


RESEARCH ARTICLE

Open Access



Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines

Clarissa Boschiero^{1,4*} , Gabriel Costa Monteiro Moreira¹, Almas Ara Gheyas², Thaís Fernanda Godoy¹, Gustavo Gasparin¹, Pilar Drummond Sampaio Corrêa Mariani¹, Marcela Paduan¹, Aline Silva Mello Cesar¹, Mônica Corrêa Ledur³ and Luiz Lehmann Coutinho¹

Abstract

Background: Meat and egg-type chickens have been selected for several generations for different traits. Artificial and natural selection for different phenotypes can change frequency of genetic variants, leaving particular genomic footprints throughout the genome. Thus, the aims of this study were to sequence 28 chickens from two Brazilian lines (meat and white egg-type) and use this information to characterize genome-wide genetic variations, identify putative regions under selection using Fst method, and find putative pathways under selection.

Results: A total of 13.93 million SNPs and 1.36 million INDELs were identified, with more variants detected from the broiler (meat-type) line. Although most were located in non-coding regions, we identified 7255 intolerant non-synonymous SNPs, 512 stopgain/loss SNPs, 1381 frameshift and 1094 non-frameshift INDELs that may alter protein functions. Genes harboring intolerant non-synonymous SNPs affected metabolic pathways related mainly to reproduction and endocrine systems in the white-egg layer line, and lipid metabolism and metabolic diseases in the broiler line. Fst analysis in sliding windows, using SNPs and INDELs separately, identified over 300 putative regions of selection overlapping with more than 250 genes. For the first time in chicken, INDEL variants were considered for selection signature analysis, showing high level of correlation in results between SNP and INDEL data. The putative regions of selection signatures revealed interesting candidate genes and pathways related to important phenotypic traits in chicken, such as lipid metabolism, growth, reproduction, and cardiac development.

Conclusions: In this study, Fst method was applied to identify high confidence putative regions under selection, providing novel insights into selection footprints that can help elucidate the functional mechanisms underlying different phenotypic traits relevant to meat and egg-type chicken lines. In addition, we generated a large catalog of line-specific and common genetic variants from a Brazilian broiler and a white egg layer line that can be used for genomic studies involving association analysis with phenotypes of economic interest to the poultry industry.

Keywords: Fat deposition, Fst, Genetic variants, Next generation sequencing, INDEL, Selection signatures, SNP, Poultry

* Correspondence: clarissaboschi@yahoo.com

¹Animal Biotechnology Laboratory, Animal Science Department, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, SP 13418-900, Brazil

⁴Noble Research Institute, 2510 Sam Noble Parkway, Ardmore, Oklahoma 73401, USA

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Chickens are one of the most important animals in the world, not only because of the intensive production of meat and eggs, but also because they are a model for developmental and genomic studies. The chicken genome has become an important tool for the worldwide avian research community since 2004, with the release of the first draft of the genome from a single female red jungle fowl, which is the wild ancestor of domestic chicken [1].

The domestic chicken has hundreds of different breeds, but commercial chickens can be divided into two main groups – broilers (meat-type) and layers (egg-type), which have been artificially selected for centuries. These two types have different phenotypic and genotypic profiles as a consequence of distinct genetic backgrounds and intensive genetic selection for different traits [2]. Broiler selection is focused on growth and muscle deposition, e.g. body weight, feed conversion, and breast weight. On the other hand, selection of layer chickens is focused mainly on reproductive traits, e.g. egg production and egg quality.

Different selection criteria applied for broiler and layer lines resulted in considerable differences in growth, development, and metabolic mechanisms during embryogenesis and hatching [3]. This intensive selection, which resulted in important changes in chicken phenotypes, can be detected by selection signatures in the genome [4, 5].

Recently, millions of SNPs and INDELs have been identified from different chicken genomes based on next generation sequencing data (NGS) [6–8], and it is now possible to detect regions of selection signatures on a genome-wide scale. Recent studies in chickens identified regions under selection using only broiler lines [9, 10], only layer lines [11, 12], or pooled sample data [4, 7].

In Brazil, Embrapa Swine and Poultry National Research Center maintains several chicken lines under multi-trait selection in Brazilian climatic and nutritional conditions for more than 20 years. Two of those lines, a white-egg type (layer) and a meat-type (broiler), due to considerable difference in their growth rates and carcass yields, were utilized to generate an F2 experimental population for QTL mapping studies [13, 14]. The CC layer line used in this study is a white-egg laying hen originated from the White Leghorn, and it has been selected since 1989 mainly for reproductive traits such as egg production and fertility, besides egg quality traits [15]. The broiler TT is a paternal line originated from the White Plymouth Rock, New Hampshire and White Cornish breeds, and it has been selected since 1992 mainly for growth and meat traits such as body weight, carcass, and cuts yields [15]. The average weight of TT broilers (~2.4 kg) was almost five-fold higher than that of CC layers (~0.5 kg), when reared as broilers, and also presented higher abdominal fat percentage and breast

yield compared to CC [16]. Moreover, the layer line has the egg production average at 70 weeks of about 210 eggs and the first egg at 140 days [17]. Genomic characterization of these two lines and their comparison with existing commercial chicken lines can provide valuable information about artificial and natural selection in these lines.

In order to achieve a greater understanding of the genetic differences between these two Brazilian white-egg layer and broiler lines, a deep catalog of genetic variants (SNPs and INDELs) were generated by sequencing the genome of 28 chickens (14 per line) at medium sequencing coverage. SNPs and, for the first time in chicken, INDELs were utilized for detecting potential selection signatures using *Fst* approach. Genes located in candidate regions under selection and genes with variants of potential functional effect were further analyzed to consider their possible influence on economically important traits in chickens, such as fat deposition, growth and reproduction. In summary, this study provides novel insights into selection footprints that can help elucidate the functional mechanisms underlying important traits relevant to broiler and layer chicken lines. In addition, this study presents detailed information about variants in the Brazilian broiler and layer chicken lines that could be important for genetic studies involving association analysis with relevant phenotypes.

Results

Sequencing and alignment

Approximately 5 billion short reads were generated from 14 broilers, and 14 white-egg layers. After quality trimming, ~78% of the reads were retained. About ~98.8% of the quality-processed reads could be aligned to the chicken reference genome (*Gallus_gallus*4.0). After removal of PCR duplicates, the average sequencing coverage for the 28 individuals was ~11.2 X (average of 11.4 X for broiler line and 10.9 X for white-egg layer line).

Polymorphism identification

The initial variant call resulted in 15,944,063 SNPs and 1,997,771 INDELs from 28 individuals analyzed, including autosomes 1–28 and 32, two linkage groups, mitochondrial, sex chromosomes (W/Z), and unplaced scaffolds. The genome-wide average density of variants considering variants from all 28 chickens was 16.84 SNPs/kb and 1.72 INDELs/kb. SNP densities for each line were 13.15 SNPs/kb for broilers and 10.82 SNPs/kb for layers. INDEL densities for each line were 1.41 INDELs/kb for broilers and 1.20 INDELs/kb for layers. Additional file 1 shows the frequency of substitution types in the SNPs that were initially called. The most frequent substitutions were G to A (18.5%) and C to T (18.4%).

Filtration of the variants based on several criteria (see Methods) removed around 4% of SNPs and 15% of INDELs, resulting in the final list of 13,927,521 SNPs and 1,361,946 INDELs from the 28 individuals. Unplaced scaffolds were not considered for the subsequent analyses, hence 478,182 SNPs and 26,057 INDELs were excluded from these regions. Also, for all the downstream analyses, only the filtered datasets of variants were considered.

The transition and transversion (TS/TV) ratio for SNPs initially detected in the whole-genome was 2.17, while the ratio in the filtered set was 2.31.

More variants were detected in the broiler ($n = 11,856,959$ SNPs and 1,200,710 INDELs in the filtered set) than in the layer line ($n = 9,780,747$ SNPs and 1,046,645 INDELs). About 26% (3,668,592) of all the SNPs were detected exclusively in the broiler line while ~12% (1,592,380) were detected only in the layer line (Fig. 1a). Similarly, about 22% (289,244) of the INDELs were specific to broiler, while ~10% (135,179) were specific to the layer line (Fig. 1b).

Moreover, a higher proportion of heterozygous variants were identified in the broiler line (on average 53.9% of SNPs and 46.6% of INDELs per individual) than in the layer line (average 44.8% of SNPs and 38.8% of INDELs) (Additional file 2). For both lines, INDEL variants showed greater level of homozygosity than SNPs. Considering all 28 chickens together from both lines, on average, 50.6% of the SNPs and 57.3% of the INDELs were homozygous per chicken, while the rest were heterozygous.

Frequency distributions of alternate allele (AAF) of SNPs and INDELs were estimated within the two lines. Our data shows that most of the SNPs (54%) and INDELs (46.3%) from the broiler and the layer lines had low frequency (≤ 0.3) (Fig. 2). However, some variants

had reached near fixation (alternative allele frequency ≥ 0.9) within lines. About 9.6% of the SNPs and 8.8% of the INDELs in the layer line were fixed, while the broiler line showed 5.9% and 5.8% of the SNPs and INDELs, respectively, as fixed.

Functional annotation of polymorphisms

Annotation showed that most of the variants belonged to non-coding regions of the genome, such as intergenic (38.5% SNPs and 39.4% INDELs), and intronic (42.1% SNPs and 42.9% INDELs) regions. Some of the variants (11.6% SNPs and 10.9% INDELs) had multiple annotations as those could be classified into multiple categories. Only about 1.2% of the SNPs were located in exonic regions, including synonymous (69.9%), non-synonymous (29.8%) and stop gain/loss SNPs (0.3%) (Table 1). Compared to SNPs, only 0.2% of the INDELs were located in exonic regions, including frameshift (54%) and non-frameshift INDELs (42.8%) (Table 1).

In the present study, it was possible to predict the potential effect for 85% of the non-synonymous SNPs using the SIFT algorithm [18] which predicted 7255 SNPs (16%) to be evolutionary intolerant and 38,363 (84%) tolerant. Out of the 7255 intolerant SNPs, 2562 were present exclusively in the broiler line and 1302 were present only in the white-egg layer line.

Metabolic pathway analyses of polymorphisms of potential functional impact

Metabolic pathways of genes harboring variants of high potential functional effects (e.g. intolerant SNPs and frameshift and non-frameshift INDELs) were investigated using the QIAGEN's Ingenuity® Pathway Analysis software (IPA) [19].

