

Pier Kenji Rauschkolb Katsuda Ito

Trabalho de Conclusão de Curso

Araçatuba

2016



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Araçatuba

Trabalho de Conclusão de Curso

Trabalho de Conclusão de Curso de Graduação apresentado à Faculdade de Medicina Veterinária da Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de Araçatuba, para obtenção do grau de Médico Veterinário.

Aluno: Pier Kenji Rauschkolb Katsuda Ito

Supervisor: José Fernando Garcia

Araçatuba

2016

ENCAMINHAMENTO

Encaminhamos o presente Trabalho de Conclusão de Curso, para que a Comissão de Estágios tome as providências cabíveis.

Estagiário

Supervisor

ARAÇATUBA
Novembro de 2016

Sumário

Relatório de Estágio Supervisionado	6
1. Introdução	8
2. Descrição geral do local de estágio.....	8
3. Descrição das atividades do estágio	10
3.1. Amostras e Recepção	10
3.2. Extração de DNA e Genotipagem	11
3.3. Controle de qualidade e envio de resultados	12
3.4. Análises realizadas	13
3.4.1. Testes de paternidade.....	13
3.4.2. Reavaliação de parâmetros de qualidade	15
3.4.3. Acurácia de imputação	16
3.4.4. Imputação da raça em amostras da espécie bovina	18
4. Recomendações	19
5. Conclusões.....	19
6. Referências	19
Assinaturas de seleção em dados genotípicos de SNP Chip do Projeto HapMap Bovino	21
Resumo	23
7. Introdução	24
8. Material e Métodos	28
8.1. Amostras	28
8.2. Controle de Qualidade (QC).....	28
8.3. Análise de assinaturas de seleção	29
8.4. Validação	30
9. Resultados.....	31

10. Discussão.....37

11. Conclusão38

12. Referências39

Pier Kenji Rauschkolb Katsuda Ito

Relatório de Estágio Supervisionado

Araçatuba

2016



Relatório de Estágio

Relatório de Estágio Supervisionado, como parte do Trabalho de Conclusão de Curso apresentado à Faculdade de Medicina Veterinária da Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de Araçatuba, para obtenção do grau de Médico Veterinário.

Aluno: Pier Kenji Rauschkolb Katsuda Ito

Supervisor: José Fernando Garcia

Araçatuba

2016

1. Introdução

Adequação às habilidades pessoais, afinidade ao tipo de atividade, bem como aplicabilidade das competências adquiridas durante o período de graduação, compõem motivadores que impulsionaram a escolha pela Clínica Veterinária como área de atuação, com especialização em Genética e Melhoramento Animal. Os objetivos de tal escolha foram aprofundar conhecimentos previamente adquiridos e participar ativamente da pesquisa e desenvolvimento de novos produtos que encontrem aplicabilidade direta em referida área.

Os motivadores pela escolha da empresa Deoxi Biotecnologia Ltda para realização de todo o período de estágio encontram sinergismo naqueles descritos acima, bem como incluem a oportunidade de colocação no mercado de trabalho, desenvolvimento pessoal e profissional e adequação à área de especialização. Somam-se ainda aos objetivos descritos anteriormente, a definição, estabelecimento e expansão das competências necessárias ao desenvolvimento da capacidade produtiva no campo de trabalho.

2. Descrição geral do local de estágio

A Deoxi Biotecnologia Ltda, fundada em 2008 na cidade de Araçatuba/SP é uma empresa privada que nasceu a partir desenvolvimento de um projeto do programa de Pesquisa Inovativa em Pequenas Empresas - PIPE financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, em atendimento à necessidade do mercado nas áreas de genética e biologia molecular com uso de tecnologia de ponta para desenvolvimento e aplicabilidade de novos produtos. A Deoxi presta serviços laboratoriais na área de biologia molecular por meio da genotipagem com uso de SNP Chips, que avaliam marcadores moleculares do tipo polimorfismo de sítio único (SNP), e do sequenciamento de nova geração (NGS).

Em abril de 2016, a Deoxi foi adquirida pela multinacional Neogen Corp., maior empresa mundial quanto à prestação de serviços nas áreas de segurança alimentar, segurança animal, genômica, ciências da vida e toxicologia, processando mais de 10.000 amostras por dia. O objetivo da aquisição foi expandir a área de genômica no

mercado brasileiro, oferecendo produtos destinados principalmente à agricultura e pecuária.

A Deoxi em sua configuração atual, conta com 13 funcionários, distribuídos nos setores de gerência (2), comercial (3), recepção de amostras (2), financeira (1), laboratorial (4) e análise (1). A empresa possui estreita relação com os meios acadêmico e científico, mantendo vínculo estreito de colaboração com o Laboratório de Bioquímica e Biologia Molecular Animal (LBBMA) da UNESP/Araçatuba e com o Centro de Estudos do Genoma Humano da USP/São Paulo. Nos setores de gerência, comercial, financeiro e recepção de amostras a empresa executa seus trabalhos por meio de sistema informatizado de controle, mantendo 3 computadores desktop e 5 notebooks. No setor laboratorial, a empresa conta com todos os equipamentos necessários à execução de análises de biologia molecular, que incluem 4 termocicladores, 1 centrífuga de placas, 1 centrífuga de tubos, 1 sistema automatizado para extração de DNA (*Biosprint*®), vidrarias para preparação de reagentes, 1 capela, 1 *Nanodrop*®, 1 *Qubit*™, 3 cubas de eletroforese, 2 agitadores de tubos Vortex, 4 agitadores de placas, 4 seladoras térmicas para placas, 1 *iScan*®, 1 *MySeq*®, bem como todos os equipamentos que acompanham estes últimos, como computadores de alta capacidade, 2 fornos de hibridação, 3 *Heatblocks*® e 1 cuba para corar os chips e blocos de hibridação. O laboratório conta ainda com 8 freezers e 4 geladeiras para manutenção de reagentes e amostras. No setor de análise, os dados gerados são armazenados em 1 servidor NAS e analisados em 1 *workstation* de alta capacidade.

O layout da empresa é distribuído em 1 sala onde funcionam os setores de recepção de amostras e financeiro, 1 sala da gerência e 1 sala do setor comercial. O setor laboratorial é dividido em 4 salas onde ocorrem a extração de DNA, genotipagem e sequenciamento, avaliação da qualidade do DNA extraído e armazenagem de reagentes, bem como das amostras recebidas. O setor de análise é mantido em ambiente climatizado junto aos fornos de hibridação, equipamento de leitura dos chips, sequenciador e termocicladores.

3. Descrição das atividades do estágio

Durante a realização do estágio, as atribuições seguiram fluxo distinto àquele disposto nos tópicos a seguir, já que inicialmente ocupou-se função imediata no processo de avaliação da qualidade, análise e organização dos dados resultantes de todo processo de genotipagem. Porém, para manter a coerência do texto e compreensão do leitor, os tópicos descrevem as atividades desenvolvidas segundo o fluxo que engloba a rotina da empresa (**Figura 1**), seguindo a entrada das amostras, seu processamento e obtenção do resultado final destinado ao cliente.

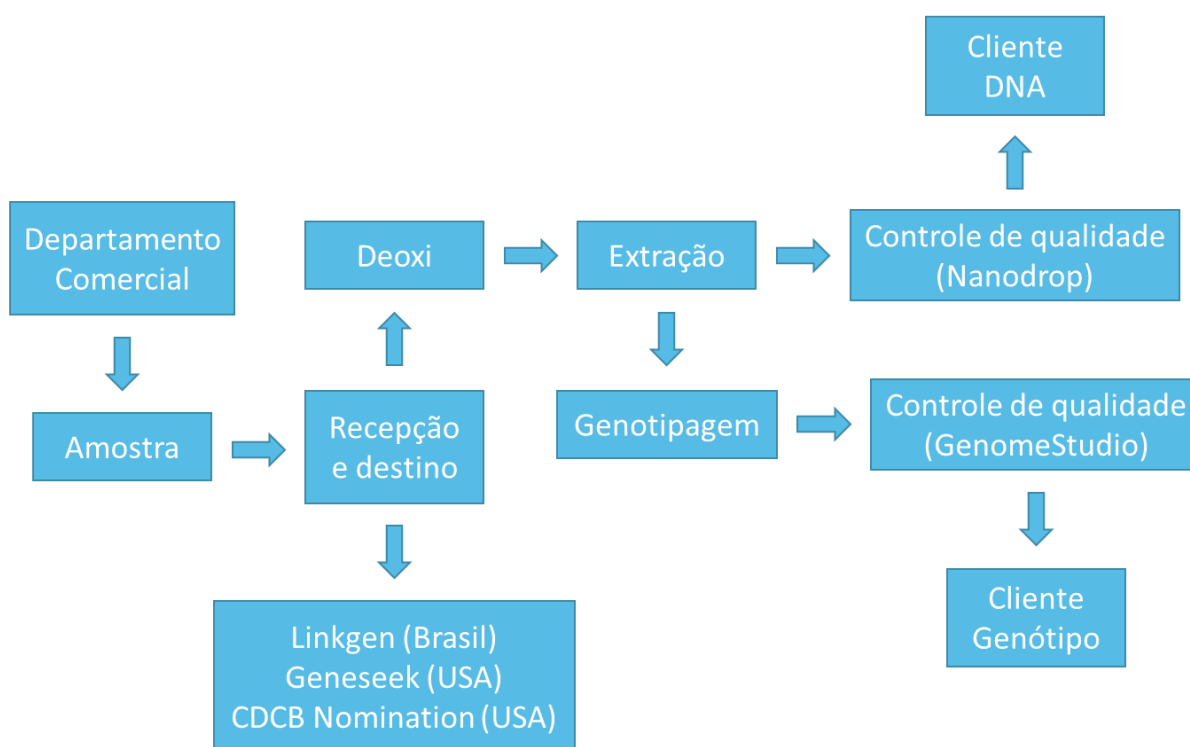


Figura 1 - Fluxograma de Trabalho

3.1 Amostras e Recepção

Apesar de oferecer o serviço de genotipagem em diferentes espécies animais e vegetais, mais de 95% dos serviços prestados, são efetuados são destinados à espécie bovina. Dentre as demais espécies atendidas pelo serviço, as que possuem frequência constante na rotina laboratorial são a humana, suína e caprina. Segundo orientação dada aos clientes, nas espécies domésticas recomenda-se o envio de amostras de pêlo (bulbo capilar), sangue, DNA previamente extraído e sêmem, na

espécie humana é feita a coleta por meio de swab oral ou sangue, porém independente de espécie são processadas amostras de tecido de plantas, material de aborto ou mesmo amostras de carne advindas de abatedouros frigoríficos. Geralmente as amostras chegam até a empresa por meio do serviço postal dos correios, após orientação de envio dada ao cliente por um dos representantes comerciais. As amostras chegam acondicionadas em cartões de coleta de pêlos e sangue (FTA Card), palhetas de sêmem, tubos de sangue, placas de 96 poços com DNA extraído, envelopes, dentre outros.

