



UNIVERSIDADE FEDERAL DO PIAUÍ - UFPI

CAMPUS UNIVERSITÁRIO MINISTRO PETRÔNIO PORTELLA

CENTRO DE CIÊNCIAS AGRÁRIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

FRANCISCO DE ASSIS DINIZ SOBRINHO

**VARREDURA GENÔMICA EM *Capra hircus* SOB SELEÇÃO NATURAL NO
SEMIÁRIDO**

TERESINA-PI/2019

FRANCISCO DE ASSIS DINIZ SOBRINHO

**VARREDURA GENÔMICA EM *Capra hircus* SOB SELEÇÃO NATURAL NO
SEMIÁRIDO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência Animal da Universidade Federal do Piauí (UFPI), como requisito para obtenção do grau de Mestre em Ciência Animal. Área de Concentração: Ciências Agrárias

Orientadora: Dra. Adriana de Melo Araújo

Coorientador: Dr. Miklos Maximiliano Bajay

TERESINA-PI

2019

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Setorial do Centro de Ciências Agrárias
Serviço de Processamento Técnico

D585v Diniz Sobrinho, Francisco de Assis
Varredura genômica em *capra hircus* sob seleção natural no
semiárido. / Francisco de Assis Diniz Sobrinho - 2019.
51 f. : il.

Dissertação (Mestrado) – Universidade Federal do Piauí, Pro-
grama de Pós-Graduação em Ciência Animal, 2019.
Orientação: Prof^a. Dr^a. Adriana de Melo Araújo

1. Diversidade genética 2. Fst 3. *Loci Outliers* 4. Marota 5.
SNP I. Título.

CDD 636.39083

**VARREDURA GENÔMICA EM *CAPRA HIRCUS* SOB SELEÇÃO NATURAL
NO SEMIÁRIDO**

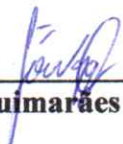
FRANCISCO DE ASSIS DINIZ SOBRINHO

Dissertação aprovada em: 16/01/2019

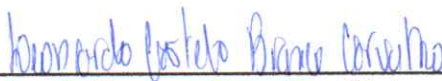
Banca Examinadora:



Profa. Dra. Adriana Mello de Araújo (Presidente) / Embrapa Meio-Norte



Prof. Dr. José Elivalto Guimarães Campelo (Interno) / DZO/CCA/UFPI



Prof. Dr. Leonardo Castelo Branco Carvalho (Externo) / PNPd-CAPES/UFPI



Prof. Dr. Miklos Maximiliano Bajay (Externo) / USP/ESALQ

“Eu sou a luz do mundo; quem me segue não andará em trevas, mas terá a luz da vida. ”

Jesus Cristo

A Deus pelo dom da vida.

A minha esposa Luíza de Sousa Borges Diniz e aos meus pais, Francisco Alves Diniz, Maria das Neves Ribeiro Diniz, exemplos de dedicação e honestidade.

Aos meus filhos, Caio Augusto Borges Diniz, Raíssa Borges Diniz, Yasmin Borges Diniz, Isabella Borges Diniz e Enzo Benjamim Borges Diniz, razão maior da minha dedicação.

A todos que contribuíram para a realização desse trabalho.

Dedico.

Agradecimentos

Em primeiro lugar agradeço a Deus pelo dom da vida. A Ele toda honra e toda glória;

À Universidade Federal do Piauí – UFPI e ao Programa de Pós-Graduação em Ciência Animal do Centro de Ciências Agrária da Universidade Federal do Piauí, pela realização do curso;

Ao Instituto Federal do Piauí – IFPI, por ter me apoiado nesse trabalho;

À EMBRAPA Meio-Norte, por ter me permitido o estudo em animais do rebanho de conservação de Caprinos Marota, por ela mantido, e por ter contribuído na realização da coleta;

À minha orientadora, a Profa. Dra. Adriana Melo de Araújo, pela dedicação, competência, paciência e disponibilidade na orientação;

Ao Dr. Miklos Maximiliano Bajay, pela co-orientação e ajuda na elaboração da dissertação;

À Dra. Jeane de Oliveira Moura, pela paciência, conselhos, incentivo e por ser minha inspiração como profissional.

Ao Dr. Leonardo Castelo Branco, pelas contribuições valiosas e incentivo;

Aos Doutores José Lindenberg Rocha Sarmiento e José Elivalto Guimarães Campelo por suas valiosas sugestões;

Aos meus amigos, Divamélia de Oliveira Bezerra Gomes, Erick Rodolfo Soares Cavalcante, Geice Ribeiro da Silva, Ivanaldo Ribeiro de Moura, Marlúcia da Silva Bezerra Lacerda e Marcelo da Silva Cardoso Ventura, pelas valiosas contribuições;

À minha família, que sempre me incentivou;

Aos colegas do mestrado, de forma especial a Tâmara Rodrigues Pereira e Lilian Rosalina, pelos momentos de estudo e troca de conhecimentos;

A todos que de alguma forma contribuíram direta ou indiretamente para a realização desse trabalho.

Obrigado!

SUMÁRIO

LISTA DE FIGURAS	viii
LISTA DE TABELAS.....	ix
RESUMO	x
ABSTRACT	xi
1 INTRODUÇÃO GERAL	12
2 CAPÍTULO I: Aplicação de marcadores SNP em estudos da diversidade genética em <i>Capra hircus</i> : Tendências, vieses e progresso.....	16
2.1 Resumo	18
2.2 Introdução	19
2.3 Material e Métodos.....	20
2.4 Resultados	21
2.5 Discussão	25
2.6 Considerações Finais	28
2.7 Bibliografia	28
3 CAPÍTULO II: Marcadores <i>outliers</i> em <i>Capra hircus</i> : Regiões genômicas sob ação de seleção natural no semiárido	32
3.1 Resumo	33
3.2 Introdução	34
3.3 Material e Métodos.....	35
3.4 Resultados	38
3.5 Discussão	43
3.6 Conclusões	46
3.7 Agradecimentos.....	47
3.8 Referências.....	47
4 CONSIDERAÇÕES FINAIS	50
5 APÊNDICES	51
5.1 Tabela 1. Resumo dos marcadores SNP analisados e localização nos cromossomos autossomos de <i>C. hircus</i>	52
5.2 Tabela 2. Regiões genômicas sob seleção localizados na região dos 14 SNPs mais significativos (com base no P-valor 0,01 e F_{ST} detectados pela metodologia do FDIST2 no Arlequin)	53
6. ANEXOS	
6.1 Rotinas do R usadas nesse estudo	54

LISTAS DE FIGURAS

Capítulo I

- Figura 1** - Número de publicações científicas baseadas na utilização de marcadores SNP para estudo de diversidade em *Capra hircus* por ano de publicação.21
- Figura 2** - Número de publicações depositadas por autores principais de países com mais de três publicações que utilizaram o marcador SNP para estudos de diversidade genética em *Capra hircus*.
.....21
- Figura 3** - Numero de publicações depositadas por autores secundários dos países que utilizaram SNP para estudos de produção animal e caracterização da biodiversidade em *Capra hircus*.22
- Figura 4** - Número de palavras-chave referentes aos principais temas que envolve produção animal e conservação da biodiversidade em *Capra hircus*.22

Capítulo II

- Figura 1** - Cenário apresentado para os 1761 *loci outliers* que estão sofrendo ação de seleção natural direcional ou disruptiva, pontos acima da linha vermelha. Os demais marcadores são *loci neutros*.38
- Figura 2** - Dispersão dos animais em plano cartesiano relação aos dois primeiros escores dos componentes principais39
- Figura 3** - (A) Distribuição dos valores de P baseados no histograma e (B) no gráfico Q-Q
.....40
- Figura 4** - Cenário para distribuição dos *loci outliers* sem correção FDR/ PCAdapt.....41
- Figura 5** - Cenário para distribuição dos *loci outliers* com correção de 0,01 de FDR/PCAdapt.
.....41

LISTAS DE TABELAS

Capítulo I

Tabela 1 - Dez principais organizações com publicações sobre uso de SNP em *Capra hircus*. 23

Tabela 2 - Publicações depositadas por autor principal do Brasil (1) e em conjunto com autores de outros países (2) que envolvem o uso de SNP na *Capra hircus*.....23

Capítulo II

Tabela 1 - Regiões genômicas sob seleção e lista de genes localizados na região dos 14 SNPs mais significativos (com base no P-valor 0,01 detectados pela metodologia FDIST2 do Arlequin 39

Tabela 2 - SNPs outliers sob seleção natural mostrando níveis significativos (FDR 0,01) e lista de genes localizados na região dos 14 SNPs detectados pelo PCAdapt R..... 42

Apêndice

Tabela 1 - Resumo dos marcadores SNP analisados e localização nos cromossomos autossomos de *C. hircus*. Nas metodologias de F_{ST} outliers e PCAdapt 51

Tabela 2 - Regiões genômicas sob seleção localizados na região dos 14 SNPs mais significativos (com base no P-valor 0,01 e F_{ST} detectados pela metodologia do FDIST2 do Arlequin)52

RESUMO: Caprinos (*Capra hircus*) tem distribuição mundial favorecida pela sua capacidade de adaptação a condições ambientais adversas. No entanto, sacrificando seu desempenho, ajustando-se ao ambiente, em vários países, as raças locais correm o risco de desaparecer pela substituição por raças comerciais mais produtivas. Devido à importância socioeconômica das raças locais, pesquisas utilizando marcadores genéticos de polimorfismo de nucleotídeo único (SNP) foram usadas para esclarecer questões envolvendo mudanças evolutivas e para auxiliar na conservação da biodiversidade da espécie. Assim, objetivamos buscar *loci outliers*, utilizando-se Chip SNP 50K, empregando diferentes metodologias e, associar regiões do genoma de *Capra hircus* a possíveis efeitos do processo de adaptação, assim como identificar genes candidatos que estão sob ação de seleção natural. O capítulo primeiro objetivou analisar a contribuição dos SNPs na genética evolucionária. O software de pesquisa bibliométrica Bibexcel foi utilizado para verificar as publicações do Web of Science TM até agosto de 2018, rastreando artigos com foco em marcadores genéticos do SNP em países de especialização em área tropical. No segundo capítulo, utilizando dados de 96 animais genotipados por Goat Chip Illumina 50K, as metodologias de análise de componentes principais (PCAdapt) e FDIST2 foram aplicadas na detecção de regiões genômicas afetadas pela seleção natural. Das 53.000 produções científicas registradas, 390 citações empregaram os SNPs marcadores. Um total de 59 países apresentou publicações envolvendo o uso de SNPs e a maioria das publicações é de países asiáticos e europeus. Os cinco países com maior produtividade em publicações sobre o tema foram China, Índia, Itália Espanha e Estados Unidos. O Brasil ficou em décimo primeiro lugar, com um de 05 publicações. Diversas palavras-chave foram agrupadas como sendo semanticamente equivalentes e associadas a um tema específico, sendo possível identificar 23% das produções científicas associadas ao polimorfismo das proteínas do leite, 23% com aspectos de reprodução 13% ligados ao crescimento e ganho de peso; diversidade genética com um percentual de 34%. Os estudos de GWAS foram citados em 7% das produções e menos de 1% dos estudos foram associados a características de adaptação. O número de publicações está diretamente relacionado à qualificação técnica e capacidade laboratorial do país. Após a aplicação do controle de qualidade dos dados, 45.600 SNPs foram incluídos nesta investigação. O estudo identificou várias regiões localizadas em 11 cromossomos potencialmente selecionados. Além disso, foram identificados genes cuja expressão está relacionada ao crescimento corporal e desenvolvimento esquelético embrionário (FGF12, AMPD2, OSTN) e para regular o balanço osmótico de caprinos em áreas áridas e semiáridas (UTSB2, SLC5A2) onde estes animais necessitam sobreviver com pouca comida e recursos hídricos.

Palavras-chave: Diversidade genética, F_{ST} , *Loci Outliers*, Marota, SNP.

Abstract: Goats (*Capra hircus*) has worldwide distribution favored by their ability to adapt to adverse environmental conditions. However, sacrificing their performance by adjusting to the environment, in several countries local breeds are at risk of disappearing. Because of local breed's socioeconomic importance, researches applying genetic markers of single nucleotide polymorphism (SNP) had been used to clarify issues involving evolutionary changes and to assist in the conservation of the population's biodiversity. Thus, we aimed to search for discrepant loci (outliers) that exhibit high F_{ST} and associate these regions of the goat genome using a local Brazilian goat breed, Marota. At the same time, to evaluate genetic effects of the adaptation process in Tropical environment, allowing signaling candidate genes, evaluating the possibility of in-depth studies on natural selection signatures, and analyzing contribution of SNP markers as an innovative technology in studies with goats, focusing on production and genetic conservation. The chapter one attempt to the analysis of the contribution of SNPs in evolutionary genetics. Bibexcel bibliometric research software was used to verify the publications of the Web of Science TM until August 2018, tracing articles with SNP genetic marker focus in countries of Tropical area expertise. In the second chapter, using data from 96 animals genotyped by Goat Chip Illumina 50K, the principal component analysis (PCAdapt) and FDist2 methodologies were applied to detection of genomic regions affected by natural selection. Of the 53,000 scientific productions recorded, 390 citations employed the SNPs markers. A number of 59 countries showed publications using SNPs. The greater numbers of publications are from Asian and European countries. The five countries with the highest scientific productivity on the topic were China, India, Italy, Spain and the USA. Brazil was in eleventh position, with a record of five (05) publications. Several of the keywords were grouped as being semantically equivalent and associated to a specific theme, thus it was possible to identify 23% of the scientific productions associated with the polymorphism of milk proteins, 23% with aspects of reproduction 13% linked to growth and gain of weight; genetic diversity with a percentage of 34%. GWAS studies were cited in 7% of the productions and less than 1% of the studies were associated with adaptation characteristics. The number of publications was related to the technical qualification and laboratory capacity of the country. After applying data quality control, 45,600 SNPs were included in this investigation. The study identified several regions located on 11 potentially selected chromosomes. In addition, genes have been identified whose expression is related to body growth and embryonic skeletal development (FGF12, AMPD2, OSTN) and to regulate the osmotic balance of goats in arid and semi-arid areas (UTSB2, SLC5A2) where these animals need to survive with little food and water resources.

Keywords: F_{ST} , Genetical diversity, *Loci Outliers*, Marota, SNP

1 INTRODUÇÃO GERAL

Os caprinos (*Capra hircus*) são provavelmente um dos primeiros ruminantes domesticados no mundo, fato ocorrido há aproximadamente 10 mil anos, na região oeste do Irã (ZEDER; HESSE, 2000). Favorecidas pelas rotas migratórias e distribuição do ser humano, a espécie está presente em praticamente todo o globo (EGITO; MARIANTE; ALBUQUERQUE, 2002), com isso tornou-se um componente importante dos sistemas de produção agrícola e pecuária em todo o mundo, principalmente nas regiões áridas e semiáridas, onde se inseriu morfofisiologicamente como parte integrante do ambiente rústico de subsistência rural, com níveis de produção que vem contribuindo significativamente para a segurança alimentar e para amenizar a situação de pobreza entre agricultores familiares de países em desenvolvimento (LIN et al., 2013; PERIASAMY et al., 2017).

O Brasil se enquadra nesse perfil, considerando-se a atividade rural configurada em duas classes, uma com a participação de grandes e médios produtores (Agrobussines) e a outra classe composta pela grande maioria de pequenos produtores (ANDRADE et al., 2018), onde a criação de caprinos está presente de forma mais intensa. No país os caprinos foram introduzidos pelos portugueses, franceses e holandeses no período da colonização, a partir do século XVI, se estabelecendo mais na região Nordeste, onde foram adquirindo características próprias, resultando na formação da maior diversidade na espécie na América do Sul, com cerca de 21 raças catalogadas (FOOD AND AGRICULTURE ORGANIZATION 2011).

Por vários séculos esses animais passaram por processos adaptativos em diferentes ecossistemas, cujos aspectos climáticos e diversidade vegetal característicos, contribuíram para a formação de raças de animais domésticos do país, ou de grupos reconhecidos como naturalizados (MACHADO, 2011) ou localmente adaptados (ARAÚJO et al., 2009).

Portanto, dos caprinos domésticos introduzidos no período colonial brasileiro, certamente a Marota é um dos descendentes, assim como os caprinos autóctones existentes. Na perspectiva de produção animal, ela apresenta porte reduzido, é bem adaptada a ambientes com pouca disponibilidade alimentar, uma das razões de serem procuradas pelos conservacionistas selecionadores (ARAÚJO et al., 2009).

Os caracteres fenotípicos, que definem esses animais no grupo genético, são o porte pequeno e leve, com adultos pesando em média 36 quilos, pelagem branca uniforme, pelos curtos -com presença ou não de barba e brincos- pele e mucosas clara, com pigmentos na cauda e face interna da orelha, cabeça grande e chifres amarelo claro (ARAÚJO et al., 2009).