The analysis of all genes containing intolerant SNPs from the white-egg layer and the broiler lines combined

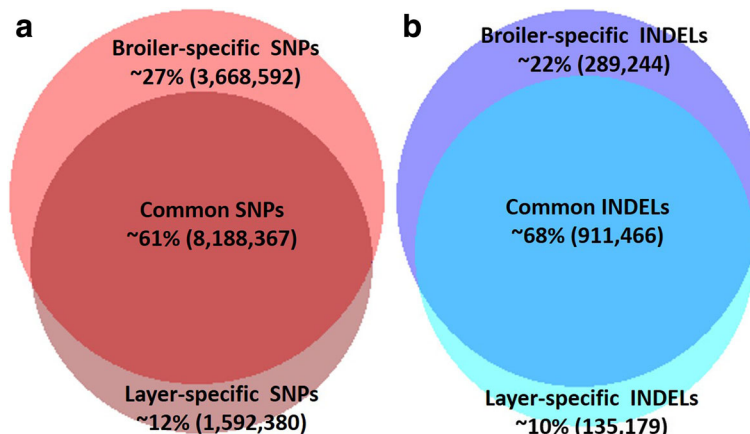


Fig. 1 Venn diagrams of SNPs (a) and INDELs (b) shared between broiler and layer lines. The size of the circles reflects the relative number of filtered variants from each line within the total number of SNPs ($n = 13,449,339$) or INDELs ($n = 1,335,889$)

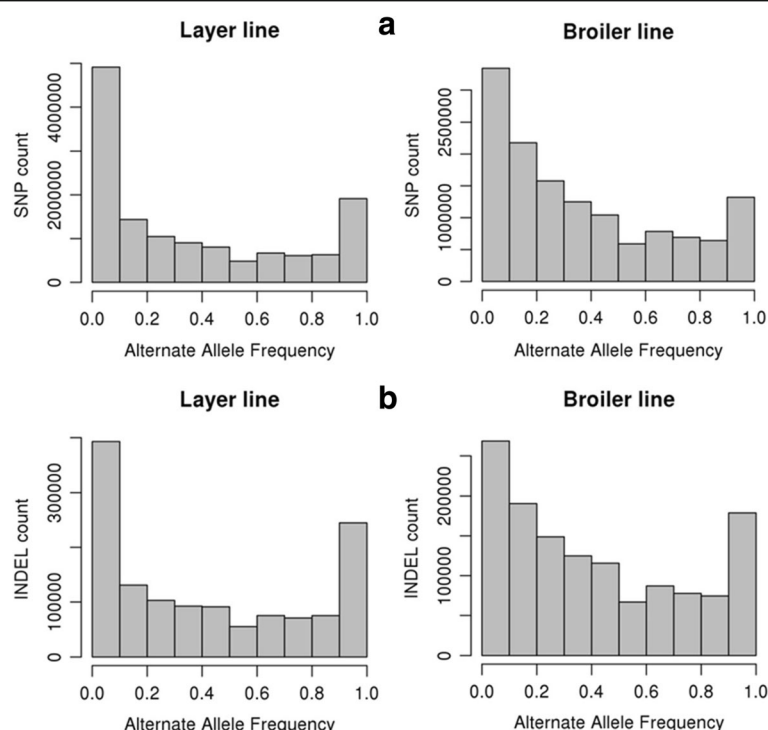


Fig. 2 Distribution of alternative allele frequencies of SNPs (a) and INDELs (b) from layers and broilers

Table 1 Summary of functional annotation of SNP and INDEL variants from layer and broiler chicken lines

Category	SNP Count	SNP %	INDEL Count	INDEL %
Total no of variants	13,449,339		1,335,889	
Total annotation	15,223,947	100	1,500,714	100
Alternative annotation	1,774,608	11.66	164,825	10.98
Intergenic	5,866,525	38.53	590,736	39.36
Intronic	6,409,486	42.10	644,103	42.92
Exonic	175,258	1.15	2555	0.17
1 kb downstream	1,274,658	8.37	123,378	8.22
1 kb upstream	1,306,400	8.58	116,815	7.78
UTR3'	142,796	0.94	17,587	1.17
UTR5'	26,064	0.17	2462	0.16
Splicing	20,703	0.14	2866	0.19
ncRNA	1814	0.012	164	0.01
miRNA	243	0.002	48	0.003
Exonic categories				
Synonymous	122,502	69.90	–	–
Non-synonymous	52,219	29.80	–	–
Stopgain	424	0.24	–	–
Stoploss/retained	88	0.05	–	–
Frameshift	–	–	1381	54.01
Non-frameshift	–	–	1094	42.82
Coding unknown	25	0.01	80	3.13

resulted in eight significant networks (Table 2). A total of 260 genes harboring intolerant SNPs participated in these eight pathways, which were related to connective and metabolic disorders, embryonic development, cardiovascular diseases, and carbohydrates metabolism, among others.

When the same analysis was performed separately for the broiler and the white-egg layer lines using genes harboring intolerant SNPs specific to a line, clearly different networks were observed (Table 3). In the layer line, seven significant networks were identified mainly related to reproductive system development and cellular functions, and also endocrine system (Table 3). In the broiler line, 12 significant networks were identified, which were mainly related to lipid and carbohydrate metabolism, metabolic and dermatologic diseases, neural development, and also organism injury (Table 3).

The metabolic pathways obtained from the analysis of coding INDELs showed a total of 257 genes participating in nine metabolic pathways, mainly associated with diseases and disorders, for example: embryonic development, connective tissue disorders, gastrointestinal, metabolic, neurological and hepatic diseases, cardiovascular development, and cancer (Additional file 3). When the IPA analysis was performed separately for the broiler and the layer lines using genes harboring line-specific coding INDELs, three significant networks for each line were detected, and different networks were observed for

Table 2 Significant networks with intolerant SNPs from the layer and broiler lines combined

Molecules ^a	Total no of molecules	Score	Top diseases and functions
AGAP3, APOLD1, CALCR, CDKL1, EDIL3, FHLS, LEPR, LEPROT, NCOR2, TGFB1	33	38	connective tissue disorders, metabolic disease, cellular assembly and organization
AMOTL2, API5, BICD1, DZIP1, ERK, FBF1, HEATR1, IFT122, NIN, TEX11	34	35	cellular assembly, organization, morphology and maintenance
GALR3, GPR6, GPR20, GPR26, GPR97, GPR98, LPHN3, LYST, MC2R, MC3R	34	35	cell signalling and interaction, cellular function and maintenance
ABTB1, BRI3BP, CHD9, DDX18, DDX31, ESR1, FKBP7, GREB1, PPIL4, WHSC1	32	33	organismal, embryonic and tissue development
ACTR10, ANK2, APC2, DCTN4, DISC1, DNAH1, DNAH3, DNAH5, FARSB, KIF3C	32	33	cellular assembly and organization, nervous system development, cancer
ACTA2, DSC2, DSG2, DTX1, DTX3L, MTCL1, NOTCH2, SERPINB12, SRC, TMCO4	33	33	cardiac arrhythmia, cardiovascular disease and congenital heart anomaly
AK7, AK8, AK9, ARC, CAPN6, CAPN8, CAPN9, HIP1, KNTC1, PIPOX	30	33	nucleic acid and carbohydrate metabolism
AP5Z1, ASTE1, FLT3, FOXK2, HSP, HSPA2, HSPH1, METAP1, TSC1, PHF3	32	31	cell cycle, developmental and hereditary disorders

^aExample of ten molecules present in each network

the broiler and layer lines, e.g. growth and cardiac development in broilers (Additional file 4).

Genome-wide scan for selection signatures

Fst values were obtained from SNP and INDEL datasets separately using an overlapping sliding window of 20 kb with 10 kb step size. A sufficient number of markers per window was obtained, ranging between 10 to 1094 SNPs

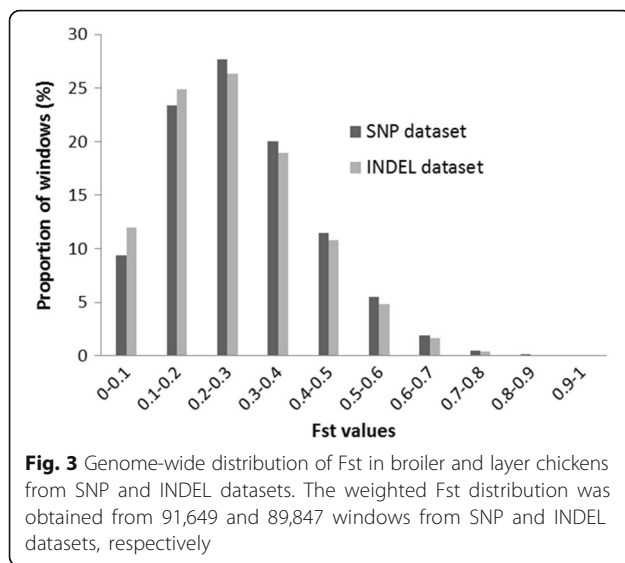
(average of 268/window) and 5 to 114 INDELs (average of 26/window).

Genome-wide weighted Fst distribution was examined (Fig. 3) based on 91,649 and 89,847 windows from SNP and INDEL separate datasets, respectively, between two lines (meat and white egg-type). The majority of the windows (60–63%) had low Fst values (< 0.3), and a few windows showed extremely high Fst values > 0.9 (11 and 3 windows for the SNP and INDEL datasets, respectively).

Table 3 Top five significant networks from intolerant SNPs exclusively from the layer or broiler line

Molecules ^a	Total no of molecules	Score	Top diseases and functions
Layer line			
BPI, DISC1, DYX1C1, IRG1, KNTC1, MAST2, MLKL, ODF2L, TELO2, TPO	34	54	cellular morphology, assembly, organization and maintenance
ACO1, FSH, ING1, LH, NEB, NPR2, PHKA2, RB1CC1, TNK2, Tropomyosin	27	37	organ morphology, organismal development and reproductive system development
AFF1, BARD1, CDC2, Cyclin A, Cyclin D, Cyclin E, EFHC1, ESR1, GNPAT, RB	27	37	developmental disorder, immunological disease and cancer
Alpha tubulin, Beta Tubulin, Dynein, EDIL3, ERK, FBN1, GPSM2, Growth Factor Receptor, INCENP, NPHP1	26	35	cellular assembly, organization, function, maintenance and cycle
ADCY, Caveolin, FHLS, GLUD1, IDH2, INSR, Insulin, LEPR, Proinsulin, Trypsin	24	31	reproductive system development, cellular growth and proliferation and endocrine system disorders
Broiler line			
AMY2A, Amylase, Apyrase, DMBT1, EDA2R, NGNL2, INF2, NKRF, NUA2, OTULIN	32	41	lipid metabolism, nervous system development and post-translational modification
AMOTL2, ERK, FBF1, IQCE, MIPOL1, SCEL, SETD5, SH3RF2, SMPD1, TIPIN	32	41	carbohydrate and drug metabolism and small molecule biochemistry
AKAP, AKAP6, AKAP9, DISC1, LXN, MAPRE3, PCNXL4, PPM1E, RPS6KA1, SCT	32	41	cellular morphology, assembly, organization, function and maintenance
DCHS1, DHX30, EFCAB11, ESR, FTSJ3, HEATR1, MGA, MRPL3, NDC1, UTRN	31	38	metabolic disease, hereditary disorder and cell cycle
Aconitase, CRISP2, LAMA1, MEF2, MYBPH, MYH2, MYO9A, NFS1, OFD1, Tubulin	30	37	cancer, dermatological diseases, organismal injury and abnormalities

^aExample of ten molecules present in each network



SNP dataset had the mean weighted F_{st} of 0.2779 (SD 0.14), and INDEL dataset had the mean weighted F_{st} of 0.2645 (SD 0.14).

