

**FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE JABOTICABAL**

**PREDIÇÃO GENÔMICA UTILIZANDO PAINÉIS DE
MARCADORES MOLECULARES COM DIFERENTES
DENSIDADES, EM BOVINOS DA RAÇA NELORE**

Fernando de Oliveira Vasconcelos

Médico Veterinário

2014

**FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE JABOTICABAL**

**PREDIÇÃO GENÔMICA UTILIZANDO PAINÉIS DE
MARCADORES MOLECULARES COM DIFERENTES
DENSIDADES, EM BOVINOS DA RAÇA NELORE**

Fernando de Oliveira Vasconcelos

Orientador: Dr. Roberto Carneiro

Coorientador: Prof. Dr. Henrique Nunes de Oliveira

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Zootecnia (Genética e Melhoramento Animal)

2014

DADOS CURRICULARES DO AUTOR

FERNANDO DE OLIVEIRA VASCONCELOS - nascido a 18 de março de 1986, na cidade de Belo Horizonte – Minas Gerais, é formado em Medicina Veterinária pela UFMG – Universidade Federal de Minas Gerais – MG, em dezembro de 2011. Em agosto de 2012, iniciou o mestrado em Genética e Melhoramento Animal, na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal, sob a orientação do Prof. Dr. Roberto Carvalheiro.

Dedico.....

**À Deus
Aos meus pais.
À minha noiva.**

AGRADECIMENTOS

Agradeço primeiramente a Deus, pois sem sua vontade e auxílio, principalmente nas horas mais difíceis, nada seria possível.

À meus pais, Helio Fidelis Vasconcelos e Maria Raimunda de Oliveira Vasconcelos, que sempre demonstraram apoio às minhas decisões e sempre se esforçaram para que eu pudesse seguir meus objetivos.

À minha noiva, Lorena Carla Vieira, que sempre esteve ao meu lado, me oferecendo amor, amizade e também conselhos quando necessário.

Aos meus sogros, Moacir e Maria Teresinha, pelo apoio e atenção.

Aos professores Dr. Henrique Nunes de Oliveira e Dr. Roberto Carvalheiro, pelos ensinamentos e orientação.

Aos professores Dr^a. Lúcia Galvão de Albuquerque, Dr. Fernando Baldi e Dr. Rusbel Raúl Aspilcueta Borquis, por compor as bancas de qualificação e defesa, contribuindo com sugestões que aprimoraram este trabalho.

Ao programa de pós-graduação em genética e melhoramento animal da UNESP, pela possibilidade oferecida de desenvolver o mestrado.

A DeltaGen pela cessão dos dados, e Capes pela concessão de bolsas.

A todos que possibilitaram de forma direta ou indireta o desenvolvimento deste trabalho.

SUMÁRIO

	Página
SUMÁRIO	iii
RESUMO	iv
ABSTRACT	v
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	4
2.1. Marcadores Moleculares	5
2.2. Seleção Genômica	6
2.3. Imputação	7
3. MATERIAL E MÉTODOS	10
3.1. Genótipos	10
3.2. Fenótipos e Pseudo-fenótipos	10
3.3. Painéis de Baixa Densidade	12
3.4. Imputação	13
3.5. Métodos de Seleção Genômica	14
3.6. Validação Cruzada	14
3.7. Critérios de Avaliação da Eficiência de Imputação	15
3.8. Critérios de Comparação da Seleção Genômica	16
4. RESULTADOS E DISCUSSÃO	17
4.1. Eficiência de Imputação	17
4.2. Habilidade de Predição Genômica	18
5. CONCLUSÃO	23
REFERÊNCIAS	24

PREDIÇÃO GENÔMICA UTILIZANDO PAINÉIS DE MARCADORES MOLECULARES COM DIFERENTES DENSIDADES, EM BOVINOS DA RAÇA NELORE

RESUMO – A seleção genômica tem sido apontada como uma tecnologia que propiciará aumentos expressivos nas taxas de progresso genético em programas de melhoramento animal. Um dos fatores limitantes para a aplicação da seleção genômica é o custo associado à necessidade de genotipagem de um número grande de animais, com painéis de alta densidade, para a obtenção de boa habilidade de predição dos valores genéticos. Uma alternativa para reduzir custos seria genotipar parte dos animais com painéis menos densos e utilizar técnicas de imputação de genótipos. Em bovinos da raça Nelore, ainda não há consenso quanto à densidade dos painéis a serem utilizados e quanto à estratégia de genotipagem e imputação. Assim, objetivou-se com o presente projeto avaliar a habilidade de predição da seleção genômica utilizando painéis de marcadores moleculares de diferentes densidades, assim como o efeito da utilização de genótipos imputados na habilidade de predição da seleção genômica, em bovinos da raça Nelore. A característica considerada foi precocidade de terminação. Um total de 2035 animais geneticamente avaliados para essa característica e genotipados com o chip Illumina® HD Bovina (780k) foram utilizados nas análises, simulando uma situação em que os animais teriam sido genotipados com chips de densidade mais baixa. Análises de seleção genômica também foram executadas utilizando parte dos marcadores do painel de 780k, disponibilizando apenas os SNPs em comum com as seguintes painéis: Illumina® BovineLD (7K), Illumina® BovineSNP50 v2 (50K) e GeneSeek® Genomic Profiler 20K e 75K para *Bos indicus*. A imputação dos genótipos foi feita com o uso do programa FImpute e as predições genômicas foram conduzidas utilizando os métodos GBLUP e LASSO Bayesiano. Ambas análises, de imputação e de predição genômica, foram conduzidas de forma repetida, seguindo um esquema de validação cruzada, com a formação de cinco grupos definidos de forma aleatória. Os resultados indicaram que painéis de 50K e 75K forneceriam, para a característica estudada, predições com acurácia semelhante à do painel 780K. Além disso, os resultados também indicaram que, se o painel 780k for considerado como o chip alvo para a aplicação de seleção genômica na raça Nelore, a sua relação custo benefício pode ser melhorada com a estratégia de genotipar parte dos animais com um chip de baixa densidade (7K ou 20K) e imputar os genótipos faltantes até 780k. Novos estudos devem ser conduzidos para avaliar a reprodutibilidade destes resultados para diferentes características e para cenários com maior número de animais na população referência e com animais de diferentes origens.

Palavras-chave: bovinos, imputação, seleção genômica

GENOMIC PREDICTION IN NELORE CATTLE USING DIFFERENT SNP CHIPS

ABSTRACT – Genomic selection has been considered as a technique that will allow significant increase in the genetic progress of animal breeding programs. A constrain for genomic selection application is the cost of genotyping several animals with high density chips in order to obtain a good prediction equation. An alternative to reduce costs is to genotype part of the animals with a lower density chip and to impute the unobserved genotypes. In Nelore cattle, there is no consensus about which chip and genotyping strategy should be adopted. Therefore, the aim of the present study was to evaluate the predictive ability of genomic selection in Nelore cattle using chips with different densities, and also to assess the effect of using imputed genotypes. The trait considered was finishing precocity. A total of 2,035 animals genetically evaluated for this trait and genotyped with the Illumina® Bovine HD chip (780K) were used in the analyses, mimicking a situation where the animals would have been genotyped with lower density chips. Genomic selection analyses were also run masking part of the 780K genotypes, making available just the SNPs in common with the following chips: Illumina® BovineLD (7K), Illumina® BovineSNP50 v2 (50K), GeneSeek® Genomic Profiler 20K and 75K for *Bos indicus*. Genomic prediction analyses were run with or without imputing the masked genotypes, using the software FImpute, and under the GBLUP and Bayesian LASSO methods. A 5-fold cross-validation scheme was adopted to perform the analyses, randomly assigning the groups. Results showed that the 50K and 75K chips presented the same predictive ability as the 780K chip. The results also indicated that if the 780K was considered as the target chip for applying genomic selection in the Nelore breed, its cost effectiveness could be improved with the strategy of genotyping part of the animals with a lower density chip (7K or 20K) and imputing their 780K missing genotypes. Further studies should be conducted to evaluate the reproducibility of these results for other traits and Nelore populations.

