**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Estratégias de imputação de genótipos de marcadores SNP para estudos de
associação genômica em animais da raça Nelore**

Fabiane de Lima Silva

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Ciência Animal
e Pastagens

**Piracicaba
2013**

Fabiane de Lima Silva
Zootecnista

**Estratégias de imputação de genótipos de marcadores SNP para estudos de associação
genômica em animais da raça Nelore**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **LUIZ LEHMANN COUTINHO**

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Ciência Animal
e Pastagens

Piracicaba
2013

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Silva, Fabiane de Lima

Estratégias de imputação de genótipos de marcadores SNP para estudos de associação genômica em animais da raça Nelore / Fabiane de Lima Silva.- - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2013. 98 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.

1. Inferência Bayesiana 2. Gado de corte 3. Genotipagem 4. Paineis de alta e baixa densidade I. Título

CDD 636.2082
S586e

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte -O autor"

DEDICATÓRIA

Ofereço

A *Deus*, pela determinação e perseverança na realização deste estudo e, principalmente, por sempre caminhar ao meu lado abrindo portas para que meus sonhos se concretizassem.

Com muito carinho aos meus pais,

Paulo Antônio da Silva e Alzira de Lima Silva, por todo incentivo e amor incondicional nos momentos mais importantes da minha vida.

Meu namorado,

Kuang Houngy por toda compreensão, companheirismo, amizade, conforto nas horas difíceis. Sem dúvida com você meus dias tornaram mais felizes.

Meus queridos e amados irmãos,

Lidiane de Lima Silva, Wilma de Lima Silva e Willames de Lima Silva pelo apoio, amizade, palavras de incentivo e compreensão.

AGRADECIMENTOS

A Deus por estar sempre presente em minha vida e tornar tudo possível. Obrigada pela força, perseverança nesta caminhada e conquista, principalmente naqueles momentos em que estive triste e enfraquecida.

À minha família, por sua capacidade de acreditar em mim. Mãe, seu cuidado e orações me deram esperança para seguir em busca dos sonhos. Pai, suas palavras de otimismo e risadas foram fundamentais. Mesmo sem a presença de vocês todos estes anos, eu sempre tive a certeza que nunca estaria sozinha.

Ao meu namorado e amigo, Kuang Hongyu, pelo carinho, amor, paciência e apoio mútuo nos momentos mais difíceis.

À Escola Superior de Agricultura “Luiz de Queiroz” e ao Programa de Pós-Graduação em “Ciência Animal e Pastagens” pela oportunidade de realização do doutorado.

Ao professor Dr. Luiz Lehmann Coutinho pela orientação, paciência, sugestões e ensinamentos transmitidos.

Ao professor Dr. Guilherme J. G. Rosa pela orientação quando estive em Madison, sugestões valiosas para este estudo, amizade, ensinamentos transmitidos, paciência e incentivo nesta caminhada. Muito obrigada!

À professora Dr^a. Angelina Bossi Fraga, pela amizade e apoio em momentos mais importantes na minha caminhada.

Ao professor Dr. Irineu Umberto Packer (*in memoriam*) pela amizade, ensinamentos transmitidos, preocupação quando cheguei à Piracicaba. Sem dúvida um grande exemplo de profissional que nunca esquecerei.

À Embrapa Pecuária Sudeste, em especial à Dra. Luciana de Almeida Regitano, pela disponibilidade do banco de dados.

À colega Ana Luiza Paço pela disponibilidade e acesso aos dados fenotípicos para a realização deste estudo.

Ao Dr. Maurício Mudadu, por todo o suporte nas análises e discussões, atenção e disponibilidade sempre que precisei de alguma ajuda.

Ao professor Dr. Gerson Barreto Mourão, pela orientação e colaboração neste trabalho.

Em especial, às queridas amigas Elizabeth Mie Hashimoto e Gláucia Tatiana Ferrari, pelo companheirismo, amizade, palavras de incentivo e convivência agradável. Amigos sempre serão amigos.....não importa a distância.

À amiga Poliana Tizioto pela troca de informação, parcerias e amizade.

Aos Amigos de Pós-graduação do Departamento de Zootecnia, Carla de Andrade, Michele Lopes do Nascimento, Aline Zampar, Gerson Oliveira Júnior, Gregori Rovadoscki, Johanna Ramírez Diaz, Mayara Salvian, Ana Luísa Conti, Simone Pértile, Juliana Petrini, Renato Prioli pela amizade e experiências compartilhadas.

Aos Amigos de Pós-graduação do Departamento de Ciências Exatas: Tiago de Almeida, Marcelino Pascoa (Popó), Ana Patrícia Peixoto, Simone Daniela Sartório, José Nilton Cruz, Pedro Amoendo, Verónika Rondinel, Sergio Arciniegas Alarcón, Marisol García Peña, Renato Nunes, Ricardo de Olinda, Gilson Silvério da Rocha, Alexandre Barbosa (*in memorian*), João Vitor Teodoro, pela proveitosa, alegre e positiva convivência.

Aos amigos do Laboratório de Biotecnologia Animal da ESALQ: Dênia Attílio, Ribamar Nunes, Aline Cesar, Gustavo Gasparin, Gabriel Costa, Tassia Mangetti, Thais Godoy, Lilian Rezende, Andrezza Maria Felício, Vinícius Henrique da Silva, pelo acolhimento, risadas e conversas nos intervalos de descanso, pois fizeram toda diferença.

Aos docentes dos Departamentos de Zootecnia e Ciências Exatas, pelos ensinamentos e grande contribuição em minha formação acadêmica.

À bibliotecária Eliana Maria Garcia, pela atenção e formatação da tese.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro concedido.

À FAPESP pela concessão da bolsa de estudos na modalidade de bolsa regular no país e no exterior (BEPE), colaborando para minha formação acadêmica e realização deste estudo.

Obrigada a todos que passaram pela minha vida durante este período em Piracicaba e de alguma forma acompanharam a concretização deste sonho. De alguma maneira todos fazem parte de mais uma etapa.

Muito Obrigada!

BIOGRAFIA

Fabiane de Lima Silva – nascida em 04 de junho de 1985 no município de São José da Laje – Alagoas, filha de Paulo Antônio da Silva e Alzira de Lima Silva. Iniciou em março de 2003 o curso de graduação em Zootecnia na Universidade Federal de Alagoas – UFAL – Campus de Maceió, e obteve título de bacharel em Zootecnia em dezembro de 2007. Durante a graduação foi bolsista da Fundação de Amparo à Pesquisa do Estado de Alagoas (UFAL) e realizou estágios na área de melhoramento genético de bovinos de corte, sob supervisão do Prof^ª. Dra Angelina Bossi Fraga e do pesquisador Dr. Maurício de Mello Alencar na unidade da Embrapa Pecuária Sudeste, São Carlos, SP. A aluna exerceu atividades de monitora em Melhoramento Genético Animal, sob orientação da Prof^ª. Dra Angelina Bossi Fraga. Em março de 2008, ingressou no curso de mestrado em Ciências pela Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz” - ESALQ/ USP, sob orientação do professor Dr. Urineu Umberto Packer (*In memoriam*), e o professor Dr. Gerson Barreto Mourão, defendendo a dissertação no início de 2010. No mestrado foi bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP. Em março de 2010, ingressou no curso de Doutorado em Ciências pela Universidade de São Paulo, Escola Superior de Agricultura 'Luiz de Queiroz' - ESALQ/ USP, sob orientação dos professores Dr. Gerson Barreto Mourão e Dr. Luiz Lehmann Coutinho. No doutorado foi bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP. Em 2013 realizou estágio supervisionado no exterior, na Universidade de Wisconsin, Madison, WI, sob supervisão do Professor Dr. Guilherme J. G. Rosa.

EPÍGRAFE

“A fé é a certeza de coisas que se esperam, e a convicção de fatos que se não vêem.”

Hebreus 11.1

“Quando os nossos dias se tornarem obscurecidos por nuvens negras e baixas, quando as nossas noites forem mais negras do que mil noites. Lembremo-nos, que no universo há um grande e benigno poder, que é capaz de abrir caminho onde não há caminho, e de transformar o ontem sombrio num luminoso amanhã”

Martin Luther King

"É melhor tentar e falhar, que preocupar-se e ver a vida passar; é melhor tentar ainda que em vão, que sentar-se fazendo nada até o final; eu prefiro nas chuvas caminhar, do que em dias tristes em casa me esconder; prefiro ser feliz embora louco, do que em conformidade viver".

Martin Luther King

SUMÁRIO

RESUMO.....	13
ABSTRACT.....	15
1 INTRODUÇÃO	17
Referências.....	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Melhoramento genético para características de temperamento	21
2.2 Desequilíbrio de ligação em população de animais	23
2.2.1 Definição e considerações iniciais	23
2.2.2 Medidas de desequilíbrio de ligação.....	24
2.3 Imputação.....	26
2.3.1 Definição e considerações iniciais	26
2.3.2 Imputação de animais relacionados	28
2.3.3 Imputação de animais não relacionados	29
2.3.4 Fatores que influenciam a acurácia de imputação	33
Referências.....	34
3 ASSOCIAÇÃO GENÔMICA PARA CARACTERÍSTICAS DE TEMPERAMENTO EM ANIMAIS DA RAÇA NELORE.....	41
Resumo	41
Abstract.....	41
3.1 Introdução	42
3.2 Material e Métodos	44
3.3 Resultados e Discussão	53
3.4 Conclusões	73
Agradecimentos	73
Referências.....	73
4 ACURÁCIA DE IMPUTAÇÃO DE GENÓTIPOS EM BOVINOS NELORE NO BRASIL	81
Resumo	81
Abstract.....	81
4.1 Introdução	82
4.2 Material e Métodos	83
4.3 Resultados e Discussão	91

4.4 Conclusões	95
Agradecimentos	95
Referências.....	96

RESUMO

Estratégias de imputação de genótipos de marcadores SNP para estudos de associação genômica em animais da raça Nelore

Temperamento em bovinos é definido como a reação dos animais em resposta ao contato com o ser humano, geralmente atribuído ao medo ocasionado no manejo. Animais agitados são mais difíceis de manejar nas fazendas. Estudos na literatura mostram que características de temperamento influenciam o desempenho produtivo dos rebanhos, devido à sua correlação com outras características dentre elas, ganho de peso diário, taxa de prenhez, qualidade e rendimento de carcaça. Todavia, estudos visando o melhor entendimento dos mecanismos biológicos, e da arquitetura genética destas características são escassos. Com a evolução das tecnologias de genotipagem em alta densidade com marcadores de polimorfismo único (SNP), observou-se um aumento de pesquisas nas áreas de seleção genômica (GS) e associação genômica (GWAS). Entretanto, devido ao alto custo que a genotipagem em larga escala apresenta, torna-se inviável a genotipagem para todos os animais candidatos à seleção. Porém, existe a possibilidade de utilizar painéis de baixa densidade de SNP e inferir estes genótipos desconhecidos para painéis de alta densidade. Neste contexto foram desenvolvidos dois estudos. No primeiro, o objetivo foi identificar regiões cromossômicas associadas com características de temperamento em animais da raça Nelore. As características avaliadas foram: velocidade de saída (VS) e mediana do escore composto (EC_mediana), em que 599 e 575 animais foram utilizados, respectivamente. Todos os animais foram genotipados com Illumina BovineHD BeadChip (800K), e para o GWAS dois modelos estatísticos foram aplicados. O primeiro modelo utilizado foi genômico de única etapa (ssGBLUP), em que os efeitos dos SNP são derivados da predição dos valores genéticos genômicos dos animais. O segundo modelo foi linear misto, similar ao modelo anterior, porém várias análises, SNP por SNP (regressão simples) foram realizadas. Nos dois modelos foram incluídos os efeitos de grupo de contemporâneo como fixo, idade do animal como covariável e efeito poligênico do animal e ambiente permanente como aleatórios. Os componentes de variância foram estimados pelo método de máxima verossimilhança restrita. Os coeficientes de herdabilidade apresentaram baixa magnitude com estimativas de 0,02 e 0,05 para VS e EC_mediana, respectivamente. Diferentes regiões cromossômicas foram associados com as características estudadas nesta população, de acordo com os modelos utilizados, contribuindo para o entendimento da arquitetura genética dessas características. No segundo, o objetivo foi avaliar a acurácia de imputação também nesta população, usando dois painéis de baixa densidade (3K e 6K) e um de média densidade (50K) imputados para o painel de alta densidade (800K). Os animais genotipados com 800K foram divididos em população de validação e população de referência. Na população de validação, os animais tiveram seus genótipos mascarados para os chips Illumina 3K, 6K e 50K. A imputação de 3K, 6K e 50K para 800K foi realizada utilizando os softwares de imputação fastPHASE e Beagle. O software fastPHASE apresentou maiores valores de acurácia de imputação em comparação ao Beagle. O painéis 6K e 50K apresentaram maiores valores de acurácia.

Palavras-chave: Inferência Bayesiana; Gado de corte; Genotipagem; Painel de alta e baixa densidade

ABSTRACT

Strategies of genotypes imputation for genome-wide association studies in Nellore cattle

Temperament in cattle is generally defined as the reaction of an animal in response to contact with human, usually attributed to the fear. Animals agitated are harder to manage in farms. Studies in the literature reported that temperamental traits have been found to influence the productive performance of herds, due to its correlation with other traits such as carcass quality, daily gains, pregnancy rate and feed efficiency. However, almost nothing is known regarding the genetic landscape controlling temperamental animal variation in cattle. With the advances in genotyping technologies for high density genotyping platforms to markers of single nucleotide polymorphisms (SNP), there was an increase in research in the areas of genomic selection (GS) and genome-wide association studies (GWAS). However, genotyping many animals with a high-density marker panel can be expensive and economically unfeasible. An alternative in this context is to use a lower density marker panel on a larger number of animals, and impute the missing genotypes. In this context two studies were developed. In the first study, the objective was to identify chromosomal regions associated with temperament traits in Nellore cattle. The temperamental traits evaluated in this study were: exit velocity (VS) and the median score temperament (EC_mediana) assessed in 599 and 575 animals, respectively. All animals were genotyped with Illumina BeadChip BovineHD (800K), and two GWAS statistical models were applied. The first model used was genomic single step (ssGBLUP) in which the effects of SNP are derived from the genomic prediction values of the animals. The second was linear mixed model, similar to the previous model, but a series of models, one for each SNP (single regression), were performed. In both models were included the effects of contemporary group as fixed, the animal's age as a covariate effect and polygenic animal and permanent environment as random effect. Variance components were estimated by restricted maximum likelihood. The heritability coefficients showed low magnitude estimative of 0.02 and 0.05 for VS and EC_mediana, respectively. Different chromosomal regions were associated with the traits studied in this population, according with the models used, and moreover to contribute in the understanding of the genetic architecture of these traits. In the second study, the objective was to evaluate the accuracy of imputation in a same population, using two panels of low-density (3K and 6K) and one medium-density (50K) panel markers imputed up to higher density of 800K. The animals genotyped with 800K markers panel were split into a reference population and validation population. The validation population the animals had markers masked to Illumina chip 3K, 6K e 50K. Imputation from 3K, 6K and 50K up to 770K markers was performed using fastPHASE and Beagle. The software fastPHASE had higher imputation accuracy compared to Beagle. The 6K and 50K panels showed higher accuracy.

Keywords: Bayesian Inference; Beef Cattle; Genotyping Panel; High and Low Density

1 INTRODUÇÃO

Temperamento em bovinos pode ser definido como a reação dos animais em resposta ao contato com o ser humano, geralmente atribuídas ao medo ocasionado no momento do manejo dos animais nas fazendas (FORDYCE; GODDARD; SEIFERT, 1982; BURROW; DILLON, 1997). Animais mais dóceis, adaptados ao manejo e mais produtivos são desejáveis nos sistemas de produção. Inúmeros estudos têm mostrado que animais mais agitados podem influenciar o desempenho produtivo de outras características devido à correlação existente, tais como qualidade de carcaça, ganho de peso diário, reprodução, bem-estar e segurança dos animais e funcionários (SANT'ANNA et al., 2012). Em sistemas intensivos de produção de bovinos de corte, a avaliação do temperamento e a seleção de animais menos agressivos pode ser ainda mais importante, pois esses animais passam por um manejo frequente de coleta de dados de pesagens, medições, controle sanitário, manejo reprodutivo e outros (SILVEIRA et al., 2008).

Diversas metodologias para avaliar o temperamento em bovinos já estão bem descritas e consolidadas na literatura, porém são difíceis de serem mensuradas. Velocidade de saída (VS) é uma característica objetiva que avalia o tempo que o animal leva para sair da balança (BURROW; SEIFERT; CORBERT, 1988). Animais que possuem VS inferior são em geral mais calmos, sendo mais fáceis de manejar nas fazendas. Escore composto (EC) é uma característica subjetiva, sendo formada por diferentes reações dos animais quando manejados na balança. A característica EC depende muito das características deslocamento e tensão. O escore composto apresenta escala de 1 até 5 (variando de animal dócil a agressivo), porém existem outras escalas para esta mesma característica. Devido aos estudos com objetivo de quantificar a variação genética aditiva, e entender os fatores genéticos e ambientais que influenciam essas características de temperamento, novos critérios de seleção foram incorporados nos programas de melhoramento genético animal. As estimativas de herdabilidade para velocidade de saída variam entre 0,11-0,35 (HOPE et al., 2010; SANT'ANNA et al., 2012). Para escore composto valores de herdabilidade variam entre 0,17-0,34 (PIOVEZAN, 1998; FIGUEIREDO et al., 2005).

Com a disponibilidade do sequenciamento completo do genoma bovino em 2009 (ELSIK; TELLAM; WORLEY, 2009), inúmeros polimorfismos de nucleotídeos únicos (SNP) foram descobertos, e o comprimento do genoma apresenta 3 bilhões de pares de bases. Em virtude disso, observou-se nos últimos anos um grande avanço nas tecnologias de genotipagem por marcadores moleculares SNP (polimorfismos de base única) e,

consequentemente, diferentes painéis de chip destes marcadores tornaram-se comercialmente disponíveis (VAN TASSEL et al., 2008; MATUKUMALLI et al., 2009). Além disso, os estudos de associação genômica (GWAS), que visam à identificação de regiões cromossômicas associadas com características econômicas em bovinos, tornaram-se possíveis. Métodos estatísticos foram desenvolvidos para fazer uso destes painéis de SNP para detectar a associação entre os SNP, bem como estimar os valores genéticos genômicos dos animais para características economicamente importantes (MEUWISSEN; HAYES; GODDARD, 2001, WANG et al., 2012).

A genotipagem de todos os animais candidatos à seleção com painéis de marcadores SNP de alta densidade ainda é muito cara, tornando-se quase que inviável sua implementação em rebanhos comerciais e sua utilização em estudos de associação. Nos estudos de GWAS faz-se necessário uma amostra representativa de animais genotipados para, assim, encontrar alguma associação com a variação causal (MISZTAL, 2011; MULDER et al., 2012). Por outro lado, a genotipagem dos animais com painéis de baixa (3K e 6K) à moderada (50K) densidade, acompanhado de métodos de imputação de genótipos de baixa densidade para alta densidade tanto em estudos de GWAS quanto em seleção genômica, têm apresentado resultados favoráveis (HAYES et al., 2011; VANRADEN et al., 2013).

A possibilidade de combinar painéis de baixa à moderada densidade de SNP e, posteriormente, imputar de tal forma que esses painéis se tornem em plataformas de alta densidade, como por exemplo, HD bead chip, disponível para bovinos com aproximadamente 777 mil SNP espalhados no genoma, pode ser utilizada para todas as características de interesse e diferentes raças (HABIER; FERNANDO; DEKKERS, 2009). Portanto, os estudos de seleção genômica tornaram-se possíveis, bem como o estudo de associação genômica para um grande número de animais provenientes de diferentes populações (ZHANG; DRUET, 2010). Em bovinos de corte, isso pode significar a genotipagem de reprodutores e ancestrais com o painel de alta densidade e, em seguida, a progênie poderá ter os seus genótipos imputados (HABIER; FERNANDO; DEKKERS, 2009; IWATA; JANNINK, 2010; MILLER, 2010).

Sendo assim, com o objetivo de contribuir para a seleção de animais menos agressivos nos sistemas de manejo nas fazendas, nesta tese propôs-se:

- (i) Identificar regiões cromossômicas associadas com características de temperamento em animais da raça Nelore;
- (ii) Avaliar a acurácia de imputação usando dois painéis de baixa densidade (3K e 6K) e um de moderada densidade (50K) para painéis de alta densidade (800K).

Referências

- BURROW, H.M.; DILLON, R.D. Relationships between temperament and growth in a feedlot and commercial carcass traits of *Bos indicus* crossbreds. **Australian Journal Experimental Agriculture**, Collingwood, v. 37, n. 4, p. 407-411, 1997
- BURROW, H.M.; SEIFERT, G.M.; CORBET, N.J. A new technique for measuring temperament in cattle. **Proceedings Australian Society of Animal Production**, Armidale, v. 17, p. 154-157, 1988.
- ELSIK, C.G.; TELLAM, R.L.; WORLEY, K.C. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, New York, v. 324, n. 5926, p. 522-528, 2009.
- FIGUEIREDO, L.G.G.; ELER, J.P.; MOURÃO, G.B.; FERRAZ, J.B.S.; BALIEIRO, J.C.C.; MATTOS, E.C. Análise genética do temperamento em uma população da raça Nelore. **Livestock Research for Rural Development**, Cali, v.17, n. 7, p. 1-7, 2005.
- FORDYCE, G.; GODDARD, M.E.; SEIFERT, G.W. The measurement of temperament in cattle and effect of experience and genotype. **Proceeding Australian Society of Animal Production**, Armidale, v. 14, p. 329-332, 1982.
- HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. Genomic selection using low-density SNP. **Genetics**, Bethesda, v. 182, p. 343–353, 2009.
- 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**, Malden, v. 43, n. 1, p. 72-80, 2011.
- HOPPE, S.; BRANDT, H.R.; KONIG, S.; ERHARDT, G.; GAULY, M. Temperament traits of beef calves measured under field conditions and their relationships to performance. **Journal of Animal Science**, Savoy, v. 88, n.6, pp.1982–1989, 2010.
- IWATA, H.; JANNINK, J.L. Marker Genotype Imputation in a Low-Marker-Density Panel with a High-Marker-Density Reference Panel: Accuracy Evaluation in Barley Breeding Lines. **Crop Science**, Madison, v. 50, n. 4, p. 1269-1278, 2010.
- MATUKUMALLI, L.K.; LAWLEY, C.T.; SCHNABEL, R.D.; TAYLOR, J.F.; ALLAN, M.F.; HEATON, M.P.; O'CONNELL, J.; MOORE, S.S.; SMITH, T.P.L.; SONSTEGARD, T.S.; VAN TASSELL, C.P. Development and characterization of a high density SNP genotyping assay for cattle. **Plos One**, San Francisco, v. 4, n. 4, p. 1-13, 2009.
- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.
- MILLER, S. Genetic improvement of beef cattle through opportunities in genomics. **Revista Brasileira de Zootecnia**, Viçosa, v. 39, p. 247-255, 2010.
- MISZTAL, I. FAQ for genomic selection. **Journal of Animal Breeding and Genetics**, Ames, v. 128, n. 4, p. 245–246, 2011.

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. **Journal of Dairy Science**, Lancaster, v. 95, n. 2, p. 876-889, 2012.

PIOVEZAN, U. **Análise de fatores genéticos e ambientais na reatividade de quatro raças de bovinos de corte ao manejo**. 1998. 51 p. Dissertação (Mestrado em Zootecnia – Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 1998.

SANT'ANNA, A.C.; PARANHOS DA COSTA, M.J.; BALDI, F.; RUEDA, P.M.; ALBUQUERQUE, L.G. Genetic associations between flight speed and growth traits in Nellore cattle. **Journal of Animal Science**, Savoy, v. 90, n.10, p. 427-3432, 2012.

SILVEIRA, I.D.B.; FISCHER, V.; FARINATTI, L.H.E.; RESTLE, J.; ALVES FILHO, D.C. Relação entre genótipos e temperamento de novilhos Charolês x Nelore em confinamento. **Revista Brasileira de Zootecnia**, Viçosa, v. 37, n. 10, p. 1808-1814, 2008.

VAN TASSELL, C.P.; SMITH, T.P.; MATUKUMALLI, L.K.; TAYLOR, J.F.; SCHNABEL, R.D.; LAWLEY, C.T.; HAUDENSCHILD, C.; MOORE, S.S.; WARREN, W.C.; SONSTEGARD, T.S. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nature Methods**, New York, v. 5, p. 247-252, 2008.

