

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

**Identificação de polimorfismos em região do cromossomo 2 da galinha
associado a deposição de músculo**

Thaís Fernanda Godoy

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

**Piracicaba
2014**

Thaís Fernanda Godoy
Engenheira Agrônoma

**Identificação de polimorfismos em região do cromossomo 2 da galinha associado a
deposição de músculo**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **LUIZ LEHMANN COUTINHO**

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

**Piracicaba
2014**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Godoy, Thaís Fernanda

Identificação de polimorfismos em região do cromossomo 2 da galinha associado a deposição de músculo / Thaís Fernanda Godoy. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.

77 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.
Bibliografia.

1. INDEL 2. Músculo de peito 3. Mutação 3. Sequenciamento de nova geração
4. SNP I. Título

CDD 636.513
G588i

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

Primeiramente a **Deus**, pois quando pedi força, Ele me deu dificuldades para me fazer forte.

Pedi sabedoria, Deus me deu problemas para resolver. Pedi prosperidade, Deus me deu cérebro e músculos para trabalhar. Pedi coragem, Deus me deu perigos para superar. Pedi amor, Deus me deu pessoas com problemas para ajudar. Pedi favores, Deus me deu oportunidades. Eu não recebi nada do que pedi, mas recebi tudo de que precisava!

E aos meus pais, **Isabete e Oswaldo**, por todo o amor, carinho, dedicação, sacrifício e luta para tornar mais este sonho possível! Vocês são os responsáveis por toda a conquista em minha vida! Amo vocês!

AGRADECIMENTOS

À **Deus**, pelo amor incondicional e pelos milagres realizados em minha vida.

Aos meus pais, **Oswaldo** e **Isabete**, e minha irmã, **Ana Flávia**, por estarem ao meu lado sempre, sendo a minha força e a minha fé nos momentos mais difíceis. Muito obrigada pelo amor, paciência, compreensão e carinho.

Aos meus amigos e companheiros caninos: **Nicole**, **Nayara**, **Maria Lina**, **Nina**, **Rubens** e, **Theodoro (Dindo)** pela companhia e fidelidade!

À gloriosa **Escola Superior de Agricultura “Luiz de Queiroz”** e todos os **professores** pela oportunidade de estudo e pelos ensinamentos concedidos.

Ao **Programa de Pós-Graduação em Ciência Animal e Pastagens**, pela oportunidade de realizar o meu mestrado com excelência.

Ao **Prof. Dr. Luiz Lehmann Coutinho**, pela orientação e ensinamentos recebidos.

À Dra. **Clarissa Boschiero**, por toda ajuda, pela paciência e por estar sempre disposta a ensinar, a corrigir e a orientar. Muito obrigada!!

À Dra. **Mônica Correa Ledur** pela orientação e apoio recebidos.

À **CAPES** e **FAPESP** pela concessão da bolsa.

Às minhas amigas, agora distantes, mas sempre presentes em meu coração: **Maria Estella (K-bô)** e **Natalia (Xiguela)** por torcerem por mim e por acreditarem na minha capacidade. E a todos os **PETianos** (PET Biotecnologia Agrícola – ESALQ) que sempre me proporcionaram imensa alegria e aprendizado.

Ao meu colega de pesquisa e AMIGO **Gabriel Costa**. Agradeço pela convivência, pelas muitas risadas (muitas mesmo), por sempre ser uma mão amiga, por me ajudar quando necessário, por me ensinar quando era preciso e por ser esta luz em minha vida.

Às minhas amigas: **Ariana** e **Tássia (Pequenina)**. Vocês entraram de mancinho em minha vida, mas se tornaram essenciais! Fazem parte desta etapa da minha vida e de muitas outras.

Ao **Renato Fonseca (Gordinho)**, que entrou em minha vida há alguns anos atrás, mas sempre se fez presente. Sendo meu amigo, meu companheiro, meu ouvinte, meu cúmplice mesmo estando distante. Porém, hoje posso dizer que está mais perto do que nunca, acompanhando mais esta vitória em minha vida.

Aos técnicos do Laboratório de Biotecnologia Animal, **Nirlei**, **Jorge**, **Ricardo** e aos colegas e amigos: **Aline**, **Andrezza**, **Áurea**, **Berna**, **Dênia**, **Gustavo**, **Fabi**, **Fábio**, **Lilian**, **Marcela**, **Millor**, **Ribamar**, **Sônia**, **Vinícius**.

Às minhas amigas e companheiras de estudo e de festas: **Liliane (Kraxá), Mayara (Minestra) e Fabiane Costa.**

Ao **Prof. Gerson** e colegas do **GEMA: Juliana, Gregori, Joana, Simone e Laiza.**

À todos que de alguma maneira contribuíram para que este sonho se tornasse realidade.

Muito Obrigada!!!

EPÍGRAFE

“Dia a dia eu ficava sabendo mais alguma coisa do Planeta, da partida, da viagem.
Mas isto devagarzinho, ao acaso das reflexões.”

O Pequeno Príncipe

SUMÁRIO

RESUMO	11
ABSTRACT	13
1 INTRODUÇÃO.....	15
2 REVISÃO BIBLIOGRÁFICA	17
2.1 Melhoramento genético na avicultura	17
2.2 Genômica no melhoramento de aves.....	18
2.3 QTL no cromossomo 2 da galinha associado com deposição de músculo de peito.....	20
2.4 Sequenciamento do genoma da galinha.....	22
2.5 Identificação de polimorfismos: SNPs e INDELS.....	25
2.6 Análise de associação e mutação causal.....	26
3 OBJETIVOS.....	29
4 MATERIAL E MÉTODOS.....	31
4.1 População experimental, extração e quantificação do DNA	31
4.2 Sequenciamento de Nova Geração	32
4.3 Preparação das amostras	32
4.4 Clusterização	33
4.5 Sequenciamento.....	34
4.6 Processamento inicial dos dados	35
4.7 Alinhamento	36
4.8 Definição da região-alvo	36
4.9 Identificação e filtragem de polimorfismos.....	37
4.10 Anotação Funcional.....	38
5 RESULTADOS E DISCUSSÃO	41
5.1 Extração e quantificação do DNA	41
5.2 Sequenciamento e alinhamento	42
5.3 Detecção e filtragem de SNPs e INDELS.....	44
5.4 Anotação Funcional.....	49
5.5 Genes relacionados ao desenvolvimento muscular	54
6 CONCLUSÃO.....	59
REFERÊNCIAS	61
APÊNDICE	75

RESUMO

Identificação de polimorfismos em região do cromossomo 2 da galinha associado a deposição de músculo

A produção brasileira de carne de frango tem uma grande importância econômica no mundo todo devido principalmente aos avanços do melhoramento genético. O surgimento de novas tecnologias de sequenciamento (sequenciamento de nova geração) tem se tornado uma ferramenta poderosa, pois por meio da identificação de SNPs (polimorfismo de nucleotídeo único) e INDELs (deleções/inserções) possibilita a adição de novas informações ao melhoramento genético. A deposição de músculo, em especial o músculo de peito, é uma das características que mais merecem destaque por causa da sua importância nutricional e econômica. Sendo assim o objetivo deste trabalho foi ressequenciar o genoma de 18 aves de duas linhagens distintas experimentais e identificar SNPs e INDELs em uma região de QTL no cromossomo 2 da galinha associado anteriormente com deposição de músculo do peito, além de caracterizar variantes potencialmente funcionais e propor mutações candidatas para estudos futuros. Para isso, dezoito galinhas de duas diferentes linhagens experimentais (corte e postura), ambas desenvolvidas pela Embrapa Suíno e Aves, foram sequenciadas pela plataforma de nova geração da Illumina. SNPs e INDELs foram identificados por meio de ferramentas de bioinformática em uma região de QTL no cromossomo 2 da galinha (105.848.755-112.648.761 pb) que foi previamente associada com deposição de músculo de peito. O sequenciamento dos 18 animais gerou em torno 2,7 bilhões de *reads* e após a filtragem por qualidade foram mantidas 77% das *reads*. Em seguida, as *reads* foram alinhadas ao genoma referência (*Gallus_gallus*-4.0, NCBI) pela ferramenta Bowtie2 e gerou em média 10,6X de cobertura de sequenciamento na região-alvo. , Foram identificados 722.832 SNPs e 63.727 INDELs para os 18 animais por meio do programa SAMtools, e após uma filtragem rigorosa, foram mantidos 77% dos SNPs (n=558.767) e 60% das INDELs (n=38.402). Com base nas variantes únicas para os 18 animais (85.765 SNPs e 7.824 INDELs) foi realizada a anotação funcional por meio da ferramenta ANNOVAR. Dentre os SNPs não sinônimos (n=153) e *stopgain* (n=3), 15 foram classificados como deletérios. Um dos SNPs deletérios que já foi depositado em banco de dados foi identificado no gene *RBICCI*, que tem sua função relacionada ao desenvolvimento do músculo de peito. Utilizando a ferramenta DAVID foi possível analisar 37 genes relacionados aos SNPs não sinônimos, *stopgain*, INDELs *frameshift* e não *frameshift*. Dentre estes genes, três (*DTNA*, *RBICCI* e *C-MOS*) foram selecionados por terem suas funções relacionadas ao desenvolvimento muscular e suas mutações foram analisadas. Sendo assim, futuros estudos podem ser realizados nestes genes candidatos e nas mutações identificadas, por meio de análises de associação e validação em populações comerciais, permitindo assim uma melhor explicação o efeito do QTL estudado.

Palavras-chave: INDEL; Músculo de peito; Mutação; Sequenciamento de nova geração; SNP

ABSTRACT

Identification of polymorphisms in the chicken chromosome 2 region associated with muscle deposition

The Brazilian chicken meat production has a great economic importance in worldwide mainly due to advances in breeding. The emergence of new techniques of sequencing (next-generation sequencing) becomes a powerful tool because through identification of SNPs (single nucleotide polymorphism) and INDELs (deletions/insertions) allows the addition of new information for genetic improvement. The muscle deposition, particularly the breast muscle, is one of the features that are most noteworthy because of its nutritional and economic importance. Therefore the aim of this study was to perform the genome resequencing of 18 chicken from two distinct experimental lines and identify SNPs and INDELs in a QTL region on chromosome 2 previously associated with breast muscle, and characterize the variants to identify potentially function ones and propose candidate mutations for future studies. To achieve these objectives, eighteen chickens of two different experimental lines (broiler and layer), both developed by Embrapa Swine and Poultry were sequenced by Illumina next-generation platform. SNPs and INDELs were identified by bioinformatic tools in a QTL region on chicken chromosome 2 (105,848,755-112,648,761 bp) which was previously associated with breast muscle deposition. Sequencing of the eighteen animals generated around 2.7 billion of reads, and 77% of the reads were retained after filtering. The reads were aligned against the chicken genome reference (*Gallus_gallus*-4.0, NCBI) by Bowtie2 tool resulting in a 10.6X coverage across the target region. Using SAMtools, 722,832 SNPs and 63,727 INDELs were identified in the all individuals, and after a stringent filtration, 77% of SNPs (n=558,767) and 60% of INDELs (n=38,402) were maintained. Based on unique variants for all the animal (85,765 SNPs and 7,828 INDELs) were performed the functional annotation by ANNOVAR tool. Among the non-synonymous SNPs (n=153) and stopgain (n=3), fifteen were predicted like a deleterious mutation. One of deleterious SNPs has already deposited in public database, and it was identified in *RBICC1* gene, which function is related to breast muscle development. Using the DAVID tool was possible to analyze the 37 genes related to the non-synonymous SNPs, stopgain, frameshift and non-frameshift INDELs. Among these genes, three (*DTNA*, *RBICC1* and *C-MOS*) were selected due their functions related to muscle development and their mutations were analyzed. Therefore, further association studies can be performed with these candidate genes and their mutations, and also validation in commercial populations, allowing a better explanation of QTL effects.

Keywords: INDEL; Breast muscle; Mutation; Next-generation sequencing; SNP

1 INTRODUÇÃO

A carne de frango, nas últimas décadas, se tornou uma carne muito consumida no mundo inteiro. No Brasil o consumo per capita em 2012 chegou a 45 kg/hab, 50,45% a mais em relação a 2000 (29,91 kg/hab) (UNIÃO BRASILEIRA DE AVICULTURA - UBABEF, 2013). Este aumento se deve a diversos fatores, tais como: a disponibilidade no mercado a um preço relativamente baixo, a ausência de obstáculos culturais ou religiosos para seu consumo e além de suas propriedades alimentares e nutricionais (VALCESCHINI, 2006).

Sua produção vem aumentando significativamente em todo o mundo e o Brasil é um dos países que mais se destacam neste setor. Os dados da União Brasileira de Avicultura (UBABEF, 2013) mostram que a produção de carne de frango de 2000 a 2012 aumentou 111%, ou seja, a produção passou de 5,98 para 12,65 milhões de toneladas. Em 2012 a produção sofreu uma redução de 3,17% em relação a 2011 devido ao aumento dos preços de milho e soja e a falta de crédito para avicultores. Entretanto, o Brasil se manteve na posição de maior exportador mundial (3,9 milhões de toneladas) e do terceiro maior produtor de carne de frango, atrás apenas dos Estados Unidos e da China.

O consumidor brasileiro é o principal cliente da carne de frango produzida pela avicultura nacional, e isto é comprovado pelos dados de 2012, onde 69% da produção foi destinada ao mercado consumidor brasileiro, que também exige os elevados padrões de qualidade e sanidade conferidos ao produto exportado (UBABEF, 2013).

A galinha doméstica não é apenas considerada importante na agroindústria, é também um modelo biológico útil para estudos científicos em diversas áreas, inclusive na genética e biologia celular (ELLEGREN, 2005; STERN, 2005; WOLPERT, 2004). Devido a esta importância biológica da galinha, vários estudos moleculares são desenvolvidos, tornando possível a aplicação da genética molecular ao melhoramento animal, proporcionando estratégias de seleção mais eficientes e promovendo maiores progressos genéticos em curto prazo.

O programa de melhoramento genético tradicional se baseia na teoria da genética quantitativa. Desta forma, o ganho genético tem sido alcançado por meio da seleção fenotípica e pela estimativa do valor genético aditivo derivado deste fenótipo. Porém, com a adição de técnicas moleculares nos programas de melhoramento, torna-se possível uma melhor compreensão da ação dos diversos genes que influenciam cada um com um pequeno efeito, como as relacionadas a crescimento, fertilidade, resistência a doenças e as

características de importância econômica, como por exemplo, as associadas ao músculo de peito (BURT; HOCKING, 2002).

De acordo com Zhou et al. (2006a) existe uma correlação positiva entre o rendimento de músculo de peito com características relacionadas à gordura. Ou seja, ao selecionarmos animais com um maior rendimento de músculo de peito, ocorre simultaneamente um maior acúmulo de gordura, o que não é desejável principalmente nos dias atuais, aonde o consumidor procura cada vez mais alimentos com menos gordura. Portanto, se faz necessário uma maior compreensão destas correlações genéticas existentes entre as características de interesse produtivo, que permita um melhoramento genético eficiente sem comprometer os ganhos genéticos já alcançados (LEDUR; PEIXOTO; SCHMIDT, 2011).

Dentre as técnicas moleculares existentes, o mapeamento de locos de características quantitativas (QTLs) permite a identificação de regiões associadas a estas características de interesse. Segundo Ikeobi et al. (2004), mapear o QTL é o primeiro passo para caracterizar a variações genéticas existentes nos genes e no DNA. Porém, estas regiões são relativamente grandes, podendo conter vários genes, o que dificulta a compreensão da atuação no fenótipo.

Além disso, as novas tecnologias de sequenciamento, denominadas sequenciamento de nova geração, vem se mostrando uma importante ferramenta e uma alternativa poderosa para estudar tanto a genômica estrutural quanto a funcional (CARVALHO; SILVA, 2010). A grande vantagem desta técnica é a possibilidade de sequenciar um genoma todo de forma muito mais econômica e eficiente em relação às técnicas anteriores (FAN et al., 2013). Desta forma, torna-se possível a identificação de SNPs (polimorfismos de base única) e INDELS (inserções e deleções), o que auxilia na identificação de possíveis mutações causais que regulam possíveis diferenças no fenótipo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Melhoramento genético na avicultura

A domesticação da galinha (*Gallus gallus domesticus*) foi a primeira forma de seleção realizada pelo homem que exigiu uma adaptação genética dos indivíduos com objetivo de gerar animais com habilidades específicas (CRAWFORD, 1990). Segundo Rubin et al. (2010) a criação destas raças especializadas (para a produção de carne ou de ovos) ocorreu para contornar o conflito existente entre o desenvolvimento de músculo e a parte reprodutiva do animal.

Sendo assim, a partir da domesticação o genoma da galinha foi diminuindo sua diversidade genética e aumentando a frequência de alelos favoráveis as características de interesse (CRAWFORD, 1990). A partir da seleção destes alelos foi que o melhoramento genético tradicional se desenvolveu inicialmente, selecionando os animais com fenótipos de interesse de uma forma intuitiva. Porém, com o tempo e com o aumento do conhecimento, o melhoramento genético passou a se basear na teoria da genética quantitativa, ou seja, com base na seleção de animais com fenótipos superiores para estimar o seu valor genético aditivo com a finalidade de obter altas taxas de ganho genético (DEKKERS, 1999).

Estudos realizados por Havenstein, Ferket e Qureshi (2003) comprovaram a importância do melhoramento genético na avicultura. Estes autores avaliaram o desempenho e a carcaça de duas linhagens de frango: uma comercial de 2001 (Ross 308) e outro controle canadense, e após o fornecimento de dietas representativas dos anos de 1957 e 2001, os autores concluíram que 85 a 90% do ganho alcançado se deve a seleção genética.

Portanto, por meio de programas de melhoramento avícola, muitos avanços foram obtidos para diversas características associadas à produção de aves, como: peso vivo, conversão alimentar, consumo de ração, rendimento de carcaça e partes, fertilidade, eclodibilidade, produção e qualidade de ovos, entre outras (DEKKERS, 1999).

Entretanto, todo este progresso foi alcançado sem o conhecimento preliminar do número e do efeito dos genes que atuam nas características de interesse (LEDUR et al., 2004). É de extrema importância ter este prévio conhecimento, pois as principais características de interesse econômico são quantitativas de efeito poligênico, ou seja, resultam da atuação de muitos genes, com pequeno efeito cada, além de serem afetadas pelo ambiente (LEDUR, 2001).

Além disto, a intensa seleção para uma maior taxa de crescimento trouxe consequências desfavoráveis em relação a algumas características, como: a qualidade da carne (DRANSFIELD; SOSNICKI, 1999), doenças metabólicas (ascite e morte súbita), aumento da deposição de gordura (KESSLER et al., 2000; GAYA et al., 2006). Isto se deve ao fato de existir correlações genéticas entre estas características, causadas principalmente pela pleiotropia, que consiste quando um gene afeta duas ou mais características.

As técnicas moleculares adicionadas ao programa de melhoramento genético se torna uma ferramenta poderosa para compreender estas correlações e assim, melhorar as características de interesse sem prejudicar os ganhos já alcançados. Além disto, o conhecimento da sequência de DNA é fundamental para compreender as variações existentes no genoma e como elas atuam no fenótipo, principalmente para as características que são difíceis ou de alto custo para serem medidas pelos métodos de seleção genética convencional (BURT, 2002).

2.2 Genômica no melhoramento de aves

Desde a década de 1990 a genômica vem contribuindo para o avanço da biotecnologia animal, que aplicado ao melhoramento genético tradicional, vem proporcionando estratégias de seleção mais eficientes e promovendo maiores progressos genéticos em curto prazo (COUTINHO; ROSÁRIO, 2010).

A genômica une diversas técnicas com o objetivo de desvendar e analisar a sequência do DNA. Dentre elas se destacam: (i) desenvolvimento de marcadores moleculares; (ii) construção de mapas de ligação; (iii) mapeamento de QTL; (iv) estudos de genes candidatos; (v) análises de expressão gênica (transcriptoma), de proteínas (proteômica) e do metabolismo (metabolômica); (vi) sequenciamento de DNA/RNA (sequenciamento de nova geração) e (viii) bioinformática (COUTINHO et al., 2010).

