

**UNIVERSIDADE FEDERAL DO CEARÁ
UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PROGRAMA DE DOUTORADO INTEGRADO EM ZOOTECNIA**

ANÁLISE GENÔMICA DAS PRINCIPAIS RAÇAS DE OVINOS BRASILEIRAS

JOÃO JOSÉ DE SIMONI GOUVEIA

**FORTALEZA – CE
ABRIL – 2013**

**UNIVERSIDADE FEDERAL DO CEARÁ
UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PROGRAMA DE DOUTORADO INTEGRADO EM ZOOTECNIA**

ANÁLISE GENÔMICA DAS PRINCIPAIS RAÇAS DE OVINOS BRASILEIRAS

**JOÃO JOSÉ DE SIMONI GOUVEIA
Médico Veterinário**

**FORTALEZA – CE
ABRIL – 2013**

JOÃO JOSÉ DE SIMONI GOUVEIA

**ANÁLISE GENÔMICA DAS PRINCIPAIS RAÇAS DE OVINOS
BRASILEIRAS**

Tese apresentada ao Programa de
Doutorado Integrado em Zootecnia
da Universidade Federal do Ceará,
Universidade Federal da Paraíba e
Universidade Federal Rural de
Pernambuco como requisito
parcial para obtenção do título de
Doutor em Zootecnia.

Profª. Dra. Sônia Maria Pinheiro de Oliveira – Orientadora
Dr. Marcos Vinícius Gualberto Barbosa da Silva – Coorientador

**FORTALEZA – CE
ABRIL – 2013**

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Ciências e Tecnologia

-
- G735a Gouveia, João José de Simoni.
Análise genômica das principais raças de ovinos brasileiras. / João José de Simoni Gouveia. –
2013.
154 f. : il. Color., enc. ; 30 cm.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Departamento de Zootecnia, Programa de Doutorado Integrado em Zootecnia, Fortaleza, 2013.
Área de Concentração: Produção Animal.
Orientação: Profa. Dra. Sônia Maria Pinheiro de Oliveira.
Coorientação: Dr. Marcos Vinicius Gualberto Barbosa da Silva.
1. Assinaturas de seleção. 2. *Ovis aries*. 3. Polimorfismos de base única. 4. Recursos genéticos animais. 5. Subestruturação de populações I. Título.

CDD 636.08

JOÃO JOSÉ DE SIMONI GOUVEIA

ANÁLISE GENÔMICA DAS PRINCIPAIS RAÇAS DE OVINOS BRASILEIRAS

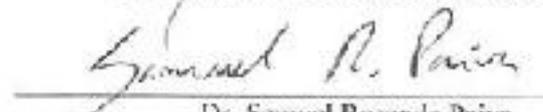
Tese apresentada ao Programa de Doutorado Integrado em Zootecnia da Universidade Federal do Ceará, Universidade Federal da Paraíba e Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do título de Doutor em Zootecnia. Área de Concentração: Produção Animal.

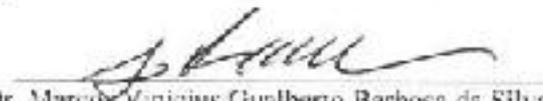
Aprovada em: 29/04/2013.

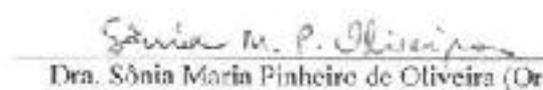
BANCA EXAMINADORA


PhD. Arlindo de Alencar Araripe Noronha Moura
Universidade Federal do Ceará (UFC)


Dra. Diana Magalhães de Oliveira
Universidade Estadual do Ceará (UECE)


Dr. Samuel Rezende Paiva
Empresa Brasileira de Pesquisa Agropecuária (Embrapa Recursos Genéticos e Biotecnologia)


Dr. Marcos Vinícius Gualberto Barbosa da Silva (Coorientador)
Empresa Brasileira de Pesquisa Agropecuária (Embrapa Gado de Leite)


Dra. Sônia Maria Pinheiro de Oliveira (Orientadora)
Universidade Federal do Ceará (UFC)

DADOS CURRICULARES DO AUTOR

JOÃO JOSÉ DE SIMONI GOUVEIA – Nascido em Fortaleza, Ceará, no dia 23 de fevereiro do ano de 1983. No ano de 2001 iniciou os estudos de Graduação em Medicina Veterinária na Universidade Estadual do Ceará (UECE). Durante o período da graduação, foi bolsista de iniciação científica através do programa PIBIC/CNPq, sob orientação da Profa. Dra. Diana Magalhães de Oliveira; monitor da disciplina de Genética Animal do Curso de Medicina Veterinária da UECE, sob orientação do Prof. Dr. Raimundo Bezerra da Costa e bolsista de iniciação científica através do programa IC/FUNCAP, sob orientação do Prof. Dr. Rodrigo Maggioni. Concluiu o curso de graduação no ano de 2005. No ano seguinte (2006), iniciou os estudos de Mestrado em Genética e Evolução pelo Programa de Pós Graduação em Genética e Evolução (PPGEV) da Universidade Federal de São Carlos (UFSCar), sob orientação da Dra. Luciana Correia de Almeida Regitano com bolsa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Concluiu o mestrado no ano de 2008, tendo defendido a dissertação intitulada: “Estudo de associação entre microssatélites localizados no cromossomo OAR3 e características de crescimento e resistência aos nematódeos gastrintestinais em ovinos”. Após a conclusão do mestrado, foi contratado pelo Instituto Superior de Teologia Aplicada (INTA) para ministrar disciplinas da área de genética e melhoramento animal para o curso de Graduação em Medicina Veterinária desta instituição. No ano seguinte (2009), iniciou os estudos de Doutorado em Zootecnia pelo Programa de Doutorado Integrado em Zootecnia (PDIZ UFC/UFPB/UFRPE), sob orientação da Profa. Dra. Sônia Maria Pinheiro de Oliveira e coorientação do Dr. Marcos Vinícius Gualberto Barbosa da Silva. Ainda no ano de 2009, foi aprovado em concurso público para o cargo de Professor Efetivo da área de Genética e Melhoramento Animal da Universidade Federal do Vale do São Francisco, onde atualmente ocupa o cargo de Professor Assistente II.

“Que sei eu do que serei, eu que não sei o que sou?
Ser o que penso? Mas penso tanta coisa!
E há tantos que pensam ser a mesma coisa que não pode haver tantos!”
Álvaro de Campos

À minha princesa, por tudo...

AGRADECIMENTOS

À minha esposa, Gisele Veneroni Gouveia, por ter tido a insanidade de casar comigo. Gostaria de agradecê-la pelo apoio, carinho e compreensão nesta fase tão única e difícil da minha vida. Sei o quanto de coisas você precisou (e quis) abrir mão para vir para Petrolina, e por isso eu te garanto que vou continuar fazendo tudo o que for possível para que você não se arrependa desta decisão.

Aos meus pais, Inácio e Beth, pela dedicação extrema que sempre tiveram com os filhos e por terem propiciado sempre um ambiente familiar perfeito para que eu pudesse sonhar. Além disso, por terem sempre ressaltado as minhas qualidades (às vezes até demais) e me estimulado a ir em busca daquilo que eu queria.

Aos meus irmãos, PP e Cla, que apesar estarem longe, sempre estão presentes. Não posso esquecer dos meus tios e primos (tanto do lado dos Simoni quanto do lado dos Gouveia) que sempre estiveram torcendo por mim e, nem da família Veneroni que tem me acolhido tão bem desde aceitaram que a filhinha fosse para tão longe.

À Profa. Dra. Sônia Maria Pinheiro de Oliveira, por ter me dado a oportunidade de desenvolver este trabalho sob sua orientação. Além disso, por ter me mostrado que boa vontade e humildade são essenciais para o sucesso de qualquer caminhada. Sônia, muito obrigado pela sua paciência e por ter confiado tanto em mim, embora a nossa convivência tenha sido muito menor do que eu gostaria que fosse, queria deixar claro que desde o início deste doutorado você tem sido uma inspiração para mim, principalmente no que diz respeito à forma como devemos encarar a vida e tratar as pessoas. Além de você ter me proporcionado a oportunidade para que eu desenvolvesse habilidades técnicas, você me proporcionou um enorme amadurecimento pessoal.

Ao Programa de Doutorado Integrado em Zootecnia (PDIZ UFC/UFPB/UFRPE) da Universidade Federal do Ceará, pela oportunidade de realizar o meu doutoramento. Quero registrar um agradecimento especial aos Docentes que compõem este programa, pelo esforço e dedicação dispensados por cada um em prol deste projeto de vida que é ser professor universitário e querer proporcionar a formação de recursos humanos qualificados e ampliação do conhecimento científico.

Ao Dr. Marcos Vinícius Gualberto Barbosa da Silva, por ter me dado a oportunidade de tê-lo como coorientador, por ter me aberto as portas do seu laboratório

e principalmente por ter proporcionado a oportunidade de desenvolvimento deste projeto.

Ao Dr. Samuel Rezende Paiva, pela oportunidade em poder trabalhar com os dados que utilizei para esta tese e por sempre ter se colocado à disposição para tirar minhas dúvidas, discutir resultados e fazer críticas extremamente construtivas.

À Profa. Dra. Ana Carolina Landim Pacheco, por ter me convidado para aquelas coletas que resultaram em uma conversa com a Profa. Sônia e na minha decisão em realizar o doutoramento na UFC.

Ao Léo, Elisângela e Dr. Wagner, do Laboratório de Bioinformática da Embrapa Gado de Leite, por terem me recebido de forma tão carinhosa durante minhas idas à “Outside’s Judge”.

À Universidade Federal do Vale do São Francisco, em especial aos colegas do Colegiado Acadêmico de Zootecnia, pelo apoio e compreensão nos momentos difíceis desta caminhada.

À todos aqueles que já foram meus professores e que tiveram a árdua tarefa de conseguir atrair minha atenção, hoje eu sei o quanto isso é difícil. Em especial àqueles que conseguiram me mostrar algo além de conhecimento técnico.

Não posso deixar de registrar o meu enorme agradecimento àqueles que tiveram a coragem de me orientar desde a graduação, em especial à Profa. Diana e à Dra. Luciana, por terem atuado como *enhancers* para que o meu amor à ciência pudesse se expressar.

Aos meus alunos, por me propiciarem a oportunidade de eu transmitir um pouco do pouco que sei. Não posso deixar de agradecer àqueles que, não satisfeitos com a disciplina, ainda me permitem exercer a deliciosa tarefa de orientar. Valeu Esdrinhas, João e Samla...

SUMÁRIO

	Página
CONSIDERAÇÕES INICIAIS.....	1
CAPÍTULO I.....	4
Identificação de assinaturas de seleção em animais de produção	4
RESUMO	5
ABSTRACT.....	6
INTRODUÇÃO	7
SELEÇÃO NATURAL E ARTIFICIAL.....	8
ASSINATURAS DE SELEÇÃO.....	9
MÉTODOS DE IDENTIFICAÇÃO DE ASSINATURAS DE SELEÇÃO	13
Testes baseados nas taxas de substituições sinônimas e não sinônimas	14
Testes baseados no espectro de frequências alélicas	14
Testes baseados no desequilíbrio de ligação.....	16
Testes baseados na diferenciação entre populações.....	17
ASSINATURAS DE SELEÇÃO EM ANIMAIS DE INTERESSE ZOOTÉCNICO....	19
Assinaturas de seleção relacionadas com a domesticação e com o desenvolvimento inicial das raças	21
Bovinos	23
Suínos.....	30
Ovinos e caprinos.....	31
CONCLUSÕES	32
REFERÊNCIAS BIBLIOGRÁFICAS.....	32
CAPÍTULO II	50
Identificação de assinaturas de seleção em ovinos de raças localmente adaptadas brasileiras	50
RESUMO	51
ABSTRACT.....	53
INTRODUÇÃO	55

MATERIAL E MÉTODOS	56
Animais e SNPs	56
Identificação de assinaturas de seleção com base na diferenciação entre populações....	57
Identificação de assinaturas de seleção com base no desequilíbrio de ligação.....	58
Anotação das regiões significativas	60
Análise funcional dos genes candidatos.....	61
RESULTADOS.....	61
DISCUSSÃO	97
CONCLUSÕES	103
REFERÊNCIAS BIBLIOGRÁFICAS.....	104
CAPÍTULO III.....	110
Identificação de subestrutura de populações em ovinos de raças localmente adaptadas brasileiras	110
RESUMO.....	111
ABSTRACT.....	113
INTRODUÇÃO	114
MATERIAL E MÉTODOS	116
Animais e SNPs	116
Diversidade genética	116
Análise baseadas nas subpopulações	117
Análises baseadas nos indivíduos	117
RESULTADOS.....	118
DISCUSSÃO	127
CONCLUSÕES	131
REFERÊNCIAS BIBLIOGRÁFICAS.....	131
CONSIDERAÇÕES FINAIS E IMPLICAÇÕES.....	135

LISTA DE TABELAS

	Página
CAPÍTULO I.....	4
Identificação de assinaturas de seleção em animais de produção	4
CAPÍTULO II	50
Identificação de assinaturas de seleção em ovinos de raças localmente adaptadas brasileiras	50
Tabela 1. Informações sobre os indivíduos utilizados	56
Tabela 2. SNPs significativos para as análises de F_{ST}	61
Tabela 3. SNPs significativos tanto para a análise de F_{ST} realizada com as três raças localmente adaptadas quanto com pelo menos uma das comparações duas a duas entre as raças.	68
Tabela 4. Regiões genômicas candidatas a conter genes sob seleção identificadas através das análises de F_{ST}	70
Tabela 5. SNPs significativos para as análises de iHS.	71
Tabela 6. Regiões genômicas candidatas a conter genes sob seleção identificadas através das análises de iHS.....	76
Tabela 7. SNPs significativos para as análises de R_{sb}	79
Tabela 8. Regiões genômicas candidatas a conter genes sob seleção identificadas através das análises de R_{sb}	88
Tabela 9. Anotação das regiões genômicas candidatas a estar sob seleção.	91
CAPÍTULO III.....	110
Identificação de subestrutura de populações em ovinos de raças localmente adaptadas brasileiras	110
Tabela 10. Informações sobre os indivíduos utilizados	116
Tabela 11. Diversidade genética nas três raças de ovinos localmente adaptadas brasileiras.	118

Tabela 12. Análise de variância molecular (AMOVA).	120
Tabela 13. Resultados do teste de Tracy-Widom para análise de componentes principais realizada considerando as três raças brasileiras de ovinos localmente adaptadas.	121
Tabela 14. Resultados do teste de Tracy-Widom para análise de componentes principais realizada para cada raça separadamente.....	121

LISTA DE FIGURAS

	Página
CAPÍTULO I.....	4
Identificação de assinaturas de seleção em animais de produção	4
CAPÍTULO II	50
Identificação de assinaturas de seleção em ovinos de raças localmente adaptadas brasileiras	50
Figura 1. Manhattan plot para a análise de F_{ST} . Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs significativos, sendo que os pontos preenchidos representam os SNPs significativos na análise envolvendo as três raças brasileiras e pelo menos uma comparação duas a duas.	66
Figura 2. Diagrama de Venn ilustrando os SNPs significativos para a análise de F_{ST} realizada com as três raças localmente adaptadas (Raças Brasileiras), e com as comparações duas a duas entre as raças (BCS x BMN, BCS x BSI e BMN x BSI).....	67
Figura 3. Manhattan plot para a análise de iHS . Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs significativos.....	75
Figura 4. Manhattan plot para a análise de R_{sb} . Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs significativos.....	87
Figura 5. Manhattan plot dos valores de R_{sb} . Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs extremos da distribuição.	102
CAPÍTULO III.....	110
Identificação de subestrutura de populações em ovinos de raças localmente adaptadas brasileiras	110
Figura 6. Heatmap ilustrando os valores de F_{ST} par a par (diagonal superior) e as distâncias genéticas de Reynolds (diagonal inferior) entre cada rebanho. Cores tendendo ao vermelho indicam valores mais próximo de zero enquanto as cores tendendo ao branco indicam valores mais distantes de zero.	119

Figura 7. Análise de coordenadas principais à partir das distâncias genéticas de Reynolds. A raça Crioula Lanada está representada pela cor vermelha, Morada Nova pela cor verde e Santa Inês pela cor azul.	120
Figura 8. Análise de componentes principais realizada considerando as três raças brasileiras de ovinos localmente adaptadas.....	122
Figura 9. Análise de componentes principais para a raça Crioula Lanada.	123
Figura 10. Análise de componentes principais para a raça Morada Nova.	123
Figura 11. Análise de componentes principais para a raça Santa Inês.....	124
Figura 12. Análise de clusterização hierárquica não paramétrica para as três raças brasileiras localmente adaptadas de ovinos. Cada cor representa uma raça: Crioula Lanada (vermelho), Morada Nova (verde) e Santa Inês (azul).	124
Figura 13. Análise de clusterização hierárquica não paramétrica para a raça Crioula Lanada.	125
Figura 14. Análise de clusterização hierárquica não paramétrica para a raça Morada Nova.	125
Figura 15. Análise de clusterização hierárquica não paramétrica para a raça Santa Inês.	126
Figura 16. Estatística Delta K para a análise de clusterização Bayesiana. A) Raças Brasileiras , B) Crioula Lanada, C) Morada Nova e D) Santa Inês.	126
Figura 17. Resultados da análise de clusterização Bayesiana para K variando de dois a cinco considerando, tanto as raças conjuntamente, como cada raça separadamente. ...	127

ANÁLISE GENÔMICA DAS PRINCIPAIS RAÇAS DE OVINOS BRASILEIRAS

RESUMO GERAL

As raças de ovinos localmente adaptadas brasileiras, também conhecidas como nativas ou crioulas, descendem de animais trazidos durante o período colonial e, desde aquela época, vêm sendo submetidas a processos evolutivos sistemáticos e não sistemáticos., o que resultou na formação de genótipos altamente adaptados às mais diversas condições ambientais brasileiras. Embora estas raças não possuam o mesmo potencial produtivo das raças exóticas melhoradas, elas são consideradas extremamente importantes devido às relações sociais e culturais que guardam com as populações do campo. Além disso, as raças localmente adaptadas possuem características adaptativas importantíssimas para a manutenção de sistemas produtivos tradicionais. A otimização da utilização dos recursos genéticos naturalizados depende de um conhecimento profundo destas populações e, portanto, a caracterização morfológica, produtiva e molecular são ferramentas imprescindíveis para o sucesso da conservação e utilização deste recurso genético. Com base nisso, o objetivo desta tese foi aprofundar os estudos de caracterização molecular das principais raças de ovinos localmente adaptadas brasileiras: Crioula Lanada, Morada Nova e Santa Inês. Assim, o capítulo I intitulado “*Identificação de assinaturas de seleção em animais de produção*” consiste em uma revisão cujo objetivo é descrever os principais efeitos da seleção natural/artificial nos genomas das espécies de animais de produção, apresentar os principais métodos de análise de assinaturas de seleção e discutir os recentes avanços nesta área de estudo. Foram realizados dois estudos que resultaram nos capítulos II e III desta tese. O capítulo II, intitulado “*Identificação de assinaturas de seleção em ovinos de raças localmente adaptadas brasileiras*”, teve como objetivo a identificação e caracterização de regiões genômicas prováveis de estar sob seleção nas três principais raças brasileiras localmente adaptadas de ovinos e caracterizar estas regiões com a finalidade de identificar genes envolvidos com diferenças produtivas/adaptativas entre estas raças. A identificação das regiões sujeitas à seleção foi feita com base em dois tipos de metodologia: diferenciação entre populações (F_{ST}) e desequilíbrio de ligação (iHS e R_{sb}). Foram identificados 78

genes candidatos envolvidos com funções como: resposta imune, desenvolvimento do sistema nervoso, percepção sensorial e desenvolvimento de pelos/lã. O capítulo III, intitulado: “*Identificação de subestrutura de populações em ovinos de raças localmente adaptadas brasileiras*”, teve como objetivos identificar e caracterizar a presença de subestruturação genética dentro das três principais raças brasileiras localmente adaptadas de ovinos. Foi observada, de uma forma geral, a presença de diferenciação genética bastante considerável ao comparar os rebanhos de cada raça analisada. Além disso, tanto na raça Crioula Lanada quanto na raça Santa Inês, pode ser observada a presença de grupos bem distintos de indivíduos, sugerindo a efetiva presença de diferentes ecótipos/linhagens dentro destas raças.

Palavras-chave: assinaturas de seleção, *Ovis aries*, polimorfismos de base única, recursos genéticos animais, subestruturação de populações

GENOMIC ANALYSIS OF THE MAJOR BRAZILIAN SHEEP BREEDS

ABSTRACT

The Brazilian locally adapted sheep breeds, also known as native and creole, are descendant from animals brought during the settlement period and, since then, are been subjected to evolutive (both systematic and non systematic) processes, what resulted in the formation of highly adapted genotypes to the diverse environmental Brazilian conditions. Although these breeds don't possess the same productive potential when compared with the exotic improved breeds, they are considered extremely important from a social and cultural point of view. Moreover, the locally adapted breeds are essential for the maintenance of traditional production systems. The optimization of the utilization of naturalized genetic resources depends on a deep knowledge of these populations, and then, the morphological, productive and molecular characterizations are essential to the success of conservation and utilization of these locally adapted genotypes. Therefore, the principal aim of this thesis was to deepen molecular the knowledge of the three main locally adapted Brazilian sheep breeds: Brazilian Creole, Morada Nova and Santa Ines. Thus, Chapter I entitled "*Identification of selection signatures in livestock species*" is a literature review that is proposed to describe the main effects of natural/artificial selection in the genomes of species of farm animals, present the main methods of signature selection analysis and discuss recent advances in this area of study. Two studies were conducted and resulted in the chapters II and III of this thesis. The chapter II, entitled "*Identification of selection signatures in brazilian locally adapted sheep breeds*", aimed the identification and characterization of putative genomic regions that underwent selection and the identification of candidate genes related to productive/adaptative differences between these breeds. The identification of signatures of selection was performed through two approaches: population differentiation (F_{ST}) and linkage disequilibrium (iHS and R_{sb}). Seventy eight genes, related to functions as: immune response, nervous system development, sensorial perception and wool/hair development were identified. The chapter III, entitled "*Identification of population substructure in brazilian locally adapted sheep breeds*",

aimed the identification and characterization of genetic substructure within the three main locally adapted Brazilian sheep breeds: Brazilian Creole, Morada Nova and Santa Ines. The level of genetic differentiation between herds of the same breeds was, in general, high. Both in the Brazilian Creole and in the Santa Ines breeds the presence of distinct groups on animals could be observed, what suggests the occurrence of different ecotypes/lineages within these breeds.

Keywords: animal genetic resources, *Ovis aries*, population substructure, signatures of selection, single nucleotide polymorphisms

CONSIDERAÇÕES INICIAIS

A crescente demanda de carne ovina no Brasil vem promovendo uma alteração considerável na ovinocultura nacional: de uma exploração extensiva e baseada em pequenas propriedades para a exploração cada vez mais empresarial. Este fato vem propiciando o aumento cada vez maior da utilização de raças exóticas melhoradas para carne.

Os benefícios da utilização de raças exóticas melhoradas para produção de carne nos sistemas mais tecnificados se contrapõem aos riscos potenciais à manutenção de um patrimônio genético altamente adaptado às mais diversas peculiaridades ambientais e culturais brasileiras, as raças localmente adaptadas.

Estas raças, também conhecidas como nativas ou crioulas, são derivadas de animais trazidos durante o período colonial e, descendem, principalmente, de animais oriundos da Península Ibérica. Desde a chegada, estes animais vem sendo alvo de vêm sendo submetidas a processos evolutivos sistemáticos e não sistemáticos, o que resultou na formação de genótipos adaptados às diversas condições ambientais brasileiras (algumas extremamente desfavoráveis).

Embora as raças localmente adaptadas não possuam o mesmo potencial produtivo das raças exóticas melhoradas, estes indivíduos possuem características adaptativas importantes, principalmente para pequenos produtores localizados em regiões de ambientes mais desfavoráveis. Além disso, existe uma relação social e cultural muito importante entre o homem do campo e estas raças.

A importância cultural e social das raças ovinas localmente adaptadas brasileiras fez com que a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) incluísse estas raças nos planos de conservação que, dentre outras estratégias, resultou na criação de núcleos de conservação *in situ*.

A melhor forma de conservação de um recurso genético é a sua utilização e, para isso, deve-se conhecer o potencial de produção das raças, a fim de se otimizar o processo produtivo para garantir um melhor retorno econômico aos produtores.

No Brasil, já existem diversos estudos com a finalidade de caracterizar as raças localmente adaptadas de ovinos (caracterização morfológica, produtiva e molecular). Os

resultados destes estudos têm servido para alicerçar os programas de conservação destas raças e para subsidiar discussões importantes em relação à utilização destas raças.

O conhecimento profundo das raças e populações alvos dos programas de conservação tem implicações diretas nas estratégias utilizadas para fazer o manejo genético destes indivíduos. Além disso, pode-se identificar peculiaridades adaptativas e produtivas (e conseqüentemente alelos únicos) nestas raças que fazem delas como fontes de variação genética extremamente importantes. Alguns exemplos de raças localmente adaptadas portadoras de alelos únicos e da sua utilização no melhoramento de outras raças são conhecidos em diversas espécies de animais de produção.

Nas raças de ovinos localmente adaptadas brasileiras já se conhece algumas características produtivas e adaptativas com um potencial interessante de utilização. Porém, para que a utilização seja otimizada, é muitas vezes importante que a base genética destas características seja decifrada.

Nos últimos anos, a genômica animal teve um desenvolvimento extraordinário em resposta a elaboração de métodos mais eficientes e baratos de sequenciamento e genotipagem de indivíduos. Estes métodos propiciaram o sequenciamento do genoma das principais espécies de interesse zootécnico, inclusive dos ovinos.

O sequenciamento de um genoma abre portas para uma infinidade de estudos com a finalidade de identificar genes/mutações responsáveis pelo controle de características de interesse econômico. O resultado do avanço da genômica animal é a publicação de diversos estudos descrevendo genes envolvidos com características como: presença/ausência de chifres, resistência contra lentivirose, prolificidade, resistência contra parasitos gastrintestinais e doenças genéticas.

Com base no exposto acima, o objetivo principal desta tese foi aprofundar o conhecimento molecular sobre três raças de ovinos brasileiras localmente adaptadas: Crioula Lanada, Morada Nova e Santa Inês através da utilização de um chip de polimorfismos de base única (SNPs) contendo aproximadamente 50 mil marcadores distribuídos ao longo de todo o genoma ovino.

Especificamente, esta tese objetivou descrever os principais efeitos da seleção natural/artificial nos genomas das espécies de animais de produção, apresentar os principais métodos de análise de assinaturas de seleção e discutir os recentes avanços nesta área de estudo (Capítulo I), identificar regiões genômicas prováveis de estar sob

seleção nas três raças citadas acima e caracterizar estas regiões com a finalidade de identificar genes envolvidos com diferenças produtivas/adaptativas entre estas raças (Capítulo II). Além disso, esta tese objetivou identificar e caracterizar a presença de subestruturação genética dentro destas raças com a finalidade de subsidiar a tomada de decisões em relação às estratégias utilizadas nos programas de conservação (Capítulo III).

CAPÍTULO I

Identificação de assinaturas de seleção em animais de produção

IDENTIFICAÇÃO DE ASSINATURAS DE SELEÇÃO EM ANIMAIS DE PRODUÇÃO

RESUMO

A identificação de regiões sujeitas à seleção é um dos principais tópicos de interesse da genética evolutiva nos dias atuais. Estes estudos podem prover, desde informações básicas sobre os processos evolutivos responsáveis por modelar os genomas, até informações funcionais sobre genes/regiões genômicas. O processo de domesticação, a criação de padrões morfológicos e funcionais que culminaram com o desenvolvimento das raças atuais e os programas de melhoramento genético possibilitaram a formação de uma quantidade muito grande de raças de animais de produção adaptadas à diversos ambientes e com diferentes aptidões produtivas. Os avanços recentes na genômica possibilitaram o desenvolvimento e utilização de diversas metodologias para identificar assinaturas de seleção, resultando, nos últimos anos, na publicação de diversos trabalhos envolvendo as espécies de interesse zootécnico. Os objetivos desta revisão são: descrever os principais efeitos da seleção natural/artificial nos genomas das espécies de animais de produção, apresentar os principais métodos de análise de assinaturas de seleção e discutir os recentes avanços nesta área de estudo.

Palavras-chave: marcas de seleção, animais domésticos, seleção artificial

IDENTIFICATION OF SELECTION SIGNATURES IN LIVESTOCK SPECIES

ABSTRACT

The identification of regions that underwent selection is one of the principal interests of evolutionary genetics nowadays. These studies can provide since information about evolutionary processes that are shaping genomes up to functional information about genes/genomic regions. Domestication process followed by breed formation and selection schemes enabled the formation of very diverse modern livestock breeds adapted to a wide variety of environments and with special productive characteristics. The recent advances in genomics enabled the development and implementation of several methods to detect selection signatures, resulting in the publication of a considerably number of studies involving livestock species in the last years. The aims of this review are to describe the principal effects of natural/artificial selection on livestock genomes, present the main methods to detect selection signatures and discuss some recent results in this area.

Keywords: footprints of selection, domestic animals, artificial selection

INTRODUÇÃO

A recente publicação dos genomas de diversas espécies de interesse zootécnico (INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM, 2004; THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009; THE INTERNATIONAL SHEEP GENOMICS CONSORTIUM, 2010; GROENEN *et al.*, 2012; DONG *et al.*, 2013) e o desenvolvimento de técnicas estatísticas tornam possível a identificação de regiões genômicas sujeitas à seleção natural/artificial nos animais de produção. Atualmente, a identificação destas regiões é um dos principais interesses dos geneticistas evolutivos pois pode auxiliar desde o entendimento de processos básicos envolvidos na evolução dos genomas até a descoberta da função de genes/regiões genômicas (NIELSEN, 2001, 2005; SCHLÖTTERER, 2003).

Por exemplo, a identificação de uma região genômica sujeita a seleção em uma população específica pode levar à inferências iniciais sobre esta região. Além disso, a identificação de assinaturas de seleção pode ser útil na identificação de genes envolvidos na manifestação de características ecológicas difíceis de mensurar em laboratório ou ainda podem auxiliar na validação de experimentos de mapeamento de QTLs (Loci de Característica Quantitativa). O último e talvez mais ambicioso dos objetivos da análise de assinaturas de seleção é identificar mutações causais que conferem vantagem adaptativa à determinadas populações ou espécies (NIELSEN, 2001; SCHLÖTTERER, 2003; HAYES *et al.*, 2008).

Apesar da sua importância, a identificação de regiões sujeitas à seleção ainda é um tópico desafiador, pois diferentes tipos de seleção podem levar à diferentes marcas genômicas. Além disso, os efeitos da seleção podem ser confundidos com alguns fenômenos demográficos como: expansões populacionais, subdivisão e efeitos de gargalo genético (BROOKFIELD, 2001; HARRIS; MEYER, 2006; TORO; MÄKI-TANILA, 2007; OLEKSYK *et al.*, 2010).

O processo de domesticação provocou mudanças acentuadas em características comportamentais e morfológicas nas espécies atuais de animais domésticos e, juntamente com o desenvolvimento das raças atuais e dos programas de melhoramento genético, propiciou o desenvolvimento da enorme variedade de raças e tipos raciais conhecidos atualmente (DIAMOND, 2002; TORO; MÄKI-TANILA, 2007; FLORI *et*

al., 2009). Estes fatos, juntamente com o aumento do conhecimento sobre o genoma das espécies de interesse zootécnico, propiciam uma excelente oportunidade tanto para a identificação de regiões genômicas sujeitas à seleção, quanto para a validação de novas metodologias para detectar estas assinaturas de seleção (HAYES *et al.*, 2008; FLORI *et al.*, 2009).

Os objetivos desta revisão são: descrever os efeitos da seleção natural/artificial nos genomas, apresentar os principais métodos de identificação de assinaturas de seleção e, finalmente, discutir os recentes avanços neste campo em relação às espécies de interesse zootécnico.

SELEÇÃO NATURAL E ARTIFICIAL

A transferência do material genético entre gerações de uma população depende de uma série de fenótipos denominados conjuntamente de componentes de sobrevivência ou “fitness”. Do ponto de vista do genoma, a variação genética pode estar relacionada à diferença nos valores de *fitness* entre os indivíduos de uma população. A seleção natural é um fenômeno influenciado pelo ambiente no qual os indivíduos portadores de genótipos específicos possuem capacidades diferentes de contribuir com a próxima geração (FALCONER; MACKAY, 1996; TEMPLETON, 2006; DRISCOLL *et al.*, 2009).

A seleção natural pode ser classificada basicamente de três formas: seleção positiva, seleção purificadora (também conhecida como estabilizadora, negativa ou seleção de fundo) e seleção balanceadora. Cada tipo de seleção é uma resposta à pressões ambientais e possui um efeito diferente na modificação das frequências alélicas e genotípicas (HARRIS; MEYER, 2006; OLEKSYK *et al.*, 2010).

A seleção positiva ocorre quando uma mutação recém surgida possui vantagem seletiva em relação às outras mutações e, assim, aumenta de frequência na população (NIELSEN, 2005; HARRIS; MEYER, 2006; OLEKSYK *et al.*, 2010).

Na seleção purificadora, a mutação recém surgida é desvantajosa, porém tende a (re)aparecer na população através de mutações recorrentes. A tendência é que estas mutações sejam eliminadas da população, assim, este tipo de seleção tende a manter a

integridade funcional das sequências de DNA (NIELSEN, 2005; HARRIS; MEYER, 2006; OLEKSYK *et al.*, 2010).

A seleção balanceadora ocorre quando o polimorfismo é beneficiado, aumentando a variabilidade genética. Processos biológicos distintos podem estar envolvidos no aparecimento deste tipo de seleção como por exemplo: a seleção sobredominante (quando o heterozigoto possui vantagem seletiva), a seleção dependente de frequência (quando diferentes alelos apresentam vantagem/desvantagem seletiva, dependendo da sua frequência na população) e também a seleção em ambientes heterogêneos (CHARLESWORTH, 2006; HARRIS; MEYER, 2006; OLEKSYK *et al.*, 2010).

Cada tipo de seleção deixa marcas específicas nos padrões de variação genética das regiões selecionadas. Estas marcas, ou assinaturas, dependem não só do tipo de seleção mas também da idade e da intensidade do evento seletivo. Além disso, diferentes tipos de seleção podem agir na mesma região genômica durante a história evolutiva de uma população (OLEKSYK *et al.*, 2010).

A seleção artificial por sua vez, é um processo mediado pelo homem no qual a próxima geração não dependerá exclusivamente (ou necessariamente) dos componentes de *fitness*, mas também de algumas características escolhidas pelo homem. A seleção artificial pode ser classificada em seleção artificial inconsciente ou seleção artificial metodológica. A primeira ocorre sem que haja um objetivo claro e definido e é sugerido que os primeiros eventos de domesticação envolveram este tipo de seleção artificial. Já o segundo tipo ocorre quando um padrão ou objetivo é claro e definido e a escolha dos pais das próximas gerações é dependente deste objetivo. Embora existam diferenças entre seleção natural e seleção artificial, as consequências genéticas de ambos os tipos de seleção são essencialmente as mesmas (AVISE; AYALA, 2009; DRISCOLL *et al.*, 2009; GREGORY, 2009).

ASSINATURAS DE SELEÇÃO

Sob neutralidade, tanto a probabilidade de fixação quanto a probabilidade de perda de um determinado alelo é dependente, exclusivamente, de sua frequência inicial.

Para um alelo recém surgido na população (o que significa dizer que apenas há neste momento uma única cópia deste alelo na população), o tempo de fixação (se esta ocorrer) será igual a $4N_e$ gerações (KIMURA, 1986).

Como os novos alelos tendem a levar muito tempo para serem fixados e o desequilíbrio de ligação (LD) criado tende a diminuir ao longo do tempo por causa da recombinação, os alelos mais antigos tendem a ser mais comuns na população e apresentar pouco desequilíbrio de ligação ao seu redor. Já os alelos recém surgidos, tendem a ser raros e podem apresentar tanto pouco quanto muito desequilíbrio de ligação ao seu redor (SABETI *et al.*, 2002).

A teoria da neutralidade propõe que as substituições neutras tendem a ser mais frequentes do que as substituições que conferem vantagem/desvantagem seletiva. Por isso, as mutações sinônimas tendem a ocorrer a taxas maiores do que as mutações não sinônimas, uma vez que uma mutação não sinônima tem maior probabilidade de estar relacionada a adaptação do que uma mutação sinônima (KIMURA, 1991; NEI, 2005). Isso faz com que a proporção de substituições não sinônimas por sítios não sinônimos (d_N) seja equivalente à proporção de substituições sinônimas por sítios sinônimos (d_S) tanto para o polimorfismo intraespecífico como quando para a variação interespecífica (YANG, 1998; YANG; NIELSEN, 1998; NEI, 2005; HARRIS; MEYER, 2006).

A ocorrência de seleção vai criar desvios em relação aos padrões de variação molecular esperados sob neutralidade. Cada forma de seleção tende a causar mudanças específicas nos padrões de variação tanto nos *loci* selecionados quanto nos *loci* neutros ligados a eles. Estas marcas deixadas pela seleção no genoma são chamadas de assinaturas de seleção e podem ser utilizadas para identificar os *loci* sujeitos à seleção (KREITMAN, 2000).

Quando a seleção positiva ocorre, os alelos recém surgidos que possuem vantagem seletiva tendem a aumentar de frequência (até alcançarem a fixação) na população, fazendo com que alguns alelos neutros ligados também aumentem de frequência. Este fenômeno é chamado de efeito carona ou arrasto seletivo (CHARLESWORTH, 2007).

O arrasto seletivo leva a uma redução na heterozigosidade das regiões próximas ao *locus* selecionado (KAPLAN *et al.*, 1989; KIM; STEPHAN, 2002) e também a uma modificação no padrão de variação das frequências alélicas (espectro de frequências

alélicas ou SFS) ocasionada pela presença de um excesso de variantes raras na região selecionada (BRAVERMAN *et al.*, 1995; KIM; STEPHAN, 2002). A magnitude destes efeitos é dependente da intensidade de seleção e também das taxas de recombinação na região (BRAVERMAN *et al.*, 1995; KAPLAN *et al.*, 1989; CHARLESWORTH, 2007; KIM; STEPHAN, 2002).

O aumento do LD, ocasionando a presença de haplótipos mais longos, também é esperado na região ao redor do *locus* selecionado (KIM; STEPHAN, 2002). O aumento do LD tende a ocorrer nas regiões próximas ao *locus* selecionado (mas não imediatamente adjacentes a este *locus*) juntamente com o aumento de frequência de alelos derivados (PRZEWORSKI, 2002; KIM; NIELSEN, 2004; MCVEAN, 2007). Porém o LD entre duas regiões separadas por um *locus* selecionado tendem a apresentar-se em níveis basais. (KIM; NIELSEN, 2004; STEPHAN *et al.*, 2006; MCVEAN, 2007).

Logo após a fixação do alelo selecionado, o LD tende a declinar e, portanto, as assinaturas de seleção baseadas no LD tendem a ser rapidamente “apagadas” (PRZEWORSKI, 2002; KIM; NIELSEN, 2004; MCVEAN, 2007). Portanto, um alelo derivado presente em alta frequência em uma população cercado por haplótipos longos com LD alto são indícios de eventos seletivos recentes (SABETI *et al.*, 2002; VOIGHT *et al.*, 2006).

Além disso, a seleção positiva tende provocar na região ao redor do *locus* selecionado, uma diminuição dos níveis de diversidade intra-populacional e o aumento dos níveis de diversidade entre as populações, levando a valores de F_{ST} mais altos do que o esperado pela teoria neutra (BEAUMONT, 2005; STORZ, 2005). Também tende a ser observada uma maior proporção de substituições não sinônimas por sítios não sinônimos (NEI, 2005; HARRIS; MEYER, 2006).

O modelo de arrasto seletivo no qual o alelo recém surgido apresenta uma vantagem seletiva grande e aumenta de frequência rapidamente até atingir a fixação é também conhecido como arrasto seletivo “forte”. Por outro lado, é possível que um alelo já presente na população em algum momento aumente de frequência por causa de uma vantagem seletiva recém adquirida. Quando isso acontece, pode-se dizer que o arrasto seletivo tende a ser do tipo “brando” e as marcas deixadas por ele tendem a ser menos claras (PRZEWORSKI *et al.*, 2005; PRITCHARD *et al.*, 2010).

As marcas deixadas pela seleção podem ser influenciadas por muitos outros fatores. Por exemplo, a idade dos alelos neutros ligados ao alelo selecionado influencia os padrões de LD. Se os alelos neutros surgirem antes do alelo selecionado, o padrão de LD tende a ser semelhante ao descrito anteriormente, porém se os alelos neutros surgirem depois do alelo selecionado, os padrões de LD tendem a ser extremamente complexos (MCVEAN, 2007).

A seleção balanceadora atua favorecendo a manutenção do polimorfismo (HARRIS; MEYER, 2006; OLEKSYK *et al.*, 2010). A persistência dos mesmos alelos na população por longos períodos evolutivos é chamada seleção balanceadora de longo prazo e, além de manter o polimorfismo no *locus* selecionado, também aumenta a diversidade em regiões neutras ligadas (CHARLESWORTH, 2006). Durante o período evolutivo, mutações neutras podem surgir no contexto dos alelos selecionados e as regiões selecionadas podem ter um aumento tanto de polimorfismo quanto do LD (NAVARRO; BARTON 2002; CHARLESWORTH *et al.*, 2003; CHARLESWORTH, 2006).

Na seleção balanceadora de longo prazo, os níveis de diversidade dentro das populações tendem a aumentar e os níveis de diversidade entre as populações tendem a diminuir para as regiões selecionadas (NAVARRO; BARTON 2002; CHARLESWORTH *et al.*, 2003; CHARLESWORTH, 2006), o que leva à diminuição dos valores de F_{ST} (BEAUMONT, 2005; STORZ, 2005). Porém em alguns casos, pode-se observar valores de F_{ST} maiores do que o esperado sob neutralidade (BEAUMONT, 2005; CHARLESWORTH, 2006).

Quando a seleção negativa ocorre, os alelos recém surgidos tendem a ser extremamente desvantajosos e, portanto, são removidos da população juntamente com a variação genética ligada a eles (INNAN; STEPHAN, 2003; HARRIS; MEYER, 2006).

Se a taxa de recombinação na região selecionada é pequena ou a população é endogâmica, a seleção negativa reduz a variabilidade ao redor dos *loci* selecionados (CHARLESWORTH *et al.*, 1993, 1995; ANDOLFATTO, 2001; STEPHAN, 2010). Um excesso de alelos com baixa frequência também é observado em populações de tamanho pequeno a moderado (CHARLESWORTH *et al.*, 1993, 1995) e a proporção de substituições não sinônimas por sítios não sinônimos tende a ser menor do que a

proporção de substituições sinônimas por sítios sinônimos (NEI, 2005; HARRIS; MEYER, 2006).

Porém, se as regiões selecionadas apresentarem taxas de recombinações consideráveis ou a população não for endogâmica, a redução na variabilidade não será observada na região ao redor do *locus* selecionado (CHARLESWORTH *et al.*, 1993, 1995; ANDOLFATTO, 2001; STEPHAN, 2010). Além disso, a seleção negativa não leva a uma mudança considerável no SFS (CHARLESWORTH *et al.*, 1993, 1995; KIM; STEPHAN, 2000; ANDOLFATTO, 2001; STEPHAN, 2010).