Recentes progressos nas tecnologias de sequenciamento de genomas completos (GWAS) associados com a metodologia de microarranjos de deoxyribonucleic acid (DNA), uso de single nucleotide polymorphism (SNP) ou polymerase chain reaction (PCR) em tempo real, proporcionam novos entendimentos sobre a história evolutiva, a diversidade genética dos caprinos e permitem a redução de custos metodológicos em larga escala (AMILLS; CAPOTE; TOSSER-KLOPP, 2017; QIAO et al., 2017).

Associar esse aparato científico tecnológico a traços fenotípicos adaptativos apresenta relevância econômica, incluindo termotolerância, funções ligadas à resposta imune a parasitas, conversão alimentar, crescimento, produção de carne e leite. Além de tornar possível a construção de mapas genéticos para um grande número de animais de interesse zootécnico (LIN et al., 2013; BRITO et al., 2017; PERIASAMY et al., 2017).

Dessa forma, a identificação no genoma caprino de genes candidatos ligados às mais diferentes características adaptativas e de produção, e assinaturas de seleção natural, torna-se passo fundamental para compreensão do processo evolutivo e da sua diversidade genética, pois variações nesses loci podem está ligada direta ou indiretamente à alterações na aptidão em diferentes ambientes. Assim, o uso dos marcadores moleculares SNP para estudos de associação genômica ampla (GWAS, Genoma Wide Association Study) torna-se uma importante ferramenta por permitir entender com maior precisão como as forças ambientais que originam a variação genética adaptativa, úteis na conservação e produção animal, além de auxiliar na identificação, melhoramento e conservação desse importante recurso natural (AMILLS; CAPOTE; TOSSER-KLOPP, 2017; QIAO et al., 2017).

Em resumo, a possibilidade de se identificar traços moleculares da seleção natural no genoma de *Capra hircus* é fundamental para compreensão de como os loci contribuem para os processos adaptativos e produtivos na espécie (BRITO et al., 2017). Para esse fim, a medida mais comum de diferenciação usada é o índice de fixação de Wright, o F_{ST} . Esse índice é frequentemente usado para identificar loci no genoma que se apresentam com diferenciação genética significativamente maior ou menor do que o esperado sob neutralidade, dessa forma são denominados como divergentes ou

discrepantes e que estão provavelmente sob ação de seleção natural (BEAUMONT; NICHOLS, 1996; BEAUMONT, 2005; FENG; JIANG.; FAN, 2015).

Para evitar que um SNP seja erroneamente identificado como outlier, as evidências do processo de adaptação por seleção foi investigada por diferentes abordagens. Nestas foi empregado um dos métodos de detecção de *loci* outliers baseado em análise dos componentes principais (PCA). A grande vantagem dessa abordagem reside no fato de não agrupar os indivíduos em populações e usar técnicas relacionadas de análise multivariadas

Objetivou-se com esta pesquisa verificar o principal foco dos estudos no mundo usando os marcadores SNPs em *Capra hircus* (caprinos), a fim de facilitar estratégias voltadas à conservação e melhoramento genético dos caprinos e detectar *loci* discrepantes (*outliers*) que exibam alto F_{ST} e associar estas regiões do genoma do caprino Marota e da raça comercial Anglonubiana com possíveis efeitos do processo de adaptação, genes candidatos e a possibilidade de estudos aprofundados sobre assinaturas de seleção natural.

Descrição da pesquisa está estruturada conforme a Resolução 001/03 CCMCA, que estabelece as normas para elaboração e apresentação de dissertação do Programa de Pós-Graduação em Ciência Animal da Universidade Federal do Piauí, essa dissertação está estruturada em: Introdução geral; capítulo I, Revisão de literatura na forma de artigo, intitulado “Aplicação de marcadores SNP em estudos da conservação dos recursos genéticos em *Capra hircus*: Tendências, vieses e progresso, redigido conforme as normas editoriais da revista indexada da área (Archivos de Zootecnia – ISSN 0004 - 0592); Capítulo II, intitulado “Marcadores *outliers* em *Capra hircus*: Regiões genômicas sob ação de seleção natural no semiárido”, elaborado de acordo com as normas editoriais da **Revista PloS One** (ISSN 1932-6203), à qual será submetido à publicação. A finalização desta pesquisa é alcançada com as Considerações Finais e anexos.

Referências

- AMILLS, M.; CAPOTE, J.; TOSSER-KLOPP, G. Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces. **Anim Genet**, v. 48, n. 6, p. 631-644, 2017.
- ANDRADE et al. Metodologias para avaliação econômica de sistemas de produção agropecuários. **Archivos de Zootecnia**, v. 67, p. 610 - 620, 2018.
- ARAÚJO, A. M. *et al.* Crescimento e mortalidade em um rebanho de conservação de caprinos Marota no Brasil. **Revista Científica de Produção Animal**, v. 11, p. 103–109, 2009.
- BEAUMONT, M. A.; NICHOLS, R. A. Evaluating Loci for Use in the Genetic Analysis of Population Structure. **Proceedings of the Royal Society B: Biological Sciences**, v. 263, n. 1377, p. 1619–1626, 1996.
- BEAUMONT, M. A. Adaptation and speciation: What can F_{st} tell us? **Trends in Ecology and Evolution**, v. 20, n. 8, p. 435–440, 2005.
- BRITO, L. F. et al. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. **BMC Genomics**, v. 18, n. 1, p. 0–20, 2017.
- DAD-IS. Domestic Animal Diversity Information System. 2011. Disponível em <<http://www.fao.org/dad-is>>. Acesso em: 20 julho 2018.
- EGITO, A. A.; MARIANTE, A. S.; ALBUQUERQUE, M. S. M. Programa brasileiro de conservação de recursos genéticos animais. **Arquivos de Zootecnia**, v. 51, p. 39-52, 2002.
- FENG, X. J.; JIANG, G. F.; FAN, Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiansu*. **Scientific Reports**, v. 5, n.13.758 , p. 1–11, 2015.
- LIN, B. Z. *et al.* Genetic diversity and structure in Asian native goat analyzed by newly developed SNP markers. **Animal Science Journal**, v. 84, n. 8, p. 579-584, 2013.
- LUU, K; BAZIN, E; BLUM ,MGB. pcadapt : an R package to perform genome scan for selection based on principal component. **Molecular Ecology Resources**. v.17, p. 67 – 77, 2017.
- MACHADO, T. ‘História das raças caprinas no Brasil’, in J Fonseca (ed.), *Produção de caprinos e ovinos de leite*, Sobral, Embrapa Caprinos e Ovinos, 2011.
- PERIASAMY, K. et al. Mapping molecular diversity of indigenous goat genetic resources of Asia. **Small Ruminant Research**, v. 148, p. 2–10, 2017.
- QIAO, H. et al. Using the KDE method to model ecological niches: A response to Blonder et al. (2017). **Global Ecology and Biogeography**, v. 26, n. 9, p. 1076–1077, 2017.
- ZEDER, M.A.; HESSE B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. **Science**. v.287, p. 2254-2257, 2000.

2 CAPÍTULO 1

Aplicação de marcadores SNP em estudos de diversidade genética em *Capra hircus*: tendências, vieses e progresso.

***Elaborado segundo normas da revista Archivos de Zootecnia (ISSN 0004 – 0592)**



- ☒ Artículo/Article
- ☐ Nota corta/Short note
- ☒ Revisión/Revisión
- ☐ Carta al Editor/Letter to the Editor
- ☐ In Memoriam

Área de genética y etnología/Area of genetics and ethnology.

Aplicação de marcadores SNP em estudos de diversidade genética em *Capra hircus*: tendências, vieses e progresso / Application of SNP markers for genetic conservation studies in *Capra hircus*: trends, biases and progress.

Sobrinho, F.A.D.^{1,2@}; Moura, J.O.²; Silva, G.R.³; Diniz, F.M.⁴; Carvalho, G.M.C⁵; Araújo, A.M.⁵

¹Departamento de Ciência Animal. Universidade Federal do Piauí. Brasil. profdiniz@ifpi.edu.br

²Instituto Federal do Piauí. Brasil. jeaneprofessora@hotmail.com

³Departamento de Ciência Animal. Universidade Federal do Piauí. Brasil. geiceamb_bio@yahoo.com.br

⁴Embrapa Caprinos e Ovinos. Brasil. fabio.diniz@embrapa.br

⁵Núcleo de Estudo em Produção Animal, Embrapa Meio-Norte. Universidade Federal do Piauí. Brasil. adriana.araujo@embrapa.br

@Email de correspondencia/Correspondence Email: profdiniz@ifpi.edu.br.

Palavras-chave

Conservação.
Genômica evolutiva.
Marcadores genéticos.
Marcador SNP.
Recursos genéticos.

Keywords

Conservation.
Evolutionary genomics.
Genetic markers.
SNP marker.
Genetic resources.

Resumo

Os caprinos (*Capra hircus*) estão distribuídos mundialmente favorecidos por sua capacidade de se adaptar a condições ambientais adversas. Entretanto, por sacrificar seu desempenho se ajustando ao meio, em vários países, raças nativas se encontram sob risco de extinção. Porém, por sua importância socioeconômica, têm recebido atenção em pesquisas com o uso de marcadores genéticos de polimorfismo de nucleotídeo único (SNP), que possibilitam esclarecer questões envolvendo mudanças evolutivas e auxiliam na conservação da biodiversidade populacional. Com esta revisão, objetivou-se analisar a contribuição de marcadores SNP como tecnologia inovadora em estudos com caprinos, com foco na produção e na conservação genética. Utilizou-se o software de pesquisa bibliométrica Bibexcel para análise das publicações da Web of Science TM, até agosto de 2018, rastreando artigos com marcador genético SNP. Das 53.000 produções científicas registradas para a espécie *C. hircus*, 390 publicações empregaram os marcadores SNPs. Um total de 59 países apresentou publicações envolvendo o uso de SNPs nessa espécie, sendo a maioria das publicações de países asiáticos e europeus. Os cinco países de maior produtividade em publicações no tópico foram a China, a Índia, a Itália, a Espanha e os EUA. O Brasil apresentou-se na décima primeira posição, com registro de 05 publicações. Várias das palavras-chave foram agrupadas por serem de forma semântica equivalentes e associadas a um tema específico; assim, foi possível identificar 23% das produções científicas associadas ao tema polimorfismo de proteínas lácteas, 23% com aspectos da reprodução, 13% ligadas ao crescimento e ganho de peso; e diversidade genética com percentual de 34%. Estudos de GWAS foram citados em 7% das produções e menos de 1% dos estudos foi associado a características de adaptação. A quantidade de publicações está diretamente relacionada com qualificação técnica e capacidade laboratorial do país.

Summary

Capra hircus is distributed worldwide due to its extraordinary adaptability and robustness as a source of food, fibers and skin. Genetic markers of single nucleotide polymorphism (SNP) allow the investigation of issues involving evolutionary changes and conservation of population biodiversity. This study aimed to verify the main focus of studies in the world using the marker SNP in *Capra hircus* (goats). Bibexcel bibliometric research software was used to analyze the publications of Web of ScienceTM until August 2018, which adopted the genetic marker SNP. Of 53,000 registered publications, only 390 employed the SNP markers. The three countries that stood out in publications were China (146 publications), India (64) and Italy (42). Brazil was in eleventh position, with a record of five (05) publications. Several of the keywords were grouped as being semantically equivalent and associated to a specific theme, thus it was possible to identify 23% of the scientific productions associated with the polymorphism of milk proteins, 23% with aspects of reproduction 13% linked to growth and gain of weight; genetic diversity with a percentage of 34%. GWAS studies were cited in 7% of the productions and less than 1% of the studies were associated with adaptation characteristics. The number of publications was related to the technical qualification and laboratory capacity of the country.

Introdução

Os caprinos (*Capra hircus*) são provavelmente um dos primeiros ruminantes domesticados no mundo, fato ocorrido há aproximadamente 10 mil anos, na região oeste do Irã (Zeder & Hesse 2000). Favorecidas pelas rotas migratórias e distribuição do ser humano, a espécie está presente em praticamente todo o globo (Egito, Mariante & Albuquerque 2002), adaptada a diversas condições climáticas, sendo um recurso natural importante dos sistemas de produção agrícola e pecuária. Em alguns países, são importantes na produção em larga escala, principalmente de leite, mas também integram a subsistência rural contribuindo significativamente para a segurança alimentar e para agregar renda a famílias rurais de países em desenvolvimento (Lin et al. 2013; Periasamy et al. 2017).

O Brasil se enquadra nesse perfil, considerando-se a atividade rural configurada em duas classes, uma com a participação de grandes e médios produtores (Agrobusiness) e a outra classe composta pela grande maioria de pequenos produtores (Andrade et al. 2018) e onde a criação de caprinos está presente de forma mais intensa. No país, os caprinos foram introduzidos pelos portugueses, franceses e holandeses no período da colonização, a partir do século XVI, se estabelecendo mais na região Nordeste, onde foram adquirindo características próprias e resultando na formação da maior diversidade da espécie na América do Sul, com cerca de 21 raças catalogadas (Food and Agriculture Organization, 2011). Por vários séculos, esses animais passaram por processos adaptativos em diferentes ecossistemas, cujos aspectos climáticos e diversidade vegetal característicos contribuíram para a formação de raças de animais domésticos do país ou de grupos reconhecidos como naturalizados (Machado 2011) ou localmente adaptados (Araújo et al. 2009).

No Brasil, há várias raças caprinas que podem ser consideradas recursos genéticos ricos, com fonte de resistência a parasitas e doenças, atribuível a seu histórico evolutivo de adaptabilidade a condições adversas de ambiente no país. Essa variedade de raças constitui um reservatório de diversidade genética valioso. No entanto, quando comparadas a raças comerciais exóticas, elas apresentam baixo valor comercial (Egito, Mariante & Albuquerque 2002), fato este que levou à importação de raças comerciais de maior importância econômica a partir de 1950, com o objetivo de aperfeiçoar os grupos genéticos naturalizados.

Entretanto, a não consolidação do uso de registro zootécnico e implementação de programas de melhoramento consistentes gerou uma miscigenação ampla de raças a partir de cruzamentos sem planejamento. Isso acarretou a ameaça de extinção dos grupos naturalizados, por erosão e diluição genética, pois traços fenotípicos relacionados à produtividade foram valorizados em detrimento das características adaptativas, essenciais para a sustentabilidade dos sistemas pecuários tradicionais (Araújo et al. 2009).

Esse contexto não é exclusivo do Brasil; assim, além da caracterização do potencial adaptativo das raças locais, se torna necessária a identificação da diversidade genética a nível molecular, para servir de suporte ou de referência para auxiliar na elaboração de políticas públicas de conservação e de reprodução sustentável desse patrimônio biológico (Colli et al. 2014). A esse respeito, avanços nas tecnologias de sequenciamento de genomas completos já estão bem adiantados na maioria das espécies de interesse zootécnico, inclusive, nos caprinos, a tendência é seguir no mesmo caminho, para se beneficiar desses recentes progressos que as tecnologias de sequenciamento de genomas completos propicia, conforme Amills, Capote & Tosser-Klopp (2017), Qiao et al. (2017).

Nas publicações científicas, constam as informações de tendências, vieses e progressos de tecnologias que compõem o estado da arte da genética moderna. Conhecer bem é um importante indicador de avaliação do desenvolvimento científico nessa área em uma espécie. Com esse fim, os métodos bibliométricos analíticos organizados que acessam de forma eficiente os bancos de dados disponíveis em plataformas consolidadas, são ferramentas importantes para avaliação do desempenho científico que, por sua vez, podem auxiliar na definição de estratégias de pesquisas e concepção de projetos (Zanotto, Vanz & Stumpf, 2017).

Dessa forma, a identificação, no genoma caprino, de genes candidatos ligados às mais diferentes características adaptativas e de produção, e assinaturas de seleção natural, torna-se passo fundamental para compreensão do processo evolutivo e da sua diversidade genética, pois variações nesses loci podem estar ligadas direta ou indiretamente a alterações na aptidão em diferentes ambientes. Assim, o uso dos marcadores moleculares SNP para estudos de associação genômica ampla (GWAS, Genome Wide Association Study) torna-se uma importante ferramenta, por permitir entender com maior precisão a forma como agem as forças ambientais que originam a variação genética adaptativa, úteis na conservação e produção animal, além de auxiliar na identificação, melhoramento e conservação desse importante recurso natural (Amills, Capote & Tosser-Klopp 2017; Qiao et al. 2017).

Recentes progressos nas tecnologias de sequenciamento de genomas completos (WGS), associados com a metodologia de microarranjos de deoxyribonucleic acid (DNA), uso de single nucleotide polymorphism (SNP) ou polymerase chain reaction (PCR) em tempo real, proporcionam novos entendimentos sobre a história evolutiva, a diversidade genética dos caprinos e permitem a redução de custos metodológicos em larga escala (Amills, Capote & Tosser-Klopp 2017; Qiao et al. 2017). Associar esse aparato científico-tecnológico a traços fenotípicos adaptativos apresenta relevância econômica, incluindo termotolerância, funções ligadas a resposta imune a parasitas, conversão alimentar, crescimento, produção de carne e leite, além de tornar possível a construção de mapas genéticos para um grande número de animais de interesse zootécnico (Lin et al. 2013; Brito et al. 2017; Periasamy et al. 2017).