VANRADEN, P.M.; NULL, D.J.; SARGOLZAEI, M.; WIGGANS, G.R.; TOOKER, M.E.; COLE, J.B.; SONSTEGARD, T.S.; CONNOR, E.E.; WINTERS, M.; VAN KAAM, J.B.C.H.M.; VALENTINI, A.; VAN DOORMAAL, B.J.; FAUST, M.A.; DOAK, G.A. Genomic imputation and evaluation using high-density Holstein genotypes. **Journal of Dairy Science**, Lancaster, v. 96, n.1, p. 668–678, 2013.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W.M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, Cambridge, v. 94, n. 2, p. 73-83, 2012.

ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **Journal of Dairy Science**, Lancaster, v. 93, n. 11, p. 5487-5494, 2010.

2 REVISÃO BIBLIOGRÁFICA

2.1 Melhoramento genético para características de temperamento

Os programas de avaliação genética de bovinos de corte no Brasil contemplam várias características de crescimento, pois são obtidas facilmente e apresentam estimativas de herdabilidade de média a alta, sugerindo elevado progresso genético pela seleção (Alencar, 2002). Além dessas características, outros critérios de seleção começam a ganhar importância nos programas de seleção, entre elas podemos destacar maciez da carne e temperamento dos animais. As características de temperamento podem responder à seleção e apresentar correlações favoráveis com outras características relacionadas à eficiência produtiva e à qualidade da carne. Hall et al. (2011) observaram uma correlação negativa ($r = -0,187$, $p = 0,030$) entre velocidade de saída (VS) com força de cisalhamento para animais cruzados Red Angus, Simental, Charolais, Hereford, Gelbviegh.

Com o intuito de geração e adaptação de tecnologias, capazes de elevar os índices produtivos da bovinocultura de corte, foi desenvolvido o projeto que está inserido na rede “Bife Qualis”, financiada pela Embrapa e liderada pela Dra. Luciana Correia de Almeida Regitano, sendo parte do plano de ação: “Avaliação de marcadores moleculares e de genes candidatos para características de qualidade da carne e de eficiência alimentar na raça Nelore”. No âmbito desta projeto, desde o ano de 2006, foi desenvolvida uma população de animais da raça Nelore, em delineamento de meios-irmãos, que foi avaliada para medidas relacionadas à eficiência alimentar, produção, qualidade da carne, temperamento entre outras. Essas características em conjunto foram utilizadas em estudos quantitativos e moleculares, visando à recomendação de critérios de seleção para melhoramento da raça Nelore.

As características relacionadas à qualidade de carcaça, vêm assumindo grande importância devido à crescente exigência dos frigoríficos e consumidores por carcaças uniformes com alta rentabilidade, poucas perdas por hematomas e bem estar animal (Barrozo, 2012). Durante o manejo pré-abate animais mais agitados podem interferir de forma direta na qualidade de carne. O manejo na fazenda, embarque, transporte, mudança de ambiente, condições de confinamento, fome, sede e machucados afetam de forma significativa a fisiologia do animal podendo o levar a uma condição de estresse (FERGUSON; WARNER, 2008). Jeremiah et al. (1988) demonstraram que a carne de animais submetidos a melhores condições pré-abate, transportados com uma menor distância e menor tempo de espera para o abate, foi melhor avaliada em um painel de aceitação pelos consumidores.

O temperamento animal está relacionado com características individuais que envolvem medo, curiosidade, docilidade (PARANHOS DA COSTA et al., 2002). Existe dificuldade em se definir uma medida de temperamento para produtores e pesquisadores que seja possível de se medir a campo (PIOVESAN, 1998) e considerar o manejo individual e em grupo (GRANDIN, 2003). Burrow e Dillon (1997) relataram que as diferenças individuais de temperamento são influenciadas pela genética, e fatores ambientais tais como idade, sexo, manejo dentre outros.

Silveira et al. (2008) observaram que animais cruzados (Angus x Nelore) mais estressados apresentaram menor ganho de peso e maior velocidade inicial da glicólise anaeróbica, afetando assim, a qualidade de carne. Burrow e Dillon (1997) avaliaram VS também em animais cruzados (5/8 Brahman x 3/8 Shorthorn) e concluíram que animais que apresentaram menor VS ganharam mais peso. Silveira et al. (2008) concluíram que independentemente do grupo genético, animais mais reativos apresentam menor ganho de peso, e que animais com maior proporção racial de Nelore são mais reativos.

As características de temperamento apresentam valores de herdabilidade de magnitude baixa à moderada. De acordo com Piovezan (1998), as estimativas de herdabilidade encontradas para as características de temperamento relatadas na literatura variam conforme a metodologia de análise estatística, a população e a raça (Tabela 1). Vários programas de melhoramento têm incluído estas características nos critérios de seleção e têm encontrado resposta à seleção

Tabela 1 - Coeficientes de herdabilidade para algumas características de temperamento avaliadas em diferentes raças de bovinos

Referência	Característica	Grupo Genético	Herdabilidade
Paranhos da Costa et al. (2002)	Escore de agitação	Nelore, Gir, Guzerá e Caracu	0,34
Figueiredo et al. (2005)	Escore de temperamento	Nelore	0,17
Carneiro et al. (2006)	Teste de docilidade	Nelore	0,20
Mourão et al. (2008)	Contenção no tronco	Zebu x Holandês	0,06-0,27
Balieiro et al. (2008)	Distância de fuga	Nelore	0,15
Hope et al. (2010)	Velocidade de saída	Angus	0,20 ± 0,08
		Charolês	0,25 ± 0,10
		Hereford	0,36 ± 0,06
		Limousin	0,11 ± 0,07
		Simental	0,28 ± 0,07
Barroso et al. (2011)	Escore de Temperamento	Nelore	0,18 ± 0,02
Sant' Anna et al. (2013)	Velocidade de saída	Nelore	0,35

2.2 Desequilíbrio de ligação em populações de animais

2.2.1 Definição e considerações iniciais

O desequilíbrio de ligação (LD) é definido como uma associação não aleatória entre alelos de diferentes locos, ocorrendo dentro de uma população, o que resulta em frequência maior do que a esperada de determinados haplótipos (BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010; BRITO et al., 2011; LEE et al., 2011). A estimação do LD é fundamental para determinar o número de marcadores que serão necessários para o mapeamento de QTL. Na Figura 1 é apresentada uma situação em que podemos observar desequilíbrio de ligação entre os alelos no mesmo cromossomo (B), e outra em que os alelos estão associados aleatoriamente, ou seja, encontram-se em equilíbrio de ligação (A).

Figura 1 – Considere dois marcadores, A e B e eles encontram-se no mesmo cromossomo. A tem os alelos A e a, e B tem os alelos B e b. Sendo que quatro haplótipos são possíveis (AB, Ab, aB, ab). Se os dois locos estão em equilíbrio de ligação (A) poderemos observar as quatro possibilidades de haplótipos na população. Por outro lado, se os locos estiverem em LD (B) apenas alguns locos em associação serão observados (AB, ab), (MACKAY, 2001)

O conhecimento da extensão e padrão do LD em regiões genômicas permite explorar o grau de diversidade entre as raças de bovinos de corte, inferir sobre a distribuição de crossing-over, e identificar regiões do genoma sob pressão de seleção. A grande resolução do mapa de LD por meio da genotipagem em larga escala pode fornecer informações benéficas sobre o tamanho do painel de SNP e poder estatístico a ser utilizado nos estudos de associação genômica (MEUWISSEN; HAYES; GODDARD, 2001, BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010; LEE et al., 2011).

A extensão do LD no genoma depende fortemente de vários fatores, entre eles podemos citar: a história e tamanho efetivo da população de interesse estudada (populações recém formadas ou isoladas apresentam maior nível de LD), densidade e distribuição de marcadores (é sabido que quanto maior a distância entre marcadores, a tendência do LD tende a diminuir devido aos eventos de recombinação), tipo de método utilizado para reconstrução de haplótipos, a utilização de apenas haplótipos maternos ou ambos os haplótipos paternos e maternos, entre outros fatores (KHATKAR, et al., 2008; MCKAY et al., 2007; BRITO et al., 2011; BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010; LEE et al., 2011).

2.2.2 Medidas de desequilíbrio de ligação

Na literatura podemos observar inúmeras medidas estatísticas propostas para calcular o desequilíbrio de ligação nas populações, entretanto, as mais utilizadas para marcadores bialélicos como os do tipo SNP são: coeficiente de correlação entre alelos de dois locos (r^2) (HILL; ROBERTSON, 1968) e valor padronizado de D $|D'|$ (LEWONTIN, 1964). Ambas as medidas variam no intervalo de 0 a 1. O r^2 é definido como:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

em que, $D^2 = f(AB) - f(A)f(B)$, representa o quadrado do parâmetro de desequilíbrio de ligação (D), e $f(AB)$, $f(A)$, $f(a)$, $f(B)$, $f(b)$ são as frequências observadas dos haplótipos AB e dos alelos A, a, B, e b, respectivamente. As estimativas para D' são dadas por:

$$D = \text{freq}(A1_B1) * \text{freq}(A2_B2) - \text{freq}(A1_B2) * \text{freq}(A2_B1)$$

em que $\text{freq}(A1_B1)$ é a frequência do haplótipo A1_B1 na população. A estatística D é dependente das frequências dos alelos individuais e por isso não é útil para LD entre os vários pares de locos (LEE et al., 2011), em que:

$$D' = \frac{|D|}{D_{\max}}$$

em que $D_{\max} = \min(\text{freq}(A1) * \text{freq}(B2), -1 * \text{freq}(A2) * \text{freq}(B1))$ se $D > 0$, caso contrário $D_{\max} = \min(\text{freq}(A1) * \text{freq}(B1), -1 * \text{freq}(A2) * \text{freq}(B2))$ se $D < 0$

A medida de LD, $|D'| < 1$, indica que o histórico da recombinação ocorreu entre os locos. $|D'| = 1$ indica ausência de recombinação entre os loci desde a ocorrência de um polimorfismo (BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010).

Embora o r^2 e $|D'|$ sejam as duas medidas estatísticas mais utilizadas nos estudos de LD em animais, alguns cuidados devem ser tomados. O $|D'|$ é mais um indicador de haplótipos perdidos do que uma medida de LD (GAULT; LONG, 2003). Além disto, $|D'|$ tende a ser fortemente inflacionada com pequenas amostras e na presença de alelos raros (TEARE et al., 2002). O r^2 é mais adequado para estimar o LD para marcadores bialélicos, como os SNP, porque o r^2 é menos sensível a frequências alélicas e tamanho da amostra do que o $|D'|$ (BRITO et al., 2011; LEE et al., 2011). O r^2 é preferido para estudos de associação, porque não há uma simples relação inversa entre o r^2 e o tamanho da amostra que é necessário para detectar a associação entre um QTL e o SNP (WALL; PRITCHARD, 2003; BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010).

LEE et al. (2011) avaliando o LD no genoma completo de animais de uma população de gado de corte nativa da Coreia (Hanwoo), os autores observaram que o valor médio de LD (r^2) variou de 0,07 à 0,17 para os cromossomos BTA28 e BTA2, respectivamente. Para $|D'|$ variou de 0,35 para o cromossomo BTA18 à 0,52 para os cromossomos BTA10, BTA11 e

BTA12. A proporção de SNP em LD ($r^2 \geq 0,25$) foi de 20% para a distância de 10 e 20 kb entre SNP.

Khatkar et al. (2008) estudaram o LD em 1546 touros da raça holandesa, originados da Austrália, usando 9195 SNP. Estes autores avaliaram o LD com três diferentes medidas: r^2 , D' e D_{vol} (volume da medida de D'). Todas as medidas foram afetadas pelo tamanho da amostra. Segundo estes autores, a medida r^2 é mais confiável com amostra pequena ($n = 75$ animais). Ainda neste estudo, Khatkar et al. (2008) concluíram que são necessários mais de 300.000 SNP, em estudos de associação, considerando a extensão de LD baseada em r^2 .

Bohmanova, Sargolzaei e Schenkel (2010) avaliaram o tamanho do viés para as medidas de LD quando utilizado $|D'|$ e r^2 , sobre haplótipos maternal, paternal e ambos. Neste estudo 887 animais pertencentes à raça holandesa e genotipados com 38.590 SNP foram avaliados. As médias dos níveis de $|D'|$ e r^2 para os marcadores foram 0,72 e 0,20 considerando uma distância de 40 e 60 kb, respectivamente. A utilização de apenas haplótipos herdados da mãe é recomendada para análises de LD em populações consistindo de grandes famílias de meios-irmãos paternos. Grande heterogeneidade no padrão e na medida de LD em bovinos holandês foi observada em ambos os autossomos e no cromossomo X. A extensão da LD foi maior no cromossomo X em comparação com os autossomos. Ainda neste trabalho, não foi observado diferenças na extensão e declínio de LD entre regiões intragênica e intergênicas.

Recentemente Espigolan et al. (2013) estudaram o LD em animais da raça Nelore. Os autores utilizaram dados de 795 animais genotipados com o chip de 800K. Neste trabalho, o desequilíbrio de ligação entre dois SNP foi avaliado utilizando o r^2 e $|D'|$. Estes autores observaram valores de 0,17 e 0,52, para r^2 e $|D'|$, respectivamente. O r^2 diminuiu com o aumento da distância física entre os marcadores de 0,34 (1 kb) à 0,11 (100 kb). Os cromossomos BTA1, BTA27, BTA28 e BTA29 apresentaram menores níveis de LD a qualquer distância entre os marcadores.

2.3 Imputação

2.3.1 Definição e considerações iniciais

Imputação de genótipos é a predição de genótipos de marcadores faltantes ou desconhecidos que não foram genotipados em uma amostra de indivíduos de uma determinada população em estudo. A imputação destes genótipos pode ser usada para

aumentar o número de marcadores que poderão ser utilizados nos estudos de associação genômica (*genome-wide association* - GWAS) (MARCHINI; HOWIE, 2010). Da mesma forma, estes genótipos imputados poderão ser utilizados nos estudos de seleção genômica, com a finalidade de preenchimento da matriz de marcadores e, posteriormente, na estimação dos efeitos dos marcadores SNP para a predição dos valores genéticos dos animais candidatos à seleção.

Os estudos de imputação em estudos de GWAS foram inicialmente realizados em humanos (MARCHINI et al., 2007; MARCHINI; HOWIE, 2010; HAO et al., 2009; CHEN et al., 2012). Recentemente, estudos de imputação têm sido aplicados em animais com o objetivo da predição do mérito genético (DRUET; SCHROOTEN; DE ROOS, 2010; DRUET, ZHANG, 2010; WEIGEL et al., 2010a; HAYES et al., 2011; HICKEY et al., 2011; MULDER et al., 2012; VANRADEN et al., 2013). A genotipagem com chip de baixa densidade (3K e 6K e outras densidades) estão sendo utilizadas amplamente na genotipagem das fêmeas e animais jovens, com posterior imputação para 50K. Uma das maiores vantagens da utilização da imputação nos programas de melhoramento genético é a redução dos custos da genotipagem por indivíduo, e o aumento da acurácia de seleção (HABIER; FERNANDO; DEKKERS, 2009; MILLER, 2010; WEIGEL et al., 2010b; HAYES et al., 2011).

Outra realidade é a imputação de dados de sequência. Entretanto, o processo de imputação para esta dimensão ainda têm sido limitado, devido ao alto custo e demanda computacional dos dados de sequências completas, porém, já tem sido objeto de estudo. Os valores das acurácias de seleção nas avaliações genéticas têm aumentado por meio do aumento do número de animais na referência e na população de validação promovendo maior eficiência nos programas de melhoramento genético animal (MEUWISSEN; HAYES; GODDARD, 2001; WEIGEL et al., 2010a).

A maioria dos métodos de imputação fornecem as probabilidades de predição para os possíveis genótipos de cada marcador molecular imputado. Estas probabilidades podem ser utilizadas para prever os valores genômicos dos animais e nos estudos de associação genômica (HAYES et al., 2011). Os softwares de imputação estão classificados em três categorias:

(i) softwares baseados em desequilíbrio de ligação (LD) dos SNP na população. Geralmente, assumem que um grande grupo de indivíduos não apresentam relacionamento familiar, sendo denominados de softwares de base populacional (HOWIE; DONNELLY; MARCHINI, 2009), por exemplo, fastPHASE (SCHEET; STEPHENS, 2006), Beagle

(BROWNING; BROWNING, 2007) e IMPUTE (HOWIE; DONNELLY; MARCHINI, 2009);

(ii) softwares de base familiar que utilizam informação de pedigree e informação de fase de ligação. Para esses softwares assumem-se que ambos os pais e pelo menos um dos filhos tenham sido genotipados com alta densidade de marcadores (LI; JIANG, 2003; ZHANG; ZHAO, 2006). Por exemplo, o *software* AlphaImpute (HICKEY et al., 2011);

(iii) softwares que levam em consideração as duas informações para realizar a imputação, tanto informação de pedigree, quanto informação de LD. Por exemplo, FImpute (SARGOLZAEI; CHESNAIS; SCHENKEI, 2011) e o findhap (VANRADEN et al., 2011). O software Beagle também pode ser utilizado nas análises de imputação de trios, quando as informações de pai e mãe encontram-se disponíveis.

Vários softwares de imputação foram propostos e continuam sendo avaliados devido aos inúmeros fatores que podem afetar a acurácia de uma imputação. As estimativas de acurácias de imputação são influenciadas pela estrutura familiar, número de animais genotipados com alta densidade, LD, densidade do painel de marcador, pedigree, erros de genotipagem, dentre outros. Diferentes experimentos avaliaram a acurácia de imputação avaliando diferentes tamanhos de população de referência, utilização de haplótipos da população de referência, tagSNP, entre outros (MARCHINI; HOWIE, 2010; BADKE et al., 2013; CHEN et al., 2013). Maiores detalhes sobre as metodologias implementadas nos softwares de imputação e fatores que afetam a acurácia de imputação serão apresentados a seguir.

2.3.2 Imputação de animais relacionados

Nessa metodologia as informações de pedigree são usadas no modelo de imputação. Neste caso, o maior número de informações de familiares genotipados com painéis de marcadores de alta densidade poderá aumentar a acurácia de imputação. Este cenário de imputação tender a ser mais realista e acurado para os programas de melhoramento genético animal, pois as informações de parentes são importantes na seleção dos animais que serão utilizados como reprodutores na próxima geração. Neste caso, os touros e algumas vacas seriam genotipados com painéis de alta densidade e demais animais com painel de baixa densidade, atualmente disponível no mercado. Uma vez genotipados, estes animais têm seus haplótipos inferidos, e são utilizados para identificar regiões cromossômicas que compartilham em comum, conhecidos como idênticos por descendência (IBD), que derivam

de um ancestral em comum (DRUET; FARNI, 2011). A probabilidade de dois alelos serem IBD é uma informação muito importante, e pode ser utilizada em seleção genômica (MEUWISSEN; GODDARD, 2001), reconstrução de haplótipos e neste caso imputação (KONG et al., 2008).

No caso específico de imputação de animais relacionados, indivíduos que compartilham alguma região IBD serão mais semelhantes entre si do que indivíduos que não compartilham IBD. Entretanto, quando apenas o pai e filho forem genotipados, ou são heterozigotos para algum determinado SNP é mais difícil inferir sobre o haplótipo. Na Figura 2 é apresentada de maneira ilustrativa a imputação quando informações de parentes estão disponíveis.

Figura 2 - Imputação de genótipos dentro de uma amostra de indivíduos aparentados. (a) Os dados observados consistem de genótipos para qualquer marcador molecular. Neste caso, um subconjunto destes marcadores foram genotipados para todos os indivíduos (vermelho), enquanto que o restante foi identificado apenas para alguns indivíduos (preto – para as duas primeiras gerações do pedigree); (b) O processo de inferir a informação de IBD por meio dos marcadores para os quais os animais foram genotipados. Cada segmento de IBD que aparece em mais de um indivíduo recebe uma única cor. Por exemplo, um segmento marcado de cor azul é compartilhado entre o primeiro indivíduo na geração dos avós (no topo do pedigree), e o primeiro indivíduo da geração parental, e os indivíduos 3 e 4 na última geração (na parte inferior do pedigree); (c) Genótipos observados e informações de IBD foram combinados para preencher os genótipos desconhecidos ausentes na geração de descendência (última geração) (LI et al., 2009)

O software AlphaImpute (HICKEY et al., 2011) é um exemplo deste tipo de imputação em que se utiliza informações de pedigree. Esse *software* utiliza análise de segregação (KERR; KINGHORN, 1996) e biblioteca de imputação de genótipos (SAHLI) para imputar

alelos e genótipos (HICKEY et al., 2011). É uma combinação dos softwares do *AlphaPhase* e *GeneProbForAlphaPhase*. O AlphaImpute é baseado na informação de genótipos de animais aparentados segundo Kong et al. (2008) método conhecido como “*long range phasing*”.

2.3.3 Imputação de animais não relacionados

A imputação para animais não aparentados é baseada na mesma teoria utilizada para animais aparentados como explicado anteriormente. Entretanto a maior diferença entre estas duas metodologias é que quando indivíduos não são aparentados, regiões dos haplótipos compartilhados (IBD) serão menores em comparação com animais aparentados que compartilham longas regiões no haplótipo. A estrutura de LD e a inferência dos haplótipos podem fornecer informações sobre os SNP desconhecidos. Neste caso, o painel de referência deve ser muito informativo para alcançar boa acurácia (LI et al., 2009; PAUSCH et al., 2013). Na Figura 3 é ilustrada a imputação para indivíduos não aparentados.

Figura 3 – Imputação de genótipos dentro de uma amostra de indivíduos não aparentados. (a) Os dados observados consistem de genótipos para cada indivíduo a ser estudado e, informações dos haplótipos para indivíduos da população referência; (b) O processo de identificação das regiões cromossômicas compartilhadas (IBD) entre os indivíduos da população de validação e indivíduos do painel de referência. Com maiores indivíduos na população de referência esperam-se maiores segmentos compartilhados; (c) Genótipos observados e haplótipos compartilhados são combinados para preencher os genótipos desconhecidos na população de validação (LI et al., 2009).

Neste estudo, iremos apresentar dois exemplos de metodologia de imputação que utilizam informações de indivíduos não aparentados: fastPHASE e Beagle.

O software fastPHASE versão 1.2 (SCHEET; STEPHENS, 2006) é baseado em um modelo de agrupamento e usa a observação que os haplótipos relacionados ou similares dos indivíduos dentro da população se aglomeram em grupos. De acordo com Marchini e Howie (2010), o modelo especifica um conjunto de K estados não observados ou um grupo que é designado para representar os haplótipos comuns significativos para representar haplótipos em comum. O k -ésimo grupo recebe um peso (α_{kl}) que denota a proporção que este grupo de haplótipo é observado no loco l :

$$\sum_k \alpha_{kl} = 1$$

sendo assim, cada grupo possui uma frequência associada (θ_{kl}) para o alelo 1 em cada loco. Os dados dos genótipos de cada indivíduo são analisados por meio de HMM sobre estes estados com transição entre os estados controlados por um conjunto de parâmetros (r) para cada SNP. Dessa forma, a probabilidade de transição é dada por:

$$P(G_i | \alpha, \theta, r) = \sum_z P(G_i | Z_i, \theta) P(Z_i | \alpha, r)$$

em que $P(G_i | Z_i, \theta)$ modela o provável genótipo observado dado os estados de transição e $P(Z_i | \alpha, r)$ é a probabilidade de transição entre os estados, mas os estados representam grupos ao invés de haplótipos de referência. Portanto, a função de verossimilhança pode ser escrita como:

$$L(G, H | \alpha, \theta, r) = \prod P(G_i | \alpha, \theta, r) \prod P(H_i | \alpha, \theta, r)$$

O software fastPHASE além de prever genótipos desconhecidos, também infere haplótipos. O modelo considera um número fixo de haplótipos e cada um destes haplótipos de referência representa um estado de Markov oculto. O algoritmo de maximização da esperança, algoritmo EM, é usado para estimar os parâmetros genéticos, frequência dos haplótipos e fase desconhecida do haplótipo. Mais detalhes sobre a metodologia consultar Scheet e Stephens (2006).

O software Beagle versão 3.2.1 (BROWNING; BROWNING, 2007) é baseado no modelo gráfico DAG (*directed acyclic graph*) de um conjunto de haplótipos localizados, que utiliza um painel de referência para estimar a fase dos haplótipos e dos genótipos

desconhecidos no conjunto de animais de validação. Similar ao software fastPHASE, apresenta um modelo de agrupamento dos haplótipos semelhantes. Entretanto, a forma como estes grupos são formados é diferente. No fastPHASE o parâmetro K (número de grupos) é fixo, porém no Beagle o número de grupos pode variar. Esse último permite inferir a fase do haplótipo em grandes conjuntos de animais não aparentados e trios de pais e filhos quando disponível. Esse método tem como *default* dez iterações do modelo de construção. Uma explicação detalhada da implementação do software é dada em Browning e Browning (2007).

Os softwares fastPHASE e Beagle fornecem as probabilidades à *posteriori* dos possíveis genótipos correspondentes ao loco AA, AB e BB. A Figura 4 apresenta de maneira comparativa os modelos de imputação no fastPHASE (LI; STEPHENS, 2003) e Beagle (BROWNING, 2006).

