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

Identificação e uso de CNV para seleção genômica em ovinos da raça Santa
Inês

Giovanni Coelho Ladeira

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

Piracicaba
2021

Giovanni Coelho Ladeira
Bacharel em Zootecnia

Identificação e uso de CNV para seleção genômica em ovinos da raça Santa Inês
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **GERSON BARRETO MOURÃO**

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

Piracicaba
2021

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

Ladeira, Giovanni Coelho

Identificação e uso de CNV para seleção genômica em ovinos Santa Inês / Giovanni Coelho Ladeira. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - -Piracicaba, 2021.

73 p.

Dissertação (Mestrado) - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Variação no número de cópias 2. Associação genômica ampla 3. Seleção genômica 4. Santa Inês I. Título

DEDICATÓRIA

Dedico esta dissertação a minha família por todo o ensinamento, alegria e amor compartilhado. Dedico especialmente à minha mãe, Maria de Fátima, ao meu pai, Sandro Ladeira, ao meu irmãozinho, Nicolas Ladeira, e ao Sandro Oliveira por tantos momentos bons e amor incondicional. Vocês são parte da minha essência.

AGRADECIMENTOS

Agradeço a Deus por todas as oportunidades e momentos vividos, todos, de alguma forma, me guiaram até aqui. Agradeço por todas as bênçãos a mim e àqueles que amo e também pelas turbulências ao longo do caminho, pois também são oportunidades de desenvolvimento.

Agradeço profundamente à minha família, tios, tias, primos e amigos pelos conselhos, ensinamentos, momentos de alegria e suporte. Agradeço especialmente a minha mãe, Maria de Fátima, ao meu pai, Sandro Ladeira, ao meu irmãozinho, Nicolas Ladeira, aos meus avós, Sonia Ladeira e José Coelho e ao Sandro Oliveira por tanto amor, por serem meu maior suporte e minha maior motivação, mas principalmente por me ensinarem a buscar felicidade, nos detalhes, em todos os momentos da vida. Todos vocês me inspiram e me fazem querer ser melhor a cada momento. Com vocês aprendi o que é ser humano de verdade, aprendi valores que carregarei comigo eternamente e que o importante é ser e fazer as pessoas que amamos felizes. Amo vocês com todo o meu ser. Obrigado!

À Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo (ESALQ/USP) e ao Programa de Pós-Graduação em “Ciência Animal e Pastagens” pela oportunidade de realização do mestrado.

Ao orientador professor Dr. Gerson Barreto Mourão pelos ensinamentos, troca de conhecimentos, oportunidades, apoio, confiança e valiosos conselhos. Obrigado pelos momentos vividos, todos eles foram de grande aprendizado.

Ao professor Dr. Luiz Lehmann Coutinho e ao professor Dr. Luís Fernando Batista Pinto pelas contribuições durante o comitê de orientação e ao longo da trajetória do mestrado. À Universidade Federal da Bahia (UFBA) pela parceria na condução de experimentos, e ao Laboratório de Biotecnologia (ESALQ/USP) pelo suporte com a genotipagem.

Aos membros do Grupo de Estudos em Estatística e Melhoramento Animal, por toda a parceria, troca de experiências, contribuições e aprendizado: Ariana Meira, Brayan Dauria, Daysi Orzuza, Fabrício Pilonetto, Fátima Bogdanski, professor Gerson Mourão, Gregori Rovadoscki, Izally Gervásio, Leonardo Mazzero, Luiz Firmino dos Santos, Juliana Petrini, Mayara Salvian e Paola Bóscollo e Sofia Silva. Agradeço especialmente ao Fabricio Pilonetto pelas trocas de ideias e contribuição durante execução deste trabalho.

Aos professores do Departamento de Zootecnia por todo conhecimento transmitido durante as aulas e excelente qualidade do trabalho desenvolvido.

Ao Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) pelo suporte à educação e pesquisa.

À Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP – Processo 2019/19213-0) pela concessão da bolsa de estudos.

A todos que passaram pela minha trajetória, com certeza todos vocês contribuíram de alguma forma para a realização deste trabalho e crescimento pessoal.

SUMÁRIO

RESUMO	7
ABSTRACT	8
1. INTRODUÇÃO.....	9
REFERÊNCIAS	13
2. ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA BASEADO EM CNV EM OVINOS SANTA INÊS	17
RESUMO	17
ABSTRACT	17
2.1. INTRODUÇÃO.....	18
2.2. MATERIAL E MÉTODOS.....	20
2.2.1. Descrição da população	20
2.2.2. Fenótipos	21
2.2.2.1. Banco de dados.....	21
2.2.2.2. Estimativas e mensurações dos fenótipos.....	21
2.2.2.2.1. Estimativa de eficiência alimentar e consumo alimentar residual.....	21
2.2.2.2.2. Mensurações de características de carcaça.....	22
2.2.3. Extração do DNA, genotipagem e controle de qualidade	23
2.2.4. Identificação de CNV	24
2.2.5. Compilação de CNVR.....	25
2.2.6. Estudo de associação genômica ampla baseado em CNV (CNV-GWAS)	25
2.2.7. Sobreposição de CNVs com QTLs conhecidos.....	26
2.2.8. Identificação de genes candidatos	27
2.3. RESULTADOS	27
2.3.1. Descrição da população e fenótipos	27
2.3.2. Identificação de CNV	27
2.3.3. Compilação de CNVR.....	28
2.3.4. Estudo de associação genômica ampla (GWAS)	28
2.3.5. CNVs sobrepostos a QTLs conhecidos	30
2.3.6. Anotação de genes	30
2.4. DISCUSSÃO.....	31
2.5. CONCLUSÃO.....	34
REFERÊNCIAS	35

3. PREDIÇÃO GENÔMICA COMBINANDO MARCADORES SNP E VARIAÇÃO NO NÚMERO DE CÓPIAS EM OVINOS SANTA INÊS	43
RESUMO	43
ABSTRACT	43
3.1. INTRODUÇÃO	44
3.2. MATERIAIS E MÉTODOS	46
3.2.1. População, genotipagem e controle de qualidade	46
3.2.2. Fenótipos	46
3.2.2.1. Estimativas das características de eficiência	47
3.2.2.2. Mensurações de características de carcaça	48
3.3. ANÁLISE GENÉTICA	48
3.3.1. Uso de marcadores CNVs para construção da matriz de relacionamento	48
3.3.2. Predição dos valores genéticos genômicos (GEBVs)	51
3.3.3. Acurácia e viés de predição	53
3.4. RESULTADOS	54
3.4.1. População e fenótipos	54
3.4.2. Matrizes de relacionamento genômico	55
3.4.3. Estimativas dos componentes de variância e herdabilidades	56
3.4.4. Acurácia e viés de predição	59
3.5. DISCUSSÃO	61
3.5.1. Relacionamento genômico e componentes de variância	61
3.5.2. Habilidade e viés de predição	63
3.6. CONCLUSÃO	65
REFERÊNCIAS	65
APÊNDICES	70

RESUMO

Identificação e uso de CNV para seleção genômica em ovinos da raça Santa Inês

Variações no número de cópias (CNV) são variações estruturais no DNA de no mínimo 1 kb e que diferem do genoma referência quanto ao número de cópias. O crescente número de estudos com CNV reportam sua associação com doenças em humanos e diversas características quantitativas em animais domésticos. Entretanto, não foram reportados estudos de associação genômica ampla baseados em CNV em ovinos Santa Inês, uma das raças mais importantes do Brasil e, ademais, informações sobre a inclusão de efeitos de CNV em modelos de predição de GEBV são escassos. Assim, os objetivos do presente estudo foram 1) identificar e quantificar CNVs em uma população de ovinos da raça Santa Inês, 2) verificar possíveis associações entre CNVs e características de interesse, 3) identificar possível sobreposição de segmentos de CNVs significativamente associados às características de interesse a QTLs e genes previamente conhecidos, e 4) avaliar possíveis ganhos em acurácia das predições dos GEBVs combinando informações de SNP e CNV ($GB_a + CNV$) comparadas às predições dos GEBVs utilizando exclusivamente informações SNP (GB_a). Dessa forma, foram utilizadas informações genômicas de 491 ovinos da raça Santa Inês (OvineSNP50 BeadChip). A identificação de CNVs foi realizada por meio do software PennCNV. As informações de CNVs foram utilizadas para a construção de uma matriz de relacionamento genômico fundamentada em CNVs que foi incorporada ao modelo misto utilizado para a predição dos GEBVs adicionalmente às tradicionais informações de SNP. Na sequência, a acurácia foi estimada pela correlação de Pearson entre os GEBVs preditos para a população de treinamento e o fenótipo corrigido para efeitos fixos na população de validação. Os GEBVs preditos foram utilizados nas análises de associação genômica ampla fundamentadas em CNVs, por meio do programa CNVRanger, para identificar regiões no genoma associadas às características de desempenho, eficiência e carcaça. Foram identificados 1.167 segmentos de CNV, entre eles segmentos significantes (p -valor<0,05) associados às características: rendimento de carcaça e consumo alimentar residual, adicionalmente, outros 5 segmentos de CNV foram considerados relevantes e indicados para estudos futuros. Os segmentos significativos foram sobrepostos a 4 QTLs e 8 genes, entre eles os genes *SPAST*, *TGFA* e *ADGRL3* com funções relacionadas a diferenciação celular e metabolismo energético. A inclusão de CNV nas análises genéticas permitiram capturar maior variância genética, entretanto a proporção da variância genética capturada por CNVs varia entre as características e não necessariamente incrementam a acurácia de predição. Entre os modelos considerados, não houve incremento na habilidade preditiva ao inserir informações de CNV. Portanto, os resultados reportados no presente estudo indicam que a inclusão de informações de CNV em modelos para predição de GEBV não deve ser prioridade.

Palavras-chave: Habilidade de predição, Matriz de relacionamento genômico, Cordeiros, Viés de predição

ABSTRACT

Identification and use of CNV for genomic selection in Santa Inês sheep

Copy number variations (CNV) are structural variations in DNA of at least 1 kb that differ from the reference genome in copy number. The growing number of studies with CNV report its association with diseases in humans and several quantitative traits in livestock animals. However, CNV-based genome-wide association studies in Santa Inês sheep, one of the most important breeds in Brazil, have not been reported and, furthermore, information on the inclusion of CNV effects in GEBV prediction models is rare. Thus, the aims of the present study were 1) to identify and quantify CNVs in a population of Santa Inês sheep, 2) to verify possible associations between CNVs and characteristics of interest, 3) to identify possible overlapping of CNV segments significantly associated with the characteristics of interest to previously known QTLs and genes, and 4) to evaluate possible gains in accuracy in GEBV predictions by combining SNP and CNV information ($GB_a + CNV$) compared to GEBV predictions using exclusively SNP information (GB_a). Thus, genomic information from 491 Santa Inês sheep (OvineSNP50 BeadChip) was used. The identification of CNVs was performed using the PennCNV software. The CNV information was used to build a genomic relationship matrix based on CNVs, which was incorporated into the mixed model used for the prediction of GEBVs in addition to the traditional SNP information. Next, accuracy was estimated by Pearson's correlation between predicted GEBVs for training population and phenotype corrected for fixed effects in the validation population. Predicted GEBVs were used in CNV-based genome-wide association analyses, through the CNVRanger program, to identify regions in the genome associated with performance, efficiency and carcass traits. 1,167 CNV segments were identified, including significant segments (p -value<0.05) associated with traits: carcass yield and residual feed intake, in addition, another 5 CNV segments were considered relevant and indicated for future studies. The significant segments overlapped 4 QTLs and 8 genes, including the *SPAST*, *TGFA* and *ADGRL3* genes with functions related to cell differentiation and energy metabolism. The inclusion of CNV in the genetic analysis allowed capturing greater genetic variance, however the proportion of genetic variance captured by CNVs varies among characteristics and does not necessarily increase the prediction accuracy. Among the models considered, there was no increase in predictive ability when entering CNV information. Therefore, the results reported in the present study indicate that the inclusion of CNV information in models to predict GEBV should not be a priority.

Keywords: Prediction ability, Genomic relationship matrix, Lambs, Prediction bias

1. INTRODUÇÃO

No cenário atual, o Brasil se mantém em destaque mundial como produtor de proteína animal, sendo o segundo maior produtor de carne bovina, o terceiro maior produtor de carne de frango e o quinto maior produtor de carne suína, com produções aproximadas em milhões de toneladas de 10,2, 13,5 e 4,2, respectivamente (FAO, 2021a). O país também se destaca mundialmente quanto às exportações de carne bovina, de frango e suína como primeiro, primeiro e quarto maior exportador, respectivamente (FAO, 2021b, 2020; USDA, 2021). Assim, a pecuária contribui de maneira positiva para o agronegócio brasileiro. Entretanto, no Brasil, existem setores da pecuária, como a ovinocultura, que apresentam números menos expressivos. Nesse contexto, a produção brasileira de carne ovina em 2020 foi de 136 mil toneladas e, em 2019, o país era apenas o vigésimo quinto no *ranking* mundial (FAO, 2021a, 2021b). Desse modo, os números indicam a prática parcial do potencial de produção de carne ovina.

O efetivo de ovinos no Brasil foi de aproximadamente 19,7 milhões para 2019, com grande parte do rebanho concentrado nas regiões Nordeste (69%) e Sul (20%) (IBGE, 2021). Entretanto, a carne ovina é pouco consumida no Brasil. O consumo per capita de carne ovino no Brasil é de 0,5 kg por ano (OECD, 2021). Mesmo com o baixo consumo, a produção nacional de carne ovina ainda não é suficiente para atender o mercado interno. Em 2018, o Brasil importou aproximadamente 7,5 mil toneladas de carne ovina *in natura* (cerca de U\$ 48 milhões), maior parte desse volume, aproximadamente 6,5 mil toneladas, foi importada do Uruguai (MAPA, 2019). Diante disso, fica evidente a possibilidade de expansão da cadeia produtiva de carne ovina e a grande porção do mercado que pode ser ocupada por produtos nacionais.

A ovinocultura ocorre em todo o território brasileiro, é destinada à atividade comercial de alto rendimento com a exploração de carne, leite, lã e couro, também é destinada à subsistência de famílias, principalmente no nordeste brasileiro (Viana, 2008). Entretanto, em um estudo de caso, Viana e Silveira (2009) retratam a ovinocultura como uma atividade rentável, mas que não remunera os custos de oportunidade. O desempenho modesto da ovinocultura nacional está associado ao sistema produtivo que é caracterizado por ser pouco organizado e tecnificado (Sorio e Rasi, 2010; Viana, 2008), com baixos índices de assistência técnica especializada (19,9%) e participação em cooperativas (11,4%) (Embrapa, 2018). Entre os entraves da ovinocultura destaca-se a subutilização do melhoramento genético, com

desenvolvimento genético limitado dos animais que melhor atendem às necessidades das comunidades que os mantêm e fornecem (Braga Lobo, 2019).

Desde a domesticação dos animais, os criadores têm direcionado a evolução dos animais domésticos. Inicialmente, as ferramentas disponíveis para os criadores eram simplesmente observações sobre o que era percebido como a aparência desejada e o acasalamento seletivo de pais escolhidos (Van Eenennaam, 2017). Desde então, houve incremento no conhecimento sobre os componentes da informação genética, hereditariedade e métodos de ação de biomoléculas, acompanhados por importantes avanços tecnológicos. Assim, observações subjetivas foram substituídas por medições objetivas e métodos estatísticos sofisticados foram implementados para isolar os efeitos genéticos hereditários das influências ambientais com incremento na taxa de ganho genético por unidade de tempo (Van Eenennaam et al., 2013).

Dessa forma, o melhoramento genético de animais domésticos se consolidou como ferramenta importante para avanços produtivos e qualitativos na produção animal. Nesse cenário, as características mais importantes são complexas, quantitativas (características afetadas por muitos genes e pelo ambiente) (Meuwissen e Goddard, 2010), de modo que a seleção para as características de interesse foi primeiramente fundamentada nas informações fenotípicas, de pedigree e na estimativa dos parâmetros genéticos (Dekkers e Hospital, 2002).

Venturosamente, os avanços tecnológicos nas áreas de genética molecular e bioinformática tornaram a seleção genômica (*Genomic Selection* - GS) uma importante ferramenta para incrementar o ganho genético por meio do melhoramento animal, especialmente para características complexas e de difícil mensuração, devido a incorporação de informações genômicas obtidas pela utilização de marcadores moleculares do tipo SNP (*Single Nucleotide Polymorphisms*) e sequenciamento completo do genoma (Stock e Reents, 2013). A seleção genômica pode incrementar acurácia na predição dos méritos genéticos dos indivíduos, revolucionou a indústria de criação de gado leiteiro, e a técnica foi implementada com vários graus de sucesso em programas de criação associados a outras indústrias de animais e plantas (Van Eenennaam, 2017). Existem diversos trabalhos que reportam abordagem genômica em ovinos (Dlamini et al., 2019; Lee et al., 2015; McLaren et al., 2018), porém trabalhos que abordam especificamente o uso da informação genômica na raça Santa Inês ainda são escassos na literatura (Alvarenga et al., 2018; Meira et al., 2019; Rovadoscki et al., 2018).

Seleção genômica pode ser definida como a seleção simultânea para muitos (dezenas, centenas de milhares) marcadores, que cobrem todo o genoma de uma maneira densa, espera-se que os genes estejam em desequilíbrio de ligação com pelo menos alguns dos marcadores (Meuwissen, 2007). A GS tornou-se viável devido ao grande número de SNPs descobertos pelo sequenciamento do genoma e novos métodos para genotipagem eficiente de grande número de SNPs (Goddard e Hayes, 2007). Os mesmos autores relatam que a GS é baseada na avaliação dos efeitos dos QTLs, utilizados para estimar os GEBVs sob equações de predição.

Ademais, para realizar a GS é necessário um conjunto de dados genotípicos de forma que os dados fenotípicos não são obrigatórios, mas acrescentam confiabilidade às predições. Segundo Bouquet e Juga (2013) é possível realizar GS com indivíduos jovens que não tenham descendentes e, apesar da menor acurácia de predição comparada à acurácia para EBVs de touros amplamente testados, o ganho genético é acumulado devido ao menor intervalo de gerações. Adicionalmente, é razoável compreender que ao realizar seleção genômica parte do custo associado ao teste de progênie é reduzido. Dessa forma, a seleção genômica contribui com a redução do intervalo de geração e ganho de acurácia.

Nesse contexto, as informações de marcadores do tipo SNP são muito úteis e amplamente utilizadas na GS. Contudo, existem outras variações no genoma que podem ser potencialmente importantes para incrementar a construção da matriz de relacionamento genômico (matriz G), bem como estarem envolvidas com características produtivas de interesse. Assim, os SNPs podem não captar todo o efeito de variações estruturais específicas, como o causado por variação no número de cópias (*Copy Number Variation* – CNV).

A variação no número de cópias (CNV) foi descrita inicialmente por Sebat et al. (2004) e logo em seguida por Iafrate et al. (2004). Os autores consideraram apenas variações em larga escala, e, por isso, classificaram CNV com o tamanho mínimo de 100 quilobases (kb). Porém, Feuk et al. (2006) descrevem CNV como um segmento cromossômico, com pelo menos 1 kb de comprimento, cujo número de cópias varia em comparação com o genoma de referência. Os mesmos autores relatam que as classes de CNVs incluem inserções, deleções e duplicações, entre outras.

Entretanto, inicialmente, os estudos sobre CNVs foram focados, principalmente, no genoma humano (Bae et al., 2010), e mais recentemente alguns estudos identificaram CNVs em aves (Fernandes et al., 2021; Rao et al., 2016; Wang et al., 2010), bovinos (da Silva et al.,

2016; Da Silva et al., 2016; Lee et al., 2020) e ovinos (Di et al., 2019; Feng et al., 2020; Wang et al., 2020). Nesse contexto, o número de trabalhos que abordam CNV em ovinos é muito baixo. Ademais, não foram encontrados registros de identificação de CNV em ovinos da raça Santa Inês, uma das principais raças presentes no Brasil.

Pesquisas indicam que variantes estruturais podem compreender milhões de nucleotídeos de heterogeneidade dentro de cada genoma, com provável contribuição importante para diversidade humana e suscetibilidade a doenças (Feuk et al., 2006). Segundo Henrichsen et al. (2009) as CNVs podem modificar as expressões de genes, seja devido ao efeito da dosagem dos genes dentro da CNV ou mudanças estruturais que interferem na expressão de genes vizinhos. Feuk et al. (2006) afirmam que segmentos de CNVs são amplamente dispersos nos genomas de mamíferos e que também contribuem independentemente para a diversidade fenotípica. Ademais, as regiões de CNVs (CNVRs) representam aproximadamente 6,5 e 6,9% dos genomas bovino e ovino, respectivamente (Da Silva et al., 2016; Yang et al., 2018).

birth weight (BW), post weaning gain (PWG), weaning gain (WG), conformation at weaning (CW), muscling at weaning (MW), precocity at weaning (PW), conformation at yearling (CY), muscling at yearling (MY) and precocity at yearling (PY)

Ademais, quantidades significativas de CNVs apresentam efeitos genéticos funcionais, esses segmentos podem estar associados aos efeitos de expressão gênica ou presentes em genes importantes para as características de interesse econômico (Redon et al., 2006; Wang et al., 2008; Yang et al., 2018; Zhou et al., 2016). Em bovinos da raça Holandesa, CNVs sobrepostos ao gene *RXFP4* (relaxina/insulina como receptor de peptídeos) foram associados ao consumo alimentar residual (Zhou et al., 2018). Adicionalmente, em uma população de bovinos da raça Nelore, foram identificados 17 CNVs significativamente associados a características de crescimento (ganho de peso e conformação ao desmame, precocidade ao desmame e ao ano, conformação ao ano, e musculatura ao desmame e ao ano) (Zhou et al., 2016). No genoma ovino, CNVs são responsáveis pela variação genética entre populações de diversas partes do globo, possivelmente refletindo o histórico de populações de diferentes raças (Yang et al., 2018).

Assim, há possibilidade de que os efeitos genéticos não se devam apenas aos SNPs, mas sejam também influenciados por outros tipos de variantes genéticas, como microssatélites e CNVs (Hay et al., 2018).

