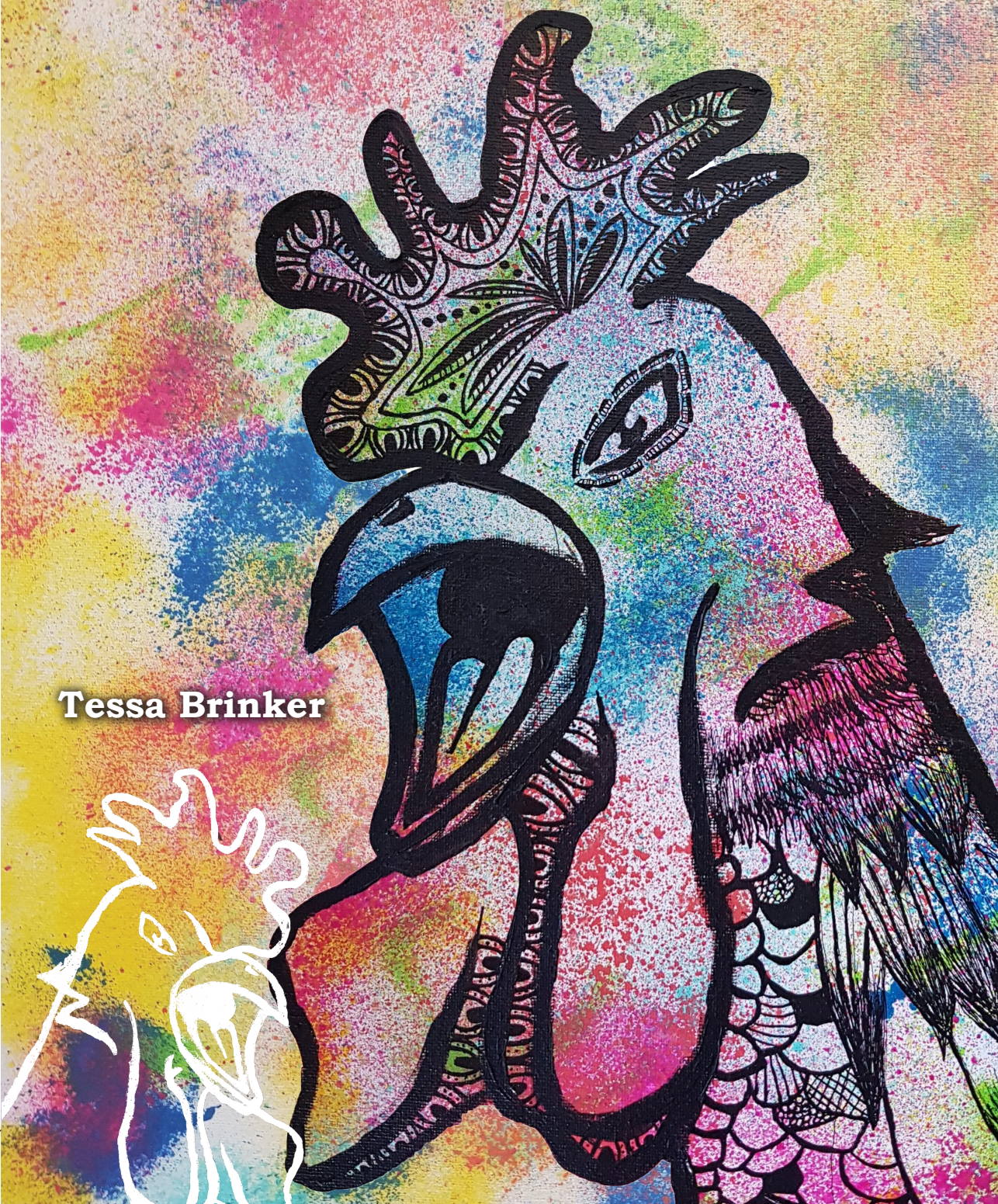


**Genetics and genomics of direct and social
genetic effects on survival time and
plumage condition in laying hens**

Tessa Brinker



**Genetics and genomics of direct and social
genetic effects on survival time and plumage
condition in laying hens**

Tessa Brinker

Thesis committee

Promotor

Prof. dr. ir. H. Bovenhuis
Professor, Animal Breeding and Genomics Centre
Wageningen University and Research

Co-promotors

Dr. P. Bijma
Assistant professor, Animal Breeding and Genomics Centre
Wageningen University and Research

Dr. E.D. Ellen
Post-doc, Animal Breeding and Genomics Centre
Wageningen University and Research

Other members

Dr. ir. J.E. Bolhuis, Wageningen University and Research
Dr. ir. E. Knol, Topigs Norsvin, NL
Prof. dr. C.H.J. van Oers, Wageningen University and Research
Prof. Dr. A. Wilson, University of Exeter, UK

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

Genetics and genomics of direct and social genetic effects on survival time and plumage condition in laying hens

Tessa Brinker

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday November 13, 2020
at 11.00 a.m. in the Aula.

Tessa Brinker

Genetics and genomics of direct and social genetic effects on survival time and plumage condition in laying hens, 188 pages.

PhD thesis, Wageningen University and Research, the Netherlands (2020)

With references, with summary in English

ISBN 978-94-6395-498-3

DOI <https://doi.org/10.18174/529196>

Abstract

With social interactions, individuals may affect each other's phenotype. In these cases, an individual's phenotype depends on the direct genetic effect of the individual itself (DGE) and the indirect genetic effects of its mates (IGE). This thesis focusses on two traits related to cannibalistic interactions among laying hens: plumage condition and survival time. This thesis shows that up to 94% of the heritable variation in plumage condition and up to 61% of the heritable variation in survival time relates to IGE. Thus, genetic selection methods incorporating both DGE and IGE offer perspectives to improve both traits. Shortcomings of studies using DGE-IGE models that focussed on survival time are that censored records were considered as exact lengths of life and models assumed that IGE were continuously expressed by all cage members, even after death. This may reduce accuracy of estimated breeding values (EBV). This thesis showed that the accuracy of EBV for survival time can be improved using models that incorporate censoring. This will also contribute to higher rates of genetic gain. Another tool to improve EBV accuracy could be to use genomic information instead of pedigree information. This thesis showed that using genomic information did not improve the accuracy of EBV for survival time. Genomic information is also used in Genome-Wide Association Studies (GWAS). Only a handful of studies extended GWAS to include IGE. In this thesis, SNPs associated with direct and indirect effects for survival time were identified. GWAS results revealed SNPs with large DGE and a link of DGE and IGE for survival time in laying hens with the GABAergic system. This supports existing evidence for the involvement of GABA in the development of abnormal behaviors. SNP effects can differ among crosses due to different levels of linkage disequilibrium between the SNP and the QTL in the parental lines. If this matters, the power of SNP detection is expected to increase when alleles are mapped specific to their allele origin. This thesis aimed to map DGE and IGE for survival time, while considering the line origin of the alleles. However, accurately assigning the line origin to alleles, appeared to be challenging: new Mendel errors were identified after assigning the line of origin to alleles (additional to the identified Mendel errors based on SNP genotypes). These errors had a large impact on the results of the GWAS analyses. Moreover, it is recommended to repeat the quality control based on Mendel errors after allele origin assignment. Finally, it is discussed that the topic of this thesis is very relevant in the light of sustainable egg production, because *i)* it addresses important societal concerns, and *ii)* it contributes to sustainable development goal "End hunger" (SDG2) of the Food and Agricultural Organization of the United Nations.

Contents

5	Abstract
9	1 – General introduction
23	2 – Plumage condition in laying hens: genetic parameters for direct and indirect effects in two purebred layer lines
43	3 – Predicting direct and indirect breeding values for survival time in laying hens using repeated measures
65	4 – Estimation of total genetic effects for survival time in crossbred laying hens showing cannibalism, using pedigree or genomic information
83	5 – The genetic architecture of socially-affected traits: a GWAS for direct and indirect genetic effects on survival time in laying hens showing cannibalism
103	6 – General discussion
123	Summary
131	References
147	Additional files
167	Curriculum vitae
173	Training and supervision plan
179	Acknowledgements
185	Colophon

1

General Introduction

1.1 Social interactions

Social interactions are an integral component of life. Interacting individuals may affect each other's characteristics. In both domestic and natural populations, social interactions occur. This includes, for example, disease transmission in *Eucalyptus globulus* (Costa e Silva et al., 2017), aggression in mink (Alemu et al., 2014), parental care in mice (Ashbrook et al., 2015), cooperation in ants (Linksvayer, 2006), growth competition in fungi (*Aspergillus nidulan*, Rode et al., 2017), and spouse effects in human (e.g. Evans et al., 2018).

Socially-affected traits are different from non-social interaction traits. With social interaction traits, there are two distinct effects: *i)* the effect of an individual on its own phenotype (direct effect) and *ii)* the effects of group mates on an individual's phenotype (indirect effect; Moore et al., 1997). Direct and indirect effects can partly be genetically determined, where the phenotype of an individual depends on both the individual's own genes (direct genetic effects; DGE) and the genes from its group mates (indirect genetic effects; IGE) (e.g. Bijma et al., 2007a; Bijma et al., 2007b; Griffing, 1967; Moore et al., 1997; Muir, 2005).

IGE have been quantified for numerous traits and populations, e.g. survival in chickens (Ellen et al., 2008; Peeters et al., 2012), aggression in field crickets (Han et al., 2018), and reproductive performance in pigs (Ki Hong et al., 2017). Estimates show that IGE can have an effect on the total heritable variation available for selection response. The total heritable variation is in fact higher or lower than what is predicted when DGE are considered alone. For example, the total heritable variation was revealed to be 30%-62% higher for survival time in laying hens when IGE were considered as well (Ellen et al., 2008; Peeters et al., 2012). An example of decrease in total heritable variation due to IGE can be found in *Eucalyptus Globulus*. The total heritable variation was reduced by 85% for growth when IGE were considered in the model (Costa e Silva et al., 2013).

This thesis focusses on the genetics of socially-affected traits in laying hens showing cannibalism. Cannibalism is one of the major welfare and economic issues occurring in the egg production sector. A common practice to prevent cannibalism is beak trimming. It has, however, received much criticism because it compromises the welfare of laying hens (Riber and Hinrichsen, 2017). Various European countries have, therefore, banned the beak trimming practice or plan to ban it in the near future. However, cannibalism will increase when hens have intact beaks and

solutions are required. Cannibalism is a multi-factorial problem caused by both environmental and animal related factors. Management solutions could be related to feed (Van Krimpen, 2005), light intensity (Kjaer and Vestergaard, 2009), provision of floor litter (Blokhuys, 1986), and group size (Bilčík and Keeling, 2000). However, there are no management solutions that can completely prevent cannibalism. Genetic selection could contribute to a more permanent solution that will allow to keep hens with intact beaks without negative effects of anti-social behaviour. In the following, I will identify research questions concerned with laying hens displaying cannibalism. I will do this following these relevant topics:

- i) The basic quantitative genetic principles of social interactions (Griffing, 1967; Muir, 2005; Bijma et al., 2007a).
- ii) The potential of methods using DNA information when assessing socially-affected traits.

1.2 Basic quantitative genetic principles of IGE

Intuitively, it makes sense to base genetic selection of socially-affected traits on behavioural observations. For example, with cannibalistic interactions in laying hens one could count the number of times an individual pecks its cage mate. This is, however, a time-consuming practice. Moreover, the collection of behavioural observations focusses on the individual performing the pecking, and disregards the victim of cannibalistic interactions. Previous research clearly showed that genetic solutions need to consider both the effect of the victim (direct genetic effect; DGE) and actor (indirect genetic effect; IGE) in order to reduce mortality due to cannibalism in laying hens (e.g. Ellen et al., 2008; Peeters et al., 2012; Alemu et al., 2016). Hence, selecting against cannibalistic interactions using behavioural observations will disregard part of the genetic variation.

Alternatively, a variance component model considers both DGE and IGE and it is not required to have knowledge about the traits underlying cannibalistic behaviour. Instead, traits that show the consequences of cannibalistic interactions are analysed. In this thesis, two traits that show the consequences of cannibalistic interactions are analysed using this variance component model: plumage condition (**Chapter 2**) and survival time (**Chapter 3-5**). Here, DGE and IGE are added as random effects to the model, and their (co-)variances are estimated based on family relationships in the dataset.

In **Figure 1.1**, the principle of DGE and IGE is illustrated following chicken 1 and its two group mates (chicken 2 and chicken 3). The phenotype of chicken 1 (P_1) is determined by its own direct effects (P_{D_1}) and indirect effects of its $n-1$ (in this example: 2) group mates ($P_{I_2} + P_{I_3}$). By substituting these components using the basic principles of genetics ($P = A + E$; see Box 1), one can see that $P_{D_1} = A_{D_1} + E_{D_1}$, and $P_{I_2} + P_{I_3} = A_{I_2} + E_{I_2} + A_{I_3} + E_{I_3}$. Hence, the phenotype of chicken 1 can be expressed as the sum of DGE, direct environmental effects, IGE of both its group mates, and indirect environmental effects of both its group mates. This theoretically shows that the social environment can contain a heritable component (IGE) which may be important for selection response. More details are in **Box 1.1**.

Box 1.1 Phenotype of socially-affected traits

The phenotype of an individual is defined as the combined effects of its genetic make-up and an environmental effect (Fisher, 1918).

$$P_i = A_i + E_i \quad (1)$$

where P_i is the phenotype of individual i , A_i the genetic effect of individual i , and E_i the environmental effect of individual i .

The environment can be broadly defined, containing everything from weather perturbances to consumed feed, but also the social environment. If the social environment contains a heritable component (*i.e.* IGE) the model can be defined as follows (Griffing, 1967; Muir, 2005; Bijma et al., 2007a):

$$P_i = \underbrace{A_{D_i} + E_{D_i}}_{\text{Direct Effects}} + \underbrace{\sum_{j \neq i}^{n-1} A_{I_j} + \sum_{j \neq i}^{n-1} E_{I_j}}_{\text{Indirect Effects}} \quad (2)$$

where A_{D_i} represents the DGE of individual i , E_{D_i} the direct environmental effect of individual i , A_{I_j} the IGE of group mate j , E_{I_j} the indirect environmental effect of individual j , and n the total number of individuals in a group (Griffing, 1967). Both the direct effects part ($A_{D_i} + E_{D_i}$) and the indirect effects part ($\sum_{j \neq i}^{n-1} A_{I_j} + \sum_{j \neq i}^{n-1} E_{I_j}$) are expressed in the phenotype of individual i .

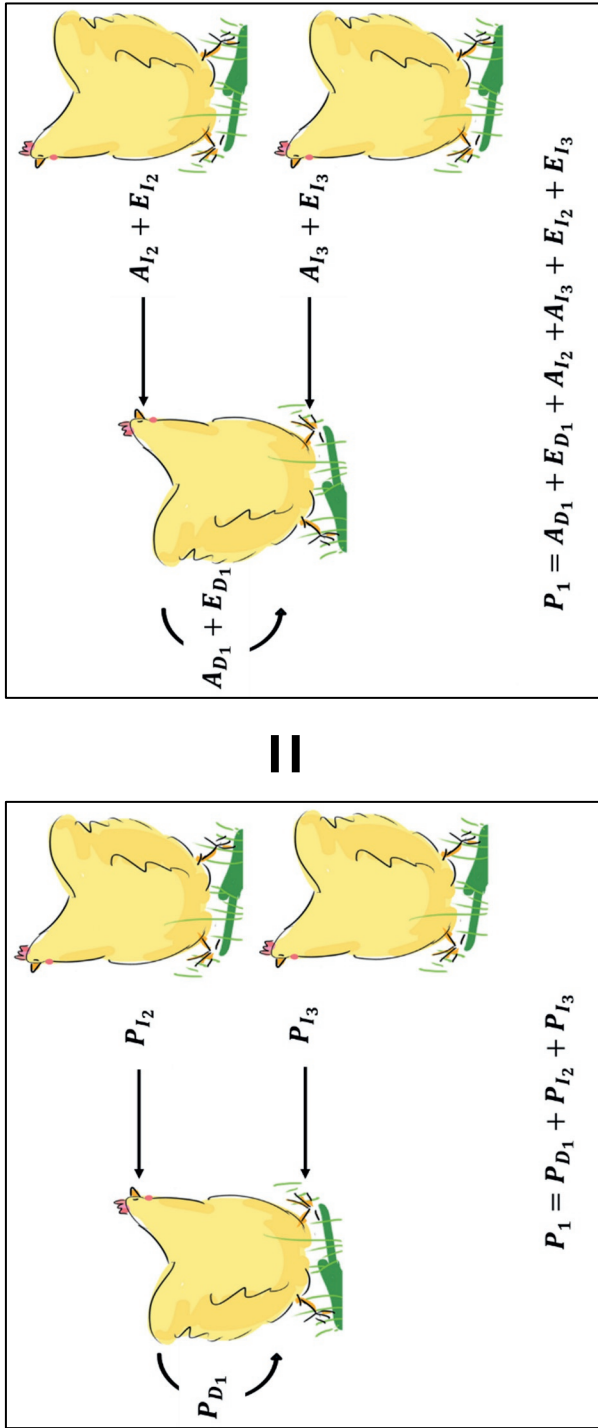


Figure 1.1 Indirect (genetic) effects explained: an example of social interactions in a group of three chickens. P represent the phenotype, subscript D refers to the direct effect, subscript I to the indirect effect, and numbers 1-3 refer to the corresponding individual. Furthermore, A is the genetic effect, and E the environmental effect.

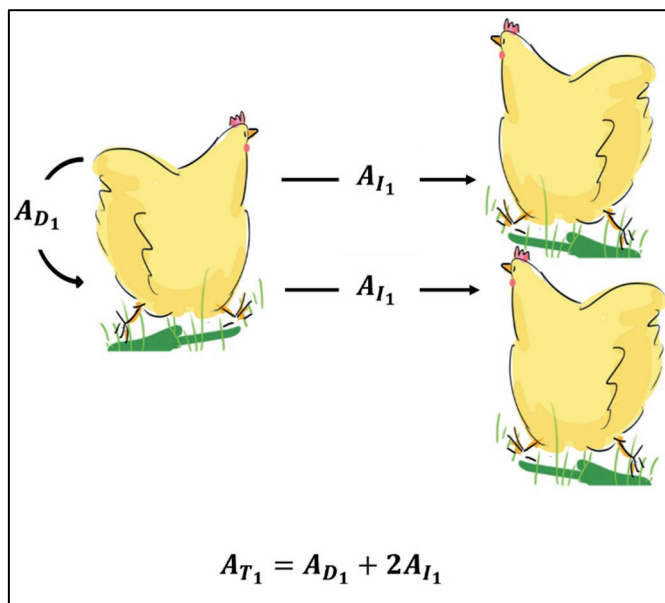


Figure 1.2 Total breeding value explained: an example of social interactions in a group of three chickens. A is the genetic effect, where subscript D refers to the direct effect, subscript I to the indirect effect, and subscript T to the total effect. Number 1 refers to individual 1.

1.2.1 Breeding values

To improve traits using genetic selection, animals are selected based on their estimated breeding values. Such estimated breeding values allow to compare individuals and choose the animals that have the best potential as parent to produce the next generation. With social interactions, the quantity relevant for selection response is the total breeding value. It is the linear combination of DGE and IGE.

Figure 1.2 illustrates the principle of the total breeding value of chicken 1. From the illustration it becomes clear that there are two distinct genetic traits expressed by chicken 1: the direct breeding value of chicken 1 (A_{D_1}) and two times an indirect breeding value (A_{I_1}). Note that the arrows from chicken 1 to its group mates in **Figure 1.2** are in the opposite direction of the arrows in **Figure 1.1**. Thus, the origin of the IGE differs depending on whether one considers the phenotype or the total breeding value, whereas the origin of the DGE is the same in both. More details are in **Box 1.2**.

Box 1.2. Total Breeding Value

With social interactions, selection should be based on the total breeding value (A_{T_i}):

$$A_{T_i} = A_{D_i} + (n - 1)A_{I_i} \quad (3)$$

where A_{T_i} is the total breeding value (also called total genetic effect) of individual i , A_{D_i} the direct breeding value of individual i (also called direct genetic effect, DGE), and A_{I_i} the indirect breeding value of individual i (also called indirect genetic effect, IGE) which is expressed $n-1$ times, *i.e.* in each group mate of i .

In order to achieve genetic progress, it is important that the total breeding value (**Box 1.2, Equation 3**) is estimated accurately. This will contribute to higher rates of genetic gain. Studies using DGE-IGE models have focussed on the trait survival time to reduce mortality due to cannibalism in laying hens. However, shortcomings of these DGE-IGE models are that censored records were considered as exact lengths of life and models assumed that IGE were continuously expressed by all cage members, irrespective of whether they were alive or dead. However, the cage composition changes when individuals die and IGE can no longer be expressed by this individual. Neglecting censoring and timing of IGE expression may reduce the accuracy of estimated breeding values. In **Chapter 3**, we considered four models to predict survival time in laying hens. One model was an analysis of survival time and the three other models treated survival in consecutive months as a repeated binomial trait. These three models can incorporate censoring and timing of IGE expression. The aim was to investigate whether the accuracy of estimated breeding values was improved using the repeated measures models compared to the model analysing survival time.