Keywords: cattle, genomic selection, imputation

1. INTRODUÇÃO

A seleção genômica consiste no uso de marcadores moleculares distribuídos densamente ao longo de todo o genoma para prever o valor genético dos animais para características poligênicas (MEUWISSEN *et al.*, 2001). A sua aplicação tornou-se possível graças ao grande avanço nas técnicas de genotipagem, permitindo que inúmeros marcadores moleculares, na forma de SNP (“*Single Nucleotide Polymorphism*”, ou polimorfismo de nucleotídeo único), sejam genotipados a um custo relativamente acessível.

O princípio básico da seleção genômica reside no fato de que painéis densos de marcadores moleculares permitem que boa parte dos QTL (locos associados a características quantitativas) esteja em desequilíbrio de ligação (LD) com pelo menos um marcador próximo a eles. Esta pressuposição é válida quando a probabilidade de recombinação entre QTL e marcador é muito baixa, caracterizando alto nível de LD entre marcadores adjacentes. Quanto maior o LD entre marcador e QTL, mais os marcadores conseguem explicar a variabilidade genética devida aos QTL, associada à expressão das características de interesse. Uma vantagem desta abordagem é que os efeitos estimados dos marcadores seriam válidos para toda a população avaliada e não apenas dentro de famílias, como no caso de esquemas de seleção assistida por marcadores baseados em análises de ligação (GODDARD; HAYES, 2007).

Em comparação à seleção com base em valores genéticos preditos a partir de informações fenotípicas e de pedigree, a seleção genômica permite pelo menos algumas das seguintes vantagens: redução no intervalo de gerações; redução nos custos de testes de progênie; aumento da acurácia de predição e da intensidade de seleção; redução do efeito da depressão endogâmica devido a um maior controle da endogamia. Em bovinos de leite, por exemplo, a seleção genômica tem mudado drasticamente o modelo de trabalho dos programas de seleção (SCHEFERS; WEIGEL, 2012) e tem sido apontada como uma tecnologia que propiciará aumentos expressivos nas taxas de progresso genético, com impacto comparado àqueles causados pela utilização da inseminação artificial e da metodologia de predição dos valores genéticos denominada como BLUP. Em bovinos de corte, a aplicação

economicamente viável da técnica ainda não está bem estabelecida, mas há trabalhos indicando oportunidade de aumentar a resposta à seleção (GARRICK; SAATCHI, 2011; SAATCHI *et al.*, 2012).

Um dos fatores limitantes para a aplicação da seleção genômica é o custo associado à necessidade de genotipagem de um número grande de animais, com painéis de alta densidade, para a obtenção de uma boa habilidade de predição dos valores genéticos (VANRADEN *et al.*, 2011). Uma alternativa para reduzir custos seria genotipar parte dos animais com painéis menos densos e utilizar técnicas de imputação de genótipos.

A imputação de genótipos consiste em inferir os genótipos de marcadores não observados de um painel de maior densidade a partir dos genótipos observados de um painel menos denso. Isso é possível devido ao parentesco entre os animais. Espera-se que indivíduos com um ancestral em comum n gerações atrás, partilhem um mesmo segmento cromossômico de tamanho médio igual a $1/n$ Morgans (DAETWYLER *et al.*, 2010). Uma vez que animais referência, representativos da população, sejam genotipados com o painel de alta densidade, é possível montar uma biblioteca de haplótipos (segmentos cromossômicos) e, usando essa informação, imputar os genótipos de animais genotipados com painéis de menor densidade. A técnica permite redução de custos de genotipagem, uma vez que painéis de menor densidade são menos onerosos.

Diferentes grupos de pesquisa estão investigando estratégias de implementação da seleção genômica em bovinos da raça Nelore, dada a importância da raça para a pecuária nacional. Entretanto, ainda não se sabe qual o número de marcadores necessários para a aplicação da técnica de forma apropriada.

Em gado Holandês, por exemplo, uma prática comum é a de genotipar animais influentes com o painel de 50 mil marcadores (50K) e os candidatos à seleção com painel de 6 mil marcadores, imputando-os para 50K (SCHEFERS; WEIGEL, 2012). Estudos evidenciaram pequeno benefício da utilização de painéis com mais de 50 mil marcadores, quanto à habilidade de predição dos valores genéticos de bovinos da raça Holandesa (VANRADEN *et al.*, 2011; KHATKAR *et al.*, 2012).

Devido à diferença no padrão de desequilíbrio de ligação das raças zebuínas (ESPIGOLAN *et al.*, 2013; PÉREZ O'BRIEN *et al.*, 2014), especula-se que seja necessário utilizar painéis de maior densidade em bovinos da raça Nelore, em comparação aos utilizados na raça Holandesa, para a obtenção de boa habilidade preditiva e eficiência de imputação. Entretanto, há necessidade de condução de pesquisa para avaliar apropriadamente essa questão.

Os objetivos do presente projeto foram os de avaliar, em bovinos da raça Nelore, a habilidade de predição da seleção genômica utilizando painéis de marcadores moleculares de diferentes densidades, assim como o efeito da utilização de genótipos imputados na habilidade de predição da seleção genômica.

2. REVISÃO DE LITERATURA

O agronegócio possui um papel fundamental na economia do Brasil, representando cerca de 7% de todo o produto interno bruto do país (www.ibge.gov.br), sendo grande parte representada pela produção e comercialização de carne bovina. O Brasil é um dos maiores produtores de carne bovina, possui um rebanho com cerca de 200 milhões de cabeças (www.agricultura.gov.br/ministerio), sendo o segundo maior rebanho do mundo e um dos maiores exportadores, responsável por cerca de um quinto de toda carne comercializada internacionalmente, sendo que no mercado de produção de bovinos, a raça Nelore possui um grande destaque, representando cerca de 80% de todo gado de corte no país.

O melhoramento genético animal no Brasil até o final da década de 1970 e início da década de 1980, era baseado na seleção pela aparência do animal. Após esse período, começaram a utilizar avaliações genéticas, utilizando modelos mistos, com informações de fenótipos e de parentesco. Nos dias atuais, informações obtidas de dados genômicos vem sendo utilizadas nos programas de melhoramento genético bovino, com aplicação da seleção assistida por marcadores (ALENCAR, 2010). Apesar da seleção tradicional ter gerado bons resultados ao longo dos anos em que o melhoramento foi implantado, existem algumas situações que a aplicação do método tradicional não consegue gerar respostas satisfatórias, como no caso de características de difícil mensuração, de baixa herdabilidade, características que são expressas tardiamente no animal ou características que são limitadas pelo sexo. Para estas características, o melhoramento depende de registros fenotípicos precisos, realizados em todos os animais disponíveis e muitas vezes de provas zootécnicas especialmente delineadas para obtenção dos dados fenotípicos estruturados, além de informações confiáveis de pedigree.

