



Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens

A. Wolc^{*,†}, J. Arango[‡], P. Settar[‡], J. E. Fulton[‡], N. P. O'Sullivan[‡], R. Preisinger[§], D. Habier[†], R. Fernando[†], D. J. Garrick[†], W. G. Hill[¶] and J. C. M. Dekkers[†]

*Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Wolynska street 33, 60-637 Poznan, Poland.

†Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, IA 50011-3150, USA. ‡Hy-Line International, Dallas Center, IA 50063, USA. §Lohmann Tierzucht GmbH, 27472 Cuxhaven, Germany. ¶Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Summary

The pioneering work by Professor Soller *et al.*, among others, on the use of genetic markers to analyze quantitative traits has provided opportunities to discover their genetic architecture in livestock by identifying quantitative trait loci (QTL). The recent availability of high-density single nucleotide polymorphism (SNP) panels has advanced such studies by capitalizing on population-wide linkage disequilibrium at positions across the genome. In this study, genomic prediction model Bayes-B was used to identify genomic regions associated with the mean and standard deviation of egg weight at three ages in a commercial brown egg layer line. A total of 24 425 segregating SNPs were evaluated simultaneously using over 2900 genotyped individuals or families. The corresponding phenotypic records were represented as individual measurements or family means from full-sib progeny. A novel approach using the posterior distribution of window variances from the Monte Carlo Markov Chain samples was used to describe genetic architecture and to make statistical inferences about regions with the largest effects. A QTL region on chromosome 4 was found to explain a large proportion of the genetic variance for the mean (30%) and standard deviation (up to 16%) of the weight of eggs laid at specific ages. Additional regions with smaller effects on chromosomes 2, 5, 6, 8, 20, 23, 28 and Z showed suggestive associations with mean egg weight and a region on chromosome 13 with the standard deviation of egg weight at 26–28 weeks of age. The genetic architecture of the analyzed traits was characterized by a limited number of genes or genomic regions with large effects and many regions with small polygenic effects. The region on chromosome 4 can be used to improve both the mean and standard deviation of egg weight by marker-assisted selection.

Keywords GWAS, egg weight, uniformity, layer chickens.

Introduction

Understanding the factors that determine the phenotype has been of interest to plant and animal breeders, geneticists and evolutionary biologists. Reliable statistical methodology has been developed to assess the proportion of differences in phenotype between individuals that is because of genetics vs. environment (Falconer & Mackay 1996; Lynch & Walsh

1998). Large changes in population means for important economic traits have been obtained by selecting on increased performance in livestock populations, including layer chickens, without knowledge of the genetic architecture of the complex quantitative traits that are of prime interest. Advances in molecular genetics have provided opportunities to identify genomic regions associated with traits, based on the pioneering work of Professor Soller *et al.* on the use of genetic markers to identify quantitative trait loci (QTL) (Soller *et al.* 1976; Soller & Genizi 1978; Beckmann & Soller 1983, 1986, 1988; Soller & Beckmann 1990), as well as on the work of other groups (see review by Weller 2009). Strategies for QTL detection in the chicken were reviewed in Soller *et al.* (2006). The development of

Address for correspondence

Jack C. M. Dekkers, Department of Animal Science, 239D Kildee Hall, Iowa State University, Ames, IA 50011-3150, USA.

E-mail: jdekke@iastate.edu

Accepted for publication 3 May 2012

methodology and availability of data has resulted in the detection of large numbers of QTL and estimates of their effects on phenotype in the chicken (Tuiskula-Haavisto *et al.* 2002; Sasaki *et al.* 2004; Honkatukia *et al.* 2005; Schreweis *et al.* 2006), as summarized in Abasht *et al.* (2006) and available online in ChickenQTLdb (<http://www.genome.iastate.edu/cgi-bin/QTLdb/GG/index>) (Hu *et al.* 2010).

Opportunities to detect genomic regions associated with traits of interest and to uncover the genetic architecture (defined as the number and size of QTL) of quantitative traits have been further advanced by the availability of genotypes from high-density panels of single nucleotide polymorphism (SNP) markers on large numbers of individuals. Such genome-wide association studies (GWAS) were pioneered primarily in human genetics (Donnelly 2008) but also have found rapid application in livestock genetics (Goddard & Hayes 2009). For example, Hayes *et al.* (2010) used mixed linear model methodology to estimate the proportion of genetic variance associated with each genomic region of 50 SNPs from the Bovine 50k Illumina SNP chip for three quantitative traits in dairy cattle. GWAS in livestock have been further facilitated by the use of Bayesian variable selection (BVS) models that were developed for the estimation of breeding values based on high-density SNP data through the concept of genomic prediction (Meuwissen *et al.* 2001). To determine the significance of effects fitted in such BVS models, Sahana *et al.* (2010) used the proportion of samples of the Monte Carlo Markov Chain (MCMC) for which the model included SNPs from a particular region of the genome. Onteru *et al.* (2011) and Fan *et al.* (2011) used a bootstrap method to derive significance levels for the proportion of genetic variance explained by genomic regions with large effects in BVS models, but these are computationally very demanding as they entail repeated analysis of 1000 bootstrap samples of the original data.

Most genetic studies focus on the genetic basis of trait means, but much less is known about the factors that determine the level of variation for a trait. Increasing uniformity of egg weight, that is, reducing variation, is desired in layer chickens because eggs of extreme sizes are not suitable for automatic packing and are not economically desirable (i.e. small eggs have low value and production of very large eggs is not efficient) and, for reproducing parent stock, hatchability of large eggs is reduced (Abiola *et al.* 2008). Several studies conducted over the last 10 years (see review by Hill & Mulder 2010) suggest that the within-individual variation has a substantial genetic coefficient of variation, although most heritability estimates are low (based on either repeated records on the same individual or between family differences in residual variance). Egg weight is known to change with age of the layer, which is one of the sources of variability in egg weight. This variability with age can be addressed by selection on estimates of breeding values from random regression models, aimed at reducing the rate of increase in egg weight

with age (Arango *et al.* 2009). However, individuals also vary in the level of uniformity of the weight of eggs laid at specific ages, which will be quantified here by the standard deviation (SD) between the weights of eggs laid by an individual at a given age. For egg weight at 30 weeks of age, Wolc *et al.* (2011) estimated heritabilities of 0.46 and 0.03 for the mean and variance (based on between family differences) and a correlation of 0.14 between estimated breeding values for the mean and variance.