O reconhecimento das marcas deixadas pelos diferentes tipos de seleção é uma tarefa crucial nos estudos visando identificar regiões genômicas que possam estar sendo selecionadas e a teoria neutra serve como um modelo importante no desenvolvimento dos testes estatísticos desenvolvidos para detectar assinaturas de seleção. Porém, em populações reais, muitos dos pressupostos da teoria neutra são comumente violados (por exemplo a presença de expansões populacionais, subdivisão e efeitos de gargalo genético), o que pode levar ao aparecimento de padrões que mimetizam o efeito da seleção nos genomas. Além disso, a interação entre os diferentes tipos de seleção e a interação entre a seleção e fatores demográficos podem dificultar a identificação de regiões genômicas sujeitas à seleção (BARTON, 1998; KIM; STEPHAN, 2000; KREITMAN, 2000; CHARLESWORTH *et al.*, 2003; HARRIS; MEYER, 2006; TORO; MÄKI-TANILA, 2007).

MÉTODOS DE IDENTIFICAÇÃO DE ASSINATURAS DE SELEÇÃO

Os métodos propostos para a identificação de assinaturas de seleção podem ser classificados de diferentes formas (HARRIS; MEYER, 2006; OLEKSYK *et al.*, 2010). Considerando os principais fatores que afetam as marcas moleculares deixadas pela seleção, Hohenlohe *et al.* (2010) propuseram a utilização de uma árvore para auxiliar a definir a(s) metodologia(s) adequada(s) para cada caso. Primariamente, deve-se levar em consideração o tempo evolutivo envolvido com o processo seletivo, porém outros fatores (como o número de populações estudado, o tipo de seleção e etc...) também

devem ser levados em consideração na escolha do método a ser utilizado para identificar assinaturas de seleção

Testes baseados nas taxas de substituições sinônimas e não sinônimas

Quando sequências de genes ortólogos são comparadas, é esperado que sob neutralidade $d_N/d_S = 1$. Quando a seleção positiva ocorre, $d_N/d_S > 1$ e quando há presença de seleção negativa, $d_N/d_S < 1$ (YANG, 1998).

Existem vários métodos para estimar d_N e d_S (NEI, 2005). Inicialmente, estes métodos eram baseados em aproximações baseadas na comparação entre duas sequências (NEI; GOJOBORI, 1986). Mais recentemente, estimativas baseadas em máxima verossimilhança têm sido propostas e podem acomodar fatores como: viés nas taxas de transições/transversões, utilização preferencial de códons, restrições seletivas (GOLDMAN; YANG, 1994) e d_N/d_S variáveis entre sítios e entre linhagens (NIELSEN; YANG, 1998; YANG *et al.*, 2000; YANG, 2002; YANG; NIELSEN, 2002; O'BRIEN *et al.*, 2009).

Porém, quando o tempo de divergência entre as linhagens estudadas é pequeno e existe uma quantidade de dados considerável para a análise, os diversos métodos disponíveis para estimar d_N e d_S tendem a apresentar resultados semelhantes (NEI, 2005, POND; FROST, 2005).

O teste de hipóteses pode ser feito através de um teste de razão de verossimilhanças, comparando o modelo que assume neutralidade com modelos alternativos (YANG, 1998; YANG; NIELSEN, 1998; YANG *et al.*, 2000, 2005). Pacotes computacionais como o MEGA (TAMURA *et al.*, 2007) e o PAML (YANG, 2007) incluem testes de seleção baseados em d_N/d_S .

Testes baseados no espectro de frequências alélicas

O parâmetro θ pode ser estimado de diferentes maneiras à partir de sequências de DNA. A comparação entre diferentes estimativas de θ é a base para vários testes desenvolvidos para identificar assinaturas de seleção (TAJIMA, 1989; FU; LI, 1993; FU, 1996, 1997).

Tajima (1989) propôs um teste baseado na diferença entre $\hat{\theta}_\pi$ (o número médio de diferenças de nucleotídeos) e $\hat{\theta}_S$ (o número de sítios segregantes ao longo de uma

sequência de DNA) pelo fato de a seleção alterar os valores estimados de $\hat{\theta}_\pi$ enquanto as estimativas de $\hat{\theta}_S$ tendem a permanecer inalteradas (TAJIMA, 1989; HARTL; CLARK, 2010).

O teste proposto é conhecido como Tajima D e corresponde à diferença padronizada entre $\hat{\theta}_\pi$ e $\hat{\theta}_S$ (TAJIMA, 1989; HARRIS; MEYER, 2006). Sob neutralidade, o valor de D tende a zero. A presença de seleção positiva ou negativa levam a uma redução na heterozigiosidade e um aumento no número de alelos raros ao redor do *locus* selecionado, fazendo com que $D < 0$ (CHARLESWORTH *et al.*, 1993, 1995; BRAVERMAN *et al.*, 1995; KAPLAN *et al.*, 1989; TAJIMA, 1989; ANDOLFATTO, 2001; KIM; STEPHAN, 2002; STEPHAN, 2010). Ao contrário, a seleção balanceadora de longo prazo leva a um aumento na diversidade ao redor do *locus* selecionado, gerando $D > 0$ (TAJIMA, 1989; NAVARRO; BARTON, 2002; CHARLESWORTH, 2006).

Muitos testes para identificar regiões genômicas sob seleção têm sido desenvolvidos com base na presença de excesso de alelos raros (FU; LI, 1993; Fu, 1996, 1997). Porém, algumas vezes estes testes não possuem uma fácil interpretação, uma vez que muitas vezes não é possível diferenciar a presença de diferentes tipos de seleção (TAJIMA, 1989; HARRIS; MEYER, 2006) e também pelo fato de que fatores demográficos podem influenciar o resultado destes testes (TAJIMA, 1989; CHARLESWORTH *et al.*, 1993; FU; LI, 1993; FU, 1996, 1997).

Enquanto a redução na heterozigiosidade e a presença de um excesso de alelos raros não podem ser tidos como características específicas da seleção, o excesso de alelos derivados é tido como uma característica específica da seleção positiva (FAY; WU, 2000).

Por causa disso, Fay e Wu (2000) propuseram a estatística H, calculada como a diferença entre $\hat{\theta}_\pi$ e $\hat{\theta}_H$ (onde $\hat{\theta}_H$ é um estimador de θ padronizado pela homozigiosidade dos alelos). A seleção positiva (mas não a negativa) leva a um aumento de $\hat{\theta}_H$ por causa do excesso de alelos derivados, fazendo com que $H < 0$. Portanto, a estatística proposta por Fay e Wu (2000) possibilita distinguir entre seleção positiva e negativa.

A redução da variabilidade causada pela seleção positiva tende a ser “quebrada” pela recombinação. Portanto, a presença de valores de heterozigiosidade reduzida são

sugeridos como sendo marcas deixadas por eventos seletivos recentes. A profundidade e extensão destes vales podem ser influenciadas por vários fatores como a intensidade da seleção, as taxas de recombinação e o tamanho efetivo da população (KIM; STEPHAN, 2002).

Por causa disso, Kim e Stephan (2002) desenvolveram uma metodologia baseada em verossimilhança para identificar eventos de seleção positiva na presença de recombinação. O teste proposto é baseado na esperança do número de sítios onde os alelos derivados estão em uma determinada frequência em uma população.

Recentemente, extensões a esta metodologia vêm sendo propostas. Estes novos métodos podem ser utilizados em dados genômicos, inclusive na presença de viés de descoberta dos marcadores moleculares utilizados (NIELSEN *et al.*, 2005; KELLEY *et al.*, 2006; WILLIAMSON *et al.*, 2007).

Testes baseados no desequilíbrio de ligação

A exploração dos padrões de LD é o foco de diversos testes para identificar assinaturas de seleção (SABETI *et al.*, 2002, 2007; VOIGHT *et al.* 2006; KIM; NIELSEN 2004; KIMURA *et al.*, 2007). Porém, estas assinaturas tendem a ser transitórias, uma vez que a recombinação tende a, rapidamente, diminuir o LD uma vez que o alelo vantajoso tenha sido fixado (PRZEWORSKI, 2002; KIM; NIELSEN, 2004; MCVEAN, 2007).

Sabeti *et al.* (2002) propuseram uma metodologia conhecida como teste do haplótipo estendido (LRH), baseada na relação entre a frequência de um alelo e o desequilíbrio de ligação ao redor dele. O teste inicia-se com a identificação dos chamados haplótipos do cerne. Então, o desequilíbrio de ligação entre o haplótipo do cerne e SNPs localizados a distâncias crescentes deste haplótipo é medido para se avaliar o decaimento do LD de acordo com a distância (SABETI *et al.*, 2002).

A mensuração do decaimento do desequilíbrio de ligação é feito através do cálculo do EHH, que pode ser definido como a probabilidade de dois cromossomos em uma população que contém um determinado haplótipo do cerne serem homozigotos para toda a região desde o haplótipo do cerne até uma distância x deste haplótipo (SABETI *et al.*, 2002).

Então o valor de EHH relativo (REHH) é calculado para comparar o decaimento do EHH para um determinado haplótipo do cerne em comparação aos outros haplótipos do cerne combinados. O teste para a presença de seleção consiste em comparar os valores de REHH (de acordo com a frequência do haplótipo do cerne referente) com os valores de REHH dos outros haplótipos do cerne. Sugere-se a presença de seleção positiva quando um determinado haplótipo do cerne possui uma combinação de alto valor de REHH e alta frequência na população (SABETI *et al.*, 2002).

O teste de LRH foi estendido por Voight *et al.* (2006). Este novo teste foi desenvolvido para varreduras genômicas utilizando chips densos de SNPs e é denominado de iHS. O valor de iHS pode ser definido como uma medida de quão incomum é o desequilíbrio de ligação ao redor de um determinado SNP quando comparado com o resto do genoma (VOIGHT *et al.*, 2006).

A metodologia proposta por Voight *et al.* (2006) inicia-se com o cálculo do EHH para cada SNP. Como os SNPs são marcadores bialélicos, pode-se assumir que um alelo será o ancestral e o outro o derivado. Portanto, a área sob a curva de um gráfico EHH x distância para cada alelo de um SNP é calculada e denominada EHH integrado (iHH). Como os alelos podem ser ancestrais ou derivados, os valores de iHH são denominados iHHA e iHHD, respectivamente. O valor final da estatística (denominada iHS) é uma padronização da razão entre iHHA e iHHD para possibilitar a comparação entre SNPs com frequências alélicas diferentes (VOIGHT *et al.*, 2006).

Tanto o LRH quanto o iHS possuem baixo poder de identificar eventos seletivos cujo alelo vantajoso já esteja fixado na população. Para lidar com esta situação, testes derivados do LRH (baseados na comparação entre duas populações) têm sido desenvolvidos (KIMURA *et al.*, 2007; SABETI *et al.*, 2007; TANG *et al.*, 2007).

Dentre estas metodologias, podem ser citadas o XP-EHH (SABETI *et al.*, 2007) e o $\ln(R_{sb})$ (TANG *et al.*, 2007), ambas baseadas na comparação entre as estimativas de EHH entre duas populações. A diferença básica entre estas duas metodologias é a forma de calcular o valor de EHH (SABETI *et al.*, 2007; TANG *et al.*, 2007).

Testes baseados na diferenciação entre populações

A base dos testes baseados na diferenciação entre populações é a estimativa de F_{st} para múltiplos *loci* e a comparação destas estimativas com as expectativas sob

neutralidade (LEWONTIN; KRAKAUER, 1973; BOWCOCK *et al.*, 1991; VITALIS *et al.*, 2001, 2003; AKEY *et al.*, 2002; BEAUMONT; BALDING, 2004; FOLL; GAGGIOTTI, 2008; EXCOFFIER *et al.*, 2009; BONHOMME *et al.*, 2010).

O primeiro teste baseado na diferenciação de populações foi proposto por Lewontin e Krakauer (1973). A idéia básica consistia no fato de que os valores de F_{st} estimados para vários *loci* sob neutralidade deveriam apresentar uma pequena heterogeneidade, porém, na presença de seleção, as estimativas de F_{st} deveriam apresentar uma variação grande. Portanto, a metodologia proposta era baseada na comparação entre a variância dos valores de F_{st} estimados e a variância esperada para os valores de F_{st} sob neutralidade (LEWONTIN; KRAKAUER, 1973).

Porém, tão logo quanto a idéia foi publicada, o teste de Lewontin & Krakauer foi duramente criticado por causa dos pressupostos assumidos para estimar a variância dos valores de F_{st} sob neutralidade (NEI; MARUYAMA, 1975; ROBERTSON, 1975). Para evitar estes inconvenientes, foi proposto que poderia-se gerar distribuições de F_{st} levando em consideração a história evolutiva das populações estudadas através de simulações computacionais (BOWCOCK *et al.*, 1991).

Recentemente, têm sido desenvolvidas metodologias para estimar distribuições de F_{st} que são robustas contra a quebra de pressupostos importantes (divergência e expansão populacional recentes, isolamento pela distância, fluxo gênico heterogêneo entre as populações) (BEAUMONT; NICHOLS, 1996; BEAUMONT; BALDING, 2004; FOLL; GAGGIOTTI, 2008; EXCOFFIER *et al.*, 2009). Estas metodologias estão disponíveis em programas computacionais gratuitos como o BayesFst (BEAUMONT; BALDING, 2004), BayeScan (FOLL; GAGGIOTTI, 2008) e o Arlequin (EXCOFFIER *et al.*, 2009).

Os métodos propostos por Beaumont e Nichols (1996) e Excoffier *et al.* (2009) tendem a ser bastante rápidos, porém a presença de alguns fatores demográficos pode levar a importantes vieses nas estimativas. Por outro lado, as metodologias baseadas em cadeias de Markov e Monte Carlo (MCMC) tendem a ser muito mais lentas mas são mais robustas na presença de desvios dos pressupostos do modelo assumido (BEAUMONT; BALDING, 2004; FOLL; GAGGIOTTI, 2008).

Outra forma de prevenir os efeitos de fatores demográficos é realizar comparações duas a duas entre as populações (TSAKAS; KRIMBAS, 1976). Baseados

nesta idéia, Vitalis *et al.* (2001) propuseram a utilização de um modelo simples de divergência entre as duas populações para estimar os valores de F_{st} e construir os intervalos de confiança destas estimativas.

Esta metodologia tem demonstrado ser bastante robusta, porém a utilização de apenas duas populações pode reduzir o poder do teste (TSAKAS; KRIMBAS, 1976; VITALIS *et al.*, 2001). Este método está implementado no software DetSel 1.0 (VITALIS *et al.*, 2003).

Todas as metodologias descritas anteriormente estão baseadas na geração de distribuições nulas de F_{st} . A presença de quebra de pressupostos nos modelos escolhidos para a geração das distribuições nulas pode levar ao aumento do número de falsos positivos. Atualmente, o número de marcadores genotipados em uma população pode ser extremamente grande e a distribuição genômica dos valores de F_{st} pode ser utilizada para construir o teste estatístico (AKEY *et al.*, 2002).

Atualmente, várias metodologias para identificar as marcas deixadas pela seleção nos genomas estão descritas e cada uma delas tem a capacidade de capturar padrões específicos da variação molecular. A utilização combinada de mais de uma metodologia é uma estratégia interessante e tem sido sugerida por diversos autores. Apesar disso, a identificação de uma determinada região por uma metodologia e não por outra não exclui a possível importância daquela região para o estudo, principalmente pelo fato de que as diferentes metodologias podem capturar sinais diferentes deixados pela seleção nos genomas (HOHENLOHE, *et al.*, 2010; OLEKSYK *et al.*, 2010).

ASSINATURAS DE SELEÇÃO EM ANIMAIS DE INTERESSE ZOOTÉCNICO

O processo de domesticação conduziu a mudanças morfológicas e comportamentais consideráveis nas espécies de interesse zootécnico. Acredita-se que nos estágios iniciais da domesticação, a seleção artificial inconsciente para características comportamentais foi importante. Este estágio inicial foi sucedido por seleção artificial metodológica na qual características especiais foram selecionadas baseadas em objetivos definidos. Como consequência, as espécies de animais domésticas tendem a possuir cérebros menores, órgãos dos sentidos menos

desenvolvidos e apresentam diferenças comportamentais e em relação à resistência a doenças quando comparados com as espécies selvagens. Mudanças no tamanho corporal, coloração de pelagem e presença/ ausência de atributos específicos (por exemplo chifres) nas raças modernas estão relacionadas com o processo de domesticação (DIAMOND, 2002; TORO; MÄKI-TANILA, 2007; GAUTIER *et al.*, 2009; GREGORY, 2009; TRUT *et al.*, 2009).

O desenvolvimento de raças especializadas, melhoradas para o aumento de produção de um determinado produto ou para atingir um determinado padrão morfológico, aumentou as diferenças entre os animais domésticos e seus parentes selvagens e também propiciou a criação de uma enorme quantidade de populações diferentes, cada uma com características específicas. Algumas destas características são quantitativas, ou seja, controladas pela interação entre muitos genes de efeito menor. Isso gera uma oportunidade para o entendimento da base molecular destas características, principalmente pelo fato de que a maioria das características de interesse econômico nos animais domésticos é quantitativa (ANDERSSON; GEORGES, 2004).

A seleção para uma característica específica pode levar a mudanças em outras características por causa da presença de correlações genéticas. Um excelente exemplo disso é a diminuição da fertilidade e da resistência contra patógenos observada em algumas raças melhoradas de bovinos leiteiros (BROTHERSTONE; GODDARD, 2005; OLTENACU; BROOM, 2010).

O advento dos marcadores moleculares veio com a promessa de se descobrir a base molecular das características quantitativas nos animais de produção. Porém, até recentemente, a quantidade de informações genômicas disponíveis para estas espécies era pequena quando comparada com a espécie humana. A publicação recente do genoma das principais espécies de interesse zootécnico e a diminuição dos custos de genotipagem têm propiciado uma explosão no conhecimento de genes e redes de interação entre genes envolvidas na manifestação de características quantitativas (ANDERSSON; GEORGES, 2004; HAYES *et al.*, 2010).

A identificação de genes sujeitos à seleção nestas espécies pode auxiliar a encontrar mutações causais em regiões previamente identificadas em experimentos de mapeamento de QTLs e podem revelar genes importantes para características adaptativas (por exemplo genes relacionados com adaptação a ambientes tropicais).

Além disso, estes estudos podem auxiliar a identificação de genes selecionados de forma divergente em diferentes populações e genes envolvidos com o aparecimento de correlações genéticas, além de genes relacionados com o processo de domesticação (SCHLÖTTERER, 2003; HAYES *et al.*, 2008; OJEDA *et al.*, 2008; FLORI *et al.*, 2009; MACEACHERN *et al.*, 2009).

Assinaturas de seleção relacionadas com a domesticação e com o desenvolvimento inicial das raças

Em algumas espécies selvagens, a expressão tanto de eumelanina quanto de feomelanina está relacionada com a produção de pelagens camufladas. Durante o processo de domesticação, os padrões de pelagem não camuflados foram selecionados (FANG *et al.*, 2009; WIENER; WILKINSON, 2011).

Nas espécies de animais domésticos, os padrões de pelagem são extremamente importantes na caracterização racial. Portanto, as sociedades de registro de raças geralmente exigem padrões de pelagem específicos para raças específicas. Fato que levou à fixação (ou quase fixação) de alelos específicos em muitas raças.

A determinação dos padrões de pelagem envolve a interação entre muitos genes. O gene para o Receptor para o Hormônio Estimulador de Melanócitos (MC1R) influencia a produção dos pigmentos eumelanina e feomelanina (Werth *et al.*, 1996; KIJAS *et al.*, 1998; FANG *et al.*, 2009; LI *et al.*, 2010b; WIENER; WILKINSON, 2011).

Em bovinos, pelo menos três alelos do gene MC1R são descritos, o alelo E^D está relacionado com a coloração preta dominante, o alelo E^+ está relacionado com o aparecimento da pelagem selvagem (um padrão rajado formado pela combinação de pelos vermelhos e marrons) e o alelo e , responsável pela pelagem vermelha (WIENER; WILKINSON, 2011).

Em suínos, o alelo selvagem E^+ está relacionado tanto com a expressão de eumelanina quanto com a expressão de feomelanina, resultando em indivíduos cuja pelagem apresentam tons variáveis de marrom (incluindo a presença de listas em animais jovens). O fenótipo dominante está relacionado com a presença de dois alelos que parecem ter evoluído de forma independente em animais de origem asiática e europeia (E^{D1} e E^{D2}). O alelo E^P está relacionado com o aparecimento do fenótipo

manchado (manchas pretas em fundo vermelho ou branco) e o alelo recessivo e é responsável pela pelagem vermelha (KIJAS *et al.*, 1998; FANG *et al.*, 2009; LI *et al.*, 2010b).

Nas espécies selvagens de suínos o alelo selvagem do MC1R (E^+), e consequentemente o fenótipo camuflado, parecem estar sendo mantidos por seleção negativa. Por outro lado, tanto em bovinos (FLORI *et al.*, 2009; STELA *et al.*, 2010) quanto em suínos (FANG *et al.*, 2009; LI *et al.*, 2010b; AMARAL *et al.*, 2011) domésticos, diferentes alelos do MC1R parecem estar sob seleção em raças específicas.

Outros genes relacionados com os padrões de pelagem são sugeridos como alvos de seleção nas espécies de animais domésticos. Assinaturas de seleção próximas ao gene KIT foram demonstradas em bovinos (STELA *et al.*, 2010; WIENER *et al.*, 2011), suínos (FONTANESI *et al.*, 2010; AMARAL *et al.*, 2011) e ovinos (KIJAS *et al.*, 2012). Também, o gene PMEL17 é sugerido como alvo de seleção em algumas raças bovinas (GAUTIER *et al.*, 2009; WIENER *et al.*, 2011).

Outra característica importante na definição de raças em algumas espécies é a presença/ausência de chifres. Em ovinos selvagens, a presença de chifres e a morfologia destes estão relacionadas com o “fitness” tanto em machos quanto em fêmeas. O mecanismo que mantém o polimorfismo em ovinos selvagens de Soay é complexo, envolvendo seleção positiva (antagônica) entre os sexos e também a presença de influência ambiental variável (JOHNSTON *et al.*, 2011).

Recentemente, o gene do Receptor 2 para Relaxina (RXFP2), localizado no cromossomo OAR10, foi associado tanto com a presença/ausência de chifres quanto com a morfologia destes em ovinos selvagens de Soay (JOHNSTON *et al.*, 2011), e um SNP próximo a este gene foi identificado como contendo um sinal forte de seleção em uma análise envolvendo 74 raças de ovinos (KIJAS *et al.*, 2012). Em bovinos, a região próxima ao *locus* mocho foi identificada também como alvo de seleção (DRÖGEMÜLLER *et al.*, 2005; LI *et al.*, 2010a; STELA *et al.*, 2010).

Mudanças comportamentais envolvidas com uma maior interação das espécies domésticas com os humanos também são creditadas à domesticação (DIAMOND *et al.*, 2002; AMARAL *et al.*, 2011; WIENER; WILKINSON, 2011). Como consequência, vários estudos sugerem a presença de assinaturas de seleção próximas a genes

envolvidos com o desenvolvimento e função do sistema nervoso (THE BOVINE HAPMAP, 2009; GAUTIER *et al.*, 2009; STELA *et al.*, 2010; AMARAL *et al.*, 2011).

Bovinos

As raças modernas de bovinos podem ser classificadas basicamente em dois grandes grupos, os taurinos e os indicinos. Dentro de cada grupo, diversas raças foram desenvolvidas com uma grande variação em aptidão produtiva, padrões morfológicos e características adaptativas (THE BOVINE HAPMAP, 2009).

A produção de leite foi a primeira característica importante incluída nos programas de melhoramento de bovinos leiteiros. Posteriormente, características de qualidade do leite (por exemplo porcentagem de proteína e gordura) foram incorporadas nos índices de seleção. Devido à presença de correlação genética, o aumento na produção de leite levou ao aumento das exigências nutricionais, diminuição em algumas características de saúde (por exemplo resistência contra mastite e laminite) e também à diminuição da fertilidade e longevidade em animais com alta produção leiteira (BROTHERSTONE; GODDARD, 2005; OLTENACU; BROOM, 2010). Como consequência, recentemente a indústria de bovinos leiteiros tem modificado os objetivos de seleção para acomodar estas características relacionadas à fertilidade e saúde, juntamente com produção e qualidade do leite (BROTHERSTONE; GODDARD, 2005).

Os programas de melhoramento genético em bovinos de corte têm focado principalmente em características relacionadas com ganho de peso e os ganhos genéticos observados são consideráveis (MIDDLETON; GIBB, 1991; BERTRAND *et al.*, 2001; GARRICK; GOLDEN, 2009; MILLER, 2010; GARRICK, 2011). Também pode ser observado ganho genético para características de qualidade de carcaça, tanto pelo fato de haver correlação genética favorável entre estas características e características de crescimento como pelo fato de alguns programas de melhoramento genético incluírem de forma direta ou indireta estas características como critérios de seleção (BERTRAND *et al.*, 2001; GARRICK; GOLDEN, 2009; MILLER, 2010; GARRICK, 2011).

Porém, animais com maior potencial de ganho de peso também tendem a apresentar maiores peso adulto e exigência nutricional, o que pode levar a uma

diminuição na rentabilidade (GARRICK, 2011). Também, pode ser observado que os ganhos genéticos para características de qualidade de carne (por exemplo maciez de carne), eficiência alimentar e performance reprodutiva ainda não são satisfatórios, o que pode ser devido alguns fatores como: dificuldade de mensuração, expressão tardia e baixa herdabilidade. O advento da seleção genômica certamente irá otimizar o potencial de ganho genético para estas características (MILLER, 2010; GARRICK; GOLDEN, 2009; GARRICK, 2011).

Diversos estudos de varredura genômica utilizando diferentes metodologias e diferentes raças têm sido realizados com o objetivo de se identificar assinaturas de seleção em bovinos (PRASAD *et al.*, 2008; HAYES *et al.*, 2009; THE BOVINE HAPMAP, 2009; FLORI *et al.*, 2009; BARENDSE *et al.*, 2009; MACEACHERN *et al.*, 2009; GAUTIER *et al.*, 2009; LI *et al.*, 2010a; STELA *et al.*, 2010; QANBARI *et al.*, 2010; HOSOKAWA *et al.*, 2011; QANBARI *et al.*, 2011; WIENER *et al.*, 2011).

Alguns trabalhos em bovinos de corte identificaram assinaturas de seleção na região centromérica do cromossomo BTA14 (HAYES *et al.*, 2009; THE BOVINE HAPMAP, 2009; WIENER *et al.*, 2011). Esta região está envolvida com deposição de gordura subcutânea e intramuscular e, genes como a Tireoglobulina (TG), o Fator de Desenvolvimento e Diferenciação 1 (DDEF1 também conhecido como ASAP1) e o DGAT1 são sugeridos como candidatos a influenciar a variação dessas características (BARENDSE, 1999; MOORE *et al.*, 2003; THALLER *et al.*, 2003; CASAS *et al.*, 2005; PANNIER *et al.*, 2010; VENERONI *et al.*, 2010). Um aumento na proporção de gordura intramuscular em animais da raça Angus da Austrália nos últimos anos associado com o efeito de marcadores nesta região em relação à característica corroboram os achados aqui descritos (HAYES *et al.*, 2009).

O produto do gene do Fator de Crescimento e Diferenciação 8 (GDF-8), também conhecido como Miostatina, é um regulador do desenvolvimento muscular. Em bovinos, pelo menos duas mutações de perda de função são conhecidas e associadas com o fenótipo musculatura dupla; uma deleção de 11bp é responsável pelo fenótipo nas raças Belgian Blue, South Devon e Asturiana e a transição de A para G (que resulta na mudança de uma cisteína para uma tirosina) é responsável pelo fenótipo nas raças Piemontês e Gasconne (BELLINGE *et al.*, 2005; WIENER; WILKINSON, 2011).

O fenótipo musculatura dupla tem sido selecionado em alguns raças de bovinos de corte e, portanto, a região próxima ao gene GDF-8 deve apresentar marcas de seleção possíveis de serem identificadas. Um estudo inicial utilizando microssatélites não conseguiu apresentar evidências conclusivas de assinaturas de seleção nesta região, porém, recentemente, alguns estudos têm demonstrado que esta região é candidata a estar sujeita à seleção em raças que apresentam musculatura dupla (THE BOVINE HAPMAP, 2009; WIENER; GUTIERREZ-Gill, 2009; WIENER; WILKINSON, 2011).

A região mediana do cromossomo BTA2 foi identificada como candidata a conter genes selecionados (BARENDSE *et al.*, 2009; THE BOVINE HAPMAP, 2009; QANBARI *et al.*, 2011). Esta região está associada com eficiência alimentar e deposição de gordura intramuscular em raças de bovinos de corte (BARENDSE *et al.*, 2007, 2009) e contém os genes R3HDM1 e ZRANB3, recentemente relacionados com eficiência alimentar (BARENDSE *et al.*, 2009; THE BOVINE HAPMAP, 2009). É interessante notar também que esta mesma região foi identificada como sujeita a seleção em humanos (SABETI *et al.*, 2007).

O cromossomo BTA6 contém pelo menos três QTLs que afetam características de produção de leite. A região entre 37.2-38.2 Mb contém um QTL que afeta as porcentagens de proteína e gordura no leite, a região 44.1-56.3 Mb contém um QTL que influencia a produção de gordura e proteína no leite e a região 88.1-88.8Mb contém um QTL que afeta produção de leite e proteína no leite (KHATKAR *et al.*, 2004; OGOREVC *et al.*, 2009; WEIKARD *et al.*, 2012). Estas regiões foram identificadas como alvos de seleção em bovinos de raças leiteiras (HAYES *et al.*, 2008; Barendse *et al.*, 2009; THE BOVINE HAPMAP, 2009; QANBARI *et al.*, 2010; SCHWARZENBACHER *et al.*, 2012).

A primeira região contém o gene ABCG2, no qual uma transversão de A para C (que leva a mudança de tirosina para serina na região extracelular da proteína) é proposta como sendo a mutação causal do QTL. O alelo A está relacionado com diminuição da produção de leite enquanto leva a um aumento das porcentagens de gordura e proteína em bovinos da raça Holandesa de vários países e em bovinos da raça Vermelha Norueguesa (COHEN-ZINDER *et al.*, 2005; OLSEN *et al.*, 2007; COLE *et al.*, 2009; WEIKARD *et al.*, 2012).

Este fato é corroborado pelo relato de que a frequência do alelo A diminuiu de 0,75 para 0,62 entre os anos de 1982 e 1990 na população de bovinos da raça Holandesa em Israel. Porém, no período entre 1990 e 2002, a frequência do alelo A aumentou para 0,77. Esta mudança nas frequências alélicas corresponde com a modificação no índice de seleção para a raça naquele país que, inicialmente, apenas direcionava a seleção para o aumento da produção de leite. Desde o ano de 1990, o índice de seleção passou a incluir porcentagem de proteína e a produção de leite passou a ter um efeito negativo no índice (COHEN-ZINDER *et al.*, 2005). Além disso, na Vermelha Norueguesa, que possui um histórico de seleção baseado em aumento da produção de leite basicamente, o decaimento do EHH para o alelo C mostrou-se menor do que o decaimento do EHH para o alelo A, sugerindo um evento seletivo recente em favor do alelo C nesta raça (HAYES *et al.*, 2008).

Apesar da associação entre o gene ABCG2 e o QTL para porcentagem de gordura e proteína no leite, o alelo A parece estar fixado (ou em frequência bastante elevada) em várias raças de bovinos leiteiros, o que implica que o ganho genético esperado com a seleção deste alelo seja limitado (RON *et al.*, 2006; TANTIA *et al.*, 2006; WEIKARD *et al.*, 2012). O gene SPP1, também conhecido como Osteopontina, é outro candidato localizado nesta região, porém nenhuma mutação neste gene é candidata a ser a mutação causal do QTL (LEONARD *et al.*, 2005; SCHNABEL *et al.*, 2005a; DE KONIG, 2006; WEIKARD *et al.*, 2012).

Esta mesma região também contém QTLs para peso e tamanho ao nascimento, probabilidade de distocia e composição de proteína e gordura na carcaça (GUTIERREZ-GIL *et al.*, 2009; OLSEN *et al.*, 2008, 2010). Diversos outros genes estão contidos nesta região como: IBSP, MEPE, LAP3, NCAPG, LCORL, MED28 e SPP1. Os genes SPP1, IBSP e MEPE estão envolvidos com a morfogênese de ossos e cartilagens e tanto o SPP1 quanto o IBSP são expressos no tecido ósseo fetal (OLSEN *et al.*, 2008, 2010).

O gene NCAPG está associado com o crescimento fetal e pós-natal e uma transversoão de T para G (que leva à troca de isoleucina para metionina na proteína) é sugerida como a mutação causal. O alelo G está associado com o aumento e o alelo T com a diminuição no peso ao nascimento, provavelmente pela expressão diferencial entre alelos na parte fetal dos placentomas (EBERLEIN *et al.*, 2009; WEIKARD *et al.*,

2010). Também foi demonstrado que este gene está envolvido na via metabólica da arginina, que é um aminoácido essencial para o desenvolvimento pré-natal (WEIKARD *et al.*, 2010).

Portanto, considerando que o gene NCAPG é um forte candidato a influenciar distocia devido ao peso e tamanho ao nascimento, a proximidade deste gene com o gene ABCG2 e o histórico de aumento na produção de leite juntamente com o aumento na ocorrência de partos distócicos, pode-se sugerir que este fato é um interessante exemplo de como genes ligados podem estar relacionados com correlações genéticas entre características e ainda mais, como a seleção para um dos genes pode afetar o outro.

A segunda região contém o gene PPARGC1A, envolvido com adipogênese, gliconeogênese e metabolismo oxidativo. Embora este gene seja um candidato potencial, ainda não foi identificada uma mutação candidata a ser mutação causal (WEIKARD *et al.*, 2005; OGOREVC *et al.*, 2009).

A terceira região contém o *cluster* dos genes da caseína, formado por quatro genes: CSN1S1, CSN2, CSN1S2, CSN3. As caseínas sensíveis ao cálcio (CSN1S1, CSN2 e CSN1S2) se ligam ao fosfato de cálcio para formar micelas de caseína e a caseína não sensível ao cálcio (CSN3) estabiliza a formação da micela interagindo com as outras caseínas (KAWASAKI *et al.*, 2011). As caseínas são responsáveis por cerca de 80% do conteúdo total de proteínas do leite bovino.

Diversos SNPs e haplótipos dentro do *cluster* da caseína foram associados com produção de leite e produção de proteína no leite, principalmente envolvendo os genes CSN2 e CSN1S2. Além disso, observa-se um alto LD entre SNPs localizados dentro dos genes CSN1S1, CSN2 e CSN1S2 (BOETTCHER *et al.*, 2004; NILSEN *et al.*, 2009; SODELAND *et al.*, 2011).

Uma transversão de C para A no gene CSN2 (que causa a troca de uma prolina para histidina na proteína) é sugerida como mutação causal para este QTL, sendo que o alelo C aumenta e o alelo A diminui a produção de leite e proteína no leite em animais da raça Vermelha Norueguesa (NILSEN *et al.*, 2009). Apesar disso, estes efeitos não são consistentes entre raças (BOETTCHER *et al.*, 2004; NILSEN *et al.*, 2009). Por causa desta controvérsia, alguns SNPs localizados em regiões regulatórias do *cluster* da caseína são candidatos a QTN, pois podem participar na regulação dos processos de transcrição e tradução (SODELAND *et al.*, 2011).

O cromossomo BTA14 possui um QTL na região centromérica com um efeito principal na porcentagem de gordura no leite (KHATKAR *et al.*, 2004; COLE *et al.*, 2009; HAYES *et al.*, 2010; JIANG *et al.*, 2010) e uma substituição de dinucleotídeo no *exon* VIII (que leva à troca de lisina para alanina na sequência de aminoácidos da proteína) do gene DGAT1 é indicado como sendo o QTN (GRISART *et al.*, 2002). O alelo K (lisina) está associado ao aumento da porcentagem de gordura e proteína e da quantidade total de gordura no leite enquanto leva a uma diminuição da produção de leite e quantidade total de proteína no leite, enquanto o alelo A (alanina) possui o efeito contrário (GRISART *et al.*, 2002, 2004; WINTER *et al.*, 2002).

Análises adicionais sugerem a presença de alelos alternativos segregando para este QTL ou a presença de um outro QTL ligado ao DGAT1 influenciando tanto a quantidade total quanto a porcentagem de proteína e gordura no leite (BENNEWITZ *et al.*, 2004; KÜHN *et al.*, 2004) e genes como CYP11B1, VPS28, NIBP são sugeridos como potenciais candidatos (KAUPE *et al.*, 2007; COLE *et al.*, 2011). Alguns estudos indicam que a região que contém o gene DGAT1 está sob seleção (HAYES *et al.*, 2009; QANBARI *et al.*, 2010; HOSOKAWA *et al.*, 2011; SCHWARZENBACHER *et al.*, 2012).

Pelo menos dois QTLs relacionados com características de produção de leite estão localizados no cromossomo BTA20. O primeiro foi mapeado próximo ao gene GHR e possui um efeito importante na porcentagem de proteína e um efeito menor na porcentagem de gordura e produção de leite, já o segundo foi mapeado na região que contém o gene PRLR e afeta produção de proteína e gordura no leite (Blott *et al.*, 2003; KHATKAR *et al.*, 2004; SCHNABEL *et al.*, 2005b; VIITALA *et al.*, 2006; COLE *et al.*, 2009; OGOREVC *et al.*, 2009; JIANG *et al.*, 2010).

Na transversão de A para T no *exon* VIII do gene GHR (que leva à troca de fenilalanina para tirosina), o alelo Y (tirosina) está relacionado com diminuição na produção de leite e aumento tanto na porcentagem de gordura quanto de proteína no leite quando comparado com o alelo F (fenilalanina) em várias raças de bovinos leiteiros. Além disso, um estudo retrospectivo em bovinos leiteiros na Nova Zelândia mostrou que o genótipo FF está associado com uma maior probabilidade de o touro ser selecionado e também que a frequência do alelo F tem aumentado nos últimos anos (BLOTT *et al.*, 2003).

Em relação ao gene PRLR, a mudança de dinucleotídeo (GT para AC) no *exon* III deste gene (que leva à troca de serina por asparagina) está associada com produção de leite, proteína e gordura no leite (VIITALA *et al.*, 2006).

As mutações descritas nos genes GHR e PRLR são sugeridas como causais para estes dois QTLs (BLOTT *et al.*, 2003; VIITALA *et al.*, 2003) e estudos recentes mostram que as regiões que contém estes dois genes estão sob seleção (HAYES *et al.*, 2009; FLORI *et al.*, 2009; THE BOVINE HAPMAP, 2009; QANBARI *et al.*, 2010, 2011; STELA *et al.*, 2010; WIENER *et al.*, 2011).

Alguns estudos demonstram a presença de um QTL relacionado com produção de proteína e gordura no leite próximo ao gene STAT1. Este gene está envolvido com o desenvolvimento da glândula mamária e, uma transição de C para T na região 3' deste gene foi recentemente associada com produção de leite, proteína e gordura no leite em animais da raça Holandesa (COBANOGLU *et al.*, 2006). Dois estudos recentes sugerem a presença de assinaturas de seleção na região próxima ao gene STAT1 (HAYES *et al.*, 2009; HOSOKAWA *et al.*, 2011).

Recentemente, um QTL no cromossomo BTA18 próximo aos genes SIGLEC-5 e ZNF577 foi associado com produtividade total, conformação, longevidade e facilidade de parto em bovinos da raça Holandesa (COLE *et al.*, 2009). Esta região também foi identificada como contendo assinaturas de seleção e, embora estas características não estejam dentro dos principais objetivos de seleção nos programas de melhoramento, a seleção contra alelos desfavoráveis pode ser responsável por estas marcas (QANBARI *et al.*, 2011).

Muitas das regiões identificadas como contendo assinaturas de seleção em bovinos não possuem genes candidatos definidos. A análise funcional destas regiões demonstrou a presença de genes envolvidos com os eixos gonadotrófico e somatotrófico, desenvolvimento muscular, crescimento, desenvolvimento do sistema nervoso e resposta imune (BARENDSE *et al.*, 2009; FLORI *et al.*, 2009; GAUTIER *et al.*, 2009; THE BOVINE HAPMAP, 2009; QANBARI *et al.*, 2010, 2011; STELA *et al.*, 2010; WIENER *et al.*, 2011).

Suínos

A domesticação dos suínos ocorreu de forma independente em diversos momentos e locais na Eurásia aproximadamente 9000 anos atrás. Os suínos domésticos são encontrados em uma enorme quantidade de ambientes e apresentam uma grande variação em características morfológicas, comportamentais e adaptativas (LARSON *et al.*, 2005; CHEN *et al.*, 2007).

A utilização dos suínos em diversos tipos diferentes de sistemas de produção e condições ambientais ao redor do mundo resultou em uma grande variedade de raças, cada uma contendo adaptações especiais. Atualmente, a maioria dos sistemas de produção de suínos é baseado na utilização de cinco raças altamente especializadas: Large White, Duroc, Landrace, Hampshire e Pietrain. É interessante ressaltar que existe uma quantidade considerável de espécies selvagens próximas aos suínos domésticos que podem ser utilizadas em estudos para conhecimento de alguns aspectos da evolução desta espécie (CHEN *et al.*, 2007).

Uma substituição no *intron* III do gene IGF2 está associada com aumento na musculabilidade e diminuição da proporção de gordura na carne. Esta característica vem sendo objeto de intenso processo de seleção nas populações comerciais de suínos (VAN LAERE *et al.*, 2003). Ojeda *et al.* (2008) demonstraram a presença de assinatura de seleção neste gene em animais das raças Pietrain, Hampshire e Duroc, comumente utilizadas como raças paternas.

Uma mutação intrônica no gene ESR está associada com tamanho da leitegada em suínos (ROTHSCHILD *et al.*, 1996; SHORT *et al.*, 1997). Embora alguns estudos tenham demonstrado resultados conflitantes (MUÑOZ *et al.*, 2007), este marcador é comumente utilizado na indústria em programas de seleção assistida por marcadores (DEKKERS *et al.*, 2004). Recentemente, Bohnomne *et al.* (2010) demonstraram que este gene está sob seleção em animais da raça Large White.

A análise funcional das regiões sob seleção em suínos mostra a presença de genes envolvidos com desenvolvimento cerebral, desenvolvimento muscular, crescimento, metabolismo, sistema olfatório e sistema imune (AMARAL *et al.*, 2011).

Ovinos e caprinos

Os ovinos e caprinos foram as primeiras espécies de animais de produção domesticadas há cerca de 9000 anos. Estas espécies são distribuídas por praticamente todas as regiões do globo e têm se adaptado às mais diversas condições ambientais, resultando em uma grande variação entre populações (DIAMOND, 2002; GENTRY *et al.*, 2004; NADERI *et al.*, 2008; CHESSA *et al.*, 2009; KIJAS *et al.*, 2009). Os ovinos vêm sendo selecionados desde a domesticação para produção de carne, leite e lã (CHESSA *et al.*, 2009; KIJAS *et al.*, 2009).

Kijas *et al.* (2012) realizaram uma varredura genômica em animais pertencentes a 74 raças de ovinos. Trinta e uma regiões foram identificadas como candidatas a estar sob seleção contendo genes relacionados com cor de pelagem, morfologia óssea, e características de crescimento e reprodução.