2.1. Marcadores Moleculares

Marcadores moleculares são polimorfismos de nucleotídeos ou de sequências de nucleotídeos com posição conhecida no genoma. Esses marcadores permitem identificar e localizar genes (ou regiões do genoma) também polimórficos que afetam a expressão de características de interesse econômico. Assim, a partir da obtenção do genótipo dos animais para os marcadores, é possível identificar animais geneticamente superiores. Sendo assim, os indivíduos podem ser selecionados com base na informação molecular, pois temos informações de que a variante alélica de interesse está presente em seu material genético.

Existem vários tipos de marcadores, mas os mais usuais e informativos são os marcadores SNP (*Single Nucleotide Polymorphism*). Estes são utilizados como marcadores, sendo que as mutações são em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina). Os SNPs são extremamente abundantes nos genomas de espécies não endogâmicas. Em bovinos podem ser mais de três milhões (LI *et al.*, 2009).

São muitas as vantagens na utilização de SNPs em comparação com os outros marcadores moleculares disponíveis, pois são muito frequentes, muitas vezes funcionalmente relevantes, identificando características de interesse econômico, possuem baixa taxa de mutação e proporcionam genotipagem de forma altamente automatizada. Uma de suas principais formas de utilização atualmente é em painéis contendo marcadores SNPs, sendo que nestes são colocadas amostras do material genético do animal a ser testado. Os marcadores contidos no painel são informativos e identificam locais específicos no DNA do animal testado. Existem painéis com diversas densidades de SNPs. Entre aqueles disponíveis no mercado, os considerados de baixa densidade são o Illumina® bovine_3k BeadChip que possui cerca de 2.900 SNPs e o Illumina® bovine_7k BeadChip que possui 6.910 SNPs distribuídos pelo genoma. Também existem os de maior densidade como o Illumina® BovineSNP50 v2 BeadChip, que é um painel que contém 54.609 SNPs. Este painel possui SNPs uniformemente distribuídos que são polimórficos em muitas raças e proporcionam um espaçamento médio de 36.9kb. O Illumina® BovineHD BeadChip, com mais de 777.000 SNPs que abrangem uniformemente todo o genoma bovino, é

um painel que permite uma ampla gama de aplicações como estudos de associação genômica ampla e mapeamento de regiões cromossômicas de interesse, e seleção genômica. A frequência média do alelo menor (MAF) em todos os loci é de 0,25 e 0,17 para *Bos taurus taurus* e para *Bos taurus indicus*, respectivamente. Mais de 749 mil SNPs foram validados através de todas as raças com um MAF > 0,05, incluindo a cobertura de SNPs autossômica, mitocondrial, e ligada ao sexo (X/Y), cobertura genômica uniforme, com uma distância média entre marcadores de 3,43 kb e mediana de 2,68 kb (Illumina Inc., San Diego, CA).

2.2. Seleção Genômica

A principal razão para o investimento em pesquisas voltadas para a genética molecular é que as informações que podem ser obtidas em nível de DNA podem conduzir a ganhos genéticos num tempo reduzido, quando comparado com os métodos tradicionais, que utilizam somente as informações fenotípicas e de pedigree (MEUWISSEN *et al.*, 2001). Com isso, consegue-se diminuir o intervalo de gerações, aumentar a intensidade de seleção e a acurácia dos valores genéticos. Também há melhor resposta em características de difícil avaliação pelos métodos tradicionais e menor taxa de endogamia por geração, além de proporcionar a utilização de novas estratégias de melhoramento.

Trabalhos de pesquisa demonstram que existem ganhos quando são incluídas informações obtidas por marcadores. Porém essas informações podem não ser úteis, porque quando os marcadores são muitos esparsos no genoma, podem não estar ligados aos *quantitative trait loci* (QTLs). Por isso, para realizar seleção genômica é necessário utilizar painéis de alta densidade, com marcadores espalhados por todo genoma, a fim de que os marcadores possam estar tão próximos aos QTLs que estejam em desequilíbrio de ligação com eles (HASTBACKA *et al.*, 1992). A grande proximidade dos marcadores contidos nestes painéis faz com que muitos deles segreguem juntos durante a meiose, formando haplótipos. Esses podem ser herdados ao longo das gerações nas famílias, podendo ocorrer que vários indivíduos aparentados tenham nos seus cromossomos segmentos que foram

herdados de um antecessor comum, sendo nesse caso haplótipos idênticos por descendência. Caso exista um QTL nesse segmento, ele será transmitido para a progênie. Por isso, quanto mais jovem é a família, ou seja, quanto mais próximos no pedigree esses animais se encontram, maiores são os segmentos de cromossomos que eles compartilham.

Na seleção genômica não é necessário a identificação das posições dos QTLs, esses são captados devido a alta densidade de marcadores espalhados pelo genoma. Os marcadores colocados em alta densidade também garantem alto desequilíbrio de ligação. O que possibilita a explicação de praticamente toda a variância genética aditiva da característica de interesse.

Conceitualmente, a seleção genômica é implementada em duas etapas, sendo que a primeira consiste em estimar o efeito dos marcadores em uma população de referência, que possui informações de genótipos e fenótipos. A seguir, com essas informações, são estimados os efeitos de todos os marcadores utilizados, e a partir destas estimativas podem ser calculados os valores genômicos dos animais desta e da população de animais candidatos à seleção (HAYES, 2009). Uma das principais limitações da seleção genômica é o alto custo de implementação e quanto mais informativa é a geração dos dados genômicos, mais dispendiosa ela é. A genotipagem tem um custo aproximado de US\$100 por animal com um painel de 50k e de aproximadamente US\$200 por animal com o painel HD. Devido a este alto custo são estudadas alternativas para tentar diminuir tais valores, sendo uma das principais alternativas a imputação de dados genômicos.

2.3. Imputação

Imputação é um termo utilizado para descrever o processo de predição de genótipos que não foram diretamente observados numa amostra de indivíduos (MARCHINI; HOWIE, 2010). Possui grande utilidade em estudos de seleção genômica, de associação genômica ampla, ou em regiões de interesse mais específicas. O principal objetivo desse processo é predizer os genótipos dos animais para os SNPs que não foram genotipados na amostra.

Amostras de indivíduos aparentados são muito informativas para a imputação de genótipos. Genótipos para um número relativamente reduzido de marcadores genéticos podem ser usados para identificar longos trechos de haplótipos compartilhados entre indivíduos aparentados. Estes trechos de haplótipos compartilhados (ou regiões "idênticos por descendência", IBD) são tipicamente usados para avaliar a evidência de ligação. Especificamente, ligação genética implica que membros da família que compartilham uma região IBD serão mais semelhantes entre si do que outra família em que os membros com o mesmo grau de parentesco que não compartilham a região IBD (LI *et al.*, 2009).