Vários estudos já foram realizados com a utilização da galinha como modelo biológico pelo fato dela possuir um curto intervalo entre gerações (gerando grandes famílias de irmãos completos em um curto prazo) e por possuir hemácias nucleadas, o que garante uma grande quantidade de DNA para estudos genômicos (CROOIJMANS et al., 1996). Outra característica importante do genoma da galinha é o fato de possuir um menor número de genes e pseudogenes comparado ao peixe-zebra, camundongo e humano (DAVEY; TICKLE, 2007). Devido à sua maior simplicidade, estudos realizados no seu genoma torna possível a

compreensão da estrutura, dos arranjos dos genes e da evolução do genoma dos animais vertebrados (DEQUÉANT; POURQUIÉ, 2005).

Em meados de 1990, mapas de ligação foram construídos com base em três populações referências: *East Lansing* (BUMSTEAD; PALYGA, 1992), *Compton* (CRITTENDEN et al., 1993) e *Wageningen* (GROENEN et al., 1998). A construção de um mapa de ligação é feita a partir da estimativa da distância entre dois marcadores ou locos, utilizando-se a taxa de recombinação como indicador quantitativo da distância. Esta distância está associada à recombinação genética (*crossing over*) existente entre os marcadores, portanto quanto maior a distância, maior a probabilidade de ocorrer um cruzamento entre as cromátides não irmãs na região entre os marcadores e produzir um número maior de recombinantes (FERREIRA; GRATAPAGLIA, 1998), sendo assim, um índice quantitativo da distância entre dois genes (GRIFFITHS et al., 1998).

Ainda na década de 90, por meio de esforços internacionais foi definido um Mapa Consenso da Galinha (GROENEN et al., 2000), integrando os três mapas de ligação, e definiu 1.889 locos em 50 grupos de ligação que cobriram 3.800 cM (centiMorgan). Posteriormente, foram incluídos 372 locos, totalizando 4.200 cM distribuídos por 53 grupos de ligação (SCHMID et al., 2005). Recentemente, um novo mapa consenso foi publicado com a inclusão de 8.599 SNPs, cobrindo 3.228 cM e incluindo microssatélites, totalizando 9.268 locos (GROENEN et al., 2009). Por meio deste aumento no número de marcadores, mais informações são geradas para o mapeamento de QTLs e consequentemente a identificação de genes de importância econômica para a avicultura (AMBO, 2007).

No Brasil, em 1999, foi desenvolvido o Projeto de Genoma da Galinha, visando inicialmente realizar o mapeamento de QTLs para características de desempenho e carcaça. Para tanto, foi estabelecida uma parceria entre a Embrapa Suínos e Aves (Concórdia/SC) e a ESALQ/USP (Piracicaba/SP), na qual duas populações referências foram desenvolvidas em esquema F_2 e denominadas TCTC e CTCT.

Ambas as populações se originaram de cruzamentos recíprocos entre uma linhagem de corte (designada TT) e outra de postura (CC). A linhagem de corte foi selecionada visando melhorar o peso corporal, rendimentos de carcaça e partes, conversão alimentar, viabilidade, fertilidade, eclodibilidade e redução de doenças e gordura abdominal. A linhagem de postura foi selecionada com objetivo de melhorar a produção, peso e qualidade dos ovos, conversão alimentar, viabilidade, maturidade sexual, fertilidade, eclodibilidade e um reduzido peso corporal (ROSÁRIO et al., 2009).

Um dos grandes objetivos de estudo de mapeamento de QTLs é compreender as bases moleculares que são responsáveis por características complexas e desta forma, utilizar as informações geradas no melhoramento genético. Para que o mapeamento de QTL seja eficiente, além da variação genética na população, deve-se levar em consideração outros fatores, como: o número e a amplitude do efeito do QTL, a herdabilidade da característica estudada, a interação entre os genes, o tamanho do genoma, a frequência da recombinação entre os QTLs e a utilização de mapas altamente saturados por marcadores informativos (TANKSLEY, 1993).

Um método proposto por Lander e Botstein (1989) para mapear QTLs é o mapeamento por intervalo, que se baseia na identificação de um QTL pontual flanqueado entre duas marcas adjacentes. Sendo assim, a posição do QTL é determinada pela posição relativa no intervalo entre os alelos e o marcador (FALCONER; MACKAY, 1996). Os métodos estatísticos utilizados para o mapeamento por intervalo são os de regressão e de máxima verossimilhança.

Ao adotar o método de regressão, dois modelos genéticos podem ser utilizados: o “line-cross” (análise de F_2) e do “half-sib” (análise de famílias de meio-irmãos). O primeiro segue a metodologia de Haley et al., (1994), supondo que as linhagens fundadoras estão fixadas para os diferentes alelos do QTL, por possuírem diferentes bases genéticas. Desta forma, é analisada a probabilidade dos genótipos dos F_2 terem sido herdados da linhagem materna ou paterna. O segundo modelo (“half-sib”) se baseia na metodologia de Knott et al., (1996) onde há uma pré-suposição sobre o número de alelos do QTL e as frequências dos mesmos dentro das linhagens fundadoras.

Nas populações da Embrapa (TCTC e CTCT) realizou-se a descrição genotípica (ROSÁRIO et al., 2009) e a construção de mapas de ligação (NONES et al., 2005; AMBO et al., 2008; ROSÁRIO et al., 2010). Diversos QTLs foram mapeados em vários cromossomos para ambas as populações. Na TCTC QTL foram mapeados no cromossomo 1 (NONES et al., 2006), 2 e 4 (BARON et al., 2010), 3 e 5 (RUY et al., 2007), 6, 7, 8, 11 e 13 (MOURA et al., 2006; BOSCHIERO, 2009), 19, 23, 24 e 26-28 (AMBO, 2007), 9, 10, 12, 14-18 (BARON et al., 2010; NONES et al., 2012). Na população CTCT foram mapeados QTLs nos cromossomos 1, 3 e 4 (ROSÁRIO et al., 2009, 2010; SILVA et al., 2011).

2.3 QTL no cromossomo 2 da galinha associado com deposição de músculo de peito

Atualmente, de acordo com o banco de QTLs (ANIMALQTLDB, 2013), 192 publicações constataram 3.919 QTLs mapeados na galinha doméstica associados com 297 características de interesse econômico. Em relação ao músculo de peito foram mapeados 119 QTLs. No cromossomo da galinha 2 (GGA2) 379 QTLs foram mapeados e associados com diversas características, como resistência a doenças (SIWEK et al., 2003; DORSHORST; SIEGEL; ASHWELL, 2011), porcentagem e peso de gordura abdominal (JENNEN et al., 2004; CAMPOS et al., 2009), produção e qualidade dos ovos (BHATTACHARYA et al., 2011; XU et al., 2011) entre outras.

Características de crescimento também foram associadas com QTLs mapeados no GGA2, tais como: porcentagem de músculo de peito (ZHOU et al., 2006b), peso de músculo de peito (IKEOBI et al., 2004; ATZMON et al., 2008; NADAF et al., 2009; TERCIC et al., 2009; NASSAR et al., 2012;) e porcentagem de peito (BARON et al., 2010). De acordo com Nassar et al. (2012), o rendimento e qualidade do músculo, em especial o músculo de peito, é uma das características de maior importância, tanto nutricional como econômica, merecendo destaque nos estudos para uma maior compreensão e posterior adição da informação ao melhoramento genético. Além disso, esta característica tem correlação genética positiva com gordura (0,34), o que dificulta a seleção de animais para esta característica (ZHOU et al., 2006b).

Além desta correlação genética, de acordo com Havenstein, Ferket e Qureshi (1994), o resultado da seleção genética em frangos de corte para maior ganho de peso resultou na diminuição do tamanho do coração, o que gera uma insuficiência na função deste órgão, prejudicando o bombeamento de oxigênio para os demais tecidos. Desta forma, torna-se fundamental uma melhor compreensão de como as características, selecionadas nos programas de melhoramento genético, em especial a de deposição de músculo de peito, se comportam e como se relacionam geneticamente com as demais características, evitando maiores prejuízos e associações negativas na integridade fisiológica das aves (GAYA, 2006).

Baron et al. (2010) realizaram o mapeamento de QTLs na população experimental F₂ da Embrapa para a análise das características de carcaça, partes de carcaça, pernas, cabeça e seus rendimentos. Utilizando 649 animais foi realizada a análise de F₂ e dentre os QTLs mapeados, merecem destaque os que foram associados com porcentagem de peito no GGA2: um significativo a 5% na posição 242 cM (entre os marcadores *MCW0185* e *MCW0264*), com 5,14% de variância fenotípica explicada pelo QTL, e um sugestivo na posição 156 cM (entre os marcadores *MCW0062* e *ADL0373*), com 3,85% de variância fenotípica. Levando em

consideração apenas o QTL significativo identificado no GGA2, a região-alvo utilizada neste estudo foi definida com base nos marcadores mencionados (*MCW0185* e *MCW0264*).

Estes resultados corroboram com os estudos realizados por Ikeobi et al. (2004); Tercic et al. (2009) e Nassar et al. (2012), que mapearam QTLs para características relacionadas ao músculo de peito na mesma região do QTL significativo mapeado por Baron et al. (2010) entre os microssatélites *MCW0185* e *MCW0264*. Porém, estas regiões de QTL são grandes e podem conter muitos genes. Nesta região de QTL, por exemplo, 66 genes foram identificados pela ferramenta *BioMart* do *Ensembl*.

2.4 Sequenciamento do genoma da galinha

O genoma da galinha foi o primeiro animal doméstico a ser sequenciado, por meio de um Consórcio Internacional (*Internacional Chicken Genome Sequencing Consortium*) (HILLIER et al., 2004). Foi utilizado o DNA de uma fêmea da espécie *Red Jungle Fowl*, que é a ancestral das linhagens e raças das galinhas domésticas atuais (FUMIHITO et al., 1994).

O genoma da galinha é relativamente pequeno (~1,07 Mb), ou seja, aproximadamente 40% do tamanho do genoma humano e do camundongo. Isto porque, existe uma baixa quantidade de sequências repetitivas, além da presença de íntrons pequenos, comparado ao genoma dos mamíferos (DEQUÉANT; POURQUIÉ, 2005).

A tecnologia de sequenciamento de DNA foi inicialmente descrita por Sanger, Nicklen e Coulson (1977), e desde então, por mais de duas décadas, o sequenciamento foi realizado fundamentado nesta metodologia e na tecnologia de eletroforese baseada na fluorescência (LI et al., 2009). O método consiste em adicionar nucleotídeos modificados (didesoxirribonucleotídeos) a uma cadeia simples de DNA por meio da enzima DNA Polimerase. Porém, o próximo nucleotídeo é impedido de ser adicionado à cadeia, devido à ausência hidroxila na extremidade 3' no nucleotídeo modificado. Após a formação de diversos fragmentos de diferentes tamanhos, os mesmos são separados por eletroforese para detectar o nucleotídeo presente (SANGER; NICKLEN; COULSON, 1977; SANGER; COULSON, 1975).

Em seguida a técnica se tornou automática, utilizando sequenciadores com eletroforese vertical em placa (*ABI377 Applied Biosystem*) ou eletroforese em capilar (*ABI3100 Applied Biosystem*) (SMITH et al., 1986). A partir de 2005 plataformas denominadas de tecnologias de sequenciamento de nova geração passaram a ser comercializadas gerando informações de milhões de pares de bases em apenas uma corrida.

As principais são: *454 FLX da Roche*, *Applied Biosystems SOLID*, *HiSeq da Illumina Genome Analyzer*, *PacBio da Pacific Biosciences* e *Ion Torrent da Life* (CARVALHO; SILVA, 2010).

Com o surgimento do sequenciamento de nova geração, dados altamente reproduzíveis e informativos e com precisão na quantificação de transcritos foram gerados (CARVALHO; SILVA, 2010), aumentando e melhorando as informações contidas nos mapas genômicos e consequentemente promovendo importantes avanços no melhoramento genético avícola.

Estas novas tecnologias de sequenciamento geram eficientemente grandes quantidades de sequências, tanto em termos de tempo, como de custo. São fundamentadas em princípios diferentes quando comparadas com o método clássico de Sanger e por isto são denominadas de segunda ou próxima geração (MARGUERAT et al., 2008; FOX et al., 2009). As plataformas de sequenciamento de nova geração são uma alternativa poderosa para estudos de genômica estrutural e funcional (CARVALHO; SILVA, 2010).

Porém, diferentemente da tecnologia de Sanger, as plataformas de sequenciamento de nova geração produzem *reads* de comprimento menor: *Roche 454* (400-500 pb), *Illumina Genome Analyzer* (até 2x100 pb), *ABI SOLID* (25-75 pb) e *Ion Torrent* (100-200 pb) (TURNER et al., 2009). Esta característica pode ser considerada desfavorável se o objetivo do trabalho for sequenciar novos organismos, ou seja, se não existe o genoma referência para auxiliar na montagem das *reads*. Isto porque, estas sequências curtas fazem pequenas sobreposições, dificultando a formação de *contigs* e consequentemente a montagem do genoma (KATO, 2009).

Ainda segundo Kato (2009), estas pequenas *reads* também geram problemas em regiões repetidas do genoma, pois elas acabam sendo alinhadas em uma mesma região, não representando a repetição existente na sequência do genoma. Além disto, para que seja realizada uma cobertura completa do genoma, é necessário obter muitas *reads*. Desta forma, se for obtida uma boa cobertura de sequenciamento do genoma, estas novas tecnologias se tornam úteis na identificação de polimorfismos, como os SNPs e INDELS.

Algumas das aplicações do sequenciamento de nova geração incluem: o sequenciamento e ressequenciamento do genoma, identificação de genes, RNAs, SNPs, deleções e inserções, análise do transcriptoma, identificação e caracterização de *splicing* alternativos e perfis de expressão gênica (FOX et al., 2009). Estas tecnologias têm em comum a geração de sequências em uma escala sem precedentes, sem a necessidade de clonagem (MARGUERAT et al., 2008).

Desta forma esta nova tecnologia tem muito a acrescentar na genômica avícola, sendo importante nos avanços no melhoramento genético (KATO, 2009). Isto porque, a partir do seu

uso, torna-se possível a identificação de muitos marcadores SNPs e INDELS para linhagens/populações específicas, e conseqüentemente a identificação de alelos relacionados às características quantitativas de interesse econômico (COUTINHO et al., 2010).

Porém, uma das grandes dificuldades encontrada ao utilizar estas plataformas de sequenciamento de nova geração é o grande número de dados gerados. Isto se deve ao fato de que estes dados são gerados em um menor tempo e menor custo, o que aumenta o interesse e a utilização em todo o mundo. Para tanto, diversos programas são utilizados para auxiliar na análise destes dados, área denominada de bioinformática.

A primeira etapa das análises de dados de sequenciamento de nova geração consiste em verificar a qualidade das *reads* obtidas, para que, em seguida seja realizada uma limpeza, com a remoção das *reads* de baixa qualidade. Em seguida as *reads* são alinhadas contra a sequência de um genoma referência, presente nos bancos de dados públicos. Após esta etapa, é possível identificar variações genéticas (SNPs, INDELS e CNVs) e então realizar a anotação funcional destas variantes.

Para a análise de qualidade das *reads*, podemos utilizar o programa chamado FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), que permite a visualização por meio de gráficos e tabelas de diferentes parâmetros de qualidade do sequenciamento, como o conteúdo de GC, número de *reads*, etc. Após esta análise é possível detectar problemas com os dados antes da realização de uma análise mais aprofundada.

Para a realização do alinhamento das *reads* contra um genoma referência, de acordo com Hamada et al. (2011), diversos programas estão disponíveis: Bowtie (LANGMEAD et al., 2009), Bowtie2 (LANGMEAD; SALZBERG, 2012), SHRiMP2 (DAVID et al., 2011), BWA (LI; DURBIN, 2010), Stampy (LUNTER; GOODSON, 2011), MAQ (LI; JUE; RICHARD, 2008), PerM (CHEN et al., 2009), entre outros.

Dentre estes, o BWA se mostra eficiente no alinhamento de seqüências curtas com uma grande seqüência referência, evitando *gaps*. Este programa é considerado de 10-20 vezes mais rápido em relação ao MAQ, mantendo a mesma precisão, além de gerar o alinhamento no formato SAM, que é útil para as análises seguintes (LI; DURBIN, 2010). Porém o Bowtie2 se destaca pela sua capacidade de resolver o problema de reconhecimento de regiões de *gaps* que podem ser erros de sequenciamento ou verdadeiras inserções ou deleções (LANGMEAD; SALZBERG, 2012).

Para a identificação de SNPs e INDELS, o programa SAMtools (LI et al., 2009) é indicado, já que permite a utilização e conversão de arquivos do tipo SAM/BAM, classifica os alinhamentos obtidos, remove as duplicadas de PCR, identifica e gera diversas informações

dos polimorfismos (SNPs e INDELS). Após a identificação de variações genéticas, torna-se possível, por meio do programa ANNOVAR (WANG; LI; HAKONARSON, 2010) e do VEP (MC LAREN et al., 2010) realizar a anotação funcional.

2.5 Identificação de polimorfismos: SNPs e INDELS

O mapeamento de QTL no genoma da galinha é o primeiro passo para que possamos compreender as variações existentes e conseqüentemente aplicá-las no melhoramento genético. Porém, para que o mapeamento de QTL seja possível um dos fatores imprescindíveis é a existência de mapas contendo um grande número de marcadores moleculares. Estes marcadores são pontos de referência no genoma e podem ou não estar associado à expressão de genes (BOSCHIERO, 2006).

Dentre os marcadores utilizados nos mapas genéticos, estão os microssatélites, SNPs e INDELS. Os microssatélites são seqüências de até seis nucleotídeos que se encontram repetidas no genoma em tandem e são co-dominantes, estando presentes nos dois alelos de um indivíduo heterozigoto (DODGSON et al., 1997). Os SNPs são mutações no DNA onde apenas uma base é alterada com uma frequência mínima de 1% de uma dada população (KWOK; GU, 1999). A princípio, os SNPs podem ser bi-, tri-, ou tetra-alélicos, porém, são mais conhecidos como marcadores bi-alélicos. As INDELS são pequenas inserções ou deleções no genoma. É o segundo tipo de polimorfismo mais comum, ficando atrás apenas dos SNPs (MULLANEY et al., 2010).

O uso dos SNPs é interessante, pois estão presentes em grande quantidade em praticamente todas as populações, mesmo sendo bialélicos, ou seja, tendo uma baixa informação polimórfica ao se comparar com os multialélicos (MEUWISSEN et al., 2001). Além disto, estão sendo utilizados cada vez mais como marcadores moleculares, gerando mapas genéticos altamente saturados e polimórficos (KUMAR; BANKS; CLOUTIER, 2012).

Na galinha, mais de 2,8 milhões de SNPs, entre outros polimorfismos, já foram identificados a partir da comparação da seqüência do genoma do ancestral da galinha com seqüências obtidas em três linhagens domesticadas: um macho de corte (*White Cornish*), uma fêmea de postura (*White Leghorn*) e uma fêmea de uma espécie ornamental (*Silkie chinesa*) (WONG et al., 2004). Rubin et al. (2010), ressequenciaram *pools* de 88 animais provenientes de quatro linhagens de postura, quatro linhagens de frangos de corte e duas populações de *Red Jungle Fowl*. Como resultado, identificaram cerca de 7 milhões de SNPs em diferentes linhagens de galinhas e quase 1.300 deleções.

Kranis et al. (2013), num grande projeto entre o Instituto Roslin (Escócia) e diversas empresas de melhoramento avícola, ressequenciaram *pools* de 243 aves de 24 diferentes linhagens (corte, postura, comerciais e experimentais) e identificaram 78 milhões de SNPs segregando em uma ou mais linhagens. Este imenso número foi reduzido para 1,8 milhões após seleção, com o objetivo final do desenvolvimento de um *chip* denso de SNPs (600K).

As INDELs, que também podem ser identificadas por meio do sequenciamento de nova geração, podem influenciar a estrutura do gene, o *splicing* no pré-RNA mensageiro, a evolução da expressão gênica, a duplicação de genes e rearranjos cromossômicos (FONTANILLAS et al., 2007). A presença de INDELs influencia o tamanho do genoma e desta forma é um parâmetro que atua na evolução do tamanho do genoma (PETROV, 2000; GREGORY, 2005).