Figura 4 – Resumo dos modelos com três marcadores (SNP $i-1$, i e $i+1$). As elipses são os estados ocultos (*hidden states*). Para Li e Stephens, esses estados são definidos pelos haplótipos de referência, enquanto que para o modelo de Browning os estados são agrupamentos de haplótipos de localizados. O modelo de Browning tende a ter menos estados em qualquer marcador que será irrestrita modelo baseado na Li e quadro Stephens, e o número de estados pode variar de marcador para marcador para o modelo de Browning, mas foi corrigido no Li e quadro Stephens. Setas entre os estados de um SNP para o próximo são transições do HMM. Em Li e Stephens, transições com maior probabilidade anterior (aqueles observados nos haplótipos de referência) estão apresentadas com setas em negrito, enquanto as setas finas permitem o histórico de recombinação. Para o modelo de Browning, existem na maioria das transições k que sai de um estado, em que k é o número de alelos na próxima marcador (isto é, 2 para os SNP), que ajuda a manter o modelo parcimonioso. Flechas que saem do topo dos estados são possíveis emissões do HMM, que são os alelos observados. Para o quadro de Li e Stephens, emissões com maior probabilidade anterior (alelos nos haplótipos de referência) são apresentadas com setas em negrito, enquanto as setas finas representam mutações para outros alelos. Os haplótipos de referência aqui são 011, 010, 101 e 001. Para o modelo de Browning, não é apenas uma das possíveis emissões a partir de cada estado, o que ajuda a manter o modelo parcimonioso (BROWNING, 2008)

2.3.4 Fatores que influenciam a acurácia de imputação

A eficiência da imputação pode ser afetada por inúmeros fatores, dentre eles enfatizam-se os seguintes:

- (i) densidade dos marcadores no chip de baixa densidade;
- (ii) relacionamento entre animais da população de teste e animais da população de referência;
- (iii) número de animais no painel de referência;
- (iv) nível de LD dentro da população (HICKEY et al., 2012);
- (v) metodologia de imputação (ZHANG; DRUET, 2010), entre outros fatores.

Pei et al. (2008) avaliaram o efeito do LD na acurácia de imputação utilizando vários *softwares*, dentre eles: MACH, IMPUTE, fastPHASE, PLINK e Beagle. Em todos os *softwares* de imputação a acurácia de imputação aumentou quando o nível de LD foi maior. Dado o tamanho da população de referência de 90 indivíduos, a acurácia foi 62,8%, 75,9% e 95,1% para os níveis baixo, médio e forte de LD para o MACH. Ainda neste trabalho, comparando os métodos utilizados de imputação os autores concluíram que MACH e IMPUTE apresentaram maiores valores de acurácia para todos os níveis de LD, com exceção do fastPHASE que apresentou maior estimativa em situação de baixo LD. Estes autores concluíram que a informação de LD é importante para obter eficiência de imputação.

Weigel et al. (2010) investigaram a acurácia de imputação em animais raça Jersey por meio do fastPHASE. Estes autores mostraram que utilizando algoritmos baseados em informação de LD e na ausência da informação de pedigree, a acurácia de imputação pode ser baixa à modesta com valores de 91, 87, 79 e 73% para os painéis de baixa densidade com 2.942, 1.468, 741 e 336 imputados para 50K, respectivamente.

Em ovinos, Hayes et al. (2011) observaram que a acurácia de imputação pode ser aumentada pelo grau de relacionamento entre os indivíduos na população de referência e os animais na população de validação. Estes autores concluíram que um aumento de 1% na média de relacionamento dos animais de referência e validação proporcionou um aumento de 10% na estimativa de acurácia. Ainda neste trabalho, verificaram que quando na referência foram usados animais mais relacionados a acurácia tornou-se muito ruim (correlação de acurácia de imputação perto de zero).

Outro resultado interessante foi apresentado por Zhang e Druet (2010), que avaliaram a acurácia de imputação em bovinos de leite usando cinco chips contendo 384, 768, 1.536,

3.000 e 6.000 SNP imputados para 50K por meio dos softwares DAGPHASE e CHROMIBD. Neste trabalho observaram que a percentagem do erro de acurácia de imputação variou de 11,7% com 384 SNP para 2% com 6.000 SNP por meio do software DAGPHASE. Da mesma forma foi observado para o software CHROMIBD percentagem de erro de 10,7% com 384 SNP com diminuição para 3,3% com 6.000 SNP. Estes autores concluíram que com o chip de 1.536 SNP o erro de imputação seria 5%.

Sun et al. (2012) compararam o desempenho de vários softwares para imputação entre eles: Beagle, IMPUTE, fastPHASE, AlphaImpute, findhap e FImpute, para imputar genótipos a partir do painel de 5K para 50K SNPs em animais de raça Angus. Os autores encontraram valores de acurácia mais altas de 0,86 e 0,98 para os programas Beagle e FImpute, respectivamente.

VanRaden et al. (2013), em um estudo semelhante com bovinos de leite estudaram a influência da densidade dos painéis sobre a acurácia de imputação a partir dos painéis de baixa densidade 3K, 6K e moderada densidade (50K) para o painel de alta densidade (800K). Os autores obtiveram aproximadamente 99, 96 e 94% de genótipos imputados corretamente do 50K, 6K e 3K para o 800K, respectivamente.

Jimenez-Montero et al. (2013) avaliaram o efeito da imputação sobre a acurácia das estimativas dos valores genéticos genômicos (GEBV) em bovinos de leite. Os autores avaliaram a acurácia de chips contendo 3K, 6K e 50K imputados para 800K, como também a imputação a partir de 3K e 6K para 800K, utilizando o software Beagle. Os autores concluíram que não houve diferença entre as eficiências de imputação nos cenários acima. Entretanto, foi observado um aumento de 4% na acurácia para o chip de 800K, quando comparado com 50K. Além disto, a soma de quadrado do resíduo foi menor em 2% na predição dos EBV no HD comparado com 50K.

Referências

BADKE, Y.M.; BATES, R.O.; ERNST, C.W.; SCHWAB, C.; VAN TASSELL, C.P.; STEIBE, J.P. Methods of tagSNP selection and other variables affecting imputation accuracy in swine. **BMC Genetics**. London, v. 14, n. 8, p. 1-14, 2013.

BALIEIRO, C.C.; ELER, J.P.; BALIEIRO, J.C.C.; FERRAZ, J.B.S.; MATTOS, E.; FIGUEIREDO, L.G. Estimativas de parâmetros genéticos para repelência, temperamento e características produtivas em bovinos da raça Nelore. In: Simpósio Nacional de Melhoramento Animal, 7, 2008. São Carlos-SP. **Anais...** São Carlos: SBMA, 2008.

BARROZO, D.; BUZANSKAS, M.E.; OLIVEIRA, J.A.; MUNARI, D. P. ; NEVES, H. H. R.; QUEIROZ, S. A. Genetic parameters and environmental effects on temperament score and reproductive traits of Nellore cattle, **Animal**, Cambridge, v. 6, n. 1, p. 36-40, 2012.

BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F.S. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, London, v. 11, n. 421, p. 1-11, 2010.

BRITO, F.V.; BRACCINI NETO, J.; SARGOLZAEI, M.; COBUCI, J.A.; SCHENKEL, F.S. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. **BMC Genetics**, London, v. 12, n. 80, p. 1-11, 2011.

BROWNING, S.R. Multilocus Mapping Using Variable-length Markov Chains. **The American Journal of Human Genetics**, Cambridge, v.78, p. 903-913, 2006.

BROWNING, S.R. Missing data Imputation and haplotype phase inference for genome-wide association studies **The American Journal of Human Genetics**, Cambridge, v. 124, n. 5, p. 439-450, 2008.

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. **The American Journal of Human Genetics**, Cambridge, v. 81, n. 5, p. 1084-1097, 2007.

BURROW, H.M.; DILLON, R. D. Relationships between temperament and growth in a feedlot and commercial carcass traits of *Bos indicus* crossbreds. **Australian Journal Experimental Agriculture**, Collingwood, v. 37, n. 4, p. 407-411, 1997.

CARNEIRO, R.L.R.; DIBIASI, N.F.; THOLON, P.; QUEIROZ, S.A.; FRIES, L.A. Estimative of heritability to temperament in Nellore cattle. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 2006, Belo Horizonte/MG. **Anais...** Belo Horizonte: SBMA/EMBRAPA/UFGM, 2006.

CHEN, J.; ZHANG, J.G.; LI, J.; PEI, Y.F.; DENG, H.W. On combining reference data to improve imputation accuracy. **PLoS ONE**, San Francisco, v. 8, n. 1, p. 1 - 7, 2013.

CHEN, M.H.; HUANG, J.; CHEN, W.M.; LARSON, M.G.; FOX, C.S.; VASAM, R.S.; SESHADRI, S.; O'DONNELL, C.J.; YANG, Q. Using family-based imputation in genome-wide association studies with large complex pedigrees: The Framingham Heart Study. **Plos One**, San Francisco, v. 7, n. 12, p. 1-11, 2012.

COOKE, R.F.; BOHNERT, D.W.; MENEGHETTI, M.; LOSI, T.C.; VASCONCELOS, J.L.M. Effects of temperament on pregnancy rates to fixed-timed AI in *Bos indicus* beef cows. **Livestock Science**, Montpellier, 142:108-113, 2011.

DRUET, T.; FARNIR, F.P. Modeling of identity-by-descent processes along a chromosome between haplotypes and their genotyped ancestors. **Genetics**, Bethesda, v. 188, n. 2, p. 409-419, 2011.

DRUET, T.; SCHROOTEN, C.; DE ROOS, A.P. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. **Journal of Dairy Science**, Lancaster, v. 93, n. 1, p. 5443-5454, 2010.

ESPIGOLAN, R.; BALDI, F.; BOLIGON, A.A.; SOUZA, F.R.P.; GORDO, D. G.M.; TONUSSI, R.L.; CARDOSO, D.F.; OLIVEIRA, H.N.; TONHATI, H.; SARGOLZAEI, M.; SCHENKEL, F.S.; CARVALHEIRO, R.; FERRO, J.A.; ALBUQUERQUE, L.G. Study of whole genome linkage disequilibrium in Nellore cattle. **BMC Genomics**, London, v. 14, n. 305, p. 1-8, 2013.

FERGUSON, D.M.; WARNER, R.D. Have we underestimated the impact of pre-slaughter stress on meat quality in ruminants? **Meat Science**, Barking, v. 80, n. 1, p. 12-19, 2008.

FIGUEIREDO, L.G.; ELER, J.P.; MOURÃO, G.B.; FERRAZ, J.B.S.; BALIEIRO, J.C.C.; MATTOS, E. Análise genética do temperamento em uma população da raça Nelore. **Livestock Research for Rural Development**, Cali, v. 17, n. 84, p. 1-7, 2005.

FORDYCE, G.; DODT, R.M.; WYTHES, J.R. Cattle temperaments in extensive beef herds in northern Queensland. 1. Factors affecting temperament. **Australian Journal of Experimental Agriculture**, East Melbourne, v. 28, n. 6, p. 683-687, 1988.

GAUT, B.S.; LONG, A.D. The lowdown on linkage disequilibrium. **The Plant Cell**, Baltimore, v. 15, n. 7, p. 1502-6, 2003.

GRANDIN, T. Behavioral agitation during handling of cattle is persistent over time. **Journal Applied Animal Behaviour Science**, Philadelphia, v. 36, n. 1, p. 1-9, 1993.

_____. Transferring results of behavioral research to industry to improve animal welfare on the farm, ranch and the slaughter plant. **Journal Applied Animal Behavior Science**, Philadelphia, n. 3, p. 215-228, 2003.

HABIER, D.; FERNANDO, R.L.; DEKKERS, J.C.M. Genomic selection using low-density marker panels. **Genetics**, Bethesda, v. 182, n. 1, p. 343-353, 2009.

HALL, N.L.; BUCHANAN, D.S.; ANDERSON, V.L.; ILSE, B.R.; CARLIN, K.R.; BERG, E.P. Working chute behavior of feedlot cattle can be an indication of cattle temperament and beef carcass composition and quality. **Meat Science**, Barking, v. 89, n. 1, p. 52-57, 2011.

HAO, K.; CHUDIN, E.; McELWEE, J.; SCHADT, E.E. Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. **BMC Genetics**, London, v. 10, n. 27, p. 1-10, 2009.

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**, Malden, v. 43, n. 1, p. 72-80, 2011.

HICKEY, J.M.; KINGHORN, B.P.; TIER, B.; WILSON, J.F.; DUNSTAN, N.; VAN DER WERF, J.H.J. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. **Genetics Selection Evolution**, London, v. 43, n. 12, p. 1-13, 2011.

HICKEY, J.M.; GROSSA, J.; BABU, R.; DE LOS CAMPOS, G. Factors Affecting the Accuracy of Genotype Imputation in Populations from Several Maize Breeding Programs. **Crop Science**, Madison, v. 52, n. 2, p. 654 – 663, 2012.

HILL, W.G.; ROBERTSON A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Heidelberg, v. 38, n. 6, p. 226-231, 1968.

HOPPE, S.; BRANDT, H.R.; KONIG, S.; ERHARDT, G.; GAULY, M. Temperament traits of beef calves measured under field conditions and their relationships to performance. **Journal of Animal Science**, Savoy, v.88, n.6, p.1982–1989, 2010.

HOWIE, B.N.; DONNELLY, P.; MARCHINI, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. **Plos Genetics**, San Francisco, v. 5, n. 6, p. 1-15, 2009.

JIMÉNEZ-MONTERO, J. A.; GIANOLA, D.; WEIGEL, K.; ALENDA, R.; GONZALOZ-RECIO, O. Assets of imputation to ultra-high density for productive and functional traits. **Journal of Dairy Science**, Lancaster, v. 96, n. 9, p. 1-12, 2013.

JEREMIAH, L.E.; NEWMAN, J.A.; TONG, A.K.W.; GIBSON, L.L. The effects of castration preslaughter stress and zeranol implants on beef: Part 1. The texture of loin steaks from bovine males. **Meat Science**, Barking, v. 22, n. 2, p. 83-101, 1988.

KERR, R.J.; KINGHORN, B.P. An efficient algorithm for segregation analysis in large populations. **Journal of Animal Breeding Genetics**, Ames, v. 113, n. 4, p. 457–469, 1996.

KHATKAR, M.S.; NICHOLAS, F.W.; COLLINS, A.R.; ZENGER, K.R.; CAVANAGH, J.A.L.; BARRIS, W.; SCHNABEL, R.D.; TAYLOR, J.F; RAADSMA, H.W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high density SNP panel. **BMC Genomics**, London, v. 9, n. 187, p. 1-18, 2008.

KONG, A.; MASSON, G.; FRIGGE, M.L.; GYLFASSON, A.; ZUSMANOVICH, P.; THORLEIFSSON, G.; OLASON, P.I.; INGASON, A.; STEINBERG, S.; RAFNAR, T.; SULEM, P.; MOUY, M.; JONSSON, F.; THORSTEINSDOTTIR, U.; GUDBJARTSSON, D.F.; STEFANSSON, H.; STEFANSSON, K. Detection of sharing by descent, long-range phasing and haplotype imputation. **Nature Genetics**, New York, v. 40, n. 9, p. 1068-1075, 2008.

LEE, S.H.; CHO, Y.M.; LIM, D.; KIM, H.C.; CHOI, B.H.; PARK, H.S.; KIM, O.H.; KIM, S.; KIM, T.H.; YOON, D.; HONG, S.K. Linkage disequilibrium and effective population size in Hanwoo Korean cattle. **Asian-Australian Journal of Animal Sciences**, Seoul, v. 24, n. 12, p. 1660-1665, 2011.

LEWONTIN, R.C. The interaction of selection and linkage. I. General considerations; heterotic model. **Genetics**, Bethesda, v. 49, n. 1, p. 49-67, 1964.

LI J.; JIANG, T. Efficient inference of haplotypes from genotypes on a pedigree. **Journal of Bioinformatics and Computational Biology**, London, v. 1, n. 1, p. 41-69, 2003.

LI, N.; STEPHENS, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. **Genetics**, Bethesda, v. 165, n. 4, p. 2213-2233, 2003.

LI, Y.; WILLER, C.; SANNA, S.; ABECASIS, G. Genotype imputation. **Annual Review of Genomics and Human Genetics**, Palo Alto, v. 10, n. 1, p. 387-406, 2009.

MACKAY, T.F.C. Quantitative trait loci in *Drosophila*. **Nature Reviews Genetic**, London, v. 2, n. 1, p. 11-20, 2001.

MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. **Nature Reviews Genetics**, London, v. 11, n. 7, 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. **Nature Genetic**, New York, v. 39, n. 7, p. 906-913, 2007.

McKAY, S.D.; SCHNABEL, R.D.; MURDOCH, B.M.; MATUKUMALLI, L.K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; DIAS NETO, E.; GILL, C.A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C.P.; WILLIAM, J.L.; TAYLOR, J.F.; MOORE, S.S. Whole genome linkage disequilibrium maps in cattle. **BMC Genetic**, London, v. 8, n. 74, p. 1-12, 2007.

MEUWISSEN, T.H.E.; GODDARD, M.E. Prediction of identity by descent probabilities from marker-haplotypes. **Genetic Selection Evolution**, London, v. 33, n. 6, p. 605-34, 2001.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.

MILLER, S. Genetic improvement of beef cattle through opportunities in genomics. **Revista Brasileira de Zootecnia**, Viçosa, v. 39, n. 1, p. 247-255, 2010.

MOURÃO, G.B.; BERGMANN, J.A.G.; FERREIRA, M.B.D. Diferenças genéticas e estimação de coeficientes de herdabilidade para temperamento em fêmeas zebras e F1 Holandês x Zebu. **Revista Brasileira Zootecnia**, Viçosa, v. 27, n. 4, p. 722-729, 1998.

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. **Journal of Dairy Science**, Lancaster, v. 95, n. 2, p. 876-889, 2012.

PARANHOS DA COSTA, M.J.R.; COSTA E SILVA, E.V.; CHIQUITELLI NETO, M.; ROSA, M.S. Contribuição dos estudos de comportamento de bovinos para implementação de programas de qualidade de carne. In: ENCONTRO ANUAL DE ETOLOGIA, 20., 2002, Natal. **Palestras...** Natal: Sociedade Brasileira de Etologia, 2002. p. 71-89.

PAUSCH, H.; AIGNER, B.; EMMERLING, R.; EDEL, C.; GOETZ, K.U.; FRIES, R. Imputation of high-density genotypes in the Fleckvieh cattle population. **Genetics Selection Evolution**, London, v. 45, n. 3, p. 1-10, 2013.

PEI, Y.; LI, J.; ZHANG, L.; PAPASIAN, C.J.; DENG, H. Analyses and comparison of accuracy of different genotype imputation methods. **PLoS ONE**, San Francisco, v. 3, n. 10, p. 1-8, 2008.

PIOVEZAN, U. **Análise de fatores genéticos e ambientais na reatividade de quatro raças de bovinos de corte ao manejo**. 1998. 51 p. Dissertação (Mestrado em Zootecnia – Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 1998.

SANT'ANNA, A.C.; PARANHOS DA COSTA, M.J.; BALDI, F.; ALBUQUERQUE, L.G. Genetic variability for temperament indicators of Nellore cattle. **Journal of Animal Science**, Savoy, v. 91, n. 8, p. 3532-3537, 2013.

SARGOLZAEI, M.; CHESNAIS, J.P.; SCHENKEL, F.S. FImpute: an efficient imputation algorithm for dairy cattle populations. **Journal of Dairy Science**, Lancaster, v. 94, n. 421, 2011.

SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **American Journal of Human Genetics**, Cambridge, v. 78, n. 4, p. 629-644, 2006.

SILVEIRA, I.D.B.; FISCHER, V.; FARINATTI, L.H.E.; RESTLE, J.; ALVES FILHO, D.C. Relação entre genótipos e temperamento de novilhos Charolês x Nelore em confinamento. **Revista Brasileira de Zootecnia**, Viçosa, v. 37, n. 10, p. 1808-1814, 2008.

SUN, C.; WU, X.; WEIGEL, K.; ROSA, G.J.M.; BAUCK, S.; WOODWARD, B.W.; SCHNABEL, R.D.; TAYLOR, J.F.; GIANOLA, G. An ensemble-based approach to imputation of moderate-density genotypes for genomic selection with application to Angus cattle. **Genetics Research**, Cambridge, v. 94, p.133-150, 2012.

TEARE, M.D.; DUNNING, A.M.; DUROCHER, F.; RENNART, G.; EASTON, D.F. Sampling distribution of summary linkage disequilibrium measures. **Annals of Human Genetics**, London, v. 66, n.3, p. 223-233, 2002.

VANRADEN, P.M.; O'CONNELL, J.R.; WIGGANS, G.R.; WEIGEL, K.A. Genomic evaluations with many more genotypes. **Genetics Selection Evolution**, London, v. 43, n. 10, p. 1-11, 2011.

VANRADEN, P.M.; NULL, D.J.; SARGOLZAEI, M.; WIGGANS, G.R.; TOOKER, M.E.; COLE, J.B.; NSONSTEGARD, T.S.; CONNOR, E.E.; WINTERS, M.; VAN KAAM, J.B.C.H.M.; VALENTINI, A.; VAN DOORMAAL, B.J.; FAUST, M.A.; DOAK, G.A. Genomic imputation and evaluation using high-density Holstein genotypes. **Journal of Dairy Science**, Lancaster, v. 96, n. 1, p. 668–678, 2013.

WALL, J.D.; PRITCHARD, J.K. Haplotype blocks and linkage disequilibrium in the human genome. **Nature Reviews Genetics**, London, v. 4, n. 8, p. 587-597, 2003.

WEIGEL, K.A.; VAN TASSELL, C.P.; O'CONNELL, J.R.; VANRADEN, P.M.; WIGGANS, G.R. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. **Journal of**

Dairy Science, Lancaster, v. 93, n. 5, p. 2229-2238, 2010a.

WEIGEL, K.A.; DE LOS CAMPOS, G.; VAZQUEZ, A.I.; ROSA, G.J.M.; GIANOLA, D.; VAN TASSELL, C.P. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. **Journal of Dairy Science**, Lancaster, v. 93, n. 11, p. 5423–5435, 2010b.

ZHANG, K.; ZHAO, H. A comparison of several methods for haplotype frequency estimation and haplotype reconstruction for tightly linked markers from general pedigrees. **Genetics Epidemiology**, New York, v. 30, n. 5, p. 423 -437, 2006.

ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **Journal of Dairy Science**, Lancaster, v. 93, n. 11, p. 5487-5494, 2010.

3 ASSOCIAÇÃO GENÔMICA PARA CARACTERÍSTICAS DE TEMPERAMENTO EM ANIMAIS DA RAÇA NELORE

Resumo

Características de temperamento em bovinos podem influenciar o desempenho produtivo dos rebanhos, pois animais mais agitados apresentam menores ganhos de peso diário, menor rendimento e inferior qualidade de carcaça e dificuldade de manejo nas fazendas. Em geral, o manejo de animais mais calmos torna o trabalho na fazenda mais fácil. O avanço tecnológico na área da biologia molecular e da bioinformática, assim como a disponibilidade do sequenciamento do genoma bovino têm permitido estudos de associação genômica (GWAS) a partir da genotipagem em larga escala de polimorfismos de nucleotídeo único (SNP). Sendo assim, o objetivo deste estudo foi identificar regiões cromossômicas associadas com características de temperamento em bovinos da raça Nelore por meio de dois diferentes modelos estatísticos. As características de temperamento avaliadas foram velocidade de saída (VS) e mediana do escore composto (EC_mediana) em que 599 e 575 animais da raça Nelore foram utilizados respectivamente. Os animais foram genotipados com Illumina BovineHD BeadChip (800K) e para GWAS dois modelos estatísticos foram aplicados. O primeiro modelo utilizado foi o genômico de única etapa (ssGBLUP), em que os efeitos dos SNP são derivados da predição dos valores genéticos genômicos dos animais. Este modelo é baseado nas informações combinadas de pedigree tradicional e genômico e das informações fenotípicas. O segundo modelo foi linear misto, similar ao modelo anterior, porém várias análises, SNP por SNP (regressão simples) foram realizadas. Nos dois modelos foram incluídos os efeitos de grupo de contemporâneo como fixo, idade do animal como covariável e efeito poligênico do animal e ambiente permanente como aleatórios. Os componentes de variância foram estimados pelo método de máxima verossimilhança restrita. Os coeficientes de herdabilidade apresentaram baixa magnitude com estimativas de 0,02 e 0,05 para VS e EC_mediana, respectivamente. Diferentes regiões cromossômicas foram associadas às características estudadas nesta população, segundo os modelos utilizados auxiliando assim no entendimento da arquitetura genética destas características.