Even though the number of INDELs per window was smaller than those in the SNP dataset, similar regions were identified in the SNP and INDEL analyses, with a significant positive correlation ($r = 0.66$; $p < 0.001$) between the F_{st} values from the same window of putative selection signatures (top 1% F_{st} values) of SNPs and INDELs analysis.

Based on the SNP dataset, 92 windows with top 0.1% F_{st} value ($F_{st} \geq 0.817$, Fig. 4) were considered as strong candidates of selection signatures while 916 windows with top 1% F_{st} ($F_{st} \geq 0.6718$, Fig. 4) were deemed as putative. These putative selection signature windows represented most of the chromosomes evaluated, except GGA16, 21, 22, 23, 25 and 27. The highest F_{st} value observed was 0.976 (GGA8: 28,220,001–28,240,000). Many windows

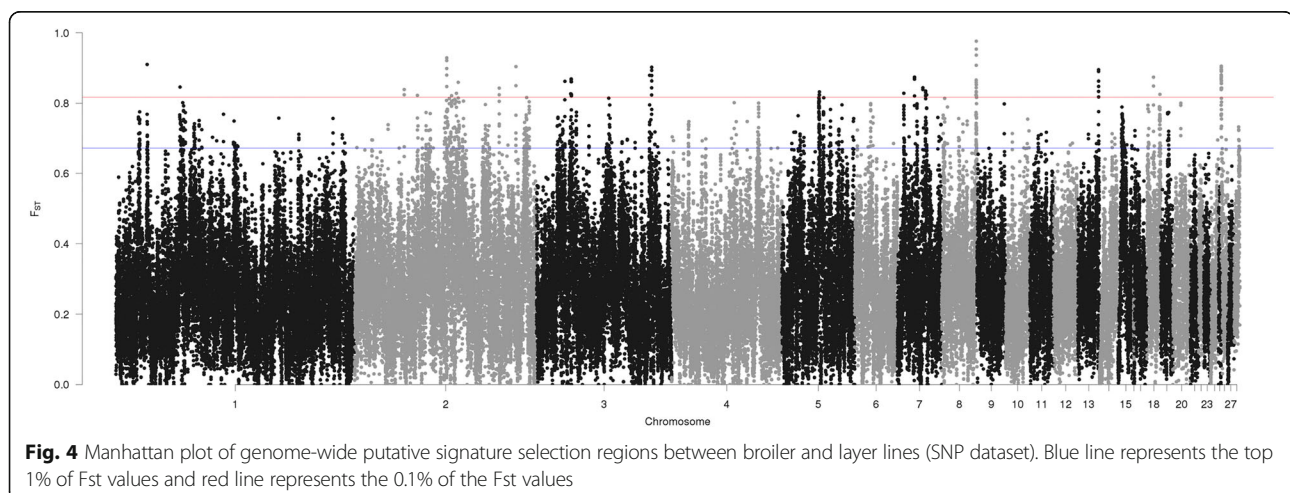
that passed the top 1% F_{st} threshold could be merged into larger regions when those were adjacent or overlapping, e.g. on GGA26, 32 windows passing the threshold could be grouped into two regions only (GGA26:10,001–350,000 and GGA26:4,090,001–4,110,000). Merging adjacent windows resulted in 345 regions (~12.8 Mb in total) for SNP based analysis. The 916 putative selection signature regions from SNP-based analysis intersected with 307 genes, including 37 novel chicken genes, eight miRNAs, and four ultra-conserved elements (Additional file 5).

When the INDEL dataset was analyzed, 90 candidate windows were identified with strong evidence of selection (top 0.1%, $F_{st} \geq 0.7865$, Fig. 5), and 896 windows as putative selection signatures (top 1%, $F_{st} \geq 0.6537$, Fig. 5). These regions represented all autosomes analyzed (chromosomes 1–28), except GGA12, 16, 22, 23, 25 and 27. The highest F_{st} value (0.951) was observed in one window on GGA13 (16,930,001–16,950,000). Merging the adjacent windows from INDEL-based analysis resulted in 425 regions (~13.5 Mb in total length). Annotation of these INDEL-based selection signatures represented 363 genes or functional features, including 48 novel chicken genes, nine miRNAs, two small nucleolar RNAs, and four ultra-conserved elements (Additional file 5).

Out of the 307 genes from the SNP-based analysis and the 363 genes from the INDEL-based analysis, 220 genes were common. When all merged signature regions (top 1% of F_{st} values) obtained from both analyses (345 and 425 regions) were compared, most of the regions ($n = 260$ with total length ~8.6 Mb) were found to be common between the two analyses.

Candidate genes under selection for fat deposition and muscle development

A number of genes potentially related to fat deposition and muscle development were identified among the genes overlapping selection signature regions (Table 4).



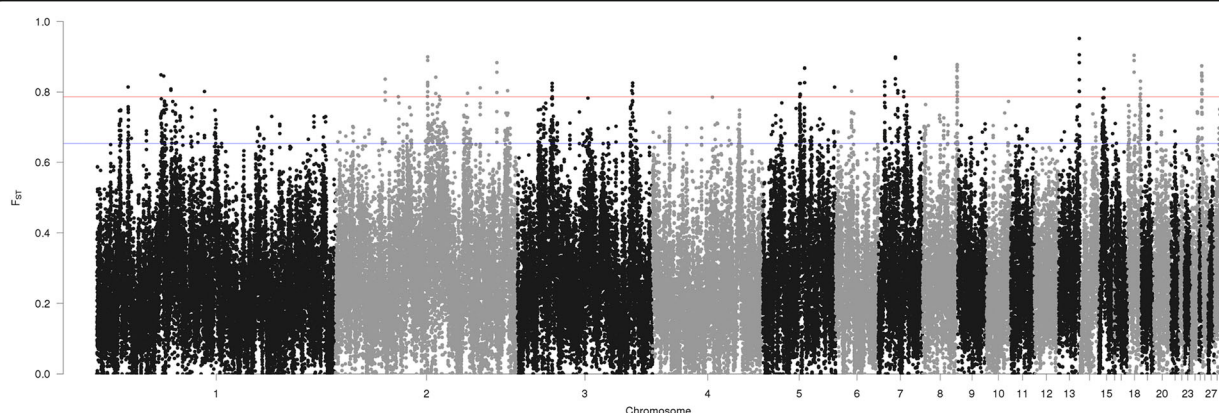


Fig. 5 Manhattan plot of genome-wide putative signature selection regions between broiler and layer lines (INDEL dataset). Blue line represents the top 1% of F_{st} values and red line represents the 0.1% of the F_{st} values

We analyzed all the genes under putative selection, 307 from SNP-based analysis and 363 from INDEL-based analysis, with GeneMANIA prediction server [20], which performs a functional enrichment analysis to find the biological functions of genes. The genes from the SNP-based selection signatures resulted in two enriched pathways: muscle system process and negative regulation of ion transport (Additional file 6). The 13 genes from the pathway related to muscle development were *ACTC1*, *AKAP6*, *ATP2A2*, *CACNA1S*, *CAV3*, *KCNMA1*, *MYOCD*, *NOS1*, *PDE4D*, *TPM4*, *TTN*, *VCL*, and *VIPR1*.

The genes from the INDEL-based selection signatures resulted in six enriched pathways related to lipid metabolism, e.g. regulation of lipase activity, and lipoprotein particle remodelling (Additional file 7). Some of the genes present in the pathways related to lipid metabolism were *AKAP6*, *APOA1/4/5*, *ADCY2/9*, *ITPR2*, *PLA2R1*, *SCARB1*, and *OPHN1*. Also, we identified one pathway related to muscle differentiation with genes that were also present in the SNP enrichment analysis, such as *ACTC1*, *AKAP6*, *NOS1*, *MYOCD*, and *TTN*.

Thus, candidate genes for fat deposition from SNP-based selection signature analysis include *ITPR2* (GGA1), *ADCY2* (GGA2), *LPGAT1* (GGA3), *AKAP6* (GGA5), *IGFBP2* and *PLA2R1* (GGA7), and *CLPS* (GGA26). Candidate genes identified for muscle development are *ACTC1* and *AKAP6* (GGA5), *IGFBP2* (GGA7), *NOS1* (GGA15), *MYOCD* (GGA18), and *MAPK13/14* (GGA26).

All the above mentioned candidate genes for fat metabolism were detected by both SNP and INDEL-based analyses, with the addition of five more candidates detected only by INDEL analysis: *LPIN1* (GGA3), *ADCY9* (GGA14), and *APOA1/4/5* (GGA 24). For muscle development, the same candidate genes were identified in SNP and INDEL analyses, except the *MYOCD* gene,

detected only in the SNP analysis, and *IGF1R*, detected only in the INDEL analysis.

Furthermore, we also checked genes containing any intolerant SNP that overlapped with selection signature regions. Sixty-three intolerant SNPs were found within selection signature regions, and they represented 39 genes (Additional file 8). Some of these genes have relevant functions, such as related to growth, e.g. *AKAP6*, *NOS1*, *MAPK13*, and *CACNA1S*.

We also identified candidate genes within the putative signature selection regions for other important traits in chicken such as cardiac, skeletal and embryonic development, broodiness, energy metabolism, and reproduction (Table 4). These candidates are described in more details in the Discussion section.

Overlap of selection signatures with known QTLs

In order to achieve greater accuracy in the detection of the putative selection signatures, the merged regions (345 and 425 from SNP and INDEL analyses, respectively) were compared with known QTL regions in chicken according to the Animal QTLdb (release 29, $n = 5462$) [21]. About 97% of the selection signature regions overlapped with one or more QTLs for different traits.

Also, a statistical test based on permutation sampling was used to access the significance of the observed overlaps between the putative selection signature regions and QTLs. Based on the regions under selection from the SNP analysis, traits related to fat deposition, body weight, muscle weight, growth, egg production, bones, and others showed significant overlaps (permutation p -value < 0.05 , Fig. 6). When the INDEL-based analysis was considered, similar traits had significant overlaps (Fig. 7).