Ao estudar amostras de indivíduos não aparentados, a mesma abordagem pode ser utilizada. A principal diferença é que, ao estudar indivíduos não aparentados, os trechos dos haplótipos compartilhados serão muito menores porque ancestrais comuns são mais distantes e, portanto, podem ser mais difíceis de identificar com confiança (LI *et al.*, 2009). Isto é feito utilizando-se informações de haplótipos já detectados naquela população. A imputação é realizada pela predição de genótipos, quando se tem indivíduos genotipados com poucos marcadores, mas existem haplótipos de referência com muita informação. Então são realizadas análises para identificar como os cromossomos foram partilhados entre a amostra individual e haplótipos de referência. Depois de estabelecida a relação de compartilhamento, os dados genotípicos observados da amostra em estudo devem ser combinados aos dados dos haplótipos de referência para completar os genótipos ausentes na amostra com genótipos faltantes (MARCHINI; HOWIE, 2010).

Para fazer a imputação dos genótipos são utilizados alguns softwares, que empregam abordagens um pouco diferentes. O Impute (MARCHINI *et al.*, 2007), MACH (LI *et al.*, 2006) e fastPHASE/BIMBAM (SCHEET; STEPHENS, 2006; SERVIN; STEPHENS, 2007) levam em conta todos os genótipos observados quando imputam cada genótipo, sempre voltando a observar e avaliar todos os genótipos para definir qual o genótipo será imputado. O Plink (PURCELL *et al.*, 2007), WHAP (ZAITLEN *et al.*, 2007) e o Beagle (BROWNING, 2006) geralmente se concentram em um pequeno número de marcadores adjacentes aos genótipos a serem imputados, formando "clusters". São formados vários subgrupos de marcadores adjacentes e a imputação dos genótipos é feita em cada um dos

subgrupos, ao final desse processo é formado o genótipo completo. O FImpute (SARGOLZAEI *et al.*, 2012) utiliza a informação de parentesco num primeiro passo, seguida pela imputação com base na informação da população, adotando-se uma técnica de “sliding window”, nesta técnica são utilizadas as informações dos haplótipos reconstituídos no primeiro passo para encontrar segmentos de cromossomos, que são compartilhados entre os indivíduos, provenientes de um ancestral comum.

Em alguns estudos de imputação foi observada alta acurácia dos dados imputados. Hayes *et al.* (2011) encontraram valores médios de acurácia de 0,801, 0,700 e 0,633, respectivamente, na imputação de dados de painéis de 5K para 50K, nas raças Border Leicester, numa composição das raças Poll Dorsets/White Face Suffolks e Merino. Já, na imputação de 50K para HD, os valores foram 0,960, 0,940 e 0,860, respectivamente. Larmer *et al.* (2012), encontraram valores de acurácia de 0,973, 0,980 e 0,992, respectivamente, para a imputação de dados genômicos de painéis de 50K para HD nas raças Guernsey, Ayshare e Holstein.

3. MATERIAL E MÉTODOS

3.1. Genótipos

Foram utilizados genótipos de 2.035 animais participantes do projeto de pesquisa “Aplicação de ferramentas genômicas no melhoramento genético da eficiência reprodutiva em bovinos da raça Nelore” (CNPq: 559631/2009-0). Neste projeto, foram genotipadas fêmeas da raça Nelore, oriundas de fazendas participantes do programa de melhoramento genético DeltaGen, desafiadas à reprodução em idade precoce (aprox. 16 meses). Todos os animais foram genotipados com o painel Illumina® BovineHD (Illumina Inc., San Diego, CA), que contém aproximadamente 780 mil SNPs (780K).

Previamente às análises de imputação e de seleção genômica foi realizado o controle de qualidade dos genótipos. Foram desconsiderados das análises os SNPs sem posição definida, em regiões não autossômicas, apresentando a mesma posição (coordenada genética) de outro SNP, com o p-valor para o teste de equilíbrio de Hardy-Weinberg menor ou igual a 10^{-5} , com *minor allele frequency* menor que 0,02 e com *call rate* menor que 0,98, sendo que apenas genótipos com *GenCall score* maior ou igual a 0,15 foram considerados como leituras válidas. Além disso, foram descartadas amostras duplicadas e com *call rate* menor que 0,90. Após o controle de qualidade dos genótipos, restaram 2.011 amostras e 516.705 SNPs para o estudo.

3.2. Fenótipos e Pseudo-fenótipos

Os fenótipos e as informações de parentesco foram obtidos da base de dados histórica das fazendas. Os animais foram criados a pasto em fazendas localizadas nas cidades de Valparaíso (SP), Naviraí (MS) e Cotegipe (BA). A característica utilizada para avaliar a habilidade de predição da seleção genômica foi a precocidade de terminação (Ps) avaliada ao sobreano por meio de escores visuais,

sendo uma medida subjetiva no qual o avaliador fornece notas de 1 a 5 aos animais. Quanto maior a nota, melhor a precocidade de terminação do animal. Os escores são definidos pela profundidade e arqueamento de costelas do animal e pela deposição de gordura como descrito por Boligon *et al.* (2013). Essa característica foi escolhida por apresentar, no estudo de Neves *et al.* (2014), alta habilidade de predição com relativamente poucos animais genotipados. O arquivo de dados histórico utilizado no presente estudo continha 132.187 animais com observação fenotípica para Ps, sendo que a matriz de parentesco apresentou um total de 213.150 animais.

Como o número de animais genotipados foi bem menor do que a quantidade de animais com fenótipo, optou-se por utilizar os valores genéticos esperados desregredidos (VGE_d) como pseudo-fenótipo para as análises de seleção genômica. Os VGE_d foram obtidos em dois passos. Primeiramente, os valores genéticos esperados foram obtidos com a aplicação de um modelo animal misto utilizando os fenótipos observados como variável dependente, a idade do animal na avaliação de Ps como covariável (efeito linear e quadrático), o grupo contemporâneo dos animais (sexo, fazenda, ano e estação de nascimento e grupo de manejo ao sobreano) como efeito fixo classificatório, e o valor genético aditivo direto e o resíduo como efeitos aleatórios. Assumiu-se independência e homocedasticidade de variância dos resíduos e a covariância dos valores genéticos foi definida pela matriz de parentesco usual multiplicada pela variância genética aditiva. Os componentes de variância e demais parâmetros do modelo foram estimados/preditos com o emprego do programa BLUPF90 e AIREMLF90 (MISZTAL, 2012).

As acurácias dos dos EBV's foram obtidas de acordo com a seguinte fórmula:

$$\text{Acc (EBV's)} = \sqrt{1 - si / \sigma^2},$$

sendo que o si é o erro padrão dos EBV's e σ^2 é a acurácia obtida pelo AIREMLF90.

Os valores genéticos esperados obtidos no primeiro passo foram então desregredidos conforme método proposto por Garrick *et al.* (2009). Nesta etapa das análises, também foram calculados os ponderadores dos VGE_d para serem utilizados nas análises de seleção genômica, seguindo as recomendações de

Garrick et al. (2009). A estimativa de herdabilidade para P_s , obtida no primeiro passo, foi de 0,26.

3.3. Painéis de Baixa Densidade

Para avaliar a habilidade de predição da seleção genômica utilizando painéis de menores densidades, apenas parte dos genótipos observados no painel 780K foram considerados, representando uma situação na qual os animais teriam sido genotipados com painéis (virtuais) de mais baixa densidade, isto é, com menor número de marcadores. Os painéis de menor densidade foram formados pelos SNPs em comum entre o painel 780K e os seguintes painéis comerciais: Illumina® BovineLD (7K), Illumina® BovineSNP50 v2 (50K) e GeneSeek® Genomic Profiler 20K e 75K para *Bos indicus*. O número de SNPs considerado em cada painel é apresentado na Tabela 1.