Diante do exposto, faz-se relevante a primeira identificação de CNVs presentes em uma população de ovinos da raça Santa Inês, bem como investigar efeitos da adição de informações de CNVs e de SNPs aos modelos mistos, mensurar seus possíveis impactos sobre a acurácia de predição dos valores genéticos genômicos (GEBV) e, adicionalmente, realizar estudo de associação genômica ampla baseado em CNVs a fim de identificar possíveis efeitos significativos de CNVs sobre características de interesse em ovinos.

Referências

- Alvarenga, A.B., Rovadoscki, G.A., Petrini, J., Coutinho, L.L., Morota, G., Spangler, M.L., Pinto, L.F.B., Carvalho, G.G.P., Mourão, G.B., 2018. Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. *Sci. Rep.* 8, 1–11. <https://doi.org/10.1038/s41598-018-27259-7>
- Bae, J.S., Cheong, H.S., Kim, L.H., NamGung, S., Park, T.J., Chun, J.Y., Kim, J.Y., Pasaje, C.F.A., Lee, J.S., Shin, H.D., 2010. Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics* 11. <https://doi.org/10.1186/1471-2164-11-232>
- Bouquet, A., Juga, J., 2013. Integrating genomic selection into dairy cattle breeding programmes: A review. *Animal* 7, 705–713. <https://doi.org/10.1017/S1751731112002248>
- Braga Lobo, R.N., 2019. Opportunities for investment into small ruminant breeding programmes in Brazil. *J. Anim. Breed. Genet.* 136, 313–318. <https://doi.org/10.1111/jbg.12396>
- da Silva, J.M., Giachetto, P.F., da Silva, L.O., Cintra, L.C., Paiva, S.R., Yamagishi, M.E.B., Caetano, A.R., 2016. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genomics* 17. <https://doi.org/10.1186/s12864-016-2752-9>
- Da Silva, V.H., De Almeida Regitano, L.C., Geistlinger, L., Pértile, F., Giachetto, P.F., Brassaloti, R.A., Morosini, N.S., Zimmer, R., Coutinho, L.L., 2016. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. *PLoS One* 11, 1–22. <https://doi.org/10.1371/journal.pone.0157711>
- Dekkers, J.C.M., Hospital, F., 2002. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3, 22–32. <https://doi.org/10.1038/nrg701>
- Di, R., Id, G., Sutera, A.M., Mastrangelo Id, S., Tolone, M., Portolano, B., Sottileid, G., Bagnato, A., Strillacciid, M.G., Sardina, M.T., 2019. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. <https://doi.org/10.1371/journal.pone.0215204>
- Dlamini, N.M., Visser, C., Snyman, M.A., Soma, P., Muchadeyi, F.C., 2019. Genomic evaluation of resistance to *Haemonchus contortus* in a South African Dohne Merino flock. *Small Rumin. Res.* 175, 117–125. <https://doi.org/10.1016/j.smallrumres.2019.04.020>
- Embrapa, C.O., 2018. Panorama da ovinocultura e da caprinocultura a partir do Censo Agropecuário 2017.
- FAO, 2021a. Food and Agriculture Organization of the United Nations. FAOSTAT, Statistical Database [WWW Document]. URL <http://www.fao.org/faostat/en/#data/QL> (accessed 6.25.21).
- FAO, 2021b. Meat market review: Overview of global meat market developments in 2020.

- FAO, 2020. Meat market review: Emerging trends and outlook [WWW Document]. URL <http://www.fao.org/3/cb3700en/cb3700en.pdf> (accessed 6.25.21).
- Feng, Z., Li, X., Cheng, J., Jiang, R., Huang, R., Wang, D., Huang, Y., Pi, L., Hu, L., Chen, H., 2020. Copy number variation of the piggy gene in sheep and its association analysis with growth traits. *Animals* 10, 688. <https://doi.org/10.3390/ani10040688>
- Fernandes, A.C., da Silva, V.H., Goes, C.P., Moreira, G.C.M., Godoy, T.F., Ibelli, A.M.G., Peixoto, J. de O., Cantão, M.E., Ledur, M.C., de Rezende, F.M., Coutinho, L.L., 2021. Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics* 22, 354. <https://doi.org/10.1186/s12864-021-07676-1>
- Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. <https://doi.org/10.1038/nrg1767>
- Goddard, M.E., Hayes, B.J., 2007. Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
- Hay, E.H.A., Utsunomiya, Y.T., Xu, L., Zhou, Y., Neves, H.H.R., Carneiro, R., Bickhart, D.M., Ma, L., Garcia, J.F., Liu, G.E., 2018. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. *BMC Genomics* 19, 1–8. <https://doi.org/10.1186/s12864-018-4787-6>
- Henrichsen, C.N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H., Reymond, A., 2009. Segmental copy number variation shapes tissue transcriptomes. <https://doi.org/10.1038/ng.345>
- Instituto Brasileiro de Geografia e Estatística (IBGE). URL <https://sidra.ibge.gov.br/tabela/3939#resultado> (accessed 6.25.21)
- lafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951. <https://doi.org/10.1038/ng1416>
- Lee, M.A., Cullen, N.G., Newman, S.A.N., Dodds, K.G., McEwan, J.C., Shackell, G.H., 2015. Genetic analysis and genomic selection of stayability and productive life in New Zealand ewes. *J. Anim. Sci.* 93, 3268–3277. <https://doi.org/10.2527/jas.2014-8259>
- Lee, Y.L., Bosse, M., Mullaart, E., Groenen, M.A.M., Veerkamp, R.F., Bouwman, A.C., 2020. Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC Genomics* 21. <https://doi.org/10.1186/s12864-020-6496-1>
- McLaren, A., Kaseja, K., Yates, J., Mucha, S., Lambe, N.R., Conington, J., 2018. New mastitis phenotypes suitable for genomic selection in meat sheep and their genetic relationships with udder conformation and lamb live weights. *Animal* 12, 2470–2479. <https://doi.org/10.1017/S1751731118000393>
- Meira, A.N., Montenegro, H., Coutinho, L.L., Mourão, G.B., Azevedo, H.C., Muniz, E.N., Machado, A.L., Sousa-Jr, L.P., Pedrosa, V.B., Pinto, L.F.B., 2019. Single nucleotide polymorphisms in the growth hormone and IGF type-1 (IGF1) genes associated with carcass traits in Santa Ines sheep. *Animal* 13, 460–468. <https://doi.org/10.1017/S1751731118001362>
- Meuwissen, T., 2007. Genomic selection : Marker assisted selection on a genome wide scale. *J. Anim. Breed. Genet.* 124, 321–322. <https://doi.org/10.1111/j.1439-0388.2007.00708.x>
- Meuwissen, T., Goddard, M., 2010. Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. *Genetics* 185, 623–631. <https://doi.org/10.1534/genetics.110.116590>

- Ministério da Agricultura, Pecuária e Abastecimento (MAPA). AGROSTAT – Estatísticas de comércio Exterior do Agronegócio Brasileiro. URL <http://indicadores.agricultura.gov.br/agrostat/index.htm> (accessed 6.25.21)
- Organization for Economic Co-operation and Development (OECD). Meat consumption (indicator). URL <https://data.oecd.org/agroutput/meat-consumption.htm> (accessed 6.25.21)
- Rao, Y.S., Li, J., Zhang, R., Lin, X.R., Xu, J.G., Xie, L., Xu, Z.Q., Wang, L., Gan, J.K., Xie, X.J., He, J., Zhang, X.Q., 2016. Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. *Poult. Sci.* 95, 1750–1756. <https://doi.org/10.3382/ps/pew136>
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwork, C., Yang, F., Zhang, Junjun, Zerjal, T., Zhang, Jane, Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. *Nature* 444, 444–454. <https://doi.org/10.1038/nature05329>
- Rovadoski, G.A., Pertile, S.F.N., Alvarenga, A.B., Cesar, A.S.M., Pértile, F., Petrini, J., Franzo, V., Soares, W.V.B., Morota, G., Spangler, M.L., Pinto, L.F.B., Carvalho, G.G.P., Lanna, D.P.D., Coutinho, L.L., Mourão, G.B., 2018. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics* 19, 1–14. <https://doi.org/10.1186/s12864-018-4777-8>
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M., 2004. Large-scale copy number polymorphism in the human genome. *Science* (80-.). 305, 525–528. <https://doi.org/10.1126/science.1098918>
- Sorio, A., Rasi, L., 2010. Ovinocultura e abate clandestino: um problema fiscal ou uma solução de mercado? *Rev. Política Agrícola* 19, 71–83.
- Stock, K., Reents, R., 2013. Genomic selection: Status in different species and challenges for breeding. *Reprod. Domest. Anim.* 48, 2–10. <https://doi.org/10.1111/rda.12201>
- USDA, 2021. Production Supply and Distribution Online. Foreign Agricultural Service [WWW Document]. URL <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery> (accessed 6.25.21).
- Van Eenennaam, A.L., 2017. Genetic modification of food animals. *Curr. Opin. Biotechnol.* <https://doi.org/10.1016/j.copbio.2016.10.007>
- Van Eenennaam, A.L., Weigel, K.A., Young, A.E., Cleveland, M.A., Dekkers, J.C.M., 2013. Applied Animal Genomics: Results from the Field. <https://doi.org/10.1146/annurev-animal-022513-114119>
- Viana, J.G.A., 2008. Panorama Geral da Ovinocultura no Mundo e no Brasil, *Revista Ovinos*, Ano.
- Viana, J.G.A., Silveira, V.C.P., 2009. Análise econômica da ovinocultura: estudo de caso na Metade Sul do Rio Grande do Sul, Brazil Economic analysis of sheep production: a case study in the south region of Rio Grande do Sul, Brazil. *Ciência Rural* 39, 1187–1192.
- Wang, K., Chen, Z., Tadesse, M.G., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., Li, M., 2008. Modeling genetic inheritance of copy number variations. *Nucleic Acids Res.* 36, 138. <https://doi.org/10.1093/nar/gkn641>

- Wang, X., Nahashon, S., Feaster, T.K., Bohannon-Stewart, A., Adefope, N., 2010. An initial map of chromosomal segmental copy number variations in the chicken. *BMC Genomics* 11. <https://doi.org/10.1186/1471-2164-11-351>
- Wang, Z., Guo, J., Guo, Y., Yang, Y., Teng, T., Yu, Q., Wang, T., Zhou, M., Zhu, Q., Wang, W., Zhang, Q., Yang, H., 2020. Genome-Wide Detection of CNVs and Association With Body Weight in Sheep Based on 600K SNP Arrays. *Front. Genet.* 11, 558. <https://doi.org/10.3389/fgene.2020.00558>
- Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J.W., Zhang, H., Li, L., Liu, G.E., 2018. Diversity of copy number variation in a worldwide population of sheep. *Genomics* 110, 143–148. <https://doi.org/10.1016/j.ygeno.2017.09.005>
- Zhou, Y., Utsunomiya, Y.T., Xu, L., Hamidi abdel Hay, E., Bickhart, D.M., Almeida Alexandre, P., Rosen, B.D., Schroeder, S.G., Carvalheiro, R., Henrique de Rezende Neves, H., Sonstegard, T.S., Van Tassell, C.P., Bento Stermann Ferraz, J., Fukumasu, H., Fernando Garcia, J., Liu, G.E., 2016. Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. <https://doi.org/10.1186/s12864-016-2461-4>

2. ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA BASEADO EM CNV EM OVINOS SANTA INÊS

RESUMO

O crescente número de estudos sobre variações estruturais do genoma e seus resultados indicam a variação no número de cópias (CNV) como importante fonte de variação genética, com frequência associada a doenças e características produtivas. Entretanto, ainda não foram reportados estudos de associação genômica ampla baseados em CNV em ovinos Santa Inês, uma das raças ovinas mais importantes do Brasil. Assim, o objetivo deste estudo foi investigar a associação entre CNVs e características de desempenho, eficiência e carcaça em ovinos. O Illumina OvineSNP50 BeadChip foi utilizado para detecção de CNVs em 491 ovinos Santa Inês. Os CNVs identificados foram utilizados no estudo de associação genômica ampla por meio de modelo linear misto, considerando uma matriz de relacionamento genômico, para dez características 1) desempenho: peso aos três (P3) e aos seis (P6) meses; 2) eficiência: conversão alimentar residual (CAR) e eficiência alimentar (EA); 3) carcaça: comprimento externo (CE), comprimento da perna (CP), rendimento de carcaça (RC), partes totais de carcaça (PTC), área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS). Um total de 1.167 CNVs autossômicos foram identificados, presentes em 438 ovinos, das quais 294 CNVs são não redundantes, variando de 21,8 a 861,9 kb, compiladas em 216 regiões distintas de variação no número de cópias (CNVRs). Um segmento de CNV no OAR3 foi associado significativamente (p -valor $<0,05$) ao RC, enquanto outro segmento de CNV no OAR6 foi associado (p -valor $<0,05$) ao CAR. Ademais, outras 5 CNVs foram consideradas relevantes e indicadas para estudos futuros. Os segmentos significativos foram sobrepostos a 4 QTLs e 8 genes, entre eles os genes *SPAST*, *TGFA* e *ADGRL3* com funções relacionadas a diferenciação celular e metabolismo energético. Assim, os resultados do presente trabalho incrementam o conhecimento sobre CNVs de ovinos, seus possíveis impactos em características de interesse produtivo e contribui com informações para investigações futuras, e serão úteis especialmente para aquelas com interesse em variações estruturais no genoma ovino.

Palavras-chave: Variação estrutural; CNV; QTLs; Cordeiro

ABSTRACT

The growing number of the genomic structural variations studies and their results indicate copy number variation (CNV) as an important source of genetic variation, frequently associated with diseases and productive traits. However, CNV-based genome-wide association studies (GWAS) in Santa Inês sheep, one of the most important sheep breeds in Brazil, have not yet been reported. Thus, the aim of this study was to investigate the association between CNVs and performance, efficiency and carcass traits in sheep. The Illumina OvineSNP50 BeadChip array was used to detect CNVs in 491 Santa Inês individuals. Then, CNV-based GWAS were

performed with a linear mixed model approach considering a genomic relationship matrix, for ten traits: 1) growth: weight at three (W3) and at six (W6) months of age; 2) efficiency: residual feed intake (RFI) and feed efficiency (FE); and 3) carcass: external length (EL), leg length (LL), carcass yield (CY), total carcass parts (TCP), rib eye area (REA) and subcutaneous fat thickness (SFT). A total of 1,167 autosomal CNVs were identified, present in 438 sheep, with 294 non-redundant CNVs, ranging from 21.8 to 861.9 kb merged into 216 distinct copy number variation regions (CNVRs). One significant CNV segment ($p < 0.05$) in OAR3 was associated with CY, while another one significant CNV segment in OAR6 was associated with RFI. In addition, another 5 CNVs segments were considered relevant for future studies. The significant segments overlapped 4 QTLs and 8 genes, including the *SPAST*, *TGFA* and *ADGRL3*, which have functions related to cell differentiation and energy metabolism. Therefore, the results of the present study increased the knowledge about CNVs in sheep, their possible impacts on productive traits and provide information for future investigations, being especially useful for those interested in structural variations in the sheep genome.

Keywords: Structural variation; CNV; QTLs; Lamb

2.1. Introdução

O consumo mundial de carne ovina é o quarto, antecedido pelo consumo das carnes de frango, bovina e suína (Mazinani e Rude, 2020). Ainda no cenário mundial, a produção de carne ovina em 2019 foi de aproximadamente 10 milhões de toneladas (FAO, 2021), números que destacam a importância econômica do setor para pecuária mundial. Nesse sentido, características de desempenho, eficiência e carcaça são amplamente estudadas em ovinos (da Silva et al., 2012; de O. Nascimento et al., 2020; Rovadoscki et al., 2018), devido aos impactos no sistema produtivo e no mercado consumidor. Assim, essas características podem ser afetadas por vários fatores, como a modulação de níveis de expressão gênica, a variação gênica é um dos mecanismos, que além de persistente é herdável (Navajas, 2014; Tortereau et al., 2020; Zhou et al., 2016).

Nesse contexto, os polimorfismos de nucleotídeos únicos (*Single Nucleotides Polimorfisms* - SNPs) são amplamente utilizados para detecção de variações no genoma, predição do valor genético e estudos de associação genômica ampla (*Genome-wide Association Studies* - GWAS). Também, são utilizados na seleção genômica (*Genomic Selection* - GS), entre outros fatores, por serem abundantes e distribuídos ao longo de todo o genoma, favorecendo que todos QTLs (*Quantitative Trait Loci*) estejam em desequilíbrio de ligação com ao menos um SNP (Meuwissen e Goddard, 2010). Ademais, após mais de uma década de GWAS, aprendemos que a maioria das características estudadas são altamente poligênicas e

influenciadas por muitas variantes genéticas, cada uma de pequeno efeito, com arquiteturas genéticas distintas entre as características (Canela-Xandri et al., 2018; Timpson et al., 2018; Watanabe et al., 2019). Entretanto, outras variações genéticas, como a variação no número de cópias (*Copy Number Variation* – CNV), também são variações estruturais importantes, responsáveis pela manutenção da diversidade, e podem influenciar fenótipos de interesse (Z. Wang et al., 2020).

A variação no número de cópias consiste em um segmento de DNA, com pelo menos 1 kilobase (kb) de comprimento, cujo número de cópias varia em comparação com o genoma referência, devido a eventos de inserções, inversões, translocações, deleções e duplicações (Feuk et al., 2006), de forma que as variações podem ocorrer também entre indivíduos de uma população (Redon et al., 2006). Ademais, CNVs podem ser herdadas através de múltiplas gerações (Redon et al., 2006), embora estudos familiares tenham sugerido que CNVs *de novo* podem ocorrer durante a meiose, com diferenças no número de cópias relatadas entre gêmeos monozigóticos (Tizaoui, 2018). Assim, CNVs apresentaram importância biológica para a manutenção da variabilidade genética populacional, podem ser responsáveis por frações herdáveis do DNA, com possível influência na predição de valores genéticos e seleção de animais. Desse modo, CNVs associadas a fenótipos de interesse econômico são alvos promissores para o melhoramento genético (Clop et al., 2012).

As variações no número de cópias foram objeto de extensa pesquisa nos últimos anos (Pös et al., 2021), especialmente no genoma humano, no qual foram inicialmente descritos (lafrate et al., 2004; Sebat et al., 2004). A identificação de CNVs pode ser realizada por diferentes análises e metodologias disponíveis, entre elas, por meio dos chips de SNPs (Carter Nigel P, 2007; Lin et al., 2013). Nesse contexto, a identificação pode ser realizada por meio de algoritmos específicos, como o PennCNV que utiliza informações da razão log de R (*Log R ratio* - LRR) e a frequência do alelo B (*B Allele Frequency* – BAF) para cada marcador (Wang et al., 2007). Adicionalmente, o PennCNV integra abordagens computacionais ajustando modelos com o conteúdo de GC (guanina-citosina) (L. Xu et al., 2013a), capaz de gerar as “ondas genômicas” (“*genomic waves*”) e interferir na qualidade da chamada (Diskin et al., 2008), também possibilita realizar o controle de qualidade de CNVs e realizar chamadas com alta qualidade (L. Xu et al., 2013b; Zhang et al., 2014).

Foram identificadas CNVs em diferentes populações de frangos (Fernandes et al., 2021; Rao et al., 2016; Wang et al., 2010), equinos (Abri et al., 2020; Ghosh et al., 2014; Solé

et al., 2019), bovinos (Da Silva et al., 2016; Hou et al., 2012; Lee et al., 2020), suínos (Chen et al., 2018; Qiu et al., 2021; Y. Wang et al., 2020), e o total de dez artigos que utilizaram populações de ovinos foram identificados (Andersson et al., 2020; Di et al., 2019; Feng et al., 2020; Fontanesi et al., 2011; Jiang et al., 2019; Liu et al., 2013; Ma et al., 2017; Toremurat et al., 2020; Z. Wang et al., 2020; Yang et al., 2018). Entretanto, a identificação de CNVs e estudos de associação baseadas na variação do número de cópias em ovinos ainda são escassos. Adicionalmente, não foram encontrados relatos de identificação de CNVs em populações de ovinos da raça Santa Inês, uma das mais importantes raças de ovinos do Brasil.

No presente estudo, identificamos CNVs no genoma de uma população de ovinos da raça Santa Inês, realizamos GWAS baseado em informações de CNVs para características de crescimento, desempenho e carcaça. Adicionalmente, detectamos sobreposição de CNVs importantes (p -valor $<0,05$ e p -valor $<0,10$) a regiões de genes conhecidos e associados às características de interesse econômico a fim de estabelecer relações entre as variações genômicas e os fenótipos.

2.2. MATERIAL E MÉTODOS

2.2.1. Descrição da população

A proposta com finalidade de pesquisa científica sob número 07/2016 foi aprovada pela Comissão de Ética no Uso de Animais (CEUA) da Escola de Medicina Veterinária e Zootecnia da Universidade Federal da Bahia. Foram utilizadas informações de 492 animais machos não castrados da raça Santa Inês, criados em sistema de pastagem até o desmame que ocorreu aproximadamente aos 110 dias de idade. As informações de *pedigree* dos cordeiros incluídos neste estudo não estão disponíveis devido aos acasalamentos ocorrerem em sistema de pastagens sem controle genealógico. Os relacionamentos genômicos, estimados por meio de 46.5535 marcadores SNPs, revelaram indivíduos não relacionados, meios-irmãos e irmãos completos, com coeficiente de relacionamento médio entre indivíduos de -0,002 (desvio-padrão = 0,0548; mínimo = -0,123; máximo = 1,223). Adicionalmente, um estudo prévio com essa população reportou baixo coeficiente de endogamia ($F = 0,04$) (Alvarenga et al., 2018).

Após o desmame os animais foram transferidos para a Fazenda Experimental da Universidade Federal da Bahia (UFBA), localizada no município de São Gonçalo dos Campos – BA, Brasil (12°23'57.51" S e 38°52'44.66" O). Os animais foram vermifugados, vacinados e

contra raiva e clostridiose, e suplementados (complexo vitamínico ADE). Os animais foram manejados em cinco lotes. O manejo alimentar foi fundamentado em milho grão, soja grão, caroço de algodão, ureia e suplemento mineral para ovinos, em adição ao feno de Tifton 85, com variações na relação volumoso-concentrado. As dietas consistiram de 2,45 Mcal de energia metabolizável por kg de matéria seca e 15% de proteína bruta e foram formuladas de acordo com o NRC (2007) para proporcionar ganho de peso diário de pelo menos 200 g.