Palavras-chave: Gado de corte; GWAS; Herdabilidade; SNP; Temperamento

Abstract

Temperamental traits have been found to influence the productive performance of herds, because animals agitated to demonstrate lower daily gains, carcass yield and quality. In general, the calm animals management in farms to supply for easier work. The advances in molecular biology and bioinformatics, as well as the availability of the bovine genome sequencing have allowed genome-wide association studies (GWAS) from the high-throughput genotyping of single nucleotide polymorphisms (SNP). Thus, the objective of this study was to identify chromosomal regions associated with temperament traits in Nelore cattle using two statistical models. The temperamental traits assessed in this study were exit velocity (VS) and the median score temperament (EC_mediana) evaluated in 599 and 575 animals, respectively. All animals were genotyped with Illumina BeadChip BovineHD (800K), and two GWAS statistical models were applied. The first model used was genomic single step (ssGBLUP) in which the effects of SNP are derived from the genomic prediction values of the animals. The second was linear mixed model, similar to the previous model, but a series of models, one for each SNP markers (single regression), were performed. In both models were

included the effects of contemporary group as fixed, the animal's age as a covariate effect and polygenic animal and permanent environment as random effect. Variance components were estimated by restricted maximum likelihood. The heritability coefficients showed low magnitude estimative of 0.02 and 0.05 for VS and EC_mediana, respectively. Results of this study shed some light regarding the genetic architecture underlying this trait in beef cattle. Different chromosomal regions were associated with the traits studied in this population, according with the models used, and moreover to contribute in the understanding of the genetic architecture of these traits.

Keywords: Beef cattle; GWAS; Heritability; SNP; Temperament

3.1 Introdução

Temperamento em bovinos é definido como a alteração do comportamento, em resposta ao medo e estresse, ocasionado no momento do manejo dos animais pelo ser humano nas fazendas ou pela mudança de ambiente (FORDYCE; DODT; WYTHES, 1988; BURROW; DILLON, 1997). É uma característica que envolve vários aspectos de comportamento, sendo influenciada por vários fatores, tais como, efeitos genéticos e ambientais (MOURÃO; BERGMANN; FERREIRA, 1998). O temperamento de um animal pode variar de dócil a agressivo, sendo que animais dóceis são preferíveis para confinamentos em que o manejo se torna mais frequente (HOPE et al., 2010). Vários estudos têm relatado a importância do temperamento em bovinos, e seu impacto em características produtivas e econômicas (PHOCAS et al., 2006; PIMENTA et al., 2009; CAFE et al., 2011; FRANCISCO et al., 2012).

Metodologias para avaliar o temperamento em bovinos estão bem descritas e consolidadas na literatura, porém são difíceis de serem mensuradas nos animais. Velocidade de saída (VS) é uma característica objetiva que avalia o tempo que o animal leva para percorrer uma distância fixa, desde a abertura da porta da balança (BURROW; SEIFERT; CORBET, 1988). Animais com menor VS são indicativos de serem mais calmos. Por outro lado, escore de comportamento (EC) é uma característica subjetiva, sendo composta por diferentes tipos de reações dos animais quando manejados dentro da balança. EC é composta de outras características, tais como, deslocamento e tensão na balança. As estimativas de herdabilidade encontradas na literatura para VS variaram entre 0,11 - 0,35 (HOPE et al., 2010; SANT'ANNA et al., 2012; SANT'ANNA et al., 2013). Para EC valores de herdabilidade variaram entre 0,17 - 0,34 (PIOVEZAN, 1998) indicando algum potencial

destas características serem incorporadas nos programas de melhoramento genético em bovinos como critérios de seleção (FIGUEIREDO et al., 2005).

Estudos têm relatado que animais com VS e EC elevados podem influenciar o desempenho produtivo e reprodutivo dos rebanhos, devido à sua correlação com outras características economicamente importantes. Animais mais agitados têm uma diminuição do ganho de peso (MULLER; VON KEYSERLINGK, 2006; SANT'ANNA, et al., 2012) e, consequentemente, redução da eficiência alimentar (PETHERICK et al., 2002). Por outro lado, animais com temperamento mais dócil tendem a apresentar maiores taxas de ganho de peso (BORBA et al., 1997). Cooke et al. (2011) observaram que vacas mais agitadas apresentaram uma diminuição na taxa de prenhez. Hall et al. (2011) encontraram uma correlação negativa ($r = -0,187$, $p = 0,03$,) entre VS e força de cisalhamento em animais cruzados de Red Angus, Simmental, Charolais, Hereford, Gelbvieh.

Alguns *quantitative trait loci* (QTLs) associados com características de temperamento foram encontrados em bovinos de corte e bovinos de leite, utilizando marcadores microsatélites. Schmutz et al. (2001) relataram QTLs localizados em seis cromossomos BTA1, BTA5, BTA9, BTA11, BTA14 e BTA15 associados com temperamento em bovinos de corte canadense, genotipados com 162 microsatélites, com intervalos de aproximadamente 20 cM. Entretanto, esses resultados são reportados utilizando mapas de baixa resolução (HÖGLUND et al., 2012).

Devido ao desenvolvimento das tecnologias de genotipagem para centenas de milhares de marcadores, em especial, os marcadores do tipo polimorfismos de base única (SNP) estudos de associação genômica (*genome-wide association studies*- GWAS) tornaram-se uma ferramenta para identificar regiões cromossômicas associadas com características complexas em diferentes espécies (PENG et al., 2010; BARSH et al., 2012). Associações significativas com características econômicas têm sido relatadas em bovinos, incluindo características de fertilidade, crescimento e qualidade da carne (ZHANG et al., 2012). Em relação às características de temperamento, quase nada se sabe sobre a arquitetura genética que controla a variação dessas características em bovinos. Desta forma, a compreensão dos mecanismos genéticos envolvidos com estas características em animais da raça Nelore pode auxiliar na identificação e mapeamento de genes que interferem na expressão do temperamento nestes animais, tendo como resultado a melhoria da produção de carne, facilidade no manejo, e outras características por meio da seleção de animais mais dóceis.

Sendo assim, o objetivo desse estudo foi identificar regiões cromossômicas relevantes associadas com características de temperamento em bovinos da raça Nelore sendo utilizados dois diferentes modelos matemáticos de associação.

3.2 Material e Métodos

Animais

Foram utilizados animais, novilhos machos, descendentes de 34 touros Nelore PO registrados, pertencentes à raça de uso comercial mais frequente no Brasil. Estes touros foram selecionados com base nos coeficientes de parentesco, com o objetivo de minimizar o grau de parentesco e representar a variabilidade genética da raça no país. As progênies foram obtidas em três estações de monta, nos anos de 2007, 2008 e 2009, por meio de inseminação artificial em tempo fixo (IATF) e criadas na Embrapa Pecuária Sudeste (CPPSE), localizada na cidade de São Carlos-SP, na Embrapa Gado de Corte (CNPGC), situada no município de Campo Grande-MS e em três propriedades particulares dos Estados de Mato Grosso do Sul e Mato Grosso.

Os novilhos foram criados e recriados a pasto nas fazendas de origem, desmamados com média de 8 meses de idade, pesados na desmama e no sobreano. Os animais foram castrados quando atingiram aproximadamente 17 meses de idade e, em seguida foram transportados para as estações experimentais, CPPSE ou CNPGC com $\pm$ 18 meses de idade, para serem confinados. Os animais produzidos nas unidades da Embrapa permaneceram nas próprias unidades e aqueles oriundos das fazendas parceiras foram divididos e alocados nas estações experimentais, em função da paternidade, para garantir sua representatividade nos diferentes locais de avaliação (Nascimento, 2011).

O período experimental de confinamento compreendido entre o início das avaliações e o abate, foi determinado em função do grau de acabamento que os animais deveriam atingir, sendo determinado em 5mm de espessura de gordura subcutânea. Em geral, teve duração de cerca de 100 dias. Nesse período foram avaliadas características de eficiência alimentar, temperamento e qualidade de carcaça. Os animais foram submetidos a um período mínimo de 28 dias de adaptação à dieta e posteriormente foram adaptados ao sistema de alimentação com cancelas eletrônicas no confinamento (Nascimento, 2011).

Avaliação do temperamento

As características de temperamento utilizadas foram escore composto (EC) e velocidade de saída da balança (VS), durante pesagens de rotina do projeto de avaliação de eficiência alimentar destes animais em confinamento (PAÇÓ, 2012). Foram utilizadas informações de 599 e 575 animais para EC e VS, respectivamente. As coletas das informações foram obtidas em quatro diferentes tempos, caracterizando em experimento com medidas repetidas no tempo, sendo que pelo menos duas vezes para cada animal. VS e EC foram coletadas a cada 28 dias, sendo que última avaliação foi realizada no dia anterior aos animais serem encaminhados para o abate, manejo que antecede o procedimento pré-abate. As metodologias de avaliação serão descritas a seguir:

1) Velocidade de saída (VS): mede a velocidade de saída com que os animais saem da balança de pesagem, e percorrem uma distância de aproximadamente 2,3 metros (BURROW et al., 1997) (Figura 1). Animais que possuem VS inferior são em geral mais calmos, e portanto, mais fáceis de manejar nas fazendas. Foi utilizado para o registro das medidas um sistema com sensores fotoelétricos acoplados a um computador que fornece o tempo (m/s) da passagem de cada animal, denominado “*flight speed*”. A VS foi avaliada nos animais após a abertura da porta da balança. A avaliação do tempo de saída foi efetuada após a pesagem dos animais e atribuição do escore composto de agitação, sem alterar o manejo rotineiro de pesagem.

Figura 1 – Avaliação da velocidade de saída. (A) Animal percorrendo uma distância fixa de 2,3 metros após a abertura da porta da balança; (B) Sistema com sensores fotoelétricos para mediar a velocidade (m/s)

2) Escore composto (EC): trata-se de uma variável categórica, em que se utiliza de uma escala numérica, sendo subjetiva. É formada a partir de outras características, como deslocamento, tensão, respiração, mugidos e posição corporal (também variáveis categóricas), avaliadas no mesmo animal quando mantidos em uma balança de pesagem, e um avaliador se aproxima do animal (Figura 2). Os escores para as características acima mencionadas foram atribuídos quatro segundos após o fechamento da porteira de entrada e de saída da balança (Tabela 1).

Figura 2- Avaliação do escore dos animais dentro da balança

O escore composto foi atribuído de acordo com metodologia adaptada de Piovezan (1998). Na avaliação de EC foram avaliados os seguintes comportamentos:

- (i) Deslocamento (DESL):** 1 = nenhum deslocamento; 2 = pouco deslocamento, parado em mais da metade do tempo de observação; 3 = deslocamentos frequentes, mas pouco vigorosos (metade do tempo de observação ou mais); 4 = movimentação constante e vigorosa e 5 = animal salta, elevando os membros superiores pelo menos 2,5 cm do solo;
- (ii) Tensão (TENS):** 1 = relaxado, quando o animal apresenta tônus muscular regular, sem movimentos bruscos de cauda e/ ou cabeça e pescoço, sem membrana esclerótica aparente no olho; 2 = alerta, quando o animal apresenta movimentos bruscos de cauda, cabeça e pescoço, membrana esclerótica do olho aparente ou não; 3 = tenso, quando o animal apresentava movimentos bruscos e contínuos de cauda, cabeça e pescoço, membrana esclerótica aparente, força a saída, faz movimentos frequentes e vigorosos e 4 = muito tenso, animal paralisado e apresentando tremor muscular;

(iii) **Respiração (RESP):** 1 = respiração normal e ritmada, 2 = bufando ou soprando, de forma não ritmada;

(iv) **Mugidos (MUG):** 1 = não mugiu e 2 = mugiu, independente da frequência ou intensidade; V) Coices (COI), considerando: 1 = ausência e 2 = presença;

(v) **Posição Corporal (PC):** 1 = em pé, 2 = ajoelhado e 3 = deitado.

Com base nos registros das categorias das variáveis de comportamento acima mencionado, foi definido a variável escore composto (EC) (Tabela 1), classificado em cinco classes, como segue (AGUILAR, 2007):

Tabela 1 - Descrição do escore composto em animais da raça Nelore

Escore Composto	Descrição
1 Calmo	DESL= 1, PC= 1, TENS= 1 ou 2, RESP= 1 ou 2, MUG= 0 ou 1 e COI= 0);
2 Ativo	DESL= 1, 2 ou 3, PC= 1, TENS= 1 ou 2, RESP= 1 ou 2, MUG= 0 ou 1 e COI= 0 ou 1
3 Inquieto	DESL= 2 ou 3, PC= 1, 2 ou 3, TENS= 2 ou 3, RESP= 1 ou 2, MUG= 0 ou 1 e COI= 0 ou 1
4 Perturbado	DESL= 3 ou 4, PC= 1, 2 ou 3, TENS= 2 ou 3, RESP= 1 ou 2, MUG= 0 ou 1 e COI= 0 ou 1
5 Muito perturbado	DESL= 4 ou 5 e TENS=3, independente dos resultados nos demais escores, intratável/perigoso

A variável EC refere-se a uma variável categórica, não apresenta distribuição normal. Devido à natureza desta característica, e tendo em vista que foi coletada ao longo do tempo, caracterizando ser uma variável categórica ordinal e longitudinal, novas variáveis foram criadas a partir dos dados originais de EC para estimar componentes de variância e herdabilidade. Com base nas estimativas de componentes de variância, e herdabilidade das novas características foi selecionado a variável mais adequada para as análises posteriores de GWAS. Sendo assim, as novas variáveis foram:

(i) **EC_Média** = trata-se da média do escore composto para cada animal, coletada em diferentes ocasiões;

(ii) **EC_Mediana** = mediana do escore composto para cada animal, coletada em diferentes ocasiões;

(iii) **PC_EC** = componente principal referente ao escore composto (EC), coletado em diferentes ocasiões;

Para a análise de componentes principais foi utilizada a matriz de correlação para o conjunto de variáveis originais do escore composto coletado em nas quatro ocasiões: EC₁, EC₂, EC₃ e EC₄. As variáveis foram transformadas em um novo conjunto de variáveis Y₁ (CP₁), Y₂ (CP₂), Y₃ (CP₃), e Y₄ (CP₄), com a função de reduzir linearmente a dimensão de um conjunto de dados, de modo que este novo conjunto de p variáveis não fossem correlacionadas, e que retenham o máximo de informações em termos de variação total (JOHNSON; WICHERN, 1998; MANLY, 2008). O primeiro componente principal selecionado para criar a nova variável PC_EC, pois apresentou maior variância ($\max \lambda_i$), e explicou o máximo de variabilidade dos dados. Para as análises de componentes principais utilizou-se os pacotes *imputation* e *princomp*, implementadas no ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2013).

As coletas das informações para as características de temperamento foram realizadas por um único observador em São Carlos. Em Campo Grande, na safra 1 as observações foram realizadas por um observador, e nas safras 2 e 3 as coletas foram realizadas por dois observadores (PAÇO, 2012).

Estimação dos componentes de variância

Os grupos de contemporâneos (GC) foram formados pela concatenação dos efeitos de safra (ano de nascimento), local de confinamento (local), safra x local, origem do animal e baia. Estes efeitos foram testados por meio do ajuste de modelo misto, utilizando o procedimento PROC MIXED do pacote estatístico SAS[®]. Para as análises de estimação dos componentes de variância foram considerados os GC, em que os efeitos que foram significativos (Tabela 2). Os GC foram iguais para todas as características de temperamento.

Tabela 2 - Resumo das análises de variância das características de temperamento avaliada nos animais da raça Nelore

Fonte de Variação	Valor de probabilidade					
	GL	VS	EC_Média	EC_Mediana	EC	CP_EC
Safra	2	0,0001	0,0261	0,0014	0,0262	0,0001
Local	1	0,0001	0,0001	0,0001	0,0001	0,0003
Origem	4	0,0001	0,0048	0,0170	0,0001	0,0001
Safra*Local	2	0,0001	0,0001	0,0001	0,0001	0,0001
Baia	1	0,2397	0,0770	0,1227	0,0159	0,1404
Idade	1	0,0044	0,0532	0,0589	0,0101	0,0010

GL: graus de liberdade; VS: velocidade de saída; EC_Média: média do escore composto; EC_Mediana: mediana do escore composto; EC: escore composto e CP_EC: componente principal do escore composto

Os componentes de variância para as variáveis VS, EC_Média, EC_Mediana PC_EC e EC foram obtidos pelo método da máxima verossimilhança restrita, utilizando-se um modelo linear misto, modelo animal, que incluiu os efeitos poligênico do animal e ambiente permanente como aleatórios, idade do animal como covariável linear e grupo de contemporâneo como efeito fixo. Para estimação foi utilizado o pacote *pedigreemm* (VAZQUEZ et al., 2010), implementado no ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2013).

Genotipagem e Controle de Qualidade

A genotipagem foi realizada no Laboratório de Genômica Funcional do Departamento de Agricultura dos Estados Unidos (BFG/USDA/EUA) e no Laboratório Multiusuários de Genômica Funcional Aplicada à Agropecuária e à Agroenergia da ESALQ – USP, financiada e realizada pela equipe do projeto Rede Bife de Qualidade, Macroprograma 1 da Embrapa. Foi utilizado o chip “Illumina BovineHD BeadChip”, que possui aproximadamente 777 mil SNP em sua plataforma, com espaçamento médio de 3,4 kb. Em seguida, os dados genômicos foram filtrados a fim de serem eliminados os resultados de baixa qualidade e aumentar a acurácia das análises (OLIPHANT et al., 2002). Os seguintes critérios foram utilizados: (i) indivíduos com “call rate” >95% (porcentagem do genoma genotipado corretamente); (ii) SNP com “call rate” > 95% (porcentagem de genotipagem de cada SNP em toda a população); (iii) *minor allelic frequency* (MAF) foi estabelecida em 1%. Foram excluídos ainda os SNP localizados no cromossomo X e no Y e aqueles sem informação de posição no

genoma. A análise de controle de qualidade foi realizada por meio do software Plink (PURCELL et al., 2007). Os resultados do controle de qualidade estão apresentados na Tabela 3.

Tabela 3 - Controle de qualidade utilizado no painel BovineHD BeadChip (800K)

Critério de controle de Qualidade	Removidos
<i>Minor allelic frequency</i> (1%)	160.872 SNP
<i>Call rate</i> /SNP (95%)	30.882 SNP
<i>Call rate</i> /animal (95%)	10 animais
Total SNP removidos	191.754
Total de SNP utilizados	560.142

Estudo de Associação Genômica (GWAS)

Dois modelos foram utilizados para a análise de associação genômica entre os genótipos dos SNP, contidos no chip HD e as variáveis de temperamento. No primeiro modelo foi usado um modelo misto (ssGBLUP) de acordo com (WANG et al., 2012).

$$\mathbf{Y} = \mathbf{1}\mu + \mathbf{X}\beta + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e} \quad (1)$$

em que $\mathbf{Y}$ é o vetor de observações para as variáveis de temperamento, μ é a média geral de registros fenotípicos, $\mathbf{1}$ é um vetor de 1s, $\mathbf{X}$ é uma matriz de incidência de efeitos fixos, β é o vetor dos efeitos fixos de grupo de contemporâneo. Os grupos de contemporâneo foram formados pelo ano de nascimento, local de nascimento e sistema de confinamento; $\mathbf{Z}$ é a matriz de incidência que aloca as informações de parentesco ao fenótipo dos animais; $\mathbf{a}$ é o vetor de valores genéticos a serem preditos $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$, em que σ_g^2 é a variância genética aditiva, e $\mathbf{G}$ é a matriz de relacionamento genômica; $\mathbf{pe}$ é o vetor de efeitos aleatórios de ambiente permanente de animal com $\mathbf{pe} \sim N(\mathbf{0}, \mathbf{I}\sigma_{pe}^2)$, $\mathbf{e}$ é o vetor de resíduos $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$.

A matriz de relacionamento genômica ($\mathbf{G}$) entre os indivíduos foi obtida da seguinte forma (VANRADEN, 2008):

$$\mathbf{G} = \frac{(\mathbf{M}-\mathbf{P})(\mathbf{M}-\mathbf{P})'}{2 \sum_{j=1}^m p_j (1-p_j)}$$

em que $\mathbf{M}$ é a matriz de genótipos com o número total de marcadores (colunas) e o número total de indivíduos genotipados (linhas), $\mathbf{P}$ é uma matriz de frequências alélicas. A matriz $\mathbf{G}$ foi construída utilizando o programa preGSf90 (AGUILAR et al., 2010), que é uma versão do

programa BLUPf90 (MIZTAL et al., 2002). Após o ajuste do modelo e predições dos valores genéticos dos animais, os efeitos dos marcadores SNP foram derivados de acordo com (WANG et al., 2012):

$$\hat{\mathbf{g}} = \mathbf{D}\mathbf{Z}'[\mathbf{ZDZ}']^{-1} \hat{\mathbf{a}}$$

em que $\mathbf{g}$ é o vetor dos efeitos dos marcadores SNP, $\mathbf{D}$ é uma matriz diagonal de pesos para os variância dos SNP, $\mathbf{a}$ é vetor de valores genéticos genômicos dos animais. O programa postGSf90 foi usado para a obtenção de soluções dos efeitos dos SNP (WANG et al., 2012).

No segundo modelo, as associações entre genótipos de SNP e as variáveis mencionadas acima foram avaliadas utilizando análise de marcador simples (SNP) como uma covariável linear (PEÑAGARICANO; WEIGEL; KHATIB, 2012). O seguinte modelo linear misto foi utilizado:

$$Y_{ijklmn} = \mu + gc_j + b_k age + b_l SNP_{il} + animal_m + PE_n + e_{ijklmn} \quad (2)$$

em que $\mathbf{Y}$ é o vetor de observações para as variáveis de temperamento; μ é a média geral de registros fenotípicos, gc_j é o efeito fixo do grupo contemporâneo. Os grupos de contemporâneo foram formados pelo ano de nascimento, local de nascimento e sistema de confinamento. A idade foi considerada como covariável linear, b_k é o coeficiente de regressão linear da variável idade, SNP_{il} é o efeito fixo da classe genotípica do marcador ($g = 0, 1, 2$), b_l é o coeficiente de regressão para o SNP; $animal_m$ representa o efeito poligênico do animal, PE_n representa o efeito de ambiente permanente do animal e e_{ijklmn} representa o resíduo para cada observação. Os efeitos aleatórios foram assumidos com distribuição normal multivariada,

$$\begin{pmatrix} a \\ pe \\ \varepsilon \end{pmatrix} \left| \sigma_a^2, \sigma_{pe}^2, \sigma_\varepsilon^2 \right. \sim N \left[0, \begin{pmatrix} G\sigma_a^2 & 0 & 0 \\ 0 & I\sigma_{pe}^2 & 0 \\ 0 & 0 & I\sigma_\varepsilon^2 \end{pmatrix} \right]$$

em que $\mathbf{a}$, $\mathbf{pe}$ e ε são os vetores de animais para os efeitos poligênico do animal, de ambiente permanente e residual, respectivamente, σ_a^2 , σ_{pe}^2 e σ_ε^2 são as variâncias do animal, efeito de ambiente permanente e residual, respectivamente; $\mathbf{G}$ representa a matriz de relacionamento genômico entre os animais (VANRADEN et al., 2008). A matriz genômica foi utilizada, pois as informações do pedigree tradicional não estavam disponíveis o suficiente (pequena estrutura de pedigree), para captar o relacionamento dos indivíduos dessa população, $\mathbf{pe}$ é um efeito aleatório de ambiente permanente, tal como $\mathbf{N}(0, \mathbf{I}\sigma^2)$, e $\mathbf{e}$ efeito residual $\mathbf{N}(0, \mathbf{I}\sigma^2)$.

Nesta abordagem, os valores de p das análises de associações entre os SNP e as características de temperamento foram constatado utilizando um teste de razão de verossimilhança (LRT) por meio da comparação com um modelo reduzido, sem o efeito do SNP, e um modelo completo, adicionando um parâmetro a mais, o efeito do SNP. O teste baseia-se na distribuição de Qui-quadrado (χ^2), com v graus de liberdade e probabilidade de erro de 5% de probabilidade (DOBSON, 2002). Dessa forma, o LRT é dado por:

$$\text{LRT} = 2\log L(F) - 2\log L(R) \quad (3)$$

em que $L(F)$ é a verossimilhança do modelo com mais parâmetros (modelo completo), e $L(R)$ a verossimilhança do modelo com menor número de parâmetros (modelo reduzido). As análises foram realizadas utilizando o pacote *pedigreemm* (VAZQUEZ et al., 2010) implementado no ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2013).

Após as análises de GWAS os valores de p e dos efeitos dos marcadores SNP foram apresentados em um gráfico (*Manhattan Plot*) para a visualização ao longo dos 29 cromossomos. O pacote *gap* implementado no ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2013) foi utilizado.