Esta análise revelou um sinal forte de seleção próximo ao gene GDF-8 quando animais da raça Texel foram comparados com animais de outras raças (KIJAS *et al.*, 2012). Além disso, Clop *et al.* (2006) demonstraram uma redução na variabilidade de microssatélites ao redor deste gene ao comparar animais da raça Texel que apresentavam musculatura dupla com animais de outras raças.

A região próxima ao gene GDF-8 contém QTLs para características de carcaça em animais da raça Texel (JOHNSON *et al.*, 2005) e uma transição de G para A na região 3' UTR deste gene é considerada como o QTN responsável pelo fenótipo de musculatura dupla em animais desta raça. A mutação em questão cria um sítio alvo para, pelo menos, dois miRNAs que estão envolvidos na regulação da produção da proteína Miostatina (CLOP *et al.*, 2006).

Kijas *et al.* (2012) também identificaram regiões sob seleção próximas a QTLs relacionados com características de pelo e lã.

Moradi *et al.* (2012) realizaram uma varredura genômica para identificar assinaturas de seleção comparando animais de raças que apresentam ou não “garupa gorda”. Este estudo revelou a presença de, pelo menos, três regiões (nos cromossomos OAR5, OAR7 e OARX) sujeitas à seleção, e é interessante ressaltar que a maioria destas regiões está associada a QTLs para características de carcaça. O desenvolvimento da anotação do genoma ovino certamente irá auxiliar na identificação e validação de genes candidatos dentro destas regiões (MORADI *et al.* 2012).

CONCLUSÕES

Os processos de domesticação e a seleção artificial certamente vêm sendo cruciais na evolução do genoma das espécies animais de interesse zootécnico. A identificação de regiões candidatas a estarem sob efeito da seleção pode auxiliar tanto no entendimento de mecanismos moleculares envolvidos na adaptação quanto na identificação de regiões genômicas associadas a características de interesse ecológico/econômico que estejam sob seleção.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKEY, J.M.; ZHANG, G.; ZHANG, K.; JIN, L.; SHRIVER, M.D. Interrogating a high-density SNP map for signatures of natural selection. **Gen. Res.**, v.12, p.1805-1814, 2002.
- AMARAL, A.J.; FERRETTI, L.; MEGENS, H-J.; CROOIJMANS, R.P.M.A.; NIE, H.; RAMOS-ONSINS, S.E.; PEREZ-ENCISO, M.; SCHOOK, L.B.; GROENEN, M.A.M. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. **PLoS ONE**, v.6, n.4, e14782, 2011.
- ANDERSSON, L.; GEORGES, M. Domestic-animal genomics: Deciphering the genetics of complex traits. **Nat. Rev. Gen.**, v.5, p.202-212, 2004.
- ANDOLFATO, P. Adaptive hitchhiking effects on genome variability. **Curr. Op. Gen. Dev.**, v.11, p.635-641, 2001.
- AVISE, J.C.; AYALA, F.J. In the light of evolution III: Two centuries of Darwin. **Proc. Nat. Acad. Sci. U.S.A.**, v.106, suppl.1, p.9933–9938, 2009.
- BARENDSE, W. Assessing lipid metabolism. International patent application PCT/AU98/00882.; International Patent Publication WO 99/23248, 1999.
- BARENDSE, W.; REVERTER, A.; BUNCH, R.J.; HARRISON, B.E.; BARRIS, W.; THOMAS, M.B. A validated whole-genome association study of efficient food conversion in cattle. **Genetics**, v.176, n.3, p.1893-1905, 2007.

- BARENDSE, W.; HARRISON, B.E.; BUNCH, R.J.; THOMAS, M.B.; TURNER, L.B. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. **BMC Genomics**, 10:178, 2009.
- BARTON, N.H. The effect of hitch-hiking on neutral genealogies. **Genet. Res.**, v.72, p.123-133, 1998.
- BEAUMONT, M.A. Adaptation and speciation: what can F_{st} tell us? **Trends Ecol. Evol.**, v.20, n.8, p.435-440, 2005.
- BEAUMONT, M.A.; BALDING, D.J. Identifying adaptive genetic divergence among populations from genome scans. **Mol. Ecol.**, v.13, p.969-980, 2004.
- BEAUMONT, M.A.; NICHOLS, R.A. Evaluating loci for use in the genetic analysis of population structure. **Proc. R. Soc. Lond. B**, v.263, p1619-1626, 1996.
- BELLINGE, R.H.S.; LIBERLES, D.A.; IASCHI, S.P.A.; O'BRIEN, P.A.; TAY, G.K. Myostatin and its implications on animal breeding: a review. **Anim. Genet.**, n.36, v.1, p.1-6, 2005.
- BENNEWITZ, J.; REINSCH, N.; PAUL, S.; LOOFT, C.; KAUPE, B.; WEIMANN, C.; ERHARDT, G.; THALLER, G.; KÜHN, C.; SCHWERIN, M.; *et al.* The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on Bovine chromosome 14. **J. Dairy Sci.**, v.87, p.431-442, 2004.
- BERTRAND, J.K.; GREEN, R.D.; HERRING, W.O.; MOSER, D.W. Genetic evaluation for beef carcass traits. **J. Anim. Sci.**, n.79, p.E190-E200, 2001.
- BLOTT, S.; KIM, J.J.; MOISIO, S.; SCHMIDT-KUNTZEL, A.; CORNET, A.; BERZI, P.; CAMBISANO, N.; FORD, C.; GRISART, B.; JOHNSON, D.; *et al.* Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition. **Genetics**, v.163, p.253-266, 2003.
- BOETTCHER, P.J.; CAROLI, A.; STELLA, A.; CHESSA, S.; BUDELLI, E.; CANAVESI, F.; GHIROLDI, S.; PAGNACCO, G. Effects of casein haplotypes on milk production traits in Italian Holstein and Brown Swiss cattle. **J. Dairy Sci.**, n.87, v.12, p.4311-4317, 2004.

- BONHOMME, M.; CHEVALET, C.; SERVIN, B.; BOITARD, S.; ABDALLAH, J.; BLOTT, S.; SANCRISTOBAL, M. Detecting selection in population trees: The Lewontin and Krakauer Test extended. **Genetics**, n.186, p.241-262, 2010.
- BOWCOCK AM.; KIDD JR.; MOUNTAIN JL.; HEBERT JM.; CAROTENUTO L.; KIDD, K.K.; CAVALLI-SFORZA, L.L. Drift admixture and selection in human evolution: A study with DNA polymorphisms. **Proc. Nat. Acad. Sci. USA**, v.88, p.839-843, 1991.
- BRAVERMAN, J.M.; HUDSON, R.R.; KAPLAN, N.L.; LANGLEY, C.H.; STEPHAN, W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. **Genetics**, n.140, p.783-796, 1995.
- BROOKFIELD, J.F.Y. Population genetics: The signature of selection. **Curr. Biol.**, v.11, p.R388–R390, 2001.
- BROTHERSTONE, S.; GODDARD, M. Artificial selection and maintenance of genetic variance in the global dairy cow population. **Phil. Trans. R. Soc. B.**, v.360, p.1479-1488, 2005.
- CASAS, E.; WHITE, S.N.; RILEY, D.G.; SMITH, T.P.L.; BRENNEMAN, R.A.; OLSON, T.A.; JOHNSON, D.D.; COLEMAN, S.W.; BENNETT, G.L.; CHASE Jr, C.C. Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle. **J. Anim. Sci.**, v.83, p,13-19, 2005.
- CHARLESWORTH B.; CHARLESWORTH. D.; BARTON, N.H. The effects of genetic and geographic structure on neutral variation. **Annu. Rev. Ecol. Evol. Syst.**, v.34, p.99-125, 2003.
- CHARLESWORTH, B. A hitch-hiking guide to the genome: a commentary on ‘The hitch-hiking effect of a favourable gene’ by John Maynard Smith and John Haigh. **Genet. Res.**, v.89, p.389–390, 2007.
- CHARLESWORTH, B.; MORGAN, M.T.; CHARLESWORTH, D. The effect of deleterious mutations on neutral molecular variation. **Genetics**, v.134, p.1289-1303, 1993.
- CHARLESWORTH, D. Balancing selection and its effects on sequences in nearby genome regions. **PLoS Genet.**, v.2, n.4:e64, 2006.

- CHARLESWORTH, D.; CHARLESWORTH, B.; MORGAN, M.T. The pattern of neutral molecular variation under the background selection model. **Genetics**, v.141, p.1619-1632, 1995.
- CHEN, K.; BAXTER, T.; MUIR, W.M.; GROENEN, M.A.; SCHOOK, L.B. Genetic Resources, Genome Mapping and Evolutionary Genomics of the Pig (*Sus scrofa*). **Int. J. Biol. Sci.**, v.3, n.3, p.153-165, 2007.
- CHESSA, B.; PEREIRA, F.; ARNAUD, F.; AMORIM, A.; GOYACHE, F.; MAINLAND, I.; KAO, R.R.; PEMBERTON, J.M.; BERALDI, D.; STEAR, M.J.; *et al.* Revealing the History of Sheep Domestication Using Retrovirus Integrations. **Science**, v.324, n.5926, p.532-536, 2009.
- CLOP, A.; MARCQ, F.; TAKEDA, H.; PIROTTINI, D.; TORDOIR, X.; BIBE, B.; BOUIX, J.; CAIMENT, F.; ELSEN, J-M.; EYCHENNE, F.; *et al.* A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. **Nat. Genet.**, v.38, n.7, p.813-818, 2006.
- COBANOGLU, O.; ZAITOUN, I.; CHANG, Y.M.; SHOOK, G.E.; KHATIB, H. Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. **J. Dairy Sci.**, v.89, n.11, p.4433-4437, 2006.
- COHEN-ZINDER, M.; SERROUSSI, E.; LARKIN, D.M.; LOOR, J.J.; EVERTS-VAN, D.E.R.; WIND, A.; LEE, J.H.; DRACKLEY, J.K.; BAND, M.R.; HERNANDEZ, A.G.; *et al.* Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. **Genome Res.**, v.15, p.936-944, 2005.
- COLE, J.B.; VANRADEN, P.M.; O'CONNELL, J.R.; VAN TASSELL, C.P.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; WIGGANS, G.R. Distribution and location of genetic effects for dairy traits. **J. Dairy Sci.**, v.92, p.2931–2946, 2009.
- COLE, J.B.; WIGGANS, G.R.; MA, L.; SONSTEGARD, T.S.; LAWLOR Jr, T.J.; CROOKER, B.A.; VAN TASSELL, C.P.; YANG, J.; WANG, S.; MATUKUMALLI, L.K.; *et al.* Genome-wide association analysis of thirty one production.; health.; reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics**, 12:408, 2011.

- DE KONING, D.J. Conflicting candidates for cattle QTLs. **Trends. Genet.**, v.22, n.6, p.301-305, 2006.
- DEKKERS, J.C.M. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. **J. Anim. Sci.**, v.82, p.E313-E328, 2004.
- DIAMOND, J. Evolution, consequences and future of plant and animal domestication. **Nature**, v.418, p.700-707, 2002.
- DONG, Y.; XIE, M.; JIANG, Y.; XIAO, N.; DU, X.; ZHANG, W.; TOSSER-KLOP, G.; WANG, J.; YANG, S.; LIANG, J.; *et al.* Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). **Nat. Biotech.**, v.31, p.135-141, 2013.
- DRISCOLL, C.A.; MACDONALD, D.W.; O'BRIEN, S.J. From wild animals to domestic pets: an evolutionary view of domestication. **Proc. Nat. Acad. Sci. U.S.A.**, v.106(suppl.1), p.9971-9978, 2009.
- DROGEMULLER, C.; WOHLKE, A.; MOMKE, S.; DISTL, O. Fine mapping of the polled locus to a 1-Mb region on bovine Chromosome 1q12. **Mamm. Genome**, v.16, p.613-620, 2005.
- EBERLEIN, A.; TAKASUGA, A.; SETOGUCHI, K.; PFUHL, R.; FLISIKOWSKI, K.; FRIES, R.; KLOPP, N.; FÜRBASS, R.; WEIKARD, R.; KÜHN, C. Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the Non-SMC Condensin I Complex, Subunit G (NCAPG) gene. **Genetics**, v.183, p.951-964, 2009.
- EXCOFFIER, L.; HOFER, T.; FOLL, M. Detecting loci under selection in a hierarchically structured population. **Heredity**, v.103, p.285-298, 2009.
- FALCONER, D.S.; MACKAY, T.F.C. Introduction to quantitative genetics. 4th edition. Longman, Essex, 478 pp, 1996.
- FANG, M.; LARSON, G.; SOARES-RIBEIRO, H.; LI, N.; ANDERSSON, L. Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. **PLoS Genet.**, 5(1): e1000341, 2009.
- FAY, J.C.; WU, C.I. Hitchhiking under positive Darwinian selection. **Genetics**, v.155, p.1405-1413, 2000.

- FLORI, L.; FRITZ, S.; JAFFRÉZIC, F.; BOUSSAHA, M.; GUT, I.; HEATH, S.; FOULLEY, J.L.; GAUTIER, M. The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. **PLoS ONE** 4(8): e6595, 2009.
- FOLL, M.; GAGGIOTTI, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. **Genetics**, v.180, p.977-993, 2008.
- FONTANESI, L.; ALESSANDRO, E.D.; SCOTTI, E.; LIOTTA, L.; CROVETTI, A.; CHIOFALO, V.; RUSSO, V. Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. **Anim. Genet.**, v.41, p.478-492, 2010.
- FU, Y.X. New statistical tests of neutrality for DNA samples from a population. **Genetics**, v.143, p.557-570, 1996.
- FU, Y.X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics**, v.147, p.915-925, 1997.
- FU, Y.X.; LI, W.H. Statistical Tests of Neutrality of Mutations. **Genetics**, v.133, p.693-709, 1993.
- GARRICK, D.J. The nature, scope and impact of genomic prediction in beef cattle in the United States. **Genet. Sel. Evol.**, v.43(1):17, 2011.
- GARRICK, D.J.; GOLDEN, B.L. Producing and using genetic evaluations in the United States beef industry of today. **J. Anim. Sci.**, n.87, p.E11-E18, 2009.
- GAUTIER, M.; FLORI, L.; RIEBLER, A.; JAFFRÉZIC, F.; LALOÉ, D.; GUT, I.; MOAZAMI-GOUDARZI, K.; FOULLEY, J.L. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. **BMC Genomics**, v.10:550, 2009.
- GENTRY, A.; CLUTTON-BROCK, J.; GROVES, C.P. The naming of wild animal species and their domestic derivatives. **J. Archaeol. Sci.**, v.31, p.645–651, 2004.
- GOLDMAN, N.; YANG, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Mol. Biol. Evol.**, v.11, n.2, p.725-736, 1994.
- GREGORY, T.R. Artificial selection and domestication: Modern lessons from Darwin's enduring analogy. **Evo. Edu. Outreach**, v.2, p.5-27, 2009.
- GRISART, B.; COPPIETERS, W.; FARNIR, F.; KARIM, L.; FORD, C.; BERZI, P.; CAMBISANO, N.; MNI, M.; REID, S.; SIMON, P.; *et al.* Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1

gene with major effect on milk yield and composition. **Genome Res.**, v.12, p.222-231, 2002.

GRISART, B.; FARNIR, F.; KARIM, L.; CAMBISANO, N.; KIM, J-J.; KVASZ, A.; MNI, M.; SIMON, P.; FRÈRE, J-M.; COPPIETTERS, W.; *et al.* Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. **Proc. Nat. Acad. Sci. U.S.A.**, v.101, n.8, p.2398-2403, 2004.

GROENEN, M.A.M.; ARCHIBALD, A.L.; UENISHI, H.; TUGGLE, C.K.; TAKEUCHI, Y.; ROTSCCHILD, M.F.; ROGEL-GAILLARD, C.; PARK, C.; MILAN, D.; MEGENS, H-J.; *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. **Nature**, v.491, p.393-398, 2012.

GUTIERREZ-GIL, B.; WILLIAMS, J.L.; HOMER, D.; BURTON, D.; HALEY, C.S.; WIENER, P. Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. **J. Anim. Sci.**, v.87, p.24-36, 2009.

HARRIS, E.E.; MEYER, D. The molecular signature of selection underlying human adaptations. **Am. J. Phys. Anthropol.**, v.131, n.S43, p.89-130, 2006.

HARTL, D.L.; CLARK, A.G. Princípios de Genética de Populações. 4ª edição. Artmed, Porto Alegre, 660pp, 2010.

HAYES, B.; PRYCE, J.; CHAMBERLAIN, A.J.; BOWMAN, P.J.; GODDARD, M.E. Genetic architecture of complex traits and accuracy of genomic prediction for Coat colour, milk-fat percentage and type in Holstein cattle as contrasting model traits. **PLoS Genet.**, 6(9):e1001139, 2010.

HAYES, B.J.; CHAMBERLAIN, A.J.; MACEACHERN, S.; SAVIN, K.; MCPARTLAN, H.; MACLEOD, I.; SETHURAMAN, .; GODDARD, M.E. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. **Anim. Genet.**, v.40, p.176-184, 2009.

HAYES, B.J.; LIEN, S.; NILSEN, H.; OLSEN, H.G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISSEN, T.H.E. The origin of selection signatures on bovine chromosome 6. **Anim. Genet.**, v.39, p.105-111, 2008.

HOHENLOHE, P.A.; PHILLIPS, P.C.; CRESKO, W.A. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. **Int. J. Plant. Sci.**, v.171, n.9, p.1059-1071, 2010.

- HOSOKAWA, D.; ISHII, A.; YAMAJI, K.; SASAZAKI, S.; OYAMA, K.; MANNEN, H. Identification of divergently selected regions between Japanese Black and Holstein cattle using bovine 50k SNP array. **Anim. Sci. J.**, 2011.
- HUSSIN, J.; NADEAU, P.; LEFEBVRE, J.F.; LABUDA, D. Haplotype allelic classes for detecting ongoing positive selection. **BMC Bioinformatics**, n.11:65, 2010.
- INNAN, H.; STEPHAN, W. Distinguishing the hitchhiking and background selection models. **Genetics**, v.165, p.2307-2312, 2003.
- INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, v.432, p.695-716, 2004.
- JIANG, L.; LIU, J.; SUN, D.; MA, P.; DING, X.; YU, Y.; ZHANG, Q. Genome wide association studies for milk production traits in Chinese Holstein population. **PLoS ONE**, 5(10): e13661, 2010.
- JOHNSON, P.L.; MCEWAN, J.C.; DODDS, K.G.; PURCHAS, R.W.; BLAIR, H.T. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. **J. Anim. Sci.**, v.83, p.988-2000, 2005.
- JOHNSTON, S.E.; MCEWAN, J.C.; PICKERING, N.K.; KIJAS, J.W.; BERALDI, D.; PILKINGTON, J.G.; Pemberton JM and Slate J Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. **Mol. Ecol.**, v.20, p.2555-2566, 2011.
- KAPLAN, N.L.; HUDSON, R.R.; LANGLEY, C.H. The "Hitchhiking Effect" revisited. **Genetics**, v.123, p.887-899, 1989.
- KAUPE, B.; BRANDT, H.; PRINZENBERG, E-M.; ERHARDT, G. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction and productive lifespan in German Holstein cattle. **J. Anim. Sci.**, v.85, p.11-21, 2007.
- KAWASAKI, K.; LAFONT, A-G.; SIRE, J-Y. The evolution of milk casein genes from tooth genes before the origin of mammals. **Mol. Biol. Evol.**, v.28, n.7, p.2053-2061, 2011.
- KELLEY, J.L.; MADEOY, J.; CALHOUN, J.C.; SWANSON, W.; AKEY, J.M. Genomic signatures of positive selection in humans and the limits of outlier approaches. **Genome Res.**, v.16, p.980-989, 2006.

- KHATKAR, M.S.; THOMSON, P.C.; TAMMEN, I.; RAADSMA, H.W. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. **Genet. Sel. Evol.**, v.36, n.2, p.163-190, 2004.
- KIJAS JW.; TOWNLEY D.; DALRYMPLE BP.; HEATON MP.; MADDOX JF.; MCGRATH A.; WILSON P.; INGERSOLL RG.; MCCULLOCH R.; MCWILLIAM S.; *et al.* A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. **PLoS ONE**, 4(3):e4668, 2009.
- KIJAS, J.M.H.; WALES, R.; TÖRNSTEN, A.; CHARDON, P.; MOLLER, M.; ANDERSSON, L. Melanocortin Receptor 1 (MC1R) Mutations and Coat Color in Pigs. **Genetics**, v.150, p.1177-1185, 1998.
- KIJAS, J.W.; LENSTRA, J.A.; HAYES, B.; BOITARD, S.; PORTO NETO, L.R.; CRISTOBAL, M.S.; SERVIN, B.; MCCULLOCH, R.; WHAN, V.; GIETZEN, K.; *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. **PLoS Biol.**, 10(2):e1001258, 2012.
- KIM, Y.; NIELSEN, R. Linkage disequilibrium as a signature of selective sweeps. **Genetics**, v.167, p.1513–1524, 2004.
- KIM, Y.; STEPHAN, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. **Genetics**, v.160, p.765–777, 2002.
- KIM, Y.; STEPHAN, W. Joint effects of genetic hitchhiking and background selection on neutral variation. **Genetics**, v.155, p.1415-1427, 2000.
- KIMURA, M. DNA and the neutral theory. **Phil. Trans. R. Soc. Lond. B.**, v.312, p.343-354, 1986.
- KIMURA, M. The neutral theory of molecular evolution: A review of recent evidence. **Jpn. J. Genet.**, v.66, p.367-386, 1991.
- KIMURA, R.; FUJIMOTO, A.; TOKUNAGA, K.; OHASHI, J. A practical genome scan for population-specific strong selective sweeps that have reached fixation. **PLoS ONE**, 2(3): e286, 2007.
- KREITMAN, M. Methods to detect selection in populations with applications to the human. **Annu. Rev. Genomics Hum. Genet.**, v.1, p.539–59, 2000.
- KÜHN, C.; THALLER, G.; WINTER, A.; BININDA-EMONDS, O.R.P.; KAUPPE, B.; ERHARDT, G.; BENNEWITZ, J.; SCHWERIN, M.; FRIES, R. Evidence for multiple

- alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. **Genetics**, v.167, n.4, p.1873-1881, 2004.
- LARSON, G.; DOBNEY, K.; ALBARELLA, U.; FANG, M.; MATISOO-SMITH, E.; ROBINS, J.; LOWDEN, S.; FINLAYSON, H.; BRAND, T.; WILLERSLEV, E.; *et al.* Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. **Science**, v.307, n.5715, p.1618-1621, 2005.
- LEONARD S.; KHATIB H.; SCHUTZKUS V.; CHANG YM and MALTECCA C Effects of the Osteopontin gene variants on milk production traits in dairy cattle. **J. Dairy Sci.**, v.88, p.4083-4086, 2005.
- LEWONTIN, R.C.; KRAKAWER, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. **Genetics**, v.74, p.175-195, 1973.
- LI, J.; YANG, H.; LI, J.R.; LI, H.P.; NING, T.; PAN, X.R.; SHI, P.; ZHANG, Y.P. Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. **Heredity**, v.105, p.274-281, 2010b.
- LI, M-H.; ISO-TOURO, T.; LAURÉN, H.; KATANEN, J. A microsatellite-based analysis for the detection of selection on BTA1 and BTA20 in northern Eurasian cattle (*Bos taurus*) populations. **Genet. Sel. Evol.**, 42:32, 2010a.
- MACEACHERN, S.; HAYES, B.J.; MCEWAN, J.; GODDARD, M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. **BMC Genomics**, 10:181, 2009.
- MCVEAN, G. The structure of linkage disequilibrium around a selective sweep. **Genetics**, v.175, p.1395–1406, 2007.
- MIDDLETON, B.K.; GIBB, J.B. An overview of beef cattle improvement programs in the United States. **J. Anim. Sci.**, v.69, p.3861-3871, 1991.
- MILLER, S. Genetic improvement of beef cattle through opportunities in genomics. **Rev. Bras. Zoot.**, v.39, p.247-255, 2010.
- MOORE, S.S.; Li, C.; BASARAB, J.; SNELLING, W.M.; KNEELAND, J.; MURDOCH, B.; HANSEN, C.; BENKEL, B. Fine mapping of quantitative trait loci and assessment of positional candidate genes for backfat on bovine chromosome 14 in a commercial line of *Bos taurus*. **J. Anim. Sci.**, v.81, p.1919-1925, 2003.

- MORADI, M.H.; NEJATI-JAVAREMI, A.; MORADI-SHAHRBABAK, M.; DODDS, K.G.; MCEWAN, J.C. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. **BMC Genetics**, 13(10), 2012.
- MUÑOZ, G.; OVILO, C.; ESTELLÉ, J.; SILIÓ, L.; FERNÁNDEZ, A.; RODRIGUEZ, C. Association with litter size of new polymorphisms on ESR1 and ESR2 genes in a Chinese-European pig line. **Genet. Sel. Evol.**, v.39, n.2, p.195-206, 2007.
- NADERI, S.; REZAEI, H.R.; POMPANON, F.; BLUM, M.G.B.; NEGRINI, R.; NAGHASH, H.R.; BALKIZ, Ö.; MASHKOURG, M.; GAGGIOTTI, O.E.; AJMONE-MARSANE, P.; *et al.* The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. **Proc. Nat. Acad. Sci. U.S.A.**, v.105, n.46, p.17659-17664, 2008.
- NAVARRO, A.; BARTON, N.H. The effects of multilocus balancing selection on neutral variability. **Genetics**, v.161, p.849-863, 2002.
- NEI, M. Selectionism and neutralism in molecular evolution. **Mol. Biol. Evol.**, v.22, n.12, p.2318–2342, 2005.
- NEI, M.; MARUYAMA, T. Letters to the editors: Lewontin-Krakauer test for neutral genes. **Genetics**, v.80, p.395, 1975.
- NEI, M.; GOJOBORI, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. **Mol. Biol. Evol.**, v.3, n.5, p.418-426, 1986.
- NIELSEN, R. Molecular Signatures of Natural Selection. **Annu. Rev. Genet.**, v.39, p.197-218, 2005.
- NIELSEN, R. Statistical tests of selective neutrality in the age of genomics. **Heredity**, v.86, p.641-647, 2001.
- NIELSEN, R.; WILLIAMSON, S.; KIM, Y.; HUBISZ, M.J.; CLARK, A.G.; BUSTAMANTE, C. Genomic scans for selective sweeps using SNP data. **Genome Res.**, v.15, p.1566-1575, 2005.
- NIELSEN, R.; YANG, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**, v.148, p.929-936, 1998.

- NILSEN, H.; OLSEN, H.G.; HAYES, B.; SEHESTED, E.; SVENDSEN, M.; NOME, T.; MEUWISSEN, T.; LIEN, S. Casein haplotypes and their association with milk production traits in Norwegian Red cattle. **Genet. Sel. Evol.**, 41:24, 2009.
- O'BRIEN, J.D.; MININ, V.N.; SUCHARD, M.A. Learning to count: Robust estimates for labeled distances between molecular sequences. **Mol. Biol. Evol.**, v.26, n.4, p.801-814, 2009.
- OGOREVC, J.; KUNEJ, T.; RAZPET, A.; DOVC, P. Database of cattle candidate genes and genetic markers for milk production and mastitis. **Anim. Genet.**, v.40, p.832-851, 2009.
- OJEDA, A.; HUANG, L.S.; REN, J.; ANGIOLILLO, A.; CHO, I.C.; SOTO, H.; LEMÚS-FLORES, C.; MAKUZ, A. S.M.; FOLCH, J.M.; PÉREZ-ENCIZO, M. Selection in the making: a worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. **Genetics**, v.178, p.1639-1652, 2008.
- OLEKSYK, T.K.; SMITH, M.W.; O'BRIEN, J. Genome-wide scans for footprints of natural selection. **Philos. Trans. R. Soc. Lond. B.**, v.365, p.185-205, 2010.
- OLSEN, H.G.; HAYES, B.J.; KENT, M.P.; NOME, T.; SVENDSEN, M.; LIEN, S. A genome wide association study for QTL affecting direct and maternal effects of stillbirth and dystocia in cattle. **Anim. Genet.**, v.41, n.3, p.273-280, 2010.
- OLSEN, H.G.; MEUWISSEN, T.H.E.; NILSEN, H.; SVENDSEN, M.; LIEN, S. Fine Mapping of Quantitative Trait Loci on Bovine Chromosome 6 Affecting Calving Difficulty. **J. Dairy Sci.**, v.91, p.4312-4322, 2008.
- OLSEN, H.G.; NILSEN, H.; HAYES, B.; BERG, P.R.; SVENDSEN, M.; LIEN, S.; MEUWISSEN, T.H.E. Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. **BMC Genetics**, 8:32, 2007.
- OLTENACU, P.A.; BROOM, D.M. The impact of genetic selection for increased milk yield on the welfare of dairy cows. **Anim. Welf.**, v.19, p.39-49, 2010.
- PANNIER, L.; MULLEN, A.M.; HAMILL, R.M.; STAPLETON, P.C.; SWEENEY, T. Association analysis of single nucleotide polymorphisms in DGAT1; TG and FABP4 genes and intramuscular fat in crossbred Bos taurus cattle. **Meat. Sci.**, v.85, n.3, p.515-518, 2010.

- POND, S.L.K.; FROST, D.W. Not so different after all: A comparison of methods for detecting amino acid sites under selection. **Mol. Biol. Evol.**, v.22, n.5, p.1208-1222, 2005.
- PRASAD, A.; SCHNABEL, R.D.; MCKAY, S.D.; MURDOCH, B.; STOTHARD, P.; KOLBEHDARI, D.; WANG, Z.; TAYLOR, J.F.; MOORE, S.S. Linkage disequilibrium and selection signatures on chromosomes 19 and 29 in beef and dairy cattle. **Anim. Genet.**, v.39, p.597-605, 2008.
- PRITCHARD, J.K.; PICKRELL, J.K.; COOP, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. **Curr. Biol.**, v.20, p.R208-R215, 2010.
- PRZEWORSKI, M. The signature of positive selection at randomly chosen loci. **Genetics**, v.160, p.1179-1189, 2002.
- PRZEWORSKI, M.; COOP, G.; WALL, J.D. The signature of positive selection on standing genetic variation. **Evolution**, v.59, n.11, p.2312-2323, 2005.
- QANBARI, S.; GIANOLA, D.; HAYES, B.; SCHENKEL, F.; MILLER, S.; MOORE, S.; THALLER, G.; SIMIANER, H. Application of site and haplotype frequency based approaches for detecting selection signatures in cattle. **BMC Genomics**, 12:318, 2011.
- QANBARI, S.; PIMENTEL, E.C.G.; TETENS, J.; THALLER, G.; LICHTNER, P.; SHARIFI, A.R.; SIMIANER, H. A genome-wide scan for signatures of recent selection in Holstein cattle. **Anim. Genet.**, v.41, p.377-389, 2010.
- ROBERTSON, A. Letters to the editors: Remarks on the Lewontin-Krakauer test. **Genetics**, v.80, p.396, 1975.
- RON, M.; COHEN-ZINDER, M.; PETER, C.; WELLER, J.I.; ERHARDT, G. A polymorphism in ABCG2 in *Bos indicus* and *Bos taurus* cattle breeds. **J. Dairy Sci.**, v.89, p.4921-4923, 2006.
- ROTHSCHILD, M.; JACOBSON, C.; VASKE, D.; TUGGLE, C.; WANG, L.; SHORT, T.; ECKARDT, G.; SASAKI, S.; VINCENT, A.; MCLAREN, D.; *et al.* The estrogen receptor locus is associated with a major gene influencing litter size in pigs. **Proc. Nat. Acad. Sci. U.S.A.**, v.93, n.1, p.201-205, 1996.
- SABETI, P.C.; REICH, D.E.; HIGGINS, J.M.; LEVINE, H.Z.P.; RICHTER, D.J.; SCHAFFNER, S.F.; GABRIEL, S.B.; PLATKO, J.V.; PATTERSON, N.J.;

- MCDONALD, G.J.; *et al.* Detecting recent positive selection in the human genome from haplotype structure. **Nature**, v.419, n.24, p.832-837, 2002.
- SABETI, P.C.; VARILLY, P.; FRY, B.; LOHMUELLER, J.; HOSTETTER, E.; COTSAPAS, C.; XIE, X.; BYRNE, E.H.; MCCARROLL, S.A.; GAUDET, R.; *et al.* Genome-wide detection and characterization of positive selection in human populations. **Nature**, v.449, n.18, p.913-919, 2007.
- SCHLÖTTERER, C. Hitchhiking mapping: functional genomics from the population genetics perspective. **Trends Genet.**, v.19, n.1, p.32-38, 2003.
- SCHNABEL, R.D.; KIM, J-J.; ASHWELL, M.S.; SONSTERGARD, T.S.; VAN TASSEL, C.P.; CONNOR, E.E.; TAYLOR, J.F. Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine Osteopontin gene. **Proc. Nat. Acad. Sci. U.S.A.**, v.102, n.19, p.6896-6901, 2005a.
- SCHNABEL, R.D.; SONSTEGARD, T.S.; TAYLOR, J.F.; ASHWELL, M.S. Whole-genome scan to detect QTL for milk production.; conformation.; fertility and functional traits in two US Holstein families. **Anim. Genet.**, v.36, p.408-416, 2005b.
- SCHWARZENBACHER, H.; DOLEZAL, M.; FLISIKOWSKI, K.; SEEFRIED, F.; WURMSER, C.; SCHLÖTTERER, C.; FRIES, R. Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. **BMC Genomics**, 13(48), 2012.
- SHORT, T.H.; ROTHSCHILD, M.F.; SOUTHWOOD, O.I.; MCLAREN, D.G.; DE VRIES, A.; VAN DER STEEN, H.; ECKARDT, G.R.; TUGGLE, C.K.; HELM, J.; VASKE, D.A.; *et al.* Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. **J. Anim. Sci.**, v.75, n.12, p.3138-3142, 1997.
- SODELAND, M.; GROVE, H.; KENT, M.; TAYLOR, S.; SVENDSEN, M.; HAYES, B.J.; LIEN, S. Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. **BMC Genetics**, 12:70, 2011.
- STELLA, A.; AJMONE-MARSAN, P.; LAZZARI, B.; BOETTCHER, P. Identification of selection signatures in cattle breeds selected for dairy production. **Genetics**, v.185, p.1451-1461, 2010.
- STEPHAN, W. Genetic hitchhiking versus background selection: the controversy and its implications. **Phil. Trans. R. Soc. B.**, v.365, p.1245-1253, 2010.

- STEPHAN, W.; SONG, Y.S.; LANGLEY, C.H. The hitchhiking effect on linkage disequilibrium between linked neutral loci. **Genetics**, v.172, p.2647–2663, 2006.
- STORZ, J.F. Using genome scans of DNA polymorphism to infer adaptive population divergence. **Mol. Ecol.**, v.14, p.671-668, 2005.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics** v.123, p.585-595, 1989.
- TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. **Mol. Biol. Evol.**, v.24, n.8, p.1596-1599, 2007.
- TANG, K.; THORNTON, K.R.; STONEKING, M. A new approach for using genome scans to detect recent positive selection in the human genome. **PLoS Biol.**, 5(7): e171, 2007.
- TANTIA, M.S.; VIJH, R.K.; MISHRA, B.P.; MISHRA, B.; KUMAR, S.T.B.; SODHI, M. DGAT1 and ABCG2 polymorphism in Indian cattle (*Bos indicus*) and buffalo (*Bubalus bubalis*) breeds. **BMC Vet. Res.** 2:32, 2006.
- TEMPLETON, A.R. Population Genetics and Microevolutionary Theory. 1st edition. John Wiley and Sons, New Jersey, 705 pp, 2006.
- THALLER, G.; KÜHN, C.; WINTER, A.; EWALD, G.; BELLMANN, O.; WEGNER, J.; ZÜHLKE, H.; FRIES, R. DGAT1.; A new positional and functional candidate gene for intramuscular fat deposition in cattle. **Anim. Genet.**, v.34, n.5, p.354-357, 2003.
- THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM; ELSIK, C.G.; TELLAM, R.L.; WORLEY, K.C. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. **Science**, v.324, n.5926, p.522-528, 2009.
- THE BOVINE HAPMAP CONSORTIUM. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, v.324, p.538-532, 2009.
- THE INTERNATIONAL SHEEP GENOMICS CONSORTIUM. The sheep genome reference sequence: a work in progress. **Anim. Genet.**, v.41, n.5, p.449-453, 2010.
- TORO, M.; MÄKI-TANILA, A. Genomics Reveals Domestication History and Facilitates Breed Development. In: Oldenbroek.; K (ed) Utilisation and Conservation of Farm Animal Genetic Resources. Wageningen Academic Publishers, Wageningen, pp 75-102, 2007.

- TRUT, L.; OSKINA, I.; KHARLAMOVA, A. Animal evolution during domestication: The domesticated fox as a model. **Bioessays**, v.31, n.3, p.349-360, 2009.
- TSAKAS, S.; KRIMBAS, C.B. Testing the heterogeneity of F values: A suggestion and a correction. **Genetics**, v.84, p.399-401, 1976.
- VAN LAERE, A.S.; NGUYEN, M.; BRAUNSCHWEIG, M.; NEZER, C.; COLLETTE, C.; MOREAU, L.; ARCHIBALD, A.L.; HALEY, C.S.; BUYS, N.; TALLY, M.; *et al.* A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. **Nature**, v.425, n.6960, p.832-836, 2003.
- VENERONI, G.B.; MEIRELLES, S.L.; GROSSI, D.A.; GASPARIN, G.; IBELLI, A.M.G.; TIZIOTO, P.C.; OLIVEIRA, H.N.; ALENCAR, M.M.; REGITANO, L.C.A. Prospecting candidate SNPs for backfat in Canchim beef cattle. **Genet. Mol. Res.**, v.9, n.4, p.1997-2003, 2010.
- VIITALA, S.; SZYDA, J.; BLOTT, S.; SCHULMAN, N.; LIDAUER, M.; MÄKI-TANILA, A.; GEORGES, M.; VILKKI, J. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. **Genetics**, v.173, p.2151-2164, 2006.
- VIITALA, S.M.; SCHULMAN, N.F.; DE KONING, D.J.; ELO, K.; KINOS, R.; VIRTA, A.; VIRTA, J.; MÄKI-TANILA, A.; VILKKI, J.H. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle **J. Dairy Sci.**, v.86, p.1828-1836, 2003.
- VITALIS, R.; DAWSON, K.; BOURSOT, P. Interpretation of variation across marker loci as evidence of selection. **Genetics**, v.158, p.1811-1823, 2001.
- VITALIS, R.; DAWSON, K.; BOURSOT, P.; BELKHIR, K. DetSel 1.0: A computer program to detect markers responding to selection. **J. Hered.**, v.94, n.5, p.429-431, 2003.
- VOIGHT, B.F.; KUDARAVALLI, S.; WEN, X.; PRITCHARD, J.K. A map of recent positive selection in the human genome. **PLoS Biol.**, v.4, n.3, p.446-458, 2006.
- WEIKARD, R.; ALTMAIER, E.; SUHRE, K.; WEINBERGER, K.M.; HAMMON, H.M.; ALBRECHT, E.; SETOGUCHI, K.; TAKASUGA, A.; KÜHN, C. Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on *Bos taurus* growth and lipid deposition. **Physiol. Genomics**, v.42, p. 79-88, 2010.

- WEIKARD, R.; KHÜN, C.; GOLDHAMMER, T.; FREYER, G.; SCHWERIN, M. The bovine PPARGC1A gene: Molecular characterization and association of an SNP with variation of milk fat synthesis. **Physiol. Genomics**, v.21, p.1-13, 2005.
- WEIKARD, R.; WIDMANN, P.; BUITKAMP, J.; EMMERLING, R.; KUEHN, C. Revisiting the quantitative trait loci for milk production traits on BTA6. **Anim. Genet.**, v.43, n.3, p.318-323, 2012.
- WERTH, L.A.; HAWKINS, G.A.; EGGEN, A.; PETIT, E.; ELDUQUE, C.; KREIGESMANN, B.; BISHOP, M.D. Melanocyte Stimulating Hormone Receptor (MC1R) maps to bovine chromosome 18. **J. Anim. Sci.**, v.74:262, 1996.
- WIENER, P.; EDRISS, M.A.; WILLIAMS, J.L.; WADDINGTON, D.; LAW, 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**, 12:65, 2011.
- WIENER, P.; GUTIÉRREZ-GIL, B. Assessment of selection mapping near the myostatin gene (GDF-8) in cattle. **Anim. Genet.**, v.40, n.5, p.598-608, 2009.
- WIENER, P.; WILKINSON, S. Deciphering the genetic basis of animal domestication. **Proc. R. Soc. B.**, v.278, p.3161-3170, 2011.
- WILLIAMSON, S.H.; HUBISZ, M.J.; CLARK, A.G.; PAYSEUR, B.A.; BUSTAMANTE, C.D.; NIELSEN, R. Localizing recent adaptive evolution in the human genome. **PLoS Genet.**, 3(6): e90, 2007.
- WINTER, A.; KRÄMER, W.; WERNER, F.A.O.; KOLLERS, S.; KATA, S.; DURSTEWITZ, G.; BUITKAMP, J.; WOMACK, J.E.; THALLER, G.; FRIES, R. Association of a lysine-232/alanine polymorphism in a bovine gene encoding Acyl-Coa-Diacylglycerol Acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. **Proc. Nat. Acad. Sci. U.S.A.**, v.99, n.14, p.9300-9305, 2002.
- YANG, Z. Inference of selection from multiple species alignments. **Curr. Opin. Genet. Dev.**, v.12, p.688-694, 2002.
- YANG, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. **Mol. Biol. Evol.**, v.15, n.5, p.568-573, 1998.
- YANG, Z. PAML 4: Phylogenetic analysis by maximum likelihood. **Mol. Biol. Evol.**, v.24, n.8, p.1586-1591, 2007.

YANG, Z.; NIELSEN, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. **Mol. Biol. Evol.**, v.19, n.6, p.908-917, 2002.

YANG, Z.; NIELSEN, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. **J. Mol. Evol.**, v.46, p.409-418, 1998.

YANG, Z.; NIELSEN, R.; GOLDMAN, N.; PEDERSEN, A.M.K. Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics**, v.155, p.431-449, 2000.

YANG, Z.; WONG, W.S.W.; NIELSEN, R. Bayes empirical bayes inference of amino acid sites under positive selection. **Mol. Biol. Evol.**, v.22, n.4, p.1107-1118, 2005.

