CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Wageningen University & Research Publications

Genetic analysis of production, immunity and behaviour in laying hens

Filippo Biscarini

Thesis committee

Thesis supervisor

Prof. Dr. Ir. Johan A.M. van Arendonk
Professor of Animal Breeding and Genetics
Wageningen University

Thesis co-supervisors

Dr. Ir. Jan J. van der Poel
Assistant Professor, Animal Breeding and Genomics Centre
Wageningen University

Dr. Ir. Henk Bovenhuis
Assitant Professor, Animal Breeding and Genomics Centre
Wageningen University

Other Members

Prof. Dr. Ir. Pim Brascamp
Wageningen University

Dr. Vincent Ducrocq
INRA, Jouy en Josas (France)

Prof. Dr. Ir. Bas Kemp
Wageningen University

Prof. Dr. Henner Simianer
University of Göttingen (Germany)

This research was conducted under the auspices of the Graduate School of the Wageningen Institute of Animal Science

Genetic analysis of production, immunity and behaviour in laying hens

Filippo Biscarini

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 13 September 2010
at 11 a.m. in the Aula

Filippo Biscarini
Genetic analysis of production, immunity and behaviour in laying hens,
132 pages.

PhD thesis: Animal Breeding and Genomics Centre, Wageningen University, Wageningen, NL (2010)

With references and including summaries in Dutch and English

ISBN 978-90-8585-786-0

Propositions

1. Pooled data can be effectively used for the genetic evaluation of farm animals (this thesis).

2. Genomic-wide association studies across lines poses statistical challenges but offer great opportunities (this thesis).

3. Providing enough food for the growing human population is a challenge that only with the help of livestock production modern agriculture can meet.

4. Professional mobility within Europe is encouraged by the EU but discouraged by the bureaucracy of individual member states.

5. Music, dancing, and arts in general are essential needs of human beings.

6. The ability to take a loss with a smile is essential for making progress

Propositions of the PhD thesis

Genetic analysis of production, immunity and behaviour in laying hens

Filippo Biscarini

Wageningen University

the 13th april 2010

Table of contents

Chapter 1

General introduction

INTRODUCTION

The farming of laying hens, with about 6.2 billions layers worldwide, is an important agricultural activity. More than 60 millions tonnes of eggs are produced each year globally, for direct consumption or to be used by the food industry in the many recipes that make use of eggs (e.g. fresh pasta, ice-creams, bakery, etc …). In The Netherlands alone there are 31.5 millions hens that produce 627000 tonnes of eggs per year for a commercial value of over 400 million euro (FAO, 2008). Behind all this there is a modern breeding industry which provides animals of high genetic merit for the development of egg production. Many different traits are analyzed in laying hens (Arthur and Albers, 2003): production (e.g. egg production per hen housed, persistency of lay), disease resistance (e.g. mortality, antibody titres), egg quality (e.g. egg weight, shell strength), metabolic traits (e.g. feed conversion) welfare-related traits (e.g. feather pecking behaviour). Based on some of these characteristics, highly specialized layer lines are developed. There are a few major breeding companies, operating in very diverse markets, facing a number of challenges, such as the changing regulations for the farming of layers and the "genomic revolution", that made, among other things, a great number of genetic markers available.

Future husbandry conditions

In the last decades layers have been mainly housed in battery cages usually consisting of four hens each (especially in Europe). In western countries, growing sectors of the public opinion have been expressing concerns about the standard husbandry conditions of laying hens, considered by many as disrespectful of their welfare. This led the EU to promulgate the directive 1999/74/EC, according to which egg production farms should move from the use of conventional cages to larger groups of birds (furnished cages, non-cage systems). Though beneficial for the welfare of layers, this will increase the antigenic pressure on the laying hens, that will share their environment with a higher number of mates, and will exacerbate feather pecking behaviour (Hughes and Duncan, 1972). Hens housed in groups tend to peck each other, causing a more or less pronounced damage to their plumage. This is recognized as detrimental for the well-being of laying hens (Craig and Muir, 1996), and has negative economic consequences. The denudation of body regions causes a substantial heat loss which can result in 20% higher energy requirements (Blokhuis and Wiepkema, 1998), and consequently greater feed consumption. Moreover, severe feather pecking can lead to cannibalism, thereby increasing the mortality rate of laying hens. Feather pecking has been traditionally kept under control by the practice of beak-trimming, which aimed at avoiding the

damage caused by the pecks. However, beak-trimming has been banned in the Scandinavian countries and in Austria, and has been strictly regulated in central European countries, where a transition period towards a complete ban has been adopted. In the Netherlands, for example, beak-trimming is expected to be abandoned by 2011. Countries in southern and eastern Europe limited themselves to the application of the directive 1999/74/EC, which allows beak-trimming only in the first 10 days of life, provided it is carried out by qualified personnel. Figure 1 illustrates the geographical situation with respect to beak-trimming (Van Horne, 2008).

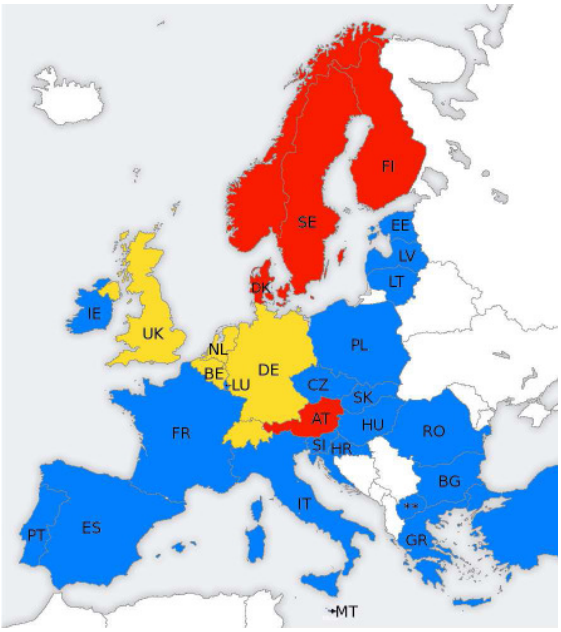


Figure 1.1. Beak-trimming ban in Europe. In countries coloured in red beak-trimming has been completely banned; in countries coloured in yellow a transition period towards a complete ban is in place; in countries coloured in blue the basic EU directive applies.

Robustness project

The research presented in this thesis is part of the larger project "The genetics of robustness in laying hens", which is a joint activity of Hendrix Genetics, a major international breeding company, and Wageningen University. The project was aimed at identifying parameters that are indicative of robust hens that could later be used in a breeding programme to improve robustness of layers. A robust hen is defined as 'an animal that has the potential to keep functioning and take short periods to recover under varying environmental circumstances' (Ellen et al., 2008).

A robust hen will be able to produce under future husbandry conditions (including ban of beak-trimming and conventional cages), thanks to the genetic background for immunological,

behavioural and physiological characteristics. Other PhD theses in the framework of this project looked at robustness of hens from the perspective of behaviour (Uitdehaag, 2008), immunity (Star, 2008) and genetics of survival and cannibalism (Ellen, 2009).

This thesis

This thesis mainly deals with two methodological aspects of layer breeding. The first originates from the structure of layer farming. Hens are housed in cages and this implies that their egg production is constituted by the pooled number of eggs laid per cage instead of the egg production of every individual hen. Chapters 2 and 3 deal with the use of such pooled data in genetic evaluations. The second methodological issue addressed in this thesis is the association between phenotypes and genetic markers in laying hens of different lines. Chapters 4 and 5 describe association studies for immunological parameters and feather pecking behaviour conducted across lines of laying hens.

In the general discussion (Chapter 6), I address the potential and the limitations of the use of pooled data in genetic evaluations, and the value of genetic markers for the estimation of genomic relationships between individuals and genetic distances between populations.

Use of pooled data in the genetic evaluation of laying hens

As pointed out, hens are usually housed in cages, and this leads to having pooled instead of individual egg records: in other words, the number of eggs laid is attributable to a cage as a whole but can not be assigned to individual hens, whose individual egg production is unknown. Current selection schemes are carried out in nucleus herds where hens are housed individually, so that their egg production can be recorded and used for genetic evaluations. Based on this information sires and dams are selected and lines of laying hens are developed. Maintaining nucleus herds where hens are housed in single cages is a considerable cost. It is however necessary, because in commercial farms, where hens are housed together in cages of usually four birds each, egg production is recorded per cage or stable and not individually. Observations recorded per group are called pooled data, and can consist of the sum of the performances of the individuals in a group, or of their average. A selection scheme based on individually housed hens introduces a discrepancy between the environment where hens are selected and the environment in which hens are kept for commercial egg production (group housing). Selecting animals in one environment and using them in a different environment might lead to the emergence of genotype x environment interactions (Besbes and Ducroq, 2003), thereby reducing the realized response to selection. The existence of G

x E interaction between the test and commercial situations was for example demonstrated in broilers (Zerehdaran et al., 2005) and in pigs (Merks, 1989). No studies so far have specifically look at G x E interaction between test and commercial environments in laying hens: it seems however reasonable to assume that G x E interaction between test and commercial situations is present also in layers. It would be therefore of interest to have the possibility of using pooled data in genetic evaluations, although the standard methods of genetic evaluation are designed for individual observations. To make this possible, first the theory of BLUP estimation of variance components, fixed effects and breeding values (Henderson, 1949; Henderson, 1950) has to be generalized in order to include pooled data. The mixed-model equations (MME: Henderson, 1950) need to be modified to take the group nature of the data into account. More than one genetic effect contributes to each pooled record, and this should be represented in the incidence matrices of the model. In the second place, this extended theory has to be applied to the genetic evaluation of farm animals, for the estimation of genetic parameters and the prediction of breeding values. In this way genetic evaluations of layers would be more flexible, being able to process pooled data, and more adherent to the practice of commercial egg production, thus avoiding potential problems of G x E interactions. Chapters 2 and 3 of this thesis deal with the use of pooled data for the estimation of heritability and the prediction of breeding values in laying hens.

Across-line association studies

The number of markers discovered and available in poultry, as well as in other livestock species, has increased dramatically over the last few years, and the cost for genotyping a single animal has in the meantime plummeted. This abundance of genomic information is a great opportunity for animal breeding, since it makes fine-mapping of QTLs and genomic selection for many different traits of zootechnical interest possible. At the same time it is also a challenge for the breeding industry, because it presents new problems in the statistical analysis of data and in the use of genomic information in selection decisions and breeding strategies.

Several QTL mapping experiments have been carried out in chickens (*Gallus gallus domesticus*), aimed at finding genes or loci having an effect on a diverse range of traits: among others, growth-related traits such as body weight, egg production traits such as number of eggs laid, age at first egg or egg weight, metabolic traits such as feed efficiency, and disease resistance traits like antibody response to specific antigens, as reviewed by Abasht et al. (2006). Initially, due to the limited number of available markers and the high costs for genotyping, analyses were performed within families, tracing the transmission of chromosomal segments from parents to offspring in order to detect segregating QTLs: this approach, known as linkage analysis, had a low resolution and the

position of QTLs could not be accurately determined. Besides, due to different association phases, results were valid exclusively within families. In recent years it has become feasible to genotype individuals for many markers at relatively low costs. This made genome-wide studies possible, not within but across the families of a single population (line or breed). There is less LD across families than within them, but the higher marker density still allows for QTL detection, at a much higher resolution compared to linkage analysis. The focus on homogeneous populations is suggested by the higher amount of LD conserved within-populations as compared to between-populations. In this thesis we propose to take one step further: association studies across lines. This requires higher marker density but increases the resolution even further.

Analysing multiple lines simultaneously poses the challenge of dealing with genealogically different populations: there is less LD between the marker and the QTL and, due to recombination events, the phase of the marker-phenotype association might be different in the different lines. These problems in conducting marker-phenotype association studies across populations were addressed in Chapters 4 and 5 of this thesis, where it was shown how to deal with multiple populations when analyzing hens from 9 different genetic lines of White Leghorn and Rhode Island Red origin genotyped for a panel of 1536 SNP (Single Nucleotide Polymorphism) markers.

The traits analysed were immunological parameters and plumage damage due to feather pecking behaviour, two classes of traits for which, given that they have low heritability and are difficult and expensive to measure, genomic information may be particularly valuable. Immunological parameters might be used in selection programmes aimed at improving disease resistance of laying hens, while information on the genetic background of feather pecking behaviour can be useful in reducing problems due to this behavioural disorder of layers. Under future husbandry conditions susceptibility to infectious diseases and feather pecking are expected to become more serious problems: both aspects of layer production are in fact highly related to stocking density, which will increase as a result of the application of the EU directive 1999/74/EC. In addition, the ban of beak-trimming will make it more difficult to control the consequences of feather pecking (plumage damage, cannibalism, mortality). Genetic selection might represent an appealing addition to the current control measures.

REFERENCES

Abasht B., J.C.M. Dekkers and S.J. Lamont (2006) Review of quantitative trait loci identified in the chicken. *Poult. Sci*. **85**:2079-2096.

Arthur J.A., M.H. Pinard and G.A.A. Albers (2003) Industrial perspective on problems and issues associated with poultry breeding. In: *Poultry Genetics, Breeding and Biotechnology* (ed. by Muir & Aggrey), pp. 1-5. CAB International, Wallingford, Oxon OX10 8DE, UK.

Besbes B. and V. Ducroq (2003) Pages 139-140 in Poultry Genetics, Breeding and Biotechnology. Muir and Aggrey editors. CAB International, Wallingford, Oxon OX10 8DE, UK.

Blokhuis H. J. and Wiepkema P. R. (1998). Studies of feather pecking in poultry. *Vet. Quart.* **20**:6-9.

Craig J.V. and W.M. Muir (1996) Group selection for adaptation to multiple-hen cages: beak-related mortality, feathering and body weight responses. *Poult. Sci.* **75**:294-302.

Ellen D.E., L. Star, K.A. Uitdehaag and F.W.A. Brom (2008) A plea to implement robustness into a breeding goal: poultry as an example. J. Agric. Environ. Ethics, 21:109-125.

Ellen D.E. (2009) Genetics of survival in cannibalistic laying hens. Wageningen University, The Netherlands.

Hughes B.O. and I.J.H. Duncan (1972) The influence of strain and environmental factors upon feather pecking and cannibalism in fowls. *Br Poult Sci* **13**:525-547

Merks J.W.M. (1989) Genotype x Environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. *Livest. Prod. Sci*. **22**:325-339

Henderson C.R. (1949) Estimation of changes in herd environment [abstract]. *J. Dairy Sci.* **32**:709.

Henderson C.R. (1950) Estimation of genetic parameters. *Ann. Math. Statist.* **21**:309.

Star L. (2008) Robustness in laying hens – Influence of genetic background, environment and early-life experiences. Wageningen University, The Netherlands.

Uitdehaag K. (2008) Effects of genetic background and social environment on feather pecking and related behavioural characteristics in laying hens. Wageningen University, The Netherlands.

Van Horne P.L.M. and T.J Achterbosch. (2008) Animal welfare in poultry production systems: impact of EU standards on world trade. *Worlds Poult. Sci. J.* **64**:40-52.

Zerehdaran S., A.L.J. Vereijken, J.A.M. van Arendonk and E. H. van der Waaij (2005) Effect of age and housing system on genetic parameters fro broiler carcass traits. *Poult. Sci*. **84**:833-838

Chapter 2

Estimation of variance components and prediction of breeding values using pooled data

F. Biscarini, H. Bovenhuis and J. A. M. van Arendonk

Animal Breeding and Genomics Centre
Wageningen University
P.O. Box 338, 6700 AH Wageningen

The Netherlands

Journal of Animal Science (2008) 86:2845-2852

ABSTRACT

Data on performance of animals are, in a number of situations, collected at group rather than individual level. Genetic evaluations in farm animals, however, are based on phenotypic information collected at the individual level. Therefore, it would be very attractive to extend genetic evaluations by incorporating information collected at group level. In this paper we show the use of data collected at group level for the estimation of variance components and the prediction of breeding values. We outline a general procedure that can be applied in different farm animal species. In the present work this procedure was applied to body weight, for which pooled as well as individual observations were available, thus allowing for a comparison of the estimates, and to egg production, for which only pooled data were available. For body weight at 19 and 27 weeks the estimated heritabilities based on individual observations where very similar to those based on pooled observations. For body weights at 43 and 51 weeks heritability estimates based on individual and pooled data were different, which can be caused by the emergence of competition effects. The accuracy of EBVs predicted from pooled observations was about 70-80% of the accuracy of EBVs predicted from individual observations. This result quantifies the loss deriving from the use of pooled instead of individual observations. Results show that the estimation of variance components and breeding values from pooled data instead of individual observations is theoretically and practically feasible.

Keywords: genetic evaluation, groups, laying hens, cage performances.

INTRODUCTION

Data on performance of animals are, in a number of situations, collected at group rather than individual level. The reason for this might be that animals are housed in groups and collection of information at individual level under these circumstances is too expensive or too complicated. For instance, information on feed intake and feed conversion rate is in many species measured at group level. Also egg production of laying hens is recorded at group level under commercial conditions. This is especially relevant for species that are routinely housed in cages, tanks or pens, such as poultry, fishes and pigs, where data can be recorded for the whole group.

Moreover, in current breeding schemes pure lines in nucleus breeding herds are usually housed individually and selected on the basis of individually recorded traits, whereas at the commercial level their offspring are kept in groups. Individual housing at the nucleus breeding herds is not only more expensive, but might also give rise to Genotype by Environment interaction (GxE) (e.g. Besbes and Ducroq, 2003) that can reduce the response to selection at the commercial level. Therefore, it would be very attractive to extend genetic evaluations of farm animals by incorporating information collected at group level. This was recognized by Olson et al. (2006) who used simulations to calculate breeding values and accuracies for pooled and individual observations.

The objective of this paper is to develop a method for estimating genetic parameters and predicting breeding values from information collected on groups of animals (pooled information) and compare its relative efficiency with the use of information collected on individual animals. For our analysis, data on individual and pooled body weights of laying hens were available. The methodology was also applied to the trait egg production, for which no individual measurements were available.

MATERIAL AND METHODS

Description of data

The animal population used in this study consisted of about 2500 laying hens from 12 distinct genetic lines housed together in a single stable of an experimental farm for a laying period of 51 weeks (weeks 19 – 69 in terms of age of the hens), from June 2004 to June 2005. These lines were either of Rhode Island Red type (RIR; laying brown eggs) or of White Leghorn type (WL; laying white eggs).

Both during the rearing and the experimental period, the hens were housed in a total of 641 cages of different size (from 2 to 5 animals); only cages with 4 hens were used in this study (≈ 560 cages). About half of the cages were composed of full sibs, while the other half had a random composition. The number of hens and cages per line and origin is presented in Table 1.

Table 1: Distribution per origin and line of the hens and cages used in this study

Rhode Island Red			White Leghorn		
line	number of hens	number of cages	line	number of hens	number of cages
B1	196	49	W1	180	45
B2	200	50	WA	196	49
B3	172	43	WB	188	47
BA	188	47	WC	200	50
BB	196	49	WD	152	38
BE	188	47	WF	188	47
total	*1140*	*285*	*total*	*1104*	*276*

The hens had intact beaks and routine vaccinations against Marek's disease, New Castle disease, infectious bronchitis, infectious bursal disease, fowl pox and avian encephalomyelitis were applied during rearing. During the experiment, feed and water were available ad libitum. From the beginning of the experiment (at 19 weeks of age) until 42 weeks of age hens were fed a standard commercial phase 1 diet (159 g/kg crude protein, 43 g/kg crude fibre and 11.17 MJ ME/kg); from 42 weeks onwards, until the end of the experiment, a standard commercial phase 2 diet (152 g/kg crude protein, 47.0 g/kg crude fibre and 11.01 MJ ME/kg) was given. Wing bands allowed individual identification of the hens. Hens arrived at the laying facility at 17 weeks of age and were kept at a 9L: 15D light scheme (light from 7.00 until 16.00), where L stands for light and D for darkness. After one week, the light period was increased by half an hour, starting at 6.30. Hereafter, the light period was increased approximately 10 min per day. From 30 weeks onwards hens received light from 00.00 until 16.00 (16L : 8D).

The data consisted of individual records of the body weight of the hens at 19 (BW19, onset of the egg production), 27 (BW27), 43 (BW43) and 51 (BW51) weeks of age. These were then pooled (i.e. summed by cage) to derive the group observations, thus allowing for the comparison between genetic evaluations from individual and pooled observations. Because of software limitations only cages with 4 hens were used. As a consequence of mortality the number of cages with 4 hens decreased over time. In addition to body weight, egg production of the entire laying period (weeks 19 - 69) of 550 cages was used. For this trait there were no individual observations, but only pooled data were available.

The 4 body weight traits were all normally distributed. To account for non-normality of egg production, the power transformation method of Box and Cox (1964) was applied to produce standardised variates, z(t), according to the following formula:

$$z(t) = \frac{y^t - 1}{tG_y^{t-1}}$$

where y is the original observation, G_y the geometric mean of the data and t is the parameter by which data are normalized. The value of t=5, empirically derived, was used to approximate a normal distribution.

From the pedigree files 4 generations of ancestors were extracted for the genetic analysis; sires have 1 to 30 daughters while dams have 1 to 5 daughters. This is consistent with a hierarchical structure in which a rooster is mated to more females whereas a female is only mated to one male.

Theoretical background

Following the work of Olson et al. (2006), we set up the mixed model equations (MME) for the prediction of breeding values and the estimation of variance components using group observations.

In the case of individual records, the model, in matrix notation, is:

$$\mathbf{y = Xb + Za + e} \qquad (1)$$

with **y** being the vector of observations, **b** and **a** the vectors of fixed and random effects, with their respective incidence matrices **X** and **Z,** and **e** a vector of random residuals. Var(**y**) = Var(**a**) + Var(**e**) which, after substitution, becomes: $Var(\mathbf{y}) = \mathbf{ZGZ'} + \mathbf{R}$, with $\mathbf{G} = \mathbf{A}\sigma_a^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$.

Solutions for fixed effects and BLUP of **a** (i.e. $\hat{\mathbf{a}}$) can be then obtained by solving the usual MME (Henderson, 1975; Henderson, 1984):

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\boldsymbol{\alpha} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

In a numerical example, we have 4 founding animals and measured individual records on 6 of their progeny, 2 from group A, 2 from group B and 2 from group C; α, the ratio between residual and genetic variance, is 2 (h^2 of 0.33). The pedigree and data for the example are shown below:

animal	sire	dam	trait	group
1	0	0	-	
2	0	0	-	
3	0	0	-	
4	0	0	-	
5	1	2	10	A
6	1	2	15	A
7	1	2	12	B
8	3	2	9	B

9	3	4	11	C
10	1	4	16	C

Only the fixed effect of the mean is considered, and the matrices for the MME have the usual appearance:

$$\mathbf{y}=\begin{bmatrix}\mathbf{10}\\ \mathbf{15}\\ \mathbf{12}\\ \mathbf{9}\\ \mathbf{11}\\ \mathbf{16}\end{bmatrix},\ \mathbf{X}=\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\end{bmatrix},\quad \mathbf{b}=\begin{bmatrix}\hat{\boldsymbol{\mu}}\end{bmatrix},\ \mathbf{Z}=\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix},\quad \mathbf{a}=\begin{bmatrix}\hat{\mathbf{a}}_1\\ \hat{\mathbf{a}}_2\\ \hat{\mathbf{a}}_3\\ \hat{\mathbf{a}}_4\\ \hat{\mathbf{a}}_5\\ \hat{\mathbf{a}}_6\\ \hat{\mathbf{a}}_7\\ \hat{\mathbf{a}}_8\\ \hat{\mathbf{a}}_9\\ \hat{\mathbf{a}}_{10}\end{bmatrix},\quad \mathbf{e}=\begin{bmatrix}\hat{\mathbf{e}}_1\\ \hat{\mathbf{e}}_2\\ \hat{\mathbf{e}}_3\\ \hat{\mathbf{e}}_4\\ \hat{\mathbf{e}}_5\\ \hat{\mathbf{e}}_6\\ \hat{\mathbf{e}}_7\\ \hat{\mathbf{e}}_8\\ \hat{\mathbf{e}}_9\\ \hat{\mathbf{e}}_{10}\end{bmatrix}$$

In this case, $Var(\mathbf{a}) = \mathbf{G} = \mathbf{A}\sigma_a^2$, where **A** is the additive relationship matrix, and $Var(\mathbf{e}) = \mathbf{R} = \mathbf{I}\sigma_e^2$, where **I** is an identity matrix, i.e. assuming that there are no residual correlations between animals of the same group. Taking the diagonal elements of the inverse of the MME (Henderson, 1975), the accuracies of predictions can be calculated with the following formula:

$$r = \sqrt{1 - d_i \alpha}$$

where d_i is the i_{th} diagonal element of the MME^{-1} and α is the ratio of variances (σ_e^2 / σ_a^2). The BLUP solutions for the mean and the breeding values of the 10 animals, obtained by direct inversion of the MME, are shown in Table 2.

Suppose the 6 phenotypes are not recorded individually, but pooled by group: in this case the observations are the group sums and the model has to be modified accordingly:

$$\mathbf{y}^* = \mathbf{X}^*\mathbf{b} + \mathbf{Z}^*\mathbf{a} + \mathbf{e}^* \qquad (2)$$

Where the * symbol indicates the modified elements of the equation. The MME therefore are:

$$\mathbf{y}^*=\begin{bmatrix}\mathbf{10+15}\\ \mathbf{12+9}\\ \mathbf{11+16}\end{bmatrix}=\begin{bmatrix}\mathbf{25}\\ \mathbf{21}\\ \mathbf{27}\end{bmatrix},\quad \mathbf{X}^*=\begin{bmatrix}\mathbf{2}\\ \mathbf{2}\\ \mathbf{2}\end{bmatrix},\quad \mathbf{Z}^*=\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}\end{bmatrix},\quad \mathbf{e}^*=\begin{bmatrix}\mathbf{e}_1+\mathbf{e}_2\\ \mathbf{e}_3+\mathbf{e}_4\\ \mathbf{e}_5+\mathbf{e}_6\end{bmatrix}$$

Groups are now the experimental units (the observations), and therefore some of the elements of the Mixed Model Equations must consequently be modified. The matrices $\mathbf{X}^*$ and $\mathbf{Z}^*$ now reflect the composition of each group: in the **X** matrix, the number of times each fixed effect is present in each group is given, while in the matrix **Z** in each row there are as many 1s as animals in the group. The vector $\mathbf{y}^*$ consists of the sums of the individual observations and $\mathbf{e}^*$ is the sum of the residuals. Vector **a** remains unchanged and in this example with only a general mean the vector **b** also remains unchanged. The variance of the genetic effect is also unmodified and equal to $\mathbf{A}\sigma_a^2$; on the other hand, because $\mathbf{e}^* \neq \mathbf{e}$, the variance of the residuals is now $Var(\mathbf{e}^*) = \mathbf{R}^* = \mathbf{D}\sigma_e^2$: assuming that there are no residual correlations between animals of the same group, **D** is not an identity matrix but a diagonal matrix with elements n_j representing the number of animals that contributed to the j_{th} pooled observation . This affects also the value of α, which is now n times that for individual observations. In our example where we have two observations in each group $\alpha^* = 2 \times \alpha = 4$. If groups are not of equal size, the relation $\alpha^* = n \times \alpha$ does not hold and R^{*-1} must be used in the MME.

Table 2: EBVs and accuracies for the animals of the numerical example when analysed individually or as pooled observations (in groups of two).

	individual		**pooled**	
effect	*solutions*	*accuracies*	*solutions*	*accuracies*
μ	12.1250		12.2228	
a_1	0.7083	0.3227	0.0891	0.1723
a_2	-0.4583	0.3227	-0.4257	0.3224
a_3	-0.7083	0.3227	-0.0891	0.1723
a_4	0.4583	0.3227	0.4257	0.3224
a_5	-0.3250	0.4387	-0.0792	0.3605
a_6	0.6750	0.4387	-0.0792	0.3605
a_7	0.0750	0.4387	-0.4703	0.2943
a_8	-1.0917	0.4463	-0.5594	0.3107
a_9	-0.3250	0.4822	0.3812	0.3690
a_{10}	1.2417	0.4463	0.4703	0.2943

With this approach, BLUP of the fixed and random effects can be obtained in the very same way from the pooled observations, setting up the MME as previously outlined and solving them either by direct inversion or iteratively. The solutions for the animals of the numerical example are reported in Table 2. For some animals EBVs from pooled data differ considerably from EBVs based on individual observations. The example illustrates that for instance full sib individuals 5 and 6,

who are in the same group (A), have the same EBV when estimates are based on pooled observations. Individual 7, which is also a full sib of individuals 5 and 6 but is in another group (B), has an EBV different from 5 and 6. Following the procedure outlined by Mrode (2005), the solutions can be partitioned into parent averages (PA), yield deviations (YD) and progeny contributions (PC); this helps explaining the differences between the two models. The main reason behind the observed differences is that the pooling of records leads to a loss of information and a reduced variation: in the case of pooled data, deviations from the mean are in general smaller than those computed from individual records. Animal 6, for instance, has an individual record of 15 which is a 20% deviation from the overall mean (12.25); its pooled observation is 25 which deviates only 2% from the overall mean of the pooled data (24.5). Its weighted yield deviation is therefore 0.575 in the individual model and 0.0616 in the pooled model and this explains the different EBVs in the two situations (0.675 vs -0.079). The same applies to animal 3, whose EBV depends entirely on the contributions of its progeny; its two offspring have YDs of -26% and -10%, and -14% and +10% in the individual and pooled models respectively, thus explaining the difference observed in its EBV in the two situations (-0.7083 vs -0.0891). The reduced variation of pooled observations not only affects the yield deviations but also the parent averages and the progeny contributions.
The loss of information and the reduced variation in the case of pooled observations as compared to individual observations leads to a lower accuracy of the estimates, especially for individuals with phenotypic observations only (and no offspring).

Data analysis

Variance components and breeding values for body weight and egg production were estimated from the MME with a REML procedure. The Asreml software package was used for the analysis (Gilmour, 2002) The function $and(\ )$ in Asreml was used to fit multiple genetic effects per observation. For the analysis of individual and pooled data, respectively, the models described in equations 1 and 2 of the previous section were used. Since all hens were kept in a single stable and on the same tier of the battery system, the only fixed effect that has been considered in this study was the effect of line.

RESULTS

Variance components

Basic statistics of the data are summarized in Table 3. The number of hens in the analysis gradually decreases in time due to mortality and the restriction that only cages with 4 hens were included. The

average body weight at 19 weeks was 1405 grams and increases gradually to 1974 gram at 51 weeks. The coefficient of variation for individual observations of body weight is approximately 13%. As pooled observations are based on 4 individuals, standard deviations of pooled observations are expected to be twice that of the individual observations, i.e. for independent observations. Table 3 shows that the standard deviations for pooled observations are 3.2 to 3.3 times the standard deviation of individual observations suggesting positive covariances among observations on cage mates.

Table 3: Descriptive statistics of the body weight traits and for egg production, both for the individual observations and for the sum of the performance of 4 hens in one cage (pooled).

	BW19[1]		**BW27**[2]		**BW43**[3]		**BW51**[4]		**EGGS**[5]
	individual	**pooled**	**individual**	**pooled**	**individual**	**pooled**	**individual**	**pooled**	**pooled**
Number	2244	561	2128	532	1932	483	1776	444	371
Mean	1405	5621	1693	6773	1844	7375	1874	7497	1113
std dev	198	658	203	678	243	781	248	789	76.5

[1]BW19 = body weight at 19 weeks; [2]BW27 = body weight at 27 weeks; [3]BW43 = body weight at 43 weeks; [4]BW51 = body weight at 51 weeks; [5]EGGS = total egg production for a 51 weeks laying period.