Foram utilizadas informações genômicas de 491 animais para as análises de identificação de variação no número de cópias (CNVs). Para as predições de valores genéticos genômicos (GEBV) e para o estudo de associação genômica ampla baseado em CNVs foram utilizadas informações genômicas e fenotípicas de 491 e 391 cordeiros, respectivamente. Os grupos de contemporâneos (n=35, compostos de 3 a 20 animais) foram formados fundamentado na composição dos lotes e dieta.

2.2.2. Fenótipos

2.2.2.1. Banco de dados

As características abordadas foram 1) desempenho: peso aos três (P3) e aos seis (P6) meses; 2) eficiência: conversão alimentar residual (CAR) e eficiência alimentar (EA); 3) carcaça: comprimento externo (CE), comprimento da perna (CP), rendimento de carcaça (RC), partes totais de carcaça (PTC), área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS).

Foi realizada a checagem de consistência dos dados, etapa na qual foram excluídos animais do banco de dados fenotípico: sem registro para todas as características analisadas, pertencentes a grupos de contemporâneos compostos por menos de três animais e com registros fora da amplitude aceitável (aqui determinada como o valor da média mais ou menos três desvios-padrão).

2.2.2.2. Estimativas e mensurações dos fenótipos

2.2.2.2.1. Estimativa de eficiência alimentar e consumo alimentar residual

O desempenho foi determinado pelo peso individual dos animais no início (P3) e final (P6) do período experimental, em média aos 110 dias e 200 dias, respectivamente, ambos mensurados sempre pela manhã após 16 horas de jejum. O ganho de peso (GP) foi calculado, para cada animal, como a diferença entre P3 e P6. Durante todo o período de confinamento,

a ingestão de matéria seca (IMS) foi calculada como a diferença entre as quantidades de alimento fornecido e de sobra, ajustadas para matéria seca.

Os valores obtidos para a IMS e o GP foram divididos pelo número de dias de confinamento para obtenção dos parâmetros diários, como o consumo de matéria seca (CMS) e ganho médio diário (GMD), os quais foram utilizados para estimar a eficiência alimentar (EA), em quilos (kg), como segue:

$$EA = GMD / CMS \quad [2.1]$$

O consumo alimentar residual (CAR) foi estimado, em kg, com base na diferença entre o consumo de matéria seca observado (CMS) e o consumo de matéria seca predito (CMSpred), como segue:

$$CAR = CMS - CMSpred \quad [2.2]$$

Em que o consumo de matéria seca predito (CMSpred) foi calculado por meio da equação de regressão:

$$y_i = \beta_0 + \beta_1 GMD + \beta_2 PVMM + \varepsilon_i \quad [2.3]$$

em que:

y_i = CMSpred para o i -ésimo animal;

β_0 = intercepto da regressão;

β_1 = coeficiente de regressão parcial de CMSpred sobre *GMD*;

β_2 = coeficiente de regressão parcial de CMSpred sobre o peso vivo metabólico médio $((P3 + P6) / 2)^{0,73}$, *PVMM*) (Knott et al., 2008).

ε_i = erro residual do consumo predito do animal i .

2.2.2.2.2. Mensurações de características de carcaça

Ao final do período experimental, os cordeiros foram transferidos para um matadouro comercial. No momento do abate, os animais foram insensibilizados com descarga elétrica, seguido de sangramento, esfolagem e evisceração. Todos os procedimentos seguiram as diretrizes de manejo e abate humanitário de animais (FAO, 2001).

As carcaças foram pesadas para determinar o peso da carcaça quente (PCQ) e o rendimento da carcaça quente ($RC = (PCQ / P6) * 100$). As medidas morfométricas foram realizadas da seguinte forma: comprimento externo da carcaça (CE) e comprimento da perna (CP), ambas mensurações realizadas com fita métrica. As carcaças foram divididas longitudinalmente em duas metades. A metade esquerda foi seccionada em cinco cortes

comerciais para avaliação da composição regional da carcaça, como segue: pescoço (entre a primeira e a sétima vértebra cervical), ombro (escápula, úmero e carpo), costelas (entre a 1ª e a 13ª vértebra torácica), lombo (ossos e músculos que compõem as vértebras lombares) e perna (entre a última vértebra lombar e a primeira vértebra sacral). Assim, as partes totais de carcaça (PTC) foram calculadas somando o peso, em quilogramas (kg), da composição regional da metade esquerda da carcaça (i.e. cauda, pescoço, paleta, costela, lombo e pernil). Os lombos direito e esquerdo de cada animal foram embalados, marcados e congelados (–20°C) para análises posteriores.

A área do olho do lombo (AOL) foi determinada por um corte transversal confeccionado entre a 12ª e 13ª vértebras torácicas e delineando a área correspondente à porção cranial do lombo, usando uma folha de transparência. Foram realizadas as seguintes medidas: comprimento (A) e profundidade máxima (B) do músculo *Longissimus dorsi*, em cm, que foram obtidas com a utilização de uma régua. A área de olho de lombo foi calculada pela seguinte fórmula de elipse (Silva Sobrinho, 1999): $AOL = (A / 2 * B / 2) \pi$, em cm². A espessura da gordura subcutânea (EGS) foi medida em mm, com paquímetro digital, a ¾ de distância da face medial do músculo *Longissimus dorsi* em direção ao processo espinhoso.

2.2.3. Extração do DNA, genotipagem e controle de qualidade

O DNA genômico foi extraído por meio de *proteinase K* a partir de amostras de tecido coletadas do músculo *Longissimus dorsi* e armazenadas a –20° C. Após a extração, a concentração e a pureza do DNA foram mensuradas pelo equipamento NanoDrop® (Thermo Fisher Scientific Inc., Waltham, MA, USA) e a integridade do DNA foi analisada por meio de eletroforese em gel de agarose 1%, com uso de corante *GelRed™* (Biotium Inc., Fremont, CA, USA) para análise do DNA genômico sob iluminação UV. Em seguida, o DNA foi diluído, de acordo com o protocolo de genotipagem, para hibridização na matriz do chip *Ovine SNP50 BeadChip* (Illumina Inc., San Diego, CA) e então submetido ao processo de genotipagem. O chip utilizado para obter os dados genômicos continha 54.241 SNPs ao longo de 26 cromossomos autossômicos e dois cromossomos sexuais. Entretanto, exclusivamente os cromossomos autossômicos foram considerados.

O controle de qualidade dos genótipos e indivíduos foi realizado por meio do programa PREGSF90 (Aguilar et al., 2014), removendo SNPs com MAF < 0,02, e desvios de *Hardy-Weinberg* (diferença entre a frequência observada e esperada de heterozigotos)

superior a 0,30, SNPs monomórficos, indivíduos e SNPs com *Call rate* menor que 95%. Dessa forma, os números efetivos de animais e SNPs no banco de dados genotípicos foram 491 e 46.535, respectivamente.

2.2.4. Identificação de CNV

Previamente à chamada de CNV, o *software* GenomeStudio 2.0 foi utilizado para mensurar, para cada SNP, *Log R Ratio* (LRR), uma medida de intensidade de sinal total normalizada, e a Frequência de Alelo B (BAF), uma medida da razão de intensidade alélica normalizada (Wang et al., 2007). A chamada de CNV foi realizada com o uso do PennCNV (v.1.0.5) (Wang et al., 2007). Para explorar as informações disponíveis para cada SNP em sua totalidade potencial, o PennCNV incorpora vários componentes em um Modelo Oculto de Markov (HMM), incluindo o LRR, a BAF, a distância entre os SNPs adjacentes e a frequência de alelo B na população (PFB) (Wang et al., 2007).

A função ‘split_illumina_report.pl’ do PennCNV foi utilizada para separar o arquivo de saída (*Final Report*) do GenomeStudio em um arquivo por amostra contendo informações de LRR e BAF para cada SNP. Então, o programa ‘illumina2preGS’ da família de programas BLUPF90 (Misztal et al., 2009) foi utilizado para converter as informações alélicas presentes no *Final Report* (A/B) em informações ajustáveis à família BLUPF90 (0, 1, 2 e 5). A PFB foi calculada baseada na BAF de cada marcador dos animais por meio da função ‘compile_pfb.pl’ do PennCNV, em seguida, apenas os SNPs que atenderam ao controle de qualidade previamente realizado (PREGSF90) foram mantidos no arquivo PFB. Na sequência, a chamada de CNV foi realizada por meio da função ‘detect_cnv.pl’, do PennCNV, com os parâmetros padrões (*default*) para o HMM. Nesse contexto, a intensidade de sinal pode ser influenciada pela concentração de GC (guanina-citosina) próxima aos marcadores SNPs, de modo que a influência para cada marcador pode ser quantificada em “fatores de onda” (“*wave factor*”) (Diskin et al., 2008). Assim, a opção ‘-gcmodel’ da função ‘detect_cnv.pl’ foi utilizada a fim de corrigir o LRR para cada marcador de todas as amostras, utilizando informações de conteúdo de GC em torno (1-Mb) de cada SNP no genoma ovino.

A filtragem de CNVs foi realizada utilizando a função ‘filter_cnv.pl’ do PennCNV, os critérios de filtragem foram: desvio-padrão (DP) para $LRR \leq 0.30$ (‘-qclrsd 0.3’), BAF drift < 0.01 (‘-qcbafdrift 0.01’), waviness fator ≤ 0.05 (‘-qcwf 0.05’), número de snps por CNV > 3 (‘-numsnp

3'), comprimento > 1 kb ('-length 1k'), número de CNV por amostra < 50 ('-qcnumcnv 50')(Feuk, 2008; Z. Wang et al., 2020).

2.2.5. Compilação de CNVR

Após a filtragem de CNVs, as regiões de CNVs (CNVR) foram determinadas por sobreposição de ao menos um par de bases entre CNVs identificadas em diferentes ovinos utilizando a função `populationRanges(grl, density=0.1)` do pacote `CNVRanger` R/Bioconductor (Da Silva et al., 2020). As regiões genômicas com densidade menores que 10% foram deletadas a fim de evitar regiões de falso positivo. As CNVRs foram classificadas em regiões de duplicação ou deleção. Adicionalmente, a compilação de segmentos de CNVs de duplicação e segmentos de CNVs de deleção originaram CNVRs denominadas complexas. A frequência de cada CNVR foi estimada baseada no número de amostras mapeadas na região da CNVRs.

2.2.6. Estudo de associação genômica ampla baseado em CNV (CNV-GWAS)

Foi realizado o estudo de associação genômica ampla entre os GEBVs preditos de cada animal para cada característica (desempenho, eficiência e carcaça) e CNVs detectados na população de ovinos Santa Inês. As predições dos GEBVs foram realizadas segundo o seguinte modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_c\mathbf{c} + \boldsymbol{\varepsilon} \quad [2.4]$$

onde $\mathbf{y}$ é o vetor de fenótipos; $\mathbf{X}$ é a matriz de incidência sobre os efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos (GC, composto por lote e dieta); $\mathbf{Z}_a$ e $\mathbf{Z}_c$ são as matrizes de incidência sobre os efeitos aleatórios, onde $\mathbf{a}$ é o vetor dos efeitos genéticos aditivos e $\mathbf{c}$ é o vetor dos efeitos de CNV; e $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais. Os componentes aleatórios do modelo seguem distribuição normal com $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$, $\mathbf{c} \sim N(0, \mathbf{G}_c\sigma_c^2)$, onde $\mathbf{G}_c$ é a matriz de relacionamento fundamentada nas informações de CNV, e $\boldsymbol{\varepsilon} \sim NID(0, \mathbf{I}\sigma_e^2)$, onde σ_a^2 e σ_c^2 são variâncias genéticas e σ_e^2 é a variância residual, respectivamente, $\mathbf{G}$ é a matriz de relacionamento genômico fundamentada em marcadores SNP, e $\mathbf{I}$ é uma matriz identidade. As matrizes $\mathbf{G}_c$ e $\mathbf{G}$ foram construídas de acordo com a metodologia proposta por VanRaden,

2008. Para as predições de GEBVs foram utilizados o total de 491 animais genotipados, dos quais 391 apresentaram informações fenotípicas.

Após a identificação dos segmentos na população de Santa Inês, a análise de associação foi desenvolvida como segue: 1) as probes dentro das CNVs foram identificadas com o correspondente número de cópias (0n, 1n, 2n, 3n); 2) todas as probes que desviaram do estado normal (2n) foram selecionadas; 3) O estudo de associação genômica ampla (GWAS) foi desenvolvido por meio do pacote CNVRanger R/Bioconductor (Da Silva et al., 2020); 4) assumimos, com base no genótipo da CNV, que probes subsequentes com genótipo idêntico ($\geq 95\%$) a nossa população foram concatenadas em segmentos de CNVs; 5) um p-valor bruto foi gerado independentemente para cada probe. Em seguida, os valores brutos de p foram corrigidos por meio da inflação genômica e, a probe com o menor valor de p foi selecionada para representar o segmento de CNV; 6) os p-valores do segmento CNV foram corrigidos para testes múltiplos, controlando a taxa de descoberta falsa (*False Discovery Rate* – FDR) (Benjamini e Hochberg, 1995); 7) então, estabelecemos dois diferentes limiares para determinar a associação utilizando o p-valor corrigido (FDR). O mais abrangente corresponde a uma associação sugestiva ($p\text{-valor} < 0,1$) e indica CNVs com relativa importância. O menos abrangente corresponde a uma associação genuinamente significativa ($p\text{-valor} < 0,05$). Os limiares servem como guias para os pesquisadores que tentam decidir quais regiões do genoma devem ser mais estudadas (Duggal et al., 2008). Assim, no presente estudo, regiões de CNVs com $p\text{-valor} < 0,1$ foram consideradas sugestivamente associadas e com potencial para serem incluídas em estudos futuros, enquanto regiões com $p\text{-valor} < 0,05$ foram consideradas significantes.

2.2.7. Sobreposição de CNVs com QTLs conhecidos

A sobreposição de segmentos de CNVs, considerados significantes ($p\text{-valor} < 0,05$) ou sugestivos ($p\text{-valor} < 0,01$), com QTLs previamente mapeados para características de crescimento, desempenho e carcaça foram determinados utilizando informações da base de dados *SheepQTLdb* (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>, (Hu et al., 2013)).

2.2.8. Identificação de genes candidatos

A anotação dos genes e a obtenção dos termos de *Gene Ontology* (GO Terms) foi realizada por meio da ferramenta Ensembl Release 104 BioMart (<https://www.ensembl.org/biomart/martview>, (Howe et al., 2021)), considerando janelas de 1Mb (500 kb em ambos os sentidos) entorno das regiões de CNVs significativas ou relevantes considerando as coordenadas da versão Oar_rambouillet_v1.0.

2.3. RESULTADOS

2.3.1. Descrição da população e fenótipos

O banco de dados fenotípicos é composto por informações de características de 1) desempenho: peso aos três meses (P3), peso aos seis meses (P6); 2) consumo: conversão alimentar residual (CAR) e eficiência alimentar (EA); 3) carcaça: comprimento externo (CE), comprimento da perna (CP), rendimento de carcaça (RC), partes totais de carcaça (PTC), área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS). A estatística descritiva para os fenótipos é apresentada na tabela 1.

Tabela 1. Estatística descritiva das características analisadas de desempenho, consumo e carcaça.

Característica ^a	N ^b	Média	Mínimo	Máximo	DP ^c
P3 (kg)	390	23,50	12,00	37,00	4,52
P6 (Kg)	389	38,94	18,80	56,60	6,97
CAR (kg/dia)	384	0,025	-0,397	0,400	0,136
EA (Kg)	388	0,178	0,058	0,334	0,051
CE (cm)	391	52,74	40,00	66,00	5,047
CP (cm)	390	38,25	28,00	48,00	3,43
RC (%)	376	43,99	32,54	56,71	3,776
PTC (kg)	389	8,779	3,726	13,690	1,787
AOL (cm ²)	386	17,66	6,79	42,38	7,912
EGS (mm)	384	2,079	0,910	4,87	0,874

^aP3: peso aos três meses; P6: peso aos seis meses; CAR: consumo alimentar residual; EA: eficiência alimentar; CE: comprimento externo; CP: comprimento da perna; RC: rendimento de carcaça; PTC: partes totais de carcaça; AOL: área de olho de lombo; EGS: espessura de gordura subcutânea. ^bNúmero de fenótipos registrados. ^cDesvio-padrão.

2.3.2. Identificação de CNV

Utilizando o Ovine SNP50 BeadChip (www.illumina.com), após o controle de qualidade inicial, 7.706 SNPs e um animal (*call rate*<95%) não atenderam aos parâmetros e foram removidos das análises. Assim, a chamada de CNV foi realizada considerando 46.535 SNPs e

491 amostras. Inicialmente, foram detectadas 1.523 CNVs presentes em 450 amostras e, após a filtragem, foram identificados 1.167 segmentos de CNVs em 438 cordeiros, sendo 332 deleções e 835 duplicações. Entretanto, uma mesma variação estrutural pode ser identificada em diferentes indivíduos, portanto a identificação de CNVs não redundantes é fundamental para realizar estudo de associação genômica ampla (GWAS) baseada em CNVs. Assim, o presente estudo identificou 294 CNVs não redundantes, sendo 65 deleções (uma 0n e sessenta e quatro 1n) e 229 duplicações (3n), as quais apresentaram tamanho mínimo, médio e máximo de 21,7, 111,6 e 861,9 kb, respectivamente. Um maior número de duplicações também foi reportado por Wang et al., 2020. Nesse contexto, as duplicações tendem a ser mais conservadas comparadas às deleções que são relativamente pobres em genes, contudo as deleções podem desempenhar importante papel na genética de características complexas (Conrad et al., 2005). As CNVs não redundantes somaram 32,799.3 kb ao longo de todo o genoma, o que corresponde a aproximadamente 1,27% do genoma (Oar_rambouillet_v1.0).

2.3.3. Compilação de CNVR

A compilação de regiões de CNV (CNVRs) foi determinada pela sobreposição de ao menos um par de base entre os CNVs detectados. Na população de ovinos da raça Santa Inês analisada, foram detectadas 216 CNVRs, dessas 121, 94 e 1 são classificadas como regiões de duplicação, deleção e complexa (contém segmentos de duplicação e deleção na mesma região), respectivamente. Assim, foram detectadas mais CNVRs de duplicação do que de deleção, entretanto as CNVRs de deleção apresentaram tamanho médio superior às CNVRs de duplicação (157,53 e 140,13 kb, respectivamente).

2.3.4. Estudo de associação genômica ampla (GWAS)

O estudo de associação genômica ampla (GWAS) foi realizado para identificar associações significativas entre os segmentos de CNVs e dez características avaliadas em cordeiros Santa Inês: P3, P6, CAR, EA, CE, CP, RC, PTC, AOL e EGS. Os Manhattan plots contendo segmentos de CNVs significativamente associados (p -valor ajustado para $FDR < 0,05$) e sugestivamente associados (p -valor ajustado para $FDR < 0,10$), segmento indicado para ser abordado em estudos futuros, ao longo dos 26 cromossomos autossômicos estão presentes na figura 1. Adicionalmente, na figura 1 podem ser notados pontos próximos (p -valor ajustado para $FDR < 0,11$) à linha azul (p -valor=0,10) que estão reportados na seção

“Apêndices” deste material. Foram reportados dois segmentos de CNVs significativos, o primeiro (3:98401885-98494873) e o segundo (6:85621179-85774838) associados às características de RC e CAR, respectivamente. Assim, não foram identificadas associações significativas entre segmentos de CNVs e as características P3, P6, EA, CE, CP, PTC, AOL e EGS. Entretanto, foi identificada um segmento de CNV (17:40959249-41028245) sugestivamente associado ao CE e, portanto, relevante para ser abordado em análises futuras. Adicionalmente, foram identificadas outras quatro CNVs com p -valores $< 0,11$. Assim, por este estudo ser o primeiro a reportar CNVs em ovinos Santa Inês, segmentos de CNV com p -valor $> 0,10$ e p -valor $< 0,11$ estão reportados no apêndice A. As CNVs significativa ou sugestivamente associadas às respectivas características, acompanhadas das respectivas probes com menor p -valor, frequências e p -valor ajustado para FDR são apresentadas na Tabela 2.

Figura 1. Manhattan plots para segmentos de CNVs nos 26 cromossomos autossômicos associados a (A) rendimento de carcaça, (B) consumo alimentar residual, (C) comprimento externo e (D) partes totais de carcaça. O eixo X representa os cromossomos autossômicos e o eixo Y indica o $-\log_{10}$ p -valor ajustados para FDR. A linha azul e a vermelha indicam os limiares de 0,10 e 0,05, respectivamente.

Tabela 2. CNVs consideradas associadas (p -valor<0,05), e indicadas para futuros estudos (p -valor<0,10), às respectivas características, rendimento de carcaça (RC), comprimento externo (CE) e consumo alimentar residual (CAR), acompanhadas das respectivas posições genômicas e probes mais significativas, frequência e p -valor ajustado para FDR. As CNVs com associações significativas estão destacadas em negrito.

Característica ^a	OAR: primeira- última posição ^b	Probe	Frequência	p -valor (FDR)
RC	3:98401885- 98494873	s18949.1	6	0.01961
CE	17:40959249- 41028245	OAR17_40959249.1	5	0.05830
CAR	6:85621179- 85774838	OAR6_85621179.1	13	0.03922

^aRC: rendimento de carcaça, CE: comprimento externo, CAR: consumo alimentar residual, PTC: partes totais de carcaça. ^bPosição do segmento de CNV baseada no genoma ovino Oar_rambouillet_v1.0 (NCBI).

2.3.5. CNVs sobrepostos a QTLs conhecidos

Foram identificados o total de 5 QTLs sobrepostos aos segmentos de CNVs com p -valor<0,10 (tabela 3). Nesse contexto, a CNV significativamente associada ao RC foi sobreposta a três QTLs relacionados a gordura interna e ao teor de ácido linoleico conjugado presente na carne. Analogamente, a CNV significativamente associada ao CAR foi sobreposta a um QTL conhecido relacionado ao peso corporal. O segmento de CNV sugestivamente associado ao CE foi sobreposto a um QTL descrito. Adicionalmente, segmentos de CNV com p -valores ajustados para FDR entre 0,10 e 0,11 são apresentados no apêndice B.