SNP and INDEL validation

SNPs were validated by comparison against two different datasets: (i) 15 M SNPs detected in a previous study by

Table 4 Genes associated with traits of interest in poultry located in putative signatures of selection regions

Chr	No of regions ^a	Selection signatures interval	No of SNP/INDEL	Mean Fst ^b	Genes	Main function related
1	4	67,980,001-68,030,000	738 SNPs	0.73	<i>ITPR2</i>	Lipid metabolism
	4		64 INDELs	0.72		
2	1	1,790,001-1,810,000	346 SNPs	0.67	<i>VIPR1</i>	Broodiness
	1	1,800,001-1,820,000	18 INDELs	0.66		
2	4	79,210,001-79,260,000	497 SNPs	0.78	<i>ADCY2</i>	Lipid metabolism
	3	79,210,001-79,250,000	38 INDELs	0.71		
2	6	83,150,001-83,220,000	629 SNPs	0.77	<i>FHOD3</i>	Cardiac development
	6	83,160,001-83,300,000	120 INDELs	0.71		
3	2	21,410,001-21,450,000	456 SNPs	0.70	<i>LPGAT1</i>	Lipid metabolism
	1	21,410,001-21,430,000	33 INDELs	0.67		
3	1	23,070,001-23,090,000	72 SNPs	0.81	<i>EML4, KCNG3</i>	Microtubule formation, heart rate regulation, insulin secretion and muscle contraction
	1		13 INDELs	0.77		
3	1	30,620,001-30,640,000	258 SNPs	0.72	<i>LTBP1</i>	Embryonic development
	1		36 INDELs	0.66		
3	1	92,370,001-92,390,000	159 SNPs	0.72	<i>TPO</i>	Growth and thyroid hormone synthesis
	3	92,340,001-92,400,000	81 INDELs	0.69		
3	1	97,030,001-97,050,000	26 INDELs	0.66	<i>LPIN1</i>	Muscle development and lipid metabolism
4	3	13,840,001-13,870,000	265 SNPs	0.69	<i>IRS4, COL4A5</i>	Growth and structural component
	1	13,850,001-13,870,000	21 INDELs	0.66		
4	3	13,920,001-13,960,000	711 SNPs	0.71	<i>COL4A5, 6</i>	Structural component
	3		66 INDELs	0.72		
5	2	10,640,001-10,670,000	33 INDELs	0.72	<i>SOX6</i>	Adipogenesis and skeletal development
5	17	30,500,001-30,690,000	1254 SNPs	0.75	<i>MEIS2</i>	Development of brain, heart, eyes, cartilage, and hematopoiesis
	16	30,490,001-30,670,000	204 INDELs	0.71		
5	2	31,540,001-31,570,000	324 SNPs	0.69	<i>ACTC1</i>	Muscle development
	1	31,540,001-31,560,000	32 INDELs	0.69		
5	8	34,070,001-34,300,000	827 SNPs	0.72	<i>AKAP6</i>	Lipid metabolism and muscle development
	8		124 INDELs	0.75		
6	1	14,950,001-14,970,000	206 SNPs	0.68	<i>VCL</i>	Muscle and heart development
7	1	14,650,001-14,670,000	127 SNPs	0.68	<i>TTN</i>	Muscle development
	1		25 INDELs	0.68		
7	1	21,330,001-21,350,000	209 SNPs	0.73	<i>PLA2R1</i>	Lipid metabolism
	1		18 INDELs	0.73		
7	2	22,790,001-22,820,000	580 SNPs	0.69	<i>IGFBP2</i>	Muscle development and lipid metabolism
	1	22,800,001-22,820,000	20 INDELs	0.71		
10	1	16,210,001-16,230,000	24 INDELs	0.70	<i>IGF1R</i>	Growth
14	1	12,640,001-12,660,000	19 INDELs	0.66	<i>ADCY9</i>	Lipid metabolism
15	1	11,300,001-11,320,000	154 SNPs	0.68	<i>NOS1</i>	Muscle development
	1		6 INDELs	0.68		
18	2	770,001-830,000	293 SNPs	0.69	<i>MYOCD</i>	Muscle and cardiac development
	3	780,001-830,000	28 INDELs	0.70		
24	1	5,100,001-5,120,000	5 INDELs	0.74	<i>APOA1, 4, 5</i>	Lipid metabolism
26	2	70,001-100,000	38 SNPs	0.84	<i>CLPS, LHFPL5, FKBP51</i>	Lipid metabolism
	2		11 INDELs	0.71		
26	6	120,001-190,000	183 SNPs	0.89	<i>MAPK14, SLC26A8, MAPK13, BRPF3</i>	Muscle development
	5		36 INDELs	0.82		
28	3	3,860,001-3,900,000	496 SNPs	0.72	<i>INSR</i>	Growth, energy metabolism, ageing and reproduction
	2	3,870,001-3,900,000	28 INDELs	0.68		

^aNumber of signature selection regions (windows) identified^bMean value of Fst considering all windows combined

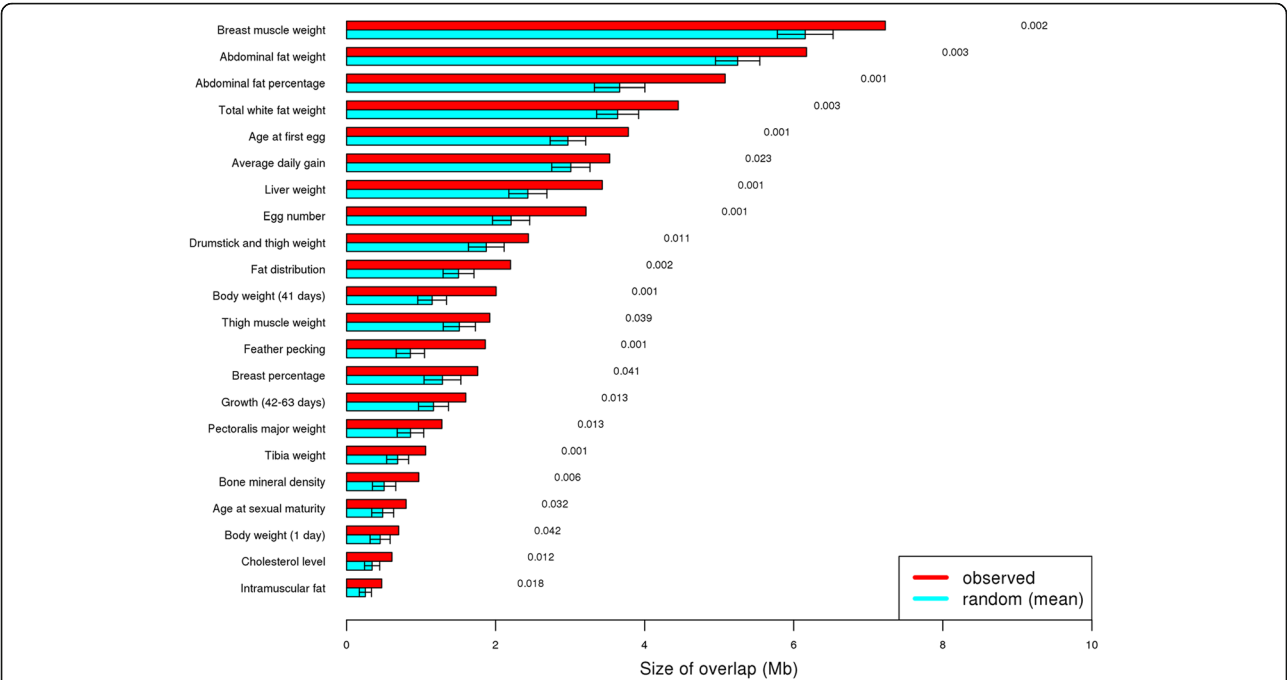


Fig. 6 Significant overlap of putative selection signature regions from SNP-based analysis with relevant chicken QTL-traits. Red bars indicate the observed overlap (selection signature regions overlapping with the respective trait), blue bars indicate the random overlap (1000 permutations), and the error bars indicate the standard deviation. The permutation *p*-values are listed on the right

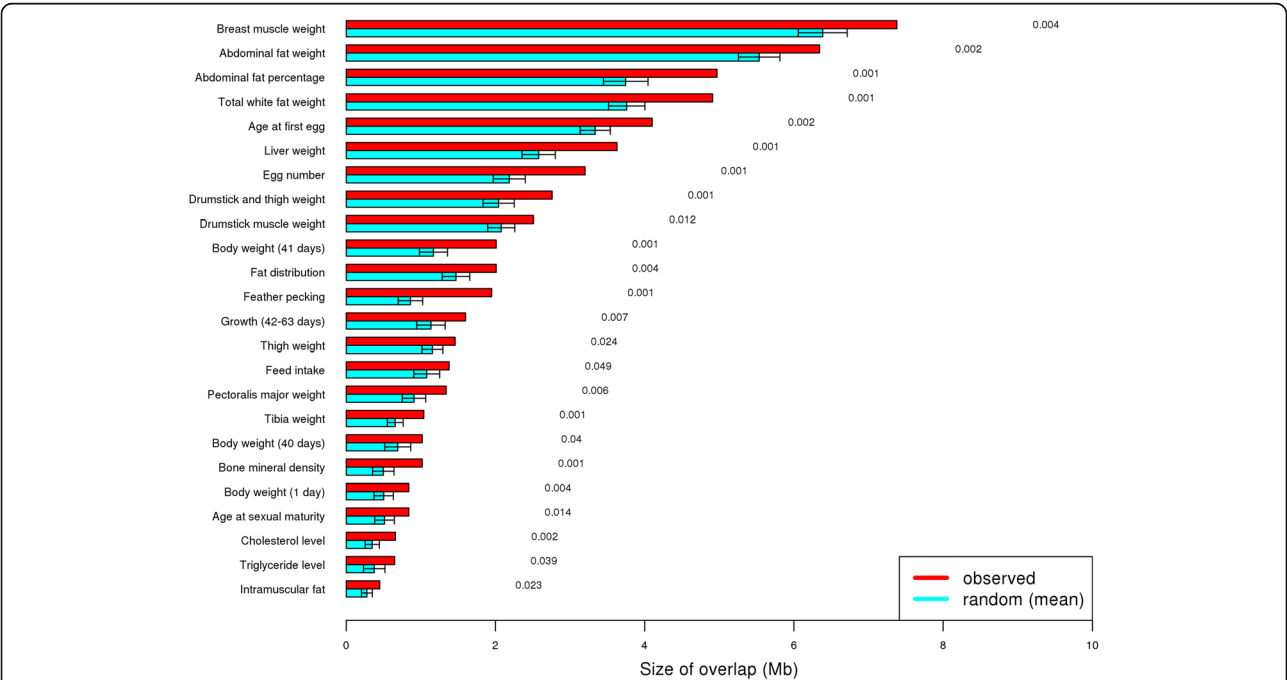


Fig. 7 Significant overlap of putative selection signature regions from INDEL-based analysis with relevant chicken QTL-traits. Red bars indicate the observed overlap (selection signature regions overlapping with the respective trait), blue bars indicate the random overlap (1000 permutations), and the error bars indicate the standard deviation. The permutation *p*-values are listed on the right

Gheyas et al. [7], and (ii) 600 K Affymetrix® Axiom® HD genotyping array data from two chickens (one male white-egg layer and one male broiler), which have also been sequenced in the present study. Comparison of the SNP catalog from Gheyas et al. [7] showed that 79.7% of our variants were common between the studies when only the genomic positions were compared, and 78.94% were common when both the position and allelic information were compared.

Comparison of the sequence genotype data against the 600 K SNP array dataset showed very high level of concordance (Additional file 9). For the two chickens used in this comparison, we could analyze genotypes of 575,599 SNPs for the layer (443,350 homozygous and 132,249 heterozygous) and 574,482 SNPs for the broiler (404,763 homozygous and 169,719 heterozygous), as these were present in the chip. Over 98% of the SNPs had the same genotypes between the NGS and array datasets (GEN-CONC), both for the broiler and the layer chicken. Slightly higher GEN-CONC was observed for homozygous SNPs than for heterozygous ones.

