1992.

Fundou a empresa GenSys Consultores Associados S/S em 1991, onde exerce atividades de consultoria e assessoria na área de melhoramento genético animal, estando temporariamente afastado em razão do curso de doutorado em Zootecnia.

Em março de 2011 ingressou no curso de Doutorado em Zootecnia na Universidade Federal do Rio Grande do Sul, em Porto Alegre-RS, sob a orientação do professor José Braccini, onde passou a desenvolver pesquisas envolvendo seleção genômica em bovinos de corte.

Entre setembro/2011 a janeiro/2012 realizou período de investigação científica junto a Faculdade de Medicina Veterinária da Universidade Técnica de Lisboa, Portugal, sob a orientação do professor Luís Telo da Gama onde trabalhou o tema de diversidade genética com as raças de origem britânicas criadas no Rio Grande do Sul.

Entre setembro e outubro de 2012 realizou missão científica de curta duração junto ao Department Animal and Poltry Science da University of Guelph-ON, Canadá, sob a supervisão do professor Flávio Scharamm Schenkel, e trabalhou no tema seleção genômica. Posteriormente, no período de setembro/2013 a agosto/2014, nesta mesma instituição, desenvolveu o Programa de Doutorado Sanduíche no Exterior - PDSE.

Em agosto de 2014 retornou ao Brasil para finalizar os trabalhos do programa de doutorado e submeter-se à avaliação de defesa da Tese no Programa de Pós-Graduação em Zootecnia, área de concentração Produção Animal da Universidade Federal do Rio Grande do Sul em Porto Alegre-RS.