

CARACTERIZAÇÃO GENÉTICA DA RAÇA CHURRA GALEGA BRAGANÇANA PRETA POR TÉCNICAS DE SEQUENCIAÇÃO



Fonte: www.ancras.pt

1. Introdução

Portugal tem um número elevado de raças autóctones de animais domésticos, que incluem 16 raças de ovinos, entre as quais a raça Churra Galega Bragançana Preta.

Estas raças representam um património genético nacional de elevada importância, que tem que ser protegido e valorizado.

Com efeito, os animais destas raças constituem recursos genéticos adaptados às condições ambientais das regiões onde são explorados, que no caso da Churra Galega Bragançana Preta inclui uma área geográfica no Nordeste de Portugal, na região de Trás-os-Montes.

Para que a preservação, valorização e utilização sustentável destes animais seja feita de forma eficiente a utilização de técnicas genómicas tem um potencial muito considerável.

Nos últimos 10 anos os avanços registados nas técnicas focadas na sequenciação de ADN têm sido muito significativos.

A capacidade de gerar dados de sequenciação de última geração aumentou sobremaneira, tendo na prática o efeito positivo de permitir a análise de genomas a um nível de resolução que anteriormente não era possível.

Este tipo de abordagem, quando aplicada no contexto das raças autóctones nacionais, como por exemplo a Churra Galega Bragançana Preta, constitui uma nova ferramenta com elevado potencial para caracterizar os genomas destes animais.

No caso dos ovinos, o genoma desta espécie foi sequenciado em 2014 (Jiang *et al.*, 2014), o que permitiu desde então a execução de vários tipos de estudos que até então não eram possíveis.

Especificamente, é agora possível caracterizar de forma muito detalhada a variação presente no genoma de qualquer raça de ovinos, principalmente através da identificação de SNPs (*Single Nucleotide Polymorphisms* – Polimorfismos de Nucleótido Único) e inserções/deleções de pequena escala.

Posteriormente a localização genómica e efeito de cada SNP pode ser avaliado e caracterizado. Por último, o conjunto de SNPs identificado permite a caracterização de vários parâmetros de genética populacional na raça, como por exemplo a extensão do desequilíbrio de ligação e a identificação de regiões homozigóticas no genoma.

Os objetivos deste trabalho foram efetuar a caracterização genética da raça Churra Galega Bragançana Preta através da utilização de técnicas de sequenciação de última geração.

2. Materiais e Métodos

Amostras de sangue foram recolhidas em três animais, fêmeas, inscritos no livro genealógico da raça Churra Galega Bragançana Preta, e provenientes de três explorações distintas.

O sangue foi utilizado para extrair ADN através da utilização de técnicas standard de biologia molecular.

Depois de extraído o ADN foi quantificado e preparado para ser utilizado na produção dos dados de sequenciação.

Os dados de sequenciação foram produzidos com recurso à técnica de resequenciação completa do genoma.

Bibliotecas específicas foram preparadas para cada animal, e sequenciadas na plataforma DNBSEQ, utilizando um comprimento para cada sequência de 150 nucleótidos.

Os dados de sequenciação foram depois alinhados contra o genoma de referência de ovinos (Jiang *et al.*, 2014) utilizando o software BWA-MEM (Li, 2013).

Os alinhamentos de menor qualidade foram removidos antes de ser iniciado o passo de identificação de SNPs, inserções e deleções, que foi realizado com a ferramenta GATK (McKenna *et al.*, 2010).

Os resultados foram depois filtrados utilizando vários parâmetros de qualidade, que incluíram o número mínimo de sequências observadas para cada genótipo, em cada SNP/inserção/deleção (mínimo de 15 sequências), assim como o índice de qualidade de cada genótipo (mínimo de 50).

Para cada SNP/inserção/deleção, os genótipos que não cumpriram estes patamares foram eliminados.

A anotação dos SNPs foi efetuada através de uma comparação direta entre todos os SNPs/inserções/deleções identificados contra a versão 102 da anotação do genoma da ovelha (NCBI, 2021, https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/102/).

Os módulos de análise implementados na ferramenta PLINK (Purcell *et al.*, 2007) foram utilizados para determinar a extensão do desequilíbrio de ligação presente nos genomas destes animais, e também para identificar regiões homocigóticas do genoma.