Com acesso a banco de dados extraídos do Web of Science™, Coleção Principal da Clarivate Analytics (WoS), objetivou-se, nesta revisão, analisar a contribuição de marcadores SNP como tecnologia inovadora em estudos com caprinos, com relação a produção e a conservação genética.

Material e Métodos

Trata-se de um estudo de revisão de literatura, no qual, por meio do software de pesquisa bibliométrica Bibexcel (Person, Danell & Schneider, 2009), analisaram-se todas as publicações registradas no portal do WoS, até agosto de 2018.

Utilizou-se a plataforma WoS por apresentar características vantajosas, como a de ser uma base multidisciplinar que indexa mais de 12.700 periódicos, nas diferentes áreas científicas, contendo informações que datam do início do século XX e são atualizadas semanalmente, além de disponibilizar endereços completos de autores e coautores de cada trabalho científico (Adriaanse & Rensleigh, 2013). Centralizou-se a prospecção da produção científica nas publicações que envolviam uso do marcador genético SNP na perspectiva de estudos gerais, genética de populações e conservação da biodiversidade na espécie *Capra hircus*.

A investigação com o software Bibexcel foi realizada acessando a opção “pesquisa avançada”. Em seguida, adotou-se uma sequência de palavras-chave, com a inclusão, de forma ordenada, das siglas TS (GOAT* AND SNP*) e CU. A sigla “TS” indica que a pesquisa foi realizada no resumo, nas palavras-chave e no título. Se for acrescido do asterisco “*” possibilita que a busca seja feita, também, utilizando as variações das palavras indicadas no plural ou palavras-compostas. O termo “AND” foi utilizado para que as palavras fossem pesquisadas. O termo “CU” funciona como delimitador do país a ser pesquisado, por exemplo, Brasil. Na revisão, adotou-se a metodologia sugerida por Moura, Dawson e Nogueira (2017), que prioriza o endereço fornecido pela instituição de pesquisa do primeiro autor, incluindo o campo geográfico dos países inseridos na abordagem. Nos casos em que o endereço do primeiro autor não foi identificado, o instituto da segunda autoria foi utilizado.

No entanto, quando o autor apresentou diferentes endereços, foi considerado, neste estudo, aquele indicado para a correspondência. Esse artifício teve como finalidade específica padronizar e organizar os dados por país de origem da instituição de vinculação do pesquisador. Além disso, partiu-se da perspectiva de considerar o primeiro e o segundo autores como os principais investigadores da publicação.

Os resultados foram salvos em arquivo de formato “txt”, para uma busca ampla, recurso encontrado na base de dados da WoS. Com o subsídio do Microsoft Excel®, foi aplicado o filtro dos pesquisadores relacionados aos países descritos na avaliação, conforme critérios e instituições aprimoradas referentes à temática descrita acima. Assim, foram realizadas, com o software, buscas pelas palavras-chave mais usadas para obter um padrão e direcionamento das pesquisas sobre a aplicação do marcador molecular SNP na conservação e produção animal na caprinocultura.

Para verificar a aplicação do uso do marcador SNP em caprinos, registrou-se a frequência de palavras-chave referentes aos principais temas que envolvem produção animal e conservação da biodiversidade: diversidade genética, associação genômica ampla (GWAS), polimorfismo das proteínas lácteas, adaptação, reprodução e crescimento. Várias das palavras-chave foram agrupadas por serem semanticamente equivalentes e associadas a um tema específico da zootecnia (Miguéis et al. 2013).

Por fim, organizou-se uma lista de instituições de ensino e pesquisa que forneceram o endereço de pesquisadores locais e de outros países, com a intenção de contribuir em estudos futuros, como guia para grupos de pesquisa voltados ao uso de marcadores SNP aplicados na produção animal e na conservação da *Capra hircus*. Nesse caso, ao ser constatado mais de um endereço apresentado do mesmo autor, utilizou-se também o fornecido para correspondência, com outra finalidade específica, a de padronizar e organizar, por países, as informações obtidas (Moura, Dawson & Nogueira 2017).

Resultados

Mundo

De acordo com a estratégia de busca, a plataforma WoS apresentou 53.000 publicações científicas envolvendo a caprinocultura (*Capra hircus*). Constatou-se que a maior parte das produções se encontra distribuída em diferentes tipos de documentos, com destaque para mais de 44.000 artigos.

Os primeiros registros de publicações mais consistentes, referentes à utilização de marcadores SNP na caprinocultura, são de 2009, com 24 trabalhos depositados no WoS. Com o tempo, aumentaram significativamente em número a cada ano, tendo o ano de 2016 como o mais produtivo, com 55 publicações (Figura 1). Portanto, verificou-se que publicações direcionadas à utilização de marcadores

SNP nos estudos sobre genética populacional, com enfoque para produção e conservação de caprinos, estão cada vez mais valorizadas, com especial contribuição de estudos de associação genômica ampla (GWAS).

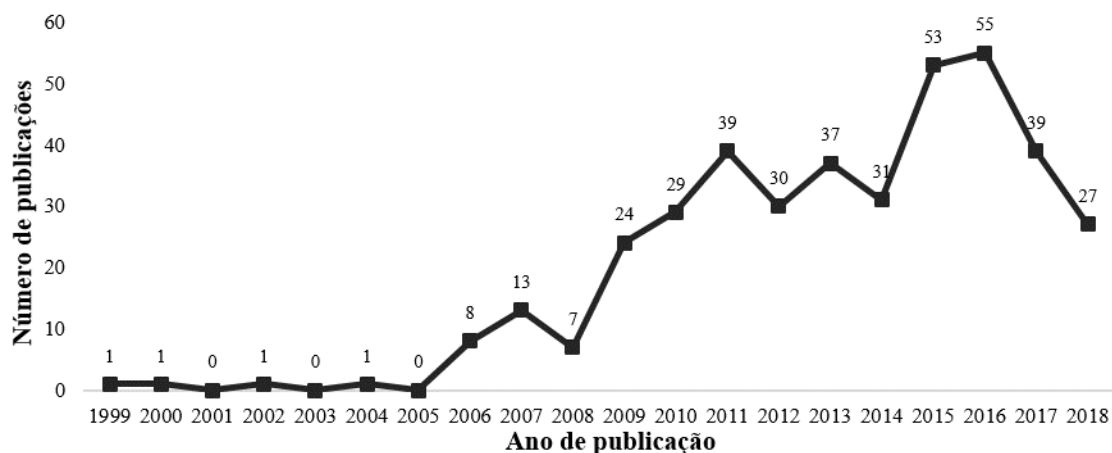


Figura 1. Número de publicações científicas baseadas na utilização de marcadores SNP para estudo de diversidade em *Capra hircus* por ano de publicação.

Pesquisadores de 59 países participaram da produção de 390 publicações envolvendo o uso dos marcadores moleculares de DNA SNP em caprinos. Dentre esses países, a maioria das publicações é oriunda de instituições asiáticas e europeias. China (146), Índia (64) e Itália (42) foram os países mais produtivos em publicações. O Brasil apresentou-se na décima primeira posição com cinco trabalhos publicados por autores principais (Figura 2).

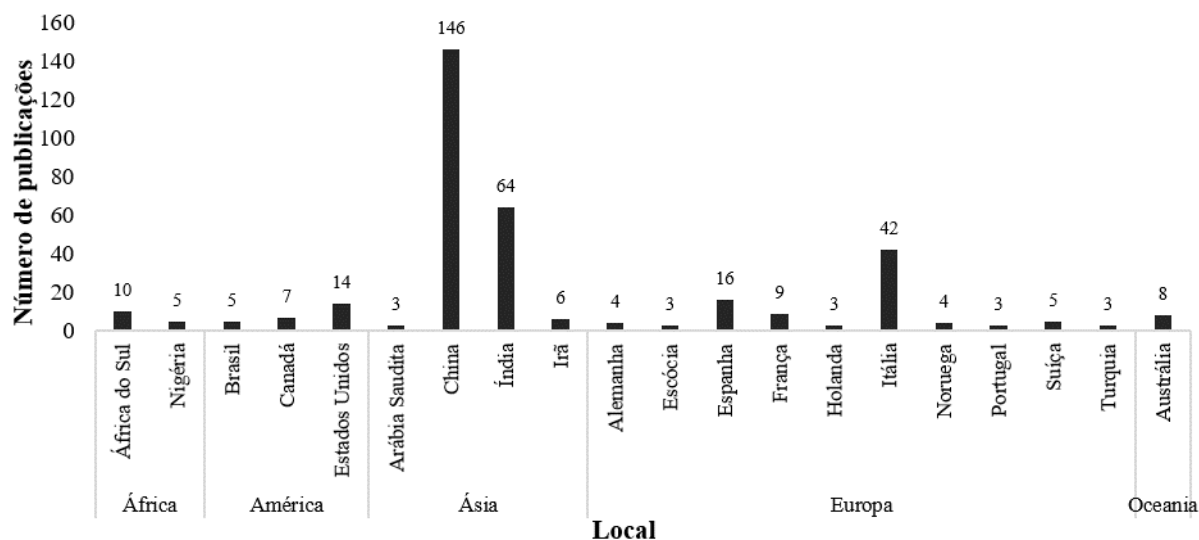


Figura 2. Número de publicações depositadas por autores principais de países com mais de três publicações que utilizaram o marcador SNP para estudos de diversidade genética em *Capra hircus*.

Nos artigos que envolveram publicação em que o autor era de uma instituição de um país que não corresponde ao seu endereço de origem nacional, incluído na busca com o critério do segundo autor,

novamente se destacaram, no cenário mundial, os países China, Índia e Itália, com 93, 36 e 28 trabalhos, respectivamente (Figura 3).

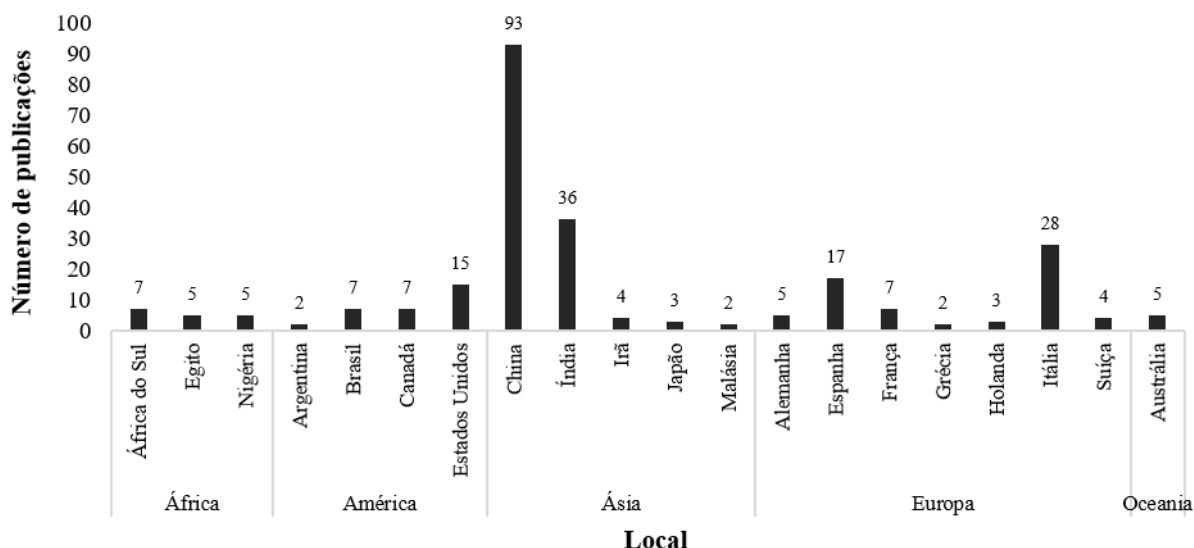


Figura 3. Número de publicações depositadas por autores secundários dos países que utilizaram SNP para estudos de produção animal e caracterização da biodiversidade em *Capra hircus*.

Em termos quantitativos, identificaram-se 393 palavras-chave que se repetiram nas publicações selecionadas. São específicas e representam os conteúdos descritos nos textos de forma eficiente. Ao se avaliar o significado das palavras-chave, identificou-se que os estudos de diversidade genética e genômica populacional foram relacionados a 162 palavras-chave (aproximadamente 41%). Nas demais áreas, constatou-se, nas publicações voltadas para a produção animal, que 23% tratavam sobre polimorfismo das proteínas lácteas (90 palavras-chave); sobre aspectos da reprodução, 23% (89 palavras-chave); e crescimento corporal, 13% (52 palavras-chave), conforme Figura 4.

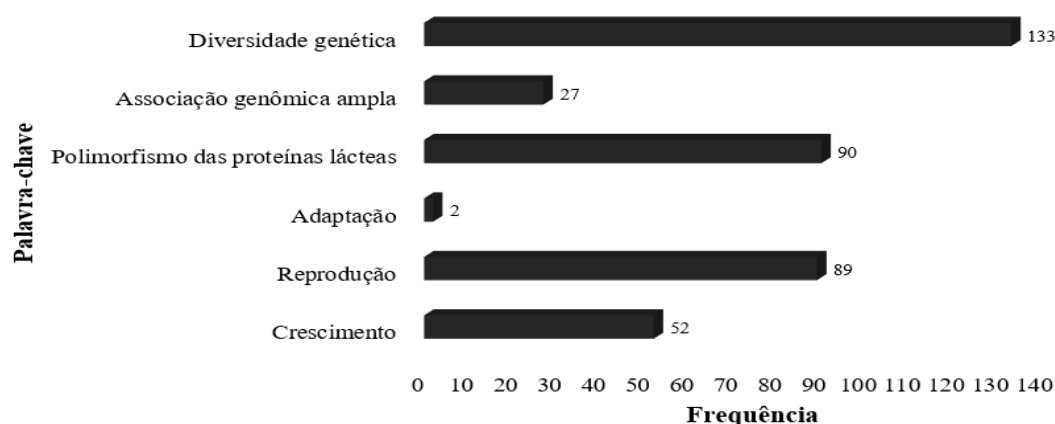


Figura 4. Frequência de palavras-chave referentes aos principais temas que envolvem produção animal e conservação da biodiversidade em *Capra hircus*.

No Brasil, cinco produções científicas foram de autores principais. Comparando esses artigos, em relação ao uso do marcador molecular SNP na caprinocultura, houve pouca abrangência contextual,

sendo que apenas uma publicação estava direcionada à diversidade gênica e resistência a endoparasitas gastrointestinais nos caprinos (Tabela 2).

Tabela 1. Dez principais organizações com publicações sobre uso de SNP em *Capra hircus*.

País	Organização	Registros	
		n	%
China	North West O. F. University China	81	30,57
	Jiangsu Normal University	27	10,19
	Chinese Academy of Agricultural Science	15	5,66
	Huazhong Agricultural University	12	4,53
Índia	Indian Council of Agricultural Research Icar	46	17,36
	Icar National Bureau of Animal Genetic Resources	36	13,58
Itália	University of Sassari	13	4,91
	University of Milan	11	4,15
França	Institut National de la Recherche Agronomique	14	5,28
Espanha	Autonomous University of Barcelona	10	3,77
Total		265	100,0

Fonte: Web of ScienceTM

Em sete publicações, o trabalho envolveu a participação de pesquisadores brasileiros em grupos de estudos da Itália, Portugal, Espanha, Argentina e Colômbia. Neles, o uso de SNP em caprinos se relacionou a estudo de diversidade genética voltada para produção de leite e conservação animal (Tabela 2).

Tabela 2. Publicações depositadas por autor principal do Brasil (1) e em conjunto com autores de outros países (2) que envolvem o uso de SNP na *Capra hircus*.

(1) Publicações de autores principais brasileiros		
Organização/país	Referência	Título da publicação
Universidade Estadual Paulista	Stafuzza et al. 2016	Sequence analysis of the S1PR1 gene in river buffalo.
	Borges et al. 2013	Clinical and molecular study of a new form of hereditary myotonia in Murrah water buffalo.
Universidade Federal Fluminense	Olivares et al. 2015	Comparison of different methods of goat sperm selection and capacitation for optimization of assisted reproductive technologies.
Embrapa Pecuária Sudeste, São Carlos, São Paulo	Bressani et al. 2014	Single nucleotide polymorphisms in candidate genes associated with gastrointestinal nematode infection in goats.
Universidade Federal do Rio Grande do Sul	Andrade, Barbosa & Driemeier 2018	Identification of single nucleotide polymorphisms in the prion protein gene in Santa Ines and Dorset.
(2) Publicações conjuntas		
Colômbia/Brasil/ Espanha/Itália/ Grécia/Irã/Romênia	Kirikçi et al. 2016	Analysing the diversity of the caprine melanocortin 1 receptor (MC1R) in goats with distinct geographic origins
Argentina/Brasil	Cardona et al. 2015	Association of SNPs in the genes for kappa-casein and beta-lactoglobulin with lactation curves in dairy goats
Itália/Brasil	Sardina et al. 2012	Polymorphisms of beta-lactoglobulin promoter region in three Sicilian goat breeds
Portugal/Brasil	Gama & Bressan 2011	Biotechnology applications for the sustainable management of goat genetic resources
Itália/Brasil	Caprera et al. 2007	Gosh: a web-based database for goat and sheep EST sequences
Itália/Brasil	Caprera et al. 2007	Gosh: a goat and sheep ESTs database

Fonte: Web of ScienceTM

Discussão

A utilização da plataforma WoS, neste estudo, possibilitou acesso a 53.000 publicações, valor que demonstra, de forma consistente, as vantagens de se recorrer a uma base multidisciplinar que indexa mais de 12.700 periódicos, nas diferentes áreas científicas. Ela contém informações que datam do início do século XX e são atualizadas semanalmente, além de dispor de endereços completos de autores e coautores de cada trabalho científico (Adriaanse & Rensleigh, 2013).