Tabela 1. Número de SNPs em comum com o painel 780K antes e após o controle de qualidade dos genótipos (CQG)

Painel	Sigla	Nº SNPs em comum antes CQG	Nº SNPs em comum após CQG
Illumina® BovineLD	7K	6.637	4.864
Illumina® Bovine SNP50 v2	50K	49.345	25.721
GeneSeek® Genomic Profiler 20K	20K	19.493	15.850
GeneSeek® Genomic Profiler 75K	75K	73.941	66.002

Para avaliar a habilidade de predição da seleção genômica utilizando genótipos imputados, foi assumido que parte dos animais (população de referência) apresentou o genótipo do painel de mais alta densidade (780K). Os demais animais (população de validação) tiveram o genótipo imputado para o painel 780K a partir dos painéis com menores densidades descritos na tabela 1. A distribuição dos animais nas populações de referência e validação obedeceu o esquema de validação cruzada descrito no item 3.6.

3.4. Imputação

Os animais virtualmente genotipados com os painéis de mais baixa densidade (7K, 20K, 50K e 75K) tiveram seus genótipos imputados para o painel de 780K. A imputação dos genótipos foi feita com o uso do programa FImpute v.2.2 (SARGOLZAEI *et al.*, 2011). No programa é utilizada a informação de parentesco num primeiro passo, seguida pela imputação com base na informação da população, adotando-se uma técnica de “sliding window”, nesta técnica são utilizadas as informações dos haplótipos reconstituídos no primeiro passo para encontrar segmentos de cromossomos, que são compartilhados entre os indivíduos, provenientes de um ancestral comum. O procedimento para encontrar esses segmentos é repetido várias vezes com alteração do tamanho das “janelas” a serem comparadas e, após esse processo, as frequências dos haplótipos na população referência são utilizadas para predizer o genótipo mais provável faltante. O FImpute foi escolhido por apresentar alta eficiência computacional e elevada acurácia de imputação dos genótipos (SARGOLZAEI *et al.*, 2011; SUN *et al.*, 2012).

3.5. Métodos de Seleção Genômica

Para cada um dos painéis testados (com genótipos imputados ou não) foram conduzidas análises de seleção genômica utilizando os métodos GBLUP e LASSO Bayesiano, conforme descrito a seguir.

3.5.1. GBLUP

O modelo utilizado neste método pode ser descrito como:

$$\mathbf{y} = \mathbf{1}_n\boldsymbol{\mu} + \sum_{j=1}^p X_j g_j + \mathbf{e},$$

onde: \mathbf{y} é o vetor com os “pseudo-fenótipos” (valor genético desregredido), de ordem n , $\boldsymbol{\mu}$ é a média geral, $\mathbf{1}_n$ é um vetor de uns de ordem n , g_j e X_j representam o efeito do j -ésimo marcador e uma matriz de incidência relacionando os dados ao efeito do marcador j , respectivamente, e \mathbf{e} é o vetor aleatório de resíduos. Foi pressuposto

que os resíduos seguem uma distribuição normal multivariada com média zero e variância igual a $\mathbf{R}\sigma_e^2$, sendo \mathbf{R} uma matriz diagonal com os ponderadores associados a cada pseudo-fenótipo (calculados no passo de desregressão dos valores genéticos esperados) e σ_e^2 é a variância residual comum para todas as observações. Além disso, assumiu-se a mesma variância (σ_g^2) para os efeitos dos p marcadores ($\sim N(0, \sigma_g^2)$). Este método foi implementado com a utilização do programa GS3 (LEGARRA *et al.*, 2012).

3.5.2. LASSO Bayesiano

Foi adotada a implementação do LASSO Bayesiano (“Improved Bayesian Least Absolute Shrinkage and Selection Operator” - IBLASSO) proposta por Legarra *et al.* (2011). Com a parametrização proposta pelos autores, as distribuições “a priori” para os efeitos dos marcadores (g_j) podem ser representadas por:

$$P(g_j|\tau_j^2) \sim N(0, \tau_j^2), \text{ e}$$

$$P(\tau_j^2|\lambda) = (\lambda^2/2) \exp(-\lambda^2|\tau_j^2|).$$

Essa parametrização implica em estimar variâncias individuais para cada marcador (τ_j^2) de forma condicional ao parâmetro de regularização λ . O programa GS3 (LEGARRA *et al.*, 2012) foi utilizado para a implementação deste método.

3.6. Validação Cruzada

As análises de seleção genômica foram conduzidas de forma repetida, seguindo um esquema de validação cruzada. Os animais genotipados que passaram pelo controle de qualidade foram divididos aleatoriamente em cinco grupos. Para cada painel de marcadores e método de seleção genômica empregados, cinco análises (repetições) foram conduzidas. A cada repetição, os animais de quatro dos cinco grupos fizeram parte da população de treinamento (referência) e os animais do grupo remanescente fizeram parte da população de validação, de forma que após as cinco repetições os animais dos cinco grupos fizeram parte da população de validação.

Nas análises de seleção genômica, apenas os animais da população de referência foram utilizados para estimar os efeitos dos marcadores e os animais da população de validação tiveram seus valores genômicos preditos com base nessas estimativas. As análises de imputação dos genótipos obedeceram a mesma divisão dos grupos, sendo que os animais da população de referência foram utilizados para montar a biblioteca de haplótipos do painel 780K e a população de validação teve os genótipos imputados dos painéis de menor densidade para o painel 780K.

Ao todo, foram realizadas 20 análises de imputação (5 repetições para cada um dos 4 painéis (7K, 20K, 50K e 75K)) e 180 análises de seleção genômica (5 repetições para cada um dos dois métodos (GBLUP e IBLASSO), dos 9 “painéis” (7K, 7K imputado, 20K, 20K imputado, 50K, 50K imputado, 75K, 75K imputado e 780K) e dos 2 tipos de análise (estimação/treinamento e predição/validação)).

3.7. Critérios de Avaliação da Eficiência de Imputação

A eficiência de imputação foi avaliada de acordo com os seguintes critérios: porcentagem de genótipos não imputados (PGNI); porcentagem de genótipos imputados corretamente (PGIC); e correlação entre o genótipo observado e o genótipo imputado (CORR).

Maiores valores de PGIC e CORR e menores valores de PGNI caracterizam maior eficiência de imputação. Diferentes critérios de avaliação da eficiência de imputação foram adotados para permitir a comparação dos resultados com diferentes trabalhos da literatura. Alguns pesquisadores preferem, por exemplo, adotar o critério CORR por ele ser menos afetado pela frequência alélica dos SNPs do que o critério PGIC (HICKEY *et al.*, 2012; KHATKAR *et al.*, 2012).

Os critérios de avaliação da eficiência de imputação foram calculados objetivando associá-los com as medidas de habilidade de predição da seleção genômica, nas análises em que genótipos imputados foram utilizados.

