Proceedings, 10th World Congress of Genetics Applied to Livestock Production

SNP Discovery in a QTL Region Associated With Breast Muscle Deposition on Chicken Chromosome 2

C. Boschiero[1*], **T.F. Godoy**[1*], **G.C.M. Moreira**[1], **A.A. Gheyas**[2], **G. Gasparin**[1], **M. Paduan**[1], **S.C.S. Andrade**[1], **H. Montenegro**[1], **D.W. Burt**[2], **M.C. Ledur**[3] **and L.L. Coutinho**[1]
[1]University of Sao Paulo (ESALQ) - Brazil, [2]The Roslin Institute, University of Edinburgh, UK,
[3]EMBRAPA Swine and Poultry National Research Center – Brazil. [*]Equal contributors.

ABSTRACT: Eighteen chickens of two experimental lines (broiler and layer) from EMBRAPA were sequenced with Illumina platform. SNPs were detected in a QTL region on chicken chromosome 2 (between microsatellite markers *MCW0185* and *MCW0264*) which was previously associated with breast muscle deposition. Initially, 722,832 SNPs were identified for the 18 chickens together, and 77% of SNPs were retained after quality filtration. Annotation was performed for the unique variants (85,765 SNPs), and from exonic regions 357 SNPs were classified as synonymous, 153 as non-synonymous and 3 as stopgain. A total of 37 genes were analyzed, and three of them were related to muscle development (*DTNA*, *RB1CC1* and *C-MOS*). We also detected 15 amino-acid changing SNPs which were predicted to be non-tolerated and those were harbored by eight genes: *MEP1B*, *PRKDC*, *NSMAF*, *TRAPPC8*, *SDR16C5*, *CHD7*, *ST18* and *RB1CC1*. Those loss-of-function mutations, and the exonic variants present in genes related to muscle deposition can be considered candidate variants for further studies in chickens.
Keywords: breast muscle; chicken; SNP

Introduction

The breast muscle is the most valuable part of the chicken carcass. To understand the molecular basis of this complex trait, analysis of Quantitative Trait Loci (QTL) has often been used as the initial step, but these QTL regions can be mapped over a large region consisting of millions of bases and containing many genes (Ahsan et al. (2013)).

Recent advances in whole genome sequencing allow detection of thousands of single nucleotide polymorphisms (SNPs), allowing characterization of the genetic variants to identify potentially functional mutations.

According to the QTL database (AnimalQTLdb 2014) 119 QTLs were associated with breast muscle. In a specific region between markers *MCW0185* and *MCW0264* on chromosome 2 (GGA2), Baron et al. (2010) mapped a QTL for breast percentage in a Brazilian F_2 Chicken Resource Population. QTLs for the same trait were also mapped in other chicken populations on GGA2. Nadaf et al. (2009) mapped one QTL associated with breast weight in the same region (*MCW0185-MCW0264*). Other studies mapped QTLs for the same trait in nearby regions (Nassar et al. (2012); Tercic et al. (2009); Ikeobi et al. (2004)).

The aim of this study was to identify SNPs from next generation sequencing data on a QTL region on chicken chromosome 2 (105,848,755-112,648,761 bp), which was previously associated with breast muscle (Baron et al. (2010)) and to suggest candidate SNPs in genes associated with muscle deposition.

Materials and Methods

Experimental population. We sequenced 18 chickens from two parental experimental lines developed by EMBRAPA Swine and Poultry National Research Center, nine from the broiler line (called TT) and nine from the layer line (called CC). The sequenced chickens were the ones used in reciprocal crosses to generate the EMBRAPA F_2 Chicken Resource Population for QTL mapping studies. More details have previously been described by Rosário et al. (2009).

QTL region. The target region selected for this study on GGA2 was based on the previous study of Baron et al. (2010), who mapped in the EMBRAPA F_2 Chicken Resource Population, one 5% genome-wide significant QTL (*F*-ratio=15.3) associated with the percentage of breast by line-cross analysis. This QTL was mapped between microsatellite markers *MCW0185* and *MCW0264* and explained 5.14% of the phenotypic variance. The target region between *MCW0185* and *MCW0264* markers was defined according to the last chicken reference genome provided by NCBI (105,848,755-112,648,761 bp, *Gallus_gallus-4.0*), and it contains 66 genes (Biomart, 2013).

Sequencing. Genomic DNA from blood samples of 18 chickens was extracted by Proteinase K (Promega). The library preparation was performed according to Nextera™ protocol DNA Sample Preparation kit (Illumina). The sequencing was performed with the Illumina platform (HiSeq1000) using TruSeq SBS kit V3 cBot-HS with 200 cycles (Illumina). The initial sequencing coverage was 18X per chicken and read length was 2 x 100 nucleotides.