Brandström e Ellegren (2007) identificaram cerca de 140.000 INDELs por meio do sequenciamento de três linhagens de galinhas realizado pelo *International Chicken Polymorphism Map Consortium*. Após filtrarem as INDELs, concluíram que a densidade de INDELs foi baixa nos microcromossomos quando comparada com os macrocromossomos e baixa no GGZ comparado com os autossomos. Os resultados indicaram também que a densidade das INDELs e SNPs estão altamente correlacionadas.

A partir da identificação destes polimorfismos em todo genoma, juntamente com as informações dos fenótipos e do *pedigree* (ZHANG et al., 2012) torna-se possível realizar análises de associação e detectar mutações causais e regiões genômicas que possam conter estas mutações de forma mais efetiva que as estratégias de mapeamento de QTL (HIRSCHHORN; DALY, 2005).

2.6 Análise de associação e mutação causal

Alguns estudos já foram realizados visando detectar polimorfismos na galinha doméstica e possíveis associações com características de interesse. Estudo de associação pode ser realizado pela varredura do genoma, por meio de estudos de genes candidatos (por posição e função) e SNPs (SHORT et al., 1997).

Estudos de genes candidatos por função se baseiam no fato de uma mutação dentro de um gene que codifica uma proteína possa estar relacionada diretamente ou indiretamente a uma característica de importância produtiva (FONTANESI, et al., 2008). Porém, a principal desvantagem desta estratégia é que são poucos os genes com característica quantitativa

conhecida e existe uma dificuldade em relacionar o efeito do gene com a variante causal, principalmente para genes com efeito menor no fenótipo (COUTINHO; ROSÁRIO, 2010).

Polimorfismos identificados na família de genes *MyoD* foram associados com características de carcaça (YIN et al., 2011). Zhang et al. (2009), identificaram dois SNPs no gene *calpaína 3* e encontram associações dos genótipos, haplótipos e diplótipos com peso vivo, peso de carcaça, peso do músculo do peito e peso do músculo de pernas. Em relação gene *hormônio do crescimento*, polimorfismos foram associados ao peso corporal e ganho de peso (NIE et al., 2005). Em bovinos, polimorfismos encontrados no gene *Diacylglycerol O-acyltransferase homolog 1 (DGATI)* foram associados a maior quantidade de gordura presente no leite (GROBET et al., 1997; WINTER et al., 2002; GRISART et al., 2002).

Na população F₂ da Embrapa diversos estudos foram realizados para identificar polimorfismos em genes candidatos e associá-los à características de interesse. Felício et al. (2013a) associaram um SNP no gene *CAPN1* com peso corporal dos 35 aos 42 dias, peso de coxas, peito e carcaça e com luminosidade da carne. Outro polimorfismo foi identificado no gene *CAPN3* e associado ao rendimento das coxas, perdas de água por cozimento da carne e força de cisalhamento. Nos genes *FGFBP1* e *FGFBP2* SNPs foram associados a desempenho, carcaça e qualidade de carne (FELÍCIO et al., 2013b).

Dentro da mesma população foi identificado um SNP no gene *KLF3* e associado aos pesos de fígado, coxas, ganhos de peso dos 35 aos 41 dias. Outro polimorfismo no gene *PPARGCIA* foi associado aos pesos das asas, cabeça, carcaça, dorso, coxas, peito, fígado e gordura abdominal (PÉRTILLE, 2013). Boschiero et al. (2013) identificaram SNPs nos genes *IGF1* e *KDM5A* e associaram à peso corporal, percentual de hematócrito, consumo de ração e porcentagem de gordura abdominal.

Em relação às mutações causais, elas são difíceis de serem encontradas e comprovadas, existindo apenas alguns resultados conhecidos (ANDERSSON, 2001). Um dos estudos mais conhecidos é o do gene *Halotano* em suínos, que é responsável pela deposição de músculo na carcaça, porém uma mutação neste gene é responsável por animais com uma carne mais pálida, mole e exsudativa (chamada PSE, em inglês *pale, soft e exsudative*) (BRIDI et al., 2006). Portanto, realizando um simples teste genético é possível identificar a mutação causal e diferenciar os animais normais dos heterozigóticos e recessivos, determinando quais animais não possuem a mutação (FUJII et al., 1991).

Uma deleção de 11 pb na sequência que codifica o gene *da miostatina (MSTN)* foi associada ao fenótipo de musculatura dupla encontrada em bovinos (GROBET et al., 1997). Essa mutação gera uma proteína não funcional, levando a um aumento na massa muscular do

animal, principalmente devido ao efeito da hiperplasia (aumento no número de fibras musculares). Porém, ocorrem prejuízos ao fenótipo do animal, tais como: a diminuição na quantidade de gordura intramuscular, responsável pelo marmoreio, problemas na fertilidade das fêmeas e em relação à tolerância ao estresse (POTTS et al., 2003). Ainda em bovinos uma mutação no éxon 8 do gene *DGATI* foi relacionado a gordura no leite (WINTER et al., 2002; GRISART et al., 2002).

Em ovinos, duas importantes mutações foram identificadas, uma no gene *BMPR1B* afetando a taxa de ovulação e outro no gene *CLPG* causando uma hipertrofia muscular, principalmente no membro pélvico do animal (quarto traseiro) (MULSANT et al., 2001; WILSON et al., 2001). Em frangos, uma mutação causal no gene *BMP12* gera o fenótipo de pescoço pelado (MOU et al., 2011). Outros estudos em animais domésticos, como bovinos, suínos e ovinos, associaram mutações causais com características de produção (BRAUNSCHWEIG, 2010).

3 OBJETIVOS

O objetivo geral deste trabalho foi identificar SNPs e INDELS (pequenas Inserções e Deleções) em uma região-alvo no cromossomo 2 da galinha (GGA2), que foi associada anteriormente com deposição de músculo de peito.

Sendo assim, os objetivos específicos foram:

(i) Realizar o ressequenciamento do genoma completo da galinha por meio do sequenciamento de nova geração de 18 aves de duas linhagens divergentes parentais experimentais, de corte e de postura, desenvolvidas pela Embrapa Suínos e Aves;

(ii) Realizar uma filtragem dos polimorfismos identificados por meio de índices de qualidades;

(iv) Realizar a anotação funcional dos polimorfismos selecionados e identificar polimorfismos potencialmente relacionados com a característica de interesse;

(v) Construir um catálogo detalhado dos polimorfismos detectados nesta região em duas populações experimentais Brasileira.

4 MATERIAL E MÉTODOS

4.1 População experimental, extração e quantificação do DNA

Para a realização deste projeto foram utilizados DNA genômico de 18 animais das linhagens parentais de corte (designada TT) e de postura (designada CC) desenvolvida pela Embrapa Suínos e Aves (Concórdia/SC). A linhagem TT se originou do cruzamento de linhagens comerciais provenientes das raças *White Plymouth Rock*, *New Hampshire* e *White Cornish*, enquanto que a CC originou se da *White Leghorn*.

Utilizando estas duas linhagens foi desenvolvida uma população experimental F₂ pela Embrapa Suínos e Aves para mapeamento de QTLs. Foram realizados cruzamentos da linhagem TT com a CC, na proporção de um macho para uma fêmea, sendo utilizados no total de sete machos e sete fêmeas. As aves foram mantidas em gaiolas individuais com controle de *pedigree* e os ovos identificados para possibilitar o anelamento dos pintos da primeira geração, chamados de F₁, ao nascer.

Para a formação da geração F₂, foram escolhidos aleatoriamente um macho e três fêmeas da geração F₁, gerando 21 famílias de irmãos completos. Em seguida, essa população F₂ foi numerada, com controle de *pedigree* individual, e suas características de crescimento e carcaça foram avaliadas. Na Figura 1 é possível visualizar a estrutura da população. Na Tabela 1 estão descritos os 18 animais das linhagens parentais que foram estudados no projeto.

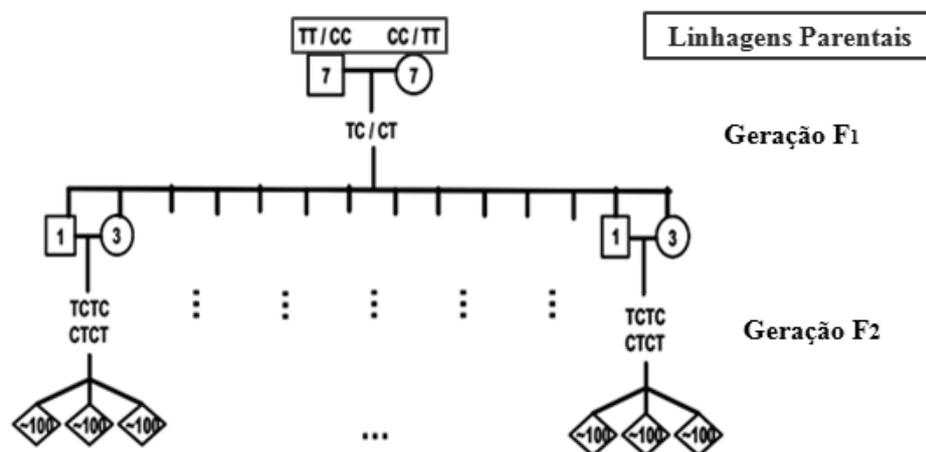


Figura 1 - Estrutura da população desenvolvida pela Embrapa Suínos e Aves (Concórdia/SC) (Adaptado de ROSÁRIO et al., 2009)

Tabela 1 – Animais parentais sequenciados, sendo nove da linhagem de corte (TT) e nove da de postura (CC)

Linhagens	Animais						
	Fêmeas				Machos		
Corte	TT5461	TT5586	TT5642	TT5561	TT5649	TT5921	TT6037
					TT6232	TT6270	
Postura	CC241	CC332	CC37	CC5	CC570	CC1	CC372
		CC88	CC886				

A extração de DNA dos 18 animais foi realizada a partir das amostras congeladas de sangue cedidas pela Embrapa Suínos e Aves. Em função de algumas amostras estarem coaguladas, empregou-se o protocolo baseado na enzima Proteinase K, conforme protocolo modificado de Maniatis et al. (1982). A concentração do DNA foi determinada pelo *NanoDrop*TM 2000 e sua integridade foi verificada em gel de agarose a 1%.

4.2 Sequenciamento de Nova Geração

O Sequenciamento de Nova Geração foi realizado em duas plataformas da Illumina: *HiScanSQ* e *HiSeq1000*, de acordo com a disponibilidade no Laboratório de Biotecnologia Animal (ESALQ/USP). Apesar da utilização de duas plataformas diferentes, a tecnologia de sequenciamento foi a mesma para os dois equipamentos (sequenciamento por síntese) e ambas realizam a *bridge amplification* para clonar os fragmentos. O que difere nestas duas plataformas é a capacidade de geração de dados, o *HiScanSQ* gera 150 Gb (Giga bases) enquanto o *HiSeq1000* gera 300 Gb.

4.3 Preparação das amostras

Antes da preparação das amostras para o sequenciamento, elas foram quantificadas novamente no *Qubit*[®] 2.0 *Fluorometer* (Life Technologies), que possibilita uma quantificação mais precisa do DNA, etapa importante para o sequenciamento de nova geração. Em seguida as amostras foram diluídas para 2,5 ng/μL, que é a concentração necessária para iniciar a preparação das amostras.

Foi utilizado o kit *Illumina Nextera*[®] *DNA Sample Preparation* (Illumina), onde a molécula de DNA é fragmentada (tamanho médio de 250 pb) por enzimas e adaptadores são adicionadas às suas extremidades. Em seguida os fragmentos são amplificados por PCR e os indexes com os adaptadores (P5 e P7) são adicionados às extremidades (Figura 2).

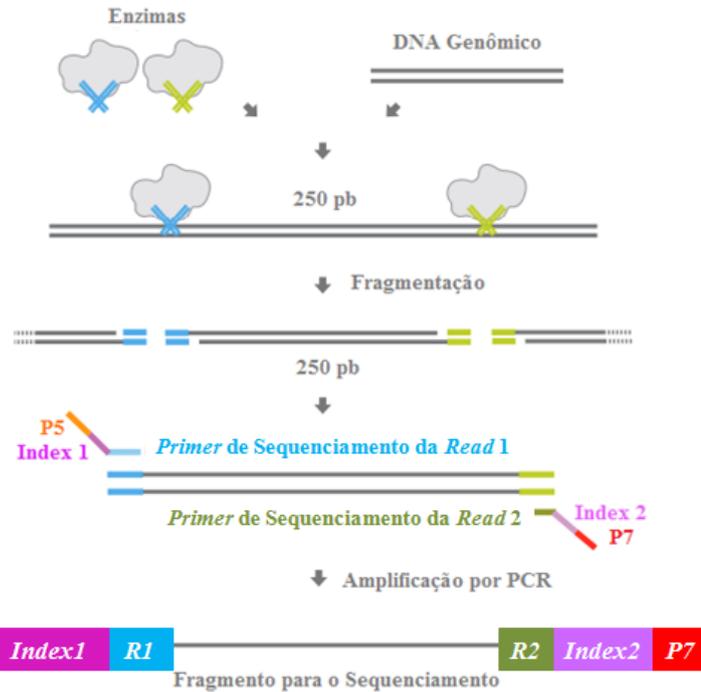


Figura 2 - Etapas da preparação das amostras: fragmentação e adição de adaptadores nas extremidades para posterior amplificação por PCR e adição dos índices com os adaptadores (P5 e P7) (Adaptado do Protocolo da *Illumina: Nextera® DNA Sample Preparation Guide*)

Após a preparação das 18 amostras, o tamanho dos fragmentos foi verificado por meio de gel de agarose (1%), e em seguida as amostras foram quantificadas por PCR em Tempo Real com o kit *KAPA Library Quantification kit* (KAPA Biosystems). A concentração de DNA nas bibliotecas foi determinada por meio de uma regressão linear determinada pelas seis amostras padrões presentes no kit.

4.4 Clusterização

Depois de calculadas as concentrações de todas as amostras, elas foram diluídas para a concentração de 20 pM e utilizando o equipamento denominado *cBot* (Illumina), estes fragmentos foram fixados à superfície das lâminas de sequenciamento (*flow cell*). A *flow cell* é dividida em oito linhas independentes e à sua superfície estão fixados oligonucleotídeos complementares aos adaptadores adicionados anteriormente às extremidades dos fragmentos de DNA (P5 e P7), fazendo com que estes fragmentos sejam fixados à *flow cell*.

Utilizando o kit *TruSeq PE Cluster kit v3-cBot-HS* (Illumina) no mesmo equipamento, foram realizadas reações sucessivas de amplificação desses fragmentos (*bridge amplification*), que ocorrem em sequência, gerando muitas cópias de cada fragmento, os quais permanecem próximos, formando os *clusters* (conjuntos de vários fragmentos idênticos entre

si) para que o sinal de fluorescência gerado tenha uma intensidade suficiente para a correta detecção pelo *laser* do equipamento. O fabricante recomenda a formação de 750 a 850 mil *clusters* por mm² da superfície da *flow cell*.

4.5 Sequenciamento

Após a clusterização, as bibliotecas foram sequenciadas no *HiScanSQ* ou no *HiSeq1000* (Illumina), de acordo com a disponibilidade dos equipamentos. Em ambos os equipamentos o comprimento de leitura foi de 2 x 100 nucleotídeos e as amostras foram distribuídas nas linhas da *flow cell* levando em consideração: a capacidade de geração de dados de cada equipamento, as combinações possíveis de *indexes* para a identificação de cada amostra e a cobertura inicial desejada por animal.

Inicialmente dois indivíduos foram sequenciados (CC88 e TT5661) no *HiScanSQ*, com o objetivo de testar e padronizar os parâmetros a serem utilizados por meio de bioinformática. Levando em consideração a capacidade deste equipamento (150 bilhões de pares de bases por corrida), e que as duas amostras foram adicionadas na mesma linha da *flow cell*, a cobertura do sequenciamento desejada inicialmente para estas duas amostras foi de 7,5X por amostra.

Em seguida, com a disponibilidade do *HiSeq1000* (que gera 300 bilhões de pares de bases por corrida), todas as 18 amostras foram sequenciadas. Porém, neste equipamento a cobertura de sequenciamento estabelecida inicialmente foi de 18X para cada amostra. Sendo assim, foram sequenciadas triplicatas de cada amostra e em cada linha da *flow cell* foram utilizadas um *pool* de seis amostras escolhidas de forma aleatória com seus respectivos *indexes* (Figura 3). Nota-se que no *pool* dos seis últimos animais (Linha 7) não foi sequenciado em triplicada. Sendo assim, foi necessária mais uma corrida de sequenciamento (realizada no *HiScanSQ*) para obter a cobertura de 18X estipulada inicialmente.

Linha 1	Linha 2	Linha 3	Linha 4	Linha 5	Linha 6	Linha 7	Linha 8
CC5	CC5	CC5	CC241	CC241	CC241	CC570	Não foi utilizada
CC37	CC37	CC37	CC332	CC332	CC332	CC886	
CC88	CC88	CC88	CC372	CC372	CC372	CC1	
TT5649	TT5649	TT5649	TT6037	TT6037	TT6037	TT5461	
TT5921	TT5921	TT5921	TT6232	TT6232	TT6232	TT5586	
TT6270	TT6270	TT6270	TT5661	TT5661	TT5661	TT5642	

Figura 3 – Esquema da distribuição das 18 amostras nas oito linhas da *flow cell* para o sequenciamento no *HiSeq1000* (Illumina). Foi realizada triplicada dos dois primeiros *pools* com seis amostras cada

Para o sequenciamento das amostras em ambos os equipamentos foram utilizados o kit *TruSeq SBS kit V3, 200 cycles (HS)* (Illumina) que realiza a incorporação dos nucleotídeos em cada uma das fitas dos *clusters*. Em seguida, o equipamento faz a leitura da superfície da *flow cell* através de imagens e reconhece o nucleotídeo incorporado pelo comprimento que o mesmo emite quando excitado pelos *lasers* do sequenciador.

4.6 Processamento inicial dos dados

Os dados resultantes do sequenciamento foram gerados em forma de imagens. O programa *Consensus Assessment of Sequence and Variation* (CASAVA, Illumina, versão 1.8.2) tem a capacidade de analisar estas imagens brutas e localizar os *clusters*, estimando a intensidade e ruído de cada um. A partir disto, o programa faz a leitura de cada base, realizando o chamado *demultiplex*, ou seja, retira os indexes que serviram para diferenciar as amostras que foram sequenciadas na mesma linha da *flow cell*. Além disto, o programa transforma os arquivos que estão no formato *.bcl para o formato *.fastq.gz (que são arquivos do tipo FASTQ compactados).

Em seguida, os arquivos foram descompactados para *.fastq e uma checagem inicial da qualidade do sequenciamento foi realizada com o programa FastQC, o que resulta num relatório da qualidade das *reads* por meio de gráficos e tabelas.

Com base na qualidade inicial das *reads*, foi realizada uma filtragem utilizando a ferramenta SeqClean (versão 1.3.12, ZHBANNIKOV; HUNTER, 2013), com *scores* mínimos de qualidade 24 e comprimento mínimo do fragmento de 65 pb.

4.7 Alinhamento

O alinhamento das *reads* contra o genoma referência da galinha (*Gallus_gallus* 4.0, NCBI) foi realizado com o *Bowtie2* (versão 2.1.0, LANGMEAD; SALZBERG, 2012), um programa rápido, preciso, com alta sensibilidade e capaz de analisar *reads* maiores que 50 pb. Segundo os mesmos autores, o *Bowtie2* permite também solucionar problemas causados pelo sequenciamento de baixa qualidade, ou quando existem regiões de *gaps* (lacunas), que podem ser verdadeiras inserções ou deleções ou erros de sequenciamento. Após o alinhamento, foi gerado um arquivo no formato *.sam (do inglês, *Sequence Alignment/Map*) ou a representação binária (*.bam) que é equivalente ao SAM, porém é mais compacto.

4.8 Definição da região-alvo

As etapas seguintes ao alinhamento, como a identificação de polimorfismos e anotação funcional, foram realizadas apenas numa região de interesse do genoma, ou seja, uma região de um QTL significativo a 5%, que foi associado com porcentagem de músculo de peito (F= 15,32 e 5,14% da variação fenotípica explicada pelo QTL, BARON et al., 2010). Esta região-alvo foi localizada no cromossomo 2 da galinha entre os marcadores microssatélites *MCW0185* e *MCW0264*.