The concept of QTL affecting the variance of quantitative traits (vQTL) has recently gained interest in plant and animal breeding, in part because of the increasing importance of uniformity. The first of recently published studies on vQTL for economically important traits was by Ordas *et al.* (2008) on recombinant inbred lines of maize, in which they found a significant vQTL for residual variation for days to flowering and a suggestive vQTL for ear height and tassel length. In animals, Yang *et al.* (2011) failed to detect significant vQTL for backfat thickness in pigs. We are not aware of studies that have reported vQTL for egg weight, which could be due to the large data sets that are needed because of low power to detect vQTL (Visscher & Posthuma 2010). Recently published methodological papers by Rönnegård & Valdar (2011), who proposed a double-hierarchical model to simultaneously scan for QTL for the mean and variance in experimental crosses, and Yang *et al.* (2011), who used MCMC methods to sample marker effects on mean and variance, reflect scientific interest in the area of vQTL detection, along with its practical importance in cases where increasing uniformity is of interest.

Against this background, the objective of this study was to use a genome-wide association study on high-density SNP data from a brown egg layer line to identify genomic regions that are associated with the mean and SD of egg weight at three different ages. BVS methods were used for analysis, and novel methods to test for the significance of associations were employed.

Materials and methods

Data

Egg weights measured to the nearest gram were collected at three ages in the production cycle on hens from six generations of a purebred brown egg layer line: on the first three eggs laid (EW1), on three to five eggs laid at 26–28 weeks of age (EW2) and on three to five eggs laid at 42–46 weeks of age (EW3). Hens were housed individually, and collected eggs were individually weighed. After pre-correcting for the station that was used to process the egg, the mean and SD of egg weight were calculated separately for each bird at each age. Means and SD were then corrected for the effect of hatch-week, using solutions from a single-trait animal model that included hatch-week as a fixed effect. The standard deviation rather than variance was used as a

measure of variation in egg weight because it is on the same scale as the mean, which makes it easier to interpret, and it has better distributional properties (less skewed) than the variance. Skew was not substantial for the mean [equal to 0.31, 0.29 and 0.20 for the means of EW1, EW2 and EW3 (EW1m, EW2m and EW3m) respectively], but distributions of the SD were skewed to the right (skew equal to 1.54, 1.76 and 2.00 for EW1SD, EW2SD and EW3SD respectively). This skewness is, however, not expected to affect results, in particular because significance levels were determined empirically (see later).

A custom Illumina SNP 42k Infinium chip, with SNPs that covered the genome, was used for genotyping. After quality checks, 24 425 markers on chromosomes 1–28, Z and two unassigned linkage groups were used for subsequent analyses. Quality checks included the proportion of missing genotypes <0.05, minor allele frequency (MAF) >0.025 and detected parent–offspring mismatches <5%. These edits accounted for 9%, 88% and 2% of removed SNPs respectively.

In total, over 2900 phenotype–genotype records from the first five available generations were used to estimate SNP effects, which will be referred to as the training data. This included over 1400 hens with individual phenotypes and a similar number of records that were represented by family means (Table 1). The latter included animals that were phenotyped but not genotyped themselves (over 10 500 hens) but that had both parents genotyped (all animals used for breeding across the six generations were genotyped, in total 530 sires and 1555 dams). Family means were linked to the average genotype of their parents with residuals weighted by $w_p = \frac{1-h^2}{(1-0.5h^2)^p}$ to account for the additional residual variance and different family sizes p (Garrick *et al.* 2009).

Validation data

A total of 287 genotyped and phenotyped progeny and 306 grand-progeny of the last generation of training

Table 1 Basic statistics (mean and standard deviation) for the phenotypic traits in the training data defined as the mean and standard deviation of egg weight (g) of the first three eggs (EW1m and EW1SD), eggs laid between 26 and 28 weeks of age (EW2m and EW2SD) and eggs laid between 42 and 46 weeks of age (EW3m and EW3SD) for genotyped animals with own phenotype and for progeny of genotyped birds (all parents were genotyped), which were incorporated as family means.

	Own phenotype			Progeny		
	N	Mean	Standard deviation	N	Mean	Standard deviation
EW1m	1398	45.6	4.50	10 482	46.0	4.99
EW2m	1482	58.0	4.12	10 796	57.5	4.87
EW3m	1483	62.5	4.32	2255	62.6	4.90
EW1SD	1395	2.72	1.77	10 468	2.69	1.74
EW2SD	1482	1.69	1.02	10 762	1.79	1.11
EW3SD	1482	1.87	1.00	2252	1.95	1.15

individuals were used to validate the estimated effects, and these will be referred to as the validation data. The confirmation of the estimated SNP effects in animals whose phenotypes were not used for estimation (training) increases the confidence in the associations detected in the training data.

Statistical analyses

Bayesian variable selection method Bayes-B (Meuwissen *et al.* 2001) was used to analyze all SNPs simultaneously, using GENSEL software (Fernando & Garrick 2009). Each trait (mean and SD within each age) was analyzed separately, with SNP allele substitution effects fitted as random effects. By fitting all SNPs simultaneously, the Bayes-B method has been shown to account implicitly for population stratification (see for example Toosi *et al.* 2010), without requiring a polygenic effect with relationships to be fitted. The Bayes-B method has been shown to map QTL precisely in simulation studies (Sun *et al.* 2011) without much overestimation of significant effects, in contrast to expectations from single SNP methods (Beavis 1998). Parameter π , which is the prior proportion of SNPs that is assumed to have no effect on the trait within an iteration of the MCMC chain, was set at 99.5%. Choice of π was motivated by allowing only regions with strong association with traits to be fitted, assuming that about 100–150 regions would explain the majority of genetic variance in the mean and SD of egg weight. Priors for genetic and residual variances for each trait were obtained from a single-trait pedigree-based REML analysis. A chain of 25 000 iterations was used, of which the first 3000 were discarded as burn-in.