SNP detection and annotation. Quality trimming of the reads was performed selecting reads with quality ≥24 and minimum length of 65 bp using the SeqyClean tool (v. 1.3.12). The alignment of the reads of each chicken was performed against the latest reference chicken genome

(Gallus_gallus-4.0, NCBI) with the Bowtie2 software (v. 2.1.0, Langmead and Salzberg (2012)).

SNP detection were performed in the target region on GGA2 (105,848,755-112,648,761 bp) with the SAMtools package (v. 0.1.19, Li et al. (2009)) using the mpileup option with mapping and base qualities ≥20. After the initial variant calling, a stringent filtration was performed to minimize false positive SNPs. The filtration was based on four criteria: quality score (≥30), minimum coverage (≥5), maximum coverage (≤mean coverage plus 3 standard deviation), and variant supported by both strands.

The functional annotation of unique SNPs (from 18 chickens) was conducted by ANNOVAR (v. 2013Aug23, Wang et al. (2010)) with default parameters using the gene annotation database from ENSEMBL release 71. We used DAVID Gene Functional Classification tool (v. 6.7, Huang et al. (2009)) with default parameters to identify genes involved in metabolic pathways. Among the identified genes, we selected those which were related to muscle development and studied their variants.

The functional effects of the exonic SNPs that caused changes in amino acids (non-synonymous and stopgain/loss) were predicted with the VEP tool (Variant Effect Predictor, Mclaren et al. (2010)). VEP software predicts SIFT (Sorting Intolerant from Tolerant) scores, which indicates whether the SNP is tolerated or non-tolerated (≤0.05) (Ng and Henikoff (2003)).

Results and Discussion

Sequencing and alignment. A total of 2,785,354,494 short reads were obtained, and after the quality trimming ~76.5% were retained. The retained reads were aligned against the chicken genome reference (Gallus_gallus-4.0). The depth of sequencing coverage for the 18 chicken ranged from 5.4X to 15.6X with an average of 10.6X for the target region studied.

SNPs detection and filtration. In the target region on GGA2 (*MCW0185-MCW0264*), we initially identified 722,832 SNPs for the 18 chickens. In total, we detected 341,816 SNPs from the broiler line and 381,016 SNPs from the layer line (Table 1). After filtration, 77% of the SNPs (n=558,767) were retained. In Table 1 we can see the number of filtered SNPs for broiler and layer lines, and also the unique number (no duplicate SNP between the lines). More SNPs were identified in the broiler than in the layer line. The average density of variants in the target QTL region for the 18 chickens was of 4.6 SNPs/kb, and these results are consistent with the global average rate found by Wong et al. (2004) of 5 SNPs/kb. We also identified the number of homozygous (presence of non-reference alleles only) and heterozygous variants in each of the 18 chickens. Except for three individuals, in all other cases from both lines, we detected more homozygous than heterozygous SNPs.

Table 1. Number of SNPs initially identified and after filtering for layer (LY) and broiler (BR) lines in target QTL region.

	Initially detected		after filtering	
Lines	SNPs	unique SNPs	SNPs	unique SNPs
LY	341,816	62,212	271,070	56,152
BR	381,016	81,514	287,697	73,666
LY and BR	722,832	94,674	558,767	85,765
Average	40,157	5,259	31,043	4,764

Functional annotation and genes related to muscle deposition. Annotation was performed for the 85,765 unique SNPs (Table 1) against the genes from ENSEMBL by ANNOVAR software. The SNPs were classified in genomic regions, such as: intergenic (~63%), intronic (~33%), exonic (~0.6%), and other regions (splicing, UTR, etc). Only 0.42% of the SNPs were classified as synonymous, and 0.18% as non-synonymous (Table 2).

Table 2. Annotation of filtered unique SNPs for 18 chickens.

Variants	Total of SNPs	SNPs %
All	85,765	100
Intergenic	54,282	63.29
Intronic	28,687	33.45
Exonic	513	0.60
Splicing	1	0.001
ncRNA	16	0.02
UTR5	78	0.09
UTR3	544	0.63
Upstream	880	1.03
Downstream	761	0.89
Exonic		
Synonymous	357	0.42
Nonsynonymous	153	0.18
Stopgain	3	0.003

The percentage was calculated based on the total annotated SNPs.

Mutations in coding regions were further analyzed using the VEP tool to predict the SIFT scores which indicates if those are functionally tolerated or not. Not tolerated SNPs are important because they are predicted to alter the protein biological function.

Three out of 37 genes on the target region with exonic variants (non-synonymous and stopgain SNPs) were related to muscle development (*DTNA, RB1CC1* and *C-MOS*), and in all of them nine non-synonymous SNPs were identified.