Table 4 shows the estimates of the variance components and heritabilities, together with their standard errors, based on individual observations and based on pooled observations from 4 hens per cage. Estimated heritability of body weight of laying hens at different ages is moderate to high and in the range of 0.4 – 0.6, which agrees with results from literature (Besbes et al., 1992; Kabir et al., 2006). Heritability estimates of body weight from individual and pooled observations were very similar at younger ages (BW19 and BW27) but started to differ at older ages (BW43 and BW51), although estimates are not significantly different. Theoretically it is expected that the residual variance estimated based on pooled observations is four times that estimated based on individual observations; results are in close agreement with this expectation for BW19 and BW27 (ratios of 4.73 and 4.27), less for BW43 and BW51 (ratios of 5.95 and 6.27). Table 3 shows that the standard error of the heritability estimates from pooled observations is about twice that when using individual observations: around 0.11 for pooled observations and around 0.05 when based on individual observations.

For egg production only pooled observations were available and therefore the comparison with the estimates from individual observations was not possible. The estimated heritability for egg production is 0.34 and this is in line with what has been reported in literature (Wei and Van der Werf, 1993; Anang et al., 2000). The standard error of the estimate is 0.13 and of the same order of magnitude as for body weight based on pooled observations.

Table 4: Variance components, h^2 estimates and standard errors for body weight traits and for egg production based on individual and pooled observations

Trait [a)]	analysis	σ_a^2	σ_e^2	$h^{2\ b)}$	*SE*
BW19	individual	10872	8574	0.56	0.05
	pooled	10747	40618	0.51	0.11
BW27	individual	11059	8052	0.58	0.05
	pooled	10901	34408	0.56	0.12
BW43	individual	16777	15682	0.52	0.06
	pooled	12219	93289	0.34	0.10
BW51	individual	21449	14488	0.60	0.06
	pooled	16247	90896	0.42	0.12
EGGS	pooled	241.32	1878.5	0.34	0.13

[a)]BW19 = body weight at 19 weeks; BW27 = body weight at 27 weeks; BW43 = body weight at 43 weeks; BW51 = body weight at 51 weeks; EGGS = total egg production for a 51 weeks laying period.

[b)] Heritability estimates for pooled observations were calculated as $\frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2/4}$ where 4 is the number of hens per cage.

Breeding values

Table 5 shows the mean accuracy of the estimated breeding values for all hens with an observation, and for sires and dams with at least 10 and 4 progeny respectively. The accuracy of EBVs from pooled observations is consistently lower than that of EBVs from individual records. The decrease in accuracy for pooled observations is greater for hens with observations than for sires with at least 10 offspring or for dams with at least 4 offspring.

Selection index theory predicts that, in the case of unrelated group mates, the accuracy of EBVs based on the cage performances of *n* individuals (pooled data) decreases with the number of individuals per cage as $r_{IH-pooled} = \sqrt{h^2/n}$, where n is the number of unrelated cage-mates. If cages consist of half sibs and full sibs, the accuracy of EBVs predicted from pooled data will asymptotically reach a value of 0.50 and 0.707 in the cases of half sibs and full sib cage mates respectively. To verify the decrease in accuracy when increasing the group size, we randomly pooled BW19 observations of 2, 3, 4, 6 and 8 animals originating from any cage, and calculated mean accuracies. The mean accuracies for hens with an observation were 0.766 when we used individual observations, 0.684 when based on pooled data of two hens, 0.569 when based on pooled data of three hens, 0.552 when based on pooled data of four hens, 0.405 when based on pooled data

of five hens and 0.330 when based on pooled data of eight hens. These results show that accuracies clearly decrease if the group size increases as was also shown by Olson *et al* (2006) and as would have been expected based on selection index calculations. However, the accuracies thus calculated decrease following a different pattern than that theoretically expected on the basis of unrelated cage mates, $r_{IH-pooled} = \sqrt{h^2/n}$. The reason is that in the experimental dataset used in this study there were not only unrelated animals but also cages of half or full sibs; moreover, additional pedigree relations were present.

Table 5: Correlations between EBVs based on individual observations and based on the pooled cage performance of 4 hens and the accuracies of the EBVs for all hens with observations, sires with more than 10 offspring and dams with more than 4 offspring.

trait[a)]	**analysis**	**correlations**			**accuracy**		
		pooled			mean		
		all hens with observations	sires ≥ 10 daug. (96)[b)]	dams ≥ 4 daug. (302)[b)]	all hens with observations	sires ≥ 10 daug[b)]	dams ≥ 4 daug[b)]
BW19	individual	0.748	0.891	0.875	0.765	0.751	0.617
	pooled				0.552	0.612	0.513
BW27	individual	0.744	0.887	0.880	0.753	0.738	0.593
	pooled				0.527	0.621	0.523
BW43	individual	0.732	0.812	0.868	0.718	0.724	0.575
	pooled				0.433	0.533	0.440
BW51	individual	0.703	0.814	0.847	0.759	0.756	0.628
	pooled				0.461	0.550	0.460
EGGS	pooled				0.417	0.504	0.450

[a)]BW19 = body weight at 19 weeks; BW27 = body weight at 27 weeks; BW43 = body weight at 43 weeks; BW51 = body weight at 51 weeks; EGGS = total egg production for a 51 weeks laying period.

[b)] the number of individuals on which these correlations were based differed between traits and varied between 86 and 96 for sires and 235 and 302 for dams

Table 5 also shows correlations between the estimated breeding values based on individual and pooled observations for three groups of individuals: all hens with an observation, sires with more than 10 daughters and dams with more than 4 daughters. The correlations between EBVs from individual and pooled observations are between 0.70 and 075 when all hens with an observation are considered. For sires with at least 10 daughters correlations are between 0.81 and 0.89 depending on the trait. For dams with 4 daughters correlations between EBVs were between 0.85 and 0.88.

DISCUSSION

Results show that the estimation of variance components and the prediction of breeding values from pooled data instead of individual observations is theoretically and practically feasible. Earlier attempts to incorporate group observations into genetic evaluations were made by Simianer and Gjerde (1991) who used the mean performances from groups of full-sibs to estimate variance components for body weight in salmons, and by Nurgiartiningsih *et al*. (2002) and Wei and Van der Werf (1995) who both used cage means in a sire model for the estimation of genetic parameters for egg production traits in laying hens. Following the work of Olson *et al* (2006), based on simulated data, we developed a method based on modified MME using an animal model, and applied it to real data. The data available in the present study made it possible to compare genetic evaluations based on individual and pooled observations.

The results show that for the same number of animals, EBVs and heritability estimates based on pooled data are less accurate than those based on individual observations. However, in practice comparison of estimates based on equal number of phenotypes rather than based on equal number of individuals might be more appropriate. The use of pooled data offers some advantages: they are often easier and cheaper to collect, they might be in some cases the only data available, and they may sometimes better reflect the commercial environment where animals are kept, thus avoiding the bias due to possible GxE interactions.

At the commercial level only pooled data are available. This data may be of great interest for breeding companies to select pure line individuals. As animals at the commercial level are almost always crossbred individuals, this aspect should be accounted for (e.g. Wei and Van der Werf, 1995)

Presence of GxE interaction for animals kept individually or in groups will reduce the response to selection at the commercial level. Merks (1989) found a genetic correlation of 0.64 for backfat thickness measured on individually housed pigs at central test stations and on groups housed pigs on commercial fattening farms. In broilers, Zerehdaran et al. (2005) estimated a genetic correlation of 0.80 between body weight in group housing and individual cages. These estimates clearly

Suggest the presence of GxE interaction for individually and group housed animals.

Therefore, selecting individuals based on group performance may be very convenient in many situations. It will make genetic evaluation of farm animals more flexible and more accurate by including information collected under commercial conditions where animals are kept in groups. The use of group sums or group means as pooled observations proved to be equivalent (results not shown).

Accuracies of Selection

The accuracy of selection based on pooled observations, for all hens with records and sires and dams with at least 10 or 4 offspring is, respectively, 65-72% and 72-88% of the accuracies that can be obtained when using individual observation (Table 5). However, this reduced accuracy should be interpreted in the context of direct versus indirect selection. The breeding goal of a breeding company is the trait under commercial conditions, i.e. group housing. If testing is under individual housing, the genetic correlation between group and individual housing is relevant. For the same selection intensity, the ratio of the selection response for direct and indirect selection is a function of the accuracies for both situations and the genetic correlation between both traits (Falconer, 1989). Similarly, the ratio between accuracies based on pooled and individual data provides a threshold for the genetic correlation between individual and group housing below which pooled data would result in a greater selection response (for the same selection intensity). Therefore, in the present study a genetic correlation below 0.72 for sires would compensate for the loss of accuracy due to pooled observations.

Fixed effects

The relationship matrix enables the estimation of the breeding values of each animal based on pooled observations. However, some differences in the definition of systematic environmental effects may arise when using either individual or pooled records. In the numerical example presented in this paper only the fixed effect of the mean was considered and the vector of fixed effects was the same in both models, but this will not always be the case. Examples are when pooled phenotypes are being produced by animals of different sex or breeds. In these situations the design matrix for the fixed effects must be modified in order to reflect the contribution of the fixed effect classes to the pooled phenotype. A consequence will be that systematic environmental effects can be estimated less accurately for pooled data when compared to a situation were individual observations are available.

Group composition

In the present study about half of the cages were composed of full sibs and the other half had a random composition. The numerical example used to outline the procedure illustrated that two full sibs in group A (animals 5 and 6) had the same EBV when estimated based on pooled data and therefore the composition of the groups will have an impact on the breeding value estimation and

the estimation of variance components. Group observations are in fact disentangled through the pedigree structure. Family groups seem to be better for productivity and social welfare (Bijma, personal communication); therefore, on one hand having groups of closely related animals pose statistical challenges, on the other hand they offer a better social environment and genetic background for animal performances and welfare. The appropriate cage composition needs therefore to be chosen in order to balance these two contrasting aspects.

Effects of selection and competition

According to theory, the residual variance estimated from pooled observations should be four times that estimated from individual observations. Results are in close agreement with this hypothesis for BW19 and BW27, less for BW43 and BW51. The reason why results diverge from theoretical expectations later in life, might in part be that we limited this study to groups with 4 hens per cage. Excluding cages with less than 4 individuals might introduce a culling bias. This may be partly accounted for by using a multivariate approach, where the unselected trait (BW19) is analyzed simultaneously with the selected traits (BW27, BW43 and BW51), and all animals have observations on at least the trait on which selection is based (Pollak et al., 1984).
An alternative explanation might be the presence of within group competitive effects (Bijma et al., 2007a,b), which have been shown to play a role in a number of traits of agricultural interest as, for instance, survival in laying hens and growth rate (Ellen et al, 2007). The phenotype of an individual might depend not only on its own genotype but also on the phenotype or the genotypes of group mates and this alters the variance structure of the model. The consequences of these competition effects might accumulate over time and in this way bias the estimated genetic parameters. If competition effects play a role, this will affect both the analyses based on individual observations as well as those based on pooled data, but they might be affected differently. Therefore, in such cases, inclusion of the associative genetic effects in the analysis might be required (Bijma et al., 2007a,b).

This paper indicates that pooled observations can be used to estimate variance components and breeding values. The use of pooled data therefore, offers interesting possibilities for breeding companies. In this study we focused on the use of pooled data from groups of the same size; there is therefore scope for the extension to groups of any size. We identified that heritability estimates based on pooled data and individual observation tended to differ at older ages which might be due to effects of selection and/or competition.

ACKNOWLEDGEMENTS

This work was conducted as part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission.

REFERENCES

Anang, A., N. Mielenz and L. Schüler. 2000. Genetic and phenotypic parameters for monthly egg production in White Leghorn hens. J. Anim. Breed. Genet. 117:407-415.

Besbes, B., V. Ducrocq, J. L. Foulley, M. Protais, A. Tavernier, M. Tixier-Boichard and C. Beaumont. 1992. Estimation of genetic parameters of egg production traits of laying hens by restricted maximum likelihood applied to a multiple-trait reduced animal model. Genet. Sel. Evol. 24:539-552.

Besbes, B. and V. Ducroq 2003. Pages 139-140 in Poultry Genetics, Breeding and Biotechnology. Muir and Aggrey editors. CAB International, Wallingford, Oxon OX10 8DE, UK.

Bijma, P., W. M. Muir and J.A.M. Van Arendonk. 2007a. Multilevel Selection 1: Quantitative Genetics of Inheritance and Response to Selection. Genetics 175:277–288.

Bijma, P., W. M. Muir, E. D. Ellen, J. B. Wolf and J.A.M. Van Arendonk. 2007b. Multilevel Selection 2: Estimating the Genetic Parameters Determining Inheritance and Response to Selection. Genetics 175:289–299.

Box, G.E.P. and D. R. Cox. 1964. An analysis of transformations. J. R. Statist. Soc. B26:211-243.

Ellen, E. D., W. M. Muir, F. Teuscher, P. Bijma. 2007. Genetic improvement of traits affected by interactions among individuals: sib selection schemes. Genetics 176:489-499.

Falconer, D. S. 1989. Pages 318-319 in Introduction to quantitative genetics. 3rd ed. Longman Scientific & Technical, New York, NY.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, R. Thompson. 2002. ASReml User Guide Release 1.0 VSN International Ltd. Hemel Hempstead, HP1 1ES, UK.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447.

Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. Guelph Press.

Kabir, M., O. O. Oni, G. N. Akpa and I. A. Adeyinka. 2006. Heritability estimates and interrelationships of body weight and shank lengths in Rhode Island Red and White chickens. Pak. J. Biol. Sci. 9(15):2892-2896.

Merks, J.W.M. 1989. Genotype x Environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. Livest. Prod. Sci. 22:325-339

Mrode, R. A. 2005. Pages 45-48 in Linear models for the prediction of animal breeding values. 2nd ed. CAB International, Wallingford, Oxon OX10 8DE, UK

Nurgiartiningsih, V., N. Mielenz, R. Preisinger, M. Schmutz and L. Schüler. 2002. Genetic parameters for egg production and egg weight of laying hens housed in single and group cages. Arch. Tierz. 45:501-508.

Olson, K. M., D. J. Garrick and R. M. Enns. 2006. Predicting breeding values and accuracies from group in comparison to individual observations. J. Anim. Sci. 84:88-92.

Pollak, E. J., J.H.J. van der Werf and R. L. Quaas. 1984. Selection bias and multiple trait evaluation. J. Dairy Sci. 67:1590-1595.

Simianer, H. and B. Gjerde. 1991. Estimating variance components from fullsib groups means. J. Anim. Breed. Genet. 108:270-279.

Wei M. and J.H.J. van der Werf. 1993. Animal model estimation of additive and dominance variances in egg production traits of poultry. J. Anim. Sci. 71:57-65.

Wei M. and J.H.J. van der Werf 1995. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. J. Anim. Sci. 73:2220-2226.

Zerehdaran, S., A.L.J. Vereijken, J.A.M. van Arendonk and E. H. van der Waaij. 2005. Effect of age and housing system on genetic parameters fro broiler carcass traits. Poult. Sci. 84:833-838

Chapter 3

Estimation of heritability and BVs for early egg production in laying hens from pooled data

F. Biscarini[1], H. Bovenhuis, E. D. Ellen, S. Addo and J. A. M. van Arendonk

Animal Breeding and Genomics Centre
Wageningen University
P.O. Box 338, 6700 AH Wageningen

The Netherlands

Poultry Science (2010) doi: 10.3382/ps.2010-00730

ABSTRACT

Under commercial conditions, data on egg production in laying hens are usually collected per cage rather than individually. In current breeding programs, genetic evaluations are, however, based on individually recorded egg production. Since commercial flocks are not maintained in single cages, this environmental difference between the breeding and commercial setting may result in genotype by environment (G×E) interaction. This study was aimed at estimating genetic parameters and predicting estimated breeding values (EBVs) for early egg production of laying hens by using pooled data (i.e. data from multiple bird cages) from pedigree birds housed in 4-bird cages. Using cage records we compared two different methods of handling pooled data: cage sums and the assignment of cage means to individual animals, referred to as the approximate method. The two methods were compared by using cross-validation. Data from three purebred White Leghorn layer lines were used. Estimated heritability for early egg production was 0.36 when cage sums were used and 0.30 with the approximate method. The correlation of EBVs between the cage sums method and the approximate method was 0.88. Cross-validation showed that the use of cage sums lead to better predictions of missing phenotypes compared to the approximate method. The results of the research demonstrate that pooled data can be used in the genetic evaluation of laying hens and show that using directly pooled records (e.g. cage sums) gives better results than assigning group means to the animals of the group thus simulating individual records.

Key words: heritability, groups, laying hens, pooled data, early egg production.

INTRODUCTION

In commercial egg production facilities layers are kept in groups. In the current battery systems, hens are housed in cages of usually four hens each. However, in the near future, communitarian legislative developments (Council Directive 1999/74/EC) will lead to housing layers in larger groups (furnished cages, non-cage systems). In these situations data on egg production are commonly collected for the whole group rather than individually. However, since standard methods for genetic evaluation are based on individual records, selection of layers is carried out in nucleus flocks where purebred layer lines are housed individually and selected on the basis of individually recorded traits. This constitutes a considerable cost for the breeding industry and might lead to genotype x environment interaction between the testing and commercial environments (Merks, 1989; Besbes and Ducroq, 2003). Using directly the information collected on groups of hens (pooled information) would therefore represent a financial benefit for the breeding companies and, depending on the magnitude of the G x E interaction, might lead to higher genetic response at the commercial level. Theoretical work on the extension of genetic evaluation methodology by incorporating pooled information has been done with simulations by Olson et al. (2006) and with actual data by Biscarini et al. (2008). Results from their research showed that genetic evaluation based on pooled data instead of individual observations is theoretically and practically feasible. Restrictions do however still exist: pooled data from groups of different size, for instance, can not yet be used, mainly due to software limitations (Biscarini et al. 2008). Estimates of BVs based on pooled data are, however, less accurate than those based on individual observations (Olson et al. 2006; Biscarini et al. 2008). The application of pooled data in the genetic evaluation of traits of agricultural interest has been very limited, so far.

Two main strategies to deal with pooled data in animal models have been proposed by Olson et al. (2006) and Biscarini et al. (2008). One is to directly use the pooled record, either the sum or average of the performance of the individuals in a group, thus taking the group structure into account in the model of analysis. The other consists in assigning the average performance of a group to each member of that group, then treating the observations as if they were individual observations. This second method, known as the approximate method, is less adherent to reality since it does not account for the group structure and systematically overestimates the accuracy of the estimates (Olson et al. 2006), but is operationally easier to implement given its similarity with the standard procedure of genetic evaluation. Besides, since current software packages for genetic evaluation have been designed for individual records, it is easier to use the approximate method, especially if groups are of different size.

The objective of the present study was to estimate heritability and predict breeding values for early egg production based on 4-bird cage records, and to compare the results of the two different methods (cage sums and approximate method) of analyzing pooled data. Data from three purebred White Leghorn layer lines were available for this study.

MATERIAL AND METHODS

Data

Data were provided by the Institut de Sélection Animale B.V. (Boxmeer, The Netherlands). For this study the same animals were used as described in Ellen et al. (2008). The animal population consisted of 15,212 laying hens from three purebred lines of White Leghorn origin (lines W1, WB, and WF). Chicks were hatched in two batches. All three lines were represented in each batch. Each batch comprised chicks of four age classes, differing by two weeks each. Newly hatched chicks were sexed and identified by means of wing bands applied to the right wing. They were vaccinated against Marek's disease and infectious bronchitis. The beaks of the chicks were kept intact. During the rearing period chicks were kept in groups of the same line and age. At 17 weeks of age, hens were transported to two different laying facilities. Both facilities had eight rows of cages separated by corridors, each consisting of three tiers (top, middle and bottom). Hens were kept in a total of 3803 cages with 4 hens of the same line and age each. Hens were assigned to cages at random. Due to chance, some hens in a cage were related to various degrees. The number of hens and cages per line is presented in Table 1. Hens were provided with water and a standard commercial layer diet *ad libitum*. The photoperiod in the laying houses was increased starting from 9 hours of light per day. Each week, the photoperiod was increased by 1 hour until 16 hours of light per day were reached, when the hens were about 26 weeks old.

Table 1: per line distribution of the hens and cages used in this study, mean and standard deviation of early egg production for the single lines and the whole population. Data are eggs produced per cage (cage sums) and cage means assigned to individual hens (approximate method)

			Early egg production (17 – 24 wk)			
			Cage sums		Approximate method	
Line	N. hens	N. cages	Mean	SD	Mean	SD
W1	5868	1467	54.3	32.7	13.6	8.2
WB	6016	1504	81.5	27.3	20.4	6.8
WF	3328	832	48.8	29.6	12.2	7.4
Total	*15212*	*3803*	*63.9*	*33.3*	*16.0*	*8.3*

The data consisted of the total eggs produced for each 4-bird cage from 17 to 24 wk of age (pooled early egg production). The number of eggs produced each day was recorded daily for each cage and were then summed over the 7 wk period (cage sums). No individual observations were available. The mean egg production of a cage was used in the approximate method, assigning it to the individual hens housed in the cage. Only cages where no mortality occurred between wk 17 and 24 were used in the analysis.

The experimental population originated from a total of 505 sires and 2331 dams. The same sires were used in the production of the hens that were housed in both laying facilities, but different sets of dams were used to breed the hens in the two facilities. On average each sire was mated to five dams resulting in 1 to 28 offspring per mating (on average 12 offspring per dam). Mating was random. Four generations of animals were extracted from the pedigree, for the calculation of the additive relationship matrix.

Data analysis

A preliminary analysis of variance was performed to determine the systematic effects having a significant influence on early egg production. All hens from the three White Leghorn lines were analysed together. The effects of laying facility, hatch week, genetic line and position of the cage in the laying facility (combination of row and tier) were included in the model. Hatch week was nested within laying facility. The following linear model was used for the estimation of variance components and breeding values:

$$y_{ijklmn} = \mu + L_i + P_j + S_k + H_{l(k)} + a_m + e_{ijklm} \qquad (1)$$

where y_{ijklm} is the early egg production (either cage sum or cage mean, for the pooled and approximate data analysis, respectively) of animal *m*, of line *i*, in cage at position *j*, hatched at week *l* in laying facility *k*; μ is the common mean, L_i is the i^{th} genetic line (W1, WB or WF), P_j is the j^{th} position of a cage in the facility (corridor x tier) in classes, S_k is the k^{th} laying facility, $H_{l(k)}$ is the l^{th} hatch week nested within the k^{th} laying facility, a_m is the random genetic effect of the m_{th} animal, and e_{ijklm} is the residual.

When using cage sums, the mixed model equations (MME) need to be modified according to Biscarini et al. (2008). The vector of observations **y** can be looked at as a vector of sums of individual egg productions, the incidence matrices **X** and **Z** reflect the group composition, and **e** is a vector of sums of the residuals. The vectors of solutions for the fixed and genetic effects, **b** and **a**,

are unmodified. For a couple of records, and considering only the line effect for the sake of simplicity, the MME in the case of cage sums are illustrated below. In this case we see two cages with 4 hens each, the first of line WB and the second of line WF.

$$\begin{bmatrix} (\mathbf{y}_1+\mathbf{y}_2+\mathbf{y}_3+\mathbf{y}_4)=\mathbf{90} \\ (\mathbf{y}_5+\mathbf{y}_6+\mathbf{y}_7+\mathbf{y}_8)=\mathbf{62} \end{bmatrix} = \begin{bmatrix} \mathbf{4} & \mathbf{0} & \mathbf{4} & \mathbf{0} \\ \mathbf{4} & \mathbf{0} & \mathbf{0} & \mathbf{4} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{W1} \\ \mathbf{WB} \\ \mathbf{WF} \end{bmatrix} +$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{...} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{...} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{a}_{s1} \\ \vdots \\ \mathbf{a}_{sn} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \\ \mathbf{a}_5 \\ \mathbf{a}_6 \\ \mathbf{a}_7 \\ \mathbf{a}_8 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3+\mathbf{e}_4 \\ \mathbf{e}_5+\mathbf{e}_6+\mathbf{e}_7+\mathbf{e}_8 \end{bmatrix}$$

The genetic variance is not affected by the modifications and is equal to $\mathbf{A}\sigma_a^2$, with **A** being the additive relationship matrix. Contrariwise, since the vector **e** is of different nature, the residual variance is equal to $\mathbf{D}\sigma_e^2$, where **D** is a diagonal matrix with group size on the diagonal and σ_e^2 is the individual residual variance. Because groups are of equal size (all cages have 4 hens), $\mathbf{D} = n \times \mathbf{I}$ and $\mathbf{R} = \mathbf{I}n\sigma_e^2$. As a result, the ratio between the residual and additive genetic variance (α) is equal to $n \times \alpha$ ($4 \cdot \alpha$ in this case), and the term $\mathbf{R^{-1}}$ can be cancelled from both sides of the MME. This implies that the residual variance estimated from cage sums is $\sigma_e^{2*} = n \times \sigma_e^2$ and that heritability should be calculated from variance components as $\frac{\sigma_a^2}{\sigma_a^2 + (\sigma_e^{2*}/n)}$, where n is group size.

With the approximate method (Olson et al., 2006), the analysis is carried out exactly as in the case of the classical MME for individual observations, with the difference that the vector **y** of observations contains cage means attributed to each individual of the cage. Using the same example as above, the vector **y** would thus look as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{90/4} \\ \mathbf{90/4} \\ \mathbf{90/4} \\ \mathbf{90/4} \\ \mathbf{62/4} \\ \mathbf{62/4} \\ \mathbf{62/4} \\ \mathbf{62/4} \end{bmatrix} = \begin{bmatrix} \mathbf{22.5} \\ \mathbf{22.5} \\ \mathbf{22.5} \\ \mathbf{22.5} \\ \mathbf{15.5} \\ \mathbf{15.5} \\ \mathbf{15.5} \\ \mathbf{15.5} \end{bmatrix}$$

The variances of the genetic and residual effects, and their ratio α, are unmodified and correspond to those of the standard MME (Henderson, 1975, 1984). Residual covariances between hens in the same cage are not taken into account.

The General Linear Model (GLM) procedure of the SAS© statistical software (SAS, 1996) was used for the determination of the significant fixed effects to include in the model. Variance components and breeding values for early egg production from both methods were estimated using Asreml (Gilmour, 2002) with a maximum likelihood approach. In the case of cage sums, the model function $and(\ )$ was used to fit multiple genetic effects per observation. Data editing and all other statistical operations were performed using the open source statistical package R.

Cross validation

The two methods of analyzing pooled data, the use of cage sums and the approximate method, were compared by means of cross-validation (Ellen et al, submitted). In the cross-validation, known phenotypes are set to missing, then their values are predicted with the model and compared to the observed phenotypes. The correlation between the predicted and observed phenotypes is a measure of the quality of the estimation method. From the total dataset, 20% of the cages (records) at a time were randomly set to missing until each cage was once removed from the dataset. Fixed effects classes were taken into account in order to create balanced subsets. This resulted in 5 subsets each containing 80% of the data and for which the missing phenotypes were predicted using either the exact or the approximate method. Heritability was re-estimated for each subset, and phenotypes were predicted summing either the estimates for both the fixed effects and the breeding values ($\hat{\mathbf{y}}_1 = \mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{a}}$) or only the estimates for the fixed effects ($\hat{\mathbf{y}}_2 = \mathbf{X}\hat{\mathbf{b}}$). Pearson's correlations were calculated between the observed phenotypes and those predicted with the two statistical methods.

RESULTS

Description of phenotypes

Basic statistics of the data are summarized in Table 1. The number of available cage records varied from 832 for line WF to 1504 for line WB. Line WB had the highest early egg production, with an average of 81.5 eggs produced per cage between 17 and 24 wk: this is about twice as much as the production of line WF, which produced an average of 48.8 eggs per cage. The overall coefficient of variation was 52%, with a minimum of 34% in line WB and a maximum of 61% in line WF. Being cage averages, values for the approximate method were exactly the same. Early egg production in the analysed dataset was normally distributed.

Variance components

Table 2 reports variance components and heritabilities for early egg production estimated from cage sums and with the approximate method. The standard error of the estimates is also reported. Estimated heritability of early egg production of laying hens is moderate, being 0.36 when using cage sums and 0.30 with the approximate method. Variance components estimated with the two methods differed: the ratio between additive genetic variance estimates based on cage sums and on the approximate method was 6.7, that for estimated residual variances was 21.4. The approximated standard error of the estimated heritability was 0.04 with cage sums and 0.02 with the approximate method.

Table 2: Variance components, h^2 estimates and standard errors for early egg production based on pooled observations

Method	σ^2_a	σ^2_e	h^2	SE
Cage Sums [a)]	36.4	263.7	0.36	0.04
Approximate Method	5.4	12.3	0.30	0.02

[a)] Heritability estimates for cage sums were calculated as $\frac{\sigma^2_A}{\sigma^2_A + \sigma^2_e / 4}$ where 4 is the number of hens per cage

Breeding values

Table 3 shows the correlation between the breeding values estimated with the two methods. The Pearson's correlation between EBVs from cage sums and from the approximate method was 0.88 for all hens with observation, 0.89 for sires with at least 10 offspring, and 0.91 for dams with at least 8 offspring. Spearman correlations were slightly lower in all three cases (see Table 3). The

ranking of animals for their genetic merit was consequently affected: when using cage sums or the approximate method 9 of the top 20 animals were the same for all hens with observations, 14 out of 20 for sires with ≥ 10 offspring, and 15 out of 20 for dams with ≥ 8 offspring.

The accuracy of EBVs was also calculated for both methods. The difference in accuracy was more marked for hens with observations (0.60 vs 0.70) and for dams (0.61 vs 0.75) than for sires with at least 10 offspring (0.78 vs 0.84).

Table 3: Pearson and rank correlations between EBVs for early egg production calculated from cage sums and with the approximate method for all hens with observations, sires with more than 10 offspring and dams with more than 8 offspring

	Correlations between EBVs		
	All Hens with observation	**Sires ≥ 10 offspring**	**Dams ≥ 8 offspring**
Pearson	0.88	0.89	0.91
Spearman	0.86	0.88	0.90

[a)]correlations were based on 15212, 166 and 996 records in the cases of all hens with observations, sires with at least 10 offspring and dams with at least 8 offspring, respectively.

Cross-validation

The Pearson's correlations between the observed phenotypes and their model predictions, for both the cage sums and the approximate method are given in Table 4. When using only fixed effects to predict phenotypes, the correlations were practically the same with cage sums or the approximate method (average correlation over all subsets of 0.86 for both methods). When phenotypes were predicted adding EBVs to the estimates of fixed effects, correlations between observations and predictions were consistently higher in all subsets when using cage sums in comparison to the approximate method. For the cage sums model, they ranged from 0.94 to 0.95 with an average of 0.94, whereas for the approximate method they had an average of 0.91, ranging from 0.91 to 0.92.