Tabela 3. Sobreposição de segmentos de CNVs, considerados significantes (em negrito) ou indicados para estudos futuros, com QTLs previamente mapeados para características de carcaça.

OAR: primeira-última posição ^a	QTL_Ids	Característica relacionada
3:98401885-98494873	QTL:14281	Quantidade de gordura interna
	QTL:14255	Quantidade de gordura interna
	QTL:17220	Teor de ácido linoleico conjugado da carne
17:40959249-41028245	QTL:57756	Produção de leite
6:85621179-85774838	QTL:14284	Peso corporal (abate)

^aPosição do segmento de CNV baseada no genoma ovino Oar_rambouillet_v1.0 (NCBI).

2.3.6. Anotação de genes

Foram identificados 8 genes dentro de uma janela de 1Mb nas regiões genômicas definidas como segmentos de CNVs significantes (p -valor<0,05) associadas ao RC e CAR,

sendo 7 referentes ao RC e 1 ao CAR. Analogamente, para o segmento de CNV relacionados ao CE foram identificados 7 genes.

Tabela 4. Lista de genes dentro da janela de 1Mb para CNVs significativos (em negrito) e indicados para estudos futuros

OAR: primeira- última posição ^a	Característica associada ^b	Gene	Ensembl ID ^c	Entrezgene ID ^d
3:98401885- 98494873	RC	NLRC4	ENSOARG000000010638	101104138
		SLC30A6	ENSOARG000000010663	101104388
		SPAST	ENSOARG000000010758	101104819
		SRD5A2	ENSOARG000000010870	100820756
		TGFA	ENSOARG000000011111	101106508
		XDH	ENSOARG000000010953	780499
		YIPF4	ENSOARG000000010622	101103885
17:40959249- 41028245	CE	ADAD1	ENSOARG000000000116	101104272
		BBS12	ENSOARG000020010513	101103758
		FGF2	ENSOARG000000017222	443306
		IL2	ENSOARG000000000044	443401
		IL21	ENSOARG000000017881	100862657
		NUDT6	ENSOARG000000017455	105602801
		SPATA5	ENSOARG000000017222	101102425
6:85621179- 85774838	CAR	ADGRL3	ENSOARG000000005854	101111884

^aPosição baseada no genoma ovino Oar_rambouillet_v1.0 (NCBI). ^bRC: rendimento de carcaça, CE: comprimento externo, CAR: consumo alimentar residual. ^cIdentificação do gene pela base de dados Ensembl (Ensembl 104 Database). ^dIdentificação do gene pela base de dados NCBI (<https://www.ncbi.nlm.nih.gov/gene>).

2.4. Discussão

Utilizando o OvineSNP50 BeadChip, foi realizada a primeira identificação de CNVs em uma população de ovinos da raça Santa Inês. Assim, devido a variabilidade genética causadas por CNVs e as distinções de padrões genéticos populacionais entre ovinos (Yang et al., 2018), o presente estudo pode contribuir também para futuros estudos funcionais e evolutivos em ovinos.

Liu et al. (2013) e Ma et al. (2015) utilizaram o OvineSNP50 BeadChip e o algoritmo PennCNV para identificar CNVs em 256 e 160 indivíduos, respectivamente. Dessa forma, Liu et al. (2013) identificaram 3,624 CNVs com tamanho médio de 144,6 kb e Ma et al. (2015) identificaram 173 CNVs, dessas 111 são não redundantes com tamanho médio de 123,84 kb. Em geral, diversos fatores podem interferir na identificação do número de cópias, entre eles: 1) o algoritmo utilizado (e.g. cnvPartition, Birdsuite, QuantiSNP e PennCNV), 2) controle de qualidade utilizado, 3) abordagem (e.g. array CGH, SNP array e sequenciamento de próxima

geração (*Next Generation Sequencing* – NGS)), 4) densidade da cobertura do genoma, entre outros (Duan et al., 2013; L. Xu et al., 2013b; Zhang et al., 2014; Zhao et al., 2012). Nesse sentido, é esperado que as quantidades de CNVs total e de CNVs não redundantes sejam variáveis entre os estudos e gerem informações complementares. Adicionalmente, o presente estudo e os desenvolvidos por Liu et al. (2013) e Ma et al. (2015) consideram números diferentes de indivíduos e raças, o que pode influenciar a identificação de CNVs devido, também, os padrões genéticos populacionais, incluindo variações no número de cópias. Entretanto, os tamanhos médios das CNVs detectadas entre os três estudos (Liu et al. (2013); Ma et al. (2015) e o presente estudo) são próximos e diferenciam dos 64,1 e 34,0 kb reportados por Ma et al. (2017) e Wang et al. (2020), que utilizaram chip 600k. Duan et al. (2013) indicam que menor cobertura dificulta a identificação de CNVs menores, assim é provável que a mesma cobertura do genoma tenha favorecido para a semelhança entre os tamanhos médios das CNVs identificadas entre os estudos que utilizaram chip de 50k. Nesse sentido, destaca-se que a compilação de regiões de CNVs (CNVRs) considera as CNVs identificadas e, portanto, segue a mesma ideia quanto à quantidade identificada e as características genômicas.

No presente estudo, o GWAS baseado em CNV revelou um segmento de CNV (3:98401885-98494873) significativamente associado ao rendimento de carcaça sobreposto a três QTLs, sendo dois QTLs biologicamente associados a quantidade de gordura interna e um ao teor de ácido graxo linoleico conjugado na carne. Nesse contexto, o rendimento de carcaça foi calculado como peso de carcaça quente em porcentagem do peso corporal ao abate, de forma que a gordura interna foi considerada no peso ao abate e não considerada para o peso de carcaça quente, portanto, variações na quantidade de gordura interna podem contribuir para variações no rendimento de carcaça. A gordura interna, caracterizada principalmente pela gordura abdominal, renal-inguinal e pélvica, é depositada precocemente comparada às gorduras intermuscular, subcutânea e intramuscular (Paulino et al., 2009; Pethick et al., 2004), sendo potencialmente o tecido adiposo em estágio mais avançado de deposição ao abate (200 dias). Adicionalmente, a gordura interna pode representar de 0,97 a 4,45% do peso de carcaça fria em ovinos de 90 a 195 dias (Westhuizen et al., 2010).

Analogamente, um segmento de CNV (6:85621179-85774838) foi significativamente associado ao consumo alimentar residual, sobreposto a um QTL biologicamente associado ao peso corporal ao abate. O CAR possibilita quantificar a variação entre animais quanto a

eficiência e é bem adequado para uso em avaliação genética e esquemas de melhoramento, porque não deriva a eficiência como uma razão entre a produção e o consumo (Martin et al., 2021). Ademais, existe uma alta correlação genética entre a característica medida no animal jovem e no adulto (Archer e Arthur, 2002), indicando que os processos biológicos que regulam o consumo e a eficiência em animais jovens são semelhantes aos processos que regulam o consumo e a eficiência em animais adultos (Bezerra et al., 2013). Adicionalmente, animais com baixo consumo alimentar residual apresentam menor ingestão de matéria seca, ingestão de energia metabolizável, produção de calor, rúmen e duodeno mais comprido quando comparados a animais de maior CAR (Asher et al., 2018; Zhang et al., 2017) e, venturosamente, não há relação entre CAR e peso corporal ou ganho de peso (Asher et al., 2018; Martin et al., 2021; Zhang et al., 2017), contribuindo para seleção independente entre eficiência e tamanho ou peso. Entretanto, peso corporal é uma característica quantitativa, influenciada por muitos *loci*, assim o QTL sobreposto a CNV (6:85621179-85774838) compõem o grupo de QTLs que atuam sobre o peso corporal. Porém, QTLs e genes localizados dentro ou nas proximidades de CNVs são de especial interesse (Fernandes et al., 2021).

Os segmentos de CNVs significativamente associados ao RC e ao CAR estão sobrepostos a 8 genes, entre eles destacam-se o *SPAST*, *TGFA* e o *ADGRL3*. O *SPAST* codifica para proteínas da família AAA, incluindo ATPases que tem papéis biológicos na dinâmica de microtúbulos, no tráfego de membrana (Kawarai et al., 2017) e catalisam a decomposição de ATP. Em bovinos, bezerros afetados por mutações no *SPAST* podem apresentar sintomas graves de desmielinização espinhal imediatamente após o nascimento (Thomsen et al., 2009). Assim, é possível que variações no gene *SPAST* influencie a síntese de proteínas da família AAA e também sejam responsáveis por danos na mielina, prejudicando o sistema nervoso. O *TGFA* (Transforming Growth Factor alpha) codifica a TGF α que pode atuar como importante fator autócrino para influenciar o crescimento e a diferenciação das células tecaais durante o desenvolvimento do folículo (Roberts e Skinner, 1991), ademais os TGFs são expressos por embriões de espécies como bovinos e ovinos, e podem afetar estágios posteriores da embriogênese ou fenótipos pós-natais (Harvey et al., 1995). Em humanos, SNPs associados ao *TGFA* foram significativamente associados à força de preensão manual (Crossland et al., 2020), entretanto os mecanismos de ação desse gene sobre a deposição de musculatura ainda precisam ser melhor compreendidos. Por sua vez, o *ADGRL3* está

sobreposto ao segmento de CNV associado ao CAR. Variantes do *ADGRL3/LPHN3* diminuem a secreção de insulina das células β do pâncreas (Röthe et al., 2019). A insulina promove a entrada de glicose nas células, atua no metabolismo de sacarídeos e, portanto, participa do metabolismo energético. Nesse contexto, a insulina desempenha papel importante no metabolismo de energia (Xi et al., 2015), atuando também nos tecidos muscular e adiposo, os quais são de grande interesse para a produção animal. Adicionalmente, pesquisas relacionam a via de sinalização da insulina com o CAR (Kelly et al., 2010; Knott et al., 2010; Xi et al., 2015). Richardson et al. (2004) relataram correlação positiva entre a concentração de insulina no plasma sanguíneo e o CAR, também indicam correlação genética positiva entre as duas características. É possível que a variação da sensibilidade à insulina entre células de diferentes tecidos (e.g. muscular e adiposo) e entre diferentes indivíduos sejam responsáveis por frações da variação fenotípica relacionadas ao CAR. Dessa forma, genes que são associados a insulina, como o *ADGRL3*, podem ser responsáveis por frações da variação genética associada ao CAR. Entretanto, os mecanismos subjacentes à influência da insulina no CAR ainda não são claros (Xi et al., 2015).

No contexto geral, utilizando o Ovine SNP50 BeadChip, foi realizada a primeira identificação de CNVs em uma população de ovinos da raça Santa Inês. Foram identificadas CNVs e as CNVRs foram compiladas, considerando os cromossomos autossômicos. Foram identificados segmentos de CNVs sobrepostas a QTLs e genes com funções biológicas descritas relacionadas a diferenciação celular e ao metabolismo de energia. Adicionalmente, foram registrados segmentos de CNVs, e suas sobreposições a QTLs e genes, considerados relevantes para estudos futuros. Assim, os resultados indicam que variações no número de cópias próximas aos genes abordados podem contribuir para variações fenotípicas, contribuir para o entendimento da arquitetura genética e compreensão das características de carcaça e desempenho em ovinos, especialmente para a raça Santa Inês.

2.5. Conclusão

Pela primeira vez em uma população de ovinos Santa Inês, foram identificadas CNVs e compiladas CNVRs, de modo que os segmentos de CNVs foram utilizados no estudo de associação genômica ampla. Dessa forma, foram identificados segmentos de CNVs significativamente associados (p -valor<0,05) ao rendimento de carcaça e ao consumo alimentar residual que revelaram estar sobrepostos a QTLs e genes previamente conhecidos e

associados a funções biológicas da diferenciação celular (células tecais) e metabolismo energético (via de sinalização da insulina). Também foram reportados QTLs e genes sobrepostos ao segmento de CNVs sugestivamente associado (p -valor $<0,10$) ao comprimento externo. Portanto, este estudo revela novas percepções sobre o efeito genético de CNVs em ovinos e seus possíveis impactos em características produtivas. Os resultados contribuem para pesquisas futuras, principalmente especialmente para aquelas com interesse em variações estruturais no genoma ovino.

Referências

- Abri, M.A. Al, Holl, H.M., Kalla, S.E., Sutter, N.B., Brooks, S.A., 2020. Whole genome detection of sequence and structural polymorphism in six diverse horses. *PLoS One* 15, e0230899. <https://doi.org/10.1371/journal.pone.0230899>
- Aguilar, Ignacio, Misztal, Ignacy, Tsuruta, Shogo, Legarra, Andres, Wang, Huiyu, Aguilar, I, Misztal, I, Tsuruta, S, Legarra, A, Wang, H, 2014. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs.
- Alvarenga, A.B., Rovadoscki, G.A., Petrini, J., Coutinho, L.L., Morota, G., Spangler, M.L., Pinto, L.F.B., Carvalho, G.G.P., Mourão, G.B., 2018. Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. *Sci. Rep.* 8, 1–11. <https://doi.org/10.1038/s41598-018-27259-7>
- Andersson, G., Upadhyay, M., Raudsepp, T., Yang, H., Wang, Z., Guo, J., Guo, Y., Yang, Y., Teng, T., Yu, Q., Wang, T., Zhou, M., Zhu, Q., Wang, W., Zhang, Q., 2020. Genome-Wide Detection of CNVs and Association With Body Weight in Sheep Based on 600K SNP Arrays. <https://doi.org/10.3389/fgene.2020.00558>
- Archer, J.A., Arthur, P., 2002. Genetic variation in feed intake and efficiency of mature beef cows and relationships with postweaning measurements Genetic technologies to reduce methane emissions from Australian beef cattle View project Studies in Performance and Meat Quality in Beef Cattle View project.
- Asher, A., Shabtay, A., Cohen-Zinder, M., Aharoni, Y., Miron, J., Agmon, R., Halachmi, I., Orlov, A., Haim, A., Tedeschi, L.O., Carstens, G.E., Johnson, K.A., Brosh, A., 2018. Consistency of feed efficiency ranking and mechanisms associated with inter-animal variation among growing calves. *J. Anim. Sci.* 96, 990–1009. <https://doi.org/10.1093/JAS/SKX045>
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Source: *Journal of the Royal Statistical Society. Series B (Methodological)*.
- Bezerra, L.R., Sarmiento, J.L.R., Neto, S.G., Paula, N.R.O. de, Oliveira, R.L., Rêgo, W.M.F. do, 2013. Residual feed intake: a nutritional tool for genetic improvement. *Trop. Anim. Heal. Prod.* 2013 458 45, 1649–1661. <https://doi.org/10.1007/S11250-013-0435-Y>
- Canela-Xandri, O., Rawlik, K., Tenesa, A., 2018. An atlas of genetic associations in UK Biobank. *Nat. Genet.* 50, 1593–1599. <https://doi.org/10.1038/s41588-018-0248-z>
- Carter Nigel P, N.P., 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. <https://doi.org/10.1038/ng2028>

- Chen, C., Liu, C., Xiong, X., Fang, S., Yang, H., Zhang, Z., Ren, J., Guo, Y., Huang, L., 2018. Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet. Sel. Evol.* 50, 72. <https://doi.org/10.1186/s12711-018-0442-6>
- Clop, A., Vidal, O., Amills, M., 2012. Copy number variation in the genomes of domestic animals. *Anim. Genet.* 43, 503–517. <https://doi.org/10.1111/j.1365-2052.2012.02317.x>
- Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., Pritchard, J.K., 2005. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 2006 381 38, 75–81. <https://doi.org/10.1038/ng1697>
- Cortez de Souza, Taiana, Cortez de Souza, Tatiana, Alberto Rovadoscki, G., Lehmann Coutinho, L., Barreto Mourão, G., Miguel Ferreira de Camargo, G., Bernal Costa, R., Giordano Pinto de Carvalho, G., Breno Pedrosa, V., Fernando Batista Pinto, L., 2021. Journal Pre-proof Genome-wide association for plasma urea concentration in sheep. <https://doi.org/10.1016/j.livsci.2021.104483>
- Crossland H., Piasecki J., McCormick D., Phillips B.E., Wilkinson D.J., Smith K., McPhee J.S., Piasecki M., Atherton P.J. Targeted genotype analyses of GWAS-derived lean body mass and handgrip strength-associated single-nucleotide polymorphisms in elite master athletes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 319: R184–R194. First published June 24, 2020; doi:10.1152/ajpregu.00110. 2020.
- da Silva, L.S.A., Fraga, A.B., da Silva, F. de L., Guimarães Beelen, P.M., de Oliveira Silva, R.M., Tonhati, H., Barros, C. da C., 2012. Growth curve in Santa Inês sheep. *Small Rumin. Res.* 105, 182–185. <https://doi.org/10.1016/j.smallrumres.2011.11.024>
- Da Silva, V., Ramos, M., Groenen, M., Crooijmans, R., Johansson, A., Regitano, L., Coutinho, L., Zimmer, R., Waldron, L., Geistlinger, L., 2020. CNVRanger: Association analysis of CNVs with gene expression and quantitative phenotypes. *Bioinformatics* 36, 972–973. <https://doi.org/10.1093/bioinformatics/btz632>
- Da Silva, V.H., De Almeida Regitano, L.C., Geistlinger, L., Pértile, F., Giachetto, P.F., Brassaloti, R.A., Morosini, N.S., Zimmer, R., Coutinho, L.L., 2016. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. *PLoS One* 11, 1–22. <https://doi.org/10.1371/journal.pone.0157711>
- de O. Nascimento, C., Santos, S.A., dos S. Pina, D., Tosto, M.S.L., Pinto, L.F.B., Eiras, D.N., de Assis, D.Y.C., Perazzo, A.F., de Araújo, M.L.M.L., Azevêdo, J.A.G., Mourão, G.B., de Carvalho, G.G.P., 2020. Effect of roughage-to-concentrate ratios combined with different preserved tropical forages on the productive performance of feedlot lambs. *Small Rumin. Res.* 182, 15–21. <https://doi.org/10.1016/j.smallrumres.2019.11.002>
- Di, R., Id, G., Sutera, A.M., Mastrangelo Id, S., Tolone, M., Portolano, B., Sottileid, G., Bagnato, A., Strillacciid, M.G., Sardina, M.T., 2019. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. <https://doi.org/10.1371/journal.pone.0215204>
- Diskin, S.J., Li, M., Hou, C., Yang, S., Glessner, J., Hakonarson, H., Bucan, M., Maris, J.M., Wang, K., 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* 36, 126. <https://doi.org/10.1093/nar/gkn556>
- Duan, J., Zhang, J.-G., Deng, H.-W., Wang, Y.-P., 2013. Comparative Studies of Copy Number Variation Detection Methods for Next-Generation Sequencing Technologies. *PLoS One* 8, 59128. <https://doi.org/10.1371/journal.pone.0059128>

- Duggal, P., Gillanders, E.M., Holmes, T.N., Bailey-Wilson, J.E., 2008. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9, 516. <https://doi.org/10.1186/1471-2164-9-516>
- Dunshea, F.R., D'Souza, D.N., Pethick, D.W., Harper, G.S., Warner, R.D., 2005. Effects of dietary factors and other metabolic modifiers on quality and nutritional value of meat. *Meat Sci.* 71, 8–38. <https://doi.org/10.1016/J.MEATSCI.2005.05.001>
- FAO, 2021. Food and Agriculture Organization of the United Nations. FAOSTAT, Statistical Database [WWW Document]. URL <http://www.fao.org/faostat/en/#data/QL> (accessed 6.25.21).
- Feng, Z., Li, X., Cheng, J., Jiang, R., Huang, R., Wang, D., Huang, Y., Pi, L., Hu, L., Chen, H., 2020. Copy number variation of the piggy gene in sheep and its association analysis with growth traits. *Animals* 10, 688. <https://doi.org/10.3390/ani10040688>
- Fernandes, A.C., da Silva, V.H., Goes, C.P., Moreira, G.C.M., Godoy, T.F., Ibelli, A.M.G., Peixoto, J. de O., Cantão, M.E., Ledur, M.C., de Rezende, F.M., Coutinho, L.L., 2021. Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics* 22, 354. <https://doi.org/10.1186/s12864-021-07676-1>
- Feuk, L., 2008. Copy number variation in the autism genome. *Expert Opin. Med. Diagn.* 2, 417–428. <https://doi.org/10.1517/17530059.2.4.417>
- Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. <https://doi.org/10.1038/nrg1767>
- Fontanesi, L., Beretti, F., Martelli, P.L., Colombo, M., Dall'Olio, S., Occidente, M., Portolano, B., Casadio, R., Matassino, D., Russo, V., 2011. A first comparative map of copy number variations in the sheep genome. *Genomics* 97, 158–165. <https://doi.org/10.1016/j.ygeno.2010.11.005>
- Food and Agriculture Organization—FAO, 2001. Guidelines for humane handling, transport and slaughter of livestock Guidelines for humane handling, transport and slaughter of livestock. Bangkok: Food and Agriculture Organization of United Nation.
- Ghosh, S., Qu, Z., Das, P.J., Fang, E., Juras, R., Cothran, E.G., McDonell, S., Kenney, D.G., Lear, T.L., Adelson, D.L., Chowdhary, B.P., Raudsepp, T., 2014. Copy Number Variation in the Horse Genome. *PLoS Genet.* 10, 1004712. <https://doi.org/10.1371/journal.pgen.1004712>
- Harvey, M.B., Leco, K.J., Arcellana-Panlilio, M.Y., Zhang, X., Edwards, D.R., Schurtz, G.A., 1995. Roles of growth factors during peri-implantation development. *Mol. Hum. Reprod.* 1, 14–20. <https://doi.org/10.1093/MOLEHR/1.1.14>
- Hay, E.H.A., Utsunomiya, Y.T., Xu, L., Zhou, Y., Neves, H.H.R., Carneiro, R., Bickhart, D.M., Ma, L., Garcia, J.F., Liu, G.E., 2018. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. *BMC Genomics* 19, 1–8. <https://doi.org/10.1186/s12864-018-4787-6>
- Hou, Y., Bickhart, D.M., Chung, H., Hutchison, J.L., Norman, H.D., Connor, E.E., Liu, G.E., 2012. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. *Funct. Integr. Genomics* 12, 717–723. <https://doi.org/10.1007/s10142-012-0295-y>