A análise do desequilíbrio de ligação (LD) foi realizada com o programa *Haploview* (BARRET et al., 2005) e a medida utilizada foi o r^2 , que mede a correlação entre alelos de dois marcadores, sendo que r^2 igual a 1 indica uma correlação perfeita entre os dois locos. Por outro lado, r^2 igual a zero indica ausência de LD. De acordo com Ardlie, Kruglyak e Seielstad (2002), o r^2 é corrigido para as respectivas frequências alélicas, sendo indicativo da quantidade de informação de um loco que é explicada pelo outro.

O banco de dados *CattleQTLdb* (HU et al., 2010) foi utilizado para buscar informações de QTLs nos cromossomos em que os SNP foram significativos ou apresentaram maiores efeitos para características de temperamento ou QTLs previamente relacionados com características de importância econômica em bovinos, nos dois modelos estatísticos estudados. Para tal etapa, estendeu-se um intervalo de busca de 500 kb para ambos os lados (1 Mb). Além disto, buscou-se fora deste intervalo QTLs relacionados apenas com características de temperamento.

3.3 Resultados e Discussão

Estatísticas descritivas e dispersão

As estatísticas descritivas e de dispersão para as características de temperamento estão apresentadas na Tabela 4. As médias para VS, EC_Média e EC_Mediana foram $1,09 \pm 0,01$, $1,94 \pm 0,04$ e $1,95 \pm 0,04$, respectivamente. Considerando a mesma população, porém utilizando informações de 800 animais, Paçó (2012) obteve menores valores médios (0,39 e 0,73) para VS e EC, respectivamente. Em parte, isso pode ser explicado pelo menor número de informações disponíveis neste estudo. Em Hope et al. (2010) os autores observaram valores médios para velocidade de saída de $1,51 \pm 0,06$; $1,78 \pm 0,06$; $1,52 \pm 0,06$; $1,87 \pm 0,07$; $1,89 \pm 0,07$ para animais Angus, Charolais, Hereford, Limousin, e Simental, respectivamente. A distribuição dos escores compostos avaliados em diferentes tempos é apresentada na Figura 1. Em geral, com base nas distribuições do EC nas quatro avaliações, o número de animais muito agitados (escores 4 e 5) não foi alto. Esses resultados sugerem que na população Nelore avaliada, os animais apresentavam temperamento que variava entre calmo e ativo.

Tabela 4 – Estatísticas descritivas e de dispersão para as características de temperamento velocidade de saída (VS), média do escore composto (EC_Média) e mediana do escore composto (EC_Mediana)

Característica	N	M	Mín	Max	Med	DV	Moda	EP	CV	Var
VS	599	1,09	0,11	4,39	0,99	0,57	0,66	0,01	0,53	0,33
EC_Média	575	1,94	0,50	5,00	1,75	1,01	1,50	0,04	0,52	1,02
EC_Mediana	575	1,95	0,75	5,00	1,75	1,03	1,00	0,04	0,53	1,00

N = Número de animais, M = Média, Mín = Mínimo, Max = Máximo, Med = Mediana, DV = Desvio padrão, Mod = Moda, EP = Erro padrão, CV = Coeficiente de variação e Var = Variância

Figura 1 - Gráfico de barras da variável resposta escore composto avaliado em diferentes ocasiões, sendo 1 = calmo; 2 = ativo; 3 = Inquieto; 4 = Perturbado e 5 = Muito perturbado

Componentes de Variâncias e Herdabilidades

Velocidade de Saída

As estimativas dos componentes de variância e dos coeficientes de herdabilidade para velocidade de saída estão apresentadas na Tabela 5. A variância genética aditiva foi menor do que as variâncias residuais e de ambiente permanente (σ_{pe}^2). Este resultado indica que o efeito σ_{pe}^2 ou seja, efeitos genéticos não aditivos apresentam alta influência sobre velocidade de saída. Assim como verificado neste estudo existe maior influência de fatores ambientais do que genéticos para velocidade de saída. Vale ressaltar que no presente estudo utilizou-se a matriz de relacionamento genômico (VANRADEN et al., 2008) entre os animais.

O coeficiente de herdabilidade para velocidade de saída foi 0,020 (Tabela 5). Essas estimativas são menores do que aquelas obtidas por outros autores. No trabalho de Burrow, Seifert e Corbet (1988) o coeficiente de herdabilidade estimado foi de alta magnitude (0,64) para velocidade de saída. Recentemente, Sant'Ana et al. (2012) relataram coeficiente de herdabilidade de $0,26 \pm 0,05$ para velocidade de saída, em animais da raça Nelore. No trabalho de Hope et al (2010) observaram estimativas de $0,20 \pm 0,08$; $0,25 \pm 0,10$; $0,36 \pm 0,06$; $0,11 \pm 0,07$ e $0,28 \pm 0,07$ para animais Angus, Charolais, Hereford, Limousin, e Simental, respectivamente.

Por sua vez, em Paçó et al. (2012) o coeficiente de herdabilidade foi nulo $0,00 \pm 0,02$ para velocidade de saída, estudando a mesma população de animais Nelore. Em Paçó (2012) o coeficiente de herdabilidade foi estimado a partir das informações fenotípicas de 800 animais, e utilizou-se uma matriz de parentesco constituída de 1672 animais. Segundo este autor, estas estimativas sugerem que velocidade de saída não apresenta resposta à seleção. Estes resultados mostram discrepância entre os resultados encontrados na literatura, como também relatados neste estudo.

A estimativa de repetibilidade para velocidade de saída foi 0,19. Estimativas similares de repetibilidade para velocidade de saída (0,18 e 0,14) foram reportadas por Burrow (2001), para animais Belmont Red. Silveira et al. (2008) em animais Nelore, obtiveram estimativas elevadas de repetibilidade de 0,73.

O coeficiente de herdabilidade no presente estudo para velocidade de saída foi afetado pelo tamanho amostral.

Tabela 5 – Estimativas dos componentes de variância genética aditiva (σ_a^2), residual (σ_e^2), fenotípica (σ_p^2), herdabilidades (h^2) e repetibilidade (r) para as características de temperamento

Característica	N	σ_a^2	σ_e^2	σ_p^2	σ_{pe}^2	h^2	r
VS	599	0,01033	0,40159	0,51308	0,10115	0,020	0,197
EC_Média	575	0,02631	0,32540	0,35172	-	0,040	-
EC_Mediana	575	0,04134	0,34533	0,38667	-	0,053	-
CP_EC	575	0,08865	1,62335	1,71200	-	0,031	-
EC	575	0,02543	0,65595	0,68138	0,16229	0,019	0,192

VS: velocidade de saída; EC_Média: média do escore composto; EC_Mediana: mediana do escore composto; CP_EC: componente principal para as medidas de escore composto; EC: escore composto

Escore Composto e demais variáveis

Foram estimados os componentes de variância e herdabilidade para as características escore composto (EC) e para as novas características formadas a partir de EC.

Os coeficientes de herdabilidade obtidos para escore composto, EC, EC_Média, EC_Mediana e CP_EC foram iguais a 0,019, 0,040, 0,053 e 0,031, respectivamente (Tabela 5). Os coeficientes de herdabilidade apresentaram baixa magnitude para todas as variáveis estudadas, quando comparados com estimativas de outros trabalhos na literatura.

Diversos resultados encontrados na literatura indicam que temperamento é uma característica que responde à seleção. Em Mourão, Bergmann e Ferreira (2008), os autores

avaliaram o temperamento em vacas Nelore, Tabapuã e Indubrasil e entre suas filhas F1 Holandês x Zebu, por meio do escore teste de contenção no tronco, com escores subjetivos (1, temperamento muito dócil a 5, muito agressivo) e obtiveram estimativas de herdabilidade que variaram de 0,06 à 0,27. Schmutz et al. (2001) registraram valores de herdabilidade para habituação (diferença entre o escore inicial na desmama, coletada com idade de 8 meses, e esta mesma medida, coletada com 12 meses) de 0,36.

Recentemente, Sant'Ana et al. (2013) relataram coeficiente de herdabilidade de 0,15 para escore de temperamento em animais da raça Nelore, utilizando 23420 informações de machos e fêmeas. Por sua vez, Barrozo et al. (2012) observaram coeficiente de $0,53 \pm 0,04$. Neste trabalho foram utilizados 37962 informações para escore de temperamento avaliados em animais da raça Nelore.

Na literatura citada neste trabalho e analisando os resultados observa-se discrepância entre as estimativas de herdabilidade. Porém, acredita-se que as estimativas de herdabilidades encontradas no presente estudo, tanto para velocidade de saída, escore compostos e demais variáveis foram afetadas pelo tamanho amostral. Além disto, pode-se afirmar que houve maior influência dos fatores ambientais do que genéticos para todas as características.

Com base nas estimativas dos coeficientes de herdabilidade destas variáveis, a característica EC_mediana ($h^2 = 0,05$) foi selecionado para as análises de GWAS.

Associação Genômica

O estudo de associação genômica (GWAS) têm o objetivo de identificar regiões cromossômicas que expliquem as diferenças fenotípicas entre os indivíduos na população em estudo. Sendo assim, na análise de GWAS foram utilizados dois modelos estatísticos. No primeiro modelo, ssGBLUP, foram estimados os efeitos dos marcadores SNP, derivados dos valores genéticos genômicos para cada animal (WANG et al., 2012). Porém, uma limitação desta metodologia seria a não utilização de um limiar para selecionar os SNP significativos. Sendo assim, adotou-se como critério de seleção dos SNP, os top 10 com maiores efeitos. No segundo modelo, utilizou-se análises de marcadores únicos, SNP por SNP. Para a seleção dos valores de p dos SNP mais significativos empregou-se cromossomo-wise, com limiar de 10%, pois os valores de p calculados através do FDR e Bonferroni não foram significativos. Para as análises de GWAS foram utilizados um total de 560.142 SNP abrangendo os 29 cromossomos. O número de SNP por cromossomo variou de 8.986 a 36.676 no BTA25 e BTA1, respectivamente. A discussão dos resultados deste estudo será decorrida para cada

característica individualmente, levando em consideração os dois modelos estatísticos separadamente.

Velocidade de Saída (VS)

No modelo 1, ssGBLUP, dos top 10 SNP selecionados com maiores valores dos efeitos destes, cinco estavam localizados no cromossomo 6. Neste modelo, foi assumido que todos os marcadores contribuem para a mesma variabilidade da característica. Na Figura 2 é apresentado o gráfico *Manhattan plot* das análises de associação genômica considerando os modelos 1 e 2 para VS. Na Figura 2A, o eixo Y representa o efeito dos SNP e no eixo X os SNP distribuídos entre os 29 cromossomos a partir do modelo 1. A Figura 2B representa no eixo do Y $-\log_{10}$ (valores de p) e no eixo do X os SNP distribuídos entre os 29 cromossomos conforme o modelo 2.

De acordo com o observado na Figura 2, os SNP selecionados nas duas metodologias de análise estatística não apresentaram concordância das posições dos SNP nos cromossomos ao longo do genoma, entre os dois modelos utilizados. Esses resultados são consistentes ao apresentado por Ober et al. (2012) os quais, observaram concordância imperfeita das posições dos SNP ao compararem o modelo GBLUP e BayesB a partir de dados de sequência completa de *Drosophila melanogaster*. Estes resultados são consequência do objetivo que cada metodologia de análise apresenta (OBER et al., 2012). No modelo 1 (ssGBLUP), o fenótipo é resultado da predição usando todas as informações disponíveis dos SNP, simultaneamente. Por outro lado, no modelo 2 (regressão de marcadores únicos), SNP por SNP, o efeito para cada um dos marcadores é estimado individualmente e a informação de LD entre os SNP não é levada em consideração. Dessa forma, neste modelo muitos dos SNP significativos em uma região genômica poderão estar em LD com o mesmo QTL, dificultando a identificação do número de QTLs associado com a característica. Por outro lado, o modelo 2 é mais conservativo em estimar o número real de QTLs, tendo como resultado a subestimação do número de SNP significativos (PRYCE et al., 2010).

Os efeitos dos marcadores estimados apresentaram uma magnitude pequena (Figura 2). Tais resultados segundo Ober et al. (2012) são consequência do número de efeitos a serem estimados, pois é muito maior do que o número de observações. Dessa forma, estes efeitos são penalizados, diminuindo a magnitude dos efeitos para zero. Outra questão importante é a densidade dos painéis de SNP, pois quanto maior for a densidade, mais SNP estarão em LD com uma possível mutação causal e, portanto, o verdadeiro efeito de substituição alélica do

marcador poderá está dividido em partes, e atribuído a vários outros SNP em um bloco de haplótipos, por que qualquer SNP em LD com o polimorfismo causal pode ter o efeito estimado. Contudo, isto pode mascarar a magnitude do efeito do SNP.

Figura 2 - Manhattan plot dos valores de p e dos efeitos dos SNP para o estudo de associação genômica para velocidade de saída (VS) em animais da raça Nelore em que: (A) o modelo ssGBLUP, e (B) o modelo SNP por SNP

No modelo 1, dos top 10 SNP selecionados com associação para VS, os mesmos encontram-se localizados nos cromossomos 6, 7, 9 e 14 (Tabela 6).

No cromossomo 6 foram encontrados cinco SNP que apresentam maiores efeitos, indicando que esta região pode conter genes que controlam VS (Tabela 6), entre os quais dois (BovineHD0600034026 e BovineHD0600034028) encontram-se em LD de alta magnitude (Figura 3). Este resultado é interessante, pois encontram-se em alto LD entre si ($r^2=0,79$). A

presença de LD entre dois SNP aumenta a chance de estudos mais aprofundados desta região. No banco de dados QTLdb para bovinos (<http://www.animalgenome.org/cgi-bin/gbrowse/cattle/#search>), foi encontrado apenas um estudo envolvendo QTL para uma característica de temperamento, similar com VS, separação-andar/correr, identificados em bovinos da raça Holândes x Charolês, genotipados para 162 marcadores microssatélites (GUTIÉRREZ-GIL et al., 2008) (Tabela 6). Neste cromossomo, considerando uma região de 1 Mb destes marcadores em LD, observou-se QTLs relacionados com peso de carcaça, em animais da raça Angus (McCLURE et al., 2010).

Figura 3 - Desequilíbrio de ligação (r^2) entre os SNP significativos no cromossomo 6 para a característica velocidade de saída. O quadrado preto representa o desequilíbrio de ligação alto entre os SNP ($r^2=0,79$)

No cromossomo 7 foram encontrados dois SNP associados com VS (BovineHD0700020072 e BovineHD0700020070). Na literatura não foram encontrados estudos para mapeamento de QTLs e GWAS com características de temperamento em animais da raça Nelore. Além disto, este é o primeiro estudo de GWAS para VS utilizando animais da raça Nelore. Gutiérrez-gil et al. (2008) avaliaram animais da raça Holândes x Charolês para características de temperamento e encontraram duas regiões para QTLs no cromossomo 7. Neste cromossomo estão descritos QTLs para área de olho de lombo, peso de carcaça e escore de marmoreio em animais da raça Angus (McCLURE et al., 2010).

No cromossomo 9, um SNP (BovineHD0900015602) foi associado com VS. Considerando uma janela de 1 Mb, não existe região de QTL descrita para características de

temperamento. Entretanto, nesta região podemos encontrar vários QTLs mapeados para escore de marmoreio (CASAS et al., 2003), e ganho de peso diário pós desmama (ALEXANDER et al., 2007). Estes resultados ressaltam a importância de um estudo futuro desta região em que foram encontrados QTLs sugestivos, associados com outras características de importância econômica. No trabalho de Silveira; Fischer; Soares (2006), os autores encontraram correlação negativa de VS com ganho médio de peso em animais da raça Angus e animais cruzados Angus x Nelore. Animais do grupo genético Angus mais agitados apresentaram menor ganho de peso. Em Hope et al. (2010), os autores observaram correlação negativa de VS e EC com ganho de peso diário (GPD) em bezerros pertencentes às raças Angus, Charolais, Hereford, Limousin e Simental. Animais com maiores VC e EC_mediana apresentaram menores GPD. O efeito genético de grupos raciais é um fator importante na determinação do temperamento (SILVEIRA; FISCHER; SOARES, 2006). Porém, estudos adicionais são necessários para validar estas associações encontradas neste estudo, antes que sejam introduzidas nos programas de melhoramento genético. Além disto, devem ser tratadas com cautela, pois estas associações podem se tratar de falso-positivos.

No cromossomo 14 foram observados dois SNP (BovineHD1400015257 e BovineHD1400015258) associados com VS localizado em 54333233 e 54334694 pb, respectivamente. Estes marcadores estão localizados em uma região com QTLs descritos para características de temperamento (Tabela 6). Além disto, várias regiões de QTLs para características de qualidade de carne foram relatadas como: rendimento de carcaça (TAKASUGA et al., 2007; MORRIS, et al., 2009; LEE et al., 2010), escore de marmoreio (CASAS et al., 2003; BARENDSE et al., 2009), gordura intramuscular (LEE et al., 2010), espessura de gordura na 12^a costela (NALAILA et al., 2012) e área de olho de lombo (TAKASUGA et al., 2007). Estes resultados de QTLs mapeados para outras características podem auxiliar no entendimento do relacionamento do temperamento dos animais com as características de crescimento e qualidade de carcaça. Estudos que tratam de GWAS com características de qualidade de carne para esta população de animais da raça Nelore são recentes (TIZIOTO et al., 2013), sendo que tal trabalho não relata o impacto ou influência do temperamento na qualidade de carne. Diante disso, faz-se necessário em estudos futuros, realizar uma junção dos dois estudos visando o melhor entendimento dos mecanismos biológicos e da arquitetura genética destas características. No trabalho de Cesar et al. (2013, não publicado) os autores encontraram três SNP associados com características de qualidade de carne para esta população de animais Nelore em estudo. Os SNP descritos neste estudo foram associados com a quantidade do ácido graxo láurico C12:0, ácido palmitoleico (C16:1

cis9) e ácido oleico (C18:1 cis9) medidos no músculo *Longissimus dorsi*, com as seguintes localizações 78993288, 78993288 e 73392111 pb, respectivamente.

No modelo 2, SNP por SNP, considerando ainda a característica VS é inicialmente apresentado o gráfico Quantil-Quantil (Q-Q plot) (Figura 4), usado para avaliar se existe estrutura de população na amostra em estudo. Dessa forma, os valores de p observados para cada SNP no modelo 2 foram plotados em ordem crescente, $-\log_{10}(p)$, contra os valores de p esperados, amostrados de uma distribuição Qui-quadrada (PEARSON; MANOLIO, 2008). Observando o Q-Q plot, podemos concluir que não existe subestrutura da população para os animais mensurados para VS, pois observando o gráfico para VS é evidente que duas distribuições apresentam alta concordância.

Figura 4 - Gráfico do Quantil-Quantil dos valores de $-\log_{10}(p)$ esperado contra observado para velocidade de saída

No modelo 2, cinco SNP foram significativos para VS, sendo estes SNP localizados nos cromossomos 3, 5, 13, 15 e 22 (Tabela 7). Para os cromossomos 3, 13 e 22, utilizando o banco de dados QTLdb para bovinos (<http://www.animalgenome.org/cgi-bin/gbrowse/bovine/>) não foi encontrado nenhum estudo de QTLs relacionado com características de temperamento (Tabela 7).

Para VS, um SNP significativo foi localizado no cromossomo 5. Hiendleder et al. (2003) estudaram uma população de bovinos de leite proveniente da Alemanha e identificaram QTLs no cromossomo 5 para a característica escore de comportamento (nervoso/agressivo ou dócil no momento da ordenha) como também tempo de ordenha. Schmutz et al. (2001) também relataram QTLs em bovinos de corte de origem canadense para

o cromossomo 5 para as características escore de agitação na balança e habituação (diferença entre o escore inicial, coletada no desmama, com o escore final). No cromossomo 15 foi relatada uma região de QTL para comportamento (Tabela 7). No modelo 2, observa-se um número menor de QTLs detectados para VS em comparação com o modelo 1. Estes resultados não podem ser interpretados como ausência de associação entre os SNP e VS, mas consequência de algumas limitações do estudo. Uma limitação seria justificada pelo modelo de análise não considerar a informação de LD entre os SNP como explicado anteriormente. Outra limitação, o tamanho amostral (n=599 animais) pode não ter sido suficiente para detectar associações significativas.

Tabela 6 - Descrição dos QTLs descritos nos cromossomos para diferentes características reportadas na literatura e possivelmente associadas com velocidade de saída usando o modelo ssGBLUP

SNP	BTA	Posição	Posição do QTL (bp)	Característica	População	Referência
BovineHD0600011000	6	40245381	11930766 - 26351244	Social separation-- Walking/running	Holândes x Charolês	Gutierrez-Gil et al., (2008)
BovineHD4100004695	6	40238284	11930766 - 26351244	Social separation-- Walking/running	Holândes x Charolês	Gutierrez-Gil et al., (2008)
BovineHD0600011001	6	40248407	11930766 - 26351244	Social separation-- Walking/running	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0600034028	6	118948929	11930766 - 26351244	Social separation-- Walking/running	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0600034026	6	118941570	11930766 - 26351244	Social separation-- Walking/running	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0700020072	7	68565231	18835910 -73909970 103765346 -103765346	Social separation-- Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0700020070	7	68563348	18835910 -73909970 103765346 -103765346	Social separation-- Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0900015602	9	57059742	46378796 – 46646401 24968872 - 47319594	Behavior QTL Social separation-- Vocalization	Holândes x Charolês bovinos de corte canadense	Schmutz et al. (2001) Gutierrez-Gil et al. (2008)
BovineHD1400015257	14	54333233	35202277 - 84648390	Behavior QTL	bovinos de corte canadense	Schmutz et al. (2001)
BovineHD1400015258	14	54334694	35202277 - 84648390	Behavior QTL	bovinos de corte canadense	Schmutz et al. (2001)

Tabela 7 - Descrição dos QTLs descritos nos cromossomos para diferentes características reportadas na literatura e possivelmente associadas com velocidade de saída usando o modelo SNP by SNP

SNP	BTA	Posição	Posição do QTL (bp)	Característica	População	Referência
Velocidade de Saída (VS)						
BTB-01168089	3	78998337	-	-	-	-
BovineHD0500016140	5	56977424	20783605 - 22374736	Behavior QTL	bovinos de corte canadense	Schmutz et al. (2001)
ARS-BFGL-NGS-41546	13	78614384	-	-	-	-
BovineHD1500012313	15	44221808	24165122 - 24165122	Behavior QTL	bovinos de corte canadense	Schmutz et al. (2001)
BovineHD2200007522	22	25748167	-	-	-	-

Mediana de escore composto (EC_mediana)

No modelo 1, os top 10 SNP com maiores efeitos para EC_mediana foram localizados nos cromossomos 2, 4, 9, 12, 18, 26 e 29 (Tabela 8).

No cromossomo 4 três SNP dos top 10 foram associados com EC_mediana. No estudo de Spelman et al. (1999) os autores identificaram QTLs neste cromossomo e relataram sua importância no temperamento em bovinos de leite, originados da Nova Zelândia. Além disto, importante resultado também foi relatado por Chan (2012) em um estudo de GWAS com um total 3140 touros das raças Holandeses, Jersey e cruzados Holandês x Jersey, genotipados com o chip de 800K. Estes autores observaram similar resultado ao encontrado neste estudo. O cromossomo 4 apresentou regiões genômicas significativas para as três raças. Tais resultados confirmam que este cromossomo pode apresentar marcadores em LD relacionadas com temperamento em animais da raça Nelore. Neste cromossomo foram identificadas regiões de QTLs para características de escore de marmoreio (YOKOUCHI et al., 2009) e altura (McCLURE et al., 2010). Estes resultados reforçam a importância para que estes SNP sejam investigados, pois poderiam ser utilizados na seleção de animais mais dóceis.

No cromossomo 9 apenas um SNP apresentou maior efeito. Schmutz et al. (2001) encontraram duas regiões de QTLs para temperamento em estudos com animais bovinos de corte de origem canadense. Gutiérrez-Gil et al. (2008) também identificaram QTLs em animais Holândes x Charolês (Tabela 8). Nenhuma região de QTL para característica relacionada com qualidade de carne e crescimento foi descrita para este cromossomo. É interessante notar que para o cromossomo 9 as características VS e EC_mediana apresentam SNP significativos por meio dos dois modelos utilizados (Tabelas 6 e 8). Por sua vez, ainda no trabalho de Hope et al. (2010) foram relatadas correlações genéticas altamente positivas entre VS e EC de $0,57 \pm 0,17$, $0,63 \pm 0,12$, $0,69 \pm 0,08$, $0,98 \pm 0,08$ nos animais Angus, Charolais, Hereford, Limousin e Simmental, respectivamente. Esses resultados obtidos sugerem que além de fatores ambientes, as características de temperamento também são dependentes de fatores genéticos e, portanto, podem ser melhoradas por meio da seleção. A correlação genética positiva é um indicativo que a seleção pode ser realizada dando enfoque somente a uma delas, visto que o melhoramento ocorrerá em ambas (FALCONER; MACKAY, 1996). De posse dessas informações, estudos futuros são necessários para entender às respostas correlacionadas de VS e EC_mediana com outras características de qualidade de carne e crescimento em animais da raça Nelore com enfoque no entendimento dos resultados provenientes das análises de GWAS. Sabe-se que as correlações genéticas são

herdáveis, sendo resultado principalmente do efeito de pleiotropia. Neste caso, um QTL pode afetar mais de uma característica, ou o desequilíbrio de ligação entre dois ou mais SNP afetam uma mesma característica (FALCONER; MACKAY, 1996).