For significance testing in a Bayesian GWAS, Onteru *et al.* (2011) and Fan *et al.* (2011) used a bootstrapping method to generate a distribution for the window variance according to the null hypothesis of no QTL within that window. Although conceptually appealing, this method is computationally very demanding. For the current study, an alternative approach was used, which has been implemented in version 4.0 of the GENSEL software (<http://biggs.ansci.iastate.edu/>). This approach was based on the property that, after burn-in, each iteration of the MCMC chain provides samples of the posterior distribution of parameters of interest. For each SNP, these samples can be used to obtain a posterior probability of inclusion (PPI), which is the proportion of samples in which a given SNP was included in the model with a non-zero effect, and to estimate the genetic variance explained by each SNP. However, because of linkage disequilibrium, the effect of a QTL may be spread over a number of neighboring SNPs, each having relatively small variance and low PPI. Thus, to quantify the combined effect of a genomic region that may be associated with QTL, the genome was divided into 1042 non-overlapping windows of 1 Mb based on Build WUGSC 2.1/galGal3 (http://genome.wustl.edu/genomes/view/gallus_gallus/#sequences_maps).

For each window and each 50th iteration of the chain, sampled values for the effects of the SNPs in the window were then used to compute a sample of the posterior distribution of the true breeding value for that window for each individual by multiplying the sampled SNP effects with the individual's SNP genotypes and summing across all SNPs in the window. The variance across individuals of the resulting sample breeding values ascribed to each window ('window variance') was then used to provide a sample of its posterior distribution. Resulting sample window variances were divided by the variance explained by all SNPs across the genome in that iteration of the MCMC chain to convert window variances to proportions of genetic variance explained by the window. The cumulative distribution of window variances ranked by size was used to infer aspects of the genetic architecture of the studied traits, namely the number and effect sizes of QTL. Windows that captured more than 0.1% of genetic variance in over 90% of the samples were declared to explain significantly more variance than expected. The cutoff of 0.1% was chosen because it is the expected percentage of genetic variance explained by each of ~1000 1-Mb windows across the genome under a pure polygenic model.

Within each of the significant windows, the SNP with the highest PPI and that explained the largest proportion of genetic variance was chosen as a candidate SNP. Significance of the candidate SNP was verified in the validation data, that is, data on progeny and grand-progeny of the last generation of training individuals. A single SNP analysis was used for this purpose, by fitting SNP genotype as a fixed effect in an animal model with ASREML (Gilmour *et al.* 2008), along with the fixed effect of hatch-week. A model fitting the interaction between SNP genotype and generation was also applied to the full data set (training plus validation) in the ASREML analysis to evaluate the consistency of estimates across generations and to check for spurious associations from potential confounding between changes in egg weight across generations and random changes in allele frequencies. The significance in validation was also evaluated for the SD of log-transformed egg weights to remove the effects of a potential mean–variance relationship that could originate from simple multiplicative effects.

Results

The estimates of heritability based on the posterior means of the proportion of genetic variance associated with all SNPs across the genome ranged from 0.38 to 0.47 for mean egg weight and from 0.035 to 0.051 for the SD of egg weight (Table 2). These estimates were at least 30% lower than estimates of heritability obtained from multi-trait pedigree-based analyses (Table 2), likely because of the high value of π used in the Bayesian analyses. Estimates of genetic correlations from pedigree-based analysis between the same trait measured at different ages were >0.8 for means and

>0.6 for SD (Table 2). Estimates of genetic correlations between the mean and SD of egg weight at a given age were positive and moderately high, ranging from 0.54 to 0.74.

A Manhattan plot of the posterior mean of the variance of true breeding values for each 1 Mb window across the genome (Fig. 1) shows that regions with large effects were consistent across the three ages for mean egg weight but less consistent for the SD of egg weight. Histograms for posterior means of window variances for EW2m and EW2SD are presented in Fig. 2. The estimate of the genetic variance explained was >1% for 12 windows for EW2m and 15 windows for EW2SD, and >2% for six windows for EW2m and seven windows for EW2SD. The largest proportion of the genetic variance explained by any window was estimated to be 28% for EW2m and 5.3% for EW2SD. Similar results were observed at other ages, with the largest effects being greater for the mean than for the SD.

The cumulative distribution of window variances when ranked by size shows that a limited number of windows were sufficient to explain over 50% of the genetic variance: about 20 for means and over 40 for SD traits (Fig. 3). The proportions of genetic variance explained by each window are not guaranteed to sum to 100% across the genome because non-zero covariances between regions are ignored in the simple sum. As the simple sums of window variances were >100%, the covariances between regions were on average negative. Most windows for which the estimates of the variance explained exceeded the polygenic expectation of 0.1% of genetic variance did so in less than 90% of posterior samples of the MCMC and were, therefore, deemed not significant. Those that exceeded the 0.1% variance expectation in 90% of posterior samples were declared significant and are summarized in Table 3.

Table 2 Estimates of heritability (bold, on the diagonal), genetic (above diagonal) and phenotypic (below diagonal) correlations between the mean (m) and standard deviation (SD) of egg weight at different ages from a multi-trait model using pedigree relationships, and estimate of heritability based on the Bayesian whole-genome marker analysis.

	EW1m	EW1SD	EW2m	EW2SD	EW3m	EW3SD
EW1m	0.662	0.54	0.88	0.65	0.83	0.72
EW1SD	0.13	0.077	0.60	0.68	0.53	0.61
EW2m	0.69	0.13	0.749	0.62	0.96	0.75
EW2SD	0.16	0.05	0.18	0.096	0.60	0.96
EW3m	0.63	0.08	0.78	0.19	0.727	0.74
EW3SD	0.14	0.04	0.18	0.10	0.19	0.069
Heritability from markers	0.383	0.035	0.467	0.051	0.447	0.048

For pedigree-based analysis, SE were 0.02–0.03 for h^2 of egg weight means and 0.01–0.02 for h^2 of standard deviation, 0.01–0.02 for genetic correlations between means, 0.05–0.06 for genetic correlations involving early SD and 0.1–0.15 for genetic correlations involving late SD.

