Discovery of SNPs Potentially Associated with Fatness in a QTL Region on Chicken Chromosome 3

G.C.M. Moreira[1], **T.F. Godoy**[1], **C. Boschiero**[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), Piracicaba, SP, Brazil, [2]The Roslin Institute, University of Edinburgh, UK, [3]EMBRAPA Swine and Poultry, Concórdia, SC, Brazil.

ABSTRACT: Fat content is an important economic trait in chickens. Quantitative trait locus (QTLs) associated with fat deposition have been detected, these regions have low resolution and the functional variants are still unknown. Eighteen parental chickens from a broiler and a layer line developed by the EMBRAPA Swine and Poultry National Research Center were re-sequenced to identify SNPs in a QTL region which was previously associated with abdominal fat weight and percentage on chicken chromosome 3. The identified SNPs were annotated. Those located in exonic regions were selected for further analysis and the genes containing those mutations were analyzed. SNPs were identified in five genes which were previously associated with lipid metabolism, namely *LOC771163*, *EGLN1*, *GNPAT*, *FAM120B*, and *THBS.* Mutations located in these genes are candidates to respond for a part of the phenotypic variation observed in abdominal fat weight and percentage in chickens.
Keywords: Abdominal fat; Re-sequencing; Polymorphisms

Introduction

Fat deposition is a negative factor in chicken meat production. According to Jennen et al. (2004), fatness in chicken meat decreases nutritional value and consequently, the commercial value of the product. Abdominal fat is a trait of high heritability (0.5-0.8) (Bihan-Duval et al. (1999)). Gaya et al. (2006) and Cruz et al. (2011) found positive genetic correlations with feed efficiency and also with body weight.

Most of carcass fat in chickens is placed under the skin, organs and the abdominal portion. Thus, the abdominal fat reflects the total fat deposition in the carcass. Higher fat deposition is observed in chickens that were selected to higher growth rate and muscle deposition. The discovery of SNPs possibly associated with this trait can contribute to selection for decreased fat content in chicken meat production, with no loss in the genetic gain already achieved for growth rate. The broiler line from the experimental population developed by EMBRAPA exhibited about 15 times more fat deposition than the layer line (Campos et al. (2009)).

The next-generation sequencing permits the discovery of a large number of SNPs with less effort and greater accuracy (Rubin et al. (2010)). This technique allows the identification of SNPs which are possibly associated with the studied trait. The aim of this research was to identify SNPs in a QTL region on chromosome 3 GGA3, previously associated with abdominal fat weight and percentage in parental chickens from the EMBRAPA F_2 Chicken Resource Population

Materials and Methods

Experimental population. Eighteen parental chickens from the Brazilian F_2 experimental population developed by EMBRAPA Swine and Poultry were resequenced (nine from the broiler and nine from the layer line). In this population, QTL mapping studies were carried out for several group of traits. The broiler line was selected for body weight, feed efficiency, carcass yield, breast weight and other traits for six generations, while the layer line was selected for egg weight and production, feed efficiency, egg quality, reduced body weight and other traits for eight generations.

Target region. A target region was defined between the microsatellite markers *LEI0161-ADL0371* (33,595,706-42,632,651 pb) comprising 9,036,945 pb. These positions were defined according the last reference sequence (Gallus_gallus-4.0, NCBI). This region was found to be associated with abdominal fat weight and percentage in a study by Campos et al. (2009). This region contains around 120 genes (BioMart, 2013).

Library preparation and sequencing. The DNA was extracted from blood samples by *proteinase K* protocol. After quantification by Qubit® fluorometer, the samples were diluted to 2.5 ng/µL. The library preparation were done using the *Nextera DNA Sample Preparation kit* (Illumina) and quantified by PCR Real Time, using *KAPA Library Quantification Kit* (KAPA Biosystems). After dilution to 16 pM, the samples were prepared to clustering using *TruSeq Kit PE Kit Cluster V3* (Illumina) and clustered by cBOT (Illumina). The sequencing was done with the Illumina platform, paired-end sequencing, with an initial estimated coverage of 18X per chicken.

SNP discovery. The reads obtained after the sequencing were trimmed by SeqyClean (v.1.3.12, Zhbannikov et al. (2013)) with q≥24 and read size ≥65 pb. The retained reads were aligned against the reference genome by Bowtie2 (v.2.1.0, Langmead and Salzberg (2012)) and the identification of SNPs was performed by SAMtools software (v.0.1.19, Li et al. (2009)). After SNP identification, a careful filtering was done based on four criteria: (1) quality score ≥ 30, (2) total depth of coverage at the variant site ≥5, (3) total depth of coverage ≤ mean coverage + 3SD (standard deviation), and (4) the evidence of a variant being supported by both forward and reverse strands (Kranis et al. (2013); Kumar et al. (2012); Amaral et al. (2009)). We also distinguished homozygous and heterozygous SNPs. A variant was referred to as homozygous when only a non-reference allele was detected and heterozygous when both the reference and non-reference alleles were detected from an individual.