In addition, we randomly checked some of the discordant genotypes, and noticed that the majority of these had poor genotype quality (<30) in NGS even though their variant qualities were high. The error rate in SNP genotype call (considering the non-concordance of genotypes between NGS and array data for SNPs that could be compared) was only 1.51% for the layer chicken and 1.81% for the broiler.

We also validated a large number of functional variants - intolerant non-synonymous and stopgain/loss SNPs against the 600 K array data. These included 757 intolerant SNPs (380 from layer and 377 from broiler chickens), and 37 stopgain/loss SNPs (18 from layer and 19 from broiler chickens). Importantly, over 41% of the 63 intolerant or stopgain/loss variants detected in the selection signature regions were present in the 600 K genotyping array and could be validated.

The INDELs identified in this study ($n = 1,335,889$) were also validated by comparison with results from two major recent studies [8, 22]. We combined these two previous datasets, and based on positions only, we observed that 77.24% ($n = 1,031,932$) of the INDELs in our study were common with the previous studies. When the alleles were also compared along with positions, 62.62% of the INDELs were validated ($n = 836,552$).

Discussion

Polymorphism identification and functional annotation

A genome-wide SNP and INDEL identification in the Brazilian white-egg layer (CC) and broiler (TT) lines was performed to obtain a detailed map of genetic variation in these lines. The initial variant call resulted in

approximately 16 million SNPs and 2 million INDELs with an average density of 16.84 SNPs/kb and 1.72 INDELs/kb. Our results are consistent with recent studies, which reported similar densities in chicken [6–8, 23].

After filtration, a total of 13.93 million SNPs and 1.36 million INDELs were retained, with more variants detected from the broiler line, including heterozygous variants.

The transition and transversion (TS/TV) ratio for SNPs increased after the filtration process, showing that SNP filtration improved the TS/TV ratio, as higher ratios indicate better accuracy and lower false-positive rates [24]. In chicken, similar TS/TV ratios were reported [25].

Heterozygous variants are more difficult to be identified than homozygous ones. It has been estimated that a sequencing coverage of 6–10 X is sufficient to detect 99% of all variants (homozygous and heterozygous); however, some studies reported that a higher coverage (>20 X) is necessary for detecting 99% of heterozygous SNPs [26, 27].

A recent study [28] analyzing the same Brazilian layer and broiler lines using reduced-representation-sequencing also reported a higher proportion of heterozygous SNPs in the broiler line. The greater variability and heterozygosity observed in the broiler line was likely due to its broad genetic background, which is similar to other commercial chickens and includes the White Plymouth Rock, New Hampshire and White Cornish breeds [15]. On the other hand, the layer line utilized in this study is a white-egg laying hen, which originated solely from the White Leghorn breed [15].

Even though there are millions of SNPs and INDELs along the genome, only a few are expected to explain the phenotypic differences observed between the two lines. As expected, annotation of the variants showed that most of them belonged to non-coding regions of the genome, and just 1.2% of the SNPs and 0.2% of the INDELs were located in exonic regions, including 122,502 non-synonymous SNPs. Non-synonymous SNPs and SNPs located in regulatory regions are known to have the highest phenotype impact [29].

Although, most of the variants were located in non-coding regions, we identified 7255 intolerant non-synonymous SNPs, 512 stopgain/loss SNPs, 1381 frameshift and 1094 non-frameshift INDELs that may alter protein functions.

Both frameshift and non-frameshift INDELs affect protein sequence by being located within coding regions. Frameshift mutations, however, can also have a major deleterious impact on protein function as these alter the amino-acid reading frame. On the other hand, non-frameshift INDELs do not change the reading frame, but

they can still alter the protein function by inserting or deleting one or more amino-acids in the protein sequence. The impact of INDELs on transcription has not been well investigated, but it has been suggested that this may be considerable [30]. Non-coding INDELs, e.g. located in 1 kb down/upstream regions, can affect gene regulation by modulating transcription or silencing genes [30].

The chicken genome is still poorly annotated for functional non-coding regions [31], which can be involved in important biological functions as was shown in humans [32]. Because of this, it is currently difficult to predict the potential effect of variants located in non-coding regions, particularly in potentially important regions e.g. UTR, up/downstream, splicing and ncRNA.

Metabolic pathway analyses of polymorphisms of potential functional impact

In order to gain insights into how the polymorphisms identified could explain the phenotypic changes observed in the two Brazilian chicken lines, metabolic pathways of genes harboring variants of high potential functional effects were investigated.

The analysis of all genes containing intolerant SNPs from the white-egg layer and the broiler lines combined resulted in eight networks related mainly to connective and metabolic disorders, embryonic development, cardiovascular diseases, and carbohydrate metabolism.

The effects of the potential disruptions of these important pathways caused by intolerant SNPs may explain various disorders observed in commercial broiler and layer chickens. Broiler chickens, for instance, have been selected for weight gain, feed efficiency, breast and carcass yields, which results in a substantial increase of growth in a short period. This particular scenario for meat-type chickens can result in heart ventricular hypertrophy and overload of the cardiopulmonary system, leading to pulmonary hypertension syndrome (ascites), which can cause chicken mortality or whole carcass condemnation during slaughter [33].

On the other hand, layer chickens, which have been mainly selected for high egg production, have a reduced body weight than the broiler chickens and can develop diseases such as fatty liver hemorrhagic syndrome (FLS) and osteoporosis [3]. FLS is a result of excessive accumulation of fat in the liver when lipoprotein transport is disrupted during high egg production and cause haemorrhage and sudden death [34].

When the analysis was performed using genes harboring intolerant SNPs specific to a line, clearly different networks were obtained for each line. In the layer line, networks were identified mainly related to reproductive system development, cellular functions, and the endocrine system. The latter in layers is particularly

important for reproductive aspects, which involve different hormones and their receptors e.g. growth hormone (GH), prolactin (PRL), thyroids, luteinizing hormone (LH), melatonin, and others [35, 36]. For instance, GH is a regulator of ovarian and oviductal functions in chicken, acting in the modulation of LH and the synthesis of insulin-like growth factors [35]. Also, prolactin is involved with different physiological processes that are important for poultry reproduction; hence, *PRL* is a candidate gene for egg production in chicken [36].

In the broiler line, networks were obtained mainly related to lipid and carbohydrate metabolism, metabolic and dermatologic diseases, neural development, and organism injury. It is well known that commercial broiler lines gain much more weight and muscle mass than layer chickens. It was observed that at 41 days of age, the Brazilian broiler chicken line weighed on average 2395 g, while the layer chicken line weighed only 513 g, when reared as broilers, resulting in about a five-fold difference between these two lines. Also, the breast yield for the broiler line was 6% higher compared to the layer line [15, 16]. This difference in muscle development and growth between broilers and layers can be observed in the early stages of chicken embryos as previously reported [3].

Another important difference between these two lines is lipid metabolism and fat deposition. Broiler chickens accumulate more fat than layers, and lipid metabolism differences were detected during chicken embryogenesis, e.g. higher triglyceride content in the liver of broilers [3]. It was observed that at 41 days of age, the Brazilian broiler chicken line presented 2.41% of abdominal fat percentage, while the layer chicken line displayed 0.16% [16]. However, so far, the differences between broiler and layer chicken lines regarding adipose growth are not well understood [37].

Frameshift and non-frameshift INDELs are very important as well, because they are located in coding regions, and both can affect protein sequence. The metabolic pathways obtained from the analysis of coding INDELs showed pathways mainly associated with diseases and disorders. Many diseases in humans have been found to be associated with INDELs (e.g. different types of cancer), and it was estimated that INDELs are responsible for 24% of the inherited diseases in humans [38]. Previous studies in chicken have also found INDELs to be associated with various health-related problems, such as retinal degeneration and embryonic mortality [39], production traits, such as egg production [40], and performance and carcass traits [41]. The frameshift and non-frameshift INDELs identified in the present study could also be involved with health problems, but further association studies need to be performed to elucidate their effects.

Genome-wide scan for selection signatures

Intensive artificial selection can cause high degree of genetic differentiation between populations in specific genomic regions, which can result in selection footprints. There are different methods and tools available to detect evidence of selection. Traditionally, the *Fst* method has been a suitable choice based on allele frequency between populations [5].

Fst values were obtained from SNP and INDEL datasets separately using an overlapping sliding window of 20 kb, which was also utilized in previous studies in chicken [42, 43]. Although some studies used 40 kb windows [4, 7, 44], we decided to use 20 kb size-windows because previous studies reported that haplotype blocks in chicken have around 10 kb. The haplotype blocks described in chicken have different sizes depending on the line analyzed. In one study, haplotype block size was different in commercial and non-commercial broiler and layer chickens, but the length was typically less than 10 kb [42]. Also, traditional and village chickens were evaluated, and a median block size of 11–12 kb was observed [43]. Another reason to use 20 kb windows was to obtain a better resolution of the regions combined with a sufficient number of variants in each window.

We obtained a sufficient number of markers per window with an average of 268 SNPs/window and 26 INDELs/window. In similar studies in chicken, Qanbari et al. [44] obtained a mean of 199 SNPs per 30–40 kb window; Stainton et al. [10] had < 30 SNPs per window, and Gholami et al. [11] analyzed windows of 40 SNPs. Detecting regions under selection with *Fst* methods requires at least 10 samples [45]. Moreover, presence of a large number of markers significantly increases the power of the analysis [45]. Even though we only analyzed 14 individuals per line, our method is robust as NGS data were used with dense marker sets.

A majority of the windows had low *Fst* values (< 0.3), although a few windows showed extremely high *Fst* values (> 0.9). Our results corroborate with similar *Fst*-based studies in chicken. For instance, Gholami et al. [11] obtained *Fst* values using the Wright method in commercial layer chickens and observed average values, depending on the breed, between 0.09 and 0.27. Stainton et al. [10], like in our study, utilized the Weir and Cockerham's pairwise *Fst* estimator and obtained mean values between 0.015 and 0.17 depending on the broiler line analyzed, reflecting the fact that each broiler line had slightly different selection criteria which resulted in a range of broiler lines with different characteristics.

Separate *Fst* analyses, with SNPs and INDELs, were performed to check if different putative signature regions would be detected. Most of the signature regions identified were found to be common between the two analyses (SNPs and INDELs), with a high level of

correlation. This shows that even though the INDEL dataset was only ~ 10% of the SNP dataset in size, both datasets detected practically the same regions under selection. Our study is the first in chicken to use INDEL variants to detect footprints of selection. There are, however, a few studies in other species, for example in humans [46]. In this study, INDELs have been used for identification of selection signatures, and they suggested that regions surrounding INDELs are more frequently involved in recent selective sweeps [46]. We believe that INDEL variants are also a good option for selection signatures analyses even with a reduced density than SNPs, but further studies are necessary to evaluate it in more detail.