A recepção é feita através da conferência das informações fornecidas pelo cliente com àquelas constantes na identificação individual de cada amostra seguida do cadastramento da amostra por meio do uso de código de barras individual quando acondicionada em outro meio que não o de placas. Quando em placas, cada placa recebe um código de barras e cada amostra é identificada pela concatenação deste código com a posição da amostra em cada poço. Após receber identificação única, é feita a ordem de serviço interna e posterior cadastramento no sistema de gerenciamento laboratorial (LIMS), definindo o destino da mesma. Quanto ao destino, as amostras podem seguir para processamento no laboratório da Deoxi ou enviada para processamento no laboratório central da Neogen em Lincoln/NE nos Estados Unidos ou para laboratórios de terceiros como a Linkgen em São Paulo/SP e LGC/EUA.

3.2 Extração de DNA e Genotipagem

Haja visto o elevado custo por amostra (R\$50,00 à R\$3.100,00), bem como a constante preocupação em manter a qualidade dos resultados gerados, a sensibilidade dos testes realizados e possibilidade de verificação por parte do consumidor final dos mesmos parâmetros e critérios utilizados para definir os padrões de qualidade, a participação nos processos de extração e genotipagem foi permitida apenas na condição de observador. O processo de extração é realizado em equipamento robotizado (*Biosprint*®) com uso de kit de extração desenvolvido pela Neogen, seguindo protocolo fornecido pela mesma. A genotipagem consiste na determinação dos alelos de cada locus cromossômico definido pela posição da forma de variação mais frequente distribuída ao longo do genoma, o polimorfismo de sítio único (SNP). O mesmo é determinado por meio da extensão de uma única base ligada

à fluoróforo específico, cuja excitabilidade produz um sinal fluorescente distinto segundo o canal de cor verde ou vermelho (**Figura 2**).

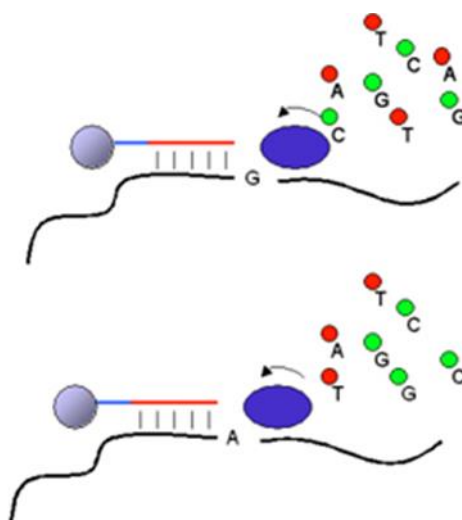


Figura 2 – Extensão de uma única base em região genômica específica, contendo fluoróforo marcado por fluorescência.

Cada microarranjo de DNA consiste em conjunto de sondas que interrogam o genoma alvo quanto aos alelos presentes em cada posição. A leitura dos chips é feita no leitor *iScan*®, que suporta diferentes tipos de chip, estes destinados a diversas espécies e com número variável de sondas. Os ensaios GGPIndicusHD™, BovineSNP50®, Z-Chip, CottonSNP63K, GoatSNP50 e CytoSNP-850K v1.1 foram realizados durante o período de estágio e destinados a distintas espécies.

3.3 Controle de qualidade e envio de resultados

Após a genotipagem, é avaliada em software próprio do equipamento de leitura dos chips (*GenomeStudio*), a taxa de determinação de genótipos (*Call Rate*) por indivíduo. O *Call Rate* é a relação entre o número de genótipos que foram determinados pelo teste e o número total de sondas avaliadas pelo mesmo. Excluindo-se algumas exceções como amostras cujo genoma é poliploide (plantas) ou ensaios em fase inicial de desenvolvimento cujas sondas não abrangem ampla população de subspecies, a amostra que apresentar relação inferior a 90% tem os resultados excluídos e o ensaio refeito.

O conjunto total de amostras de cada cliente é também verificado quanto à duplicidade de genótipos e quanto o índice de heterozigosidade individual. No primeiro avalia-se a identidade genotípica entre duas amostras, isto é, a proporção de alelos idênticos partilhados por estado entre as mesmas. Esta mensuração varia entre 0 e 1, onde valores superiores a 0.8 sugerem possível duplicidade de genótipos, amostras resultantes de clonagem, gêmeos univitelinos ou indivíduos de população com elevado índice de endogamia, cabendo maiores explicações por parte do cliente ou nova execução do ensaio quando a duplicidade foi promovida por erros laboratoriais como erros de pipetagem durante a transferência entre placas.

3.4 Análises realizadas

Durante o período de estágio foram desenvolvidas e aplicadas metodologias de análise dos resultados advindos da genotipagem com objetivo de atender solicitação do consumidor final do serviço ou necessidades internas da empresa a título de pesquisa e desenvolvimento de novos produtos.

3.4.1 Testes de paternidade

Consiste na avaliação da proporção de alelos partilhados por descendência segundo informação genotípica dos pares de indivíduos testados. Tal proporção, denominada PI HAT, é dada pela relação entre a probabilidade de duas amostras partilharem dois alelos idênticos por descendência ($IBD=2$) somada à metade da probabilidade de partilharem ao menos um alelo idêntico por descendência ($IBD=1$) (PURCELL et al., 2007). Se os alelos avaliados seguem um mecanismo de segregação mendeliana, então espera-se que a proporção de alelos partilhados entre pais e filhos seja de ao menos 50%, definindo assim, o grau de parentesco ou confirmação de paternidade quando fornecida a informação a priori dos possíveis pais de determinado indivíduo. Em rebanhos multitueros inseridos em programas de melhoramento genético, é de fundamental importância determinar com precisão a matriz de parentesco, já que esta desempenha efeito importante sob as estimativas do valor genético dos animais avaliados.

A metodologia desenvolvida consiste em execução de pequenos softwares desenvolvidos *in house*, bem como do software PLINK v1.9 (CHANG et al., 2015) que visam organizar, compilar, validar, testar e normalizar os dados genotípicos

produzidos por diferentes plataformas de genotipagem. Estes são advindos dos serviços de genotipagem de terceiros como a LGC (painel desenvolvido em colaboração com a equipe do LBBMA/UNESP), da unidade americana da Neogen, a Geneseek (painéis *TotalIndicus & Seeksire™*, *GGPHD™*, *GGPLD™* e *GGPIndicusHD™*), ou fornecidos pelo próprio cliente (painéis Illumina *BovineHD®*, *BovineSNP50®* e *BovineLD®* e *Z-Chip*) quando o mesmo já possui conjunto de animais genotipados em painéis compatíveis com o teste. A compatibilidade é dada pelo número mínimo de marcadores partilhados entre diferentes painéis (200 SNP), número mínimo necessário para determinar, com margem probabilística significativa, (99.99%) a paternidade entre produtos e touros. A **Figura 3** exemplifica o resultado gerado pela metodologia de análise.

FID1	IID1	FID2	IID2	PI_HAT	DST	IBS0	IBS1	IBS2
C	1/15_AJ	S	11/13_AJ	0.5	0.853801	0	50	121
C	1000/15_AJ	S	9/12_AJ	0.5	0.813953	1	62	109
C	1003/15_AJ	S	652312AJ	0.5	0.784091	1	74	101
C	101/15_AJ	S	88/12_AJ	0.5	0.831429	1	57	117
C	1011/15_AJ	S	127/12_AJ	0.5	0.811765	2	60	108
C	1014/15_AJ	S	829812AJ	0.5118	0.877907	2	38	132
C	1015/15_AJ	S	487512AJ	0.5	0.835227	1	56	119
C	1016/15_AJ	S	807212AJ	0.5	0.819444	1	63	116
C	1017/15_AJ	S	9/12_AJ	0.5	0.819767	1	60	111
C	1017/15_AJ	S	36/12_AJ	0.5	0.807018	2	62	107
C	1020/15_AJ	S	92/12_AJ	0.5	0.762857	4	75	96
C	1021/15_AJ	S	108/12_AJ	0.5078	0.865169	0	48	130
C	1025/15_AJ	S	115/12_AJ	0.5779	0.884393	0	40	133
C	1029/15_AJ	S	232/12_AJ	0.5	0.78869	2	67	99
C	103/15_AJ	S	116/12_AJ	0.5	0.819209	1	62	114

Figura 3 - Exemplo de resultado do teste de paternidade.

Incrementos no número de marcadores avaliados incorrem no aumento do poder do teste, necessário em casos onde a paternidade permanecer incógnita. Estes casos ocorrem em rebanhos cujo índice de endogamia é elevado, pois promove incrementos da proporção de alelos partilhados entre indivíduos por descendência indireta como por exemplos nos casos em que o suposto pai é irmão do pai verdadeiro quando o produto é advindo de linhagem com descendência direta de ambos. Nesse

caso, a probabilidade de número reduzido de alelos serem partilhados apenas por estado é maior, implicando no aumento dos valores de PI HAT.

Durante o período de estágio foram realizados mais de 6000 testes de paternidade segundo solicitação de diferentes fazendas. Destes, foram determinadas em média, a paternidade de 85% dos produtos avaliados. Os casos inconclusivos ocorreram devido à erros de identificação da amostra por parte do cliente, número insuficiente de touros amostrados, erros ou falhas de genotipagem e número reduzido de marcadores.