Table 4: Pearson's correlations between predicted ($\hat{y}$) and observed phenotypes, using either cages sums or the approximate method. 20% of the observations were set to missing in subsets S1 to S5 and then reconstructed from the effects estimated from the model. $\hat{y}_1$ and $\hat{y}_2$ refer to phenotypes predicted from estimates of both systematic and genetic effects or of systematic effects only.

method	$\hat{y}$	**S1**	**S2**	**S3**	**S4**	**S5**
cage sums	$\hat{y}_1$	0.94	0.95	0.94	0.94	0.95
	$\hat{y}_2$	0.85	0.87	0.86	0.86	0.86

approximate	$\hat{y}_1$	0.91	0.92	0.91	0.92	0.91
	$\hat{y}_2$	0.86	0.86	0.86	0.87	0.85

DISCUSSION

In this study we estimated variance components and predicted breeding values using 4-bird cage data instead of data from individual hens housed in single-bird cages. This constitutes one of the few practical applications of the theoretical work of Olson et al. (2006) and of Biscarini et al. (2008) on the use of pooled data in the genetic evaluation of farm animals. We compared two ways of using pooled data. One is the direct use of the pooled information which implies that the group is the experimental unit and its total performance is the recorded observation. The other consists in apportioning the group mean to each member of the group, thus mimicking the situation in which individual observations are available. This approach is known as the approximate method (Olson et al., 2006). Cage sums are the available data and their use allows one to take into account the group structure of the data in the MME. The approximation of assigning cage means to the individual animals of each cage makes it possible to use pooled data in the usual framework for genetic evaluation of individual observations, and might help work around software limitations. However, the correlation between EBVs calculated with the two methods is well below 1, also for top sires and dams; this leads to re-ranking of animals and to selection of different animals for reproduction. Hence the need to assess the relative quality of the two methods. We showed that there are reasons to consider the direct use of pooled records more appropriate.

Cage sums vs the approximate method

The two methods of analyzing pooled data, the use of cage sums and the approximate method, were used to estimate variance components and to predict breeding values, and the results were compared. Cross-validation was used to assess the validity of the proposed methods.

Heritabilities for early egg production estimated with the two methods were different, 0.36 with cage sums and 0.30 with the approximate method. These are not two independent estimates, since they are both estimated from the same data, but with two different methods. The standard error of the estimate was lower for the approximate method. Accuracy of breeding values was higher with the approximate method (0.70 vs 0.60 for hens with observations, 0.75 vs 0.61 for dams, and 0.84 vs 0.78 for sires). However, accuracies are obtained from the diagonal elements of the inverse of the MME (Henderson, 1975), and in the approximate method these are not modified to account for the group structure. Therefore, with the approximate method accuracies are the same as if every animal

had an individual observation, and are consequently systematically overestimated, as shown already by Olson et al. (2006). The discrepancy between true and estimated accuracy under the approximate method increases with group size (Olson et al., 2006; Biscarini et al., 2008). Therefore we used cross-validation to compare the two methods in predicting phenotypes from estimated effects. When phenotypes were predicted only from fixed effects ($\hat{\mathbf{y}}_2$), there were no differences between cage sums and the approximate method. When phenotypes were predicted from the sum of the estimates of fixed effects and EBVs ($\hat{\mathbf{y}}_1$), cage sums performed consistently better than the approximate method. This indicates that the difference between the two methods lies in the EBVs, which are better estimated by using cage sums. The results show that genetic parameters and breeding values estimated with the approximate method are less reliable and the direct use of group records can therefore be regarded as to be a more correct way of analyzing pooled data.

The differences observed between the two methods reside, besides the different incidence matrices **X** and **Z** of the MME, in the residual variances. For the 4 hens in a cage, the residual variance is $var(e_1+e_2+e_3+e_4)$ when using pooled data. In the approximate method the same residual variance is a 4x4 diagonal matrix of the 4 residual variances ($var(e_1)$, $var(e_2)$ … $var(e_4)$), where no residual covariances are modeled for the observations from the same cage. Including such residual covariances in the model could make the results from the approximate method more similar to those obtained with pooled records.

Genetic parameters

We found a moderate level of heritability for early egg production: estimates ranged from 0.30 with the approximate method to 0.36 when cage sums were used. These values do not differ substantially from heritability estimates for early egg production based on individual production records. Though variable, literature estimates are in fact all of moderate to high magnitude and fall generally in the range between 0.30 and 0.60 (Besbes et al., 1992; Wei and van der Werf 1993, 1995; Anang et al., 2000; Mizstal and Besbes, 2000; Nurgiartiningsih et al., 2004). When animals are kept in groups, as layers in cages, social interactions like competition for resources, are bound to emerge (Bijma et al., 2007b). Since we analysed pooled data and not individual records, we could not include the genetic associative effect in the model (Bijma et al., 2007a,b).

In this study heritability for early egg production was estimated across lines: data of three different lines of laying hens were combined in a single analysis and the effect of line was included in the statistical model to account for difference in genetic merit between lines. However, heritabilities were estimated also for the individual lines, using both cage sums and the approximate method. With cage sums the heritability of early egg production was 0.23, 0.38 and 0.30 in lines W1, WB

and WF. With the approximate method estimates were 0.10. 0.12 and 0.09 in lines W1, WB and WF, respectively. These estimates do not differ substantially from those obtained with the combined analysis in the case of cage sums, but do differ with the approximate method: this provides further evidence of the unreliability of the approximate method. It is also interesting to notice that the approximate method seems to systematically underestimate heritability for early egg production.

When group performances are analysed the residual variance matrix **R** is no longer $\mathbf{I}\sigma_e^2$ but $\mathbf{D}\sigma_e^2$, with **D** being a diagonal matrix reflecting the number of individuals in each group. If groups are all of equal size, then D is equal to $n \times \mathbf{I}$, where n represents the number of individuals in each group. This implies that the residual variance estimated from group sums is expected to be n-times that estimated from individual observations. In the present study therefore, with cages of 4 hens each, the residual variance should be divided by four in order to be compared with the results of previous works on heritability of early egg production. This gives a residual variance of 37.7 for individual records, which agrees with previous estimates which, though variable, are in the range of 23 to 70 (Besbes et al. 1992; Wei and van der Werf, 1993). The expectation of the genetic variance, on the other hand, remains unmodified also in the case of pooled data; our estimate of 36.4 is very close to the results of Wei and van der Werf (1993), Nurgiartiningsih et al. (2004), and mostly in line with those of the other works mentioned above. Biscarini et al. (2008) estimated heritability for body weight from individual and pooled data and found good agreement between the estimated variance components.

The variability of the estimates of variance components and the small differences between our results and those of other studies can be attributed to differences in the definition of the trait and to transformation of the data to handle lack of normality. For instance, Wei and van der Werf (1993) defined early production as the egg number laid between 18 and 25 weeks of age, while in this study early egg production was defined as the egg numbers from week 17 to the end of week 24.

Egg production traits often do not follow a normal distribution and various mathematical techniques, such as the Box-Cox transformation (Box and Cox, 1964), are used to adjust for this. Our data were however normally distributed and we did not apply any transformation to them. Also the use of different genetic lines (crossbred or purebred hens) and the difference in data structure (individual or pooled observations) can partly account for the different estimates of variance components. Estimates of additive genetic and residual variance with the approximate method deviated a bit more from results in the literature.

Pooled data have been used previously to estimate variance components. Simianer and Gjerde (1991) used the mean weight of samples of full-sibs to estimate variance components for BW in

salmon with a MIVQUE (Minimum Variance Quadratic Unbiased Estimation) methodology. Biscarini et al. (2008) adopted an approach based on modified MME using an animal model and analysed cage sums to estimate heritability for BW in laying hens. Earlier studies on the analysis of pooled data for egg production traits in laying hens were conducted by Wei and van der Werf (1995) and by Nurgiartiningsih et al. (2004), both of whom used a sire model to estimate heritabilities from cage means of daughters. As for early egg production, Wei and van der Werf (1995) found heritabilities for eggs produced between 18 and 25 wk of age of 0.51 and 0.4 from two different crosses of hens; Nurgiartiningsih et al. (2004) found heritabilities for eggs produced in the first month of lay of 0.32 and 0.38 in two lines of White Leghorn hens.

Accuracies of Selection

In this study we estimated breeding values for early egg production for 15212 hens. However, these EBVs were not based on as many individual observations but, given that only pooled data were available, they were based on 3803 cage records. This reduced number of available records, together with the fact that pooled information, being either a cage mean or a cage sum, is a poorer source of information, leads inevitably to a lower accuracy of estimates, as has been shown already by Olson et al. (2006) and by Biscarini et al. (2008). The same authors also showed that loss of accuracy increases with group size. However, increasing the number of available records, though pooled, will lead to higher accuracy of EBVs. In the present work accuracies of 0.60, 0.78 and 0.61 were obtained for hens with records, sires and dams respectively. These results can be compared with those obtained by Biscarini et al. (2008) for total egg production from 371 cage records. They calculated an accuracy of 0.42, 0.50 and 0.45 for hens with records, sires and dams respectively. It can be seen that increasing the number of records improves the accuracy of the estimates. More precisely, with ten times as many records (371 and 3803, in the two studies) the accuracy increased 45% on average, and by 55% for sires.

Besides, as mentioned by Biscarini et al. (2008), the reduction in accuracy due to the use of pooled data should also be interpreted in the context of direct versus indirect selection. In the case of laying hens, group housing at the commercial level and individual housing in the test station are two different environments, therefore the genetic correlation between EBVs estimated in the two situations is relevant for response to selection. For the same selection intensity, the ratio of the selection response for direct and indirect selection is a function of the accuracies for both situations and the genetic correlation between both traits (Falconer, 1989).

The approximate method does not take into account that the observations used in the coefficient matrix of the MME are records from a group of animals, and therefore the accuracy of prediction is systematically overestimated with this method (Olson et al., 2006).

Use of pooled data in selection

This paper illustrates a practical application of the use of pooled data in the genetic evaluation of laying hens. The results showed that the direct use of pooled data (e.g. cage sums) is preferable to simulating the availability of individual records by assigning the group mean to each of the animals in the group (approximate method). Modifying the MME to analyse pooled data gives in fact more accurate estimates of genetic parameters and genetic values. In a comparison of results from individual and pooled data, Biscarini et al. (2008) showed the potential for the use of pooled data in genetic evaluations of laying hens.

The use of pooled data entails a certain loss of accuracy that, however, can be compensated for by using a higher number of records. Also group composition plays a role: pooled observations from groups of closely related animals (e.g. full-sibs) are more difficult to attribute to the individual genotypes contributing to the observation, and make the statistical analysis less accurate (Biscarini et al., 2008). This loss of accuracy, however, might not be so negative in light of the potential response to selection under the test versus commercial conditions. Moreover the use of pooled data offers some advantages: they are often easier and cheaper to collect, they might in some cases be the only data available, and they may sometimes better reflect the commercial environment under which the offspring will be kept, thus avoiding the potential consequences of GxE interactions. Therefore, selecting individuals based on group performance may be very convenient in many situations. It will make genetic evaluation of farm animals more flexible and possibly more accurate by including information collected under commercial conditions where animals are kept in groups, and offers interesting possibilities for breeding companies.

ACKNOWLEDGMENTS

This work was conducted as part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission. Data were collected for the project “The genetics of robustness in laying hens”, which was a collaboration between Wageningen University and Hendrix Genetics, and financially supported by Senter Novem.

REFERENCES

Anang, A., N. Mielenz and L. Schüler. 2000. Genetic and phenotypic parameters for monthly egg production in White Leghorn hens. J. Anim. Breed. Genet. 117:407-415.

Besbes, B., V. Ducrocq, J. L. Foulley, M. Protais, A. Tavernier, M. Tixier-Boichard and C. Beaumont. 1992. Estimation of genetic parameters of egg production traits of laying hens by restricted maximum likelihood applied to a multiple-trait reduced animal model. Genet. Sel. Evol. 24:539-552.

Besbes, B. and V. Ducroq 2003. Pages 139-140 in Poultry Genetics, Breeding and Biotechnology. Muir and Aggrey editors. CAB International, Wallingford, Oxon OX10 8DE, UK.

Biscarini, F., H. Bovenhuis, and J. A. M. Van Arendonk. 2008. Estimation of variance components and prediction of breeding values using pooled data. J. Anim. Sci. 86:2845-2852.

Bijma, P., W. M. Muir and J. A. M. Van Arendonk. 2007a. Multilevel Selection 1: Quantitative Genetics of Inheritance and Response to Selection. Genetics 175:277–288.

Bijma, P., W. M. Muir, E. D. Ellen, J. B. Wolf and J. A. M. Van Arendonk. 2007b. Multilevel Selection 2: Estimating the Genetic Parameters Determining Inheritance and Response to Selection. Genetics 175:289–299.

Box, G. E. P. and D. R. Cox. 1964. An analysis of transformations. J. R. Statist. Soc. B26:211-243.

Ellen, E. D., J. Visscher, J. A. M. van Arendonk and P. Bijma. 2008. Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines. Poult. Sci. 87:233-239.

Ellen, E. D., V. Ducroq, B. Ducro and P. Bijma. Submitted. Combining survival analysis and a linear animal model to estimate genetic parameters for direct and associative effects on survival time in three layer lines.

Falconer, D. S. 1989. Pages 318-319 in Introduction to quantitative genetics. 3rd ed. Longman Scientific & Technical, New York, NY.

Gilmour, A. R., B. J. Gogel, B.R. Cullis, S. J. Welham, R. Thompson. 2002. ASReml User Guide Release 1.0 VSN International Ltd. Hemel Hempstead, HP1 1ES, UK.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447.

Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. Guelph Press.

Merks, J. W. M. 1989. Genotype x Environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. Livest. Prod. Sci. 22:325-339

Misztal, I. and B. Besbes. 2000. Estimates of parental-dominance and full-sib permanent environment variances in laying hens. Anim. Sci. 71:421-426

Nurgiartiningsih, V., N. Mielenz, R. Preisinger, M. Schmutz and L. Schüler. 2004. Estimation of genetic parameters based on individual and group mean records in laying hens. Brit. Poultry Sci. 45:604-610.

Olson, K. M., D. J. Garrick and R. M. Enns. 2006. Predicting breeding values and accuracies from group in comparison to individual observations. J Anim. Sci. 84:88-92.

SAS. 1996. SAS User's Manual Release 6.12. SAS Institute Inc., Cary, NC.

Simianer, H. and B. Gjerde. 1991. Estimating variance components from fullsib groups means. J. Anim. Breed. Genet. 108:270-279.

Wei, M. and J. H. J. van der Werf. 1993. Animal model estimation of additive and dominance variances in egg production traits of poultry. J. Anim. Sci. 71:57-65.

Wei, M. and J. H. J. van der Werf 1995. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. J. Anim. Sci. 73:2220-2226.

Chapter 4

Across-line SNP association study of innate and adaptive immune response in laying hens

F. Biscarini*, H. Bovenhuis*, J. A. M. van Arendonk*, H. K. Parmentier[#], A. P. Jungerius[§] and J. J. van der Poel*

** Animal Breeding and Genomics Centre, Wageningen University P.O. Box 338, 6700 AH Wageningen, The Netherlands*
 # Adaptation Physiology Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands
§ Hendrix Genetics, Research & Technology Centre, Spoorstraat 69, 5830 AC Boxmeer, The Netherlands

Animal Genetics (2010) 41:26-38

ABSTRACT

The aim of the present study was to detect quantitative trait loci (QTLs) for innate and adaptive immunity in laying hens. For this purpose the associations between 1022 single nucleotide polymorphism (SNP) markers and immune traits were studied in 583 hens from nine different layer lines. Immune traits were NAbs for keyhole limpet hemocyanin (KLH) and lipopolysaccharide (LPS) at 20, 40 and 65 wk, acquired antibodies to the vaccinal virus of New Castle disease at 20 wk and complement activity measured on sheep and bovine red blood cells at 20, 40 and 65 wk. We adopted a novel approach based on across-line analysis and testing of the SNP-by-line interaction. Among lines linkage disequilibrium is conserved at shorter distances than in individual lines; therefore, SNPs significantly associated with immune traits across lines are expected to be near the functional mutations. In the analysis, the SNPs that had a significant across-line effect but did not show significant SNP-by-line interaction were identified, to test that the association was consistent in the individual lines. Ultimately, 59 significant associations between SNPs and immune traits were detected. Our results confirmed some previously identified QTLs and identified new QTLs potentially involved in the immune function. We found evidence for a role of IL17A (chromosome 3) in natural and acquired antibody titres and in the classical and alternative pathways of complement activation. The major histocompatibility genes on chromosome 16 showed significant association with natural and acquired antibody titres and classical complement activity. The IL12B gene on chromosome 13 was associated with natural antibody titres.

Keywords SNPs, immunological traits, laying hens, across-line association study.

Introduction

Immunity in birds in general and hens in particular is not as well studied or understood as in mammals. A good understanding of the immune response is required to improve the health of laying hens, an important issue in poultry breeding. Immunological characteristics such as antibody titres have been shown to be heritable in poultry (e.g., Pinard *et al.* 1992; Lamont *et al.* 2003), indicating the possibility of selecting for immune response and disease resistance. Immunity is composed of both innate and acquired components and can be measured in a number of ways, such as antibody titres and the activity of the complement system in the blood. Natural antibodies (NAbs) and the complement system define innate humoral immunity and constitute a substantial part of the basic capacity of an organism to respond to danger (Matzinger 1994). Acquired antibodies neutralize pathogens with which the organism has already had contact and form part of the adaptive humoral immunity. Both the innate and adaptive immune systems have a cellular component composed of lymphocytes, macrophages, antigen-presenting cells and other types of cells. Natural antibodies and complement titres have been associated with survival in laying hens (Star *et al.* 2007). Therefore, determining the genetic bases of these immunological parameters is of considerable interest as this information could be used to select for animals with superior immune response. Previous studies identified microsatellite markers located in chromosomal regions affecting immune traits in chickens (Yonash *et al.* 1999, 2001; Kaiser *et al.* 2002; Yunis *et al.* 2002; Zhou *et al.* 2003; Siwek *et al.* 2003a, 2003b, 2006). QTL mapping was initially based on linkage analysis. More recently, the increasing availability of SNPs has allowed researchers to exploit linkage disequilibrium (LD) in association studies and perform fine mapping of previously detected QTLs.

In this study, we used a new approach based on the analysis of multiple lines and test of SNP-by-line interaction in a population of laying hens measured for several immune traits. If the SNP showed a significant association with the immune traits in the across-line analysis and no SNP-by-line interaction, it was considered to be consistently associated with the phenotype in all lines, irrespective of their pedigree. Because of the reduced extent of LD conserved across lines, these SNPs are probably closer to QTLs for immune responsiveness than are those from previous analyses of single lines or crosses. Thirteen immune traits involved in natural and acquired antibody titres and complement activity were measured in nine lines of laying hens. These traits were associated with a set of SNPs distributed over several chromosomes of the chicken genome. A single SNP across-line association study has been performed for this purpose: methods of analysis and results are hereby presented.

Materials and methods

Experimental Population

The animal population used in this study consisted of 583 laying hens chosen at random from nine genetic lines, four of Rhode Island Red type (brown layers) and five of White Leghorn type (white layers). The hens' blood was sampled for genotyping and measuring immune traits (Table 1).

All hens were housed in battery cages in the same stable; cages contained four hens from the same line, either full-sibs or randomly mixed. Hens arrived at the laying facility at 17 wk of age and remained in the stable for the entire laying period of 52 wk. The hens had intact beaks and received routine vaccinations against Marek's disease (d 1), New Castle disease (NCD; wk 2, 6, 12, 15), infectious bronchitis (d1, wk 2, 10, 12, 15), infectious bursal disease (wk 3, 15), fowl pox (wk 15), and avian encephalomyelitis (wk 15).

Feed and water were available *ad libitum*. From the beginning of the experiment (at 19 wks of age) until 42 wks of age, hens were fed a standard starter diet (159 g/kg crude protein, 43 g/kg crude fibre, and 11.17 MJ ME/kg). From 42 wks of age until the end of the experiment, a standard grower diet (152 g/kg crude protein, 47 g/kg crude fibre, and 11.01 MJ ME/kg) was provided. Wing bands allowed identification of individuals. The hens were kept on a standard light schedule.

Every rooster was mated on average with 2.3 females while each hen was mated with only one male. Four generations of ancestors were extracted from the pedigree file.

Phenotypes

The following immune traits were measured in blood samples: NAbs against the exo-antigen keyhole limpet hemocyanin and the environmental antigen lipopolysaccharide at 20, 40, and 65 wk (KLH20, KLH40, KLH65 and LPS20, LPS40 and LPS65, respectively); acquired antibodies to the vaccinal virus of NCD at 20 wk (NCD20); classical and alternative pathways for complement activation measured, respectively, on sheep and bovine red blood cells at 20, 40 and 65 wk (SRBC20, SRBC40, SRBC65, and BRBC20, BRBC40, BRBC65). Antibody titres were determined with indirect ELISA by Star *et al.* (2007) and expressed as the $-\log_2$ of corresponding dilutions. Complement activity was determined with a haemolytic assay by Star *et al.* (2007) and reported as the hemolytic dose 50, expressed in units per millilitre of serum (CH50 U/ml). Depending on the trait, the number of available observations ranged from 277 for BRBC20 to 583 for LPS40 (Table 1). All traits followed approximately a normal distribution. Details about the traits analyzed in this study can be found in Star *et al.* (2007).

Genotypes

Genotyping was done in a 1536-plex format using the GoldenGate assay (Illumina, San Diego) by a commercial genotyping facility (ServiceXS, Leiden, NL). As part of a bigger experiment, SNPs were selected to cover QTL regions for immune and behavioural traits identified in previous mapping studies, as well as specific candidate genes. Only association with immune traits were subject of this study. Twenty-four of the 39 chromosomes of the chicken genome were partly covered (Table 2). Only SNPs showing three distinct clusters in the allelic discrimination plot as provided by the Illumina Beadstudio software were considered reliable and in this way 180 SNPs were excluded. Thus, the initial set included 1356 SNPs. A further check based on The Hardy-Weinberg (HW) equilibrium was applied to these 1356 SNPs. Deviations from HW equilibrium were assessed within line using Fisher's exact test (Wigginton *et al.* 2005) at the Bonferroni-corrected 0.05 significance level: the Bonferroni correction was applied separately for each chromosome. Forty-five SNPs presented genotype frequencies that deviated from HW equilibrium in five or more hen lines. Their allelic discrimination plot was carefully re-checked for cluster separation and 14 of these SNPs showed no clear separation between heterozygotes and homozygotes and were therefore excluded from further analysis. The remaining 31 were kept in the panel leading to a total of 1342 SNPs. Fixed SNPs and SNPs with a minor allele frequency ≤ 0.05 were not used in the analysis: these were 320 SNPs. Finally,1022 SNPs were used in the association study.

Data Analysis

A two-step procedure was used for data analysis. First, an association study was performed without accounting for relationships between animals. Potentially interesting SNPs from the first step were then analyzed in more detail using a mixed model and taking additive genetic relationships among animals into account.

The association between genotypes and phenotypes was tested across lines using a single SNP approach. The across-line analysis increases the number of observations per genotype but picks up only markers that are in LD with the QTL across lines. In the first step, single-trait analyses were performed across lines on a chromosome-by-chromosome basis using the following statistical model:

$$y_{ijk} = \mu + SNP_i + line_j + (line \times SNP)_{ij} + e_{ijk} \qquad (1)$$

where y_{ijk} represents the observation of animal k, with SNP genotype i, in line j; μ is the overall mean; SNP_i is the effect of the SNP genotype, either AA, AB or BB; $line_j$ is the line effect (nine classes); $(line \times SNP)_{ij}$ is the interaction between SNP genotype and line; and e_{ijk} are the residuals.

This model (equation [1]) was run twice, once without the line-by-SNP interaction term to obtain p-values for the SNP effect and once with the line-by-SNP term to determine the significance of the interaction. Adapting the approach of Saccone *et al*. (2008), we looked for SNPs that had an across-line significant effect in model [1] ($p \leq 0.05$) and did not show a significant line-by-SNP interaction ($p > 0.05$) , to test that the association was consistent in all lines. We focused on SNPs that met at least one of the following 3 criteria:

1. Significant SNP effect with $p \leq 0.001$ and no line-by-SNP interaction;
2. Significant effect ($p \leq 0.05$) consistent over time (at 20, 40 and 65 weeks of age) for a trait (KLH, LPS, SRBC or BRBC) and no line-by-SNP interaction;
3. Significant effect ($p \leq 0.05$) in four or more traits, irrespective of which trait and time, and no line-by-SNP interaction: these SNPs might point to regions involved in multiple aspects of the immune response.

The SNPs that satisfied at least one of these conditions were selected for the second step of the association study in which the polygenic effect was added to the model to account for family relationships among animals. Ignoring relationships in cases where animals are related would inflate significance levels and lead to biased estimates of gene effects, especially in populations undergoing directional selection (Kennedy *et al.* 1992). The following mixed model was used:

$$y_{ijk} = \mu + SNP_i + line_j + a_k + e_{ijk} \qquad (2)$$

where all the terms are as specified in equation [1] except a_k, which is the random genetic effect of the k_{th} animal. Var(**a**) = **G** = **A** σ_a^2 , with **A** being the additive relationship matrix and Var(**e**) = R = I σ_e^2 . The ratio between residual and genetic variances was fixed using heritabilities previously estimated on the entire available population, including hens whose phenotype was measured but that were not genotyped (~ 900 hens). Heritabilities for single traits had high standard errors; therefore, heritabilities for the traits measured at 20, 40, and 65 weeks were averaged and used in the analysis: 0.21 for KLH20, KLH40 and KLH65; 0.26 for LPS20, LPS40 and LPS65; 0.16 for SRBC20, SRBC40, and SRBC65; 0.22 for BRBC20, BRBC40 and BRBC65; and 0.25 for NCD20.

The SNPs that still showed a significant effect ($p \leq 0.05$) on the traits from model [2] were considered to be in a genomic region of interest.

Data editing, analyses with model [1], and all other statistical analyses were performed using the open source statistical package R. The R library GenABEL (Aulchenko *et al*. 2007) was used to test for HW equilibrium. Variance components for the immunological traits, polygenic effects, and SNP effects as described in model [2] were estimated with a REML procedure using the Asreml software package (Gilmour *et al.* 2002).

Results

Descriptive statistics of the traits

Descriptive statistics of the immunological traits are summarized in Table 1. Antibody titres and complement activity generally increased with age in all lines. The overall across-line coefficient of variation ranged from 31% (LPS40 and LPS65) to 68% (SRBC20), indicating considerable variability in the immunological traits of hens. There were substantial differences in antibody titres and complement activity between lines: e.g., from 28.3 (line WB) to 155.4 (WF) CH50 U/ml for BRBC20, and from 1.4 (line WA) to 3.1 (line WF) $-\log_2$ of antibody titre for KLH20. Phenotypic correlations (results not shown) among the traits were generally weak: 88% of the correlations were between -0.17 and 0.24. The highest correlations were found between NAb titres for the same antigen measured at different ages (0.69 between KLH40 and KLH65, and 0.62 between LPS40 and LPS65).

SNPs

The SNPs used in this study were located on 24 of the 39 chromosomes of the chicken genome. Positions of the SNPs were derived from the NCBI database (Galgal2.1 build 128). The average interval between SNPs varied from 25.29 Kbps (~ 0.06 cM) on chromosome 16 to 5740.74 Kbps (~14.35 cM) on chromosome 2. For practical interpretation, the positions in base pairs were converted to centimorgans on the basis that 4×10^5 bps roughly corresponds to 1 cM in chickens (International chicken genome sequencing consortium, 2004). On average, the proportion of SNPs that deviated significantly from HW equilibrium within lines was 7.1% (Table 2), with the lowest percentages in lines W1 (2.7%) and B3 (5.7%) and the highest percentages in lines WA (13.1%) and BB (9.7%). Table 2 also reports the number of monomorphic (fixed) loci for the various chromosomes in the different lines. There were more fixed loci in the white layers (28.2%) than in the brown layers (22.6%). This is compatible with White Leghorn hens' longer history of artificial

selection, which results in higher homozygosity and lower genetic polymorphism (Hillel *et al.* 2003).

Association study

In the first step of the association analysis, 799 across-line SNP-trait significant associations were detected ($p \leq 0.05$), of which 481 showed no significant genotype-by-line interaction. Among these 481, there were 68 SNPs with associations that met at least one of the three established criteria and were used in the second step of the association analysis. After accounting for relationships between animals, fifty-nine SNPs showed significant associations and are reported in Table 3. Overall, there was good agreement between the significance levels of models 1 and 2 (Pearson's correlation coefficient between p-values = 0.93): only nine of the SNPs found significant in Model 1 were not significant when the analysis was refined to take relationships among animals into account.

The detected associations comprised: six SNPs for KLH20, 10 for KLH40, 11 for KLH65, three for LPS20, 12 for LPS40, nine for LPS65, 12 for NCD20, three for SRBC20, 11 for SRBC40, 13 for SRBC65, three for BRBC20, 14 for BRBC40 and 10 for BRBC65. Twenty-six of these SNPs were associated with only a single immunological trait; the other 33 were associated with more than one trait. Fourteen SNPs were associated with two immunological traits, 15 with three traits, two with four traits, and two SNPs were found to be significantly associated with as many as five different immunological traits.

The strongest associations were found for BRBC40 ($-\log_{10}(p\_value) = 3.79$) and NCD20 ($-\log_{10}(p\_value) = 3.34$) on chromosome 5, and SRBC40 ($-\log_{10}(p\_value) = 3.68$) on chromosome 4.

Based on significant SNP associations, potential QTLs for immunity in laying hens have been found. Examples include QTLs for the production of NAbs (SNPs rs13513104 and rs13513581) and a QTL for NAb and complement activity at SNPs rs13520611, rs13520637, rs13520872 and rs13520980, all on chromosome 4. On chromosome 5, potential QTLs for alternative complement activation (rs13755931 and rs13756481) and NAbs to exo-antigens (SNPs rs13584901) were detected. On chromosome 7, a QTL for NAbs to exo-antigens was located at SNPs rs13596817 and rs13596877; this QTL also showed significant association with complement activation. On chromosome 16, a QTL for titres of NAbs was detected at SNP rs15788216; this QTL was also associated with acquired antibodies. Table 3 contains the complete list of SNPs for immunity detected in this study. In many cases, neighbouring SNPs provided additional evidence supporting

the results for an individual QTL but did not meet the criteria set in this study and their significance decreased gradually as distance increased (results not shown).

The genotypic effects of SNPs with –log(p-value) > 2.5 were estimated as deviations from a reference genotype and are reported in Table 4. The average percentage of the variance explained by the SNPs was 2.4%, with most of the SNPs falling in the range between 1.5% and 6% of the variance. All sorts of possible degrees of dominance were observed: overdominance, complete dominance, partial dominance and absence of dominance (codominance or additivity). The cases of overdominance were observed not only in the across-line analysis but also within lines (results not shown). For instance, the effect on BRBC40 of SNP rs13596817 (chromosome 7) was overdominant in four of the five lines in which it was estimable, and the effect on LPS20 of SNP rs15677371 (chromosome 13) was overdominant in six lines out of nine. The standard errors of the estimates of the genotypic effects (Table 4) ranged from 0.10 to 0.38 σ_p and had an average of 0.17 σ_p.

Discussion

Methodology

In this study we presented an original approach to detect SNPs associated with immune traits. The method is based on the simultaneous analysis of multiple lines using a multiple-step procedure. In the across-line analysis, the SNPs detected were expected to be near the QTLs for immune traits due to the reduced extent of LD conserved across lines. Building on the work by Saccone *et al.* (2008), we tested for the SNP-by-line interaction to ensure consistency of the association across lines. A possible source of false positive associations due to population stratification was avoided by including a line effect in the model. Consequently, SNPs explaining part of the between-line variation could not be detected in this approach. Family relationships within lines could be another source of false positive associations which was dealt with by including a polygenic effect in the model that accounted for the effects of all other genes on the trait. Significance levels of SNP effects from model [2] correlated well with those from model [1] and were generally lower, which agrees with the results of Hassen *et al*. (2009). The analysis proved to be robust to variations in heritability: heritabilities were varied (from -0.1 to +0.2 around the average) with limited impact on the significance of the results This agrees with the results of Hassen *et al*. (2009). This approach led to the detection of 59 SNPs significantly associated with immune traits in laying hens.