1.2.2 T^2 and associated parameters.

In classical genetic theory, the heritability is a measure that indicates which part of the phenotypic variation in a population is due to heritable variation. It is an indicator of a trait's potential to respond to selection. As an analogy, the total heritable variation ($\sigma_{A_T}^2$) can be expressed relative to the phenotypic variance (σ_P^2) of a social interaction trait, referred to as T^2 (Bergsma et al., 2008). The $\sigma_{A_T}^2$ can be larger or smaller than expected based on predictions considering DGE only. When the sum of $(n - 1)^2\sigma_{A_I}^2$ and $2(n - 1)\sigma_{A_D A_I}$ yields a positive value, the $\sigma_{A_T}^2$ will be larger compared to considering DGE alone. When this sum yields a negative value (*i.e.* when

$\sigma_{A_D A_I}$ is sufficiently negative), the $\sigma_{A_T}^2$ will be smaller compared to considering DGE alone. Also, the phenotypic variance is affected by IGE (**Box 1.3**).

Knowledge of these genetic parameters underlying a trait is required in order to estimate breeding values. In other words, it requires the separation of phenotypic (co-)variances into variance components, *e.g.* (in-)direct genetic (co-)variances. The availability of these parameters will, moreover, provide more insight in the potential contribution of IGE to selection response. In this thesis, genetic parameters of plumage condition (**Chapter 2**) and survival time (**Chapter 3 and Chapter 4**) are presented.

Box 1.3. T^2 and associated parameters

T^2 is an analogy of the heritability and is expressed as to total heritable variation relative to the phenotypic variance:

$$T^2 = \frac{\sigma_{A_T}^2}{\sigma_P^2} \quad (4)$$

with,

$$\sigma_{A_T}^2 = \sigma_{A_D}^2 + 2(n-1)\sigma_{A_D A_I} + (n-1)^2\sigma_{A_I}^2 \quad (5)$$

where, $\sigma_{A_D}^2$ is the direct genetic variance, $\sigma_{A_D A_I}$ the direct-indirect genetic covariance, and $\sigma_{A_I}^2$ the indirect genetic variance. The $\sigma_{A_D A_I}$ is a measure for the relationship between DGE and IGE. For example, if individuals that have a negative effect on their group mates on average have better performance, then the $\sigma_{A_D A_I}$ will be negative.

Also, phenotypic variance is affected by IGE:

$$\sigma_P^2 = \sigma_{A_D}^2 + \sigma_{E_D}^2 + (n-1)(\sigma_{A_I}^2 + \sigma_{E_I}^2) + r(n-1)(2\sigma_{A_D A_I} + (n-2)\sigma_{A_I}^2) \quad (6)$$

where, $\sigma_{E_D}^2$ is the direct environmental variance, $\sigma_{E_I}^2$ the indirect environmental variance, and r the average genetic relatedness in a group.

1.3 DNA information (SNP Genotypes)

Genotyping is the process of determining the DNA base pairs for a certain number of locations on the genome. It has become a common practice in animal breeding. This technique identifies small variations in the DNA within populations, *e.g.* single-nucleotide polymorphisms (SNPs). SNPs are single base-pair changes in the DNA that occur at specific places.

Knowledge of SNP genotypes provides:

- i) New opportunities to estimate genetic parameters when traits are affected by social interactions.
- ii) Opportunities to perform genomic selection: a form of genetic selection where all SNP markers are used to select for a specific trait. It relies on the idea that all quantitative trait loci (QTL) are in linkage disequilibrium with at least one SNP marker.
- iii) Opportunities to perform Genome-Wide Association Studies (GWAS): detect associations between markers across the DNA and traits of interest with the possibility to identify causal genes that reveal more about the underlying biology of a specific phenotype. This could also be useful in genomic selection, *e.g.* giving more weight to the relevant SNPs.

1.3.1 New opportunities to estimate genetic parameters

Separate estimation of DGE and IGE is not always possible. Using pedigree information, DGE and IGE can only be estimated when hens are housed in groups composed of multiple families (Bijma, 2011; Ellen et al., 2008; Muir, 2005). Layer breeding organizations, however, often use recurrent testing, where hens are housed in sire-family groups and dam pedigree is unknown. In this scenario, DGE and IGE are fully confounded. In this case, a sire model can be used to estimate the total breeding value of the sire, which is the linear combination of the sire DGE and the sire IGE (Peeters, 2015).

Thus, there are two issues:

- The dam pedigree is unknown.
- Separate estimation of DGE and IGE is not possible using pedigree information.

Genomic information may solve both issues. First, genomic information will provide the opportunity to reconstruct the dam pedigree. With this reconstructed pedigree,

it is possible to distinguish full-sibs from half-sibs. This, in turn, may allow the separate estimation of DGE and IGE, even when a group has the same sire (but not the same dam). Second, genomic information provides information on the actual genetic relationship between individuals (VanRaden, 2008; Yang et al., 2010). Actual genetic relationships between individuals vary around their expected value based on pedigree information because of linkage and Mendelian sampling (Hill & Weir, 2011). It is expected that total breeding value prediction can be improved using the actual genetic relationships calculated from genomic data. (Meuwissen, 2007).

Thus, genomic information may provide new opportunities for estimation of genetic parameters when traits are affected by social interactions. In **Chapter 4** it was, therefore, investigated whether DGE and IGE of survival time can be estimated separately for crossbred laying hens housed in sire-family groups using genomic information.

1.3.2 Genomic selection

Many traits have successfully been improved using selection based on phenotypes and pedigree information. Genetic progress depends on the selection intensity, the accuracy of selection, the genetic variation, and the generation interval. However, genetic progress has been limited by a long generation interval or limited accuracy for traits that are expressed late in life, are difficult to measure (e.g. behavior), or traits that cannot be measured on selection candidates (sex-limited traits like egg production; Meuwissen et al., 2001; Muir, 2007).

Increasing the accuracy of selection and selection intensity, and reducing the generation interval can enhance genetic progress. Genomic data provides opportunities for this because:

- Genomic data provides information on actual genetic relationships between individuals rather than expected genetic relationships using pedigree information (Yang et al., 2010). It is, therefore, expected that the accuracy of selection will increase compared to using pedigree information.
- Genomic data is available early in life, which makes it possible to differentiate between individuals at young age, regardless of sex or lack of phenotypic information. This could reduce the generation interval for traits that are difficult to measure or measured late in life. Survival time and plumage condition (**this thesis**) are typical examples of traits that are measured late in life.

For this reason, using genomic selection is increasingly implemented in laying hen breeding. For socially-affected traits like plumage condition and survival time in laying hens, it is clear that both DGE and IGE are important (e.g. Ellen et al., 2008; Peeters et al., 2012; Alemu et al., 2016). Genomic selection methods should, therefore, be extended to include IGE as well.

For genomic selection, a reference population is required. Both phenotypes and genotypes are available for individuals in a reference population. In order to perform genomic selection on survival time and/or plumage condition in laying hens, a specific reference population needs to be established. Namely, purebred hens that are kept to collect phenotypes on eggs, are housed in single-bird cages. However, cannibalistic interactions only occur in groups. Hence, setting up a reference population may be challenging.

Alternatively, family groups could be used to estimate total genomic breeding values. In **Chapter 4**, we used recurrent test data, where crossbred hens are kept in sire-family groups. The genotype of the sires was available and we could, therefore, fit a genomic selection model that yields sire total genomic breeding values. The aim of **Chapter 4** was to investigate whether genomic information of the sires increased the accuracy of total breeding values compared with using pedigree information. The chapter focussed on survival time in laying hens showing cannibalism.

1.3.3 Genome-Wide Association studies

Genome-Wide Association Studies (GWAS) are widely applied to a variety of traits and populations, but have been mainly focussed on DGE. Results from GWAS on DGE show that most quantitative traits in livestock are highly polygenic, and that variants tend to be associated with more than one trait (pleiotropy; Hill et al., 2008). The genetic architecture of IGE may, however, differ from the genetic architecture of DGE. With IGE, the phenotype of other individuals are affected rather than the phenotype of the individual itself (*i.e.* DGE). Selection targets such IGE only in the presence of feed-back mechanisms, using *e.g.* group selection or kin selection (Bijma and Wade, 2008). IGE may, therefore, be less exposed to natural selection or artificial selection when not accounting for social interactions. In other words, selection has not exhausted variation due to IGE. It is, therefore, expected that some loci may have large IGE (Denison et al., 2013).

However, only a handful of studies extended GWAS to include IGE (e.g. Biscarini et al., 2010; Mutic and Wolf, 2007; Baud et al., 2018). Biscarini et al. (2010) conducted a GWAS using only 1022 SNPs. They identified 81 SNPs for IGE associated with plumage condition in laying hens. The number of observations used in Biscarini et al. (2010) was also limited; 662 laying hens originating from 9 lines were used for analyses. Mutic and Wolf (2007) found 13 QTL for IGE associated with size, developmental, and fitness related traits in *Arabidopsis*. They analysed SNPs for IGE that were already associated with DGE to maximize the power of IGE detection, but ignored potential other SNPs with IGE. More recently Baud et al. (2018), performed a GWAS on 170 different behavioural, physiological, and morphological phenotypes in nearly 2000 mice. Among those 170 phenotypes, authors identified 21 SNPs associated with IGE for 17 phenotypes. With the exception of Baud et al. (2018), the genetic architecture of social interaction traits remains largely unknown.

In **Chapter 5**, the aim was, therefore, to identify SNPs associated with direct and indirect effects for survival time in laying hens that show cannibalism. Moreover, GWAS results from analyses of survival time versus using repeated measures models (**Chapter 3**) were compared.

1.4 Summary thesis outline

This thesis is part of the STW-project entitled “Genomic solutions for socially-affected traits: Genetic architecture and improvement of survival in cannibalistic laying hens”. The aim of the project was to increase our understanding of the genetic architecture of socially-affected traits. This thesis focussed on the genetics of socially-affected traits in laying hens and can be separated in roughly three sections and two traits (survival and plumage condition):

1. Genetic parameter estimation (**Chapter 2-4**)
2. Breeding value prediction (**Chapter 3-4**).
3. Genome Wide Association Study (**Chapter 5**).

Moreover, in the general discussion (**Chapter 6**), I address the following three topics:

1. Scientific relevance: here I discuss the most important findings of this thesis.
2. Data quality issues: here I discuss the use of phased data in GWAS on crossbred layers.
3. Societal relevance: here I discuss the importance of including DGE and IGE in layer breeding programs to reduce mortality due to cannibalism in laying hens.

2

Plumage condition in laying hens: genetic parameters for direct and indirect effects in two purebred layer lines

Tessa Brinker¹, Piter Bijma¹, Jeroen Visscher², T Bas Rodenburg³, Esther D Ellen¹

¹Animal Breeding and Genomics Centre, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands; ²Institut de Sélection Animale B.V., a Hendrix Genetics Company, P.O. Box 114, 5830 AC Boxmeer, The Netherlands; ³Behavioural Ecology Group, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

Abstract

Feather pecking is a major welfare issue in the laying hen industry that leads to mortality. Due to a ban on conventional cages in the EU and on beak trimming in some countries of the EU, feather pecking will become an even bigger problem. Its severity depends both on the victim receiving pecking and on its group mates inflicting pecking (indirect effects), which together determine plumage condition of the victim. Plumage condition may depend, therefore, on both the direct genetic effect of an individual itself and on the indirect genetic effects of its group mates. Here, we present estimated genetic parameters for direct and indirect effects on plumage condition of different body regions in two purebred layer lines, and estimates of genetic correlations between body regions.

Feather condition scores (FCS) were recorded at 40 weeks of age for neck, back, rump and belly and these four scores were added-up into a total FCS. A classical animal model and a direct-indirect effects model were used to estimate genetic parameters for FCS. In addition, a bivariate model with mortality (0/1) was used to account for mortality before recording FCS. Due to mortality during the first 23 weeks of laying, 5363 (for W1) and 5089 (for WB) FCS records were available.

Total heritable variance for FCS ranged from 1.5% to 9.8% and from 9.8% to 53.6% when estimated respectively with the classical animal and the direct-indirect effects model. The direct-indirect effects model had a significantly higher likelihood. In both lines, 70% to 94% of the estimated total heritable variation in FCS was due to indirect effects. Using bivariate analysis of FCS and mortality did not affect estimates of genetic parameters. Genetic correlations were high between adjacent regions for FCS on neck, back, and rump but moderate to low for belly with other regions.

Our results show that 70% to 94% of the heritable variation in FCS relates to indirect effects, indicating that methods of genetic selection that include indirect genetic effects offer perspectives to improve plumage condition in laying hens. This, in turn could reduce a major welfare problem.

Key words: laying hens, feather score, indirect genetic effects, plumage condition

2.1 Introduction

Feather pecking (FP) is a major welfare issue in commercial laying hens. Depending on the severity of the pecking it can result in feather loss or damage, or skin or muscle injuries (Savory, 1995). If the latter results in death, FP is referred to as cannibalism (Blokhuys and Arkes, 1984). When hens are not beak trimmed, the incidence of cannibalistic FP is higher in non-cage systems than in cage systems (Fossum, 2009). Since the beginning of 2012, cage systems have been prohibited in the European Union and the problem of FP is expected to increase (Rodenburg et al., 2012). Therefore, it is necessary to find a solution to prevent or reduce FP, especially when beak treatments are banned or will be banned in the future.

FP is a multi-factorial problem caused by both animal-related and environmental factors (Hughes and Duncan, 1972). A common procedure to limit the consequences of FP is beak trimming. There are plans to ban beak trimming since this procedure has welfare implications, such as chronic pain (Gentle et al., 1990). In some European countries, beak trimming is already prohibited or regulated (Van Horne and Achterbosch, 2008). Other management solutions could be feed-related (Van Krimpen et al., 2005). Furthermore, it has been shown that FP behaviour can be influenced by light intensity (Kjaer and Vestergaard, 1999), provision of floor litter (Blokhuys, 1986), group size (Bilčík and Keeling, 2000), and stocking density (Nicol et al., 1999). So far, there are no management solutions that can completely prevent FP. An additional measure to reduce FP is genetic selection (Rodenburg et al., 2010; Kjaer and Sørensen, 2001).

FP is a social interaction phenomenon, that involves both the victim that receives the pecking and its group mates that inflict the pecking (Buitenhuis et al., 2003; Biscarini et al., 2010; Keeling et al., 2004; Riber and Forkman, 2007). When traits are affected by social interactions among group members, the genetic effects that underlie individual phenotypes can be partitioned into a direct genetic effect (DGE) of the genotype of the individual itself, and the indirect genetic effects (IGE) of the genotypes of its group mates (Griffing, 1967; Moore et al., 1997; Wolf et al., 1998; Bijma et al., 2007a; Bijma et al., 2007b; Muir, 2005). IGE can contribute to the heritable variation in a trait. For example, in laying hens showing cannibalistic behaviour, IGE contribute 33% to 87% of the total heritable variation in survival time (Ellen et al., 2008; Peeters et al., 2012). Genetic selection for survival time, using a selection method that takes IGE into account, could reduce FP behaviour. Unfortunately, survival time is only known at the end of the laying period.

Therefore, it is necessary to define a trait that can be collected early in the laying period and that is correlated with FP behaviour.

To measure FP, behavioural observations are needed. However, several authors have used a feather condition score (FCS) to assess plumage condition as an alternative to behavioural observations (Hughes and Duncan, 1972; Allen and Perry, 1975; Bilčík and Keeling, 1999). Damage to the plumage is strongly related to the incidence of severe FP behaviour (Bilčík and Keeling, 1999). One disadvantage of using FCS instead of direct observations of pecking behaviour is that one can only detect the victim of FP and not the animal that actually inflicts the pecking. However, using methods that take IGE into account allows us to estimate both the breeding value for pecker-effect (the IGE) and for the victim-effect (the DGE) in group-housed laying hens (Muir, 2005; Bijma, 2010a). Estimating a breeding value for the pecker-effect is less accurate than using behavioural observations but, in animal breeding, it is not feasible to record a behavioural observation for each individual but it is possible to estimate DGE and IGE for each individual.

So far, most studies that focus on the plumage condition of laying hens have ignored the effect of group mates. However, it is expected that IGE contribute significantly to the heritable variation of plumage condition in laying hens, especially in the case of damaging behaviour such as FP. To improve plumage condition in laying hens, knowledge of the genetic parameters for both direct and indirect effects is required.

In the present study, genetic parameters of plumage condition in laying hens were estimated using a classical animal model and a model that combined both direct and indirect genetic effects. Plumage condition was measured on four body regions; neck, back, rump and belly using the FCS. In addition, genetic correlations between the different body regions were estimated.

2.2 Methods

For this study, data from the experiment that is described in Ellen et al. (2008) were used except that FCS were used instead of survival time data. The main characteristics are summarized below. Further details are in Ellen et al. (2008).

2.2.1 Population and pedigree

Data were collected under control of Institut de Sélection Animale B.V., the layer breeding division of Hendrix Genetics. Hendrix Genetics complies with the Dutch law on animal wellbeing. Data on two purebred White Leghorn layer lines were provided by the Institut de Sélection Animale B.V.. The two lines were coded W1 and WB. Data from line WF were not used in this study because fewer observations were available for this line and its mortality due to cannibalism was low (in comparison with the other two lines), which would lead to inaccurate genetic parameters for direct and indirect effects of survival time (Ellen et al., 2008).

Within a line, sires and dams were mated at random. Matings were done in two batches with a six-month time period between the two batches. Sires used for both batches were largely the same (89% for line W1 and 94% for line WB), while dams were all different. For each batch, sires (36 of line W1 and 35 of line WB) were mated to approximately eight dams each, and each dam contributed on average 12.3 female offspring. For both lines, observations from a single generation were used. Chickens of both lines were hatched in two batches and each batch consisted of four consecutive age groups that differed by a two-week period each. All 12 192 chickens had intact beaks.

2.2.2 Housing

When the hens were on average 17 weeks old, they were transported to two laying houses with traditional four-bird-battery cages. The two batches were each placed in separate laying houses, termed 1 and 2. In both laying houses, the 17-week-old hens were randomly allocated to laying cages, with four birds of the same line and age in a cage. The individuals making up a cage were combined at random, without taking size of the hens into account. In both laying houses, cages were grouped into eight double rows. Each row consisted of three levels (top, close to the light; middle; and bottom). A feeding trough was in front of the cages, and each pair of back-to-back cages shared two drinking nipples. A standard commercial layer diet and water were provided *ad libitum*.

In both laying houses, the hens started with a light period of 9 hours/day. Every week the light period was increased by 1 hour until 16 hours/day, when the hens were on average 26 weeks of age.

2.2.3 Data

Plumage condition was measured at 40 weeks of age on all hens alive. It was measured at eight time points, each separated by a two-week period, starting with the oldest hens, so that all birds were measured at the same age. To quantify plumage condition, the feather condition score (FCS) described in Bilčík and Keeling (1999) was used, as modified by Uitdehaag et al. (2008). The body of the hen was divided into four regions: neck, back, rump, and belly. These body regions were chosen because they are expected to be the regions to receive the largest number of pecks and plumage condition in those regions is less affected by abrasion (Bilčík and Keeling, 1999). Each body region was inspected and given a score from 0 (intact feathers) to 5 (completely denuded area). For further analysis, observations with a score 0 or 1 were combined into the score 1 class. FCS were recorded by four persons. Before performing the FCS, a single set of 153 birds were scored by the four persons to estimate the correlation between scores of different persons (between-observer correlation). The between-observer correlation ranged from 0.84 to 0.94 for neck, back, and rump but from 0.66 to 0.83 for belly (E.D. Ellen, unpublished data).

All hens were observed daily. Dead hens were removed, and wing band number, cage number, and cause of death were recorded. Cause of death was determined subjectively without dissection. Thus for all hens in the dataset, their status alive (0) or dead (1) was known at the time of feather scoring. A total of 12 192 hens were present in the dataset composed of 5920 hens of line W1 and 6272 hens of line WB. Due to mortality during the first 23 weeks of laying, FCS were unavailable for 9.4% of the W1 hens and for 18.9% of the WB hens, which resulted in 5363 FCS records for line W1 and 5089 for line WB.

2.2.4 Data analysis

Model

A preliminary data analysis was performed using the SAS statistical program (SAS, 2004). The GLM procedure was used to identify significant fixed effects to be included in the model for subsequent analysis. Analysis of FCS was done for each line and body region separately. The four body regions were summed into a total FCS, which was also analysed. The most significant fixed effects identified were the interaction between laying house-row-level and the person carrying out the scoring. Age and batch were fully confounded with laying house and row, and therefore not included in the model.