3.4.2 Reavaliação de parâmetros de qualidade

Atendendo necessidade interna da empresa, foram realizadas análises para melhor estabelecer parâmetros capazes de determinar o sucesso de genotipagem das amostras destinadas ao ensaio *Infinium CytoSNP-850K v1.1 BeadChip*. O mesmo, avalia em torno de 850 mil marcadores distribuídos ao longo de todo genoma humano perfazendo cobertura ampla de 3.262 genes de conhecida relevância citogenética em aplicações como identificação de enfermidades de ordem genética e oncologia. A tecnologia do ensaio utiliza sondas longas (50 nucleotídeos) com elevada especificidade do SNP alvo e elevada redundância (15x) para determinar a chamada de variações no número de cópias (CNV). Tal redundância também auxilia a reduzir a relação ruído/sinal produzido para cada SNP, permitindo determinação precisa desse tipo de variação. Apesar da sensibilidade do teste, o sucesso do mesmo é dependente da intensidade gerada nos canais verde e vermelho produzida pela excitação dos respectivos fluoróforos. A taxa de determinação de genótipos (CR), definida por estas intensidades, estima de maneira robusta a performance geral do teste. Porém, ao avaliar número significativo de amostras, verificou-se que apenas esse parâmetro é insuficiente para determinar o sucesso da genotipagem, pois amostras com elevado *Call Rate* apresentaram elevado nível de ruído quando comparadas aquelas com CR dentro da mesma faixa de valores. O ruído então foi estimado a partir do desvio médio apresentado pelos canais de intensidade ($\Delta \text{LogR Ratio}$) e ilustrado a partir do gráfico de dispersão entre este e o CR de cada amostra (**Figura 1**).

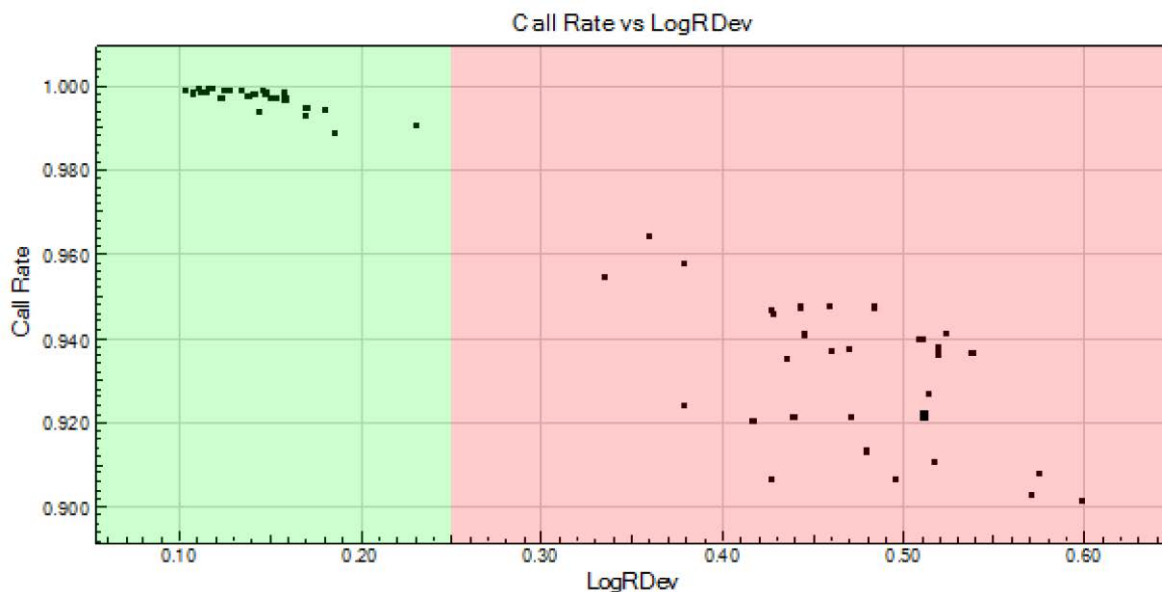


Figura 4 - Dispersão das amostras segundo a taxa de determinação de genótipos e o desvio da intensidade.

Na figura, amostras com alta taxa de ruído ou maior desvio médio da intensidade (região vermelha do gráfico), apresentaram inconsistência na determinação de regiões genômicas que albergam CNV determinando o fracasso do teste. O limiar adotado para determinar sucesso do ensaio foi de 0,25 desvios (região verde do gráfico) e passou a compor a rotina de análise para verificação e envio de resultados na empresa.

3.4.3 Acurácia de imputação

Imputação genotípica consiste na determinação de genótipos desconhecidos a partir da informação daqueles que são conhecidos em determinado indivíduo. A partir dessa definição, foi efetuada análise por solicitação do cliente para avaliar a diferença entre a acurácia de imputação entre dois painéis de baixa densidade destinados a subgrupos raciais distintos (Z-Chip/Zebuinos - 27.533 SNP e GGPLD/Taurinos - 41.273 SNP), para um painel de densidade de marcadores mais elevada (BovineSNP50 - 44.609 SNP), nas raças Braford e Hereford, ambas taurinas. Para tanto foram utilizados os genótipos de 899 animais, genotipados no painel alvo (BovineSNP50) e mascarados os genótipos de 10 indivíduos ao acaso a cada interação, efetuando-se a imputação de 100 indivíduos para cada painel de baixa densidade. Mascarar o genótipo corresponde a definir como valor desconhecido os

genótipos do painel de alta densidade até que restem apenas os marcadores determinados nos painéis de baixa densidade.

Foi desenvolvido código próprio em *bash script* no Unix para efetuar cada interação, efetuando as recodificações dos genótipos no software PLINK v1.9 (CHANG et al., 2015) e formatação dos dados para etapa de imputação no software FImpute v2.2 (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014). Este utiliza estratégia de janelas deslizantes com sobreposição para a relação de semelhança entre os haplótipos dos indivíduos alvo (genótipos desconhecidos) e de referência (genótipos conhecidos) (**Figura 5**).

	1 BP	14 BP	31 BP	79BP	112BP	121BP	142BP
7 SNP	A	A	B	A	B	B	A
7 SNP	A	A	B	A	B	B	A
7 SNP	A	A	B	A	B	B	A
7 SNP	A	A	B	A	B	B	A
7 SNP	A	?	B	A	B	B	A
7 SNP	B	B	A	B	B	A	A

Figura 5 - Exemplo de janela deslizante avaliada pelo software FImpute.

Após a imputação dos indivíduos em cada painel, avaliou-se a concordância entre genótipos reais e àqueles imputados, sendo definida a acurácia de imputação como a proporção de genótipos imputados idênticos aos reais, dado o número total de genótipos no painel de referência (de maior densidade). A **Tabela 1** sumariza esses resultados.

Tabela 1 - Acurácia de imputação segundo a raça e o painel utilizado.

Raça_Painel	Min.	1st Quartil	Mediana	Média	3rd Quartil	Maximo
BRAFORD_ZCHIP	0.8482	0.8944	0.922	0.9197	0.9538	0.9678
HERORD_ZCHIP	0.8645	0.9199	0.9414	0.9352	0.9587	0.9662
BRAFORD_GGPLD	0.902	0.9409	0.9484	0.947	0.9624	0.9731
HEREFORD_GGPLD	0.9494	0.9619	0.9687	0.97	0.9772	0.9889

A acurácia do painel Z-Chip, apesar de inferior comparada ao painel GGPLD, oferece relação custo benefício muito superior ao último, haja visto o custo relativo do chip dada a densidade de cada painel, levando o cliente a optar pelo painel de menor densidade e conseqüentemente de menor custo.

3.4.4 Imputação da raça em amostras da espécie bovina

Estas análises foram efetuadas para atender solicitação da Neogen quanto a organização de todos os animais genotipados pela Deoxi desde o início de suas atividades, totalizando cerca de 22.000 animais da espécie bovina genotipados em quatro painéis (GGPIndicusHD, BovineHD, BovineSNP50 e Z-Chip). Como a informação de raça das amostras é em sua grande maioria omitida pelo cliente, mais de 17.000 genótipos eram desprovidos de tal informação. Assim, foi desenvolvida metodologia de análise capaz de identificar a raça de determinado indivíduo a partir da informação genotípica do mesmo. A metodologia na análise de componentes principais (PCA) efetuada na matriz de parentesco genômico cuja variância foi estandardizada, isto é, efetuar a redução da dimensionalidade do conjunto de dados a partir da decomposição geométrica das distâncias entre indivíduos dada pela matriz de parentesco (CHANG et al., 2015). Na **Figura 6**, observa-se a distribuição dos indivíduos segundo coordenadas definidas pelo primeiro (PC1) e segundo (PC2) componentes principais, definindo os grupamentos raciais.

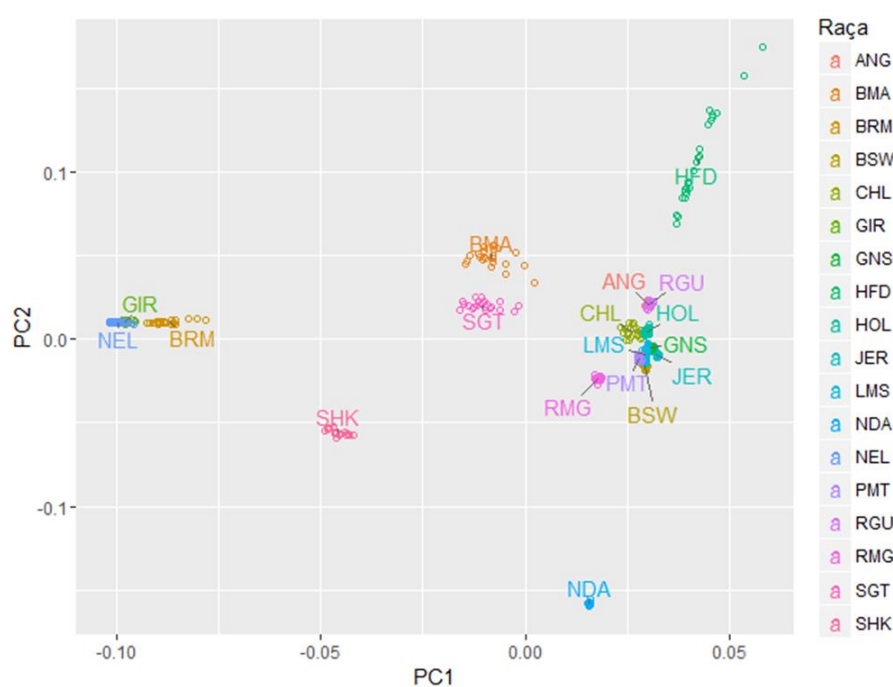


Figura 6 - Resultado da análise de componentes principais (PCA).

A análise de componentes principais foi realizada no software PLINK v1.9 (CHANG et al., 2015) e a determinação da raça foi efetuada a partir do agrupamento das amostras pelo método de particionamento em torno de medoids (REYNOLDS et

al., 2006), alocando indivíduos um a um a cada grupo racial. Esta última etapa foi realizada com auxílio do pacote *cluster* (MAECHLER et al., 2016) no R (IHAKA; GENTLEMAN, 1996).