Functional annotation. The unique SNPs (with no duplicates) were annotated by ANNOVAR software (v.2013aug23, Wang et al. (2010)) with default parameters using the gene annotation database from Ensembl release 71. After annotation, the functional effects of exonic SNPs (non-synonymous and stopgain/loss) were predicted by VEP tool online (McLaren et al., 2010). VEP tool predicts SIFT (Sorting Intolerant from Tolerant) scores, which predicts whether the SNP is tolerated or not (≤0.05) (Ng and Henikoff, 2003). DAVID Gene Functional Classification tool (v. 6.7, Huang et al. (2009)) was used to identify genes involved in metabolic pathways with default parameters, and those related to the lipid metabolism were selected.

Results and Discussion

The average sequencing coverage for the 18 chickens after quality trimming and the removal of PCR duplicates was ~10X for the whole genome, and ~11X for the target region studied. Approximately 2.8 billion reads of 100 bp were obtained from the 18 chickens, and after quality trimming ~77% of the reads were retained.

We identified 136,054 unique SNPs in the target region (GGA3: 33,595,706–42,632,651) of the 18 chickens. All SNPs were classified according to their respective quality bins based on *phred* score (Figure 1). The filtration of the SNPs based on four criteria resulted in the removal of ~22% of the unique SNPs (n=12,646) for the 18 chickens.

The variant density for the target region on GGA3 was 13.7 SNPs/kb, calculated from the ratio between the total number of unique filtered SNPs for the 18 chickens and the size of the target region in base pairs. Previous studies have reported highly variable densities of variants in chicken, ranging from ~5 to ~78 SNPs/kb across the genome (Kranis et al. (2013); Rubin et al. (2010); Wong et al. (2004)). The high variability in average density depends on the number of chickens sequenced in these studies. We also identified homozygous and heterozygous SNPs, and on an average 59% of the SNPs were homozygous per chicken. This higher proportion of homozygous variants was observed in both lines.

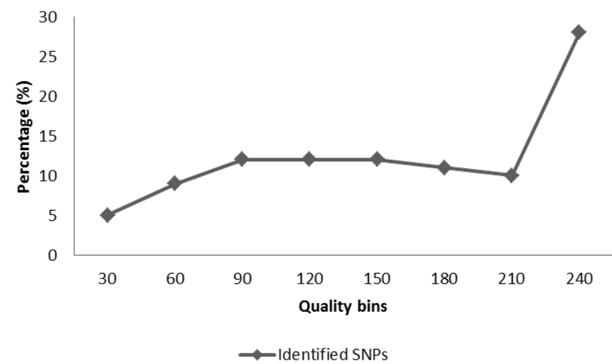

Figure 1: Percentage of SNPs classified according to their quality.

After annotation of unique SNPs (Table 1), 1,226 SNPs in exonic regions were classified as synonymous, non-synonymous and stopgain to identify the genes that contain potential mutations related to the variation in fat deposition. In the total, 77 genes were identified and from them, five were related to lipid metabolism, namely: *LOC771163*, *EGLN1*, *GNPAT*, *FAM120B* and *THBS2*. Five not tolerated SNPs were identified in the genes *LOC771163*, *EGLN1*, *GNPAT* and *FAM120B*, while two tolerated non-synonymous SNPs were identified in the *THBS2* gene (Table 2). Not tolerated SNPs are important because they are predicted to alter the biological function of the protein. Considering all the not tolerated SNPs identified across the five genes, six were novel and were submitted to dbSNP (NCBI).

Table 1: Annotation of unique SNPs filtered for 18 chickens (layer and broiler lines).

Variants	Total SNPs	Percentage (%)
All variants[1]	*124,509*	*100*
intronic	61,129	49.10
intergenic	58,099	46.67
exonic	1,226	0.98
splicing	7	0.01
ncRNA	14	0.01
5' UTR	211	0.17
3' UTR	1,016	0.82
upstream (1 kb)	1,393	1.12
downstream (1 kb)	1,392	1.12
Exonic		
synonymous SNP	840	0.68
non-synonymous SNP	376	0.30
stopgain SNP	10	0.008

[1]To calculate the percentage, the total number of annotated SNPs was considered.

Table 2: Non-synonymous SNPs identified in genes related to lipid metabolism.

Gene (Symbol)	SNP (ID from NCBI)	Line (N[1])
LOC771163	c.575C>T (ss947428828)	Broiler (1)
	c.541G>A (ss947428829)	Broiler (2)
EGLN1	c.29C>A (ss947428846)	Layer (1)
		Broiler (3)
GNPAT	c.377C>T (ss947428835)	Layer (1)
FAM120B	c.1496A>C (rs13675883)	Layer (3)
		Broiler (1)
THBS2	c.52C>T (ss947428847)	Layer (1)
	c.1553T>C (ss947428848)	Broiler (2)

[1]Number of chickens in which the mutation was identified in the line.