Genetic drift is a well-known factor that can also cause divergences in genomic regions. It is almost impossible to determine if a putative selection signature was caused by drift [47]. Besides, a combination of genetic drift and selection could be responsible for eliciting signals similar to selection footprints [48]. Despite that, there are different ways to minimize the false discovery of selective sweeps, such as the use of a stringent *Fst* cut-off or checking the overlap of those regions with QTLs [47, 49]. In our study, different steps were utilized to minimize the detection of false positive regions due to drift: (i) stringent *Fst* cut-off was applied (> 99% *Fst*); (ii) windows with less than 10 SNPs or 5 INDELs were excluded and weighted *Fst* values were used; (iii) putative selection signature regions were overlapped with known QTL regions from chicken QTLdb and permutation tests were applied to check the significance of these overlaps (results presented and discussed in sections below); (iv) putative selection signature regions were further interpreted based on different downstream analyses such as annotation and pathway enrichment analysis of genes within the signature regions (presented and discussed in sections below); and (v) *Fst* analyses were performed utilizing SNP and INDEL datasets separately to obtain more accurate results.

Candidate genes under selection for fat deposition and muscle development

We analyzed all genes under putative selection using network and gene enrichment analyses, and gene function by literature search, and we identified candidate genes for fat deposition and muscle development. There were 307 from SNP-based analysis and 363 from INDEL-based analysis. Candidate genes associated with fat deposition were *ADCY2/9*, *APOA1/4/5*, *AKAP6*, *CLPS*, *IGFBP2*, *ITPR2*, *LPGAT1*, *LPIN1*, and *PLA2R1*. Candidate genes associated with muscle development were *ACTC1*, *AKAP6*, *IGFBP2*, *IGF1R*, *NOS1*, *MAPK13/14* and *MYOCD*.

The *AKAP6* gene is present in both muscle development and lipid metabolism pathways. A recent study suggests that this gene is an important regulator of muscle regeneration and myoblast differentiation in mice [50], in addition to its role in cardiac functions [51]. This gene seems to play a major role in the regulation of different processes, and further studies in chicken are necessary to investigate its role in greater details.

The 30 kb selective sweep region (GGA7:22,790,001–22,820,000), covering the gene *IGFBP2* identified in our study (with variants mainly fixed in the layer line), was previously reported in a comparison of various layer chickens [11]. This gene is known to be related to growth [52] and fat deposition [53] from previous studies in chicken.

A putative selection signature region of 20 kb size was identified (GGA10: 16,210,001–16,230,000) with 24 INDELs overlapping the gene *IGF1R* with variants mainly fixed in the layer line. Recent studies in chicken, using the same method described here, also identified the same gene under selection in a commercial broiler [10] and layer [12] lines. This gene plays an important role in chicken growth [54] and may also be related to reproduction in layers [55].

One selective sweep of 40 kb (GGA28: 3,860,001–3,900,000) overlapping the *insulin receptor* gene (*INSR*) was identified in our study with variants mainly fixed in the broiler line (Table 4). This region has been previously reported in studies with layers [44], broilers [4], and both broilers and layers [7]. Also, *INSR* was associated with growth in chicken in a previous study [56].

Candidate genes under selection for other important traits in poultry

Apart from fat deposition and muscle development traits, genes within the putative signature selection regions were found to be candidates for other important traits in chicken, such as cardiac, skeletal and embryonic development, broodiness, energy metabolism, and reproduction. For instance, the *MEIS2* gene is related to the development of brain, heart, eyes, cartilage and hematopoiesis [57]. In our study, 17 adjacent windows were detected as putative selection signatures (GGA5: 30,490,001–30,690,000) harboring the *MEIS2* gene, which were merged into one large region of 200 kb, possibly indicating a strong selection region. In the study of Gheyas et al. [7], they identified broiler-specific SNPs in the *MEIS2* gene.

In addition, two genes, *ITPR2* and *VIPRI*, were identified in the putative selection signature regions, which are possibly related to layer reproductive traits. One recent GWAS study in chicken revealed that the *ITPR2* gene was

associated with eggshell ultrastructure [58]. On the other hand, the *VIPRI* gene was found to be associated with broodiness [59], and egg number in chicken [60].

Certain regions or genes are known to be selected in domesticated chickens, such as *TSHR* (affecting reproductive behaviors) and the *BCDO2* (for yellow skin) as have been discussed in a number of previous studies [4, 7, 44]. These genes, however, were not picked up in our study as we applied Fst-based approach for detection of selection footprints between meat-type (broiler) and white-egg layer lines. However, both regions were checked in more detail in both lines. In the *TSHR* and *BCDO2* genes, there were 812 and 187 SNPs, respectively and 57 and 23 INDELs, respectively. Most of these variants were fixed (homozygous) in both lines, especially in the *BCDO2* gene, showing that these selection signatures are clearly present in both chicken lines.

Overlap of selection signatures with known QTLs

In order to gain confidence in detecting selection signatures and linking these regions to potential phenotypes, we compared the identified selection signature regions with known QTL regions according to the Animal QTLdb [21].

Approximately 97% of the selection signature regions overlapped with one or more QTLs for different traits. This provides an independent support for the validity of the selection signature regions detected in this study. These QTLs corresponded to 190 traits related to fat deposition, growth, carcass, egg production and quality, bones, blood parameters, and organs, among others.

In addition, a statistical test was performed to check if overlaps of the selection signature regions with QTLs were significant and not just by chance. This comparison revealed that a higher proportion of the putative signature regions identified in this study were located in QTL regions associated with relevant traits in chicken, and should be further investigated.

The results showed that almost all the selection signature regions overlapped with different QTLs, providing additional confidence of the putative selection regions identified, and also may be an indication of genetic correlation between traits due to pleiotropy.

SNP and INDEL validation

Availability of variant information from previous NGS studies and high throughput genotype data from 600 K arrays offer an excellent opportunity to carry out cross-study and cross-platform (NGS vs array) comparisons to achieve greater confidence in the variant and genotype calls. We took this opportunity to validate SNPs and INDELs detected in the present study.

SNPs were validated by comparing NGS data against two datasets: 15 M SNPs detected in a previous study by

Gheyas et al. [7] and SNP data from 600 K Affymetrix® Axiom® HD genotyping array from two chickens.

We validated more than 78% of our SNPs by comparison with the catalog from Gheyas et al. [7]. This provides a strong validation of the SNPs detected in our study.

In addition, we observed high levels of concordance with 600 K SNP array data. We noticed a slightly higher percentage of genotype concordance for homozygous SNPs than for heterozygous ones, which were also observed previously in a study with bovine variants [61].

The majority of the discordant genotypes checked had poor genotype quality in our NGS data even though their variant qualities were high. A possible explanation of this could be that the sequencing coverage was not sufficient in these cases for accurate prediction of the genotype.

We observed low error rates in our study (considering the non-concordance of genotypes between NGS and array data for SNPs), showing that alignment and SNP calling were efficient.

The 600 K array allowed us to validate a large number of functional variants, intolerant non-synonymous and stopgain/loss SNPs, thereby providing confidence in their detection.

Moreover, INDELs identified in this study were also validated by comparison with two major studies [8, 22], which have generated the largest catalogs of INDELs in chicken. We observed that 77% of the INDELs reported in this study were common with the two previous studies evaluated, and 62% had the same alleles. Similar results were observed before based on the same type of comparison [8]. The lower percentage of concordance, when allele information was considered, is an indication that the INDEL alleles differ among different chicken populations, probably because a large proportion of INDELs actually consists of microsatellite like tandem repeats. This comparison, with the two major INDEL studies in chicken [8, 22], shows that the methodologies used in this study were efficient for detecting and filtering INDELs, and ~ 23% of novel INDELs were detected.

Conclusions

In this study, we performed a genome-wide identification and characterization of genetic variations, and found putative genomic footprints of selection from two different chicken lines, a white egg layer and a meat-type (broiler) line, which have been under multi-trait selection in Brazilian climatic and nutritional conditions for more than 20 years. Approximately 15 million genetic variants were identified. Of which, over 10 thousand are expected to alter protein functions. Important pathways related to lipid metabolism, growth, and reproductive traits were detected from the analysis of SNPs and INDELs, especially the line-specific variants.

Fst-based analyses were used to identify putative regions of selection showing population differentiation. This approach revealed a number of highly plausible candidate genes and pathways under selection for fat metabolism, growth, reproduction, and cardiac development. Moreover, this is the first study to utilize genome-wide INDEL variants in chicken to identify selection signatures, showing high level of correlation in results between the SNP and INDEL based data.

To minimize the effect of genetic drift, different approaches were applied to identify high confidence regions of putative selection. However, despite the effort to avoid detecting regions caused by random drift, some of the regions detected in this study may represent false positives, especially the ones that did not overlap with genes having known phenotypic association with QTLs. Further analysis in different populations would be required to confirm the accuracy of those regions.

In summary, this study provides novel insights into selection footprints that can help elucidate the functional mechanisms underlying important phenotypic traits relevant to broiler and layer chicken lines. In addition, we have presented a detailed genetic catalog of variants both line-specific and common from two Brazilian broiler and layer chicken lines that can be used for future genomic studies involving association analysis with relevant phenotypes, and consequently, facilitate the identification of causative mutations in chicken, and an important resource for marker-assisted or genomic selection for important traits in chicken.

Methods

Genome re-sequencing of experimental chicken lines

Twenty eight chickens from two different experimental lines with pedigree control and multi-trait selection by Embrapa Swine and Poultry National Research Center (Brazil) were sequenced: 14 individuals from a paternal broiler line called TT (7 females and 7 males), and 14 from a layer line called CC (7 females and 7 males). Previously, these two experimental lines were used to generate a reciprocal F₂ Resource population for QTL mapping studies [13, 14].

The broiler line (TT), originating from the White Plymouth Rock, New Hampshire and White Cornish breeds, has been selected since 1992 for body weight, feed conversion, carcass and parts yields, chick viability, fertility, hatchability of fertile eggs, and reduced fat deposition and metabolic disorders [15]. On the other hand, the white-egg layer line (CC), originating from the White Leghorn, has been selected since 1989 for egg production, egg weight and quality, chick viability, feed conversion, sexual maturity, fertility, hatchability of fertile eggs, and decreased body weight [15]. Chickens were selected based on their performance for the traits

mentioned above, and each line was raised under specific feeding regimes [16].

Each chicken was sequenced individually using a paired-end protocol in the HiSeq2500 sequencer (Illumina) with paired read length of 101 bases. Further details on library preparation and sequencing can be obtained from Moreira et al. [23].

SNP and INDEL identification and filtration

First, the quality of the sequencing reads was checked with FastQC tool [62]. Read quality trimming was performed using SeqClean tool (v.1.3.12) [63] to select reads with average Phred score quality ≥ 24 and minimum length of 65 bp. The reads were aligned against the *Gallus_gallus*-4.0 chicken reference genome [64] using Bowtie2 v.2.1.0 [65]. PCR duplicates were removed using Picard [66] (v.1.112). Then, SNP and INDEL identification was performed with SAMtools v.1.2 [67] using the *mpileup* option with mapping and base qualities (Phred score) ≥ 20 . All 28 sample BAM files were analyzed together to improve the variant calling in low coverage regions. After the initial calling, different filtration criteria were applied to reduce the number of false-positives and to avoid copy number variation regions. The filtration criteria included (1) Phred-based variant quality score of at least 40; (2) minimum depth of coverage at the variant site of 5; (3) maximum depth of coverage of not more than mean coverage plus 3 standard deviations; (4) variant supported by both forward and reverse strands (at least one read on the forward strand and one on the reverse strand); (5) variant supported by at least 3 reads, and (6) SNP clusters (> 10 SNPs in 50 bp), and INDEL clusters (gap of 1 bp between two INDELs) were removed.