4. Recomendações

A Deoxi Biotecnologia oferece oportunidade para àqueles que desejam ingressar em um campo de trabalho dinâmico, inovador e que mantém laços estreitos com as áreas de pesquisa e desenvolvimento, bem como o meio acadêmico. Para tanto, o estagiário deve ser assíduo, atento e focado nas tarefas designadas, pois erros cometidos durante qualquer das etapas do fluxo de trabalho, acarretam em perdas financeiras expressivas e incredulidade do cliente quanto à qualidade dos serviços prestados pela empresa.

5. Conclusões

O período de estágio integralizado na Deoxi possibilitou exposição das deficiências pessoais e profissionais, as quais serão supridas com a aquisição de experiência no campo de trabalho. Demonstrou ainda, competências e habilidades que antes permaneciam ocultas aos olhos deste autor e que passaram a ser valorizadas e exploradas com o intuito de ampliar o potencial de cada uma delas. A realização do estágio cumpriu com todos os objetivos propostos culminando no reconhecimento do esforço empreendido através da colocação no mercado de trabalho.

6. Referências

CHANG, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. **GigaScience**, v. 4, n. 1, p. 7, 25 dez. 2015.

IHAKA, R.; GENTLEMAN, R. R: A language for data analysis and graphics. **Journal of Computational and Graphical Statistics**, v. 5, n. 3, p. 299–314, set. 1996.

MAECHLER, M. et al. **cluster: Cluster Analysis Basics and Extensions**, 2016.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, set. 2007.

REYNOLDS, A. P. et al. Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. **Journal of Mathematical Modelling and Algorithms**, v. 5, n. 4, p. 475–504, 3 dez. 2006.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, v. 15, n. 1, p. 478, jan. 2014.

Pier Kenji Rauschkolb Katsuda Ito

**Assinaturas de seleção em dados genotípicos
de SNP Chip do Projeto HapMap Bovino**

Araçatuba

2016

Assinaturas de seleção em dados genotípicos de SNP Chip do Projeto HapMap Bovino

Trabalho Científico, como parte do Trabalho de Conclusão de Curso apresentado à Faculdade de Medicina Veterinária da Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de Araçatuba, para obtenção do grau de Médico Veterinário.

Aluno: Pier Kenji Rauschkolb Katsuda Ito

Supervisor: José Fernando Garcia

Araçatuba

2016

Assinaturas de seleção em dados genotípicos de SNP Chip do Projeto HapMap Bovino

Pier Kenji Rauschkolb Katsuda Ito

Resumo

É manifesta a necessidade de pesquisas que levem ao desenvolvimento de tecnologias que possam prover acréscimos eficazes na produtividade dos rebanhos. Iniciativas como o sequenciamento do genoma bovino e o projeto HapMap, permitiram detectar as “pegadas” deixadas no genoma pelos mecanismos de seleção, que atuam alterando as frequências alélicas deixando uma marca na estrutura de variação genotípica. Este tipo de estudo é comumente chamado de Assinatura de Seleção. Nos bovinos, devido à forte seleção artificial, geralmente a localização destas assinaturas estará associada à uma região controladora de características de produção. Do exposto, o presente trabalho adaptou e validou metodologia para procurar regiões indicativas de assinaturas de seleção nos genomas de diferentes raças bovinas. Para tanto, foram utilizados genótipos de 137 animais das raças Angus, Holandesa, Hereford e Pardo Suíço obtidos por meio do ensaio Illumina® BovineHD Genotyping BeadChip, interrogando 786.799 marcadores distribuídos no genoma. A metodologia consiste em adaptação do método ZHp, que avalia o decaimento da heterozigosidade local ao longo do genoma indicando possível assinatura de seleção. Para validar o método, os achados foram contrastados com assinaturas nos genes MC1R e KIT, conhecidas da literatura. A metodologia proposta detectou evidências de seleção nas regiões de validação, bem como sinalizou assinaturas de seleção em diferentes regiões cromossômicas. Em suma, o método ZHp quando aplicado a dados genotípicos de alta densidade na espécie bovina, obtidos pelo uso de SNP chips comercialmente disponíveis, é válido para a detecção de assinaturas de seleção recente conforme hipotetizado.

Palavras-chave: assinatura de seleção, melhoramento genético, SNP, genótipo, bovino.

7. Introdução

No 1º trimestre de 2016, 7,9 milhões de cabeças de bovinos foram abatidas no Brasil, número inferior ao 1º trimestre do ano anterior, totalizando peso acumulado de carcaças de 1,795 milhão de toneladas. Do total produzido, 287 mil toneladas foram exportadas como carne in natura sendo o faturamento da exportação de 1,1 bilhão de dólares, tendo Hong Kong como principal destino das exportações (IBGE, 2016).

As projeções de 2013 a 2023 mostram que o setor deverá crescer 2,0% ao ano, totalizando crescimento de 22,5% para o período. Mostram ainda a preferência nacional pela carne bovina, com aumento no consumo de 3,6% ao ano, o que implica em aumento de 42,8% nos próximos 10 anos. A taxa de crescimento anual para as exportações de carne bovina deve manter-se em 2,5%, caso se consolidem os mercados atendidos atualmente (MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO, 2013).

Haja visto que a bovinocultura atua como ponderador do cenário econômico nacional e mundial, é manifesta a necessidade de pesquisas que levem ao desenvolvimento de tecnologias capazes de atender a demanda por incrementos eficazes na produtividade dos rebanhos.

O papel histórico, socioeconômico e cultural desempenhado pela bovinocultura, deu origem a iniciativas como o sequenciamento do genoma bovino (ELSIK et al., 2009), que forneceu recursos para identificar e compreender a estrutura genética dessa espécie e permitir a elucidação dos mecanismos biológicos que atuam na expressão fenotípica de características de importância econômica. Esse conhecimento deu origem a uma série de ferramentas capazes de acelerar o melhoramento genético dos rebanhos, de maneira a atender à necessidade das indústrias da carne e do leite em aumentar sua produtividade.

Uma destas ferramentas baseia-se nos marcadores genéticos do tipo polimorfismo de base única (SNP do termo em inglês *Single-nucleotide Polymorphism*), a forma mais frequente de variação ao longo do genoma. Em trabalho seminal, ao interrogar simultaneamente 37.470 SNP, foi possível avaliar a estrutura genética e os processos de domesticação e seleção que deram origem às diferentes raças bovinas (GIBBS et al., 2009). Com a popularização destes ensaios teve início a era genômica nos bovinos. A aplicação desta tecnologia deu origem a novos estudos

que permitiram reavaliar a maneira como os processos de domesticação e adaptação desta espécie estão correlacionados à história humana (AJMONE-MARSAN et al., 2010).

Posteriormente, o advento dos testes de SNP de alta densidade, possibilitou a detecção e o mapeamento fino de variações na sequência de DNA que afetam características fenotípicas complexas (MATUKUMALLI et al., 2009). Esta tecnologia possibilitou a popularização dos estudos de “assinaturas de seleção”. Nestes, são detectadas as “pegadas” deixadas no genoma pelos mecanismos de seleção. Tais mecanismos atuam alterando as frequências alélicas em resposta a uma pressão externa, deixando uma marca no padrão das variações genotípicas (OLEKSYK; SMITH; O'BRIEN, 2010). Um destes mecanismos é exemplificado na **Figura 7**, em que um alelo vantajoso (vermelho) aumenta em frequência quando positivamente selecionado. Os alelos próximos a este (marrom) acabam pegando “carona” e são segregados ao longo das gerações juntos com os primeiros, efeito chamado de “carona genética”, do inglês *genetic hitchhiking*. A fixação de determinada variante genética, em região específica do genoma, provocada pelo aumento da frequência desta variante numa dada população ao longo da evolução (ou da seleção artificial), denomina-se assinatura de seleção (FAY; WU, 2000).

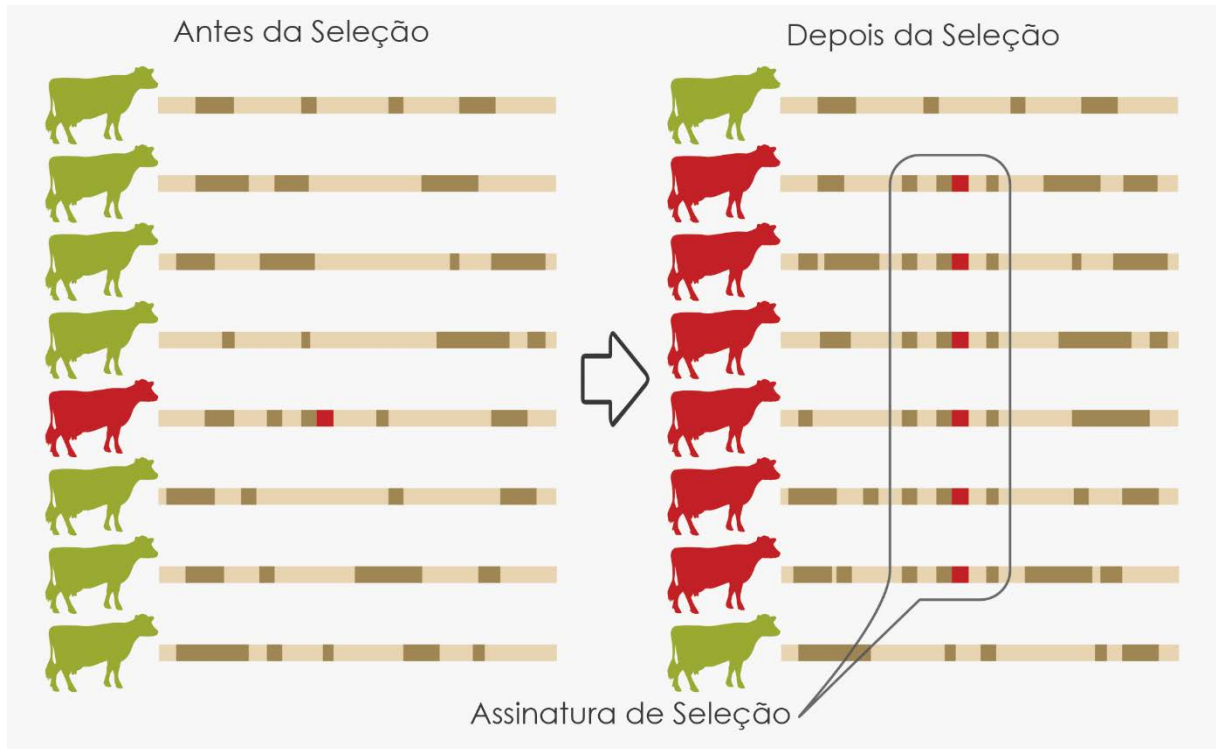


Figura 7 - Exemplo ilustrativo do processo de geração de “assinaturas de seleção” numa dada população.