- Howe, K.L., Achuthan, P., Allen, James, Allen, Jamie, Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., Gall, A., Garcia Giron, C., Grego, T., Guijarro-Clarke, C., Haggerty, L., Hemrom, A., Hourlier, T., Izuogu, O.G., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le, T., Lemos, D., Gonzalez Martinez, J., Marugán, J.C., Maurel, T., McMahon, A.C., Mohanan, S., Moore, B., Muffato, M., Oheh, D.N., Paraschas, D., Parker, A., Parton, A., Prosovetskaia, I., Sakthivel, M.P., Salam, A.I.A., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Steed, E., Szpak, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Walts, B., Winterbottom, A., Chakiachvili, M., Chaubal, A., De Silva, N., Flint, B., Frankish, A., Hunt, S.E., Ilesley, G.R., Langridge, N., Loveland, J.E., Martin, F.J., Mudge, J.M., Morales, J., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, S.J., Cunningham, F., Yates, A.D., Zerbino, D.R., Flicek, P., 2021. Ensembl 2021. *Nucleic Acids Res.* 49, D884–D891. <https://doi.org/10.1093/NAR/GKAA942>
- Hu, Z.-L., Park, C.A., Wu, X.-L., Reecy, J.M., 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 41, D871–D879. <https://doi.org/10.1093/nar/gks1150>
- lafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951. <https://doi.org/10.1038/ng1416>
- Jiang, R., Cheng, J., Cao, X.K., Ma, Y.L., Chaogetu, B., Huang, Y.Z., Lan, X.Y., Lei, C.Z., Hu, L.Y., Chen, H., 2019. Copy number variation of the SHE gene in sheep and its association with economic traits. *Animals* 9. <https://doi.org/10.3390/ani9080531>
- Kawarai, T., Montecchiani, C., Miyamoto, R., Gaudiello, F., Caltagirone, C., Izumi, Y., Kaji, R., Orlacchio, A., 2017. Spastic paraplegia type 4: A novel SPAST splice site donor mutation and expansion of the phenotype variability. *J. Neurol. Sci.* 380, 92–97. <https://doi.org/10.1016/J.JNS.2017.07.011>
- Kelly, A.K., McGee, M., Crews, D.H., Sweeney, T., Boland, T.M., Kenny, D.A., 2010. Repeatability of feed efficiency, carcass ultrasound, feeding behavior, and blood metabolic variables in finishing heifers divergently selected for residual feed intake. *J. Anim. Sci.* 88, 3214–3225. <https://doi.org/10.2527/JAS.2009-2700>
- Knott, S.A., Cummins, L.J., Dunshea, F.R., Leury, B.J., 2010. Feed efficiency and body composition are related to cortisol response to adrenocorticotropin hormone and insulin-induced hypoglycemia in rams. *Domest. Anim. Endocrinol.* 39, 137–146. <https://doi.org/10.1016/J.DOMANIEND.2010.03.003>
- Knott, S.A., Cummins, L.J., Dunshea, F.R., Leury, B.J., 2008. The use of different models for the estimation of residual feed intake (RFI) as a measure of feed efficiency in meat sheep. *Anim. Feed Sci. Technol.* 143, 242–255. <https://doi.org/10.1016/j.anifeedsci.2007.05.013>
- Lee, Y.L., Bosse, M., Mullaart, E., Groenen, M.A.M., Veerkamp, R.F., Bouwman, A.C., 2020. Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC Genomics* 21. <https://doi.org/10.1186/s12864-020-6496-1>
- Lin, C.-F., Naj, A.C., Wang, L.-S., 2013. Analyzing Copy Number Variation Using SNP Array Data: Protocols for Calling CNV and Association Tests, in: *Current Protocols in Human Genetics*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 1.27.1–1.27.15. <https://doi.org/10.1002/0471142905.hg0127s79>

- Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X., Zhang, S., Zhou, X., Wei, C., Zhao, F., Du, L., 2013. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14. <https://doi.org/10.1186/1471-2164-14-229>
- Ma, Q., Liu, X., Pan, J., Ma, L., Ma, Y., He, X., Zhao, Q., Pu, Y., Li, Y., Jiang, L., 2017. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array. *Sci. Rep.* 7, 1–10. <https://doi.org/10.1038/s41598-017-00847-9>
- Martin, P., Ducrocq, V., Faverdin, P., Friggens, N.C., 2021. Invited review: Disentangling residual feed intake—Insights and approaches to make it more fit for purpose in the modern context. <https://doi.org/10.3168/jds.2020-19844>
- Mazinani, M., Rude, B., 2020. Population, World Production and Quality of Sheep and Goat Products. *Am. J. Anim. Vet. Sci.* 15, 291–299. <https://doi.org/10.3844/ajavsp.2020.291.299>
- McKinney, C., Merriman, M.E., Chapman, P.T., Gow, P.J., Harrison, A.A., Highton, J., Jones, P.B.B., McLean, L., O'Donnell, J.L., Pokorny, V., Spellerberg, M., Stamp, L.K., Willis, J., Steer, S., Merriman, T.R., 2008. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann. Rheum. Dis.* 67, 409–413. <https://doi.org/10.1136/ard.2007.075028>
- Meuwissen, T., Goddard, M., 2010. Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. *Genetics* 185, 623–631. <https://doi.org/10.1534/genetics.110.116590>
- Misztal, I., Legarra, A., Aguilar, I., 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. <https://doi.org/10.3168/jds.2009-2064>
- Moro, C., Cornette, R., Vieaud, A., Bruneau, N., Gourichon, D., Bed'hom, B., Tixier-Boichard, M., 2015. Quantitative effect of a CNV on a morphological trait in chickens. *PLoS One* 10, 1–13. <https://doi.org/10.1371/journal.pone.0118706>
- Navajas, E.A., 2014. ANIMAL BREEDING AND GENETICS | DNA Markers and Marker-Assisted Selection in the Genomic Era, in: *Encyclopedia of Meat Sciences*. Elsevier, pp. 12–18. <https://doi.org/10.1016/B978-0-12-384731-7.00002-7>
- Paulino, P.V.R., Valadares Filho, S. de C., Detmann, E., Valadares, R.F.D., Fonseca, M.A., Marcondes, M.I., 2009. Deposição de tecidos e componentes químicos corporais em bovinos Nelore de diferentes classes sexuais. *Rev. Bras. Zootec.* 38, 2516–2524. <https://doi.org/10.1590/S1516-35982009001200030>
- Pethick, D.W., Harper, G.S., Oddy, V.H., 2004. Growth, development and nutritional manipulation of marbling in cattle: a review. *Aust. J. Exp. Agric.* 44, 705–715. <https://doi.org/10.1071/EA02165>
- Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., Szemes, T., 2021. Copy number variation: Characteristics, evolutionary and pathological aspects. *Biomed. J.* <https://doi.org/10.1016/j.bj.2021.02.003>
- Qiu, Y., Ding, R., Zhuang, Z., Wu, J., Yang, M., Zhou, S., Ye, Y., Geng, Q., Xu, Z., Huang, S., Cai, G., Wu, Z., Yang, J., 2021. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs. *BMC Genomics* 22, 1–16. <https://doi.org/10.1186/s12864-021-07654-7>

- Rao, Y.S., Li, J., Zhang, R., Lin, X.R., Xu, J.G., Xie, L., Xu, Z.Q., Wang, L., Gan, J.K., Xie, X.J., He, J., Zhang, X.Q., 2016. Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. *Poult. Sci.* 95, 1750–1756. <https://doi.org/10.3382/ps/pew136>
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, Junjun, Zerjal, T., Zhang, Jane, Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. *Nature* 444, 444–454. <https://doi.org/10.1038/nature05329>
- Richardson, E.C., Herd, R.M., Archer, J.A., Arthur, P.F., 2004. Metabolic differences in Angus steers divergently selected for residual feed intake. *Aust. J. Exp. Agric.* 44, 441–452. <https://doi.org/10.1071/EA02219>
- Roberts, A.J., Skinner, M.K., 1991. Transforming Growth Factor- α and - β Differentially Regulate Growth and Steroidogenesis of Bovine Thecal Cells during Antral Follicle Development. *Endocrinology* 129, 2041–2048. <https://doi.org/10.1210/ENDO-129-4-2041>
- Röthe, J., Thor, D., Winkler, J., Knierim, A.B., Binder, C., Huth, S., Kraft, R., Rothmund, S., Schöneberg, T., Prömel, S., 2019. Involvement of the Adhesion GPCRs Latrophilins in the Regulation of Insulin Release. *Cell Rep.* 26, 1573–1584.e5. <https://doi.org/10.1016/J.CELREP.2019.01.040>
- Rovadoscki, G.A., Pertile, S.F.N., Alvarenga, A.B., Cesar, A.S.M., Pértile, F., Petrini, J., Franzo, V., Soares, W.V.B., Morota, G., Spangler, M.L., Pinto, L.F.B., Carvalho, G.G.P., Lanna, D.P.D., Coutinho, L.L., Mourão, G.B., 2018. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics* 19, 1–14. <https://doi.org/10.1186/s12864-018-4777-8>
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M., 2004. Large-scale copy number polymorphism in the human genome. *Science* (80-.). 305, 525–528. <https://doi.org/10.1126/science.1098918>
- Silva Sobrinho, A.G., 1999. Body composition and characteristics of carcasses from lambs of different genotypes and age at slaughter.
- Solé, M., Ablondi, M., Binzer-Panchal, A., Velie, B.D., Hollfelder, N., Buys, N., Ducro, B.J., François, L., Janssens, S., Schurink, A., Viklund, Å., Eriksson, S., Isaksson, A., Kultima, H., Mikko, S., Lindgren, G., 2019. Inter- And intra-breed genome-wide copy number diversity in a large cohort of European equine breeds. *BMC Genomics* 20, 1–12. <https://doi.org/10.1186/s12864-019-6141-z>
- Thapar, A., Cooper, M., 2013. Copy Number Variation: What Is It and What Has It Told Us About Child Psychiatric Disorders? *J. Am. Acad. Child Adolesc. Psychiatry* 52, 772. <https://doi.org/10.1016/J.JAAC.2013.05.013>
- Thomsen, B., Nissen, P.H., Agerholm, J.S., Bendixen, C., 2009. Congenital bovine spinal dysmyelination is caused by a missense mutation in the SPAST gene. *neurogenetics* 2009 112 11, 175–183. <https://doi.org/10.1007/S10048-009-0214-0>

- Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J., Richards, J.B., 2018. Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* <https://doi.org/10.1038/nrg.2017.101>
- Tizaoui, K., 2018. De novo vs. inherited copy number variations in multiple sclerosis susceptibility. *Cell. Mol. Immunol.* 15, 812–814. <https://doi.org/10.1038/cmi.2017.166>
- Toremurat, Z., Ibrahim, E., Huang, Y.-Z., Lan, X., Pi, L., Chaogetu, B., Hu, L., Chen, H., 2020. Animal Biotechnology Copy number variations of TOP2B gene are associated with growth traits in Chinese sheep breeds Copy number variations of TOP2B gene are associated with growth traits in Chinese sheep breeds. <https://doi.org/10.1080/10495398.2020.1773490>
- Tortereau, F., Marie-Etancelin, C., Weisbecker, J.L., Marcon, D., Bouvier, F., Moreno-Romieux, C., François, D., 2020. Genetic parameters for feed efficiency in Romane rams and responses to single-generation selection. *Animal* 14, 681–687. <https://doi.org/10.1017/S1751731119002544>
- VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665–1674. <https://doi.org/10.1101/gr.6861907>
- Wang, X., Nahashon, S., Feaster, T.K., Bohannon-Stewart, A., Adefope, N., 2010. An initial map of chromosomal segmental copy number variations in the chicken. *BMC Genomics* 11. <https://doi.org/10.1186/1471-2164-11-351>
- Wang, Y., Zhang, T., Wang, C., 2020. Detection and analysis of genome-wide copy number variation in the pig genome using an 80 K SNP Beadchip. *J. Anim. Breed. Genet.* 137, 166–176. <https://doi.org/10.1111/jbg.12435>
- Wang, Z., Guo, J., Guo, Y., Yang, Y., Teng, T., Yu, Q., Wang, T., Zhou, M., Zhu, Q., Wang, W., Zhang, Q., Yang, H., 2020. Genome-Wide Detection of CNVs and Association With Body Weight in Sheep Based on 600K SNP Arrays. *Front. Genet.* 11, 558. <https://doi.org/10.3389/fgene.2020.00558>
- Watanabe, K., Stringer, S., Frei, O., Umičević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., Posthuma, D., 2019. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* 51, 1339–1348. <https://doi.org/10.1038/s41588-019-0481-0>
- Xi, Y.M., Yang, Z., Wu, F., Han, Z.Y., Wang, G.L., 2015. Gene expression profiling of hormonal regulation related to the residual feed intake of Holstein cattle. *Biochem. Biophys. Res. Commun.* 465, 19–25. <https://doi.org/10.1016/J.BBRC.2015.07.092>
- Xu, L., Hou, Y., Bickhart, D., Song, J., Liu, G., 2013a. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. *Microarrays* 2, 171–185. <https://doi.org/10.3390/microarrays2030171>
- Xu, L., Hou, Y., Bickhart, D.M., Song, J., Liu, G.E., 2013b. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data 2, 171–185. <https://doi.org/10.3390/microarrays2030171>
- Xu, Y., Zhang, L., Shi, T., Zhou, Y., Cai, H., Lan, X., Zhang, C., Lei, C., Chen, H., 2013. Copy number variations of MICAL-L2 shaping gene expression contribute to different phenotypes of cattle. <https://doi.org/10.1007/s00335-013-9483-x>

- Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J.W., Zhang, H., Li, L., Liu, G.E., 2018. Diversity of copy number variation in a worldwide population of sheep. *Genomics* 110, 143–148. <https://doi.org/10.1016/j.ygeno.2017.09.005>
- Zhang, X., Du, R., Li, S., Zhang, F., Li, J., Wang, H., 2014. Evaluation of copy number variation detection for a SNP array platform. <https://doi.org/10.1186/1471-2105-15-50>
- Zhang, X., Wang, W., Mo, F., La, Y., Li, C., Li, F., 2017. Association of residual feed intake with growth and slaughtering performance, blood metabolism, and body composition in growing lambs. *Sci. Reports* 2017 7, 1–11. <https://doi.org/10.1038/s41598-017-13042-7>
- Zhao, M., Wang, Qingguo, Wang, Quan, Jia, P., Zhao, Z., 2012. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives 4–7. <https://doi.org/10.1186/1471-2105-14-S11-S1>
- Zhou, Y., Utsunomiya, Y.T., Xu, L., Hay, E.H. abdel, Bickhart, D.M., Alexandre, P.A., Rosen, B.D., Schroeder, S.G., Carvalheiro, R., de Rezende Neves, H.H., Sonstegard, T.S., Van Tassell, C.P., Ferraz, J.B.S., Fukumasu, H., Garcia, J.F., Liu, G.E., 2016. Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. *BMC Genomics* 17, 1–9. <https://doi.org/10.1186/s12864-016-2461-4>
- Zhu, C., Fan, H., Yuan, Z., Hu, S., Ma, X., Xuan, J., Wang, H., Zhang, L., Wei, C., Zhang, Q., Zhao, F., Du, L., 2016. Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays. *Sci. Rep.* 6, 1–9. <https://doi.org/10.1038/srep27822>

3. PREDIÇÃO GENÔMICA COMBINANDO MARCADORES SNP E VARIAÇÃO NO NÚMERO DE CÓPIAS EM OVINOS SANTA INÊS

RESUMO

Variações no número de cópias (CNV) são variações estruturais no DNA de no mínimo 1 kb que diferem do genoma referência quanto ao número de cópias. O crescente número de estudos com CNV reportam sua associação com doenças em humanos e diversas características quantitativas em animais domésticos. Entretanto, pode ser que marcadores SNPs, amplamente utilizados para predição de GEBVs, não sejam suficientes para capturar o efeito de CNV em sua totalidade. Portanto, o objetivo deste estudo foi avaliar o ganho em acurácia devido à inclusão de efeito de CNV no modelo de predição de GEBV. Foram utilizadas informações genômicas (OvineSNP50 BeadChip) de 491 cordeiros da raça Santa Inês e 14 características entre desempenho, eficiência e carcaça foram analisadas. Foram comparados dois modelos GBLUP, um contendo exclusivamente informações de SNP (GB_a) e outro contendo informações de SNP e CNV ($GB_a + CNV$). A acurácia de predição foi estimada pela correlação de Pearson entre os GEBVs preditos para a população de treinamento e o fenótipo ajustado para efeitos fixos no grupo de validação. A inclusão de CNV nas análises genéticas permitiram capturar maior variância genética, entretanto a proporção da variância genética capturada por CNVs varia entre as características e não necessariamente incrementa a acurácia de predição. Assim, entre os modelos considerados, não houve incremento na habilidade preditiva ao combinar informações de SNP e CNV. Portanto, os resultados reportados no presente estudo indicam que a inclusão de informações de CNV em modelos para predição de GEBV não deve ser prioridade.

Palavras-chave: Seleção genômica; Habilidade de predição; Viés de predição; Características complexas

ABSTRACT

Copy number variations are structural variations in DNA of at least 1 kb that differ from the reference genome in copy number. The growing number of studies with CNV report its association with diseases in humans and several quantitative traits in livestock animals. However, it is possible that SNPs markers, widely used to predict GEBVs, are not fully to capture the CNV effect. Therefore, the aim of this study was to evaluate the gain in accuracy due to the inclusion of the CNV effect in the GEBV prediction model. Genomic information (OvineSNP50 BeadChip) of 491 Santa Inês lambs were used and 14 traits among performance, efficiency and carcass were analyzed. Two GBLUP models were compared, one containing only SNP information (GB_a) and other one containing SNP and CNV information ($GB_a + CNV$). Prediction accuracy was estimated by Pearson's correlation between predicted GEBVs for training population and the phenotype adjusted for fixed effects in the validation dataset. The inclusion of CNV in the genetic analysis allowed captured

greater genetic variance, however the proportion of genetic variance captured by CNVs differs among traits and does not necessarily increase the prediction accuracy. Among the models considered, there was no increase in predictive ability when combining SNP and CNV information. Therefore, the results reported in the present study indicate that the inclusion of CNV information in models to predict GEBV should not be a priority.

Keywords: Genomic selection; Prediction ability; Prediction bias; Complex traits

3.1. Introdução

As espécies de interesse produtivo foram domesticadas pela primeira vez há cerca de 10.000 anos devido a capacidade de transformar forragem em carne, leite e ovos, todas fontes de proteína de alta qualidade, lipídios e micronutrientes que permitiram aos humanos sobreviver em ambientes variados (Hayes et al., 2013). Desde então, a evolução dos animais domésticos se associou aos objetivos produtivos e foi conduzida, inicialmente, com técnicas simples por meio de observações sobre o que era percebido como a aparência desejada e o acasalamento seletivo de pais escolhidos (Van Eenennaam, 2017). Ao longo do tempo houve incremento no conhecimento a respeito dos sistemas produtivos e suas importantes áreas como nutrição, reprodução, bem-estar, genética e melhoramento genético. Mais especificamente, relacionado ao melhoramento genético, o conhecimento sobre hereditariedade, modelos mistos, inversão de matrizes de relacionamento, componentes de variância e informações genômicas (Henderson, 1975, 1963; Meuwissen et al., 2001; VanRaden, 2008), acompanhado por importantes avanços tecnológicos que permitiram acessá-las e utilizá-las. Assim, observações subjetivas foram substituídas por medições objetivas e métodos estatísticos sofisticados foram implementados para isolar os efeitos genéticos hereditários das influências ambientais com incremento na taxa de ganho genético por unidade de tempo (Van Eenennaam et al., 2013).

Nesse contexto, o melhoramento genético apresenta importante contribuição quanto a evolução produtiva dos animais domésticos. O principal objetivo no melhoramento animal é selecionar indivíduos que tenham melhores valores genéticos (*Breeding Value* – BV) para características de interesse como pais para produzir a próxima geração e fazê-lo o mais rápido possível (Dekkers, 2012). Por exemplo, estudos reportam a evolução produtiva em aves e bovinos leiteiros, com destaque para a raça Holandesa (Havenstein et al., 2003; Wiggans et al., 2017; Zuidhof et al., 2014).

Para animais de produção, as características mais importantes são complexas, quantitativas (características afetadas por muitos genes e pelo ambiente) (Meuwissen e Goddard, 2010), de modo que a seleção para as características de interesse foi primeiramente fundamentada nas informações fenotípicas, de pedigree, e na estimativa dos parâmetros genéticos (Dekkers e Hospital, 2002). Diferentes abordagens e informações (e.g. pedigree, genômicas, ambas simultaneamente) podem ser utilizadas para a predição de valores genéticos estimados (*Estimated Breeding Value* – EBV), valores genéticos genômicos estimados (*Genomic Estimated Breeding Value* – GEV) e estimativas de parâmetros genéticos, podendo resultar em diferenças quanto a acurácia de predição (Daetwyler et al., 2012; Putz et al., 2018).

Atualmente o amplo acesso às informações genômicas permite incorporá-las nas análises genéticas, com possíveis ganhos em acurácia. As informações genômicas podem ser utilizadas para substituir a matriz de relacionamento baseada no pedigree (**A**) (Henderson, 1975) pela matriz de relacionamento genômico (**G**) (VanRaden, 2008) ou pela matriz de relacionamento baseada nos relacionamentos genômicos e de pedigree (**H**) (Aguilar et al., 2010). Nesse contexto, a seleção genômica (*Genomic Selection* – GS) é amplamente aplicada (Blasco e Toro, 2014; Stock e Reents, 2013; Wiggans et al., 2017) e pode ser definida como a seleção simultânea para muitos (dezenas, centenas de milhares) marcadores, que cobrem todo o genoma de uma maneira densa, e espera-se que todos os genes estejam em desequilíbrio de ligação (*Linkage Disequilibrium* – LD) com pelo menos alguns dos marcadores (Meuwissen, 2007). Assim, a GS revolucionou os programas de melhoramento, principalmente devido a aplicação de marcadores de polimorfismo de nucleotídeo único (*Single Nucleotide Polymorphism* – SNP). Entretanto, outras variações genéticas, como a variação no número de cópias (*Copy Number Variation* – CNV), também são responsáveis pela manutenção da diversidade e podem influenciar fenótipos de interesse em diferentes espécies como frangos, bovinos e ovinos (Da Silva et al., 2016; Fernandes et al., 2021; Redon et al., 2006; Yang et al., 2018).