A linear animal model was used in ASReml to estimate genetic parameters for FCS (Gilmour et al., 2009). First, genetic parameters were estimated by using a classical animal model,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{V}\mathbf{c} + \mathbf{e}, \quad (1)$$

in which \mathbf{y} is a vector of observed FCS; \mathbf{b} is a vector of fixed effects, with incidence matrix \mathbf{X} linking observations to the fixed effects; \mathbf{a} is a vector of the breeding values, with incidence matrix \mathbf{Z} linking the observations on individuals to their breeding value; \mathbf{c} is a vector of independent random cage effects, with incidence matrix \mathbf{V} linking the observations to the random cage effect; and \mathbf{e} is a vector of random residuals. The variance structure of the model terms are: $\text{var}[\mathbf{a}] = \mathbf{A}\sigma_A^2$, $\text{var}[\mathbf{c}] = \mathbf{I}\sigma_c^2$, and $\text{var}[\mathbf{e}] = \mathbf{I}\sigma_e^2$. Matrix \mathbf{A} is the matrix of additive genetic relationships between individuals based on five generations of pedigree, σ_A^2 the genetic variance, \mathbf{I} an identity matrix, σ_c^2 the cage variance, and σ_e^2 the residual variance. To avoid pedigree errors, hens with an unknown or double identification were coded as having an unknown pedigree ($n=63$). The observations on these hens were included in the analysis to better estimate fixed effects.

Second, genetic parameters were estimated for both the direct and indirect genetic effects using a direct-indirect effects model (Muir, 2005; Bijma et al., 2007b),

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{V}\mathbf{c} + \mathbf{e}, \quad (2)$$

where \mathbf{a}_D is a vector of direct breeding values, with incidence matrix \mathbf{Z}_D linking observations on individuals to their direct breeding value, \mathbf{a}_S is a vector of indirect breeding values, with incidence matrix \mathbf{Z}_S linking observations on individuals to the indirect breeding values of their group mates (the other three individuals in the same cage; see (Muir, 2005)). The covariance structure of the genetic terms is $\text{var} \begin{bmatrix} \mathbf{a}_D \\ \mathbf{a}_S \end{bmatrix} = \mathbf{C} \otimes \mathbf{A}$, where \otimes is the Kronecker product of matrices, and $\mathbf{C} = \begin{bmatrix} \sigma_{A_D}^2 & \sigma_{A_{DS}} \\ \sigma_{A_{DS}} & \sigma_{A_S}^2 \end{bmatrix}$, where $\sigma_{A_D}^2$ is the direct genetic variance, $\sigma_{A_S}^2$ is the indirect genetic variance, and $\sigma_{A_{DS}}$ is the direct-indirect genetic covariance.

Model comparison

The classical animal model (Equation 1) and the direct-indirect effects model (Equation 2) were statistically compared using a log-likelihood ratio test. The

2 Genetic parameters for plumage condition

classical animal model was compared with a model without random effects (null model) and a model with only a random cage effect to test additive genetic variance. The direct-indirect effects model was compared with the classical animal model to test indirect genetic (co)variance.

Heritable variation

In the classical animal model, the heritability is the ratio of heritable variance (σ_A^2) and phenotypic variance (σ_P^2):

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}. \quad (3)$$

In the direct-indirect effects model, the total heritable variance (σ_{TBV}^2) available for response to selection is $\sigma_{TBV}^2 = \sigma_{A_D}^2 + 2(n-1)\sigma_{A_{DS}} + (n-1)^2\sigma_{A_S}^2$, where n is the number of individuals in a group (Bijma et al., 2007a). Phenotypic variance is $\sigma_P^2 = \sigma_{A_D}^2 + (n-1)\sigma_{A_S}^2 + \sigma_e^2$. The term $(n-1)$ in both expressions refers to the $n-1$ group mates with which the individual interacts. For socially affected traits, the ratio of total heritable variance and phenotypic variance (T^2) is:

$$T^2 = \frac{\sigma_{TBV}^2}{\sigma_P^2}. \quad (4)$$

A comparison of h^2 and T^2 reflects the impact of IGE on heritable variation.

Selection bias

Estimates of genetic parameters for single traits such as FCS can be biased when the data represent a selected subset of the population (Pollak et al., 1984). In lines W1 and WB, the percentage of animals that died before FCS was recorded was 9.4% and 18.9% respectively. These dead animals can bias the estimated genetic parameters, since they are expected to have a higher FCS (more damage). Such selection bias can be reduced by using multiple-trait analysis (Pollak et al., 1984). For this reason, a bivariate analysis (Equation 5), including both FCS and mortality at 40 weeks of age (0/1) was applied to both the classical animal model and the direct-indirect effects model. In the bivariate analysis, the model for mortality at 40 weeks of age (0/1) included only a DGE, since models that included both DGE and IGE failed to converge.

Genetic correlations

Genetic correlations between the different body regions were estimated using a classical animal and the direct-indirect effects model. To estimate genetic correlations using the direct-indirect effects model, the bivariate model of Peeters et al. (2012) was used,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{1D} & 0 \\ 0 & \mathbf{Z}_{2D} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1D} \\ \mathbf{a}_{2D} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{1S} & 0 \\ 0 & \mathbf{Z}_{2S} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1S} \\ \mathbf{a}_{2S} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_1 & 0 \\ 0 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \quad (5)$$

where subscripts 1 and 2 denote FCS on two different body regions, e.g. neck and back. All other terms are the same as for Equation 2. The corresponding covariance

structure of the genetic terms is $\text{var} \begin{bmatrix} \mathbf{a}_{1D} \\ \mathbf{a}_{2D} \\ \mathbf{a}_{1S} \\ \mathbf{a}_{2S} \end{bmatrix} = \mathbf{C} \otimes \mathbf{A}$, with

$$\mathbf{C} = \begin{bmatrix} \sigma_{A_{1D}}^2 & \sigma_{A_{12D}} & \sigma_{A_{1DS}} & \sigma_{A_{1D^2S}} \\ \sigma_{A_{12D}} & \sigma_{A_{2D}}^2 & \sigma_{A_{2D^1S}} & \sigma_{A_{2DS}} \\ \sigma_{A_{1DS}} & \sigma_{A_{2D^1S}} & \sigma_{A_{1S}}^2 & \sigma_{A_{12S}} \\ \sigma_{A_{1D^2S}} & \sigma_{A_{2DS}} & \sigma_{A_{12S}} & \sigma_{A_{2S}}^2 \end{bmatrix}.$$

Thus there are four genetic variances and six genetic covariances; $\sigma_{A_{12D}}$ and $\sigma_{A_{12S}}$ are the direct and indirect genetic covariances between two body regions, $\sigma_{A_{1DS}}$ and $\sigma_{A_{2DS}}$ are the genetic covariances between direct and indirect effects of one of the body regions, and $\sigma_{A_{1D^2S}}$ and $\sigma_{A_{2D^1S}}$ are the genetic covariances between the direct effect of one body region and the indirect effect of another body region. For all body regions and lines, genetic correlations were estimated between the DGE (r_{12D}), the IGE (r_{12S}), and the TBV (total breeding value) (r_{12T}). Correlation r_{12T} depends on the total heritable variance within body regions ($\sigma_{TBV_1}^2$ and $\sigma_{TBV_2}^2$), and the total genetic covariance between body regions ($\sigma_{TBV_{12}}$) and is given by

$$r_{12T} = \frac{\sigma_{A_{12D}} + (n-1)\sigma_{A_{1D^2S}} + (n-1)\sigma_{A_{2D^1S}} + (n-1)^2\sigma_{A_{12S}}}{\sqrt{(\sigma_{A_{1D}}^2 + 2(n-1)\sigma_{A_{1DS}} + (n-1)^2\sigma_{A_{1S}}^2)(\sigma_{A_{2D}}^2 + 2(n-1)\sigma_{A_{2DS}} + (n-1)^2\sigma_{A_{2S}}^2)}} \quad (\text{Peeters et al., 2012}).$$

In the bivariate analyses, the random cage effect and residual were also allowed to be correlated between body regions.

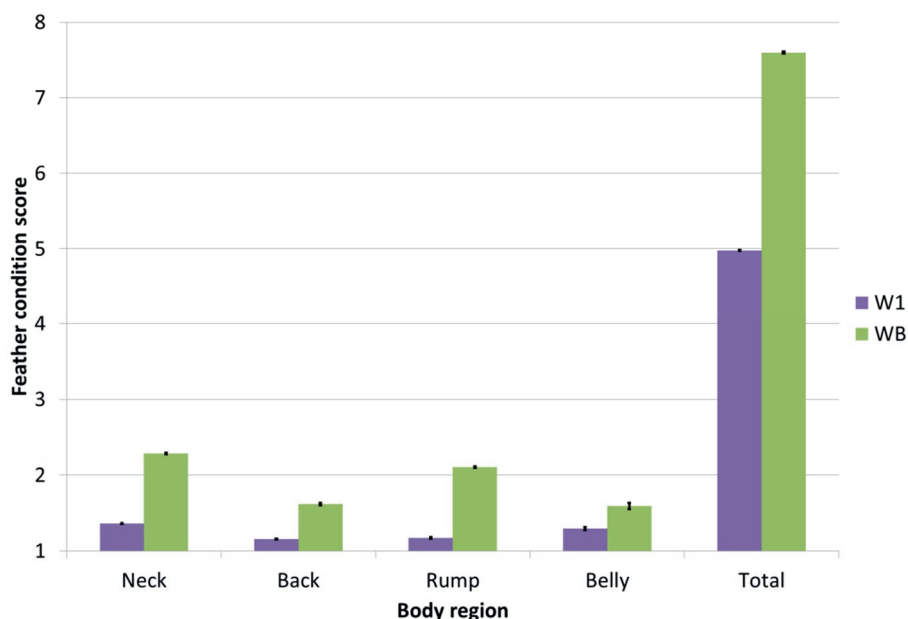


Figure 2.1 FCS of each body region and total FCS for two lines (W1,WB), with corresponding SE.

2.3 Results and Discussion

2.3.1 Feather score

Average plumage condition differed between body regions (**Figure 2.1**) and between the two lines ($p < 0.001$). Overall, line WB yielded the highest average FCS (worst plumage condition), ranging from 1.6 (belly) to 2.3 (neck). Average FCS by region for line W1 ranged from 1.1 (rump) to 1.4 (neck). This is in line with results of Ellen et al. (2008), who found that average survival was lowest in line WB. Overall, the plumage condition was worst for the neck region and best for the back region. This is in contradiction with results of Bilčík and Keeling (1999) who used the same scoring method, but, in their study, hens were kept in groups of 15 birds. They found that plumage condition for the hybrid layer line Hisex white was worst for the belly region, and best for the neck and back regions. In addition, they showed that, although belly had the worst plumage condition, most peaks were targeted at the rump and back. These discrepancies with our results might be due to differences in line, age, and housing conditions such as density, size of groups, and light intensity. One major difference between the studies is that Bilčík and Keeling birds were housed in floor pens, while in the present study birds were

housed in conventional cages. In conventional cages, much higher levels of abrasion of neck feathers are observed, due to contact with the cage door while feeding. The abrasion of neck feathers could stimulate the feather pecking behaviour of group mates, which can result in higher levels of neck damage.

The FCS was significantly different between the two laying houses, except for belly ($p = 0.055$). In laying house 2, back and rump had the lowest FCS, whereas in laying house 1 neck had the lowest FCS. Line WB had the worst plumage condition in both laying houses. Furthermore, significant differences in FCS between the three levels of each row and between the corridors were found (**Table 2.1**), except for the effect of level on rump, belly, and total FCS ($p > 0.05$). Overall, the lowest FCS (best plumage condition) were recorded on birds located on the bottom level, whereas the highest FCS were recorded on birds on the top and middle levels. Upper levels had higher light intensities, which stimulate FP behaviour and thus result in higher FCS (worst plumage condition) (Hughes and Duncan, 1972; Ellen et al., 2008). Overall, the highest FCS were observed for rows 7 and 8 without any clear explanation.

Table 2.1 Number of hens (n) and least square means (\pm SE) for FCS

	n	Neck	Back	Rump	Belly	Total
Laying house						
1	7376	1.80 \pm 0.01	1.41 \pm 0.01	1.75 \pm 0.01	1.44 \pm 0.01	6.40 \pm 0.03
2	4816	1.89 \pm 0.02	1.36 \pm 0.01	1.50 \pm 0.02	1.47 \pm 0.01	6.21 \pm 0.05
Level						
Top ¹	2400	1.87 \pm 0.02	1.39 \pm 0.02	1.64 \pm 0.02	1.46 \pm 0.02	6.36 \pm 0.06
Middle	4772	1.85 \pm 0.01	1.41 \pm 0.01	1.64 \pm 0.01	1.44 \pm 0.01	6.33 \pm 0.04
Bottom	5020	1.81 \pm 0.01	1.36 \pm 0.01	1.60 \pm 0.01	1.46 \pm 0.01	6.23 \pm 0.04
Row						
1	1096	1.69 \pm 0.03	1.19 \pm 0.03	1.29 \pm 0.03	1.32 \pm 0.02	5.49 \pm 0.09
2	1316	1.57 \pm 0.03	1.15 \pm 0.02	1.28 \pm 0.03	1.41 \pm 0.02	5.41 \pm 0.08
3	1524	1.56 \pm 0.02	1.25 \pm 0.02	1.41 \pm 0.03	1.35 \pm 0.03	5.58 \pm 0.07
4	1648	1.60 \pm 0.02	1.23 \pm 0.02	1.38 \pm 0.02	1.27 \pm 0.03	5.48 \pm 0.07
5	1632	1.83 \pm 0.02	1.42 \pm 0.02	1.79 \pm 0.02	1.52 \pm 0.03	6.57 \pm 0.07
6	1616	1.93 \pm 0.02	1.45 \pm 0.02	1.79 \pm 0.03	1.65 \pm 0.03	6.82 \pm 0.07
7	1700	2.11 \pm 0.02	1.63 \pm 0.02	2.07 \pm 0.02	1.62 \pm 0.03	7.44 \pm 0.07
8	1660	2.43 \pm 0.02	1.78 \pm 0.02	1.99 \pm 0.03	1.49 \pm 0.03	7.68 \pm 0.07

Least square means are shown for the fixed effect levels in the model for analysis of FCS of different body regions (neck, back, rump, belly) and total FCS. FCS is feather condition score, for each body region FCS ranges from 1 to 5, and total FCS ranges from 4 to 20; ¹only laying house 1.

2.3.2 Genetic parameters

Classical animal model

Table 2.2 shows the results of the likelihood ratio test. For all lines and body regions, including a random cage effect or a random animal effect improved the goodness of fit significantly (all P-values < 0.001 or < 0.01, respectively). **Table 2.3** and **Additional file 2.1: Table S2** show the estimated genetic parameters obtained with the classical animal model. Using univariate analysis, heritabilities (h^2) for FCS of the different regions and for total FCS were low, ranging from 1.5 to 9.8%. In previous studies, h^2 estimates for total FCS ranged from 22 to 54% (Craig and Muir, 1989; Damme and Pirchner, 1984; Kjaer and Sørensen, 1997). To our knowledge, no h^2 estimates for separate body regions have been reported in the literature.

There are several reasons that could explain the difference in heritabilities between the present and previous studies. In the present study, FCS were recorded when the hens were 40 weeks of age, whereas in previous studies, they were recorded when hens were between 51 and 56 weeks of age (Craig and Muir, 1989; Kjaer and Sørensen, 1997). Kjaer and Sørensen (1997) reported that h^2 of FP behaviour (inflicting and receiving FP) increased when hens grew older (69 weeks compared to 38 weeks). Using younger birds could explain the lower h^2 .

A second reason for the lower h^2 observed in this study relates to the use of individual records vs. records pooled by cage. We used individual records on four random hens of the same line kept in one cage, while Craig and Muir (1989) used the average FCS of cages of three full sibs, which has two effects. First, it averages residuals over cage members, which reduces residual variance ($\sigma_e^2 < \sigma_e^2$) and thus increases heritability. Second, as demonstrated by Peeters et al. (2013), the analysis of cage averages yields an estimate of the total heritable variation (σ_{TBV}^2), rather than of the ordinary (direct) additive genetic variance (σ_A^2). Thus, the estimate of Craig and Muir (1989) refers to σ_{TBV}^2 instead of σ_A^2 . Together, those effects may explain the substantially higher h^2 found by Craig and Muir (1989).

Direct-Indirect Effects Model

Statistical comparison of the direct-indirect effects model (Equation 2) and the classical animal model (Equation 1), showed a significant improvement of the goodness of fit for both lines and for all body regions (**Table 2.2**; all p -values < 0.001 except $p = 0.13$ for belly in line W1), providing evidence for indirect genetic effects on FCS for almost all body regions.

Table 2.3 and **Additional file 2.1: Table S1** show the estimated genetic parameters obtained with the direct-indirect effects model. Except for the direct-indirect genetic correlations, most of the genetic parameters were significantly different from zero for both lines. As expected, the standard deviation of the direct breeding value (σ_{AD}) for the different body regions was of similar magnitude as σ_A from the classical animal model (see **Additional file 2.1: Tables S1 and S2**). The magnitude of σ_{AD} is not affected, because groups are composed of non-relatives (Peeters et al., 2013). Overall, estimates for σ_{AD} and σ_{AS} were highest in line WB. For all body regions in both lines, the standard deviation of the total breeding value (σ_{TBV}) was 1.3 to 4.0-fold larger than σ_{AD} , which indicates substantial indirect genetic effects. Again, line WB yielded the highest σ_{TBV} . For the body regions with significant IGE, non-direct genetic effects explained 70 (neck line W1) to 94% (rump line WB) of the total heritable variation in FCS (**Table 2.4**). In both lines, the contribution of non-direct genetic effects was highest for the rump region, explaining approximately 93% of the total heritable variation. Using univariate analysis, the total heritable variance expressed as the proportion of phenotypic variance (T^2) ranged from 9.8 to 53.6% (**Table 2.3**). Line WB yielded the highest T^2 .

In this study, line WB had the lowest FCS, whereas line W1 had the highest FCS. Therefore, it was expected that the contribution of indirect effects would be highest in line WB, because FCS depends on the behaviour of group mates. The estimates of genetic parameters for direct and indirect effects found here were indeed in line with those expectations. Furthermore, estimated genetic parameters were in agreement with the results of Ellen et al. (2008), who showed that the total heritable variation in survival time was substantially larger than suggested by the classical animal model. The estimated breeding values obtained in our analysis also provide an elegant way of discriminating between individuals that inflict FP and have high EBV for indirect effects, and individuals that are victims of FP and have high EBV for direct effects, as was previously suggested by Biscarini et al. (2010).

Bivariate analysis with mortality

Estimating genetic parameters using the bivariate classical animal model and the bivariate direct-indirect effects model, both with mortality (0/1) at 40 weeks of age, did not result in significant changes of the estimated genetic parameters compared to the univariate analyses (**Table 2.3**). Comparing log likelihoods of the bivariate model with a model in which the variance components were fixed to the estimates obtained from the univariate analysis, did not change the log likelihoods

2 Genetic parameters for plumage condition

Table 2.2 Model comparison for analysis of FCS using a likelihood ratio test

Line	Region	Cage ³		Classical ⁴		Direct-indirect ⁵	
		LR ¹	P (vs null ²)	LR	P (vs cage ³)	LR	P (vs classical ⁴)
W1	Neck	604.0	<0.001	64.8	<0.001	16.3	<0.001
	Back	209.2	<0.001	70.4	<0.001	22.9	<0.001
	Rump	1025.4	<0.001	8.5	0.004	11.5	<0.001
	Belly	326.2	<0.001	105.2	<0.001	4.1	0.128
	Total	862.3	<0.001	83.7	<0.001	29.5	<0.001
WB	Neck	1220.9	<0.001	156.3	<0.001	48.9	<0.001
	Back	1124.8	<0.001	45.2	<0.001	66.2	<0.001
	Rump	1698.9	<0.001	57.7	<0.001	66.6	<0.001
	Belly	401.0	<0.001	55.8	<0.001	15.4	<0.001
	Total	1565.2	<0.001	132.8	<0.001	90.7	<0.001

Model comparison was done for FCS in different regions of the body and for the total FCS in two lines (W1, WB); ¹LR is the likelihood ratio test (LR is two times the difference in log likelihood between the complex model and the less complex model); ²null is a model without random effects; ³Cage is a model with only a random cage effect; ⁴Classical is a model with an additive genetic effect and a random cage effect (Equation 1); ⁵Direct-indirect is a model with both direct and indirect genetic effects (Equation 2).