Diversas metodologias foram desenvolvidas para a detecção de assinaturas de seleção, sendo que a maioria baseia-se na comparação das distribuições das frequências alélicas para um dado marcador (SNP), verificando se estas diferem de maneira estatisticamente significativa daquela esperada ao acaso (STELLA et al., 2010).

Em humanos, os estudos de assinatura de seleção permitiram elucidar a maneira pela qual alterações culturais e ambientais atuaram moldando o genoma ao longo dos séculos (AKEY, 2002; SABETI et al., 2002). Permitiram ainda, avaliar a filogenia dos alelos após a ocorrência do evento de seleção (BROOKFIELD, 2001) e identificar genes que contribuem para ou são causadores de defeitos genéticos (BERSAGLIERI et al., 2004; SABETI, 2006). As metodologias utilizadas nestes estudos passaram a ser aplicadas em outras espécies, e serviram de molde para o desenvolvimento de novos métodos.

Rubin et al. (2010), ao trabalhar com dados de sequenciamento em galinhas, identificaram genes que sofreram o efeito da domesticação e que estão diretamente ligados a características produtivas. Mais recentemente, esta metodologia também foi aplicada em suínos e ilustrou como as mudanças estruturais no DNA contribuíram

para a evolução fenotípica, alterando diferentes características tais como: comportamento, composição corporal e aspectos reprodutivos (RUBIN et al., 2012).

Nos bovinos, a domesticação teve início há mais de 8 mil anos, e desde então a espécie tem se adaptado às necessidades humanas (ELSIK et al., 2009). A seleção artificial promovida pelo homem, deixou impressas assinaturas detectáveis no genoma desta espécie (GIBBS et al., 2009). O estudo dessas “impressões” permitiu identificar no genoma regiões que potencialmente afetam características produtivas, tais como produção de leite e carne, bem como padrões destas assinaturas que permitiram a alta especialização das raças (HAYES et al., 2009). Flori et al. (2009), realizaram estudo envolvendo raças especializadas na produção de leite e encontraram assinaturas de seleção que evidenciam o antagonismo entre características produtivas e reprodutivas. Na espécie bovina, devido à forte seleção artificial, a localização destas assinaturas estará geralmente associada à uma região controladora de características de produção (HAYES et al., 2008). Portanto, o mapeamento destas regiões pode auxiliar o melhoramento genético das espécies de produção, à medida que genes de interesse podem ser rastreados na população, facilitando o direcionamento dos programas de acasalamento (ELFERINK et al., 2012).

Segundo a FAO (2013), até 2050 a população mundial deverá ultrapassar os 9 bilhões de pessoas sendo necessário o incremento de 60% na produção agropecuária, comparada a seus níveis de 2005 - 2007. O aumento no consumo de produtos de origem animal devido às mudanças na dieta e nos padrões de consumo, tem impulsionado a pecuária mundial. Nesse cenário, a baixa capacidade produtiva é a principal causa da estagnação do setor produtivo da carne bovina, onde o crescimento da produção per capita é insignificante quando comparado a produção de carne suína e de frango. Nesse contexto, a necessidade crescente de ferramentas capazes de aumentar de maneira eficiente e em tempo recorde a produtividade dos rebanhos, requer o desenvolvimento e aplicação de novas tecnologias capazes de sobrepujar estes desafios. A identificação de assinaturas de seleção em animais domésticos utilizando informação genômica com alta densidade de marcadores SNP tem sido a meta de diversos grupos de pesquisa. O uso de dados de sequenciamento genômico (consequentemente de milhões de SNPs) para a detecção de assinaturas de seleção empregando o método ZHp, foi recentemente descrito para aves (RUBIN et al., 2010) e suínos (RUBIN et al., 2012). Entretanto, os ainda altos custos

associados com a geração de dados genômicos completos e a existência de grande quantidade de informação genotípica na forma de SNP chips de diferentes densidades (particularmente na espécie bovina), aponta para a necessidade do desenvolvimento de métodos alternativos que usem essas últimas informações para a geração de informações fidedignas sobre assinaturas de seleção, para futuras aplicações no melhoramento animal.

Sendo assim, o presente trabalho objetiva, adaptar e validar metodologia capaz de encontrar regiões indicativas de assinaturas de seleção por meio do uso de dados SNP de painel comercial de marcadores da espécie bovina de maneira a fornecer subsídios técnicos e científicos que auxiliem os programas de seleção e melhoramento genético na aplicação de novas tecnologias de maneira a promover ganhos na produtividade de seus rebanhos.

8. Material e Métodos

8.1 Amostras

Os genótipos de 137 animais pertencentes às raças Angus (27), Holandesa (59), Hereford (27) e Pardo Suíço (24) da espécie bovina foram cedidos pelo Consórcio HapMap Bovino (GIBBS et al., 2009). A genotipagem das amostras foi feita com o ensaio Illumina® BovineHD Genotyping BeadChip (MATUKUMALLI et al., 2009) e corresponde à informação genotípica de 786.799 marcadores (SNP) distribuídos ao longo do genoma.

8.2 Controle de Qualidade (QC)

Os critérios adotados para a inclusão de SNP e indivíduos nas análises foram descritos por UTSUNOMIYA et al., (2013) e visam reduzir a ocorrência de associações espúrias, minimizar efeitos por erros de amostragem, mitigar o viés provocado pelo parentesco entre os indivíduos e evitar possíveis erros de genotipagem. O procedimento foi realizado para o conjunto de dados genotípicos de cada raça, fazendo uso do software *PLINK* (PURCELL et al., 2007). Foram incluídos apenas os marcadores dos cromossomos autossômicos e removidos os SNP com p-valor para o

equilíbrio de Hardy-Weinberg inferior a 1×10^{-6} ($EHW < 1 \times 10^{-6}$) e todos aqueles com taxa de determinação de genótipos (*call rate*) inferior a 90% ($CR_{SNP} < 90\%$). Animais cujo *call rate* foi inferior a 90% ($CR_{ANIM} < 90\%$) também foram removidos. Para manter o maior número possível de animais pouco aparentados, foram removidos de maneira sistemática animais com coeficiente de parentesco genômico superior a 40% ($PI_HAT > 0.4$), segundo a proporção de alelos compartilhados de forma idêntica por descendência (IBD).

8.3 Análise de assinaturas de seleção

A detecção de evidências de assinaturas de seleção recente consiste em adaptação do método utilizado por Rubin et al. (2010; 2012), que avaliou dados de sequência genômica completa em aves e suínos. A metodologia avalia o decaimento da heterozigosidade esperada quando comparada a heterozigosidade média do genoma. Para tanto foi feita a contagem dos alelos contidos em uma janela deslizante com tamanho em quilo bases (Kb) variando entre 100Kb e 1000Kb (100Kb, 200Kb, 400Kb, 500Kb, 600Kb, 700Kb, 800Kb, 900Kb e 1000Kb). A janela percorreu os cromossomos de cada raça dando passos com um quarto de seu tamanho (25Kb, 50Kb, 100Kb, 125Kb, 150Kb, 175Kb, 200Kb, 225Kb e 250Kb) cujo ponto médio demarca a coordenada genômica (**Figura 8**).

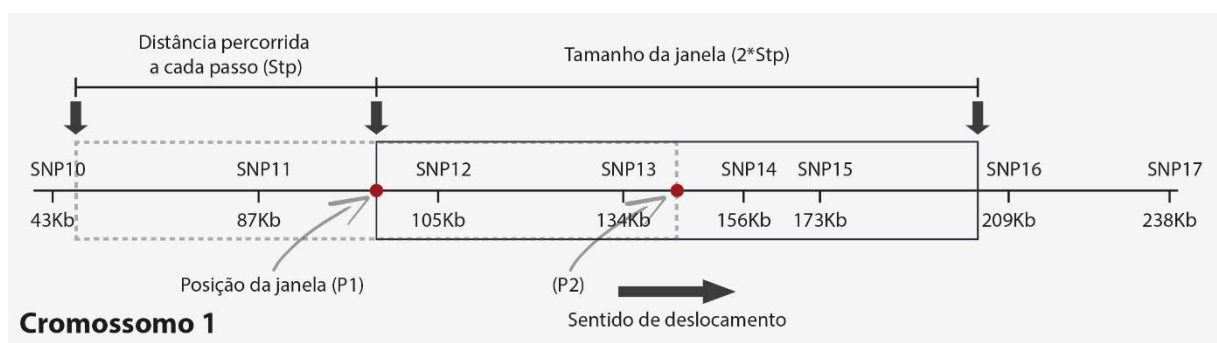


Figura 8 - Ilustração do processo de análise por janela deslizante percorrendo a informação genotípica ao longo do cromossomo.

Para cada SNP dentro da janela, foram contabilizados o número de leituras correspondentes ao alelo mais e menos frequente (n_{MAJ} e n_{MIN} , respectivamente) e calculada a heterozigosidade esperada (H_p) segundo a fórmula:

$$H_p = \frac{2\sum n_{MAJ}\sum n_{MIN}}{(\sum n_{MAJ} + \sum n_{MIN})^2}$$

Onde $\sum n_{MAJ}$ e $\sum n_{MIN}$ são os somatórios das contagens de alelos para cada janela. Após a obtenção dos valores de H_p , efetuou-se a padronização destes por meio da transformação Z :

$$ZH_p = \frac{(H_p - \mu_{H_p})}{\sigma_{H_p}}$$

Assim, o ZH_p é o escore padronizado da heterozigosidade esperada para dada janela, indicando seu distanciamento da heterozigosidade média do genoma permitindo comparações entre diferentes raças. Para evitar sinais de fixação espúria, ocasionados pela adição de janelas que contenham poucos SNP, foram omitidas janelas com menos de 10 marcadores. Os algoritmos para executar o método descrito foram desenvolvidos e implementados na linguagem R (IHAKA; GENTLEMAN, 1996) com o auxílio do ambiente de desenvolvimento integrado (IDE) RStudio (RACINE, 2012). Janelas com um valor de ZH_p inferior a -3, representam a extremidade inferior da distribuição, portanto indicativas de uma possível assinatura de seleção associada à sua posição.