CAPÍTULO II

Identificação de assinaturas de seleção em ovinos de raças localmente adaptadas brasileiras

IDENTIFICAÇÃO DE ASSINATURAS DE SELEÇÃO EM OVINOS DE RAÇAS LOCALMENTE ADAPTADAS BRASILEIRAS

RESUMO

As raças localmente adaptadas podem ser uma fonte importante de informações que levem à descoberta/validação de genes/regiões genômicas envolvidos com o controle de características de interesse econômico e ecológico. Além disso, quanto maior o conhecimento sobre estas raças, maior o seu potencial de utilização nos sistemas de produção e menor o risco de extinção desta raça. As raças localmente adaptadas de ovinos brasileiras apresentam um enorme potencial de utilização e possuem especificidades ecológicas e produtivas que podem ser exploradas do ponto de vista genético. Portanto, este trabalho teve o objetivo de identificar e caracterizar regiões genômicas candidatas a estar sob seleção e que possam explicar estas particularidades. Animais pertencentes às raças Crioula Lanada, Morada Nova e Santa Inês foram genotipados para mais de 50 mil SNPs distribuídos ao longo do genoma ovino utilizando o *ovine SNP50 BeadChip* (Illumina®). A identificação de assinaturas de seleção foi realizada com base em dois grupos de metodologias: diferenciação entre populações (F_{ST}) e desequilíbrio de ligação (iHS e Rsb). A anotação das regiões genômicas identificadas possibilitou a identificação de 78 genes candidatos, sendo que destes, três foram identificados pelas três metodologias conjuntamente (GNS, NIPBL e SPEF2) e seis foram identificados através de duas metodologias (LOC100336620, KRT8, FUT9, WDR70, MGC134093 e LOC784368). Além disso, os genes CNTNAP5, DIS3L2, KIF1B, RXFP2 e TARBP1 foram identificados no presente estudo e em outros estudos envolvendo ovinos e bovinos, o que sugere que estes genes sejam candidatos importantes. A análise funcional possibilitou a identificação de genes relacionados com resposta imune (IL17C, IL18R1, IL7, IL1RL1, SMAD6, GPR65, SKAP2, THY1), desenvolvimento do sistema nervoso (SLITRK1, FUT9, ST8SIA4, CNTNAP2, CNTN3, THY1), diferenciação neuronal (SLITRK1, CNTNAP2, THY1), reprodução (SPAG6, RXFP2, MSTN) e percepção sensorial (TRPM8, OR13F1, COL1A1, TTC8). Dentre os genes identificados neste estudo podem ser citados o gene RXFP2, que

recentemente foi associado com presença/ausência e também com a morfologia dos chifres em ovinos; o gene EDARADD, importante na morfogênese e desenvolvimento do folículo piloso; o gene TRPM8, envolvido com a regulação da temperatura corporal em baixas temperaturas; os genes DIS3L2, PLAG1 e NIPBL, associados à variação de altura e por fim, os genes SPEF2 e SPAG6, importantes para a espermatogênese. Embora o presente estudo não possibilite a validação das relações genótipo - fenótipo sugeridas, abre-se uma nova possibilidade para o conhecimento de genes relacionados às diferenças/particularidades ecológicas e produtivas entre as raças de ovinos localmente adaptadas brasileiras.

Palavras-chave: arrasto seletivo, ambientes desfavoráveis, conservação

IDENTIFICATION OF SELECTION SIGNATURES IN BRAZILIAN LOCALLY ADAPTED SHEEP BREEDS

ABSTRACT

The locally adapted breeds can be an important source of information to the discovery/validation of genes/genomic regions involved in the control of economic and ecological important traits. Moreover, as much as the knowledge about a these increases, its extinction risk tends to decrease. The Brazilian locally adapted sheep breeds present an enormous potential and posses ecological and productive specificities that can be explored from a genetic point of view. Therefore, the aim of this study was to identify and characterize candidate genomic regions under selection that can explain these particularities. Individuals belonging the Brazilian Creole, Morada Nova and Santa Ines breeds were genotyped for approximately 50k SNP, distributed along the whole ovine genome with the *ovine SNP50 BeadChip* (Illumina®). The identificiation of signatures of selection was performed through two approaches: population differentiation (F_{ST}) and linkage disequilibrium (iHS and Rsb). The annotation of the genomic regions enabled the identification of 79 candidate genes. From these, three were detected using the three methods (GNS, NIPBL e SPEF2) and six were detected using two methods (LOC100336620, KRT8, FUT9, WDR70, MGC134093 e LOC784368). Furthermore, the genes CNTNAP5, DIS3L2, KIF1B, RXFP2 and TARBP1 were also suggested as candidates to be under selection by other studies involving ovine and bovine species, what suggests that these genes are important candidates. The functional analysis revealed genes involved with immune response (IL17C, IL18R1, IL7, IL1RL1, SMAD6, GPR65, SKAP2, THY1), nervous system development (SLITRK1, FUT9, ST8SIA4, CNTNAP2, CNTN3, THY1), neuronal differentiation (SLITRK1, CNTNAP2, THY1), reproduction (SPAG6, RXFP2, MSTN) and sensory perception (TRPM8, OR13F1, COL1A1, TTC8). Among the genes identified, some can be pinpointed such as the gene RXFP2, that recently was associated with horn presence/absence and horn development in sheep; EDARADD, which is important to the morphogenesis and development of hair follicle; TRPM8,

which is involved with the body temperature regulation; DIS3L2, PLAG1 and NIPBL, related to height variation and the genes SPEF2 and SPAG6, involved with spermatogenesis. Although this study could not validate the genotype – phenotype relationships suggested, it opens the possibility to deepen the knowledge about genes related to adaptative/productive traits that can be important to the Brazilian locally adapted sheep breeds.

Keywords: selective sweep, harsh environments, conservation

INTRODUÇÃO

A identificação de regiões genômicas sujeitas à seleção em animais de produção, pode auxiliar tanto no entendimento dos processos envolvidos na evolução dos genomas destas espécies, quanto na descoberta e validação de regiões genômicas envolvidas na manifestação de características de interesse econômico e ecológico (HAYES *et al.*, 2008; NIELSEN, 2001, 2005; SCHLÖTTERER, 2003).

O processo de domesticação associado com o desenvolvimento de raças e a evolução dos programas de melhoramento genético, propiciaram a modificação das espécies de interesse zootécnico do ponto de vista morfológico, comportamental, produtivo, adaptativo e, conseqüentemente genético, levando ao surgimento de uma enorme variedade de raças e tipos raciais adaptados às mais diversas condições ambientais e com variadas aptidões produtivas (DIAMOND, 2002; TORO; MÄKI-TANILA, 2007; FLORI *et al.*, 2009).

As raças de ovinos localmente adaptadas brasileiras são descendentes de indivíduos trazidos ao Brasil durante o período de colonização e, desde então, vêm sendo submetidas a processos evolutivos sistemáticos e não sistemáticos. Isso fez com que fossem desenvolvidos genótipos adaptados às diversas condições ambientais brasileiras (MARIANTE *et al.*, 2009).

Além da adaptação às diferentes condições ambientais brasileiras, estas raças apresentam algumas características interessantes do ponto de vista econômico. As raças deslanadas brasileiras, por exemplo, apresentam peles de melhor qualidade (JACINTO *et al.*, 2004, 2011) e menores níveis de infecção por nematódeos gastrintestinais quando comparadas com animais de raças lanadas exóticas (BUENO *et al.*, 2002; AMARANTE *et al.*, 2004).

O sucesso das estratégias de conservação de genótipos localmente adaptados depende primariamente do conhecimento do potencial de utilização destes genótipos (e da efetiva utilização destes indivíduos nos sistemas de produção) e de fatores que envolvem a estrutura genética destas populações. Felizmente, diversos estudos vêm sendo realizados no Brasil para subsidiar os programas de conservação e dar suporte aos programas de melhoramento genético que vêm sendo desenvolvidos para as raças de ovinos localmente adaptadas (MARIANTE *et al.*, 2009, MCMANUS *et al.*, 2010).

A recente publicação do genoma ovino (ISGC *et al.*, 2010) e os avanços nas tecnologias de genotipagem em larga escala vêm propiciando um grande aumento na capacidade de identificar genes/ regiões genômicas associadas à características de interesse econômico na espécie ovina (KIJAS *et al.*, 2012).

Portanto, os principais objetivos do presente estudo são identificar regiões genômicas candidatas a estar sob seleção em animais de raças ovinas localmente adaptadas brasileiras que possam explicar diferenças/particularidades ecológicas e produtivas entre estes indivíduos e, caracterizar estas regiões com a finalidade de propor genes candidatos.

MATERIAL E MÉTODOS

Animais e SNPs

Noventa e dois animais pertencentes às três principais raças de ovinos localmente adaptadas brasileiras (Crioula Lanada, Morada Nova e Santa Inês) foram genotipados como parte do Consórcio Internacional do Genoma Ovino (KIJAS *et al.*, 2012), utilizando o *ovine SNP50 BeadChip* (Illumina®).

Destes, oitenta e sete indivíduos não aparentados (22 Crioula Lanada, 20 Morada Nova e 45 Santa Inês) foram utilizados para a realização deste trabalho. As informações referentes aos locais de coleta dos indivíduos estão descritas na Tabela 1.

Tabela 1. Informações sobre os indivíduos utilizados

Raça	N	Rebanho	Local
Crioula Lanada	5	BCS_CAS	Caçapava do Sul, RS
	6	BCS_EMB	Bagé, RS
	5	BCS_LAG	Lages, SC
	6	BCS_POA	Ponte Alta, SC
Morada Nova	13	BMN_EBD	Itapetinga, BA
	7	BMN_EMB	Sobral, CE
Santa Inês	6	BSI_EBD	Itapetinga, BA
	5	BSI_ECO	Sobral, CE
	13	BSI_EMN	Teresina, PI
	10	BSI_ETC	Aracaju, SE
	11	BSI_GAA	Inhumas, GO

O *ovine SNP50 BeadChip* (Illumina®) contém 54241 SNPs distribuídos por todo o genoma ovino. Após a genotipagem, os dados passaram por um pré-processamento, restando 49034 SNPs com alta qualidade (KIJAS *et al.*, 2012).

Os dados então foram submetidos a um processamento adicional que consistiu na manutenção apenas dos SNPs localizados nos cromossomos autossômicos, de acordo com a versão 3.1 do genoma ovino (ISGC *et al.*, 2010, disponível em: <http://www.livestockgenomics.csiro.au/sheep/oar3.1.php>), e com MAF > 0,05 (considerando os 87 indivíduos utilizados neste estudo). Ao final deste processamento, restaram 44850 SNPs que foram utilizados para a realização deste trabalho.

Identificação de assinaturas de seleção com base na diferenciação entre populações

Esta análise foi realizada utilizando o software BayeScan v2.1 (FOLL; GAGGIOTTI, 2008). A estimativa dos valores de F_{ST} para cada *locus* é feita com base na distribuição de Dirichlet e a identificação dos *loci* candidatos a estarem sob seleção é feita à partir da decomposição destes valores de F_{ST} utilizando regressão logística, expressa por:

$$\log\left(\frac{F_{st}}{1 - F_{st}}\right) = \alpha_i + \beta_j$$

A probabilidade de um *locus* estar sob seleção é estimada à partir da distribuição posterior obtida de dois modelos alternativos utilizando um algoritmo de MCMC. O primeiro modelo leva em consideração a presença de seleção e inclui os parâmetros α_i e β_j , já o segundo não leva em consideração a presença da seleção e, portanto, inclui apenas o parâmetro β_j (FOLL; GAGGIOTTI, 2008). Valores de α_i diferentes de zero indicam que a seleção está atuando. Além disso, valores de α_i positivos indicam seleção positiva enquanto valores negativos indicam seleção balanceadora (FOLL; GAGGIOTTI, 2008).

A escolha do modelo (incluindo ou não seleção) deve ser feita com base na probabilidade posterior. Portanto, para controlar o número de falsos positivos, a probabilidade *a priori* do modelo neutro em relação ao modelo assumindo seleção foi considerada igual a 10 e a definição dos SNPs significativos foi realizada considerando o q-valor (o FDR mínimo para um *locus* ser considerado significativo) igual a 0,01.

A análise de F_{ST} foi realizada considerando as três populações conjuntamente e também separadamente para cada comparação duas a duas.

Identificação de assinaturas de seleção com base no desequilíbrio de ligação

Primeiramente, os haplótipos de cada indivíduo foram reconstruídos separadamente para cada cromossomo dentro de cada raça, utilizando o software fastPHASE (SCHEET; STEPHENS, 2006).

Então, as taxas de recombinação populacionais ($\rho = 4 \times N_e \times r$) foram estimadas separadamente para cada cromossomo dentro de cada raça, utilizando o software LDhat (MCVEAN *et al.*, 2004). Estas estimativas foram convertidas em distâncias genéticas (cM) considerando o tamanho efetivo populacional de 371 indivíduos para Crioula Lanada, 327 para Morada Nova e 520 para Santa Inês (KIJAS *et al.*, 2012). Um mapa genético de referência foi então criado à partir da média aritmética dos mapas de cada raça.

A informação dos alelos ancestrais necessária para estas análises foi obtida do Consórcio Internacional do Genoma Ovino (KIJAS *et al.*, 2012).

A identificação de assinaturas de seleção com base no desequilíbrio de ligação foi realizada utilizando duas metodologias complementares, o iHS, proposto por Voight *et al.* (2006) e o Rsb, proposto por Tang *et al.* (2007).

Estes dois métodos são derivados do teste do haplótipo longo (LRH), proposto por Sabeti *et al.* (2002) e se baseiam em uma medida de desequilíbrio de ligação denominada de haplótipo homozigoto estendido (EHH).

Considerando um marcador bialélico (por exemplo um SNP), o EHH pode ser definido como a probabilidade de dois cromossomos em uma população, que possuem um determinado alelo deste marcador, serem homozigotos para a região desde o SNP analisado até uma distância x deste marcador (SABETI *et al.*, 2002).

Matematicamente, o EHH pode ser representado por:

$$EHH_t = \frac{\sum_{i=1}^s \binom{e_{ti}}{2}}{\binom{c_t}{2}}$$

Onde c é a probabilidade de um SNP possuir um determinado alelo, e é a probabilidade de ocorrência de um determinado haplótipo estendido e s é o número de haplótipos estendidos possíveis (SABETI *et al.*, 2002).

Se assumirmos que um marcador bialélico possui um alelo ancestral e um alelo derivado, podemos calcular o EHH para o alelo ancestral e para o alelo derivado deste marcador.

Baseado nisso, Voight *et al.* (2006) propôs uma metodologia para comparar a taxa de decaimento do EHH entre o alelo ancestral e o alelo derivado de um determinado SNP.

O teste proposto por Voight *et al.* (2006), inicia-se com o cálculo do EHH para cada alelo de um determinado SNP a distâncias crescentes deste SNP. Então calcula-se a área sob a curva do gráfico EHH x distância, representada por iHH_A ou iHH_D (dependendo se o alelo é o ancestral ou o derivado, respectivamente).

Para possibilitar a comparação entre marcadores com frequências alélicas diferentes, deve-se padronizar os valores de iHS . Primeiramente divide-se os marcadores em classes de frequência (foram utilizadas classes de 0,05) e são calculadas a média e desvio padrão do iHS dentro de cada classe de frequência, então o valor padronizado de iHS é dado por:

$$iHS = \frac{\ln\left(\frac{iHH_A}{iHH_D}\right) - E_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}{SD_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}$$

A análise de iHS foi realizada utilizando o pacote rehh (GAUTIER; VITALIS, 2012). Para identificar os *loci* sujeitos à seleção, os valores de iHS foram transformados em $piHS$, através de:

$$piHS = -\log\left[1 - 2 \times (\Phi_{(iHS)} - 0,5)\right]$$

Onde $\Phi_{(x)}$ representa a função Gaussiana acumulada. Segundo Gautier e Vitalis (2012), assumindo que os valores de iHS são distribuídos de forma normal sob neutralidade, os valores de $piHS$ podem ser interpretados como $-\log_{10}(P - valor)$. A análise de iHS foi realizada separadamente para cada uma das três raças estudadas e foram considerados significativos, os SNPs que apresentaram $piHS \geq 4$.

A metodologia proposta por Tang *et al.* (2007), por sua vez, consiste na comparação do decaimento do EHH de um determinado SNP (EHHS) entre duas populações.

O EHHS pode ser definido como média do EHH para os dois alelos do SNP, ponderada pelo quadrados das respectivas frequências alélicas. Então calcula-se a área sob a curva do gráfico EHHS x distância para cada população (iES_{pop1} e iES_{pop2}) e o valor de Rsb é dado por:

$$Rsb = \frac{\log\left(\frac{iES_{pop1}}{iES_{pop2}}\right) - med\left[\log\left(\frac{iES_{pop1}}{iES_{pop2}}\right)\right]}{sd\left[\log\left(\frac{iES_{pop1}}{iES_{pop2}}\right)\right]}$$

A análise de Rsb foi realizada utilizando o pacote *rehh* (GAUTIER; VITALIS, 2012). Para identificar os *loci* sujeitos à seleção, os valores de Rsb foram transformados em $pRsb$, através de:

$$pRsb = -\log[1 - 2 \times (\Phi_{(Rsb)} - 0,5)]$$

Onde $\Phi_{(x)}$ representa a função Gaussiana acumulada. Segundo Gautier e Vitalis (2012), assumindo que os valores de Rsb são distribuídos de forma normal sob neutralidade, os valores de $pRsb$ podem ser interpretados como $-\log_{10}(P - valor)$. A análise de Rsb foi realizada para cada comparação duas a duas entre as raças estudadas e foram considerados significativos os SNPs que apresentaram $pRsb \geq 4$.

Anotação das regiões significativas

Os SNPs considerados significativos nas análises de F_{ST} , iHS e Rsb foram utilizados para identificar genes candidatos sob seleção. A identificação destes genes foi feita com base na versão 3.1 do genoma ovino (ISGC *et al.*, 2010, disponível em: <http://www.livestockgenomics.csiro.au/sheep/oar3.1.php>). Para cada análise, os SNPs significativos distantes menos de 500kb entre si foram agrupados em uma única região e o SNP com maior valor da estatística (F_{ST} , iHS ou Rsb) dentro de cada região foi considerado como o SNP significativo da região.

Quando um SNP significativo estava contido dentro de um gene, este gene foi considerado como candidato a estar sob seleção. Para os SNPs significativos não localizados dentro de genes, foram considerados candidatos os genes mais próximos deste SNP até uma distância máxima de 250kb.

Análise funcional dos genes candidatos

A plataforma DAVID (HUANG *et al.*, 2009, disponível em: <http://david.abcc.ncifcrf.gov/>) foi utilizada para realizar a análise funcional dos genes identificados como candidatos à estarem sob seleção. Além disso, foram utilizadas duas ferramentas de *text mining* para auxiliar a busca de literatura científica relacionada aos genes candidatos: IHOP (HOFFMAN; VALENCIA, 2004, disponível em: <http://www.ihop-net.org/>) e CHILIBOT (CHEN; SHARP, 2004, disponível em: <http://www.chilibot.net/>).

RESULTADOS

A análise de F_{ST} revelou a presença de 58 SNPs significativos ($q\text{-valor} \leq 0,01$) para a análise considerando as três raças, 33 SNPs significativos considerando a comparação Crioula Lanada x Morada Nova, 53 SNPs significativos para Crioula Lanada x Santa Inês e 23 SNPs significativos para Morada Nova x Santa Inês. As informações referentes aos SNPs significativos na análise de F_{ST} estão representadas na Tabela 2.

Tabela 2. SNPs significativos para as análises de F_{ST} .

Comparação	SNP	Cromossomo	Posição	q-valor	α	F_{ST}
Raças Brasileiras	s49628	1	6832517	0,0005	1,6314	0,3091
Raças Brasileiras	s00481	1	15853511	0,0063	1,4902	0,2837
Raças Brasileiras	OAR1_89287467	1	84029037	0,0053	1,3466	0,2564
Raças Brasileiras	OAR1_107797839	1	100477080	0,0055	1,4547	0,2774
Raças Brasileiras	OAR1_121001304	1	111963966	0,0040	1,3868	0,2634
Raças Brasileiras	OAR1_187375309_X	1	173910690	0,0051	1,3672	0,2603
Raças Brasileiras	OAR1_194627962	1	180296161	0,0080	1,5004	0,2869
Raças Brasileiras	OAR2_55853730	2	52254402	0,0001	1,6954	0,3213
Raças Brasileiras	OAR2_55861669	2	52266474	0,0002	1,7106	0,3244
Raças Brasileiras	s20468	2	52409379	0,0000	1,8421	0,3499
Raças Brasileiras	OAR2_146111824	2	137403728	0,0091	1,4261	0,2728
Raças Brasileiras	OAR2_226474680	2	213965644	0,0036	1,4622	0,2783
Raças Brasileiras	OAR3_105506271	3	99246368	0,0067	0,1344	0,2558
Raças Brasileiras	s09462	3	104064729	0,0068	1,3542	0,2583
Raças Brasileiras	s26177	3	154006814	0,0051	1,3553	0,2577

Comparação	SNP	Cromossomo	Posição	q-valor	α	F _{ST}
Raças Brasileiras	OAR3_164788310	3	154069702	0,0013	1,4079	0,2668
Raças Brasileiras	s16949	3	154107772	0,0007	1,4878	0,2806
Raças Brasileiras	s69653	3	154156748	0,0001	1,6445	0,3099
Raças Brasileiras	OAR3_165050963	3	154252449	0,0000	1,8209	0,3444
Raças Brasileiras	OAR3_173071993	3	161858143	0,0057	1,3615	0,2591
Raças Brasileiras	s62626	3	217731909	0,0007	1,6546	0,3147
Raças Brasileiras	s32131	4	21287297	0,0083	1,4604	0,2788
Raças Brasileiras	OAR4_25990541	4	24787710	0,0002	1,5022	0,2836
Raças Brasileiras	OAR4_63158847	4	59789032	0,0021	1,4085	0,2675
Raças Brasileiras	OAR4_73170316	4	69208152	0,0078	1,4892	0,2846
Raças Brasileiras	OAR4_114392312	4	106938140	0,0028	1,5242	0,2897
Raças Brasileiras	OAR4_117150483	4	109733886	0,0075	1,5053	0,2871
Raças Brasileiras	OAR5_54542816	5	50275508	0,0031	1,5428	0,2935
Raças Brasileiras	s73701	6	71533740	0,0010	1,5477	0,2938
Raças Brasileiras	OAR7_13638892	7	13297119	0,0012	1,5339	0,2911
Raças Brasileiras	OAR7_21409209	7	20556382	0,0067	1,3845	0,2634
Raças Brasileiras	OAR7_106207879	7	97408830	0,0004	1,6551	0,3135
Raças Brasileiras	OAR8_40885595	8	37942516	0,0026	0,1499	0,2857
Raças Brasileiras	OAR9_38099807	9	36163638	0,0019	1,4421	0,2737
Raças Brasileiras	OAR9_57751897	9	55056846	0,0040	1,5215	0,2896
Raças Brasileiras	s31334	9	83420640	0,0098	1,3302	0,2545
Raças Brasileiras	OAR10_29538398	10	29502667	0,0002	1,7589	0,3346
Raças Brasileiras	OAR10_86298190	10	79129227	0,0071	1,4505	0,2768
Raças Brasileiras	OAR11_38437597	11	35982089	0,0075	0,1344	0,2564
Raças Brasileiras	OAR13_27814492	13	25164063	0,0008	1,6352	0,3105
Raças Brasileiras	s58362	14	13401157	0,0045	1,3635	0,2601
Raças Brasileiras	OAR14_14862862	14	14657550	0,0005	1,5291	0,2892
Raças Brasileiras	OAR15_45152619	15	43007145	0,0088	1,3675	0,2615
Raças Brasileiras	OAR16_39840363	16	36642330	0,0036	1,3403	0,2551
Raças Brasileiras	OAR16_39888776	16	36697783	0,0007	1,5109	0,2856
Raças Brasileiras	OAR16_41681642	16	38418208	0,0085	1,3187	0,2521
Raças Brasileiras	s29275	16	44099474	0,0042	1,5111	0,2877
Raças Brasileiras	OAR18_26087486	18	25173878	0,0093	0,1468	0,2807
Raças Brasileiras	s14823	19	15080853	0,0047	1,5072	0,2872
Raças Brasileiras	OAR19_28888123	19	27281255	0,0002	1,6609	0,3148
Raças Brasileiras	OAR20_45964534	20	42334266	0,0019	1,5268	0,2905
Raças Brasileiras	s27532	22	48966180	0,0096	1,4498	0,2774
Raças Brasileiras	DU388965_586	25	6062510	0,0061	1,3654	0,2597
Raças Brasileiras	s22440	25	7171415	0,0024	1,4025	0,2661
Raças Brasileiras	s30024	25	7356301	0,0003	1,5424	0,2913
Raças Brasileiras	s25195	25	7392689	0,0000	1,6199	0,3057
Raças Brasileiras	s09722	25	9173663	0,0001	1,6167	0,3048
Raças Brasileiras	s45791	26	35828422	0,0059	1,3922	0,2648

Comparação	SNP	Cromossomo	Posição	q-valor	α	F _{ST}
BCS x BMN	OAR1_63040872	1	59625317	0,0099	1,5651	0,3270
BCS x BMN	OAR1_107588571	1	100269214	0,0081	1,4266	0,3016
BCS x BMN	OAR1_107797839	1	100477080	0,0011	1,7442	0,3577
BCS x BMN	OAR1_121001304	1	111963966	0,0066	1,4778	0,3106
BCS x BMN	OAR2_55853730	2	52254402	0,0002	1,9249	0,3908
BCS x BMN	OAR2_55861669	2	52266474	0,0005	1,9230	0,3904
BCS x BMN	OAR2_146111824	2	137403728	0,0013	1,7249	0,3546
BCS x BMN	s09462	3	104064729	0,0039	1,4850	0,3115
BCS x BMN	OAR3_173071993	3	161858143	0,0047	1,4686	0,3086
BCS x BMN	s32131	4	21287297	0,0055	1,5978	0,3322
BCS x BMN	OAR4_25990541	4	24787710	0,0006	1,6540	0,3407
BCS x BMN	s45195	4	105363106	0,0028	1,6732	0,3455
BCS x BMN	s73701	6	71533740	0,0008	1,8334	0,3739
BCS x BMN	OAR6_95454480	6	87098156	0,0076	1,5701	0,3279
BCS x BMN	s33707	6	113378122	0,0071	1,5718	0,3279
BCS x BMN	s16493	6	115855240	0,0034	1,6400	0,3395
BCS x BMN	OAR7_13638892	7	13297119	0,0011	1,7965	0,3673
BCS x BMN	OAR7_106207879	7	97408830	0,0031	1,6389	0,3393
BCS x BMN	OAR8_40885595	8	37942516	0,0003	1,7793	0,3638
BCS x BMN	OAR9_57751897	9	55056846	0,0017	1,7151	0,3527
BCS x BMN	OAR10_86298190	10	79129227	0,0015	1,7618	0,3612
BCS x BMN	OAR12_4615833	12	1255316	0,0090	1,5675	0,3273
BCS x BMN	s35770	12	17914935	0,0095	1,5756	0,3287
BCS x BMN	s23962	13	7638271	0,0020	1,7267	0,3549
BCS x BMN	OAR14_14862862	14	14657550	0,0025	1,5205	0,3175
BCS x BMN	OAR15_76864698	15	70944492	0,0063	1,6236	0,3368
BCS x BMN	OAR17_38962935	17	36065164	0,0063	1,6082	0,3341
BCS x BMN	OAR19_28888123	19	27281255	0,0043	1,5969	0,3320
BCS x BMN	DU388965_586	25	6062510	0,0025	1,5155	0,3165
BCS x BMN	OAR25_6527496	25	6923544	0,0055	1,5932	0,3315
BCS x BMN	s30024	25	7356301	0,0008	1,6575	0,3413
BCS x BMN	s25195	25	7392689	0,0005	1,7124	0,3510
BCS x BMN	s25158	26	27431180	0,0086	1,5642	0,3268
BCS x BSI	s49628	1	6832517	0,0011	1,7014	0,3036
BCS x BSI	OAR1_33544806	1	32810494	0,0002	1,6354	0,2878
BCS x BSI	s11273	1	175176338	0,0085	1,3922	0,2448
BCS x BSI	s30608	1	214872825	0,0067	1,3865	0,2432
BCS x BSI	OAR1_236278124	1	219029998	0,0092	1,3738	0,2421
BCS x BSI	OAR1_241765164	1	224216189	0,0001	1,7365	0,3093
BCS x BSI	OAR2_55853730	2	52254402	0,0020	1,4685	0,2574
BCS x BSI	OAR2_55861669	2	52266474	0,0018	1,4713	0,2577
BCS x BSI	OAR2_76547571	2	71844617	0,0062	1,3997	0,2456
BCS x BSI	OAR2_145195113	2	136512593	0,0099	1,4222	0,2507

Comparação	SNP	Cromossomo	Posição	q-valor	α	F _{ST}
BCS x BSI	OAR2_226474680	2	213965644	0,0000	1,6599	0,2920
BCS x BSI	OAR3_44188902	3	41281473	0,0090	1,3737	0,2416
BCS x BSI	OAR3_105506271	3	99246368	0,0023	1,4729	0,2586
BCS x BSI	s69653	3	154156748	0,0002	1,7093	0,3013
BCS x BSI	OAR3_165050963	3	154252449	0,0003	1,6564	0,2924
BCS x BSI	s62626	3	217731909	0,0000	1,9992	0,3621
BCS x BSI	OAR4_62927453	4	59566544	0,0006	1,5609	0,2746
BCS x BSI	OAR4_63158847	4	59789032	0,0038	1,4733	0,2592
BCS x BSI	OAR5_57265711	5	52840602	0,0021	1,5044	0,2638
BCS x BSI	OAR5_96703012	5	88661821	0,0046	1,5616	0,2774
BCS x BSI	OAR6_110447914	6	100354946	0,0054	1,4135	0,2484
BCS x BSI	OAR7_21170284	7	20314928	0,0043	1,3986	0,2456
BCS x BSI	OAR8_34350484	8	31560321	0,0064	1,4013	0,2460
BCS x BSI	OAR9_64748830	9	61608126	0,0057	1,5709	0,2798
BCS x BSI	OAR9_76865315	9	72430891	0,0016	1,4760	0,2583
BCS x BSI	OAR10_29538398	10	29502667	0,0000	2,0882	0,3807
BCS x BSI	OAR13_27814492	13	25164063	0,0015	1,6575	0,2951
BCS x BSI	s58362	14	13401157	0,0003	1,5619	0,2738
BCS x BSI	OAR14_14862862	14	14657550	0,0003	1,6204	0,2845
BCS x BSI	OAR15_39364437	15	37497280	0,0082	1,3937	0,2448
BCS x BSI	OAR15_39396547	15	37529233	0,0032	1,4345	0,2513
BCS x BSI	OAR15_45152619	15	43007145	0,0015	1,4781	0,2588
BCS x BSI	s68098	15	51474825	0,0009	1,5196	0,2661
BCS x BSI	OAR15_78548290	15	72470215	0,0099	1,3698	0,2409
BCS x BSI	s27911	15	72836601	0,0049	1,4244	0,2494
BCS x BSI	OAR16_30364510	16	27985457	0,0069	1,5672	0,2787
BCS x BSI	OAR16_39840363	16	36642330	0,0025	1,4617	0,2564
BCS x BSI	OAR16_39888776	16	36697783	0,0003	1,5810	0,2775
BCS x BSI	OAR16_39987451	16	36797401	0,0062	1,3885	0,2436
BCS x BSI	OAR16_40589074	16	37359778	0,0072	1,4297	0,2515
BCS x BSI	OAR16_41681642	16	38418208	0,0007	1,5151	0,2656
BCS x BSI	OAR16_42159705	16	38851083	0,0041	1,4357	0,2517
BCS x BSI	OAR16_43758108	16	40258646	0,0035	1,4609	0,2570
BCS x BSI	s43481	16	40936379	0,0075	1,3761	0,2417
BCS x BSI	s29275	16	44099474	0,0082	1,5448	0,2748
BCS x BSI	OAR17_25214825	17	22801020	0,0099	1,3738	0,2417
BCS x BSI	OAR20_12280341	20	11752113	0,0090	1,4346	0,2535
BCS x BSI	OAR20_54332894	20	49862860	0,0030	1,4403	0,2529
BCS x BSI	OAR22_18486216	22	15001318	0,0077	1,3819	0,2427
BCS x BSI	s08282	24	40179300	0,0028	1,6148	0,2873
BCS x BSI	s09722	25	9173663	0,0001	1,7381	0,3070
BCS x BSI	s58657	26	5777905	0,0004	1,5715	0,2760
BCS x BSI	s45791	26	35828422	0,0051	1,4651	0,2584

Comparação	SNP	Cromossomo	Posição	q-valor	α	F_{ST}
BMN x BSI	s49070	1	109251246	0,0067	1,5309	0,1998
BMN x BSI	OAR1_159357393	1	147515992	0,0090	1,4778	0,1935
BMN x BSI	OAR1_175474366	1	162709463	0,0030	1,7827	0,2322
BMN x BSI	s20468	2	52409379	0,0000	2,1610	0,2820
BMN x BSI	OAR2_56383269	2	52536647	0,0074	1,5086	0,1971
BMN x BSI	OAR2_57196512_X	2	53304271	0,0024	1,6155	0,2097
BMN x BSI	OAR3_159338437	3	149423233	0,0041	1,7416	0,2273
BMN x BSI	OAR3_165050963	3	154252449	0,0011	1,6041	0,2080
BMN x BSI	s65009	3	221659279	0,0006	1,9153	0,2492
BMN x BSI	s55866	3	221672349	0,0100	1,4555	0,1909
BMN x BSI	OAR4_12662221	4	12464561	0,0008	1,8189	0,2365
BMN x BSI	s32131	4	21287297	0,0027	1,7942	0,2338
BMN x BSI	OAR4_73170316	4	69208152	0,0013	1,8451	0,2403
BMN x BSI	s11164	4	70385766	0,0021	1,8326	0,2386
BMN x BSI	OAR4_117150483	4	109733886	0,0003	1,9631	0,2556
BMN x BSI	OAR6_39029427	6	34919579	0,0018	1,5948	0,2069
BMN x BSI	OAR7_18037906	7	17392897	0,0060	1,6958	0,2216
BMN x BSI	OAR9_28089762	9	26786744	0,0047	1,7359	0,2265
BMN x BSI	OAR9_38099807	9	36163638	0,0006	1,7571	0,2272
BMN x BSI	OAR9_44636659	9	42465998	0,0053	1,5332	0,1997
BMN x BSI	OAR11_38437597	11	35982089	0,0081	1,4906	0,1949
BMN x BSI	OAR18_65638912	18	61611739	0,0036	1,7742	0,2313
BMN x BSI	s68206	20	44373915	0,0033	1,7604	0,2295

Os sinais de seleção identificados estão distribuídos por 23 dos 26 cromossomos autossômicos do genoma ovino. Apenas não foram identificados SNPs significativos pela análise de F_{ST} nos cromossomos OAR12, OAR21 e OAR24 (Figura 1). É interessante notar que nos cromossomos OAR2, OAR3, OAR16 e OAR25 há a presença de *clusters* de SNPs significativos (Figura 1), o que nos leva a acreditar que estas regiões podem ser fortes candidatas a conter genes sob seleção.

Os SNPs identificados para cada análise individualmente foram agrupados de modo que apenas os SNPs significativos tanto para a análise considerando as três raças como para pelo menos uma comparação duas a duas foram utilizados para as análises posteriores (Tabela 3). Sendo assim, restaram 42 SNPs distribuídos ao longo dos cromossomos OAR1, OAR2, OAR3, OAR4, OAR6, OAR7, OAR8, OAR9, OAR10, OAR11, OAR13, OAR14, OAR15, OAR16, OAR19, OAR 25 e OAR26 (Figura 2), que foram utilizados para definir as regiões genômicas a serem anotadas em busca de genes candidatos a estar sob seleção.

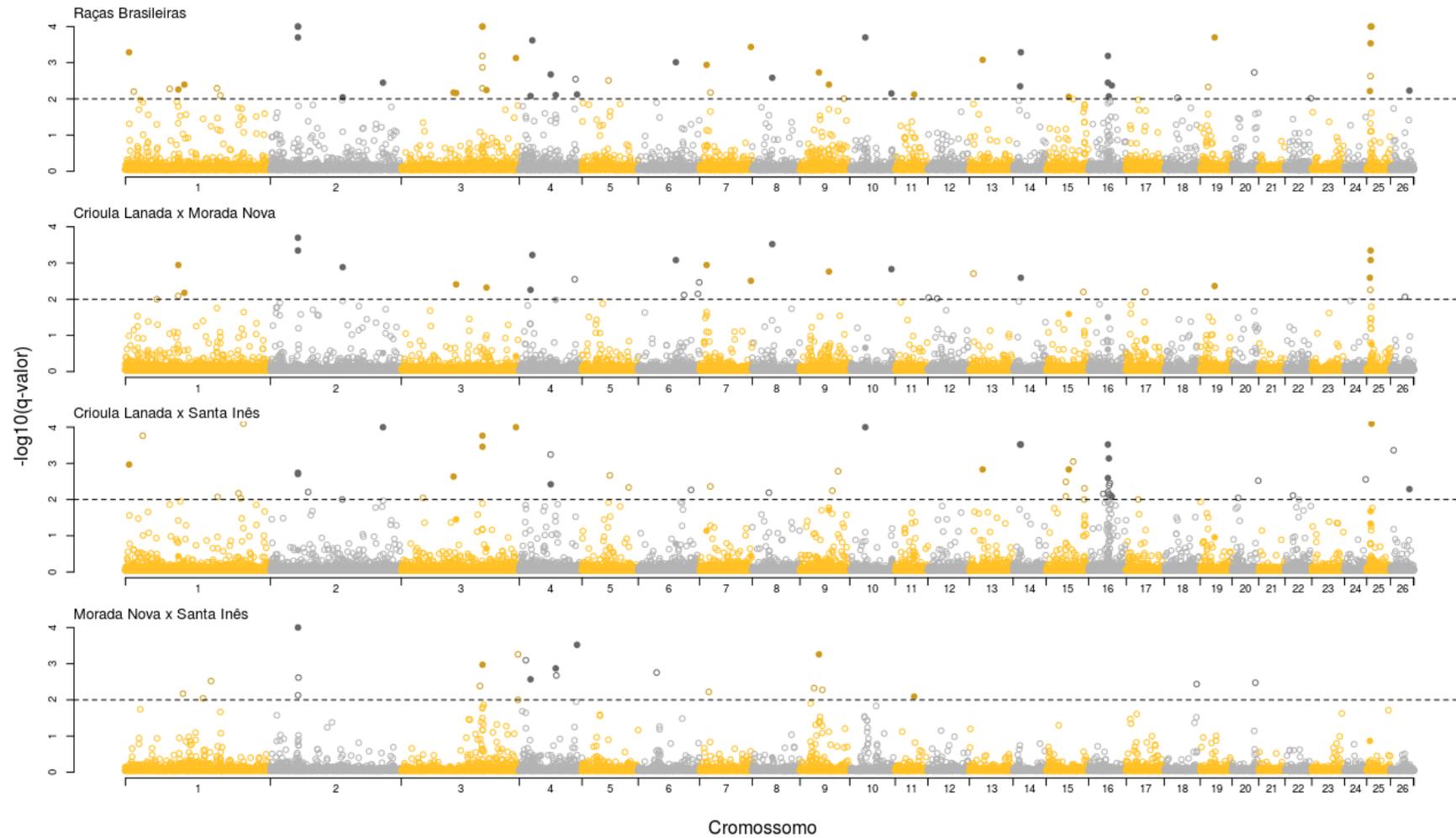


Figura 1. Manhattan plot para a análise de F_{ST} . Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs significativos, sendo que os pontos preenchidos representam os SNPs significativos na análise envolvendo as três raças brasileiras e pelo menos uma comparação duas a duas.

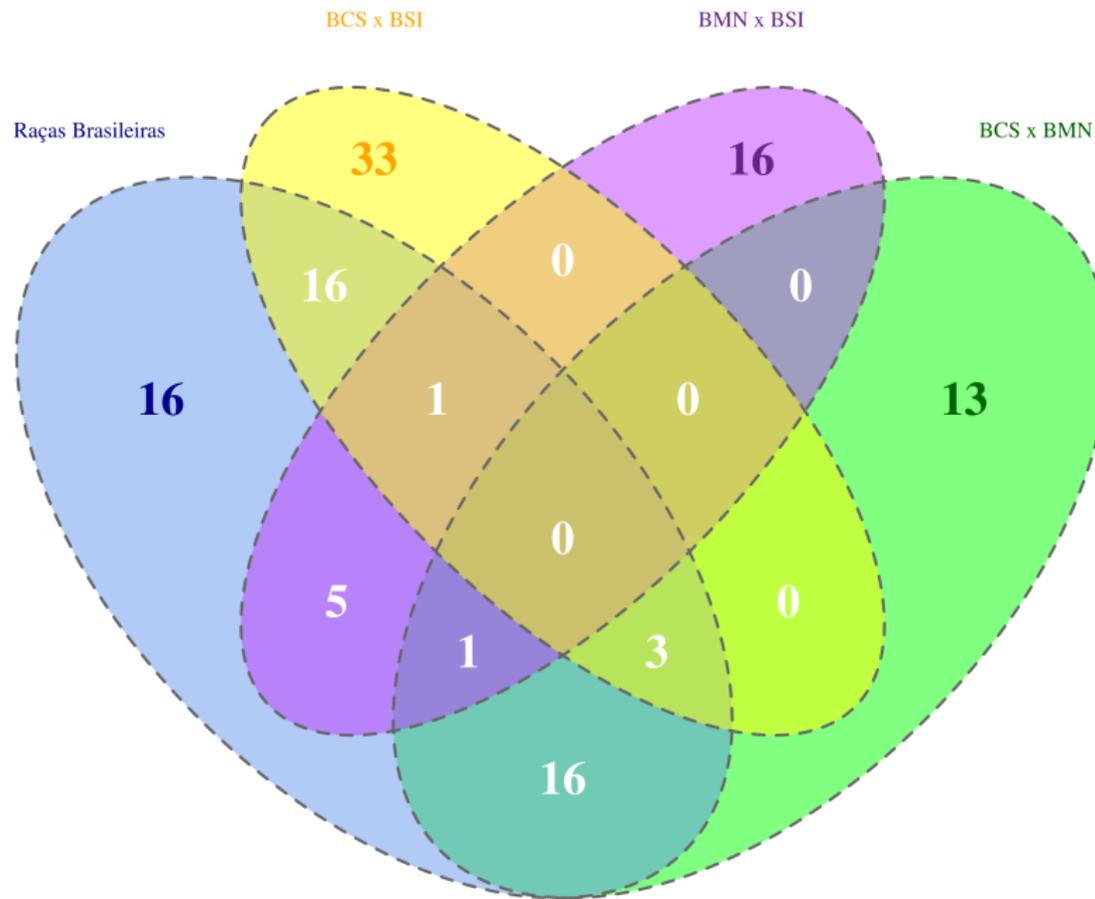


Figura 2. Diagrama de Venn ilustrando os SNPs significativos para a análise de F_{ST} realizada com as três raças localmente adaptadas (Raças Brasileiras), e com as comparações duas a duas entre as raças (BCS x BMN, BCS x BSI e BMN x BSI).

Tabela 3. SNPs significativos tanto para a análise de F_{ST} realizada com as três raças localmente adaptadas quanto com pelo menos uma das comparações duas a duas entre as raças.