Table 2.3 Estimates of heritability from classical animal model ($h^2 \pm SE$) and direct-indirect effects model ($T^2 \pm SE$) for FCS

Line	Region	$h^2_{univariate}$	$h^2_{bivariate}$	$T^2_{univariate}$	$T^2_{bivariate}$
W1	Neck	0.059 ± 0.015	0.056 ± 0.014	0.195 ± 0.077	0.187 ± 0.074
	Back	0.074 ± 0.018	0.077 ± 0.019	0.257 ± 0.083	0.237 ± 0.077
	Rump	0.015 ± 0.007	0.015 ± 0.007	0.098 ± 0.066	0.094 ± 0.063
	Belly	0.079 ± 0.018	0.077 ± 0.017	0.143 ± 0.061	0.134 ± 0.058
	Total	0.060 ± 0.014	0.060 ± 0.014	0.251 ± 0.089	0.249 ± 0.089
WB	Neck	0.098 ± 0.018	0.102 ± 0.020	0.371 ± 0.117	0.406 ± 0.121
	Back	0.052 ± 0.014	0.048 ± 0.013	0.349 ± 0.110	0.397 ± 0.114
	Rump	0.048 ± 0.012	0.046 ± 0.012	0.456 ± 0.131	0.482 ± 0.133
	Belly	0.063 ± 0.017	0.068 ± 0.017	0.253 ± 0.091	0.255 ± 0.090
	Total	0.095 ± 0.018	0.093 ± 0.017	0.536 ± 0.140	0.588 ± 0.145

Estimates are shown for FCS in each body region and for total FCS in two lines (W1, WB); h^2 is the heritability obtained with the classical animal model using an univariate model ($h^2_{univariate}$) or a bivariate model with mortality ($h^2_{bivariate}$); T^2 is the total heritable variance expressed as a proportion of the phenotypic variance obtained with the direct-indirect effects model using an univariate model ($T^2_{univariate}$) or a bivariate model with mortality ($T^2_{bivariate}$); FCS is feather condition score, for each body region FCS ranges from 1 to 5, and total FCS ranges from 4 to 20.

Table 2.4 Contribution of direct and indirect genetic effects to total heritable variation of FCS

Line	Region	Contribution to the total heritable variation ¹ (%)			
		Direct ²	Indirect ³	Covariance ⁴	Total non-direct ⁵
W1	Neck	30.5	56.0	13.5	69.5
	Back	29.4	57.2	13.4	70.6
	Rump	8.4	98.4	-6.8	91.6
	Belly	55.6	34.9	9.4	44.4
	Total	22.0	68.9	9.1	78.0
WB	Neck	20.6	94.8	-15.4	79.4
	Back	7.5	114.7	-22.2	92.5
	Rump	6.3	96.0	-2.3	93.7
	Belly	27.5	50.7	21.8	72.5
	Total	10.1	106.6	-16.7	89.9

Contribution of direct genetic effects (direct), indirect genetic effects (indirect), the covariance between direct and indirect genetic effects (covariance), and total non-direct genetic effects (total non-direct) to total heritable variation of FCS in different body regions for two lines (W1, WB); ¹Total heritable variation ($\sigma_{TBV}^2 = \sigma_{AD}^2 + 2(n-1)\sigma_{ADS} + (n-1)^2\sigma_{AS}^2$); ²direct = σ_{AD}^2 ; ³indirect = $(n-1)\sigma_{AS}^2$; ⁴covariance = $2(n-1)\sigma_{ADS}$; ⁵total non-direct = $2(n-1)\sigma_{ADS} + (n-1)^2\sigma_{AS}^2$.

substantially (data not shown). Therefore, for this dataset, it is not necessary to use bivariate analyses with mortality to estimate genetic parameters for FCS of the different body regions.

Genetic correlations between body regions

Estimates of genetic correlations from the direct-indirect effects model are in **Table 2.5** and **Additional file 2.1: Table S3**. In both lines, genetic correlations (r_{12T}) were positive. The highest estimated genetic correlations were found between adjacent regions, whereas the lowest estimates were found between any region and belly (except for neck-belly in line WB). So far, there are no studies that report genetic correlations for FCS between different body regions. In previous studies, FCS of different body regions were combined and analyses were done on the total FCS (Craig and Muir, 1989; Damme and Pirchner, 1984; Kjaer and Sørensen, 1997). In the present study, the high genetic correlations between adjacent regions (neck, back, and rump) suggest that these regions could be combined into a single total FCS. The on average low genetic correlations with belly damage suggest that belly is a distinct trait (except for neck-belly in line WB, which suggests that neck and belly could be combined in a single FCS for this line). FP in the belly region may be

2 Genetic parameters for plumage condition

Table 2.5 Estimates of genetic correlations between total breeding values for FCS

Region	Region			
	Neck	Back	Rump	Belly
Neck		0.81 ± 0.13	NC ¹	0.52 ± 0.24
Back	NC ¹		>0.99	0.34 ± 0.25
Rump	0.87 ± 0.07	0.95 ± 0.04		0.70 ± 0.24
Belly	0.85 ± 0.13	0.72 ± 0.15	0.46 ± 0.19	

Estimates of genetic correlations between total breeding values (r_{12_r}) for FCS are shown for the four body regions, and for the two lines, line W1 (above diagonal) and line WB (below diagonal); ¹NC = not converged.

closely related with cannibalistic vent pecking. Vent pecking and FP are caused by different internal and external factors (Savory, 1995). Furthermore, FP is thought to be a redirected foraging behaviour (Blokhuis, 1986), whereas vent pecking seems to be a separate form of damaging behaviour (Savory, 1995; Newberry, 2004). In addition, when comparing the contribution of direct effects to the total heritable variation (**Table 2.4**), belly has a larger contribution of direct effects than the other body regions, while the contribution of indirect effects is lower for belly. This could indicate that a different behaviour is associated with pecking on the belly. Therefore, the belly region should be analysed separately and should not be included in the total FCS, as also reported by Parmentier et al. (2009). Furthermore, Bilčík and Keeling (1999) showed that feathers of the belly were pulled out more easily, which might overestimate the FCS due to FP.

2.3.4 Effect of person recording FCS

FCS was recorded by four persons, which could introduce bias due to incomplete correlation of FCS scores. The phenotypic correlation between observers was greater than 0.84 for neck, back, and rump, but was as low as 0.64 for belly (data not shown). The largest difference in means of the persons recording FCS was found for neck (for further details see **Additional file 2.1: Table S4**). Adams et al. (1978) reported a mean correlation of 0.88 between total FCS recorded by three persons.

To investigate the effect of the person recording the FCS on genetic parameters: (1) heterogeneity of residual variance was evaluated by fitting a separate residual variance for person and (2) genetic correlations between persons were estimated using the classical animal model (data not shown). Accounting for heterogeneous residual variances did not change the estimated genetic parameters. Therefore,

homogeneous residual variance was assumed. Genetic correlations of total FCS between persons were greater than 0.80 for both lines, which indicates only minor differences in FCS between persons at the genetic level.

2.3.5 Future breeding program

In this study, we showed that a substantial part of the total genetic variation in plumage condition of different body regions in two purebred layer lines is due to IGE. Accounting for IGE in the genetic analysis showed that total heritable variation was up to 9-fold greater than suggested by results of the classical animal model. Thus, using breeding programs that exploit the heritable variation due to IGE can considerably accelerate response to selection on FCS. In previous studies, FP was significantly related with lower FCS (Bilčik and Keeling, 1999; Kjaer and Sørensen, 2001). Since FP can result in death (referred to as cannibalism) (Blokhus and Arkes, 1984), it is worthwhile to investigate whether FCS at 40 weeks of age can be used as a predictor for survival at the end of the laying period. In this study, we showed that genetic correlations are high between FCS for adjacent body regions, whereas FCS for belly can be considered as a distinct trait, which suggests that neck, back, and rump can be combined into a total FCS. However, before drawing this conclusion, it is important to investigate whether total FCS, total FCS (without belly), or separate body regions can be used as predictor for survival at the end of the laying period. This could contribute to reducing an important welfare problem in laying hens. Regardless, both breeding and management solutions should be applied to prevent FP.

Measuring the plumage condition by recording FCS is time-consuming. Moreover, laying hens need to be at least 40 weeks before FCS can be recorded, and FCS measured on selection candidates is not very useful, because selection candidates are housed individually. Hence, selection would have to be based on sib or progeny information. These obstacles can be overcome by using genomic selection (Meuwissen et al., 2001; Wolc et al., 2011). Our results clearly show that genetic improvement of plumage condition cannot rely on DGE only. Therefore, genomic selection methods must be extended to accommodate IGE. A challenge is how to design a reference population that is suitable for genomic selection on plumage condition and survival time in laying hens.

2.4 Conclusions

Social interactions have a substantial effect on plumage condition in laying hens. This study shows that, for the different body regions (neck, back, rump, and belly), the total heritable variance of FCS, expressed as the proportion of the phenotypic variance (T^2) ranges from 9.8 (rump line W1) to 53.6% (total FCS). A substantial part, 70 to 94%, of the total heritable variation relates to IGE. Thus, it is expected that including both direct and indirect effects in a genetic selection program will contribute to a reduction in FP behaviour, one of the major welfare issues in the laying hen industry.

2.5 Acknowledgements

We thank the employees of the laying houses. Furthermore, we thank Laura Star, Koen Uitdehaag and Patrick Wissink for recording feather condition scores, and Bart Ducro and Han Mulder for statistical advice. This research was part of a joint project of Institut de Sélection Animale B.V., a Hendrix Genetics Company, and Wageningen University on “Genetics of robustness in laying hens”, which was financially supported by SenterNovem. Both EDE and PB are financially supported by the Dutch science council (NWO) and part of this work was co-ordinated by the Netherlands Technology Foundation (STW).

3

Predicting direct and indirect breeding values for survival time in laying hens using repeated measures

Tessa Brinker¹, Esther D. Ellen¹, Roel F. Veerkamp¹, Piter Bijma¹

¹ Animal Breeding and Genomics Centre, Wageningen University and Research, P.O.
Box 338, 6700 AH, Wageningen, The Netherlands

Genetics Selection Evolution (2015) 47:75

Abstract

Minimizing bird losses is important in the commercial layer industry. Selection against mortality is challenging because heritability is low, censoring is high, and individual survival depends on social interactions among cage members. With cannibalism, mortality depends not only on an individual's own genes (direct genetic effects; DGE) but also on genes of its cage mates (indirect genetic effects; IGE). To date, studies using DGE-IGE models have focussed on survival time but their shortcomings are that censored records were considered as exact lengths of life and models assumed that IGE were continuously expressed by all cage members even after death. However, since dead animals no longer express IGE, IGE should ideally be time-dependent in the model. Neglecting censoring and timing of IGE expression may reduce accuracy of estimated breeding values (EBV). Thus, our aim was to improve prediction of breeding values for survival time in layers that present cannibalism.

We considered four DGE-IGE models to predict survival time in layers. One model was an analysis of survival time and the three others treated survival in consecutive months as a repeated binomial trait (repeated measures models). We also tested whether EBV were improved by including timing of IGE expression in the analyses. Approximate EBV accuracies were calculated by cross-validation. The models were fitted to survival data on two purebred White Leghorn layer lines W1 and WB, each having monthly survival records over 13 months.

Including the timing of IGE expression in the DGE-IGE model reduced EBV accuracy compared to analysing survival time. EBV accuracy was higher when repeated measures models were used. However, there was no universal best model. Using repeated measures instead of analysing survival time increased EBV accuracy by 10 to 21 and 2 to 12% for W1 and WB, respectively. We showed how EBV and variance components estimated with repeated measures models can be translated into survival time.

Our results suggest that prediction of breeding values for survival time in laying hens can be improved using repeated measures models. This is an important result since more accurate EBV contribute to higher rates of genetic gain.

Key words: laying hens, repeated measures, survival, indirect genetic effects

3.1 Introduction

Minimizing bird losses in the commercial layer industry is important, both from welfare and economic points of view. Thus, selection against mortality has been of interest to researchers (Craig and Muir, 1996; Muir, 1996; Ellen et al., 2014) but has not always been effective (Preisinger, 1998). Genetic improvement of mortality in poultry breeding is challenging for several reasons. In addition to having a low heritability, one of the main complications is that the time until death is often not observed because most laying hens are still alive at the end of the recording period (Lagakos, 1979; Ducrocq et al., 2000). Hence, only a lower bound of the true survival time is known for most hens, which is referred to as censoring (Lagakos, 1979). Excluding censored records from analyses or considering the lower bound as the actual record is expected to reduce the accuracy of estimated breeding values (EBV).

The fact that commercial laying hens live in groups complicates selection for lower mortality even more. Group housing allows social interactions between group members, such that survival time in laying hens might be adversely affected by harmful social behaviours such as feather pecking (Savory, 1995; Rodenburg et al., 2013). In these cases, survival time depends on both the genes of the potential victim (known as the direct genetic effect; DGE) and on the genes of its cage mates (known as the indirect genetic effect; IGE) (Muir, 1996; Moore et al., 1997; Wolf et al., 1998; Muir, 2005; Bijma et al., 2007a. Bijma et al., 2007b). In other words, the environment that individuals experience contains a heritable component (IGE), expressed by the cage mates. Such IGE can affect response to selection considerably and neglecting IGE when selecting for lower mortality can even result in a negative response to selection (Craig and Muir, 1996; Griffing, 1967).

Ellen et al. (2008) and Peeters et al. (2012) investigated the contribution of IGE to heritable variation in survival time of laying hens. These two studies used a DGE-IGE linear mixed model to estimate genetic parameters. Shortcomings of this model are that censored records were considered as exact lengths of life and it assumed that IGE were continuously expressed by all individuals in a cage, irrespective of whether they were alive or dead. The latter assumption is invalid because cage composition changes over time due to death of animals, as dead animals no longer express IGE on their cage mates. Thus, to increase the accuracy of estimates of DGE and IGE for survival time, methods that can cope with censoring and timing of IGE expression should be explored.

Ellen et al. (2010) investigated opportunities of survival analysis models with DGE and IGE to account for censoring. Survival analysis models exploit information on both censored and uncensored records properly by accounting for the non-linearity of survival time data and can also include time-dependent effects (Kalbfleisch and Prentice, 1980; Ducrocq and Casella, 1996). Ellen et al. (2010) used a two-step approach that combined survival analysis and a DGE-IGE linear mixed animal model for survival time because it was not possible to estimate the variance of correlated genetic effects with existing survival analysis software. However, Ellen et al. (2010) showed that the accuracy of EBV was not improved with the two-step approach compared to the DGE-IGE linear mixed animal model for data in which all surviving animals were censored at the same time.

Several other statistical techniques for analysing survival data have been proposed that can consider censoring and time-dependent effects, including repeated measures models (Madgwick and Goddard, 1989; Jairath et al., 1998; Veerkamp et al., 2001; Schaeffer, 2004; Jamrozik et al., 2008). In repeated measures models, a survival indicator is used to circumvent censoring. Survival is measured as a binomial trait (0/1), which indicates that an individual is dead (0) or alive (1) at specific time points (Jamrozik et al., 2008), or that an individual has survived or not during a specific time period (Veerkamp et al., 2001). In the first case, survival has no missing records, whereas in the second case, records for time periods after death of an individual are set to missing. The trait definition differs in each approach since the method of Jamrozik et al. (2008) approximates the survival function of the proportional hazard model, whereas the method of Veerkamp et al. (2001) approximates the hazard function of the proportional hazard model. Modelling the hazard function enables to estimate covariances between independent time intervals whereas modelling the survival function enables to estimate covariances between cumulative averages of time intervals (Van Pelt et al., 2014).

In a study on survival in dairy cows, Veerkamp et al. (2001) found that the repeated measures model was robust to censoring since the correlations between EBV from uncensored and randomly censored data were high. Ødegård et al. (2006) confirmed this finding in a comparative analysis of different models for survival data in Atlantic salmon. They showed that the repeated measures model had a greater predictive ability than survival analysis (Ødegård et al., 2006). Thus, repeated measures models appear to be an appropriate tool for analysing survival

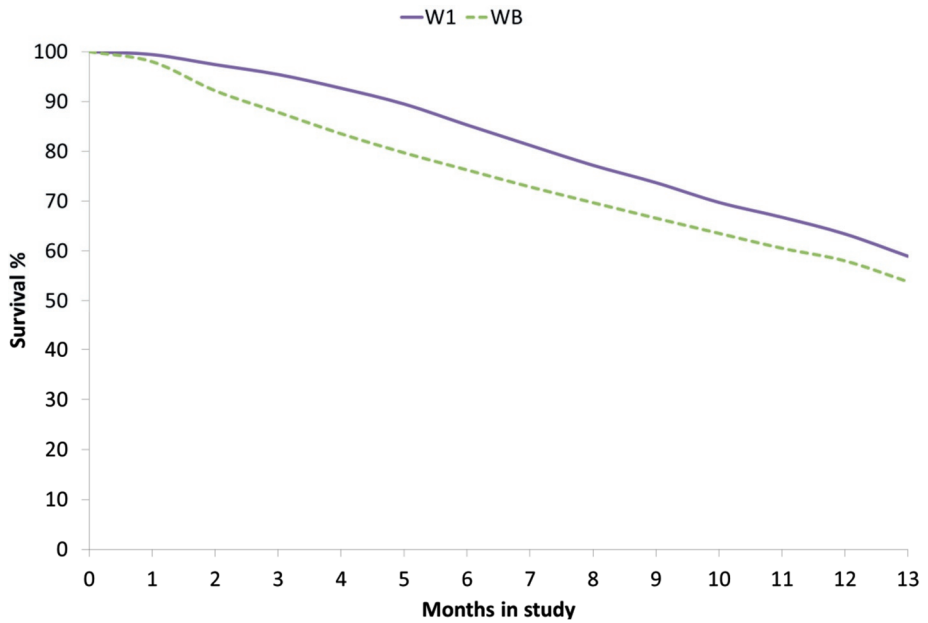


Figure 3.1 Percentage of survival of layer chickens for lines W1 and WB throughout the experiment (max = 13 months).

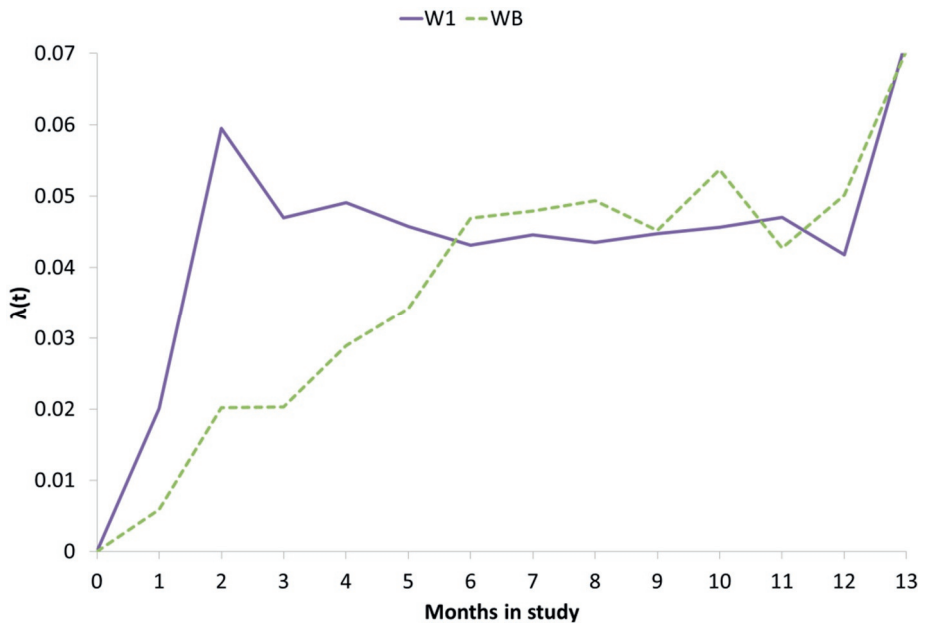


Figure 3.2 Hazard function $\lambda(t)$ of layer chickens for lines W1 and WB throughout the experiment (max = 13 months).

data since they can account for censoring and timing of IGE expression. However, to date, the potential of repeated measures models for estimating DGE and IGE on survival time has not been investigated.

The aim of this study was to improve prediction of breeding values for direct and indirect effects on survival time in two purebred White Leghorn layer lines in which the level of mortality was high due to cannibalism (Ellen et al., 2008). For this purpose, we compared EBV for survival (0/1) from repeated measures models to EBV for survival time from a linear mixed model. The predictive ability of EBV from both models was assessed by cross-validation.

3.2 Methods

3.2.1 Populations and pedigree

Data were collected under the control of the Institut de Sélection Animale B.V. (ISA), the layer breeding division of Hendrix Genetics. Hendrix Genetics complies with the Dutch law on animal welfare. ISA provided data on two purebred White Leghorn layer lines, denoted W1 and WB (Ellen et al., 2008).