No cromossomo 18, foram relatadas duas regiões de QTLs com a característica de temperamento (Tabela 8), envolvendo animais da raça Holândes x Charolês, genotipados para 162 marcadores microssatélites (GUTIÉRREZ-GIL et al., 2008). Estudos relataram regiões de QTLs para área de olho de lombo e peso corporal na maturidade para este cromossomo (McCLURE et al., 2010).

No cromossomo 26, foi observado um SNP associado com EC_mediana. No banco de dados QTLdb para bovinos (<http://www.animalgenome.org/cgi-bin/gbrowse/cattle/#search>), (Tabela 8) foi encontrado um estudo com QTL para uma característica denominada Social/Separação/Vocalização. Além disto, inúmeros QTLs foram identificados para características de qualidade de carne: teor de ácido graxo mirístico, ácidos graxos saturados (MORRIS et al., 2010), percentagem de gordura e grau de acabamento (CASAS et al., 2003) e características de crescimento, tais como peso ao nascer, peso à desmama, peso na maturidade e altura neste cromossomo (McCLURE et al., 2010).

No cromossomo 29 também foi observado um SNP associado com EC_mediana (Tabela 8). Neste cromossomo há um QTL identificado para velocidade de alimentação e QTLs para as características de rendimento de carcaça (KIM et al., 2003; McCLURE et al., 2010), espessura de gordura na 12^a costela, escore de marmoreio (McCLURE et al., 2010).

Para os cromossomos 2 e 12 utilizando o banco de dados QTLdb (<http://www.animalgenome.org/cgi-bin/gbrowse/bovine/>) não foi encontrado nenhum estudo envolvendo as características de temperamento considerando uma janela de 1 Mb. Entretanto, Taxis (2011) relatou SNP associados com as características VS e habituação (diferença entre duas avaliações para VS) em progênie descendentes de touros das raças Simental e Angus, genotipados com o chip 50K no cromossomo 12. Neste estudo apenas 2 SNP (0,006%) dos 42.51 foram associados com VS e habituação. O SNP associado com EV localizado no cromossomo 12 apresentou localização de 30.967.371 pb. Para habituação, o SNP foi localizado no cromossomo 12, a 59.185.443 pb. Tais resultados segundo o autor, foram influenciados pelo tamanho amostral (n = 599 animais para VS e n = 587 animais para habituação).

Figura 5 - Manhattan plot dos valores de p e dos efeitos dos SNP para o estudo de associação genômica para mediana do escore composto (EC_mediana) em animais da raça Nelore em que: (A) o modelo ssGBLUP, e (B) o modelo SNP por SNP

Para o modelo 2, observando-se na Figura 6 o gráfico Quantil-Quantil (Q-Q plot) para EC_mediana, podemos concluir que não existe subestrutura entre as amostras da população, uma vez que os valores observados encontram-se distribuídos de modo uniforme, próximos a diagonal. Também é possível observar ausência de associação da maioria dos SNP com a característica estudada. Os valores observados acompanham uma linha abaixo dos esperados, evidenciando o baixo poder de associação devido ao pequeno número de indivíduos na população genotipados para uma característica de baixa herdabilidade (n=575 animais).

Figura 6 - Gráfico Quantil-Quantil dos valores de $-\log_{10}(p)$ esperado *versus* observado para as características de EC_mediana

No modelo 2, SNP por SNP, três SNP foram significativos com EC_mediana, sendo um SNP localizado no cromossomo 7 e dois SNP localizados nos cromossomos 17 (Tabela 9). A Figura 6 apresenta o gráfico *Manhattan plot* da análise de associação global do genoma, representada pelo $-\log_{10}$ (valores de p) no eixo do Y e o SNP no eixo do X, por cromossomo, considerando a característica EC_mediana.

No cromossomo 7 foi encontrado um SNP associado com EC_mediana e duas regiões de QTLs (Tabela 9). Um resultado importante a ser discutido considerando o cromossomo 7 é a localização de dois SNP com maiores efeitos no modelo 1 para a característica VS, resultado já discutido. Os gráficos *Manhattan plot* para as duas características são apresentados nas Figuras 2 e 5. Estes resultados são muito importantes, tendo em vista que estas características apresentam correlação genética positiva (HOPPE et al., 2010). Os dois modelos conseguiram detectar SNP para o mesmo cromossomo, embora com posições diferentes. Para VS os SNP apresentaram localização de 68565231 e 68563348 pb, enquanto

para EC_mediana a localização é 103.981.034 pb. Sendo assim, estes resultados sugerem que os SNP podem afetar em conjunto a mesma característica, pois correlação genética sugere mecanismos biológicos semelhantes às variáveis (CURLEY et al., 2004).

Outro resultado interessante ainda no cromossomo 7 é a identificação de vários QTLs com características de qualidade de carne e crescimento: cor a*, cor b*, cor L* e pH do músculo (REARDON et al., 2010), escore de maciez (ALLAIS et al., 2011), peso de carcaça (McCLURE et al., 2010), força de cisalhamento (ALLAIS et al., 2011), gordura subcutânea (COLLIS et al., 2012), compressão do músculo (DAVIS et al., 2007), ácido graxo linoleico conjugado, ácido graxo monoinsaturados (ALEXANDER et al., 2007) e características de crescimento em diferentes idades (McCLURE et al., 2010; PINTOS; CORVA, 2011). Uma possível explicação seria o efeito de pleiotropia, em que SNP estão envolvidos na variação destas características, sendo importante serem discutidas e analisadas em conjunto. Em Hall et al. (2011), a força de cisalhamento (FC) foi correlacionada negativamente ($r = -0,187$, $p = 0,03$) com VS (quando coletada no início do confinamento) avaliado em novilhos cruzados Angus, Red Angus, Simmental, Charolais, Hereford, Gelbviegh, South Devon, Chianina, Maine Anjou, e Shorthorn. Animais que saíram mais devagar da balança apresentaram menor força de cisalhamento do que animais com maiores valores para VS. Silveira, Fischer e Soares (2006) reportaram que a cor L* esteve negativamente correlacionada com EC e pH24 ($r = -0,54$ e $-0,39$, respectivamente) e a cor b*, com pH às 3 horas e pH às 7 horas ($r = -0,39$ e $-0,45$, respectivamente). Ainda neste trabalho, animais com maiores valores de EC apresentaram maior reatividade e movimentação, o que resultou em aumento da taxa de consumo de glicogênio, diminuindo o pH inicial. Novilhos azebuados e mais agitados apresentam ganho de peso inferior e menores alterações quanto ao pH inicial e à cor da carne.

Para o cromossomo 17, dois SNP foram significativos, porém utilizando o banco de dados QTLdb (<http://www.animalgenome.org/cgi-bin/gbrowse/bovine/>) não foi encontrado nenhum estudo envolvendo as características de temperamento. O baixo poder de detecção da análise pode ter sido em decorrência do pequeno tamanho da amostra para uma característica de baixa herdabilidade ou pode se tratar de um falso-positivo. A utilidade dos marcadores SNP encontrados para características complexas, como as características de temperamento, são limitadas devido à natureza poligênica destes fenótipos (KEMPER; GODDARD, 2012)

No presente estudo foi possível a identificação de algumas regiões cromossômicas associadas com temperamento em animais da raça Nelore. Dentre os modelos estudados, o modelo 1 apresentou mais SNP associados tanto com VS quanto com EC_mediana. Entretanto, deve-se ter cautela com os resultados de GWAS provenientes de diferentes

metodologias de análise estatística. Estudos sobre temperamento, e consequentemente a identificação de SNP associados para VS e EC_mediana surgem como uma ferramenta que possibilita o melhor entendimento da arquitetura genética das características relacionadas à temperamento e ligadas à produção e qualidade do carne, facilidade no manejo, auxiliando assim, os programas de melhoramento genético na seleção de animais mais dóceis. Em suma, estes resultados devem ser validados na população, pois uma das grandes limitações deste estudo foi o tamanho amostral mensurados para VS e EC_mediana. Sabe-se que o tamanho amostral pode influenciar o número de casos de falsos-positivos ou associações negativas (WIENER et al., 2011). Segundo Stranger, Stahl e Raj (2011) maiores tamanhos populacionais permitem identificar regiões com menores efeitos.

Tabela 8 - Descrição dos QTLs descritos nos cromossomos para diferentes características reportadas na literatura e possivelmente associadas com EC_mediana usando o modelo ssGBLUP

(continua)

SNP	BTA	Posição	Posição do QTL (bp)	Característica	População	Referência
BovineHD0200011184	2	38325938	-	-	-	-
BovineHD0400031659	4	110611553	54975023 – 93254589 54975023 – 93254589	Social separation--Standing alert Social separation--Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0400031660	4	110612685	54975023 – 93254589 54975023 – 93254589	Social separation--Standing alert Social separation--Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0400031662	4	110614351	54975023 – 93254589 54975023 – 93254589	Social separation--Standing alert Social separation--Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD0900030978	9	105606912	46378796 – 46646401 24968872 – 47319594	Behavior QTL	bovinos de corte canadense Holândes x Charolês	Schmutz et al. (2001) Gutierrez-Gil et al. (2008)
BovineHD1200018096	12	66358707	-	-	-	-
BovineHD1200018108	12	66305151	-	-	-	-
BovineHD1800000692	18	2622724	2590139 – 10548225	Social separation--Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD2600007747	26	28905556	3202623 – 25229124	Social separation--Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD2900003371	29	11661654	33367872 - 47655654 33367872 – 47655654	Flight from feeder	Holândes x Charolês	Gutierrez-Gil et al. (2008)

Tabela 9 - Descrição dos QTLs descritos nos cromossomos para diferentes características reportadas na literatura e possivelmente associadas com com EC_mediana usando o modelo SNP por SNP

(conclusão)

SNP	BTA	Posição	Posição do QTL (bp)	Característica	População	Referência
BovineHD0700030364	7	103981034	18835910 -73909970 103765346 -103765346	Social separation-- Vocalization	Holândes x Charolês	Gutierrez-Gil et al. (2008)
BovineHD1700013457	17	47867274	-	-	-	-
BovineHD1700013737	17	48709448	-	-	-	-

3.4 Conclusões

(i) Vários marcadores SNP identificados nos cromossomos foram associados com VS e EC_mediana nessa população da raça Nelore, o que demonstra a existência de um componente genético contribuindo com a variação dessa característica.

(ii) Os resultados obtidos neste trabalho devem ser explorados com outras características de importância econômicas favorecendo o poder de detecção de SNP de menor efeito, além de outras metodologias. Estes resultados devem ser validados na população, pois uma das grandes limitações deste estudo foi o tamanho amostral mensurados para VS e EC_mediana.

Agradecimentos

À Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e à Rede BifeQuali. Em especial, a Bill Taylor e ao pessoal do Sistema de Computação do projeto “*Condor System*” da Universidade de Wisconsin, Madison, Estados Unidos, pelo auxílio e treinamento para o acesso do cluster dos computadores, os quais tornaram possíveis as análises desse estudo.

Referências

AGUILAR, I.; MISZTAL, I.; JOHNSON, D.L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Lancaster, v. 93, n. 2, p. 743–752, 2010.

AGUILAR, N.M.A. **Avaliação da reatividade de bovinos de corte e sua relação com caracteres reprodutivos e produtivos**. 2007. 69 p. Dissertação (Mestrado em Zootecnia) - Faculdade de Ciências Agrárias e Veterinárias. Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2007.

ALEXANDER, L.J.; GEARY, T.W.; SNELLING, W.M.; MACNEIL, M.D. Quantitative trait loci with additive effects on growth and carcass traits in a Wagyu-Limousin F2 population. **Animal genetics**, Malden, v. 38, n. 4, p. 413-416, 2007.

ALEXANDER, L.J.; MACNEIL, M.D.; GEARY, T.W.; SNELLING, W.M.; RULE, D.C.; SCANGA, J.A. Quantitative trait loci with additive effects on palatability and fatty acid composition of meat in a Wagyu-Limousin F2 population. **Animal Genetics**, Malden, v. 38, n. 5, p. 506-513, 2007.

ALLAIS, S.; JOURNAUX, L.; LEVÉZIEL, H.; PAYET-DUPRAT, N.; RAYNAUD, P.; HOCQUETTE, J.F.; LEPETIT, J.; ROUSSET, S.; DENOYELLE, C.; BERNARD-CAPEL, C.; RENAND, G. Effects of polymorphisms in the calpastatin and μ -calpain genes on meat tenderness in 3 French beef breeds. **Journal of Animal Science**, Savoy, v. 89, n. 1, p. 1-11, 2011.

ARDLIE, K.G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Nature Reviews Genetics**, London, v. 3, n. 4, p. 299-310, 2002.

BARENDSE, W.; BUNCH, R.J.; THOMAS, M.B.; HARRISON, B.E. A splice site single nucleotide polymorphism of the fatty acid binding protein 4 gene appears to be associated with intramuscular fat deposition in longissimus muscle in Australian cattle. **Animal Genetics**, Malden, v. 40, n. 5, p. 770-3, 2009.

BARRETT, J.C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, Oxford, v. 21, n. 2, p. 263-265, 2005.

BARROZO, D.; BUZANSKAS, M.E.; OLIVEIRA, J.A.; MUNARI, D. P. ; NEVES, H. H. R.; QUEIROZ, S. A. Genetic parameters and environmental effects on temperament score and reproductive traits of Nellore cattle, **Animal**, Cambridge, v. 6, n. 1, p. 36-40, 2012.

BARSH, G.S.; COPENHAVER, G.P.; GIBSON, G.; WILLIAMS, S.M. Guidelines for genome-wide association studies. **PLoS Genetics**, San Francisco, v. 8, n. 7, p. 1-2, 2012.

BORBA, L.H.F.; PIOVEZAN, U.; PARANHOS DACOSTA, M.J.R. Uma abordagem preliminar no estudo de associação entre escores de reatividade e características produtivas de bovinos de corte. In: ENCONTRO NACIONAL DE ETOLOGIA, 1997, São Carlos. **Anais...** São Carlos: SBET, 1997. p. 388.

BOURDON, R.M. **Understanding animal breeding**. 2nd ed. New Jersey: Prentice Hall Upper Saddle River, 2000. 538 p.

BURROW, H.M. Variances and covariances between productive and adaptive traits and temperament in composite breed of tropical beef cattle. **Livestock Production Science**, Montpellier, v. 70, p. 213-233, 2001.

BURROW, H.M.; DILLON, R. D. Relationships between temperament and growth in a feedlot and commercial carcass traits of *Bos indicus* crossbreds. **Australian Journal Experimental Agriculture**, Collingwood, v. 37, n. 4, p. 407-411, 1997.

BURROW, H.M.; SEIFERT, G.M.; CORBET, N.J. A new technique for measuring temperament in cattle. **Proceedings Australian Society of Animal Production**, Armidale, v. 17, p. 154-157, 1988.

CAFE, L.M.; ROBINSON, D.L.; FERGUSON, D.M.; McINTYRE, B.L.; GEESINK, G.H.; GREENWOOD, P.L. Cattle temperament: Persistence of assessments and associations with productivity, efficiency, carcass and meat quality traits. **Journal of Animal Science**, Savoy, v. 89, n. 5, p. 1452-1465, 2011.

CASAS, E.; KEELE, J.W.; SHACKELFORD, S.D.; KOOHMARAIE, M.; STONE, R.T. Identification of Quantitative Trait Loci for Growth and Carcass Composition in Cattle. **Animal Genetics**, Malden, v. 34, n. 1, p. 2-6, 2003.

CASAS, E.; SHACKELFORD, S.D.; KEELE, J.W.; KOOHMARAIE, M.; SMITH, T.P.L.; STONE, R.T. Detection of Quantitative Trait Loci for Growth and Carcass Composition in Cattle. **Journal of Animal Science**, Savoy, v. 81, n.12, p. 2976-2983, 2003.

CESAR, A.S.M.; REGITANO, L.C.A.; TULLIO, R.R.; LANNA, D.P.D.; NASSU, R.T.; MUDADO, M.A.; OLIVEIRA, P.S.N.; NASCIMENTO, M.L. do; CHAVES, A.S.; ROSA, A.N.; ALENCAR, M.M.; SONSTEGARD, T.; GARRICK, D.J.; REECY, J.M.; COUTINHO, L.L. Genome-wide association study for intramuscular fat deposition and composition in Nellore breed, 2013 In press.

CHAN, F.Y. **Genome wide association studies for temperament in New Zealand dairy cattle**. 2012.106 p. Thesis (Master of Science in Animal Science) – Massey University, New Zealand, 2012.

COLLIS, E.; FORTES, M.R.S.; ZHANG, Y.; TIER, B.; SCHUTT, K.; BARENDSE, W.; HAWKEN, R. Genetic variants affecting meat and milk production traits appear to have effects on reproduction traits in cattle. **Animal Genetics**, Malden, v. 43, n. 4, p. 442-446, 2012

COOKE, R.F.; BOHNERT, D.W.; M.; LOSI, T.C.; VASCONCELOS, J.L.M. Effects of temperament on pregnancy rates to fixed-timed AI in *Bos indicus* beef cows. **Livestock Science**, Montpellier, v. 142, p. 108-113, 2011.

CURLEY JR., K.O.; NEUENDORFF, D.A.; LEWIS, A.W. **Evaluation of temperament and stress physiology may be useful in breeding programs**. Beef cattle research in Texas publication, section physiology. Texas: College Station, Texas A& M University, p. 1-4, 2004.

DAVIS, G.P.; MOORE, S.S.; SHORTHORSE, W.R.; LOXTON, I.D.; BARENDSE, W.; HETZEL, D.J.S. QTL for meat tenderness in the M. longissimus lumborum of cattle. **Animal Genetics**, Malden, v. 39, n. 1, p. 40-45, 2007.

DOBSON, A.J. **An introduction to generalized linear models**. 2nd ed. Boca Raton: Chapman & Hall; CRC Press, 2002. 231p.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 5th ed. London: Longman, 1996. 464 p.

FIGUEIREDO, L.G.; ELER, J.P.; MOURÃO, G.B.; FERRAZ, J.B.S.; BALIEIRO, J.C.C.; MATTOS, E. Análise genética do temperamento em uma população da raça Nelore. **Livestock Research for Rural Development**, Cali, v. 17, n. 84, p. 1-7, 2005.

FORDYCE, G.; DODT, R.M.; WYTHES, J.R. Cattle temperaments in extensive beef herds in northern Queensland. Factors affecting temperament. **Australian Journal Experimental Agriculture**, Collingwood, v. 28, n. 6, p. 683-687, 1988.

FRANCISCO, C.L.; JORGE, A.M.; REZENDE, F.D.; SCHMIDEK, A.; BENATTI, J.M.B.; FARIA, M.H.; OBA, E.; COOKE, R.F. Effects of temperament on physiological responses, feedlot performance, and carcass characteristics of Nelore steers. **Journal of Animal Science**, Savoy, v. 90. E-Suppl. 3, p. 427, 2012.

GUTIÉRREZ-GIL, B.; BALL, N.; BURTON, D.; HASKELL, M.; WILLIAMS, J.L.; WIENER, P. Identification of quantitative trait loci affecting cattle temperament. **Journal of Heredity**, Washington, v. 99, n. 6, p. 629-638, 2008.

HALL, N.L.; BUCHANAN, D.S.; ANDERSON, V.L.; ILSE, B.R.; CARLIN, K.R.; BERG, E.P. Working chute behavior of feedlot cattle can be an indication of cattle temperament and beef carcass composition and quality. **Meat Science**, Barking, v. 89, n. 1, p. 52–57, 2011.

HEARNshaw, H.; MORRIS, C.A. Genetic and environmental effects on a temperament score in beef cattle. **Australian Journal of Agricultural Research**, Melbourne, v. 35, n. 5, p. 723-733, 1984.

HIENDLEDER, H.; THOMSEN, H.; REINSCH, N.; BENNEWITZ, J.; LEYHE-HORN, B.; LOOFT, C.; XU, N.; MEDJUGORAC, I.; RUSS, I.; KÜHN, C.; BROCKMANN, G.A.; BLÜMEL, J.; REINHARDT, F.; REENTS, R.; AVERDUNK, G.; SCHWERIN, M.; FÖRSTER, M.; KALM, E.; ERHARDT, G. Mapping of QTL for body conformation and behavior in cattle. **Journal of Heredity**, Washington, v. 94, n. 6, p. 496-506, 2003.

HÖGLUND, J.K.; GULDBRANDTSEN, B.; LUND, M.S.; SAHANA, G. Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. **BMC Genetics**, London, v. 13, n. 71, p. 1-9, 2012.

HOPPE, S.; BRANDT, H.R.; KÖNIG, S.; ERHARDT, G.; GAULY, M. Temperament traits of beef calves measured under field conditions and their relationships to performance. **Journal of Animal Science**, Savoy, v. 88, n. 6, p. 1982–1989, 2010.

HU, Z.L.; PARK, C.A.; FRITZ, E.C.; REECY, J.M. QTLdb: a comprehensive database tool building bridges between genotypes and phenotypes. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 9., 2010, Leipzig.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 1998. 816 p.

KEMPER, K.E.; GODDARD, M.E. Understanding and predicting complex traits: knowledge from cattle. **Human Molecular Genetics**, San Francisco, v. 21, n. R1, p. R45-R51, 2012.

KIM, J.J.; FARNIR, F.; SAVELL, J.; TAYLOR, J.F. Detection of Quantitative Trait Loci for Growth and Beef Carcass Fatness Traits in a Cross Between *Bos taurus* (Angus) and *Bos indicus* (Brahman) Cattle. **Journal of Animal Science**, Savoy, v. 81, n. 8, p. 1933-1942, 2003.

KINSELLA, R.J.; KÄHÄRI, A.; HAIDER, S.; ZAMORA, J.; PROCTOR, G.; SPUDICH, G.; ALMEIDA-KING, J.; STAINES, D.; DERWENT, P.; KERHORNou, A.; KERSEY, P.; FLICEK, P. **Ensembl BioMart**: a hub for data retrieval across taxonomic space. Database.

Oxford, 201. Disponível em: <<http://www.ensembl.org/index.html>>. Acesso em: 10 set. 2013.

LEE, S.H.; VAN DER WERF, J.H.; LEE, S.H.; PARK, E.W.; OH, S.J.; GIBSON, J.P.; THOMPSON, J.M. Genetic polymorphisms of the bovine Fatty acid binding protein 4 gene are significantly associated with marbling and carcass weight in Hanwoo (Korean Cattle). **Animal Genetics**, Malden, v. 41, n. 4, p. 442-4, 2010.

MANLY, B.J.F. **Métodos estatísticos multivariados: uma introdução**. 3. ed. Porto Alegre: Bookman, 2008. 229 p.

McCLURE, M.C.; MORSCI, N.S.; SCHNABEL, R.D.; KIM, J.W.; YAO, P.; ROL, M.M.; MCKAY, S.D.; GREGG, S.J.; CHAPPLE, R.H.; NORTHCUTT, S.L.; TAYLOR, J.F. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. **Animal Genetics**, Malden, v. 41, n. 6, p. 597-607, 2010.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D.H. BLUPF90 and related programs BGF90. In: THE WORLD CONGRESS GENETICS APPLICATION LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings...** Montpellier :INRA;CIRAD, 2002. p. 467-468.

MORRIS, C.A.; BOTTEMA, C.D.; CULLEN, N.G.; HICKEY, S.M.; ESMAILIZADEH, A.K.; SIEBERT, B.D.; PITCHFORD, W.S. Quantitative trait loci for organ weights and adipose fat composition in Jersey and Limousin back-cross cattle finished on pasture or feedlot. **Animal Genetics**, Malden, v. 41, n. 6, p. 589-596, 2010.

MORRIS, C.A.; PITCHFORD, W.S.; CULLEN, N.G.; ESMAILIZADEH, A.K.; HICKEY, S.M.; HYNDMAN, D.; DODDS, K.G.; AFOLAYAN, R.A.; CRAWFORD, A.M. Quantitative trait loci for live animal and carcass composition traits in Jersey and Limousin back-cross cattle finished on pasture or feedlot. **Animal Genetics**, Malden, v. 40, n. 5, p. 648-54, 2009.

MOURÃO, G.B.; BERGMANN, J.A.G.; FERREIRA, M.B.D. Diferenças genéticas e estimação de coeficientes de herdabilidade para temperamento em fêmeas zebu e F1 Holandês x Zebu. **Revista Brasileira Zootecnia**, Viçosa, v. 27, n. 4, p. 722-729, 1998.