Utilizando os pares de *primers* flanqueadores de cada marcador (obtidos no *Ensembl*) (Tabela 2) e a ferramenta BLAST (*Basic Local Alignment Search Tool*) foi possível determinar a posição exata da região-alvo no genoma referência da galinha (*Gallus_gallus*-4.0) em três bancos de dados: *NCBI* (<http://www.ncbi.nlm.nih.gov/>), *Ensembl* (<http://www.ensembl.org/index.html>) e *Pré-Ensembl* (<http://pre.ensembl.org/index.html>). Desta forma, foi determinada a região do QTL no genoma: 105.848.755-112.648.761 pb, que é composta por 66 genes (*BioMart - Ensembl*).

Tabela 2 – *Primers* direto e reverso dos marcadores microssatélites que delimitaram a região-alvo estudada no GGA2 obtidos do *Ensembl*

Marcadores	<i>Primer</i> Direto	<i>Primer</i> Reverso
<i>MCW0185</i>	TGAATAGATTTTCAGTGAGTGC	GATCTACTGTCATTTTAGTTT
<i>MCW0264</i>	CTTACTTTTCACGACAGAAGC	AGACTGAGTCACACTCGTAAG

4.9 Identificação e filtragem de polimorfismos

Após o alinhamento das *reads* e com a posição exata no genoma da região-alvo a ser estudada (105.848.755-112.648.761 pb), foi possível identificar os polimorfismos com o programa SAMtools (versão 1.4, LI et al., 2009). Mas antes disso, alguns procedimentos foram necessários, como: (i) remover as duplicatas de PCR; (ii) realizar a ordenação posicional e indexação e (iii) checar o arquivo BAM inicial, quanto ao número de *reads* mapeadas e as presentes na fita direta ou reversa (LI et al., 2009).

Para a detecção de polimorfismos (SNPs e pequenas INDELS) o programa SAMtools (opção *mpileup*) utiliza diferentes tipos de informações, tais como: o número de *reads* que apresentam *mismatch* em relação ao genoma referência, qualidade da sequência e taxa de erro de sequenciamento. Quando o programa detecta uma *mismatch* é mais provável que esta seja devido aos erros no sequenciamento do que realmente devido à uma mutação (LI et al., 2009). Antes da identificação de polimorfismos, foram testadas várias opções do SAMtools (*mpileup*) para a definição dos melhores parâmetros (Tabela 3).

Tabela 3 – Diferentes parâmetros testados para definir o comando utilizado para a identificação de SNPs e INDELS com o programa SAMtools e opção *mpileup*

Parâmetros	Definição
C50	Reduz o efeito das <i>reads</i> com excesso de <i>mismatches</i>
m2	Mínimo de duas <i>reads</i> alinhadas para predição da INDEL
E	Computação BAQ estendida
D100	Limita a cobertura a no máximo de 100 <i>reads</i>
B	Desativa o realinhamento probabilístico para o cálculo da qualidade da base de alinhamento (BAQ), reduzindo os falsos positivos
A	Não remove os pares anômalos na identificação de variante
h	Coeficiente de erros de homopolímeros
S	Cálculo de <i>strand bias</i>

Com base nos resultados dos testes efetuados, foi definido o seguinte comando para a identificação dos polimorfismos:

```
samtools mpileup -q20 -Q20 -AB -r Chr2:105,848,755-112,648,761 -ugf genoma_Ggallus.fa
arquivo.bam | bcftools view -bvcg - > arquivo.raw.bcf.
```

A opção `-q20` significa uma qualidade mínima de mapeamento de 20 e o `-Q20` significa a qualidade mínima da base de 20. As opções `-A` e `-B` estão descritas na Tabela 3. As demais opções utilizadas no comando estão descritas na Tabela 4.

Tabela 4 - Descrição das opções utilizadas para a identificação de polimorfismos por meio do programa SAMtools (opção *mpileup*)

Opção	Definição
<i>mpileup</i>	
f	Arquivo referência no formato FASTA
g	Calcular as probabilidades genóticas e gerar no formato BCF
u	Semelhante à opção g, porém gera o formato BCF descompactado
<i>bcftools view</i>	
b	Saída no formato BCF
v	Saída com as variantes genéticas e respectivas posições
c	Chamada de variantes usando a inferência Bayesiana
g	Chamada dos genótipos das variantes para cada amostra

Após a detecção dos polimorfismos foi necessário realizar uma filtragem precisa dos mesmos, e para isso foram utilizados quatro parâmetros: qualidade dos SNPs/INDELs ($Q \geq 30$, onde o valor 30 representa o *phred score*, ou seja, aceitação de 1 erro a cada 1.000 bases); cobertura total mínima da mutação ($DP4 \geq 5$, o DP4 representa a soma total de *reads* na fita direta e reversa de cada mutação); SNPs/INDELs presentes em ambas as fitas (direta e reversa, para evitar a chamada *strand bias*) e remoção dos SNPs/INDELs com cobertura máxima maior que três vezes o desvio padrão da cobertura média (também levando em consideração o valor de DP4).

4.10 Anotação Funcional

Após a filtragem dos polimorfismos foi realizada a anotação funcional dos SNPs e INDELs com o programa ANNOVAR (*Annotate Variation*) (versão 2013aug23, WANG et al., 2010). Esta etapa consistiu em identificar a localização dos SNPs/INDELs e consequentemente possíveis efeitos funcionais dos polimorfismos detectados em regiões codificadoras dos genes. O ANNOVAR fornece a localização no genoma dos polimorfismos (região *downstream/upstream*, 3' UTR/5' UTR, intergênica, éxon, íntron, RNA não codificante (ncRNA), *splicing*, etc) e seu efeito funcional: SNP sinônimo ou não sinônimo, SNP *stopgain* (causa *stop* códon) e *stoploss* (muda o *stop* códon) e INDEL *frameshift* ou não *frameshift*, como está ilustrado na Figura 4.

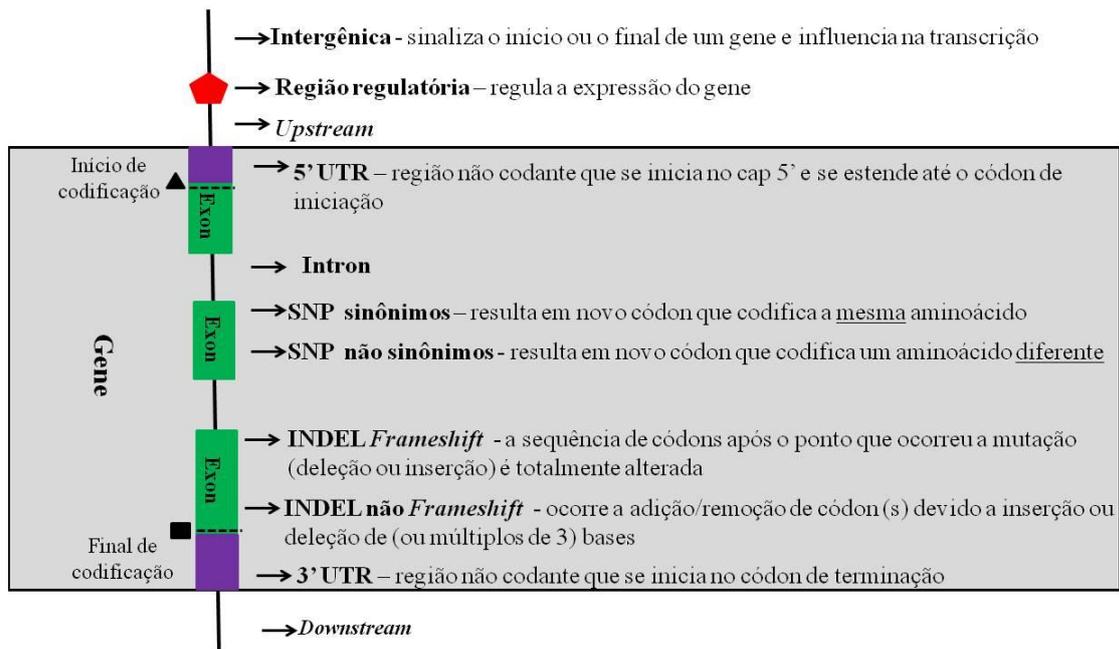


Figura 4 – Esquema ilustrativo da estrutura do DNA e o efeito de cada polimorfismo (SNP ou INDEL) em relação à sua ação (Adaptado de MCLAREN et al., 2010)

Após a anotação com o ANNOVAR foi utilizada a ferramenta VEP (*Variant Effect Predictor*) do *Ensembl* (<http://www.ensembl.org/info/docs/tools/vep/index.html>) para determinar o efeito dos SNPs encontrados nos genes. Estes efeitos foram verificados com base em *scores* do SIFT (*Sorting Intolerant From Tolerant*), que leva em consideração a posição do SNP e o tipo de aminoácido que é alterado, ou seja, prediz se a troca irá afetar a função da proteína. O SIFT *score* indica se a mutação foi não tolerável ($\leq 0,05$) ou tolerável ($> 0,05$).

Um SNP não sinônimo deletério é aquele que ocorre em regiões muito conservadas de uma proteína. Segundo Ng e Henikoff (2003), se em determinada região de uma família de proteína sempre existir o aminoácido leucina, presumi-se que a leucina possui uma função importante nesta proteína, e desta forma, qualquer mudança (mutação) será prejudicial, e classificado como mutação não tolerada. Outra forma de ser considerado um SNP deletério é se ocorrer uma mudança de aminoácido para outro com polaridade diferente, por exemplo, uma troca de aminoácido hidrofóbico para um hidrofílico. Um SNP tolerado é aquele em que a mudança de aminoácido ocorre em uma posição que não está relacionada com a estrutura e função da proteína (NG; HENIKOFF, 2001).

Além da anotação (SNPs e INDELs) e da predição dos *scores* SIFT (SNPs), também foram verificadas as funções biológicas dos genes onde estão presentes os SNPs (não sinônimos, *stopgain* e *stoploss*) e INDELs (*frameshift* e *não frameshift*). Para isto foi

utilizado o programa DAVID (*Database for Annotation, Visualization, and Integrated Discovery*, HUANG; SHERMAN; LEMPICKI, 2003a,b).

5 RESULTADOS E DISCUSSÃO

5.1 Extração e quantificação do DNA

Após a extração do DNA das 18 amostras do sangue congelado, foi realizada a quantificação do DNA no espectrofotômetro *NanoDrop*TM 2000 (Thermo Scientific). Além da concentração de ácido nucleico (DNA), foram obtidos os valores da absorbância em 260 e 280 nm, que é respectivamente o comprimento de onda que os ácidos nucleicos e proteínas absorvem luz.

Por meio do cálculo da razão da absorbância (260/280) pode-se verificar a pureza das amostras. Segundo o próprio fabricante a razão de pureza (260/280), o valor aceitável para DNA é de 1,8–2,0 (Tabela 5). Após esta etapa, o DNA foi diluído para a concentração de 2,5 ng/μL para a construção das bibliotecas, clusterização e sequenciamento, conforme descrito anteriormente.

Tabela 5 - Concentração do DNA, valores da absorbância (260 e 280 nm) e a razão 260/280 para os 18 animais realizada no espectrofotômetro *NanoDrop*TM 2000

Animal	Concentração de DNA (ng/μl)	A260	A280	260/280
CC1	74,6	1,491	0,929	1,61
CC5	617,0	12,34	6,482	1,90
CC37	28,8	0,575	0,313	1,84
CC88	43,8	0,876	0,436	2,01
CC241	75,2	1,503	0,799	1,88
CC332	89,8	1,797	0,922	1,95
CC372	50,9	1,018	0,517	1,97
CC570	284,6	5,692	3,042	1,87
CC886	240,8	4,816	2,501	1,93
TT5461	98,7	1,973	1,083	1,82
TT5586	87,8	1,876	1,011	1,85
TT5642	895,2	17,904	9,896	1,81
TT5649	18,7	0,374	0,188	1,98
TT5921	417,2	8,345	4,395	1,90
TT6037	76,1	1,522	0,770	1,98
TT6073	931,4	18,628	10,022	1,86
TT6232	570,6	11,412	6,145	1,86
TT6270	195,0	3,900	2,104	1,85

5.2 Sequenciamento e alinhamento

Após o sequenciamento dos 18 animais foi obtido um total de 2.785.354.494 *reads* (1.417.232.967 da *read 1* e 1.368.121.527 da *read 2*). Utilizando a ferramenta FASTQC foi avaliada a qualidade inicial de ambas as *reads* em relação a diferentes parâmetros de qualidade de sequenciamento, como conteúdo GC, número, qualidade e comprimento das *reads*, duplicatas e etc. Na Figura 5 (a e b), a fim de exemplificar, observa-se a qualidade das *reads 1* do animal CC1 antes e após a filtragem. Antes da filtragem (Figura 5a) nota-se que as bases no final da *read* (90-100 pb) tiveram uma qualidade menor, pois na maioria das plataformas a qualidade vai diminuindo ao longo da corrida. Após a filtragem as *reads* se mantiveram com alta qualidade (Figura 5b).

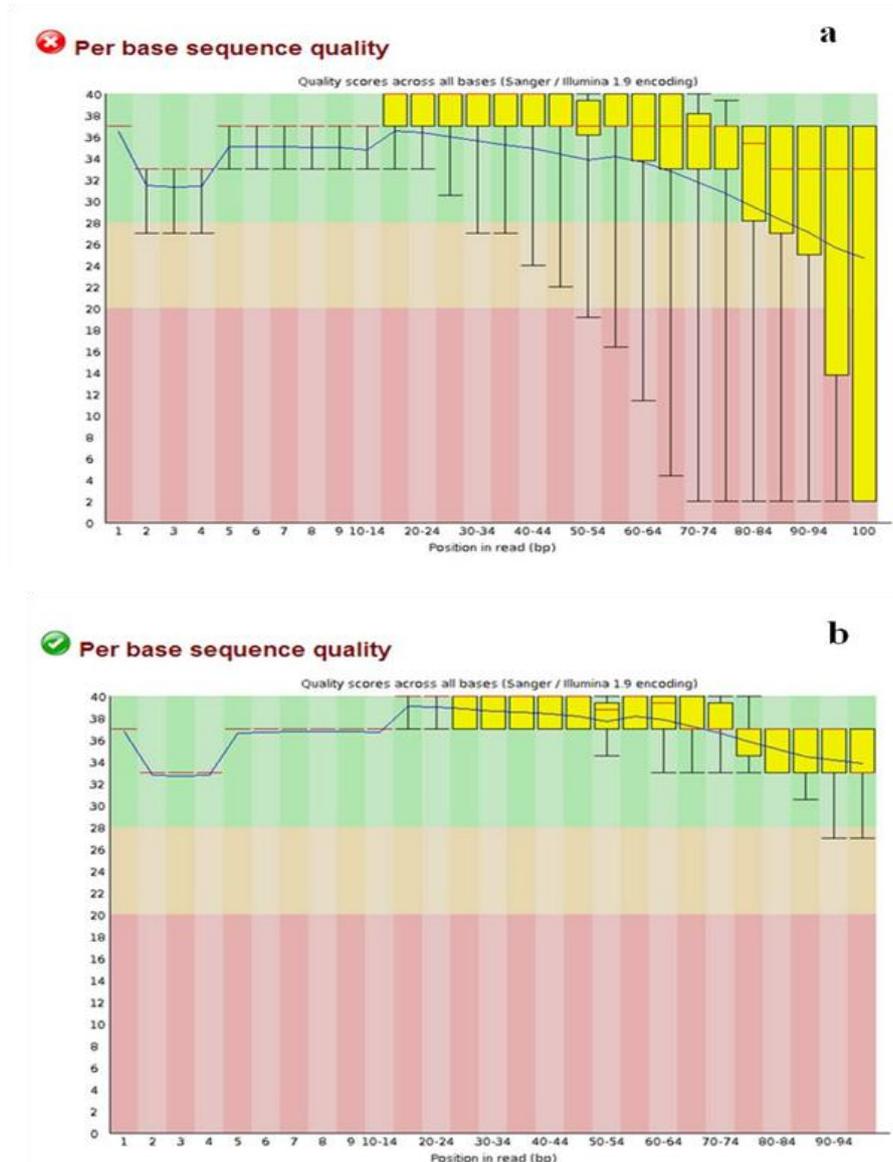


Figura 5 – Qualidade por base obtida pela ferramenta FASTQC da sequência da *read* 1 do animal CC1, antes (Figura 5a) e após (Figura 5b) a filtragem pelo *SeqyClean*. O eixo x do gráfico representa o *score* de qualidade, dividido em alta qualidade (verde, 28 a 40), média (laranja, 20 a 28) e baixa (rosa, 0 a 20). O eixo y representa a posição da base nas *reads* (0 a 94 pb). A linha central vermelha é o valor mediano, a linha azul é a qualidade média, a caixa amarela representa o intervalo interquartil (25-75%), os segmentos verticais representam o maior e menor valor observado

A filtragem das *reads* foi realizada por meio do *software SeqyClean*, mantendo um total de 2.132.638.003 *reads* para os 18 animais, ou seja ~77% do número inicial foi mantido, pois permaneceu apenas as *reads* com *score* mínimo de qualidade 24 e comprimento mínimo do fragmento de 65 pb. Portanto nota-se na Figura 5 que depois da limpeza o tamanho do fragmento diminuiu e *reads* permaneceram com alta qualidade (28-40).

Em média os 18 animais apresentaram 10,5X de cobertura de sequenciamento, sendo que a cobertura máxima foi de 15,6X (animal TT6270), e a mínima foi de 5,4X (CC372) (Figura 6). A região-alvo estudada, entre as posições 105.848.755–112.648.761 pb,

apresentou uma cobertura de 10,6X. Uma das possíveis justificativas para a variação de cobertura de sequenciamento entre os animais são as possíveis variações que tenham ocorrido durante a construção das bibliotecas e na etapa de clusterização.

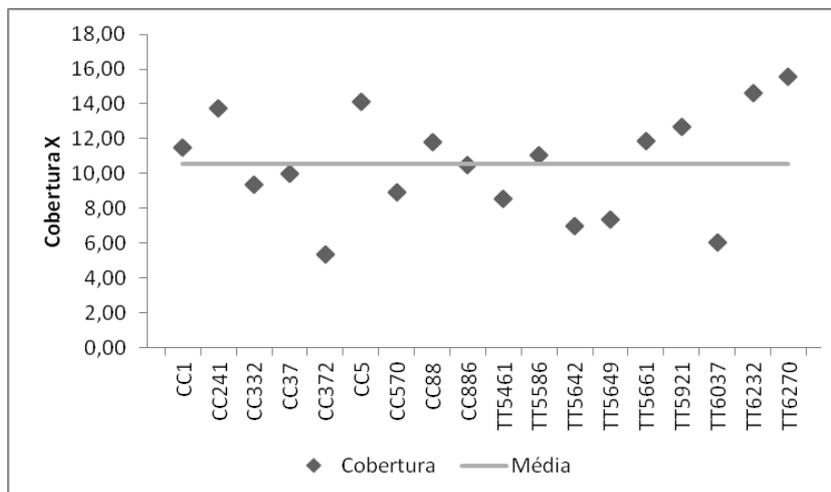


Figura 6 – Média da cobertura de sequenciamento do genoma todo e a média para cada um dos 18 animais

Os animais que foram sequenciados na mesma linha da *flow cell* apresentaram uma cobertura bem variada, não apresentando nenhum padrão. Por exemplo, os animais CC241, CC332, CC372, TT5661, TT6037 e TT6232 foram sequenciados na mesma linha e a cobertura de sequenciamento para cada animal foi bem variada, por exemplo, o animal CC372 obteve a menor cobertura (5,4X) e o TT6270 obteve a maior cobertura (15,6X).

5.3 Detecção e filtragem de SNPs e INDELS

A identificação de SNPs e INDELS foi realizada na região-alvo de interesse (105.848.755–112.648.761 pb), determinada anteriormente pelos marcadores microssatélites que delimitaram a região de QTL associada à porcentagem de músculo de peito (Baron et. al., 2010).

Utilizando a opção *mpileup* do SAMtools foram identificados inicialmente o total de 722.832 SNPs e 63.727 INDELS para os 18 animais juntos. Em média foram identificados 40.157 SNPs e 3.540 INDELS para cada animal, e em relação às linhagens, na de corte foram detectados em média 42.335 SNPs e 3.639 INDELS e na de postura, 37.980 SNPs e 3.442 INDELS. A linhagem de corte apresentou um maior número médio de mutações em relação à de postura, indicando uma alta variabilidade.