For each line, matings between sires and dams were randomly assigned and occurred in two batches with a 6-month interval. For each batch, each sire (36 for line W1 and 35 for line WB) was mated to approximately eight dams, resulting in an average of 12.3 female offspring per dam. Each batch was partitioned into four age groups that differed in age by 2 weeks. Laying hens had intact beaks.

3.2.2 Housing

Laying hens of the same age were randomly allocated to four-bird battery cages approximately 17 weeks after hatching. Each batch was transported to a different laying house, coded as 1 and 2. The laying houses had eight double rows of cages and each row comprised three levels (top, middle, and bottom). Hens in laying house 2 were not placed on the top level. A standard commercial layer diet and water were provided ad libitum in the front and back parts of each cage, respectively. Light intensity was stronger in laying house 2 than in laying house 1 (Ellen et al., 2008). Further details are in Ellen et al. (2008).

3.2.3 Data

Dead hens were removed daily. After death, wing band number, cage number, date of death, and cause of death were recorded. The latter was done subjectively by

the employees of ISA without dissection. The study was terminated when hens were on average 75 weeks old. In total, 59% W1 and 54% WB laying hens survived (**Figure 3.1**). Survival rates of W1 and WB hens differed most during the first 4 months of the experiment (**Figure 3.2**). Most hens died because of cannibalism; only 37 W1 and 15 WB hens died for other reasons, e.g. some hens were killed by mink. Observations on hens that died for other reasons were removed from the dataset because the objective was to investigate death from cannibalism. However, the identification numbers were retained in their cage mates' observations for IGE modelling.

Survival time was defined as the number of days from entry in the laying house to death. Since each batch consisted of four age groups that differed in age by 2 weeks, the maximum number of survival days differed between age groups. Thus, the maximum number of survival days was cut off at 416 days, which is the maximum survival time of the youngest age group, which means that all hens that were still alive at 416 days of age had censored records on survival time. In the statistical analysis of survival time, those hens were given a value of 416 days. In total, records on 6276 and 6916 hens were used for statistical analysis of survival time for lines W1 and WB, respectively.

To define survival, the laying period was divided into 13 months. For each month, survival was coded as 1 if the laying hen was alive at the end of that month, and as 0 if not. Thus, a survival record (0/1) was available for each month. This resulted in a total of 81 588 and 89 908 monthly records for lines W1 and WB, respectively.

3.2.4 Statistical models

Four statistical models were compared: a linear mixed model for survival time, two linear mixed models for survival (0/1), and a generalized linear mixed model for survival (0/1). Five generations of pedigree were included in all genetic analyses. All models were implemented using ASReml (Gilmour et al., 2014).

Survival time model STM

DGE and IGE for survival time were estimated using a survival time model (STM) (Muir, 2005; Bijma, 2007b) with the following linear mixed model:

$$y_{ijk} = fixed + A_{Di} + \sum_{j \neq i}^{n-1} A_{Ij} + cage_k + e_{ijk}, \quad (1)$$

3 Predicting direct and indirect breeding values

where y_{ijk} is the observed survival time (days) for individual i , with cage mates j , in cage k , the *fixed* term is the fixed effect of the combination of laying house-row-level, A_{Di} is the random DGE of individual i , $\sum_{j \neq i}^{n-1} A_{Ij}$ is the sum of the $n - 1$ random IGE of the cage mates j , with n denoting cage size at the start of the experiment ($n = 4$), $cage_k$ is the random cage effect, and e_{ijk} is the residual. Individuals that were still alive at the end of the recording period were assigned a survival time record of 416 days. Genetic effects were assumed to follow a normal distribution $\sim N(0, \mathbf{C} \otimes \mathbf{A})$, with $\mathbf{C} = \begin{bmatrix} \sigma_{AD}^2 & \sigma_{ADI} \\ \sigma_{ADI} & \sigma_{AI}^2 \end{bmatrix}$, the Kronecker product of matrices \otimes , a relationship matrix \mathbf{A} , direct genetic variance σ_{AD}^2 , indirect genetic variance σ_{AI}^2 , and direct-indirect genetic covariance σ_{ADI} .

Residuals of cage members may be correlated because of non-heritable indirect effects, with: $\rho = (2\sigma_{EDI} + (n - 2)\sigma_{EI}^2) / \sigma_e^2$ (Bijma, 2007b). In cases where cage members are ‘similar’, i.e. when ρ is positive, a random cage effect can be fitted instead of fitting correlated residuals, with $\sigma_{cage}^2 = 2\sigma_{EDI} + (n - 2)\sigma_{EI}^2 = \rho\sigma_e^2$. Based on a previous study (Ellen et al., 2008), the correlation was estimated to be positive, and a random cage effect was therefore fitted in this study. The cage effect cannot be fitted as a fixed effect because the indirect genetic variance is not statistically identifiable when a fixed cage effect is included (Cantet and Cappa, 2008).

Repeated measures model RMM.t

Monthly survival (0/1) was analysed using a repeated measures model that included random DGE and IGE regressions on time (hence RMM.t) based on a sire-dam model. No genetic effects for intercept were fitted, because there is no phenotypic variation at the start of the experiment ($t = 0$), since all hens were alive at time 0. After that, phenotypic variance in survival increases over time until it reaches a maximum at 50% mortality, and then declines again. In this experiment, mortality was less than 50%, so phenotypic variance only increased over time. This increase in variance over time is consistent with a model with random regressions on time, for which the variance is proportional to the square of time. The model with random regressions on time was:

$$y_{ijklm} = \text{fixed} + A_{sd_{Di}} \cdot t_m + \sum_{j \neq i}^{n-1} A_{sd_{Ij}} \cdot t_m + cage_{km} + cage_k \cdot t_m + PE_l \cdot t_m + e_{ijklm}, \quad (2)$$

where y_{ijklm} is the observed survival (0/1) for individual l , offspring of sire-dam combination i , with cage mates from sire-dam combination j , in cage k , at time t_m measured in months since entry of the experiment, the *fixed* term is the fixed interaction effect of laying house-row-level, which was fitted with a sixth-order polynomial of time, $A_{sd_{D_i}}$ is the DGE of the sire-dam combination as a function of time, $\sum_{j \neq i}^{n-1} A_{sd_{I_j}}$ is the sum of the $n-1$ IGE of the sire-dam combination of the cage mates as a function of time, $cage_{km}$ is the random effect of cage k at time m , $cage_k$ is the random effect of cage k as a function of time, PE_l is the random permanent environmental effect of individual l as a function of time, t_m is the time, and e_{ijklm} is the residual. A separate residual variance was estimated for each month. The DGE and IGE were allowed to be correlated. Sire-dam effects were assumed to follow a normal distribution $\sim N(0, \mathbf{C} \otimes \mathbf{A})$, with $\mathbf{C} = \begin{bmatrix} \sigma_{A_{Dsd}}^2 & \sigma_{A_{Dsd}I_{sd}} \\ \sigma_{A_{Dsd}I_{sd}} & \sigma_{A_{I_{sd}}}^2 \end{bmatrix}$, direct genetic sire-dam variance $\sigma_{A_{Dsd}}^2$, indirect genetic sire-dam variance $\sigma_{A_{I_{sd}}}^2$, and direct-indirect genetic sire-dam covariance $\sigma_{A_{Dsd}I_{sd}}$.

This model contains two random cage effects: one is a random cage effect at each time, $cage_{km}$, and the other, $cage_k \cdot t_m$, is a random regression on time. The $cage_{km}$ effect accounts for covariances among cage members at specific time points. A single variance was estimated for this effect. The $cage_k \cdot t_m$ term accounts for covariances between records on the same cage at different time points, and for increasing variance over time. Both random cage effects were very significant and when excluding one or the other, the variance explained by cage was not fully covered.

Repeated measures model RMM.p

The regressions on time in the RMM.t model (Equation 2) imply that variances are a quadratic function of time; e.g. $var(a \cdot t) = t^2 var(a)$, where a is the genetic effect and t is time. However, the true variance for binomial traits equals $p(1-p)$, where p is the probability for an individual to survive until time t . In other words, p is the mean survival at time t . A quadratic function of time does not fit a $p(1-p)$ function well because the slope of t^2 increases with t , whereas the slope of $p(1-p)$ decreases with p . To better fit the variance, the regression on time was replaced by a regression on $\sqrt{p(1-p)}$, so that fitted variances were proportional to $\sqrt{p(1-p)}$; e.g. $var(a \cdot \sqrt{p(1-p)}) = p(1-p) \cdot var(a)$. In other words, monthly survival (0/1) was analysed using a repeated measures model including

3 Predicting direct and indirect breeding values

DGE and IGE regressions on a function of mean survival (RMM.p), rather than on time. The model was:

$$y_{ijklm} = fixed + A_{sd_{D_i}} \cdot x_m + \sum_{j \neq i}^{n-1} A_{sd_{I_j}} \cdot x_m + cage_{km} + cage_k \cdot x_m + PE_l \cdot x_m + e_{ijklm}, \quad (3)$$

where $x_m = \sqrt{p_m(1-p_m)}$, p_m denoting the mean survival at time m . For each time m , the p_m was calculated separately for each fixed effects class. The other terms are the same as for Equation 2.

Generalized linear mixed model GLMM

To account for the binomial distribution of monthly survival, DGE and IGE for survival were estimated using a generalized linear mixed model (GLMM) with a logit link function. The GLMM was:

$$\eta(E(y_{ijklm})) = fixed + A_{sd_{D_i}} + \sum_{j \neq i}^{n-1} A_{sd_{I_j}} + cage_{km} + PE_l, \quad (4)$$

where $\eta(\)$ is the logit link function that links the probability p of surviving to the linear predictor, $E(y_{ijklm})$ is the probability of survival for individual l , offspring of sire-dam combination i , with cage mates from sire-dam combination j , in cage k , at time m . The other terms are the same as for Equation 2.

In contrast to the survival models, the GLMM only includes a genetic intercept rather than a regression on time. This is because the non-linear link function takes the change in variance over time into account. Hence, at the beginning of the recording period, the variance of survival probabilities can be (near) zero even when $var[\eta(E(y_{ijklm}))]$ is substantially greater than 0.

3.2.5 Time-dependent IGE

The data structure of the survival models (RMM.t, RMM.p, and GLMM) allows inclusion of time-dependent random effects. The composition of the cage changes over time because animals die. Thus, we tested if the prediction of breeding values could be improved by including timing of IGE expression in the model. For this purpose, the cage mates that were alive at the beginning of each month were indicated for each monthly survival observation. Hence, in models RMM.t, RMM.p

and GLMM, the $\sum_{j \neq i}^{n-1} A_{sd_{ij}}$ was time-dependent, and included only sires-dam combinations of cage mates that were still alive at the beginning of the month. Thus, the sires and dams of cage mates that were no longer alive were set to missing. For time periods after death of the focal individual, the cage composition was kept identical to that at the time of death because these observations were no longer impacted by changes in cage composition.

3.2.6 Cross-validation

The quality of EBV from each model was assessed by cross-validation, which is a technique of model validation where the correlation between predicted and observed phenotypes serves as a quality measure (Stone, 1974). With this procedure, known phenotypes are set to missing, their values are predicted, and finally the predicted values are compared to the observed values. In this study, five mutually exclusive subsets were created, where survival phenotypes for approximately 20% of the cages were removed. Those cages were selected at random. These subsets were used to predict the phenotypes of the individuals that belonged to the 20% cages that were removed. All fixed effect classes were present for each subset. A bivariate analysis in ASReml (Gilmour et al., 2014) of the ranks of observed and predicted phenotypes, with a fixed effect for each subset (from 1 to 5), was used to calculate the Spearman rank correlation and corresponding standard error across the five validation sets (Gilmour et al., 2014). We used rank correlations because survival phenotypes were unknown for the censored individuals (see below). The following sections describe how observed and predicted phenotypes were obtained.

Observed and predicted phenotypes

For the individuals that were not censored, observed phenotypes were the observed survival days. First, these were adjusted for fixed effects using a linear model with fixed effects only; $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ with \mathbf{y} being a vector of observed survival times, \mathbf{X} being an incidence matrix linking survival time observations to fixed interaction effects of laying house-row-level, \mathbf{b} being a vector of the fixed effects, and \mathbf{e} is the residual term. Hence, the residual of this model represents the adjusted phenotypes. For the censored individuals, the observed phenotype is unknown but these observations contain important information because they correspond to hens with the largest number of survival days. To allow calculation of the rank correlation between predicted and observed phenotypes, we followed the approach presented in (Ellen et al., 2010), which assumes that censored individuals

died in random order after surviving up to 416 days. Under this assumption, censored individuals can be given the average rank of all censored individuals. For example, when five out of 10 individuals are censored, then the average rank of the censored individuals is $(6+7+8+9+10)/5 = 8$ and all censored individuals are given a rank of 8, while uncensored individuals were given their observed rank (after correction for fixed effects).

Predicted phenotypes were the rank of predicted survival times of individuals. Phenotypes were predicted by combining the estimated DGE (\hat{A}_{D_i}) of the individual itself and the estimated IGE of its cage mates ($n = 3$) that were present at the start of the experiment ($\sum \hat{A}_{I_j}$) (see **Additional file 3.1** for details).

Approximate accuracy

Based on the method described in Ellen et al. (2010), approximate accuracies of EBV were calculated for all models. These were approximations because accuracy refers to the ranks rather than to the phenotypic values and EBV themselves. If EBV underlying the predicted phenotypes were estimated with an accuracy of 1, the expected rank correlation would be:

$$\sqrt{r^2} = \sqrt{(\sigma_{A_D}^2 + (n-1)\sigma_{A_I}^2)/\sigma_P^2}, \quad (5)$$

where the numerator is the genetic component of phenotypic variance. The approximate accuracy was calculated as:

$$\hat{r}_{IH} = \text{corr}[\text{rank}(P_i - \bar{P})_i, \text{rank}(\hat{P}_i)]/\sqrt{r^2}, \quad (6)$$

with $P_i - \bar{P}$ represents the observed phenotype corrected for fixed effects, and \hat{P}_i represents the predicted phenotype. Genetic parameter estimates from model STM were used to calculate $\sqrt{r^2}$ (see **Table 3.1**).

3.2.7 Genetic parameters

In addition to EBV, we were interested in estimating genetic parameters for survival time, which can, e.g., give an indication of the amount of genetic progress that can be made for a trait. To make genetic parameter estimates from the different models comparable, they were transformed to the survival time scale, as described in the following section for RMM.t. For RMM.p and GLMM, the

translation of genetic parameters to the survival time scale involves tedious integrals and, thus, genetic parameters from these models are not presented here.

For STM, parameters for survival time follow directly from the estimates. In the presence of social interactions, where each individual interacts with $n - 1$ cage mates, the total heritable variance for response to selection is given by (Bijma et al., 2007b; Bijma, 2011):

$$\sigma_{TBV}^2 = \sigma_{A_D}^2 + 2(n - 1)\sigma_{A_{DI}} + (n - 1)^2\sigma_{A_I}^2, \quad (7)$$

phenotypic variance equals:

$$\sigma_P^2 = \sigma_{A_D}^2 + (n - 1)\sigma_{A_I}^2 + \sigma_{cage}^2 + \sigma_e^2, \quad (8)$$

and the ratio of heritable variance and phenotypic variance is equal to:

$$T^2 = \frac{\sigma_{TBV}^2}{\sigma_P^2}, \quad (9)$$

T^2 is an analogy of heritability that expresses the total heritable variance that is available for response to selection relative to phenotypic variance (Bijma et al., 2007a; Bijma, 2011).

To obtain genetic parameters for survival time from the models for monthly survival (0/1), survival has to be translated into survival time. Survival time (ST) of an individual is the sum of its survival records ($S = 0,1$) for each day over time,

$$ST_i = c \cdot \sum_{t=start}^{t=end} (S_i)_t, \quad (10)$$

where c is a multiplication factor that translates monthly survival (0/1) into days; $c = 30.4$ days. With this relationship it is possible to translate the estimated genetic parameters for survival to the survival time scale (See **Additional file 3.2**). Hence, a survival model can be used to estimate genetic parameters and breeding values for survival time.

3.3 Results and discussion

3.3.1 Cross-validation

The rank correlations between observed and predicted phenotypes, $\text{corr}[\text{rank}(P_i - \bar{P})_i, \text{rank}(\hat{P}_i)]$ and the corresponding approximate accuracy (\hat{r}_{IH}) of the four models (STM, RMM.t, RMM.p, and GLMM) are in **Table 3.2**. Rank correlations ranged from 0.135 to 0.162 for line W1, and from 0.170 to 0.190 for line WB. Corresponding \hat{r}_{IH} ranged from 0.44 to 0.53 for line W1, and from 0.46 to 0.52 for line WB. Although rank correlations appear low, they are in fact within our range of expectations because if breeding values were predicted with an accuracy of 1, the rank correlation would be equal to the square root of the proportion of phenotypic variance explained by the genetic variance, i.e. $\sqrt{r^2}$ (Equation 5) (Ellen et al., 2010). Using the genetic parameters from STM (**Table 3.1**), $\sqrt{r^2}$ is equal to 0.31 and 0.37 for lines W1 and WB, respectively. These values are the upper bounds of the rank correlations in **Table 3.2**.

Model comparison

Previous studies analysed survival time with an animal model (Ellen et al., 2008; Peeters et al., 2012; Ellen et al., 2010). Our objective was to investigate whether predictions from such models could be improved, and for this reason, STM was analysed with an animal model, while RMM.t, RMM.p, and GLMM were analysed with a sire-dam model. Sire-dam models were used because animal models may result in biased genetic parameter estimates because of the “extreme category problem” when analysing binomial/categorical data (Hoeschele and Tier, 1995). At the start of the experiment no variation exists, but as hens start to die the variation within classes of fixed effects starts to change. Moreover, for some classes of fixed effects, the variation is not apparent until later in the experiment. It is not an option to remove these fixed-effect classes because they contain important information on survival. Hence, we found sire-dam models more appropriate than animal models to predict breeding values and genetic parameters.

To justify these comparisons with STM, we investigated whether analysis of survival time using an animal model or a sire-dam model were comparable. Genetic parameters of survival time (see STM in **Table 3.1**) and rank correlations between observed and predicted phenotypes (see STM in **Table 3.2**) were the same for the animal model and the sire-dam model. In addition, correlations between EBV that were predicted for the cross-validation sets by the animal and sire-dam models

were higher than 0.99 for both lines. Thus, sire and animal models for survival time gave very similar results.

Comparing EBV from RMM.t, RMM.p, and GLMM with EBV from STM showed that accuracies increased for all three models (**Table 3.2**), by up to 20% for line W1, and 12% for line WB. Model RMM.p resulted in the highest predictive ability for line W1 but led to little improvement over STM for line WB. Based on theory, RMM.p was expected to improve prediction of breeding values for both lines compared to RMM.t because RMM.p fits a variance that better agrees with the true binomial variance. However, the results using survival data on line WB were not in line with this expectation, which might be caused by differences in survival curves between both lines.

A comparison of GLMM with RMM.t indicated that the predictive ability of these two models was quite similar, although the GLMM resulted in 1% (line W1) to 3% (line WB) greater accuracies than RMM.t. The ranks of predicted phenotypes from GLMM and RMM.t were highly correlated: 0.97 for line W1 and 0.98 for line WB (**Table 3.3**), which means that there was almost no re-ranking by using GLMM or

Table 3.1 Estimates of genetic parameters (\pm SE) for survival time using models STM, RMM.t, and RMM.t with time-dependent indirect genetic effects (RMM.t-td) for two layer lines W1 and WB.

	W1			WB		
	STM	RMM.t	RMM.t – td	STM	RMM.t	RMM.t – td
σ_{A_D}	28 \pm 3	28 \pm 3	29 \pm 3	41 \pm 4	38 \pm 4	41 \pm 4
σ_{A_I}	10 \pm 2	11 \pm 2	33 \pm 2	16 \pm 3	12 \pm 2	20 \pm 1
$\sigma_{A_{DI}}$	57 \pm 67	57 \pm 64	255 \pm 129	-158 \pm 120	-111 \pm 87	-311 \pm 112
σ_{TBV}	45 \pm 8	46 \pm 7	109 \pm 8	55 \pm 9	46 \pm 8	58 \pm 7
σ_P	107 \pm 1	107 \pm 1	114 \pm 1	135 \pm 1	123 \pm 1	128 \pm 1
T^2	0.18 \pm 0.06	0.19 \pm 0.06	0.93 \pm 0.11	0.16 \pm 0.05	0.14 \pm 0.05	0.21 \pm 0.05
r_A	0.20 \pm 0.22	0.19 \pm 0.20	0.26 \pm 0.13	-0.24 \pm 0.18	-0.24 \pm 0.19	-0.38 \pm 0.13

Estimates of genetic parameters are provided for survival time in days for both W1 and WB lines. σ_{A_D} , σ_{A_I} , and $\sigma_{A_{DI}}$ are the direct genetic standard deviation, indirect genetic standard deviation, and direct-indirect genetic covariance. σ_{TBV} is the total genetic standard deviation, σ_P is the phenotypic standard deviation, T^2 is the total heritable variance relative to the phenotypic variance, and r_A is the genetic correlation between direct and indirect genetic effects. **Additional file 3.1** describes the procedure to translate genetic parameters of RMM.t to survival days.