MÜLLER, R.; VON KEYSERLINGK, M.A.G. Consistency of flight speed and its correlation to productivity and to personality in *Bos Taurus* beef cattle. **Applied Animal Behaviour Science**, Amsterdam, v. 99, n. 3, p. 193-204, 2006.

NASCIMENTO, M.L. **Eficiência alimentar e suas associações com lucro, características de carcaça e qualidade de carne de bovinos Nelore**. 2011. 119 p. Tese (Doutorado em Ciências) – Escola Superior de Agricultura “Luiz de Queiroz”. Universidade de São Paulo, Piracicaba, 2011.

NALAILA, S.M.; STOTHARD, P.; MOORE, S.S.; LI, C.; WANG, Z. Whole-genome QTL scan for ultrasound and carcass merit traits in beef cattle using Bayesian shrinkage method. **Journal of Animal Breeding and Genetics**, Berlin, v. 129, n. 2, p. 107-119, 2012.

OBER, U.; AYROLES, J.F.; STONE, E.A.; RICHARDS, S.; ZHU, D.; GIBBS, R.A.; STRICKER, C.; GIANOLA, D.; SCHLATHER, M.; MACKAY, T.F.C.; SIMIANER, H. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. **Plos Genetics**, San Francisco, v. 8, n. 5, p. 1-14, 2012.

OLIPHANT, A.; BARKER, D.L.; STUELPNAGEL, J.R.; CHEE, M.S. Bead Array genotyping. **BioTechniques**, Natick, v. 32, n. 1, p.56-61, 2002.

PAÇÓ, A.L. **Estudo de fatores genéticos e de ambiente para reatividade na raça Nelore**. 2012. 58 p. Dissertação (Mestrado em Zootecnia) - Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2012.

PEARSON, T.A.; MANOLIO, T.A. How to interpret a genome-wide association study. **Journal of the American Medical Association**, Chicago, v. 299, n. 11, p. 1335-1344, 2008.

PEÑAGARICANO, F.; WEIGEL, K.A.; KHATIB, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. **Animal Genetics**, Malden, v. 43, n. 1, p. 65-71, 2012.

PENG, G.; LUO, L.; SIU, H.; ZHU, Y.; HU, P.; ZHAO, J.; ZHOU, X.; REVEILLE, J.D.; AMOS, C.I.; XIONG, M. Gene and pathway-based second-wave analysis of genome-wide association studies. **European Journal of Human Genetics**, London, v. 18, n. 1, p. 111-117, 2010.

PETHERICK, J.C.; HOLROYD, R.G.; DOOGAN, V.J.; VENUS, B.K. Productivity, carcass and meat quality for lot fed *Bos indicus* cross steers grouped according to temperament. **Australian Journal Experimental Agriculture**, Collingwood, v. 42, n. 4, p. 389-398, 2002.

PHOCAS, F.; BOIVIN, X.; SAPA, J.; TRILLAT, G.; BOISSY, A.; NEINDRE, L.P. Genetic correlations between temperament and breeding traits in Limousin heifers. **Journal of Animal Science**, Savoy, v. 82, n. 6, p. 805–811, 2006.

PINTOS, D.; CORVA, P.M. Association between molecular markers for beef tenderness and growth traits in Argentinian angus cattle. **Animal Genetics**, Malden, v. 42, n. 3, p. 329-332, 2011.

PIOVESAN, U. **Análise de fatores genéticos e ambientais na reatividade de quatro raças de bovinos de corte ao manejo**. 1998. 51 p. Dissertação (Mestrado em Zootecnia) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 1998.

PRYCE, J.E.; BOLORMAA, S.; CHAMBERLAIN, A.J.; BOWMAN, P.J.; SAVIN, K.; GODDARD, M.E.; HAYES, B.J. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. **Journal of Dairy Science**, Lancaster, v. 93, n. 7, p. 3331-3345, 2010.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P.I.; DALY, M.J.; SHAM, P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, Cambridge, v. 81, n. 3, p. 559-575, 2007.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013.

REARDON, W.; MULLEN, A.M.; SWEENEY, T.; HAMILL, R.M. Association of polymorphisms in candidate genes with colour, water-holding capacity, and composition traits in bovine *M. longissimus* and *M. semimembranosus*. **Meat Science**, Barking, v. 86, n. 2, p. 270-275, 2010.

SANT'ANNA, A.C.; PARANHOS DA COSTA, M.J.; BALDI, F.; RUEDA, P.M.; ALBUQUERQUE, L.G. Genetic associations between flight speed and growth traits in Nellore cattle. **Journal of Animal Science**, Savoy, v. 90, n. 10, p. 427-3432, 2012.

SANT'ANNA, A.C.; PARANHOS DA COSTA, M.J.; BALDI, F.; ALBUQUERQUE, L.G. Genetic variability for temperament indicators of Nellore cattle. **Journal of Animal Science**, Savoy, v. 91, n. 8, p. 3532-3537, 2013.

SCHMUTZ, S.M.; STOOKEY, J.M.; WINKELMAN-SIM, D.C.; WALTZ, C. S.; PLANTE, Y.; BUCHANAN, F.C. A QTL study of cattle behavioral traits in embryo transfer families. **Journal of Heredity**, Washington, v. 92, n. 3, p. 290-292, 2001.

SILVEIRA, I.D.B.; FISCHER, V.; JORGE, G.; SOARES, D. Relação entre o genótipo e o temperamento de novilhos em pastejo e seu efeito na qualidade da carne. **Revista Brasileira de Zootecnia**, Viçosa, v. 35, n. 2, p. 519-526, 2006.

SILVERA, I.D.B.; FISCHER, V.; FARINATTI, L.H.E.; RESTLE, J.; ALVES FILHO, D.C. Relação entre genótipos e temperamento de novilhos Charolês x Nelore em confinamento. **Revista Brasileira de Zootecnia**, Viçosa, v. 37, n. 10, p. 1808-1814, 2008.

SPELMAN, R.J.; HUISMA, A.E.; SINGIREDDY, W.; COPPIETERS, J.; ARRANZ, M.G.; GARRICK, D.J. Short communication: quantitative trait loci analysis on 17 Non production traits in the New Zealand dairy population. **Journal of Dairy Science**, Lancaster, v. 82, p. 2514-2516, 1999.

STRANGER, B.B.; STAHL, E.A.; RAJ, T. Progress and promise of genome-wide association studies for human complex trait genetics, **Genetics**, Cambridge, v. 18 n. 23, p. 67-83, 2011.

TAKASUGA, A.; WATANABE, T.; MIZOGUCHI, Y.; HIRANO, T.; IHARA, N.; TAKANO, A.; YOKOUCHI, K.; FUJIKAWA, A.; CHIBA, K.; KOBAYASHI, N.; TATSUDA, K.; OE, T.; FURUKAWA-KUROIWA, M.; NISHIMURA-ABE, A.; FUJITA, T.; INOUE, K.; MIZOSHITA, K.; OGINO, A.; SUGIMOTO, Y. Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. **Mammalian Genome**, New York, v. 18, n. 2, p. 125-36, 2007.

TAXIS, T.M. **Phenotypic and genetic effects of disposition on beef tenderness and quality attributes**. 2011. 98 p. Thesis (Master of Science) - University of Missouri. Missouri, 2011.

TIZIOTO, P.C.; DECKER, J.E.; TAYLOR, J.F.; SCHNABEL, R.D.; MUDADU, M.A.; SILVA, F.L.; MOURÃO, G.B.; COUTINHO, L.L.; THOLON, P.; SONSTEGARD, T.S.; ROSA, A.N.; ALENCAR, M.M.; TULLIO, R.R.; MEDEIROS, S.R.; NASSU, R.T.; FEIJÓ,

G.L.D.; SILVA, L.O.C.; TORRES, R.A.; SIQUEIRA, HIGA, F.; R.; REGITANO, L.C.A. A genome scan for meat quality traits in Nellore beef cattle. **Physiological Genomics**, Bethesda, v. 10, p. 1–23, 2013.

TM TECHNOLOGY. **Enabling an accurate, cost-effective approach to high-throughput** Upper Saddle River, 2000.

VANRADEN, P.M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Lancaster, v. 91, n. 11, p. 4414–4423, 2008.

VAZQUEZ, A.I.; BATES, D.M.; ROSA, G.J.M.; GIANOLA, D.; WEIGEL, K.A. Technical note: an R package for fitting generalized linear mixed models in animal breeding. **Journal of Animal Science**, Savoy, v. 88, n. 2, p. 497–504, 2010.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W.M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, Cambridge, v. 94, n. 2, p. 73–83, 2012.

WIENER, P; EDRISS, M.A; WILLIAMS, J. L.; WADDINGTON, D.; LAW A.; JOHN A WOOLLIAMS, J.A.; GUTIÉRREZ-GIL, B. Information content in genome-wide scans: concordance between patterns of genetic differentiation and linkage mapping associations. **BMC Genomics**, London, v. 12, n. 65, p. 1–10, 2011.

YOKOUCHI, K.; MIZOGUCHI, Y.; WATANABE, T.; IWAMOTO, E.; SUGIMOTO, Y.; TAKASUGA, A. Identification of a 3.7-Mb region for a marbling QTL on bovine chromosome 4 by identical-by-descent and association analysis. **Animal Genetics**, Malden, v. 40, n. 6, p. 945–951, 2009.

ZHANG, J.; BENAVENTE, C.A.; McEVOY, J.; FLORES-OTERO, J.; DING, L.; CHEN, X.; ULYANOV, A.; WU, G.; WILSON, M.; WANG, J.; BRENNAN, R.; RUSCH, M.; MANNING, A.L.; MA, J.; EASTON, J.; SHURTLEFF, S.; MULLIGHAN, C.; POUNDS, S.; MUKATIRA, S.; GUPTA, P.; NEALE, G.; ZHAO, D.; LU, C.; FULTON, R.S.; FULTON, L.L.; HONG, X.; DOOLING, D.J.; OCHOA, K.; NAEVE, C.; DYSON, N.J.; MARDIS, E.R.; BAHRAMI, A.; ELLISON, D.; WILOSN, R.K.; DOWNING, J.R.; DYER, M.A. A novel retinoblastoma therapy from genomic and epigenetic analyses. **Nature**, London, v. 481, n. 7381, p. 329–334, 2012.

4 ACURÁCIA DE IMPUTAÇÃO DE GENÓTIPOS EM BOVINOS DA RAÇA NELORE

Resumo

Estudos de associação genômica (GWAS) e seleção genômica (GS), geralmente requerem que um grande número de indivíduos sejam ambos fenotipados e densamente genotipados para marcadores SNP. No entanto, a genotipagem de muitos animais com painéis de marcadores de alta densidade pode ser cara, e economicamente inviável. Uma alternativa, neste contexto, é a utilização de chips de baixa ou média densidade de marcadores SNP, e posteriormente a predição para chip de alta densidade. O objetivo deste estudo foi avaliar a acurácia de imputação em bovinos da raça Nelore, usando dois painéis de baixa densidade (3K e 6K) e um de média densidade (50K) para marcadores de alta densidade (800K). Foram utilizados dados de 779 novinhos, pertencentes a 34 famílias de meios-irmãos, nascidos no período de 2007 e 2009 nas fazendas experimentais da Embrapa Pecuária Sudeste e Embrapa Gado de Corte e três fazendas adicionais no Estado de Mato Grosso do Sul, e Mato Grosso, Brasil. Estas famílias de meios-irmãos paternos representam amplamente a raça Nelore, e os touros têm sido utilizados em diversos programas de melhoramento no Brasil. Todos os animais foram genotipados usando o chip de alta densidade Illumina BovineHD BeadChip (800K). Os animais genotipados para alta densidade foram divididos em população de validação e população de referência. Na população de validação, os animais tiveram seus genótipos mascarados para os chips Illumina Bovine3K, Illumina BovineLD 6K BeadChip e BeadChip Illumina BovineSNP50, de forma a obter painéis semelhantes de baixa densidade e média densidade disponíveis no mercado para comercialização e, posteriormente imputados para o chip de alta densidade (800K). Na população de referência alguns animais permaneceram genotipados para o chip de 800K. A imputação de 3K, 6K e 50K para 800K foi realizada utilizando os softwares de imputação fastPHASE e Beagle. O software fastPHASE apresentou maiores valores de acurácia em comparação com o Beagle. Os painéis 6K e 50K apresentaram maiores valores de acurácia, porém os painéis 3K e 50K podem ser utilizados em animais da raça Nelore. O cenário com 34 touros no painel de referência apresentou maior valor de acurácia, sendo o cenário mais adequado em termos de ganho em acurácia e economia na genotipagem de animais.

Palavras-chave: Gado de corte; Imputação; Nelore; SNP

Abstract

Genome-wide association studies (GWAS) and genomic selection (GS) generally requires large numbers of individuals that are both phenotyped and densely genotyped for markers across the genome. However, genotyping many animals with a high-density marker panel can be expensive and economically unfeasible. An alternative in this context is to use a lower density marker panel on a larger number of animals, and impute the missing genotypes. Thus, the objective this study was to evaluate the accuracy of imputation in a Nelore cattle population, using different panels of low-density (3K and 6K) and medium-density (50K) markers. The data included 779 Nelore males belonging to 34 half-sib families born between 2007 and 2009 at experimental farms of Embrapa Cattle-Southeast and Embrapa Beef Cattle, and three additional farms in the State of Mato Grosso do Sul and Mato Grosso, Brazil. These paternal half-sib families represented Nelore bulls widely used in several breeding programs in Brazil. Genotyped animals were split into a reference and validation population. All

animals were genotyped using the Illumina BovineHD BeadChip (800K). The validation population had markers from the high-density marker panel masked in order to obtain low-density and medium-density panels. Imputation from 3K, 6K and 50K up to 800K markers was performed using fastPHASE and Beagle. The fastPHASE was the greatest in accuracy compared to Beagle. The panels 6K and 50K showed higher accuracy, however 3K and 50K panels might be recommend cattle. The scenario with 34 sires in the reference panel had higher accuracy, being the most suitable in conditions of gain in accuracy and economic.

Keywords: Beef cattle; Imputation; Nellore, SNP

4.1 Introdução

A seleção genômica (GS) proposta por Meuwissen, Hayes e Goddard, (2001) permite a seleção dos indivíduos baseados nos seus respectivos valores genéticos genômicos (GEBV). Os GEBV são estimados através da soma dos efeitos dos marcadores, ou haplótipos inferidos no genoma com a finalidade de capturar todos os *quantitative trait loci* (QTLs) que contribuem para a variação da característica (GODDARD; HAYES, 2007; VAN TASSELL et al., 2008; HAYES et al., 2009; HEFFENER et al., 2009; LUAN et al., 2009; HAYES et al., 2011). Os efeitos de QTLs uma vez inferidos a partir de inúmeros marcadores do tipo polimorfismos de base única (SNP) ou haplótipos na primeira geração, em uma população de referência contendo informação fenotípica, permitirá que nas próximas gerações apenas as informações de marcadores sejam suficiente para calcular os GEBV dos animais (HAYES et al., 2009).

A aplicação de GS nos programas de melhoramento genético ainda é muito cara quando utilizada plataformas de marcadores de alta densidade, tornando-se quase que inviável sua implementação em rebanhos comerciais para predição dos GEBV. Habier, Fernando e Dekkers (2009) propuseram a predição de GEBVs dos candidatos à seleção utilizando painéis de baixa densidade e, posteriormente, utilizar metodologia de imputação para prever genótipos desconhecidos/faltantes para o chip de alta densidade. A abordagem de imputação de genótipos pode ser uma alternativa eficaz para reduzir os custos de genotipagem de todos os animais, com milhares de marcadores SNP, especialmente nos casos em que características apresentam estimativas de herdabilidade baixa (HABIER; FERNANDO; DEKKERS, 2009, MILLER, 2010, WEIGEL et al., 2010, HAYES et al., 2011). Na prática, os touros seriam genotipados com painéis de alta densidade (800K), e seus descendentes genotipados com painéis de baixa densidade (3K, 6K, dentre outros).

A imputação de genótipos provenientes de painéis de baixa e média densidade para alta densidade pode ser utilizada em diferentes situações: (i) aumentar o número de SNP nos

estudos de associação genômica (*genome-wide association* – GWAS) (MARCHINI; HOWIE, 2010; MISZTAL, 2011; BOICHARD et al., 2012; CHEN et al., 2012; DASSONNEVILLE et al., 2012; MULDER et al., 2012; SUN et al., 2012;); (ii) combinar diferentes plataformas de marcadores (LI et al., 2009; KHATKAR et al., 2012); (iii) imputar genótipos que não foram obtidos por motivos de falha da técnica de genotipagem. A imputação de genótipos tem sido utilizada em estudos com humanos (HAO et al., 2009; MARCHINI; HOWIE, 2010) e, recentemente, com animais (DRUET; SCHROOTEN; DE ROOS, 2010; WEIGEL et al., 2010; ZHANG; DRUET, 2010; HAYES et al., 2011; HICKEY et al., 2011; MULDER et al., 2012; VANRADEN et al., 2013). Em estudos com humanos, os genótipos não obtidos ou faltantes são inferidos principalmente pela informação de desequilíbrio de ligação (LD) entre os SNP (SUN et al., 2012) empregando os softwares fastPHASE (SCHEET; STEPHENS, 2006), Beagle (BROWNING; BROWNING, 2009) e IMPUTE (HOWIE; DONNELLY; MARCHINI, 2009).

A eficiência dos métodos de imputação pode ser afetada por vários fatores, dentre eles: (i) a densidade de marcadores (KHATKAR et al., 2012; SEGELKE et al., 2012); (ii) parentesco entre animais da população de referência e animais da população de validação (HAYES et al., 2011; PAUSCH et al., 2013); (iii) o número de animais no painel de referência (KHATKAR et al., 2012); (iv) nível de LD na população (WEIGEL et al., 2010; DRUET; SCHROOTEN; DE ROOS, 2010); (v) método de imputação utilizado (ZHANG; DRUET, 2010).

Sendo assim, o objetivo desse estudo foi avaliar a acurácia de imputação em bovinos da raça Nelore, considerando dois painéis de baixa densidade (3K e 6K) e um de média densidade (50K) para marcadores de alta densidade (800K).

4.2 Material e Métodos

Animais

Foram utilizados 779 animais, novilhos machos, descendentes de 34 touros Nelore registrados, pertencentes às principais linhagens da raça de uso comercial mais frequente no Brasil. Estes touros foram selecionados com base nos coeficientes de parentesco, com o objetivo de minimizar o grau de parentesco e representar a variabilidade genética da raça no país. As progênies foram geradas em três estações de monta, nos anos de 2007, 2008 e 2009, por meio de inseminação artificial em tempo fixo (IATF) e criados na Embrapa Pecuária

Sudeste (CPPSE), localizada na cidade de São Carlos-SP, na Embrapa Gado de Corte (CNPGC), situada no município de Campo Grande-MS e propriedades particulares dos Estados de Mato Grosso do Sul e Mato Grosso.

Os novilhos foram criados e recriados a pasto nas fazendas de origem, desmamados com média de 8 meses de idade, pesados na desmama e no sobreano. Os animais foram castrados quando atingiram aproximadamente 17 meses de idade e, em seguida foram transportados para as estações experimentais, CPPSE ou CNPGC com $\pm$ 18 meses de idade, para serem confinados. Os animais produzidos nas unidades da Embrapa permaneceram nas próprias unidades e aqueles oriundos das fazendas parceiras foram divididos e alocados nas estações experimentais, em função da paternidade, para garantir sua representatividade nos diferentes locais de avaliação (Nascimento, 2011).

O período experimental de confinamento compreendido entre o início das avaliações e o abate, foi determinado em função do grau de acabamento que os animais deveriam atingir, sendo determinado em 5mm de espessura de gordura subcutânea. Em geral, teve duração de cerca de 100 dias. Nesse período foram avaliadas características de eficiência alimentar, temperamento e qualidade de carcaça. Os animais foram submetidos a um período mínimo de 28 dias de adaptação à dieta e posteriormente foram adaptados ao sistema de alimentação com cancelas eletrônicas no confinamento (Nascimento, 2011).

Genotipagem e Controle de Qualidade

A genotipagem foi realizada no Laboratório de Genômica Funcional do Departamento de Agricultura dos Estados Unidos (BFGL/USDA/EUA), e no Laboratório Multiusuários de Genômica Funcional Aplicada à Agropecuária e à Agroenergia da ESALQ – USP, financiada pelo projeto Rede Bife de Qualidade, Macroprograma 1 da Embrapa. Foi utilizado o chip “Illumina BovineHD BeadChip”, que possui aproximadamente 777 mil SNP em sua plataforma, com espaçamento médio de 3,4 kb. Em seguida, os dados genômicos foram filtrados a fim de serem eliminados os resultados de baixa qualidade e aumentar a acurácia das análises (OLIPHANT et al., 2002). Os seguintes critérios foram utilizados: (i) indivíduos com “call rate” >95% (porcentagem do genoma genotipado corretamente); (ii) SNP com “call rate” > 95% (porcentagem de genotipagem de cada SNP em toda a população); (iii) *minor allelic frequency* (MAF) foi estabelecida em 1%. Foram excluídos ainda os SNP localizados no cromossomo X e no Y, e aqueles sem informação de posição no genoma. A análise de

controle de qualidade foi realizada por meio do software Plink (PURCELL et al., 2007). Os resultados do controle de qualidade estão apresentados na Tabela 1.

Tabela 1 - Controle de qualidade utilizado no painel BovineHD BeadChip (800K)

Critério de controle de Qualidade	Removidos
<i>Minor allelic frequency</i> (1%)	160.872 SNP
<i>Call rate</i> /SNP (95%)	30.882 SNP
<i>Call rate</i> /animal (95%)	10 animais
Total SNP removidos	191.754
Total de SNP utilizados	560.142

Criação de painéis de baixa e média densidade de marcadores SNP

Dois painéis de baixa densidade (3K e 6K) e um de média densidade (50K) foram criados a partir do painel de 800K com a finalidade de avaliar a acurácia de imputação na população Nelore. Esses painéis de densidade criados foram baseados nos chips de SNP disponíveis no mercado para bovinos de corte (Tabela 2). As informações sobre a identificação dos SNP e posição no genoma podem ser encontradas em [http://animalgenome.org/repository/cattle/ repositório](http://animalgenome.org/repository/cattle/). Em geral, a maior parte do SNP nos painéis 3K, 50K e 6K estão presentes no chip de alta densidade (800K). O número de SNP dos painéis 3K, 6K, 50K para 800K são apresentados na Tabela 2. A imputação foi realizada para todos os 29 cromossomos autossômicos. Em geral 95 à 99% dos SNP contidos no chip de 800K foram mascarados para a criação de novos chips (Tabela 3).

Tabela 2 - Número de SNP presente em diferentes chips de baixa e média densidade de SNP e no BovineHD BeadChip (800K)

Chip	Descrição	Número de SNP no chip	Número de SNP no chip com o QC
3K	Illumina Bovine3K BeadChip	2.661	2.249
6K	Illumina BovineLD BeadChip	6.855	5.398
50K	Illumina Bovine50K BeadChip	49.915	29.009
800K	Illumina 800K BovineHD beadChip	742.906	560.142

QC = controle de qualidade

Tabela 3 – Descrição dos cenários de imputação para os diferentes chips de SNP utilizados no estudo para avaliar a acurácia de imputação

Chip	N SNP total	N SNP mascarados	% N SNP mascarados
3K/800K	560.142	557.893	99
6K/800K	560.142	554.744	99
50K/800K	560.142	531.133	95

3K/800K: genótipos imputados a partir do 3K para 800K; 6K/800K: genótipos imputados a partir do 6K para 800K; 50K/800K: genótipos imputados a partir do 50K para 800K

Cenários de imputação

Animais genotipados com o painel de alta densidade de SNP (800K) foram divididos em população de validação e população de referência. Os animais incluídos na população de validação para baixa (3K e 6K) e média densidade (50K) tiveram seus genótipos derivados do chip de 800K. Foram estudados dois cenários de imputação: (1) 34 touros genotipados com 800K foram incluídos na população de referência e 779 progênieis na população de validação. Esta estratégia de imputação foi avaliada com o objetivo de verificar a acurácia de imputação em situação prática, em que os touros geralmente são genotipados com painel de alta densidade e progênieis são genotipados, com chip de baixa ou média densidade; (2) 215 animais (34 touros e 181 progênieis) foram utilizados na população de referência, e 598 animais na população de validação. Na Figura 1 é apresentado o esquema dos cenários de imputação utilizados neste estudo. Em geral, 74 à 96% dos 813 animais tiveram seus genótipos mascarados (Tabela 4). Deve-se ressaltar que o objetivo deste estudo não foi selecionar o melhor cenário de imputação, mas avaliar dentro de cada cenário o painel mais adequado para fins de imputação (3K, 6K ou 50K) para outros painéis de alta densidade, e possivelmente serem utilizados nos estudos de GWAS ou seleção genômica para animais da raça Nelore.

