Genome-wide association study for body weight at 35 days in broilers

C. Dadousis[1], A. Somavilla[1], P. Gottardo[1], A. Whalen[1], D. Wilson[1], D. De Burca[1], A. Kranis[2,1] & J.M. Hickey[1]

[1] The Roslin Institute, University of Edinburgh, Midlothian, UK
[2] Aviagen Ltd, EH28 8SZ, Midlothian, UK
christos.dadousis@roslin.ed.ac.uk (Corresponding Author)

Summary

Our objective was to conduct a genome-wide association analysis (GWAS) for chicken body weight measured at 35 days (BW35). A large dataset of more than 43,000 broilers (meat-type poultry) from a commercial line with imputed 600,000 single nucleotide polymorphisms (SNPs) genotypes was analysed. Single marker regression was fitted with a genomic relationship matrix in the GEMMA software. Prior to GWAS, BW35 was adjusted for the non genetic effects of hatch, mating group, pen and sex. In total, 136 significant SNPs ($P < 5\times10^{-7}$) were identified across 7 *Gallus gallus* autosomes (GGA) and the Z chromosome. A sharp peak was detected on GGA4 at 67,440,745bp ($P = 3.91\times10^{-28}$; *Gallus_gallus-4.0*). Highly significant SNPs [$-\log_{10}(P) > 10$] were also found on GGA1 (~142.7Mb), GGA2 (~111.9Mb), GGA3 (~49.82Mb) and the Z chromosome (~68.6Mb).

Keywords: body weight, broilers, genome-wide association study

Introduction

Body weight (BW) is a trait of economic importance in poultry breeding and genomic evaluations for this trait are routinely applied. Identification of genomic regions related to BW might improve selection efficiency and genetic gain in poultry breeding, and provide new insights into the genetic background of the trait. However, as a complex polygenic trait, BW is expected to be affected by a large number of small-effect quantitative trait loci (QTL), thus a large sample size is required in genome-wide association analysis (GWAS) to detect them.

Our aim was to conduct a GWAS for body weight measured at 35 days of age (BW35) in a commercial broiler (meat-type chicken) line. A dataset consisting of >43,000 individuals from a purebred broiler line with phenotypes and imputed 600k genotypes was studied.

Material and methods

Phenotype and genotyping

The trait of interest in this study was BW35 and phenotypic records (n=43,914) were obtained from a purebred commercial broiler line owned by Aviagen. Different low and high density SNP (single nucleotide polymorphism) panels (600k, 50k, 42k, 3k and 384) were used to genotype the animals [see Wolc et al. (2016) for a detailed description of the SNP panels used]. All genotypes were imputed to 600k with the AlphaImpute software (Hickey et al., 2012; Antolín et al., 2017). After SNP quality control that included: (1) call rate >95%, (2) minor allele frequency > 0.005, and (3) no extreme deviation of Hardy-Weinberg proportions (P < 0.00001), 309,174 SNPs distributed over all but the 16th *Gallus gallus* autosomes (GGA) and the Z chromosome were retained. The Ensembl *Gallus_gallus-4.0* database was used to identify the SNPs positions on the genome (http://www.ensembl.org/index.html).

Genetic analysis

Variance component and heritability were estimated using an animal model with a pedigree relationship matrix in ASReml.v3 (Butler et al., 2009) as follows:

(1)

where is a vector containing the measures of BW35; is a vector of the non-genetic effects of hatch week (381 levels), pen (25 levels), mating group (325 levels) and sex] and and Z are the design matrices linking each measure to each effect in and , respectively. The random effects consisted of animal, , and residual, , where is the pedigree relationship matrix, is an identity matrix, and and are the additive genetic and residual variances, respectively. The coefficient of heritability was estimated as .

Genome-wide association analysis

Prior to the GWAS analysis, the phenotypes were adjusted for the non-genetic factors described in (1). The analysis was conducted with the GEMMA software (Zhou and Stephens, 2012) using the following statistical model:

(2)

where is a vector containing the adjusted phenotypes, and and are the additive genomic and residual random terms, respectively. For the random terms the following normal distributions were assumed: $g \sim N(0, \mathbf{G}\sigma_g^2)$ and $e \sim N(0, \mathbf{I}\sigma_e^2)$, where G is the genomic relationship matrix calculated following the first method of VanRaden (2008) and σ_g^2 and σ_e^2 denote the additive genomic and residual variances, respectively. *P*-values smaller than 5×10^{-7} were declared as significant associations (Burton et al., 2007). Manhattan and quantile-quantile (Q-Q) plots of the GWAS results were generated using the *qqman* R package (R Team, 2013; Turner, 2014).