Essa base de dados presta serviço de indexação de referências baseado em assinaturas e propicia a realização de pesquisa ampla sobre citações, ao permitir o acesso a vários bancos de dados relacionados com pesquisas de abrangência multidisciplinar. Nesse sentido, o uso do WoS vem crescendo e ajudando o mundo acadêmico a entender, com mais precisão, as publicações científicas (Liu et al. 2018).

O conhecimento detalhado das publicações permite aos pesquisadores avaliar tendências ou o foco principal dos trabalhos depositados nessa base de dados. Assim, nas áreas de produção e conservação animal, é imperativo ter o conhecimento detalhado da diversidade genética dentro das populações e entre elas, bem como ter ciência dos manejos adequados, da capacidade adaptativa e produtiva das raças e dos grupos genéticos caprinos locais ou naturalizados (Toro, Fernández & Caballero 2009). Dessa forma, discute-se a situação geral das publicações com enfoque no uso dos marcadores SNP para os estudos de produção e conservação dos recursos genéticos em *Capra hircus*, no mundo e no Brasil.

Produção científica no Mundo

Os caprinos estão entre os pequenos ruminantes mais importantes dentre os animais de interesse zootécnico. Presentes mais nos países em desenvolvimento, desempenham, de forma eficiente, papel socioeconômico e cultural nos sistemas de produção da agricultura familiar (Onzima et al. 2018).

Em um contexto geral, Groeneveld et al. (2010) asseguram que as informações disponíveis sobre a caracterização genômica de recursos genéticos, dentre eles os caprinos, em todo o mundo, aumentaram nos últimos anos. Neste estudo, o levantamento de publicações direcionadas à utilização de marcadores SNP, nos estudos sobre genômica evolutiva com enfoque para produção animal e conservação de caprinos, corroboram essa afirmação.

Os primeiros registros de publicações mais consistentes, referentes à utilização de marcadores SNP na caprinocultura, são de 2009, com 24 trabalhos depositados no WoS, e aumentaram significativamente em número a cada ano, tendo o ano de 2016 como o mais produtivo, com 55 publicações (Figura 1).

Esses resultados são decorrentes do progresso nas tecnologias de sequenciamento de genomas completos (WGS) associados com a metodologia de microarranjos de deoxyribonucleic acid (DNA), uso de single nucleotide polymorphism (SNP) ou polymerase chain reaction (PCR) em tempo real, que proporcionaram novos entendimentos sobre a história evolutiva e a diversidade genética dos caprinos. A automação tornou possível a redução de custos de sequenciamento em larga escala (Amills, Capote & Tosser-Klopp 2017; Qiao et al., 2017).

Associar esse aparato científico-tecnológico a traços fenotípicos, por exemplo, adaptativos, funções ligadas à resposta imune a parasitas, conversão alimentar, crescimento, produção de carne e leite,

agrega relevância econômica suficiente para justificar essa evolução. Tornar possível a construção de mapas genéticos dos animais de interesse zootécnico (Lin et al. 2013; Brito et al. 2017; Periasamy et al. 2017) impactou positivamente na quantidade de publicações e isso foi detectado neste estudo.

A utilização de abordagens estatísticas mais robustas, associada ao desenvolvimento de tecnologias de sequenciamento genômico mais recentes, tornou possível a compreensão mais clara da diversidade genética em caprinos e, como inovação, também pode influenciar a ocorrência de número crescente de publicações na área (Benjelloun et al. 2015; Ahrens et al. 2018).

Percebeu-se uma grande diferença na quantidade de trabalhos publicados por países asiáticos em relação aos demais (europeus, americanos e africanos), envolvendo o uso de marcadores moleculares SNP na produção animal e conservação dos caprinos. Em uma abordagem geral, a China e a Índia ocupam posição de destaque, pois respondem por mais de 54% das publicações depositadas no WoS, no período avaliado.

Esses resultados não podem ser justificados exclusivamente como sendo uma consequência da influência do efetivo de animais do país, pelo fato de o maior rebanho caprino do planeta estar localizado nos países da Ásia e da África, nos quais a caprinocultura é uma significativa atividade econômica e representa uma das mais importantes fontes de proteína animal para a alimentação humana. Mas, principalmente, por dispor de recursos humanos habilitados, aliado à capacidade tecnológica. A consistência dessa afirmação está embasada no fato de países com produção científica na área possuírem centros especializados para formação de pessoal qualificado.

Nesse contexto, a North West O. F. University China, Indian Council of Agricultural Research Icar e a University of Milan são instituições públicas, localizadas em países da Ásia e da Europa, que podem ser consideradas centros de alta produção científica e, portanto, fomentadores de conhecimento agropecuário de destaque no cenário mundial, sendo referência para pesquisadores e instituições de outras nações, conforme Tabela 1.

Ao se estabelecer um paralelo, em termos quantitativos, entre rebanho caprino e produtividade científica, o continente europeu se apresenta como o segundo mais produtivo, com registros de mais de 32% dos trabalhos depositados na base de dados. No entanto, a Europa tem um rebanho bastante reduzido (1,65% do efetivo mundial), com a produção voltada ao ramo leiteiro. A Itália, a Espanha e a França são merecedoras de destaque, pois produziram 49, 22 e 18 publicações, respectivamente. Conforme Aziz (2010), a Espanha, a França e a Grécia são os países que estabeleceram programas eficientes de seleção, processamento e comercialização do leite de cabra, fato evidenciado pelo uso das palavras-chave que envolvem termos referentes ao leite, presentes nas publicações depositadas no WoS por países europeus.

Apesar de o continente africano possuir um rebanho caprino de cerca de 35% do quantitativo mundial, percentual relevante no cenário global, não apresenta um número de trabalhos depositados de forma representativa, visto que foi comprovada a ocorrência de apenas 41 publicações. Confrontando a relevância numérica do rebanho caprino e o número de publicações de pesquisadores africanos, o valor pode ser considerado pequeno. Mesmo assim, países, como a África do Sul (14 publicações), o Egito (07) e a Nigéria (07), respectivamente, superam países da América do Sul e se destacam no continente.

Os números apresentados podem ser justificados pela falta de centros especializados em formação de pessoal qualificado no desenvolvimento de tecnologias genéticas (Tabela 1). Além disso, tendo em vista os equipamentos e tecnologias caros, não se destacam políticas públicas de financiamento para

equipar laboratórios, tampouco transferência de conhecimento científico, fundamentais no estudo de diversidade genética aplicada na produção e conservação dos rebanhos caprinos. É possível que os pesquisadores africanos produzam seus trabalhos em parceria com autores de outros continentes, devido à melhor infraestrutura dos laboratórios e maior incentivo governamental.

Ao se analisarem os dados com base no significado das palavras-chave indexadoras, observa-se que, nos países mais produtivos em publicações, vários temas foram explorados, destacando-se: diversidade genética, associação genômica ampla, polimorfismo das proteínas lácteas, crescimento e reprodução. Portanto, há, nesses países, preocupação institucional e de pesquisadores na caracterização e entendimento da diversidade genética caprina (Guo et al. 2018; Ming-Xing et al. 2018).

Além disso, constata-se concordância entre pesquisadores quanto ao fato de o uso do SNP, no estudo da diversidade genética, ter sido fundamental, pois fornece informações sobre a estrutura populacional dos grupos genéticos e de raças, bem como redução de diversidade, que são essenciais para o estabelecimento de metas de conservação e produção animal (Toro, Fernández & Caballero 2009; Brito et al. 2017).

Em síntese, a promoção e o desenvolvimento da pesquisa dentro de um país são dependentes da presença de instituições fortes e de profissionais qualificados para o desenvolvimento de tecnologias genômicas, programas de formação de pessoal qualificado, assim como a necessidade de políticas de financiamento para equipar laboratórios e a difusão de conhecimento científico. Nesse aspecto, acredita-se que essas sejam as principais diferenças entre o que ocorre nos continentes em relação ao estudo de caprinos.

Situação no Brasil

Apesar de possuir um rebanho caprino de aproximadamente 9,6 milhões de cabeças, distribuído principalmente na região Nordeste, a produção brasileira de carne, leite e derivados dessa espécie é baixa, requerendo atenção política para esse ramo da pecuária. Essencialmente, o conhecimento da estrutura populacional e da diversidade genética dos grupos caprinos locais, no sentido de otimizar a produtividade animal e conservação de seu patrimônio genético, pode favorecer a expansão do setor, sendo os marcadores SNP ferramentas de auxílio a essa atividade (Onzima et al. 2018).

Embora o Brasil seja, nesse tema, o país mais produtivo da América do Sul, apenas 11 artigos foram publicados e depositados no WoS até o período de recuperação das publicações. Desses trabalhos, cinco envolveram pesquisadores clasificados, no levantamento realizado, como primeiro autor, e os outros seis foram produzidos com colaboração de pesquisadores de outros países, como a Colômbia, Itália, Espanha e Portugal. Das cinco publicações que envolvem a aplicação de marcadores SNP na caprinocultura, apenas um efetivamente caracterizou o seu estudo no tema principal, o que coloca esse campo de pesquisa brasileiro promissor. Em contrapartida, os outros seis trabalhos que têm a cooperação de grupos estrangeiros e de pesquisadores brasileiros centralizaram seus estudos no conhecimento de diversidade genética populacional para melhoria na produção e conservação animal dos grupos genéticos caprinos.

Além disso, é possível que essa parceria científica entre pesquisadores brasileiros e instituições de outros países decorra da melhor infraestrutura laboratorial e do incentivo governamental para apoiar a

pesquisa científica e a formação de pessoal qualificado encontradas nas instituições dos países parceiros. Parcerias são fundamentais para formação de pesquisadores mais qualificados.

Entretanto, é fundamental despertar, nas instituições públicas ou privadas, o interesse em fomentar o desenvolvimento e a disseminação das pesquisas científicas voltadas para os diferentes campos de interesse da sociedade, com vistas ao alcance de desenvolvimento social, econômico e sustentável do país. Essa visão despertou, nas instituições de diferentes países, a necessidade de um olhar mais direcionado ao desenvolvimento das ciências e tecnologias.

Considerações finais

Nos últimos 10 anos, inovações metodológicas envolvendo o marcador molecular SNP ocorreram e foram utilizadas em estudos genômicos e de estrutura populacional na caprinocultura, gerando informações aplicadas na conservação e produção animal. Nesta revisão, a partir de 2009, verificou-se o interesse crescente dos pesquisadores pela aplicação dos SNPs na caprinocultura, com destaque para China, Índia e Itália, que despontaram com maior número de publicações, com prevalência de estudos voltados para diversidade genética populacional, polimorfismo das proteínas lácteas e aspectos reprodutivos.

As pesquisas com caprinos usando os marcadores SNP, direcionadas a aspectos produtivos e reprodutivos, relacionam a importância no mercado mundial da carne, principalmente para atender ao consumo asiático, ao do leite e derivados, onde se destaca a França. Embora já seja conhecido o genoma completo, percebe-se que existem ainda muitas limitações para a aplicação dos marcadores moleculares SNP em caprinos, principalmente de ordem econômica, como a falta de laboratórios e recursos humanos especializados, representando uma barreira para estudos genéticos baseados nessa tecnologia nos países em desenvolvimento, como o Brasil. Para superar limitações, políticas públicas voltadas para o aparelhamento de laboratórios e formação de pessoal qualificado devem ser priorizadas.

Bibliografia

- Adriaanse, L & Rensleigh, C 2013, 'Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison', *The Electronic Library*, vol. 31, no. 6, pp.727-44.
- Ahrens, C, Rymer, P, Stow, A, Bragg, J, Dillon, S, Umbers, K & Dudaniec, R 2018, 'The search for loci under selection: trends, biases and progress', *Molecular ecology*, vol. 27, no. 6, pp. 1342-56.
- Amills, M, Capote, J & Tosser-Klopp, G 2017, 'Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces', *Animal Genetics*, vol. 48, no. 6, pp. 631-44.
- Andrade, C, Barbosa, J & Driemeier, D 2018, 'Identification of single nucleotide polymorphisms in the prion protein gene in Santa Ines and Dorset sheep', *Pesquisa Veterinária Brasileira*, vol. 38, no. 4, pp. 624-8.
- Araújo, A, Beffa, L, Almeida, M, Abreu, U, Cavalcante, D, Leal, T & Paiva, S 2009, 'Crescimento e Mortalidade em um Rebanho de Conservação de Caprinos Marota no Brasil', *Revista Científica de Produção Animal*, vol. 11, no. 2, pp.103-9.
-

- Aziz, M 2010, 'Present status of the world goat populations and their productivity', *Lohmann Information*, vol. 45, no. 2, pp. 42-52.
- Benjelloun, B, Alberto, F, Streeter, I, Boyer, F, Coissac, E, Stucki, S, BenBati, M, Ibnelbachyr, M, Chentouf, M, Bechchari, A, Leempoel, K, Alberti, A, Engelen, S, Chikhi, A, Clarke, L, Flicek, P, Joost, S, Taberlet, P, Pompanon, F & NextGen Consortium 2015, 'Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data', *Frontiers In Genetic*, vol. 6, pp. 1-14.
- Borges, A, Barbosa, J, Resende, L, Mota, L, Amorim, R, Carvalho, T, Garcia, J, Oliveira-Filho, J, Oliveira, C, Souza, J & Winand, N 2013, 'Clinical and molecular study of a new form of hereditary myotonia in Murrah water buffalo'. *Neuromuscular Disorders*, vol. 23, no. 3, pp. 206-13.
- Bressani, F, Tizioto, P, Giglioti, R, Meirelles, S, Coutinho, R, Benvenuti, C, Malagó, W, Mudadu, M, Vieira, L, Zaros, L, Carrilho, E & Regitano, L 2014, 'Single nucleotide polymorphisms in candidate genes associated with gastrointestinal nematode infection in goats', *Genetics and Molecular Research Journal*, vol. 13, no. 4, pp. 8530-6.
- Brito, L, Kijas, J, Ventura, R, Sargolzaei, M, Porto-Neto, L, Cánovas, A, Feng, Z, Jafarikia, M & Schenkel, F 2017, 'Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers', *BMC Genomics*, vol. 18, no. 1, pp. 1-20.
- Caprera, A, Lazzari, B, Stella, A, Merelli, I, Caetano, A & Mariani, P 2007, 'GoSh: a web-based database for goat and sheep EST sequences', *Bioinformatics*, vol. 23, no. 8, pp. 1043-5.
- Caprera, A, Lazzari, B, Stella, A, Merelli, I, Caetano, A & Mariani, P 2007, 'GoSh: a goat and sheep ESTs database', *Italian Journal of Animal Science*, vol. 6, pp. 60-2.
- Cardona, S, Álvarez, J, Sarmiento, J, Herrera, L, & Cadavid, H 2015, 'Associação de SNPs nos genes para κ -caseína e β -lactoglobulina com curvas de lactação em cabras leiteiras', *Pesquisa Agropecuária Brasileira*, vol. 50, no. 3, pp. 224-32.
- Colli, L, Joost, S, Negrini, R, Nicoloso, L, Crepaldi, P, Ajmone-Marsan, P & ECONOGENE Consortium 2014, 'Assessing the spatial dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers', *PLoS ONE*, vol. 9, no. 1, pp. 1-12.
- Food and Agriculture Organization 2011, *Domestic Animal Diversity Information System (DAD-IS)*, viewed 20 July 2018, <http://www.fao.org/dad-is/en/>
- Egito, A, Mariante, A & Albuquerque, M 2002, 'The Brazilian Genetic Resources Conservation Programm', *Archivos de Zootecnia*, vol. 51, pp. 39-52.
- Gama, L & Bressan, M 2011, 'Biotechnology applications for the sustainable management of goat genetic resources'. *Small Ruminant Research*, vol. 98, no. 1-3, pp. 133-46.
- Groeneveld, L, Lenstra, J, Eding, H, Toro, M, Scherf, B, Pilling, D, Negrini, R, Finlay, E, Jianlin, H, Groeneveld, E, Weigend, S & GLOBALDIV Consortium 2010, 'Genetic diversity in farm animals--a review', *Animal Genetics*, vol. 41, pp. 6-31.
- Guo, J, Tao, H, Li, P, Li, L, Zhong, T, Wang, L, Ma, J, Chen, X, Song, T & Zhang, H 2018, 'Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds'. *Scientific Reports*, vol. 8, no. 10405, pp. 1-11.
- Kirikçi, K, Noce, A, Zidi, A, Serradilla, J, Carrizosa, J, Urrutia, B, Pilla, F, D'Andrea, M, Capote, J, Bizelis, I, Balteanu, V, Cardoso, T, Eghbalsaiied, S, Pons, A, Álvarez, L, Pazzola, M, Vacca, G, Obexer-Ruff, G & Amills, M 2016, 'Analysing the diversity of the caprine melanocortin 1 receptor (MC1R) in goats with distinct geographic origins', *Small Ruminant Research*, vol. 45, pp. 7-11.