SNP	Cromossomo	Posição	q-valor ¹	α^1	F_{ST}^1	q-valor ²	α^2	F_{ST}^2	q-valor ³	α^3	F_{ST}^3	q-valor ⁴	α^4	F_{ST}^4
s49628	1	6832517	0,0005	1,6314	0,3091	0,7207	0,0997	0,1174	0,0011	1,7014	0,3036	0,9002	0,0129	0,0590
OAR1_107797839	1	100477080	0,0055	1,4547	0,2774	0,0011	1,7442	0,3577	0,3677	0,3306	0,1089	0,6290	0,1572	0,0714
OAR1_121001304	1	111963966	0,0040	1,3868	0,2634	0,0066	1,4778	0,3106	0,5675	0,1486	0,0875	0,6622	0,1370	0,0689
OAR2_55853730	2	52254402	0,0001	1,6954	0,3213	0,0002	1,9249	0,3908	0,0020	1,4685	0,2574	0,8019	0,0663	0,0633
OAR2_55861669	2	52266474	0,0002	1,7106	0,3244	0,0005	1,9230	0,3904	0,0018	1,4713	0,2577	0,7976	0,0618	0,0629
s20468	2	52409379	0,0000	1,8421	0,3499	0,8640	0,0468	0,1107	0,2502	0,5401	0,1342	0,0000	2,1610	0,2820
OAR2_146111824	2	137403728	0,0091	1,4261	0,2728	0,0013	1,7249	0,3546	0,4188	0,2652	0,1010	0,6738	0,1365	0,0695
OAR2_226474680	2	213965644	0,0036	1,4622	0,2783	0,3053	0,4958	0,1696	0,0000	1,6599	0,2920	0,8775	-0,0380	0,0565
OAR3_105506271	3	99246368	0,0067	1,3440	0,2558	0,8979	0,0468	0,1099	0,0023	1,4729	0,2586	0,9033	0,0083	0,0586
s09462	3	104064729	0,0068	1,3542	0,2583	0,0039	1,4850	0,3115	0,0353	1,2215	0,2194	0,9058	0,0147	0,0589
s69653	3	154156748	0,0001	1,6445	0,3099	0,9012	0,0179	0,1069	0,0002	1,7093	0,3013	0,5631	0,1946	0,0737
OAR3_165050963	3	154252449	0,0000	1,8209	0,3444	0,8280	-0,0345	0,1035	0,0003	1,6564	0,2924	0,0011	1,6041	0,2080
OAR3_173071993	3	161858143	0,0057	1,3615	0,2591	0,0047	1,4686	0,3086	0,2233	0,6107	0,1428	0,8811	0,0533	0,0617
s62626	3	217731909	0,0007	1,6546	0,3147	0,3759	0,3840	0,1551	0,0000	1,9992	0,3621	0,8900	-0,0098	0,0578
s32131	4	21287297	0,0083	1,4604	0,2788	0,0055	1,5978	0,3322	0,8938	-0,0181	0,0705	0,0027	1,7942	0,2338
OAR4_25990541	4	24787710	0,0002	1,5022	0,2836	0,0006	1,6540	0,3407	0,4856	0,1905	0,0920	0,4268	0,3926	0,0920
OAR4_63158847	4	59789032	0,0021	1,4085	0,2675	0,4992	0,2410	0,1366	0,0038	1,4733	0,2592	0,8705	-0,0360	0,0565
OAR4_73170316	4	69208152	0,0078	1,4892	0,2846	0,0104	1,5632	0,3266	0,8831	-0,0245	0,0702	0,0013	1,8451	0,2403
OAR4_117150483	4	109733886	0,0075	1,5053	0,2871	0,6251	0,1451	0,1234	0,9009	0,0184	0,0729	0,0003	1,9631	0,2556
s73701	6	71533740	0,0010	1,5477	0,2938	0,0008	1,8334	0,3739	0,1665	0,7230	0,1565	0,6510	0,1572	0,0715
OAR7_13638892	7	13297119	0,0012	1,5339	0,2911	0,0011	1,7965	0,3673	0,0730	1,0256	0,1934	0,7540	0,1004	0,0663
OAR7_106207879	7	97408830	0,0004	1,6551	0,3135	0,0031	1,6389	0,3393	0,3546	0,4013	0,1215	0,6424	0,1520	0,0703
OAR8_40885595	8	37942516	0,0026	1,4990	0,2857	0,0003	1,7793	0,3638	0,6302	0,1176	0,0833	0,3860	0,5087	0,1048
OAR9_38099807	9	36163638	0,0019	1,4421	0,2737	0,8954	0,0375	0,1090	0,8842	0,0488	0,0756	0,0006	1,7571	0,2272
OAR9_57751897	9	55056846	0,0040	1,5215	0,2896	0,0017	1,7151	0,3527	0,0195	1,3109	0,2319	0,8753	0,0451	0,0617
OAR10_29538398	10	29502667	0,0002	1,7589	0,3346	0,2185	0,6921	0,1969	0,0000	2,0882	0,3807	0,8889	-0,0108	0,0580
OAR10_86298190	10	79129227	0,0071	1,4505	0,2768	0,0015	1,7618	0,3612	0,4672	0,2088	0,0942	0,5784	0,2017	0,0757
OAR11_38437597	11	35982089	0,0075	1,3440	0,2564	0,8846	-0,0165	0,1044	0,4104	0,2737	0,1022	0,0081	1,4906	0,1949
OAR13_27814492	13	25164063	0,0008	1,6352	0,3105	0,8409	0,0632	0,1128	0,0015	1,6575	0,2951	0,9051	0,0315	0,0600

SNP	Cromossomo	Posição	q-valor ¹	α^1	F_{ST}^1	q-valor ²	α^2	F_{ST}^2	q-valor ³	α^3	F_{ST}^3	q-valor ⁴	α^4	F_{ST}^4
s58362	14	13401157	0,0045	1,3635	0,2601	0,7830	0,0673	0,1126	0,0003	1,5619	0,2738	0,9036	-0,0019	0,0578
OAR14_14862862	14	14657550	0,0005	1,5291	0,2892	0,0025	1,5205	0,3175	0,0003	1,6204	0,2845	0,8992	-0,0226	0,0571
OAR15_45152619	15	43007145	0,0088	1,3675	0,2615	0,0256	1,3355	0,2877	0,0015	1,4781	0,2588	0,8883	-0,0274	0,0570
OAR16_39840363	16	36642330	0,0036	1,3403	0,2551	0,2444	0,6084	0,1854	0,0025	1,4617	0,2564	0,8910	-0,0327	0,0568
OAR16_39888776	16	36697783	0,0007	1,5109	0,2856	0,0316	1,2724	0,2781	0,0003	1,5810	0,2775	0,8981	-0,0267	0,0570
OAR16_41681642	16	38418208	0,0085	1,3187	0,2521	0,7613	0,0781	0,1138	0,0007	1,5151	0,2656	0,8961	-0,0141	0,0575
s29275	16	44099474	0,0042	1,5111	0,2877	0,8727	0,0556	0,1120	0,0082	1,5448	0,2748	0,9028	0,0223	0,0593
OAR19_28888123	19	27281255	0,0002	1,6609	0,3148	0,0043	1,5969	0,3320	0,1099	1,0149	0,2009	0,8122	0,0666	0,0627
DU388965_586	25	6062510	0,0061	1,3654	0,2597	0,0025	1,5155	0,3165	0,8946	0,0472	0,0755	0,1358	0,9917	0,1469
s30024	25	7356301	0,0003	1,5424	0,2913	0,0008	1,6575	0,3413	0,0461	1,1568	0,2108	0,7778	0,0705	0,0631
s25195	25	7392689	0,0000	1,6199	0,3057	0,0005	1,7124	0,3510	0,0210	1,3284	0,2358	0,7850	0,0680	0,0630
s09722	25	9173663	0,0001	1,6167	0,3048	0,1732	0,8233	0,2151	0,0000	1,7381	0,3070	0,8837	-0,0215	0,0574
s45791	26	35828422	0,0059	1,3922	0,2648	0,8737	0,0528	0,1109	0,0051	1,4651	0,2584	0,9058	0,0180	0,0590

¹Raças Brasileiras

²Crioula Lanada x MoradaNova

³Crioula Lanada x Santa Inês

⁴Morada Nova x Santa Inês

O agrupamento dos 42 SNPs resultou na definição de 37 regiões genômicas que foram consideradas para a anotação e identificação dos genes candidatos a estar sob seleção (Tabela 4).

As principais regiões identificadas pelas análises de F_{ST} foram: uma região no cromossomo OAR2 contendo um *cluster* de três SNPs (OAR2_55853730, OAR2_55861669 e s20468), uma região no cromossomo OAR3 contendo um *cluster* de dois SNPs (s69653, OAR3_165050963), uma região no cromossomo OAR16 contendo um *cluster* de dois SNPs (OAR16_39840363, OAR16_39888776) e uma região do cromossomo OAR25 contendo um *cluster* de dois SNPs (s30024, s25195).

Tabela 4. Regiões genômicas candidatas a conter genes sob seleção identificadas através das análises de F_{ST} .

Região	SNP	Cromossomo	Posição	Início	Final
1	s49628	1	6832517	6582517	7082517
2	OAR1_107797839	1	100477080	100227080	100727080
3	OAR1_121001304	1	111963966	111713966	112213966
4	OAR2_55853730; OAR2_55861669; s20468*	2	52409379	52004402	52659379
5	OAR2_146111824	2	137403728	137153728	137653728
6	OAR2_226474680	2	213965644	213715644	214215644
7	OAR3_105506271	3	99246368	98996368	99496368
8	s09462	3	104064729	103814729	104314729
9	s69653; OAR3_165050963*	3	154252449	153906748	154502449
10	OAR3_173071993	3	161858143	161608143	162108143
11	s62626	3	217731909	217481909	217981909
12	s32131	4	21287297	21037297	21537297
13	OAR4_25990541	4	24787710	24537710	25037710
14	OAR4_63158847	4	59789032	59539032	60039032
15	OAR4_73170316	4	69208152	68958152	69458152
16	OAR4_117150483	4	109733886	109483886	109983886
17	s73701	6	71533740	71283740	71783740
18	OAR7_13638892	7	13297119	13047119	13547119
19	OAR7_106207879	7	97408830	97158830	97658830
20	OAR8_40885595	8	37942516	37692516	38192516
21	OAR9_38099807	9	36163638	35913638	36413638
22	OAR9_57751897	9	55056846	54806846	55306846

23	OAR10_29538398	10	29502667	29252667	29752667
24	OAR10_86298190	10	79129227	78879227	79379227
25	OAR11_38437597	11	35982089	35732089	36232089
26	OAR13_27814492	13	25164063	24914063	25414063
27	s58362	14	13401157	13151157	13651157
28	OAR14_14862862	14	14657550	14407550	14907550
29	OAR15_45152619	15	43007145	42757145	43257145
30	OAR16_39840363; OAR16_39888776*	16	36697783	36392330	36947783
31	OAR16_41681642	16	38418208	38168208	38668208
32	s29275	16	44099474	43849474	44349474
33	OAR19_28888123	19	27281255	27031255	27531255
34	DU388965_586	25	6062510	5812510	6312510
35	s30024; s25195*	25	7392689	7106301	7642689
36	s09722	25	9173663	8923663	9423663
37	s45791	26	35828422	35578422	36078422

*SNP com maior valor de F_{ST} dentro da região.

A análise de iHS para a raça Crioula Lanada possibilitou a identificação de 13 SNPs significativos ($\pi iHS \geq 4$) distribuídos pelos cromossomos OAR2, OAR3, OAR6, OAR9, OAR12, OAR13, OAR21 e OAR25. Na raça Morada Nova foram identificados oito SNPs significativos distribuídos pelos cromossomos OAR3, OAR5, OAR8, OAR10 e OAR16. Já na raça Santa Inês, foram identificados 70 SNPs significativos distribuídos pelos cromossomos OAR1, OAR2, OAR3, OAR6, OAR9, OAR15 e OAR16. As informações referentes aos SNPs significativos na análise de iHS estão representados na Tabela 5. É interessante notar a presença de clusters de SNPs significativos nos cromossomos OAR2, OAR3 e OAR16 (Figura 3), o que sugere que estas regiões podem ser fortes candidatas a conter genes sob seleção.

Tabela 5. SNPs significativos para as análises de iHS .

Raça	SNP	Cromossomo	Posição	πiHS	iHS
Crioula Lanada	OAR2_119411966	2	111172962	4,0714	3,9303
Crioula Lanada	OAR2_126354465	2	118180362	4,3559	4,0850
Crioula Lanada	OAR2_137134681	2	128762227	4,3880	4,1022
Crioula Lanada	OAR2_200920481	2	189539157	5,5552	-4,6861
Crioula Lanada	OAR2_205533418	2	193984737	4,1680	3,9834
Crioula Lanada	OAR3_151019080	3	141203065	4,6118	4,2199

Raça	SNP	Cromossomo	Posição	piHS	iHS
Crioula Lanada	OAR6_57182975	6	51814853	4,7955	4,3142
Crioula Lanada	OAR9_10077769	9	10082948	4,2158	4,0095
Crioula Lanada	OAR9_67039654	9	63523344	4,5752	-4,2008
Crioula Lanada	s69940	12	41343197	5,0094	4,4218
Crioula Lanada	OAR13_49971490	13	46647627	4,0629	-3,9256
Crioula Lanada	s44390	21	18528589	4,6555	-4,2425
Crioula Lanada	OAR25_23974308	25	22965788	4,5279	-4,1761
Morada Nova	OAR3_10304371	3	9783340	4,0283	-3,9064
Morada Nova	s42902	3	10378941	4,0819	-3,9361
Morada Nova	OAR3_11033570	3	10515273	4,5818	-4,2042
Morada Nova	OAR3_12289153	3	11712721	4,3197	4,0656
Morada Nova	OAR5_106230044	5	97597361	4,0718	3,9305
Morada Nova	OAR8_44102022	8	41015484	4,2546	4,0306
Morada Nova	OAR10_66804684	10	64726524	4,7902	-4,3116
Morada Nova	OAR16_51422323	16	47245495	4,1708	-3,9850
Santa Inês	OAR1_146213099	1	135459627	4,0206	-3,9021
Santa Inês	s36668	2	37506606	4,2740	4,0411
Santa Inês	s35831	2	50934632	4,4882	4,1553
Santa Inês	OAR2_63699523	2	59231945	4,3243	-4,0681
Santa Inês	DU184544_326	3	108058182	4,1548	3,9762
Santa Inês	s52741	3	129451837	6,3581	5,0515
Santa Inês	OAR3_138290871	3	129648614	6,8681	5,2712
Santa Inês	OAR3_141611507	3	132505054	4,6860	4,2582
Santa Inês	OAR3_142056197	3	132852500	4,8472	-4,3405
Santa Inês	s33601	3	133314721	4,0236	3,9038
Santa Inês	s20793	3	133380467	6,0338	4,9069
Santa Inês	OAR3_165549468_X	3	154735486	4,3094	-4,0601
Santa Inês	OAR3_165741097	3	154933986	5,1969	4,5142
Santa Inês	OAR3_165766186	3	154959719	4,1501	-3,9737
Santa Inês	OAR3_165801399	3	154994981	5,1196	4,4764
Santa Inês	OAR3_168114192	3	157096210	4,4806	-4,1512
Santa Inês	OAR3_168639353	3	157562244	4,8819	4,3580
Santa Inês	OAR3_168915114	3	157864029	4,4746	4,1481
Santa Inês	OAR3_198298678	3	184106820	4,0817	3,9360
Santa Inês	OAR6_40311379	6	36104954	4,4736	4,1476
Santa Inês	s22461	9	29885716	5,0150	-4,4246
Santa Inês	s33234	15	29495444	4,1240	-3,9593
Santa Inês	OAR16_38040585	16	34884982	5,6100	-4,7119
Santa Inês	OAR16_38221926	16	35037917	6,0404	4,9099
Santa Inês	OAR16_38414913	16	35232372	6,0219	-4,9015
Santa Inês	OAR16_38988953	16	35781385	4,3209	-4,0663
Santa Inês	s24799	16	36323883	4,3460	4,0798
Santa Inês	s40074	16	36472371	5,0499	4,4420

Raça	SNP	Cromossomo	Posição	piHS	iHS
Santa Inês	OAR16_39706312	16	36510923	4,3379	4,0754
Santa Inês	OAR16_39780130	16	36586383	5,8112	4,8054
Santa Inês	DU529574_332	16	36626318	4,6899	4,2602
Santa Inês	OAR16_39840363	16	36642330	7,0661	-5,3543
Santa Inês	OAR16_39888776	16	36697783	6,1353	4,9526
Santa Inês	OAR16_39987451	16	36797401	5,7642	-4,7837
Santa Inês	OAR16_40016364	16	36820258	5,4238	4,6237
Santa Inês	OAR16_40191916	16	36988703	4,7963	4,3146
Santa Inês	OAR16_40558553	16	37331737	5,1352	-4,4840
Santa Inês	OAR16_40846568_X	16	37622186	4,3436	4,0785
Santa Inês	OAR16_41617317	16	38365952	5,7997	4,8001
Santa Inês	OAR16_41681642	16	38418208	5,7997	4,8001
Santa Inês	OAR16_42159705	16	38851083	5,9437	-4,8661
Santa Inês	OAR16_42184944	16	38878086	5,4630	4,6424
Santa Inês	s42786	16	38880548	5,2156	-4,5234
Santa Inês	OAR16_42312325	16	38943467	5,6602	4,7354
Santa Inês	OAR16_45680234	16	42070192	4,8083	-4,3208
Santa Inês	OAR16_46450104	16	42763057	5,3303	4,5789
Santa Inês	OAR16_47088252	16	43222980	4,1694	3,9842
Santa Inês	OAR16_47146922	16	43273047	9,0485	6,1272
Santa Inês	OAR16_47204051	16	43340338	5,4140	4,6190
Santa Inês	OAR16_47351258	16	43480573	4,9007	-4,3675
Santa Inês	OAR16_47980776	16	44034414	4,1386	-3,9673
Santa Inês	s29275	16	44099474	5,1421	-4,4874
Santa Inês	OAR16_48105833	16	44145304	5,9603	-4,8736
Santa Inês	s17664	16	44214840	5,0551	4,4445
Santa Inês	OAR16_48362074	16	44414891	4,5203	4,1721
Santa Inês	OAR16_48444618	16	44501487	5,0057	-4,4200
Santa Inês	OAR16_48453032_X	16	44507527	4,1500	3,9736
Santa Inês	OAR16_49094655	16	45123144	5,5650	-4,6907
Santa Inês	s66605	16	45882078	4,9806	-4,4075
Santa Inês	OAR16_50214709	16	45992215	4,1916	3,9964
Santa Inês	OAR16_50476073	16	46190260	6,2901	5,0215
Santa Inês	OAR16_50955052	16	46685424	6,1322	-4,9512
Santa Inês	OAR16_51134755	16	46887875	5,0006	-4,4175
Santa Inês	OAR16_51292717	16	47038828	6,7951	-5,2403
Santa Inês	OAR16_51422323	16	47245495	8,3086	-5,8501
Santa Inês	OAR16_51697426	16	47524384	8,5558	-5,9440
Santa Inês	OAR16_53844475	16	49585602	9,1806	6,1754
Santa Inês	OAR16_53901725	16	49624463	4,0646	3,9265
Santa Inês	OAR16_53929465	16	49651677	8,8665	6,0601
Santa Inês	OAR16_54540482	16	50008827	4,0828	3,9366

Os SNPs significativos nas análises de iHS foram agrupados, formando 13 regiões para Crioula Lanada, sete regiões para Morada Nova e 23 regiões para Santa Inês (Tabela 6).

As principais regiões identificadas pelas análises de iHS estão localizadas nos cromossomos OAR3 e OAR16. No cromossomo OAR3 foi identificada uma região contendo um *cluster* de dois SNPs significativos (s42902 e OAR3_11033570) na raça Morada Nova e quatro regiões adjacentes contendo *clusters* de dois (s52741 e OAR3_138290871), três (OAR3_141611507, OAR3_142056197, s33601 e s20793), quatro (OAR3_165549468_X, OAR3_165741097, OAR3_165766186 e OAR3_165801399) e três SNPs (OAR3_168114192, OAR3_168639353 e OAR3_168915114) na raça Santa Inês.

Já no cromossomo OAR16, foram identificadas seis regiões adjacentes contendo *clusters* de SNPs significativos na raça Santa Inês. As regiões continham *clusters* de 12 (s24799, s40074, OAR16_39706312, OAR16_39780130, DU529574_332, OAR16_39840363, OAR16_39888776, OAR16_39987451, OAR16_40016364, OAR16_40191916, OAR16_40558553 e OAR16_40846568_X), seis (OAR16_41617317, OAR16_41681642, OAR16_42159705, OAR16_42184944, s42786 e OAR16_42312325), cinco (OAR16_46450104, OAR16_47088252, OAR16_47146922, OAR16_47204051 e OAR16_47351258), sete (OAR16_47980776, s29275, OAR16_48105833, s17664, OAR16_48362074, OAR16_48444618 e OAR16_48453032_X), oito (s66605, OAR16_50214709, OAR16_50476073, OAR16_50955052, OAR16_51134755, OAR16_51292717, OAR16_51422323 e OAR16_51697426) e quatro SNPs significativos (OAR16_53844475, OAR16_53901725, OAR16_53929465 e OAR16_54540482).

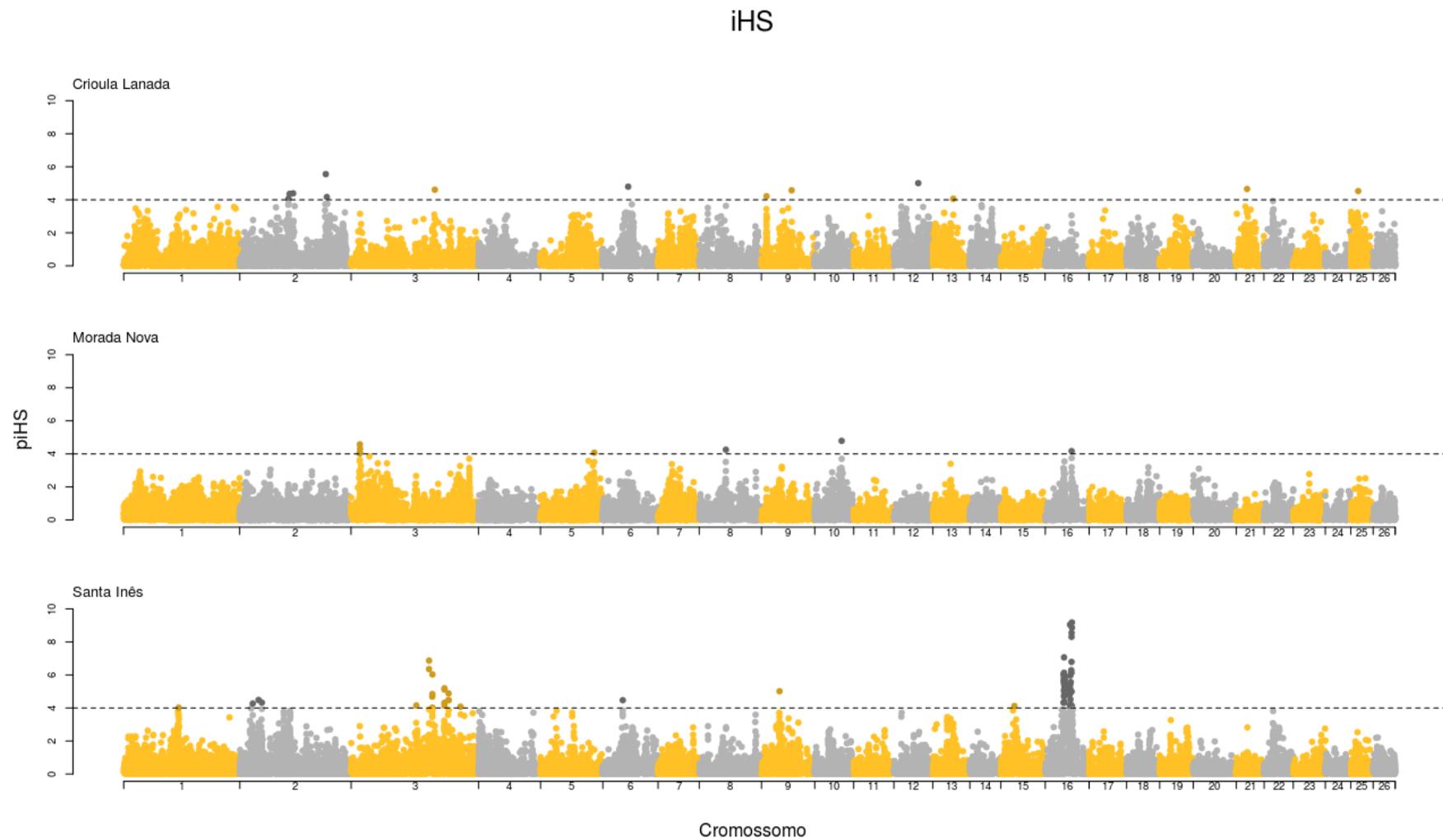


Figura 3. Manhattan plot para a análise de iHS. Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs significativos.

Tabela 6. Regiões genômicas candidatas a conter genes sob seleção identificadas através das análises de iHS.

Região	Raça	SNP	Cromossomo	Posição	Início	Final
1	Crioula Lanada	OAR2_119411966	2	111172962	110922962	111422962
2	Crioula Lanada	OAR2_126354465	2	118180362	117930362	118430362
3	Crioula Lanada	OAR2_137134681	2	128762227	128512227	129012227
4	Crioula Lanada	OAR2_200920481	2	189539157	189289157	189789157
5	Crioula Lanada	OAR2_205533418	2	193984737	193734737	194234737
6	Crioula Lanada	OAR3_151019080	3	141203065	140953065	141453065
7	Crioula Lanada	OAR6_57182975	6	51814853	51564853	52064853
8	Crioula Lanada	OAR9_10077769	9	10082948	9832948	10332948
9	Crioula Lanada	OAR9_67039654	9	63523344	63273344	63773344
10	Crioula Lanada	s69940	12	41343197	41093197	41593197
11	Crioula Lanada	OAR13_49971490	13	46647627	46397627	46897627
12	Crioula Lanada	s44390	21	18528589	18278589	18778589
13	Crioula Lanada	OAR25_23974308	25	22965788	22715788	23215788
14	Morada Nova	OAR3_10304371	3	9783340	9533340	10033340
15	Morada Nova	s42902; OAR3_11033570*	3	10515273	10128941	10765273
16	Morada Nova	OAR3_12289153	3	11712721	11462721	11962721
17	Morada Nova	OAR5_106230044	5	97597361	97347361	97847361
18	Morada Nova	OAR8_44102022	8	41015484	40765484	41265484
19	Morada Nova	OAR10_66804684	10	64726524	64476524	64976524
20	Morada Nova	OAR16_51422323	16	47245495	46995495	47495495
21	Santa Inês	OAR1_146213099	1	135459627	135209627	135709627
22	Santa Inês	s36668	2	37506606	37256606	37756606

Região	Raça	SNP	Cromossomo	Posição	Início	Final
23	Santa Inês	s35831	2	50934632	50684632	51184632
24	Santa Inês	OAR2_63699523	2	59231945	58981945	59481945
25	Santa Inês	DU184544_326	3	108058182	107808182	108308182
26	Santa Inês	s52741; OAR3_138290871*	3	129648614	129201837	129898614
27	Santa Inês	OAR3_141611507; OAR3_142056197; s33601; s20793*	3	133380467	132255054	133630467
28	Santa Inês	OAR3_165549468_X; OAR3_165741097*; OAR3_165766186; OAR3_165801399	3	154933986	154485486	155244981
29	Santa Inês	OAR3_168114192; OAR3_168639353*; OAR3_168915114	3	157562244	156846210	158114029
30	Santa Inês	OAR3_198298678	3	184106820	183856820	184356820
31	Santa Inês	OAR6_40311379	6	36104954	35854954	36354954
32	Santa Inês	s22461	9	29885716	29635716	30135716
33	Santa Inês	s33234	15	29495444	29245444	29745444
34	Santa Inês	OAR16_38040585; OAR16_38221926*; OAR16_38414913	16	35037917	34634982	35482372
35	Santa Inês	OAR16_38988953	16	35781385	35531385	36031385
36	Santa Inês	s24799; s40074; OAR16_39706312; OAR16_39780130; DU529574_332; OAR16_39840363*; OAR16_39888776; OAR16_39987451; OAR16_40016364; OAR16_40191916; OAR16_40558553; OAR16_40846568_X	16	36642330	36073883	37872186
37	Santa Inês	OAR16_41617317; OAR16_41681642; OAR16_42159705*; OAR16_42184944; s42786; OAR16_42312325	16	38851083	38115952	39193467

Região	Raça	SNP	Cromossomo	Posição	Início	Final
38	Santa Inês	OAR16_45680234	16	42070192	41820192	42320192
39	Santa Inês	OAR16_46450104; OAR16_47088252; OAR16_47146922*; OAR16_47204051; OAR16_47351258	16	43273047	42513057	43730573
40	Santa Inês	OAR16_47980776; s29275; OAR16_48105833*; s17664; OAR16_48362074; OAR16_48444618; OAR16_48453032_X	16	44145304	43784414	44757527
41	Santa Inês	OAR16_49094655	16	45123144	44873144	45373144
42	Santa Inês	s66605; OAR16_50214709; OAR16_50476073; OAR16_50955052; OAR16_51134755; OAR16_51292717; OAR16_51422323; OAR16_51697426*	16	47524384	45632078	47774384
43	Santa Inês	OAR16_53844475*; OAR16_53901725; OAR16_53929465; OAR16_54540482	16	49585602	49335602	50258827

*SNP com maior valor de iHS dentro da região.

A análise de Rsb revelou a presença de 21 SNPs significativos ($p_{Rsb} \geq 4$) para a comparação Crioula Lanada x Morada Nova, distribuídos pelos cromossomos OAR1, OAR2, OAR3, OAR8, OAR10, OAR13, OAR16 E OAR22. Para a comparação Crioula Lanada x Santa Inês foram identificados 204 SNPs significativos, distribuídos pelos cromossomos OAR2, OAR3, OAR9, OAR12, OAR13 e OAR16. Já na comparação Morada Nova x Santa Inês, foram identificados 26 SNPs significativos, distribuídos pelos cromossomos OAR1, OAR3, OAR8, OAR9, OAR10 e OAR16. As informações referentes aos SNPs significativos na análise de Rsb estão representados na Tabela 7.

Tabela 7. SNPs significativos para as análises de Rsb.

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BCS x BMN	OAR1_127962239	1	118639747	-4,0712	4,3300
BCS x BMN	s72107	2	18584559	-4,8918	6,0005
BCS x BMN	OAR2_19756148	2	19436541	-4,1827	4,5404
BCS x BMN	OAR3_215849996	3	200608549	-3,9084	4,0320
BCS x BMN	OAR8_43900420	8	40813422	-3,9641	4,1326
BCS x BMN	OAR8_44102022	8	41015484	-3,9550	4,1162
BCS x BMN	OAR10_66804684	10	64726524	-3,9443	4,0967
BCS x BMN	s60887	13	4314690	4,0076	4,2123
BCS x BMN	s62936	13	23148764	4,2440	4,6584
BCS x BMN	OAR13_25941147	13	23407041	3,9955	4,1900
BCS x BMN	OAR16_39062204	16	35852290	-4,1847	4,5443
BCS x BMN	OAR16_39780130	16	36586383	-3,9407	4,0903
BCS x BMN	OAR16_40558553	16	37331737	-4,3968	4,9592
BCS x BMN	s32951	16	38301584	-4,3078	4,7828
BCS x BMN	OAR13_34211514	16	38342057	-4,2104	4,5935
BCS x BMN	OAR16_41617317	16	38365952	-3,9356	4,0810
BCS x BMN	OAR16_41681642	16	38418208	-5,5456	7,5332
BCS x BMN	OAR16_45550332	16	41943519	-4,0885	4,3623
BCS x BMN	OAR16_47146922	16	43273047	-3,9679	4,1397
BCS x BMN	OAR16_51134755	16	46887875	-4,1185	4,4186
BCS x BMN	OAR22_23392099	22	19477317	4,5738	5,3197
BCS x BSI	OAR2_245597717	2	232510155	3,9525	4,1115
BCS x BSI	OAR2_245680195	2	232591652	4,0198	4,2348
BCS x BSI	OAR2_246045571	2	232957928	4,6077	5,3903
BCS x BSI	OAR2_251955154	2	238567116	3,9269	4,0653
BCS x BSI	s23721	2	238643447	3,9020	4,0204
BCS x BSI	s67306	2	239564082	3,9883	4,1769
BCS x BSI	OAR3_114955879	3	107958569	-4,1960	4,5659
BCS x BSI	OAR3_141611507	3	132505054	-3,9621	4,1290
BCS x BSI	OAR3_165741097	3	154933986	-4,1387	4,4568

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BCS x BSI	OAR9_1007769	9	10082948	-4,2580	4,6856
BCS x BSI	OAR12_17169201	12	14458685	4,2249	4,6215
BCS x BSI	OAR12_17296831	12	14612434	3,9599	4,1250
BCS x BSI	OAR13_1628753	13	1147107	3,9843	4,1696
BCS x BSI	OAR13_1832147	13	1366306	4,0006	4,1994
BCS x BSI	OAR13_9356390	13	8385113	4,0956	4,3757
BCS x BSI	OAR13_25099166	13	22555622	3,9360	4,0817
BCS x BSI	s62936	13	23148764	4,0850	4,3557
BCS x BSI	OAR16_36156919	16	33249488	-4,0567	4,3030
BCS x BSI	OAR16_36192133	16	33284858	-4,0306	4,2546
BCS x BSI	s45149	16	33447141	-4,5499	5,2702
BCS x BSI	OAR16_36378409	16	33470561	-4,0686	4,3252
BCS x BSI	OAR16_36410113	16	33504980	-4,4327	5,0312
BCS x BSI	OAR16_36452841	16	33545078	-4,2269	4,6255
BCS x BSI	OAR16_36498470	16	33584033	-3,9034	4,0230
BCS x BSI	OAR16_36508382	16	33596117	-3,9716	4,1464
BCS x BSI	OAR16_36743952	16	33798516	-4,6570	5,4936
BCS x BSI	OAR16_36821541	16	33872546	-4,2576	4,6848
BCS x BSI	OAR16_36833073	16	33885737	-4,8324	5,8700
BCS x BSI	OAR16_36890689	16	33941127	-5,0581	6,3732
BCS x BSI	OAR16_36987741	16	34034559	-4,1410	4,4612
BCS x BSI	s50295	16	34092264	-4,1570	4,4916
BCS x BSI	OAR16_37226202	16	34279856	-4,5451	5,2603
BCS x BSI	OAR16_37614079	16	34625175	-4,3401	4,8464
BCS x BSI	OAR16_37779529	16	34679175	-4,5380	5,2457
BCS x BSI	OAR16_38040585	16	34884982	-5,2946	6,9234
BCS x BSI	OAR16_38083665	16	34927121	-4,2631	4,6954
BCS x BSI	OAR16_38221926	16	35037917	-5,6397	7,7686
BCS x BSI	s40592	16	35098287	-4,2254	4,6225
BCS x BSI	OAR16_38355680	16	35171594	-4,2219	4,6157
BCS x BSI	OAR16_38414913	16	35232372	-5,5933	7,6521
BCS x BSI	s03406	16	35684132	-5,0039	6,2504
BCS x BSI	OAR16_38988953	16	35781385	-5,4063	7,1916
BCS x BSI	s74559	16	35801129	-4,7595	5,7120
BCS x BSI	OAR16_39062204	16	35852290	-6,5120	10,1298
BCS x BSI	s25960	16	35866332	-5,0266	6,3016
BCS x BSI	OAR16_39143709	16	35930429	-5,0146	6,2746
BCS x BSI	s49984	16	36028239	-5,2110	6,7263
BCS x BSI	s35691	16	36235813	-4,9697	6,1736
BCS x BSI	s24799	16	36323883	-5,7161	7,9626
BCS x BSI	s40074	16	36472371	-5,6739	7,8551
BCS x BSI	OAR16_39706312	16	36510923	-5,8915	8,4172
BCS x BSI	OAR16_39780130	16	36586383	-5,9477	8,5654

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BCS x BSI	DU529574_332	16	36626318	-5,7111	7,9497
BCS x BSI	OAR16_39840363	16	36642330	-6,4494	9,9498
BCS x BSI	OAR16_39888776	16	36697783	-6,3517	9,6716
BCS x BSI	OAR16_39987451	16	36797401	-6,1053	8,9888
BCS x BSI	OAR16_40016364	16	36820258	-5,4522	7,3032
BCS x BSI	OAR16_40191916	16	36988703	-5,3055	6,9493
BCS x BSI	s48729	16	37034665	-5,2613	6,8446
BCS x BSI	OAR16_40269571	16	37062919	-5,1906	6,6787
BCS x BSI	OAR16_40313142	16	37106181	-5,1983	6,6966
BCS x BSI	OAR16_40345735	16	37138240	-5,1983	6,6966
BCS x BSI	OAR16_40426113	16	37217962	-5,2617	6,8455
BCS x BSI	OAR16_40503533	16	37284426	-5,5381	7,5146
BCS x BSI	OAR16_40558553	16	37331737	-5,8570	8,3268
BCS x BSI	OAR16_40589074	16	37359778	-5,0249	6,2978
BCS x BSI	s34284	16	37464470	-5,4364	7,2648
BCS x BSI	OAR16_40715706_X	16	37488524	-5,6635	7,8288
BCS x BSI	OAR16_40749182	16	37521227	-5,5954	7,6573
BCS x BSI	s11229	16	37571144	-5,8088	8,2012
BCS x BSI	OAR16_40846568_X	16	37622186	-5,1906	6,6786
BCS x BSI	OAR16_41320024	16	38088215	-4,7322	5,6535
BCS x BSI	s32951	16	38301584	-5,4769	7,3636
BCS x BSI	OAR13_34211514	16	38342057	-6,7829	10,9290
BCS x BSI	OAR16_41617317	16	38365952	-5,9684	8,6206
BCS x BSI	OAR16_41681642	16	38418208	-7,6978	13,8577
BCS x BSI	OAR16_41711634	16	38448143	-6,0083	8,7271
BCS x BSI	OAR16_41737441	16	38473232	-5,2958	6,9263
BCS x BSI	OAR16_41779659	16	38512493	-4,3479	4,8618
BCS x BSI	OAR16_41804913	16	38538612	-4,3479	4,8618
BCS x BSI	s67705	16	38606820	-4,4149	4,9954
BCS x BSI	OAR16_42159705	16	38851083	-4,9666	6,1667
BCS x BSI	OAR16_42172047	16	38862399	-5,3465	7,0474
BCS x BSI	OAR16_42184944	16	38878086	-4,7289	5,6463
BCS x BSI	s42786	16	38880548	-4,7289	5,6463
BCS x BSI	s33178	16	38899653	-5,4239	7,2343
BCS x BSI	OAR16_42312325	16	38943467	-4,3895	4,9446
BCS x BSI	s06837	16	40024589	-5,1426	6,5672
BCS x BSI	OAR16_43573089	16	40080673	-4,9302	6,0854
BCS x BSI	OAR16_43715998	16	40219105	-4,0587	4,3068
BCS x BSI	OAR16_43747463	16	40248940	-4,9832	6,2037
BCS x BSI	OAR16_43758108	16	40258646	-4,4142	4,9941
BCS x BSI	OAR16_43833978	16	40339621	-6,1758	9,1817
BCS x BSI	OAR16_43894833	16	40390788	-5,9382	8,5404
BCS x BSI	OAR16_43916302	16	40405148	-5,5993	7,6670

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BCS x BSI	OAR16_44015351	16	40503694	-5,4019	7,1810
BCS x BSI	s20727	16	40527757	-5,1630	6,6144
BCS x BSI	OAR16_44102697	16	40583350	-5,8928	8,4205
BCS x BSI	OAR16_44166560	16	40650762	-4,9699	6,1739
BCS x BSI	OAR16_44325630	16	40797822	-4,4879	5,1432
BCS x BSI	s23014	16	40837120	-4,7463	5,6838
BCS x BSI	s43481	16	40936379	-5,8988	8,4364
BCS x BSI	s37118	16	41036884	-5,9579	8,5926
BCS x BSI	OAR16_44680056	16	41134921	-5,2137	6,7326
BCS x BSI	OAR16_44759148	16	41206143	-6,6822	10,6283
BCS x BSI	OAR16_44884811	16	41295748	-4,9600	6,1518
BCS x BSI	250506CS3900211600001_1041	16	41355381	-5,1405	6,5623
BCS x BSI	s53151	16	41463528	-5,6867	7,8877
BCS x BSI	s45551	16	41497759	-5,1640	6,6168
BCS x BSI	s22329	16	41557371	-4,9174	6,0569
BCS x BSI	s29682	16	41680378	-6,8815	11,2275
BCS x BSI	s59948	16	41755334	-3,9691	4,1418
BCS x BSI	OAR16_45398511	16	41803341	-6,6075	10,4079
BCS x BSI	s17055	16	41853736	-5,3066	6,9519
BCS x BSI	s27395	16	41904978	-4,4354	5,0367
BCS x BSI	OAR16_45550332	16	41943519	-6,7277	10,7634
BCS x BSI	OAR16_45590568	16	41983615	-4,7029	5,5909
BCS x BSI	OAR16_45631306	16	42020885	-4,1317	4,4436
BCS x BSI	OAR16_45680234	16	42070192	-5,3311	7,0105
BCS x BSI	OAR16_45863887	16	42238684	-4,0975	4,3792
BCS x BSI	OAR16_45922214	16	42296648	-4,2833	4,7348
BCS x BSI	OAR16_45960993	16	42336937	-4,2833	4,7348
BCS x BSI	OAR16_46027470	16	42361787	-4,0467	4,2844
BCS x BSI	OAR16_46251400	16	42600749	-4,1775	4,5306
BCS x BSI	OAR16_46290531	16	42639280	-5,2403	6,7951
BCS x BSI	OAR16_46325523	16	42670945	-5,1133	6,4995
BCS x BSI	OAR16_46450104	16	42763057	-4,9455	6,1195
BCS x BSI	OAR16_46544413	16	42811161	-5,0598	6,3770
BCS x BSI	OAR16_46761904	16	42853485	-5,6548	7,8069
BCS x BSI	OAR16_46789766	16	42884453	-4,4411	5,0482
BCS x BSI	OAR16_46955783	16	43093912	-4,6354	5,4483
BCS x BSI	OAR16_46984845	16	43124647	-4,8131	5,8280
BCS x BSI	s60183	16	43166209	-5,2152	6,7361
BCS x BSI	OAR16_47088252	16	43222980	-5,3912	7,1550
BCS x BSI	OAR16_47146922	16	43273047	-6,3212	9,5858
BCS x BSI	OAR16_47204051	16	43340338	-5,4547	7,3093
BCS x BSI	OAR16_47351258	16	43480573	-5,3507	7,0574
BCS x BSI	OAR16_47697777	16	43745925	-5,1695	6,6295

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BCS x BSI	OAR16_47932818	16	43992109	-4,9862	6,2105
BCS x BSI	OAR16_47980776	16	44034414	-4,4891	5,1455
BCS x BSI	s29275	16	44099474	-5,7690	8,0984
BCS x BSI	OAR16_48105833	16	44145304	-4,9556	6,1419
BCS x BSI	OAR16_48150525	16	44187150	-5,1284	6,5343
BCS x BSI	s17664	16	44214840	-4,9232	6,0698
BCS x BSI	OAR16_48258378	16	44300442	-5,4140	7,2103
BCS x BSI	OAR16_48313493	16	44354865	-4,8378	5,8818
BCS x BSI	OAR16_48362074	16	44414891	-5,3058	6,9502
BCS x BSI	OAR16_48400221_X	16	44458369	-4,7803	5,7569
BCS x BSI	OAR16_48444618	16	44501487	-4,9518	6,1335
BCS x BSI	OAR16_48453032_X	16	44507527	-4,8225	5,8484
BCS x BSI	OAR16_48510905	16	44557185	-4,4376	5,0410
BCS x BSI	OAR16_48529126	16	44572270	-4,7093	5,6046
BCS x BSI	OAR16_48633427	16	44679745	-4,1162	4,4144
BCS x BSI	OAR16_48703340	16	44746250	-4,2344	4,6398
BCS x BSI	OAR16_48760523	16	44784202	-4,6427	5,4635
BCS x BSI	OAR16_48866451	16	44908520	-4,7344	5,6581
BCS x BSI	OAR16_49003509	16	45044422	-4,6717	5,5247
BCS x BSI	OAR16_49094655	16	45123144	-5,0069	6,2571
BCS x BSI	OAR16_49670092	16	45449574	-4,0569	4,3033
BCS x BSI	OAR16_49886669	16	45663897	-4,7929	5,7842
BCS x BSI	OAR16_49926380	16	45698202	-5,1107	6,4936
BCS x BSI	OAR16_49974943	16	45758412	-4,7420	5,6744
BCS x BSI	s05750	16	45811094	-4,5486	5,2676
BCS x BSI	s66605	16	45882078	-5,4080	7,1958
BCS x BSI	OAR16_50103993	16	45894112	-4,7082	5,6022
BCS x BSI	OAR16_50214709	16	45992215	-4,4922	5,1518
BCS x BSI	OAR16_50235560	16	46011155	-3,9315	4,0735
BCS x BSI	s60314	16	46057863	-4,7314	5,6516
BCS x BSI	OAR16_50476073	16	46190260	-5,2408	6,7961
BCS x BSI	OAR16_50546916	16	46258699	-4,5128	5,1940
BCS x BSI	OAR16_50565808	16	46287228	-4,3538	4,8736
BCS x BSI	OAR16_50643262_X	16	46363196	-4,1781	4,5317
BCS x BSI	OAR16_50704170	16	46422567	-4,4115	4,9886
BCS x BSI	s58708	16	46452686	-4,6333	5,4439
BCS x BSI	OAR16_50765168	16	46489946	-5,5159	7,4598
BCS x BSI	OAR16_50897399	16	46625066	-5,0189	6,2843
BCS x BSI	OAR16_50955052	16	46685424	-6,3532	9,6760
BCS x BSI	OAR16_50979998	16	46708448	-5,0449	6,3431
BCS x BSI	OAR16_51020636	16	46753849	-5,3983	7,1722
BCS x BSI	OAR16_51057548	16	46790392	-5,5342	7,5050
BCS x BSI	OAR16_51134755	16	46887875	-5,9776	8,6450

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BCS x BSI	OAR16_51194558	16	46955536	-4,6607	5,5014
BCS x BSI	OAR16_51292717	16	47038828	-6,1788	9,1898
BCS x BSI	OAR16_51365222	16	47189578	-4,4324	5,0307
BCS x BSI	OAR16_51422323	16	47245495	-5,9851	8,6650
BCS x BSI	OAR16_51483488	16	47311748	-4,5222	5,2132
BCS x BSI	OAR16_51559181	16	47383354	-5,7363	8,0142
BCS x BSI	OAR16_51697426	16	47524384	-5,4965	7,4118
BCS x BSI	OAR16_51795310	16	47620181	-4,5348	5,2391
BCS x BSI	OAR16_51977009	16	47764553	-4,2191	4,6103
BCS x BSI	OAR16_51992311	16	47785437	-4,0993	4,3825
BCS x BSI	OAR16_52082065	16	47864819	-5,0663	6,3920
BCS x BSI	OAR16_52323728	16	48069203	-4,8234	5,8504
BCS x BSI	OAR16_52348833	16	48099002	-4,4368	5,0395
BCS x BSI	OAR16_52790463	16	48581196	-4,4507	5,0675
BCS x BSI	OAR16_53203108	16	49007700	-3,9854	4,1716
BCS x BSI	OAR16_53803957	16	49554749	-4,0555	4,3007
BCS x BSI	OAR16_53844475	16	49585602	-5,0852	6,4351
BCS x BSI	OAR16_53901725	16	49624463	-4,2420	4,6545
BCS x BSI	OAR16_53929465	16	49651677	-4,2894	4,7468
BCS x BSI	OAR16_53970846_X	16	49684342	-4,0389	4,2700
BCS x BSI	OAR16_54049558	16	49782274	-4,1230	4,4271
BCS x BSI	OAR16_54446295	16	49947010	-4,2770	4,7226
BCS x BSI	OAR16_54540482	16	50008827	-4,0940	4,3726
BMN x BSI	s16767	1	39829283	3,9480	4,1034
BMN x BSI	OAR3_55935023	3	52899326	4,3800	4,9256
BMN x BSI	OAR3_56021384	3	52976986	4,2319	4,6350
BMN x BSI	s22657	3	132678490	-3,9432	4,0947
BMN x BSI	s68180	3	132689753	-4,7683	5,7309
BMN x BSI	OAR3_142190582	3	132984366	-3,9526	4,1118
BMN x BSI	s62934	3	133025037	-3,8921	4,0027
BMN x BSI	s20793	3	133380467	-5,7316	8,0022
BMN x BSI	s21200	3	133517016	-4,2618	4,6930
BMN x BSI	OAR8_43915897	8	40828066	4,3528	4,8716
BMN x BSI	OAR8_44102022	8	41015484	4,4851	5,1374
BMN x BSI	OAR8_44275328_X	8	41182198	4,0056	4,2086
BMN x BSI	OAR8_44479113	8	41314334	4,3448	4,8558
BMN x BSI	OAR8_44532355	8	41350703	4,1295	4,4395
BMN x BSI	OAR8_44575878_X	8	41398540	4,2173	4,6068
BMN x BSI	OAR8_44684641	8	41478607	4,1118	4,4060
BMN x BSI	OAR9_38670488	9	36746818	-4,2062	4,5856
BMN x BSI	OAR10_60531037	10	59095153	4,1048	4,3929
BMN x BSI	OAR10_60624505	10	59172145	3,9345	4,0791
BMN x BSI	s54545	10	64182232	4,2267	4,6251

Comparação	SNP	Cromossomo	Posição	Rsb	pRsb
BMN x BSI	OAR10_66415688	10	64333687	4,0607	4,3104
BMN x BSI	OAR10_66533600	10	64461296	4,0384	4,2691
BMN x BSI	OAR10_66804684	10	64726524	4,8622	5,9352
BMN x BSI	OAR10_67301196	10	65063529	4,2377	4,6462
BMN x BSI	s29682	16	41680378	-3,9329	4,0761
BMN x BSI	OAR16_54540482	16	50008827	-4,0810	4,3483

É interessante notar a presença de *clusters* de SNPs significativos nos cromossomos OAR2, OAR3, OAR8, OAR10, OAR13 e OAR16 (Figura 4), o que sugere que estas regiões podem ser fortes candidatas a conter genes sob seleção.