Nestas análises foram utilizados apenas os SNPs com um genótipo válido para todos os animais.

3. Resultados, discussão e aplicação prática

3.1 Resequenciação completa dos genomas

Os resultados relativos aos dados de sequenciação produzidos para cada animal, assim como a informação relativa a cada animal, estão indicados na tabela 1.

Nº Amostra	LG	Sexo	Nº Sequências	Q20 (%)	GC (%)
38170	404/ED37A	Fêmea	185,908,710	95.92	43.61
38171	606/ED65Q	Fêmea	185,051,250	95.75	43.39
38196	802/ED65Q	Fêmea	185,312,760	95.85	43.46

Tabela 1 – Informação sobre os animais estudados e resultados descritivos dos dados de sequenciação produzidos. Q20 diz respeito à percentagem de nucleótidos com um índice de qualidade superior a 20. A percentagem GC representa a proporção de nucleótidos G e C, entre todos os nucleótidos.

O número de sequências produzidas foi aproximadamente 185 milhões de sequências, para todos os animais analisados.

Este volume de sequenciação revelou-se apropriado para identificar o primeiro conjunto de identificação massiva de SNPs no genoma de ovelhas da raça Churra Galega Bragançana Preta. A percentagem Q20 obtida foi bastante boa, tendo a esmagadora maioria dos dados de sequenciação apresentado índices de qualidade muito elevados.

Relativamente à percentagem GC, os valores obtidos estão dentro dos parâmetros normais para a espécie ovina.

3.2 Identificação de SNPs

O número total de SNPs identificado foi de 2,426,990, depois de aplicados os critérios de filtragem.

Em qualquer espécie, normalmente o número de SNPs identificados está diretamente relacionado com o tamanho de cada cromossoma.

Desta forma, o número de SNPs identificados é maior nos cromossomas de maior tamanho, resultado que também foi observado na Churra Galega Bragançana Preta.

O número de SNPs identificado por cromossoma variou de 36,631 (cromossoma 24) a 236,563 (cromossoma 1).

A densidade de SNPs por cromossoma foi determinada através da normalização do número de SNPs tendo em conta o tamanho de cada cromossoma, uniformizando o número de SNPs identificados em cada milhão de nucleótidos (1 megabase – Mb).

A densidade de SNPs por cromossoma variou de 805 (cromossoma 5) a 1,436 (cromossoma 14). Estes resultados estão ilustrados graficamente na Figura 1, enquanto que os resultados determinados para cada cromossoma estão incluídos na Tabela Suplementar 1.

A lista completa de todos os SNPs identificados é fornecida num ficheiro suplementar.