- Lin, B, Kato, T, Kaneda, M, Matsumoto, H, Sasazaki, S & Mannen, H 2013, 'Genetic diversity and structure in Asian native goat analyzed by newly developed SNP markers', *Animal Science Journal*, vol. 84, no. 8, pp. 579-84.
- Liu, W, Ni, M, Jia, W, Wan, W & Tang, J 2018, 'Evidence-based medicine in neurosurgery: an academic publication view', *Neurosurgical Review*, vol. 41, no. 1, pp. 55-65.
- Machado, T 2011, 'História das raças caprinas no Brasil', in J Fonseca (ed.), *Produção de caprinos e ovinos de leite*, Sobral, Embrapa Caprinos e Ovinos.
- E, GX, Zhao, YJ, Chen, LP, Ma, YH, Chu, MX, Li, XL, Hong, QH, Li, LH, Guo, JJ, Zhu, L, Han, YG, Gao, HJ, Zhang, JH, Jiang, HZ, Jiang, CD, Wang, GF, Ren, HX, Jin, ML, Sun, YZ, Zhou, P & Huang, YF 2018, 'Genetic diversity of the Chinese goat in the littoral zone of the Yangtze River as assessed by microsatellite and mtDNA', *Ecology and Evolution*, vol. 8, no. 10, pp. 5111-23.
- Miguéis, A, Neves, B, Silva, A, Trindade, A & Bernardes, J 2013, 'A importância das palavras-chave dos artigos científicos da área das Ciências Farmacêuticas, depositados no Estudo Geral: estudo comparativo com os termos atribuídos na MEDLINE', *InCID: Revista de Ciência da Informação e Documentação*, vol. 4, no. 2, pp. 112-5.
- Moura, R, Dawson, D & Nogueira, D 2017, 'The use of microsatellite markers in neotropical studies of wild birds: A literature review. *Anais da Academia Brasileira de Ciências*, vol. 89, no. 1, pp. 145-54.
- Olivares, C, Fonseca, J, Camargo, L, Souza-Fabjan, J, Rodrigues, A & Brandão, F 2015, 'Comparison of different methods of goat sperm selection and capacitation for optimization of assisted reproductive technologies', *Small Ruminant Research*, vol. 107, pp. 44-9.
- Onzima, R, Upadhyay, M, Mukiibi, R, Kanis, E, Groenen, M & Crooijmans, R 2018, 'Genome-wide population structure and admixture analysis reveals weak differentiation among Ugandan goat breeds', *Animal Genetics*, vol. 49, no. 1, pp. 59-78.
- Periasamy, K, Vahidi, S, Silva, P, Faruque, M, Naqvi, A, Basar, M, Cao, J, Zhao, S, Thuy, L, Pichler, R, Podesta, M, Shamsuddin, M, Boettcher, P, Garcia, J, Han, J, Marsan, P, Diallo, A & Viljoen, G 2017, 'Mapping molecular diversity of indigenous goat genetic resources of Asia'. *Small Ruminant Research*, vol. 148, pp. 2-10.
- Persson, O, Dannell, R, Schneider, J,W, "How to use Bibexcel for various Types of bibliometric analysis. In Celebration scholarly communication studies: A festschrift for olle person at his 60th Birthday, ed. F. Astron, R. Danell, B. Larsen, J. Schneider, 2009. Leuven, Belgium: International society for scientometrics and informetrics, 2009, p.9-24.
- Qiao, H, Escobar, L, Saupe, E, Ji, L & Soberón, J 2017, 'Using the KDE method to model ecological niches: A response to Blonder et al. (2017)', *Global Ecology and Biogeography*, vol. 26, no. 9, pp. 1076-7.
- Sardina, M, Rosa, A, Davoli, R, Braglia, S & Portolano, B 2012, 'Polymorphisms of beta - lactoglobulin promoter region in three Sicilian goat breeds', *Molecular Biology Reports*, vol. 39, no. 3, pp. 3203-10.
- Stafuzza, N, Naressi, B, Borges, M & Amaral-Trusty M 2016, 'Sequence analysis of the S1PR1 gene in river buffalo', *Genetics and Molecular Research Journal*, vol. 15, no. 1, pp. 1-8.
- Toro, M, Fernández, J & Caballero, A 2009, 'Molecular characterization of breeds and its use in conservation', *Livestock Science*, vol. 120, no. 3, pp. 174-95.
- Zanotto, S, Vanz, S, & Stumpf, I 2017, 'Fator de difusão: uma medida da difusão do conhecimento através das citações', *Investigación bibliotecológica*, vol. 31, no. spe, pp. 101-22.

485 Zeder, M & Hesse, B 2000, 'The initial domestication of goats (*Capra hircus*) in the Zagros mountains
486 10,000 years ago', *Science*, vol. 287, no. 5461, pp. 2254-7.

3 CAPÍTULO 2

Marcadores *outliers* em *Capra hircus*: Regiões genômicas sob ação de seleção natural no semiárido

Marcadores *outliers* em *Capra hircus*: Regiões genômicas sob ação de seleção natural

**Francisco de A. D. Sobrinho^{1,2*}; Jeane de O. Moura², Miklos M. Bajay³; Leonardo Castelo Branco⁴
Adriana M. de Araújo⁵**

1 Programa de pós-graduação em Ciência Animal/UFPI, 2 Instituto Federal do Piauí, Teresina, Brasil, 3 Universidade Estadual de São Paulo, Brasil, 4 Universidade Federal do Piauí, Teresina, Piauí, Brasil, 5 Embrapa Meio-Norte, Teresina, Piauí, Brasil

***profdiniz@ifpi.edu.br**

Resumo

Os caprinos (*Capra hircus*) estão distribuídos mundialmente devido à extraordinária adaptabilidade e robustez, servindo como fonte de alimento, fibras e pele. Os marcadores genéticos de polimorfismo de nucleotídeo único (SNP) possibilitam investigar e esclarecer questões envolvendo mudanças evolutivas e conservação da biodiversidade populacional. Os objetivos do presente trabalho foram: buscar *loci* discrepantes (*outliers*) que exibam alto F_{ST} e PCA; associar as regiões do genoma dos caprinos Marota e Anglonubiana a possíveis efeitos do processo de adaptação, genes candidatos e aproveitar a possibilidade de estudos aprofundados sobre assinaturas de seleção natural. A busca por regiões adaptativas entre Anglonubiana e Marota foi realizada usando dados de genotipagem de um total de 96 animais. Depois de aplicar controle de qualidade de dados, 45.600 SNPs foram incluídos nesta investigação. A detecção de regiões genômicas afetadas pela seleção natural foi realizada utilizando as metodologias de análise de componentes principais (PCAdapt) e FDist2. O estudo identificou várias regiões localizadas em 11 cromossomos potencialmente selecionadas. Além disso, identificaram-se genes cuja expressão está relacionada ao crescimento corporal e a desenvolvimento esquelético embrionário (FGF12, AMPD2, OSTN) e de regulação do equilíbrio osmótico dos caprinos nas áreas áridas e semiáridas (UTSB2, SLC5A2), onde esses animais precisam sobreviver com pouco recurso alimentar e hídrico. Percebeu-se que a determinação da localização e caracterização dos *outliers* no genoma caprino é fundamental para compreender a função e o envolvimento dessas regiões no processo evolutivo dos grupos genéticos locais ou naturalizados, e que os Marcadores SNPs permitem criar uma robusta fundamentação conceitual e mapas de seleção natural de alta resolubilidade, que facilitarão identificar genes candidatos e uma melhor compreensão do histórico evolutivo da *Capra hircus* no Nordeste brasileiro.

Palavras-chave: F_{ST} , genes candidatos, *Loci Outliers*, Marota, SNP.

Introdução

A cabra (*Capra hircus*) é provavelmente um dos primeiros ruminantes domesticados, fato ocorrido a aproximadamente 10 mil anos [1] e tornou-se um componente importante dos sistemas de produção agrícola e pecuária em todo o mundo, principalmente nas regiões áridas e semiáridas, onde se inseriu morfofisiologicamente como parte integrante do ambiente rústico de subsistência rural, com níveis de produção que vêm contribuindo significativamente para a segurança alimentar e para amenizar a situação de pobreza entre agricultores familiares de países em desenvolvimento [2, 3].

Conduzidos pelas migrações humanas e rotas comerciais, os caprinos se espalharam pelo mundo, inicialmente na Ásia, na Europa e na África e, a partir da Península Ibérica, foram levados à América. Essa distribuição foi favorecida pela versatilidade genética e grande capacidade adaptativa como a resistência a diferentes e adversas condições ambientais, expressando resistência a doenças e principalmente, adequação térmica e adaptabilidade nutricional [4, 5, 6].

Nas Américas, e, em particular, no Brasil, a partir do século XVI, onde os caprinos foram introduzidos pelos portugueses, franceses e holandeses, o sistema de manejo e as condições de ambiente quente a que foram submetidos, moldaram geneticamente e fenotipicamente esses animais, criando variação suficiente para tornar o Brasil com a maior diversidade de *C. hircus* da América do Sul, com 21 raças, segundo a Domestic Animal Diversity Information System [7].

Os caprinos passaram por processos adaptativos no Brasil durante vários séculos, originando as raças e grupos com padrão fenotípico uniforme, que são reconhecidos como naturalizados ou localmente adaptados [8]. Do ponto de vista de conservação da biodiversidade na espécie, estes animais são tratados como patrimônio cultural e também como patrimônio biológico de importância regional. Porém, as técnicas moleculares atuais podem comprovar se são únicos e, conseqüentemente, indicar riscos da perda de genes e de genótipos adaptativos que podem ser favoráveis para a produção animal, frente a possíveis mudanças climáticas globais, decorrentes da extinção desses grupos genéticos [9].

Na concepção conceitual de genética de populações, a frequência dos *loci* gênicos sofrem influências de forma sistemática de mutações, seleção e dos processos migratórios e de forma dispersiva pela seleção natural e deriva genética. Assim, é imperativo para os sistemas produtivos e conservacionistas, entender que proporção do genoma de uma raça ou quais genes estão sob ação dos fatores de seleção, principalmente da natural ou mesmo dos processos evolutivos ocasionais, para assim, quantificar o papel da adaptação na história evolutiva das espécies [10, 11, 12, 13].

Na prática, é crucial que seja identificado com precisão o genoma das populações naturalizadas de caprinos do país, para gerenciá-las com perspectivas de produção e conservação genética. Em ambas, o conhecimento do genoma é fundamental para a identificação de *loci* adaptáveis às condições ambientais [5], pois tanto o homem como o ambiente interferem muito na variação fenotípica e genética, principalmente sob condições adversas de temperatura e nutrição, comuns em regiões semiáridas, onde

genes específicos podem estar sob ação de seleção natural, detectáveis por metodologias disponíveis atualmente.

A esse respeito, o desenvolvimento de métodos e tecnologias de sequenciamento de genomas completos (GWAS), a partir de genotipagem com uso de painéis de polimorfismo de nucleotídeo único (SNP), facilitou o rastreamento da variabilidade genética das espécies, bem como identificar assinaturas adaptativas no genoma, além de detectar associações entre regiões genômicas e características de interesse. Com a automação, ocorreu a redução de custos, possibilitando o uso dessas metodologias em larga escala [14,15,6,16,17], e assim abrir caminhos para sua utilização na seleção genômica, além de proporcionar melhor entendimento sobre a história evolutiva dos caprinos, que se encontra mais atrasado que nas demais espécies de interesse zootécnico [17].

Em relação à estatística, uma medida de diferenciação genética usada é a abordagem F_{ST} -heterozigosidade, proposta por Beaumont e Nichols [18]. Esse enfoque, aplicado a dados moleculares, é um indicador de diferenciação genética que permite a comparação da estrutura populacional de diferentes organismos. É usado para identificar *loci* no genoma que se apresentam com diferenciação genética significativamente maior ou menor do que o esperado sob condições de equilíbrio, que são denominados de *loci outliers*, divergentes ou discrepantes, e estão possivelmente sob ação de seleção natural, sendo favorecidos ou não pelo ambiente [19, 20, 21, 13 e 22].

A estatística F_{ST} é um método muito utilizado, mas pode ocorrer um número elevado de *loci* falso positivo, por considerar os indivíduos agrupados em populações [23,24]. Uma alternativa tem sido combinar com abordagens baseadas em indivíduos, pois, com isso aumenta-se a confiabilidade nos estudos sobre assinaturas de seleção natural [25,13].

Existem vários métodos de varreduras do genoma baseados em indivíduos que utilizam técnicas relacionadas em análises multivariadas. O uso do PCAdapt baseado em componentes principais é uma alternativa plausível para validar ou identificar assinaturas de seleção [26].

Diante da exposição, objetivou-se, com o presente trabalho, buscar *loci outliers*, utilizando-se Chip SNP 50K, empregando diferentes metodologias e, associar regiões do genoma de *Capra hircus* a possíveis efeitos do processo de adaptação, assim como identificar genes candidatos que estão sob ação de seleção natural.

Material e Métodos

Declaração de ética

Todos os procedimentos realizados com os animais estão de acordo com normas do comitê de Ética e Pesquisa da Universidade Federal do Piauí, número de registro 058/14.

Amostragem

A pesquisa foi realizada com amostra de 96 indivíduos: do grupo genético Marota (n=86) (localmente adaptado), pertencentes a dois criatórios, (n=76) localizados no município Castelo do Piauí-PI (latitude 05°19'20" sul e longitude 41°33'09" oeste), e (n=10) no município de Elesbão Veloso-PI (latitude 06°12'07" sul e longitude 42°08'25" oeste). Os demais indivíduos da raça Anglonubiana (n=10) são de um rebanho localizado em Teresina-PI (05°02'39,95" sul, longitude 42°47'03,70" oeste), serviram como referência comparativa relativa à diversidade genética

Genotipagem

A genotipagem foi realizada utilizando o Chip SNPs caprino 50K, contendo 53.347 SNPs uniformemente espaçados nos cromossomos. O chip foi fabricado de acordo com a tecnologia Infinium da IlluminaTM, usando a plataforma iScan. O protocolo de genotipagem foi seguido conforme estabelecido pelo fabricante (disponível em www.illumina.com).

Para avaliação da qualidade dos dados genotípicos, utilizou-se o programa Plink versão 1.9 [27]. Foram considerados apenas os SNPs localizados em cromossomos autossômicos. Eliminaram-se das análises as amostras com *Call Rate* inferior a 80% e com Alelo de Menor Frequência (MAF) < 0,05 e que se mostraram em equilíbrio de Hardy-Weinberg, determinado pelo teste exato de Fisher com 1000 permutações e considerando $P > 0,05$.

Após a filtragem e o controle de qualidade conduzido nas populações, foram excluídas amostras que criteriosamente apresentavam: *Call Rate* < 90%, ou seja, apresentaram menos de 90% de genótipos determinados pelo painel de genotipagem; apresentar heterozigosidade acima de três desvios padrões em relação à média; genótipos idênticos (>99,5%) e erro de identificação do sexo, no caso de indivíduos identificados como machos apresentarem genótipos heterozigotos para marcadores no cromossomo X. O número de SNPs remanescentes, após o controle de qualidade foi 45.600 [28,29].

Determinação de regiões genômicas sob seleção

Para a detecção de *loci* sob seleção, a subpopulação Marota foi analisada. A Marota é um caprino típico da região semiárida, que está ameaçada de extinção [8] devido à constante introdução de material genético comercial na região. A conservação *in situ* da Marota tem-se prolongado durante décadas devido a projetos governamentais e também a parcerias privadas. Nesses rebanhos fechados (denominados Núcleos) não ocorre migração e nem seleção artificial direcional. Devido à estratégia utilizada *in situ*, a diversidade genética dentro de cada subpopulação é mantida sob efeito apenas da seleção natural [8].

Nesse sentido, uma maneira comum de identificar *loci*, supostamente sob seleção em estudos de genômica populacional, é investigar os que têm alta diferenciação F_{ST} em relação à sua heterozigosidade esperada no equilíbrio de Hardy-Weinberg [18]. O índice F_{ST} , corresponde à probabilidade de dois alelos

serem sorteados aleatoriamente, em uma população descendente de um alelo ancestral presente na mesma população. Para isso, combinaram-se abordagens diferentes e alternativas para dar maior confiabilidade aos estudos de se localizar regiões do genoma sob ação de seleção natural.

Então, a detecção de SNPs potencialmente sob seleção foi realizada através de dois métodos de investigação de *outliers*. Essa análise foi conduzida pelo método padrão do FDIST2, baseado no software ARLEQUIN v 3.5.2.1 [30] e no pacote fsthet R. Também, identificaram-se *loci outliers* em uma análise de componentes principais (PCA) usando o pacote do R, denominado de PCAdapt R [26].