A SNP/INDEL was referred to as homozygous when only a non-reference allele was observed and heterozygous when both the reference and non-reference alleles were observed. The alternate allele frequencies (AAF) of SNPs/INDELs were calculated with VCFtools v. 0.1.12 [68], and AAF was estimated based on the number of times an alternate allele (SNP or INDEL) appears over all individuals at that site, divided by the total number of non-missing alleles at that site.

Functional annotation and effect prediction

Sets of unique SNPs and INDELs from the 28 chickens were annotated using Variant Effect Predictor tool v.75 [69] against the gene annotation database from Ensembl (release 71) for chicken. For SNP prediction, we also used the SIFT option from VEP tool [18], which predicts functional effects of SNPs and classifies them either as intolerant (affects protein function) or tolerant (functionally

neutral) based on amino acid properties and sequence homology (degree of evolutionary conservation).

Metabolic pathway analyses

QIAGEN's Ingenuity® Pathway Analysis (IPA®) software [19] was used with default parameters to find metabolic pathways of genes with relevant biological functions from SNPs and INDELs. First, all genes with putative functional coding variants present in both lines were analyzed together to find a global result. Then, variants, exclusively from broilers or layers, were analyzed separately to find metabolic pathways affected in a particular line. Different sets of variants were utilized, viz. genes containing (1) intolerant SNPs (3746 unique genes); (2) intolerant SNPs exclusively from broilers (1825 unique genes) or layers (1057 unique genes); (3) frameshift and non-frameshift INDELs (1462 unique genes), and (4) frameshift and non-frameshift INDELs exclusively from broilers (406 unique genes) or layers (182 unique genes). IPA computes a p-score (network score) derived from *p*-values and calculated using Fisher's exact test. A network score of > 3 (*p*-value < 0.001) indicates a $> 99.9\%$ confidence that a network is not generated by random chance; however, a network score ≥ 30 was considered for determining significant results. Similar cut-off scores were utilized in previous studies to ensure that only highly significant networks were identified [70, 71].

Genome-wide scan for selection signatures

The Weir and Cockerham pairwise *F_{st}* (Fixation index) method [72] was applied to estimate the genome-wide genetic differentiation between broiler and layer lines. Chromosomes W/Z, unplaced, random, and mitochondrial were not considered in this study. This method was performed using VCFtools v. 0.1.12 software [69] with SNP ($n = 12,806,643$) and INDEL ($n = 1,273,210$) datasets considered separately and using overlapping windows of 20 kb and a step size of 10 kb. Weighted *F_{st}* values were calculated for windows with at least 10 SNPs or 5 INDELs. Windows with the top 1% *F_{st}* values were considered as candidates of selection signatures, while the extreme top 0.1% windows were considered candidates with strong evidence of selection. The selection signature regions obtained were annotated against the chicken gene database from Ensembl (release 84). All genes from regions under selection (top 1%) were further analyzed to predict the function of these genes with GeneMANIA prediction server [20] considering human database (chicken not available). Gene enrichment analysis was performed considering Q-values calculated from a FDR test, and pathways with Q-values < 0.1 were considered significant.

QTL overlap analysis

QTL overlap analysis was carried out using the *regi-oneR* package [73]. From 190 traits, a subset of 62 representing the most relevant in poultry were selected for this analysis. A permutation test was performed to evaluate the significant associations between the genomic regions (merged signature regions) and the QTL regions by random permutations ($n = 1000$). In every permutation, the overlap with the QTLs was recomputed based on the total genomic size (Mb) that was overlapped. The observed number of overlaps was compared with the empirical distribution to obtain the p -values. A p -value < 0.05 was considered to determine significant associations.

SNP and INDEL validation

The SNPs and INDELs detected in the present study were compared with variants detected from different populations in previous NGS studies as a means of validation. The filtered SNP set was compared with the 15 M SNPs detected in Gheyas et al. [7], which had generated the largest catalog of SNPs in chicken. Similarly, the filtered INDEL set was compared against the combined dataset from two major recent studies: 883,840 INDELs from Boschiero et al. [8] and 1,343,782 INDELs from Yan et al. [22]. Comparisons were made based on both positions and allelic information.

The accuracy of the NGS-based SNP genotype calls was investigated by comparing sequence data from two chickens against the genotype calls from 600 K Affymetrix® Axiom® HD genotyping array from the same chickens, which have been previously genotyped with the array. The chickens chosen for this analysis had intermediate sequencing coverage of ~ 11 X/chicken.

From the NGS dataset, we compared 13,218,246 SNPs with 11.2 X coverage from the layer and 13,229,324 SNPs with 11.5 X average coverage from the broiler chicken. From the chip dataset, the total of 575,599 (layer) and 574,482 (broiler) SNPs were used, excluding the SNPs located on the linkage groups. In both datasets, the chromosomes analyzed were GGA1–28 and sex W/Z. Both heterozygous and homozygous genotypes were considered for comparison of the two datasets. The comparison was performed based on a methodology previously proposed [61] by calculating two types of concordance: (i) POS-CONC - SNPs with same position in both datasets, and (ii) GEN-CONC- only POS-CONC SNPs with same genotypes between the two datasets. The error rate in genotype call from NGS data was estimated as the percentage of POS-CONC SNPs which had non-concordant genotypes between NGS and array data, i.e. error rate $\% = [100 - (\text{number of GEN-CONC SNPs} / \text{number of POS-CONC SNPs}) * 100]$.

Additional files

Additional file 1: A figure with the substitution types of SNPs initially identified from the 28 chickens (broiler and layer chickens combined). (PDF 1108 kb)

Additional file 2: A figure with the proportion of heterozygous and homozygous SNPs (A) and INDELs (B) observed in each individual of layer and broiler lines. (PDF 7169 kb)

Additional file 3: An excel file with the top nine networks from frameshift and non-frameshift INDELs from the layer and broiler lines combined. QIAGEN's Ingenuity® Pathway Analysis (IPA®) software (<http://www.ingenuity.com/>) with default parameters were used to find metabolic pathways of genes with relevant biological functions. (XLSX 10 kb)

Additional file 4: An excel file with the top six networks from frameshift and non-frameshift INDELs exclusively from the layer or broiler line. QIAGEN's Ingenuity® Pathway Analysis (IPA®) software (<http://www.ingenuity.com/>) with default parameters were used to find metabolic pathways of genes with relevant biological functions. (XLSX 10 kb)

Additional file 5: An excel file with the common putative genes under selection from SNP and INDEL based analyses. (XLSX 31 kb)

Additional file 6: An excel file with enrichment analysis results including the enriched GO terms from genes located in top 1% regions under selection (SNP dataset). The enrichment analysis was performed with GeneMANIA prediction server considering human database. (XLSX 9 kb)

Additional file 7: An excel file with enrichment analysis results including the enriched GO terms from genes located in top 1% regions under selection (INDEL dataset). The enrichment analysis was performed with GeneMANIA prediction server considering human database. (XLSX 9 kb)

Additional file 8: An excel file with the intolerant SNPs located within genes under selection (top 1%). (XLSX 11 kb)

Additional file 9: An excel file with the SNP validation comparison results showing the number and percentages of SNPs with same position (POS-CONC) and genotype (GEN-CONC) from broiler and layer chickens of NGS and chip dataset. (XLSX 9 kb)

Abbreviations

AAF: Alternate Allele Frequency; FLS: Fatty liver hemorrhagic syndrome; GEN-CONC: SNPs with same genotypes; GGA: *Gallus gallus*; GH: Growth hormone; GWAS: Genome-Wide Association Study; INDEL: Insertions and Deletions; IPA: Ingenuity Pathway Analysis; LH: Luteinizing hormone; NGS: Next Generation Sequencing; PCR: Polymerase Chain Reaction; POS-CONC: SNPs with same positions; PRL: Prolactin; QTL: Quantitative Trait Locus; SNP: Single Nucleotide Polymorphism; TS/TV: Transition/Transversion

Acknowledgements

The authors are grateful to Prof. Sônia CS Andrade (USP), Prof. Gabriel RA Margarido (USP) and Choon-Kiat Khoo (The Roslin Institute) for helpful suggestions.

Funding

CB received a fellowship from the program Science Without Borders - National Council for Scientific and Technological Development (CNPq, grant 370620/2013-5). GCMM and TFG received fellowships from São Paulo Research Foundation (FAPESP, grants 14/21380-9 and 15/00616-7). LLC is recipient of productivity fellowship from CNPq. This project was funded by São Paulo Research Foundation (FAPESP) - thematic project (2014/08704-0).

Availability of data and materials

All SNPs and INDELs were submitted to dbSNP (NCBI) with the submitter handle "LBA_ESALQ" (https://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1062434).

Authors' contributions

CB designed the study, analyzed and interpreted the data, and wrote the manuscript with the valuable help of AAG. GCMM performed the SNP validation. AAG helped with the data interpretation, analyses, and manuscript

review. TFG submitted all variants to NCBI. GCMM and TFG performed the genotyping. GG, MP and PDSCM performed the sequencing. ASMC performed the QTL permutation analysis. MCL provided the biological material, helped in the interpretation of results and reviewed the manuscript. LLC designed the study, reviewed the manuscript and provided overall supervision. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All animal experimental protocols were carried out in accordance with resolution number 010/2012 approved by the Embrapa Swine and Poultry Ethics Committee on Animal Utilization to ensure compliance with international guidelines for animal welfare.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Animal Biotechnology Laboratory, Animal Science Department, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, SP 13418-900, Brazil. ²Department of Genetics and Genomics, The Roslin Institute and Royal School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. ³Embrapa Suínos e Aves, Concórdia, SC 89715-899, Brazil. ⁴Noble Research Institute, 2510 Sam Noble Parkway, Ardmore, Oklahoma 73401, USA.