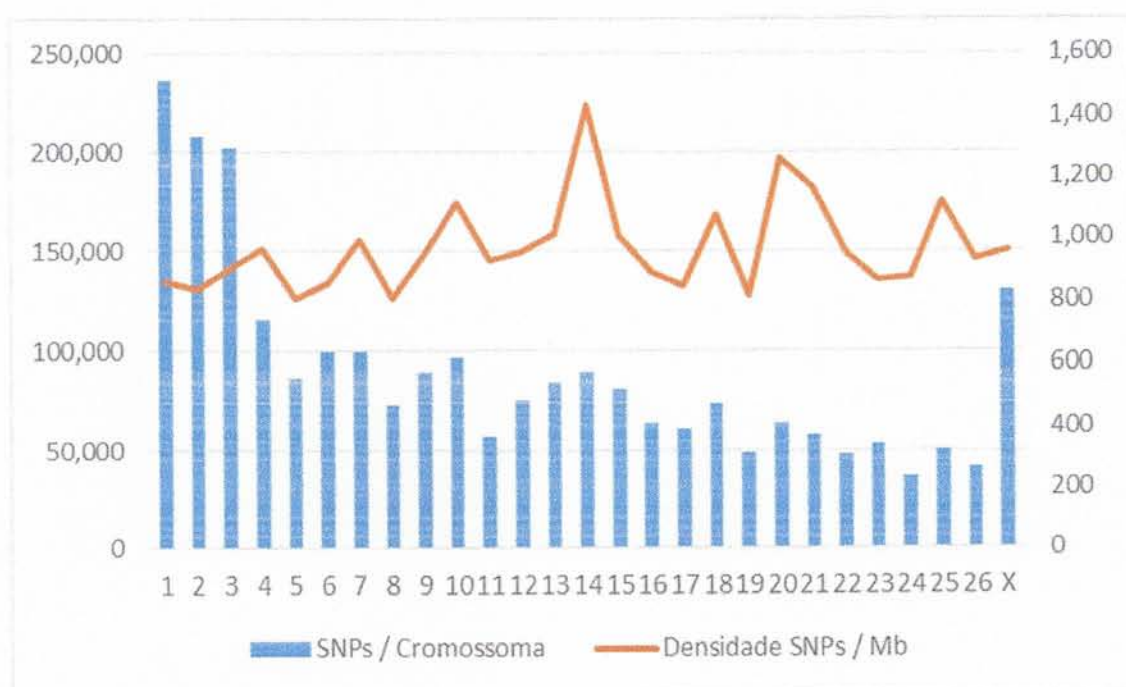


Figura 1 – Número de SNPs identificados por cromossoma, com valores apresentados para os 26 autossomas e o cromossoma X. A linha laranja representa a densidade de SNPs por Mb determinada para cada cromossoma.

Tendo por base este conjunto de análises foi determinado o primeiro conjunto de SNPs para a raça Churra Galega Bragançana Preta, para um total de SNPs que ultrapassou os 2.4 milhões.

Este número foi alcançado implementando critérios bastante exigentes para a filtragem dos SNPs, opção que é justificada pela necessidade de imediatamente disponibilizar para a raça um conjunto de SNPs de grau de confiança muitíssimo elevado, para serem utilizados em outras aplicações práticas de interesse para a raça.

No entanto, se os critérios de filtragem forem menos exigentes o número de SNPs identificado irá certamente aumentar.

Esta abordagem será implementada no futuro, quando for realizada uma nova ação de caracterização genética da raça por técnicas de sequenciação.

Quando os dados de sequenciação da nova ação estiverem disponíveis serão analisados conjuntamente com os dados produzidos para os três animais incluídos nesta ação realizada em 2020.

Desta forma, o aumento do número de animais analisados irá funcionar como uma garantia adicional para aplicar critérios de filtragem menos exigentes, aumentando o número de SNPs identificado para a raça, mas mantendo a níveis mínimos SNPs que poderiam eventualmente ser falsos positivos.

3.3 Identificação de inserções e deleções

Há vários tipos de variação que podem ser identificados nos genomas de todos os seres vivos.

Os SNPs representam o tipo de variação mais abundante, mas quando dados de resequenciação de genomas estão disponíveis é também possível identificar variações estruturais, como por exemplo inserções e deleções de pequeno tamanho.

Uma inserção é uma adição de um ou mais nucleótidos a uma sequência de DNA, enquanto que numa deleção ocorre a remoção de um ou mais nucleótidos da sequência de DNA.

Nas análises efetuadas neste trabalho, uma inserção representa uma região do genoma da Churra Galega Bragançana Preta onde ocorreu uma adição de nucleótidos, quando comparada com o genoma de referência, enquanto que as deleções representam as regiões do genoma da Churra Galega Bragançana Preta onde se verificaram fenómenos de remoção de nucleótidos, comparativamente ao genoma de referência.

No total foram identificadas 122,371 inserções e 146,994 deleções, distribuídas por todos os cromossomas.

Atualmente as técnicas de biologia molecular e/ou sequenciação ainda não permitem a genotipagem de inserções/deleções de forma massiva e económica, pelo que esta razão é o

principal obstáculo para aumentar o número de animais incluídos em trabalhos focados na genotipagem de inserções/deleções.

No entanto, este trabalho deixa estabelecido o primeiro conjunto de resultados alcançados na identificação deste tipo de variação na raça Churra Galega Bragançana Preta.

Posteriormente estes dados estarão disponíveis para serem utilizados em qualquer tipo de estudo.

A distribuição deste tipo de variação por cromossoma está ilustrada na Figura 2, enquanto que os resultados globais estão incluídos na Tabela S2.

A lista completa de todas as inserções e deleções identificadas é fornecida num ficheiro suplementar.



Figura 2 – Número de inserções e deleções identificadas por cromossoma, com valores apresentados para os 26 autossomas e o cromossoma X.