Results and discussion

The average BW35 was 1830g, ranging from 1080dg to 2740g and the estimated coefficient of heritability was 0.42. The Q-Q plot (Figure 1) showed extreme departure on the tail of the distribution providing extra evidence for associations of the GWAS analyses. The information on significant SNPs is presented in Table 1 and Figure 1. The association analysis revealed

136 significant SNPs ($P < 5 \times 10^{-7}$) across 8 chromosomes (GGA1 – 4, GGA8, GGA10, GGA13 and the Z). A clear signal was found on GGA4 with a peak at ~67.4Mb.

More than 65% of the significant SNPs were detected in nine regions on GGA4. The regions 4a (~13.42Mb), 4c (~52.7Mb) and 4i (~72.7Mb) were represented by only one SNP each. In the region 4b (~50.2Mb), five SNP were present, with *P*-values around the significance threshold. The 4d region (~54.4 – 55.6Mb) contained 23 SNPs with a peak at ~55.3Mb. Nineteen significant SNPs were found between ~59.6-60Mb. In close proximity, a sharp peak was detected at 67,440,745bp, spanning between ~65.4-67.4Mb.
Strong signals [$-\log_{10}(P) > 10$] were also detected on GGA1, GGA2, GGA3 and the Z chromosome. More precisely, on GGA1 five sub-regions were identified, each one represented by one SNP. The strongest signal was identified at 142,670,238bp, while for the rest of the sub-regions the signals were on the significance threshold. Three sub-regions were detected on GGA2 (~18.9, 21.5 and ~111.0-111.9Mb) with a peak at 111,900,608bp. On GGA3 a QTL was identified at 49,817,660bp in a region spanning between ~49.8-54.1Mb. A second signal was identified at ~27.7Mb, albeit with a weaker strength. Three regions were detected on the Z chromosome, spanning between ~67.1-72.4Mb, with a peak at 68.35Mb. Weaker associations were found on three more GGA. On GGA8, 2 sub-regions were detected at ~22.96 and 27.25Mb. Two SNPs were significantly associated to BW35 at ~10.00 and ~10.81Mb on GGA10. At the tail of GGA13 a weak signal was detected at ~16.81Mb.

In line to previous studies (Liu et al., 2013; Pértille et al., 2015, 2017), the importance of GGA4 for body weight in broilers has been highlighted in our analysis. Although our highest signal on GGA4 was not in the area of *CCKAR* (Dunn et al. 2013), which expression is involved in the regulation of growth and BW in chickens, the region 4i is ~80.7kb downstream, laying within the *STIM2* (stromal interaction molecule 2). Moreover, our results support the importance of GGA1 to 3 and the Z chromosome (Sewalem et al., 2002; Ikeobi et al., 2004) for BW.

Conclusion

To the best of our knowledge, this is the largest GWAS analysis that has been conducted for body weight in chickens. We have identified specific regions on five chromosomes (GGA 1 – 4 and the Z chromosome) with clear associations to BW35. Our results support the small-effect QTL hypothesis for BW35 and the need of large sample size GWAS and will help in fine mapping of the QTLs related to one of the most important traits in the poultry industry. Moreover, the identified regions associated to BW35 might serve as a substrate in genomic predictions for constructing weighted genomic relationship matrices accounting for the trait genomic architecture.

Table 1. Summary of the significant results of the genome wide association analysis.