The gene *THBS2* was differentially expressed in chicken abdominal fat in a study using chicken lines with lower and higher fat deposition (Resnyk et al. (2013)). Variants in this gene might be related to a higher abdominal fat deposition in chickens. In the *THBS2* gene, two novel tolerated SNPs from the layer and broiler lines were identified; a heterozygous SNP in one chicken from the layer line and a homozygous SNP in two chickens from the broiler line (Table 2).

The increase in gene expression of *LOC771163* may result in increased fat deposition in humans (Rosmond et al. (2001)). Two novels not tolerated SNPs were detected in *LOC771163* gene only in chickens from the broiler line (all heterozygous SNPs). The *EGLN1* gene is related to the regulatory mechanism of cell differentiation into adipocytes in humans (Wang et al. (2012)). One novel not tolerated SNP was identified in this gene in three chickens from the broiler line, and in one from the layer line. The *GNPAT* gene is responsible for synthesis of enzymes associated with lipid synthesis (Mizuno et al. (2013)). In this gene, one novel not tolerated SNP was identified, but it was detected only in one chicken from the layer line (heterozygous SNP). In a study with mice, Li et al. (2007) detected the *FAM120B* gene expressed in different tissues including adipose tissue. In this gene, one not tolerated SNP (*rs13675883*) was identified in the broiler and layer lines (heterozygous SNP). All the SNPs described above are in Table 2

The not tolerated mutations identified in each of those genes related to lipid are important candidate variants for further association studies. In chickens that the QTL are segregating (Campos et al. (2009)), those mutations should be further analyzed.

Conclusion

This study allowed the characterization of variants from a QTL region on GGA3 that was previously found to be associated with abdominal fat deposition in chickens. Seven SNPs on exonic regions in genes previously associated with fat deposition (lipid metabolism) were identified (six are novel SNPs). These SNPs should be further investigated, validated in other chicken populations, and/or used in future association studies.

Aknowledgements

T.F. Godoy, G.C. Moreira, H. Montenegro and S.C.S. Andrade received fellowships from São Paulo Research Foundation – FAPESP (2012/15294-7, 2012/14965-5, 2012/02906-4 and 2012/21784-7). C. Boschiero received a fellowship from the program Science Without Borders - National Council for Scientific and Technological Development (CNPq, grant 370620/2013-5).

Literature Cited

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

Amaral, A. J., Megens, H. J. et al. (2009) BMC Genomics, 10:374-384.

BioMart. http://central.biomart.org/ Accessed in November 2013.

Campos, R. L. R., Nones, K., Ledur, M. C. et al. (2009) Anim. Genet., 40:729–736. , 2009.

Cruz, V. A. R., Venturini, G. C. et al. (2013) http://ainfo.cnptia.embrapa.br/digital/bitstream/item/53466/1/estimativas-de-parametros-geneticos-para-peso-corporal.pdf. Accessed in November 2013.

Gaya, L. D. G., Mourão, G. B., Ferraz, J. B. S. (2006) Cienc. Rural, 36: 709-716.

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

Jennen, D. G. J., Vereijken, A. L. J., Bovenhuis, H. et al. (2004) Poult. Sci., 83:295–301.

Kranis, A., Gheyas, A. A., Boschiero, C. et al. (2013) BMC Genomics, 14:59.

Kumar, S., Banks, T. W., Cloutier, S. (2012) Int. J. Plant. Genomics, 2012: 1-15.

Langmead, B., Salzberg, S. L. (2012) Nat. Methods, 6:357-359.

Bihan-Duval, E.L., Millet, N., Remignon, H. (1999) Poult. Sci., 78: 822–826.

Li, H., Durbin, R. (2009) Bioinformatics, 25:1754–1760.

McLaren, W., Pritchard, B., Rios, D. et al. (2010) Bioinformatics, 26:2069–2070.

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

Mizuno, Y., Ninomiya, Y., Nakachi, Y. et al. (2013) PLoS Genetics 9: 1-15.

Resnyk, C.W., Carré, W., Wang, X., et al. (2013) BMC Genomics 14:1-26.

Rosário, M. F., Ledur, M. C., Moura, A. S. A. M. T. et al. (2009) Sci. Agric., 66: 150–158.

Rosmond, R., Chagnon, M., Bouchard, C., Bjorntorp, P. (2001) J. Clin. Endocrinol. Metab., 86:2178-2180.

Rubin, C.J., Zody, M. C., Eriksson, J., Meadows, J. R. S. et al. (2010) Nat., 464: 587–591.

Wang, K., Li, M., Hakonarson, H. (2010) Nucleic. Acids. Res., 16:1-7.

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

Zhbannikov, I. Y., Hunter, S. S., Settles, M. L. (2013) http://sourceforge.net/projects/seqclean/files/ Accessed in June 2003.