Em seguida foi realizada a filtragem destes polimorfismos utilizando quatro critérios mencionados anteriormente (qualidade da mutação, cobertura mínima e máxima e mutação

presente nas duas fitas). Em relação a qualidade *phred score* ($Q \geq 30$), apenas 5% dos SNPs inicialmente detectados nos 18 animais estavam abaixo deste critério, porém as INDELS apresentaram uma maior porcentagem (18,8%) abaixo da qualidade 30 (Figura 7).

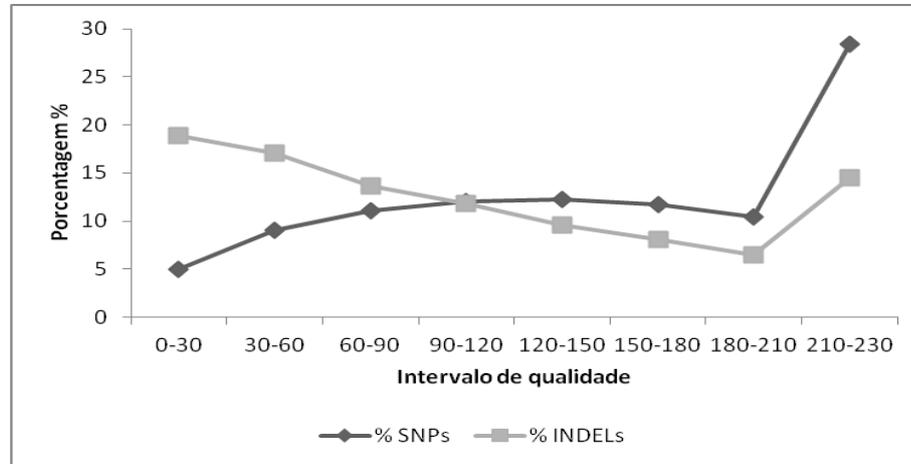


Figura 7 – Porcentagem de SNPs e de INDELS em relação aos intervalos de qualidades (0 a 230)

Segundo Li et al. (2008), esta maior presença de INDELS de baixa qualidade (0-30), quando comparado a porcentagem de SNPs, pode ser explicada devido a maior dificuldade em detectar as INDELS. Esta dificuldade se deve ao fato de que, primeiramente, o número de INDELS é cerca de oito vezes menor que SNPs (LUNTER; GOODSON, 2007; CARTWRIGHT, 2009). Há também, uma maior complexidade ao mapear *reads* em regiões de INDELS (LI et al., 2009), sendo necessárias ferramentas que possam inferir de forma correta a sua presença (NEUMAN; ISAKOV; SHOMRON, 2012). O realinhamento local é uma destas ferramentas, e o SAMtools a realiza, permitindo um realinhamento das *reads* em torno de regiões com possíveis INDELS (LI; HOMER, 2010). Além disso, muitas vezes as INDELS não podem ser mapeadas de forma única no genoma referência devido à existência de repetições, ou seja, uma deleção/inserção pode apresentar várias unidades de repetição, que resulta em um haplótipo alternativo igual à referência (ALBERTS et al., 2011).

Após a filtragem foi mantido um total de 77% dos SNPs ($n=558.767$) e 60% das INDELS ($n=38.402$) nos 18 animais. Na Tabela 6 é possível observar as mutações (SNPs e INDELS) detectadas inicialmente e após a filtragem para cada ave e a cobertura de sequenciamento para cada animal.

Tabela 6 – Número dos SNPs e INDELs identificados inicialmente pelo SAMtools e após a filtragem, a média e o número de polimorfismos únicos identificados para os 18 animais

Animais	Nº SNPs detectados	Nº SNPs após a filtragem	Nº INDELs detectados	Nº INDELs após a filtragem
CC1	39.932	33.536	3.537	2.281
CC5	41.819	37.158	4.102	2.986
CC37	37.297	29.581	3.485	2.118
CC88	40.359	33.628	3.860	2.473
CC241	36.460	32.162	3.768	2.639
CC332	38.100	29.236	3.393	1.902
CC372	30.689	16.248	2.523	887
CC570	39.195	29.060	3.114	1.751
CC886	37.965	30.461	3.193	1.984
TT5461	35.719	26.953	3.086	1.743
TT5586	43.793	31.590	3.321	1.849
TT5642	41.143	26.319	2.975	1.367
TT5649	40.645	26.547	3.407	1.653
TT5661	44.428	36.652	3.989	2.599
TT5921	46.552	38.833	4.371	2.917
TT6037	35.923	20.319	3.047	1.201
TT6232	46.973	40.261	4.136	2.808
TT6270	45.840	40.223	4.420	3.244
Média	40.157	31.043	3.540	2.133
Únicos	94.674	85.765	10.448	7.824

Após a filtragem, na linhagem de corte foram mantidos 81.514 SNPs únicos e 8.592 INDELs únicas. Na de postura, foram mantidos 62.212 SNPs únicos e 6.791 INDELs únicas. Portanto, mais SNPs e INDELs foram identificados na linhagem de corte, indicando uma maior variabilidade nesta região para os nove animais da linhagem de corte. Nesta região-alvo do GGA2 o número médio de SNPs e INDELs por Kb foi de 4,56 e 0,31, respectivamente. Este resultado é consistente com a média de SNPs encontrada por Wong et al. (2004), de 5 SNPs.kb⁻¹ e a densidade de INDEL de 0,19 INDELs.kb⁻¹ (Brandstrom & Ellegren 2007).

Dentre os polimorfismos mantidos após a filtragem, alguns estavam presentes em apenas uma das linhagens, ou seja, eram exclusivos. Na linhagem de corte foram identificados 73.666 SNPs e 6.189 INDELs exclusivos, já na de postura, 56.152 SNPs e 4.909 INDELs foram exclusivos. Estas variantes exclusivas para determinada linhagem são interessantes para estudo posterior.

Ao analisar cada um dos animais separadamente, nota-se que no animal TT5661 foi identificado um maior número de SNPs exclusivos (n=2.101) e no TT5921 o maior número de INDELs exclusivas (291). O menor número de polimorfismos exclusivos foi encontrado no animal CC372, 40 SNPs e 15 INDELs (Figura 9 a e b).

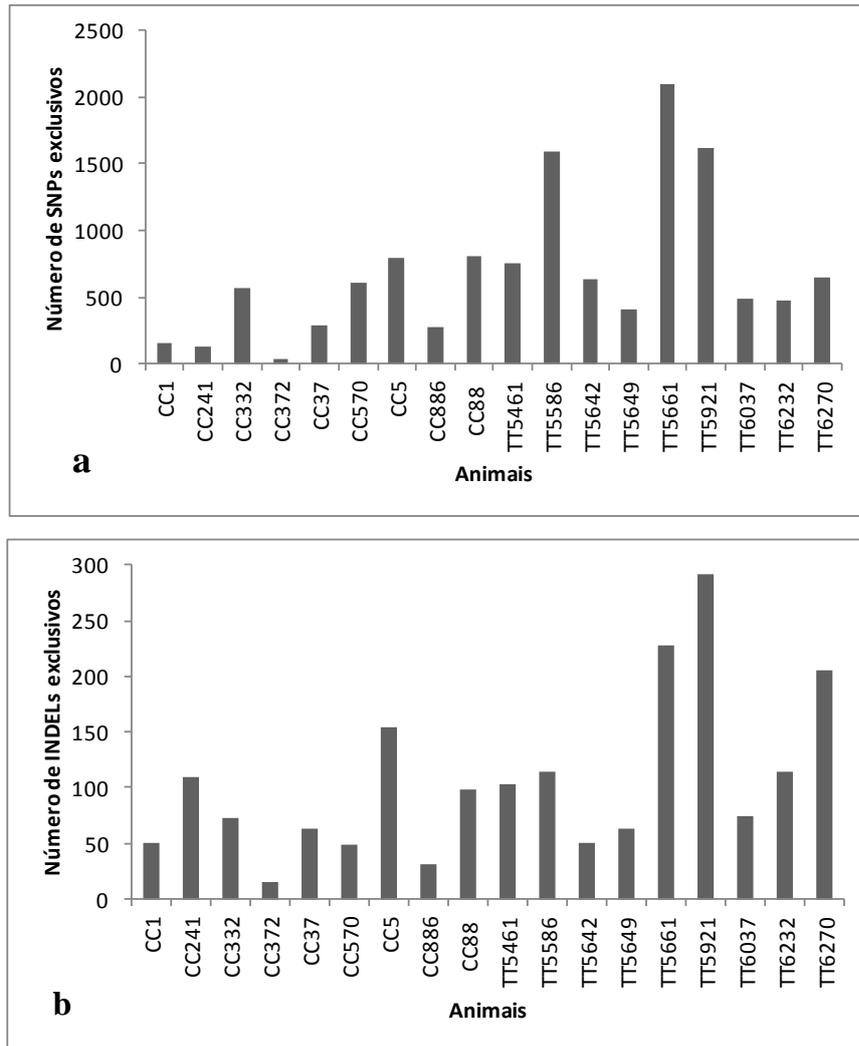


Figura 9 – Número de SNPs (a) e INDELs (b) exclusivos para cada um dos 18 animais

Dentro das informações geradas para cada mutação existe o parâmetro do SAMtools chamado GT que nos informa se um polimorfismo é heterozigoto em relação genoma referência (0/1) ou homozigoto (1/1). Sendo assim, foi identificada para cada animal a distribuição de SNPs e INDELs homozigotos e heterozigotos (Figura 10a e 10b).

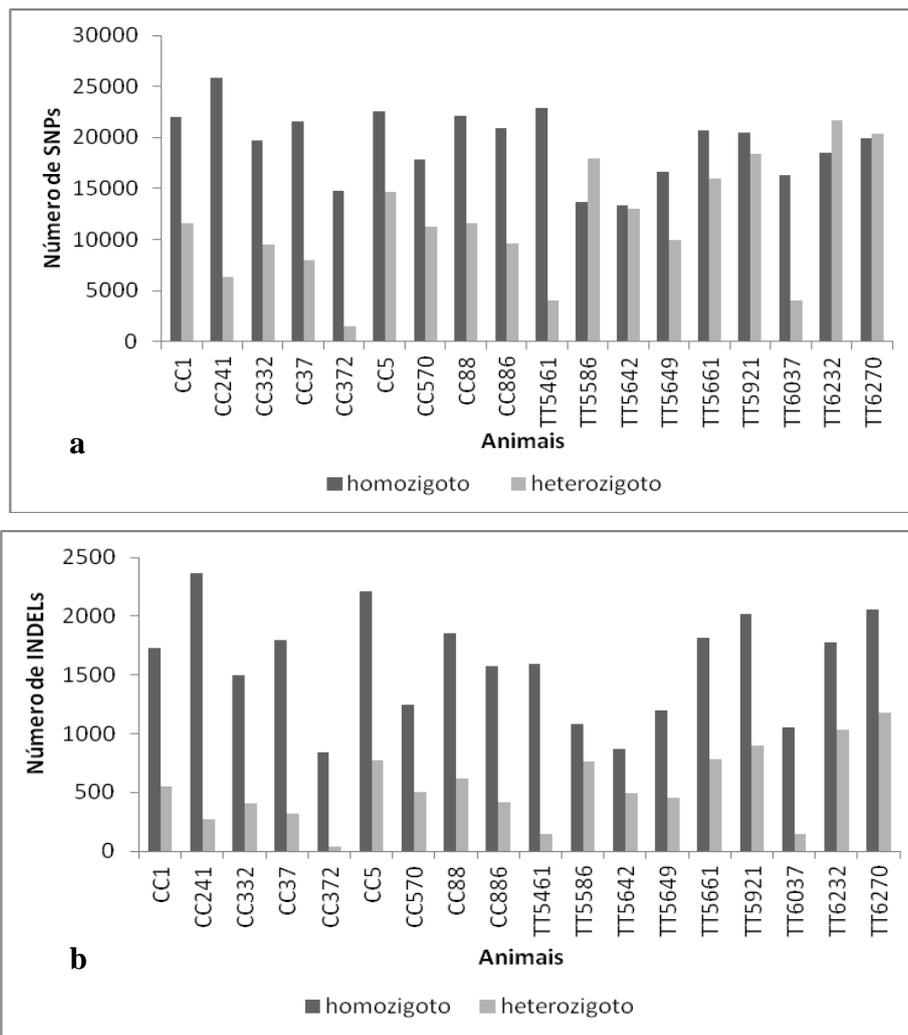


Figura 10– Porcentagens de SNPs (a) e INDELs (b) homozigotos e heterozigotos para cada animal

Em ambas as linhagens, foram identificadas mais SNPs e INDELs homozigotos do que heterozigotos. Segundo Bentley et al. (2008), para detectar uma variante heterozigota é necessário pelo menos o dobro de cobertura em relação ao homozigoto e para ambos os alelos. De acordo com os mesmos autores, quanto maior a cobertura de sequenciamento, maior é a detecção de SNPs. O maior acúmulo de heterozigotos ocorre com a cobertura de sequenciamento em torno de 30X (Figura 10). Eck et al. (2009), também afirmaram que, quanto maior a cobertura de sequenciamento, maior é a detecção de SNPs heterozigotos, diminuindo a diferença entre homozigotos e heterozigotos.

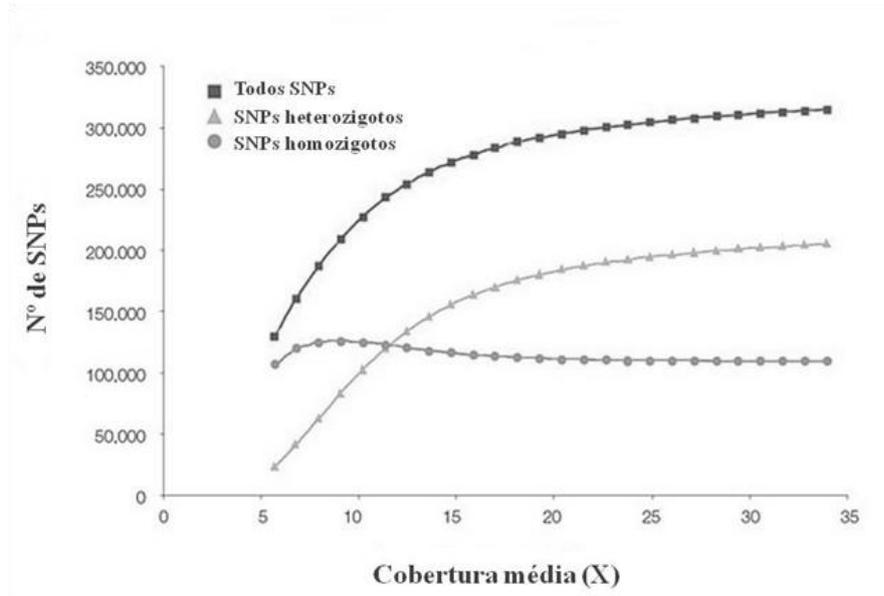


Figura 10 – Relação entre o número de SNPs heterozigotos e homozigotos e sua respectiva cobertura de sequenciamento (Adaptado de BENTLEY et al., 2008)

Todos os animais de postura apresentaram uma maior proporção de SNPs homozigotos do que heterozigotos. Porém, dentro da linhagem de corte, três animais (TT5586, TT6232, TT6270) apresentaram maior proporção de SNPs heterozigotos, o que sugere que nesta linhagem há SNPs segregando. Em relação às INDELs, todos os animais apresentaram mais homozigotos que heterozigotos.

Chen et al. (2007) explicaram que as INDELs homozigotas são mais fáceis de serem localizadas por meio da identificação de lacunas (*gaps*) durante o alinhamento das sequências, já os heterozigotos, são mais difíceis e muitas vezes são confundidos com erros de sequenciamentos.

5.4 Anotação Funcional

A anotação funcional foi realizada para as variantes únicas filtradas para os 18 animais (sem duplicatas), sendo 85.765 SNPs únicos e 7.824 INDELs únicas, por meio do *software* ANNOVAR. Inicialmente, os SNPs anotados (n=85.765) foram classificados nas regiões do genoma, tais como: regiões intergênicas (54.282), em regiões de íntrons (28.687), de éxons (513), entre outras (*splicing*, ncRNA, UTR5, UTR3, *upstream* e *downstream*). Em seguida, dentro da região de éxon, os SNPs foram classificados em: não sinônimos (n=153), sinônimos (357), *stopgain* (3) e *stoploss* (0) (Tabela 7).

Tabela 7 – Anotação dos SNPS únicos filtrados para os 18 animais

Variantes	Total	Porcentagem (%) ¹
Todos SNPs	85.765	100
Intergênico	54.282	63,29
Íntron	28.687	33,45
Éxon	513	0,60
<i>Splicing</i>	1	0,001
ncRNA	16	0,02
UTR5	78	0,09
UTR3	544	0,63
<i>Upstream</i>	880	1,03
<i>Downstream</i>	761	0,89
Éxon		
Não sinônimo	153	0,18
Sinônimo	357	0,42
<i>Stopgain</i>	3	0,003

¹A porcentagem foi calculada com base no total de SNPs anotados

As INDELS ao serem anotadas também foram classificadas em relação às regiões do genoma: regiões intergênicas (n=4.809), regiões de íntrons (2.788), de éxons (7), entre outras. As INDELS presentes em regiões de éxon foram classificadas em: *frameshift* deleção (n=1) e inserção (3) e não *frameshift* deleção (2) e inserção (1) (Tabela 8).

Tabela 8 - Anotação das INDELS únicas filtradas para os 18 animais

Variantes	Total	Porcentagem (%) ¹
Todas INDELS	7.824	100
Intergênico	4.809	61,46
Íntron	2.788	35,63
Éxon	7	0,09
<i>Splicing</i>	0	0,00
ncRNA	4	0,05
UTR5	4	0,05
UTR3	55	0,70
<i>Upstream</i>	73	0,93
<i>Downstream</i>	84	1,07
Éxon		
<i>Frameshift</i> deleção	1	0,01
<i>Frameshift</i> inserção	3	0,04
Não <i>frameshift</i> deleção	2	0,03
Não <i>frameshift</i> inserção	1	0,01

¹A porcentagem foi calculada com base no total de INDELS anotadas

A anotação de SNPs como não sinônimos e *stopgain* são mutações em regiões codantes que alteram a sequência dos aminoácidos na proteína, podendo afetar a sua função. A predição do SIFT *score* utilizando a ferramenta SIFT (*Sorting Intolerant from Tolerant*) é importante para classificar as mutações (SNPs não sinônimos, *stopgain* e *stoploss*) em toleradas, quando não afetam a função dos genes, e não toleradas, quando afetam a função gênica. Para realizar esta predição, o SIFT considera a posição em que ocorreu a mutação e o tipo de aminoácido que foi trocado, por exemplo, se ocorreu uma troca de um aminoácido hidrofóbico para hidrofílico ou ao contrário (NG; HENIKOFF, 2003).

Utilizando a ferramenta VEP foi realizada a predição do SIFT para os 153 SNPs não sinônimos e três *stopgain*. Dentre estes SNPs analisados, 15 foram classificados como deletérios, ou seja, um SIFT *score* igual ou menor que 0,05. Na Tabela 9 e 10 é possível visualizar a posição destes SNPs no genoma referência, o gene ao qual pertence, a troca de aminoácido, o SIFT *score*, se este SNP já foi depositado no banco de dados público de SNPs (NCBI) e em quais animais ele está presente.

Dentre os SNPs deletérios, sete foram identificados em ambas as linhagens (Tabela 9), cinco foram exclusivos da linhagem de corte e três exclusivos da postura (Tabela 10). Oito destes SNPs ainda não foram depositados no banco de dados (NCBI) e podem ser exclusivos das linhagens Brasileira avaliada, principalmente aqueles encontrados em vários animais, tais como: o SNP c.1295A>G (presente em 12 dos 18 animais), c.2446T>C (presente em 8 animais) e o c.797C>T (em 9 animais).