GGA	No. of SNP	Interval, Mb	*P*-value (range)	Top SNP location, bp	Effect (SE)	Top SNP MAF
1a	1	-		51,608,497	-0.56 (0.11)	0.20
1b	1	-		127,040,305	0.51 (0.10)	0.50
1c	1	-		142,670,238	-0.61 (0.09)	0.32
1d	1	-		151,527,237	-0.63 (0.12)	0.19
1e	1	-		168,712,594	0.72 (0.14)	0.35
2a	2	18.89-18.92	)	18,895,734	-0.72 (0.11)	0.21
2b	1	-		21,496,060	0.47 (0.09)	0.44
2c	5	111.04-111.92	)	111,900,608	-1.22 (0.16)	0.09
3a	2	27.69-27.70		27,689,724	0.82 (0.15)	0.14
3b	5	49.82-54.10		49,817,660	-0.70 (0.10)	0.47
4a	1	-		13,417,959	0.59 (0.10)	0.28
4b	4	50.23-50.33	)	50,231,486	-0.94 (0.18)	0.09
4c	1	-		52,734,745	-0.78 (0.10)	0.30
4d	23	54.47-55.63	)	55,349,682	-1.60 (0.22)	0.08
4e	19	59.56-60.07	)	59,684,668	-1.41 (0.24)	0.06
4f	6	65.40-65.42	)	65,423,538	-1.25 (0.21)	0.06
4g	33	66.45-67.44		67,440,745	-1.10 (0.10)	0.44
4h	5	69.52-70.64	)	69,617,303	-0.82 (0.12)	0.34
4i	1	-		72,737,359	-0.86 (0.14)	0.43
8a	11	22.80-23.06		22,963,270	-1.11 (0.19)	0.07
8b	4	27.22-27.25		27,246,458	0.93 (0.18)	0.21
10a	1	-		10,008,797	-0.62 (0.11)	0.53
10b	1	-		10,806,057	-0.72 (0.11)	0.24
13	2	16.71-16.81		16,810,184	0.80 (0.16)	0.37
Za	1	-		67,142,554	-0.39 (0.08)	0.44
Zb	1	-		68,352,353	1.11 (0.15)	0.08
Zc	2	72.45-72.54		72,453,986	-0.43 (0.07)	0.42

GGA = *Gallus gallus* autosome chromosome; Z = The sex chromosome; #SNP = number of the single nucleotide polymorphisms (SNP)

significantly associated to the trait; Interval = The region on the chromosome spanned among the significant SNP (in base pairs); *P*-value (range) = The *P*-value of the highest significant SNP adjusted for genomic control and the range of the *P*-values when multiple SNP were significantly associated to one trait; Top SNP location (bp) = position of the highest significant SNP on the chromosome in base pairs on Galgal4 (http://www.ensembl.org/index.html); Effect = the effect of the top SNP. In parenthesis the standard error; Top SNP MAF = minor allele frequency of the top SNP.