8.4 Validação

Para validar o método, janelas que apresentaram valores de ZH_p abaixo do limiar de corte (janelas candidatas) foram quanto à proximidade da sua localização às regiões dos genes *receptor de melanocortina 1* (MC1R - ENSBTAG00000023731) e do *receptor do fator de crescimento de mastócitos e células tronco* (KIT - ENSBTAG00000002699), frequentemente utilizados na literatura como loci de confirmação para validar achados em estudos de assinaturas de seleção (FLORI et al., 2009; GIBBS et al., 2009; HAYES et al., 2008, 2009; STELLA et al., 2010; UTSUNOMIYA et al., 2013). Para tanto foram criados com o auxílio do software PLINK, quatro conjuntos de dados correspondentes aos genótipos das raças Angus (ANG), Holandês (HOL), Hereford (HFD) e Pardo Suíço (BSW). A implementação do algoritmo foi efetuada em cada conjunto separadamente e verificada a presença de

assinaturas de seleção nas regiões de validação descritas. Devido à dimensão do conjunto de dados, os resultados foram representados na forma de *Manhattan plots* (GIBSON, 2010), compondo o mapeamento do decaimento da heterozigidade esperada local para cada raça do presente estudo.

9. Resultados

Na **Tabela 2**, *P_{snp}* e *P_{anim}* indicam o percentual de SNP e animais respectivamente, que passaram no controle de qualidade, totalizando 120 animais que prosseguiram para etapas de análise subsequentes.

Tabela 2 - Número de SNP e animais antes e após o controle de qualidade segundo a raça.

Raça	ID	T _{snp}	T _{anim}	QC _{snp}	QC _{anim}	P _{snp} (%)	P _{anim} (%)
Angus	ANG	786799	27	733187	25	93.19	92.59
Hereford	HFD	786799	27	733531	25	93.23	92.59
Holandês	HOL	786799	59	733282	46	93.20	77.97
Pardo Suíço	BSW	786799	24	742910	24	94.42	100.00

A **Tabela 3** exemplifica o resultado observado com o emprego do software PLINK, onde *CHR* indica o cromossomo, *SNP* corresponde à identificação do marcador, *A1* e *A2* são os alelos encontrados para dado marcador, *C1* e *C2* indicam as contagens dos respectivos alelos e *G0* indica o número de animais com genótipos indeterminados.

Tabela 3 - Seis primeiras linhas da contagem de frequências alélicas na raça Angus.

CHR	SNP	A1	A2	C1	C2	G0
1	BovineHD0100000005	A	B	7	43	0
1	BovineHD0100000015	B	A	8	42	0
1	BovineHD0100000024	0	A	0	50	0
1	BovineHD0100000026	B	A	5	45	0
1	BovineHD0100000027	A	B	7	43	0
1	BovineHD0100046367	B	A	7	43	0

Na **Tabela 4**, observam-se exemplos dos resultados da implementação do algoritmo de detecção de assinaturas de seleção, que compreendem a identificação do cromossomo em que a janela se encontra (*CHR*), heterozigidade agrupada da

janela (*HP*), escore padronizado (*ZHp*), posição do centro da janela no cromossomo (*WCEN*), número de marcadores avaliados na janela (*WSNPS*), posição de início (*WSTT*) e fim da janela (*WEND*), e tamanho da mesma (*WLEN*).

Tabela 4 - Seis primeiras linhas do resultado gerado pelo algoritmo para detecção de assinaturas de seleção.

CHR	HP	ZHp	WCEN	WSNPS	WSTT	WEND	WLEN
1	0.30	-0.60	316947	170	16947	616947	6.00E+05
1	0.27	-0.95	466947	188	166947	766947	6.00E+05
1	0.28	-0.78	616947	206	316947	916947	6.00E+05
1	0.27	-1.10	766947	190	466947	1066947	6.00E+05
1	0.25	-1.29	916947	187	616947	1216947	6.00E+05
1	0.28	-0.79	1066947	191	766947	1366947	6.00E+05

A **Figura 9** ilustra a distribuição dos valores do *ZHp* quando usados 600Kb como tamanho da janela e passo de 150Kb, segundo a raça. Apresenta distribuição aproximadamente normal onde valores de *ZHp* do extremo esquerdo da distribuição ($ZHp \leq -3$), aponta para janelas cuja coordenada genômica sugere região putativa de albergar possível assinatura de seleção.

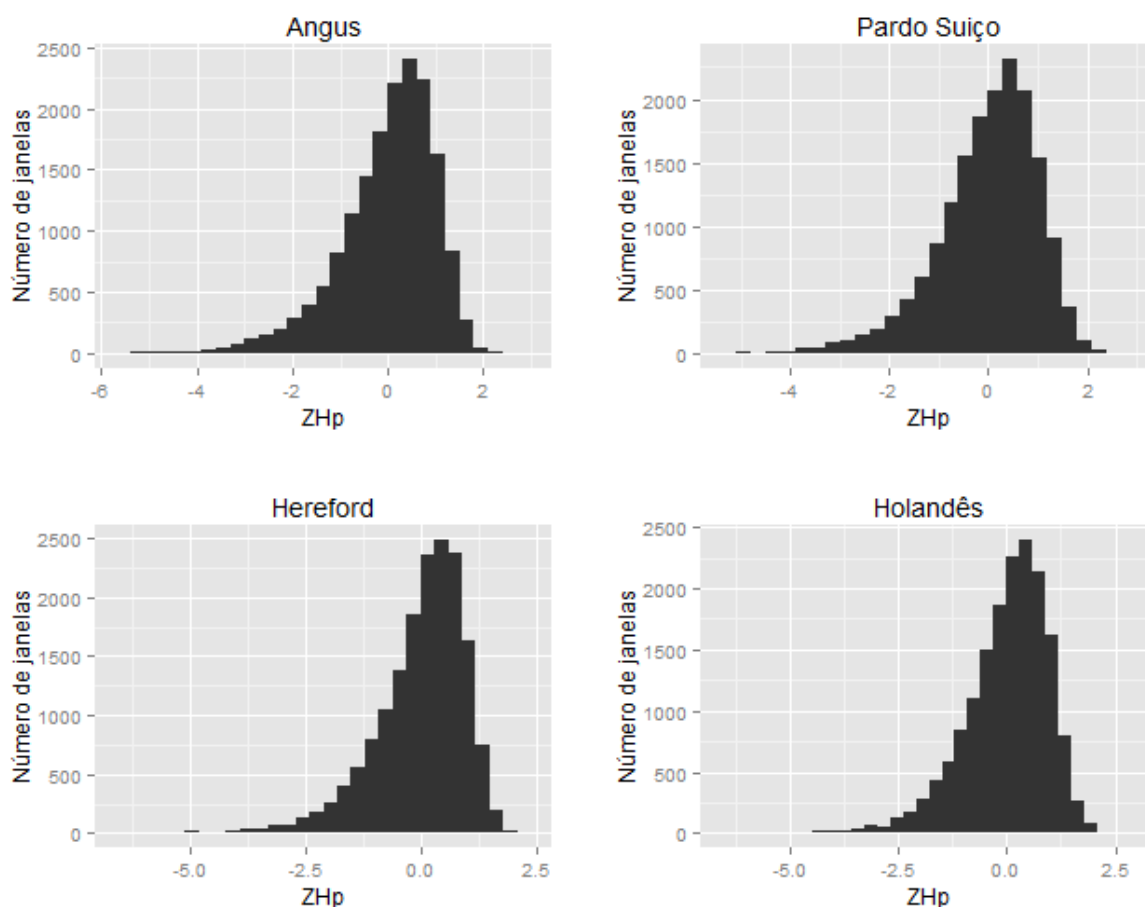


Figura 9 - Distribuição do ZHp segundo a raça.

Nas **Figuras 10, 11, 12 e 13**, os pontos representam os valores obtidos para os escores padronizados da heterozigosidade esperada ZHP (eixo Y), em função da localização no genoma (eixo X). Foram considerados apenas os desvios à esquerda da distribuição ($ZHp < 0$). A linha tracejada indica o limiar de detecção (-3 desvios) e os picos observados (pronunciamento dos pontos em direção ao limiar e que ultrapassam o mesmo) sugerem regiões genômicas sugestivas de assinatura de seleção.

Nas **Figuras 10 e 11** é observada a distribuição do ZHp ao longo do genoma na raça Angus quando utilizados diferentes tamanhos de janela, apontando incremento progressivo do desvio da heterozigosidade esperada segundo aumento progressivo do tamanho da janela.