3.4 Anotação das variantes (SNPs, inserções e deleções) identificadas

O processo de anotação das variantes detetadas teve como objetivo identificar as variantes cujas posições coincidiam com as regiões génicas do genoma, que são aquelas que contêm os genes. Estas regiões são depois subdivididas em regiões codificantes (exões) e não-codificantes (intrões e regiões UTR).

Esta análise foi efetuada de forma conjunta para as raças Churra Galega Bragançana Branca e Churra Galega Bragançana Preta.

Foram detetadas 67,496 variantes localizadas na região génica do genoma, das quais 60,672 eram SNPs, 3,092 inserções, e 3,732 deleções. Estes resultados estão sumarizados na Tabela 2.

Cromossoma	SNPs	Inserções	Deleções	Total
1	5,596	307	382	6,285
2	3,648	227	263	4,138
3	4,214	238	279	4,731
4	3,844	179	234	4,257
5	2,059	109	114	2,282
6	2,700	117	175	2,992
7	1,660	77	108	1,845
8	984	43	70	1,097
9	1,178	47	70	1,295
10	485	19	31	535
11	1,671	69	99	1,839
12	1,742	93	104	1,939
13	2,171	96	141	2,408
14	1,840	224	132	2,196
15	700	28	37	765
16	534	28	42	604
17	867	50	43	960
18	901	40	36	977
19	3,297	160	203	3,660
20	1,764	125	139	2,028
21	12,975	517	657	14,149
22	355	10	11	376
23	519	11	17	547
24	1,125	78	87	1,290
25	1,905	77	108	2,090
26	1,384	89	104	1,577
X	554	34	46	634
TOTAL	60,672	3,092	3,732	67,496

Tabela 2 – Número de variantes identificadas nas regiões génicas do genoma, incluindo SNPs, inserções e deleções. O total de cada tipo de variante identificado em cada cromossoma é também apresentado.

Uma vez que nos genomas dos organismos superiores a região codificante representa apenas uma percentagem muito reduzida, normalmente à volta de 1-2%, é normal o número de variantes localizadas nesta regiões representar uma percentagem reduzida do total de variantes.

Regra geral, o número identificado destas variantes esteve diretamente relacionado com o tamanho do cromossoma.

No entanto, o cromossoma 21 constituiu uma exceção a esta regra, uma vez que neste cromossoma foram identificadas 14,149 variantes localizadas na região génica.

Este conjunto de SNPs identificados nas regiões génicas do genoma, incluindo várias centenas localizados em exões, que são a região codificante dos genes, constituem o primeiro conjunto de variantes com maior interesse para serem analisadas de forma mais detalhada em ações futuras.

As 67,496 variantes foram identificadas num total de 751 genes distintos, o que significou que uma percentagem muito significativa dos genes apresentou mais do que uma variante.

Com efeito, em apenas 48 genes foi detetada uma única variante.

Adicionalmente, foram também identificados genes que apresentaram mais do que um tipo de variante, dada a sobreposição para o número de genes identificado para SNPs (746), inserções (439) e deleções (451).

Estes resultados estão sumarizados na Tabela 3.

Cromossoma	SNPs	Inserções	Deleções
1	80	49	53
2	51	36	36
3	85	57	52
4	29	18	21
5	46	24	19
6	27	16	20
7	25	18	16
8	12	8	8
9	16	9	9
10	9	5	7
11	43	20	20
12	34	14	16
13	31	24	19

14	25	12	11
15	21	8	14
16	6	4	4
17	18	10	11
18	20	12	12
19	22	13	15
20	32	20	20
21	23	15	17
22	10	4	5
23	13	7	6
24	18	11	12
25	9	6	6
26	12	8	7
X	29	11	15
TOTAL	746	439	451

Tabela 3 – Número de genes nos quais foram identificadas variantes, incluindo SNPs, inserções e deleções. O número de genes identificado em cada cromossoma é também apresentado, para cada tipo de variante.

A ontologia de cada variante foi determinada, para caracterizar o tipo de efeito encontrado.

A maior percentagem foi observada para as variantes intrónicas, que representaram 65,151 variantes.

Este resultado está de acordo com as expectativas, uma vez que dentro das regiões génicas as zonas não codificantes representam a maior percentagem.

Foram identificadas 709 variantes sinónimas, que são variantes onde a alteração do nucleótido não causa uma mudança no aminoácido, enquanto que o número de variantes não sinónimas observadas, aquelas onde a alteração do nucleótido provoca uma alteração do aminoácido e consequentemente da sequência da proteína produzida pelo gene, foi de 780.