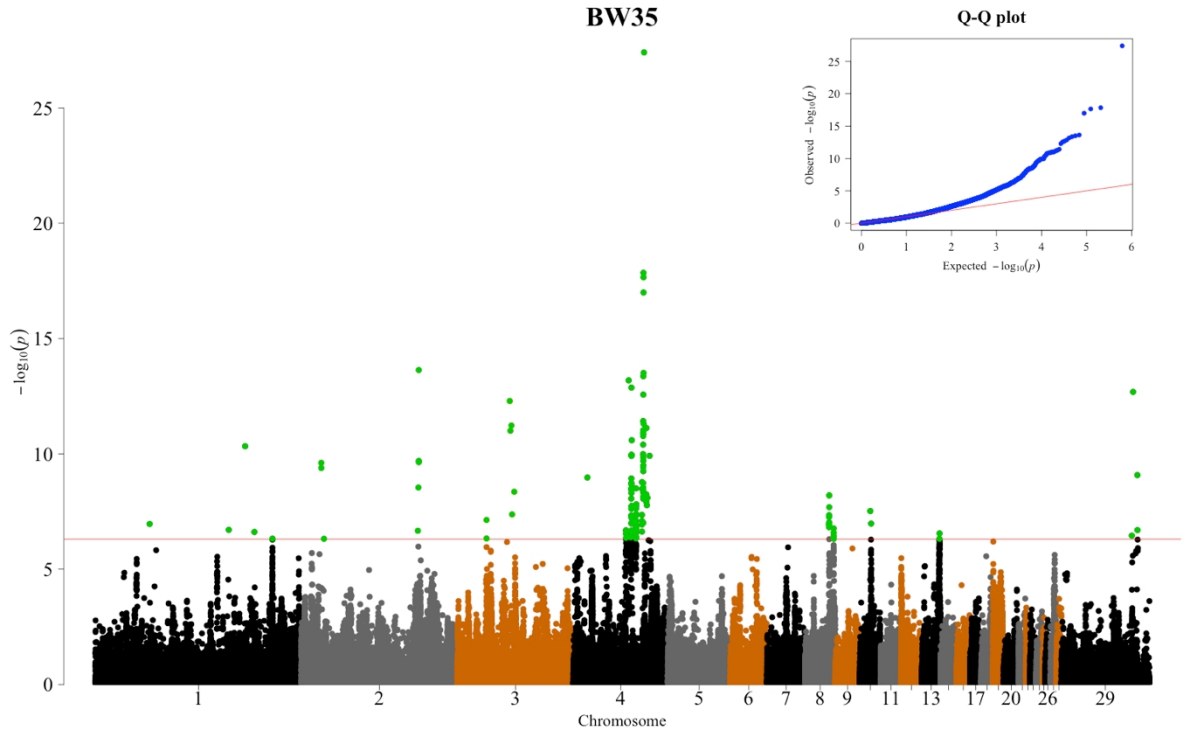

Figure 1. Manhattan plot of for the genome-wide association study (GWAS). The (red) horizontal line indicates a of 6.30 (corresponding to P-value = 5×10^{-7}). Chromosome 29 denotes the Z (sex) chromosome. On top right the quantile-quantile (Q-Q) plot of the observed test statistics of the genome wide association study.

List of References

Antolín, R., C. Nettelblad, G. Gorjanc, D. Money, and J.M. Hickey. 2017. A hybrid method for the imputation of genomic data in livestock populations. Genet. Sel. Evol. 49:30. doi:10.1186/s12711-017-0300-y.

Burton, P.R., D.G. Clayton, L.R. Cardon, N. Craddock, P. Deloukas, A. Duncanson, D.P. Kwiatkowski, M.I. McCarthy, W.H. Ouwehand, and N.J. Samani. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678.

Butler, D.G., B.R. Cullis, A.R. Gilmour, and B.J. Gogel. 2009. ASREML-R Reference Manual. Release 3.0. VSN International Ltd, Hemel Hempstead.

Dunn, I.C., S.L. Meddle, P.W. Wilson, C.A. Wardle, A.S. Law, V.R. Bishop, C. Hindar, G.W. Robertson, D.W. Burt, S.J. Ellison, and others. 2013. Decreased expression of the satiety signal receptor CCKAR is responsible for increased growth and body weight during the domestication of chickens. Am. J. Physiol.-Endocrinol. Metab. 304:E909–E921.

Hickey, J.M., B.P. Kinghorn, B. Tier, J.H. van der Werf, and M.A. Cleveland. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Genet. Sel. Evol. 44:9. doi:10.1186/1297-9686-44-9.

Ikeobi, C.O.N., J.A. Woolliams, D.R. Morrice, A. Law, D. Windsor, D.W. Burt, and P.M. Hocking. 2004. Quantitative trait loci for meat yield and muscle distribution in a broiler layer cross. Livest. Prod. Sci. 87:143–151.

Liu, R., Y. Sun, G. Zhao, F. Wang, D. Wu, M. Zheng, J. Chen, L. Zhang, Y. Hu, and J. Wen. 2013. Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. PLoS One 8:e61172.

Pértille, F., G.C.M. Moreira, R. Zanella, J. de R. da Silva Nunes, C. Boschiero, G.A. Rovadoscki, G.B. Mourão, M.C. Ledur, and L.L. Coutinho. 2017. Genome-wide association study for performance traits in chickens using genotype by sequencing approach. Sci. Rep. 7.

Pértille, F., R. Zanella, A.M. Felício, M.C. Ledur, J.O. Peixoto, and L.L. Coutinho. 2015. Identification of polymorphisms associated with production traits on chicken (Gallus gallus) chromosome 4. Genet. Mol. Res. 14:10717–10728.

Sewalem, A., D.M. Morrice, A. Law, D. Windsor, C.S. Haley, C.O. Ikeobi, D.W. Burt, and P.M. Hocking. 2002. Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross. Poult. Sci. 81:1775–1781.

Team, R.C. 2013. R: A Language and Environment for Statistical Computing.

Turner, S.D. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv 5165.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423.

Wolc, A., A. Kranis, J. Arango, P. Settar, J.E. Fulton, N.P. O'Sullivan, A. Avendano, K.A. Watson, J.M. Hickey, G. de los Campos, R.L. Fernando, D.J. Garrick, and J.C.M. Dekkers. 2016. Implementation of genomic selection in the poultry industry. Anim. Front. 6:23–31. doi:10.2527/af.2016-0004.

Zhou, X., and M. Stephens. 2012. Genome-wide Efficient Mixed Model Analysis for Association Studies. Nat. Genet. 44:821–824. doi:10.1038/ng.2310.