Método FDIST2 padrão/fsthet-R

A metodologia FDIST2, proposta por Beaumont e Nichols [18], é utilizada para obter a distribuição nula de F_{ST} versus heterozigosidade esperada sob um modelo de ilha por simulações de coalescência, portanto essa distribuição muda de forma, com base na demografia da população.

Utilizou-se o programa fsthet-R, que calcula quantis suavizados a partir do conjunto de dados existentes na genotipagem para se identificarem *loci outliers* com valores extremos de F_{ST} em relação à sua heterozigosidade no equilíbrio de H-W [31]. Nesse caso, considera-se que os marcadores identificados como *outliers*, apresentam potencial adaptativo [19,20,22].

Método FDIST2/ Arlequin

Em outra abordagem, utilizou-se o método padrão FDIST2 [18], implementado no software ARLEQUIN 3.5 [30]. A metodologia aplicada frequentemente usa o modelo hierárquico de ilhas. Para cada *locus*, as frequências alélicas são usadas para calcular os valores de F_{ST} condicionados à sua heterozigosidade no EHW e para calcular os valores de P para cada *locus* [20, 23].

Para gerar a distribuição conjunta de F_{ST} versus heterozigosidade, realizaram-se 50.000 simulações coalescentes com 50 grupos de 100 demes. Com base no percentil de 0,01 foi construído o limite de intervalo de confiança de 99% para a distribuição esperada da relação do F_{ST} e heterozigosidade. Loci mostrando comportamento de diferenciação atípica (F_{ST}) e localizado fora da neutralidade foram detectados como *outliers* [20].

Método do PCAdapt (versão 2.0 do pacote R) [26]

Identificou-se o F_{ST} *outlier*, usando a abordagem do pacote PCAdapt R, que executa varreduras genômicas e detecta genes sob seleção, utilizando dados baseados em indivíduos e em técnicas de análise multivariadas, dentre elas a análise de componentes principais (PCA).

Inicialmente, o PCAdapt foi executado com um número de componentes principais $K=1$, o número K explicando a maior parte da variação, conforme recomendado pelos autores do software [26].

Em seguida, as distâncias de Mahalanobis foram usadas como teste estatístico para procurar SNPs discrepantes e transformados em valores de p para realizar múltiplos testes de hipóteses.

Finalmente, o corte para detectar *Loci outliers* foi escolhido com base no procedimento de valor

q implementado no pacote q -value R (R Core Team 2013), usando 0,01% como o limite da taxa de descoberta falsa.

A distribuição dos valores de p foi verificada usando: Um gráfico Quantil-Quantil dos valores p esperados versus valores de p observados e um histograma [26].

Taxa de Descobertas Falsas (FDR)

A taxa falsa-positiva, conceitualmente, corresponde ao número de *loci* neutros significativos (falsos positivos) dividido pelo número total de *loci* neutros testados. Da mesma forma, entende-se que o FDR pode ser definido pelo número de *locus* neutros falsos-positivos dividido pelo número total de resultados positivos [32]. Transformou-se os resultados em q -values para corrigir as comparações múltiplas.

Adiante, utilizou-se um corte de valor q 0,01 para definir um resultado positivo [33]. O valor q de um *locus* corresponde à proporção esperada de falsos-positivos entre todos os *locus* com valores de p iguais ou menores do que o *locus* observado. Nesse caso, entende-se que os *loci* que tem valores de q iguais a 0,01, devem ter um FDR esperado de 1% [33, 26], usando o q -values do pacote R (R Core Team 2013), que transforma os valores- p em valores- q .

Conteúdo gênico de regiões identificadas como sob seleção

Localizou-se os genes das regiões genômicas significativas usando o genoma de cabra de referência v1.0 (<http://www.ncbi.nlm.nih.gov/genome?term=capra%20hircus>). Fez-se a busca gênica para os marcadores SNPs localizados dentro de alguns genes, usando o Ensembl Comparative Genomics Resources e o banco de dados NCBI Gene. Em alguns casos, devido à anotação incompleta do genoma da cabra, a ferramenta BioMart de Ensembl (www.ensembl.org/biomart) foi usada para determinar os bovinos ortólogos (*Bos Taurus*), ovinos (*Ovis aries*), suínos (*Sus scroffa*) e humanos (*Homo sapiens*) e com IDs genéticos para cada gene detectado.

Fez-se uma pesquisa na literatura e no banco de dados de QTL de bovinos, suínos e ovinos (disponível online em <http://www.animalgenome.org/cgi-bin/QTLdb/index>) para identificar fenótipos conhecidos.

Resultados

Valor de F_{ST} global

O valor de F_{ST} igual a 0,16 é indicativo da existência de diferenciação genética entre o grupo genético Marota e a raça comercial Anglonubiana.

Loci sob seleção e identificação de genes sob seleção natural

Método FDIST2

O método do FDIST2, implementado no software fsthet R, detectou uma lista de 1761 *loci outliers* que apresentam um F_{ST} acima da zona de neutralidade, com média de 0.29 e uma heteroziguidade média de 0,35. Estes estão localizados em todos os cromossomos autossômicos, sendo o cromossomo 6 o que apresentou maior quantidade de *outliers*, conforme anexo 1 e Fig 1.

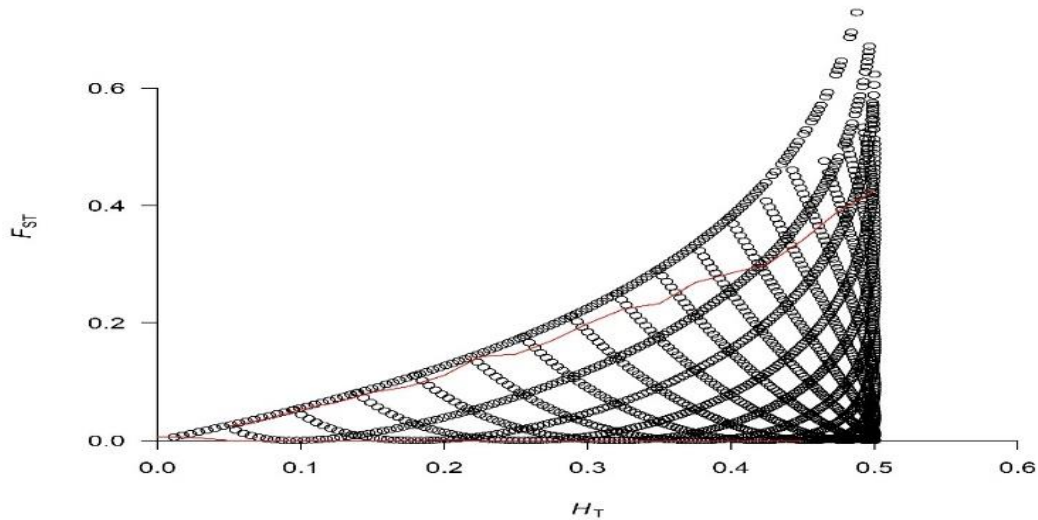


Fig 1. Cenário apresentado para os 1761 *loci outliers* que estão sofrendo ação de seleção natural direcional ou disruptiva, pontos acima da linha vermelha. Os demais marcadores são *loci* neutros.

Como o pacote fsthet do R, identificou muitos *loci outliers*, o software Arlequin, por ter uma estatística mais precisa foi utilizado, pois associa um p-valor aos *loci outliers*, no qual foi escolhido um alfa de 0.01 para detectar os SNPs *outliers*. Nesse caso, foram identificados 14 marcadores *outliers*, conforme Tab 1.

Tabela 1. Regiões genômicas sob seleção e lista de genes localizados na região dos 14 SNPs mais significativos (com base no P-valor 0,01 detectado pela metodologia FDIST2 do Arlequin).

Cromossomo	SNP outlier	P-value	Distribuição	Genes
06	snp31742-scaffold354-420831*	0,006	420831	Região intergênica
	snp24457-scaffold248-1468420*	0,006	74829520	Região intergênica
08	snp38279-scaffold4816-76898*	0,006	94632773	Região intergênica
	snp33341-scaffold391-3245653*	0,006	3665139	Região intergênica
	snp33316-scaffold391-2266280*	0,006	4644512	GALNTL6
09	snp43771-scaffold588-1668067*	0,005	14724319	NKAIN2
	snp36869-scaffold447-3340669*	0,009	80064765	Região intergênica
11	snp15213-scaffold1621-1365230*	0,006	58948520	LRRTM4
15	snp53215-scaffold801-192969*	0,007	9907114	Região intergênica
	snp41670-scaffold542-11277*	0,007	47407147	OTOG
21	snp18963-scaffold191-316399*	0,007	44518722	CFL2
25	snp44135-scaffold6-477013*	0,005	27505380	SLC5A2, AHSP, ARMC5, LOC 102180368
	snp57108-scaffold909-714756*	0,006	4282276	LOC102181181
26	snp55604-scaffold862-687700*	0,006	26426028	Região intergênica SORCS3

P-valor, distribuição e genes presentes nas regiões dos marcadores SNPs;

*sobreposição em duas abordagens metodológicas;

Abordagem do PCAdapt R

Tendo como referência, a origem de duas populações de caprinos analisadas, o primeiro componente principal (PC), agrupou geneticamente os indivíduos (**Fig 2**).

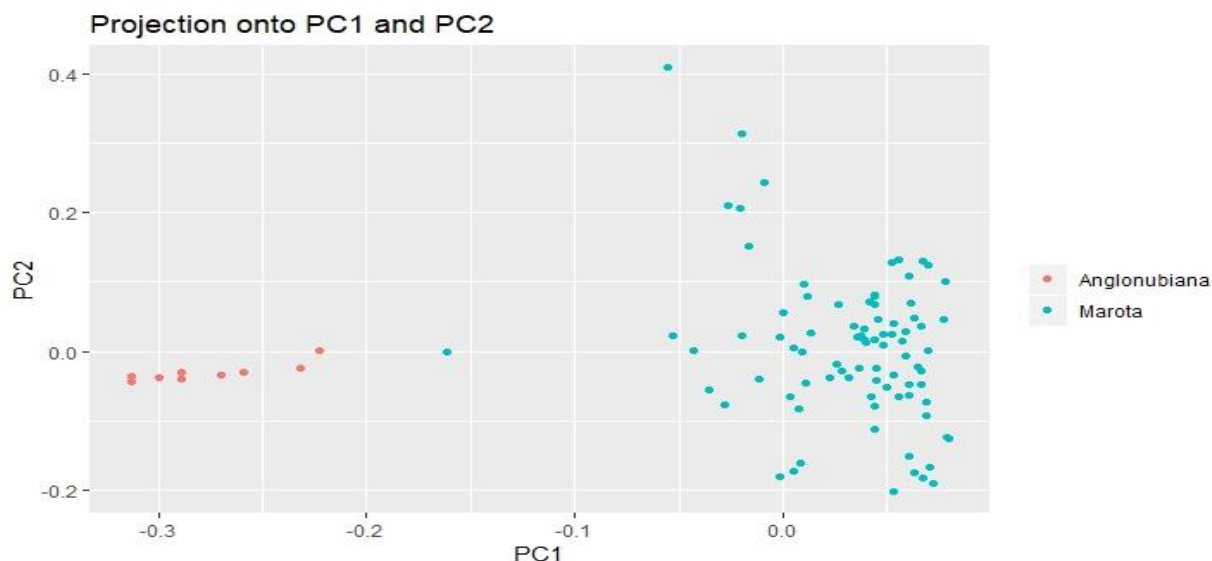
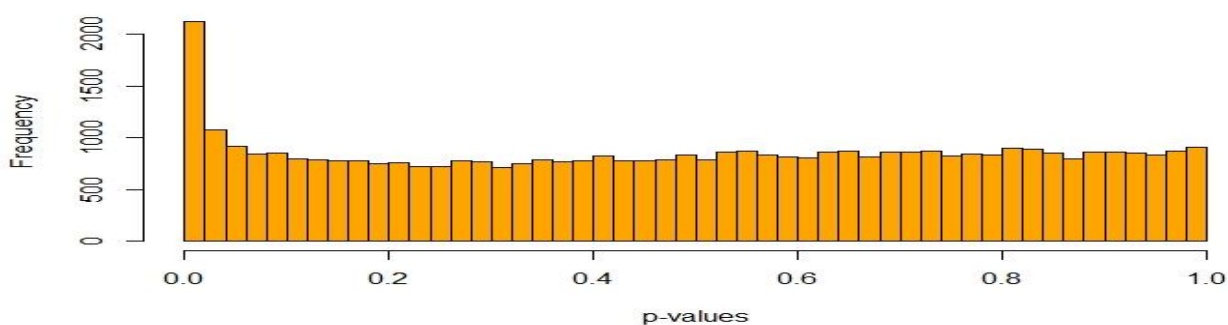


Fig 2. Dispersão dos animais em plano cartesiano relação aos dois primeiros escores dos componentes principais.

Em seguida, foi construído um histograma, no qual é perceptível que os valores de p estão bem calibrados, pois estão distribuídos de maneira uniforme. No entanto, fica evidente, na distribuição, que existe um pico em torno de 0, o que constitui um forte indício da presença de *loci outliers* (**Fig 3A**).

Para os marcadores SNPs outliers analisados, foi apresentado um gráfico quantil–quantil, que é normalmente usado para verificar o ajuste da distribuição da frequência dos dados a uma distribuição de probabilidades [34]. Nesse sentido, o conjunto de dados amostrais foi aproveitado em análises do PCAdapt R para detectar SNPs outliers. Então, a análise do gráfico Q-Q dos p-valores esperados e observados indica que a maioria dos valores de p segue a distribuição uniforme esperada, com níveis mais baixos de significância (0,01) enquanto os p-valores (0,01) são menores que o esperado, portanto, confirma-se a presença de loci outliers. Com base na taxa de falsa descoberta (FDR) de 0,01, quatorze SNPs foram detectados como outliers (**Fig 3B**).

(A)



(B)

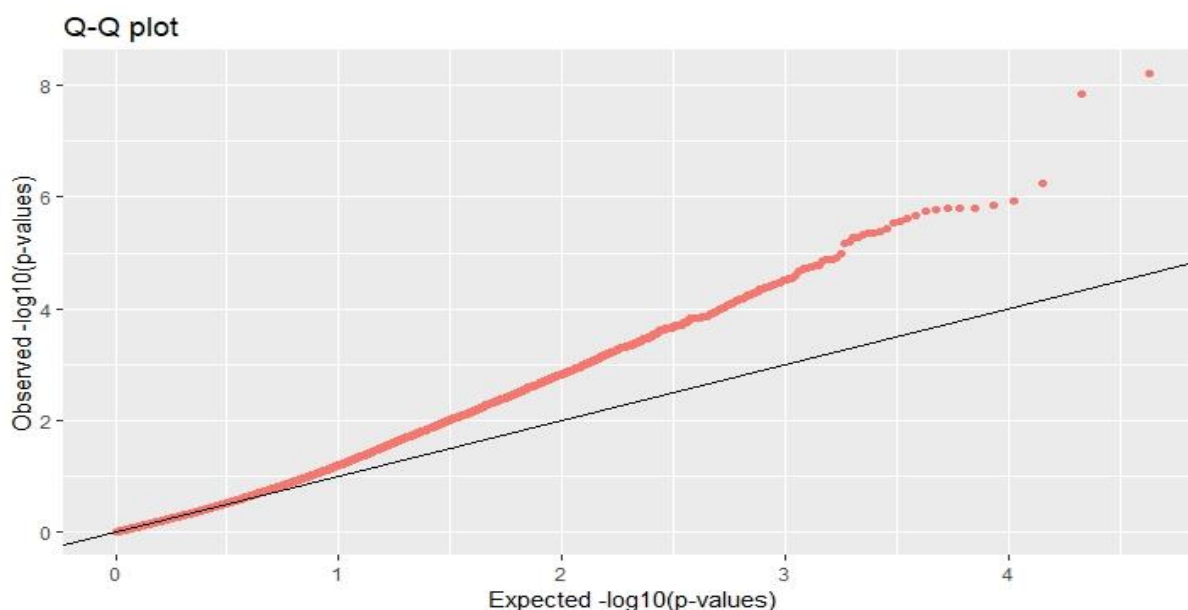


Fig 3. (A) Distribuição dos valores de P baseados no histograma e (B) no gráfico Q-Q

Sob a metodologia baseada em Componentes Principais (PCAdapt), foram detectados 1372 marcadores, localizados acima da zona de neutralidade, com um p-valor maior que 2,0 e sem correção FDR. Portanto, os marcadores *outliers* foram identificados acima da linha de normalidade (**Fig 4**). É fundamental ressaltar que, para as duas metodologias, os resultados abrangem todos os 29 cromossomos autossomos, com o número de SNPs significativos variando de 11 e 13 (cromossomo 27) a 112 e 169 (cromossomo 06) (**Figs 3 e 4**).