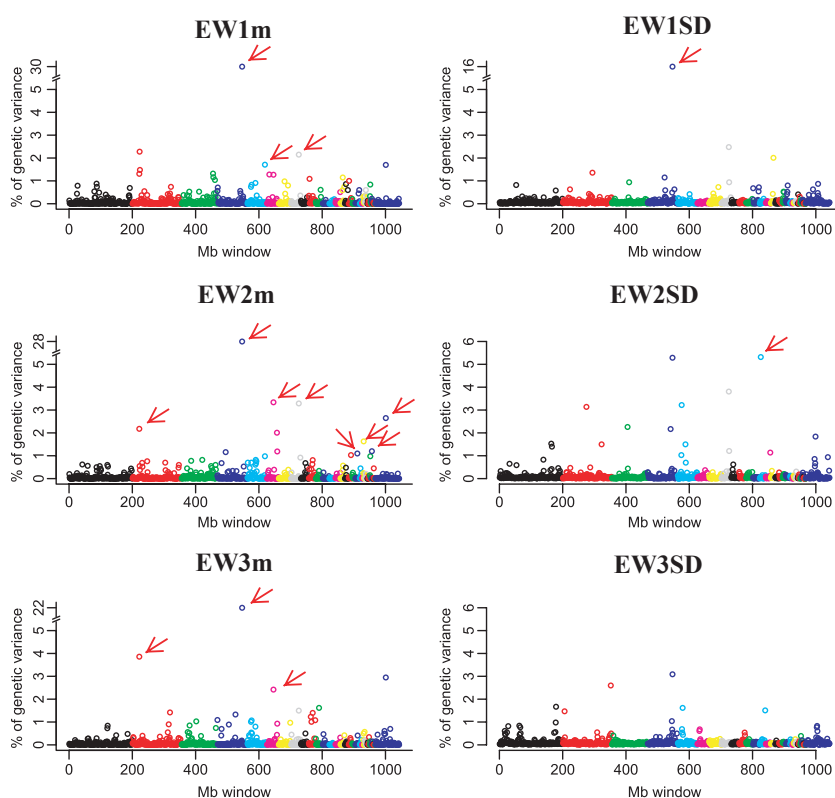


Figure 1 Proportion of genetic variance explained by each 1-Mb SNP window across the genome for the mean and standard deviation of egg weight of the first three eggs (EW1m and EW1SD), eggs laid between 26 and 28 weeks of age (EW2m and EW2SD) and eggs laid between 42 and 46 weeks of age (EW3m and EW3SD). Arrows point to the windows that explain a significant amount of variation and different colors code the consecutive chromosomes.

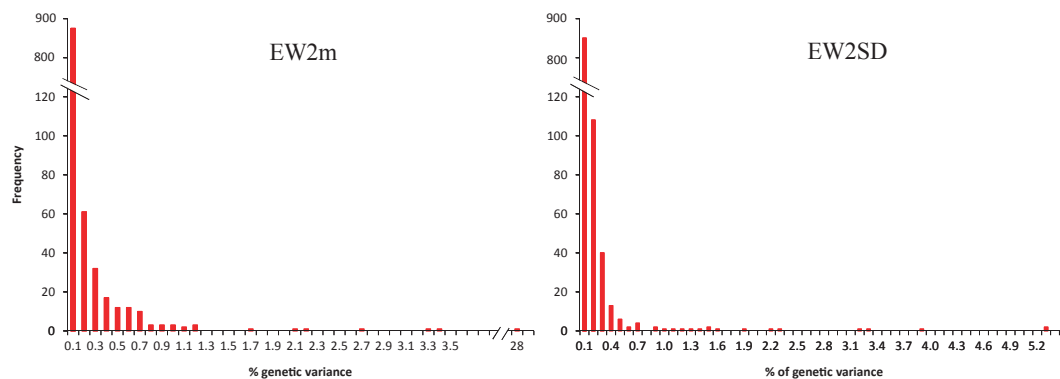


Figure 2 Histogram of the proportion of genetic variance explained by 1-Mb SNP windows across the genome for the mean and standard deviation of early egg weight.

The top SNP from each significant window (listed in Table 3) were fitted one at a time as a fixed effect in an animal model using ASREML. Models that fitted both an additive and a dominance effect of the SNP did not reveal significant deviations from additivity ($P > 0.05$) for any SNP, and therefore, only additive allele substitution effects were fitted. Estimates by generation of allele substitution effects fitted to the full data set (training plus validation) are shown in Fig. 4 and were consistent for most of the top SNPs. Only the SNP on chromosome 13 for EW2SD and SNP *rs14491030* on chromosome 4 for EW3m showed significant interactions with generation (P -values 0.05 and 0.03

respectively). In the following, results for the mean and SD of egg weight for specific regions are described in further detail.

Egg weight mean

The highest proportion of genetic variance (30%) was explained by a window on chromosome 4 (Fig. 1) for mean weight of the first three eggs, with the highest PPI for two adjacent SNPs (PPI = 1 and 0.34, Table 3). These SNPs were approximately 50 Kbp apart but were not in high linkage disequilibrium with each other ($r^2 = 0.25$). The

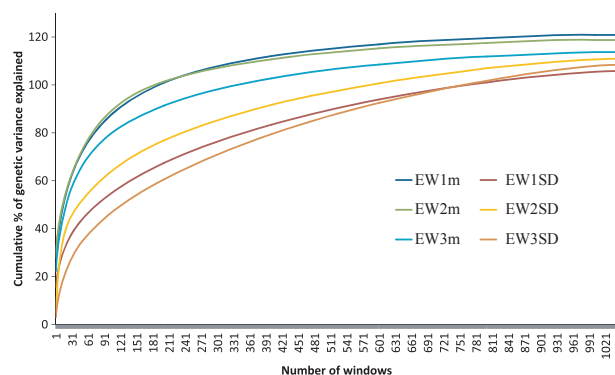


Figure 3 Cumulative proportion of variance explained by 1 Mb windows, ranked by size. The asymptote exceeds 100% because the accumulation assumes that the window effects are independent, but they may be correlated.

same region explained 28% of the genetic variation for EW2 and 22% for EW3. The frequency of the favorable alleles at these two SNPs increased over the five analyzed generations from 0.73 to 0.82 for *rs14491030* and from 0.41 to 0.56

for *rs14699480*. To avoid spurious associations owing to potential confounding of changes in allele frequencies and egg weight over generations, SNP effects were also estimated within each generation by including the interaction of SNP and generation in an ASREML analysis. The SNP effect estimates were consistent across generations (Fig. 4), and the interaction for the top SNP was not significant for mean egg weight at any age ($P > 0.05$). This region was also highly significant ($P < 2 \times 10^{-9}$) in the validation data.