CNV é um segmento de DNA de no mínimo 1 kb e está presente em um número de cópias variável em comparação com um genoma de referência (Feuk et al., 2006). Nos últimos anos os estudos de CNV vem aumentando, foram relatadas quantidades importantes de segmentos de CNV com efeitos genéticos funcionais podendo estar associados à expressão gênica, diversidade e variação fenotípica (Feuk et al., 2006; Henrichsen et al., 2009; Redon et

al., 2006; Wang et al., 2008; Yang et al., 2018; Zhou et al., 2016). Assim, segmentos de CNVs associados a fenótipos de interesse econômico são alvos promissores para o melhoramento genético (Clop et al., 2012). Dessa forma, há possibilidade de que os efeitos genéticos não se devam apenas aos SNPs, mas sejam também influenciados por outros tipos de variantes genéticas, como CNVs, que podem ter os efeitos apenas parcialmente captados por marcadores SNPs (Hay et al., 2018; Lee et al., 2020; Taylor, 2014).

Diante do exposto, no presente estudo foram realizadas predições de valores genéticos genômicos considerando exclusivamente marcadores SNP, modelo GB_a , e combinando marcadores SNP e CNV, modelo $GB_a + CNV$, por meio da inclusão de uma matriz de relacionamento baseada em CNV na análise genética. Ademais, foram comparadas as acurácias dos modelos GB_a e $GB_a + CNV$.

3.2. Materiais e Métodos

3.2.1. População, genotipagem e controle de qualidade

A descrição da população referente às análises genéticas do presente estudo e as informações sobre genotipagem e controle de qualidade são apresentadas nos tópicos 2.2.1 (Descrição da população) e 2.2.3 (Extração do DNA, genotipagem e controle de qualidade) deste material.

3.2.2. Fenótipos

Foram consideradas 14 características: 1) desempenho: peso aos três (P3) e aos seis (P6) meses, ganho de peso (GP), e ganho médio diário (GMD); 2) eficiência: consumo alimentar residual (CAR), conversão alimentar (CA) e eficiência alimentar (EA); 3) carcaça: comprimento externo (CE), comprimento da perna (CP), rendimento de carcaça (RC), partes totais de carcaça (PTC), área de olho de lombo (AOL), espessura de gordura subcutânea (EGS) e peso de carcaça quente (PCQ).

A consistência dos dados fenotípicos foi checada, e animais sem registro para todas as características analisadas, pertencentes a grupos de contemporâneos compostos por menos de três animais ou com registros fora da amplitude aceitável (aqui determinada como o valor da média mais ou menos três desvios-padrão) foram excluídos.

Na sequência, foi realizada uma análise de variância, ANOVA, por meio do software R (R Core Team, 2020) para testar o nível de significância das variáveis lote e dieta. Ambas

foram significativamente diferentes de zero pelo teste F, portanto as variáveis foram agrupadas para compor o grupo de contemporâneos (GC).

3.2.2.1. Estimativas das características de eficiência

O desempenho foi determinado pelo peso individual dos animais, mensurados sempre pela manhã após 16 horas de jejum, no início (P3) e final (P6) do período experimental, em média aos 110 dias e 200 dias, respectivamente. O ganho de peso (GP) foi calculado, para cada animal, como a diferença entre P3 e P6. Durante todo o período de confinamento, a ingestão de matéria seca (IMS) foi calculada como a diferença entre as quantidades de alimento fornecido e de sobra, ajustadas para matéria seca.

Os valores obtidos para a IMS e o GP foram divididos pelo número de dias de confinamento para obtenção dos parâmetros diários, como o consumo de matéria seca (CMS) e ganho médio diário (GMD), os quais foram utilizados para estimar a eficiência alimentar (EA), em quilos (kg), como descrito abaixo:

$$EA = GMD / CMS \quad [3.1]$$

A conversão alimentar (CA) foi estimada, em kg, pela razão entre o CMS e o GMD, como descrito abaixo:

$$CA = CMS / GMD \quad [3.2]$$

O consumo alimentar residual (CAR) foi estimado, em kg, com base na diferença entre o consumo de matéria seca observado (CMS) e o consumo de matéria seca predito (CMSpred), como descrito abaixo:

$$CAR = CMS - CMSpred \quad [3.3]$$

Em que o consumo de matéria seca predito (CMSpred) foi calculado por meio da regressão abaixo:

$$y_i = \beta_0 + \beta_1 GMD + \beta_2 PVMM + \varepsilon_i \quad [3.4]$$

em que:

y_i = CMSpred para o i -ésimo animal;

β_0 = intercepto da regressão;

β_1 = coeficiente de regressão parcial de CMSpred sobre GMD ;

β_2 = coeficiente de regressão parcial sobre o peso vivo metabólico médio $((P3 + P6) / 2)^{0,73}$, $PVMM$) (Knott et al., 2008).

ε_i = erro residual do consumo predito do animal i .

3.2.2.2. Mensurações de características de carcaça

Ao final do período experimental, os cordeiros foram transferidos para um matadouro comercial. No momento do abate, os animais foram insensibilizados com descarga elétrica, seguido de sangramento, esfolagem e evisceração. Todos os procedimentos seguiram as diretrizes de manejo e abate humanitário de animais (FAO, 2001).

As carcaças foram pesadas para determinar o peso da carcaça quente (PCQ) e o rendimento da carcaça quente ($RC = (PCQ / P6) * 100$). As medidas morfométricas foram realizadas da seguinte forma: comprimento externo da carcaça (CE) e comprimento da perna (CP), ambas mensurações realizadas com fita métrica. As carcaças foram divididas longitudinalmente em duas metades. A metade esquerda foi seccionada em cinco cortes comerciais para avaliação da composição regional da carcaça, a saber, pescoço (entre a primeira e a sétima vértebra cervical), ombro (escápula, úmero e carpo), costelas (entre a 1ª e a 13ª vértebra torácica), lombo (ossos e músculos que compõem as vértebras lombares) e perna (entre a última vértebra lombar e a primeira vértebra sacral). Assim, as partes totais de carcaça (PTC) foram calculadas somando o peso, em kg, da composição regional da metade esquerda da carcaça (i.e. cauda, pescoço, paleta, costela, lombo e pernil). Os lombos direito e esquerdo de cada animal foram embalados, marcados e congelados (-20°C) para análises posteriores.

A área do olho do lombo (AOL) foi determinada por um corte transversal confeccionado entre a 12ª e 13ª vértebras torácicas e delineando a área correspondente à porção cranial do lombo, usando uma folha de transparência. Foram realizadas as seguintes medidas: comprimento (A) e profundidade máxima (B) do músculo *longissimus dorsi*, em cm, que foram obtidas com a utilização de uma régua. A área de olho de lombo foi calculada pela seguinte fórmula de elipse (Silva Sobrinho, 1999): $AOL = (A / 2 * B / 2) \pi$, em cm². A espessura da gordura subcutânea (EGS) foi medida em mm, com paquímetro digital, a ¾ de distância da face medial do músculo *longissimus dorsi* em direção ao processo espinhoso.

3.3. Análise genética

3.3.1. Uso de marcadores CNVs para construção da matriz de relacionamento

A estrutura da matriz de marcadores CNVs é semelhante à estrutura de uma matriz de marcadores SNPs, assim a matriz é $m \times n$, onde m representa o número de animais (linhas) e n o número de marcadores não-redundantes (colunas). Definida a estrutura da matriz, os

pacotes ‘data.table’, ‘dplyr’ e ‘tidyr’ do programa R (R Core Team, 2020) foram utilizados para a criação da matriz de marcadores CNVs.

Após a detecção e filtragem de CNVs, o programa PennCNV proporciona diferentes arquivos de saída com informações variadas e completas. O arquivo gerado por meio da opção ‘-out’ da função ‘filter_cnv.pl’, aqui denominado como ‘sampleall_adjusted.goodcnv’, contém informações (e.g. localização, número de SNPs, comprimento, estado (número de cópias, $2 <$ ou > 2), identificação da amostra e start e end probes). Adicionalmente, o arquivo de genótipo (SNP *marker file*) ‘snps2pregs_clean’, gerado após o controle de qualidade dos marcadores SNPs contém informações de identificação dos indivíduos (ID, $n = 491$) e genotípicas fundamentadas em marcadores SNPs. Portanto, o arquivo ‘snps2pregs_clean’ contém informações de identificação (*m*) de animais com ou sem CNVs identificadas no genoma, e o arquivo ‘sampleall_adjusted.goodcnv’ contém informações de CNVs (*n*), ambas informações são necessárias para a construção da matriz de marcadores CNVs.

Inicialmente, os arquivos ‘sampleall_adjusted.goodcnv’ e ‘snps2pregs_clean’ foram carregados. Na sequência, as informações utilizadas do arquivo ‘sampleall_adjusted.goodcnv’ foram ID (exclusivamente para indivíduos com CNV), posição do segmento de CNV (e.g. chr4:24941706-25119971) e números de cópias (CN, variando de 0 a 3), compondo o arquivo aqui denominado ‘cnv.good’. Analogamente, as informações genotípicas do arquivo ‘snps2pregs_clean’ foram removidas.

Uma matriz de marcadores consiste em informações genotípicas de diferentes indivíduos, assim cada indivíduo tem, se não houver informações perdidas, o genótipo descrito de pontos genômicos determinados por marcadores. Nesse sentido, é fundamental que os genótipos estejam corretamente organizados, ou seja, informações genotípicas e marcadores corretamente associados na matriz. A fim de ordenar as CNVs e garantir a integridade das informações dentro da matriz de marcadores, a posição da CNV foi fragmentada em três informações: cromossomo, início e fim, todas elas classificadas como ‘integer’. Por exemplo, ‘chr4:24941706-25119971’ foi fragmentada em ‘4’, ‘24941706’ e ‘25119971’. Então, o arquivo foi ordenado, combinando as funções ‘with’ e ‘order’, considerando o cromossomo e o início de cada CNV. Na sequência, combinando as funções ‘data.frame’, ‘mutate’ e ‘unique’, foi criada uma coluna contendo números sequenciais únicos para cada combinação das informações de cromossomo, início e fim. Portanto, foi criada uma coluna contendo as posições (*n*) dentro da matriz de marcadores para cada CNV não redundante. Depois, de

forma semelhante, o arquivo foi ordenado por ID dos indivíduos e uma coluna contendo números únicos para cada ID foi criada, assim, foi criada uma coluna contendo as posições (m) de cada indivíduo na matriz de marcadores. Adicionalmente, os eventos de deleção ($CN < 2$), composto por uma única deleção completa ($0n$) e deleções parciais ($1n$), e duplicação ($CN > 2$) foram recodificados para 0 e 2, respectivamente. Portanto, o arquivo 'cnv.good' apresenta o código para variações no número de cópias e as coordenadas de onde cada código deve ser inserido na matriz de marcadores CNVs.

As informações de ID presentes no arquivo 'snps2pregs_clean' foram utilizadas para criar um banco de dados ('without.cnv'), com colunas rotuladas como em 'cnv.good', contendo animais sem CNVs identificados ($CN = 2$), com número de cópias recodificado para 1, portanto ausentes em 'sampleall_adjusted.goodcnv'. Na sequência, foi feita a junção dos arquivos 'cnv.good' e 'without.cnv', gerando o arquivo 'cnv.491' com reordenação e inclusão das posições de cada indivíduo na matriz de marcadores. Depois, uma matriz de uns de dimensão $m \times n$ foi criada, e, por meio de loop, as informações de posição de linha (m) e coluna (n) foram utilizadas para realizar a substituição dos valores dentro da matriz de uns pelos valores recodificados (0 e 2) das variações nos números de cópias, gerando, portanto, um uma matriz contendo 294 CNVs não redundantes.

Diante do exposto, a matriz de CNVs não redundantes apresenta as informações genômicas de indivíduos devidamente ordenados nas linhas da matriz. As informações presentes em cada coluna (294 CNVs não redundantes) foram concatenadas (sem espaços entre os códigos 0, 1 e 2) e utilizadas para compor o arquivo de marcadores de CNVs (CNV *marker file*), caracterizado por uma coluna de IDs únicos, ordenados conforme a matriz de marcadores, e outra coluna com informações genotípicas fundamentadas no estado (deleção, normal e duplicação) da variação no número de cópias. Assim, o arquivo de marcadores de CNVs apresenta formatação semelhante ao arquivo de marcadores SNPs, amplamente utilizados em programas da família BLUPF90.

Por fim, utilizando o programa PREGSF90 e a metodologia proposta por VanRaden (2008), foram criadas as matrizes $\mathbf{G}$ e $\mathbf{G}_c$ utilizando os arquivos de marcadores SNP, com controle de qualidade (descritos em 3.3.2), e CNV, sem controle de qualidade adicional nesta etapa, respectivamente.

Em suma, por meio do programa PREGSF90 foi construída a matriz de relacionamento genômico fundamentada em CNV, denominada $\mathbf{G}_c$, conforme VanRaden (2008), utilizando

informações de variações estruturais genômicas de deleção, ausência de variação no número de cópias (normal) e duplicação codificadas como 0, 1 e 2, respectivamente. Portanto, o relacionamento genômico foi estimado com foco na presença ou ausência de variações estruturais, e não nos seus comprimentos (número de pares de base).

3.3.2. Predição dos valores genéticos genômicos (GEBVs)

Foram utilizados os programas da família BLUPF90 (Aguilar et al., 2018) para realizar as análises genéticas considerando os cromossomos autossômicos de 492 ovinos Santa Inês. O controle de qualidade dos genótipos e indivíduos foi realizado por meio do programa PREGSF90 (Aguilar et al., 2014), removendo SNPs com $MAF < 0,02$, e desvios de *Hardy-Weinberg* (diferença entre a frequência observada e esperada de heterozigotos) superior a 0,30, indivíduos e SNPs com *Call rate* menor que 95%. Dessa forma, os números efetivos de animais e SNPs foram 491 e 46.535, respectivamente.

O padrão dos programas da família BLUPF90 é realizar análises genômicas em etapa única (ssGBLUP), que consiste em combinar os relacionamentos de pedigree e genômico, respectivamente inseridos nas matrizes **A** e **G**, em uma matriz única de relacionamento, denominada **H** (Legarra et al., 2009; Lourenco et al., 2020; Misztal et al., 2009). A inversa de **H** pode ser expressa como:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \quad [3.5]$$

em que **A** é a matriz de relacionamento aditivo de pedigree, **G** é a matriz de relacionamento genômico, **A**₂₂ uma submatriz de **A** para os animais genotipados.

Entretanto, a população considerada não possui registro de pedigree e, na ausência da informação de pedigree, o RENUMF90, por padrão, gera um arquivo de pedigree (“dummy” pedigree) contendo a identificação de cada animal genotipado e as informações de pai e mãe zerados, assumindo parentesco desconhecidos. Dessa forma, o método GBLUP foi mimetizado fazendo o uso de um “falso” *single-step* GBLUP (Aguilar, 2014). A matriz **G** foi calculada conforme VanRaden (2008) utilizando as informações dos marcadores moleculares, na qual o numerador é análogo à matriz de relacionamento aditivo de pedigree:

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{2\sum_i p_j(1-p_j)} \quad [2.6]$$

em que $\mathbf{Z}$ contém os genótipos ajustados com a frequência alélica e p_j a frequência dos alelos do marcador j . Aqui, o coeficiente de endogamia genômico individual é amostrado como $\mathbf{G}_{jj} - 1$, e o relacionamento entre os indivíduos j e k são obtidos pela divisão dos elementos $\mathbf{G}_{jk}$ pela raiz quadrada da diagonal $\mathbf{G}_{jj}$ e $\mathbf{G}_{kk}$.

Os componentes de (co)variância foram estimados usando o método de máxima verossimilhança restrita com o algoritmo de informação média utilizando a matriz $\mathbf{G}$, no software AIRMLF90 (Misztal et al., 2018). Então, as herdabilidades genômicas foram estimadas e os GEBVs preditos por meio de análise univariada seguindo o modelo, $\mathbf{GB}_a$, de melhor predição linear genômica não viesada (GBLUP):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon} \quad [3.7]$$

onde $\mathbf{y}$ é o vetor de fenótipos; $\mathbf{X}$ é a matriz de incidência sobre os efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos (GC, composto por lote e dieta); $\mathbf{Z}$ é a matriz de incidência sobre os efeitos aleatórios, onde $\mathbf{a}$ é o vetor dos efeitos genéticos aditivos; e $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais. Os componentes aleatórios do modelo seguem distribuição normal com $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$ e $\boldsymbol{\varepsilon} \sim NID(0, \mathbf{I}\sigma_e^2)$, onde σ_a^2 é a variância genética e σ_e^2 é a variância residual, respectivamente, $\mathbf{G}$ é a matriz de relacionamento genômico fundamentada em marcadores SNP, e $\mathbf{I}$ é uma matriz identidade. A matriz $\mathbf{G}$ foi construída de acordo com a metodologia proposta por VanRaden (2008). Para as predições de GEBVs foram utilizados o total de 491 animais genotipados, dos quais 391 apresentaram informações fenotípicas.

Analogamente, também foi utilizado o modelo $\mathbf{GB}_a + \mathbf{CNV}$, com o efeito de CNV incorporado, incluindo a matriz $\mathbf{G}_c$ como arquivo extra no cartão de parâmetros ("user_file"), como segue:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_c\mathbf{c} + \boldsymbol{\varepsilon} \quad [3.8]$$

onde $\mathbf{y}$ é o vetor de fenótipos; $\mathbf{X}$ é a matriz de incidência sobre os efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos (GC, composto por lote e dieta); $\mathbf{Z}_a$ e $\mathbf{Z}_c$ são as matrizes de incidência sobre os efeitos aleatórios, onde $\mathbf{a}$ é o vetor dos efeitos genéticos aditivos e $\mathbf{c}$ é o vetor dos efeitos de CNV; e $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais. Os componentes aleatórios do modelo

seguem distribuição normal com $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$, $\mathbf{c} \sim N(0, \mathbf{G}_c\sigma_c^2)$, onde $\mathbf{G}_c$ é a matriz de relacionamento fundamentada nas informações de CNV, e $\boldsymbol{\varepsilon} \sim NID(0, \mathbf{I}\sigma_e^2)$, onde σ_a^2 e σ_c^2 são variâncias genéticas e σ_e^2 é a variância residual, respectivamente, $\mathbf{G}$ é a matriz de relacionamento genômico fundamentada em marcadores SNP, e $\mathbf{I}$ é uma matriz identidade. As matrizes $\mathbf{G}_c$ e $\mathbf{G}$ foram construídas de acordo com a metodologia proposta por VanRaden (2008). Para as predições de GEBVs foram utilizados o total de 491 animais genotipados, dos quais 391 apresentaram informações fenotípicas.

3.3.3. Acurácia e viés de predição

A fim de comparar a acurácia e viés de predição dos modelos GB_a e $GB_a + CNV$, além do banco de dados completo representando a população de treinamento e contendo informações genotípicas e fenotípicas para 491 e 391 indivíduos, respectivamente, foi criado um subconjunto de dados no qual 1/3 das informações fenotípicas foram aleatoriamente removidas, por meio da função ‘sample’ do software R, representando a população de validação. Dessa forma, o conjunto de dados da população de validação contém informações genotípicas e fenotípicas para 491 e 260 indivíduos, respectivamente.

A acurácia foi estimada pela correlação de Pearson (r) entre os GEBVs preditos para a população de treinamento e o fenótipo ajustado para efeitos fixos no grupo de validação (Legarra et al., 2008):

$$r = \text{cor}(GEBV, y^*) \quad [3.9]$$

em que r é o valor de correlação entre o valor genético genômico predito e o fenótipo ajustado para o efeito fixo de grupo de contemporâneos.

O fenótipo foi determinado pelo ajuste do fenótipo realizado por meio de um modelo linear simples, considerando um vetor de média dos valores fenotípicos da população de treinamento e o grupo de contemporâneos como o efeito fixo. Assim, o fenótipo ajustado foi obtido por:

$$y^* = \mu + y - X\beta - \varepsilon \quad [3.10]$$

em que y^* é o fenótipo ajustado; μ vetor de média geral para o fenótipo da população de treinamento; y é o vetor de fenótipos da população de treinamento; X é a matriz de

incidência para os efeitos fixos; β é o vetor de efeitos fixos (grupo de contemporâneos); ε é o vetor de resíduos.

O viés de predição foi obtido por meio de um modelo de regressão linear univariada do fenótipo ajustado, em função do GEBV predito na população de treinamento, para cada característica (Aliloo et al., 2016). Dessa forma, valores iguais ou próximos a 1 indicam ausência ou baixa deflação (valores maiores que 1) ou inflação (valores menores que 1) nas predições.

3.4. RESULTADOS

3.4.1. População e fenótipos

A estatística descritiva para as características peso aos três (P3) e aos seis (P6) meses, ganho de peso (GP), ganho médio diário (GMD), conversão alimentar residual (CAR), conversão alimentar (CA), eficiência alimentar (EA), comprimento externo (CE), comprimento da perna (CP), rendimento de carcaça (RC), partes totais de carcaça (PTC), área de olho de lombo (AOL), espessura de gordura subcutânea (EGS) e peso de carcaça quente (PCQ) são apresentadas na tabela 1. Nesse contexto, os intervalos e os valores médios registrados para as características de desempenho, eficiência e carcaça estão dentro dos limites biológicos para as características analisadas. As variações quanto ao número de registros entre as características são devidas, principalmente, a análise de consistência e limpeza do banco de dados fenotípicos.

Tabela 5. Estatística descritiva das características analisadas de desempenho, consumo e carcaça.