Os SNPs significativos nas análises de Rsb foram agrupados, formando 16 regiões para Crioula Lanada x Morada Nova, 15 para Crioula Lanada x Santa Inês e nove para Morada Nova x Santa Inês (Tabela 8).

Para a comparação Crioula Lanada x Morada Nova, as principais regiões identificadas na análise de Rsb foram: um *cluster* contendo dois SNPs, localizado no cromossomo OAR8 (OAR8_43900420 e OAR8_44102022), um *cluster* contendo dois SNPs, localizado no cromossomo OAR13 (s62936, OAR13_25941147) e um *cluster* contendo quatro SNPs, localizado no cromossomo OAR16 (s32951, OAR13_34211514, OAR16_41617317 e OAR16_41681642).

Para a comparação Crioula Lanada x Santa Inês, as principais regiões identificadas foram: dois *clusters* de SNPs significativos localizados no cromossomo OAR2, um contendo três (OAR2_245597717, OAR2_245680195 e OAR2_246045571) e o outro contendo dois SNPs (OAR2_251955154 e s23721), um *cluster* contendo dois SNPs significativos, localizado no cromossomo OAR12 (OAR12_17169201 e OAR12_17296831), um *cluster* contendo dois SNPs, localizado no cromossomo OAR13 (OAR13_1628753 e OAR13_1832147). Além disso, foram identificados três *clusters* adjacentes no cromossomo OAR16, um contendo 70 SNPs, um contendo 109 SNPs e, por fim, um terceiro contendo oito SNPs.

Para a comparação Morada Nova x Santa Inês, as principais regiões identificadas foram: um *cluster* contendo dois SNPs, localizado no cromossomo OAR3 (OAR3_55935023 e OAR3_56021384), um *cluster* contendo seis SNPs, também localizado no cromossomo OAR3 (s22657, s68180, OAR3_142190582, s62934, s20793 e s21200), um *cluster* contendo oito SNPs, localizado no cromossomo OAR8

(OAR8_43915897, OAR8_44102022, OAR8_44275328_X, OAR8_44479113, OAR8_44532355, OAR8_44575878_X e OAR8_44684641), um *cluster* contendo dois SNPs, localizado no cromossomo OAR10 (OAR10_60531037 e OAR10_60624505) e um *cluster* contendo cinco SNPs, também localizado no cromossomo OAR10 (s54545, OAR10_66415688, OAR10_66533600, OAR10_66804684 e OAR10_67301196).

Considerando conjuntamente todas as análises realizadas, foram identificados sinais de seleção distribuídos ao longo de 21 dos 26 cromossomos autossômicos do genoma ovino. Apenas não foram encontrados sinais de seleção nos cromossomos OAR17, OAR18, OAR20, OAR23 e OAR24.

As regiões identificadas por cada uma das três análises individualmente (F_{ST} , iHS e Rsb) foram alinhadas em busca de regiões coincidentes. Foram identificados sinais coincidentes por mais de uma metodologia nos cromossomos OAR3, OAR8, OAR9, OAR10, OAR13 e OAR16, resultando em 92 regiões genômicas identificadas como candidatas a estar sob seleção. Destas, quatro foram identificadas pelas três metodologias, 11 por duas e 77 por apenas uma das metodologias (Tabela 9).

A anotação destas regiões revelou a presença de 78 genes candidatos, e destes, 40 continham o SNP significativo localizado dentro do gene (Tabela 9). A análise funcional permitiu a identificação de genes relacionados com diversas funções biológicas, dentre elas: resposta imune (IL17C, IL18R1, IL7, IL1RL1, SMAD6, GPR65, SKAP2, THY1), desenvolvimento do sistema nervoso (SLITRK1, FUT9, ST8SIA4, CNTNAP2, CNTN3, THY1), diferenciação neuronal (SLITRK1, CNTNAP2, THY1), reprodução (SPAG6, RXFP2, MSTN) e percepção sensorial (TRPM8, OR13F1, COL1A1, TTC8).

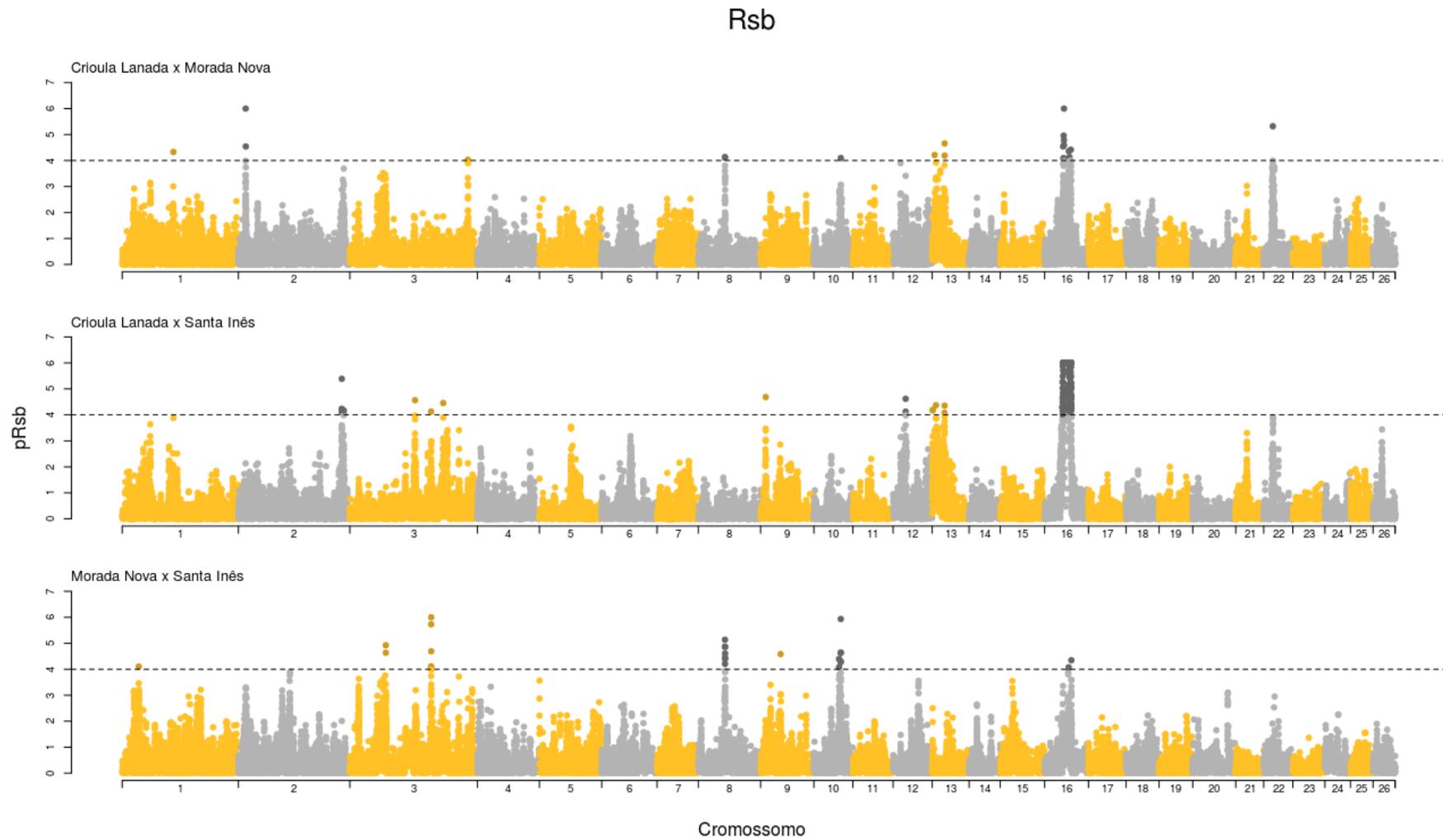


Figura 4. Manhattan plot para a análise de Rsb. Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs significativos.

Tabela 8. Regiões genômicas candidatas a conter genes sob seleção identificadas através das análises de Rsb.

Região	Comparação	SNP	Cromossomo	Posição	Início	Final
1	BCS x BMN	OAR1_127962239	1	118639747	118389747	118889747
2	BCS x BMN	s72107	2	18584559	18334559	18834559
3	BCS x BMN	OAR2_19756148	2	19436541	19186541	19686541
4	BCS x BMN	OAR3_215849996	3	200608549	200358549	200858549
5	BCS x BMN	OAR8_43900420*; OAR8_44102022	8	40813422	40563422	41265484
6	BCS x BMN	OAR10_66804684	10	64726524	64476524	64976524
7	BCS x BMN	s60887	13	4314690	4064690	4564690
8	BCS x BMN	s62936*; OAR13_25941147	13	23148764	22898764	23657041
9	BCS x BMN	OAR16_39062204	16	35852290	35602290	36102290
10	BCS x BMN	OAR16_39780130	16	36586383	36336383	36836383
11	BCS x BMN	OAR16_40558553	16	37331737	37081737	37581737
12	BCS x BMN	s32951; OAR13_34211514; OAR16_41617317; OAR16_41681642*	16	38418208	38051584	38668208
13	BCS x BMN	OAR16_45550332	16	41943519	41693519	42193519
14	BCS x BMN	OAR16_47146922	16	43273047	43023047	43523047
15	BCS x BMN	OAR16_51134755	16	46887875	46637875	47137875
16	BCS x BMN	OAR22_23392099	22	19477317	19227317	19727317
17	BCS x BSI	OAR2_245597717; OAR2_245680195; OAR2_246045571*	2	232957928	232260155	233207928
18	BCS x BSI	OAR2_251955154*; s23721	2	238567116	238317116	238893447
19	BCS x BSI	s67306	2	239564082	239314082	239814082
20	BCS x BSI	OAR3_114955879	3	107958569	107708569	108208569
21	BCS x BSI	OAR3_141611507	3	132505054	132255054	132755054
22	BCS x BSI	OAR3_165741097	3	154933986	154683986	155183986

23	BCS x BSI	OAR9_10077769	9	10082948	9832948	10332948
24	BCS x BSI	OAR12_17169201*; OAR12_17296831	12	14458685	14208685	14862434
25	BCS x BSI	OAR13_1628753; OAR13_1832147*	13	1366306	897107	1616306
26	BCS x BSI	OAR13_9356390	13	8385113	8135113	8635113
27	BCS x BSI	OAR13_25099166	13	22555622	22305622	22805622
28	BCS x BSI	s62936	13	23148764	22898764	23398764
29	BCS x BSI	OAR16_36156919; OAR16_36192133; s45149; OAR16_36378409; OAR16_36410113; OAR16_36452841; OAR16_36498470; OAR16_36508382; OAR16_36743952; OAR16_36821541; OAR16_36833073; OAR16_36890689; OAR16_36987741; s50295; OAR16_37226202; OAR16_37614079; OAR16_37779529; OAR16_38040585; OAR16_38083665; OAR16_38221926; s40592; OAR16_38355680; OAR16_38414913; s03406; OAR16_38988953; s74559; OAR16_39062204; s25960; OAR16_39143709; s49984; s35691; s24799; s40074; OAR16_39706312; OAR16_39780130; DU529574_332; OAR16_39840363; OAR16_39888776; OAR16_39987451; OAR16_40016364; OAR16_40191916; s48729; OAR16_40269571; OAR16_40313142; OAR16_40345735; OAR16_40426113; OAR16_40503533; OAR16_40558553; OAR16_40589074; s34284; OAR16_40715706_X; OAR16_40749182; s11229; OAR16_40846568_X; OAR16_41320024; s32951; OAR13_34211514; OAR16_41617317; OAR16_41681642*; OAR16_41711634; OAR16_41737441; OAR16_41779659; OAR16_41804913; s67705; OAR16_42159705; OAR16_42172047; OAR16_42184944; s42786; s33178; OAR16_42312325 s06837; OAR16_43573089; OAR16_43715998; OAR16_43747463; OAR16_43758108; OAR16_43833978; OAR16_43894833; OAR16_43916302; OAR16_44015351; s20727; OAR16_44102697; OAR16_44166560; OAR16_44325630; s23014; s43481; s37118; OAR16_44680056; OAR16_44759148; OAR16_44884811; 250506CS3900211600001_1041; s53151; s45551; s22329; s29682*; s59948; OAR16_45398511; s17055; s27395; OAR16_45550332; OAR16_45590568; OAR16_45631306; OAR16_45680234; OAR16_45863887; OAR16_45922214; OAR16_45960993; OAR16_46027470; OAR16_46251400; OAR16_46290531; OAR16_46325523; OAR16_46450104; OAR16_46544413; OAR16_46761904; OAR16_46789766; OAR16_46955783; OAR16_46984845; s60183; OAR16_47088252; OAR16_47146922;	16	38418208	32999488	39193467
30	BCS x BSI		16	41680378	39774589	49257700

		OAR16_47204051; OAR16_47351258; OAR16_47697777; OAR16_47932818; OAR16_47980776; s29275; OAR16_48105833; OAR16_48150525; s17664; OAR16_48258378; OAR16_48313493; OAR16_48362074; OAR16_48400221_X; OAR16_48444618; OAR16_48453032_X; OAR16_48510905; OAR16_48529126; OAR16_48633427; OAR16_48703340; OAR16_48760523; OAR16_48866451; OAR16_49003509; OAR16_49094655; OAR16_49670092; OAR16_49886669; OAR16_49926380; OAR16_49974943; s05750; s66605; OAR16_50103993; OAR16_50214709; OAR16_50235560; s60314; OAR16_50476073; OAR16_50546916; OAR16_50565808; OAR16_50643262_X; OAR16_50704170; s58708; OAR16_50765168; OAR16_50897399; OAR16_50955052; OAR16_50979998; OAR16_51020636; OAR16_51057548; OAR16_51134755; OAR16_51194558; OAR16_51292717; OAR16_51365222; OAR16_51422323; OAR16_51483488; OAR16_51559181; OAR16_51697426; OAR16_51795310; OAR16_51977009; OAR16_51992311; OAR16_52082065; OAR16_52323728; OAR16_52348833; OAR16_52790463; OAR16_53203108				
31	BCS x BSI	OAR16_53803957; OAR16_53844475*; OAR16_53901725; OAR16_53929465; OAR16_53970846_X; OAR16_54049558; OAR16_54446295; OAR16_54540482	16	49585602	49304749	50258827
32	BMN x BSI	s16767	1	39829283	39579283	40079283
33	BMN x BSI	OAR3_55935023*; OAR3_56021384	3	52899326	52649326	53226986
34	BMN x BSI	s22657; s68180; OAR3_142190582; s62934; s20793*; s21200	3	133380467	132428490	133767016
35	BMN x BSI	OAR8_43915897; OAR8_44102022*; OAR8_44275328_X; OAR8_44479113; OAR8_44532355; OAR8_44575878_X; OAR8_44684641	8	41015484	40578066	41728607
36	BMN x BSI	OAR9_38670488	9	36746818	36496818	36996818
37	BMN x BSI	OAR10_60531037*; OAR10_60624505	10	59095153	58845153	59422145
38	BMN x BSI	s54545; OAR10_66415688; OAR10_66533600; OAR10_66804684*; OAR10_67301196	10	64726524	63932232	65313529
39	BMN x BSI	s29682	16	41680378	41430378	41930378
40	BMN x BSI	OAR16_54540482	16	50008827	49758827	50258827

*SNP com maior valor de Rsb dentro da região.

Tabela 9. Anotação das regiões genômicas candidatas a estar sob seleção.

Região	Metodologia	Chr	Pico	Início	Final	RefSeq	Sigla do gene	Nome do gene
1	FST	1	6832517*	6582517	7082517	NM_001206066	TRPM8	transient receptor potential cation channel, subfamily M, member 8
2	Rsb (BMN x BSI)	1	39829283	39579283	40079283	XM_002686328 XM_616138	CACHD1	cache domain containing 1
3	FST	1	100477080*	100227080	100727080	NM_001098956	SNX27	sorting nexin family member 27
4	FST	1	111963966	111713966	112213966	NM_001083790	C3H1orf226	chromosome 3 open reading frame, human C1orf226
5	Rsb (BCS x BMN)	1	118639747*	118389747	118889747	NM_001192122	SFT2D2	SFT2 domain containing 2
6	iHS (BSI)	1	135459627	135209627	135709627	-	-	-
7	Rsb (BCS x BMN)	2	18584559	18334559	18834559	XM_002689886 XM_001254354	OR13F1	olfactory receptor, family 13, subfamily F, member 1
8	Rsb (BCS x BMN)	2	19436541	19186541	19686541	-	-	-
9	iHS (BSI)	2	37506606	37256606	37756606	NM_001079794	AQP3	aquaporin 3 (Gill blood group)
10	iHS (BSI)	2	50934632	50684632	51184632	NM_001034408	POLR1E	polymerase (RNA) I polypeptide E, 53kDa
11	FST	2	52409379	52004402	52659379	XM_002689701 XM_001250503	FAM221B	family with sequence similarity 221, member B
12	iHS (BSI)	2	59231945	58981945	59481945	XM_002689669 XM_864964	FOXB2	forkhead box B2
13	iHS (BCS)	2	111172962	110922962	111422962	XM_003582498 XM_002689443	PALLD	palladin, cytoskeletal associated protein
14	iHS (BCS)	2	118180362	117930362	118430362	NM_001001525	MSTN	myostatin
15	iHS (BCS)	2	128762227	128512227	129012227	XM_002685271 XM_001249354	CWC22	CWC22 spliceosome-associated protein homolog (S. cerevisiae)
16	FST	2	137403728	137153728	137653728	XM_002685281 XM_617685	TLK1	tousled-like kinase 1
17	iHS (BCS)	2	189539157*	189289157	189789157	XM_002685462	CNTNAP5	contactin associated protein-like 5

Região	Metodologia	Chr	Pico	Início	Final	RefSeq	Sigla do gene	Nome do gene
18	iHS (BCS)	2	193984737	193734737	194234737	XM_003581848 -	-	-
19	FST	2	213965644*	213715644	214215644	XM_002685559 XM_003581859 XM_002685560 XM_003581861 XM_003581860 XM_003585767	IKZF2	IKAROS family zinc finger 2 (Helios)
20	Rsb (BCS x BSI)	2	232957928*	232260155	233207928	NM_001192506	DIS3L2	DIS3 mitotic control homolog (S. cerevisiae)-like 2
21	Rsb (BCS x BSI)	2	238567116*	238317116	238893447	NM_001098991	FGR	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog
22	Rsb (BCS x BSI)	2	239564082*	239314082	239814082	NM_001035346	DHDDS	dehydrodolichyl diphosphate synthase
23	iHS (BMN)	3	9783340	9533340	10033340	XM_869668	FAM125B	family with sequence similarity 125, member B
24	iHS (BMN)	3	10515273*	10128941	10765273	NM_001081603	MAPKAP1	mitogen-activated protein kinase associated protein 1
25	iHS (BMN)	3	11712721*	11462721	11962721	XM_003586671	LOC100848819	uncharacterized LOC100848819
26	Rsb (BMN x BSI)	3	52899326	52649326	53226986	- NM_001206302	- IL1RL1	- interleukin 1 receptor-like 1
27	FST	3	99246368*	98996368	99496368	XM_002691182 XM_590497	IL18R1	interleukin 18 receptor 1
28	FST	3	104064729*	103814729	104314729	NM_001102086	STARD7	StAR-related lipid transfer (START) domain containing 7
29	iHS (BSI); Rsb (BCS x BSI)	3	108008375,5	107808182	108208569	XM_003582172	LOC100336620	tryptophan 5-hydroxylase 2-like
30	iHS (BSI)	3	129648614*	129201837	129898614	NM_001037622	MRPL42	mitochondrial ribosomal protein L42
31	iHS (BSI); Rsb (BCS x BSI); Rsb (BMN x	3	133380467*	132428490	132755054	NM_001033610 XR_083785	KRT8 LOC515823	keratin 8 keratin, type II cytoskeletal 8-like

Região	Metodologia	Chr	Pico	Início	Final	RefSeq	Sigla do gene	Nome do gene
	BSI)					XR_082839		
32	iHS (BCS)	3	141203065	140953065	141453065	NM_001075461	DBX2	developing brain homeobox 2
33	iHS (BSI); Rsb (BCS x BSI); FST	3	154933986*	154683986	155183986	NM_001075562	GNS	glucosamine (N-acetyl)-6-sulfatase
34	iHS (BSI)	3	157562244*	156846210	158114029	XM_002684155 XM_002687401	FAM19A2	family with sequence similarity 19 (chemokine (C-C motif)-like), member A2
35	FST	3	161858143*	161608143	162108143	NM_001038091	MARS	methionyl-tRNA synthetase
36	iHS (BSI)	3	184106820	183856820	184356820	-	-	-
37	Rsb (BCS x BMN)	3	200608549	200358549	200858549	XM_003582255 XM_003586107	LOC100848080	uncharacterized LOC100848080
38	FST	3	217731909*	217481909	217981909	NM_001046468	PACSLIN2	protein kinase C and casein kinase substrate in neurons 2
39	FST	4	21287297	21037297	21537297	-	-	-
40	FST	4	24787710*	24537710	25037710	XM_002686679 XM_002686680	ISPD	isoprenoid synthase domain containing
41	FST	4	59789032	59539032	60039032	XM_002686815 XM_583123	GPR141	G protein-coupled receptor 141
42	FST	4	69208152*	68958152	69458152	NM_001038214	SKAP2	src kinase associated phosphoprotein 2
43	FST	4	109733886*	109483886	109983886	XM_002687137 XM_003582156	CNTNAP2	contactin associated protein-like 2
44	iHS (BMN)	5	97597361	97347361	97847361	NM_001001163	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4
45	iHS (BSI)	6	36104954*	35854954	36354954	NM_001083663	HERC3	HECT and RLD domain containing E3 ubiquitin protein ligase 3
46	iHS (BCS)	6	51814853	51564853	52064853	-	-	-
47	FST	6	71533740*	71283740	71783740	XM_001251014 XM_002688279	LOC782369	uncharacterized LOC782369
48	FST	7	13297119	13047119	13547119	NM_001206145	SMAD6	SMAD family member 6

Região	Metodologia	Chr	Pico	Início	Final	RefSeq	Sigla do gene	Nome do gene
49	FST	7	97408830*	97158830	97658830	NM_001101303 XM_002691089 XM_867514	GPR65 TTC8	G protein-coupled receptor 65 tetratricopeptide repeat domain 8
50	FST	8	37942516*	37692516	38192516	NM_001076059	FBXL4	F-box and leucine-rich repeat protein 4
51	iHS (BMN); Rsb (BCS x BMN); Rsb (BMN x BSI)	8	41015484	40765484	41265484	NM_174735	FUT9	fucosyltransferase 9 (alpha (1,3) fucosyltransferase)
52	iHS (BCS); Rsb (BCS x BSI)	9	10082948	9832948	10332948	-	-	-
53	iHS (BSI)	9	29885716	29635716	30135716	NM_001206406 XM_585648	ZHX2	zinc fingers and homeoboxes 2
54	FST	9	36163638	35913638	36413638	XM_002692655	PLAG1	pleiomorphic adenoma gene 1
55	Rsb (BMN x BSI)	9	36746818	36496818	36996818	NM_001046335	IMPAD1	inositol monophosphatase domain containing 1
56	FST	9	55056846	54806846	55306846	NM_173924	IL7	interleukin 7
57	iHS (BCS)	9	63523344	63273344	63773344	NM_001105499	CSMD3	CUB and Sushi multiple domains 3
58	FST	10	29502667*	29252667	29752667	XM_002684281 XM_002691757 XM_003586722 XM_867800	RXFP2	relaxin/insulin-like family peptide receptor 2
59	Rsb (BMN x BSI)	10	59095153	58845153	59422145	XM_003582877 XM_002691909	SLITRK1	SLIT and NTRK-like family, member 1
60	iHS (BMN); Rsb (BCS x BMN); Rsb (BMN x BSI)	10	64726524	64476524	64976524	-	-	-
61	FST	10	79129227	78879227	79379227	-	-	-
62	FST	11	35982089	35732089	36232089	NM_001034039	COL1A1	collagen, type I, alpha 1
63	Rsb (BCS x BSI)	12	14458685	14208685	14862434	-	-	-
64	iHS (BCS)	12	41343197*	41093197	41593197	XM_003583233	KIF1B	kinesin family member 1B

Região	Metodologia	Chr	Pico	Início	Final	RefSeq	Sigla do gene	Nome do gene
						XM_002694045		
65	Rsb (BCS x BSI)	13	1366306	897107	1616306	NM_174817	PLCB1	phospholipase C, beta 1 (phosphoinositide-specific)
66	Rsb (BCS x BMN)	13	4314690	4064690	4564690	-	-	-
67	Rsb (BCS x BSI)	13	8385113*	8135113	8635113	NM_001105029	MACROD2	MACRO domain containing 2
68	Rsb (BCS x BSI)	13	22555622*	22305622	22805622	NM_001038185	SPAG6	sperm associated antigen 6
69	Rsb (BCS x BMN); Rsb (BCS x BSI)	13	23148764	22898764	23398764	NM_001206130	MSRB2	methionine sulfoxide reductase B2
70	FST	13	25164063*	24914063	25414063	NM_001206442	GPR158	G protein-coupled receptor 158
71	iHS (BCS)	13	46647627*	46397627	46897627	NM_001078046	CDS2	CDP-diacylglycerol synthase (phosphatidate cytidyltransferase) 2
72	FST	14	13401157*	13151157	13651157	XM_869815 XM_002694857	IL17C	interleukin 17C
73	FST	14	14657550	14407550	14907550	XM_003587208 XM_003583369	LOC100847136	uncharacterized LOC100847136
74	iHS (BSI)	15	29495444	29245444	29745444	NM_001034765	THY1	Thy-1 cell surface antigen
75	FST	15	43007145*	42757145	43257145	XM_002693106 XM_001252217	DENND5A	DENN/MADD domain containing 5A
76	iHS (BSI); Rsb (BCS x BMN); Rsb (BCS x BSI)	16	36683961	35602290	36031385	NM_001076238	WDR70	WD repeat domain 70
77	iHS (BSI); Rsb (BCS x BMN); Rsb (BCS x BSI); FST	16	37086176	36392330	36836383	NM_001206584	NIPBL	Nipped-B homolog (Drosophila)
78	iHS (BSI); Rsb (BCS x BMN); Rsb (BCS x BSI)	16	37464091,67	37081737	37581737	NM_001077070	MGC134093	uncharacterized protein MGC134093
79	iHS (BSI); Rsb (BCS x BMN); Rsb (BCS x BSI); FST	16	38418208*	38168208	38668208	XM_002696381	SPEF2	sperm flagellar 2
80	iHS (BSI); Rsb (BCS	16	41680378*	41820192	41930378	XM_001252695	LOC784368	golgi phosphoprotein 3-like

Região	Metodologia	Chr	Pico	Início	Final	RefSeq	Sigla do gene	Nome do gene
	x BMN); Rsb (BCS x BSI); Rsb (BMN x BSI)					XM_002696405		
81	iHS (BSI); Rsb (BCS x BMN); Rsb (BCS x BSI)	16	43273047	43023047	43523047	-	-	-
82	iHS (BSI); Rsb (BCS x BSI); FST	16	43308385,33	43849474	44349474	-	-	-
83	iHS (BSI); iHS (BMN); Rsb (BCS x BMN); Rsb (BCS x BSI)	16	45834533	46995495	47137875	-	-	-
84	iHS (BSI); Rsb (BCS x BSI); Rsb (BMN x BSI)	16	49585602	49758827	50258827	-	-	-
85	FST	19	27281255*	27031255	27531255	XM_002696970 XM_605070	CNTN3	contactin 3 (plasmacytoma associated)
86	iHS (BCS)	21	18528589*	18278589	18778589	XM_602972	LOC524642	glycerophosphodiester phosphodiesterase domain containing 4-like
87	Rsb (BCS x BMN)	22	19477317	19227317	19727317	XM_003587871 XM_003584084	HPSE2	heparanase 2
88	FST	25	6062510*	5812510	6312510	XM_002698790 XM_607409	PCNXL2	pecanex-like 2 (Drosophila)
89	FST	25	7392689*	7106301	7642689	XM_002698809	TARBP1	TAR (HIV-1) RNA binding protein 1
90	FST	25	9173663	8923663	9423663	XM_596273	EDARADD	EDAR-associated death domain
91	iHS (BCS)	25	22965788*	22715788	23215788	NM_001079638	CTNNA3	catenin (cadherin-associated protein), alpha 3
92	FST	26	35828422	35578422	36078422	NM_001080280	SLC20A2	solute carrier family 20 (phosphate transporter), member 2

*SNP significativo contido dentro do gene candidato.

DISCUSSÃO

Considerando cada análise individualmente, foram identificadas 120 regiões genômicas candidatas a conter genes sob seleção. Após o agrupamento das regiões coincidentes, restaram 92 regiões genômicas, distribuídas ao longo de 21 dos 26 cromossomos autossômicos do genoma ovino, que foram consideradas para a identificação de genes candidatos.

A quantidade de regiões genômicas candidatas identificadas no presente estudo é maior do que o relatado em outros estudos envolvendo a espécie ovina (KIJAS *et al.*, 2012; MORADI *et al.* 2012; FARIELLO *et al.*, 2013), o que pode estar relacionado ao fato de que este trabalho se baseou na utilização de três metodologias distintas para identificar assinaturas de seleção enquanto os demais trabalhos foram baseados em apenas uma metodologia.

Das 92 regiões genômicas identificadas, a grande maioria (77) foi identificada utilizando apenas uma metodologia, mas houve 11 sinais coincidentes entre duas metodologias e quatro sinais coincidentes entre as três metodologias. Esta discrepância entre os sinais de seleção encontrados utilizando metodologias diferentes se dá pelo fato de que cada uma das metodologias utilizadas captura marcas diferentes deixadas pela seleção no genoma (SABETI *et al.*, 2002; BEAUMONT, 2005; STORZ, 2005; VOIGHT *et al.*, 2006; TANG *et al.*, 2007).

Por conta disso, a identificação de assinaturas de seleção em uma determinada região por mais de uma metodologia pode ser encarada como uma forte evidência da ação da seleção sobre aquele gene, porém o fato de um gene ter sido identificado como sujeito a seleção por uma metodologia e não por outra não exclui a possibilidade daquele gene estar sob seleção (HOHENLOHE *et al.*, 2010; OLEKSYK *et al.*, 2010).

A estratégia adotada no presente trabalho para definir os genes candidatos dentro de cada região consistiu em: 1) para as regiões identificadas por apenas uma metodologia e que apenas um SNP foi considerado significativo num intervalo de 500kb, o gene mais próximo deste SNP foi considerado como candidato; 2) para as regiões identificadas por apenas uma metodologia e que apresentaram mais de um SNP significativo no intervalo de 500kb, o gene mais próximo do SNP com maior valor da estatística foi considerado como candidato; 3) para as regiões coincidentes para mais de

uma metodologia e que o mesmo SNP foi significativo para mais de uma metodologia, o gene mais próximo deste SNP foi considerado como candidato e 4) para as regiões coincidentes para mais de uma metodologia mas que os SNPs significativos para cada metodologia eram diferentes, foi calculado o ponto médio entre os SNPs significativos e este foi utilizado para definir o gene candidato.

O tamanho das regiões genômicas identificadas variou entre 110186 e 1267819 bp, o que resultou em algumas regiões que não continham nenhum gene anotado, mas também foram identificadas regiões contendo muitos genes anotados. O tamanho das regiões identificadas e a quantidade de genes contidos dentro destas regiões, torna muitas vezes difícil a identificação de genes candidatos que expliquem o sinal de seleção.

Deve-se ressaltar também que a identificação de um gene como candidato a estar sob seleção não é, por si só, uma prova irrefutável de que o processo seletivo ocorreu de fato. Além disso, não se pode definir uma relação de causa e efeito à partir deste tipo de estudo. No entanto, a identificação de um gene como candidato a estar sob seleção pode ter um papel importante na identificação de regiões genômicas envolvidas na manifestação de características de interesse econômico e ecológico.

A anotação das regiões coincidentes entre as três metodologias (F_{ST} , iHS e Rsb) possibilitou a identificação de 78 genes candidatos. Destes, três foram identificados pelas três metodologias conjuntamente (GNS, NIPBL e SPEF2) e seis foram identificados através de duas metodologias (LOC100336620, KRT8, FUT9, WDR70, MGC134093 e LOC784368), o que faz com que estes genes possam ser considerados como fortes candidatos a estar sob seleção.

Também, os genes CNTNAP5, DIS3L2, KIF1B, RXFP2 e TARBP1 foram identificados no presente estudo e em outros estudos envolvendo ovinos e bovinos (FLORI et al., 2009; GAUTIER et al., 2009; KIJAS *et al.*, 2012; STELA *et al.*, 2010), o que também pode ser encarado como uma evidência adicional de que estes genes sejam candidatos interessantes.

O processo de domesticação provocou mudanças tanto morfológicas quanto comportamentais nas espécies de animais domésticos. Sabe-se que as espécies domésticas apresentam cérebros menores, órgãos do sentido menos desenvolvidos e diferenças de resistência contra patógenos quando comparados com seus parentes

selvagens. Além disso, o processo de seleção artificial consciente levou à fixação de características em raças específicas, como por exemplo presença/ausência de chifres e padrões de pelagem, além de promover diferenciação em características de interesse econômico e adaptativas (DIAMOND, 2002; TORO; MÄKI-TANILA, 2007; GAUTIER *et al.*, 2009; GREGORY, 2009; TRUT *et al.*, 2009).

O gene RXFP2, identificado neste estudo como candidato a seleção através da análise de F_{ST} (Tabela 9), foi recentemente associado com presença/ausência e também com a morfologia dos chifres em ovinos (JOHNSTON *et al.*, 2011). Além disso, este gene foi considerado por Kijas *et al.* (2012) como responsável pelo principal sinal de seleção em um estudo envolvendo 74 raças de ovinos distribuídas ao redor do mundo (que também incluiu os animais utilizados para a realização do presente trabalho).

Este sinal de seleção (F_{ST}) foi identificado tanto na comparação considerando as três raças brasileiras conjuntamente quanto na comparação Crioula Lanada x Santa Inês (Figura 1) e pode ser explicado pelo fato de que tanto para a raça Morada Nova quanto para a Raça Santa Inês, o padrão racial indica que os animais devam ser mochos, porém na raça Morada Nova a presença de chifres é considerada um defeito para fins de registro genealógico enquanto para a Raça Santa Inês é permitida a presença de chifres rudimentares. Já na raça Crioula Lanada, são aceitos animais tanto mochos quanto chifrudos, além do mais, nesta raça aceita-se, para fins de registro genealógico, animais que apresentem mais de um par de chifres (policerismo).

A importância da presença/ausência de chifres do ponto de vista de definição racial pode ser ressaltada pelo fato de que, também em bovinos, a região próxima ao *locus* responsável por esta característica foi identificada como sujeita à seleção (STELA *et al.*, 2010).

Em ovinos, existe variação genética tanto para a presença/ausência de lã quanto para características relacionadas a sua produção e qualidade (PURVIS; FRANKLIN, 2005). As raças Morada Nova e Santa Inês foram desenvolvidas na região Nordeste do Brasil e são consideradas raças deslanadas (sendo que na raça Santa Inês permite-se a presença de resquírios de lã), já a raça Crioula Lanada é classificada como uma raça lanada e predomina na região Sul do país.

O gene EDARADD foi identificado pela análise de F_{ST} tanto na comparação considerando as três raças brasileiras conjuntamente quanto na comparação Crioula

Lanada x Santa Inês (Figura 1). Este gene está localizado no cromossomo OAR25 em uma região que contém QTLs para tamanho das fibras e peso do velo (PONZ *et al.*, 2001; HU *et al.*, 2013) e participa da via metabólica do TNF (fator de necrose tumoral), uma das mais importantes na morfogênese e desenvolvimento do folículo piloso. Além disso, mutações no gene EDARADD estão associadas com fenótipos relacionados com desenvolvimento anormal de pêlos em várias espécies de mamíferos (CUI; SCHLESSINGER, 2006; SHIMOMURA; CRISTIANO, 2010).

Um gene candidato muito interessante identificado pela análise de F_{ST} foi o TRPM8 (Tabela 9). Este gene está envolvido com a regulação da temperatura corporal em baixas temperaturas através tanto de mecanismos de autonômicos (indução da atividade termogênica do tecido adiposo marrom e vasoconstrição na pele) quanto de alterações comportamentais (ALMEIDA *et al.*, 2012; GAVVA *et al.*, 2012; TAJINO *et al.*, 2011).

O sinal de seleção foi identificado tanto na comparação considerando as três raças brasileiras conjuntamente quanto na comparação Crioula Lanada x Santa Inês (Figura 1), o que indica que o gene TRPM8 deve ser um candidato interessante para futuras investigações, uma vez que a raça Crioula Lanada foi desenvolvida na região Sul do Brasil e que as condições climáticas nesta região são bastante diferentes da região semiárida, onde foram desenvolvidas as raças deslanadas Morada Nova e Santa Inês. Com isso, pode-se sugerir que este gene esteja relacionado com a capacidade dos animais Crioula Lanada regular a temperatura corporal em situações de baixa temperatura e que esta assinatura de seleção esteja relacionada com adaptabilidade às condições de temperaturas mais baixas encontradas na região Sul do país.

Entre as raças analisadas existe uma variação considerável em relação ao tamanho adulto dos animais, enquanto a raça Santa Inês é considerada de porte grande, as raças Crioula Lanada e Morada Nova são consideradas de pequeno porte. É interessante notar que foram identificados neste estudo alguns genes (DIS3L2, PLAG1, NIPBL) associados à variação de altura.