Foram ainda observadas, em menor número, como esperado, variantes com efeitos mais dramáticos sobre os genes, como por exemplo variantes que provocaram inserções/deleções disruptivas, ou que provocaram a perda/ganho de um codão stop.

Estes resultados estão sumarizados na Tabela 4.

Tipo de variante	Número
3' UTR	501

5' UTR	47
Deleção disruptiva	3
Inserção disruptiva	5
Variante downstream	1
Variante mudança fase	81
Deleção na fase	6
Inserção na fase	11
Variante iniciadora de codão	3
Intrónica	65,151
Não sinónima	780
Variante aceitadora de splice	22
Variante dadora de splice	11
Variante região splice	140
Ganho codão stop	18
Perda codão stop	5
Retenção codão stop	2
Sinónima	709

Tabela 4 – Número de observações registadas para o tipo de efeito provocado por cada variante localizada na região génica do genoma.

3.5 Variação encontrada em genes de interesse

A variação identificada na região génica dos genomas das ovelhas Churras Galegas Bragançanas Pretas analisadas incluiu variantes detetadas em genes previamente associados, ou com ação biológica direta, com vários tipos de características fenotípicas.

Uma revisão bibliográfica recente (Gebreselassie *et al.*, 2019) sumariza os principais genes candidatos encontrados até à data.

Comparando os genes indicados nessa revisão com a lista de genes identificados neste trabalho encontramos genes que na Churra Galega Bragançana Preta apresentam variação relevante.

No que respeita às características de crescimento e qualidade da carcaça, foi detetada variação em genes como por exemplo a calpaína (CAPN1), calpastatina (CAST), hormona do crescimento (GH), recetor da hormona do crescimento (GHR), fator de crescimento semelhante à insulina tipo 1 (IGF1), vários genes que codificam para proteínas de ligação do IGF1, recetor de leptina (LEPR) e proteína desacopladora 1 (UCP1).

Foram também identificadas variantes em genes associados a características reprodutivas, como é o caso da proteína morfogenética óssea (BMP15), recetor tipo 1B da proteína morfogenética óssea (BMPR1B), recetores da melatonina 1A (MTNR1A) e 1B (MTNR1B), e da proteína transmembramática 154 (TMEM154).

Relativamente a genes relacionados com produção e composição do leite, estas ovelhas Churras Galegas Bragançanas Pretas apresentaram variação em genes como o DGAT1 (Diacilglicerol O-Aciltransferase 1), POU1F1 (fator de transcrição positivo tipo 1, pituitária-específico), assim como nos genes que codificam para as proteínas do leite alfa-s1, alfa-s2, beta e kapa caseína (CSN1S1, CSN1S2, CSN2, CSN3, respetivamente) e alfa-lactalbumina (LALBA).

Esta variação fica imediatamente disponível para ser caracterizada num maior número de animais da raça, idealmente com registos fenotípicos disponíveis, de forma a permitir a identificação de marcadores genéticos associados com este tipo de fenótipos.

3.6 Caracterização do desequilíbrio de ligação existente na raça

O desequilíbrio de ligação (LD – *linkage disequilibrium*) é a associação não aleatória de alelos de *loci* diferentes.

Este parâmetro de genética de populações é influenciado por fatores como seleção, sistemas de reprodução, e estrutura da população, entre outros.

Em populações de animais domésticos o LD é um indicador muito valioso para monitorizar a evolução, ao nível global do genoma, da aplicação de medidas de seleção ou de conservação genética, constituindo então uma ferramenta importante para a gestão destas populações, como por exemplo a raça Churra Galega Bragançana Preta. Por exemplo, um dos possíveis efeitos da aplicação de esquemas de seleção genética pode ser o aumento do LD que, se for excessivo, pode posteriormente implicar várias consequências negativas para a raça.

Logo, monitorizar o LD existente terá que ser feito de forma regular para evitar este tipo de problemas.

O LD foi determinado utilizando o conjunto de SNPs para o qual todos os animais tinham genótipo disponível, apenas para os SNPs que não estavam separados por mais de 100 SNPs, e analisando uma extensão máxima do genoma de 1 Mb (análise feita dividindo o genoma em intervalos de 1 Mb). No total, o LD foi avaliado para 1,175,593 combinações entre SNPs.