The *DTNA* gene is a member of the dystrophin family and is involved in the maintenance of homeostasis and neuromuscular junction (Kapsa et al. (2003)). Four tolerated non-synonymous SNPs were identified in this gene, and among them, three were present only in the broiler line, which has higher muscle deposition. The *RB1CC1* gene is involved in muscle development because it is associated with cell proliferation and growth (Nishimura et al. (2011)). Three non-synonymous SNPs were detected in this gene. SNP *rs312375447* (tolerated) was detected in both lines. SNP *c.3318C>G* (novel and tolerated) was present only in the layer line, and SNP *rs314633726* (non-tolerated) was identified in homozygosity in only one chicken of the layer line. The *C-MOS* gene generates transcripts highly expressed in reproductive tissues (Yew et al. (1993)). Two non-synonymous tolerated SNPs were identified in this gene. A novel SNP (*c.904G>A*) and SNP *rs80770298* were detected only in the broiler line.

Prediction of SIFT scores on 153 non-synonymous and three stopgain SNPs present in the genes of the target region were also performed. SIFT prediction was possible for 145 SNPs, and 15 were classified as not tolerated. Besides the non-tolerated SNP identified in *RB1CC1* gene, the other 14 non-tolerated SNPs were identified in seven different genes involved in various important functions (*MEP1B, PRKDC, NSMAF*, *TRAPPC8*, *SDR16C5*, *CHD7* and *ST18*).

Among the non-tolerated SNPs, seven were identified in both lines, five were detected exclusively in the broiler line, and two were present exclusively in the layer line. Although these seven genes were not directly related to muscle development, non-tolerated variants in these genes need to be better investigated. Muscle development is a polygenic trait, and multiple functional mutations from different genes might explain the phenotypic variance.

Conclusion

A large number of SNPs were detected in a QTL region associated with breast muscle in chickens, many of which are novel. By annotating the variants and predicting their functional effects, we suggest a list of candidate functional mutations for future studies. Also, SNPs were identified in genes related to muscle deposition (*DTNA*, *RB1CC1* and *C-MOS*), including non-tolerated and novel SNPs. These candidate variants should be further investigated, used in association studies, and validated in other chicken populations.

Acknowledgements

T.F. Godoy, G.C. Moreira, S.C.S. Andrade and H. Montenegro received fellowships from São Paulo Research Foundation – FAPESP (2012/15294-7, 2012/14965-5, 2012/21784-7 and 2012/02906-4). C. Boschiero received a fellowship from the program Science Without Borders - National Council for Scientific and Technological Development (CNPq, grant 370620/2013-5). L.L. Coutinho is recipient of productivity fellowship from CNPq. We are indebted to M. Brandão for access to the computational resources of the Thunder Cluster.

Literature Cited

Ahsan, M., Li, X., Lundberg, A. E. et al. (2013). Front. Genet. 4:1-8.

AnimalQTLdb. http://www.animalgenome.org/cgi-bin/QTLdb/index Accessed on February 2014.

Baron, E. E., Moura, A. S. A. M. T., Ledur, M. C. et al. (2010). Anim. Genet. 42:117-124.

Biomart. http://www.central.biomart.org/ Accessed on November 2013.

Huang D.W., Sherman B.T., Lempicki R.A. (2009). Nat. Protoc. 4: 44-57.

Ikeobi, C. O. N., Woolliams, J. A., Morrice, D. R. et al. (2004). Livest. Prod. Sci.87: 143-151.

Kapsa, R., Kornberg A. J., Byrne, E. (2003) Lancet Neurol. 2: 299-310.

Langmead, B., Salzberg, S. L. (2012). Nat. Methods 9: 357-359.

Li, H., Handsaker, B., Wysoker, A. et al. (2009). Bioinformatics **25:**2078–9.

Mclaren, W., Pritchard, B., Rios, D. et al. (2010). Bioinformatics 16: 2069–2070.

Nadaf, J., Pitel, F., Gilbert, H. et al. (2009). Physiol. Genomics 38: 241–249.

Nassar, M.K., Goraga, Z.S., Brockmann, G.A. (2012). Anim. Genet. 43: 739-745.

Ng, P.C., Henikoff, S. (2003). Nucleic Acids Res. 13:3812–3814.

Nishimura, I., Chano, T., Kita, H. et al. (2011). Journal Biol. Chem. 286:43925-43932.

Rosário, M. F., Ledur, M. C., Moura, A. S. A. M. T. et al. (2009). Sci. Agri. 2:150–158.

Tercic, D., Holcman, A., Dovc, P. (2009). Anim. Genet. 40:743-748.

Wang, G.K., Li M., Hakonarson, H. (2010). Nucleic Acids Re. 38:164.

Wong, G.K.S., Liu, B., Wang, J. et al. (2004). Nature 432:717-722.

Yew, N., Strobel, M., Vande Woude, G.F. (1993). Curr. Opin. Genet. Dev. 3:19-25.