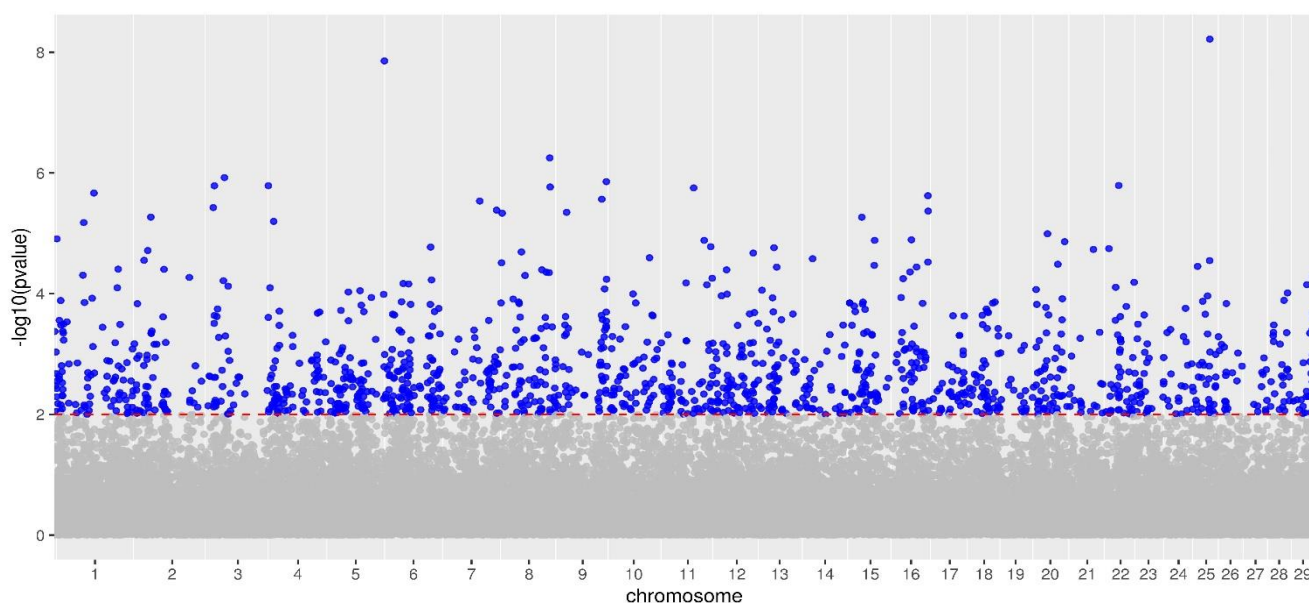


Fig 4. Cenário para distribuição dos *loci outliers* sem correção FDR/ PCAdapt

Devido ao grande número de marcadores SNPs *outliers*, localizados nas diversas regiões cromossômicas do grupo genético Marota, surgiu a necessidade de se obterem dados mais confiáveis e robustos para os cenários que envolvem a busca de *loci outlier*. Assim, aplicou-se FDR com q-valor de

0,01, nesse caso, identificaram-se 14 marcadores SNPs, ou seja, com possibilidades de estarem sob ação de seleção (**Fig 5**) e com maiores detalhes na **Tab 2**.

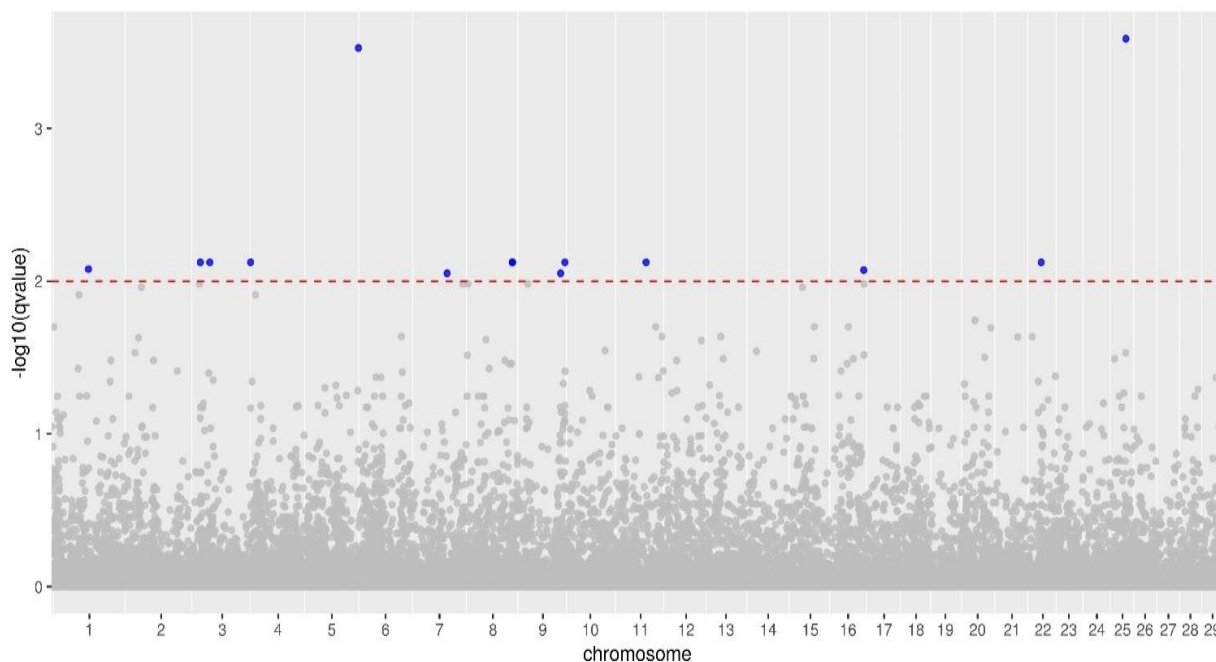


Fig 5. Cenário para distribuição dos *loci outliers* com correção de 0,01 de FDR/ PCAadapt.

Tab 2. SNPs outliers sob seleção natural mostrando níveis significativos (FDR 0,01) e lista de genes localizados na região dos 14 SNPs detectados pelo PCAadapt R.

Cromossomo	SNP outlier	Posição	P-value	Genes
01	snp36295-scaffold4351695548*	74579244	0,044	FGF12, UTS2B, OSTN, CCDC50
03	snp6320-scaffold1223-315475*	12096695	0,086	Região intergênica
	snp25340-scaffold261 1087491*	31688141	0,088	DAB1, AMPD2, LOC102185621, LOC100861222
04	snp46417-scaffold641-1278697*	1278697	0,098	DNAJB6, UBE3C
06	snp59993-scaffold999-426450*	1736025	0,127	MARCH1, TMA16, NPY1R, NPY5R, TKTL2,
07	snp41464-scaffold54-493580*	68458023	0,158	Região intergênica
08	snp38279-scaffold4816-76898**	94632773	0,182	Região intergênica
	snp59201-scaffold972-291675*	95443929	0,182	Região intergênica
09	snp36869-scaffold447-3340669**	80064765	0,204	Região intergênica
	snp8962-scaffold1324-311436*	89207416	0,206	Região intergênica
11	snp18060-scaffold18510803188	62255976	0,238	PELI1
16	snp52580-scaffold785-203051*	71615383	0,318	HHAT
22	snp21240-scaffold2073-421602*	25482065	0,400	Região intergênica
25	snp44135-scaffold6-477013**	27505380	0,445	SLC5A2, AHSP, ARMC5, LOC 102180368, LOC102181181

P-valor, distribuição e genes presentes nas regiões dos marcadores SNPs.

*Sobreposição em duas abordagens metodológicas.

**Sobreposição em três abordagens metodológicas.

Mapeando Regiões selecionadas positivamente para anotações do genoma

Vários *loci outliers*, em diferentes regiões cromossômicas, foram detectadas no genoma dos caprinos Marota e Anglonubiana. A princípio, percebe-se que essas regiões contêm genes que estão

envolvidos em importantes processos biológicos adaptativos e/ou produtivos, como produção carne ou fibras, tamanho corporal e controle do equilíbrio hidrossalino. Essas regiões candidatas estão distribuídas nos diferentes cromossomos autossômicos (**Tab 1 e 2**) [13, 35].

As regiões genômicas identificadas (**Tab 1 e/ou 2**) possuem marcadores *outliers* localizados dentro de sequências gênicas, o que constitui um forte indício de que esses genes estejam sob ação de seleção natural. Dentre os genes identificados, destacam-se: OTOG, CFL2, LRRTM4, NKAIN2, GALNTL6, LOC102183030, HHAT, PELI1, DAB1, FGF12, DNAJB6, SORCS3. Além desses genes, as análises realizadas, detectaram nas regiões próximas, *loci* gênicos essenciais à adaptação dos caprinos ao semiárido, incluindo-se nesses: UTS2B, OSTN, SLC5A2.

Discussão

Os caprinos introduzidos no Brasil foram criados, por muito tempo, para diversos propósitos e, gradualmente, acumularam traços fenotípicos adaptativos, tornando-se bem ajustados aos ambientes tropicais. Então, entender os mecanismos genéticos e evolutivos responsáveis por essas características, deve ser um princípio norteador para elaboração de programas de conservação e melhoramento genético animal. Portanto, o desenvolvimento das tecnologias de sequenciamento de genomas completos e o uso de Chip de SNP de média densidade permitem a compreensão dessa diversidade genética populacional [5].

Análise do F_{ST} global

Tendo como referência os resultados de outros trabalhos apresentados, constatou-se maior diversidade genética entre animais naturalizados e raças comerciais do que comparando animais de raças comerciais entre si (Grasso et al., 2014) [36]. Nesse aspecto, tendo um F_{ST} global de 0,16 e a raça Anglonubiana como referência, existe um forte indício de que o grupo genético Marota apresenta uma considerável diversidade genética [37].

Análise dos componentes principais

Para detecção de *loci outliers* no genoma, o pacote do R, o PCAdapt tem uma vantagem adicional em relação ao método do F_{ST} *outlier* do FDIST2, pois analisa a população, sem a necessidade de subdividi-la, ou seja, a análise dos dados é baseada em indivíduos [38].

A princípio, as análises de estruturação genética, baseadas nos componentes principais, revelaram que o número de agrupamento que explica com mais clareza os dados da variação genômica, entre o grupo genético Marota e raça e o da Anglonubiana, foi de $K=2$ para análise de marcadores SNPs. Nesse sentido, quando as subpopulações são representadas no plano cartesiano, constituído pelos dois primeiros componentes principais, observou-se uma separação nítida entre os grupos de animais (**Fig 2**), com um indivíduo dos grupos distintos proximamente relacionado geneticamente, o que indica haver participação genética da raça Anglonubiana no grupo naturalizado Marota.

Na perspectiva da distribuição do p-valor no histograma e no gráfico quantil-quantil, a uniformidade apresentada demonstra calibração, pois os mesmos estão bem distribuídos. No entanto, a presença de *outliers* no pico em torno do zero é um forte indicador de *loci* sob ação de seleção [39].

Determinação de regiões genômicas sob seleção

Os resultados das tabelas 1 e 2, confirmam achados anteriores para populações de cabras, em especial para os aspectos adaptativos, como a termotolerância, equilíbrio osmótico e processos de produção animal [13,35]. Em relação às diferentes abordagens de detecção de *outliers*, houve sobreposição entre os marcadores presentes nas regiões cromossômicas identificadas pelos métodos de modelo de ilha hierárquica baseada na metodologia do FDIST2, implementada no Arlequin e no fsthet R, que identifica *loci outliers* com valores extremos de F_{ST} em relação à heterozigosidade e do PCAdapt, que detecta genes sob seleção, utilizando dados genômicos baseados em indivíduos e em técnicas de análise multivariada, dentre elas, a análise de componentes principais (PCA). Portanto, a metodologia usada foi eficiente para buscar no genoma caprino *loci* sob ação de seleção.

Há expectativa de que a adaptação aos ambientes tenha moldado a diversidade genética dos grupos nativos ou naturalizados de caprinos na região semiárida, dentre eles destacando-se a Marota. Da mesma forma, espera-se identificar evidências dos diferentes tipos de seleção natural desse grupo, como respostas ao processo migratório, a seleção natural e a deriva genética. Então, identificar o número e a localização dos *loci* envolvidos no histórico adaptativo local dos caprinos é um ponto crucial e inicial para compreender e caracterizar a sua base genômica evolutiva [20,13,16,22].

Para esse fim, foi fundamental a utilização recente dos marcadores SNPs que, devido a sua distribuição ampla nas diversas regiões do genoma, podem ser localizados em regiões codificadoras ou com funções regulatórias. Portanto, os SNPs fornecem oportunidades para se identificarem variantes genéticas sob seleção ou assinaturas genômicas de seleção. Assim, esse marcador molecular permite localizar *loci outliers* ou de valores discrepantes entre populações, como medido pela estatística F_{ST} e em diferentes abordagens [13,16,22].

Para compreender o uso de SNPs como marcador para se identificar *loci* discrepantes ou *outliers* no genoma, vale ressaltar que esse marcador está associado a outros SNPs no cromossomo, no qual ocorreu a mutação. Nesse sentido, o conjunto específico de SNPs em um cromossomo denomina-se haplótipo. Assim, os marcadores SNPs estão ligados e são herdados juntos. Então, quando esse marcador está fisicamente próximo a um *locus* (adaptativo ou não), ele tenderá a ser herdado juntamente com o mesmo, portanto, se um alelo em um haplótipo for favorecido pela seleção natural, esse alelo terá maior frequência na população [40, 41].

Para evitar que um SNP seja erroneamente identificado como *outlier*, as evidências do processo de adaptação por seleção foram investigadas por meio de diferentes abordagens. Empregou-se um dos

métodos de detecção de *loci outliers* baseado em análise dos componentes principais (PCA). A grande vantagem dessa abordagem reside no fato de não agrupar os indivíduos em populações e usar técnicas relacionadas de análises multivariadas [26]. Há achados que mostram, de forma evidente, que a diminuição do número de amostras tem efeito insignificante na identificação dos resultados positivos confiáveis ao utilizar abordagens multivariadas [22,42]. Além disso, houve resultados práticos expressos em simulações que envolviam apenas duas populações, portanto, as demais abordagens são metodologicamente parecidas e utilizam o modelo de ilha hierárquica implementado no FDIST2 [31,43].

Nas duas metodologias deste estudo, houve sobreposições de alguns marcadores distribuídos nos cromossomos autossômicos das populações amostradas. Neste caso, em decorrência do histórico evolutivo da Marota, pode ter ocorrido gargalo populacional ou mesmo expansão populacional, que poderia resultar em excesso de alelos raros [37]. No entanto, a identificação de picos, tais como os localizados nos cromossomos 06 e 25, que foram identificados por mais de um método, são excelentes indicadores de *loci* sob ação de seleção natural, ou seja, dificilmente se caracterizariam como falsos-positivos [13].

Nesse sentido, os *loci outliers* mostraram padrão de distribuição diferente no genoma, com localização que abrange diversas regiões em todos os cromossomos autossômicos, o que caracteriza o processo de ligação gênica. Dentre essas regiões, os cromossomos de números 01, 04, 06 e 08, respectivamente, destacaram-se, com quantidade superior de *loci outliers*. Nesse sentido, fica documentado que os marcadores para diversos genes não estão localizados de maneira uniforme pelo genoma e, com isso, podem ser agrupados em regiões com significado relevante para traços fenotípicos produtivos e/ou adaptativos, relatados em outros estudos [13,36].

Trabalhos de independentes autores, documentaram regiões do genoma da *C. hircus* e identificaram genes responsáveis pela produção de algum componente de fibras da pelagem, do leite e da carne [13,36]. No presente estudo, os achados também identificaram genes de importância. Por exemplo, a região gênica do cromossomo 06, corrobora as informações contidas em outro estudo [13]. Assim, a existência dessas regiões, sob seleção no genoma do grupo genético Marota e da raça Anglonubiana, fornece oportunidades para determinar características adaptativas e até produtivas, que podem estar correlacionadas ao modo de vida do animal caprino nas regiões tropicais semiáridas do Nordeste brasileiro, fato documentado em trabalhos em caprinos de regiões semiáridas [44].

Descrição da função biológica prevista do *loci outlier*

Genes candidatos que foram encontrados dentro das regiões genômicas mais significativas, foram avaliados para entendimento funcional em processos e funções biológicas e moleculares específicas. Então os genes FGF12 (fibroblast growth fator 12), AMPD2 (adenosine monophosphate deaminase 2), OSTN (osteocrin), expressam proteínas que influenciam no tamanho corporal e o desenvolvimento esquelético e embrionário [43,13,35]. O gene SLC5A2 (solute carrier family 5 member

2) codifica proteína que atua como transportador de glicose e sais minerais nos rins [45]. O gene UTSB2 (urotensina 2), atua como potente vaso constritor [46].

Sabe-se que o tamanho reduzido e a formação corporal do caprino Marota estão intimamente relacionados com a capacidade de regulação eficiente da temperatura e isso explica o tamanho corpóreo pequeno dos vários grupos genéticos naturalizados nas regiões áridas e semiáridas, onde as pastagens são geralmente de baixo valor nutritivo e as temperaturas são elevadas, pois estes processos são controlados por genes sob ação de seleção natural, como os genes FGF2, OSTN [45, 47, 37,13]. Neste estudo, destaca-se que a característica de crescimento dos caprinos é influenciada pela seleção artificial direcional, em oposição ao processo de seleção natural em ambientes tropicais.