Para uma percentagem muito elevada destas combinações (aproximadamente 93.1%) não foi detetado qualquer LD entre os SNPs (Tabela 5).

LD completo foi detetado para 42,732 combinações de SNPs (representando 3.6% do total), tendo sido ainda registadas 38,795 combinações com valor de LD inferior.

Estes resultados estão de acordo com o que seria esperado para a raça, uma vez que não seria expectável que os esquemas de seleção genética aplicados na raça até à data tivessem exercido uma intensidade suficiente para aumentar de forma significativa os níveis globais de LD.

Uma análise comparativa com os resultados obtidos para a Churra Galega Bragançana Branca indica que a extensão de LD é ligeiramente inferior.

De futuro este tipo de análise será imprescindível para monitorizar a evolução da extensão de LD na raça, à medida que começarem a ser aplicados esquemas de seleção de outra natureza.

LD	Nº observações
0	1,094,066
0.25	38,795
1	42,732
TOTAL	1,175,593

Tabela 5 – Caracterização do desequilíbrio de ligação (LD) existente no genoma de animais da raça Churra Galega Bragançana Preta.

3.7 Identificação de regiões homozigóticas no genoma

A identificação de regiões homozigóticas no genoma (ROH – *Runs of homozygosity*) foi efetuada utilizando os SNPs para os quais todos os animais apresentaram um genótipo válido, tendo sido avaliada em intervalos genómicos com um mínimo de 10 SNPs e 5 kb de tamanho. ROHs são regiões contíguas de genótipos homozigóticos presentes no genoma de um indivíduo, como resultado de haplótipos idênticos transmitidos pelos progenitores à sua descendência.

Em raças de animais domésticos as ROH são uma ferramenta muito útil para avaliar a ancestralidade a nível individual e populacional, assim como o impacto na raça dos esquemas de seleção genética que são aplicados.

Desta forma, as ROH são um bom indicador do nível global de consanguinidade ao nível do genoma.

Na Churra Galega Bragançana Preta o número de ROHs detetadas por animal variou de 7 a 12 (Tabela 6).

A menor ROH identificada apresentou um tamanho de 5.1 kb, tendo a maior ROH apresentado um tamanho de 404.5 kb. O número de SNPs envolvido nas ROH variou de 179 a 250.

A média observada para o tamanho das ROH identificadas (106.6 kb) indica que a extensão destas regiões não é muito elevada na Churra Galega Bragançana Preta, sinal de que a consanguinidade global para a raça não é muito elevada.

Estes resultados estão de acordo com os observados em 11 raças Espanholas de ovinos (Luigi-Sierra *et al.*, 2019), onde números reduzidos foram igualmente detetados, mas em contraste com resultados obtidos em 21 raças Italianas (Mastrangelo *et al.*, 2018), que aparentam ter níveis muito superiores de consanguinidade.

Tamanho das ROH (kb)					
Animal	Nº ROH	Mínimo	Máximo	Média	Nº SNPs
404/ED37A	12	5.1	395.1	107.3	250
606/ED65Q	10	5.1	335.3	66.1	207
802/ED65Q	7	5.4	404.5	146.4	179

Tabela 6 – Caracterização das regiões homozigóticas (ROH) existentes no genoma de animais da raça Churra Galega Bragançana Preta.

4. Comentários finais

Neste estudo foram analisados os primeiros genomas de ovinos da raça Churra Galega Bragançana Preta, tendo sido identificado o primeiro conjunto de SNPs de grande dimensão para a raça, que incluiu 2,426,990 SNPs.

Adicionalmente, foram também identificadas 122,371 inserções e 146,994 deleções, distribuídas por todos os cromossomas.

Este conjunto global de variantes fica imediatamente disponível para os vários tipos de aplicações de interesse para a raça onde este tipo de informação seja necessário.

A anotação das variantes identificadas revelou a presença de 67,496 variantes localizadas em 751 genes distintos.

Estas variantes foram identificadas em genes que desempenham um papel biológico relevante, ou previamente associados, com características fenotípicas de interesse para a raça, como por exemplo o crescimento e a reprodução. Os resultados determinados para o desequilíbrio de ligação e identificação de regiões homozigóticas revelam um cenário de pouca homozigotia e consanguinidade no genoma da raça.