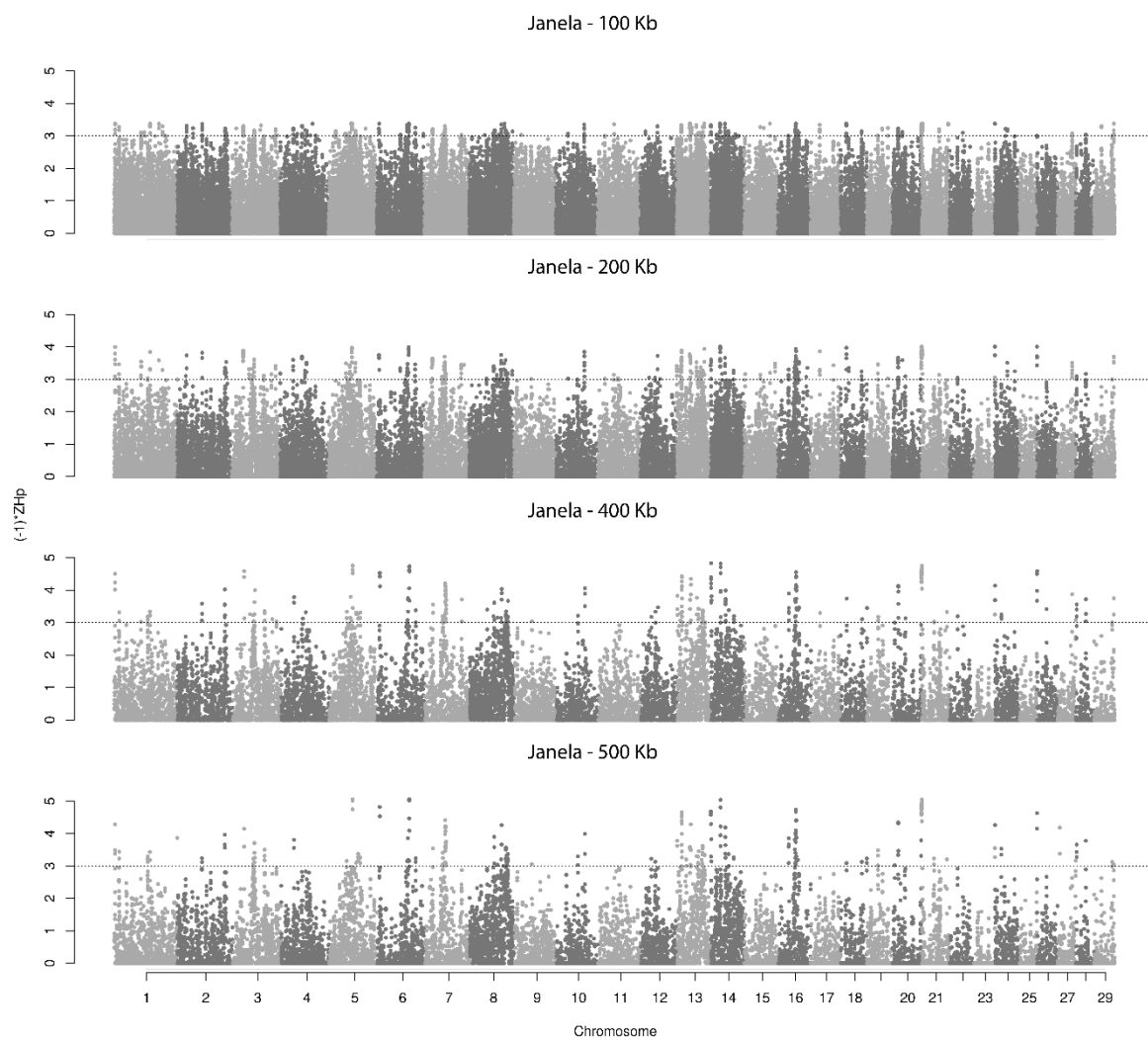


Figura 10 - Manhattan Plots dos valores de $Z_{Hp} < 0$ segundo a posição da janela no cromossomo e o tamanho das mesmas (100 Kb – 500 Kb) na raça Angus.

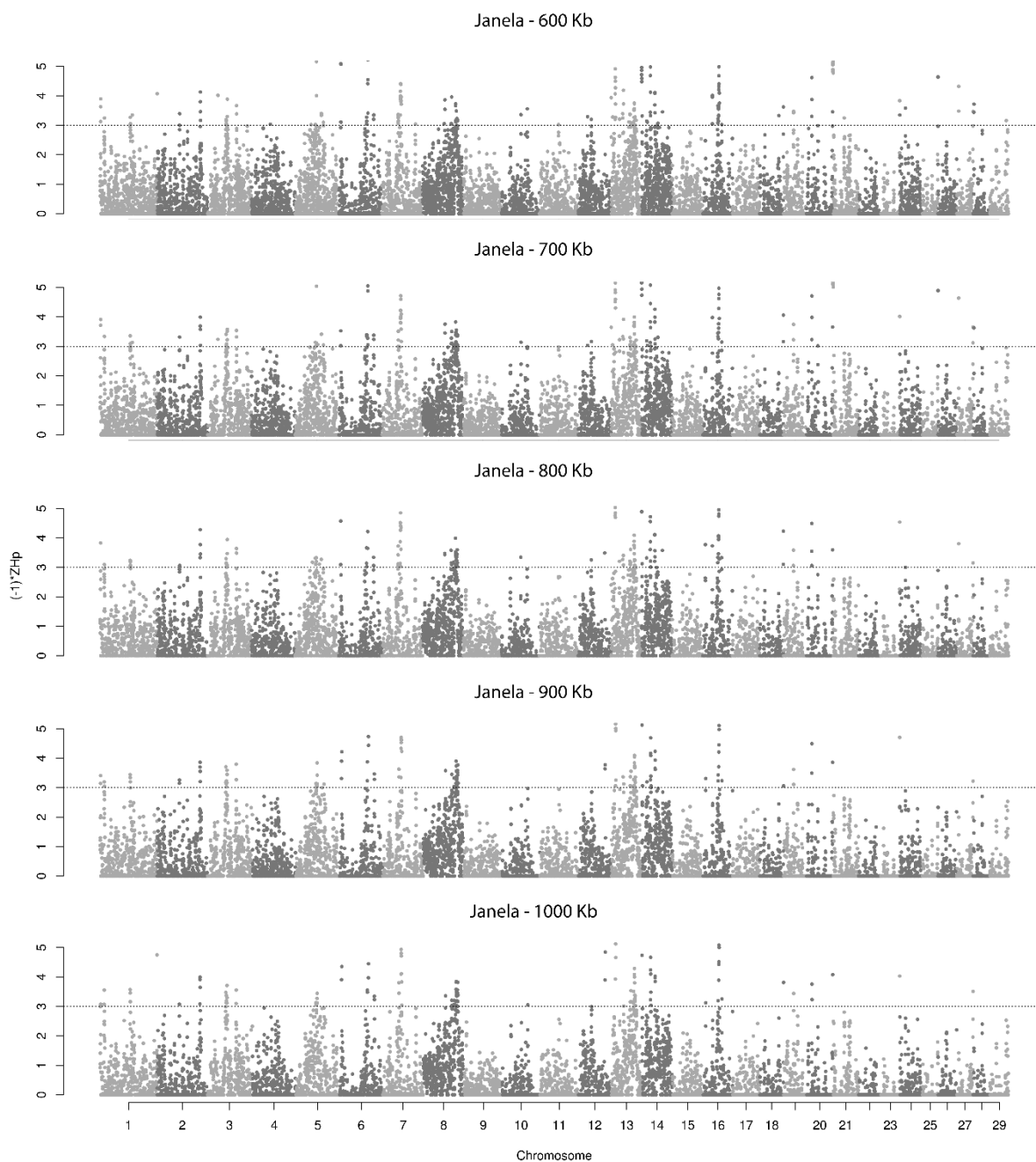


Figura 11 - Manhattan Plots dos valores de $ZHp < 0$ segundo a posição da janela no cromossomo e o tamanho das mesmas (600 Kb – 1000 Kb) na raça Angus.

Assinaturas de seleção evidenciadas nas raças do conjunto de dados de validação, que recaem na região dos genes MC1R (BTA18: 14757332-14759082 UMD v3.1 *assembly*) e KIT (BTA6:71796318-71917431 UMD v3.1 *assembly*), utilizados para validar a metodologia, cujas posições são apontadas pelas setas vermelhas (**Figuras 12 e 13**).

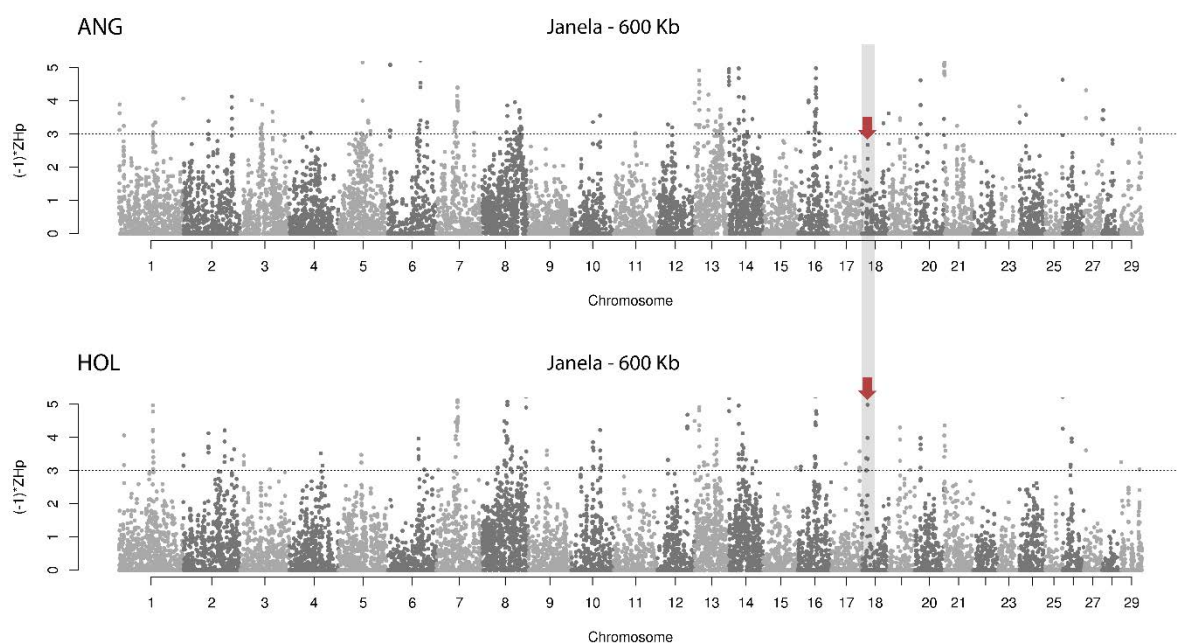


Figura 12 - Manhattan Plots evidenciando a região do gene MC1R e os valores de ZHp nas raças Angus e Holandês.