Table 3.2 Rank correlations between observed and predicted phenotypes (\pm SE) and approximate accuracies for lines W1 and WB.

Model	Time-dependent	Rank correlation		% Improvement*		WB		% Improvement*		Approximate accuracy	
		W1		W1		WB		WB		W1	WB
STM	-	0.135 \pm 0.012	-	0.170 \pm 0.012	-	0.170 \pm 0.012	-	0.44	0.46		
RMM.t	No	0.148 \pm 0.012	+10	0.185 \pm 0.012	+9	0.185 \pm 0.012	+9	0.48	0.51		
RMM.p	No	0.162 \pm 0.012	+20	0.174 \pm 0.012	+2	0.174 \pm 0.012	+2	0.53	0.47		
GLMM	No	0.150 \pm 0.012	+11	0.190 \pm 0.012	+12	0.190 \pm 0.012	+12	0.49	0.52		
RMM.t	Yes	0.063 \pm 0.013	-53	0.134 \pm 0.012	-21	0.134 \pm 0.012	-21	0.20	0.37		
RMM.p	Yes	0.049 \pm 0.013	-64	0.124 \pm 0.012	-27	0.124 \pm 0.012	-27	0.16	0.34		
GLMM	Yes	0.081 \pm 0.013	-41	0.149 \pm 0.012	-12	0.149 \pm 0.012	-12	0.26	0.41		

Rank correlations between observed and predicted phenotypes and approximate accuracies using STM, RMM.t, RMM.p, and GLMM are provided for lines W1 and WB. All models were analysed by either excluding (time-dependent = no) or including (time-dependent = yes) timing of IGE expression. *compared to STM.

Table 3.3 Rank correlations between predicted phenotypes from models without timing of IGE expression (\pm SE).

Model	STM	RMM.t		RMM.p		GLMM	
STM		0.876 \pm 0.003		0.789 \pm 0.005		0.883 \pm 0.003	
RMM.t	0.910 \pm 0.002			0.878 \pm 0.003		0.976 \pm 0.001	
RMM.p	0.795 \pm 0.005	0.876 \pm 0.003				0.881 \pm 0.003	
GLMM	0.926 \pm 0.002	0.973 \pm 0.001	0.874 \pm 0.003				

Rank correlations between predicted phenotypes from STM, RMM.t, RMM.p, and GLMM excluding timing of IGE expression are provided for lines W1 (below the diagonal) and WB (above the diagonal).

RMM.t to predict breeding values. This is in line with other studies that found no advantage of analysing binomial/categorical data with binomial/categorical models compared to using linear models (Olesen et al., 1994; Vazquez et al., 2009).

Survival analysis is a commonly used method to deal with censoring and time-dependent effects (Kalbfleisch and Prentice, 1980; Ducrocq and Casella, 1996). However, it is not possible to estimate the variance for correlated genetic effects with the current software for survival analysis. Veerkamp et al. (2001) used a method that approximates survival analyses by analysing survival as dead (0), alive (1) or missing (for time periods after the death of an individual) with a repeated measures model regressed on time and an intercept. This is different from our study, in which we coded survival as either dead (0) or alive (1), without missing records, following Jamrozik et al. (2008). The method of Veerkamp et al. (2001) has similarities with the hazard function used in survival analysis. For our data, we found that rank correlations obtained with the method of Veerkamp et al. (2001) were lower than those obtained with RMM.t, RMM.p and GLMM and were similar or lower than those obtained with STM. When coding survival as dead (0), alive (1) or missing, rank correlations were equal to 0.132 and 0.157 for lines W1 and WB.

Ellen et al. (2010) explored the potential of survival analysis by applying a two-step approach that combined survival analysis and STM. Similar to the Veerkamp et al. (2001) method, the Ellen et al. (2010) method approximates survival analysis. Applying the method of Veerkamp et al. (2001) to our data yielded results that were in line with those of Ellen et al. (2010), who observed no improvement in rank correlations using the two-step approach compared to STM.

Time-dependent IGE

Including timing of IGE expression in the repeated measures models had a substantial negative effect on rank correlations; compared to the results obtained with STM, rank correlations decreased by 12 and 64% for lines WB and W1, respectively (**Table 3.2**). Lipschutz-Powell et al. (2012) performed a simulation study to investigate whether DGE-IGE models applied to infectious disease data could accommodate the dynamic nature of such data. The non-infected individuals in Lipschutz-Powell et al. (2012) are equivalent to dead individuals in our study, since both traits do not express the IGE. Lipschutz-Powell et al. (2012) found that accounting for timing of IGE expression in the DGE-IGE model inflated the variance for IGE, similar to what we observed (**Table 3.1**). They indicated that the problem probably arises from the fact that modification of the incidence matrices that link

observations to IGE directly depends on the observations (2012). As a consequence, a considerable amount of the phenotypic variance will be explained by IGE and IGE variance will be overestimated. Other methods to incorporate timing of IGE expression need to be explored.

Censoring

Our results indicate that it is important to use methods that incorporate censoring when analysing survival time. With STM, only a lower bound of the true survival time is known for censored records, which results in reduced accuracy. In the repeated measures models, this problem was circumvented by using a survival indicator, i.e. dead (0) or alive (1) at a given time. In this case, accuracy of predictions of breeding values increased by up to 20 and 12% for lines WB and W1, respectively, compared to accuracies from STM. In this study, censored records were obtained at the same time for all individuals, which is often the case in layer breeding programs. Censoring at various times during the laying period will result in different accuracies compared to what was presented in this study. We investigated the effect on predicted breeding values when 50% of the censored individuals were censored half way during the laying period using STM and RMM.t. Compared to analyses where all hens were censored at the same time, rank correlations using STM decreased by 16 to 25%, while rank correlations using RMM.t decreased by only 3 to 6% (results not shown). Thus, RMM.t was more robust to censoring at various times than STM. Similar results are expected for RMM.p and GLMM compared to STM. Thus, the benefits of RMM.t, RMM.p, and GLMM observed here are conservative estimates because all individuals were censored at the same time.

3.3.2 Genetic parameters

Estimates of genetic parameters for DGE and IGE from STM (Equation 1) and RMM.t (Equation 2) are in **Table 3.1**. Genetic parameters were expressed on the survival time scale (See **Additional file 3.1**), which demonstrates that it is possible to translate the variance components estimated with RMM.t to STM. Genetic parameters of survival time were very similar for the two models but prediction of breeding values was improved by using RMM.t compared to STM.

Estimates of genetic parameters for the same data were slightly different than those reported in Ellen et al. (2008). The total heritable variance relative to the phenotypic variance, T^2 , was estimated at 19% in Ellen et al. (2008), and 18% in our study. In our study, the maximum number of survival days was cut off at 416 days,

which was the maximum survival time of the youngest age group. In Ellen et al. (2008), the maximum survival time was 447 days. Furthermore, Ellen et al. (2008) included the average survival time of the back neighbours as a fixed covariate because hens shared drinking nipples. In our study, the fixed effect of survival of the back neighbours was excluded because this effect may indirectly result from the focal cage itself; if a back neighbour effect exists, then mortality in the focal cage will affect mortality of back neighbours, and *vice versa*, creating a feedback loop. Consequently, fitting a fixed effect for mortality of back neighbours might indirectly correct for the mortality observed in the focal cage itself, at least partly.

As in the study of Ellen et al. (2008), covariances between direct and indirect genetic effects were positive for line W1 and negative for line WB. As described by these authors, with death due to cannibalism it is expected that the covariance between DGE and IGE will be negative because of strong competition. A positive covariance would mean that hens benefit from not harming others (Bijma et al., 2007b). However, in our study, covariance estimates were not significantly different from zero.

3.4 Conclusions

Our results indicate that including timing of IGE expression in analysis of survival reduces the accuracy of EBV for survival. Moreover, our results show that repeated measures models improve accuracy of EBV for survival time in laying hens. Although there was no universal best method, accuracies of EBV increased by up to 20 and 12 % for lines WB and W1, respectively. Thus, it is important to use methods that can incorporate censoring when analysing survival data, such as using a repeated measures model instead of a general linear mixed model to analyse survival data. This is an important finding since more accurate EBV contribute to increased rates of genetic gain.

3.5 Acknowledgements

We would like to thank the employees of the laying houses for taking care of the hens and collecting the data on survival. The data of this research was part of a joint project of the Institut de Sélection Animale B.V., a Hendrix Genetics Company, and Wageningen University on “Genetics of robustness in laying hens”, which was financially supported by SenterNovem. Furthermore, this research was supported by the Dutch Technology Foundation STW, which is part of the Netherlands

3 Predicting direct and indirect breeding values

Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

4

Estimation of total genetic effects for survival time in crossbred laying hens showing cannibalism, using pedigree or genomic information

Tessa Brinker¹, Biaty Raymond¹, Piter Bijma¹, Addie Vereijken², Esther D Ellen¹

¹Animal Breeding and Genomics Centre, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands; ²Hendrix Genetics, Research and Technology Centre, P.O. Box 114, 5830 AC Boxmeer, The Netherlands;

Journal of Animal Breeding and Genetics (2016) 134:60-68

Abstract

Mortality of laying hens due to cannibalism is a major problem in the egg-laying industry. Survival depends on two genetic effects: the direct genetic effect of the individual itself (DGE) and the indirect genetic effects of its group mates (IGE). For hens housed in sire-family groups, DGE and IGE cannot be estimated using pedigree information, but the combined effect of DGE and IGE is estimated in the total breeding value (TBV). Genomic information provides information on actual genetic relationships between individuals and might be a tool to improve TBV accuracy. We investigated whether genomic information of the sire increased TBV accuracy compared with pedigree information, and we estimated genetic parameters for survival time. A sire model with pedigree information (BLUP) and a sire model with genomic information (ssGBLUP) were used. We used survival time records of 7290 crossbred offspring with intact beaks from four crosses. Cross-validation was used to compare the models. Using ssGBLUP did not improve TBV accuracy compared with BLUP which is probably due to the limited number of sires available per cross (~ 50). Genetic parameter estimates were similar for BLUP and ssGBLUP. For both BLUP and ssGBLUP, total heritable variance (T^2), expressed as a proportion of phenotypic variance, ranged from 0.03 ± 0.04 to 0.25 ± 0.09 . Further research is needed on breeding value estimation for socially affected traits measured on individuals kept in single-family groups.

Key words: BLUP, laying hens, social interactions, ssGBLUP, total genetic effect

4.1 Introduction

Mortality of laying hens due to cannibalism is one of the major welfare and economic problems in the egg-laying industry (Hocking et al., 2004). An increase in mortality is expected, because cage systems are banned within the European Union since 2012 and beak trimming is expected to be banned in 2018. There is therefore an urgent need for methods to reduce cannibalism in laying hens. Genetic selection, alongside improved management, is a promising tool that can lead to a gradual but permanent reduction of cannibalism (Craig and Muir, 1996).

Cannibalism is affected by social interactions among individuals. With social interactions, the phenotype of an individual depends on two genetic effects: the direct genetic effect (DGE) of the genotype of the individual itself and the indirect genetic effect (IGE) of the genotype of its group mates (e.g. Griffing, 1967; Muir, 2005; Bijma et al., 2007a). An IGE is therefore a heritable effect of an individual on phenotypes of its social partners. The use of methods that consider only DGE when performing genetic selection on traits affected by social interactions may yield unfavourable results and have sometimes even resulted in selection responses in the opposite direction (Muir et al., 2013). To improve traits affected by social interactions, it is therefore important to consider both DGE and IGE.

Indirect genetic effects contribute 33 - 76% of the heritable variance in survival time in purebred and crossbred laying hens (Ellen et al., 2008; Peeters et al., 2012). When hens are housed in groups composed of multiple families, then DGE and IGE can be estimated from genetic relationships between individuals, using pedigree information (Muir, 2005; Ellen et al., 2008). Layer breeding companies, however, often use recurrent testing, where hens are housed in sire-family groups and dam pedigree is unknown. When hens are housed in sire-family groups, DGE and IGE cannot be estimated separately using pedigree information, because DGE and IGE are fully confounded. Layer breeding companies use a sire model instead, where the total genetic effect of the sire can be estimated. The total genetic effect is the linear combination of the sire DGE and the sire IGE and is the quantity relevant for response to selection (Peeters, 2015).

Genomic data provide information on the actual genetic relationships between individuals (Yang et al., 2010). Actual genetic relationships between individuals vary around their expected value based on pedigree information because of linkage and Mendelian sampling (Hill and Weir, 2011). It is therefore expected that total genetic

effect predictions can be improved using actual genetic relationships calculated from genomic information (Meuwissen, 2007). Moreover, in recurrent test data where only the sire pedigree is known, the dam pedigree can be reconstructed from genomic data. Hence, this reconstructed pedigree distinguishes full-sibs from half-sibs, which may allow estimation of DGE and IGE even when cage mates all have the same sire (but not the same dam).

The aim of this study was to evaluate the potential benefit of using genomic information in estimation of genetic parameters and breeding values for total genetic effects for survival time of laying hens kept in sire-family groups. We used phenotypic data on survival time (days) of four crossbred layer populations with intact beaks, and genomic data on both the crossbreds and their parents. First, we investigated whether direct and indirect genetic parameters could be estimated from the reconstructed dam pedigree or from the actual genomic relationships. However, the majority of these analyses did not converge and we therefore used a sire model which directly estimates the total breeding value. Estimates from a sire model using genomic information were compared to those from a sire model using pedigree information. We used cross-validation to assess predictive abilities of both models. As a second objective, we present estimates of genetic parameters for survival time in crossbred laying hens with intact beaks, for which few estimates are available at present.

4.2 Materials and methods

4.2.1 Genetic stock and pedigree

Data on four crossbred White Leghorn layer lines were provided by Institut de Sélection Animale (ISA) B.V., the layer breeding division of Hendrix Genetics. Hendrix Genetics complies with the Dutch law on animal well-being. Crossbreds descended from one sire line (W1) and four dam lines (WA, WB, WC and WD). The four crosses were coded W1*WA, W1*WB, W1*WC and W1*WD.

A total of 209 sires and 4275 dams were used, with 48-57 sires per cross (**Table 4.1**). Each sire was mated to approximately twenty randomly selected dams, resulting in approximately two female offspring per dam. The sire pedigree was recorded for all offspring. Dam pedigree was unknown.

4.2.2 Housing conditions

Chickens of the four crosses hatched simultaneously in the Netherlands. One-day old chickens were vaccinated, wing banded, and transported to Canada. During rearing, chickens were kept in seven pens containing approximately 2000 chickens each. Chickens of the same cross were placed in the same pen. Chickens had intact beaks. Chickens were transported to a laying house at approximately 17 weeks of age. The laying house consisted of two wings: wing 1 and wing 2. In each wing, rows were grouped into six double rows containing two levels (top and bottom), with a corridor between each double row to allow access to cages. In each wing, hens were allocated to battery cages. Each cage contained five individuals of the same sire because this allows the egg-laying industry to measure the number of eggs per sire. Thus, cage mates were either paternal half-sibs or full-sibs. A feeding trough was in front of the cages and each cage had its own drinking nipples. Hens were fed a standard commercial layer diet. The diet and water were provided *ad libitum*. The light period was 10.5 h/day at 16 weeks of age, and increased every week by 1/2 h, until 15 h/day. The light intensity was higher at the top level compared with the bottom level due to closer proximity to light sources.

4.2.3 Data

Individual survival (dead or alive, 1/0) was recorded daily. Dead hens were removed and not replaced. The study was terminated when hens were approximately 75 weeks of age. Survival time was defined as the number of days from the start of the laying period until either death or the end of the study, with a maximum of 402 days. Hence, hens that were alive at the end of the laying period received a survival time of 402 days. Cages with less than five hens and cages with mistakes in composition were removed from the dataset. **Table 4.1** summarizes the data used for analyses.

4.2.4 Genotyping and SNP quality

Institut de Sélection Animale genotyped parents using DNA extracted from blood. Birds were genotyped using a custom-made Illumina 60 K chicken SNP BeadChip, which has 52 232 SNPs across chromosomes 1 through 28, Z, W, two unmapped linkage groups and some unassigned. PLINK (Purcell et al., 2007) was used for the quality control of genotypes. SNPs with a missing rate > 0.30, MAF < 0.005, SNPs with no physical position on the genome, SNPs located on sex chromosomes, and SNPs that did not have all genotypes present in the data were removed. Individuals with a call rate < 90% were also removed. The number of genotyped sires and number of SNPs available after quality control for each cross are shown in **Table 4.1**.

4 Estimation of total genetic effects

Table 4.1 Breeding scheme of four crossbred layer lines and number of genotypes¹

Cross		n	#Cages	#Sires	#Dams	#Maternal grandsires	#Genotyped sires	#SNP
Sire line	Dam line							
W1	WA	1930	386	54	1097	53	51	34 411
W1	WB	1940	388	57	1146	69	55	35 222
W1	WC	1670	334	48	993	56	42	34 649
W1	WD	1750	350	50	1039	80	48	34 942

¹Number of sires and dams used and total number of crossbred offspring are shown for crosses W1*WA, W1*WB, W1*WC and W1*WD. The number of genotyped sires and number of SNPs available are also shown.

4.2.5 Statistical analysis

Data were analysed separately for each cross, because average mortality differed clearly between crosses (see Results). First, we investigated whether direct and indirect genetic parameters could be estimated separately. This might be feasible, because reconstruction of the dam pedigree or the use of actual genomic relationships creates variation in relatedness among cage mates, which is required to separate DGE and IGE (Peeters, 2015). Hence, we fitted an animal model with both DGE and IGE, using either a pedigree relationship matrix or a genomic relationship matrix. However, most ReML analyses failed to converge. We therefore moved to sire models, using either genomic or pedigree relationships, which were implemented in ASReml (Gilmour et al., 2014).

Pedigree-based sire model (BLUP)

The model was,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{s} + \mathbf{V}\mathbf{c} + \mathbf{e}, \quad (1)$$

where \mathbf{y} is a vector of individual survival time records (days), \mathbf{b} a vector of fixed effects of the row, level and wing of the laying house; \mathbf{s} a vector of sire breeding values, \mathbf{c} a vector of random cage effects, and \mathbf{e} a vector of residuals. \mathbf{X} , \mathbf{Z} and \mathbf{V} are design matrices. All random effects, including the residuals, were assumed to be normally distributed. The covariance structures for model terms were: $\text{var}[\mathbf{s}] = \mathbf{A}\sigma_s^2$, $\text{var}[\mathbf{c}] = \mathbf{I}\sigma_c^2$, and $\text{var}[\mathbf{e}] = \mathbf{I}\sigma_e^2$, where \mathbf{A} is a pedigree relationship matrix including five generations of pedigree, σ_s^2 is the sire variance, \mathbf{I} an identity matrix, σ_c^2 the cage variance, and σ_e^2 the residual variance.

As cage mates had the same sire, the sire effect estimated from this model contains both the direct genetic effect (A_D) and the indirect genetic effect (A_I) of the sire and is an estimate of the total breeding value (A_T) of the sire (Peeters, 2015),

$$s = \frac{1}{2}A_T = \frac{1}{2}[A_D + (n-1)A_I], \quad (2)$$

where n denotes group size ($n = 5$ here). Thus, the sire variance is an estimate of one quarter of the total genetic variance,

$$\sigma_s^2 = \frac{1}{4}\sigma_{A_T}^2 = \frac{1}{4}[\sigma_{A_D}^2 + 2(n-1)\sigma_{A_{DI}} + (n-1)^2\sigma_{A_I}^2], \quad (3)$$

where $\sigma_{A_D}^2$ ($\sigma_{A_I}^2$) is the direct (indirect) genetic variance and $\sigma_{A_{DI}}$ is the covariance. The total genetic variance represents the potential of a population to respond to selection (see Bijma et al., 2007b) The random cage effect was fitted to account for non-genetic covariance between phenotypes of cage members, and to avoid overestimation of the genetic variance.

Single-step sire model (ssGBLUP)

As genotypes were lacking for 13 sires, data were analysed by combining genomic and pedigree information using the single-step procedure (ssGBLUP). The model is similar to Equation 1 except that relationship matrix \mathbf{A} was replaced by the single-step relationship matrix \mathbf{H} (Aguilar et al., 2010; Christensen and Lund ,2010). \mathbf{H}^{-1} was constructed with `calc_grm` (Calus, 2015), using the method of Aguilar et al. (2010) and Christensen and Lund (2010),

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (\tilde{\mathbf{G}}^{-1} - \mathbf{A}_{22}^{-1}) \end{bmatrix}, \quad (4)$$

where $\tilde{\mathbf{G}}$ is the genomic relationship after regressing it to the pedigree relationship matrix, $\tilde{\mathbf{G}} = \alpha\mathbf{G} + (1 - \alpha)\mathbf{A}_{22}$, with $\alpha=0.80$ for line W1*WA, $\alpha=0.95$ for lines W1*WB, W1*WC, and W1*WD. \mathbf{G} was constructed following Yang et al. (2010).