Diferentemente das populações estudadas por outros autores [13,43], cabe a este trabalho realizar a discussão de *loci outliers* ligados à adaptação ao ambiente tropical. Para tanto, destacam-se os *loci* SLC5A2 (solute carrier family 5 member 2) que codifica a proteína que atua como transportadora de glicose e sais minerais nos rins [45] e o *loci* UTSB2 (urotensina 2) que atua como potente vaso constritor [46]. Muito provavelmente esses *loci* atuam regulando o equilíbrio osmótico dos caprinos nas áreas áridas e semiáridas, onde os animais precisam sobreviver com pouco recurso hídrico.

Toda essa diversidade de *loci outliers* mostrou forte evidência de seleção, mesmo na ausência de um sinal de associação com traços fenotípicos conhecidos. Os *loci* identificados podem estar controlando características para as quais não existe anotação dos genes, em virtude da anotação incompleta do genoma dos caprinos, mesmo se utilizando ortólogos da *Capra hircus* [13].

Conclusões

Considerando esta pesquisa para os caprinos (*Capra hircus*) sob seleção natural nos trópicos, pode-se concluir que:

- Os marcadores SNPs permitem criar uma robusta fundamentação conceitual e mapas de seleção natural de alta resolubilidade, que facilitarão identificar genes candidatos e uma melhor compreensão do histórico evolutivo da espécie;
- As metodologias de detecção de *loci outliers* foram eficientes e complementares na elucidação de regiões genômicas dos caprinos pesquisados;
- Com essa metodologia, foi possível identificar, no processo evolutivo ao ambiente semiárido, traços fenotípicos relacionados ao crescimento corporal;
- O balanço hídrico e as características que privilegiam o menor consumo de água nesses pequenos ruminantes foram também revelados como relevantes no histórico evolutivo dos caprinos nas regiões tropicais.

Agradecimentos

Fapepi- Fundação de Apoio à Pesquisa do Estado do Piauí

Referências

1. Zeder MA, Hesse B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science*. 2000; 287:2254–2257.
2. Lin BZ, Kato T, Kaneda M, Matsumoto H, Sasazaki S, Mannen H. Genetic diversity and structure in Asian native goat analyzed by newly developed SNP markers. *Animal Science Journal*. 2013;84: 579–584. doi:10.1111/asj.12039
3. Periasamy K, Vahidi SMF, Silva P, Faruque MO, Naqvi AN, Basar M, et al. Mapping molecular diversity of indigenous goat genetic resources of Asia. *Small Ruminant Research*. Elsevier B.V.; 2017;148: 2–10. doi:10.1016/j.smallrumres.2016.12.035
4. Aziz MA. Present status of the world goat populations and their productivity. *Lohmann Information*. 2010;45: 42–52.
5. Benjelloun B. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. 2015;6. doi:10.3389/fgene.2015.00107
6. Amills M, Capote J, Tosser-Klopp G. Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces. *Animal Genetics*. 2017;48: 631–644. doi:10.1111/age.12598
7. Dad-is. Domestic Animal Diversity Information System. 2011. Disponível em <http://www.fao. Org/dad-is>. Acesso em: 20 julho 2018.
8. Araújo AM de, Beffa LM, Jacob M, Almeida O, Abreu P, Cavalcante DH, et al. Crescimento e Mortalidade em um Rebanho de Conservação de Caprinos Marota no Brasil1. *Rev Cient Prod Anim*. 2009;11: 103–109.
9. Barros EA, Ribeiro MN, Almeida MJO, Araújo AM. Estrutura populacional e variabilidade genética da raça caprina Marota. *Archivos de Zootecnia*. 2011;60: 543–552. doi:10.4321/S0004-05922011000300043
10. Cavalli-Sforza, L. L. 1966. Population structure and human evolution. *Proc. Roy. SOC., London, B*. 164: 362–379.
11. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*. 1973;74: 175–195.
12. Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, et al. Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics*. 2010;186: 241–262. doi:10.1534/genetics.110.117275
13. Brito LF, Kijas JW, Ventura V, Sargolzaei M, Laercio R. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genetics*. 2017;186: 241–262. doi:10.1186/s12864-017-3610-0
14. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Research*. 2005;15: 1566–1575. doi:10.1101/gr.4252305
15. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and characterization of a 52K SNP chip for goats. *PLoS ONE*. 2014;9. doi:10.1371/journal.pone.0086227
16. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Lowry DB, Poss ML, et al. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *HHS Public Access*. 2017;188: 379–397. doi:10.1086/688018.Finding
17. Qiao H, Escobar LE, Saupe EE, Ji L, Soberón J. Using the KDE method to model ecological niches: A response to Blonder et al. (2017). *Global Ecology and Biogeography*. 2017;26: 1076–1077. doi:10.1111/geb.12610

18. Beaumont MA, Nichols RA. Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society B: Biological Sciences*. 1996;263: 1619–1626. doi:10.1098/rspb.1996.0237
19. Beaumont MA. Adaptation and speciation: What can F_{ST} tell us? *Trends in Ecology and Evolution*. 2005;20: 435–440. doi:10.1016/j.tree.2005.05.017
20. Pariset L, Joost S, Marsan PA, Valentini A, Abo-Shehadeh M, Jamil AT, et al. Landscape genomics and biased F_{ST} approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics*. 2009;10: 1–8. doi:10.1186/1471-2156-10-7
21. Feng XJ, Jiang GF, Fan Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiangsu*. *Scientific Reports*. Nature Publishing Group; 2015;5. doi:10.1038/srep13758
22. Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, et al. The search for loci under selection: trends, biases and progress. *Molecular Ecology*. 2018;27: 1342–1356. doi:10.1111/mec.14549
23. Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. *Heredity*. Nature Publishing Group; 2009;103: 285–298. doi:10.1038/hdy.2009.74
24. Bierne N, Roze D, Welch JJ. Pervasive selection or is it.? Why are F_{ST} outliers sometimes so frequent? *Molecular Ecology*. 2013;22: 2061–2064. doi:10.1111/mec.12241
25. Simianer H, Qanbari S. Statistical Problems in Livestock Population Genomics. Conferência. 2014.
26. Luu, K, Bazin, E, Blum MGB. PCAdapt : an R package to perform genome scans for selection based on principal component. *Molecular Ecology Resources*. 2017;17: 67–77.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007;81: 559–575. doi:10.1086/519795
28. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris P, Zondervan KT. Europe PMC Funders Group Data quality control in genetic case-control association studies. *Nature Protocols*. 2011;5: 1564–1573. doi:10.1038/nprot.2010.116.Data
29. Kijas JW, Ortiz JS, McCulloch R, James A, Brice B, Swain B, et al. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Animal Genetics*. 2013;44: 325–335. doi:10.1111/age.12011
30. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010;10: 564–567. doi:10.1111/j.1755-0998.2010.02847.x
31. Flanagan SP, Jones AG. Constraints on the F_{ST} -Heterozygosity Outlier Approach. *Journal of Heredity*. 2017;108: 561–573. doi:10.1093/jhered/esx048
32. Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*. 2014;23: 2178–2192. doi:10.1111/mec.12725
33. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *PNAS*. 2003;100: 9940–9945.
34. Settles M, Zanella R, McKay SD, Schnabel RD, Taylor JF, et al. (2009) A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. paratuberculosis infection status in US holstein cattle. *Anim Genet* 40: 655–662.
35. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. Signatures of selection and environmental adaptation across the goat genome post - domestication. *Genetics Selection Evolution*. BioMed Central; 2018; 1–24. doi:10.1186/s12711-018-0421-y

36. Grasso, A.N.; Goldberg, V.; Navajas, E.A.; Iriarte, W.; Gimeno, D.; Aguilar, I.; Medrano, J.F.; Rincón, G.; Ciappesoni, G. Genomic variation and population structure detected by single nucleotide polymorphism arrays in Corriedale, Merino and Creole sheep. *Genetics and Molecular Biology*, v.37, p.389-395, 2014.
37. Moura JO, Campelo JE., Bajay MM, Costa M., Cavalcante DH, Araújo AM. Diversidade Genética Em Caprinos Localmente Adaptados no Brasil Utilizando o Beadchip 50K. *Actas Iberoamericanas de Conservación Animal*. 2015;6: 92–97. doi:10.13140/RG.2.1.3239.2161
38. Duforet-Frebourg N, Bazin E, Blum MGB. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*. 2014;31: 2483–2495. doi:10.1093/molbev/msu182.
39. Moravčíková N, Simčič M, Mészáros G, Sölkner J. Genomic Response to Natural Selection within Alpine Cattle Breeds. *Czech Journal of Animal Science*. 2018;2018: 136–143. doi:10.17221/62/2017-CJAS
40. Pierce B. *Genética: um enfoque conceitual*. 3.ed. Rio de Janeiro: Guanabara Koogan, 2011.
41. Sanders MF, Bowman JL. *Análise Genética: uma abordagem integrada*. São Paulo: Pearson Education do Brasil, 2014.
42. Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*. 2018;27: 2215–2233. doi:10.1111/mec.14584
43. Willing E, Dreyer C, Oosterhout C Van. Estimates of Genetic Differentiation Measured by F_{ST} do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLoS ONE*. 2012;7: 1–7. doi:10.1371/journal.pone.0042649
44. Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*. Nature Publishing Group; 2016;116: 255–264. doi:10.1038/hdy.2015.94
45. Kaji K, Nishimura N, Seki K, Sato S, Saikawa S, Nakanishi K. Attenuates liver cancer cell growth and angiogenic activity. *International Journal of Cancer*. 2018;142: 1712–1722. doi:10.1002/ijc.31193
46. Debiec R, Christofidou P, Denniff M, Bloomer LD, Bogdanski P, Wojnar L, et al. Urotensin-II System in Genetic Control of Blood Pressure and Renal Function. 2013;8: 6–13. doi:10.1371/journal.pone.0083137
47. Mcmanus C, Paiva S, Corrêa PS. *Estatísticas para descrever Genética de Populações*. INCT: Informação Genético-Sanitária da Pecuária Brasileira. 2011; 1–50.

4 CONSIDERAÇÕES FINAIS

Nos últimos 10 anos, inovações metodológicas envolvendo o marcador molecular SNP ocorreram e foram utilizadas em estudos genômicos e de estrutura populacional na caprinocultura, geraram informações aplicadas nos estudos de diversidade genética e produção animal. Nesta revisão, a partir de 2009 verificou-se o interesse crescente dos pesquisadores pela aplicação dos SNP na caprinocultura, com destaque para China, a Índia e a Itália, que despontaram com maior número de publicações, com prevalência de estudos voltados para diversidade genética populacional, polimorfismo das proteínas lácteas e aspectos reprodutivos.

Nesse aspecto, os marcadores moleculares SNP, se mostraram úteis e eficientes para se determinar o conhecimento da variabilidade genética do caprino Marota, pois permitiram criar uma fundamentação conceitual e mapas de seleção natural, que facilitarão identificar genes candidatos e uma melhor compreensão do histórico evolutivo da espécie. Também, as metodologias de detecção de *loci outliers* foram eficientes e complementares na elucidação de regiões genômicas dos caprinos pesquisados, pois alguns traços fenotípicos relacionados ao crescimento corporal e ao equilíbrio hidrossalino foram identificados, o que se caracteriza como relevantes no histórico adaptativo e evolutivo desse grupo genético.

Estudos adicionais são necessários para validar a identificação de *loci outliers* encontrados neste estudo, para que sejam introduzidas nos programas de melhoramento e conservação da diversidade dos caprinos, visando o melhor entendimento dos mecanismos biológicos e da arquitetura genética destas características.

Nesse contexto, a utilização de programas de conservação dos recursos genéticos, mantidos por diferentes centros de pesquisas como as universidades públicas e privadas, acrescidas de parceria com criadores, podem gerar tecnologias e informações de fundamental importância para a pecuária, pois recursos genéticos animais são essenciais para a segurança alimentar, redução das desigualdades sociais, erradicação da fome e geração de emprego e renda em diversas partes do mundo.

5. APÊNDICES

5.1 Tabela que informa a localização cromossômica dos marcadores SNP analisados nas populações Marota e Anglonubiana.

Tabela 3. Resumo dos marcadores SNP analisados e localização nos cromossomos autossomos de *C. hircus*. Nas metodologias de F_{ST} outliers e PCAdapt.

Cromossomo	Contagem F_{ST} outliers	% (N=1731)	Contagem PCAdapt outliers	% (N=1372)
1	106	6,01930721	88	6,413994169
2	82	4,6564452	56	4,081632653
3	61	3,46394094	35	2,551020408
4	103	5,84894946	70	5,102040816
5	82	4,6564452	73	5,320699708
6	169	9,59681999	112	8,163265306
7	47	2,6689381	50	3,644314869
8	92	5,22430437	77	5,612244898
9	58	3,29358319	50	3,644314869
10	80	4,54287337	53	3,862973761
11	69	3,91822828	66	4,810495627
12	73	4,14537195	69	5,029154519
13	54	3,06643952	52	3,790087464
14	43	2,44179443	34	2,478134111
15	54	3,06643952	56	4,081632653
16	52	2,95286769	54	3,935860058
17	59	3,35036911	37	2,696793003
18	54	3,06643952	48	3,498542274
19	19	1,07893242	16	1,166180758
20	48	2,72572402	51	3,717201166
21	50	2,83929585	19	1,38483965
22	62	3,52072686	46	3,352769679
23	21	1,19250426	22	1,603498542
24	42	2,38500852	23	1,67638484
25	49	2,78250994	32	2,332361516
26	25	1,41964793	17	1,239067055
27	13	0,73821692	11	0,801749271
28	42	2,38500852	24	1,749271137
29	42	2,38500852	27	1,967930029
0	9	0,51107325	2	0,145772594
Total	1.731	100	1.372	100

5.2 Regiões genômicas sob seleção e lista de genes localizados na região dos 14 SNPs mais significativos

Tabela 4. Regiões genômicas sob seleção localizados na região dos 14 SNPs mais significativos (com base no P-valor 0,01 e F_{ST} detectados pela metodologia do FDIST2 do Arlequin).

Cromossomo	SNP outlier	P- value	F_{ST}	Distribuição
06	snp31742-scaffold354-420831	0,0062949653	0,82568939	420831
	snp24457-scaffold248-1468420	0,0063936756	0,051949921	74829520
	snp38279-scaffold4816-76898*	0,0061357017	0,85323761	94632773
08	snp33341-scaffold391-3245653	0,0062949653	0,82568939	3665139
	snp33316-scaffold391-2266280	0,0065023118	0,83521023	4644512
09	snp43771-scaffold588-1668067	0,0058954371	0,82663897	14724319
	snp36869-scaffold447-3340669*	0,0092420383	0,79529459	80064765
11	snp15213-scaffold1621-1365230	0,0063936756	0,051949921	58948520
15	snp53215-scaffold801-192969	0,0076511185	0,81502999	9907114
	snp41670-scaffold542-11277	0,0076511185	0,81502999	47407147
21	snp18963-scaffold191-316399	0,0076511185	0,81502999	44518722
25	snp44135-scaffold6-477013*	0,0056004485	0,85584861	27505380
	snp57108-scaffold909-714756	0,006294965	0,82568939	4282276
26	snp55604-scaffold862-687700	0,0063936756	0,051949921	26426028

6. ANEXOS

6.1 Rotinas do Programa R utilizadas nesse estudo

Rotina do Arlequin

```
Library (fsthet)
gpop<-my.read.genepop("cap2.gen")

gpop<-gsub()

fsts<-calc.actual.fst(gpop)
plot(fsts$Ht, fsts$Fst,xlab="Ht",ylab="Fst",pch=19)
write.table(fsts, file = "fsts.txt")
quant.out<-fst.boot(gpop, bootstrap = T)
quant.list<-ci.means(quant.out[[3]])
jpeg("out.jpg", width=7, height=7, units="in", res=600)
plotting.cis(df=fsts,ci.df=quant.list,make.file=F)
dev.off()
outliers<-find.outliers(fsts,boot.out=quant.out)
write.table(outliers, file = "outliers.txt")
```

ROTINA DO PCADAPT R

```

K <- 25
x <- pcadapt(pca_genotype, K = K)
plot(x, option = "screeplot") # 19 groups seems to be the correct value
plot(x, option = "scores", pop = sel[, 1]) # how populations are shared among the 19 groups
K <- 19
x <- pcadapt(pca_genotype, K = K, min.maf = 0)

summary(x) # numerical quantities obtained after performing a PCA
plot(x, option = "manhattan")
plot(x, option = "qqplot", threshold = 0.1)
plot(x, option = "stat.distribution") # Distribution of Mahalanobis distances.
qval <- qvalue(x$pvalues)$qvalues
alpha <- 0.1
outliers_pcadapt <- which(qval < alpha)
print(outliers_pcadapt)
length(outliers_pcadapt) # 14 outliers
alpha <- 0.01 # use of a more stringent threshold to detect outliers
outliers <- which(qval < alpha)
print(outliers)
length(outliers) # 14 outliers
ind <- paste("pop", sel[, 1]) # vector with the name of population

```