3.8. Critérios de Comparação da Seleção Genômica

Os resultados das diferentes análises de seleção genômica foram comparados de acordo com três estatísticas referentes ao VGE_d e o valor genômico predito (VGP) dos animais do grupo de validação, sendo elas: i) correlação de Pearson entre o VGE_d e o VGP, como medida empírica da acurácia de predição; ii) correlação de Pearson entre o VGE_d e o VGP, ponderada pela acurácia média do VGE_d, como medida empírica da acurácia de predição ajustada pelo grau de confiança em VGE_d; e iii) inclinação da regressão linear do VGE_d sobre o VGP, para avaliar o grau de inflação das predições genômicas. No geral, maiores correlações e valores de inclinação próximos a 1 estão associados com melhor habilidade de predição.

4. RESULTADOS E DISCUSSÃO

4.1. Eficiência de Imputação

As medidas de eficiência de imputação para os diferentes painéis são apresentadas na Tabela 2. A porcentagem de genótipos não imputados não consta na tabela pois ela foi igual a zero para todos os casos, isto é, o FImpute imputou todos os genótipos faltantes nos painéis de baixa densidade. As médias da porcentagem de genótipos imputados corretamente (PGIC) e da correlação entre os genótipos imputados e observados (CORR) foram altas para todos os painéis, sendo os menores valores observados para o painel 7k. Conforme esperado, os valores de CORR foram proporcionalmente um pouco maiores dos que os valores de PGIC, uma vez que a medida PGIC, em comparação com a CORR, aplica uma penalização maior para o caso de apenas um dos alelos imputados incorretamente. Assim como em Badke et al. (2013), a eficiência de imputação aumentou com o aumento do número de SNPs nos painéis de mais baixa densidade.

Tabela 2. Eficiência de imputação de painéis de menor densidade para o painel referência (Illumina® BovineHD – 780K) – médias das repetições da validação cruzada

Painel	PGIC* (%)			CORR**		
	Mínimo	Média	Máximo	Mínimo	Média	Máximo
7K	79,59	92,29	99,54	0,8224	0,9386	0,9967
20K	90,17	98,23	99,92	0,9203	0,9858	0,9994
50K	90,85	98,33	99,91	0,9265	0,9872	0,9993
75K	94,94	99,40	99,97	0,9574	0,9955	0,9998

* Porcentagem de genótipos imputados corretamente.

** Correlação entre o genótipo observado e o genótipo imputado.

O ganho em eficiência de imputação foi proporcionalmente maior de 7K para 20K do que de 20K para as demais densidades, uma vez que os valores de PGIC e CORR já foram bastante altos para 20K. Resultado semelhante foi encontrado por

Khatkar et al. (2012) que, ao imputar genótipos do painel 50K de animais da raça Holandesa, observaram um maior ganho relativo em eficiência de imputação ao aumentar a densidade dos painéis de 3K para 7K, do que de 7K para 10K.

No presente estudo, mesmo apresentando maior densidade, o painel 50K apresentou resultados similares ao painel 20K. Este resultado pode ser explicado pelo fato do painel 50K ter sido desenvolvido, principalmente, para aplicações genômicas em animais de raças taurinas, evidenciando a importância de se utilizar painéis customizados para populações ou grupos raciais específicos. A eficiência de imputação para o painel 75K foi superior aos demais sobretudo quanto aos valores mínimos de PGIC e CORR (Tabela 2).

Embora os animais da raça Nelore apresentem um menor nível de desequilíbrio de ligação entre marcadores mais próximos do que animais de raças taurinas (ESPIGOLAN et al., 2013; PÉREZ O'BRIEN *et al.*, 2014), a eficiência de imputação observada no presente estudo foi similar à reportada na literatura para raças taurinas. Ma *et al.* (2013), por exemplo, observaram valor de CORR ao redor de 0,97 ao imputar os genótipos de 50K para 780K de animais *Swedish and Finnish Red cattle*. Esse resultado pode estar em parte associado ao fato dos animais genotipados no presente estudo serem de apenas duas fazendas que fazem parte do mesmo programa de melhoramento genético, portanto possivelmente apresentem um maior grau de parentesco. Conforme reportado na literatura (ZHANG; DRUET, 2010; BERRY; KEARNEY, 2011; HICKEY *et al.*, 2012; MA *et al.*, 2013), a eficiência de imputação é influenciada pelo grau de parentesco dos animais da validação com os animais referência.

4.2. Habilidade de Predição Genômica

As medidas de habilidade de predição genômica para os diferentes painéis e métodos são apresentadas na Tabela 3. Em todos os casos, a correlação ponderada entre o VGE_d e o VGP (acurácia ponderada) resultou em maiores valores de acurácia de predição do que a correlação não ponderada (acurácia). Este resultado era esperado uma vez que a acurácia média do VGE_d dos animais da população de

validação, utilizada como ponderador, não era perfeita (i.e. era menor do que um). De acordo com os resultados do estudo de simulação de Neves (2013), a acurácia ponderada é uma medida mais adequada para calcular a acurácia empírica de predição quando o VGE_d é utilizado para estimar os efeitos dos marcadores e para validar a predição genômica, por gerar resultado mais próximo ao da acurácia verdadeira (correlação entre o valor genético verdadeiro/simulado e o VGP).

Tabela 3. Habilidade de predição genômica para os diferentes painéis de marcadores e métodos de seleção genômica utilizados – médias das repetições da validação cruzada

Painel	Acurácia [†]		Acurácia ponderada [†]		Regressão [‡]	
	GBLUP	IBLASSO	GBLUP	IBLASSO	GBLUP	IBLASSO
7K	0,1948	0,1941	0,3541	0,3529	0,7371	0,8352
7K imputado	0,2225	0,2205	0,4046	0,4009	0,9670	1,1198
20K	0,2073	0,2067	0,3769	0,3759	0,8588	0,8582
20K imputado	0,2238	0,2212	0,4070	0,4021	0,9533	1,1140
50K	0,2310	0,2305	0,4199	0,4191	0,9512	0,8912
50K imputado	0,2265	0,2238	0,4117	0,4069	0,9619	1,1222
75K	0,2160	0,2144	0,3928	0,3898	0,9144	0,9144
75K imputado	0,2263	0,2240	0,4115	0,4073	0,9610	1,1254
780K	0,2282	0,2258	0,4149	0,4105	0,9701	1,1286

[†] Correlação de Pearson entre o valor genético esperado desregredido (VGE_d) e o valor genômico predito (VGP) dos animais da população de validação.

[†] Correlação de Pearson entre o VGE_d e o VGP ponderada pela acurácia média do VGE_d dos animais da população de validação.

[‡] Coeficiente de regressão linear do VGE_d sobre o VGP dos animais da população de validação.

Ao comparar os métodos de predição utilizados, é possível observar na Tabela 3 que o GBLUP e o IBLASSO geraram resultados bem semelhantes em termos de acurácia de predição. Os métodos diferiram quanto ao grau de inflação das predições genômicas, sendo que para alguns casos o IBLASSO apresentou predições genômicas deflacionadas (i.e. coeficientes de regressão maiores que um).

A inflação/deflação das predições genômicas se torna importante quando há interesse em comparar a avaliação genética de animais genotipados com a de animais não genotipados (VITEZICA *et al.*, 2011). Predições deflacionadas, por exemplo, fariam com que os animais genotipados tivessem seus valores genéticos subestimados em relação aos valores genéticos dos animais não genotipados. No geral, o GBLUP gerou coeficientes lineares de regressão mais próximos de um, em comparação ao IBLASSO, indicando melhor escala das predições genômicas fornecidas pelo GBLUP.