Cross-validation

Cross-validation was used to validate estimated total breeding values (TBV) of sires from the two models. With this procedure, known phenotypes are set to missing, their values are predicted, and the correlation between predicted and observed

values measures the quality of the predictions. Mutually exclusive subsets were created by setting offspring phenotypes of a single sire to missing, resulting in 54 subsets for W1*WA, 57 for W1*WB, 48 for W1*WC, and 50 for W1*WD (**Table 4.1**).

For hens still alive at the end of the laying period, true phenotypes were unknown. These individuals still provide important information as they had the highest survival time. To make as few assumptions as possible on their order of death, we followed Ellen et al. (2010). In summary, it was assumed that censored animals died in random order after surviving up to 402 days. Under this assumption, censored animals can be given the average rank of all censored animals.

Observed phenotypes for uncensored individuals were simply the ranks of observed survival time corrected for fixed effects ($P_i - \bar{P}$), which were estimated using a linear model with only fixed effects. Hence, we used ranks instead of phenotypes to combine information on both censored and non-censored individuals (Ellen et al., 2010). Note that the objective was to validate estimated total breeding values (TBV) of sires. When cages consist of half-sibs, the sire TBV surfaces in the average phenotype of the cage (Ellen et al., 2007; Peeters, 2015). Thus, validation was based on the average ranks of cages ($rank(\bar{P}_c) = \frac{1}{n_w} \sum_1^{n_w} rank(P_i - \bar{P})$; n_w is the number of individuals within each cage, $n_w=5$), rather than on individual phenotypes. Therefore, the final observed phenotype of a sire (P_{obs}) was the average of the cage-average ranks of its offspring ($P_{obs} = \frac{1}{n_c} \sum_1^{n_c} rank(\bar{P}_c)$; n_c is the number of cages per sire). The predicted phenotype (\hat{P}) of a sire was the rank of half the estimated TBV of the sire.

Approximate accuracies of estimated TBV of sires were calculated for both models from the Pearson correlation of predicted and observed ranks (hereafter rank correlation; Ellen et al., 2010). The expected rank correlation is approximately the product of the accuracy of the estimated TBV and the correlation between the true TBV and the phenotype. Thus, the approximate accuracy of estimated TBV of sires (\hat{r}_{IH}) was calculated by dividing the rank correlation by the correlation between the true TBV of the sire and the mean phenotype of its offspring (r),

$$\hat{r}_{IH} = corr[P_{obs}, \hat{P}] / r, \quad (5)$$

Where

$$r = \sqrt{\frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2 / \bar{n}_c + \sigma_e^2 / (\bar{n}_c / n_w)}}, \quad (6)$$

where \bar{n}_c is the average number of cages per sire ($\bar{n}_c=7.2$ for WA, 6.8 for WB, 7.0 for WC and 7.0 for WD). The numerator of Equation 6 represents the variance in progeny averages due to the sire and the denominator the total variance of progeny averages, and estimates were taken from BLUP. Note that the latter does not affect model comparison, as the same value was used for both models.

4.3 Results and Discussion

4.3.1 Survival and survival time

Large differences in survival time (days) and survival (%) were found between the four crosses (**Figure 1, Table 4.2**). Cross W1*WC had the lowest survival ($63.1 \pm 1.2\%$) and lowest mean survival time (324 ± 3 days). Cross W1*WA had the highest survival

Table 4.2 Number of birds (n), average survival time¹ (d) with SE and mean survival rate² (%) with SE of four crossbred layer lines and fixed effects

	n	Survival time ¹ (d)	Survival ² (%)
Cross			
W1*WA	1930	364.6 ± 2.0	78.3 ± 0.9
W1*WB	1940	349.8 ± 2.4	75.1 ± 1.0
W1*WC	1670	324.0 ± 3.0	63.1 ± 1.2
W1*WD	1750	543.3 ± 2.6	78.0 ± 1.0
Wing			
1	2335	353.8 ± 2.1	76.0 ± 0.9
2	4955	346.6 ± 1.6	72.9 ± 0.6
Row			
1	905	349.7 ± 3.6	74.7 ± 1.4
2	1150	360.6 ± 2.8	78.3 ± 1.2
3	1100	355.8 ± 2.9	75.7 ± 1.3
4	1160	350.9 ± 3.1	73.7 ± 1.3
5	1290	338.4 ± 3.3	71.0 ± 1.3
6	1285	347.5 ± 3.0	72.8 ± 1.2
7	400	327.1 ± 6.5	67.8 ± 2.3
Level			
1	2885	337.0 ± 2.2	69.3 ± 0.9
2	4405	356.7 ± 1.5	76.9 ± 0.6

¹Survival time is the average number of days from the start of the study (on average 17 weeks old) till either death or the end of the study. ²Survival rate is the percentage of laying hens still alive at the end of the study.

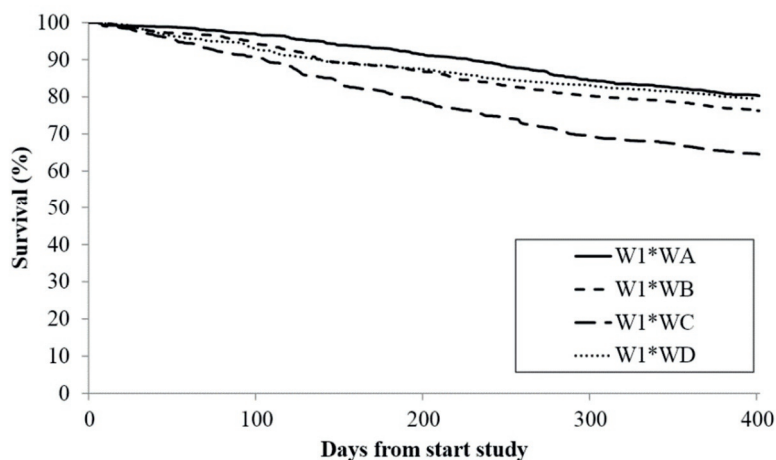


Figure 4.1 Survival curve of four crossbred layer lines, W1*WA, W1*WB, W1*WC and W1*WD

($78.3 \pm 0.9\%$) and highest mean survival time (365 ± 2 days). These results agree with previous studies showing that feather pecking differed between strains or breeds (Hughes and Duncan, 1972; Hocking et al., 2004) and that layer lines differ in survival and survival time (Ellen et al., 2008; Peeters et al. 2012; Alemu et al. 2016). Peeters et al. (2012) reported a higher mortality in crossbred lines compared with purebred lines. In their study, a higher mortality in crossbred lines was found compared with our study.

Differences in survival and survival time were found between wings, levels and rows of the laying house (**Table 4.2**). The top level had a lower survival and survival time than the bottom level. Previous studies have shown that light intensity affects survival and survival time of laying hens (e.g. Hughes and Duncan, 1972). In our study, the top level was in closer proximity to light sources, which might have caused the lower survival and survival time compared with the bottom level.

Table 4.3 Estimates of genetic parameters with SE for total genetic effect on survival time in four crossbred layer lines¹ using a pedigree-based sire model (BLUP)

Parameters ²	W1*WA	W1*WB	W1*WC	W1*WD
σ_s^2	50 ± 74	306 ± 169	925 ± 352	594 ± 245
$\sigma_{A_T}^2$	200 ± 297	1225 ± 676	3700 ± 1408	2377 ± 981
σ_{A_T}	14 ± 10	35 ± 10	61 ± 12	49 ± 10
σ_c^2	750 ± 175	1557 ± 281	1896 ± 389	2005 ± 317
σ_p^2	7665 ± 255	11224 ± 391	15067 ± 601	11081 ± 439
T^2	0.03 ± 0.04	0.11 ± 0.06	0.25 ± 0.09	0.21 ± 0.09

¹Estimates of genetic parameters are shown for survival time in four crossbred lines W1*WA, W1*WB, W1*WC and W1*WD. ² σ_c^2 , σ_s^2 , $\sigma_{A_T}^2$, and σ_p^2 are cage variance, genetic variation due to sire, total genetic variation ($\sigma_{A_T}^2 = 4\sigma_s^2$) and phenotypic variation ($\sigma_p^2 = \sigma_s^2 + \sigma_c^2 + \sigma_e^2$), respectively. T^2 is total heritable variance relative to phenotypic variance; $T^2 = \frac{\sigma_{A_T}^2}{\sigma_p^2}$.

Table 4.4 Estimates of genetic parameters with SE for total genetic effect on survival time in four crossbred layer lines¹ using a genomic-and pedigree-based sire model (ssGBLUP)

Parameters ²	W1*WA	W1*WB	W1*WC	W1*WD
σ_s^2	53 ± 73	247 ± 159	960 ± 365	586 ± 250
$\sigma_{A_T}^2$	213 ± 293	990 ± 636	3838 ± 1460	2346 ± 1000
σ_{A_T}	15 ± 10	31 ± 10	62 ± 12	48 ± 10
σ_c^2	747 ± 175	1610 ± 286	1880 ± 386	2031 ± 320
σ_p^2	7665 ± 255	11218 ± 389	15087 ± 608	11099 ± 441
T^2	0.03 ± 0.04	0.09 ± 0.06	0.25 ± 0.09	0.21 ± 0.09

¹Estimates of genetic parameters are shown for survival time in four crossbred lines W1*WA, W1*WB, W1*WC and W1*WD. ² σ_c^2 , σ_s^2 , $\sigma_{A_T}^2$, and σ_p^2 are cage variance, genetic variation due to sire, total genetic variation ($\sigma_{A_T}^2 = 4\sigma_s^2$) and phenotypic variation ($\sigma_p^2 = \sigma_s^2 + \sigma_c^2 + \sigma_e^2$), respectively. T^2 is total heritable variance relative to phenotypic variance; $T^2 = \frac{\sigma_{A_T}^2}{\sigma_p^2}$.

4.3.2 Genetic parameters

Table 4.3 shows estimated genetic parameters from BLUP and **Table 4.4** shows estimated genetic parameters from ssGBLUP. Results from both methods showed that the four crosses differed in $\sigma_{A_T}^2$. Genetic parameter estimates were similar for BLUP and ssGBLUP. Total genetic variance, expressed as a proportion of phenotypic variance (T^2), ranged from 0.03 for cross W1*WA to 0.25 for cross W1*WC. Because our cages were composed of family members, the genetic parameters in **Table 4.3**

and **Table 4.4** refer to total genetic effects, including both DGE and IGE (see Methods). Total genetic standard deviations ranged from 14 days for W1*WA to 62 days for W1*WC. Peeters et al. (2012) found values of ~ 60 days in crossbred laying hens. Together with the theoretical work of Ellen et al. (2007), Peeters et al. (2013), and Muir et al. (2013), these results suggest that mortality due to cannibalism can be reduced by selection using a sire model, even though underlying DGE and IGE are unknown, because sire models capture the TBV. This is supported by evidence of Muir (1996) that showed that hens housed in sire-family groups can successfully be selected against mortality.

Because we used a sire model, the estimates refer to genetic variation for crossbred performance within the sire line. The sire line was the same for all crosses. Nevertheless, variance components differed considerably between crosses (**Table 4.3** and **Table 4.4**). The large differences in the sire variance between crosses correspond, however, reasonably well with the large differences observed in mean survival and in phenotypic variance (**Table 4.2** and **Figure 4.1**). When mean survival increases, a larger fraction of individuals is censored which reduces phenotypic variance (as censored individuals were all given the maximum survival time). This reduces phenotypic variance, and also heritabilities. Although we cannot mathematically prove it, we think that the latter is similar to the situation with a 0/1 trait, where heritabilities on the observed scale decrease when prevalence is approaching zero (Dempster and Lerner, 1950).

4.3.3 Cross-validation

Table 4.5 shows the rank correlations between observed and predicted phenotypes, $\text{corr}[P_{\text{obs}}, \hat{P}]$, for BLUP and ssGBLUP. Most rank correlations were not significantly different from zero, except for W1*WD with sire BLUP and for W1*WC with ssGBLUP. Results varied between crosses and showed no convincing pattern.

Using data on brown layer lines, Alemu et al. (2016) also compared BLUP to ssGBLUP using a sire model. They found ~33 % improved accuracy with ssGBLUP compared with BLUP, whereas our study yielded inconsistent results. Alemu et al. (2016) also investigated the benefit of ssGBLUP for response to selection and found that increases in response were substantially larger (~90%) than those in accuracy. A major contribution of this increase originated from higher accuracy of selection in females.

Table 4.5 Rank correlations between observed phenotype and predicted phenotype with SE

Cross ²	Models - sire ¹		Approximate accuracy	
	Sire BLUP	ssGBLUP	Sire BLUP	ssGBLUP
W1*WA	0.162 ± 0.134	0.154 ± 0.134	0.42	0.41
W1*WB	-0.020 ± 0.134	-0.152 ± 0.131	-0.03	-0.25
W1*WC	0.196 ± 0.140	0.390 ± 0.124	0.25	0.50
W1*WD	0.412 ± 0.119	0.143 ± 0.140	0.57	0.20

¹Rank correlations between observed and predicted phenotypes using a pedigree-based sire model (BLUP) and genomic-and pedigree-based sire model (ssGBLUP). ²Rank correlations are shown for four crosses W1*WA, W1*WB, W1*WC and W1*WD.

One of the factors that may have benefited ssGBLUP in the research of Alemu et al. (2016) is that multiple generations of phenotypes and sire genotypes were used, five for one line and three for the other line. Research of Muir (2007) showed that with an increasing number of generations of data, GEBV accuracies increase as well, whereas BLUP accuracies plateau. In our research, we had only a single generation of phenotypes that have sire genotypes available. We, however, think that the difference between results from Alemu et al. (2016) and results from our study is mainly due to the large difference in the number of sires between both studies. We had ~50 sires per cross, whereas Alemu et al. (2016) had ~500 for one sire line and ~280 for another sire line. The training population used in Alemu et al. (2016) was consequently larger compared with our training populations. The size of the training data partly determines the success of genomic prediction (Daetwyler et al., 2008). We investigated whether it is theoretically expected to find significant accuracies, given the number of sires. First, we calculated the theoretically expected accuracies (Daetwyler et al., 2008),

$$r_{\hat{g}g} = \sqrt{\frac{r^2}{r^2 + \frac{M_e}{N_p}}} \quad (7)$$

where r^2 is the reliability of sire TBV which is the square of Equation 6, M_e is the effective number of chromosome segments (M_e is 565 for W1*WA, 579 for W*WB, 771 for W1*WC and 415 for W1*WD), and N_p is the number of sires in the training population. Secondly, we calculated the corresponding standard error ($SE(\hat{r}_{\hat{g}g})$) (Bijma and Bastiaansen, 2014):

$$SE(\hat{r}_{gg}) = \frac{1-r_{gg}^2}{\sqrt{N_p-1}}, \quad (8)$$

Theoretically expected accuracies and corresponding SE were 0.12 ± 0.14 for W1*WA, 0.19 ± 0.13 for W1*WB, 0.19 ± 0.14 for W1*WC, and 0.24 ± 0.14 for W1*WD. Thus, none of the theoretically expected accuracies are expected to be significantly different from 0. This corresponds with our cross-validation results where most rank correlations were not significantly different from 0. This implies that our study would benefit from an increased training population. An option would be to combine phenotypic data of all four crosses, which may be feasible because the crosses originate from a single sire line resulting in ~ 200 genotyped sires.

4.3.4 Model

Survival time was analysed using a linear mixed model. Shortcomings of this model are that censored records are considered as true survival time records and that the non-normality of survival time records violates the normality assumption of the linear mixed model. Muranty et al. (2015) observed that genomic prediction accuracy is highly influenced by phenotypic distribution. The choice of model may therefore have affected the outcome of our results.

Survival analysis, for example, can deal with both censoring and non-normality of survival data (Kalbfleisch and Prentice, 1980). Ellen et al. (2010) found, however, no benefit of survival analysis compared with analysing survival time using a linear mixed model when social interactions are included in the model. Brinker et al. (2015) proposed to analyse survival as a repeated binomial trait to account for censoring. TBV accuracies improved up to 21% compared with analysing survival time using a linear mixed model. We applied this model to our survival data of W1*WB (worst ssGBLUP predictions) and W1*WC (best ssGBLUP predictions). We analysed the repeated binomial survival records with a logit link function to account for the non-linearity of the trait. Accuracies from BLUP were 0.018 ± 0.134 for cross W1*WB, and 0.290 ± 0.134 for cross W1*WC. Accuracies from ssGBLUP were -0.125 ± 0.132 for W1*WB, and 0.395 ± 0.123 for W1*WC. Accuracies improved for both BLUP and ssGBLUP when analysing repeated binomial survival compared with analysing survival time. However, both methods lead to the same conclusion: using ssGBLUP did not improve accuracies for W1*WB, but it did for W1*WC.

We have investigated the benefit of genomic information for the prediction of breeding values for a socially affected trait. Despite the large number of genotyped

and phenotyped individuals, we found little benefit of genomic information. This result originated from the difficulty in estimating DGE and IGE when cage mates are related, even when genomic data are available. In our data, social groups (i.e. the cages) consisted of a mix of full-sibs and half-sibs. All cage mates had the same sire, but in most cases a different dam. Genotypes were available on nearly all individuals, and on both sires and dams, and dam pedigrees were reconstructed based on the genotypes. While this provided some variation in relatedness among cage mates, this appeared insufficient to separate DGE from IGE, as illustrated by the convergence problems of the DGE-IGE animal models.

An alternative approach would have been to fit an animal model with only the DGE of the individuals. However, when cages are composed of a mix of relatives, the interpretation of the estimates of a DGE-model is unclear. In an animal model, the DGE of an individual will partly pick up the IGE of its group mates when group mates are related (Muir et al., 2013; Peeters, 2015). Hence, EBV from such a model are a mix of DGE and IGE. It is unclear how an estimate of the TBV can be obtained from those estimates, and how cross-validation can be done. Similar issues occur in a sire-dam model. For those reasons, we used a sire model. The sire model yields an estimate of the TBV, because all cage mates descended from the same sire. However, with a sire model, only ~50 genotyped individuals (i.e. sires) were available per cross as a reference population, which proved insufficient to find a benefit of genomic information. In conclusion, because of the close relationships between cage mates in our data, we did not manage to utilize the large amount of genotypic data (~7000 reference individuals) for the genomic prediction of TBV. An option to utilize the data better would be to use pooled cage observations for variance components estimation and cross-validation. Analyses of pooled cage observations when data are housed in half-sib families yield estimates of the TBV and its variance (Peeters et al., 2013). This approach may better reveal the benefit of genomic selection, because we can utilize ~1400 pooled records of phenotypes and genotypes of cage mates. Further research is needed on how genomic information can be used for breeding value estimation on socially affected traits measured on individuals kept in groups composed of a mix of close relatives.

4.4 Conclusion

The estimated total genetic variation available for selection was similar for BLUP compared with ssGBLUP. Cross-validation results showed no improvement in the accuracy of breeding value predictions, probably because of the limited number of

sires and the use of a sire model. Further research is needed on the estimation of breeding values for socially affected traits measured on individuals kept in groups composed of a mix of close relatives.

4.5 Acknowledgements

We would like to thank the employees of the laying houses for taking care of hens and data collection on survival. Data of this research was part of a joint project of Institut de Sélection Animale B.V., a Hendrix Genetics Company and Wageningen University on “Genomic solutions for socially-affected traits”, which was financially supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO) and which is partly funded by the Ministry of Economic Affairs.

5

The genetic architecture of socially-affected traits: a GWAS for direct and indirect genetic effects on survival in laying hens showing cannibalism

Tessa Brinker¹, Piter Bijma¹, Addie Vereijken² and Esther D Ellen¹

¹Animal Breeding and Genomics Centre, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands; ²Hendrix Genetics, Research and Technology Centre, P.O. Box 114, 5830 AC Boxmeer, The Netherlands;

Abstract

Cannibalism is an important welfare problem in the layer industry. Cannibalism is a social behaviour where individual survival is affected by direct genetic effects (DGE) and indirect genetic effects (IGE). Previous studies analysed repeated binomial survival, instead of survival time, which improved accuracies of breeding value predictions. Our study aimed at identifying SNPs associated with DGE and IGE for survival time, and comparing results from models that analyse survival time and repeated binomial survival.

Survival data of three layer crosses (W1*WA, W1*WB, and W1*WC) were used. Each individual had one survival time record and 13 monthly survival (0/1) records. Approximately 30 000 single nucleotide polymorphisms (SNPs) were included in the genome-wide association study (GWAS), using a linear mixed model for survival time, a linear mixed model and a generalized linear mixed model for repeated binomial survival (0/1). Backwards elimination was used to determine phenotypic and genetic variance explained by SNPs.