Característica ^a	N ^b	Média	Mínimo	Máximo	DP ^c
P3 (kg)	390	23,50	12,00	37,00	4,52
P6 (Kg)	389	38,94	18,80	56,60	6,97
GP (Kg)	390	15,36	3,40	27,60	5,00
GMD (kg/dia)	390	0,198	0,05	0,378	0,06
CAR (kg)	384	0,025	-0,397	0,400	0,136
CA (kg)	378	5,58	2,66	10,38	1,55
EA (kg)	388	0,178	0,058	0,334	0,051
CE (cm)	391	52,74	40,00	66,00	5,047
CP (cm)	390	38,25	28,00	48,00	3,43
RC (%)	376	43,99	32,54	56,71	3,776
PTC (kg)	389	8,779	3,726	13,690	1,787
AOL (cm ²)	386	17,66	6,79	42,38	7,912
EGS (mm)	384	2,079	0,910	4,87	0,874
PCQ (kg)	388	17,13	6,90	26,90	3,71

^aP3: peso aos três meses; P6: peso aos seis meses; GP: ganho de peso; CAR: consumo alimentar residual; CA: conversão alimentar; EA: eficiência alimentar; CE: comprimento externo; CP: comprimento da perna; RC: rendimento de carcaça; PTC: partes totais de carcaça; AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; PCQ: peso de carcaça quente. ^bNúmero de fenótipos registrados. ^cDesvio-padrão.

3.4.2. Matrizes de relacionamento genômico

As estatísticas descritivas das matrizes de relacionamento genômico fundamentadas em marcadores SNPs e CNVs são apresentadas nas tabelas 2 e 3, respectivamente.

Tabela 6. Estatística descritiva da matriz de relacionamento genômico fundamentada em marcadores SNPs (**G**)

	N ^a	Média	Mínimo	Máximo	Variância
Diagonal	491	1,005	0,859	1,462	0,009
Off-diagonal	240590	-0,002	-0,123	1,223	0,003

^aNúmero de registros

Tabela 7. Estatística descritiva da matriz de relacionamento genômico fundamentada em marcadores CNVs (**G_c**)

	N ^a	Média	Mínimo	Máximo	Variância
Diagonal	491	1,000	0,428	3,849	0,171
Off-diagonal	240590	0,000	-0,360	0,747	0,007

^aNúmero de registros

A matriz de relacionamento **G** indicou relacionamentos entre irmãos completos e meio irmãos, também revelou indivíduos não correlacionados, com relacionamento médio entre indivíduos de -0,002 (desvio-padrão = 0,0547). Rovadoscki et al. (2018) investigaram a

arquitetura genética da composição de ácidos graxos utilizando uma fração dos animais ($n = 216$) da mesma população utilizada no presente estudo e reportaram relacionamento médio entre indivíduo de $-0,005$ com desvio-padrão de $0,0648$. Nesse sentido, os indivíduos da população considerada apresentam baixo relacionamento. Adicionalmente, Alvarenga et al. (2018) avaliaram LD na mesma população e relataram baixo coeficiente de endogamia ($F=0,04$).

Houveram leves mudanças nas estimativas médias de relacionamento entre a matriz de relacionamento genômico $\mathbf{G}$ e a matriz de relacionamento genômico $\mathbf{G}_c$. Entretanto, houveram variações consideráveis quanto aos valores máximos e mínimos, especialmente da diagonal. Os valores mais extremos observados na matriz $\mathbf{G}_c$ provavelmente estão relacionados à estrutura genômica da CNV, ao número baixo, 294, de marcadores utilizados e à presença de indivíduos sem variações no número de cópias identificadas no genoma.

3.4.3. Estimativas dos componentes de variância e herdabilidades

Os componentes de variância foram estimados com o objetivo de verificar o impacto da inclusão dos efeitos de CNV em um modelo de predição genômica tradicional e o possível aporte para as estimativas dos componentes de variância. As estimativas de (co)variâncias para os modelos GB_a e $GB_a + CNV$ estão apresentadas nas tabelas 8 e 9, respectivamente.

Tabela 8. Variância genética aditiva associada ao efeito de SNP (σ_{aSNP}^2), variância fenotípica (σ_p^2), variância residual (σ_e^2), herdabilidade (h_a^2) e respectivos erros-padrão, entre parênteses, para as características analisadas por meio do modelo reduzido (GB_a)

Característica ^a	σ_{aSNP}^2	σ_p^2	σ_e^2	h_a^2
P3	2,88 (3,444)	16,51 (1,74)	13,63 (2,445)	0,17 (0,201)
P6	8,93 (9,298)	28,27 (5,144)	19,34 (4,861)	0,31 (0,325)
GP	4,60 (4,405)	16,46 (2,283)	11,86 (2,648)	0,27 (0,255)
GMD	1,04E-03 (1,03E-03)	2,88E-03 (5,90E-04)	1,85E-03 (5,04E-04)	0,36 (0,362)
CAR	3,61E-04 (2,41E-04)	1,73E-02 (1,29E-03)	1,69E-02 (1,29E-03)	0,02 (0,014)
CA	0,94 (1,058)	3,23 (0,313)	2,29 (1,266)	0,29 (0,356)
EA	6,57E-04 (6,12E-04)	1,78E-03 (1,70E-04)	1,12E-03 (5,12E-04)	0,36 (0,333)
CE	2,49 (1,711)	9,52 (0,924)	7,036 (1,241)	0,26 (0,168)
CP	1,34 (0,763)	5,45 (0,436)	4,117 (0,681)	0,24 (0,135)
RC	1,48E-04 (1,93E-04)	1,36E-03 (1,13E-04)	1,21E-03 (1,73E-04)	0,11 (0,142)
PTC	0,26 (0,247)	1,74 (0,145)	1,48 (0,215)	0,15 (0,139)
AOL	1,79 (0,514)	7,45 (0,901)	5,67 (1,275)	0,24 (0,094)
EGS	0,11 (0,077)	0,37 (0,038)	0,26 (0,054)	0,28 (0,195)
PCQ	1,82 (1,21)	6,97 (0,647)	5,15 (0,916)	0,26 (0,164)

^aP3: peso aos três meses; P6: peso aos seis meses; GP: ganho de peso; GMD: ganho médio diário; CAR: consumo alimentar residual; CA: conversão alimentar; EA: eficiência alimentar; CE: comprimento externo; CP: comprimento da perna; RC: rendimento de carcaça; PTC: partes totais de carcaça; AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; PCQ: peso de carcaça quente.

Tabela 9. Variância genética aditiva associada ao efeito de SNP (σ_{aSNP}^2), variância genética aditiva associada ao efeito de CNV (σ_{aCNV}^2), variância aditiva total ($\sigma_{aSNP+CNV}^2$), estimativa percentual da variância genética capturada por CNV ($\frac{\sigma_{aCNV}^2}{\sigma_{aSNP+CNV}^2}$), variância fenotípica (σ_p^2), variância residual (σ_e^2), herdabilidade (h_a^2) e respectivos erros-padrão, entre parênteses, para as características analisadas por meio do modelo $GB_a + CNV$

Crt. ^a	σ_{aSNP}^2	σ_{aCNV}^2	$\sigma_{aSNP+CNV}^2$	$\frac{\sigma_{aCNV}^2}{\sigma_{aSNP+CNV}^2}$	σ_p^2	σ_e^2	h_a^2
P3	2,91 (3,429)	0,28 (0,840)	3,19	8,88	16,51 (1,742)	13,32 (2,562)	0,19 (0,205)
P6	8,93 (9,297)	0,0001 (0,00949)	8,93	0,00	28,27 (5,098)	19,34 (4,860)	0,31 (0,314)
GP	4,53 (4,411)	0,32 (0,824)	4,86	6,63	16,43 (2,269)	11,57 (2,746)	0,29 (0,253)
GMD	1,02E-03 (1,03E-03)	4,34E-05 (1,29E-04)	1,07E-03	4,07	2,88E-03 (0,0005)	1,81E-03 (0,0005)	0,37 (0,348)
CAR	2,38E-04 (6,16E-03)	1,09E-04 (8,41E-04)	3,47E-04	31,39	1,73E-02 (0,003)	1,69E-02 (0,003)	0,02 (0,411)
CA	1,14 (1,516)	0,09 (0,185)	1,24	7,57	4,02 (0,867)	2,79 (0,758)	0,30 (0,398)
EA	8,03E-04 (6,24E-04)	8,91E-06 (1,07E-05)	8,12E-04	1,10	2,13E-03 (0,0003)	1,31E-03 (0,0002)	0,38 (0,261)
CE	3,01 (2,093)	6,46E-06 (1,04E-04)	3,01	0,00	9,87 (1,109)	6,86 (1,362)	0,30 (0,191)
CP	1,35 (0,766)	6,67E-02 (0,291)	1,42	4,70	5,45 (0,439)	4,03 (0,745)	0,26 (0,145)
RC	1,68E-04 (2,48E-04)	6,99E-05 (9,88E-05)	2,38E-04	29,36	1,38E-03 (0,0002)	1,14E-03 (0,0002)	0,17 (0,183)
PTC	0,26 (0,247)	1,72E-06 (3,92E-04)	2,63E-01	0,00	1,74 (0,145)	1,48 (0,215)	0,15 (0,137)
AOL	1,80 (0,515)	1,98E-01 (0,544)	1,99	9,94	7,44 (0,894)	5,45 (1,380)	0,26 (0,122)
EGS	1,18E-01 (8,33E-02)	2,32E-06 (1,74E-04)	1,18E-01	0,00	0,37 (0,041)	0,25 (0,056)	0,31 (0,199)
PCQ	1,97 (1,304)	4,73E-07 (0,00001)	1,97	0,00	7,06 (0,686)	5,09 (0,947)	0,27 (0,169)

^aP3: peso aos três meses; P6: peso aos seis meses; GP: ganho de peso; GMD: ganho médio diário; CAR: consumo alimentar residual; CA: conversão alimentar; EA: eficiência alimentar; CE: comprimento externo; CP: comprimento da perna; RC: rendimento de carcaça; PTC: partes totais de carcaça; AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; PCQ: peso de carcaça quente.

As estimativas de variância genética aditiva associadas ao efeito de SNP foram levemente maiores para o modelo $GB_a + CNV$, exceto para a característica GP, que foi sutilmente menor, e para P6 e PTC que foram iguais comparadas às estimativas de variância genética aditiva associada ao efeito de SNP para o modelo GB_a . No geral, a inclusão do efeito

de CNV interferiu de forma sutil na estimativa de variância aditiva associada aos SNPs. Entretanto, as estimativas de variância genética aditiva associada aos CNVs foram baixas, variando entre próximo de 0,00 e 0,32. Entretanto, a variância genética aditiva total, no geral, foi maior para o modelo $GB_a + CNV$, exceto para CAR e PTC que apresentaram estimativas de variância total ligeiramente menor e igual ao modelo GB_a , respectivamente. Ademais, P3, GP, CA, CE e AOL destacam-se com maiores incrementos devido a variância aditiva captada por CNV e, portanto, apresentam maior proporção da variância aditiva associada a CNVs que variou, no geral, entre 0,00 e 31,39%.

As variâncias fenotípicas e residuais pouco variaram entre os modelos, sendo a primeira levemente menor para o modelo GB_a e a segunda sutilmente menor para o modelo $GB_a + CNV$. Entretanto, a herdabilidade restrita, variância aditiva total em relação à variância total, foi maior, para todas as características consideradas, quando estimada por meio do modelo $GB_a + CNV$. Assim, mesmo que o componente genético de CNV tenha sido pouco captado pelo modelo, parte relevante da variância fenotípica foi explicada pelo efeito genético de CNV. Nesse contexto, a maior variação entre os modelos foi observada para RC que apresentou herdabilidade de 0,11 e 0,17 quando estimada por meio dos modelos GB_a e $GB_a + CNV$, respectivamente. Entretanto, os erros associado às estimativas de herdabilidade foram levemente maiores para o modelo $GB_a + CNV$.

3.4.4. Acurácia e viés de predição

A acurácia foi estimada por meio da correlação entre o fenótipo ajustado e o GEBV. Dessa forma, ambos os modelos apresentaram acurácias elevadas e próximas entre os modelos, com a diferença mais expressiva de 0,082 entre os modelos para eficiência alimentar, e com maior e menor acurácias de 0,936 e 0,369 associadas a eficiência alimentar e ao rendimento de carcaça, respectivamente, por meio do modelo GB_a . Assim, para ambos os modelos, houveram variações quanto as acurácias entre as características. Para a maioria dos fenótipos, o modelo GB_a apresentou sutil vantagem quanto as acurácias comparado ao modelo $GB_a + CNV$, exceto para peso de carcaça quente e rendimento de carcaça para as quais as acurácias foram iguais entre os modelos e maior para o modelo $GB_a + CNV$, respectivamente.

O viés de predição foi calculado como o coeficiente de regressão do fenótipo ajustado em função do GEBV, de forma que coeficientes de regressão iguais ou próximos de 1 indicam

ausência ou baixo viés de predição, respectivamente. No geral, os dois modelos apresentaram elevados coeficientes de regressão para todas as características analisadas, indicando alto viés de predição e subestimação dos GEBVs preditos, com maior discrepância entre os modelos para EA, e maior semelhança para AOL. Os maiores e menores coeficientes de regressão foram para EA e AOL, respectivamente, variando entre 0,969 e 3,754 e 1,021 e 5,588 para os modelos GB_a e $GB_a + CNV$, respectivamente. Assim, os modelos GB_a e $GB_a + CNV$ apresentaram coeficientes de regressão próximos a 1 para a característica AOL, indicando baixo viés de predição. Entretanto, para todas as outras características analisadas houve subestimação dos GEBVs independente do modelo. A tabela 10 apresenta as acurácias e os coeficientes de regressão, associados a cada característica, para os modelos reduzido e completo.

Tabela 10. Acurácia e coeficientes de regressão do fenótipo ajustado em função do GEBV (by^* , GEBV) e erro-padrão, entre parênteses, para as características consideradas nas análises para os modelos reduzido (GB_a) e completo ($GB_a + CNV$)

Característica ^a	Acurácia		Viés de predição	
	GB_a	$GB_a + CNV$	GB_a	$GB_a + CNV$
P3	0,858	0,824	2,968 (0,157)	3,433 (0,209)
P6	0,843	0,807	3,594 (0,203)	4,345 (0,281)
GP	0,828	0,799	3,025 (0,181)	3,700 (0,246)
GMD	0,860	0,847	3,754 (0,197)	4,710 (0,261)
CAR	0,873	0,862	3,327 (0,164)	4,728 (0,245)
CA	0,909	0,827	2,227 (0,090)	4,823 (0,289)
EA	0,936	0,903	2,866 (0,095)	5,588 (0,235)
CE	0,865	0,846	2,851 (0,146)	3,606 (0,201)
CP	0,896	0,853	2,633 (0,116)	3,030 (0,164)
RC	0,369	0,372	3,024 (0,653)	4,133 (0,885)
PTC	0,815	0,807	2,370 (0,149)	2,784 (0,180)
AOL	0,570	0,551	0,969 (0,123)	1,021 (0,136)
EGS	0,755	0,737	2,390 (0,182)	2,964 (0,238)
PCQ	0,645	0,645	2,829 (0,294)	3,491 (0,363)

^aP3: peso aos três meses; P6: peso aos seis meses; GP: ganho de peso; GMD: ganho médio diário; CAR: consumo alimentar residual; CA: conversão alimentar; EA: eficiência alimentar; CE: comprimento externo; CP: comprimento da perna; RC: rendimento de carcaça; PTC: partes totais de carcaça; AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; PCQ: peso de carcaça quente.

3.5. Discussão

3.5.1. Relacionamento genômico e componentes de variância

A utilização de grandes bancos de dados com informações confiáveis contribui para predições acuradas de valores genéticos. Contudo, não há um grande conjunto de dados para uso em esquemas de seleção de ovinos no Brasil e o processo de construção leva tempo (Medrado et al., 2021). Assim, este estudo contribui com incremento de informações de componentes de variância e herdabilidade para a raça Santa Inês, uma das principais raças de ovinos do Brasil, e proporciona novas perspectivas sobre a inclusão de variações no número de cópias em modelos de predição genômica.

Todos os cordeiros da população de ovinos Santa Inês considerada para as análises genéticas foram genotipados (Ovine SNP50 BeadChip, www.illumina.com), proporcionando a inclusão de todos os animais na matriz de relacionamento genômico e predição de GEBV para cada cordeiro seguindo dois modelos de melhor predição linear genômica não viesada (GBLUP), GB_a e $GB_a + CNV$. No geral, é comum que apenas uma fração dos animais seja genotipada e isso dificulta a seleção por meio de modelos GBLUP, pois animais genotipados e não genotipados apresentam GEBV e EBV, respectivamente (Lourenco et al., 2020), tornando a comparação mais complexa. Adicionalmente, os efeitos de marcadores SNPs e CNVs foram considerados por meio das matrizes de relacionamento genômico $\mathbf{G}$ e $\mathbf{G}_c$, respectivamente. As matrizes $\mathbf{G}$ e $\mathbf{G}_c$ apresentaram médias baixas de relacionamento entre os indivíduos e muito próximas da registrada (-0,005) por Rovadoscki et al. (2018), que consideraram uma fração dos indivíduos aqui abordados. Entretanto, o valor máximo da diagonal de $\mathbf{G}_c$ é superior ao de $\mathbf{G}$, indicando maior desvio do relacionamento médio provavelmente devido à estrutura genômica da CNV e ao baixo número de marcadores. Dessa forma, a influência de cada marcador para o relacionamento é maior comparado a $\mathbf{G}$ e há maior probabilidade, também devido a não detecção de CNV em 53 indivíduos, de alguns animais apresentarem a mesma sequência de marcadores. Assim, no CNV *marker file* existem 53 animais com todos as variantes codificadas como 1 (2n, indivíduos sem variação no número de cópias). Entretanto, pelo mesmo motivo, fora da diagonal era esperado valor máximo superior em $\mathbf{G}_c$.

As matrizes $\mathbf{G}$ e $\mathbf{G}_c$ foram utilizadas para estimar herdabilidade. Estimativas confiáveis de parâmetros genéticos são relevantes para melhorar a predição dos valores genéticos (Rojas De Oliveira et al., 2018). Porém, as estimativas podem ser influenciadas por inúmeros fatores, incluindo a proporção da variância aditiva captada. As herdabilidades estimadas pelo

modelo GB_a para as 14 características consideradas estão, no geral, próximas às reportadas por Brito et al. (2017), Medrado et al. (2021) e Tortereau et al. (2020) e diferiram sutilmente das estimadas por meio do modelo $GB_a + CNV$. Nesse cenário, as estimativas de herdabilidade foram iguais entre os modelos para P6, CAR e PTC, e apresentaram maior diferença de 0,06 para RC. Indicando que a inclusão de efeito de CNV, no geral, contribui pouco para a herdabilidade, possivelmente por reduzir ligeiramente o componente de variância aditiva associado aos SNPs e apresentar baixo componente de variância associado a CNV. A proporção da variação aditiva captada por CNV variou entre 0,00 e 31,11%, apresentando maior contribuição para CAR. Nesse contexto, segmentos de CNV foram associados ao CAR de animais domésticos (Pu et al., 2016; Zhou et al., 2018), destaca-se a associação de segmento de CNV com o CAR na população considerada neste estudo, como demonstrado no primeiro capítulo deste material. Corroborando com a importância dos efeitos de segmentos de CNV sobre CAR, o componente da variância aditiva associado aos segmentos de CNV foi o proporcionalmente maior entre as 14 características analisadas.

A proporção da variação aditiva captada por CNV para RC foi de 29,11%, contudo diversos fatores podem estar relacionados ao rendimento de carcaça. Escores visuais são geneticamente correlacionados com características de carcaça e importantes para a avaliação de RC (Cancian et al., 2014; Walmsley et al., 2020; Zhou et al., 2016), principalmente na ausência de mensurações mais complexas, como as realizadas por ultrassonografia. Nesse contexto, CNVs foram associados a conformação, musculatura e precocidade em bovinos Nelore (Zhou et al., 2016), indicando que CNVs influenciam a deposição muscular. Adicionalmente, destaca-se a associação de segmento de CNV com o RC na população considerada neste estudo, como demonstrado no primeiro capítulo deste material. Corroborando com a proporção da variância aditiva total explicada pelo componente de variância aditiva associado a CNV. No geral, as variâncias aditivas totais foram maiores quando estimadas pelo modelo $GB_a + CNV$ comparado às estimadas pelo modelo GB_a , pois esse modelo provavelmente capta maiores efeitos aditivos associados aos segmentos de CNV. Portanto, os efeitos de CNV não foram completamente captados pela variância aditiva associada aos marcadores SNPs, assim como indicado em estudos com bovinos (Chen et al., 2021; Hay et al., 2018; Lee et al., 2020). No entanto, modelos incluindo efeitos de CNV podem capturar mais variância, mas isso não se reflete necessariamente em ganhos de capacidade preditiva (Hottis Lyra et al., 2019).

No presente estudo, houve abrangência quanto às características abordadas a fim de incrementar o conhecimento sobre a inclusão de efeitos de CNV em modelos de predição de valores genéticos para características de desempenho, eficiência e carcaça. No geral, CNVs são responsáveis por uma fração da variância fenotípica em ovinos, assim como em bovinos (Chen et al., 2021). O modelo $GB_a + CNV$ seguiu o mesmo padrão de estimativas de componentes de variâncias, herdabilidade e predição de GEBV que o modelo GB_a para as três classes de características. Adicionalmente, os modelos GB_a e $GB_a + CNV$ não apresentaram grandes discrepâncias quanto aos erros-padrão associados às estimativas de herdabilidade, que foram altos para os dois modelos. É provável que, principalmente, o número de animais e o baixo relacionamento médio entre eles tenham contribuído para a inflação do erro associado às estimativas de herdabilidade.