Os SNPs deletérios foram identificados em oito genes diferentes, como mostrado nas Tabelas 9 e 10. Sendo eles: *TRAPPC8*, *MEP1B*, *PRKDC*, *ST18*, *RBICC1*, *SDR16C5*, *NSMAF* e *CHD7*, que estão envolvidos em diversas funções, tais como: transporte de vesícula intracelular (*TRAPPC8*) (SCRIVENS et al., 2011); hidrólise de peptídeos e substâncias proteicas (*MEP1B*) (NORMAN et al., 2003); homeostase da célula (*PRKDC*) (LEES-MILLER, 1992); ligação do DNA a fatores de transcrição (*ST18*) (YANG et al., 2008); autofagia, apoptose, proliferação e crescimento celular (*RBICC1*) (NISHIMURA et al., 2011); relacionado à doença inflamatória na pele (*SDR16C5*) (MATSUZAKA et al., 2002); regulador a produção de ceramida (*NSMAF*) (ADAM-KLAGES et al., 1996) e regulador de transcrição e da biogênese do RNA ribossômico (*CHD7*) (ZENTNER et al., 2010).

Tabela 9 - Descrição dos SNPs deletérios em relação ao genoma referência (Gallus_gallus-4.0, NCBI) encontrados em animais de ambas as linhagens

Gene e posição (pb)	SNP	Mudança de aminoácido	SIFT score	REFSNP (NCBI) ^a	Animais
<i>MEP1B</i> (106.699.529)	A/C	Aca/Cca	0,04	c.1295A>C	CC1, CC5, CC37, CC88*, CC241, CC372, TT5586*, TT5642, TT5649, TT5661, TT6232*, TT6270,
<i>MEP1B</i> (106.700.492)	A/G	Aat/Gat	0,02	rs313118741	CC1, CC5, CC37, CC88*, CC241, TT5586, TT5649, TT5661, TT6232*, TT6270*
<i>PRKDC</i> (107.639.579)	G/A	Cat/Tat	0,02	rs14684568	CC1*, CC5*, CC37, CC241, CC332*, CC886*, TT5461, TT5586, TT5921*, TT6037, TT6270, TT6232
<i>PRKDC</i> (107.694.985)	G/A	tCg/tTg	0,03	c.116G>A	CC5, CC241, CC886, TT5586*, TT5661, TT6232*
<i>SDR16C5</i> (111.055.988)	C/T	Gca/Aca	0,00	rs312478265	CC5*, CC37*, CC241*, TT5661*
<i>NSMAF</i> (111.775.956)	T/C	aAa/aGa	0,01	c.2446T>C	CC5*, CC37, CC241, CC332, CC372, CC886*, TT5462*, TT6037
<i>CHD7</i> (112.613.760)	C/T	gCt/gTt	0,02	c.797C>T	CC5*, CC88*, CC570*, TT5642, TT5649*, TT5921*, TT6037, TT6232*, TT6270*

^aPara os SNPs ainda não depositados no banco de dados do NCBI (sem RefSNP) foi gerada a nomenclatura padrão com sua posição no cDNA e sua respectiva troca de bases. *Animais heterozigotos para o SNP

Tabela 10 - Descrição dos SNPs deletérios em relação ao genoma referência (*Gallus_gallus*-4.0, NCBI) exclusivos para animais da linhagem de corte e para a de postura

Gene e posição (pb)	SNP	Mudança de aminoácido	SIFT score	REFSNP (NCBI) ^a	Animais
<i>TRAPPC8</i> (105.986.285)	G/C	cCt/cGt	0,02	rs313738880	TT5921*, TT6232*, TT6270*
<i>MEP1B</i> (106.071.008)	C/A	tCc/tAc	0,00	c.773C>A	TT5921*, TT6232*, TT6270*
<i>PRKDC</i> (107.628.373)	C/T	Gtg/Atg	0,05	c.9841C>T	TT5461, TT5586*
<i>ST18</i> (109.283.869)	C/T	cGt/cAt	0,02	c.2354C>T	TT5642*, TT5921*
<i>NSMAF</i> (111.767.261)	C/A	Gtt/Ttt	0,01	rs313109240	TT5461*, TT5921*, TT6232*, TT6270
<i>TRAPPC8</i> (105.994.460)	G/A	Cgt/Tgt	0,00	rs314904296	CC332
<i>MEP1B</i> (106.071.068)	A/G	cAa/cGa	0,03	c.833A>G	CC88*
<i>RBICCI</i> (109.659.296)	G/T	tCt/tAt	0,04	rs314633726	CC37

^aPara os SNPs ainda não depositados no banco de dados do NCBI (sem RefSNP) foi gerada a nomenclatura padrão com sua posição no cDNA e a respectiva troca de bases.*Animais heterozigotos para o SNP

Dentre os 15 SNPs deletérios, sete já foram depositados em banco de dados (dbSNP), sendo então, uma confirmação na população estudada. Três SNPs dos sete já depositados estão presentes em ambas as linhagens, dois estão somente em animais da linhagem de corte, e dois na de postura. O SNP rs314633726 está presente apenas em um animal da linhagem de postura (CC37) de forma homozigota e merece destaque por estar posicionado no gene *RBICCI1*, que é responsável pela autofagia, proliferação e crescimento celular e apoptose (morte celular), e conseqüentemente relacionado ao desenvolvimento do músculo esquelético (NISHIMURA et al., 2011). Este SNP foi identificado em apenas um dos 18 animais avaliados, mas merece destaque por estar presente num gene de função relacionada ao QTL mapeado nesta região, e por isso, deve ser melhor estudado em outros animais da população experimental Brasileira.

Em relação aos novos SNPs identificados, quatro estão presentes tanto na linhagem de corte, quanto na de postura: c.1295A>G (*MEP1B*), c.116G>A (*PRKDC*), c.2446T>C (*NSMAF*) e c.797C>T (*CHD7*). E os outros quatro SNPs, três estão apenas na linhagem de corte: c.773C>A (*MEP1B*), c.9841C>T (*PRKDC*) e c.2354C>T (*ST18*), e apenas o c.833A>G (*MEP1B*) na de postura. Sendo que os SNPs c.773C>A e c.2354C>T estão somente na forma heterozigota, porém o primeiro está presente no gene *MEP1B* que é responsável pela hidrólise de vários peptídeos e substâncias proteicas (BUTLER et al., 1987). E o segundo está localizado no gene *ST18*, que tem como função ligar fatores de transcrição a sequência específica de DNA (JEAN et al., 1996). Desta forma, estes genes não estão diretamente relacionados com o desenvolvimento muscular, porém podem estar ligados a outros genes, devido ao efeito de ligação gênica, por exemplo, e devem ser melhor investigados.

5.5 Genes relacionados ao desenvolvimento muscular

A partir da anotação que foi realizada pelo ANNOVAR, obtivemos os genes (*genes ID*) onde estão localizados os polimorfismos (SNPs e INDELS). No total, 37 genes foram identificados pelo ANNOVAR por meio das variantes, sendo que 31 foram relacionados somente aos SNPs não sinônimos, três aos *stopgain* e três genes foram relacionados tanto a SNPs não sinônimos quanto às INDELS *frameshift* e não *frameshift*.

Utilizando o programa DAVID foi possível analisar os 37 genes e identificar quais são as funções biológicas e conseqüentemente selecionar aqueles que são relacionados ao desenvolvimento muscular. Com base neste critério, três genes foram escolhidos (*DTNA*,

RBICCI e *C-MOS*), sendo que em todos eles foram localizados os SNPs não sinônimos (Tabela 11).

Tabela 11 – SNPs não sinônimos identificados nos genes relacionados ao desenvolvimento muscular e suas respectivas posições presentes em animais da linhagem de postura e de corte

Gene	REFSNP (NCBI) ^a	Posição (pb) ^b	Linhagem de Postura	Linhagem de Corte
<i>DTNA</i>	c.206A>G	107.162.693	-	TT5921*, TT6037*, TT6232*, TT6270*
	c.995A>G	107.181.256	-	TT5461, TT5586*, TT5921*
	rs317583066	107.202.913	-	TT5586*, TT6037*, TT6232*, TT6270*
	rs314738751	107.204.378	CC1, CC5, CC37, CC88, CC241, CC886*	TT5649, TT5661
<i>RB1CC1</i>	rs314633726 ^c	109.659.296	CC37	-
	c.3318C>G	109.659.438	CC37, CC570*	-
	rs312375447	109.659.713	CC1, CC5, CC88, CC241, CC332, CC570, CC886	TT5461, TT5586*, TT5649, TT5661*, TT5921*, TT6270*
<i>C-MOS</i>	rs80770298	110.921.092	CC1, CC5*, CC37*, CC88, CC332*, CC570*, CC886	TT5461, TT5586, TT5642, TT5649, TT5921, TT6232, TT6270
	c.904G>A	110.921.122	-	TT5661*, TT6037

^aPara os SNPs ainda não depositados no banco de dados do NCBI (sem RefSNP) foi gerada a nomenclatura padrão com sua posição no cDNA e sua respectiva troca de bases.

^bPosição dos SNPs em relação ao genoma referência (Gallus_gallus-4.0, NCBI). ^cSNP deletério.*Animais heterozigotos para o SNP

O gene *DTNA* (*Alfa-distrobrevina*) é um membro da família de proteínas Distrofina e é um componente importante do complexo de proteína Distrofina (DPC). Este complexo é fundamental para a ligação da matriz celular com o citoesqueleto (KAPSA et al., 2003), o que gera uma estabilidade na fibra muscular e no sarcolema.

O gene *DTNA* está envolvido na manutenção da junção neuromuscular e na homeostase do músculo, pois se liga ao C terminal da Distrofina e da Sintrofina, unindo diretamente a matriz extracelular ao citoesqueleto do músculo (KAPSA et al., 2003). Segundo Grady et al. (1999), uma mutação neste gene em mamíferos poderia gerar uma deficiência da proteína, levando a distrofia muscular. Porém, alguns estudos vêm sendo realizados visando utilizar a *DTNA* como tratamento terapêutico para pacientes com distrofia muscular. Pois, segundo Gieseler, Bessou e Segalat (2002), doses extras de *DTNA* poderia compensar parcialmente a ausência de Distrofina.

Na Tabela 11 nota-se que os quatro SNPs não sinônimos identificados no gene *DTNA* são tolerados e dentre eles, três ocorreram somente em animais da linhagem de corte, que tem maior deposição muscular. Sendo que duas destas mutações (c.206A>G e rs317583066) são heterozigotas para os quatro animais. Estas três mutações exclusivas nos animais de corte devem ser melhor investigadas em futuros estudos de associação, sendo variantes candidatas para estudos de deposição muscular em galinhas.

O gene *RBICCI* (*RBI-inducible coiled-coil 1*) está envolvido diretamente no desenvolvimento muscular e durante a ossificação do embrião humano a sua expressão é baixa, porém sua expressão vai aumentando conjuntamente com o aumento da calcificação (CHANO et al., 2002). De acordo com Nishimura et al. (2011), mais estudos sobre o gene *RBICCI* se faz necessário para compreender a sua ação no desenvolvimento da cartilagem e do músculo esquelético. E desta forma, entender como são gerados indivíduos de diferentes taxas de crescimento e peso.

No gene *RBICCI* foram detectados três SNPs não sinônimos, sendo que apenas um na posição 109.659.713 pb foi detectado em ambas as linhagens, sendo heterozigoto para alguns dos animais da linhagem de corte. Os outros dois SNPs estão presentes apenas na linhagem de postura, porém um deles (rs314633726) é deletério. Desta forma, é um gene candidato interessante para estudo, pois poderia estar atuando de forma diferenciada nestas duas linhagens, já que é responsável por gerar indivíduos de diferentes pesos. Além disto, devido a relação que a expressão deste gene tem com o período de ossificação, seria interessante seu estudo, já que o melhoramento genético avícola proporcionou um rápido crescimento nas aves o que afetou a composição mineral dos ossos (BERNARDI, 2011).

O gene *C-MOS* gera transcritos altamente expressos em tecidos reprodutivos (testículos e ovários) e também em embriões que codifica a Serina/Treonina quinase, tem como função a regulação da maturação meiótica da metáfase II (YEW et al., 1993). Durante a fase de desenvolvimento em ratos, ocorre o acúmulo do transcrito do C-Mos no músculo esquelético atuando como uma proteína quinase, sugerindo ter um papel importante no desenvolvimento do músculo esquelético (LEIBOVITCH et al., 1991). Sendo assim, o *C-MOS* tem sua função relacionada com o desenvolvimento muscular, por isso, os SNPs localizados neste gene devem ser melhores estudados, uma vez que poderiam estar associados ao desenvolvimento de músculo de peito.

Foram identificados dois SNPs não sinônimos no gene *C-MOS*, um na posição 110.921.092 pb (c.934C>G) e um outro na posição 110.921.122 pb (c.904G>A) (Tabela 11), sendo que o primeiro está presente em 16 animais, sendo oito da linhagem de postura (quatro heterozigotos) e oito na de corte. O segundo apenas foi observado em dois animais da linhagem de corte, sendo apenas um heterozigoto.

Porém para confirmar as hipóteses levantadas, seria interessante que futuros estudos de associação das respectivas mutações identificadas neste estudo fossem realizados nestes genes candidatos. Principalmente, devido à escassez de estudos dos genes citados no trabalho em frangos.

6 CONCLUSÃO

Na região de QTL estudada no GGA2 foi possível identificar diversos SNPs e INDELS de boa qualidade por meio de ferramentas de bioinformática. Dentre estes, foram identificados SNPs e INDELS em regiões de éxons, que alteram ou não a proteína, além de mutações em regiões UTR e de *splicing*. Por meio de diferentes análises foi obtida uma lista de SNPs e INDELS candidatos com possíveis efeitos funcionais (deletérios) e SNPs presentes nos genes *DTNA*, *RBICCI* e *C-MOS*, que estão relacionados ao desenvolvimento muscular. Futura validação destas mutações poderá ser realizada em populações comerciais ou por meio de sequenciamento do tipo Sanger, além de estudos de associação.

REFERÊNCIAS

- ADAM-KLAGES, S.; ADAM, D.; WIEGMANN, K.; STRUVE, S.; KOLANUS, W.; SCHNEIDER-MERGENER, J.; KRONKE, M. FAN, a novel WD-repeat protein, couples the p55 TNF-receptor to neutral sphingomyelinase. **Cell**, Cambridge, v. 86, p. 937-947, 1996.
- ALBERS, C.A.; LUNTER, G.; MACARTHUR, D.G.; MCVEAN, G.; OUWEHAND, W.H.; DURBIN, R. Dindel: accurate indel calls from short-read data. **Genome research**, New York, v. 21, n. 6, p. 961–973, 2011.
- AMBO, M. **Mapeamento de locos de características quantitativas (QTLs) associados a desempenho nos cromossomos 19, 23, 24, 26, 27, 28 de *Gallus gallus***. 2007. 83 p. Dissertação (Mestre em Ciência Animal e Pastagens) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, 2007.
- AMBO, M.; CAMPOS, R.L.R.; MOURA, A.S.A.M.T.; BOSCHIERO, C.; ROSÁRIO, M.F.; LEDUR, M.C.; NONES, K.; COUTINHO, L.L. Genetic linkage maps of chicken chromosomes 6, 7, 8, 11 and 13 from a Brazilian resource population. **Scientia Agricola**, Piracicaba, v. 65, p. 447-452, 2008.
- ANDERSSON, L. Genetic dissection of phenotypic diversity in farm animals. **Nature**, London, v. 2, p. 130-138, 2001.
- ANIMALQTLDB. **Animal Quantitative Trait Loci database**. Disponível em: <<http://www.animalgenome.org/cgi-bin/QTLdb/index>>. Acesso em: 04 nov. 2013].
- ATZMON, G.; SHULA B.; MARC F.; AVIGDOR C.; URI L.; JOSSI H. QTLs detected in a multigenerational resource chicken population. **The Journal of Heredity**, Washington, v. 99, n. 5, p. 528–538, 2008.
- BARON, E.E; MOURA, A.S.A.M.T.; LEDUR, M.C.; PINTO, L.F.B.; BOSCHIERA, C.; RUY, D.C.; NONES, K.; ZANELLA, E.L.; ROSARIO, M.F.; BURT, D.W.; COUTINHO, L.L.. QTL for percentage of carcass and carcass parts in a broiler x layer cross. **Animal Genetics**, Oxford, v. 42, p. 117-124, 2010.
- BENTLEY, D.R.; BALASUBRAMANIAN, S.; SWERDLOW, H.P.; SMITH, G.P.; MILTON, J.; BROWN, C.G.; HALL, K.P.; EVERS, D.J.; BARNES, C.L.; HELEN, R.; BOUTELL, J.M.; BRYANT, J.; CARTER, R.J.; CHEETHAM, R.K.; COX, A.J.; ELLIS, D.J.; FLATBUSH, M.R.; GORMLEY, N.A.; SEAN, J.; IRVING, L.J.; KARBELASHVILI, M.S.; KIRK, S.M.; LI, H.; MAISINGER, K.S.; MURRAY, L.J.; OBRADOVIC, B.; OST, T.; MICHAEL, L.; PRATT, M.R.; RASOLONJATOVO, I.M.J.; REED, M.T.; RIGATTI, R.; RODIGHIERO, C.; ROSS, M.T.; SABOT, A.; SANKAR, S.V.; SCHROTH, G.P.; SMITH, M.E.; SMITH, V.P.; SPIRIDOU, A.; PETA, E.; TZONEV, S.S.; VERMAAS, E.H.; WALTER, K.; WU, X.; ZHANG, L.; BANERJEE, S.; BARBOUR, S.G.; BAYBAYAN, P.A.; BENOIT, V.A.; BRIDGHAM, A.; BROWN, R.C.; BROWN, A.A.; BUERMANN, D.H.; BUNDU, A.A.; COOLEY, R.N.; CRAKE, N.R.; DADA, O.O.; KONSTANTINOS, D. Accurate whole human genome sequencing using reversible terminator chemistry. **Nature**, London, v. 456, n. 7218, p. 53-59, 2008.

REFERÊNCIAS

- ADAM-KLAGES, S.; ADAM, D.; WIEGMANN, K.; STRUVE, S.; KOLANUS, W.; SCHNEIDER-MERGENER, J.; KRONKE, M. FAN, a novel WD-repeat protein, couples the p55 TNF-receptor to neutral sphingomyelinase. **Cell**, Cambridge, v. 86, p. 937-947, 1996.
- ALBERS, C.A.; LUNTER, G.; MACARTHUR, D.G.; MCVEAN, G.; OUWEHAND, W.H.; DURBIN, R. Dindel: accurate indel calls from short-read data. **Genome research**, New York, v. 21, n. 6, p. 961–973, 2011.
- AMBO, M. **Mapeamento de locos de características quantitativas (QTLs) associados a desempenho nos cromossomos 19, 23, 24, 26, 27, 28 de *Gallus gallus***. 2007. 83 p. Dissertação (Mestre em Ciência Animal e Pastagens) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, 2007.
- AMBO, M.; CAMPOS, R.L.R.; MOURA, A.S.A.M.T.; BOSCHIERO, C.; ROSÁRIO, M.F.; LEDUR, M.C.; NONES, K.; COUTINHO, L.L. Genetic linkage maps of chicken chromosomes 6, 7, 8, 11 and 13 from a Brazilian resource population. **Scientia Agricola**, Piracicaba, v. 65, p. 447-452, 2008.
- ANDERSSON, L. Genetic dissection of phenotypic diversity in farm animals. **Nature**, London, v. 2, p. 130-138, 2001.
- ANIMALQTLDB. **Animal Quantitative Trait Loci database**. Disponível em: <<http://www.animalgenome.org/cgi-bin/QTLdb/index>>. Acesso em: 04 nov. 2013].
- ATZMON, G.; SHULA B.; MARC F.; AVIGDOR C.; URI L.; JOSSI H. QTLs detected in a multigenerational resource chicken population. **The Journal of Heredity**, Washington, v. 99, n. 5, p. 528–538, 2008.
- BARON, E.E; MOURA, A.S.A.M.T.; LEDUR, M.C.; PINTO, L.F.B.; BOSCHIERA, C.; RUY, D.C.; NONES, K.; ZANELLA, E.L.; ROSARIO, M.F.; BURT, D.W.; COUTINHO, L.L.. QTL for percentage of carcass and carcass parts in a broiler x layer cross. **Animal Genetics**, Oxford, v. 42, p. 117-124, 2010.
- BENTLEY, D.R.; BALASUBRAMANIAN, S.; SWERDLOW, H.P.; SMITH, G.P.; MILTON, J.; BROWN, C.G.; HALL, K.P.; EVERS, D.J.; BARNES, C.L.; HELEN, R.; BOUTELL, J.M.; BRYANT, J.; CARTER, R.J.; CHEETHAM, R.K.; COX, A.J.; ELLIS, D.J.; FLATBUSH, M.R.; GORMLEY, N.A.; SEAN, J.; IRVING, L.J.; KARBELASHVILI, M.S.; KIRK, S.M.; LI, H.; MAISINGER, K.S.; MURRAY, L.J.; OBRADOVIC, B.; OST, T.; MICHAEL, L.; PRATT, M.R.; RASOLONJATOVO, I.M.J.; REED, M.T.; RIGATTI, R.; RODIGHIERO, C.; ROSS, M.T.; SABOT, A.; SANKAR, S.V.; SCHROTH, G.P.; SMITH, M.E.; SMITH, V.P.; SPIRIDOU, A.; PETA, E.; TZONEV, S.S.; VERMAAS, E.H.; WALTER, K.; WU, X.; ZHANG, L.; BANERJEE, S.; BARBOUR, S.G.; BAYBAYAN, P.A.; BENOIT, V.A.; BRIDGHAM, A.; BROWN, R.C.; BROWN, A.A.; BUERMANN, D.H.; BUNDU, A.A.; COOLEY, R.N.; CRAKE, N.R.; DADA, O.O.; KONSTANTINOS, D. Accurate whole human genome sequencing using reversible terminator chemistry. **Nature**, London, v. 456, n. 7218, p. 53-59, 2008.