The multiple steps, consisting of the analyses with and without accounting for animal relationships, the test for the SNP-by-line interaction and the three selection criteria, served as

progressive refinements of the analysis with the aim of avoiding false associations. In addition we also calculated the false discovery rate (FDR) based on the p-values from model 1 for 1022 SNPs and 13 traits. The FDR for the SNPs reported in table 4 ranged from 0.11 to 0.42.

Detected associations

Some of the results of the present work confirmed findings from previous QTL mapping studies. QTLs for acquired antibodies on chromosome 3 were also detected by Siwek *et al.* (2003, 2006b), Zhou *et al.* (2003), and Yunis *et al.* (2002), on chromosomes 4 and 16 by Siwek *et al.* (2003, 2006b), and on chromosome 5 by Zhou *et al.* (2003), Yonash *et al.* (2001) and Kaiser *et al.* (2002). Siwek *et al.* (2006a) also found QTLs for NAbs to environmental antigens on chromosomes 3, 5 and Z, which were confirmed in this study.

Several SNPs were associated with phenotypic variation in more than one immunological trait, suggesting pleiotropic effects. SNPs associated with NAbs to both KLH and LPS were found on chromosomes 4, 5 and 7, which agrees with the genetic correlation of 0.81 estimated for these traits by Star (PhD thesis, 2008). Siwek *et al.* (2006a) estimated positive genetic correlations between natural and acquired antibody titres and we identified four SNPs associated with antibodies to NCD and either KLH or LPS. This result provides insight into those of Star *et al.* (2007), who observed that the rankings of the lines for NAbs binding to KLH or LPS were similar and did not differ significantly from the rankings for complement activity and specific antibodies against the NCD virus.

Clear patterns of associations couldn't always be observed: closely linked SNPs do not always confirm associations neither did traits measured at different ages always result in a confirmation of SNP associations. Phenotypic correlations between traits were in general low, even for traits measured at different ages. Highest correlations were observed between klh40 and klh65 (0.69) and lps40 and lps65 (0.62). Other traits were almost completely independent with correlations close to 0, e.g. between srbc20 and srbc40 or brbc20 and brbc65. For such traits, the probability of replicating significant associations depends upon the power. In our study, we calculated a power of 80% to detect a gene explaining 8% of the additive genetic variance with 583 records. For the the effects estimated in this study (Table 4) power ranges from 20% to 67%. LD between a functional mutation and a SNP might be variable, even at short distances. Andreescu et al. (2007), showed that more than 30% of the markers within 0.25 cM had an r^2 between 0 (no LD) and 0.2. Therefore, a significant effect of one SNP does not necessarily result in a significant effect of a nearby SNP. This will be a function of the actual LD between the nearby SNP and the functional mutation, and

of the variance explained by the functional mutation. Such variable power and LD may well account for the absence of regular patterns of association.

Sometimes, however, patterns of associations were observed. Star *et al.* (2007) showed that chickens tend to be immunologically consistent over time and that hens that had high titres of NAbs for KLH and LPS (at any age) were also high responders for NCD20; both findings are supported by the results of the present study that described several SNPs that were associated with the same class of trait over time or with both natural and specific antibodies. These results also suggest a genetic basis for functional relationships between innate and adaptive immune competence. Almost all of the SNPs that were associated with classical complement activity (SRBC20, SRBC40, SRBC65), with few exceptions, were also associated with natural or acquired antibody titres, confirming that the formation of antibody-antigen complexes is the main mechanism for activating the complement cascade (Walport 2001).

The lowest number of significant SNPs was found for SRBC20 and BRBC20, the traits with the lowest number of available observations. With fewer observations the power of the statistical analysis is lower and therefore fewer associations between SNPs and immune traits can be detected. The power to detect the effects reported in Table 4 with 277 records ranged from 6% to 36%.

Candidate genes

On chromosome 3, two SNPs (rs13526054 and rs15458146) were situated in the gene for interleukin 17A (*IL17A*) and were associated with the production of natural and acquired antibodies as well as the classical and alternative pathways of complement activation. Three surrounding SNPs, only 0.1 cM distant, were also associated with the mentioned traits, although just below the criteria set in the present study. The IL-17 family of cytokines has been implicated in several immune functions in mice and humans, such as the onset and mediation of phlogosis and the stimulation of the production of other cytokines, chemokines and prostaglandins in many cell types (like fibroblasts, macrophages and endothelial cells) (Iwakura *et al.*, 2008). The IL-17 family has been linked to the induction and stimulation of chemotaxis and to the function of T helper 17 lymphocytes. To our knowledge, no previous work has highlighted a role for the *IL17A* gene on NAbs and complement activity in laying hens. On chromosome 16, which contains the genes for the Major Histocompatibility Complex (*MHC*), we detected two SNPs associated with the synthesis of natural and acquired antibodies with additive effects of around 0.5 σ_p. Subliminal effects on antibody titres were also noticed for two neighbouring SNPs. The SNP rs14050302 located in the *IL12B* (interleukin 12B) gene on chromosome 13 was linked to the production of NAbs at different

ages. Two SNPs within 0.1 cM of this locus (rs15677371 and rs15677377) were also associated with NAb levels. These results underscore the role of *IL12B* in mediating the immune response.

These findings suggest an important but unexpected involvement of the *MHC* region, and *IL-12B* and *IL-17A* cytokines in the regulation of NAbs. To our knowledge this is the first indication either in poultry or mammals that NAbs levels are regulated by these genes. Star *et al.* (2007) reported that NAbs are related to survival in laying hens; the authors claimed that most of the animals used in the experiment died of non-specific causes (neither specific disease nor cannibalism). It was postulated that innate immunity might play an important role in preventing non-specific death in hens. Low titres of NAbs to KLH at various ages were detected in those hens that did not survive through the experiment. Parmentier *et al.* (2004b) reported that low levels of innate immunity, either cellular or humoral, might be related to disease susceptibility, whereas high levels might be related to disease resistance. Natural antibodies might be important for the maintenance of homeostasis and disease resistance (Lutz *et al.* 2008), and an imbalance in NAb levels might lead to higher disease susceptibility (Lutz 2006). Taken together, these results suggest that innate immunity may be important in selection for disease resistance, which is a relevant issue in layer breeding. Indeed, NAbs are present from birth, do not require time to be developed in response to infection, do not rely on the progressively vanishing immunological memory, and are active against a broad spectrum of antigens (Lammers *et al*., 2004; Parmentier *et al*., 2004a).

The genes for the serotonin receptors *HTR2C* and *HTR3A* on chromosomes 4 and 24 were found to be associated with classical complement activity and NAb titres, respectively. They both showed codominant effects on the respective traits. Serotonin is a neurotransmitter that modulates several aspects of behaviour in mammals and chickens (Bolhuis *et al*., 2009), and these results suggest correlations between behaviour and immunity. A SNP in the sequence of the *IL10* (Interleukin 10) gene on chromosome 26 was significantly associated with LPS40 and SRBC20. *IL10* is a cytokine with anti-inflammatory and immunoregulatory activities. In broilers, Ghebremicael et al. (2008) found a strong association between SNPs in the *IL10* region and bacterial burden in the spleen and cecum following exposure to *Salmonella enteritidis*. LPS is a major component of the outer membrane of gram-negative germs. Thus, the association of *IL10* with NAb titres for LPS provides further evidence of the role of *IL10* in the response to gram-negative bacterial infections in poultry and points to its possible use in selecting for resistance to carrier state in bacterial diseases such as salmonellosis.

Linkage disequilibrium

LD between marker pairs in an animal population decreases exponentially with distance (Sved, 1971). Ardlie *et al.* (2002) established that the minimum amount of LD for QTL mapping, as measured by the statistic r^2, is 0.3. In laying hens, Aerts *et al*. (2007) estimated that an LD of 0.3 extends for about 4 cM, and Heifez *et al.* (2005) showed that about 60% of the markers separated by less than 5 cM have an LD $\geq$ 0.2, as measured by χ^2. Both studies looked at LD within lines. Between lines, LD is conserved at shorter distances and is a function of the correlation between LD measured in the two lines. The more diverse the lines, the shorter the distance at which LD is conserved across lines. Andreescu *et al.* (2007) estimated an average correlation of 0.52 between LD patterns in nine lines of broilers. In the present work, nine lines of layers were used; therefore, the across-line extent of LD was expected to be shorter than the 4 cM estimated within lines. Using the square of the average correlation coefficient between lines of broilers (0.52) as estimated by Andreescu *et al.* (2007), an approximation of the LD conserved across lines in the present work is ¼ of the within-line LD, or 1 cM. Thus, SNPs found to be significantly associated with a trait in the across-line analysis are likely to lie within 1 cM of the QTL. Therefore, SNPs spaced by a maximum of 2 cM could point to the same QTL. For example, the SNPs rs13584362 and rs13584670 on chromosome 5 lie within 2 cM of one another and are associated with both NAbs and complement activity; thus, a single QTL might control these immunological traits. On the same chromosome, SNPs rs13756966 and rs13757293 are separated by 1.4 cM and could also represent a unique QTL for NAbs and classical complement activity. Other examples of SNPs showing such tight spacing were found on chromosomes 4, 7, 13 and Z.

Andreescu *et al.* (2007) also showed that there is almost perfect correspondence between genetic distances between lines and correlations between their LD patterns. This means that knowing the genetic distance between lines provides a good estimate of the amount of LD conserved across lines. We estimated genetic distances in our lines using the method of Nei (1972). Genetic distance was smallest between lines B1 and B2 (0.05) and highest between lines B1 and W1 (0.26). Lines of White Leghorn and Rhode Island Red origin formed two distinct phylogenetic clusters, suggesting that LD patterns are better conserved within white and brown layers than between them. The analysis of white and brown layers separately showed that in fact, in addition to those detected by the across-line analysis, other and different SNPs were associated with immunity in the two subsets. For instance, rs13520872 and rs13520980 on chromosome 4 were found to be associated with SRBC40 in the across-line analyses. The same two SNPs were also significantly associated with the trait when analyzing white and brown layers separately. In addition, these separate analyses showed that two adjacent SNPs were significantly associated with SRBC40 in the white layers, and that four surrounding SNPs and one intermediate SNP were significantly

associated with SRBC40 in the brown layers. Similar patterns were observed on chromosome 5 with the SNP rs13586560 for KLH20 and with the SNPs rs15669480 and rs15669488 for BRBC65. These results illustrate the power of across-line analysis in refining association signals (Saccone *et al*., 2008).

Conclusions

The results of this work shed new light on the genetic background of immune response in birds. Several potential QTLs for immunity were identified, and the roles of *IL17A*, *IL12B* and *MHC* genes in immune function have been clarified by the identification of SNPs associated with natural and acquired antibody titres and complement activity. To our knowledge, these data constitute the first report of a simultaneous effect of a single locus on both antibody titres and complement activity, suggesting coordinated control of these distinct pathways. The involvement of the genes for the serotonin receptors suggests a fascinating relationship between behaviour and immunity.

The use of denser marker sets for genome-wide scans will help research move from the determination of SNP associations with QTLs, as reported in this paper, to the identification of genes and causal mutations that define the genetic basis of immune function.

Acknowledgements

This work was conducted as part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission.

References

Aerts J., Megens H.J., Veenendaal T., Ovcharenko I., Crooijmaans R., Gordon L., Stubbs L. & Groenen M. (2007) Extent of linkage disequilibrium in chicken. *Cytogenetic and Genome Research* **117**, 338-345.

Andreescu C., S. Avendano, S.R. Brown, A. Hassen, S.J. Lamont and J.C.M. Dekkers (2007) Linkage disequilibrium in related breeding lines of chickens. *Genetics* **177**, 2161-2169.

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**, 299-309.

Aulchenko Y.S., Ripke S., Isaacs A. & van Duijn C.M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296.

Bolhuis E.J., Ellen E.D., Van Reenen C.G., De Groot J., Ten Napel J., Koopmanschap R.E., De Vries Reilingh G., Uitdehaag K.A, Kemp B. & Rodenburg T.B. (2009) Effects of genetic group selection against mortality on behaviour and peripheral serotonin in domestic laying hens with trimmed and intact beaks. *Physiology & Behaviour* **97**, 470-475.

Fernando R.L. & Grossman M. (1989) Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* **21**, 467-477.

Ghebremicael S.B., Hasenstein J.R. & Lamont S.J. (2008) Association of Interleukin-10 cluster genes and *Salmonella* response in the chicken. *Poultry Science* **87**, 22-26.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. & Thompson R. (2002). ASReml User Guide Release 1.0 VSN International Ltd. Hemel Hempstead, HP1 1ES, UK

Hassen A., Avendano S., Hill W.G., Fernando R.L., Lamont S.J. & Dekkers J.C.M. (2009) The effect of heritability estimates on high-density single nucleotide polymorphism analyses with related animals. *Journal of Animal Sciences* **87**, 868-875.

Heifetz E.M., Fulton J.E., O'Sullivan N., Zhao H., Dekkers J.C.M. & Soller M. (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* **171**,1173-1181.

Hillel J., Groenen M.A.M., Tixier-Boichard M., Korol A.B., David L., Kirzhner V.M., Burke T., Barre-Dirie A., Crooijmans R.P.M.A., Elo K., Feldaman M.W., Freidlin P.J., Mäki-Tanila A., Oortwijn M., Thomson P., Vignal A., Wimmers K. & Weigend S. (2003) Biodiversity of

52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics Selection Evolution* **35**, 533-557.

International chicken genome sequencing consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-717.

Iwakura Y., Nakae S., Saijo S. & Ishigame H. (2008) The roles of IL-17A in inflammatory responses and host defense against pathogens. *Immunological reviews* **226**, 57-79.

Kaiser M.G., Deeb N. & Lamont S.J. (2002) Microsatellite markers linked to *Salmonella enterica* serovar enteritidis vaccine response in young F_1 broiler-cross chicks. *Poultry Science* **81**, 193-201.

Kennedy B.W., Quinton M.& van Arendonk J.A.M. (1992) Estimation of effects of single genes on quantitative traits. *Journal of Animal Science* **70**, 2000-2012.

Lammers A., Klomp M.E.V, Nieuwland M.G.B, Savelkoul H.F.J. & Parmentier H.K. (2004) Adoptive transfer of natural antibodies to non-immunized chickens affects subsequent antigen-specific humoral and cellular immune responses. *Developmental & Comparative Immunology* **28**, 51-60.

Lamont S.J., Pinard M.H., Cahaner A., van der Poel J. & Parmentier H.K. (2003) Selection for disease resistance: direct selection on the immune response. In: *Poultry Genetics, Breeding and Biotechnology* (ed. by Muir & Aggrey), pp. 399-414. CAB International, Wallingford, Oxon OX10 8DE, UK.

Lutz H.U. (2006) Homeostatic roles of naturally occurring antibodies: An overview. *Journal of Autoimmunity* **29**, 287-294.

Lutz H.U., Binder C.J. & Kaveri S. (2008) Naturally occurring auto-antibodies in homeostasis and disease. *Trends in Immunology* **30**, 43-51.

Matzinger P. (1994) Tolerance, danger, and the extended family. *Annual Review of Immunology* **12**, 991-1045.

Nei, M. 1972. Genetic distance between populations. *American Naturalist* **106**, 283-292.

Parmentier H.K.**,** Lammers A., Hoekman J.J, de Vries Reilingh G., Zanen I. & Savelkoul H.F.J. (2004a) Different levels of natural antibodies in chickens divergently selected for specific antibody responses. *Developmental & Comparative Immunology* **28**, 39-49.

Parmentier H.K., Baelmans R., Savelkoul H.F.J., Dorny P., Demey F. & Berkvens D. (2004b) Serum haemolytic complement activities in 11 different MHC (B) typed chicken lines. *Veterinary Immunology and Immunopathology* **100**, 25-32.

Pinard M.H., van Arendonk J.A.M., Nieuwland M.G.B. & van der Zijpp A.J. (1992) Divergent selection for immune responsiveness in chickens: estimation of realized heritability with an animal model. *Journal of Animal Science* **70**, 2986-2993.

Saccone N.L., Saccone S.F., Goate A.M., Gruzca R.A., Hinrichs A.L., Rice J.P. & Bierut L.J. (2008) In search of causal variants: refining disease association signals using cross-population contrasts. *BMC Genetics* **9**:58.

Siwek M., Buitenhuis A.J., Cornelissen S.J.B., Nieuwland M.G.B., Bovenhuis H., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K. & van der Poel J.J. (2003) Detection of quantitative trait loci for antibody responses to keyhole limpet hemocyanin and Mycobacterium butyricum in two unrelated populations of laying hens. *Poultry Science* **82**, 1845-1852.

Siwek M., Buitenhuis B., Cornelissen S., Nieuwland M., Knol E.F., Crooijmans R., Groenen M., Parmentier H.& van der Poel J. (2006a) Detection of QTL for innate: antibody levels binding LPS and LTA in two independent populations of laying hens. *Developmental & Comparative Immunology* **30**, 659-666.

Siwek M., Cornelissen S.J.B., Nieuwland M.G.B., Buitenhuis A.J., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K. & van der Poel J.J. (2006b) Erratum. Corrections for: Detection of QTL for immune response to sheep red blood cells in laying hens. *Animal Genetics* **37**, 608.

Star L., Frankena K., Kemp B., Nieuwland M.G.B. & Parmentier H.K. (2007) Natural humoral immune competence and survival in layers. *Poultry Science* 86, 1090-1099.

Sved J.A. (1971) Linkage disequilibrium and homozygousity of chromosome segments in finite populations. *Theoretical Population Biology* **2**, 125-141.

Walport M.J. (2001) Complement; second of two parts. *New England Journal of Medicine* **344**, 1140-1144.

Wigginton J.E., Cutler D.J. & Abecasis G.R. (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics* **76**, 887-893.

Yonash N., Bacon L.D., Witter R.L. & Cheng H.H. (1999) High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. *Animal Genetics* **30**, 126-135.

Yonash N., Cheng H.H., Hillel J., Heller D.E. & Cahaner A. (2001) DNA microsatellites linked to quantitative trait loci affecting antibody response and survival rate in meat-type chickens. *Poultry Science* **80**, 22-28.

Yunis R., Heller E.D., Hillel J. & Cahaner A (2002) Microsatellite markers associated with quantitative trait loci controlling antibody response to *Escherichia coli* and *Salmonella enteritidis* in young broilers. *Animal Genetics* 33, 407-414.

Zhou H., Li H. & Lamont S.J. (2003) Genetic markers associated with antibody response kinetics in adult chickens. *Poultry Science* 82, 699-708.

Table 1 Per-line distribution of the hens used in the study and mean and standard deviation of traits for the single lines and the overall population[1].

				Innate humoral immunity[2]						Adaptive humoral immunity[2]	Complement activity[3]					
				Exogenous antigen			Environmental antigen				Classical pathway			Alternative pathway		
Breed	**Line**	**N**		**KLH20**	**KLH40**	**KHL65**	**LPS20**	**LPS40**	**LPS65**	**NCD20**	**SRBC20**	**SRBC40**	**SRBC65**	**BRBC20**	**BRBC40**	**BRBC65**
Rhode Island Red	B1	68		3.00	3.16	4.01	4.71	4.65	5.40	7.42	113.1	216.7	402.6	68.2	79.2	227.2
	B2	67		2.55	2.86	3.89	4.01	4.56	5.00	5.89	197.1	199.8	730.5	50.8	95.5	341.3
	B3	71		2.05	2.51	2.71	3.67	3.74	4.25	6.11	203.7	311.6	560.0	53.9	87.8	373.1
	BB	62		1.76	2.74	2.25	2.94	3.72	4.56	5.00	195.9	207.5	511.3	59.0	58.4	320.2
White Leghorn	W1	55		2.96	4.09	4.50	3.63	4.96	5.00	6.41	256.0	196.2	563.2	78.1	67.9	217.1
	WA	70		1.38	2.60	3.05	3.52	4.23	3.96	5.69	282.9	163.7	433.7	86.8	31.9	108.0
	WB	65		2.12	2.51	2.34	2.87	3.49	3.39	4.53	110.7	175.9	420.3	28.3	51.1	126.7
	WC	56		3.00	2.45	2.37	3.54	4.28	4.20	6.04	229.5	204.7	520.4	58.4	60.4	219.3
	WF	69		3.08	4.12	4.28	3.13	5.75	4.93	5.17	372.8	226.5	442.7	155.4	130.8	284.8
			n	572	582	575	575	583	577	564	340	566	505	277	427	414
	Total	583	mean	2.41	3.00	3.26	3.58	4.37	4.52	5.83	238.6	212.7	511	81.7	81.7	255.7
			SD	1.59	1.47	1.59	1.44	1.36	1.40	2.13	162.8	113.7	277.8	53.8	42.4	165.0

[1]KLH20, KLH40, KLH65 = natural antibody titres for keyhole limpet hemocyanin at 20, 40, and 65 wk of age; LPS20, LPS40, LPS65 = natural antibody titres for lipopolysaccharide at 20, 40, and 65 wk of age; NCD20 = acquired antibody titres for the Newcastle disease virus at 20 wk of age; SRBC20, SRBC40, SRBC65 = classical complement activity at 20, 40, and 65 wk of age; BRBC20, BRBC40, BRBC65 = alternative complement activity at 20, 40, and 65 wk of age.

[2]Expressed as $-\log_2$ of dilutions that gave 50% extinction.

[3]Expressed as hemolytic doses 50 per ml.

Table 2 Distribution of SNPs per chromosome along the chicken genome, number of SNPs deviating from Hardy-Weinberg equilibrium, and homozygosity in the nine genotyped lines of layers[1].

chr	Size (Mbp)	SNPs		FL (across lines)	Within-line values																	
		All	Used		B1		B2		B3		BB		W1		WA		WB		WC		WF	
					HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL
1	201	68	61	6	4	19	1	17	1	12	3	16	0	24	8	21	0	20	0	15	1	23
2	155	24	22	2	1	6	0	7	1	3	1	5	0	10	2	8	1	10	1	9	0	9
3	114	140	121	12	3	31	0	35	2	28	7	22	0	41	9	40	0	34	0	27	0	37
4	94	421	371	65	9	102	1	126	6	120	13	105	2	141	24	142	11	117	8	126	16	126
5	62	285	265	25	9	48	6	56	7	46	22	41	3	78	23	58	4	64	20	62	8	76
6	37	27	22	1	2	3	2	3	0	3	3	2	0	4	3	2	2	4	1	3	1	3
7	38	175	149	18	5	33	1	35	5	30	9	35	3	48	14	42	5	43	10	39	8	43
8	31	6	6	0	2	1	1	0	0	1	2	0	0	3	2	2	2	2	0	0	0	0
9	26	12	10	0	0	2	0	2	0	2	0	5	1	4	0	1	0	3	0	3	0	6
10	22.6	4	3	0	0	1	0	1	2	1	0	2	1	0	0	0	1	0	0	1	0	1
11	21.9	8	6	3	0	3	0	3	0	4	0	4	0	4	1	4	1	4	0	4	0	4
12	20.5	7	7	2	0	4	0	3	0	4	0	3	1	3	0	5	0	4	2	3	0	3
13	18.9	37	32	3	2	6	0	9	0	9	3	5	0	9	9	7	1	8	4	7	3	7
14	15.8	7	6	0	0	0	0	1	0	0	0	1	1	1	0	0	2	0	0	1	0	1
15	13	7	5	1	0	2	0	2	0	2	0	1	0	4	0	4	1	2	1	2	0	1
16	0.43	17	12	3	0	3	2	3	3	3	0	3	1	4	1	4	2	4	1	3	0	4
17	11.2	7	7	0	0	1	0	1	0	1	1	1	0	1	2	1	0	0	1	1	0	1
19	9.9	27	20	4	2	6	2	6	2	5	2	6	1	9	2	6	1	7	2	8	0	6
21	7	4	4	1	0	1	0	1	0	2	0	1	0	2	1	3	2	2	1	2	0	2
22	3.9	3	2	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0
23	6	4	2	1	0	1	0	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1
24	6.4	16	15	3	0	6	0	4	0	4	1	5	2	3	2	5	1	5	0	4	1	4
26	5.1	60	50	4	6	12	2	12	2	12	3	11	2	16	8	15	1	12	4	17	0	11
Z	75	160	138	21	0	47	0	49	0	47	0	50	0	76	0	61	0	60	0	63	0	62
-		8	6	0	1	2	1	2	0	2	1	2	1	2	0	1	1	2	1	1	0	1
Total		1534	1342	175	46	340	19	379	31	342	71	327	20	488	113	433	40	408	57	402	39	432

[1]HW = loci not in Hardy-Weinberg equilibrium, FL = fixed loci. The symbol "-" refers to SNPs that were not assigned to any chromosome. Size refers to the size of the whole chromosome as derived from the NCBI chicken genome database.

Table 3 (continued on the next page) SNPs significantly associated with immunological traits. $-\log_{10}$ of the p-values are reported in the columns[1].

chr	SNP	kbps	cM[2]	KLH20	KLH40	KLH65	LPS20	LPS40	LPS65	NCD20	SRBC20	SRBC40	SRBC65	BRBC20	BRBC40	BRBC65
chr 3	rs13775078	3583	9.0													2.20
	rs13773493	4459	11.1			1.79				1.41			1.39			
	rs13503401	9873	24.7											2.88		
	rs14082130	17015	42.5						1.37			2.36			1.38	
	rs13526054 (*IL17A*)[3]	110370	275.9					2.11					1.45		1.42	
	rs15458146 (*IL17A*)[3]	110371	275.9					2.44		2.30						
chr 4	rs15475503 (*HTR2C*)[3]	2796	7.0										3.06			
	rs13513104	31200	78.0	1.57			1.41		1.43							
	rs13513581	32423	81.1				1.85	1.47	1.94							
	rs13515125	35502	88.8							1.39						2.23
	rs13517985	44847	112.1			1.32				1.84					2.05	
	rs13520611	50650	126.6					1.52				1.46			1.57	
	rs13520637	50702	126.8					1.85								
	rs13520872	51260	128.1			1.93						3.63				
	rs13520980	51437	128.6									3.68				
	rs13521841	53421	133.6							2.57		1.96				2.01
chr 5	rs13756966	13353	33.4			2.08						1.38				
	rs13757293	13936	34.8			2.33						2.76				
	rs15669480 (*TOLLIP*)[3]	15790	39.5													2.34
	rs15669488 (*TOLLIP*)[3]	15792	39.5													1.72
	rs15676823	21363	53.4							3.34						
	rs13755931	27450	68.6												3.70	
	rs13756481	28876	72.2												3.79	
	rs13584362	30234	75.6					1.74	2.05			1.31		1.39	2.00	
	rs13584670	30998	77.5						1.43				1.80			
	rs13584901	31328	78.3	2.96												
	rs13585105	32180	80.5	1.56	1.59											
	rs13585538	36870	92.2		1.51	1.36			1.79							
	rs15703226	39904	99.8									1.57	2.35			
	rs13586560	40114	100.3	2.47												

	rs13586776	40720	101.8						2.10			2.21		1.77	2.60
chr 6	rs14580491 (*CXCL12*)[3]	20117	50.3											2.01	2.62

Table 3 (continued): SNPs with significant associations with immunological parameters. –log10 of the p-values are reported in the columns.

chr	SNP	kbps	cM^2	KLH20	KLH40	KLH65	LPS20	LPS40	LPS65	NCD20	SRBC20	SRBC40	SRBC65	BRBC20	BRBC40	BRBC65
chr 7	rs13596817	27272	68.2			2.28							1.70		3.43	2.76
	rs13596877	27403	68.5	3.42	1.66	1.54										
	rs13599559	33001	82.5		2.15	1.56		1.57				2.45			1.80	
	rs13600367	35125	87.8													2.27
chr 9	rs15986720	23820	59.6							1.39			1.69		1.61	
chr 13	rs15677371	7927	19.8				3.26									
	rs15677377	7927	19.8	1.40	2.33	1.69										
	rs14050302 (*IL12B*)[3]	7917	19.8		2.31	1.70										
	rs14064765 (*GMCSF*)[3]	17233	43.1										3.02			
	rs14064896 (*IRF1*)[3]	17452	43.6						1.34							3.30
	rs14064900 (*IRF1*)[3]	17452	43.6							1.53	1.62		1.47			
chr 16	snp.gga16Tapasin.180144exon1TC (*TAPBP*)[3]	65	0.2							2.26		1.89	1.68			
	rs15788216 (*MHC / BLB1*)[3]	70	0.2					3.29		1.47						
chr 19	rs14119843 (*HSPB1*)[3]	4218	10.5												2.94	
	rs15047494	4759	11.9		2.58											
chr 21	rs16180997	4089	10.2												1.81	
chr 22	rs16183803	3810	9.5					2.38		1.75			1.45			
chr 24	rs16197918 (*HTR3A*)[3]	4490	11.2						2.64							
chr 26	rs14298900 (*IL10*)[3]	2375	5.9					1.61			1.46					
	rs13606449	3332	8.3										1.36	1.64		
chr Z	rs16099653	18236	45.6		2.04											
	rs13676733	18253	45.6		2.04											
	rs13676747	18402	46.0		2.07											
	rs16102303	11935	29.8					1.93								

rs13615389	unmapped												1.37					
rs13734043	23316	58.3									2.02							
rs16106323	28051	70.1						2.22										

[1]KLH20, KLH40, KLH65 = natural antibody titres for keyhole limpet hemocyanin at 20, 40, and 65 wk of age; LPS20, LPS40, LPS65 = natural antibody titres for lipopolysaccharide at 20, 40, and 65 wk of age; NCD20 = acquired antibody titres for the Newcastle disease virus at 20 wk of age; SRBC20, SRBC40, SRBC65 = classical complement activity at 20, 40, and 65 wk of age; BRBC20, BRBC40, BRBC65 = alternative complement activity at 20, 40, and 65 wk of age.

[2]1cM = 4 x 10^5 bps

[3]*IL17A* = gene encoding interleukin 17A; *HTR2C* = gene encoding the serotonin receptor 2C; *IL12B* = gene encoding interleukin 12B; *TOLLIP* = gene encoding TLR interaction protein; *CXCL12*= gene encoding the CXCL12 chemokine; *GMCSF* = gene encoding granulocyte macrophage colony-stimulating factor (cytokine); *IRF1* = gene encoding interferon regulatory transcription factor; *HSPB1* = gene encoding heat shock 27kD protein; *MHC* = gene encoding major histocompatibility complex; *BLB1* = gene encoding class II beta chain 1; *TAPBP* = gene encoding TAP binding protein; *HTR3A* = gene encoding serotonin receptor 3A; *IL10* = gene encoding interleukin 10.

Table 4 Genotypic effects for the SNPs with a –log(p_value) > 2.5. Effects are expressed as deviations from the BB genotypes in phenotypic standard deviations of the trait. Standard errors are reported in brackets.