The second important region for mean egg weight, explaining 2 and 3% of genetic variance for EW1 and EW2 respectively was located on chromosome 8 (Table 3). The SNP with the highest PPI in this region was significant in validation for EW2m ($P = 0.002$) but not for EW1m ($P = 0.11$).

Another two regions, on chromosomes 2 and 6, were found to explain a significant proportion of variation for EW2 and EW3, but only the first was confirmed in validation (Table 3). Other windows that explained a substantial proportion of variation in mean egg weight appeared to be age specific. The two significant SNPs on chromosome Z for EW2m were in perfect linkage disequilibrium ($r^2 = 1$). None

Table 3 One mega base windows that explain a significant proportion of variation (probability > 90% of window variance > 0.1%) and results for the most significant SNPs within these windows and P -values of the effect of these SNP in validation.

Trait	Significant 1 Mb window					Most significant SNP within the window				
	Chromosome	1 Mb window number	% genetic variance	Number of SNPs	Probability > 0.1	SNP	Position (kb)	Posterior Probability of Inclusion	% genetic variance SNP	P -value in validation
EW1m	4	78	30.3	20	1.000	<i>rs14491030</i>	78775527	1.00	38.5	9.9×10^{-14}
	5	56	1.7	29	0.995	<i>rs13592733</i>	56230958	0.88	1.8	0.046
	8	24	2.2	28	0.991	<i>rs14652932</i>	24651218	0.95	2.7	0.11
EW1SD	4	78	15.7	20	0.984	<i>rs14491030</i>	78775527	0.97	21.3	0.005
EW2m	2	21	2.2	25	0.950	<i>rs13541562</i>	21545338	0.93	2.6	0.002
	4	78	28.0	20	1.000	<i>rs14491030</i>	78775527	1.00	24.4	1.4×10^{-15}
						<i>rs14699480</i>	78724797	0.99	4.6	2.2×10^{-10}
	6	21	3.3	27	1.000	<i>rs14581563</i>	21376191	0.99	4.2	0.21
						<i>rs14762454</i>	32441589	0.56	0.3	5.4×10^{-5}
	6	32	2.0	29	1.000	<i>rs10722133</i>	32368872	0.99	2.5	0.96
	6	33	1.2	24	0.966	<i>rs14593588</i>	33178622	0.65	0.7	0.44
	8	24	3.3	28	1.000	<i>rs14652932</i>	24651218	0.97	3.8	0.002
	20	7	1.1	22	0.957	<i>rs13632881</i>	6412107	0.78	0.9	0.54
	23	2	1.6	18	0.943	<i>rs13529891</i>	1576845	0.65	0.8	0.67
				0.943	<i>rs14288877</i>	1516675	0.19	0.06	0.56	
	28	2	1.2	35	0.979	<i>rs13546044</i>	1991551	0.98	1.5	0.33
Z	32	2.7	25	0.977	<i>rs14762500</i>	32384275	0.40	0.7	5.4×10^{-5}	
EW2SD	13	4	5.3	20	0.920	<i>rs14055479</i>	3501201	0.87	5.8	0.20
EW3m	2	21	3.9	25	0.989	<i>rs13541562</i>	21545338	0.73	2.5	0.005
	4	78	22.3	20	1.000	<i>rs14491030</i>	78775527	1.00	22.7	2.02×10^{-9}
						<i>rs14699480</i>	78724797	0.57	1.1	6.9×10^{-8}
	6	21	2.4	27	0.948	<i>rs14581563</i>	21376191	0.92	2.5	0.68

Chromosome and position in Mb of the significant windows, number of SNPs in the window, % of genetic variance explained by each window in training, probability of window variance > 0.1%, rs name and chromosome position of the most significant SNP in the window, posterior probability of inclusion and proportion of genetic variance explained by the most significant SNP in training, P -value for significance of additive SNP effect in the validation set.

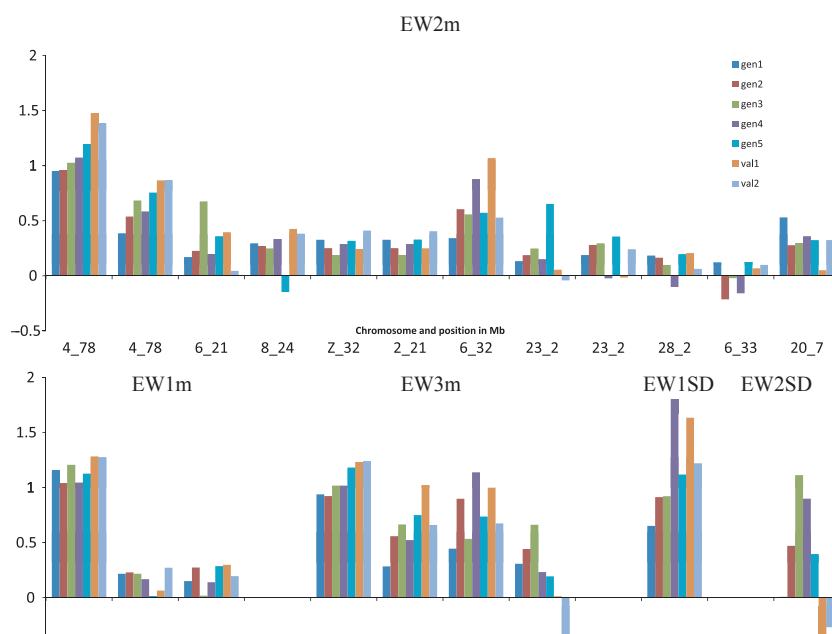


Figure 4 Estimates of allele substitution effects within each generation for the SNP from windows that explain a significant proportion of variation. SNPs are identified on the horizontal axis by chromosome number_1 Mb window on the chromosome, corresponding to Table 3. Estimates of allele substitution effects are standardized by the marker-based genetic standard deviation for the trait.

of the regions that explained <2.5% of genetic variance were confirmed in validation (Table 3).