A vantagem da utilização de métodos Bayesianos em comparação ao GBLUP, observada em estudos de simulação (MEUWISSEN *et al.*, 2001; CLARK *et al.*, 2011), nem sempre ocorre em dados reais, exceto para características afetadas por genes de efeito maior como o caso do teor de gordura no leite (VANRADEN *et al.*, 2009; COLOMBANI *et al.*, 2013). Os resultados do presente estudo não permitem afirmar, entretanto, que a característica precocidade de terminação não é afetada por genes de efeito maior. Estudos de associação genômica ampla estão sendo conduzidos para investigar mais apropriadamente esta questão.

A acurácia de predição genômica obtida no presente estudo com o painel 780K (0,41) foi bem inferior à obtida por Neves *et al.* (2014) para touros Nelore genotipados com o mesmo painel (0,74). Este resultado pode ser em parte explicado pela menor acurácia dos pseudo-fenótipos utilizados no presente estudo para treinar os efeitos dos marcadores. Neves *et al.* (2014) utilizaram os genótipos de touros influentes da raça, com grande número de filhos avaliados.

Exceto para o painel de 50K, as acurácias de predição dos painéis de menor densidade não imputados foram menores do que as acurácias apresentadas pelo painel 780K (Tabela 3). As acurácias dos painéis 7K e 20K, por exemplo, foram em torno de 85% e 91%, respectivamente, da acurácia apresentada pelo painel 780K. Por sua vez, os painéis 50K e 75K não imputados resultaram em acurácia de predição semelhante ao painel 780K. Não está claro quanto o tamanho e a composição da população referência afetou os resultados. De acordo com VanRaden *et al.* (2011), o benefício de utilizar painéis mais densos para a seleção genômica se torna mais evidente a medida que o número de animais na população referência aumenta. Uma vez que os animais genotipados no presente estudo são

provenientes de apenas três fazendas, que participam do mesmo programa de melhoramento genético, possivelmente apresentem alto grau de parentesco. A similaridade dos resultados encontrados entre os painéis 50K (ou 75K) e 780K também pode estar relacionada ao fato da variância capturada pelos marcadores estar sendo mais afetada pelo parentesco aditivo e pela cosegregação do que propriamente pelo desequilíbrio de ligação entre os marcadores e os QTL (HABIER *et al.*, 2013).

A imputação dos genótipos fez com que os painéis de menor densidade que haviam apresentado menor acurácia de predição (7K e 20K), alcançassem acurácia semelhante ao painel 780K (Tabela 3). No presente estudo, os painéis 7K e 20K imputados apresentaram acurácias de predição similares (Tabela 3), embora a eficiência de imputação tenha sido menor para o painel 7K (Tabela 2). VanRaden *et al.* (2011), Khatkar *et al.* (2012) e Mulder *et al.* (2012) observaram influência da eficiência de imputação na acurácia das predições genômicas. Mulder *et al.* (2012) derivaram uma equação determinística para predizer a acurácia das predições genômicas em função da acurácia de imputação e observaram uma relação linear entre ambas medidas. Daetwyler *et al.* (2011) sugeriram que o declínio na acurácia das predições genômicas é ligeiramente inferior ao declínio da acurácia de imputação.

A imputação dos genótipos teve o efeito de melhorar a escala das predições genômicas para o método GBLUP. Para o IBLASSO, a imputação dos genótipos ocasionou deflação das predições genômicas. Em ambos os métodos, as predições genômicas se apresentaram inflacionadas nos painéis não imputados (Tabela 3). Alguns estudos na literatura também reportaram diferenças entre métodos de predição genômica quanto à escala das predições (ex.: SAATCHI *et al.*, 2011; COLOMBANI *et al.*, 2013). Entretanto, não há consenso a respeito do método que fornece as predições com melhor escala.

Uma das principais dificuldades para a aplicação mais efetiva da seleção genômica em gado de corte é a falta da definição de estratégias que permitam a aplicação da técnica de uma forma economicamente viável. Em gado de leite, por exemplo, há estudos indicando que os custos com genotipagem seriam compensados pelo aumento do progresso genético e pela redução dos custos de

teste de progênie (SCHAEFFER, 2006). Para animais da raça Nelore, ainda não está bem definido se o painel 780K deve ser considerado como o painel de alta densidade alvo para a aplicação da seleção genômica. Os resultados do presente estudo indicaram que os painéis de 50K e 75K forneceria, para a característica estudada, predições com acurácia semelhante à do painel 780K. Os resultados também indicaram que a estratégia de genotipar parte dos animais com painéis de baixa densidade (7K ou 20K) e imputar os genótipos faltantes apresentaria habilidade de predição em conformidade com a estratégia mais onerosa de genotipar os animais apenas com o painel mais denso. De acordo com os valores atuais de mercado, o custo de genotipagem seria em torno de três vezes menor ao utilizar um painel de 20K em comparação ao 780K (R\$ 125,00 x R\$ 380,00).

Novos estudos devem ser conduzidos para avaliar a reprodutibilidade destes resultados para diferentes características e para cenários com maior número de animais na população referência e com animais de diferentes origens/fazendas. O presente estudo não permite inferir se estes resultados se reproduziriam para outras características e para animais de outros rebanhos.

5. CONCLUSÕES

O GBLUP e o IBLASSO geraram resultados bem semelhantes em termos de acurácia de predição. A imputação de genótipos melhorou a acurácia de predição em quase todos os painéis, excetuando-se o painel de 50K, fazendo com que eles alcançassem acurácia semelhante ao painel de 780K. Este acréscimo foi mais evidente nos painéis de 7K e 20K, pois os painéis de 50K e 75K já apresentaram acurácias próximas ao de 780K, no caso do painel de 50K, a acurácia foi mais alta em mais baixa densidade do que quando foi imputado.

A seleção genômica da característica precocidade de terminação ao sobreano para os animais Nelore do rebanho estudado poderia ser aplicada de forma menos onerosa com a genotipagem de parte dos animais com o painel 7K ou 20K. Esta estratégia, utilizada em combinação com a imputação de genótipos, resultará em predições genômicas com acurácia semelhante à fornecida pelo painel 780K.

REFERÊNCIAS

- ALENCAR, M. M. Cruzamentos para a produção de carne bovina. In: **Simpósio sobre manejo e nutrição de gado de corte**, 2000, Goiânia.
- BADKE, Y. M.; BATES, R. O.; ERNST, C. W., *et al.* Methods of tagSNP selection and other variables affecting imputation accuracy in swine. **BMC Genetics**, v. 14, n. 8, 2013.
- BERRY, D. P.; KEARNEY, J. F. Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. **Animal**, v. 5, n. 8, p. 1162–1169, 2011.
- BOLIGON, A. A.; CARVALHEIRO, R.; ALBUQUERQUE, L. G. Evaluation of mature cow weight: Genetic correlations with traits used in selection indices, correlated responses, and genetic trends in Nelore cattle. **J. Anim. Sci.**, v. 91, p. 20-28, 2013.
- BROWNING, S. R. Multilocus association mapping using variable-length Markov chains. **Am. J. Hum. Genet.**, v. 78, p. 903–913, 2006.
- CLARK, S. A.; HICKEY, J. M.; VAN DER WERF, J. H. J. Different models of genetic variation and their effect on genomic evaluation. **Gen. Sel. Evol.**, v. 43, n. 18, 2011.
- COLOMBANI, C.; LEGARRA, A.; FRITZ, S., *et al.* Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC π methods for genomic selection in French Holstein and Montbéliarde breeds. **J. Dairy Sci.**, v. 96, p. 575–591, 2013.
- DAETWYLER, H. D.; PONG-WONG, R.; VILLANUEVA, B. *et al.* The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021-1031, 2010.
- DAETWYLER H. D.; WIGGANS, G. R.; HAYES, B. J.; WOOLLIAMS, J. A.; GODDARD, M. E. Imputation of missing genotypes from sparse to high density using long-range phasing. **Genetics**, v. 189, p. 317–327, 2011.