Chr	SNP	cM[1]	Trait[2]	AA		A/B	
chr 3	rs13503401	18.1	BRBC20	0.61	(0.26)	0.46	(0.14)
chr 4	rs15475503 (*HTR2C*)[3]	6.6	SRBC65	-0.63	(0.22)	-0.46	(0.14)
	rs13520872	125.8	SRBC40	0.55	(0.13)	0.24	(0.11)
	rs13520980	126.2	SRBC40	0.76	(0.21)	0.44	(0.13)
	rs13521841	131.1	NCD20	-1.31	(0.38)	-1.18	(0.38)
chr 5	rs13757293	26.0	SRBC40	-0.31	(0.13)	-0.38	(0.11)
	rs15676823	41.9	NCD20	0.52	(0.16)	-0.08	(0.12)
	rs13755931	56.3	BRBC40	-0.46	(0.17)	-0.47	(0.11)
	rs13756481	60.0	BRBC40	0.57	(0.20)	0.04	(0.20)
	rs13584901	64.8	KLH20	-0.71	(0.21)	-0.51	(0.16)
	rs13586776	87.9	BRBC65	-0.55	(0.21)	-0.45	(0.15)
chr 6	rs14580491 (*CXCL12*)[3]	45.5	BRBC65	0.42	(0.17)	-0.04	(0.16)
chr 7	rs13596817	66.0	BRBC40	0.03	(0.33)	-0.55	(0.14)
			BRBC65	-0.37	(0.35)	-0.56	(0.16)
	rs13596877	66.3	KLH20	0.55	(0.14)	0.44	(0.13)
chr 13	rs15677371	3.6	LPS20	0.21	(0.16)	0.48	(0.13)
	rs14064765 (*GMCSF*)[3]	39.3	SRBC65	-0.50	(0.14)	-0.09	(0.10)
	rs14064896 (*IRF1*)[3]	39.8	BRBC65	-0.63	(0.17)	-0.10	(0.13)
chr 16	rs15788216 (*MHC* / *BLB1*)[3]	0.4	LPS40	-0.67	(0.17)	-0.15	(0.12)
chr19	rs14119843 (*HSPB1*)[3]	10.0	BRBC40	0.45	(0.15)	0.05	(0.12)
	rs15047494	11.2	KLH40	-0.52	(0.20)	0.04	(0.16)
chr 24	rs16197918 (*HTR3A*)[3]	10.8	LPS65	0.53	(0.15)	0.16	(0.12)

[1]1cM = 4 x 10^5 bps

[2]KLH20, KLH40, KLH65 = natural antibody titres for keyhole limpet hemocyanin at 20, 40, and 65 wk of age; LPS20, LPS40, LPS65 = natural antibody titres for lipopolysaccharide at 20, 40, and 65 wk of age; NCD20 = acquired antibody titres for the Newcastle disease virus at 20 wk of age; SRBC20, SRBC40, SRBC65 = classical complement activity at 20, 40, and 65 wk of age; BRBC20, BRBC40, BRBC65 = alternative complement activity at 20, 40, and 65 wk of age.

[3]*HTR2C* = gene e ncoding serotonin receptor 2C; *CXCL12*= gene encoding the CXCL12 chemokine; *GMCSF* = gene encoding granulocyte macrophage colony-stimulating factor (cytokine); *IRF1* = gene encoding interferon regulatory transcription factor; *MHC* = gene encoding major histocompatibility complex; *HSPB1* = gene encoding heat shock 27kD protein; *HTR3A* = gene encoding serotonin receptor 3A.

Chapter 5

Across-line SNP association study for direct and associative effects on feather damage in laying hens

F. Biscarini*, H. Bovenhuis*, J. van der Poel*, T. B. Rodenburg*, A. P. Jungerius[#] and J. A. M. van Arendonk*

** Animal Breeding and Genomics Centre, Wageningen University P.O. Box 338, 6700 AH Wageningen, The Netherlands*
 # Hendrix Genetics, Boxmeer, The Netherlands

Behavior Genetics (2010) doi: 10.1007/s10519-010-9370-0

ABSTRACT

An association study between SNP markers and feather condition score on the back, rump and belly of laying hens was performed. Feather condition score is a measure of feather damage, which has been shown to be closely related to feather pecking behaviour in hens housed in groups. A population of 662 hens was genotyped for 1536 SNPs of which 1022 could be used for the association study. The analysis was conducted across 9 different lines of White Leghorn and Rhode Island Red origin. Across lines linkage disequilibrium is conserved at shorter distances than within lines; therefore, SNPs significantly associated with feather condition score across lines are expected to be closer to the functional mutations. The SNPs that had a significant across-line effect but did not show significant SNP-by-line interaction were identified, to test that the association was consistent across lines. Both the direct effect of the individual's genotype on its plumage condition, and the associative effect of the genotype of the cage mates on the individual's plumage condition were analysed. The direct genetic effect can be considered as the susceptibility to be pecked at, whereas the associative genetic effect can be interpreted as the propensity to perform feather pecking. Finally, 11 significant associations between SNPs and behavioural traits were detected in the direct model, and 81 in the associative model. A role of the gene for the serotonin receptor 2C (HTR2C) on chromosome 4 was found. This supports existing evidence of a prominent involvement of the serotonergic system in the modulation of this behavioural disorder in laying hens. The genes for IL9, IL4, CCL4 and NFKB were found to be associated to plumage condition, revealing relationships between the immune system and behaviour.

Key words: feather pecking, plumage condition, association study, SNPs, laying hens, direct and associative effects

INTRODUCTION

Feather pecking (FP) is one of the most serious behavioural disorders of laying hens. Severe FP, the type of pecking that causes most feather damage, consists of the forceful pecking and pulling of feathers of other birds (Savory 1995). FP is a multifactorial problem caused by both genetic and environmental factors. There is evidence of line differences in FP (Kjaer et al. 2001; Uitdehaag et al. 2008), and it has been demonstrated that FP is influenced by group size, light intensity, diet and type of litter (Hughes and Duncan 1972; Blokhuis and Arkes 1984; Savory 1995). As for its aetiology, FP has been considered to be redirected ground pecking (Blokhuis, 1986), abnormal dustbathing behaviour (Vestergaard and Lisborg, 1993), or the consequence of a more general hyperactivity disorder (Kjaer 2009). Most of the evidence point at the redirected ground pecking theory (Huber-Eicher and Wechsler 1997), with active or even hyperactive birds having the highest risk of developing FP (Newberry et al. 2007; Kjaer 2009). The serotonergic system has been shown to play an important role in the modulation of FP (van Hierden et al. 2002; van Hierden et al. 2004a, b; Buitenhuis et al. 2006). Especially gentle FP is viewed as a stereotyped behaviour, which shows similarities with obsessive compulsive disorders in other species, in which the serotonergic system plays a comparable role (Pigott 1996). Feather pecking, which can lead to feather damage and cannibalism, thereby causing mortality and economic losses for the laying industry, has been traditionally controlled through the husbandry practice of beak-trimming. However, the EU laying hens directive 1999/74/EC is causing member states to move from conventional cages to larger groups (furnished cages, non-cage systems) which will make the problem more difficult to control; and in some countries beak trimming, as preventive measure, is or will be prohibited (Jendral and Robinson 2004). Selection of more sociable animals with a less pronounced tendency to peck each other might therefore be highly beneficial to the farming of layers. Feather pecking has already been shown to be heritable (Kjaer and Sorensen 1997; Rodenburg et al. 2003), and it has been demonstrated that individual selection against FP is feasible (Kjaer et al. 2001). Social interactions have been revealed to play a role in survival related to FP and cannibalism in laying hens (Ellen et al. 2008). This associative effect due to the genotypes of group mates can contribute substantially to the total heritable variation (Bijma et al. 2007a, b). However, measuring feather pecking requires direct observations which are time consuming and expensive. A convenient indirect way of measuring FP is looking at plumage condition: Bilčík and Keeling (1999) showed that feather condition scores used to assess plumage condition are related to feather pecking activity. Few genetic studies on plumage condition have been carried out until now (Jensen et al. 2005). Previous studies detected microsatellite markers located in chromosomal regions involved in feather pecking (Buitenhuis et al. 2003a, b).

The aim of this study was to detect associations between mutations in the genome and feather damage across lines of laying hens, focusing both on the feather peckers and on the victims of feather pecking. We performed an association study between SNP markers and feather condition score across 9 lines of layers of White Leghorn (white feathered) and Rhode Island Red (brown feathered) origin, looking at the interaction between the SNP and line effects (Saccone et al. 2008; Biscarini et al. 2010). SNPs showing a significant across line association and no SNP-by-line interaction were considered to be consistently associated with the phenotype. Both the direct genetic effect of the individual and the associative genetic effect of cage mates on feather condition scores were analysed. The direct effect can be considered the susceptibility to receive feather pecking, while the associative effect reflects the propensity to express a pecking behaviour. To our knowledge this is the first time that the associative genetic effect is analysed in an association study.

MATERIAL AND METHODS

Experimental Population

The animal population used in this study consisted of 662 laying hens randomly chosen from 9 genetic lines, 4 of Rhode Island Red type (RIR, brown feathered) and 5 of White Leghorn type (WL, white feathered). The number of hens per line is reported in Table 1. The birds originated from mating 175 roosters with 401 dams. Every rooster was mated on average with 2.3 females while each hen was mated with only one male. Four generations of ancestors were extracted from the pedigree file for the calculation of the additive genetic relationships among the birds.

All hens were housed in battery cages (44 cm height x 46 cm depth x 39 cm width) within the same stable; cages comprised 4 hens from the same line, either full-sibs or randomly mixed. Hens arrived at the laying facility at 17 wk of age and remained in the stable for the entire laying period of 52 wk.

The hens had intact beaks and received routine vaccinations against Marek's disease (d 1), New Castle disease (wk 2, 6, 12, 15), infectious bronchitis (d1, wk 2, 10, 12, 15), infectious bursal disease (wk 3, 15), fowl pox (wk 15) and avian encephalomyelitis (wk 15).

During the experiment, feed and water were available ad libitum. From the beginning of the experiment (at 19 wk of age) until 42 wk of age hens were fed a standard commercial phase 1 diet (159 g/kg crude protein, 43 g/kg crude fibre, and 11.17 MJ ME/kg); from 42 wk onwards, until the end of the experiment, a standard commercial phase 2 diet (152 g/kg crude protein, 47.0 g/kg crude fibre, and 11.01 MJ ME/kg) was given. Wing bands allowed individual identification of the hens. After arrival in the stable, hens were kept on a 9L:15D light scheme (light from 7.00 until 16.00), where L stands for light and D for darkness. After 1 wk, the light period was increased by 30 min,

starting at 6.30. Thereafter, the light period was increased approximately 10 min per day. From 30 wk onwards hens received light from 00.00 until 16.00 (16L:8D). This is a standard light regime.

Phenotypes

Feather damage was assessed at two ages (51 wk and 69 wk) by assigning a score to plumage condition on the back, rump and belly of the hens. Damage to these regions is unlikely due to abrasion and these regions are a frequent target of feather pecking (Bilčík and Keeling 1999). The classification of Bilčík and Keeling (1999) was followed, with a range going from 0 to 5 (higher scores indicate more severe damage). Damage to the rump and back area was combined into a single score: back and rump feather scores were summed to give a backrump (BR) score ranging from 0 to 10, as previously described in Uitdehaag et al. (2008). Feather condition scores measured at 51wk and 69 wk of age were used in the analysis (Belly51, BR51, Belly69 and BR69). The number of available observations ranged from 655 at 69 wk to 662 at 51 wk (see Table 1). Part of the phenotypes used in this study were previously analysed by Uitdehaag et al. (2008).

Genotypes

Genotyping was done in a 1536-plex format using the GoldenGate assay (Illumina, San Diego) by a commercial genotyping facility (ServiceXS, Leiden, NL). As part of a bigger experiment, SNPs were selected to cover QTL regions for behavioural and immune traits identified in previous mapping studies catalogued in the QTL database (http://www.genome.iastate.edu/cgi-bin/QTLdb/GG/index, accessed May 2006). In addition, specific candidate genes, which from literature (mouse/human) have a known or expected effect on immune or behavioural traits (e.g. genes for interleukins or serotonin receptors), were considered for the choice of the SNPs. Per selected gene 2 to 4 SNPs were chosen; for a QTL region SNPs equally spread over the QTL region were chosen. Twenty-four of the 39 chromosomes of the chicken genome were partly covered (Table 2). The SNPs that did not show three distinct clusters in the allelic discrimination plot as provided by the Illumina Beadstudio software (194 SNPs) and the SNPs with a minor allele frequency ≤ 0.05 (320 SNPs) were excluded from the analysis. Thus, 1022 SNPs were used in the association study. Details on the SNP editing procedure are described by Biscarini et al. (2010).

Data Analysis

The direct genetic effect of the individual's SNP genotype and the associative genetic effect of the SNP genotypes of cage mates on feather condition scores were analysed. A two-step procedure was used for data analysis. First, an association study was performed without accounting for

relationships between animals. Potentially interesting SNPs from the first step were then analyzed in more detail using a mixed model thus taking additive genetic relationships among animals into account.

The association between genotypes and phenotypes was tested across lines using a single SNP approach. The across-line analysis picks up only markers that are in LD with the QTL across lines.

The direct genetic effect refers to the association of the SNP genotype of a hen and its plumage condition score. In the first step analyses were performed across lines using the following statistical model:

$$y_{ijk} = \mu + SNP_i + line_j + (line \times SNP)_{ij} + e_{ijk} \qquad (1)$$

where y_{ijk} represents the feather damage score of animal k, with SNP genotype i, in line j; μ is the overall mean; SNP_i is the effect of the SNP genotype; $line_j$ is the line effect (nine classes); $(line \times SNP)_{ij}$ is the interaction between SNP genotype and line; and e_{ijk} are the residuals.

Model [1] was run twice, once without the line-by-SNP interaction term to obtain p-values for the SNP effect and once with the line-by-SNP term to determine the significance of the interaction. Adapting the approach of Saccone et al. (2008), we looked for SNPs that had an across-line significant effect in model [1] ($p \leq 0.05$) and did not show a significant line-by-SNP interaction ($p > 0.05$) , to test that the association was consistent in all lines.

The false discovery rate (FDR, Benjamini and Hochberg 1995) was calculated for all the SNPs tested in the association study. SNP-phenotype associations from model [1] with a FDR ≤ 0.15 were selected for the second step of the association study in which a polygenic effect was added to the model to account for family relationships among animals. The following mixed model was used:

$$y_{ijk} = \mu + SNP_i + line_j + a_k + e_{ijk} \qquad (2)$$

where all the terms are as specified in equation [1] except a_k, which is the random genetic effect of the k_{th} animal. Var(**a**) = $\mathbf{A}\sigma_a^2$, with **A** being the additive relationship matrix and Var(**e**) = $\mathrm{I}\sigma_e^2$. The ratio between residual and genetic variances was fixed using heritabilities for feather condition score in laying hens estimated by van der Winkel on 17009 White Leghorn hens (unpublished results), averaged over the three lines (6324, 7018 and 3667 hens, respectively) used in that study: 0.03 for BR51, 0.08 for Belly51, 0.17 for BR69 and 0.20 for Belly69. The SNPs that still showed a significant effect ($p \leq 0.05$) on the traits from model [2] were reported.

The associative genetic effect refers to the effect of the SNP genotypes of cage mates on plumage condition of individual hens. In the analysis, individual feather condition scores were regressed on the allele frequency of cage mates. Note that the SNP genotype of the animal itself is not included in the analysis. Cages with 3 or 4 hens (hence 2 or 3 cage mates) were considered in the analysis. The following model was used:

$$y_{ijk} = \mu + \beta_1 p_k + line_j + \beta_{kj}(line \times p)_{kj} + e_{ijk} \qquad (3)$$

where y_{ijk} represents the feather condition score of animal i of line j in cage k; μ is the overall mean; p_k is SNP allele frequency of the cage mates of animal i; $line_j$ is the line effect (nine classes); $(line \times p)_{kj}$ is the interaction between SNP allele frequency of cage mates and line; and e_{ijk} are the residuals, weighted for the number of group mates present in each cage (either 2 or 3). β_1 and β_{kj} are the regression coefficients. Equation [3] was also run twice, with and without the interaction term. SNPs with an across-line significant associative genetic effect ($p \leq 0.05$) and no significant line-by-allele frequency interaction ($p > 0.05$) were considered to be consistently associated with the phenotypes in all lines.

Also in the case of the associative genetic effect, SNP-phenotype associations from model [3] with a FDR ≤ 0.15 were selected for the second step of the association study in which family relationships among animals were accounted for. The following mixed model was used:

$$y_{ijk} = \mu + \beta_1 p_k + line_j + a_i + e_{ijk} \qquad (4)$$

where all the terms are as specified in equation [3] except a_i, which is the random genetic effect of the i_{th} animal. The variance structure is as specified for model [2].

Data editing, analyses with models [1] and [3], and all other statistical analyses were performed using the open source statistical package R. The polygenic effects, and SNP effects as described in models [2] and [4] were estimated with a REML procedure using the Asreml software package (Gilmour et al. 2002).

RESULTS

Descriptive statistics

Descriptive statistics of individual feather condition scores are summarized in Table 1. The damage to the plumage due to feather pecking cumulates over time; therefore feather condition scores at older ages (BR69, Belly69) are about 1.5 to 2 times higher than feather condition scores at younger

ages (BR51, Belly51). There was more feather damage on the back-rump region than on the belly area (feather damage score of 1.84 vs 1.22 at 51 wk, and of 2.71 vs 2.57 at 69 wk, after correcting for the different scale). The overall across-line coefficient of variation ranged from 39% for BR69 to 111% for Belly51, indicating considerable variability in the feather condition scores of hens. Also within lines there was substantial variability: the coefficient of variation ranged from 25% for BR69 in line B3 to 532% for Belly51 in line B2, with an average of 75%. There were ample differences in feather condition scores between lines: e.g., from 1.09 (line B2) to 5.86 (WC) for BR51, and from 1.68 (line B2) to 5.66 (line W1) for Belly69. Brown layers showed less feather damage than white layers, for both regions and ages (on average more than 2 times lower), as described by Uitdehaag et al. (2008). Phenotypic correlations (results not shown) among the traits were weak, ranging from 0.14 between BR69 and Belly51, to 0.53 between both BR51 and Belly51, and BR69 and Belly69.

SNPs

The SNPs used in this study were located on 24 of the 39 chromosomes of the chicken genome. Positions of the SNPs were derived from the NCBI database (Galgal2.1 build 128). The average interval between SNPs varied from 25.29 Kbps (~ 0.06 cM) on chromosome 16 to 5740.74 Kbps (~14.35 cM) on chromosome 2. On average, the proportion of SNPs that deviated significantly (at the Bonferroni-corrected 0.05 level) from HW equilibrium within lines was 7.1% (Table 2), with the lowest percentages in lines W1 (2.7%) and B3 (5.7%) and the highest percentages in lines WA (13.1%) and BB (9.7%). Table 2 also reports the number of monomorphic (fixed) loci for the various chromosomes in the different lines. There were more fixed loci in the white layers (28.2%) than in the brown layers (22.6%). This is compatible with White Leghorn hens' longer history of artificial selection, is expected to result in higher homozygosity and lower genetic polymorphism (Hillel et al. 2003).

Association study

In the direct analysis where individual SNP genotypes were associated with individual feather condition scores, 321 significant across-line SNP-trait associations were detected at the 5% significance level: of these 275 showed no significant genotype-by-line interaction. Among these 275, there were 11 SNPs with a FDR lower than 0.15. After accounting for relationships among animals in the model, all the 11 SNPs still showed significant associations. The results of the analysis are shown in Table 3. The reported p-values come from model [2].

In the associative analysis where the allele frequency of the cage mates is related to the individual feather condition score, 478 significant across-line SNP-trait associations were detected ($p \leq 0.05$) in the first step of the association study: of these 357 showed no significant genotype-by-line interaction. Eighty-one (81) of these had a FDR ≤ 0.15. After accounting for relationships between animals, 57 of these SNP showed significant associations ($p \leq 0.05$) and are given in Table 4. The reported p-values come from model [4].

In the analysis of direct genetic effects, the detected associations comprised 4 SNPs for BR51, 2 for Belly51 and 5 for BR69. All SNPs reported in Table 3 had effect on one trait, with the exception of SNP rs15385785 on chromosome 1, which had an effect on the plumage condition of the back-rump region both at 51 and 69 wk. In many cases, neighbouring SNPs also showed effects, but had a FDR > 0.15 and are therefore not reported in Table 3.

In the associative analysis there were 27 SNPs associated to BR51, 19 to Belly51, 15 to BR69 and 7 to Belly69. Forty-seven of the SNPs reported in Table 4 were associated with only one trait, 9 with two traits and 1 SNP was associated with three different traits.

Two SNPs proved to be significantly associated with feather condition score both in the direct and associative model: SNPs rs13640917 on chromosome 4 and rs14999300 on chromosome 13. They were both associated with plumage condition on the back and rump regions at 51 wk. The same allele of SNP rs13640917, in the sequence of the serotonin receptor 2C (HTR2C), was associated with more feather damage on the back and rump at 51 wk in the direct as well as in the associative analysis, with an effect of approximately 0.5 σ_p in both analyses. SNP rs16340917 is a SNP at position 2798627 on chromosome 4 of the chicken genome: it is an intronic SNP within the HTR2C gene. As for SNP rs14999300, alternative alleles were associated with greater damage on the back and rump at 51 wk in the direct and associative analysis, with an effect of about 0.75 σ_p in the direct analysis and 0.6 σ_p in the associative analysis.

The strongest associations from the direct analysis were found for Belly51 ($-\log_{10}(p\_value) = 4.12$) and BR51 ($-\log_{10}(p\_value) = 3.56$) on chromosome 4. With the associative model the strongest associations were with BR51 on chromosome 3 ($-\log_{10}(p\_value) = 6.70$) and on chromosome 5 ($-\log_{10}(p\_value) = 4.27$), and with BR69 on chromosome 5 ($-\log_{10}(p\_value) = 4.67$).

Based on this association study, some genomic regions of interest for feather pecking behaviour in laying hens have been identified. From the associative model, the SNPs rs13717237, rs13717379, rs13717382, rs13717441 and rs13717447 on chromosome 3, all lying in less than 1 cM, were associated with the traits BR51 and Belly51. The average across lines LD between these SNPs, measured by r^2, was 0.45. Again on chromosome 3, the SNPs rs13717686, rs13717773 and

rs13717778, all in a range of 0.5 cM and with an r^2 of 0.15 and 0.34 respectively, were found to have an effect on feather condition of the back and rump region at 51 and 69 wk. These two regions on chromosome 3 were not in LD: the average r^2 between them was lower than 0.01. This suggests that these are two distinct regions with different genes influencing feather pecking behaviour. On chromosome 7 a region with effect on BR51 and Belly69 was identified at SNPs rs13598049, rs13598125 and rs13598160, spanning for 0.5 cM (see Table 4). The LD between these two pairs of SNPs was 0.11 and 0.18, as measured by r^2.

In the direct model, the percentage of the phenotypic variance explained by the SNPs reported in Table 3 ranged between 1% and 4%, with an average of 1.65%. The SNPs explaining the highest proportion of the phenotypic variance were rs13773912 on chromosome 3 (2.93% of σ_p^2) and rs13640917 on chromosome 4 (4% of σ_p^2).

For the SNP effects of Table 4 (associative model), the regression coefficients of allele frequency on feather condition score were estimated (results not shown): these had on average a magnitude of 0.6 σ_p, with a maximum of 1.22 σ_p and a minimum of 0.22 σ_p. For example, the SNP rs13640917 in the gene for the serotonin receptor 2C (HTR2C) had an effect of 0.54 σ_p on BR51, and the SNP rs13717447 on chromosome 3 had an effect of 1.22 $\boldsymbol{\sigma_p}$ on Belly51**.** Allelic frequencies for cage mates vary from 0 to 1, therefore the estimated regression coefficients reflect the maximum values when all cage mates are homozygous for the same allele.

DISCUSSION

In the present study we looked at genetic marker effects on plumage condition from two different perspectives: the direct effect of the genotype of the individual on its feather condition, and the associative effect of the allele frequency of its cage mates on the individual's feather condition. They reflect the genetic influence of, respectively, the individual hen and the cage mates on individual feather condition. Since feather damage on the back, rump and belly regions can be attributed almost exclusively to feather pecking (Bilčík and Keeling, 1999), the genetic direct and associative effects as described in this paper can be interpreted as the propensity to either receive or perform feather pecking. However, interactions among cage mates may result in a feather condition score different than expected, as for instance in cages where all hens are peckers.

With the associative model more significant results than with the direct model were found and significance levels from the associative model were generally higher than those from the direct model. This can be explained in part by the fact that performing feather pecking is more heritable than receiving it: heritability estimates for performing and receiving FP ranged from 0.12 to 0.56,

and from 0.00 to 0.15, respectively (Kjaer and Sorensen 1997, Rodenburg et al. 2003). This suggests a larger basis for associative rather than direct genetic effects. Furthermore, associative effects can contribute substantially to the total heritable variation, and have been shown to be often bigger in magnitude than direct genetic effects (Bijma et al. 2007a, b; Ellen et al. 2008).

In our experiment mortality was not high: blood samples for genotyping were collected at 40 wk of age, and only 13 of the genotyped hens died before having a feather condition score or between recording it at 51 and 69 wk. We looked at their SNP allelic frequencies and noticed that they generally had higher frequencies of the unfavourable alleles for feather damage, compared to the hens with both genotype and phenotypic observation (survivors). For SNP rs13640917 (HTR2C) on chromosome 4, for instance, the unfavourable allele frequency was 0.92 in the 13 dead hens and 0.62 in all females. This might imply that in those cases feather pecking led to cannibalism and death.

Methodology

In this study we simultaneously analysed multiple lines using a two-step procedure. In the across-line analysis, the SNPs detected were expected to be closer to the genes for behavioural traits due to the reduced extent of LD conserved across lines. Building on the work by Saccone et al. (2008), we tested for the SNP-by-line interaction to ensure consistency of the association across lines. The method is visually illustrated in Figure 1, where a region of chromosome 1 is reported. Histogram 1A reports the results of the analysis when all the lines are included. Histograms 1B and 1C report the results for the white and brown layers. These constitute two subpopulations of different origin (White Leghorn and Rhode Island Red, respectively), both comprising different lines. White and brown layers form two distinct phylogenetic clusters (Biscarini et al. 2010), implying that LD patterns are better conserved within than across them (Andreescu et al. 2007). When analysing the two subpopulations separately many SNPs result to be significantly associated with the phenotype (BR51): 3 in the white layers and 8 in the brown layers. This reflects the larger extent of LD conserved within homogenous populations in which fewer recombinations have occurred and more surrounding SNPs are linked to the QTL. When combining the two subpopulations in the across-lines analysis, most of those SNP association signals are lost, due to either lower significance of association or significant SNP-by-line interaction. Finally, only 1 SNP (rs15385785) is consistently associated with the phenotype in all lines, and due to the lower extent of LD conserved across all lines, this is likely to be closer to the QTL for the trait. The same procedure was applied by Biscarini et al. (2010) in an association study of immune response in laying hens.

A possible source of false positive associations due to population stratification was avoided by including a line effect in the model. Consequently, SNPs explaining part of the between-line variation could not be detected in this approach. Family relationships within lines could be another source of false positive associations which was dealt with by including a polygenic effect in the model that accounted for the effects of all other genes on the trait. After taking into account family relationships, the significance levels as measured by the opposite of the logarithm of p values, decreased on average by about 10% in the direct model and by about 20% in the associative model. The analysis proved to be robust to variations in heritability: heritabilities were varied with limited impact on the significance of the results This agrees with the results of Hassen et al. (2009).

Detected associations

Some of the results of the present work confirmed findings from previous QTL mapping studies. Buitenhuis et al. (2003a, b) detected a QTL for receiving FP on chromosome 5, and QTLs for performing FP on chromosomes 1, 4, 13 and 24. QTLs for fear related behaviours were found on chromosome 1 by Schütz et al. (2004) and Buitenhuis et al. (2004), and on chromosomes 3 and 4 by Buitenhuis et al. (2004). A relation between fearfulness and FP and its consequences has been revealed by several studies (Jones et al. 1995; Rodenburg et al. 2004; Uitdehaag et al. 2008).

Several of the associations detected in the present study were in the sequences of candidate genes that could play a role in behaviour. These include the genes for the monoamine oxidase of type A (MAO-A) and the serotonin receptor (HTR2C) from the direct model, and the genes for the cation channels (TRPM3, TRPC3), the neuronal transcription factor NPAS3 and again the serotonin receptor (HTR2C) from the associative model. In the case of SNP rs13640917, in the sequence of the serotonin receptor 2C (HTR2C), the same allele was associated with more feather damage on the back and rump at 51 wk in the direct as well as in the associative analysis. Theoretically it is expected that the opposite SNP alleles will be associated with greater feather damage in the direct analysis (receiving FP) and in the associative analysis (performing FP). That the same allele of SNP rs13640917 is associated with feather damage both in the associative and direct analysis is therefore against expectations. If one allele is associated with more feather damage in the direct analysis (feather pecking received) it should be associated with lower feather damage in the associative analysis (FP performed). Unless, in the case of a gene that leads to higher FP behaviour, all animals in a cage have the "positive" allele (more FP). Then there will be more feather pecking in the cage, and also peckers (and not only receivers) will show higher feather damage. It is not uncommon that feather peckers get pecked themselves. If there are 4 peckers in a cage this may even be likely.

Associations with feather condition score have been found also for the genes of the interleukins 4 and 9 (IL4, IL9), of the nuclear factor KB (NFKB) and of the CCL4 chemokine: these are cytokines involved in the mediation of the immune response. Relations between behaviour and immunity have been suggested also in other works. Biscarini et al. (2010) detected associations between the serotonin receptors HTR2C and HTR2A and, respectively, complement activity and antibody titres. Combining their results with the ones of the present study, we see that in the case of HTR2C higher feather damage corresponds to lower complement activity. Buitenhuis et al. (2006) found higher IgG titres and lower leukocyte concentration, CD4+ lymphocytes percentage and MHC I expression in high FP compared to low FP lines. Parmentier et al (2009) also found a relationship between immunity and feather pecking: birds challenged with the antigen human serum albumin (HuSA) at young age were more likely to develop feather damage at a later age, compared to unchallenged birds. This points to complex and interesting links between behaviour and immune system.

From the analysis of associative effects, some SNPs on the sex chromosome Z have been associated to feather condition score on the back-rump and belly regions. These findings on the sex chromosome may relate to previous observations that feather pecking is affected by gonadal hormones and is more common in females than in males (Hughes 1973; Jensen et al. 2005).
When looking at gene effects, we consistently observed a higher frequency of the alleles linked to more feather damage in hens of White Leghorn origin as compared to Rhode Island Red origin (results not shown). For SNP rs15385785 on chromosome 1 (MAOA), for instance, the frequency of the allele associated with greater feather damage was 0.79 in brown layers and 0.94 in white layers. For SNP rs13640917 on chromosome 4 (HTR2C), it was 0.35 in brown layers and 0.84 in white layers. This is consistent with reports of more feather damage on the back and rump of white layers in comparison with brown layers (Uitdehaag et al. 2008), which indicates higher incidence of feather pecking behaviour in White Leghorns. Differences in fearfulness and in the metabolism of the neurotransmitters serotonin and dopamine between white and brown layers have also been observed (Uitdehaag, personal communication).

Social interactions

Feather pecking is a trait in which social interactions between group mates play an important role (Ellen et al. 2008). In our study we looked separately at the direct genetic effects (receiving FP) and at the associative effects (performing FP). The existence of social interaction raises the question of group composition. Groups of closely related animals (e.g. full sibs) on one hand tend to have less negative social interactions, but on the other hand pose statistical challenges. In laying hens, for

instance, feather pecking and cannibalism are reduced in cages of full-sibs (Bijma et al. 2007a, b), but from such data associative effects can not be estimated (Bijma et al. 2007a, b) and variance components might be more difficult to estimate (Biscarini et al. 2008). Depending on the case, groups of animals with similar or different personality traits might perform better (Rodenburg et al. in press). We took the SNPs for the MAOA and HTR2C and looked at the unadjusted feather condition scores. We saw that from a mere phenotypic point of view having more heterozygous hens in a cage leads to increased feather damage. In both cases one allele is associated to increased feather damage and the other to lower feather damage. However, the effect of the number of heterozygous hens per cage was not significant when added to the statistical model. Probably the relation between feather damage and cage composition is too weak to be revealed by this dataset size and experimental design. Higher feather damage can reflect either propensity to peck or docility. Hens that tend to peck can in fact be involved in numerous fights damaging their own plumage. Other hens might be so docile that they do not respond to the pecking insult and might be preferred target for peckers. So genes linked to high or low feather damage may reflect both types of situation. Homozygous hens for the mentioned SNPs might therefore be either active peckers or docile animals. Heterozygotes will be somewhere in the middle. When only docile hens are in a cage, not much is likely to happen. When only active peckers are in a cage, they might be afraid of each other and refrain to fight. Intermediate hens might on the contrary give rise to fights resulting in higher feather damage in those cages.