Egg weight standard deviation

The same chromosomal region on chromosome 4 as for the mean showed the strongest association for SD of egg weight for EW1SD (16% of genetic variance) and EW3SD (3%) and the second strongest association for EW2SD (5%) (Fig. 1). However, this window reached the significance level (based on window variance >0.1% in over 90% of iterations) only for EW1SD and was also validated only for EW1SD (Table 3), but the direction of the effect was consistent across ages. An additional region on chromosome 13 was significant for EW2SD (Table 3), but the SNP with the strongest signal in this window was not significant in the validation data.

Discussion

Genomic regions associated with the mean and standard deviation of egg weight.

In this study, a region on chromosome 4 was found to have a strong association with both the mean and SD of egg weight at three specific ages, although its effects on SD at the two later ages were not significant (Fig. 1, Table 3). Estimates of effects of this region were consistent across generations and in independent validation data (Fig. 4). A similar region on chromosome 4 was previously reported to be associated with egg weight by Tuiskula-Haavisto *et al.* (2002), Sasaki *et al.* (2004) and Schreiweis *et al.* (2006) in F2 crosses using microsatellite markers and by Arango *et al.*

(2008) in a White Leghorn line. To our knowledge, however, this is the first study reporting genomic regions for SD in egg weight, including this important vQTL on chromosome 4.

Alleles at SNPs in the chromosome 4 region that increased the mean also increased the SD, but effects on the SD were greater than could be explained by simple scale effects. For example, for EW2, the effect of the QTL on EW2m was about 4% of the mean egg weight, whereas its effect on EW2SD was about 5% of the mean SD. In addition, in the training data, this region was also significant ($P < 0.03$) as a main effect across generations for the SD when the SDs of log-transformed data for EW2 and EW1 were analyzed.

Several of the other chromosomal regions showing association also were identified in previously reported QTL studies. A region on chromosome Z with significant effects for egg weight reported by Tuiskula-Haavisto *et al.* (2002) was located close to one identified in our study. A QTL for late egg weight (46–61 weeks) on chromosome 2 also was suggested by Tuiskula-Haavisto *et al.* (2002) and Honkatukia *et al.* (2005). Sasaki *et al.* (2004) found a suggestive association with egg weight between microsatellites in a different region than we identified on chromosome 5. The ChickenQTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/GG/index>) does not show any egg weight QTL on chromosome 6. However, in the chromosome 6 region that we identified for egg weight, several studies have reported QTL for body weight (Sewalem *et al.* 2002; Carlborg *et al.* 2003; Siwek *et al.* 2004; Le Rouzic *et al.* 2008), a trait strongly correlated with egg weight. The impact of the QTL for egg weight identified in this study on body weight requires further investigation.

Significance testing and genetic architecture

Most of the SNPs with large effects were found to have consistent effects across generations and were confirmed in the validation data. However, SNPs that explained less than 2.5% of genetic variance in the training data were not confirmed in validation. This could indicate false positives or that the validation data set was not of sufficient size to confirm the association found in the training data. In some cases, the effect of the 1-Mb window was captured by multiple SNPs. For example, the most significant SNP from the window on chromosome 28 for EW2m (Table 3) was only marginally significant ($P = 0.04$) in the whole data set, but other SNPs with sizable PPI in the same window could have contributed to the overall significance of the window.

In general, the proportion of variance explained by all SNPs across the genome was over half that of pedigree-based estimates of heritability for these traits in this population (Table 2) and in other layer lines (Wolc *et al.* 2011). This suggests that the fitted markers captured a sizeable portion of the additive variation through linkage disequilibrium with QTL and/or relationships. If more markers were allowed to be fitted in the model (here 99.5% of markers were assumed to have no effect within a given iteration of the MCMC chain), the proportion of variance explained by markers increased and approximated pedigree-based estimates of heritability (details not shown). In human genetics, the missing heritability problem (Maher 2008), that is, the general finding that markers explain only a small proportion of genetic variance estimated from pedigree, even for complex traits with high heritability such as height, has garnered much attention in recent years. Yang *et al.* (2010) showed that this is in part explained by initial studies focusing only on SNPs with significant effects. Nevertheless, even when fitting all SNPs simultaneously, as in the present analyses, Yang *et al.* (2010) were able to account for only 45% of the genetic variance estimated from pedigree for human height. The additional missing heritability was attributed to incomplete linkage disequilibrium between markers and QTL. Yang *et al.* (2010) based their analysis on nominally unrelated individuals, in stark contrast to our data set, which included strong family relationships. In data sets such as ours, markers also explain relationships (Habier *et al.* 2007), which likely explains our finding that nearly all genetic variance could be accounted for when more markers were included in the model with a low value of π .

To explore the genetic structure of three traits in Holstein dairy cattle, Hayes *et al.* (2010) estimated the distribution of genetic variance explained by chromosomal segments of 50 SNPs from the 50-K bovine SNP panel, with contrasting results by trait. A small proportion of the genome (<2%) explained over 80% of the genetic variance for fat percentage in milk (their Fig. 5), mainly because of the presence of one major gene (*DGAT1*). Three windows with large

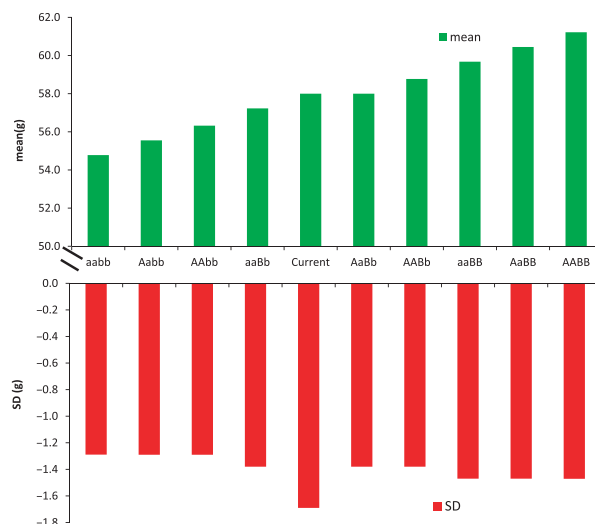


Figure 5 Least square means for the mean (above horizontal axis) and for the standard deviation (SD, below horizontal axis) of early egg weight in birds that carry specific genotype combinations at the two SNPs on chromosome 4 in the region affecting the mean and SD of egg weight. Genotype combinations are ordered on the horizontal axis by their impact on the mean. 'Current' refers to the current population means for the mean and SD of egg weight.