O gene NIPBL foi identificado pela análise de F_{ST} tanto na comparação entre as três raças quanto na comparação Crioula Lanada x Santa Inês (Figura 1). Além disso, a análise de iHS revelou que, na raça Santa Inês, a região genômica que contém estes genes apresentam um aumento do desequilíbrio de ligação em relação ao resto do

genoma (Figura 3). Já a análise de Rsb sugere um aumento do desequilíbrio de ligação destas regiões genômicas nas raças Santa Inês e Morada Nova quando comparadas com a raça Crioula Lanada (Figura 5). Pode-se sugerir, portanto, que a raça Santa Inês seja a mais provável pelos sinais de seleção identificados. Mutações neste gene estão associadas a síndrome Cornelia de Lange, uma desordem do desenvolvimento que pode ser caracterizada por deformações faciais e nas extremidades superiores, hirsutismo, defeitos cardíacos, baixa estatura e retardo mental (GILLIS *et al.*, 2004; KEMPER *et al.*, 2012).

O gene PLAG1 foi identificado pela análise de F_{ST} , tanto para a comparação considerando as três raças brasileiras conjuntamente, quanto na comparação Morada Nova x Santa Inês (Figura 1). Estudos recentes indicam que este gene está associado a variação de estatura em bovinos (KARIM *et al.*, 2011; LITTLEJOHN *et al.*, 2011).

Já o gene DIS3L2 foi identificado pela análise de Rsb na comparação Santa Inês x Crioula Lanada (Figura 5). Este gene também é candidato a estar sob seleção na espécie bovina (GAUTIER *et al.*, 2009) e, em humanos está associado a síndrome de Perlman, uma desordem caracterizada por crescimento exagerado e suscetibilidade ao desenvolvimento de tumores (ASTUTI *et al.*, 2012).

O tamanho adulto do animal é uma característica diretamente relacionada com eficiência produtiva pelo fato de animais maiores demandarem um maior gasto de energia de manutenção. A identificação de genes candidatos a estar sob seleção na raça Santa Inês que possam ser relacionados a variação de estatura corrobora os relatos de que a estatura média da raça aumento nos últimos anos e abre possibilidades para estudos de validação da associação entre estes genes e a característica.

Alguns genes identificados no presente estudo estão envolvidos no processo reprodutivo, como os genes SPEF2 e SPAG6 que são importantes para a espermatogênese (ESCALIER, 2006; SIRONEN *et al.*, 2010, 2012). Mutações nestes genes estão associadas à infertilidade em machos em várias espécies de mamíferos (ESCALIER, 2006; SIRONEN *et al.*, 2006, 2011) e o gene SPAG6 é sugerido como candidato a seleção também em humanos (BARREIRO *et al.*, 2008; WILLIAMSON *et al.*, 2007).

Rsb

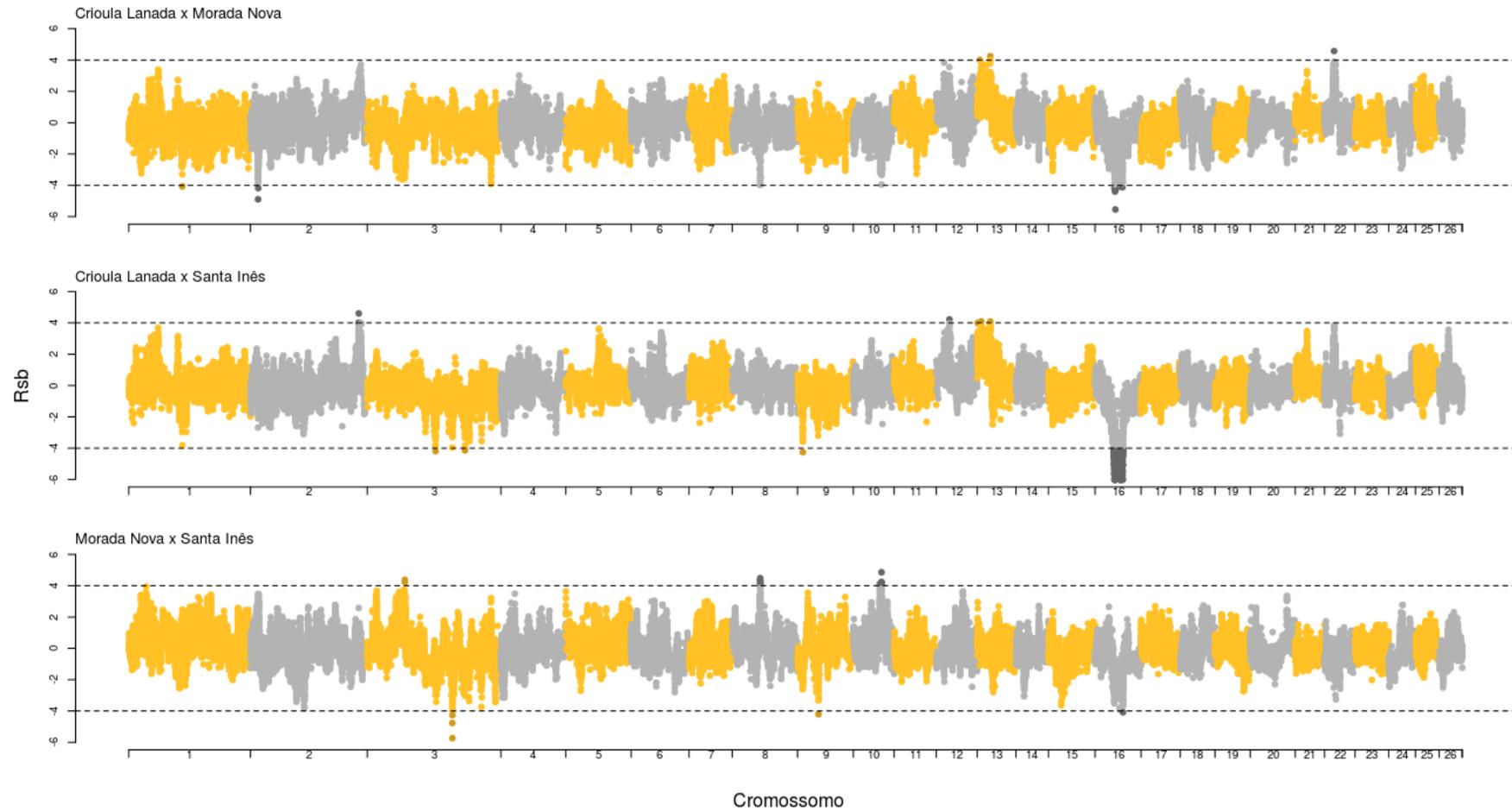


Figura 5. Manhattan plot dos valores de Rsb. Cromossomos ímpares estão representados em amarelo enquanto os cromossomos pares estão representados em cinza. Os pontos acima da linha tracejada indicam os SNPs extremos da distribuição.

O gene SPEF2 foi identificado na análise de F_{ST} tanto na comparação entre as três raças quanto na comparação Crioula Lanada x Santa Inês (Figura 1). Além disso, a análise de iHS revelou que, na raça Santa Inês, a região genômica que contém este gene apresenta um aumento do desequilíbrio de ligação em relação ao resto do genoma (Figura 3). Já a análise de Rsb sugere um aumento do desequilíbrio de ligação desta região genômica nas raças Santa Inês e Morada Nova quando comparadas com a raça Crioula Lanada (Figura 5). Pode-se sugerir, portanto, que a raça Santa Inês seja a mais provável pelo sinal de seleção identificados, embora, haja a possibilidade de que na raça Morada Nova estas regiões também estejam sendo selecionadas.

Já para o gene SPAG6, a análise de Rsb sugere um aumento do desequilíbrio de ligação desta região genômica na raça Crioula Lanada quando comparada com a raça Santa Inês (Figura 5), sugerindo que a primeira é responsável pelo sinal da seleção.

Embora muitos genes tenham sido sugeridos como alvos de seleção e alguns destes apresentem evidências de participação em mecanismos fisiológicos importantes, o presente estudo não tem o objetivo de definir uma relação de causa e efeito entre genótipo e fenótipo. Além disso, existe a possibilidade de que alguns dos sinais identificados neste estudos não sejam devido à seleção e sim a outros processos evolutivos (sistemáticos ou não sistemáticos).

Apesar disso, o presente estudo abre a possibilidade de uma investigação mais detalhada no sentido de se validar associações genótipo - fenótipo que possam explicar diferenças/particularidades ecológicas e produtivas entre as raças de ovinos localmente adaptadas brasileiras.

CONCLUSÕES

A combinação de três metodologias diferentes de identificação de assinaturas de seleção em ovinos pertencentes às principais raças localmente adaptadas brasileiras possibilitou a identificação de inúmeros genes candidatos a estar sob seleção nestas raças. Alguns sinais de seleção previamente identificados tanto em ovinos quanto em bovinos puderam ser identificados e além disso, foram identificados inúmeros novos candidatos. Dentre os genes identificados como candidatos, alguns parecem estar

associados a processos fisiológicos importantes do ponto de vista adaptativo/produtivo e embora não seja possível, através deste estudo, confirmar as relações genótipo - fenótipo, abre-se uma nova possibilidade de investigação para tal validação.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, M.C.; HEW-BUTLER, T.; SORIANO, R.N.; RAO, S.; WANG, W.; WANG, J.; TAMAYO, N.; OLIVEIRA, D.L.; *et al.* Pharmacological blockade of the cold Receptor TRPM8 attenuates autonomic and behavioral cold defenses and decreases deep body temperature. **J. Neurosci.**, v.32, n.6, p.2086-2099, 2012.
- AMARANTE, A.F.T.; BRICARELLO, P.A.; ROCHA, R.A.; GENNARI, S.M. Resistance of Santa Ines, Suffolk and Ile de France sheep to naturally acquired gastrointestinal nematode infections. **Vet. Parasit.**, 120(1-2): 91-106, 2009
- ASTUTI, D.; MORRIS, M.R.; COOPER, W.N.; STAALS, R.H.J.; WAKE, N.C.; FEWS, G.A.; GILL, H.; GENTLE, D.; SHUIB, S.; RICKETTS, C.J.; *et al.* Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. **Nat. Genet.**, v.44, p.277-284, 2012.
- BARREIRO, LB.; LAVAL, G.; QUACH, H.; PATIN, E.; QUINTANA-MURCI, L. Natural selection has driven population differentiation in modern humans. **Nature Gen.**, v.40, n.3, p.340-345, 2008.
- BEAUMONT, M.A. Adaptation and speciation: what can Fst tell us? **Trends Ecol. Evol.**, v.20, n.8, p.435-440, 2005.
- BUENO, M.S.; CUNHA, E.A.; VERÍSSIMO, C.J.; SANTOS, L.E.; LARA, M.A.C.; OLIVEIRA, S.M.; SPÓSITO FILHA, E.; REBOUÇAS, M.M. Infecção por nematodos en razas de ovejas cárnicas criadas intensivamente en la región del sudeste del Brasil. **Arch. Zootec.**, 51:271-278, 2002
- CHEN, H.; SHARP, B.M. Content-rich biological network constructed by mining PubMed abstracts. **BMC Bioinf.**, v.8, n.5, 2004.
- COUTINHO, M.F.; LACERDA, L.; ALVES, S. Glycosaminoglycan storage disorders: a review. **Biochem. Res. Int.**, 2012.

- CUI, C-Y.; SCHLESSINGER, D. EDA signaling and appendage development. **Cell Cyc.**, v.5, n.21, p.2477-2483, 2006.
- DIAMOND, J. Evolution, consequences and future of plant and animal domestication. **Nature**, v.418, p.700-707, 2002.
- ESCALIER, D. Knockout mouse models of sperm flagellum anomalies. **Hum. Rep. Update**, v.12, n.4, p.449-461, 2006.
- FARIELLO, M.I.; BOITARD, S.; NAYA, H.; SANCRISTOBAL, M.; SERVIN, B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. **Genetics**, v.193, n.3, p.929-941, 2013.
- FLORI, L.; FRITZ, S.; JAFFRÉZIC, F.; BOUSSAHA, M.; GUT, I.; HEATH, S.; FOULLEY, J.L.; GAUTIER, M. The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. **PLoS ONE** 4(8): e6595, 2009.
- FOLL, M.; GAGGIOTTI, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. **Genetics**, v.180, p.977-993, 2008.
- GAUTIER, M.; FLORI, L.; RIEBLER, A.; JAFFRÉZIC, F.; LALOÉ, D.; GUT, I.; MOAZAMI-GOUDARZI, K.; FOULLEY, J.L. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. **BMC Genomics**, v.10:550, 2009.
- GAUTIER, M.; VITALIS, R. rehh : an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. **Bioinformatics**, v.28, n.8, p.1176-1177, 2012.
- GAVVA, N.R.; DAVIS, C.; LEHTO, S.G.; RAO, S.; WANG, W.; ZHU, D.X.D. Transient receptor potential melastatin 8 (TRPM8) channels are involved in body
- GILLIS, L.A.; MCCALLUM, J.; KAUR, M.; DESCIPPIO, C.; YAEGER, D.; MARIANI, A.; KLINE, A.D.; LI, H-H.; DEVOTO, M.; JACKSON, L.G.; KRANTZ, I.D. NIPBL Mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. **Am. J. Hum. Gen.**, v.75, n.4, p.610-623, 2004.
- GOSH, A.K. Factors involved in the regulation of type I collagen gene expression: implication in fibrosis. **Exp. Biol. Med.**, v.227, p.301-314, 2002.
- GREGORY, T.R. Artificial selection and domestication: Modern lessons from Darwin's enduring analogy. **Evo. Edu. Outreach**, v.2, p.5-27, 2009.

- HAYES, B.J.; LIEN, S.; NILSEN, H.; OLSEN, H.G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISSEN, T.H.E. The origin of selection signatures on bovine chromosome 6. **Anim. Genet.**, v.39, p.105-111, 2008.
- HOFFMANN, R.; VALENCIA, A.A. Gene network for navigating the literature. **Nat. Genet.**, 36(664), 2004.
- HOHENLOHE, P.A.; PHILLIPS, P.C.; CRESKO, W.A. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. **Int. J. Plant. Sci.**, v.171, n.9, p.1059-1071, 2010.
- HU, Z-L.; PARK, A.A.; WU, X-L.; REECY, J.M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nuc. Ac. Res.**, n.41, v.D1, p. D871-D879, 2013.
- HUANG, D.W.; SHERMAN, B.T.; LEMPICKI, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. **Nature Protoc.**, n.4, v.1, p44-57, 2009.
- JOHNSTON, S.E.; MCEWAN, J.C.; PICKERING, N.K.; KIJAS, J.W.; BERALDI, D.; KIJAS, J.W.; LENSTRA, J.A.; HAYES, B.; BOITARD, S.; PORTO NETO, L.R.; CRISTOBAL, M.S.; SERVIN, B.; MCCULLOCH, R.; WHAN, V.; GIETZEN, K.; *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. **PLoS Biol.**, 10(2):e1001258, 2012.
- JOHNSTON, S.E.; MCEWAN, J.C.; PICKERING, N.K.; KIJAS, J.W.; BERALDI, D.; PILKINGTON, J.G.; PEMBERTON, J.M.; SLATE, J. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. **Mol. Ecol.**, v.20, p.2555-2566, 2011.
- KARIM, L.; TAKEDA, H.; LIN, L.; DRUET, T.; ARIAS, J.A.C.; BAURAIN, D.; CAMBISANO, N.; DAVIS, S.R.; FARNIR, F.; GRISART, B; *et al.* Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. **Nat. Genet.**, v.43, p.405-413, 2011.
- KEMPER, K.E.; VISSCHER, P.M.; GODDARD, M.E. Genetic architecture of body size in mammals. **Gen. Biol.**, v.13, n.244, 2012.
- LITTLEJOHN, M.; GRALA, T.; SANDERS, K.; WALKER, C.; WAGHORN, G.; MACDONALD, K.; COPPIETERS, W.; GEORGES, M.; SPELMAN, R.; HILLERTON, E.; *et al.* Genetic variation in PLAG1 associates with early life body

weight and peripubertal weight and growth in *Bos taurus*. **Anim. Genet.**, v.43, p.591-594, 2011.

MARIANTE, A.S.; ALBUQUERQUE, M.S.M.; EGITO, A.A.; MCMANUS, C.; LOPES, M.A.; PAIVA, S.R. Present status of the conservation of livestock genetic resources in Brazil. **Livestock Sci.**, v.120, n.3, p.204-212, 2009.

MCMANUS, C.; PAIVA, S.R.; ARAÚJO, R.O. Genetics and breeding of sheep in Brazil. **R. Bras. Zootec.**, v.39, p.236-246, 2010.

MCVEAN, G.A.T.; MYERS, S.; HUNT, S.; DELOUKAS, P.; BENTLEY, D.; DONNELLY, P. The fine-scale structure of recombination rate variation in the human genome. **Science.**, v.304, p.581-584, 2004.

NIELSEN, R. Molecular Signatures of Natural Selection. **Annu. Rev. Genet.**, v.39, p.197-218, 2005.

NIELSEN, R. Statistical tests of selective neutrality in the age of genomics. **Heredity**, v.86, p.641-647, 2001.

OLEKSYK, T.K.; SMITH, M.W.; O'BRIEN, J. Genome-wide scans for footprints of natural selection. **Philos. Trans. R. Soc. Lond. B.**, v.365, p.185-205, 2010.

PONZ, R.; MORENO, C.; ALLAIN, D.; ELSEN, J.M.; LANTHIER, F.; LANTIER, I.; BRUNEL, J.C.; PÉREZ-ENCISO, M. Assessment of genetic variation explained by markers for wool traits in sheep via a segment mapping approach. **Mamm. Gen.**, v.12, p.569-572, 2001.

PURVIS, I.W.; FRANKLIN, I.R. Major genes and QTL influencing wool production and quality: a review. **Genet. Sel. Evol.**, v.37, p.S97-S107, 2005.

SABETI, P.C.; REICH, D.E.; HIGGINS, J.M.; LEVINE, H.Z.P.; RICHTER, D.J.; SCHAFFNER, S.F.; GABRIEL, S.B.; PLATKO, J.V.; PATTERSON, N.J.; MCDONALD, G.J.; *et al.* Detecting recent positive selection in the human genome from haplotype structure. **Nature**, v.419, n.24, p.832-837, 2002.

SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. **Am. J. Hum. Gen.**, v.78, p.629-44, 2006.

SCHLÖTTERER, C. Hitchhiking mapping: functional genomics from the population genetics perspective. **Trends Genet.**, v.19, n.1, p.32-38, 2003.

- SHIMOMURA, Y.; CHRISTIANO, A.M. Biology and genetics of hair. **Annu. Rev. Genomics Hum. Genet.**, v.11, p.109-132, 2010.
- SIRONEN, A.; HANSEN, J.; THOMSEN, B.; ANDERSSON, M.; VOLKKI, J.; TOPPARI, J.; KOTAJA, N. Expression of SPEF2 during mouse spermatogenesis and identification of IFT20 as an interacting protein. **Biol. Reprod.**, v82, p.580-590, 2010.
- SIRONEN, A.; KOTAJA, N.; MULHERN, H.; WYATT, T.A.; SISSON, J.H.; PAVLIK, J.A.; MIILUNIEMI, M.; FLEMING, M.D.; LEE, L. Loss of SPEF2 function in mice results in spermatogenesis defects and primary ciliary dyskinesia. **Biol. Reprod.**, v.85, p.690-701, 2011.
- SIRONEN, A.; THOMSEN, B.; ANDERSSON, M.; HOLA, V.; VILKKI, J. An intronic insertion in KPL2 results in aberrant splicing and causes the immotile short-tail sperm defect in the pig. **Proc. Nat. Acad. Sci. U.S.A.**, v.103, n.13, p.5006-5011, 2006.
- SIRONEN, A.; UIMARI, P.; ISO-TOURU, T.; VILKKI, J. L1 insertion within SPEF2 gene is associated with increased litter size in the Finnish Yorkshire population. **J. Anim. Breed. Genet.**, v.129, p.92-97, 2012.
- STELLA, A.; AJMONE-MARSAN, P.; LAZZARI, B.; BOETTCHER, P. Identification of selection signatures in cattle breeds selected for dairy production. **Genetics**, v.185, p.1451-1461, 2010.
- STORZ, J.F. Using genome scans of DNA polymorphism to infer adaptive population divergence. **Mol. Ecol.**, v.14, p.671-668, 2005.
- TAJINO, K.; HOSOKAWA, H.; MAEGAWA, S.; MATSUMURA, K.; DHAKA, A.; KOBAYASHI, S. Cooling-sensitive TRPM8 is thermostat of skin temperature against temperature regulation. **Molec. Pain**, 8:36, 2012.
- TANG, K.; THORNTON, K.R.; STONEKING, M. A new approach for using genome scans to detect recent positive selection in the human genome. **PLoS Biol.**, 5(7): e171, 2007.
- THE INTERNATIONAL SHEEP GENOMICS CONSORTIUM (ISGC), ARCHIBALD, A.L.; COCKETT, N.E.; DALRYMPLE, B.P.; FARAUT, T.; KIJAS, J.W.; MADDOX, J.F.; MCEWAN, J.C.; HUTTON ODDY, V.; RAADSMA, H.W.; WADE, C.; WANG, J.; WANG, W.; XUN, X. The sheep genome reference sequence: a work in progress. **Anim. Genet.**, n.41, p.449-453, 2010.

TORO, M.; MÄKI-TANILA, A. Genomics Reveals Domestication History and Facilitates Breed Development. In: Oldenbroek.; K (ed) Utilisation and Conservation of Farm Animal Genetic Resources. Wageningen Academic Publishers, Wageningen, pp 75-102, 2007.

TRUT, L.; OSKINA, I.; KHARLAMOVA, A. Animal evolution during domestication: The domesticated fox as a model. **Bioessays**, v.31, n.3, p.349-360, 2009.

VOIGHT, B.F.; KUDARAVALLI, S.; WEN, X.; PRITCHARD, J.K. A map of recent positive selection in the human genome. **PLoS Biol.**, v.4, n.3, p.446-458, 2006.

WILLIAMSON, S.H.; HUBISZ, M.J.; CLARK, A.G.; PAYSEUR, B.A.; BUSTAMANTE, C.D.; NIELSEN, R, Localizing recent adaptive evolution in the human genome. **PLoS Genet.**, 3(6): e90, 2007.

CAPÍTULO III

**Identificação de subestrutura de populações em ovinos de raças
localmente adaptadas brasileiras**

IDENTIFICAÇÃO DE SUBESTRUTURA DE POPULAÇÕES EM OVINOS DE RAÇAS LOCALMENTE ADAPTADAS BRASILEIRAS

RESUMO

As raças brasileiras localmente adaptadas são descendentes de animais trazidos durante o período de colonização e, desde então, vêm sendo submetidas a processos evolutivos sistemáticos e não sistemáticos, o que resultou na formação de genótipos adaptados às diversas condições ambientais brasileiras. A definição da melhor estratégia de conservação depende de um conhecimento profundo das populações alvo dos programas de conservação. Existem, pelo menos, duas situações nas raças de ovinos localmente adaptadas brasileiras, que sugerem a presença de subestruturação importante, a presença dos ecótipos Serrana e Fronteira na raça Crioula Lanada e a presença das linhagens “velha” e “nova” dentro da raça Santa Inês. Com base nisso, o objetivo do presente estudo foi identificar e caracterizar a presença de subestruturação genética nas três principais raças de ovinos localmente adaptadas brasileiras: Crioula Lanada, Morada Nova e Santa Inês. De uma forma geral, foi observada uma diferenciação genética bastante considerável ao comparar os rebanhos de cada raça analisada, algumas vezes excedendo valores de diferenciação entre diferentes raças ovinas. Tanto na raça Crioula Lanada quanto na raça Santa Inês, pode ser observada a presença de grupos bem distintos de indivíduos, sugerindo a efetiva presença de diferentes ecótipos/linhagens dentro destas raças. Os resultados obtidos no presente estudo indicam que o programa de conservação da raça Crioula Lanada deve incluir animais pertencentes aos dois ecótipos e que seja utilizada esta informação na definição dos acasalamentos. Além disso, uma discussão junto às associações de criadores a fim de se definir critérios no registro genealógico dos animais para identificação de qual ecótipo os animais pertencem deve ser iniciada. Para a raça Santa Inês, deve-se manter no programa de conservação, pelo menos, animais pertencentes aos rebanhos BSI_ECO, BSI_EMN e BSI_EBD, uma vez que estes são geneticamente distintos do rebanho BSI_GAA, que pode ser considerado como representativo do rebanho comercial. O presente estudo, portanto, demonstra que o aprofundamento do conhecimento sobre as populações alvo

pode evitar esforços (e gastos) desnecessários e também pode ser decisivo para a inclusão de populações adicionais da mesma raça no contexto dos programas de conservação.

Palavras-chave: conservação, miscigenação, ovinos nativos

IDENTIFICATION OF POPULATION SUBSTRUCTURE IN BRAZILIAN LOCALLY ADAPTED SHEEP BREEDS

ABSTRACT

The Brazilian locally adapted breeds are descendant of animals brought during the settlement period, and since then, submitted to systematic and non systematic evolutive processes. As a consequence, genotypes adapted to Brazilian diverse environmental conditions were created. The definition of the best conservation strategy depends on a deep knowledge of the target populations. There are, at least, two situations in the Brazilian locally adapted sheep breeds that suggest the presence of an important breed substructure: the occurrence of Serrana and Fronteira ecotypes in the Brazilian Creole breed and the existence of the “old” and the “new” Santa Ines lineages. Therefore, the aim of this study was to identify and characterize the presence of genetic substructure within the three main locally adapted Brazilian sheep breeds: Brazilian Creole, Morada Nova and Santa Ines. The level of genetic differentiation between herds of the same breeds was, in general, high, sometimes exceeding the comparison between breeds. Both in the Brazilian Creole and in the Santa Ines breeds the presence of distinct groups on animals could be observed, what suggests the occurrence of different ecotypes/lineages within these breeds. Our results suggest that the conservation program of Brazilian Creole breed must include individuals from both ecotypes and also that the information about the origin of the individuals should be considered in the definition of the breeding strategies. Furthermore, it is important to the breeders association to include the definition of ecotype in the genealogical records. For the Santa Ines breed, it can be suggested that, at least, the herds BSI_ECO, BSI_EMN and BSI_EBD must be maintained in the conservation program. This can be justified by the fact that these herds are genetically different from the herd BSI_GAA, which can be considered as representative from the commercial herd. This study, therefore, demonstrates that a deep knowledge about the target populations is essential for the effective development of conservation schemes.

Keywords: conservation, admixture, native sheep

INTRODUÇÃO

As raças brasileiras localmente adaptadas de ovinos são descendentes de indivíduos trazidos durante o período de colonização do Brasil e, desde então, vêm sendo submetidas a processos evolutivos sistemáticos e não sistemáticos. Como resultado, surgiram genótipos adaptados às mais diversas condições ambientais (algumas extremamente desfavoráveis) do país. A importância (social, cultural e econômica) destes animais aliada a possibilidade de perda deste valioso material genético fizeram com que a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) incorporasse estas raças em seus planos de conservação de recursos genéticos o que possibilitou, dentre outras coisas, a criação de núcleos de conservação (MARIANTE *et al.*, 2009).

Kijas *et al.* (2012b) sugerem que a definição de raça transcende o ponto de vista genético, uma vez que fatores culturais, políticos, históricos e geográficos podem influenciar o agrupamento de indivíduos dentro das raças. Apesar disso, a definição de raças/tipos raciais é de fundamental importância para o desenvolvimento de programas eficientes de conservação. Alguns estudos demonstram haver subestruturação genética considerável dentro de raças ovinas (PAIVA *et al.*, 2005a; KIJAS *et al.*, 2009; GONÇALVES *et al.*, 2010; LI *et al.*, 2011; DUMASY *et al.*, 2012; KIJAS *et al.*, 2012b). Além disso, Kijas *et al.* (2012b) demonstraram que, baseado em dados moleculares, a raça Gulf Coast Native pode ser dividida em duas raças diferentes.

Existem duas situações importantes nas raças de ovinos brasileiras localmente adaptadas que sugerem a presença de uma subestruturação genética. A primeira é o caso da raça Crioula Lanada, que possui, pelo menos, dois ecótipos geneticamente e morfologicamente distintos (GONÇALVES *et al.*, 2010; SILVA *et al.*, 2013). O segundo caso é a Raça Santa Inês, que atualmente parece estar dividida em duas diferentes linhagens, a “Velha Santa Inês” e a “Nova Santa Inês” (PAIVA *et al.*, 2005a; CARNEIRO *et al.*, 2010; SOUZA *et al.*, 2012).

O processo de formação destes diferentes ecótipos/linhagens dentro das raças brasileiras localmente adaptadas de ovinos não parece ter sido influenciado apenas por efeito fundador, preferências de acasalamentos e distâncias geográficas. É sugerido que

a introgressão de genes oriundos de raças exóticas melhoradas também tenha contribuído com este processo (CARNEIRO *et al.*, 2010; MCMANUS *et al.*, 2010).

A identificação de subestruturação dentro de populações tem uma implicação importante nas estratégias de manutenção de variabilidade genética em programas de conservação (TORO; CABALLERO, 2005; FERNÁNDES *et al.*, 2008). Além disso, quando não identificada, a presença de subestruturação pode afetar a acurácia de estudos de associação entre marcadores moleculares e características de interesse econômico (QUIGNON *et al.*, 2007).

Ainda, o fato da definição de raças poder ser afetada por inúmeros fatores não genéticos, faz com que seja possível agrupar animais dentro de uma determinada raça que apresentem diferenças genéticas bastante consideráveis. Por causa disso, a caracterização genética é essencial para a identificação de grupos homogêneos para serem incluídos dentro dos esquemas de conservação (MARIANTE *et al.*, 2009; KIJAS *et al.*, 2012b).

Embora as principais raças localmente adaptadas de ovinos brasileiras já estejam incluídas em programas de conservação, algumas questões cruciais ainda devem ser respondidas para possibilitar o sucesso efetivo destes. Por exemplo, considerando que os recursos financeiros para a manutenção de programas de conservação são geralmente escassos, o conhecimento profundo sobre as populações alvo pode evitar esforços (e gastos) desnecessários e além disso, pode ser decisivo para a inclusão de populações adicionais da mesma raça mas que possuam características únicas (GANDINI; OLDENBROEK, 2007; MARIANTE *et al.*, 2009).

Portanto, o objetivo do presente trabalho foi identificar e caracterizar a presença de subestruturação genética nas três principais raças de ovinos localmente adaptadas brasileiras: Crioula Lanada, Morada Nova e Santa Inês e, a partir daí, propor estratégias para otimizar os programas de conservação destas raças.

MATERIAL E MÉTODOS

Animais e SNPs

Noventa e dois animais pertencentes às três principais raças de ovinos localmente adaptadas brasileiras (Crioula Lanada, Morada nova e Santa Inês) foram genotipados como parte do Consórcio Internacional do Genoma Ovino (KIJAS *et al.*, 2012a), utilizando o *SNP50 BeadChip* (Illumina®).

Destes, oitenta e sete indivíduos não aparentados (22 Crioula Lanada, 20 Morada Nova e 45 Santa Inês) foram utilizados para a realização deste trabalho. As informações referentes aos locais de coleta dos indivíduos estão descritas na Tabela 3.

Tabela 10. Informações sobre os indivíduos utilizados

Raça	N	Rebanho	Local
Crioula Lanada	5	BCS_CAS	Caçapava do Sul, RS
	6	BCS_EMB	Bagé, RS
	5	BCS_LAG	Lages, SC
	6	BCS_POA	Ponte Alta, SC
Morada Nova	13	BMN_EBD	Itapetinga, BA
	7	BMN_EMB	Sobral, CE
Santa Inês	6	BSI_EBD	Itapetinga, BA
	5	BSI_ECO	Sobral, CE
	13	BSI_EMN	Teresina, PI
	10	BSI_ETC	Aracaju, SE
	11	BSI_GAA	Inhumas, GO

O ovino *SNP50 BeadChip* (Illumina®) contém 54241 SNPs distribuídos por todo o genoma ovino. Após a genotipagem, os dados passaram por um pré-processamento, restando então 49034 SNPs com alta qualidade (KIJAS *et al.*, 2012a).

Os dados então foram submetidos a um processamento adicional que consistiu na manutenção apenas dos SNPs localizados nos cromossomos autossômicos, de acordo com a versão 2.0 do genoma ovino (ISGC *et al.*, 2010, disponível em: <http://www.livestockgenomics.csiro.au/sheep/oar2.0.php>), e com $MAF > 0,05$ (considerando os 87 indivíduos utilizados neste estudo). Ao final deste processamento, restaram 41296 SNPs que foram utilizados para a realização deste trabalho.

Diversidade genética

A proporção de *loci* polimórficos (PP), as heterozigozidades observada (H_O) e esperada (H_E) foram calculados com o software ARLEQUIN 3.5.1.3 (EXCOFFIER;

LISCHER, 2010). A riqueza alélica (A_R) e a riqueza alélica privada (P_A) foram calculadas utilizando o método de rarefação implementado no software ADZE 1.0 (SZPIECH *et al.*, 2008).

Análise baseadas nas subpopulações

As estatísticas F globais (F_{IT} , F_{ST} e F_{IS}) foram calculadas com o software GENEPOP 4.1.4 (ROUSSET, 2008). O F_{ST} par a par, as distâncias genéticas de Reynolds e a análise de variância molecular (AMOVA) foram calculados com o software ARLEQUIN 3.5.1.3 (EXCOFFIER; LISCHER, 2010) com 10^3 permutações para testar a significância.

Análises baseadas nos indivíduos

A análise de componentes principais (PCA) foi realizada utilizando o software SMARTPCA (PATTERSON *et al.*, 2006). Primeiramente a análise de PCA foi realizada utilizando o conjunto completo de SNPs (41296), então a análise foi realizada utilizando um subconjunto (de 5265 SNPs) obtido através do comando --indep-pairwise 20 5 0.05 do software PLINK (PURCELL *et al.*, 2007), para testar equivalência de resultados entre o conjunto completo e o subconjunto de SNPs. A significância foi testada utilizando o teste de Tracy-Widom (PATTERSON *et al.*, 2006).

Duas análises adicionais foram realizadas utilizando o subconjunto de 5265 SNPs, um método de clusterização hierárquica não paramétrico (GAO; STARMER, 2007) e um método de clusterização Bayesiana (PRITCHARD *et al.*, 2000).

Para a realização da análise de clusterização hierárquica não paramétrica, foi utilizado o software AWclust (GAO; STARMER, 2008). A inferência do número mais provável de subpopulações (K) foi feita utilizando a estatística gap (TIBSHIRANI *et al.*, 2001), obtida à partir de 100 simulações com K variando de um a dez.

A análise de clusterização Bayesiana foi realizada utilizando o software STRUCTURE (PRITCHARD *et al.*, 2000). Para cada valor de K , foram realizadas dez corridas independentes com 10^6 iterações e período de burn-in igual a 10^6 . As análises do STRUCTURE foram realizadas no cluster de computadores Bioportal da Universidade de Oslo (KUMAR *et al.*, 2009). O número mais provável de subpopulações (K) foi definido com base no método proposto por Evanno *et al.* (2005),

implementado no software STRUCTURE HARVESTER (EARL; VONHOLDT, 2012). Os resultados das corridas independentes do STRUCTURE foram alinhados utilizando o software CLUMPP 1.1.2 (JAKOBSSON; ROSENBERG, 2007) e os gráficos foram construídos utilizando o software DISTRUCT 1.1 (ROSENBERG, 2004).

RESULTADOS

Os valores globais de F_{IT} , F_{ST} e F_{IS} estimados foram 0,1321; 0,0848 e 0,0517, respectivamente. Os valores par a par de F_{ST} foram: 0,12609 (Crioula Lanada x Morada Nova); 0,08619 (Crioula Lanada x Santa Inês) e 0,06361 (Santa Inês x Morada Nova) e, os valores de F_{IS} específicos para cada raça foram: 0,1194 (Crioula Lanada); 0,0299 (Morada Nova) e 0,0260 (Santa Inês).

As medidas de diversidade genética para cada rebanho estão apresentadas na Tabela 11 e os valores de F_{ST} par a par e as distâncias genéticas de Reynolds estão representadas na Figura 6.

Tabela 11. Diversidade genética nas três raças de ovinos localmente adaptadas brasileiras.

População	PP	H_O^*	H_E^*	F_{IS}	A_R	P_A
BCS_CAS	0,87456	0,37301 ± 0,21144	0,42330 ± 0,12415	0,1317	1,87456	0,00005
BCS_EMB	0,89679	0,41658 ± 0,20580	0,41117 ± 0,12753	-0,0145	1,87337	0,00001
BCS_LAG	0,84882	0,42864 ± 0,21710	0,42210 ± 0,12541	-0,0175	1,84882	0,00001
BCS_POA	0,74613	0,40424 ± 0,20982	0,38670 ± 0,13690	-0,0501	1,71882	0,00001
BMN_EBD	0,92299	0,36734 ± 0,17745	0,36519 ± 0,14127	-0,0106	1,82087	0,00001
BMN_EMB	0,80105	0,37924 ± 0,19506	0,37845 ± 0,13816	-0,0023	1,75312	0,00001
BSI_EBD	0,83584	0,44101 ± 0,21016	0,39561 ± 0,13435	-0,1277	1,80813	0,00001
BSI_ECO	0,80695	0,43228 ± 0,21579	0,40999 ± 0,12821	-0,0616	1,80695	0,00000
BSI_EMN	0,94193	0,37518 ± 0,17407	0,37598 ± 0,13664	0,0022	1,84990	0,00001
BSI_ETC	0,94360	0,38730 ± 0,17979	0,38742 ± 0,13259	0,0003	1,87196	0,00001
BSI_GAA	0,95196	0,38779 ± 0,18003	0,38291 ± 0,13402	-0,0134	1,87190	0,00000

*média ± desvio padrão.

Os valores de F_{ST} par a par dentro da raça Crioula Lanada variaram entre 0,08659 (BCS_EMB x BCS_CAS) e 0,19024 (BCS_POA x BCS_CAS). Para a raça Morada Nova, o valor estimado foi de 0,06793 e para a raça Santa Inês, os valores estimados variaram entre 0,01809 (BSI_ETC x BSI_GAA) e 0,13092 (BSI_ECO x BSI_EBD).

Na maioria dos rebanhos, o valor estimado de F_{IS} tendia a zero. Porém, o rebanho BCS_CAS apresentou um valor de F_{IS} grande e positivo (0,1317) enquanto o rebanho BSI_EBD apresentou um valor grande e negativo (-0,1277).

Deve ser ressaltado que em duas situações, os valores de F_{ST} par a par entre rebanhos da mesma raça foram maiores do que a comparação entre raças. O rebanho BCS_POA apresentou uma grande diferenciação genética quando comparado com os outros rebanhos da raça Crioula Lanada. E os rebanhos BSI_ECO e BSI_EBD também apresentaram um valor bastante alto de F_{ST} entre si.

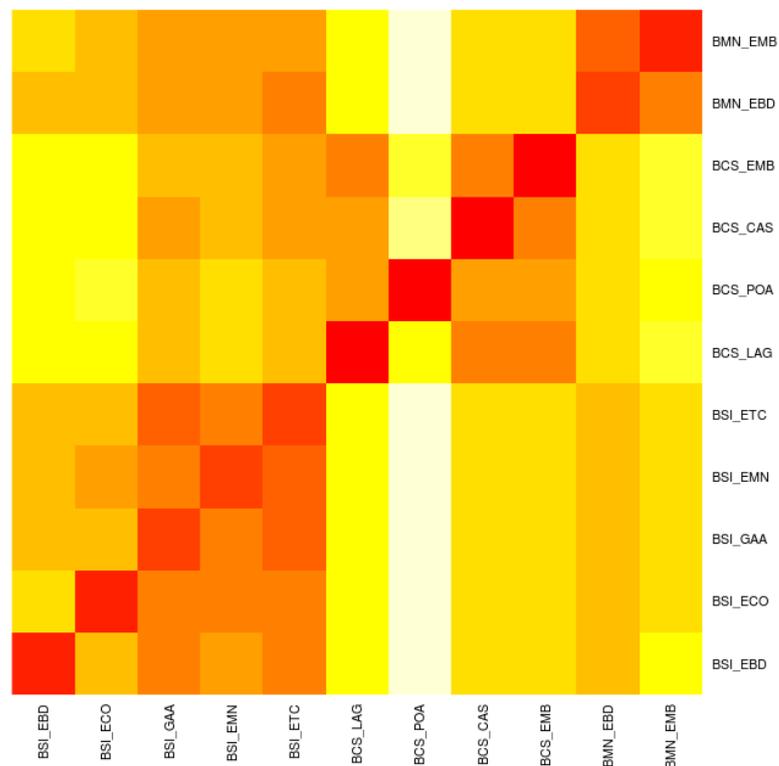


Figura 6. Heatmap ilustrando os valores de F_{ST} par a par (diagonal superior) e as distâncias genéticas de Reynolds (diagonal inferior) entre cada rebanho. Cores tendendo ao vermelho indicam valores mais próximo de zero enquanto as cores tendendo ao branco indicam valores mais distantes de zero.

Para melhor visualizar as diferenças entre rebanhos, uma análise de coordenadas principais foi realizada à partir das distâncias genéticas de Reynolds. O primeiro eixo separa claramente as raças deslanadas (Morada Nova e Santa Inês) da raça Crioula Lanada e o segundo eixo permite a separação entre as duas raças deslanadas (Figura 7).

A combinação entre os três eixos principais permite visualizar uma questão interessante em relação às diferenças entre rebanhos da mesma raça. Para as raças

Crioula Lanada e Morada Nova, existe uma clara separação entre os rebanhos dentro de cada raça. É interessante ressaltar que o rebanho BCS_POA parece se separar de forma considerável dos outros rebanhos da raça Crioula Lanada. Para a raça Santa Inês, a análise não permite separar os rebanhos BSI_ETC e BSI_GAA, mas fica evidente a separação entre os outros rebanhos (Figura 7).

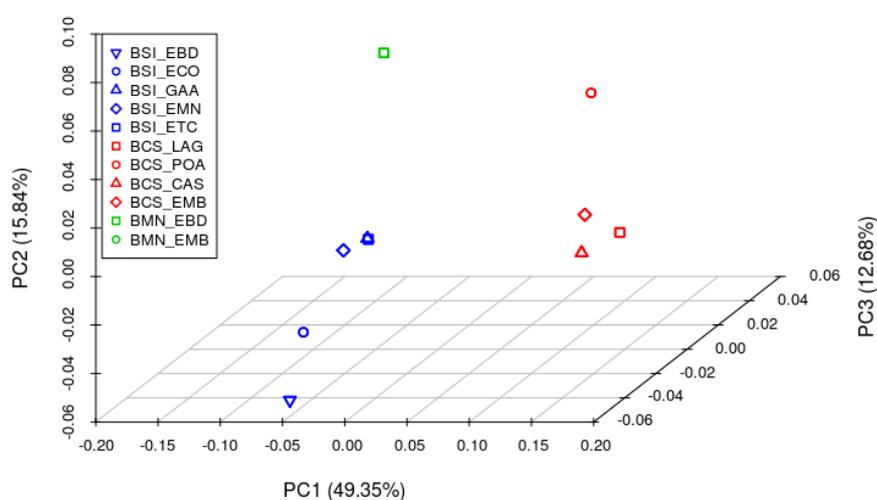


Figura 7. Análise de coordenadas principais à partir das distâncias genéticas de Reynolds. A raça Crioula Lanada está representada pela cor vermelha, Morada Nova pela cor verde e Santa Inês pela cor azul.