The inclusion of the associative effects in the model, for traits influenced by social interactions such as FP in laying hens, can lead to a considerable increase in the genetic response to selection, thanks to the additional heritable genetic variation of the associative effects (Ellen et al., 2007).

Bijma et al. (2007b) showed that non-genetic covariance among group mates can bias the estimates of genetic associative effects. Although we modelled social interactions differently, we fitted a random group effect to asses the magnitude of such non-genetic covariance in our analysis (see Bergsma et al. 2008). The significance of the estimates of the SNPs effects from models [3] and [4] decreased only fractionally and did not affect the presented results at all.

Serotonin

Substantial scientific evidence of the role of the serotonergic system in the development and modulation of feather pecking behaviour in laying hens has accrued over the last years (van Hierden et al., 2002; van Hierden et al., 2004a, b; Bolhuis et al., 2009). These studies related the occurrence of feather pecking with serotonin concentration and activity either in the brain or peripherally (circulatory system or peripheral nervous system), predominantly suggesting that lower levels of

serotonin are associated with predisposition to perform feather pecking (van Hierden et al., 2004a, b; Bolhuis et al., 2009). We detected an association between the gene for the serotonin receptor HTR2C and feather condition score in the back-rump region both in the direct (receiving FP) and associative (performing FP) analyses. Flisikowski et al. (2008) associated the gene for a regulatory factor of the serotonergic system (DEAF1) on chromosome 5 with feather pecking behaviour. The same authors postulated that finding the same association in populations of Rhode Island Red and White Leghorn origin indicates that the origin of the allele predisposing to feather pecking predates the breeding activity of at least the last 50 years. They suggest that detecting the same genetic markers in different populations implies that they are close to the functional mutation, which agrees with the theory of across-line association studies (Biscarini et al. 2010). Therefore feather pecking in laying hens can seemingly be controlled by modulating their serotonergic system, by means of genetic selection or husbandry practices, either pharmacologically or dietary (van Hierden et al. 2004a). Interestingly, the recent study by Bolhuis et al. (2009) shows that genetic selection for low mortality, using the social models, leads to changes in the serotonergic system, already in the second generation of selection.

The results of this work contribute to a better understanding of the genetic background of feather pecking behaviour in laying hens. The analysis of both direct and associative genetic effects confirmed that social interactions play an important role in the emergence of feather pecking, and is therefore a valuable tool for the investigation of this behavioural characteristic of birds. To our knowledge this was the first time that the associative effect was addressed in an association study in laying hens. The gene for the serotonin receptor (HTR2C) was found to be associated with feather damage, which adds to existing evidence of the role of the serotonergic system in the modulation of feather pecking. The involvement of the genes for interleukins (IL4, IL9) and chemokines (CCL4) points at fascinating relationships between behaviour and immunity.

ACKNOWLEDGEMENTS

We would like to thank ISA B.V., the layer breeding division of Hendrix Genetics, for providing animals and facilities. We would also like to thank the staff at the ISA test facility 'Stevensbeek' for taking excellent care of the laying hens. Further, we would like to thank Esther Ellen, Laura Star, Koen Uitdehaag and Patrick Wissink for conduction the plumage condition scoring.

This work was conducted as part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission.

REFERENCES

Andreescu C, Avendano S, Brown SR, Hassen A, Lamont SJ, Dekkers JCM (2007) Linkage disequilibrium in related breeding lines of chicken. Genetics 177:2161-2169

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57:289-300

Bergsma R, Kanis E, Knol EF, Bijma P (2008) The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics 178:1559-1570

Bilčik B, Keeling LJ (1999) Changes in feather condition in relation to feather pecking and aggressive behaviour in laying hens. Br Poult Sci 40:444-451

Bijma P, Muir WM, Van Arendonk JAM (2007a) Multilevel Selection 1: Quantitative Genetics of Inheritance and Response to Selection. Genetics 175:277–288

Bijma P, Muir WM, Ellen ED, Wolf JB, Van Arendonk JAM (2007b) Multilevel Selection 2: Estimating the Genetic Parameters Determining Inheritance and Response to Selection. Genetics 175:289–299

Biscarini F, Bovenhuis H, van Arendonk JAM (2008) Estimation of variance components and prediction of breeding values using pooled data. J Anim Sci 86:2845-2852.

Biscarini F, Bovenhuis H, van Arendonk JAM, Parmentier HK, Jungerius AP, van der Poel JJ (2010) Across-line SNP association study of innate and adaptive immune response in laying hens. Anim Gen 41:26-38

Blokhuis HJ, Arkes JG (1984) Some observations on the development of feather-pecking in poultry. Appl Anim Behav Sci 12:145-157

Blokhuis HJ (1986) Feather-pecking in poultry: its relation with ground-pecking. Appl Anim Behav Sci 16:63-67.

Bolhuis JE, Ellen ED, Van Reenen CG, De Groot J, Ten Napel J, Koopmanschap RE, De Vries Reilingh G, Uitdehaag KA, Kemp B, Rodenburg TB (2009) Effects of genetic group selection against mortality on behaviour and peripheral serotonin in domestic laying hens with trimmed and intact beaks. Physiol Behav 97:470-475

Buitenhuis AJ, Rodenburg TB, van Hierden YM, Siwek M, Cornelissen SJB, Nieuwland MGB, Crooijmans RPMA, Groenen MAM, Koene P, Korte SM, Bovenhuis H, van der Poel JJ (2003a) Mapping quantitative trait loci affecting feather pecking behavior and stress response in laying hens. Poult Sci 82:1215-1222

Buitenhuis AJ, Rodenburg TB, Siwek M, Cornelissen SJB, Nieuwland MGB, Crooijmans RPMA, Groenen MAM, Koene P, Bovenhuis H, van der Poel JJ (2003b) Identification of quantitative trait loci for receiving pecks in young and adult laying hens. Poult Sci 82:1661-1667

Buitenhuis AJ, Rodenburg TB, Siwek M, Cornelissen SJB, Nieuwland MGB, Crooijmans RPMA, Groenen MAM, Koene P, Bovenhuis H, van der Poel JJ (2004) Identification of QTLs involved in open-field behaviour in young and adult laying hens. Behav Genet 34:325-333

Buitenhuis AJ, Kjaer JB, Labouriau R, Juul-Madsen HR (2006) Altered circulating levels of serotonin and immunological changes in laying hens divergently selected for feather pecking behaviour. Poult Sci 85:1722-1728

Ellen ED, Muir WM, Teuscher F, Bijma P (2007) Genetic improvement of traits affected by interactions among individuals: sib selection schemes. Genetics 176:489-499

Ellen ED, Visscher J, van Arendonk JAM, Bijma P (2008) Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines. Poult Sci 87:233-239

Flisikowski K, Schwarzenbacher H, Wysocki M, Weigend S, Preisinger R, Kjaer JB, Fries R (2008) Variation in neighbouring genes of the dopaminergic and serotonergic systems affects feather pecking behaviour of laying hens. Anim Gen 40:192-199

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK

Hassen A, Avendano S, Hill WG, Fernando RL, Lamont SJ, Dekkers JCM (2009) The effect of heritability estimates on high-density single nucleotide polymorphism analyses with related animals. J Anim Sci 87:868-875.

Hillel J, Groenen MAM, Tixier-Boichard M. et al. (2003) Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genet Sel Evol 35:533-557

Huber-Eicher B, Wechsler B (1997) Feather pecking domestic chicks: Its relation to dustbathing and foraging. Animal Behaviour 54, 757-768

Hughes BO, Duncan IJH (1972) The influence of strain and environmental factors upon feather pecking and cannibalism in fowls. Br Poult Sci 13:525-547

Hughes BO (1973) The effect of implanted gonadal hormones on feather pecking and cannibalism in pullets. Br Poult Sci 14:341-348

Jendral MJ, Robinson FE (2004) Beak trimming in chickens: historical, economical, physiological and welfare implications, and alternatives for preventing feather pecking and cannibalistic activity. Avian Poult Biol Rev 15:9-23

Jensen P, Keeling L, Schutz K, Andersson L, Mormede P, Brandstrom H, Forkman B, Kerje S, Fredriksson R, Ohlsson C, Larsson S, Mallmin H, Kindmark A (2005) Feather pecking in chickens is genetically related to behavioural and developmental traits. Physiol Behav 86:52-60

Jones RB, Blokhuis HJ, Beuving G (1995) Open-field and tonic immobility responses in domestic chicks of two genetic lines differing in their propensity to feather peck. Br Poult Sci 36:525-530

Kjaer J, Sørensen P (1997) Feather pecking behaviour in White Leghorns, a genetic study. Br Poult Sci 38:335-343

Kjaer JB, Sørensen P, Su G (2001). Divergent selection on feather pecking behaviour in laying hens (*Gallus gallus domesticus*). Appl Anim Beahv Sci. 71:229-239

Kjaer JB (2009) Feather pecking in domestic fowl is genetically related to locomotor activity levels: implications for a hyperactivity disorder model of feather pecking. Behav Genet 39:564-570

Newberry RC, Keeling LJ, Estevez I, Bilcik B (2007) Behaviour when young as a predictor of severe feather pecking in adult laying hens: the redirected foraging hypothesis revisited. Appl Anim Behav Sci 107:262-274

Parmentier HK, Rodenburg TB, De Vries Reilingh G, Beerda B, Kemp B (2009) Does enhancement of specific immune responses predispose laying hens for feather pecking? Poult Sci 88:536-542

Pigott TA (1996) OCD: Where the serotonin selectivity story begins. J. Clin Psychiatry 57:11-20

Rodenburg TB, Buitenhuis AJ, Ask B, Uitdehaag KA, Koene P, van der Poel JJ, Bovenhuis H (2003). Heritability of feather pecking and open-field response of laying hens at two different ages. Poult Sci 82:861-867

Rodenburg TB, Buitenhuis AJ, Ask B, Uitdehaag KA, Koene P, van der Poel JJ, van Arendonk JAM, Bovenhuis H (2004) Genetic and phenotypic correlations between feather pecking and open-field response in laying hens at two different ages. Behav Genet 34:407-415

Rodenburg TB, Bijma P, Ellen ED, Bergsma R, de Vries S, Bolhuis JE, Kemp B, Van Arendonk JAM (2010) Breeding amiable animals? Improving farm animal welfare by including social effects in breeding programs. Animal Welfare 19(S):77-82.

Saccone NL, Saccone SF, Goate AM, Gruzca RA, Hinrichs AL, Rice JP, Bierut LJ (2008) In search of causal variants: refining disease association signals using cross-population contrasts. BMC Genetics 9:58

Savory CJ (1995) Feather pecking and cannibalism. Worlds Poult Sci J 51:215-219

Schütz KE, Kerje S, Jacobsson L, Forkman B, Carlborg Ö, Andersson L, Jensen P (2004) Major growth QTLs in fowl are related to fearful behavior: possible genetic links between fear response and production traits in a Red Junglefowl x White Leghorn intercross. Behav Genet 34:121-130

Uitdehaag KA, Komen H, Rodenburg TB, Kemp B, van Arendonk JAM (2008) The novel object test as predictor of feather damage in cage-housed Rhode Island Red and White Leghorn laying hens. Appl Anim Behav Sci 109:292-305

Van Hierden YM, Korte SM, Ruesink EW, van Reenen CG, Engel B, Korte-Bouws GAH, Koolhaas JM, Blokhuis HJ (2002) Adrenocortical reactivity and central serotonin and dopamine turnover in young chick from a high and low feather pecking line of laying hens. Physiol Behav 75:653-659

Van Hierden YM, Koolhaas JM, Korte SM (2004a) Chronic increase of dietary L-tryptophan decreases gentle feather pecking behavior. App Anim Behav Sci 89:71-84

Van Hierden YM, de Boer SF, Koolhaas JM, Korte SM (2004b) The control of feather pecking by serotonin. Behav Neurosci 118:575-583

Vestergaard KS, Lisborg L (1993) A model for feather pecking development which relates to dustbathing in the fowl. Behaviour 126:291-308

Table 1: number of hens per line, average feather scores for the single lines, and mean and standard deviations of the traits in the overall population[1].

				Individual Feather Scores			
breed	**line**	**n**		**BR51**	**Belly51**	**BR69**	**Belly69**
Rhode Island Red	B1	81		2.28	0.34	4.42	2.09
	B2	76		1.09	0.04	3.88	1.68
	B3	75		2.79	0.45	4.83	2.23
	BB	66		2.02	0.88	4.70	2.53
White Leghorn	W1	68		5.53	2.09	6.63	3.24
	WA	77		4.93	1.56	5.60	3.16
	WB	77		4.18	1.41	4.61	2.96
	WC	63		5.86	2.67	6.65	3.51
	WF	79		2.23	1.87	3.92	2.03
			n	662	662	655	655
Total		662	mean	3.37	1.22	4.96	2.57
			sd	2.49	1.35	1.93	1.13

[1]BR51, BR69 = sum of the individual feather scores for the back and rump regions at 51 and 69 wk of age (scale 0-10); Belly51, Belly69 = individual feather scores for the belly region at 51 and 69 wk of age (scale 0-5).

Table 2: number of SNPs per chromosome along the chicken genome, number of SNPs deviating from Hardy-Weinberg equilibrium, and homozygosity in the nine genotyped lines of layers[1].

chr	Size (Mbp)	SNPs		FL (across lines)	Within-line values																	
		All	Used		B1		B2		B3		BB		W1		WA		WB		WC		WF	
					HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL	HW	FL
1	201	68	61	6	4	19	1	17	1	12	3	16	0	24	8	21	0	20	0	15	1	23
2	155	24	22	2	1	6	0	7	1	3	1	5	0	10	2	8	1	10	1	9	0	9
3	114	140	121	12	3	31	0	35	2	28	7	22	0	41	9	40	0	34	0	27	0	37
4	94	421	371	65	9	102	1	126	6	120	13	105	2	141	24	142	11	117	8	126	16	126
5	62	285	265	25	9	48	6	56	7	46	22	41	3	78	23	58	4	64	20	62	8	76
6	37	27	22	1	2	3	2	3	0	3	3	2	0	4	3	2	2	4	1	3	1	3
7	38	175	149	18	5	33	1	35	5	30	9	35	3	48	14	42	5	43	10	39	8	43
8	31	6	6	0	2	1	1	0	0	1	2	0	0	3	2	2	2	2	0	0	0	0
9	26	12	10	0	0	2	0	2	0	2	0	5	1	4	0	1	0	3	0	3	0	6
10	22.6	4	3	0	0	1	0	1	2	1	0	2	1	0	0	0	1	0	0	1	0	1
11	21.9	8	6	3	0	3	0	3	0	4	0	4	0	4	1	4	1	4	0	4	0	4
12	20.5	7	7	2	0	4	0	3	0	4	0	3	1	3	0	5	0	4	2	3	0	3
13	18.9	37	32	3	2	6	0	9	0	9	3	5	0	9	9	7	1	8	4	7	3	7
14	15.8	7	6	0	0	0	0	1	0	0	0	1	1	1	0	0	2	0	0	1	0	1
15	13	7	5	1	0	2	0	2	0	2	0	1	0	4	0	4	1	2	1	2	0	1
16	0.43	17	12	3	0	3	2	3	3	3	0	3	1	4	1	4	2	4	1	3	0	4
17	11.2	7	7	0	0	1	0	1	0	1	1	1	0	1	2	1	0	0	1	1	0	1
19	9.9	27	20	4	2	6	2	6	2	5	2	6	1	9	2	6	1	7	2	8	0	6
21	7	4	4	1	0	1	0	1	0	2	0	1	0	2	1	3	2	2	1	2	0	2
22	3.9	3	2	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0
23	6	4	2	1	0	1	0	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1
24	6.4	16	15	3	0	6	0	4	0	4	1	5	2	3	2	5	1	5	0	4	1	4
26	5.1	60	50	4	6	12	2	12	2	12	3	11	2	16	8	15	1	12	4	17	0	11
Z	75	160	138	21	0	47	0	49	0	47	0	50	0	76	0	61	0	60	0	63	0	62
-		8	6	0	1	2	1	2	0	2	1	2	1	2	0	1	1	2	1	1	0	1
Total		1534	1342	175	46	340	19	379	31	342	71	327	20	488	113	433	40	408	57	402	39	432

[1]HW = loci not in Hardy-Weinberg equilibrium, FL = fixed loci. The symbol "-" refers to SNPs that were not assigned to any chromosome. Size refers to the size of the whole chromosome as derived from the NCBI chicken genome databass

Table 3: SNPs significantly associated with feather scores (FS) in the analysis of direct genetic effects. $-\log_{10}$ of the *P*-values are reported in the columns, with corresponding FDR in brackets.[1]

				Individual FS							
chr	**SNP**	**kbps**	**cM**[2]	**BR51**		**Belly51**		**BR69**		**Belly69**	
chr 1	rs15385785 (*MAOA*)[3]	114912046	287.3	3.15	(0.10)			2.76	(0.12)		
chr 3	rs13773912	5595342	14.0					3.20	(0.07)		
chr 4	rs13640917 (*HTR2C*)[3]	2798627	7.0	3.25	(0.12)						
	rs13788969	20659524	51.6	3.56	(0.10)						
	rs13517693	43865427	109.7			3.07	(0.17)				
	rs13522023	54013816	135.0			4.12	(0.08)				
chr 5	rs15692150	30998335	77.5					2.10	(0.10)		
	rs15707740	42231488	105.6					3.23	(0.10)		
chr 6	rs16558389	unmapped						3.25	(0.10)		
chr 13	rs14999300 (*IL9*)[3]	15574216	38.9	3.43	(0.10)						

[1]BR51, BR69 = sum of the individual feather scores for the back and rump regions at 51 and 69 wk of age (scale 0-10); Belly51, Belly69 = individual feather scores for the belly region at 51 and 69 wk of age (scale 0-5)
[2]Assuming 1 cM = 4 x 10^5 bps
[3]*MAOA* = gene for the mono-amino oxidase A; *HTR2C* = gene for the serotonin receptor 2C; *IL9* = gene for the interleukin 9.

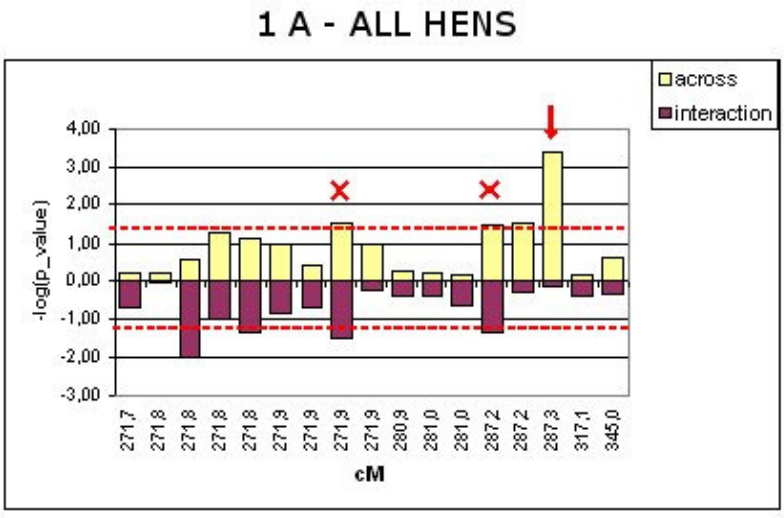
Table 4: significance of SNP allele frequency of cage mates on the feather score of individual hens. $-\log_{10}$ of the *P*-values are reported in the columns, with corresponding FDR in brackets.[1]

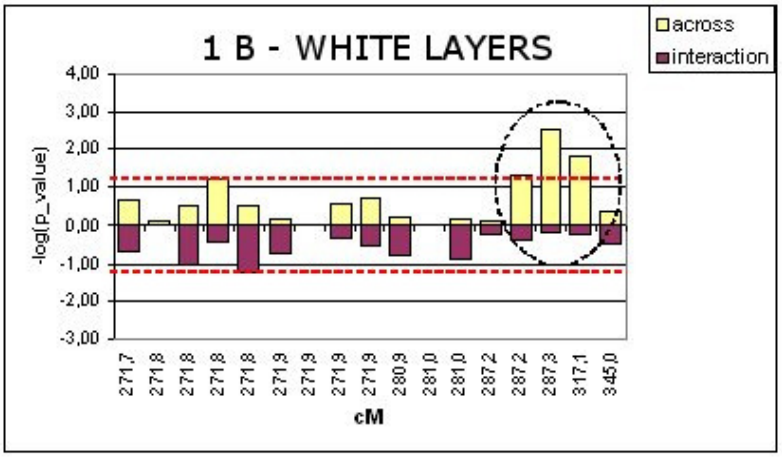
				Social effects							
chr	**snp**	**map**	**cM**[2]	**BR51s**		**Belly51s**		**BR69s**		**Belly69s**	
chr Z	rs16101283	20655488	51.6	3.15	(0.06)						
	rs16101484	20904552	52.3	2.93	(0.07)						
	rs16105159	23468761	58.7	2.20	(0.10)	2.91	(0.06)			1.90	(0.07)
	rs16104871 (*GFM2*)[3]	23690970	59.2			2.35	(0.10)				
	rs16106976	33770169	84.4					1.67	(0.10)		
	rs13762897 (*TRPM3*)[3]	34827204	87.1	2.57	(0.11)						
chr 1	rs14810117	36931526	92.3							2.63	(0.06)
chr 3	rs13503459	9757137	24.4	1.85	(0.12)						
	rs13503401	9873492	24.7	2.94	(0.09)						
	rs13503220	10297909	25.7	2.28	(0.12)			1.95	(0.11)		
	rs13717237	18943533	47.4	2.89	(0.06)						
	rs13717379	19142921	47.9			2.92	(0.10)				
	rs13717382	19146715	47.9			2.74	(0.14)				
	rs13717441	19232318	48.1			2.51	(0.14)				
	rs13717447	19243861	48.1			3.13	(0.06)				
	rs13717645 (*PFN3*)[3]	19755775	49.4	3.05	(0.07)			2.05	(0.14)		
	rs13717686	19841213	49.6	2.40	(0.07)						
	rs13717773	20014822	50.0					1.79	(0.13)		
	rs13717778	20015177	50.0	6.70	(0.00)			3.25	(0.06)		
	rs13717881	20357976	50.9					1.61	(0.13)		
chr 4	rs13640917 (*HTR2C*)[3]	2798627	7.0	3.39	(0.03)			2.30	(0.01)		
	rs13512983	30896422	77.2					2.74	(0.11)		
	rs13514279	33836739	84.6					2.34	(0.08)		
	rs13515243 (PPP2CB)[3]	35781204	89.5	2.70	(0.10)						
	rs13517937 (GALNT7)[3]	44712609	111.8	2.69	(0.13)						
	rs13521963	53833480	134.6			2.51	(0.06)				
	rs13522188	54500061	136.3			2.90	(0.06)				
	rs13522598 (TRPC3)[3]	55438724	138.6			2.40	(0.08)			1.96	(0.07)
	rs13522688 (TRPC3)[3]	55714579	139.3	2.21	(0.08)						
	rs13523367	57888982	144.7	2.28	(0.10)			1.48	(0.14)		
	rs16422070	62651897	156.6			2.03	(0.12)				
chr 5	rs15661619	11950758	29.9			2.09	(0.08)				
	rs13758305	16664011	41.7					1.43	(0.03)		
	rs13794185	23057797	57.6	2.44	(0.12)	3.16	(0.06)				
	rs15681243 (CKAP5)[3]	25554863	63.9			2.48	(0.09)				
	rs13756469 (RGS6)[3]	28876011	72.2	4.27	(0.01)						
	rs13585105	32180230	80.5			3.39	(0.06)				
	rs13585316	33526797	83.8			2.62	(0.10)				
	rs13585357	33758339	84.4	3.82	(0.03)			2.08	(0.10)		
	rs13585704 (NPAS3)[3]	37788082	94.5	2.72	(0.11)						
	rs13585761 (EGLN3)[3]	38086335	95.2	2.66	(0.14)						
	rs13586409	39786663	99.5			2.37	(0.12)				

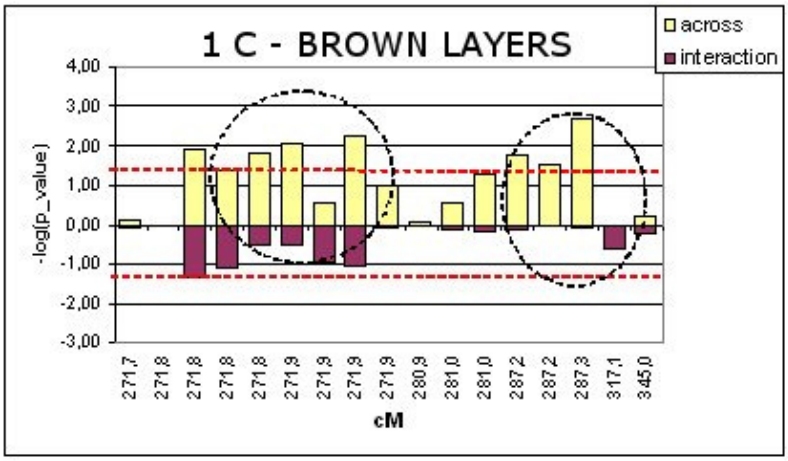
	rs13587250	42038263	105.1					1.90	(0.14)		
	rs15707740	42231488	105.6					4.67	(0.01)		
chr 6	rs13562501	5575686	13.9	2.18	(0.10)						
	rs16548180 (NFKB2)[3]	18021166	45.1	2.95	(0.06)			2.14	(0.03)		
chr 7	rs13781704	23592646	59.0							1.67	(0.06)
	rs13596168	25491840	63.7							1.92	(0.13)
	rs13598049 (SNX4)[3]	29698020	74.2							1.76	(0.08)
	rs13598125 (PARP14)[3]	29835983	74.6	1.56	(0.11)						
	rs13598160	29915809	74.8	2.52	(0.06)						
	rs13601268	37224680	93.1			2.00	(0.14)				
chr 13	rs14999300 (IL9)[3]	15574216	38.9	2.27	(0.07)						
	rs15709659 (IL4)[3]	17534793	43.8			3.16	(0.06)				
chr 19	rs13573020 (CCL4)[3]	376195	0.9							1.30	(0.08)
	rs14119838 (HSPB1)[3]	4216458	10.5	3.16	(0.10)						
chr 24	rs15209193 (TIRAP)[3]	430067	1.1			1.82	(0.14)				

[1]BR51, BR69 = sum of the individual feather scores for the back and rump regions at 51 and 69 wk of age (scale 0-10); Belly51, Belly69 = individual feather scores for the belly region at 51 and 69 wk of age (scale 0-5)
[2]1 cM = 4 x 10^5 bps
[3]*GFM2* = gene encoding elongation factor G2; *TRPM3* = gene encoding transient receptor potential cation channel, subfamily M, member 3; *PFN3* = gene encoding profilin; *HTR2C* = gene encoding serotonin receptor 2C; *PPP2CB* = gene encoding protein phosphatase 2; *GALNT7* = gene encoding GalNAc transferase 7; *TRPC3* = gene encoding ransient receptor potential cation channel, subfamily C, member 3; *CKAP5* = gene encoding cytoskeleton associated protein 5; *RGS6* = gene encoding regulator of G-protein signaling 6; *NPAS3* = gene encoding Neuronal PAS domain protein 3; *EGLN3* = gene encoding Egl nine homolog 3; *NFKB2* = gene encoding Nuclear factor NF-kappa-B p100 subunit; *SNX4* = gene encoding sorting nexin-4; *PARP14* = gene encoding poly (ADP-ribose) polymerase family, member 14; *IL9* = gene encoding interleukin 9; *IL4* = gene encoding interleukin 4; *CCL4* = gene encoding chemokine CCL4; *HSPB1* = gene encoding heat shock 27kDa protein 1; *TIRAP* = gene encoding toll-interleukin 1 receptor.

Fig 1: significance of SNP (-log(p value)) and of the SNP-by-line interaction (log(p value)) for the genomic region surrounding SNP rs15385785 on chromosome 1. The combined analysis for all lines and the separate analyses for five white and four brown layers lines are presented. The dashed lines are the threshold of significance (-log(0.01)≈1.3).




1 B - WHITE LAYERS
across
interaction
4,00
3,00
2,00
1,00
0,00
-1,00
-2,00
-3,00
-log(p_value)
271,7
271,8
271,8
271,8
271,8
271,9
271,9
271,9
271,9
280,9
281,0
281,0
287,2
287,2
287,3
317,1
345,0
cM


1 C - BROWN LAYERS
across
interaction
4,00
3,00
2,00
1,00
0,00
-1,00
-2,00
-3,00
-log(p_value)
271,7
271,8
271,8
271,8
271,8
271,9
271,9
271,9
271,9
280,9
281,0
281,0
287,2
287,2
287,3
317,1
345,0
cM

Chapter 6

General discussion

Introduction

During the last decades many theoretical and technological advancements have occurred in the field of animal breeding and genetics. The application of more sophisticated statistical models and increased computer power have led to more accurate estimates of genetic parameters and breeding values. Genetic theory has been extended allowing for more complete genetic models, such as with the inclusion of the effect of social interactions when animals are kept in groups (Bijma et al., 2007a,b), as is the case of laying hens. The availability of the complete chicken genome sequence and the reduction in genotyping costs opened new opportunities to identify genes that contribute to genetic variation in traits of zootechnical interest, and to estimate genomic assisted breeding values. As a result, estimation of breeding values no longer relies exclusively on observations of phenotypes and pedigree. The use of genomic information can substantially increase the response to selection, for instance –at least in some species- by reducing the generation interval, given that genotypes, unlike phenotypes, can be recorded much earlier in life and in both sexes. Genomic information might be especially useful for those traits that are hard or expensive to measure, or that have low heritabilities, like immune parameters and behavioural characteristics, i.e. traits that are more difficult to improve with the standard methods (Goddard and Hayes, 2009). In this thesis, the use of statistical models that exploit information at group rather than individual level, and the application of genetic markers to genetic evaluations of laying hens were addressed. These topics constitute at the same time a challenge and an opportunity for the layer breeding industry. The potential and limitations of the use of pooled data and genetic markers in layer breeding will be discussed hereafter.

Use of pooled data in genetic evaluations

Chapters 2 and 3 of this thesis deal with the use of pooled data in genetic evaluations. The research in this field is still at an initial stage, and many related issues remain to be resolved. However, based on the findings reported in this thesis, it is possible to discuss the potential applications and limitations of using pooled data in the genetic evaluation of laying hens and other livestock species. Pooled data are in principle of interest for all livestock species that in farming practice are housed in groups, like poultry (both layers and broilers, but also other fowl species), pigs, rabbits and

aquaculture species, and where performance is recorded at group level. To assess the value of using pooled data, it is important also to look at which traits are recorded for the whole group and could not be easily measured individually. Laying hens provide a good example, since the number of eggs produced can be recorded easily enough per group of hens housed in the same cage. It would be more time consuming and technically challenging -and therefore expensive- to record egg production for every individual hen when they are kept in groups. On the other hand, body weight in layers can be always measured individually, although broilers are kept in groups. In aquaculture, body weight can be measured on a sample of the fishes in a tank, and the average weight of the sample can be used as pooled record for the entire group (Simianer and Gjerde, 1991). Feed consumption of a group of pigs provides another example where collection of pooled data is easier than the collection of individual data. Feed intake is an important trait in animal breeding which is difficult to measure in species housed in groups (like pigs, fishes, poultry): in such cases pooled data might be used in the genetic evaluation for feed intake.