effects were found for black coat color, but about 50% of the genome was needed to explain 80% of the genetic variance. No windows with major effects were identified for overall type, for which 60% of the genome was needed to explain over 80% of the genetic variance. In general, they concluded that, apart from a few genes of major effect, the genetic architecture of the three traits they studied was characterized by a large number of genes of very small effects, in agreement with results for complex traits in humans (Yang *et al.* 2010). This general conclusion was also confirmed by our study of egg weight and SD of egg weight, in which a small number of 1-Mb windows with relatively large effects accounted for about 50% of the genetic variance for the mean and 35% for the SD, with the remaining variation spread over multiple windows with small effects. To explain over 80% of the genetic variance (Fig. 3), over 10% of the genome was needed for the mean but over 25% for the SD.

Our results for the distribution of effects of genomic regions may not be directly comparable with those of Hayes *et al.* (2010) because of differences in methodology. Their results may be affected by double counting because the variance contributed by each region was estimated in separate analyses; indeed, the sum of variances across regions was substantially greater than the total genetic variance (twofold for fat percentage and proportion black) (Hayes *et al.* 2010). Although we avoid this double counting by fitting all regions simultaneously, in our study the sum of window variances also exceeded the total genetic variance, but by <20%, which was likely caused by covariance

between window variances. A sum greater than 100% is consistent with the presence of negative covariances between genes affecting a trait that is expected as a result of selection, that is, the Bulmer effect (Bulmer 1971), but could also reflect negative sampling covariances owing to confounding between neighboring windows. The greater overestimation by the sum for mean egg weight compared to SD could be consistent with stronger past selection on the mean than on the SD.

Possible selection strategies

Two SNPs on chromosome 4 explained sizeable proportions of variation for both the mean and the SD of egg weight and could potentially be used for marker-assisted selection (Soller 1978). First, the SD of eggs produced by a flock could be reduced by producing commercial birds that have the same genotype at this region, thereby removing the variance created by having birds with different genotypes in the same flock. The impact of this for different genotypes under an additive model at the two SNPs is illustrated in Fig. 5. For the current population, this would reduce the between-bird SD of egg weight at a given age by 0.31 g, a 6.6% reduction in the SD. Birds used for commercial production are typically the result of a four-way cross between parental lines. By fixing parental lines for different combinations of alleles at the two SNPs in this region, in principle any combination of genotypes could be generated in the four-way cross. Which genotype to target in the commercial cross depends on its mean egg weight relative to the optimal weight range of market eggs and the importance of increasing uniformity beyond what would already be achieved by fixing the genotype. For example, if the objective was to increase egg weight, the AABB genotype would be preferred, but this would also result in an increase in the within bird SD by 5.3% (Fig. 5). Production of AABb birds would be expected to enable an increase in the mean by 0.77 g without increasing SD. Genotypes that decrease SD are generally associated with a decrease in mean egg weight, with the AAbb genotype resulting in the smallest reduction in the mean (1.78 g) and a 5.3% decrease in SD. As always, selection for specific marker genotypes within parental lines must be weighed against the loss in selection pressure on other genes and other traits that would accompany such selection and is best accomplished by including marker genotypes in an index (Settar *et al.* 2002). Uniformity could be further improved by within-line selection for increased uniformity by standard quantitative genetic means or using additional markers or whole-genome genomic selection predictors for SD.

Acknowledgements

This study was supported by Hy-Line Int., the EW group and Agriculture and Food Research Initiative competitive grants

2009-35205-05100 and 2010-65205-20341 from the USDA National Institute of Food and Agriculture Animal Genome Program. The authors also acknowledge the contributions made to the field of animal breeding and genetics by Prof Morris Soller and the rewarding and stimulating interactions and collaborations several of us have had with him over the years.

Conflicts of interest

The author has no conflicts of interest to declare.

References

- Abasht B., Dekkers J.C.M. & Lamont S.J. (2006) Review of quantitative trait loci identified in the chicken. *Poultry Science* **85**, 2079–96.
- Abiola S.S., Meshioye O.O., Oyerinde B.O. & Bamgbose M.A. (2008) Effect of egg size on hatchability of broiler chicks. *Archivos de Zootecnia* **57**, 83–6.
- Arango J., Settar P., Fulton J.E., Arthur J.A. & O'Sullivan N.P. (2008). From markers to BLUP prediction in a layer line. *Proceedings of the XXIIIth World Poultry Congress*, Brisbane, Australia. 4 pp.
- Arango J., Turkmur L., Settar P. & O'Sullivan N.P. (2009) A random regression approach to model egg weight trajectory in two layer populations. *Proceedings of the 6th European Poultry Genetics Symposium*, Będlewo, Poland. p. 102.
- Beavis W.D. (1998). QTL analyses: power, precision and accuracy. In: *Molecular Dissection of Complex Traits* (Ed. by H. Paterson), pp. 143–62. CRC Press, Boca Raton, FL.
- Beckmann J.S. & Soller M. (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theoretical and Applied Genetics* **67**, 35–43.
- Beckmann J.S. & Soller M. (1986) Restriction fragment length polymorphisms and genetic improvement in agricultural species. *Euphytica* **35**, 111–24.
- Beckmann J.S. & Soller M. (1988) Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. *Theoretical and Applied Genetics* **76**, 228–36.
- Bulmer M.G. (1971) The effect of selection on genetic variability. *The American Naturalist* **105**, 201–11.
- Carlborg O., Kerje S., Schütz K., Jacobsson L., Jensen P. & Andersson L. (2003) A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Research* **13**, 413–21.
- Donnelly P. (2008) Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–31.
- Falconer D.S. & Mackay T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longmans Green, Harlow, Essex, UK.
- Fan B., Onteru S.K., Du Z.-Q., Garrick D.J., Stalder K.J. & Rothschild M.F. (2011) Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PLoS One* **6**, e14726. doi: 10.1371/journal.pone.0014726
- Fernando R.L. & Garrick D.J. (2009) GENSEL – User Manual. <http://big.ansci.iastate.edu/>
- Garrick D.J., Taylor J.F. & Fernando R.L. (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics, Selection, Evolution* **41**, 55.