- ESPIGOLAN, R.; BALDI, F.; BOLIGON, A. A., *et al.* Study of whole genome linkage disequilibrium in Nellore cattle. **BMC Genomics**, v. 14, p. 305, 2013.
- EUCLIDES FILHO, K. Evolução do melhoramento genético de bovinos de corte no Brasil. **Rev. Ceres.**, v. 56, n. 5, p. 620-626, 2009.
- GARRICK, D.; SAATCHI, M. Opportunities and challenges for genomic selection of beef cattle. **R. Bras. Zootec.**, v. 40, p. 310-316, 2011 (supl. especial).
- GARRICK, D.; TAYLOR, J. F.; FERNANDO, R. L. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Gen. Sel. Evol.**, v. 41, p. 55-62, 2009.
- GODDARD, M. E.; HAYES, B. J. Genomic selection. **J. Anim. Breed. Genet.**, v. 124, p. 323-330, 2007.
- HABIER, D.; FERNANDO, R. L.; GARRICK, D. J. Genomic BLUP decoded: A look into the black box of genomic prediction. **Genetics**, v. 194, p. 597-607, 2013.
- HAYES, B. J.; BOWMAN, P. J.; DAETWYLER, H. D.; KIJAS, J. W.; VAN DER WERF, J. H. J. Accuracy of genotype imputation in sheep breeds. **Animal Genetics**, v. 43, p. 72-80, 2011.
- HENDERSON, C. R. Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario. 423p. 1984.
- HENDERSON, C. R. Estimates of changes in herd environment. **J. Dairy Sci.**, v. 32, p. 706, 1949.
- HICKEY, J.; CROSSA, J.; BABU, R.; CAMPOS, G. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. **Crop Science**, v. 52, p. 654-663, 2012.
- KHATKAR, M.; MOSER, G.; HAYES, B. J.; RAADSMA, H. W. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. **BMC Genomics**, v. 13, p. 538-549, 2012.

- LARMER, S.; SARGOLZAEI, M.; VENTURA, R.; SCHENKEL, F. Imputation accuracy from low to high density using within and across breed reference populations in Holstein, Guernsey and Ayrshire cattle. Disponível em: <http://www.cdn.ca/Articles/GEBMAR2012>
- LEGARRA, A.; RICARDI, A.; FILANGI, O. GS3 – Genomic Selection, Gibbs Sampling, Gauss-Seidel (and BayesCπ and Bayesian Lasso) - Version 2.0. 2012. Disponível em: <http://snp.toulouse.inra.fr/~alegarra>.
- LEGARRA, A.; ROBERT-GRANIÉ, C.; CROISEAU, P., *et al.* Improved Lasso for genomic selection. **Genet. Res., Camb.**, v. 93, n. 1, p. 77-87, 2011.
- LI, Y.; DING, J.; ABECASIS, G. R. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. **Am. J. Hum. Genet.**, v. 79, S2290, 2006.
- LI, Y.; WILLER, C.; SANNA, S.; ABECASIS, G. Genotype imputation. **Annu. Rev. Genomics Hum. Genet.**, v. 10, p. 387-406, 2009.
- MA, P.; BRØNDUM, R. F.; ZHANG, Q., *et al.* Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish red Cattle. **J. Dairy Sci.**, v. 96, p. 4666–4677, 2013.
- MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. **Genetics**, v. 11, p. 499-511, 2010.
- MARCHINI, J.; HOWIE, B.; MYERS, S.; McVEAN, G.; DONNELLY, P. A new multipoint method for genome-wide association studies by imputation of genotypes. **Nat. Genet.**, v. 39, p. 906–913, 2007.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, p. 1819–1829, 2001.
- MISZTAL, I. BLUPF90 - a flexible mixed model program in Fortran 90. 2012. Disponível em: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90.pdf>.

- MULDER, H. A.; CALUS, M. P. L.; 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.**, v. 95, p. 876–889, 2012.
- NEVES, H. H. R. Strategies to improve the efficiency of genomic selection in animal breeding programs. Unesp-Jaboticabal, 162p., 2013. (Tese de Doutorado)
- NEVES, H. H. R.; CARVALHEIRO, R.; PÉREZ O'BRIEN, A. M., *et al.* Accuracy of genomic predictions in *Bos indicus* (Nelore) cattle. **Gen. Sel. Evol.**, v. 46, p. 17, 2014.
- PÉREZ O'BRIEN, A. M.; MÉSZÁROS, G.; UTSUNOMIYA, Y. T., *et al.* Linkage disequilibrium levels in *Bos Indicus* and *Bos Taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. **Livest. Sci.**, 2014. (submetido).
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R. *et al.* PLINK: a toolset for whole genome association and population-based linkage analyses. **Am. J. Hum. Genet.**, v. 81, p. 559–575, 2007.
- SAATCHI, M.; MCCLURE, M. C.; MCKAY, S. D., *et al.* Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Gen. Sel. Evol.**, v. 43, p. 40, 2011.
- SAATCHI, M.; SCHNABEL, R. D.; ROLF, M. M., *et al.* Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. **Gen. Sel. Evol.**, v. 44, p. 38-47, 2012.
- SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. Accuracy of imputed 50k genotypes from 3k and 6k chips in dairy cattle breeds using FImpute. **In: International Plant and Animal Genome Conference**, San Diego, 2012.
- SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. FImpute – An efficient imputation algorithm for dairy cattle populations. **J. Dairy Sci.**, v. 94(E-Suppl.1), p. 421, 2011. (Abstr.)
- SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **J. Anim. Breed. Genet.**, v. 123, n. 4, p. 218–223, 2006.

- SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **Am. J. Hum. Genet.**, v. 78, p. 629–44, 2006.
- SCHEFERS, J. M.; WEIGEL, K. A. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. **Animal Frontiers**, v.2, n. 1, p. 4-9, 2012.
- SERVIN, B.; STEPHENS, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. **PLoS Genet.**, v. 3, e114, 2007.
- SUN, C.; WU, X.; WEIGEL, K., *et al.* An ensemble-based approach to imputation of moderate-density genotypes for genomic selection with application to Angus cattle. **Genet. Res., Camb.**, v. 94, p. 133–150, 2012.
- VANRADEN, P. M.; O'CONNELL, J. R.; WIGGANS, G. R.; WEIGEL, K. A. Genomic evaluations with many more genotypes. **Gen. Sel. Evol.**, v. 43, p. 10, 2011.
- VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R., *et al.* Invited review: Reliability of genomic predictions for North American Holstein bulls. **J. Dairy Sci.**, v. 92, p. 16–24, 2009.
- VITEZICA, Z. G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. **Genet. Res., Camb.**, v. 93, p. 357–366, 2011.
- ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **J. Dairy Sci.**, v. 93, n. 11, p. 5487–5494, 2010.