Subsequently, attention should be paid to the size and composition of the group on which traits are recorded. As groups get larger, the accuracy of the estimated breeding values of individuals decreases (Olson et al., 2006; Biscarini et al., 2008). Based on these results, it is also likely that genetic parameters are estimated with lower accuracy when data from larger groups are used. In Chapters 2 and 3 I showed that for groups consisting of 4 to 5 hens breeding values can be estimated rather accurately. Samples of fishes and litters of piglets are examples of intermediate group size for which the estimation of breeding values based on pooled data is still feasible, though with progressively lower accuracies. Using simulations, it has been shown in Chapter 2 that the mean accuracy of EBVs from pooled records was 0.684 with cages of 2 hens, 0.569 with cages of 3 hens, 0.552 with cages of 4 hens, 0.405 with cages of 5 hens, and 0.330 with cages of 8 hens. This decrease in accuracy follows approximately a logarithmic pattern, and it can be extrapolated that for groups of 20 animals the accuracy of EBVs would be below 0.15. This implies that for very large groups – e.g. feed intake recorded at farm level for dairy cows - the estimation of breeding values from pooled data becomes very inaccurate and virtually impossible. Group composition is also relevant for the accuracy. For groups of closely related animals it is more difficult to dissect the pooled observation into genetic and non-genetic components through the pedigree: when all full sibs are in the same group, for

instance, they all obtain the same EBV (Biscarini et al., 2008). In order to effectively use pooled data in genetic evaluations it is therefore better to have groups of half-sibs or randomly mixed animals. This can be a limitation in pigs (unless cross-fostering is practiced), where litters consisting of full-sibs are housed together, whereas in laying hens groups are often mixed.

Current software packages have so far permitted the estimation of heritabilities and breeding values only from groups of equal size. There is no theoretical impediment to the use of pooled data from groups of different size. In the case of unequal group size, the matrix of the residual variances R can not be factored out of the mixed model equations and should, therefore, be used in the estimation process. Computer programmes for estimating genetic parameters and breeding values, though, have been designed for the elaboration of individual records: it is not trivial to use them with pooled observations where the number of individuals in a group varies. The extension of the method for estimating breeding values from pooled data to situations with unequal group size is relatively easy. For this, in fact, only the inversion of the MME is required: provided that the correct R matrix with the dimension of each group along its diagonal is supplied, this would give the correct solutions. The estimation of genetic parameters is a computationally more demanding task and the corresponding software modification is less straightforward. It would be very helpful if the analysis of pooled data could be extended to groups of any size. In pigs, for instance, the number of piglets in a litter is usually variable. Also in laying hens, due to mortality, cages might host different numbers of hens. In general, groups of unequal size occur in many species of agricultural interest: poultry, pigs, rabbits, aquaculture species. One solution to temporarily work around this limitation in the estimation of genetic parameters is to use only groups of the same size in the analysis. The estimated parameters can then be used in the estimation of breeding values for the entire population, irrespective of group size. Attention has nonetheless to be paid to checking that selection of groups of equal size does not introduce a bias due to non-random sampling from the population, i.e. the problem that parameters do not refer to the general population. This could, for example, be the case if variation in group size would result from variation in mortality.

When only pooled observations per group are available, the environmental effect of the group itself can't usually be estimated. For example, in the case of laying hens the cage as fixed effect can not be estimated. This can be overcome by modeling the

environmental effect of a cage as a random effect, and a spatial (co)variance structure between cages that depends on the position of cages relative to each other. Under this model, it is feasible to estimate the environmental variance due to cages and the correlation between cages at different positions. In this way, the environmental effect of cages can be included in the genetic analysis. In general, depending on the trait and the group size, the environmental group effect might be more or less relevant. For small groups kept in homogeneous and controlled environments, as is the case for laying hens in battery cages, the bias due to ignoring the environmental group effect in the statistical model is expected to be negligible. In other situations, where groups are larger and environmental conditions more heterogeneous, not including the effect of the group environment in the model is expected to be a serious drawback which would lead to unreliable estimates of genetic parameters and biased breeding values.

Type of pooled data

Pooled data are the overall performance of groups of individuals (Olson et al., 2006; Biscarini et al., 2008). These group performances might be mainly of two types. The first is the total performance of the group, represented by the sum of the performances of the members of the group. For instance, all the eggs produced by the hens in a cage are added together to form the pooled record for the cage. The second type of pooled data is the average performance of the group, consisting of the mean performance of the members of the group. An example can be the mean egg production of hens in a cage. In some cases, however, pooled data are neither the sum nor the average of the individual performances of a group, but simply the performance of the group. This is the case, for example, for the response to the novel object test (Uitdehaag et al., 2008). In the novel object test a score is assigned to the group based on the reaction of the entire group to the presence of a new and unknown object.

Since group sums and group averages are merely a mathematical transformation of one into another, there is no fundamental difference in using either type of pooled data. They should yield the same results, provided that the necessary operational adjustments are adopted to take into account the difference in the mathematical nature of sums and means. The following example with three groups of two individuals each will illustrate this for a model where the general mean is the only systematic environmental effect.

In the case of group sums, the vector **y** contains the sums of the performances of the individuals in each group. The design matrices **X** and **Z** reflect the group composition and vector **e** contains the sums of the residuals (see chapter 2 of this thesis for more details).

$$\mathbf{y^*} = \begin{bmatrix} \mathbf{15+14} \\ \mathbf{9+11} \\ \mathbf{10+9} \end{bmatrix} = \begin{bmatrix} \mathbf{29} \\ \mathbf{20} \\ \mathbf{19} \end{bmatrix}, \qquad \mathbf{X^*} = \begin{bmatrix} \mathbf{2} \\ \mathbf{2} \\ \mathbf{2} \end{bmatrix}, \qquad \mathbf{Z^*} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{bmatrix},$$

$$\mathbf{e^*} = \begin{bmatrix} \mathbf{e_1 + e_2} \\ \mathbf{e_3 + e_4} \\ \mathbf{e_5 + e_6} \end{bmatrix}$$

The vectors of solutions for fixed and genetic effects, **b** and **a**, are the same as in the usual mixed model equations (Henderson, 1975, 1984), since the objective is to estimate the same number of parameters with a different source of information (pooled instead of individual records).

The use of sums affects also the variance structure of the model. The genetic variance-covariance matrix remains unchanged, being $\mathbf{G} = \mathbf{A}\sigma_a^2$. However, due to the different composition of vector **e***, the residual variance-covariance matrix changes: $Var(e^*) = \mathbf{R^*}=\mathbf{D}\sigma^2_{e^*}$. Assuming that there are no residual correlations between animals in different groups, **D** is not an identity matrix but a diagonal matrix with diagonal element n_j representing the number of animals that contributed to the j^{th} pooled observations (group size). In the case of groups with the same size, $\mathbf{D} = n \times \mathbf{I}$, where n is the group size. The variance ratio α is, therefore, also affected and, assuming no covariance between animals in the same group, becomes n times that of individual observations:

$$\alpha^* = \sigma^2 e^* \big/ \sigma^2 a = n \cdot \sigma^2 e \big/ \sigma^2 a = n \times \alpha,$$

This relation, however, holds only when groups are of equal size. When groups are of different sizes, $\mathbf{R^{*-1}}$ must be used in the mixed model equations. Notice that

$\sigma_e^2* = \text{var}(e_1 + ... + e_n)$ is only equal to $n \cdot \sigma_e^2$ if there is no covariance between the residuals of group mates.

When group means are used, the same line of reasoning is followed, but vector **y*** contains in this case the mean of the performances of the individuals in a group. The design matrices **X*** and **Z*** still reflect the group composition, but their row totals add up to 1. Vector **e*** contains the mean of the residuals of individuals in a group.

$$\mathbf{y}^* = \begin{bmatrix} (15+14)/2 \\ (9+11)/2 \\ (10+9)/2 \end{bmatrix} = \begin{bmatrix} 14.5 \\ 10 \\ 9.5 \end{bmatrix}, \mathbf{X}^* = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{Z}^* = \begin{bmatrix} 0 & 0 & 0 & 0 & .5 & 0 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .5 & .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & .5 \end{bmatrix},$$

$$\mathbf{e}^* = \begin{bmatrix} (\mathbf{e}_1 + \mathbf{e}_2)/2 \\ (\mathbf{e}_3 + \mathbf{e}_4)/2 \\ (\mathbf{e}_5 + \mathbf{e}_6)/2 \end{bmatrix}$$

Vectors **a** and **b** of solutions for random and fixed effects and the genetic variance-covariance matrix do not change compared to a model for individual observations, while the residual variance is $Var(\mathbf{e}^*) = \mathbf{R}^* = \mathbf{D}^* \sigma^2_e$ where **D*,** assuming residuals are not correlated, is a diagonal matrix with the reciprocals of group size ($1/n_j$) along the diagonal. For groups of equal size, $\mathbf{D}^* = 1/n \times \mathbf{I}$ and $\sigma^2_{e^*} = 1/n \sigma^2_e$ (whose square root is the standard error of the mean, $\sigma_e/\sqrt{n}$). The ratio of the residual and genetic variances, α, becomes $1/n$ times that of individual observations:

$$\alpha^* = \sigma^2 e^* / \sigma^2 a = {}^1/_n . \sigma^2 e / \sigma^2 a = 1/n \times \alpha$$

Again, if groups are of different sizes $\mathbf{R}^{*-1}$ must be used in the mixed model equations.

Whole group responses, such as that to the novel object test, can be treated as group averages.

Estimation of covariances

In multivariate analysis, multiple traits measured on the same animals are evaluated simultaneously (Henderson and Quaas, 1976). This allows for the estimation of genetic and residual covariances between traits from which genetic parameters for the implementation of genetic improvement schemes (e.g. heritabilities and genetic correlations) can be calculated. Covariances are usually estimated from individual observations: however, also pooled data might serve the purpose. In the case of a bivariate analysis the model in matrix notation is:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

The genetic variance-covariance matrix is:

$$\mathbf{G} = \mathbf{Var} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}g_{11} & \mathbf{A}g_{12} \\ \mathbf{A}g_{21} & \mathbf{A}g_{22} \end{bmatrix}$$

where **A** is the additive relationship matrix, g_{11} and g_{22} are the genetic variances of the two traits and $g_{12} = g_{21}$ is the genetic covariance between the traits. The residual variance-covariance matrix is:

$$\mathbf{R} = \mathbf{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}\varepsilon_{11} & \mathbf{I}\varepsilon_{12} \\ \mathbf{I}\varepsilon_{21} & \mathbf{I}\varepsilon_{22} \end{bmatrix}$$

with **I** being the identity matrix ε_{11} and ε_{22} are the residual variances of the two traits and $\varepsilon_{12} = \varepsilon_{21}$ is the residual covariance between the traits.

In the analysis of pooled data the matrix **G** is unaffected since the vectors of genetic values $\mathbf{a_1}$ and $\mathbf{a_2}$ are the same as for individual observations. Contrariwise, in the matrix **R** the identity matrices need to be substituted by the diagonal matrix **D**, whose elements depend on group size:

$$\mathbf{R}^* = \mathbf{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{D}\varepsilon_{11} & \mathbf{D}\varepsilon_{12} \\ \mathbf{D}\varepsilon_{21} & \mathbf{D}\varepsilon_{22} \end{bmatrix}$$

For a group of two animals, individual 1 and individual 2, the residual covariance between trait x and y would be:

$$Cov(e_{1x} + e_{2x}, e_{1y} + e_{2y}) = Cov(e_{1x}, e_{1y}) + Cov(e_{1x}, e_{2y}) + Cov(e_{2x}, e_{1y}) + Cov(e_{2x}, e_{2y}),$$

which, assuming that residuals between animals in a group are uncorrelated, reduces to:

$$Cov(e_{1x} + e_{2x}, e_{1y} + e_{2y}) = Cov(e_{1x}, e_{1y}) + Cov(e_{2x}, e_{2y}) = 2 \cdot Cov(e_x, e_y) = 2r_e\sigma_{e_x}\sigma_{e_y}$$

When group sums are used as pooled data, therefore, assuming all groups have the same size, the residual covariance between two traits is equal to that for individual records multiplied by the group size. Consequently, the residual covariance from pooled data will be $Cov(e_1^*, e_2^*) = n \times Cov(e_1, e_2)$. Recalling that $Var(e^*) = n \times Var(e)$ (see chapter 2), $Cov(p_1, p_2) = Cov(a_1, a_1) + Cov(e_1^*, e_2^*)/n$, and $Var(p) = Var(a) + Var(e^*)/n$. The genetic and phenotypic correlations based on pooled data are therefore $r_g = \frac{Cov(a_1, a_2)}{\sigma_{a1} \cdot \sigma_{a2}}$ and $r_p = \frac{Cov(p_1, p_2)}{\sigma_{p1} \cdot \sigma_{p2}}$.

Based on these theoretical considerations, I estimated genetic and phenotypic correlations for pooled body weight at 19, 27, 43 and 51 wk of age. The pooled observations were obtained by adding the individual observations by cage (i.e. cage sums: see Chapter 2 for more details on the traits). This analysis allows a comparison of the estimates of correlations based on individual and pooled data. Estimates for BW traits were based on about 2000 hens and 500 cages, for individual and pooled records respectively. Results are reported in Table 6.1 (genetic correlations) and Table 6.2 (phenotypic correlations).

With individual observations, genetic correlations for BW traits ranged from 0.61 between BW19 and BW51 to 0.90 between BW27 and BW43, and phenotypic correlations ranged from 0.47 between BW19 and BW51 to 0.82 between BW27 and BW43. With pooled data, genetic correlations for BW traits ranged from 0.43 between BW19 and BW51 to 0.88 between BW27 and BW43, and phenotypic correlations ranged from 0.32 between BW19 and BW51 to 0.79 between BW27 and BW43.

Correlations between BW43 and BW51 were not estimable both with individual and pooled records. Standard errors of correlations estimated from pooled data were generally bigger than with individual observations.

Table 6.2: genetic correlations for BW traits of laying hens estimated from individual (above diagonal) and pooled (below diagonal) records. Standard errors of the estimates are between brackets.

Corr	BW19		BW27		BW43		BW51	
BW19			0.74	(0.04)	0.66	(0.05)	0.61	(0.05)
BW27	0.73	(0.08)			0.90	(0.02)	0.86	(0.02)
BW43	0.69	(0.12)	0.88	(0.05)			not estimable	
BW51	0.43	(0.17)	0.52	(0.16)	not estimable			

Table 6.3: phenotypic correlations for BW traits of laying hens estimated from individual (above diagonal) and pooled (below diagonal) records. Standard errors of the estimates are between brackets

Corr	BW19		BW27		BW43		BW51	
BW19			0.58	(0.02)	0.49	(0.02)	0.47	(0.02)
BW27	0.55	(0.03)			0.82	(0.01)	0.80	(0.01)
BW43	0.46	(0.04)	0.79	(0.02)			not estimable	
BW51	0.32	(0.04)	0.46	(0.04)	not estimable			

Correlations from individual and pooled data were generally in good agreement, except for correlations with BW measured later in life. While genetic covariances from individual and pooled data should more or less correspond, residual covariances from pooled data are expected to be 4 times those from individual observations (cages comprise 4 hens each). Table 6.3 shows genetic and residual covariances from individual and pooled data. The ratio between the residual covariance estimated from pooled and individual records was about 4 for the covariance between BW19 and BW27, slightly higher for the covariances between BW19 and BW43, and BW27 and BW43. However, the ratio was much larger than 4 for estimates of covariances between BW19 and BW51, and BW27 and BW51. In the derivations of variances and covariances for pooled data showed previously, it was assumed that residuals within a group are uncorrelated. This might not be true at later time points, when competition effects (Bijma et al., 2007a,b) may emerge. Social interactions are in fact likely to play a bigger role later in time. At 19 wk of age hens enter the cage and their weight can not be influenced by interactions with cage mates; at 51 wk of age social interactions may have been operating for 32 wk. If ignored in the model, social

interactions can lead to incorrect estimation of variance and covariance components. The same pattern was observed for residual variances, as described in Chapter 2.

Table 6.4: genetic and residual covariances for BW traits estimated with individual and pooled records

	INDIVIDUAL		POOLED	
	cov(A)	cov(E)	cov(A)	cov(E*)
BW19-27	8007	3193	7962	12369
BW19-43	8746	3396	7937	18274
BW19-51	8830	3524	8855	34054
BW27-43	12866	8043	10599	40829
BW27-51	12369	7962	10466	56637

The same approach was then used to estimate genetic and phenotypic correlations between production traits, for which only pooled data were available (see Chapter 3 for details on the data). For egg production traits genetic correlations were estimated from 3803 cage records. Covariances were estimated also between early and total egg production, for which only pooled data were available. Results are summarized in Table 6.4. Van der Berg et al. (unpublished results) estimated covariances between egg production traits from pooled data: between early and total egg production she found a genetic correlation of 0.51 (s.e. 0.22) and a phenotypic correlation of 0.23 (s.e. 0.03). Again between early and total egg production, Nurgiartiningsih et al. (2004) estimated genetic correlations of 0.83 and 0.84, and of 0.81 and 0.63 from individual and pooled records in two lines of laying hens.

Table 6.5: covariances and correlations between early and total egg production estimated from pooled data[1] Standard errors for genetic and phenotypic correlations are reported between brackets

	cov(A)	cov(E*)	cov(P)	r(G)		r(P)	
EPr - TotPr	74.2	1204.9	375.4	0.24	(0.11)	0.27	(0.02)

[1)]EPr: early egg production; TotPr: total egg production; cov(E*): residual covariance from the analysis of pooled data; cov(P) = cov(A) + cov(E)/4.

In conclusion, the results shown in this thesis demonstrate that it is possible to use pooled data in the BLUP methodology for single- and multiple-trait genetic evaluation of farm animals. The use of group housing data instead of individual housing data in genetic models closer resembles the commercial situation. This might lead to higher response to selection. Besides, when pooled data are use, more animals can be tested with the same level of costs, thus contributing to increase genetic progress resulting

from breeding programmes. Software packages however need to be upgraded in order to extend the current methodology for estimating genetic parameters to groups of unequal size.

Association studies based on pooled data

At the start of the genomic era genotyping was expensive relative to the collection of phenotypes. This led to the development and application of the so-called grand-daughter design in QTL-mapping experiments. In this experimental design genotypes were only collected on parents, and phenotypes were recorded on larger numbers of offspring. Over time, genotyping has become considerably cheaper. For some traits the cost of measuring phenotypes has therefore become the limiting factor in designing experiments. If costs of genotyping continue to decrease, this might be the case for all traits in the future. In many situations it is economically attractive to genotype all the animals and measure the phenotypes on the genotyped animals directly. It might even become attractive to record phenotypes at group rather than individual level, like eggs produced per cage or feed intake in litters of piglets. These pooled data can be used in genomic analyses like, for instance, association studies: the average performance of a group of animals can be associated with the allele frequency of genetic markers in that group. Using the data presented in chapter 5, I analysed the average feather damage per cage associating it with the allele frequency of SNP markers in the cage. Genomic regions with an effect on feather pecking behaviour could thus be identified. The results were comparable with those from the analysis of individual feather damage (see Chapter 5): three of the 5 most significant marker-phenotype associations from the pooled data analysis were also among the most significant marker-phenotype associations in the individual feather scores. However, much fewer significant associations were detected (12 vs ~90) in total, and the significance levels were considerably lower (average FDR of 25% vs 8%). This can be partly explained taking into account that with pooled data some information is lost compared to individual data. The analysis of pooled data was used also to check the consequences of violating the assumption of independence of the residuals. When individual records are analysed, residuals are usually assumed to be independent. In the case of laying hens housed in the same cage, this assumption might not hold, and the analysis of individual and group performances may yield very different results. In our analysis, a certain degree of correspondence was found between the results of the

association study of individual and pooled feather s cores. This means that in this specific case ignoring that residuals in a cage are not independent did not invalidate the results.

Across-line association studies

In Chapters 4 and 5 the methodology to conduct association studies across lines of laying hens has been presented. To my knowledge, this is the first time that association studies across populations have been performed in farm animals. When multiple lines are analysed simultaneously, there is less LD between the marker and the QTL so that, due to recombination events, the phase of the marker-phenotype association might be different in the different lines. Therefore, a higher marker density is required in order to detect associations, and attention has to be paid to ensure that the phase of the associations is the same in every line. On the other hand, the lower extent of LD conserved across lines increases the resolution of the association study, thus allowing to detect markers that are much closer to the QTL. For most of the linkage studies the confidence interval for QTL location covers over 20 cM (Soller et al., 2006). Compared to linkage analysis, association studies reduce this confidence interval, to an extent which depends on the amount of LD. In this thesis a step forward was taken, by moving to across-line association studies. This reduces the confidence interval for QTL location even further: in Chapter 4 I showed that, with the given set of SNPs and layer lines, a marker found to be significantly associated with a trait is likely to lie within 1cM of the QTL. This is a considerable improvement compared to within-line association studies and QTL mapping experiments.

The genetic markers (SNPs) used in the across-line association studies described in Chapters 4 and 5 have been selected from QTL regions and candidate genes relevant for the studied traits reported in the scientific literature. This restrictive approach in choosing genetic markers from specific regions or genes for the across-line association study proved to be successful and led to the detection of genes and genomic regions associated with the immune response and feather pecking behaviour of laying hens. Some of the detected genes are indeed correlated in biological sense with the examined phenotypes, like for instance the genes for several interleukins (IL10, IL12B, IL17A) and the MHC (major histocompatibility complex) genes in the

case of the immune response, and the gene for the serotonin receptor HTR2C in the case of feather pecking behaviour.

A better coverage of the chicken genome, e.g. using the current 60k SNP-chip, would lead to higher resolution of the association studies and would probably detect additional genes – and perhaps unexpected – genes that play a role in the analysed phenotypes.

Estimation of relationships from genetic markers

Genetic markers – microsatellites, RFLP, SNPs, etc … - can not only be used to identify loci influencing the phenotypic expression of specific physiological or behavioural animal characteristics, but also to study the genetic relationships between animals and populations. Genetic markers can be used to infer the genomic relationships between animals, with which the genomic relationship matrix (G-matrix) can be obtained (Van Raden, 2008). Genetic markers can be used to estimate the genetic distance between populations (species, breeds, lines), and to construct phyologenetic trees. These two approaches can be combined by extending the construction of the G-matrix by including information on the relationships between different populations (e.g. lines or breeds). Relationships between populations can not be derived from pedigree files –or are so small that are supposed to be equal to 0- and are consequently ignored in the construction of the pedigree-based relationship matrix (A-matrix).

Estimation of within- and between-line variance based on genomic relationships

The additive relationship matrix (A-matrix) used in the classical MME approach to genetic evaluations of farm animals reports the average proportion of alleles identical by descent (IBD) shared by a pair of individuals predicted on the basis of their pedigree relationship (Henderson, 1976). Due to Mendelian sampling the true proportion of genes shared by a pair of individuals can be different from the average value inferred from pedigree information (Hill, 1993; Guo, 1996). For instance, two full-sibs are expected to have on average 50% of the genes in common but, depending on which alleles they inherit from their parents at the moment of meiosis, this proportion can theoretically vary from 0 to 100%. Around the mean value of 50%, Guo (1996) estimated a standard deviation of 4% for additive relationships between full-sibs in a species with 30 chromosomes of 1M length each. The availability of

genomic information, therefore, allows for a more accurate estimation of the actual proportion of alleles IBD for each pair of individuals. The use of genomic instead of pedigree relationships leads to higher accuracy of the EBVs (Villanueva et al., 2005; Hayes et al., 2009), thus increasing response to selection. Also genetic parameters are expected to be estimated more accurately with genomic relationships. So far the use of genomic relationship matrices for estimating breeding values or genetic parameters has been limited to studies within populations. Using the 1031 SNPs with a minor allele frequency > 5% from the 1536 SNP chip described in Chapters 4 and 5, the genomic relationship matrix (G-matrix) was obtained for 675 hens originating from 9 genetic lines of White Leghorn and Rhode Island Red origin. The first of the 3 methods described by Van Raden (2008) was used in this study to estimate genomic relationships. Allele frequencies were calculated across lines. More details on material and methods can be found in Biscarini et al. (2010). The G-matrix was used to estimate the heritability and EBVs for body weight at 19 wk (BW19) in laying hens. These results were compared to those from the classical MME where the pedigree-based A-matrix was used (Table 6.5). The estimated heritabilities using both approaches looked very similar. However, they are not fully comparable, since the A-matrix model estimates only the within line genetic variation, whereas the G-matrix model includes also the between-line genetic variation. The heritability estimated using the G-matrix had lower standard errors (roughly half) compared to the heritability estimated using the A-matrix.

Genomic relationships may be extended by including relatedness between different populations. The information on relatedness of populations can be incorporated in the estimation of the genetic variance between populations. In this way both the within-line and between-line genetic variation can be estimated. Variance components for body weight at 19 wk (BW19) were estimated using the extended G-matrix. These estimates were compared with those obtained from a model using the pedigree-based A-matrix and population fitted as random effect. Results are shown in Table 6.5.

Tabel 6.5: variance components for BW19 estimated using the G-matrix and the A-matrix. Var(line) is the between-line variance estimated with the line effect fitted as random.

BW19	var(A)	var(E)	var(line)	h2	s.e.
A-matrix	9117	8972	21149	0.504	0.103
G-matrix	15653	15299		0.506	0.054

In the analysis with the extended G-matrix the total genetic variance is estimated. The total genetic variance is the combination of the within- and between-line variances. This estimate is higher than the additive genetic variance (9117 g^2) estimated with the A-matrix. However, the G-matrix includes also the between-line variance, whereas the A-matrix model estimates the within-line genetic variance. In the latter analysis, the between-line variance is estimated separately by fitting line as random effect. The variance of line fitted as random effect amounts to 21149 g^2. This allows for a different definition of heritability, required in order to make a fair comparison between the two models. A total heritability could be defined that includes both the within-line and between-line variation. For the two models, this would be as follows:

$$A-matrix: h^2_{Total} = \frac{\sigma^2_a + \sigma^2_{line}}{\sigma^2_a + \sigma^2_{line} + \sigma^2_e} = \frac{9117 + 21149}{9117 + 21149 + 8972} = 0.77$$

$$G-matrix: h^2_{Total} = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_e} = \frac{15653}{15653 + 15299} = 0.51$$

A larger heritable variation is estimated with the A-matrix model. This suggests that the G-matrix model doesn't seem to capture very accurately the between-line variation. This is illustrated also by the smaller between-line variance estimated with the G-matrix model obtained by subtracting the additive genetic variance estimated with the A-matrix model from the total genetic variance estimated with the G-matrix model. This gives a value of 6536 g^2 for the between-line variance estimated from the G-matrix model, approximately 30% of the between-line variance estimated with the A-matrix model. This large difference is indeed unexpected. It might be due in part to the choice of the allele frequencies used to derive the G-matrix. In our analysis the across-line allele frequencies were used. It might be worthwhile to explore other possibilities: for instance, within-line allele frequencies could be used to calculate within-line genomic relationships, and across-line allele frequencies could be used to derive relatedness between lines. It might also be that it is better to estimate between-line variance (including relationships between lines) and within-line variance separately. However, the results of Table 6.5 show that there is a substantial contribution of the between-line variation to the total genetic variance: from 30% (G-matrix model) to over half of the total genetic variance (54% with the A-matrix model). If different populations are analysed simultaneously, the fixed effect of lines

or genetic groups is usually included in the model; genomic relationships offer an alternative for modeling the joint analysis of multiple populations. There are anyway, some issues that still need to be addressed, such as the much larger residual variance estimated with the G-matrix model as compared to the A-matrix model, and the number of markers used to reconstruct genomic relationships: in this study 1031 SNPs were used, but this might not be enough, especially if relationships across lines are reconstructed.

Genetic distances between lines

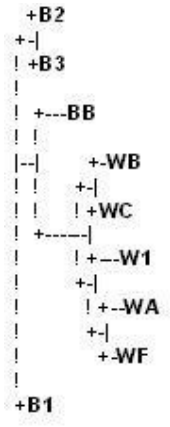
Genetic markers can be used to infer relationships between populations. Based on the allelic frequencies of genetic markers it is possible to measure the genetic distance between two populations. The method of Nei (1972) was applied to 675 hens from 9 different lines genotyped for 1536 SNPs: the genetic distances thus obtained are reported in Table 6.6. The computer programme Phylip was used to estimate genetic distances.

Tabel 6.6: Genetic distances between 9 lines of laying hens estimated from marker allele frequencies according to Nei (1972).

Line	B1	B2	B3	BB	W1	WA	WB	WC	WF
B1	-	**0.049**	**0.051**	**0.146**	*0.264*	*0.252*	*0.255*	*0.256*	*0.246*
B2	**0.049**	-	**0.028**	**0.155**	*0.264*	*0.250*	*0.252*	*0.263*	*0.249*
B3	**0.051**	**0.028**	-	**0.147**	*0.263*	*0.246*	*0.250*	*0.259*	*0.245*
BB	**0.146**	**0.155**	**0.147**	-	*0.252*	*0.247*	*0.238*	*0.227*	*0.246*
W1	*0.264*	*0.264*	*0.263*	*0.252*	-	**0.143**	**0.131**	**0.121**	**0.109**
WA	*0.252*	*0.250*	*0.246*	*0.247*	**0.143**	-	**0.138**	**0.140**	**0.100**
WB	*0.255*	*0.252*	*0.250*	*0.238*	**0.131**	**0.138**	-	**0.057**	**0.115**
WC	*0.256*	*0.263*	*0.259*	*0.227*	**0.121**	**0.140**	**0.057**	-	**0.102**
WF	*0.246*	*0.249*	*0.245*	*0.246*	**0.109**	**0.100**	**0.115**	**0.102**	-

Genetic distances are expressed as the opposite of the logarithm of the normalized identity of genes between any two populations (Nei, 1972): therefore smaller values correspond to shorter genetic distances and closer relationships between lines. From genetic distances the phylogenetic tree visualized in figure 6.1 could be drawn.

Figure 6.1: Phylogenetic tree of 9 lines of laying hens of White Leghorn and Rhode Island Red origin. Ramifications are derived from the genetic distances calculated according to Nei (1972).



Approximate genetic distances can be obtained also from the average coefficients of relationship between-lines in the G-matrix: in this case higher values indicate a closer relationship and a shorter genetic distance between populations. Average coefficients of relationships were obtained from the G-matrix built using the 1536 SNP markers genotyped on the same 675 laying hens mentioned above. Results are shown in Table 6.7. The rank correlation between the average coefficients of relationship derived from the G-matrix and the genetic distances estimated with the method of Nei is 1, indicating perfect correspondence between the two measures of relatedness between populations. Results from both methods show that lines of White Leghorn and of Rhode Island Red origin are closer between them than across them.