- Gilmour A.R., Gogel B.J., Cullis B.R. & Thompson R. (2008) ASREML User Guide Release 3.0. VSN Int Ltd, Hemel Hempstead, UK.
- Goddard M.E. & Hayes B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**, 381–91.
- Habier D., Fernando R.L. & Dekkers J.C.M. (2007) The impact of relationship information genome-assisted breeding values. *Genetics* **177**, 2389–97.
- Hayes B.J., Pryce J., Chamberlain A.J., Bowman P.J. & Goddard M.E. (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genetics* **6**, e1001139. doi: 10.1371/journal.pgen.1001139
- Hill W.G. & Mulder H.A. (2010) Genetic analysis of environmental variation. *Genetics Research* **92**, 381–95.
- Honkatukia M., Tuiskula-Haavisto M., de Koning D.J., Virta A., Maki-Tanila A. & Vilkki J. (2005) A region on chicken chromosome 2 affects both egg white thinning and egg weight. *Genetics Selection Evolution* **37**, 563–77.
- Hu Z.-L., Park C.A., Fritz E.R. & Reecy J.M. (2010) QTLdb: a comprehensive database tool building bridges between genotypes and phenotypes. *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*. Leipzig, Germany. <http://www.kongressband.de/wcgalp2010/assets/html/0017.htm>
- Le Rouzic A., Alvarez-Castro J.M. & Carlborg O. (2008) Dissection of the genetic architecture of body weight in chicken reveals the impact of epistasis on domestication traits. *Genetics* **179**, 1591–9.
- Lynch M. & Walsh B. (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- Maher B. (2008) Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21.
- Meuwissen T.H.E., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–29.
- Onteru S.K., Fan B., Du Z.-Q., Garrick D.J., Stalder K.J. & Rothschild M.F. (2011) A whole-genome association study for pig reproductive traits. *Animal Genetics* **43**, 18–26.
- Ordas B., Malvar R.A. & Hill W.G. (2008) Genetic variation and quantitative trait loci associated developmental stability and the environmental correlation between traits in maize. *Genetical Research* **90**, 385–95.
- Rönnegård L. & Valdar W. (2011) Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics* **188**, 435–47.
- Sahana G., Guldbbrandtsen B., Janss L. & Lund M.S. (2010) Comparison of association mapping methods in a complex pedigree population. *Genetic Epidemiology* **34**, 455–62. doi: 10.1002/gepi.20499
- Sasaki O., Odawara S., Takahashi H. *et al.* (2004) Genetic mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chickens. *Animal Genetics* **35**, 188–94.
- Schreiweis M.A., Hester P.Y., Settar P. & Moody D.E. (2006) Identification of quantitative trait loci associated with egg quality, egg production, and body weight in an F2 resource population of chickens. *Animal Genetics* **37**, 106–12.
- Settar P., Dekkers J.C.M. & van der Steen H.A.M. (2002) Control of QTL frequency in breeding populations. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*. Montpellier, France, Session 23, pp. 1–4. 04
- Sewalem A., Morrice D.M., Law A., Windsor D., Haley C.S., Ikeobi C.O., Burt D.W. & Hocking P.M. (2002) Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross. *Poultry Science* **81**, 1775–81.
- Siwek M., Cornelissen S.J., Buitenhuis A.J., Nieuwland M.G., Bovenhuis H., Crooijmans R.P., Groenen M.A., Parmentier H.K. & van der Poel J.J. (2004) Quantitative trait loci for body weight in layers differ from quantitative trait loci specific for antibody responses to sheep red blood cells. *Poultry Science* **83**, 853–9.
- Soller M. (1978) The use of loci associated with quantitative effects in dairy cattle improvement. *Animal Production* **27**, 133–9.
- Soller M. & Beckmann J.S. (1990) Marker-based mapping of quantitative trait loci using replicated progenies. *Theoretical and Applied Genetics* **80**, 205–8.
- Soller M. & Genizi A. (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* **34**, 47–55.
- Soller M., Brody T. & Genizi A. (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**, 35–9.
- Soller M., Weigend S., Romanov M.N., Dekkers J.C.M. & Lamont S.J. (2006) Strategies to assess structural variation in the chicken genome and its associations with biodiversity and biological performance. *Poultry Science* **85**, 2061–78.
- Sun X., Habier D., Fernando R.L., Garrick D.J. & Dekkers J.C.M. (2011) Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian Methods. *BMC Proceedings* **5**(Suppl 3), S13.
- Toosi A., Fernando R.L. & Dekkers J.C.M. (2010) Genomic selection in admixed and crossbred populations. *Journal of Animal Science* **88**, 32–46.
- Tuiskula-Haavisto M., Honkatukia M., Vilkki J., de Koning D.J., Schulman N.F. & Maki-Tanila A. (2002) Mapping of quantitative trait loci affecting quality and production traits in egg layers. *Poultry Science* **81**, 919–27.
- Visscher P.M. & Posthuma D. (2010) Statistical power to detect genetic loci affecting environmental sensitivity. *Behavior Genetics* **40**, 728–33.
- Weller J.I. (2009) *Quantitative Trait Loci Analysis in Animals*, 2nd edn, p. 272. CABI, Wallingford, Oxford.
- Wolc A., Lisowski M., Hill W.G. & White I.M.S. (2011) Inheritance of variation in production traits of laying hens. *British Poultry Science* **52**, 537–40.
- Yang J., Benyamin B., McEvoy B.P. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565–9.
- Yang Y., Christensen O.F. & Sorensen D. (2011) Use of genomic models to study genetic control of environmental variance. *Genetics Research* **93**, 125–38.