3.5.2. Habilidade e viés de predição

No presente estudo, as correlações entre os fenótipos ajustados e os GEBVs foram sutilmente superiores para o modelo GB_a comparado ao modelo $GB_a + CNV$. As acurácias dos modelos GB_a e $GB_a + CNV$ foram altas e variaram entre 0,36 e 0,93, e 0,37 e 0,90, respectivamente. Mesmo com maior variância genética aditiva total captada pelo modelo $GB_a + CNV$, não houve ganho em acurácia com a inclusão de efeitos de CNV para a predição de GEBV comparando os modelos GB_a e $GB_a + CNV$. Nesse contexto, Hay et al. (2018), trabalhando com uma população de Nelore, avaliaram a inclusão de efeitos de CNV como covariável em modelos de predição de valores genéticos considerando BayesA, BayesB e GBLUP com e sem CNV. Os autores reportaram acurácia de predição variando entre 0,12 e 0,62 entre nove características estudadas, com ganhos mínimos em acurácia ao incluir informações de CNV nos modelos. Considerando modelos GBLUP, o modelo considerando SNPs e todas as deleções e duplicações bialélicas (SNP + todas CNVs) apresentou ganho em acurácia de predição exclusivamente para musculatura e precocidade ao ano, com ganhos de 0,01 e 0,02, respectivamente. Ademais, duas características não diferiram quanto a acurácia de predição e outras cinco foram maiores para o modelo considerando exclusivamente SNPs. Mesmo com diferenças na inclusão dos efeitos de CNV nos modelos de predição (covariável; efeito aditivo extra), espécie e características abordadas, os resultados do presente estudo apresentam padrão semelhante aos reportados por Hay et al. (2018). No presente estudo, não houve variação quanto a acurácia de predição para PCQ e houve ganho de 0,003 para RC

ao incluir efeitos de CNV no modelo. Para as outras 12 características consideradas na análise o modelo GB_a apresentou melhores acurácia de predição. Esses resultados indicam que não houve diferença importante em acurácia de predição ao incluir efeitos de CNV no modelo e que o padrão de variação em acurácia foi semelhante ao utilizar CNV como covariável ou efeito aditivo extra.

Chen et al. (2021), compararam dois modelos GBLUP, um exclusivamente com efeitos de SNPs e outro com efeitos de SNPs e variações estruturais (SV, entre elas CNV) ambas incluídas como valores genéticos aleatórios estimados por meio de matriz de relacionamento genômico por meio de SNPs e CNVs, respectivamente. Utilizando informações das raças Holandesa e Jersey, reportaram acurácias de predição variando entre 0,262 e 0,838 para produção de leite, gordura e proteína, fertilidade e tipo (conformação). Entre as características analisadas, os autores reportaram o maior ganho em acurácia de 0,003 ao incluírem efeito de CNV no modelo. Assim, a inclusão de variantes estruturais no modelo de predição genômica não elevou a acurácia de predição (Chen et al., 2021). Apesar das diferenças metodológicas entre o presente estudo e os reportados por Chen et al. (2021) e Hay et al. (2018), observa-se ausência ou baixa variação quanto a acurácia de predição devida a inclusão de efeitos de CNV nos modelos de predição. Assim, a inclusão de efeito de CNV no modelo não necessariamente significa aumento em acurácia de predição para todas as características e, portanto, a inclusão dos efeitos de CNV em modelos de predição de GEBVs não deve ser prioridade.

Adicionalmente, foi calculado o viés de predição associado a cada característica para os modelos GB_a e $GB_a + CNV$. Para modelos não viesados são esperados coeficientes de regressão iguais a 1. Nesse contexto, os modelos GB_a e $GB_a + CNV$ apresentaram coeficientes de regressão alto para todas as características, variando de 0,96 a 5,588, exceto para AOL que apresentou coeficiente de regressão de 0,96 e 1,02 para os modelos GB_a e $GB_a + CNV$, respectivamente. Assim, para AOL os modelos apresentam baixo viés com leves superestimação e subestimação dos GEBVs quando considerados os modelos GB_a e $GB_a + CNV$, respectivamente. Entretanto, no geral, utilizando os modelos GB_a e $GB_a + CNV$ as predições de GEBVs foram subestimadas para a maioria das características na população de ovinos considerada.

3.6. Conclusão

Por meio de um arquivo de marcadores exclusivamente composto por CNVs foi construída uma matriz de relacionamento genômico utilizada para incorporar os efeitos de CNV ao modelo com informações de SNPs para predição de GEBVs e o ganho em acurácia foi comparado a um modelo utilizando exclusivamente informações de SNPs. Nesse contexto, a inclusão de CNV nas análises genéticas permitiram capturar maior variância genética, entretanto a proporção da variância genética explicada por CNVs varia entre as características. Adicionalmente, não foram encontrados estudos que utilizaram a mesma metodologia para incorporar efeitos de CNV em modelos de predição de GEBV para avaliação de características de animais domésticos. Entre os modelos considerados, não houve incremento na habilidade preditiva ao inserir informações de CNV. Portanto, os resultados reportados no presente estudo indicam que a inclusão de informações de CNV em modelos para predição de GEBV não deve ser prioridade.

REFERÊNCIAS

- Aguilar, I., 2014. Short introduction to BLUPF90 family programs. Retrieved on 31 January 2018, from http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=slides_2014_lva.pdf
- Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., Lawlor, T.J., 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score 1 93, 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D.A.L., Legarra, A., Misztal, I., 2018. BLUPF90 suite of programs for animal breeding with focus on genomics. 11th World Congr. Genet. Appl. to Livest. Prod. Auckland, New Zeal. 11.751.
- Aguilar, Ignacio, Misztal, Ignacy, Tsuruta, Shogo, Legarra, Andres, Wang, Huiyu, Aguilar, I, Misztal, I, Tsuruta, S, Legarra, A, Wang, H, 2014. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs.
- Blasco, A., Toro, M.A., 2014. A short critical history of the application of genomics to animal breeding. *Livest. Sci.* 166, 4–9. <https://doi.org/10.1016/j.livsci.2014.03.015>
- Boerner, V., Johnston, D.J., Tier, B., 2014. Accuracies of genomically estimated breeding values from pure-breed and across-breed predictions in Australian beef cattle. *Genet. Sel. Evol.* 46, 1–11. <https://doi.org/10.1186/s12711-014-0061-9>
- Brito, L.F., McEwan, J.C., Miller, S., Bain, W., Lee, M., Dodds, K., Newman, S.-A., Pickering, N., Schenkel, F.S., Clarke, S., 2017. Genetic parameters for various growth, carcass and meat quality traits in a New Zealand sheep population. *Small Rumin. Res.* 154, 81–91. <https://doi.org/10.1016/j.smallrumres.2017.07.011>

- Cancian, P.H., Gomes, R. da C., Manicardi, F.R., Ianni, A.C., Bonin, M. de N., Leme, P.R., Silva, S. da L. e, 2014. Correlations of visual scores, carcass traits, feed efficiency and retail product yield in Nellore cattle. *Sci. Agric.* 71, 17–22. <https://doi.org/10.1590/S0103-90162014000100002>
- Chen, L., Pryce, J.E., Hayes, B.J., Daetwyler, H.D., 2021. Investigating the Effect of Imputed Structural Variants from Whole-Genome Sequence on Genome-Wide Association and Genomic Prediction in Dairy Cattle. *Anim.* 2021, Vol. 11, Page 541 11, 541. <https://doi.org/10.3390/ANI11020541>
- Clop, A., Vidal, O., Amills, M., 2012. Copy number variation in the genomes of domestic animals. *Anim. Genet.* 43, 503–517. <https://doi.org/10.1111/j.1365-2052.2012.02317.x>
- Da Silva, V.H., De Almeida Regitano, L.C., Geistlinger, L., Pértille, F., Giachetto, P.F., Brassaloti, R.A., Morosini, N.S., Zimmer, R., Coutinho, L.L., 2016. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. *PLoS One* 11, 1–22. <https://doi.org/10.1371/journal.pone.0157711>
- Daetwyler, H.D., Swan, A.A., Werf, J.H. van der, Hayes, B.J., 2012. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genet. Sel. Evol.* 2012 441 44, 1–11. <https://doi.org/10.1186/1297-9686-44-33>
- Dekkers, J.C.M., 2012. Application of Genomics Tools to Animal Breeding, *Current Genomics*.
- Fernandes, A.C., da Silva, V.H., Goes, C.P., Moreira, G.C.M., Godoy, T.F., Ibelli, A.M.G., Peixoto, J. de O., Cantão, M.E., Ledur, M.C., de Rezende, F.M., Coutinho, L.L., 2021. Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics* 22, 354. <https://doi.org/10.1186/s12864-021-07676-1>
- Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. <https://doi.org/10.1038/nrg1767>
- Food and Agriculture Organization—FAO, 2001. Guidelines for humane handling, transport and slaughter of livestock Guidelines for humane handling, transport and slaughter of livestock. Bangkok: Food and Agriculture Organization of United Nation.
- Havenstein, G.B., Ferket, P.R., Qureshi, M.A., 2003. Carcass composition and yield of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult. Sci.* 82, 1509–1518. <https://doi.org/10.1093/PS/82.10.1509>
- Hay, E.H.A., Utsunomiya, Y.T., Xu, L., Zhou, Y., Neves, H.H.R., Carneiro, R., Bickhart, D.M., Ma, L., Garcia, J.F., Liu, G.E., 2018. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. *BMC Genomics* 19, 1–8. <https://doi.org/10.1186/s12864-018-4787-6>
- Hayes, B.J., Lewin, H.A., Goddard, M.E., 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29, 206–214. <https://doi.org/10.1016/J.TIG.2012.11.009>
- Henderson, C.R., 1975. Rapid Method for Computing the Inverse of a Relationship Matrix. *J. Dairy Sci.* 58, 1727–1730. [https://doi.org/10.3168/jds.S0022-0302\(75\)84776-X](https://doi.org/10.3168/jds.S0022-0302(75)84776-X)
- Henderson, C.R., 1963. Selection index and expected genetic advance. *Stat. Genet. Plant Breed.* 982, 141–163.
- Henrichsen, C.N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H., Reymond, A., 2009. Segmental copy number variation shapes tissue transcriptomes. <https://doi.org/10.1038/ng.345>

- Hottis Lyra, D., Galli, G., Couto Alves, F., Stefanine Correia Granato, Í., Suzane Vidotti, M., Bandeira Sousa, M., Silva Morosini, J., Crossa, J., Fritsche-Neto, R., 2019. Modeling copy number variation in the genomic prediction of maize hybrids 132, 273–288. <https://doi.org/10.1007/s00122-018-3215-2>
- Knott, S.A., Cummins, L.J., Dunshea, F.R., Leury, B.J., 2008. The use of different models for the estimation of residual feed intake (RFI) as a measure of feed efficiency in meat sheep. *Anim. Feed Sci. Technol.* 143, 242–255. <https://doi.org/10.1016/j.anifeedsci.2007.05.013>
- Lee, Y.L., Bosse, M., Mullaart, E., Groenen, M.A.M., Veerkamp, R.F., Bouwman, A.C., 2020. Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC Genomics* 21. <https://doi.org/10.1186/s12864-020-6496-1>
- Legarra, A., Aguilar, I., Misztal, I., 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. <https://doi.org/10.3168/jds.2009-2061>
- Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., Misztal, I., 2020. Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. *Genes (Basel)*. 11, 790. <https://doi.org/10.3390/genes11070790>
- Medrado, B.D., Pedrosa, V.B., Pinto, L.F.B., 2021. Meta-analysis of genetic parameters for economic traits in sheep. *Livest. Sci.* 247, 104477. <https://doi.org/10.1016/j.livsci.2021.104477>
- Meuwissen, T., 2007. Genomic selection : Marker assisted selection on a genome wide scale. *J. Anim. Breed. Genet.* 124, 321–322. <https://doi.org/10.1111/j.1439-0388.2007.00708.x>
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Misztal, I., Legarra, A., Aguilar, I., 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. <https://doi.org/10.3168/jds.2009-2064>
- Mrode, R., Tarekegn, G.M., Mwacharo, J.M., Djikeng, A., 2018. Invited review: Genomic selection for small ruminants in developed countries: How applicable for the rest of the world? *Animal* 12, 1333–1340. <https://doi.org/10.1017/S1751731117003688>
- Pu, L., Zhang, L.C., Zhang, J.S., Song, X., Wang, L.G., Liang, J., Zhang, Y.B., Liu, X., Yan, H., Zhang, T., Yue, J.W., Li, N., Wu, Q.Q., Wang, L.X., 2016. Porcine MAP3K5 analysis: molecular cloning, characterization, tissue expression pattern, and copy number variations associated with residual feed intake. *Genet. Mol. Res.* 15. <https://doi.org/10.4238/gmr.15037998>
- Putz, A.M., Tiezzi, F., Maltecca, C., Gray, K.A., Knauer, M.T., 2018. A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data. *J. Anim. Breed. Genet.* 135, 5–13. <https://doi.org/10.1111/JBG.12302>
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shaperro, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, Junjun, Zerjal, T., Zhang, Jane, Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. *Nature* 444, 444–454. <https://doi.org/10.1038/nature05329>

- Rojas De Oliveira, H., Ventura, H.T., Costa, E.V., Pereira, M.A., Veroneze, R., Duarte, M.D.S., Dias De Siqueira, O.H.G.B., Fonseca E Silva, F., 2018. Meta-analysis of genetic-parameter estimates for reproduction, growth and carcass traits in Nellore cattle by using a random-effects model. *Anim. Prod. Sci.* 58, 1575–1583. <https://doi.org/10.1071/AN16712>
- Rovadoscki, G.A., Pertile, S.F.N., Alvarenga, A.B., Cesar, A.S.M., Pértile, F., Petrini, J., Franço, V., Soares, W.V.B., Morota, G., Spangler, M.L., Pinto, L.F.B., Carvalho, G.G.P., Lanna, D.P.D., Coutinho, L.L., Mourão, G.B., 2018. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics* 19, 1–14. <https://doi.org/10.1186/s12864-018-4777-8>
- Rupp, R., Mucha, S., Larroque, H., McEwan, J., Conington, J., 2016. Genomic application in sheep and goat breeding. *Anim. Front.* 6, 39–44. <https://doi.org/10.2527/af.2016-0006>
- Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123, 218–223. <https://doi.org/10.1111/j.1439-0388.2006.00595.x>
- Sena, L.S., Figueiredo Filho, L.A.S., Santos, G.V. dos, Sousa Júnior, A. de, Santos, N.P. da S., Britto, F.B., Sarmiento, J.L.R., Brito, L.F., 2020. Genetic evaluation of tropical climate-adapted sheep for carcass traits including genomic information. *Small Rumin. Res.* 188. <https://doi.org/10.1016/j.smallrumres.2020.106120>
- Silva Sobrinho, A.G., 1999. Body composition and characteristics of carcasses from lambs of different genotypes and age at slaughter.
- Stock, K., Reents, R., 2013. Genomic selection: Status in different species and challenges for breeding. *Reprod. Domest. Anim.* 48, 2–10. <https://doi.org/10.1111/rda.12201>
- Taylor, J.F., 2014. Implementation and accuracy of genomic selection. *Aquaculture.* 421, 8–14. <https://doi.org/10.1016/j.aquaculture.2013.02.017>
- Tortereau, F., Marie-Etancelin, C., Weisbecker, J., Marcon, D., Bouvier, F., Moreno-Romieux, C., François, D., 2020. Genetic parameters for feed efficiency in Romane rams and responses to single-generation selection. *Animal* 14, 681–687. <https://doi.org/10.1017/S1751731119002544>
- VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Walmsley, B.J., Cafe, L.M., Wilkins, J.F., McPhee, M.J., Walmsley, B.J., Cafe, L.M., Wilkins, J.F., McPhee, M.J., 2020. Selection for increased visual muscling increases carcass leanness without compromising predicted Meat Standards Australia eating-quality index. *Anim. Prod. Sci.* 61, 294–299. <https://doi.org/10.1071/AN20157>
- Wang, K., Chen, Z., Tadesse, M.G., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., Li, M., 2008. Modeling genetic inheritance of copy number variations. *Nucleic Acids Res.* 36, 138. <https://doi.org/10.1093/nar/gkn641>
- Weller, J.I., Ezra, E., Ron, M., 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* 100, 8633–8644. <https://doi.org/10.3168/jds.2017-12879>
- Wiggans, G.R., Cole, J.B., Hubbard, S.M., Sonstegard, T.S., 2017. Genomic Selection in Dairy Cattle: The USDA Experience. *Annu. Rev. Anim. Biosci.* 5, 309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>
- Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J.W., Zhang, H., Li, L., Liu, G.E., 2018. Diversity of copy number variation in a worldwide population of sheep. *Genomics* 110, 143–148. <https://doi.org/10.1016/j.ygeno.2017.09.005>

- Zhou, Y., Connor, E.E., Wiggans, G.R., Lu, Y., Tempelman, R.J., Schroeder, S.G., Chen, H., Liu, G.E., 2018. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. *BMC Genomics* 19, 314. <https://doi.org/10.1186/s12864-018-4699-5>
- Zhou, Y., Utsunomiya, Y.T., Xu, L., Hay, E.H. Abdel, Bickhart, D.M., Alexandre, P.A., Rosen, B.D., Schroeder, S.G., Carvalheiro, R., de Rezende Neves, H.H., Sonstegard, T.S., Van Tassell, C.P., Ferraz, J.B.S., Fukumasu, H., Garcia, J.F., Liu, G.E., 2016. Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. *BMC Genomics* 17, 419. <https://doi.org/10.1186/s12864-016-2461-4>
- Zuidhof, M.J., Schneider, B.L., Carney, V.L., Korver, D.R., Robinson, F.E., 2014. Growth, efficiency, and yield of commercial broilers from 1957, 1978, and 2005. *Poult. Sci.* 93, 2970–2982. <https://doi.org/10.3382/PS.2014-04291>

APÊNDICES

APÊNDICE A. CNVs com p -valor $>0,10$ e p -valor $<0,11$ associadas às respectivas características, consumo alimentar residual (CAR) e partes totais de carcaça (PTC), acompanhadas das posições e respectivas probes mais significativas, frequências e p -valores.

Característica ^a	OAR: primeira- última posição ^b	Probe	Frequência	p -valor
CAR	6:63210772- 63292850	s29556.1	8	0.10724
CAR	1:85202342- 85261298	OAR1_85202342.1	5	0.10724
PTC	17:43258867- 43356226	OAR17_43258867.1	8	0.10229
PTC	21:19511069- 19613155	OAR21_19511069.1	11	0.10229

^aRC: rendimento de carcaça, CE: comprimento externo, CAR: consumo alimentar residual, PTC: partes totais de carcaça. ^bPosição do segmento de CNV baseada no genoma ovino Oar_rambouillet_v1.0 (NCBI).

APÊNDICE B. Sobreposição de segmentos de CNVs, com p -valor $>0,10$ e p -valor $<0,11$, com QTLs previamente mapeados para características de crescimento, desempenho e carcaça.

OAR: primeira-última posição ^a	QTL_ids	Característica relacionada
6:63210772-63292850	QTL:14284	Peso corporal (abate)
1:85202342-85261298	QTL:14238	Deposição de músculo na terceira lombar
	QTL:14241	Deposição de músculo na terceira lombar
	QTL:14239	Gordura na terceira lombar
	QTL:17216	Conteúdo de ácido docosapentaenóico da carne
	QTL:14275	Peso do osso na carcaça
	QTL:14276	Peso do músculo na carcaça
	QTL:14278	Porcentagem de rendimento de carne magra
	QTL:14277	Porcentagem de gordura na carcaça
	QTL:13885	Peso corporal (8 semanas)
	QTL:17215	Conteúdo de ácido eicosapentaenóico da carne
	QTL:17221	Teor de ácidos graxos poliinsaturados da carne
17:43258867-43356226	-	-
21:19511069-19613155	QTL:17218	Conteúdo de ácido docosapentaenóico da carne
	QTL:13946	Ganho médio diário
	QTL:13930	Peso corporal (56 semanas)
	QTL:13952	Ganho médio diário
	QTL:17207	Teor de ácido gadoléico da carne
	QTL:13925	Peso corporal (43 semanas)
	QTL:13936	Peso corporal (83 semanas)
	QTL:13959	Ganho médio diário
	QTL:17201	Teor de ácido mirístico da carne
	QTL:17202	Teor de ácido palmítico da carne
	QTL:17211	Teor de ácido linoleico da carne
	QTL:17203	Teor de ácido esteárico da carne
	QTL:17205	Teor de ácido oleico da carne
	QTL:17206	Teor de ácido cis-vacênico da carne
	QTL:17209	Teor de ácido linoleico da carne
	QTL:17212	Teor de ácido arcadiônico da carne

^aPosição do segmento de CNV baseada no genoma ovino Oar_rambouillet_v1.0 (NCBI).

APÊNDICE C. Lista de genes dentro da janela de 1Mb para CNVs com p -valor $>0,10$ e p -valor $<0,11$

OAR: primeira- última posição ^a	Característica associada ^b	Gene	Ensembl ID ^c	Entrezgene ID ^d
6:63210772- 63292850	CAR	<i>FAM114A1</i>	ENSOARG000000010269	101112224
		<i>KLF3</i>	ENSOARG000000010144	101111966
		<i>PGM2</i>	ENSOARG000000009837	101111285
1:85202342- 85261298	CAR	<i>COL11A1</i>	ENSOARG000000018591	101117548
		<i>RNPC3</i>	ENSOARG000000018719	101118319
17:43258867- 43356226	PTC	<i>FSTL5</i>	ENSOARG000000001443	101105281
21:19511069- 19613155	PTC	<i>ALG8</i>	ENSOARG000000006601	101109656
		<i>AQP11</i>	ENSOARG000000007178	101111499
		<i>GAB2</i>	ENSOARG000000006340	101115743
		<i>INTS4</i>	ENSOARG000000006741	101110460
		<i>KCTD14</i>	ENSOARG000000015282	101116522
		<i>KCTD21</i>	ENSOARG000000006482	101109401
		<i>NARS2</i>	ENSOARG000000006276	101108871
		<i>RSF1</i>	ENSOARG000000006933	101116780
		<i>TENM4</i>	ENSOARG000000006069	101108614
		<i>THRSP</i>	ENSOARG000000015271	101110189
		<i>USP35</i>	ENSOARG000000006396	101109133

^aPosição baseada no genoma ovino Oar_rambouillet_v1.0 (NCBI). ^bCAR: consumo alimentar residual, PTC: partes totais de carcaça. ^cIdentificação do gene pela base de dados Ensembl (Ensembl 104 Database). ^dIdentificação do gene pela base de dados NCBI (<https://www.ncbi.nlm.nih.gov/gene>)