CAMPOS, R.L.R.; NONES, K.; LEDUR, M.C.; MOURA, A.S.A.M.T.; PINTO, L.F.B.; AMBO, M.; BOSCHIERO, C. Quantitative trait loci associated with fatness in a broiler-layer cross. **Animal genetics**, Malden, v. 40, n. 5, p. 729–736, 2009.

CARTWRIGHT, R.A. Problems and solutions for estimating indel rates and length distributions. **Molecular Biology and Evolution**, Oxford, v. 26, p. 473–480, 2009.

CARVALHO, M.C.C.G.; SILVA, D.C.G. da. Next generation DNA sequencing and its applications in plant genomics, **Ciência Rural**, Santa Maria, v. 40, n. 3, p. 735–744, 2010.

CHANO, T.; SAEKI, Y.; SERRA, M.; MATSUMOTO, K.; OKABE, H. Preferential expression of *RBI*-inducible coiled-coil 1 in terminal differentiated musculoskeletal cells. **American Journal of Pathology**, Bethesda, v. 161, n. 2, p. 359-364, 2002.

CHEN, C.; TREES-JUEN, C.; BEN-YANG, L.; FENG-CHI, C. Scanning for the signatures of positive selection for human-specific insertions and deletions. **Genome Biology and Evolution**, Oxford, v. 1, p. 415–419, 2009.

CHEN, K.; MCLELLAN, M.D.; DING, L.; WENDL, M.C.; KASAI, Y.; WILSON, R.K.; MARDIS, E.R. PolyScan : an automatic indel and SNP detection approach to the analysis of human resequencing data. **Genome Research**, New York, v. 17, n. 314, p. 659-666, 2007.

COUTINHO, L.; ROSARIO, M.F. Biotecnologia animal. **Estudos Avançados**, São Paulo, v. 24, n. 70, p. 123-147, 2010.

COUTINHO, L.L.; JORGE, E.C.; ROSÁRIO, M.F.; REGITANO, L.C.A. A genômica na bovinocultura de corte. In: PIRES, A.V. (Ed.) **Bovinicultura de corte**. Piracicaba: FEALQ, 2010. v. 2, p. 813-826.

CRAWFORD, R.D. Origin and history of poultry species. In: _____. **Poultry breeding and genetics**. Amsterdam: Elsevier Science, 1990. chap. 1, p 10-82.

CROOIJMANS, R.P.; GROEN, A.F.; VAN KAMPEN, A.J.; VAN DER BEEK, S.; VAN DER POEL, J.J.; GROENEN, M.A. Microsatellite polymorphism in commercial broiler and layer lines estimated using pooled blood samples. **International Journal of Poultry Science**, Faisalabad, n. 75, p. 904-909, 1996.

DAVEY, M.G.; TICKLE, C. The chicken as a model for embryonic development. **Cytogenetic and Genome Research**, Basel, v. 117, n. 1/4, p. 231-239, 2007.

DAVID, M.; DZAMBA, M.; LISTER, D.; ILIE, L.; BRUDNO, M. SHRiMP2: sensitive yet practical short read mapping. **Bioinformatics**, Oxford, v. 27, p. 1011–1012, 2011.

DEKKERS, J.C.M. Breeding in the 21st century: application of molecular technology. **Proceeding Association for the Advancement of Animal Breeding and Genetics**, Western Australia, v. 13, p. 1-16, 1999.

DEQUÉANT, M.L.; POURQUIÉ, O. Chicken genome: new tools and concepts. **Developmental Dynamics**, New York, v. 232, n. 4, p. 883-886, 2005.

DODGSON, J.B.; DELANY, M.E.; CHENG, H.H. Poultry genome sequences: progress and outstanding challenges. **Cytogenetic and genome research**, Basel, v. 1, n. 134, p. 19–26, 2011.

DORSHORST, B.J.; SIEGEL, P B.; ASHWELL, C. M. Genomic regions associated with antibody response to sheep red blood cells in the chicken. **Animal Genetics**, Oxford, v. 42, p. 300-308, 2011.

DRANSFIELD, E.; SOSNICKI, A.A. Relationship between muscle growth and poultry meat quality. **Poultry Science**, London, v. 78, p. 743-746, 1999.

ECK, S.H.; BENET-PAGÈS, A.; FLISIKOWSKI, K.; MEITINGER, T.; FRIES, R.; STROM, T.M. Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. **Genome Biology**, London, v. 8, n. 10, p. 1-8, 2009.

ELLEGREN, H. The avian genome uncovered. **Trends in ecology & evolution**, Amsterdam, v. 20, n. 4, p. 180-186, 2005.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th ed. London: Prentice Hall, 1996. 464 p.

FAN, W.L.; NG, C.S.; CHEN, C.F.; LU, M.Y.J.; CHEN, Y.H.; LIU, C.J.; WU, S.M.; ET AL. Genome-wide patterns of genetic variation in two domestic chickens. **Genome Biology and Evolution**, Oxford, v. 7, n. 5, p. 1376–1392, 2013.

FELÍCIO, A.M.; BOSCHIERO, C.; BALIEIRO, J.C.C. LEDUR, M.C.; FERRAZ, J.B.S.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Polymorphisms in *FGFBP1* and *FGFBP2* genes associated with carcass and meat quality traits in chickens. **Genetics and Molecular Research**, Ribeirão Preto, v. 12, n. 1, p. 208-222, 2013a.

FELÍCIO, A.M.; BOSCHIERO, C.; BALIEIRO, J.C.C.; LEDUR, M.C.; FERRAZ, J.B.S.; MICHELAN FILHO, T.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Identification and association of polymorphisms in *CAPN1* and *CAPN3* candidate genes related to performance and meat quality traits in chickens. **Genetics and Molecular Research**, Ribeirão Preto, v. 12, n. 1, p. 472-482, 2013b.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: EMBRAPA, 1998. 220 p.

FONTANESI, L.; TAZZOLI, M.; SCOTTI, E.; RUSSO, V. Analysis of candidate genes for meat production traits in domestic rabbit breeds. In: WORLD RABBIT CONGRESS, 9., 2008, Verona. **Proceedings...** Verona: Genetics, 2008. p. 79-84.

FONTANILLAS, P.; HARTL, D.L.; REUTER, M. Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. **PLoS Genetics**, San Francisco, v. 3, p. e210, 2007.

FOX, S.; FILICHKIN, S.; MOCKLER, T.C. Applications of ultra-high-throughput sequencing. **Methods in Molecular Biology**, Cambridge, v. 553, p. 79-108, 2009.

FUJII, J.; OTSU, K.; ZORZATO, F.; DE LEON, S.; KHANNA, V.K.; WEILER, J.E.; O'BRIEN, P.J.; MACLENNAN, D.H. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. **Science**, Washington, v. 253, p. 448-451, 1991.

FUMIHITO, A.; MIYAKE, T.; SUMI, S.I.; TAKADA, M.; OHNO, S.; KONDO, N. One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. **The Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 91, n. 26, p. 12505-12509, 1994.

GAYA, L.D.G.; MOURÃO, G.B.; FERRAZ, J.B.S. Estudo genético da deposição de gordura abdominal e de características de desempenho, carcaça e composição corporal em linhagem macho de frangos de corte. **Ciência Rural**, Santa Maria, v. 36, n. 2, p. 709-716, 2006.

GIESELER, K.; BESSOU, C.; SEGALAT, L. Dystrobrevin- and dystrophinlike mutants display similar phenotypes in the nematode *Caenorhabditis elegans*. **Neurogenetics**, London, v. 2, p. 87-90, 1999.

GRADY, R.M.; GRANGE, R.W.; LAU, K.S.; MAIMONE, M.M.; NICHOL, M.C.; STULL, J.T.; SANES, J.R. Role for α -dystrobrevin in the pathogenesis of dystrophin-dependent muscular dystrophies. **Nature Cell Biology**, London, v. 1, p. 215-220, 1999.

GREGORY, T.R. Synergy between sequence and size in large- scale genomics. **Nature Review Genetics**, London, v. 6, p. 699-708, 2005.

GRIFFITHS, A.J.F.; MILLER J.H.; SUZUKI, D.T.; LEWONTIN, R.C.; GERBALT, W.M. **Introdução à genética**. Rio de Janeiro: Guanabara Koogan, 1998. 856 p.

GRISART, B.; COPPIETERS, W.; FARNIR, F.; KARIM, L.; FORD, C.; BERZI, P.; CAMBISANO, N.; MNI, M.; REID, S.; SIMON, P.; SPELMAN, R.; GEORGES, M.; SNELL, R. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. **Genome Research**, New York, v. 12, p. 222-231, 2002.

GROBET, L. A deletion in the bovine myostatin gene causes the double- muscled phenotype in cattle. **Nature Genetics**, New York, v. 17, n. 1, p. 71-74, 1997.

GROENEN, M.A.; WAHLBERG, P.; FOGLIO, M. A highdensity SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. **Genome Research**, New York, n. 19, p. 510-519, 2009.

GROENEN, M.A.M.; CHENG, H.H.; BUMSTEAD, N.; BENKEL, B.F.; BRILES, W.E.; BURKE, T.; BURT, D.W.; CRITTENDEN, L.B.; DODGSON, J.; HILLEL, J.; LAMONT, S.; PONCE DE LEON, F.A.; SOLLER, M.; TAKAHASHI, H.; VIGNAL, A. A consensus linkage map of the chicken genome. **Genome Research**, Woodbury, v. 10, n. 1, p. 137-147, 2000.

HALEY, C.S.; KNOTT, S.A.; ELSE, J.M. Mapping quantitative trait loci in crosses between outbred lines using least squares, **Genetics**, Bethesda, v. 136, p. 1195-1207, 1994.

HAVENSTEIN, G.B.; FERKET, P.R.; QURESHI, M.A. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. **Poultry science**, London, v. 82, n. 10, p. 1500-1508, 2003.

HAVENSTEIN, G.B.; FERKET, P.R.; SCHEIDELER, S.E.; LARSON, B.T. Growth, livability, and feed conversion of 1957 vs. 1991 broilers when fed “ typical” 1957 and 1991 broiler diets. **Poultry Science**, Salvoy, v. 73, p. 1785-1794, 1994.

HILLIER, L.W.; MILLER, W.; BIRNEY, E.; WARREN, W.; HARDISON, R.C.; PONTING, C.P.; BORK, P.; BURT, D.W.; GROENEN, M.A.; DELANY, M.E.; DODGSON, J.B.; CHINWALLA, A.T.; CLIFTEN, P.F.; CLIFTON, S.W.; DELEHAUNTY, K.D.; FRONICK, C.; FULTON, R.S.; GRAVES, T.A.; KREMITZKI, C.; LAYMAN, D.; MAGRINI, .V; MCPHERSON, J.D.; MINER, T.L.; MINX, P.; NASH, W.E.; NHAN, M.N.; NELSON, J.O.; ODDY, L.G.; POHL, C.S.; RANDALL-MAHER, J.; SMITH, S.M.; WALLIS, J.W.; YANG, S.P.; ROMANOV, M.N.; RONDELLI, C.M.; PATON, B.; SMITH, J.; MORRICE, D.; DANIELS, L.; TEMPEST, H.G.; ROBERTSON, L.; MASABANDA, J.S.; GRIFFIN, D.K.; VIGNAL, A.; FILLON, V.; JACOBBSOON, L.; KERJE, S.; ANDERSSON, L.; CROOIJMANS, R.P.; AERTS, J.; VAN DER POEL, J.J.; ELLEGREN, H.; CALDWELL, R.B.; HUBBARD, S.J.; GRAFHAM, D.V.; KIERZEK, A.M.; MCLAREN, S.R.; OVERTON, I.M.; ARAKAWA, H.; BEATTIE, K.J.; BEZZUBOV, Y.; BOARDMAN, P.E.; BONFIELD, J.K.; CRONING, M.D.; DAVIES, R.M.; FRANCIS, M.D.; HUMPHRAY, S.J.; SCOTT, C.E.; TAYLOR, R.G.; TICKLE, C.; BROWN, W.R.; ROGERS, J.; BUERSTEDDE, J.M.; WILSON, S.A.; STUBBS, L.; OVCHARENKO, I.; GORDON, L.; LUCAS, S.; MILLER, M.M.; INOKO, H.; SHIINA, T.; KAUFMAN, J.; SALOMONSEN, J.; SKJOEDT, K.; WONG, G.K.; WANG, J.; LIU, B.; WANG, J.; YU, J.; YANG, H.; NEFEDOV, M.; KORIABINE, M.; DEJONG, P.J.; GOODSTADT, L.; WEBBER, C.; DICKENS, N.J.; LETUNIC, I.; SUYAMA, M.; TORRENTS, D.; VON MERING, C.; ZDOBNOV, E.M.; MAKOVA, K.; NEKRUTENKO, A.; ELNITSKI, L.; ESWARA, P.; KING, D.C.; YANG, S.; TYEKUCHEVA, S.; RADAKRISHNAN, A.; HARRIS, R.S.; CHIAROMONTE, F.; TAYLOR, J.; HE, J.; RIJNKELS, M.; GRIFFITHS- JONES, S.; URETA-VIDAL, A.; HOFFMAN, M.M.; SEVERIN, J.; SEARLE, S.M.; LAW, A.S.; SPEED, D.; WADDINGTON, D.; CHENG, Z.; TUZUN, E.; EICHLER, E.; BAO, Z.; FLICEK, P.; SHTEYNBERG, D.D.; BRENT, M.R.; BYE, J.M.; HUCKLE, E.J.; CHATTERJI, S.; DEWEY, C.; PACHTER, L.; KOURANOV, A.; MOURELATOS, Z.; HATZIGEORGIOU, A.G.; PATERSON, A.H.; IVARIE, R.; BRANDSTROM, M.; AXELSSON, E.; BACKSTROM, N.; BERLIN, S.; WEBSTER, M.T.; POURQUIE, O.; REYMOND, A.; UCLA, C.; ANTONARAKIS, S.E.; LONG, M.; EMERSON, J.J.; BETRAN, E.; DUPANLOUP, I.; KAESSMANN, H.; HINRICHS, A.S.; BEJERANO, G.; FUREY, T.S.; HARTE, R.A.; RANEY, B.; SIEPEL, A.; KENT, W.J.; HAUSSLER, D.; EYRAS, E.; CASTELO, R.; ABRIL, J.F.; CASTELLANO, S.; CAMARA, F.; PARRA, G.; GUIGO, R.; BOURQUE, G.; TESLER, G.; PEVZNER, P.A.; SMIT, A.; FULTON, L.A.; MARDIS, E.R.; WILSON, R.K. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, London, v. 432, n. 7018, p. 695-716, 2004.

HIRSCHHORN, J.N.; DALY, M.J. Genome-wide association studies for common diseases and complex traits. **Nature Review Genetics**, London, v. 6, p. 95–108, 2005.

HUANG, D.W.; SHERMAN, B.T.; LEMPICKI, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Research**, London, v. 37, n. 1, p. 1-13, 2009a.

_____. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature Protocols**, London, v. 4, n. 1, p. 44-57, 2009b.

IKEOBI, C.O.N.; WOOLLIAMS, J.A.; MORRICE, D.R.; LAW, A.; WINDSORA, D.; BURT, D.W.; HOCKING, P.M. Quantitative trait loci for meat yield and muscle distribution in a broiler layer cross. **Livestock Production Science**, Amsterdam, v. 87, p. 143–151, 2004.

JENNEN, D.G.J.; VEREIJKEN, A.L.J.; BOVENHUIS, H.; CROOIJMANS, R.P.M.; VEENENDAAL, A.; VAN DER POEL, J.J.; GROENEN, M.A.M. Detection and localization of quantitative trait loci affecting fatness in broilers. **Poultry Science**, Savoy, v. 3, n. 83, p. 295–301, 2004.

KAPSA, R.; KORNBERG, A. J.; BYRNE, E. Novel therapies for Duchenne muscular dystrophy. **Lancet Neurology**, London, v. 2, n. 5, p. 299-310, 2003.

KATO, K. Review Article Impact of the next generation DNA sequencers. **International Journal Clinical and Experimental Medicine**, Madison, v. 2, p. 193–202, 2009.

KESSLER, A.M.; SNIZEK JR.; P.N., BRUGALLI, I. **Manipulação da quantidade de gordura na carcaça de frangos**. In: CONFERÊNCIA APINCO DE CIÊNCIA E TECNOLOGIA AVÍCOLAS, 2000, Campinas. **Anais...** Campinas: FACTA, 2000. p. 107-133.

KNOTT, S.A.; ELSEN, J.M.; HALEY, C.S. Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. **Theoretical and Applied Genetics**, Berlin, v. 93, p. 71-80, 1996.

KRANIS, A.; GHEYAS, A.A; BOSCHIERO, C.; TURNER, F.; YU, L.; SMITH, S.; TALBOT, R. Development of a high density 600K SNP genotyping array for chicken. **BMC Genomics**, London, v. 1, n. 14, p. 59, 2013.

KUMAR, S.; BANKS, T.W.; CLOUTIER, S. SNP discovery through next-generation sequencing and its applications. **International Journal of Plant Genomics**, Ohio, v. 2012, p. 1-15, 2012.

KWOK, P.Y.; GU, Z. Single nucleotide polymorphism libraries: why and how are we building them? **Molecular Medicine Today**, London, v. 5, n. 12, p. 538-543, 1999.

LANDER, E.S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Bethesda, v. 121, p. 185–199, 1989.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, London, v. 9, p. 357-359, 2012.

LANGMEAD, B.; TRAPNELL, C.; POP, M.; SALZBERG, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, Oxford, v. 10, R25, 2009.

LEDUR, M.C. Genoma do frango: mapeamento de QTL. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 38, 2001, Piracicaba. **Anais...** Piracicaba: SBZ, 2001. P. 1-9.

LEDUR, M.C.; PEIXOTO, J.O.; SCHMIDT, G.S. Novos rumos da genética no desenvolvimento avícola. **Avicultura Industrial**, São Paulo, v. 102, n. 1, p. 12-16, 2011.

LEDUR, M.C.; NONES, K.; ALVES, J.H.; BERTANI, G.R. A importância do conhecimento do genoma para a produção de aves. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 41, 2004, Campo Grande. **Anais...** Campo Grande: SBZ, 2004. p. 379-389.

LEES-MILLER, S.P.; GODBOUT, R.; CHAN, D.W.; WEINFELD, M.; DAY, R.S.; BARRON, G.M.; ALLALUNIS-TURNER, J. Absence of p350 subunit of DNA-activated protein kinase from a radiosensitive human cell line. **Science**, Washington, v. 267, p. 1183-1185, 1995.

LEIBOVITCH, S.A.; GUILLIER, M.; LENORMAND, J.L.; LEIBOVITCH, M.P. Accumulation of the c-mos protein is correlated with postnatal development of skeletal muscle. **Oncogene**, Basingstoke, v. 6, p. 1617-1622, 1991.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, Oxford, v. 25, n. 14, p. 1754-1760, 2010.