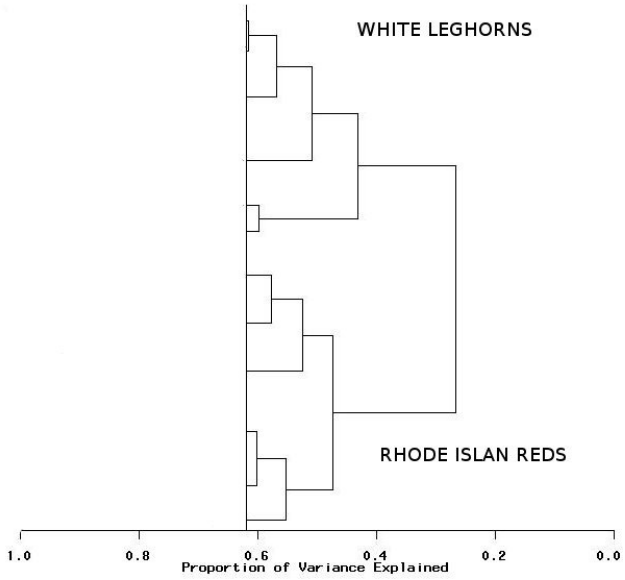
Table 6.7: Average coefficients of relationship within- and between-lines derived from the G-matrix.

	B1	B2	B3	BB	W1	WA	WB	WC	WF
B1	-	**0.49**	**0.406**	**0.032**	*-0.367*	*-0.382*	*-0.423*	*-0.297*	*-0.364*
B2		-	**0.462**	**0.071**	*-0.355*	*-0.363*	*-0.397*	*-0.281*	*-0.355*
B3			-	**0.017**	*-0.326*	*-0.328*	*-0.328*	*-0.2*	*-0.311*
BB				-	*-0.025*	*-0.102*	*-0.131*	*-0.072*	*-0.053*
W1					-	**0.078**	**0.107**	**0.083**	**0.152**
WA						-	**0.195**	**0.199**	**0.298**
WB							-	**0.339**	**0.198**
WC								-	**0.239**
WF									-

Additionally, based on the SNP allele frequencies a cluster analysis was performed. At first data were let free to cluster (Figure 6.2 and Table 6.8); later the maximum number of cluster was set to 9, the number of genotyped lines (Table 6.9). The procedure TREE of the software package SAS® 9.1.3 was used (SAS Institute Inc., 2002-2003). In the first analysis we see that based on the genotype data it was

possible to clearly distinguish and recognize the two main clusters, which corresponded to white and brown layers (Figure 6.2).

Figure 6.2: Phylogenetic tree reconstructed from the reuslts of cluster analysis. Branches of the tree represent clusters, which do not correspond perfectly with lines.



From Table 6.8 it can be seen also that some lines were almost uniquely attributed to one cluster (B1, WA, WB, WC, WF), while others fitted in a couple of main clusters (B2, B3, W1), sometimes overlapping between two lines (e.g. cluster14 for lines B2 and B3). Line BB appeared to be the most difficult to attribute to specific clusters, and hens belonging to this line were scattered over several clusters. Outlying animals falling in clusters not corresponding to lines as coded in the pedigree file might actually represent pedigree errors, i.e. animals incorrectly registered as pertaining to a specific line. According to Hendrix Genetics (personal communication), the total amount of pedigree errors may be as high as 10%, and a portion of these refers to hens wrongly assigned to a specific line. In addition, errors might be introduced also when blood samples are taken for genotyping and not correctly assigned to animals.

Table 6.8: cluster analysis of laying hens based on their genotype data. No maximum number of cluster was specified.

		clusters											
Line	**Hens**	**c9**	**c11**	**c13**	**c14**	**c16**	**c17**	**c18**	**c19**	**c20**	**c21**	**c22**	**c23**
B1	83	**77**	1							1	3	1	
B2	77		1		**34**			6	4		1	1	**30**

B3	77				**34**				1	1		11	**30**	
BB	67			2	4	5	5	6	7	**24**	2	4		8
W1	74				**22**			**43**	7	1			1	
WA	78	2	**71**		1				1		1	2		
WB	77			1	1		1				**70**	4		
WC	63	3		2		4		1	4	3	3	**41**	2	
WF	79					1	**66**		7	4	1			

When limiting the number of clusters to 9 (the number of lines present in the study), similar patterns were observed (Table 7.10). Most of the clusters corresponded well and uniquely to specific lines (B1, B2, B3, W1, WA, WB, WF), with the exception of line BB whose hens were scattered over several clusters, overlapping both with brown and white layer lines. This matches with results shown in Table 6.9; besides, line BB was also the less related with other brown layers lines and the closest, among them, to the white layer lines (Tables 6.6 and 6.7; Figure 6.1). The explanation is that, unlike other brown layer lines, line BB originates from crosses with White Leghorn lines (Hendrix Genetics, personal communication). From pedigree it is also known that lines B2 and B3 are the most closely related among Rhode Island Reds, and lines WB and WC are the most closely related among White Leghorns. This is confirmed by the results of the estimation of genetic distances (Tables 6.6 and 6.7; Figure 6.1) and of cluster analysis (Tables 6.8 and 6.9).

Table 6.9: cluster analysis of laying hens base don their genotype data. The maximum number of cluster was limited to 9, the number of genetic lines present in the study.

		clusters								
Line	**Hens**	**c5**	**c7**	**c9**	**c12**	**c13**	**c14**	**c15**	**c16**	**c17**
B1	83		4		**75**		1		1	1
B2	77	7	2			**67**	1			
B3	77	1	7		1				**31**	**37**
BB	67	**27**	12	8		10	2	8		
W1	74	8		**64**	2					
WA	78	1	4			1	**66**		1	
WB	77	2	**71**	1		2	1			
WC	63	8	**38**	1	4		5	1	4	2
WF	79	8	3					**68**		

Summarizing, genetic markers can be used effectively to estimate genetic distances between populations. The estimation of genetic distances between genetic lines or breeds is of interest and use, for instance, in reconstructing the phylogenetic history of animal populations, but also, as we see in Chapter 4, to exploit LD in genomic selection and to interpret the results of associations studies and QTL mapping

experiments. Together with cluster analysis and its comparison with pedigree data, it can be a valid help also to identify and correct pedigree errors which would otherwise affect the results of genetic analysis and evaluation.

Overall this thesis touched upon interesting frontiers of poultry genetics, both quantitative and molecular. First, the possibility of using group performances (pooled data) instead of individual performances for the estimation of genetic parameters and the prediction of breeding values was investigated. This opens the opportunity to use genetic models closer to the reality of commercial laying facilities, thus leading probably to higher response to selection. The possibility of using pooled data gives also greater flexibility to the methodology of genetic evaluation allowing for the use of different kinds of data. Secondly, genetic markers were used in across-line association studies. This constitutes a new approach to the analysis of genetic marker data, that have been so far limited to within-line analyses. Analysing multiple lines simultaneously is challenging from a statistical point of view, but offers the possibility of exploiting the LD conserved across lines to detect QTLs with higher resolutions as compared to within-line association studies. This approach was successfully applied to the analysis of immune response and feather pecking behaviour in laying hens. Several genomic regions of interest were identified, and the role of some genes (e.g. IL17A, MHC and HTRC2C) in the modulation of the immune response and the developing of feather packing behaviour in laying hens.

REFERENCES

Biscarini F., H. Bovenhuis and J. A. M. van Arendonk (2008) Estimation of variance components and predictions of BVs using pooled data. *J. Anim. Sci*. **86**:2845-2852.

Biscarini F., H. Bovenhuis, J. A. M. van Arendonk, H. K. Parmentier, A. P. Jungerius and J. J. Van der Poel (2010) Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim. Genet*. **41**:26-38.

Bijma P., W. M. Muir and J.A.M. van Arendonk (2007a) Multilevel Selection 1: Quantitative Genetics of Inheritance and Response to Selection. *Genetics* **175**:277–288.

Bijma P., W. M. Muir, E. D. Ellen, J. B. Wolf and J.A.M. van Arendonk (2007b) Multilevel Selection 2: Estimating the Genetic Parameters Determining Inheritance and Response to Selection. *Genetics* **175**:289–299.

Goddard M. E. and B. J. Hayes (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* **10**:381:391.

Guo S. W. (1996) Variation in genetic identity among relatives. *Hum. Hered*. **46**:61-70.

Hayes B. J., P. M. Visscher and M. E. Goddard (2009) Increased accuracy of artificail selection by using the realized relationship matrix. *Genet. Res*. **91**:47-60.

Henderson C. R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**:423-447.

Henderson C. R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in predicting breeding values. *Biometrics* **32**:69-83.

Henderson C. R. and R. L. Quaas (1976) Multiple trait evaluation using relatives' records. *J. Anim. Sci*. **43**:1188-1197.

Henderson C. R. (1984) Applications of linear models in animal breeding. Univ. Guelph Press.

Hill W. G. (1993) Variation in genetic identity within kinships. *Heredity* **71**:652-653.

Nei M. (1972) Genetic distance between populations. *Am. Nat*. **106**:283-292.

Nurgiarrtiningsih V. M. A., N. Mielenz, R. Preisinger, M. Schmutz and L. Schüler (2004) Estimation of genetic parameters based on individual and group mean records in laying hens. *Br. Poult. Sci*. **45**:604-610.

Olson, K. M., D. J. Garrick and R. M. Enns (2006) Predicting breeding values and accuracies from group in comparison to individual observations. *J. Anim. Sci.* **84**:88-92.

Simianer, H. and B. Gjerde (1991) Estimating variance components from fullsib groups means. *J. Anim. Breed. Genet*. **108**:270-279.

Soller M., S. Weigend, M. N. Romanov, J. C. Dekkers and S. J. Lamont (2006) Strategies to assess structural variation in the chicken genome and its

associations with biodiversity and biological performance. *Poult. Sci*. **85**:2061-2078.

Uitdehaag K. A., H. Komen, T. B. Rodenburg, B. Kemp and J. A. M. van Arendonk (2008) The novel object test as predictor of feather damage in cage-housed Rhode Island Red and White Leghorn laying hens. *Poult. Sci.* **109**:292-305

Van Raden P. M. (2008) Efficient methods to compute genomic predictions. *J. Dairy Sci*. **91**:4414-4423.

Villanueva B., R. Pong-Wong, J. Fernández and M. A. Toro (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. *J. Anim. Sci.* **83**:1747-1752.

Summary

The new regulations about the husbandry of laying hens and the so-called genomic revolution offer both opportunities and challenges for the breeding of layers. Hens are currently housed mainly in battery cages of 4 individuals each. Following recent developments of the communitarian legislation, many countries will soon adopt furnished cages or non-cage systems, which will lead to larger groups of hens. Also, beak-trimming will be prohibited in EU countries in the near future. Advancements in sequencing technology are making an always greater number of genetic markers available at increasingly cheaper prices, making genome-wide studies possible and helping geneticists to start unraveling the mystery of the genetic make-up of animals, which until a few years ago was considered a black-box. This thesis touches upon the impact of such innovations on the breeding of laying hens.

Use of pooled data in the genetic evaluation of laying hens

Hens are usually housed in cages and therefore pooled instead of individual egg records are often available: a pooled egg record is the total production of a cage, when the egg production of the individual hens is unknown. Current selection schemes are carried out in nucleus herds where hens are housed individually, so that egg production of individual birds can be recorded and used for genetic evaluations. Based on this information sires and dams are selected. Such a selection scheme based on individually housed hens introduces a discrepancy between the environment where hens are selected and the environment in which hens are kept for commercial egg production (group housing). Selecting animals in one environment and using them in a different environment might lead to genotype x environment interaction (Besbes and Ducroq, 2003), thereby reducing the realized response to selection. Future husbandry conditions, with larger groups of hens or hens housed in furnished cages might make this problem even worse. A method to use pooled data in the genetic evaluation of laying hens would therefore be of interest. In Chapters 2 and 3 of this thesis it is described how to use pooled records for the estimation of heritability and breeding values. In chapter 2 the use of individual and pooled observations is compared. Individual body weights of hens at different ages were available: these were then

pooled by cage in order to create pooled records. Heritabilities estimated from pooled and individual data correlated well: the standard error of estimates based on pooled records was however about twice that of estimates based on individual records. The accuracy of EBVs from pooled data is lower than the accuracy of EBVs from individual data; in the case of sires with at least 10 offspring the reduction in accuracy was about 23%. This loss of precision in estimating genetic parameters and breeding values is understandable considering that pooled records are a less detailed of information. However, this lower accuracy should be interpreted in the context of direct vs indirect selection. The breeding goal is the trait under commercial conditions (group housing), and if testing is under individual housing, the genetic correlation between group and individual housing is relevant. The ratio of the selection response for direct and indirect selection is a function of the accuracies for both situations, the standard deviations of the traits and the genetic correlation between the traits (Falconer, 1989). Similarly, the ratio between accuracies based on pooled and individual data provides a threshold for the genetic correlation between individual and group housing below which pooled data would result in a greater selection response. In practical breeding also the costs of individual housing relative to the costs of group housing are relevant. Since group housing is cheaper than individual housing, more selection candidates could be tested for the same level of costs. This would in turn result in higher selection intensity and larger response to selection.

In chapter 3 the method of analyzing pooled data developed in chapter 2 was compared with an approximation consisting in assigning cage means to each hen in a cage, then treating them as individual observations. Cross-validation was used to compare the two methods: the method developed in Chapter 2 performed consistently better than the approximate method in terms of predicting ability.

In the general discussion, finally, it was described how to estimate genetic and phenotypic correlations from pooled data.

Across-line association studies for immune response and feather pecking behaviour

The great number of genetic markers available at increasingly lower prices has been fostering developments in genomic research. Association studies between genetic markers and phenotypes are typically conducted within populations (breeds, or lines): the amount of LD conserved in a population is exploited using high marker density,

such as SNP chips, and markers relatively close to QTLs are expected to show significant effects in association studies. In this thesis we propose to take it one step further and perform association studies across lines. This requires higher marker density but increases the resolution. The amount of LD conserved across lines is expected to be lower than within lines and the phase of the marker-phenotype association might be different in the different lines. On the other hand markers that happen to show significant effects in an across-line association study are likely to be close to the QTL. These issues in conducting marker-phenotype association studies across populations were addressed in Chapters 4 and 5 of this thesis, where it was shown how to deal with multiple populations when analyzing hens from 9 different genetic lines of White Leghorn and Rhode Island Red origin genotyped for a panel of 1536 SNP (Single Nucleotide Polymorphism) markers.

The traits analysed were immunological parameters and plumage damage due to feather pecking behaviour, two classes of traits for which, given that they have relatively low heritability and are difficult and expensive to measure, genomic information may be particularly valuable. Immunological parameters might be used in selection programmes aimed at improving disease resistance of laying hens, while information on the genetic background of feather pecking behaviour can be useful in reducing problems due to this behavioural disorder of layers. Under future husbandry conditions susceptibility to infectious diseases and feather pecking are expected to become more serious problems: both aspects of layer production are in fact related to the number of individuals that interact with each other, which will increase as a result of the application of the EU directive 1999/74/EC. In addition, the ban of beak-trimming will make it more difficult to control the consequences of feather pecking (plumage damage, cannibalism, mortality). Genetic selection might represent an appealing addition to the current control measures. The association studies identified several regions of interest. The gene for interleukin 17 (IL17), on chromosome 3, was found to be associated with natural and acquired antibody titres, and with the classical and alternative pathways of complement activation. The major histocompatibility complex (MHC) genes on chromosome 16 showed significant association with natural and acquired antibody titres and classical complement activity. The interleukin 12B gene (IL12B) on chromosome 13 was associated with natural antibody titres. As for feather pecking behaviour, a role of the gene for the serotonin receptor 2C (HTR2C) on chromosome 4 was found. This supports existing evidence of a prominent

involvement of the serotonergic system in the modulation of this behavioural disorder in laying hens. The genes for IL9, IL4, CCL4 and NFKB were found to be associated to plumage condition, revealing relationships between the immune system and behaviour.

Samenvatting

Nieuwe regels met betrekking tot huisvesting van leghennen en de zogenaamde genomische revolutie bieden zowel kansen als uitdagingen voor de fokkerij van leghennen. Momenteel worden hennen voornamelijk in legbatterijen gehouden met 4 individuen in een kooi. Om aan wetgeving te voldoen zullen veel landen op korte termijn overschakelen op verrijkte kooien en niet-kooi systemen, wat zal leiden tot grotere groepen hennen. Bovendien zal het in EU landen in de nabije toekomst verboden worden om snavels te knippen. Ontwikkelingen in sequencing technologie zorgen voor steeds meer genetische merkers tegen afnemende kosten, hierdoor is het mogelijk om studies uit te voeren over het gehele genoom en het bied genetici de mogelijkheid om het mysterie van de genetische architectuur van dieren, die tot voor kort nog werd beschouwd als een 'black box', op te lossen. Deze thesis gaat over de impact van zulke innovaties op de fokkerij van leghennen.

Gebruik van pooled data in genetische evaluatie van leghennen

Hennen zijn meestal met meerderen in een kooi gehuisvest waardoor enkel pooled data beschikbaar is aangezien het totale aantal eieren per kooi gemeten wordt in plaats van de individuele ei productie per hen. Huidige selectie schema's worden uitgevoerd in nucleus stallen waar hennen individueel gehuisvest zijn, zodat individuele ei productie geregistreerd kan worden en kan worden gebruikt voor genetische evaluatie. Aan de hand van deze informatie worden ouder dieren geselecteerd. In zo'n selectie schema, gebaseerd op individuele ei productie, komt de huisvesting van de selectie kandidaten niet overeen met de (groeps)huisvesting van de leghennen op commerciële bedrijven. Selectie van dieren in het ene huisvesting systeem en productie in een ander huisvesting systeem kan leiden tot genotype maal omgeving interactie (Besbes en Ducroq, 2003), waardoor de gerealiseerde response in selectie lager kan uitvallen dan verwacht. Toekomstige huisvestingssystemen met grotere groepen hennen of verrijkte kooien vergroten dit probleem nog verder. Methoden om pooled data te gebruiken in genetische evaluatie zijn daarom interessant. In hoofdstuk 2 en 3 van deze thesis wordt omschreven hoe pooled data gebruikt kan worden voor het schatten van erfelijkheidsgraden en fokwaarden. In hoofdstuk 2 wordt het gebruik van

individuele data en pooled data met elkaar vergeleken. Individueel lichaamsgewicht van hennen van verschillende leeftijd was beschikbaar: de lichaamsgewichten waren samengevoegd per kooi tot een pooled record. Erfelijkheidsgraden geschat van pooled data en individuele data hadden een sterke correlatie: de standaard fout van de schattingen gebaseerd op pooled data was echter ongeveer twee keer zo groot als de schattingen gebaseerd op individuele data. De nauwkeurigheid van fokwaarden gebaseerd op pooled data is lager dan de nauwkeurigheid van fokwaarden gebaseerd op individuele data. Voor hanen met minstens 10 nakomelingen was de nauwkeurigheid met 23% gedaald. Deze daling in nauwkeurigheid van genetische parameters en fokwaarden is begrijpelijk omdat pooled data minder gedetailleerde informatie bevat. Echter, deze lagere nauwkeurigheid zou geïnterpreteerd moeten worden als directe versus indirecte selectie. Als het fokdoel het kenmerk onder commerciële omstandigheden (groepshuisvesting) is en het kenmerk getest wordt onder individuele huisvesting is de genetische correlatie tussen de groep en individuele huisvesting van belang. De ratio van selectie response voor directe over indirecte selectie is een functie van de nauwkeurigheden van beide situaties, de standaard deviatie van het kenmerk en de genetische correlatie tussen beide kenmerken (Falconer, 1989). De ratio tussen nauwkeurigheid gebaseerd op pooled data en individuele data verschaft een drempel voor de genetische correlatie tussen individuele huisvesting en groepshuisvesting waaronder pooled data zal resulteren in een grotere selectie respons. In commerciële fokkerij zijn ook de kosten van individuele huisvesting ten opzichte van groepshuisvesting van belang. In groepshuisvesting kunnen meer selectie kandidaten getest worden, omdat groepshuisvesting goedkoper is dan individuele huisvesting. Op zijn beurt resulteert dit in een hogere selectie intensiteit en grotere respons op selectie.

In hoofdstuk 3 wordt de methode om pooled data te analyseren die in hoofdstuk 2 is ontwikkeld vergeleken met een benadering die bestaat uit het toekennen van een kooi gemiddelde aan iedere hen in een kooi en deze te beschouwen als een individuele observatie. Beide methoden zijn met elkaar vergeleken door middel van kruislingse validatie: de methode ontwikkeld in hoofdstuk 2 had constant een beter voorspellingsvermogen dan de benaderingsmethode van kooi gemiddelden.

Tenslotte is in de algemene discussie omschreven hoe genetische en fenotypische correlaties geschat kunnen worden met pooled data.

Associatie studies voor immuun respons en veren pikken over lijnen

Ontwikkeling in genomisch onderzoek worden aangemoedigd door het groeiende aantal genetische merkers beschikbaar tegen steeds lager wordende kosten. Associatie studies tussen genetische merkers en fenotypes worden voornamelijk uitgevoerd binnen populaties (rassen, of lijnen). De hoeveelheid LD geconserveerd in een populatie wordt benut door het gebruik van een hoge merker dichtheid, zoals SNP (single nucleotide polymorphism) chips en merkers relatief dicht bij QTL worden geacht een significant effect te tonen in associatie studies. In deze thesis stellen we voor om nog een stap verder te gaan en een associatie uit te voeren over lijnen. Hiervoor is een hogere merker dichtheid nodig maar het verhoogd de resolutie. Verwacht wordt dat de hoeveelheid LD geconserveerd over lijnen lager is dan binnen lijnen en dat de fase van de merker-fenotype associatie verschillend kan zijn in de verschillende lijnen. Aan de andere kant zullen significante merkers in een associatie over lijnen waarschijnlijk dicht bij de QTL zitten. Deze punten in merker-fenotype associatie studies over lijnen worden behandelt in hoofdstuk 4 en 5 van deze thesis. In deze hoofdstukken wordt een aanpak voor meerdere populaties getoond door 9 verschillende genetische lijnen, die afstammen de Witte Leghoorn en Rhode Island Rood, te analyseren die gegenotypeerd waren voor 1536 SNP merkers.

De kenmerken in de analyse waren immunologische parameters en beschadigingen aan verenkleed door veren pikken. Genomische informatie kan voor deze twee groepen kenmerken zeer waardevol zijn, gezien hun relatief lage erfelijkheidsgraad en moeilijke en dure meetmethoden. Immunologische parameters kunnen worden gebruikt in selectie programma's om weerstand tegen ziektes in leghennen te verhogen terwijl informatie over de genetisch achtergrond van veren pikken nuttig kan zijn om gedragsproblemen in leghennen te verminderen. Vatbaarheid voor infectueuze ziekten en veren pikken zijn gerelateerd aan het aantal individuen die interactie met elkaar kunnen hebben, dit aantal zal toenemen onder uitvoering van de EU richtlijn 1999/74/EC. Daarom worden serieuze problemen verwacht ten aanzien van beide aspecten onder toekomstige huisvestingsystemen. Bovendien zal het moeilijker worden om de consequenties van veren pikken (veder beschadiging, kannibalisme, sterfte) onder controle te houden door het verbod op snavel knippen. Genetische selectie kan een aantrekkelijke aanvulling zijn op de huidige controle maatregelen. Uit de associatie studies zijn verschillende interessante regio's naar voren gekomen. Het gen voor interleukine 17 (IL17), op chromosoom 3, toonde

associatie met natuurlijke en verworven antilichaam titers, en met de klassieke en alternatieve routes voor complement activering. De major histocompatibilieits complex (MHC) genen op chromosoom 16 toonden significante associatie met natuurlijke en verworven antilichaam titers en met de klassieke complement activiteit. Het interleukine 12B (IL12B) gen op chromosoom 13 was geassocieerd met natuurlijke antilichaam titers. Voor veren pik gedrag was er een rol voor de serotonine receptor 2C (HTR2C) op chromosoom 4 gevonden. Dit ondersteund bestaand bewijs voor een prominente betrokkenheid van het serotonergisch systeem in de modulatie van dit gedragsprobleem in leghennen. De genen voor IL9, IL4, CCL4 en NFKB waren geassocieerd met bevederings conditie, wat aan het licht brengt dat er relaties zijn tussen het immuun systeem en gedrag.

Curriculum Vitae

Filippo Biscarini was born in Roma (Italy). He graduated in Veterinary Medicine at the University of Perugia (Italy) in 2002. In 1999 he spent a 6 months study period at the University of Utrecht (The Netherlands) in the framework of the Erasmus european programme. In 2002 he worked as web-programmer and data base analyst at the ICBF (Irish Cattle Breeders Federation) in Bandon (Ireland). From 2003 to 2006 he was geneticist at the Italian Holstein Association (ANAFI) in Cremona (Italy). In 2005 he spent a 4 months working period as visiting scientist at the University of Guelph, in Canada. From September 2006 to February 2010 he worked on his PhD at the University of Wageningen (The Netherlands). Since March 2010 he is post-doc at the University of Göttingen (Germany). He is married with the musician Luisa Di Giacomo and has a child named Giovanni.

List of publications

Peer-reviewed articles

F. Biscarini, H. Bovenhuis, E. Ellen, S. Addo and J. A. M. van Arendonk (2010) "Estimation of heritability and BVs for early egg production in laying hens from pooled data". *Poultry Science*, doi:10.3382/ps.2010-00730

F. Biscarini, H. Bovenhuis, J.J. van der Poel, A. Jungerius and J.A.M. van Arendonk (2010) "Across-line SNP association study for direct and associative effects on feather damage in laying hens". *Behaviour Genetics*, doi:10.1007/s10519-010-9370-0

F. Biscarini, H. Bovenhuis, J. van Arendonk, H. Parmentier, A. Jungerius and J. van der Poel. (2009) "Across-line SNP association study of innate and adaptive immune response in laying hens", *Animal Genetics*, **41**:26-38.

F. Biscarini, H. Bovenhuis and J. van Arendonk. (2008) "Estimation of variance components and prediction of breeding values using pooled data", *Journal of Animal Science*, **86**:2845-2852.

Bulletins, posters and conference papers

F. Biscarini, H. A. Mulder, H. Bovenhuis and M. P. L. Calus (2010) "Estimating between and within line variation based on pedigree and genomic relationship matrix in laying hens", *WCGALP9*, 2010, Leipzig (Germany).

F. Biscarini, H. Bovenhuis and J.A.M. van Arendonk (2008) "Estimation of variance components & prediction of breeding values from pooled data", poster at the SABRE conference of September 2008 in Foulum, Denmark.

F. Biscarini, S. Biffani, F. Canavesi (2006) "The consequences of biases in the international genetic evaluations", Acts of the 8th World Congress on Genetics Applied to Livestock Production (WCGALP), 2006, Belo Horizonte (Brazil).

F. Biscarini, S. Biffani, F. Canavesi (2006) "Developing a data validation test based on Mendelian Sampling deviations", Interbull Bulletin n. 35.

S. Biffani, F. Biscarini, M. Marusi and F. Canavesi (2005) "Developing a genetic evaluation for fertility using angularity and milk yield as correlated traits", Interbull Bulletin n. 33.

F. Biscarini, S. Biffani and F. Canavesi. (2004) "Does selection for production affect type traits genetic evaluation in Italian Holsteins?", Interbull Bulletin n. 32.

F. Biscarini, S. Biffani and F. Canavesi (2003) "Genetic analysis of type traits for the Italian Jersey breed", Interbull Bulletin n. 31.

Acknowledgements

The three years and a half spent in Wageningen have been a very pleasant and rewarding time for me and my wife. Therefore my first acknowledgement goes to The Netherlands, a friendly country that has welcomed us and where we really enjoyed living, and that gave us plenty of opportunities to pursue our interests and to develop ourselves.

I am very grateful to Johan van Arendonk, Jan van der Poel and Henk Bovenhuis, my supervisors, for the guidance and support they gave me during my PhD. From them I learnt a lot, not only in the field of science but also in that, more complex and more important, of professional and personal life. Thanks also to Roel Veerkamp and his group at Lelystad, with whom I had the pleasure of cooperating.

A sincere thank you to all the colleagues at the ABGC: they contributed to create a cheerful and fruitful atmosphere in the working group, and many of them have been good friends during my stay in Wageningen.

I woluld also like to acknowledge Hendrix Genetics for providing the data and for the technical collaboration, the SABRETRAIN Project, of which this thesis work has been part, and the funding received by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission.

Finally, I would like to thank all the friends and nice people I met in The Netherlands, with whom I shared many unforgettable moments, my wife Luisa who accompanied me in this adventure, and our son Giovanni, who came to life while I was in Wageningen, giving us an incommensurable joy.

Training and Supervision Plan **Graduate School WIAS**

Name PhD student	Filippo Biscarini
Group	Animal Breeding and Genomics Centre
Daily supervisor(s)	Dr. Jan van der Poel and Dr. Henk Bovenhuis
Supervisor(s)	Prof. Dr. Johan van Arendonk
Project term	September 2006 to February 2010

The Basic Package		year	credits
WIAS Introduction Course		2007	1.5
WIAS course 'Biology underpinning animal sciences: Broaden your Horizon		2007	1.5
	Subtotal		**3**
Scientific Exposure			
International conferences			
EADGENE/SABRE Conference on "Genomics for Animal Health", Utrecht, The Netherlands		2007	1.5
SABRE Conference "Welfare and quality genomics", Foulum, Denmark		2008	0.6
QTL/MAS workshop, Wageningen, The Netherlands		2009	0.6
EAAP, Barcelona, Spain		2009	1.2
Seminars and workshops			
F&G Connection Day, Vught, The Netherlands		2008	0.6
WIAS Science Day		2009	0.3
Presentations			
Poster at the SABRE Conference, Foulum, Denmark		2008	1.0
Oral presentation at the F&G Connection Day 2008		2008	1.0
Oral presentation at the WIAS Science Day		2009	1.0
Oral presentation at the QTL/MAS workshop in Wageningen		2009	1.0
Oral presentation at the EAAP 2009 in Barcelona		2009	1.0
	Subtotal		**10**
In-Depth Studies			
Disciplinary and interdisciplinary courses			
Study of Resistance Mechanisms in Animal Infectious Diseases - Liège, Belgium		2006	2.0
Science meets society		2006/07	1.5
Fortran 95 for Animal Breeding		2007	1.5
QTL mapping, MAS and Genomic Selection		2008	1.5
Applied statistical methods in animal genomics		2008	1.5
Advanced statistics courses			
Modern Statistics in life sciences (modules)		2007	6.0
Bayesian statistics course		2007	2.0
WIAS Advanced Statistics course Experimental Design		2008	1.0
PhD students' discussion groups			
Quantitative Genetics study group		2006/09	2.0
R workshop		2007	2.0
	Subtotal		**21**
Professional Skills Support Courses			
Dutch course II		2007	1.0
SABRETRAIN Training Course in Industrially Relevant Skills		2007	1.0
Dutch Course III		2007	1.0
Teaching and supervising thesis students		2009	1.0
WUR Sprint UP Becoming a teacher		2009/10	2.0
	Subtotal		**6**
Research Skills Training			
Preparing own PhD research proposal		2006	6.0
Reviewer of a paper submitted to JAS (Journal of Animal Science)		2008	0.5

	Subtotal		**7**
Didactic Skills Training			
Lecturing			
ABG (Animal Breeding and Genetics) course		2009	1.0
WUR Sprint UP Becoming a teacher		2009/10	2.0
Supervising theses			
Supervising theses MSc major thesis		2008	2.0
Tutorship			
ABG (Animal Breeding and Genetics) course		2009	4.0
Preparing course material			
ABG (Animal Breeding and Genetics) course		2009	1.0
	Subtotal		**10**
Management Skills Training			
Organisation of seminars and courses			
Organization of the WIAS Science Day 2009		2009	1.5
Membership of boards and committees			
Representative of SABRETRAIN students		2007-09	0.5
	Subtotal		**2**
Education and Training Total			**58**

Colophon

The research in this thesis was part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission.

The printing of the thesis was funded by the Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands.

Printed by: Wöhrmann Print Service, Zutphen, The Netherlands.