A AMOVA revelou que o componente dentro de rebanhos explica a maior parte da variação genética, todavia, os outros dois componentes também explicam uma porcentagem significativa da variância genética total (Tabela 12).

Tabela 12. Análise de variância molecular (AMOVA).

FV¹	SQ²	DF³	CV⁴	% Var⁵	P-valor
Entre raças	91246,096	2	510,65301	6,17151	< 0,00001
Entre rebanhos / dentro de raças	134655,101	8	641,62801	7,75441	< 0,00001
Dentro de rebanhos	1160898,889	163	7122,07907	86,07408	< 0,00001
Total	1386800,086	173	8274,36009	-	-

¹ Fonte de variação

² Soma de quadrados

³ Graus de liberdade

⁴ Componentes de variância

⁵ Proporção da variância total

As análises de componentes principais realizadas tanto com o conjunto completo de SNPs quanto com o subconjunto apresentaram resultados concordantes (Tabela 13 e Tabela 14), portanto apenas serão apresentados e discutidos os resultados relativos à análise considerando o subconjunto de SNPs.

Tabela 13. Resultados do teste de Tracy-Widom para análise de componentes principais realizada considerando as três raças brasileiras de ovinos localmente adaptadas.

41296 SNPs			5265 SNPs		
Componente	Eigenvalor	P-valor	Componente	Eigenvalor	P-valor
1	6,81767	<0,00001*	1	5,66585	<0,00001*
2	3,95422	<0,00001*	2	4,00660	<0,00001*
3	2,97015	<0,00001*	3	2,70350	<0,00001*
4	2,37862	<0,00001*	4	2,48644	<0,00001*
5	2,11805	<0,00001*	5	2,21236	<0,00001*
6	1,99118	<0,00001*	6	2,01590	<0,00001*
7	1,84915	<0,00001*	7	1,91153	<0,00001*
8	1,84341	<0,00001*	8	1,82564	<0,00001*
9	1,78329	<0,00001*	9	1,77035	<0,00001*
10	1,71120	<0,00001*	10	1,69713	<0,00001*
11	1,37260	<0,00001*	11	1,45147	<0,00001*
12	1,33212	<0,00001*	12	1,29455	0,158132 ^{NS}
13	1,23330	0,12255 ^{NS}	13	1,22025	1,00000 ^{NS}

* Significativo

^{NS} Não significativo

Tabela 14. Resultados do teste de Tracy-Widom para análise de componentes principais realizada para cada raça separadamente.

	41296 SNPs			5265 SNPs	
	Componente	Eigenvalor	P-valor	Eigenvalor	P-valor
Crioula Lanada	1	2,91987	0,00008*	2,98605	0,00006*
	2	1,87462	0,01510*	1,84637	0,02452*
	3	1,72718	0,00054*	1,72757	0,00076*
	4	1,28722	0,12650 ^{NS}	1,31909	0,06408 ^{NS}
Morada Nova	1	2,31262	0,01516*	2,32295	0,03226*
	2	2,10199	0,00019*	2,21088	0,00013*
	3	1,38443	0,21578 ^{NS}	1,34654	1,00000 ^{NS}
	4	1,19100	1,00000 ^{NS}	1,17927	1,00000 ^{NS}
Santa Inês	1	2,65911	<0,00001*	2,71005	<0,00001*
	2	2,36473	<0,00001*	2,37679	<0,00001*
	3	2,00504	<0,00001*	2,09213	<0,00001*
	4	1,55329	0,00109*	1,59745	0,00017*

* Significativo

^{NS} Não significativo

De acordo com Paterson *et al.* (2006), considerando que o número de marcadores é suficientemente grande, o número de componentes significativos

identificados pelo teste de Tracy-Widom mais um é igual ao número mais provável de populações distintas na amostra.

A análise realizada com as três raças brasileiras localmente adaptadas de ovinos revelou a presença de 11 componentes significativos (Tabela 13). A análise dos três principais componentes permite a separação das três raças e além disso permite a identificação de uma aparente subestruturação tanto para a raça Crioula Lanada quanto para a raça Santa Inês (Figura 8).

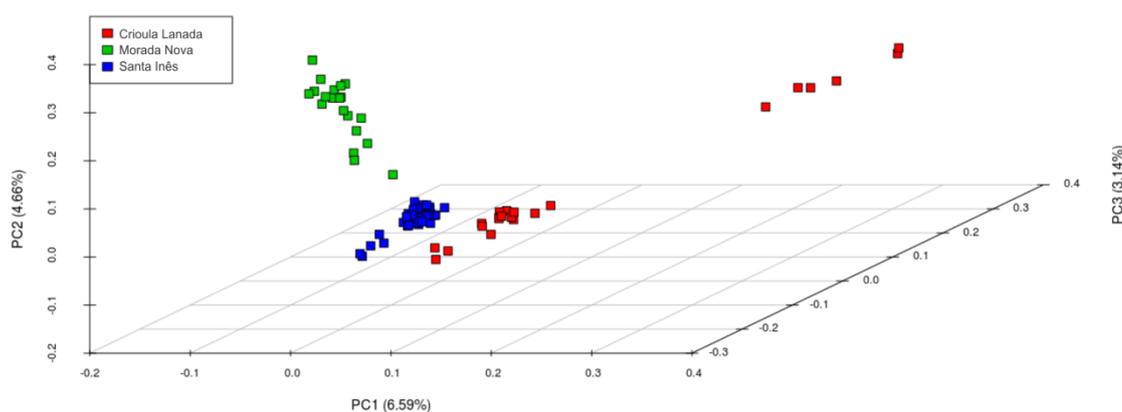


Figura 8. Análise de componentes principais realizada considerando as três raças brasileiras de ovinos localmente adaptadas.

Para identificar uma possível subestruturação dentro das raças, a PCA também foi realizada separadamente para cada raça e revelou a presença de três componentes significativos para a raça Crioula Lanada, dois para Morada Nova e quatro para Santa Inês (Tabela 14).

Para a raça Crioula Lanada, o primeiro eixo permite a separação do rebanho BCS_POA dos outros três rebanhos estudados da raça. A análise dos três principais componentes permite a identificação de quatro *clusters*. É interessante ressaltar que há um aparente agrupamento de indivíduos BCS_CAS no *cluster* referente ao rebanho BCS_LAG (Figura 9).

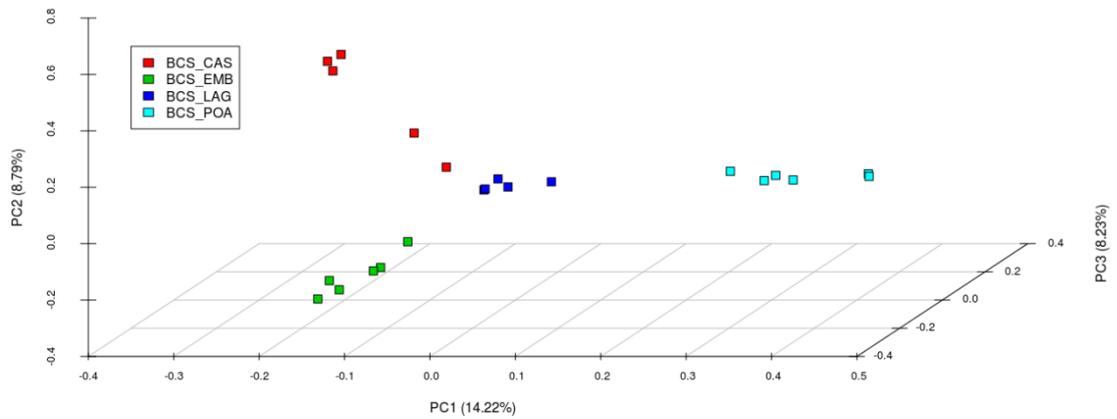


Figura 9. Análise de componentes principais para a raça Crioula Lanada.

A observação dos componentes significativos (Tabela 14) permite a separação dos indivíduos da raça Morada nova em três *clusters* (Figura 10). Aparentemente, o rebanho BMN_EBD pode ser separado em três *clusters* e o rebanho BMN_EMB agrupa com um dos *clusters* do rebanho BMN_EBD.

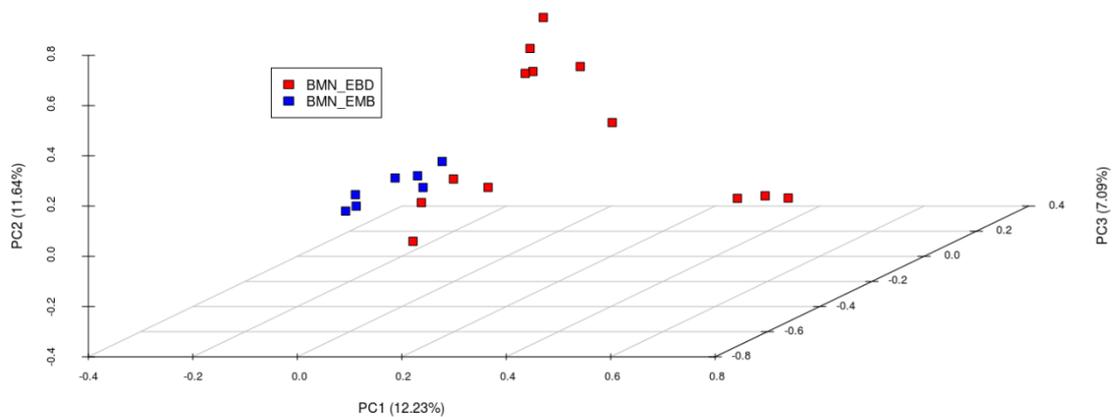


Figura 10. Análise de componentes principais para a raça Morada Nova.

Para a raça Santa Inês, a análise de PCA sugere a presença de cinco *clusters* distintos (Tabela 14). Os três primeiros *clusters* parecem ser formados apenas por indivíduos pertencentes a um único rebanho (BSI_EBD, BSI_GAA ou BSI_EMN). Um *cluster* parece ser formado por indivíduos do rebanho BSI_ECO e um indivíduo pertencente ao rebanho BSI_EMN e, o outro *cluster* parece ser formado por indivíduos dos rebanhos BSI_ETC, BSI_GAA, BSI_ECO e BSI_EBD (Figura 11).

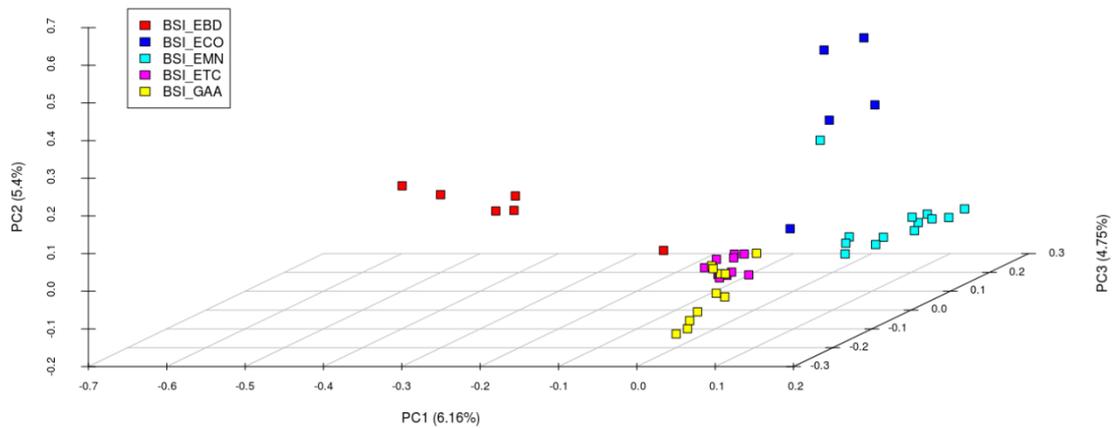


Figura 11. Análise de componentes principais para a raça Santa Inês.

A análise de clusterização hierárquica não paramétrica foi realizada tanto com as três raças conjuntamente quanto para cada raça separadamente. Gao & Starmer (2007) sugerem que o número mais provável de clusters (K) pode ser definido com base na identificação do pico que representa a maximização da distância entre as curvas observadas e esperadas de $\log(W_k)$.

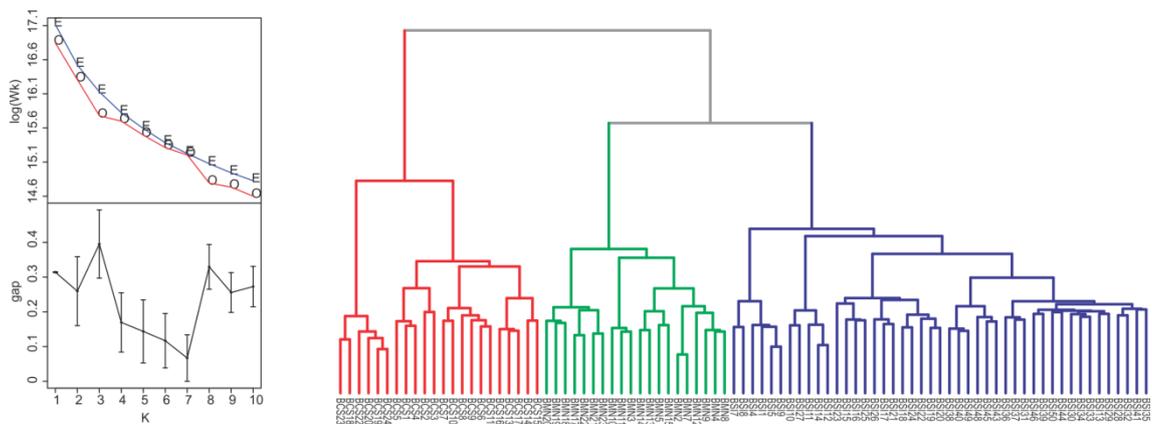


Figura 12. Análise de clusterização hierárquica não paramétrica para as três raças brasileiras localmente adaptadas de ovinos. Cada cor representa uma raça: Crioula Lanada (vermelho), Morada Nova (verde) e Santa Inês (azul).

Para as três raças brasileiras em conjunto, a análise de clusterização hierárquica não paramétrica possibilitou a identificação de três *clusters* distintos representando cada uma das raças analisadas (Figura 12).

Para a raça Crioula Lanada, foram identificados quatro *clusters* distintos com base na estatística gap (Figura 13). Pode-se observar que todos os indivíduos puderam ser agrupados de acordo com seu rebanho de origem. Além disso, pode-se observar

também que o rebanho BCS_POA é mais distante dos outros três rebanhos da raça Crioula Lanada estudados.

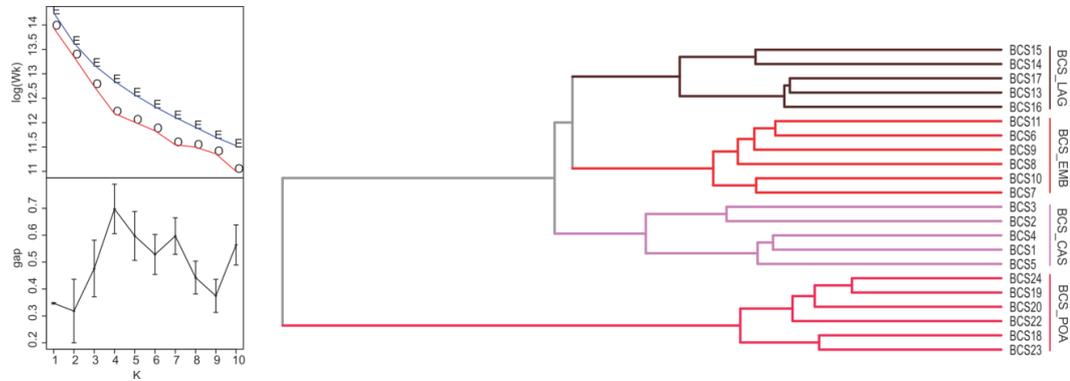


Figura 13. Análise de clusterização hierárquica não paramétrica para a raça Crioula Lanada.

Para a raça Morada Nova, a curva da estatística gap apresentou uma tendência de crescimento para K variando de um a dez. Por causa disso, a análise de clusterização hierárquica não paramétrica foi realizada novamente com K variando de um a 16 (o número máximo de *clusters* permitido no software AWclust). Ainda assim, a definição do número mais provável de *clusters* não foi possível, uma vez que a curva da estatística gap continuou apresentando tendência de aumento (Figura 14).

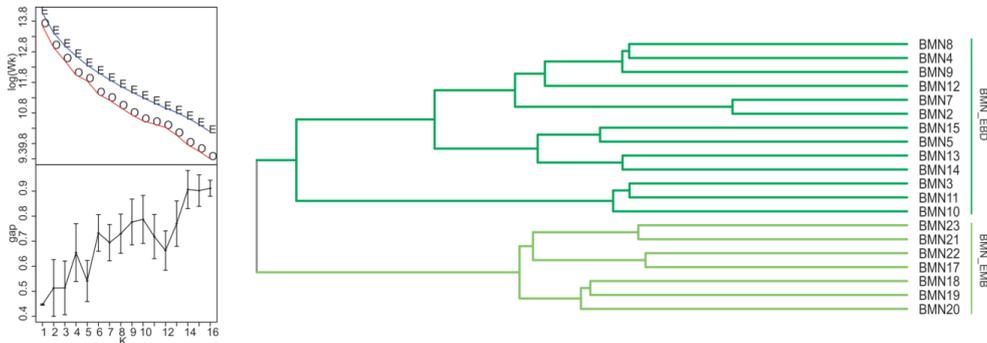


Figura 14. Análise de clusterização hierárquica não paramétrica para a raça Morada Nova.

Para a raça Santa Inês, a análise da curva da estatística gap sugeriu que o número mais provável de *clusters* fosse igual a um. Porém, Tibshirani *et al.* (2001) sugerem que a observação completa da curva da estatística gap pode revelar padrões adicionais de clusterização. Portanto, a análise da curva da estatística gap indica um pico adicional em K=5, o que sugere que podem haver cinco subpopulações. É interessante notar que apenas para o rebanho BSI_EBD, os indivíduos agruparam corretamente, para todos os

outros rebanhos pode-se observar a presença de indivíduos agrupados de forma incorreta (Figura 15).

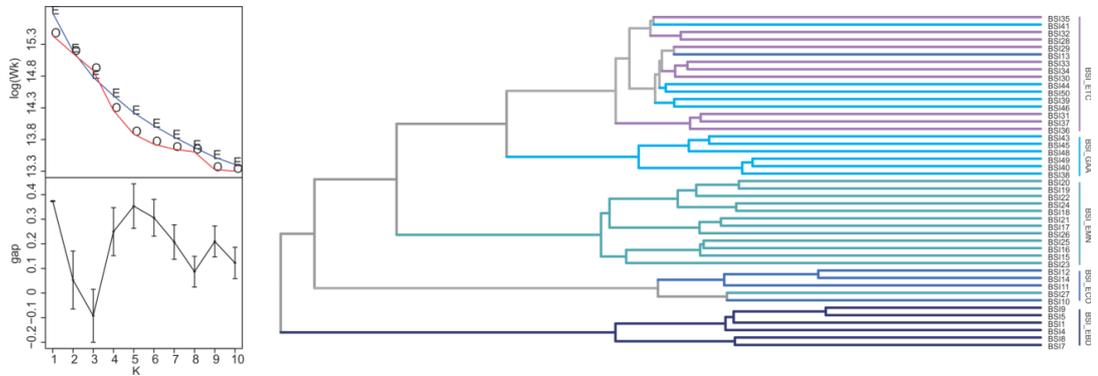


Figura 15. Análise de clusterização hierárquica não paramétrica para a raça Santa Inês.

A análise de clusterização bayesiana considerando as raças localmente adaptadas de ovinos brasileiras conjuntamente, sugere a presença de dois *clusters* (Figura 16A). Assumindo $K=2$ é possível identificar claramente a separação entre as raças deslanadas nordestinas (Morada Nova e Santa Inês) e a raça Crioula Lanada (Figura 17).

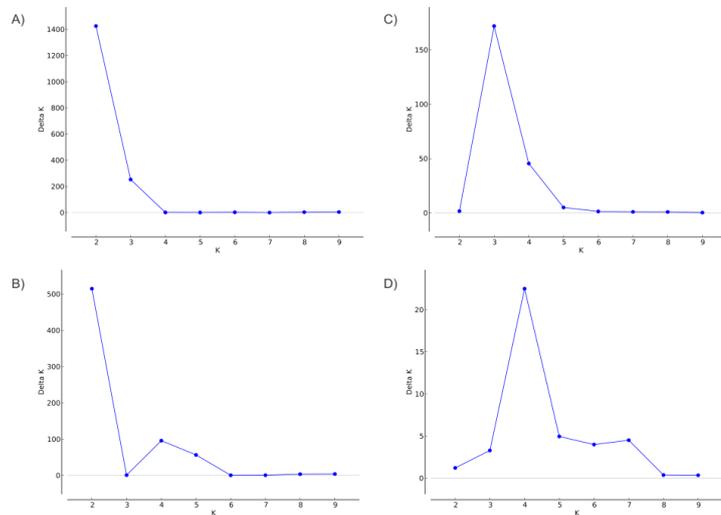


Figura 16. Estatística Delta K para a análise de clusterização Bayesiana. A) Raças Brasileiras , B) Crioula Lanada, C) Morada Nova e D) Santa Inês.

O método proposto por Evanno *et al.* (2005) geralmente identifica o nível mais alto de subestruturação, o que no presente caso representa a diferenciação entre as raças deslanadas e a raça lanada estudadas. Observando $K=3$, é possível identificar

claramente a separação entre as raças deslanadas e observando $K > 3$ pode-se sugerir níveis adicionais de estruturação dentro de raças (Figura 17).

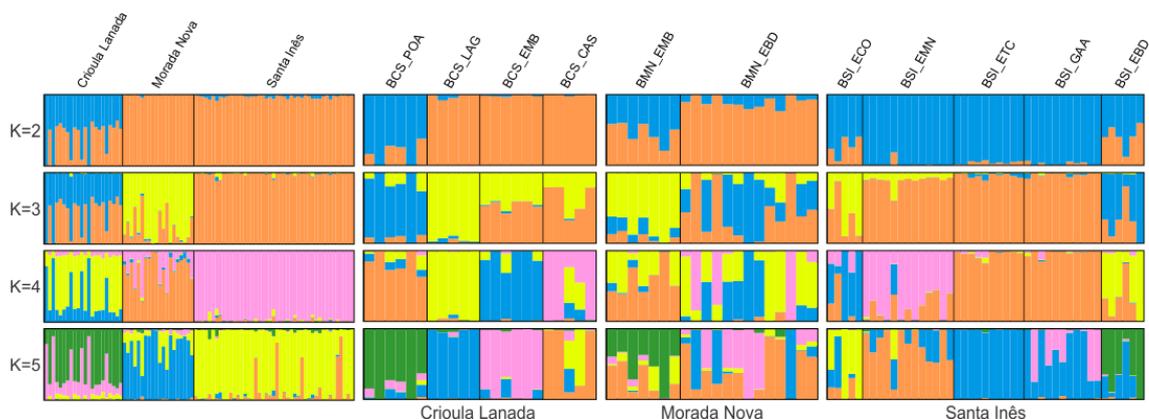


Figura 17. Resultados da análise de clusterização Bayesiana para K variando de dois a cinco considerando, tanto as raças conjuntamente, como cada raça separadamente.

Para a raça Crioula Lanada, o número mais provável de *clusters* foi estimado em dois (Figura 16B) e pode ser observado que é clara a diferenciação entre o rebanho BCS_POA e os outros rebanhos da raça Crioula Lanada (Figura 17). A análise do Delta K (Figura 16B) permite a identificação de um pico adicional em $K=4$ e, portanto, é possível observar a separação entre os quatro rebanhos estudados da raça.

Para a raça Morada Nova, o número de *clusters* estimado foi três (Figura 16C) e a observação da Figura 17 permite a observação de, pelo menos, dois grupos distintos dentro do rebanho BCS_EBD. Já na raça Santa Inês, o número mais provável de clusters estimados foi igual a quatro (Figura 16D). Pode-se observar uma clara homogeneidade entre os indivíduos BSI_ETC e BSI_GAA. Observando $K=5$ pode-se sugerir uma influência do rebanho BSI_ETC nos outros rebanhos estudados da raça Santa Inês (Figura 17).

DISCUSSÃO

O presente estudo evidenciou que o nível de diferenciação genética entre as três raças estudadas é moderado (HARTL; CLARK, 2010), o que é consistente com estudos que analisaram raças ovinas pertencentes ao mesmo país ou continente (KIJAS *et al.*, 2009; KIJAS *et al.*, 2012b), e também com os registros históricos que sugerem que as

três raças estudadas são descendentes de animais trazidos da Península Ibérica durante o período de colonização (MARIANTE *et al.*, 2009; MCMANUS *et al.*, 2010). Além disso todas as três raças pertencem ao mesmo haplótipo mitocondrial de raças europeias (PAIVA *et al.*, 2005b; MCMANUS *et al.*, 2010; LOBO *et al.*, 2013).

A separação entre as raças deslanadas e a raça lanada pode ser proposta como o nível mais alto de estruturação entre as raças analisadas, o que ressalta o fato de as raças deslanadas do Nordeste (Morada Nova e Santa Inês) possuírem um menor tempo de divergência quando comparadas com a raça Crioula Lanada (KIJAS *et al.*, 2012a) e também a participação da raça Morada Nova no processo de formação da Santa Inês (PAIVA *et al.*, 2005a).

Felizmente, os níveis de diversidade genética e o tamanho efetivo destas raças é bastante satisfatório (KIJAS *et al.*, 2012a). Além disso, os níveis de diversidade genética dentro de cada rebanho são bastante satisfatórios quando comparados com outros estudos envolvendo a espécie ovina (KIJAS *et al.*, 2009; KIJAS *et al.*, 2012a; KIJAS *et al.*, 2012b). também, as três raças estudadas já estão incluídas em programas de conservação e, pelo menos duas delas (Morada nova e Santa Inês) estão atualmente inseridas em programas de melhoramento ou pré-melhoramento, o que faz com que o risco de extinção destas raças possa ser considerado bastante baixo.

Apesar disso, algumas questões cruciais ainda devem ser respondidas para possibilitar a otimização destes programas, principalmente no que diz respeito à manutenção do número de animais e rebanhos ideal tanto do ponto de vista de manutenção da variabilidade genética quanto do ponto de vista de custo de manutenção destes rebanhos.

Para cada raça analisada, os valores de F_{ST} par a par mostram que, em geral, existe uma diferenciação genética considerável entre rebanhos da mesma raça. A única exceção clara é a comparação entre BSI_ETC e BSI_GAA. Os níveis de diferenciação genética encontrados são maiores do que os valores encontrados para comparação entre animais da mesma raça que apresentam cor de pelagem específicos ou que foram selecionados para objetivos específicos (leite ou carne) ou ainda mantidos em países diferentes (KIJAS *et al.*, 2009; LI *et al.*, 2011; KIJAS *et al.*, 2012b).

Os resultados da AMOVA ressaltam a importância do componente dentro de rebanhos, o que corrobora outros estudos envolvendo ovinos de raças brasileiras

(PAIVA *et al.*, 2005a) e exóticas (LAWSON HANDLEY *et al.*, 2005; DUMASY *et al.*, 2012). Deve-se observar também que o fato de o componente entre rebanhos/dentro de raças ser significativo e maior do que o componente entre raças, pode sugerir a presença de subestruturação dentro de raça.

Tomadas em conjunto, a análise de componentes principais, a análise de clusterização hierárquica não paramétrica e a análise de clusterização Bayesiana possibilitam a observação de questões adicionais envolvendo as populações analisadas.

Para a raça Crioula Lanada observou-se uma clara diferenciação entre o rebanho BCS_POA e os demais rebanhos da raça. Deve-se ressaltar que o rebanho BCS_EMB é pertencente ao núcleo de conservação da Embrapa enquanto os demais rebanhos (BCS_POA, BCS_LAG e BCS_CAS) pertencem a criadores particulares.

Este fato nos leva a acreditar, de fato, na existência de dois ecótipos geneticamente distintos (GONÇALVES *et al.*, 2010; SILVA *et al.*, 2013) e que, embora não haja diferenciação entre ecótipos para fins de registro genealógico, e os criadores da raça realizem intercâmbio de reprodutores pertencentes a diferentes ecótipos, existe uma subestruturação bastante importante nesta raça.

Para a raça Crioula Lanada há a sugestão de que existam, pelo menos, dois ecótipos geneticamente distintos, denominados de Serrana e Fronteira (GONÇALVES *et al.*, 2010; SILVA *et al.*, 2013). O fato de o rebanho BCS_POA ter se mostrado geneticamente muito diferente dos outros três rebanhos da raça Crioula Lanada indica que a subestruturação identificada no presente estudo possivelmente é um reflexo de diferenças genéticas entre os ecótipos, que podem estar relacionadas a fatores ambientais, preferência de criadores, isolamento geográfico e eventos de introgressão de genes de raças exóticas (GONÇALVES *et al.*, 2010; SILVA *et al.*, 2013).

Do ponto de vista prático pode-se sugerir que o programa de conservação da raça Crioula Lanada inclua animais pertencentes aos dois ecótipos e que seja utilizada esta informação no direcionamento dos acasalamentos. Além disso, deve-se iniciar uma discussão junto às associações de criadores a fim de se definir critérios no registro genealógico dos animais para identificação de qual ecótipo os animais pertencem.

Também, devem ser realizados estudos a fim de se verificar a ocorrência (ou não) de diferenças produtivas e adaptativas entre os ecótipos de modo a otimizar a utilização destes animais dentro dos sistemas de produção.

Para a raça Santa Inês existe a hipótese da presença de duas linhagens, a “velha Santa Inês” e a “nova Santa Inês”. A primeira compreende animais menores e mais rústicos, já a segunda, foi originada provavelmente no estado de Sergipe e compreende animais maiores e com melhor musculatura na região posterior (PAIVA *et al.*, 2005a; CARNEIRO *et al.*, 2010; SOUZA *et al.*, 2012).

A “nova Santa Inês” se espalhou por todo o país e, atualmente, praticamente todos os criadores da raça utilizam reprodutores “melhorados”. Porém, existe a hipótese de que para a criação desta nova linhagem tenham ocorrido eventos de introgressão com raças exóticas (CARNEIRO *et al.*, 2010; MCMANUS *et al.*, 2010). Este fato parece estar relacionado com o aumento das exigências nutricionais destes animais e com a diminuição de adaptabilidade em relação ao *stress* térmico e resistência contra parasitas gastrintestinais (CARNEIRO *et al.*, 2010; MCMANUS *et al.*, 2010).

No presente estudo, observou-se para a raça Santa Inês, uma clara diferenciação entre quatro dos cinco rebanhos analisados, sendo que os rebanhos BSI_ETC e BSI_GAA praticamente não apresentaram diferenciação genética entre si. Vale ressaltar que os rebanhos BSI_ECO, BSI_EMN, BSI_ETC e BSI_EBD pertencem ao núcleo de conservação da raça Santa Inês e o rebanho BSI_GAA pertence a um criador particular.

Estes resultados nos levam a acreditar que, de fato, haja a existência de duas linhagens dentro da raça Santa Inês. Possivelmente os rebanhos BSI_ETC e BSI_GAA representam a linhagem “nova Santa Inês”, enquanto os demais rebanhos representam a linhagem “velha Santa Inês”. Porém, observa-se que todos os rebanhos pertencentes ao núcleo de conservação apresentam algum grau de *admixture* com a “nova Santa Inês”.

Com base nestes resultados, pode-se sugerir que o programa de conservação da Raça Santa Inês deve manter, pelo menos, animais pertencentes aos rebanhos BSI_ECO, BSI_EMN e BSI_EBD, uma vez que estes são geneticamente distintos do rebanho BSI_GAA, que pode ser considerado como representativo do rebanho comercial.

Os resultados do presente estudo demonstram que o aprofundamento do conhecimento sobre as populações alvo pode evitar esforços (e gastos) desnecessários e além disso, pode ser decisivo para a inclusão de populações adicionais da mesma raça mas que possuam características únicas (GANDINI; OLDENBROEK, 2007; MARIANTE *et al.*, 2009).

CONCLUSÕES

Existe subestruturação genética significativa nas raças ovinas localmente adaptadas brasileiras. A hipótese de que existem diferentes ecótipos/linhagens dentro das raças Crioula Lanada e Santa Inês pode ser sugerida e portanto, devem ser tomadas medidas para evitar a perda de material genético importante no contexto dos programas de conservação destas raças.

REFERÊNCIAS BIBLIOGRÁFICAS

- CARNEIRO, H.; LOUVADINI, H.; PAIVA, S.R.; MACEDO, F.; MERNIES, B.; MCMANUS, C. Morphological characterization of sheep breeds in Brazil, Uruguay and Colombia. **Small Rum. Res.**, v.94, p.58-65, 2010.
- DUMASY, J-F.; DANIAUX, C.; DONNAY, I.; BARET, P.V. Genetic diversity and networks of exchange: a combined approach to assess intra-breed diversity. **Gen. Sel. Evol.**, 44:17, 2012.
- EARL, D.A.; VONHOLDT, B.M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. **Cons. Gen. Res.**, v.4, n.2, p.359-361, 2012.
- EVANNO, G.; REGNAUT, S.; GOUDET J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Mol. Ecol.**, v.14, n.8, p.2247-2620, 2005.
- EXCOFFIER, L.; LISCHER, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Mol. Ecol. Res.**, v.10, p.564-567, 2010.
- FERNÁNDEZ, J.; TORO, M.A.; CABALLERO, A. Management of Subdivided Populations in Conservation Programs: Development of a Novel Dynamic System. **Genetics**, v.179, p.683-692, 2008.
- GANDINI, G.; OLDENBROEK, K. Strategies for moving from conservation to utilisation. In: Utilisation and Conservation of Farm Animal Genetic Resources (ed. By K. OLDENBROEK), pp. 29-54. Wageningen Academic Publishers, Wageningen, 2007.

- GAO, X.; STARMER, J. AWclust: point-and-click software for non-parametric population structure analysis. **BMC Bioinf.**, 9:77, 2008.
- GAO, X.; STARMER, J. Human population structure detection via multilocus genotype clustering. **BMC Genet.**, 8:34, 2007.
- GONÇALVES, G.L.; MOREIRA, G.R.P.; FREITAS, T.R.O.; HEPP, D.; PASSOS, D.T.; WEIMER, T.A. Mitochondrial and nuclear DNA analyses reveal population differentiation in Brazilian Creole sheep. **Anim. Genet.**, v.41, n.3, p.308:310, 2010.
- HARTL, D.L.; CLARK, A.G. Princípios de Genética de Populações. 4ª edição. Artmed, Porto Alegre, 660pp, 2010.
- JAKOBSSON, M.; ROSENBERG, N.A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. **Bioinformatics**, v.23, n.14, p.1801-1806, 2007.
- KIJAS, J.W.; LENSTRA, J.A.; HAYES, B.; BOITARD, S.; PORTO NETO, L.R.; CRISTOBAL, M.S.; SERVIN, B.; MCCULLOCH, R.; WHAN, V.; GIETZEN, K.; *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. **PLoS Biol.**, 10(2):e1001258, 2012.
- KIJAS, J.W.; MILLER, J.E.; HADFIELD, T.; MCCULLOCH, R.; GARCIA-GAMEZ, E.; *et al.* Tracking the Emergence of a New Breed Using 49,034 SNP in Sheep. **PLoS ONE**, 7(7): e41508, 2012b.
- KIJAS, J.W.; TOWNLEY, D.; DALRYMPLE, B.P.; HEATON, M.P.; MADDOX, J.F.; *et al.* A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. **PLoS ONE**, 4(3):e4668, 2009.
- KUMAR, S.; SKJÆVELAND, Å.; ORR, R.J.S.; ENGER, P.; RUDEN, T.; MEVIK, B.-H.; BURKI, F.; BOTNEN, A.; SHALCHIAN-TABRIZI, K. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. **BMC Bioinf.**, 10:357, 2009.
- LAWSON HANDLEY, L-J.; BYRNE, K.; SANTUCCI, F.; TOWNSEND, S.; TAYLOR, M.; *et al.* Genetic structure of European sheep breeds. **Heredity**, v.99, p.620-631, 2007.
- LI, M-H.; STRANDEN, I.; TIIRIKKA, T.; SEVON-AIMONEN, M-L.; KANTANEN, J. A comparison of approaches to estimate the inbreeding coefficient and pairwise

relatedness using genomic and pedigree data in a sheep population. **PLoS ONE**, 6(11):e26256, 2011.

LOBO, F.P.; YAMAGISHI, M.E.B.; CAETANO, A.R.C.; MCMANUS, C.M.; CARNEIRO, P.L.; FACÓ, O.; SOUZA, C.J.H.; PAIVA, S.R. Genetic origin of brazilian local adapted sheep (*Ovis aries*) breeds by complete mitochondrial genome data analysis. *In: PLANT AND ANIMAL GENOME CONFERENCE, XXI, 2013, San Diego*. Disponível em: <<http://www.intlpag.org/2013/>>. Acesso em: 24 jan. 2013.

MARIANTE, A.S.; ALBUQUERQUE, M.S.M.; EGITO, A.A.; MCMANUS, C.; LOPES, M.A.; PAIVA, S.R. Present status of the conservation of livestock genetic resources in Brazil. **Livestock Sci.**, v.120, n.3, p.204-212, 2009.

MCMANUS, C.; PAIVA, S.R.; ARAÚJO, R.O. Genetics and breeding of sheep in Brazil. **R. Bras. Zootec.**, v.39, p.236-246, 2010.

PAIVA, S.R.; SILVÉRIO, V.C.; EGITO, A.A.; MCMANUS, C.; FARIA, D.A.; *et al.* Genetic variability of the Brazilian hair sheep breeds. **Pesq. Agropec. Bras.**, v.40, n.9, p.887-893, 2005a.

PAIVA, S.R.; SILVÉRIO, V.C.; PAIVA, D.A.F.; MCMANUS, C.; EGITO, A.A.; *et al.* Origin of the main locally adapted sheep breeds of Brazil: a PCR-RFLP molecular analysis. **Arch. Zootec.**, v.54, p.395-399, 2005b.

PATTERSON, N.; PRICE, A.L.; REICH, D. Population structure and eigenanalysis. **PLoS Genet.**, 2(12): e190, 2006.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of Population Structure Using Multilocus Genotype Data. **Genetics**, v.155, p.945-959, 2000.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P.I.W.; DALY, M.J.; SHAM, P.C. PLINK: a toolset for whole-genome association and population-based linkage analysis. **Am. J. Hum. Gen.**, 81, 2007.

QUIGNON, P.; HERBIN, L.; CADIEU, E.; KIRKNESS, E.F.; HÉDAN, B.; *et al.* Canine Population Structure: Assessment and Impact of Intra-Breed Stratification on SNP-Based Association Studies. **PLoS ONE**, 2(12): e1324, 2007.

ROSENBERG, N.A. DISTRUCT: a program for the graphical display of population structure. **Molec. Ecol. Notes**, v.4, p.137-138, 2004.

- ROUSSET, F. GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. **Molec. Ecol. Res.**, v.8, n.1, p.103-106, 2008.
- SILVA, M.C.; LOPES, F.B.; VAZ, C.M.S.; PAULINI, F.; MONTESINOSA, I.S.; FIORAVANTIA, M.C.S.; MCMANUS, C.; SERENO, J.R.B. Morphometric traits in Crioula Lanada ewes in Southern Brazil. **Small Rum. Res.**, v.110, p.15-19, 2013.
- SOUZA, C.A.; PAIVA, S.R.; MCMANUS, C.M.; AZEVEDO, H.C.; MARIANTE, A.S.; GRATTAPAGLIA D. Genetic diversity and assessment of 23 microsatellite markers for parentage testing of Santa Inês hair sheep in Brazil. **Gen. Mol. Res.**, v.11, n.2, p.1217-1229, 2012.
- SZPIECH, Z.A.; JAKOBSSON, M.; ROSENBERG, N.A. ADZE: a rarefaction approach for counting alleles private to combinations of populations. **Bioinformatics**, v.24, n.21, p.2498-2504, 2008.
- THE INTERNATIONAL SHEEP GENOMICS CONSORTIUM (ISGC), ARCHIBALD, A.L.; COCKETT, N.E.; DALRYMPLE, B.P.; FARAUT, T.; KIJAS, J.W.; MADDOX, J.F.; MCEWAN, J.C.; HUTTON ODDY, V.; RAADSMA, H.W.; WADE, C.; WANG, J.; WANG, W.; XUN, X. The sheep genome reference sequence: a work in progress. **Anim. Genet.**, n.41, p.449–453, 2010.
- TIBSHIRANI, R.; WALTHER, G.; HASTLE T. Estimating the number of clusters in a dataset via the gap statistic. **J. R. Stat. Soc. B.**, v.63, n.2, p.411-423, 2001.
- TORO, M.A.; CABALLERO, A. Characterization and conservation of genetic diversity in subdivided populations. **Phil. Trans. R. Soc. B.**, v.360, n.1459, p.1367-1378, 2005.

CONSIDERAÇÕES FINAIS E IMPLICAÇÕES

As raças brasileiras localmente adaptadas de ovinos certamente possuem uma enorme importância social e cultural. Além disso, estes indivíduos possuem características adaptativas importantíssimas, principalmente para pequenos produtores localizados em regiões de ambientes mais desfavoráveis.

Os programas de conservação têm tido um papel fundamental na manutenção destes genótipos adaptados e na disseminação de uma cultura de valorização deste patrimônio genético. Porém, para que os programas de conservação sejam otimizados, é necessário um conhecimento profundo das raças/populações alvos destas iniciativas.

Esta tese, portanto, teve o objetivo de contribuir com o conhecimento das três principais raças brasileiras localmente adaptadas de ovinos: Crioula Lanada, Morada Nova e Santa Inês. Para isso foi realizada uma revisão de literatura que descreveu os principais efeitos da seleção natural/artificial nos genomas das espécies de animais de produção, apresentou os principais métodos de análise de assinaturas de seleção e discutiu os recentes avanços nesta área de estudo. Também foram realizados dois estudos, um objetivando identificar e caracterizar assinaturas de seleção e o outro objetivando identificar e caracterizar a presença de subestruturação genética dentro destas raças.

Com o primeiro estudo, pôde-se propor alguns genes candidatos a estar sob seleção que possam influenciar características adaptativas importantes como: resposta imune, desenvolvimento do sistema nervoso, percepção sensorial e desenvolvimento de pêlos/lã. Estes resultados abrem um caminho interessante para futuras investigações, principalmente no entendimento da base genética envolvida na adaptação à ambientes desfavoráveis (como a região semiárida do nordeste brasileiro).

Já o segundo estudo pôde demonstrar que existe uma considerável diferenciação genética entre rebanhos das principais raças brasileiras localmente adaptadas de ovinos: Crioula Lanada, Morada Nova e Santa Inês. Este resultado implica que os estudos de associação entre marcadores moleculares e características fenotípicas devem sempre atentar para a presença da subestruturação genética. Além disso, os estudos sugerem que nas raças Crioula Lanada e Santa Inês existem grupos de animais consideravelmente distintos, o que possivelmente reflete o fato de que nestas raças existe a presença de

diferentes ecótipos/linhagens. A presença destes diferentes ecótipos/linhagens deve ser discutida e, certamente, deve ser considerada na tomada de decisões em relação aos programas de conservação e sociedades de registro genealógico para evitar a perda de parte deste patrimônio genético.