LI, H.; HOMER, N. A survey of sequence alignment algorithms for next-generation sequencing. **Briefings in Bioinformatics**, London, v. 11, n. 5, p. 473-483, 2010.

LI, H.; JUE, R.; RICHARD, D. Mapping short DNA sequencing reads and calling variants using mapping quality scores. **Genome Research**, New York, v. 18, n. 11, p. 1851-1858, 2008.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNEL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. 1000 genome project data processing subgroup. The sequence alignment/map format and SAM tools. **Bioinformatics**, Oxford, v. 25, n. 16, p. 2078-2079, 2009.

LUNTER, G.; GOODSON, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. **Genome Research**, New York, v. 21, n. 6, p. 936-939, 2011.

MANIATIS, T.; FRITSCH, E.F.; SAMBROOK, J. **Molecular cloning**. New York: Cold Spring Harbor Laboratory, 1982, 545 p.

MARGUERAT, S.; WILHELM, B.T.; BÄHLER, J. Next-generation sequencing: applications beyond genomes. **Biochemical Society Transactions**, London, v. 36, p. 1091-1096, 2008.

MATSUZAKA, Y.; OKAMOTO, K.; TSUJI, H.; MABUCHI, T.; OZAWA, A.; TAMIYA, G.; INOKO, H. Identification of the hRDH-E2 gene, a novel member of the SDR family, and its increased expression in psoriatic lesion. **Biochemical and Biophysical Research Communications**, Orlando, v. 297, n. 5, p. 1171-1180, 2002.

- MCLAREN, W.; PRITCHARD, B.; RIOS, D.; CHEN, Y.; FLICEK, P.; CUNNINGHAM, F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. **Bioinformatics**, Oxford, v. 16, n. 26, p. 2069–2070, 2010.
- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.
- MOU, C.; PITEL, F.; GOURICHON, D.; VIGNOLES, F.; TZIKA, A.; TATO, P.; YU, L.; BURT, D.W.; BED'HOM, B.; TIXIER-BOICHARD, M.; PAINTER, K.J.; HEADON, D.J. Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. **PLoS Biology**, San Francisco, v. 9, n. 3, p. e1001028, 2011.
- MOURA, A.S.A.M.T; BOSCHIERO, C.; CAMPOS, R.L.R.; AMBO, M.; NONES, K.; LEDUR, M.C. ROSARIO, M.F.; MELO, C.M.R.; BURT, D.W.; COUTINHO, L.L. Mapping QTL for performance and carcass traits in chicken chromosomes 6, 7, 8, 11 and 13. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Anais...** Belo Horizonte: Instituto Prociência, 2006. 1 CD-ROM.
- MULLANEY, J.M.; MILLS, R.E.; PITTARD, W.S.; DEVINE, S.E. Small insertions and deletions (INDELs) in human genomes. **Human Molecular Genetics**, Oxford, v. 19, n. R2, p. R131–R136, 2010.
- MULSANT, P.; LECERF, F.; FABRE, S.; SCHIBLER, L.; MONGET, P.; LANNELUC, I.; PISSELET, C.; RIQUET, J.; MONNIAUX, D.; CALLEBAUT, I.; CRIBIU, E.; THIMONIER, J.; TEYSSIER, J.; BODIN, L.; COGNIE, Y.; CHITOUR, N.; ELSEN, J-M. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. **The Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 98, p. 5104–5109, 2001.
- NADAF, J.; PITEL, F.; GILBERT, H.; DUCLOS, M.J.; VIGNOLES, F.; BEAUMONT, C.; VIGNAL, A.; PORTER, T.E.; LARRY, A.C.; AGGREY, S.E.; SIMON, J.; BIHAN-DUVAL, E.L. QTL for several metabolic traits map to loci controlling growth and body composition in an F2 intercross between high- and low-growth chicken lines. **Physiological Genomics**, Bethesda, v. 38, p. 241–249, 2009.
- NASSAR, M.K.; GORAGA, Z.S.; BROCKMANN, G.A. Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: II. Muscle weight and carcass composition. **Animal Genetics**, Oxford, v. 43, p. 739-745, 2012.
- NEUMAN, J.; ISAKOV, O.; SHOMRON, N. Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection. **Briefings in Bioinformatics**, London, v. 14, n. 1, p. 46-55, 2012.
- NG, P.C.; HENIKOFF, S. Predicting deleterious amino acid substitutions predicting deleterious amino acid substitutions. **Genome Research**, New York, v. 11, p. 863-874, 2001.
- _____. SIFT: predicting amino acid changes that affect protein function. **Nucleic Acids Research**, London, v. 31, n. 31, p. 3812–3814, 2003.

NIE, Q.; SUN, B.; ZHANG, D.; LUO, C.; ISHAG, N.A.; LEI, M.; YANG, G.; ZHANG, X. High diversity of the chicken growth hormone gene and effects on growth and carcass traits. **The Journal of heredity**, Washington, v. 96, n. 6, p. 698-703, 2005.

NISHIMURA, I.; CHANO, T.; KITA, H.; MATSUSUE, Y.; OKABE, H. RB1CC1 protein suppresses type II collagen synthesis in chondrocytes and causes dwarfism. **The Journal of Biology Chemistry**, New York, v. 286, p. 43925-43932, 2011.

NONES, K.; LEDUR, M.C.; RUY, D.C.; BARON, E.E.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Genetic linkage map of chicken chromosome 1 from a Brazilian resource population. **Scientia Agricola**, Piracicaba, v. 62, p. 12-17, 2005.

NONES, K.; LEDUR, M.C.; ZANELLA, E.L.; KLEIN, C.; PINTO, L.F.B.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Quantitative trait loci associated with chemical composition of the chicken carcass, **Animal Genetics**, New York, v. 43, p. 570–576, 2012.

NONES, K.; LEDUR, M.C.; RUY, D.C.; BARON, E.E.; MELO, C.M.R.; MOURA, A.S.A.M.T.; ZANELLA, E.L.; BURT, D.W.; COUTINHO, L.L. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. **Animal Genetics**, New York, v. 37, p. 95-100, 2006.

NORMAN, L.P.; JIANG, W.; HAN, X.; SAUNDERS, T.L.; BOND, J.S. Targeted disruption of the meprin beta gene in mice leads to underrepresentation of knockout mice and changes in renal gene expression profiles. **Molecular and Cellular Biology**, Washington, v. 23, p. 1221-1230, 2003.

PÉRTILLE, F. **Identificação de polimorfismos associados às características de desempenho e carcaça no cromossomo 4 da galinha (*Gallus gallus*)**. 2013. 103 p. Dissertação (Mestre em Ciência Animal e Pastagens) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo, Piracicaba, 2013.

PETROV, D.A.; SANGSTER, T.A.; JOHNSTON, J.S.; HARTL, D.L.; SHAW, K.L. Evidence for DNA loss as a determinant of genome size. **Science**, Washington, v. 287, p. 1060–1062, 2000.

POTTS, J.K.; ECHTERNKAMP, S.E.; SMITH, T.P.; REECY, J.M. Characterization of gene expression in double-muscled and normal- muscled bovine embryos. **Animal Genetics**, New York, v. 34, p. 438-444, 2003.

ROSÁRIO, M.F. do; LEDUR, M.C.; MOURA, A.S.A.M.T.; COUTINHO, L.L.; GARCIA, A.A.F. Genotypic characterization of microsatellite markers in broiler and layer selected chicken lines and their reciprocal F1s. **Scientia Agricola**, Piracicaba, v. 2, n. 66, p. 150–158, 2009.

ROSÁRIO, M.F.; MARGARIDO, G.R.; BOSCHIERO, C.; MOURA, A.S.; LEDUR, M.C.; COUTINHO, L.L.; GARCIA, A.A. Precision of distances and ordering of microsatellite markers in consensus linkage maps of chromosomes 1, 3 and 4 from two reciprocal chicken populations using bootstrap sampling. **Genetic Molecular Research**, Ribeirão Preto, v. 9, n. 3, p. 1357-1376, 2010.

RUBIN, C.J.; ZODY, M. C.; ERIKSSON, J.; MEADOWS, J. R. S.; SHERWOOD, E.; WEBSTER, M. T.; JIANG, L. Whole-genome resequencing reveals loci under selection during chicken domestication. **Nature**, London, v. 464, n. 7288, p. 587–591, 2010.

RUY, D.C.; MOURA, A.S.A.M.T.; NONES, K. BARON, E.E.; LEDUR, M.C.; CAMPOS, R.L.R.; AMBO, M.; MELO, C.M.R.; COUTINHO, L.L. Detection of QTL for performance, fatness and carcass traits on chicken chromosomes 3 and 5. In: QUANTITATIVE TRAIT LOCO MARKER ASSISTED SELECTION, 11., 2007, Toulouse. **Proceedings...** Toulouse: Institut National de la Recherche Agronomique, 2007, p 26-29.

SANGER, F.; COULSON, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **Journal of Molecular Biology**, London, v. 94, n. 3, p. 441-446, 1975.

SANGER, F.; NICKLEN, S.; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. **The Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 74, p. 5463–5467, 1977.

SCHMID, M.; NANDA, I.; HOEHN, H.; SCHARTL, M.; HAAF, T.; BUERSTEDDE, J.M.; ARAKAWA, H.; CALDWELL, R.B.; WEIGEND, S.; BURT, D.W.; SMITH, J.; GRIFFIN, D.K.; MASABANDA, J.S.; GROENEN, M.A.; CROOIJMANS, R.P.; VIGNAL, A.; FILLON, V.; MORISSON, M.; PITEL, F.; VIGNOLES, M.; GARRIGUES, A.; GELLIN, J.; RODIONOV, A.V.; GALKINA, S.A.; LUKINA, N.A.; BEN-ARI, G.; BLUM, S.; HILLEL, J.; TWITO, T.; LAVI, U.; DAVID, L.; FELDMAN, M.W.; DELANY, M.E.; CONLEY, C.A.; FOWLER, V.M.; HEDGES, S.B.; GODBOUT, R.; KATYAL, S.; SMITH, C.; HUDSON, Q.; SINCLAIR, A.; MIZUNO, S. Second report on chicken genes and chromosomes 2005. **Cytogenetic and Genome Research**, Basel, v. 109, n. 4, p. 415-479, 2005.

SCRIVENS, P.J.; NOUEIHED, B.; SHAHRZAD, N.; HUL, S.; BRUNET, S.; SACHER, M. C4orf41 and TTC-15 are mammalian TRAPP components with a role at an early stage in ER-to-Golgi trafficking. **Molecular and Cellular Biology**, Washington, v. 22, p. 2083-2093, 2011.

SHORT, T.H.; ROTHSCHILD, M.F.; SOUTHWOOD, O.I.; MCLAREN, D.G.; DE VRIES, A.; VAN, D.S.H.; ECKARDT, G.R.; TUGGLE, C.K.; HELM, J.; VASKE, D.A.; MILEHAM, A.J.; PLASTOW, G.S. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. **Journal of Animal Science**, Savoy, v. 75, n. 12, p. 3138-3142, 1997.

SILVA, F.E.D.J.; GUIDO, L.N.; GAZAFFI, R.; GARCIA, A.A.F.; LEDUR, M.C.; COUTINHO, L.L.; ROSÁRIO, M.F. do. Regiões genômicas associadas a características de desempenho e carcaça no cromossomo 5 de linhagens Brasileiras de galinha. **Pesquisa Agropecuária Brasileira**, Brasília, v. 46, n. 3, p. 229–238, 2011.

SIWEK, M.; CORNELISSEN, S.J.B.; NIEUWLAND, M.G.B.; BUITENHUIS, A.J.; BOVENHUIS, H.; CROOIJMANS, R.P.M.A.; GROENEN, M.A.M.; DE VRIES- REILINGH, PARMENTIER, H.K.; VAN DER POEL, J.J. Detection of QTL for immune response to sheep red blood cells in laying hens. **Animal Genetics**, Oxford, v. 34, p. 422-428, 2003.

SMITH, L.M.; SANDERS, J.Z.; KAISER, R.J.; HUGHES, P.; DODD, C.; CONNELL, C.R.; HEINER, C.; KENT, S.B.H.; HOOD, L.E. Fluorescence detection in automated DNA sequence analysis. **Nature**, London, v. 321, n. 6071, p. 674–679, 1986.

STERN, C.D. The chick; a great model system becomes even greater. **Developmental Cell**, Cambridge, v. 8, n. 1, p. 9-17, 2005.

TANKSLEY, S.D. Mapping polygenes. **Annual Review of Genetics**, Palo Alto, v. 27, p. 205-233, 1993.

TERCIC, D.; HOLCMAN, A.; DOVC, P. Identification of chromosomal regions associated with growth and carcass traits in an F(3) full sib intercross line originating from a cross of chicken lines divergently selected on body weight. **Animal Genetics**, Oxford, v. 40, p. 743-748, 2009.

TURNER, D.J.; THOMAS, M.K.; IAN, S.; DAVID, J.A. Next-generation sequencing of vertebrate experimental organisms. **Mammalian Genome: Official Journal of the International Mammalian Genome Society**, New York, v. 20, n. 6, p. 327–338, 2009.

UNIÃO BRASILEIRA DE AVICULTURA. **Relatório anual 2013**. Disponível em: <<http://www.ubabef.com.br/publicacoes>>. Acesso em: 04 nov. 2013.

VALCESCHINI, E. Poultry meat trends and consumer attitudes. In: EUROPEAN POULTRY CONFERENCE, 12., 2006, Verona. **Proceedings...** Verona: Verona Fiere, 1 CD-ROM.

WANG, G.K.; LI, M.; HAKONARSON, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. **Nucleic Acids Research**, London, v. 38, n.16, p. 164, 2010.

WILSON, T.; WU, XI-YANG; JUENGEL, J.L.; ROSS, I.K.; LUMSDEN, J.M.; LORD, E.A.; DODDS, K.G.; WALLING, G.A.; MCEWAN, J.C.; O'CONNELL, A.R.; MCNATTY, K.P.; MONTGOMERY, G.W. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. **Biology of Reproduction**, Champaign, v. 64, p. 1225–1235, 2001.

WINTER, A.; KRAMER, W.; WERNER, F. A.; KOLLERS, S.; KATA, S.; DURSTEWITZ, G.; BUITKAMP, J.; WOMACK, J. E.; THALLER, G.; FRIES, R. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. **The Proceedings of the National Academy of Sciences of the United States of America**, Wahinston, v. 99, p. 9300–9305, 2002.

WOLPERT, L. Much more from the chicken's egg than breakfast: a wonderful model system. **Mechanisms of Development**, Chicago, v. 121, n. 9, p. 1015-1017, 2004.

WONG, G.K.S.; LIU, B.; WANG, J.; ZHANG, J.; YANG, X.; ZHANG, Z.; MENG, Q.; ZHOU, J.; LI, D.; ZHANG, J.; NI, P.; LI, S.; RAN, L.; LI, H.; ZHANG, J.; LI R.; LI, S.; ZHENG, H.; LIN, W.; LI, G.; WANG, X.; ZHAO, W.; LI, Z.; YE, C.; DAI, M.; RUAN, J.; ZHOU, Y.; LI, Y.; HE, Y.; ZHANG, Y.; WANG, J.; HUANG, X.; TONG, W.; CHEN, J.;

YE, J.; CHEN, C.; WEI, N.; LI, G.; DONG, L.; LAN, F.; SUN, Y.; ZHANG, Z.; YANG, Z.; YU, Y.; HUANG, Y.; HE, D.; XI, Y.; WEI, D.; QI, Q.; LI, W.; SHI, J.; WANG, M.; XIE, F.; WANG, J.; ZHANG, X.; WANG, P.; ZHAO, Y.; LI, N.; YANG, N.; DONG, W.; HU, S.; ZENG, C.; ZHENG, W.; HAO, B. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. **Nature**, London, v. 432, p.717–722, 2004.

XU, Y.; LI, X.; JIN, L.; ZHEN, Y.; LU, Y.; LI, S.; YOU, J.; WANG, L. Application of chicken egg yolk immunoglobulins in the control of terrestrial and aquatic animal diseases: A review. **Biotechnology Advances**, Oxford, v. 29, n. 6, p. 860-868, 2011.

YANG, J.; SIQUEIRA, M. F.; BEHL, Y.; ALIKHANI, M.; GRAVES, D. T. The transcription factor ST18 regulates proapoptotic and proinflammatory gene expression in fibroblasts. **The FASEB Journal**, Bethesda, v. 22, n. 11, p. 3956-3967, 2008.

YEW, N.; STROBEL, M.; VANDE WOUDE, G.F. Mos and the cell cycle: the molecular basis of the transformed phenotype. **Current Opinion in Genetics & Development**, London, v. 3, p. 19-25, 1993.

YIN, H.; ZHANG, Z.; LAN, X.; ZHAO, X.; WANG, Y.; ZHU, Q. Association of *MyF5*, *MyF6* and *MyOG* gene polymorphisms with carcass traits in Chinese meat type quality chicken populations. **Journal of Animal and Veterinary Advances**, Faisalabad, v. 10, n. 6, p. 704-708, 2011.

ZENTNER, G.E.; HURD, E.A.; SCHNETZ, M.P.; HANDOKO, L.; WANG, C.; WANG, Z.; WEI, C.; TESAR, P.J.; HATZOGLOU, M.; MARTIN, D.M.; SCACHERI, P.C. CHD7 functions in the nucleolus as a positive regulator of ribosomal RNA biogenesis. **Human Molecular Genetics**, Oxford, v. 19, p. 3491-3501, 2010.

ZHANG, Z.R.; YI-PING, L.; YONG-GANG, Y.; XIAO-SONG, J.; HUA-RUI, D.; QING, Z.H.U. Identification and association of the single nucleotide polymorphisms in calpain3 (CAPN3) gene with carcass traits in chickens. **BMC Genetics**, London, v. 10, n. 1, p. 10, 2009.

ZHBANNIKOV, I.Y.; HUNTER, S.S.; SETTLES, M.L. **SeqyClean user manual**. Disponível em: < <http://sourceforge.net/projects/seqclean/files/>>. Acesso em: 03 jun. 2013.

ZHOU, H.; DEEB, N.; EVOCK-CLOVER, C.M.; ASHWELL, C.M.; LAMONT, S.J. Genome-wide linkage analysis to identify chromosomal regions affecting phenotypic traits in the chicken. I. Growth and average daily gain. **Poultry Science**, Savoy, v. 85, n. 10, p. 1700-1711, 2006a.

ZHOU, H.; DEEB, N.; EVOCK-CLOVER, C.M.; ASHWELL, C.M.; LAMONT, S.J. Genome-wide linkage analysis to identify chromosomal regions affecting phenotypic traits in the chicken. II. Body composition. **Poultry science**, London, v. 85, n. 10, p. 1712-1721, 2006b.

APÊNDICE

**APÊNDICE A - PROTOCOLO DE EXTRAÇÃO DE DNA (PROTEINASE K)
MÉTODO RÁPIDO DE EXTRAÇÃO E PURIFICAÇÃO DE DNA
EXTRAÇÃO COM SAL**

Protocolo modificado para extração de DNA de sangue de galinha

1. Misturar 50 μ L de sangue total colhido em EDTA com 1mL de tampão de lise de células vermelhas;
2. Centrifugar por 1 minuto a 13000rpm. Descartar o sobrenadante. Repetir o passo 1 e 2 (4 vezes);
3. Lavar o pellet uma vez com 0,5mL de água destilada (13000rpm, 1 minuto);
4. Ressuspender o pellet em 80 μ L do tampão de proteinase K (5X), 7 μ L de proteinase K (20mg/ml), 10 μ L de SDS 20% e 282 μ L de água destilada;
5. Incubar a 55°C por 1 hora;
6. Adicionar 120 μ L de NaCl 5M e agitar no vortex por 15 segundos. Centrifugar a 13000 rpm por 5 minutos;
7. Transferir o sobrenadante, secar o pellet, ressuspender em 100 μ L H₂O UP e dissolver passando no vortex por 30 segundos.

Tampão de Lise

1%

Triton X

Tampão de Proteinase K (5X)

0,32M sucrose

12nM Tris-HCl pH7,5

0,375M NaCl

5mM MgCl₂

0,12M EDTA (ph 8,0)