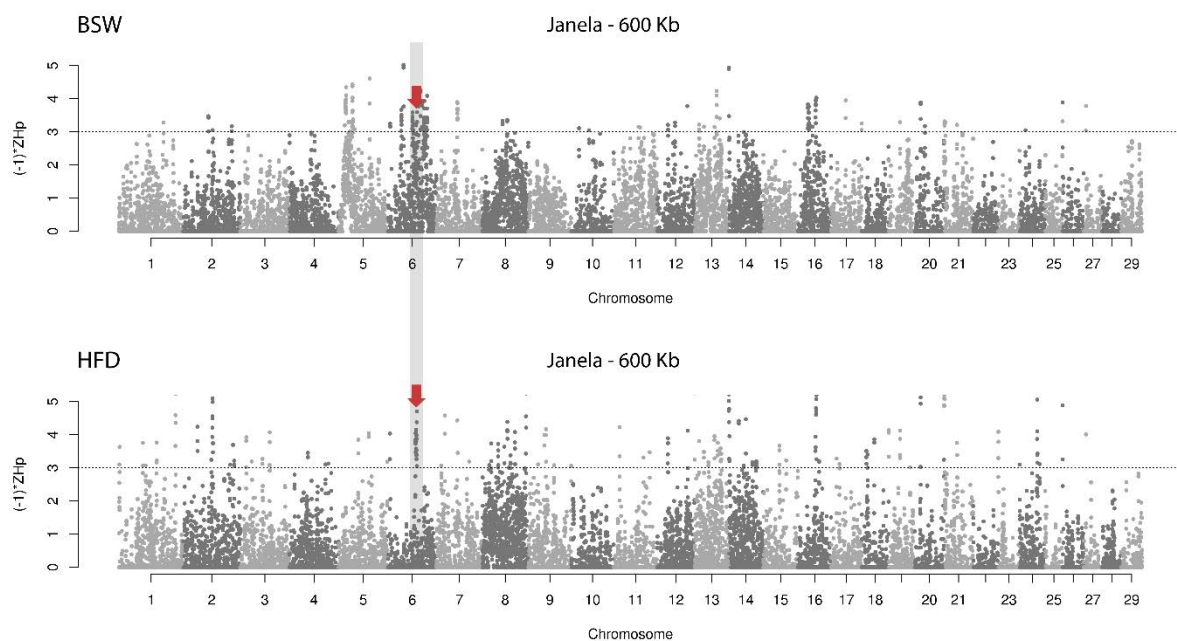


Figura 13 - Manhattan Plots evidenciando a região do gene KIT e os valores de ZHp nas raças Pardo Suíço e Hereford.

A **Tabela 5** descreve as regiões de validação cuja janela apresentou o maior desvio à esquerda da distribuição (menor heterozigiosidade local) segundo a raça avaliada.

Tabela 5 - Janelas apresentando os menores valores de ZHp na região dos genes de validação.

Raça	CHR	ZHp	NSNP	WSTT	WEND
Holandês	18	-5.6	105	14335883	14935883
Angus	18	-2.7	122	14485883	15085883
Hereford	6	-4.71	154	71281280	71881280
Pardo Suíço	6	-3.6	167	71731280	72331280

10. Discussão

Conforme sumarização dos resultados para o controle de qualidade de SNP e indivíduo (**Tabela 2**), denota ampla cobertura do genoma com distância média entre marcadores de 4,35Kb segundo número total de marcadores que prosseguiram para análise. Apesar do número expressivo de marcadores, essa informação em números absolutos está muito aquém do número de marcadores utilizado por (RUBIN et al., 2010) para desenvolver a metodologia. Estes, trabalharam com dados de sequenciamento cuja densidade de informação por quilo base do genoma é superior (1,03 SNP por Kb do genoma) à utilizado no presente estudo (0,23 SNP por Kb do genoma), muito embora o custo da informação seja diretamente relacionado à densidade da mesma. Incrementos desta densidade favorecem a significância bem como o mapeamento fino dos achados, porém levam à problemas computacionais importantes dadas as dimensões do conjunto de dados exemplificado nas **Tabelas 3** e **4**. As tecnologias necessárias para trabalhar com os mesmos dados utilizados pelos últimos autores, são de custo elevado e inviabilizam a aplicação das mesmas fora do âmbito acadêmico (BROOKFIELD, 2001).

Na **Figura 9** as distribuições dos escores padronizados da heterozigidade esperada seguem distribuição aproximadamente normal, com uma cauda pesada à esquerda, indicando o limiar de detecção das regiões de assinatura de seleção, permitindo inferir quanto ao distanciamento local da heterozigidade em relação à média. Distribuição semelhante também foi observada por Rubin et al. (2010; 2012), assim como foi notada a dependência da variabilidade dos valores de ZHp ao número total de marcadores avaliados por janela (**Figura 10** e **11**), onde incrementos sucessivos no tamanho da janela, implicam na redução do número total de janelas ao longo do genoma e correspondente incremento do número de SNP avaliados, resultando no aumento da variabilidade e conseqüente desvio da média. Logo, o uso

de janelas maiores leva a um incremento dos níveis de significância (janelas que ultrapassam o limiar de detecção), porém implicam em perda de resolução local, isto é, a inferência a respeito das possíveis assinaturas de seleção acaba recaindo sobre uma região maior do que aquela observada quando utilizadas janelas de menor tamanho. Assim, dentre os distintos tamanhos de janela avaliados, aqueles com tamanho entre 600Kb e 1000Kb conferiram a melhor relação entre resolução e significância. Logo, o número total de SNP avaliados, o cromossomo e a posição no mesmo dentro da janela, o tamanho desta e seu passo ao percorrer o genoma são fatores afetaram o número de SNP e conseqüentemente a variabilidade do ZHp.

O tamanho de 600Kb adotado para a validação da metodologia correspondeu ao menor tamanho de janela onde foi possível obter o maior desvio da heterozigosidade esperada e a melhor resolução nas regiões de validação (**Figuras 12 e 13**). Nestas regiões, a metodologia identificou assinatura de seleção próxima aos genes MC1R, localizado no cromossomo 18, determina a cor da pelagem preta em bovinos das raças Holandês e Angus (KLUNGLAND et al., 1995; MATUKUMALLI et al., 2009; STELLA et al., 2010), e ao gene KIT, localizado no cromossomo 6, responsável pelo aspecto malhado característico da pelagem de animais das raças Hereford e Pardo Suíço (GROSZ; MACNEIL, 1999; STELLA et al., 2010). Na **Figura 12**, as janelas sobre a região do gene MCR1 na raça Angus não ultrapassaram o limiar de detecção, entretanto é possível observar o pronunciamento das janelas vizinhas no sentido da significância estipulada, indicando claro decaimento da heterozigosidade local. É possível que indivíduos com a pelagem vermelha tenham sido introduzidos na amostragem, afetando estes resultados. Nas demais raças os picos de decaimento nas regiões de validação foram claramente observados, e a **Tabela 5** sumariza os achados mais significativos para cada uma destas regiões.

11. Conclusão

O método ZHp quando aplicado a dados genotípicos de alta densidade na espécie bovina, obtidos pelo uso de SNP chips comercialmente disponíveis, é válido para a detecção de assinaturas de seleção recente conforme hipotetizado. As ferramentas e resultados obtidos no presente trabalho cumpriram com os objetivos

propostos e sua disponibilização poderá auxiliar o entendimento dos mecanismos biológicos pelos quais ocorre a expressão fenotípica das características de interesse.

12. Referências

AJMONE-MARSAN, P. et al. On the Origin of Cattle: How Aurochs Became Cattle and Colonized the World. **Evolutionary Anthropology**, v. 19, n. 4, p. 148–157, 26 ago. 2010.

AKEY, J. M. Interrogating a High-Density SNP Map for Signatures of Natural Selection. **Genome Research**, v. 12, n. 12, p. 1805–1814, 1 dez. 2002.

BERSAGLIERI, T. et al. Genetic signatures of strong recent positive selection at the lactase gene. **American Journal of Human Genetics**, v. 74, n. 6, p. 1111–1120, jun. 2004.

BROOKFIELD, J. F. Y. Population genetics: The signature of selection. **Current Biology**, v. 11, n. 10, p. 388–390, maio 2001.

ELFERINK, M. G. et al. Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. **PLoS ONE**, v. 7, n. 2, p. e32720, jan. 2012.

ELSIK, C. G. et al. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. **Science**, v. 324, n. 5926, p. 522–528, 24 abr. 2009.

FAY, J. C.; WU, C. I. Hitchhiking under positive Darwinian selection. **Genetics**, v. 155, n. 3, p. 1405–1413, jul. 2000.

FLORI, L. et al. The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. **PLoS ONE**, v. 4, n. 8, p. e6595, jan. 2009.

FOOD AND AGRICULTURE ORGANISATION. **FAO Statistical Yearbook World food and agriculture**. vii ed. Rome: FAO, 2013.

GIBBS, R. A. et al. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. **Science**, v. 324, n. 5926, p. 528–532, 24 abr. 2009.

GIBSON, G. Hints of hidden heritability in GWAS. **Nature Genetics**, v. 42, n. 7, p. 558–560, jul. 2010.

GROSZ, M. D.; MACNEIL, M. D. The “spotted” locus maps to bovine chromosome 6 in a Hereford- cross population. **Journal of Heredity**, v. 90, n. 1, p. 233–236, 1 jan. 1999.

HAYES, B. J. et al. The origin of selection signatures on bovine chromosome 6.

Animal Genetics, v. 39, n. 2, p. 105–111, abr. 2008.

HAYES, B. J. et al. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. **Animal Genetics**, v. 40, n. 2, p. 176–184, abr. 2009.

IBGE. **Indicadores IBGE - Estatística da Produção Pecuária 2016**. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 13 jun. 2016.

IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. **Journal of computational and graphical ...**, v. 5, n. 3, p. 299–314, set. 1996.

KLUNGLAND, H. et al. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. **Mammalian Genome**, v. 6, n. 9, p. 636–639, set. 1995.

MATUKUMALLI, L. K. et al. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. **PLoS ONE**, v. 4, n. 4, p. e5350, jan. 2009.

MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Pecuária e Abastecimento. Projeções do Agronegócio : Brasil 2012/2013 a 2022/2023**. Brasília: Mapa/ACS, 2013.

OLEKSYK, T. K.; SMITH, M. W.; O'BRIEN, S. J. Genome-wide scans for footprints of natural selection. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 365, n. 1537, p. 185–205, 12 jan. 2010.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, set. 2007.

RACINE, J. S. RStudio: A Platform-Independent IDE for R and Sweave. **Journal of Applied Econometrics**, v. 27, n. 1, p. 167–172, 26 jan. 2012.

RUBIN, C.-J. et al. Whole-genome resequencing reveals loci under selection during chicken domestication. **Nature**, v. 464, n. 7288, p. 587–591, 25 mar. 2010.

RUBIN, C.-J. et al. Strong signatures of selection in the domestic pig genome. **Proceedings of the National Academy of Sciences**, v. 109, n. 48, p. 19529–19536, 27 nov. 2012.

SABETI, P. C. et al. Detecting recent positive selection in the human genome from haplotype structure. **Nature**, v. 419, n. 6909, p. 832–837, 24 out. 2002.

SABETI, P. C. Positive Natural Selection in the Human Lineage. **Science**, v. 312, n. 5780, p. 1614–1620, 16 jun. 2006.

STELLA, A. et al. Identification of Selection Signatures in Cattle Breeds

Selected for Dairy Production. **Genetics**, v. 185, n. 4, p. 1451–1461, ago. 2010.

UTSUNOMIYA, Y. T. et al. Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods. **PLoS ONE**, v. 8, n. 5, p. e64280, jan. 2013.