Figura 1 – Diagrama dos cenários de imputação para a população de referência e população de validação

Tabela 4 - Descrição dos cenários de imputação para os diferentes chips de SNP utilizados no estudo para avaliar a acurácia de imputação

Cenário	Chip	Animais mascarados validação (%)	N animais total	N animais validação	N animais referência
1	3K/800K	96	813	779	34
2	3K/800K	74	813	598	215
1	6K/800K	96	813	779	34
2	6K/800K	74	813	598	215
1	50K/800K	96	813	779	34
2	50K/800K	74	813	598	215

3K/800K: genótipos imputados a partir do 3K para 800K; 6K/800K: genótipos imputados a partir do 6K para 800K; 50K/800K: genótipos imputados a partir do 50K para 800K

Métodos de imputação

Neste estudo foi avaliada a acurácia de imputação utilizando duas metodologias de imputação baseadas em estrutura populacional: fastPHASE versão 1.2 (SCHEET;

STEPHENS, 2006) e Beagle versão 3.2.1 (BROWNING; BROWNING, 2009). (Beagle 3.3.2 e fastPHASE 2.0) em bovinos da raça Nelore. Ambos os softwares são baseados em cadeia de markov oculta (HMM).

O software fastPHASE versão 1.2 (SCHEET; STEPHENS, 2006) é baseado em um modelo de agrupamento e usa a observação que os haplótipos relacionados ou similares dos indivíduos dentro da população se aglomeraram em grupos. De acordo com Marchini e Howie (2010) o modelo especifica um conjunto de K estados não observado ou um grupo que é designado para representar os haplótipos comuns significativos para representar haplótipos em comum. O k -ésimo grupo recebe um peso (α_{kl}) que denota à proporção que este grupo de haplótipo é observado no loco l :

$$\sum_k \alpha_{kl} = 1$$

sendo assim, cada grupo possui uma frequência associada (θ_{kl}) para o alelo l em cada loco. Os dados dos genótipos de cada indivíduo são analisados por meio de HMM sobre estes estados com transição entre os estados controlados por um conjunto de parâmetros (r) para cada SNP. Dessa forma, a probabilidade de transição é dada por:

$$P(G_i | \alpha, \theta, r) = \sum_z P(G_i | Z_i, \theta) P(Z_i | \alpha, r)$$

em que $P(G_i | Z_i, \theta)$ modela o provável genótipo observado dado os estados de transição e $P(Z_i | \alpha, r)$ é a probabilidade de transição entre os estados, mas os estados representam grupos ao invés de haplótipos de referência. Portanto, a função de verossimilhança pode ser escrita como:

$$L(G, H | \alpha, \theta, r) = \prod_i P(G_i | \alpha, \theta, r) \prod_i P(H_i | \alpha, \theta, r)$$

O software fastPHASE além de predizer genótipos desconhecidos, também infere haplótipos. O modelo considera um número fixo de haplótipos e cada um destes haplótipos de referência representa um estado de markov oculto. O algoritmo de maximização da esperança, algoritmo EM, é usado para estimar os parâmetros genéticos, frequência dos haplótipos e fase desconhecida do haplótipo. Mais detalhes sobre a metodologia são relatados por Scheet e Stephens (2006).

O software Beagle versão 3.2.1 (BROWNING; BROWNING, 2007) é baseado no modelo gráfico DAG (*directed acyclic graph*) de um conjunto de haplótipos localizados, em

que utiliza um painel de referência para estimar a fase haplótipo e genótipos desconhecidos no conjunto de animais de validação. Similar ao software fastPHASE, apresenta um modelo de agrupamento dos haplótipos semelhantes, entretanto a forma como estes grupos são formados é diferente. No fastPHASE o parâmetro K (número de grupos) é fixo, porém no Beagle o número de grupos pode variar. Permite inferir a fase do haplótipo em grandes conjuntos de animais não aparentados e trios de pais e filhos quando disponível. Esse método tem como *default* dez iterações do modelo de construção. Uma explicação detalhada da implementação do algoritmo é dada em Browning e Browning (2007).

Ambos os softwares fornecem as probabilidade à *posteriori* dos possíveis genótipos correspondentes AA, AB e BB e, imputaram todos os genótipos faltantes dos marcadores SNP, tanto presentes na população de referência quanto na população de validação. As análises foram realizadas individualmente para cada cromossomo nos diferentes cenários de imputação definidos anteriormente. O número de SNP presentes nos chips de baixa e média densidade nos 29 cromossomos após o controle de qualidade é apresentado na Tabela 5. Os resultados de imputação apresentados neste trabalho foram obtidos a partir dos parâmetros de *default* para o software Beagle (java --Xmx2048 --jar unphased=unphased1.input.bgl unphased=unphased2.input.bgl markers=marker.ids missing=0 niterations=10 out=out_file) e para o fastPHASE (./fastPHASE_Linux -oX X.inp)

Tabela 5- Número de SNP para os diferentes painéis utilizado na avaliação da acurácia de imputação dos 29 cromossomos

Cromossomo	Número de SNP na população de validação 3K	Número de SNP na população de validação 6K	Número de SNP na população de validação 50K	Número de SNP na população de referência 800K
BTA1	149	336	1974	36527
BTA2	125	270	1495	29875
BTA3	102	257	1372	27192
BTA4	117	259	1406	26637
BTA5	106	231	1140	25038
BTA6	112	275	1585	28244
BTA7	104	233	1276	25415
BTA8	100	243	1273	26033
BTA9	95	235	1204	24733
BTA10	96	211	1201	22316
BTA11	90	234	1270	24290
BTA12	77	169	881	18773
BTA13	67	166	899	17310
BTA14	74	190	1131	20545
BTA15	74	195	1020	19465
BTA16	69	167	919	18452
BTA17	66	161	902	16886
BTA18	58	147	728	14883
BTA19	57	149	736	13436
BTA20	65	162	869	16369
BTA21	66	155	790	16677
BTA22	62	152	767	13252
BTA23	45	130	622	11866
BTA24	63	137	720	14058
BTA25	34	103	495	8959
BTA26	48	115	636	11915
BTA27	40	103	554	9912
BTA28	42	102	549	10266
BTA29	46	111	595	10818
Total	2249	5398	29009	560142

Acurácia de imputação

A acurácia de imputação foi calculada para cada animal como o número de genótipos imputados corretamente divididos pelo número de genótipos imputados. A proporção de genótipos imputados corretamente foi baseada nas probabilidades à posteriori para os três genótipos possíveis (AA, AB e BB) dos SNP fornecidos para cada animal.

4.3 Resultados e Discussão

Softwares

Na Tabela 6 é apresentada os resultados dos valores de acurácia de imputação para os dois softwares utilizados neste estudo. Os valores de acurácia de imputação são apresentados como a média de genótipos imputados corretamente para todos os cromossomos. O número de genótipos imputados corretamente é apenas uma das medidas utilizadas para avaliar a acurácia neste estudo. Pode-se verificar maior acurácia de imputação para o software fastPHASE em comparação com o Beagle para todos os painéis e cenários avaliados.

Na literatura vários estudos compararam a acurácia de imputação de genótipos utilizando vários softwares (PEI et al., 2008, NHO et al., 2011, SUN et al., 2012). Apesar dos resultados na literatura demonstrarem que o software Beagle apresentou melhores valores de acurácia em diferentes populações de bovinos, há alguns relatos e explicações para os resultados aqui relatados neste estudo. A acurácia depende de vários fatores entre eles podemos destacar a informação de LD, tamanho da amostra, relacionamento entre os indivíduos, metodologia, dentre outros fatores.

Hayes et al. (2011) reportaram que o software fastPHASE apresentou maiores valores de acurácia de imputação em comparação com o Beagle quando imputado do 50K para o genoma completo em ovinos. Segundo estes autores, uma explicação para este resultado seria que o fastPHASE foi desenvolvido para ser utilizado em situações em que a informação de LD na população é mais importante do que outros parâmetros como recombinação e mutação, utilizados no procedimento de imputação do software Beagle, capturados nos indivíduos em estudo. Estes parâmetros reduzem o espaço de busca do software, tendo como resultado um aumento na acurácia de imputação. De acordo com Browning e Browning (2007) o software Beagle apresenta maior desempenho quando uma população de referência grande é usada (por exemplo, mais de 1000 indivíduos). Por outro lado, o fastPHASE apresenta melhor desempenho quando a população de referência é menor (KHATKAR et al., 2012). No presente estudo utilizamos menor número de animais na população de referência (34 e 215 animais) do que na população de validação (779 e 598 animais).

Os resultados encontrados neste estudo corroboram com Khatkar et al. (2012). Neste estudo, foram avaliados vários cenários de imputação utilizando o cromossomo BTA20 em bovinos de leite, utilizando painéis de baixa densidade 3.000 (igualmente espaçado), 5.000, 10.000, 20.000 e 35.000 SNP e 3K (comercial) imputados para 50K. Nos cenários 27 animais foram usados no painel de referência e 2.700 animais na população de validação. Para todos

os cenários o fastPHASE foi superior ao software Beagle. A acurácia de imputação variou de 74 à 93%.

Outro estudo, avaliando o fastPHASE foi conduzido por Weigel et al. (2010) que investigaram a acurácia de imputação em animais raça Jersey. Estes autores mostraram que utilizando algoritmos baseados em informação de LD e na ausência da informação de pedigree, a acurácia de imputação pode ser baixa à modesta com valores de 91, 87, 79 e 73% para os painéis de baixa densidade com 2.942, 1.468, 741 e 336 imputados para 50K, respectivamente.

Pausch et al. (2013) relataram que para alcançar valores altos de acurácia de imputação, a escolha do software é o fator mais importante do que a seleção dos animais que serão utilizados na população de referência em situações em que o número de animais que serão utilizados na população de referência é pequeno. Dessa forma, os resultados de acurácia de imputação por meio do software fastPHASE podem ser considerados como boa para animais da raça Nelore.

Tabela 6 - Acurácia de imputação usando dois painéis de baixa e um de média densidade em dois cenários de imputação

Painel	Genótipos imputados corretamente			
	FastPHASE		Beagle	
	34 animais (%)	215 animais (%)	34 animais (%)	215 animais (%)
3K	68	70	54	52
6K	70	72	55	53
50K	79	81	56	54

3K: genótipos imputados a partir do 3K para 800K; 6K: genótipos imputados a partir do 6K para 800K; 50K: genótipos imputados a partir do 50K para 800K

Densidade dos painéis

A acurácia de imputação aumentou quando o número de SNP aumentou nos painéis (Tabela 6), sendo maior para o painel 50K seguido dos painéis 6K e 3K. A imputação a partir do 50K para 800K apresentou um ganho em acurácia de 16 e 13% em relação aos painéis 3K e 6K, respectivamente, para os dois cenários (34 touros e 215 animais na população de referência), utilizando o software fastPHASE. Neste caso, a utilização do 50K foi adequada e poderia ser incorporada nos estudos de imputação em bovinos da raça Nelore. O aumento do número de marcadores nos painéis pode melhorar a predição genômica e o poder nos estudos de associação genômica (PAUSCH et al., 2013).

Considerando os dois painéis de baixa densidade (3K e 6K) a escolha da genotipagem dos animais Nelore com o painel 3K é mais conveniente do que o painel 6K, tendo em vista que a densidade do painel é um fator importante a ser considerada devido ao custo da genotipagem do número de animais a serem genotipados na população. Comparando o ganho em acurácia para estes dois painéis observa-se um aumento de 2% em termos de ganho. O painel de 3K pode ser utilizado na pré-seleção de animais jovens e em larga escala na seleção genômica de fêmeas (DASSONNEVILLE et al., 2012).

ZHANG e DRUET (2010) observaram a eficiência da imputação com o aumento do número de SNP no painel. Neste estudo, os autores avaliaram a imputação com painéis contendo 384, 768, 1.536, 3.000 e 6.000 SNP imputados para 50K em bovinos de leite, por meio dos softwares DAGPHASE e CHROMIBD. A percentagem do erro de acurácia de imputação variou de 11,7% com 384 SNP para 2% com 6.000 SNP, utilizando DAGPHASE. Para CHROMIBD, a percentagem de erro variou de 10,7% com 384 SNP com diminuição para 3,3% com 6.000 SNP.

VanRaden et al. (2013), em um estudo semelhante com bovinos de leite estudaram a influência da densidade dos painéis sobre a acurácia de imputação a partir dos painéis de baixa densidade 3K, 6K e moderada densidade (50K) para o painel de alta densidade (800K). Os autores obtiveram aproximadamente 99, 96 e 94% de genótipos imputados corretamente do 50K, 6K e 3K para o 800K, respectivamente.

Número de animais no painel de referência

No cenário 1, com 34 animais (touro) na população de referência utilizando o software fastPHASE, a acurácia de imputação variou entre 68 à 79% entre os painéis de SNP. No cenário 2, com 215 animais (incluindo touros e progênie), a acurácia de imputação variou entre 70 à 81% para o fastPHASE (Tabela 6).

Verificou-se acréscimo de 2% em acurácia quando o número de animais aumentou na população de referência de 34 animais para 215 animais, utilizando o software fastPHASE e painel de 50K. Neste caso, esse aumento não representou ganho significativo em acurácia. Além disto, seria necessário a genotipagem de 191 animais a mais, representando um custo inviável de genotipagem. Estes resultados discordam com Pausch et al. (2013), que avaliaram quatro cenários de imputação utilizando 50, 100, 200 e 400 animais na população de referência. Observaram aumento da acurácia de 84, 92, 97 e 98%, para os cenários, respectivamente.

Imputação por cromossomos

Observa-se que houve variação na acurácia de imputação entre os 29 cromossomos nos dois cenários (Tabela 7). Discutiremos os resultados apenas para o software fastPHASE e cenário 1 (34 animais), tendo em vista que apresentaram os melhores resultados. Os valores de acurácia entre os cromossomos variaram entre 65% (BTA14) e 70% (BTA19) com o painel 3K, e 76% (BTA8) e 81% (BTA22), considerando o painel de 50K, utilizando apenas 34 animais no painel de referência. Os valores médios de LD para os cromossomos acima BTA8, BTA14, BTA19 e BTA22 foram: $0,20 \pm 0,25$, $0,18 \pm 0,24$ e $0,20 \pm 0,26$ e $0,16 \pm 0,23$ (Espigolan et al., 2013).

Os resultados observados neste estudo foram discordantes aos observados por Druet, Schooten e De Roos (2010) que avaliaram a influência da densidade dos marcadores nos cromossomos na acurácia de imputação com animais da raça Holandesa. Estes autores observaram que o cromossomo com maior densidade apresentou menores erros de imputação, e os piores resultados correspondiam ao cromossomo com a menor densidade de marcadores.

Segundo Zhang e Druet (2010) estudos de imputação são importantes para responder questões que direcionem qual densidade de marcadores de SNP e metodologia mais adequada para obter eficiência na predição dos valores genéticos genômicos na população. Ainda segundo estes autores, os fatores que influenciam a acurácia de imputação devem ser estudados antes de se definir um painel adequado, pois considerações econômicas e técnicas devem ser levadas em consideração para decidir a utilização de painel.

Tabela 7 – Acurácia de imputação de alta densidade SNP nos cromossomos para fastPHASE

Cromossomo	Cenários de Imputação					
	N = 34 animais			N = 215 animais		
	3K	6K	50K	3K	6K	50K
BTA1	0,69	0,70	0,79	0,71	0,72	0,82
BTA2	0,69	0,70	0,78	0,70	0,72	0,81
BTA3	0,68	0,70	0,79	0,70	0,72	0,81
BTA4	0,69	0,70	0,79	0,71	0,73	0,81
BTA5	0,68	0,69	0,78	0,70	0,71	0,80
BTA6	0,66	0,68	0,80	0,69	0,71	0,83
BTA7	0,68	0,70	0,79	0,71	0,72	0,81
BTA8	0,66	0,67	0,76	0,68	0,69	0,79
BTA9	0,66	0,68	0,77	0,69	0,70	0,80
BTA10	0,69	0,70	0,79	0,71	0,72	0,82
BTA11	0,69	0,71	0,80	0,72	0,73	0,82
BTA12	0,68	0,69	0,78	0,70	0,71	0,80
BTA13	0,67	0,69	0,77	0,69	0,71	0,80
BTA14	0,65	0,67	0,78	0,68	0,70	0,81
BTA15	0,69	0,71	0,79	0,72	0,73	0,82
BTA16	0,68	0,69	0,78	0,70	0,71	0,81
BTA17	0,68	0,69	0,79	0,70	0,72	0,81
BTA18	0,69	0,70	0,79	0,71	0,72	0,81
BTA19	0,70	0,71	0,79	0,72	0,73	0,81
BTA20	0,69	0,70	0,79	0,71	0,73	0,82
BTA21	0,69	0,70	0,79	0,71	0,73	0,81
BTA22	0,69	0,71	0,81	0,71	0,73	0,83
BTA23	0,67	0,69	0,78	0,69	0,71	0,81
BTA24	0,68	0,70	0,78	0,70	0,72	0,81
BTA25	0,68	0,70	0,78	0,70	0,72	0,80
BTA26	0,69	0,70	0,79	0,71	0,72	0,81
BTA27	0,68	0,69	0,78	0,69	0,71	0,79
BTA28	0,69	0,70	0,80	0,71	0,73	0,82
BTA29	0,69	0,71	0,79	0,71	0,73	0,81

4.4 Conclusões

(i) A acurácia de imputação foi maior nos dois cenários utilizando o software fastPHASE. Os painéis 6K e 50K apresentaram maiores valores de acurácia. Entretanto, os painéis 3K e 50K podem ser recomendados para serem utilizados em animais da raça Nelore.

(ii) O cenário com 34 touros no painel de referência apresentou maior valor de acurácia, sendo o cenário mais adequado em termos de ganho em acurácia e economia na genotipagem dos animais.

(iii) É possível imputar os genótipos dos marcadores SNP em bovinos da raça Nelore, entretanto, os valores de acurácia dependem da densidade dos painéis de marcadores SNP e do software empregado.

Agradecimentos

À Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e à Rede BifeQualis. Em especial, a Bill Taylor e o pessoal do Sistema de Computação do projeto “*Condor System*” da Universidade de Wisconsin, Madison, Estados Unidos, pelo auxílio e treinamento ao acesso do cluster de computadores, os quais se fizeram possíveis às análises desse estudo.

Referências

ALENCAR, M.M. Critérios de seleção em bovinos de corte no Brasil. In: Simpósio Nacional de Melhoramento Animal, 4, 2002. Campo Grande - MS. **Anais...** Campo Grande: SBMA, 2002.

BOICHARD, D.; CHUNG, H.; DASSONNEVILLE, R.; DAVID, X.; EGGEN, A.; FRITZ, S.; GIETZEN, K.J.; HAYES, B.J.; LAWLEY, C.YT.; SONSTEGARD, T.S.; VAN TASSELL, C.P.; VANRADEN, P.M.; VIAUD-MARTINEZ, K.A.; WIGGANS, G.R. Design of a bovine low-density SNP array optimized for imputation. **Plos One**, San Francisco, v. 7, n.3, p. 1-10, 2012.

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. **The American Journal of Human Genetics**, Cambridge, v. 84, n. 2, p. 210-223, 2009.

BROWNING, B.L.; YU, Z. Simultaneous genotype calling and haplotype phase inference improves genotype accuracy and reduces false positive associations for genome-wide association studies. **The American Journal of Human Genetics**, Cambridge, v. 85, n. 6, p. 847-861, 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. **The American Journal of Human Genetics**, Cambridge, v. 81, n. 5, p. 1084-1097, 2007.

CHEN, M.H.; HUANG, J.; CHEN, W.M.; LARSON, M.G.; FOX, C.S.; VASAM, R.S.; SESHADRI, S.; O'DONNELL, C.J.; YANG, Q. Using family-based imputation in genome-wide association studies with large complex pedigrees: the Framingham heart study. **Plos One**, San Francisco, v. 7, n. 12, p. 1-11, 2012.

DASSONNEVILLE, R.; FRITZ, S.; DUCROCQ, V.; BOICHARD, D. Short communication: Imputation performances of 3 low-density marker panels in beef and dairy cattle. **Journal of Dairy Science**, Lancaster, v. 95, n. 7, p. 4136-4140, 2012.

DRUET, T.; SCHROOTEN, C.; DE ROOS, A.P. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. **Journal of Dairy Science**, Lancaster, v. 93, n. 1, p. 5443-5454, 2010.

HABIER, D.; FERNANDO, R.L.; DEKKERS, J.C.M. Genomic selection using low-density SNP. **Genetics**, Bethesda, v. 182, p. 343-353, 2009.

HAO, K.; CHUDIN, E.; McELWEE, J.; SCHADT, E.E. Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. **BMC Genetics**, London, v. 10, n. 27, p. 1-10, 2009.

HAYES, B.J.; BOWMAN, P.J.; CHAMBERLAIN, A.J.; GODDARD, M.E. Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Lancaster, v. 92, n. 2, p. 433-443, 2009.

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**, Malden, v. 43, n. 1, p. 72-80, 2012.

HICKEY, J.M.; KINGHORN, B.P.; TIER, B.; WILSON, J.F.; DUNSTAN, N.; VAN DER WERF, J.H.J. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. **Genetics Selection Evolution**, London, v. 43, n. 12, p. 1-13, 2011.

HOWIE, B.N.; DONNELLY, P.; MARCHINI, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. **Plos Genetics**, San Francisco, v. 5, n. 6, p. 1-15, 2009.

KHATKAR, M.S.; MOSER, G.; HAYES, B.J.; RAADSMA, H.W. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. **BMC Genomics**, London, v. 13, n. 538, p. 1-12, 2012.

LI, Y.; WILLER, C.; SANNA, S.; ABECASIS, G. Genotype imputation. **Annual Review of Genomics and Human Genetics**, Palo Alto, v. 10, p. 387-406, 2009.

MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. **Nature Reviews Genetics**, London, v. 11, n. 7, p. 499-511, 2010.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.

MILLER, S. Genetic improvement of beef cattle through opportunities in genomics. **Revista Brasileira de Zootecnia**, Viçosa, v. 39, p. 247-255, 2010.

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. **Journal of Dairy Science**, Lancaster, v. 95, n. 2, p. 876-889, 2012.

PAUSCH, H.; AIGNER, B.; EMMERLING, R.; EDEL, C.; GOETZ, K.U.; FRIES, R. Imputation of high-density genotypes in the Fleckvieh cattle population. **Genetics Selection Evolution**, London, v. 45, n. 3, p. 1-10, 2013.

SARGOLZAEI, M.; CHESNAIS, J.P.; SCHENKEL, F.S. FImpute: an efficient imputation algorithm for dairy cattle populations. **Journal of Dairy Science**, Lancaster, v. 94, n. 421, 2011.

SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **The American Journal of Human Genetics**, Cambridge, v. 78, n. 4, p. 629-644, 2006.

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. **Journal of Dairy Science**, Lancaster, v. 95, n. 9, p. 5403-5411, 2012.

SUN, C.; WU, X.L.; WEIGEL, K.A.; ROSA, G.J.M.; BAUCK, S.; WOODWARD, B.W.; SCHNABEL, R.D.; TAYLOR J.F.; GIANOLA, D. An ensemble-based approach to imputation of moderate-density genotypes for genomic selection with application to Angus cattle. **Genetics Research**, Cambridge, v. 94, n. 3, p. 133-150, 2012.

VANRADEN, P.M.; O'CONNELL, J.R.; WIGGANS, G.R.; WEIGEL, K.A. Genomic evaluations with many more genotypes. **Genetics Selection Evolution**, London, v. 43, n. 10, p. 1-11, 2011.

VANRADEN, P.M.; NULL, D.J.; SARGOLZAEI, M.; WIGGANS, G.R.; TOOKER, M.E.; COLE, J.B.; SONSTEGARD, T.S.; CONNOR, E.E.; WINTERS, M.; VAN KAAM, J.B.C.H.M.; VALENTINI, A.; VAN DOORMAAL, B.J.; FAUST, M.A.; DOAK, G.A. Genomic imputation and evaluation using high-density Holstein genotypes. **Journal of Dairy Science**, Lancaster, v. 96, n.1, p. 668-678, 2013.

WEIGEL, K.A.; VAN TASSELL, C.P.; O'CONNELL, J.R.; VANRADEN, P.M.; WIGGANS, G.R. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. **Journal of Dairy Science**, Lancaster, v. 93, n. 5, p. 2229-2238, 2010.

ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **Journal of Dairy Science**, Lancaster, v. 93, n. 11, p.5487-5494, 2010.

ZHANG, B.; ZHI, D.; ZHANG, K.; GAO, G.; LIMDI, N.N.; LIU, N. Practical consideration of genotype imputation: Sample size, window size, reference choice, and untyped rate. **Stat Interface**, Alexandria, v. 4, n. 3, p. 339-352, 2011.