Futuramente, este tipo de análise deverá ser novamente utilizado para monitorizar, ao nível global do genoma, a evolução destes parâmetros de genética populacional, que serão muito úteis para a gestão do efetivo e a aplicação de esquemas de seleção genética que incorporem informação derivada de marcadores moleculares.

5. Referências bibliográficas

Gebreselassie G, Berihulay H, Jiang L, Ma Y. Review on Genomic Regions and Candidate Genes Associated with Economically Important Production and Reproduction Traits in Sheep (*Ovis aries*). *Animals (Basel)*. 2019 Dec 23;10(1):33. doi: 10.3390/ani10010033. PMID: 31877963; PMCID: PMC7022721.

Li H, 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

Luigi-Sierra MG, Cardoso TF, Martínez A, Pons A, Bermejo LA, Jordana J, Delgado JV, Adán S, Ugarte E, Arranz JJ, Casellas J, Amills M. Low genome-wide homozygosity in 11 Spanish ovine breeds. *Anim Genet*. 2019 Oct;50(5):501-511. doi: 10.1111/age.12832. Epub 2019 Aug 8. PMID: 31393638

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton JA, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristensen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC, Kijas J, Wang J, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*. 2014 Jun 6;344(6188):1168-1173. doi: 10.1126/science.1252806. PMID: 24904168; PMCID: PMC4157056.

Mastrangelo S, Ciani E, Sardina MT, Sottile G, Pilla F, Portolano B; Bi.Ov. Ita Consortium. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. *Anim Genet*. 2018 Feb;49(1):71-81. doi: 10.1111/age.12634. Epub 2018 Jan 15. PMID: 29333609.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010 Sep;20(9):1297-303. doi: 10.1101/gr.107524.110. Epub 2010 Jul 19. PMID: 20644199; PMCID: PMC2928508.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559-75. doi: 10.1086/519795. Epub 2007 Jul 25. PMID: 17701901; PMCID: PMC1950838.

6. Material Suplementar

6.1 Tabela S1 – Número de SNPs identificado e densidade de SNPs/Mb para cada cromossoma.

Cromossoma	SNPs / Cromossoma	Tamanho Cromossoma (bp)	Densidade SNPs / Mb
1	236,563	275,406,953	859
2	208,150	248,966,461	836

3	202,309	223,996,068	903
4	115,377	119,216,639	968
5	86,825	107,836,144	805
6	99,658	116,888,256	853
7	99,643	100,009,711	996
8	73,180	90,615,088	808
9	89,708	94,583,238	948
10	96,365	86,377,204	1,116
11	57,476	62,170,480	924
12	75,476	79,028,859	955
13	84,264	82,951,069	1,016
14	89,827	62,568,341	1,436
15	81,367	80,783,214	1,007
16	63,811	71,693,149	890
17	61,187	72,251,135	847
18	73,782	68,494,538	1,077
19	49,228	60,445,663	814
20	64,300	51,049,468	1,260
21	58,193	49,987,992	1,164
22	48,332	50,780,147	952
23	53,779	62,282,865	863
24	36,631	41,976,827	873
25	50,679	45,223,504	1,121
26	41,010	44,047,080	931
X	129,870	135,185,801	961
TOTAL	2,426,990	2,584,815,894	939

6.2 Tabela S2 – Número de inserções e deleções identificadas para cada cromossoma.

Cromossoma	Inserções	Deleções
1	11,985	13,887
2	9,707	11,834
3	10,361	12,433
4	6,010	7,269
5	4,302	5,091

6	4,572	5,578
7	5,292	6,455
8	3,280	4,126
9	4,292	5,272
10	4,906	6,009
11	2,824	3,500
12	3,964	4,534
13	4,325	5,209
14	5,649	6,600
15	3,990	4,965
16	2,843	3,487
17	2,942	3,770
18	3,800	4,490
19	2,086	2,589
20	3,394	4,070
21	2,594	3,108
22	2,410	2,804
23	2,358	2,925
24	1,816	2,157
25	2,560	2,922
26	1,884	2,254
X	8,225	9,656
TOTAL	122,371	146,994

Mangualde, 30 de Dezembro de 2020

P'la Star Genomics



STAR GENOMICS,
Unipessoal Lda
NIPC 516 232 738
A Gerência