Received: 2 April 2017 Accepted: 10 January 2018

Published online: 25 January 2018

References

- Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 2004;432(7018):695–716.
- Hocking PM. Developments in poultry genetic research 1960–2009. *Br Poult Sci*. 2010;51:44–51.
- Buzala M, Janicki B, Czarnecki R. Consequences of different growth rates in broiler breeder and layer hens on embryogenesis, metabolism and metabolic rate: a review. *Poult Sci*. 2015;94:728–33.
- Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010;464(7288):587–U145.
- Cadzow M, Boocock J, Nguyen HT, Wilcox P, Merriman TR, Black MA. A bioinformatics workflow for detecting signatures of selection in genomic data. *Front Genet*. 2014;5:293.
- Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*. 2013;14:59.
- Gheyas AA, Boschiero C, Eory L, Ralph H, Kuo R, Woolliams JA, et al. Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Res*. 2015;22(3):205–17.
- Boschiero C, Gheyas AA, Ralph HK, Eory L, Paton B, Kuo R, et al. Detection and characterization of small insertion and deletion genetic variants in modern layer chicken genomes. *BMC Genomics*. 2015;16:562.
- Zhang H, Hu X, Wang Z, Zhang Y, Wang S, Wang N, et al. Selection signature analysis implicates the PC1/PCSK1 region for chicken abdominal fat content. *PLoS One*. 2012;7(7):e40736.
- Stainton JJ, Haley CS, Charlesworth B, Kranis A, Watson K, Wiener P. Detecting signatures of selection in nine distinct lines of broiler chickens. *Anim Genet*. 2015;46:37–49.
- Gholami M, Erbe M, Gärke C, Preisinger R, Weigend A, Weigend S, et al. Population genomic analyses based on 1 million SNPs in commercial egg layers. *PLoS One*. 2014;9:e94509.
- Gholami M, Reimer C, Erbe M, Preisinger R, Weigend A, Weigend S, et al. Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information. *PLoS One*. 2015;10:e0130497.
- Nones K, Ledur MC, Ruy DC, Baron EE, Melo CM, Moura AS, et al. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. *Anim Genet*. 2006;37:95–100.
- Moura AS, Ledur MC, Boschiero C, Nones K, Pinto LF, Jaenisch FR, et al. Quantitative trait loci with sex-specific effects for internal organs weights and hematocrit value in a broiler-layer cross. *J Appl Genet*. 2016;57(2):215–24.
- Jorge EC, Figueira A, Ledur MC, Moura ASAMT, Coutinho LL. Contributions and perspectives of chicken genomics in Brazil: from biological model to export commodity. *Worlds Poul Sci J*. 2007;63(04):597–610.
- Ledur MC, Peixoto JO, Nones K, Coutinho LL. XXIV World's poultry congress. Salvador; 2012. <http://www.facta.org.br/wpc2012-cd/papers/>. Accessed 15 Aug 2017.
- Venturini GC, Savegnago RP, Nunes BN, Ledur MC, Schmidt GS, El Faro L, Munari DP. Genetic parameters and principal component analysis for egg production from White Leghorn hens. *Poult Sci*. 2013;92(9):2283–9.
- Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–4.
- QIAGEN's Ingenuity® Pathway software. 2017. <http://www.ingenuity.com/>. Accessed 12 Nov 2016.
- Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucl Acids Res*. 2010;38:W214–20.
- Animal Quantitative Trait Loci Database (Animal QTLdb). 2016. <http://www.animalgenome.org/cgi-bin/QTLdb/index>. Accessed 24 Oct 2016.
- Yan Y, Yi G, Sun C, Qu L, Yang N. Genome-wide characterization of insertion and deletion variation in chicken using next generation sequencing. *PLoS One*. 2014;9(8):e104652.
- Moreira GC, Godoy TF, Boschiero C, Gheyas A, Gasparin G, Andrade SC, et al. Variant discovery in a QTL region on chromosome 3 associated with fatness in chickens. *Anim Genet*. 2015;46:141–7.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–8.
- Oh D, Son B, Mun S, Oh MH, Oh S, Ha J, et al. Whole genome re-sequencing of three domesticated chicken breeds. *Zool Sci*. 2016;33:73–7.
- Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 2008;18:1851–8.
- Sims D, Sudbery I, Iltott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*. 2014;15:121–32.
- Pértille F, Guerrero-Bosagna C, Silva VH, Boschiero C, Nunes JR, Ledur MC, et al. High-throughput and cost-effective chicken genotyping using next-generation sequencing. *Sci Rep*. 2016;6:26929.
- Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30:3894–900.
- Haraksingh RR, Snyder MP. Impacts of variation in the human genome on gene regulation. *J Mol Biol*. 2013;425:3970–7.
- Schmid M, Smith J, Burt DW, Aken BL, Antin PB, Archibald AL, et al. Third report on chicken genes and chromosomes 2015. *Cytogenet Genome Res*. 2015;145(2):78–179.
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
- Wideman RF, Chapman ME, Hamal KR, Bowen OT, Lorenzoni AG, Erf GF, Anthony NB. An inadequate pulmonary vascular capacity and susceptibility to pulmonary arterial hypertension in broilers. *Poult Sci*. 2007;86(5):984–98.
- Trott KA, Giannitti F, Rimoldi G, Hill A, Woods L, Barr B, et al. Fatty liver hemorrhagic syndrome in the backyard chicken: a retrospective histopathologic case series. *Vet Pathol*. 2014;51(4):787–95.
- Hrabia A. Growth hormone production and role in the reproductive system of female chicken. *Gen Comp Endocrinol*. 2015;220:112–8.
- Li HF, Shu JT, Du YF, Shan YJ, Chen KW, Zhang XY, et al. Analysis of the genetic effects of prolactin gene polymorphisms on chicken egg production. *Mol Biol Rep*. 2013;40(1):289–94.
- Chen P, Suh Y, Choi YM, Shin S, Lee K. Developmental regulation of adipose tissue growth through hyperplasia and hypertrophy in the embryonic leghorn and broiler. *Poult Sci*. 2014;93:1809–17.
- Ball EV, Stenson PD, Abeyasinghe SS, Krawczak M, Cooper DN, Chuzhanova NA. Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum Mutat*. 2005;26:205–13.

39. Tummala H, Ali M, Getty P, Hocking PM, Burt DW, Inglehearn CF, et al. Mutation in the guanine nucleotide-binding protein beta-3 causes retinal degeneration and embryonic mortality in chickens. *Invest Ophthalmol Vis Sci*. 2006;47(11):4714–8.
40. Cui JX, Du HL, Liang Y, Deng XM, Li N, Zhang XQ. Association of polymorphisms in the promoter region of chicken prolactin with egg production. *Poult Sci*. 2006;85:26–31.
41. Wang R, Wang T, Lu W, Zhang W, Chen W, Kang X, et al. Three indel variants in chicken LPIN1 exon 6/flanking region are associated with performance and carcass traits. *Br Poult Sci*. 2015;56(6):621–30.
42. Megens HJ, Crooijmans RP, Bastiaansen JW, Kerstens HH, Coster A, Jalving R, et al. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. *BMC Genet*. 2009;10:86.
43. Wragg D, Mwacharo JM, Alcalde JA, Hocking PM, Hanotte O. Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity*. 2012;109(1):6–18.
44. Qanbari S, Strom TM, Haberer G, Weigend S, Gheyas AA, Turner F, et al. A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. *PLoS One*. 2012;7(11):e49525.
45. Willing E, Dreyer C, Oosterhout C. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One*. 2012;7:e42649.
46. Chen CH, Chuang TJ, Liao BY, Chen FC. Scanning for the signatures of positive selection for human-specific insertions and deletions. *Genome Biol Evol*. 2009;1:415–9.
47. Lillie M, Sheng Z, Honaker CF, Dorshorst BJ, Ashwell CM, Siegel PB, et al. Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens. *BMC Genomics*. 2017;18:99.
48. Bateson ZW, Whittingham LA, Johnson JA, Dunn PO. Contrasting patterns of selection and drift between two categories of immune genes in prairie-chickens. *Mol Ecol*. 2015;24(24):6095–106.
49. Johnson JL, Wittgenstein H, Mitchell SE, Hyma KE, Temnykh SV, Kharlamova AV, et al. Genotyping-By-Sequencing (GBS) detects genetic structure and confirms behavioral QTL in tame and aggressive foxes (*Vulpes vulpes*). *PLoS One*. 2015;10(6):e0127013.
50. Lee SW, Won JY, Yang J, Lee J, Kim SY, Lee EJ, et al. AKAP6 inhibition impairs myoblast differentiation and muscle regeneration: positive loop between AKAP6 and myogenin. *Sci Rep*. 2015;5:16523.
51. Passariello CL, Li J, Dodge-Kafka K, Kapiloff MS. mAKAP-a master scaffold for cardiac remodeling. *J Cardiovasc Pharmacol*. 2015;65(3):218–25.
52. Li ZH, Li H, Zhang H, Wang SZ, Wang QG, Wang YX. Identification of a single nucleotide polymorphism of the insulin-like growth factor binding protein 2 gene and its association with growth and body composition traits in the chicken. *J Anim Sci*. 2006;84(11):2902–6.
53. Leng L, Wang S, Li Z, Wang Q, Li H. A polymorphism in the 3'-flanking region of insulin-like growth factor binding protein 2 gene associated with abdominal fat in chickens. *Poult Sci*. 2009;88(5):938–42.
54. Lei M, Peng X, Zhou M, Luo C, Nie Q, Zhang X. Polymorphisms of the IGF1R gene and their genetic effects on chicken early growth and carcass traits. *BMC Genet*. 2008;9:70.
55. Armstrong DG, Hogg CO. Insulin-like growth factor I (IGF-I), IGF-II and type-I IGF receptor gene expression in the ovary of the laying hen. *J Reprod Fertil*. 1996;106:101–6.
56. Chen B, Xu J, He X, Xu H, Li G, Du H. A genome-wide mRNA screen and functional analysis reveal FOXO3 as a candidate gene for chicken growth. *PLoS One*. 2015;10:e0137087.
57. Machon O, Masek J, Machonova O, Krauss S, Kozmik Z. Meis2 is essential for cranial and cardiac neural crest development. *BMC Dev Biol*. 2015;15:40.
58. Duan Z, Sun C, Shen M, Wang K, Yang N, Zheng J, et al. Genetic architecture dissection by genome-wide association analysis reveals avian eggshell ultrastructure traits. *Sci Rep*. 2016;6:28836.
59. Zhou M, Lei M, Rao Y, Nie Q, Zeng H, Xia M, et al. Polymorphisms of vasoactive intestinal peptide receptor-1 gene and their genetic effects on broodiness in chickens. *Poult Sci*. 2008;87(5):893–903.
60. Xu HP, Zeng H, Zhang DX, Jia XL, Luo CL, Fang MX, et al. Polymorphisms associated with egg number at 300 days of age in chickens. *Genet Mol Res*. 2011;10(4):2279–89.
61. Eck SH, Benet-Pagès A, Flisikowski K, Meitinger T, Fries R, Strom TM. Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biol*. 2009;10(8):R82.
62. FastQC tool. 2016. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 3 Feb 2016.
63. SeqClean tool. 2016. <http://sourceforge.net/projects/seqclean/files/>. Accessed 3 Feb 2016.
64. NCBI Chicken genome sequence. ftp://ftp.ncbi.nih.gov/genomes/Gallus_gallus (2014). Accessed 10 Feb 2015.
65. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
66. Picard tool. 2016. <http://broadinstitute.github.io/picard/>. Accessed 13 Mar 2016.
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
68. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
69. 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*. 2010;26(16):2069–70.
70. Pellatt AJ, Slattery ML, Mullany LE, Wolff RK, Pellatt DF. Dietary intake alters gene expression in colon tissue: possible underlying mechanism for the influence of diet on disease. *Pharmacogenet Genomics*. 2016;26(6):294–306.
71. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight*. 2016;1(20):e90151.
72. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
73. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*. 2016;32(2):289–91.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