The same quantitative trait loci were identified with all models. A SNP associated with DGE was found in cross W1*WA, with an allele substitution effect of 22 days. This SNP explained 3% of the phenotypic variance, and 36% of the total genetic variance. Four SNPs associated with DGE were found in cross W1*WB, with effects ranging from 16 to 35 days. These SNPs explained 1 to 6% of the phenotypic variance and 9 to 44% of the total genetic variance. Our results suggest a link of DGE and IGE for survival time in layers with the gamma-aminobutyric acid (GABA) system, since a SNP located near a gene for a GABA receptor was associated with DGE and with IGE (not significant).

This is one of the first large studies investigating the genetic architecture of a socially-affected trait. The power to detect SNP associations was relatively low and thus we expect that many effects on DGE and IGE remained undetected. Yet, GWAS results revealed SNPs with large DGE and a link of DGE and IGE for survival time in layers with the GABAergic system, which supports existing evidence for the involvement of GABA in the development of abnormal behaviours.

Key words: GWAS, laying hens, social interactions

5.1 Background

Mortality due to cannibalism has important welfare and economic implications in the commercial laying hen industry. Brinker et al. (2016) reported mortality to be between 22 and 37% in crossbred chickens with intact beaks, while beak-trimmed hens of the same crosses showed a mortality of 2 to 3% at the end of the laying period (personal communication J. Visscher), which indicates that a substantial part of the mortality is due to cannibalism. Previous research revealed that survival time is affected both by an individual's own genes (direct genetic effects; DGE) and by genes of its group mates (indirect genetic effects; IGE). It was found that IGE contribute 33 to 76% of the heritable variation in survival time in purebred and crossbred laying hens with intact beaks (Peeters et al., 2012; Ellen et al., 2008). However, the genetic architecture of survival time in laying hens that show cannibalism remains largely unknown.

The availability of genomic information has increased our understanding of complex traits, but studies have mainly focussed on DGE. Results from genome-wide association studies (GWAS) on DGE show that most quantitative traits in livestock are highly polygenic and that variants tend to be associated with more than one trait (Hill et al., 2008). However, the genetic architecture of IGE may differ from the genetic architecture of DGE. For example, IGE are less exposed to natural selection compared to DGE (Denison et al., 2003), and therefore we expect that some loci may have large effects for IGE. A few studies have investigated the genetic architecture of IGE. Biscarini et al. (2010) conducted an association study using 1022 single nucleotide polymorphisms (SNPs) and identified 81 SNPs that were associated with IGE for plumage condition in laying hens. However, the number of observations used was limited; 662 laying hens originating from nine lines were used for analyses. Mutic and Wolf (2007) identified 13 quantitative trait loci (QTL) for IGE associated with size, development, and fitness related traits in *Arabidopsis*. To increase the power of IGE detection, they did not consider loci that did not have DGE in their IGE analyses.

From a statistical point of view, survival time is a difficult trait because many laying hens are still alive at the end of the recording period. For these hens, true survival time cannot be observed but is known to exceed the length of the recording period (censored). Several statistical techniques have been proposed to deal with survival data and IGE, including survival time analysis (Ducrocq and Casella, 1996; Ellen et al., 2011) and the use of repeated binomial survival records (0/1) (Brinker et al., 2015). Ellen et al. (2011) showed that survival time analysis did not improve breeding value

predictions compared to analysing survival time with an ordinary mixed linear model when censoring occurs at the same moment in time. Compared to a linear mixed model analysis of survival time, the use of repeated binomial survival records (0/1) by Brinker et al. (2015) improved accuracies of breeding value predictions up to 21%. Hence, the use of repeated binomial survival records (0/1) may also be beneficial for the identification of direct and indirect SNP associations in GWAS.

This study had two aims: (1) to identify SNPs associated with direct and indirect effects for survival time in laying hens that show cannibalism; and (2) to compare GWAS results from analysis of survival time versus repeated binomial survival (0/1).

5.2 Methods

5.2.1 Genetic stock and pedigree

Data were collected under the control of Hendrix Genetics. Hendrix Genetics complies with the Dutch law on animal welfare. Hendrix Genetics provided data on three crossbred White Leghorn layer lines. Crossbreds descended from one sire line (W1) and three dam lines (WA, WB, and WC), and were coded W1*WA, W1*WB, and W1*WC. In contrast to Brinker et al. (2016), data from cross W1*WD were not used in this study because the quality of the genomic data was insufficient.

A total of 159 sires and 3218 dams were used, with 48 to 57 sires per cross. For each cross, matings between sires and dams were randomly assigned, which resulted in approximately two female offspring per dam. The sire pedigree was recorded for all offspring. The dam's pedigree was initially unknown but a reconstructed pedigree, based on genomic information, was provided by Hendrix Genetics.

5.2.2 Housing conditions

Chickens of the three crosses hatched simultaneously in the Netherlands. One-day old chickens were vaccinated, wing-banded, and transported to Canada. Chickens had intact beaks. Chickens were transported to a laying house at approximately 17 weeks of age. The laying house consisted of two wings. In each wing, rows were grouped into six double rows that each contained two levels, with a corridor between each double row to allow access to cages. Each cage contained five individuals of the same sire. Thus, cage mates were either paternal half-sibs or full-sibs. A standard commercial layer diet and water were provided ad libitum. A feeding trough was in front of the cages and each cage had its own drinking nipples. The light

Table 5.1 Data description of three crossbred layer lines, number of crossbreds phenotyped and genotyped, and number of SNPs after quality control

Sire line	Dam line	Mean survival days \pm SD	Number of phenotypes	Survival time	Survival (0/1)	Number of genotyped crossbreds	Number of SNPs
W1	WA	364.6 \pm 87.9	1920	24 960	1889	27 204	
W1	WB	349.8 \pm 107.0	1875	24 375	1816	32 473	
W1	WC	323.9 \pm 123.2	1620	21 060	1580	38 588	

intensity was stronger at the top level compared to the bottom level, due to closer proximity to light sources. Further details are in Brinker et al. (2016).

5.2.3 Data

Dead hens were removed daily but the cause of death was not determined. The wing band number and cage number were recorded after death. The study was terminated when hens were approximately 75 weeks old. Survival at the end of the study was 78, 75, and 63% for crosses W1*WA, W1*WB, and W1*WC respectively.

Cages with initially less than five hens and cages with mistakes in their composition (e.g. with hens descending from multiple sires) were removed from the dataset. Cages with hens descending from multiple sires were identified based on genomic information.

Survival time was defined as the number of days from the start of the laying period until either death or the end of the study, with a maximum of 402 days. Hens that were alive at the end of the laying period were given a survival time of 402 days. In total, records on 1920, 1875, and 1620 laying hens were used for the statistical analyses for crosses W1*WA, W1*WB, and W1*WC respectively (**Table 5.1**).

To generate repeated binomial survival records (0/1), the laying period was divided into 13 months. For each month, survival was coded 1 if the laying hen was alive at the end of that month and as 0 if not. Thus, a survival record (0/1) was available for each month. In total, 24 960, 24 375, and 21 060 monthly records were available for crosses W1*WA, W1*WB, and W1*WC respectively (**Table 5.1**).

5.2.4 Genotyping and SNP quality

Birds were genotyped based on DNA extracted from blood, using a custom made Illumina 60 K chicken SNP BeadChip, which included 52 232 SNPs across chromosomes 1 through 28, Z, W, and two unmapped linkage groups, along with some unassigned SNPs. PLINK (Chang et al., 2015; Purcell and Chang, 2015) was used for the quality control of genotypes. SNPs with a missing rate higher than 0.30 and a MAF lower than 0.005 were removed. Individuals with a missing rate higher than 0.10 were also removed. SNPs that deviated from Hardy-Weinberg Equilibrium ($p < 10^{-5}$) in the parental population were removed from the full dataset. SNPs contributing more than one Mendelian error and individuals contributing more than five Mendelian errors were also removed from the dataset. Remaining Mendelian errors were set to missing.

The number of genotyped crossbreds and the number of SNPs available after quality control for each cross are in **Table 5.1**. Some individuals were not genotyped (max. 4 %) due to death before blood sampling (~1 month after the start of the laying period) or due to poor DNA quality. To allow cages that contained individuals without genotype information to be included in the statistical analyses, missing genotypes were replaced by the “parental mean”, which was the average of the allele count of the sire and the average allele count of its mates (~ 20 dams). In case parental genotypes were missing, the line average was used. The same procedure was applied for missing genotype information, *i.e.* for genotypes associated with Mendelian errors.

5.2.5 Statistical analysis

Data were analysed separately for each cross. Three statistical models were compared: a linear mixed model for survival time (STM), a linear mixed model for repeated binomial survival (0/1; RMM.t), and a generalized linear mixed model for repeated binomial survival (0/1; GLMM). All models were implemented using ASReml (Gilmour et al., 2014).

First, genetic parameters were estimated without SNP effects in the model (see models below). Five generations of (reconstructed) pedigree information on sires and dams were included in all genetic analyses. In our data, cages consisted of paternal half-sibs, with an occasional full-sib. Therefore, direct and indirect polygenic effects were strongly confounded (Brinker et al., 2016; Peeters, 2015). With cages composed of families, an animal model with DGE only will also pick up IGE and consequently yields genetic parameters estimates that refer to the total breeding

value (Peeters, 2015). Thus, an animal model with DGE only was used to account for population stratification (*i.e.* family structure in the population). This animal model incorporated the genetic covariance structure across individuals, to avoid spurious SNP associations due to relatedness.

Second, SNP effects were estimated one by one, including both the direct SNP effect of the individual and the summed indirect SNP effects of its cage mates in the model. For all models, variance components were fixed to the estimated values from the corresponding model without the SNP effect. Direct and indirect SNP effects were fitted simultaneously, because a GWAS with direct SNP effects only would also capture part of the indirect effect of the SNP, as the related group mates have an above-average probability to carry the same alleles.

Survival time model STM

Survival time records were analysed using the following linear mixed model:

$$y_{ik} = \text{fixed} + \beta_D \cdot \text{SNP}_i + \beta_I \cdot \sum_{j \neq i}^{n-1} \text{SNP}_{j(k)} + a_i + \text{cage}_k + e_{ijk}, \quad (1)$$

where y_{ik} is the observed survival time (days) for individual i in cage k , with cage mates j , n is the number of cage members ($n=5$), and *fixed* represents the fixed effect of the combination of wing-row-level; β_D and β_I are regression coefficients of y_{ik} on SNP genotypes fitted as fixed effects, where β_D is the fixed direct effect of the SNP of individual i , and β_I is the fixed indirect effect of the same SNP in cage mates j ; SNP_i is the allele count (0,1,2) for the individual, $\sum_{j \neq i}^{n-1} \text{SNP}_{j(k)}$ is the summed SNP allele counts (0-8) of cage mates j , a_i is the random polygenic effect of individual i , cage_k is the random cage effect, and e_{ijk} is the residual. All random effects were assumed to be normally distributed. The covariance structures for the model terms were: $\text{var}(a) = \mathbf{A}\sigma_a^2$, $\text{var}(\text{cage}) = \mathbf{I}\sigma_{\text{cage}}^2$, and $\text{var}(e) = \mathbf{I}\sigma_e^2$, where \mathbf{A} is a pedigree relationship matrix, σ_a^2 is the additive genetic variance, \mathbf{I} is an identity matrix, σ_{cage}^2 is the cage variance, and σ_e^2 is the residual variance.

Repeated measures model RMM.t

Repeated binomial records on survival (0/1) were analysed using a repeated measures model that included random regressions on time (hence RMM.t). Following Brinker et al. (2015), the model was:

$$y_{ikm} = fixed + \beta_D \cdot t_m \cdot SNP_i + \beta_I \cdot t_m \cdot \sum_{j \neq i}^{n-1} SNP_{j(k)} + a_i \cdot t_m + cage_{km} + cage_k \cdot t_m + PE_i \cdot t_m + e_{ijkm}, \quad (2)$$

where y_{ikm} is the observed survival (0/1) for individual i in cage k and month m , with cage mates j , at time t_m measured in months since the start of the experiment; *fixed* represents the interaction effect of wing-row-level with time, which was a sixth-order polynomial of time, fitted as a fixed effect, which was used to model the survival curve across time and to allow this curve to depend on location (i.e., on the wing-row-level combination); $cage_{km}$ is the random effect of cage k at time m , which accounts for covariances between cage members at specific time points; $cage_k$ is the random permanent effect of cage k , with $cage_k \cdot t_m$ accounting for covariances between records on the same cage at different time points, and for increasing variance over time. Together, the $cage_{km}$ and $cage_k$ effects account for similarity of cage mates due to shared environment, which is essential to avoid inflation of genetic estimates in the analysis of socially-affected traits (Bijma et al., 2007b). Finally, PE_i is the random permanent environmental effect of individual i , t_m is the time, and e_{ijkm} is the residual. A separate residual variance was estimated for each month. Other terms are the same as in STM. More details on this model are in Brinker et al. (2015).

Generalized linear mixed model GLMM

To account for the binomial distribution of survival (0/1), we used a generalized linear mixed model with a logit link function. ASReml uses approximate likelihood techniques, of which the limitations are discussed in the Results and discussion section (Gilmour et al., 2014). The model was:

$$\eta(E(y_{ikm})) = fixed + \beta_D \cdot SNP_i + \beta_I \cdot \sum_{j \neq i}^{n-1} SNP_{j(k)} + a_i + cage_{km} + PE_i, \quad (3)$$

Where $\eta()$ is the logit link function that links the probability of surviving to the linear predictor, and $E(y_{ikm})$ is the probability of surviving for individual i in cage k , with cage mates j , at time m . The other terms are the same as in STM and RMM.t.

The GLMM includes only a genetic intercept and no regression on time because the non-linear link function takes the change in variance over time into account. Consequently, at the beginning of the recording period, the variance of the survival probabilities can be (near) zero even when $var[\eta(E(y_{ikm}))]$ is greater than zero. More details on this model are in Brinker et al. (2015).

5.2.6 Model fit

The three models, STM, RMM.t, and GLMM were compared by reviewing $-\log_{10} p$ values of SNP effects and the size of the inflation factor for p values (λ ; see below). Pearson correlations between $-\log_{10} p$ values were calculated to quantify the agreement between the three models.

5.2.7 Genomic control

A quantile-quantile plot (Q-Q plot) was used to investigate the distribution of the observed p values compared to their expected distribution under the null hypothesis that the SNP has no association with the trait. The extent of deviation of the observed distribution from the expected distribution was expressed as λ , where a λ equal to 1 means no deviation (Devlin and Roeder, 1999). In cases where λ was larger than 1.10, the so-called “genomic control” was applied to avoid spurious associations with the trait by dividing F values by λ before calculating p values (Devlin and Roeder, 1999).

Multiple testing was accounted for by controlling the false discovery rate (FDR) using the q -value package (Storey and Tibshirani, 2003) in R (R Core Team, 2015). The FDR is the expected proportion of false-positives among those that were called significant under the distribution of the p values. The q -value package calculates an FDR based on the distribution of p values, which represents the minimum FDR when the SNP effect is called significant, which was set to 0.3. This is a liberal threshold that also reveals suggestive SNP associations with survival time and was chosen because little is known about the background of DGE and IGE for survival time, this study being one of the first large ones.

5.2.8 Phenotypic and genetic variance explained by SNPs

A backwards elimination method was used to obtain an estimate for phenotypic and genetic variance explained by SNPs. Backwards elimination involved including all direct and indirect SNP effects below the genome-wide FDR threshold ($q < 0.3$) in the model to account for possible linkage disequilibrium (LD) between them, testing their model fit, and dropping the least significant SNP effect. This was repeated until all SNPs reached the FDR threshold. Then, for each remaining SNP, we calculated the genetic variance explained by the SNP following Falconer and Mackay (1996) as $V = 2p(1 - p)\alpha^2$, with p being the major allele frequency and α is the estimated direct allele substitution effect from the model with all remaining SNPs. LD between fitted SNP effects was not considered in the calculation of V .

5 GWAS for direct and indirect genetic effects

Table 5.2 Estimates¹ of genetic parameters for survival time in three crossbred layer lines using STM

	W1*WA	W1*WB	W1*WC
σ_A^2	576 ± 326	1415 ± 583	5310 ± 1386
σ_c^2	763 ± 173	1813 ± 287	1832 ± 374
σ_P^2	7645 ± 260	10 781 ± 389	15 102 ± 647
T^2	0.08 ± 0.04	0.13 ± 0.05	0.35 ± 0.08

¹Estimates of genetic parameters are shown for survival time, σ_c^2 , σ_A^2 , σ_P^2 , T^2 are cage variance, genetic variance and phenotypic variance ($\sigma_P^2 = \sigma_A^2 + \sigma_c^2 + \sigma_e^2$), $T^2 = \frac{\sigma_A^2}{\sigma_P^2}$ (Bergsma et al., 2008), respectively. All variances are in days²

Table 5.3 Pearson Correlations* between $-\log_{10} p$ -values of the three models for direct and indirect SNP effects for each cross

Effect	Cross	STM-RMM.t	STM-GLMM	RMM.t-GLMM
Direct	W1*WA	0.98	0.96	0.97
	W1*WB	0.97	0.93	0.93
	W1*WC	0.96	0.96	0.93
Indirect	W1*WA	0.98	0.95	0.96
	W1*WB	0.98	0.97	0.96
	W1*WC	0.97	0.97	0.96

* All standard errors were less than 0.01

Table 5.4 Inflation factor λ^* for all crosses and models

Effect	Model	W1*WA	W1*WB	W1*WC
Direct	STM	1.09	1.04	0.91
	RMM.t	1.13	1.18	1.09
	GLMM	1.13	1.02	0.91
Indirect	STM	0.92	1.26	1.10
	RMM.t	0.94	1.43	1.32
	GLMM	0.93	1.22	1.13

*All standard errors were less than 0.01

The proportions of phenotypic variance (σ_P^2) and of genetic variance (σ_A^2) explained by the direct effects of SNPs were calculated as $\frac{V}{\sigma_P^2}$ and $\frac{V}{\sigma_A^2}$, respectively. The σ_A^2 is an estimate of the total genetic variance since group members are related (Bijma et al., 2007b).

5.3 Results and Discussion

5.3.1 Model comparison

Genetic parameters were estimated without SNP effects in the model. For all models, variance components were fixed to the estimated values from the corresponding model without SNP effects. The σ_P^2 and σ_A^2 values are in **Table 5.2**. Variance components from RMM.t and GLMM are not presented in **Table 5.2** because they can be translated to the survival time scale, for which estimates are in **Table 5.2** (Brinker et al., 2015). The total genetic standard deviation (σ_A) was 24 days for cross W1*WA, 38 days for cross W1*WB, and 73 days for cross W1*WC.

There were no evident differences between GWAS results from STM, RMM.t, and GLMM (**Tables 5.3, 5.4**) and [see **Additional files 5.1, 5.2** and **5.3**]. The same QTL were identified with all three models. **Table 5.3** shows that correlations between $-\log_{10} p$ values for direct and indirect SNP effects from the three models were higher than 0.9, thus similar SNPs were identified as having weak(-er) or strong(-er) associations in all models. This is in line with findings of Rönnegård et al. (2016), who investigated the benefit of using uncensored repeated measures in GWAS in a simulation study with direct effects only. In a design with an equal number of observations per individual, as in our study, they found that the correlation between $-\log_{10} p$ values between results from a model fitting average phenotypes and a model fitting repeated measures was higher than 0.9 (Rönnegård et al., 2016). Rönnegård et al. (2016) concluded that a repeated measures model in GWAS was most beneficial when individual phenotypes were very different across time or when the number of observations varied among individuals.

Table 5.4 shows the inflation factor λ and the corresponding Q-Q plots are in **Additional files 5.1** and **5.2**. There was no clear pattern for λ across models. Moreover, there were no large differences in the number of SNPs with $p < 0.001$ between the three models after genomic control (Note: the aim was to compare models here, not to identify SNPs) [see **Additional file 5.3**].

In this study, we used a generalized linear mixed model (GLMM) for GWAS, which was fitted with ASReml (Gilmour et al., 2014). ASReml uses an approximate likelihood technique (the penalized quasi-likelihood), which has not been studied well for hypothesis testing. Gilmour et al. (2014) recommend to use GLMM in ASReml with caution. However, when comparing p values from the GLMM to those

of the other two models, STM and RMM.t, we found that the Q-Q plots from the GLMM behaved well and were similar to those for STM and RMM.t [see **Additional files 5.1 and 5.2**], $-\log_{10} p$ values from the GLMM were highly correlated with those from STM and RMM.t (**Table 5.3**), and Manhattan plots from GLMM showed a similar pattern as those from STM and RMM.t [see **Additional files 5.3 and 5.4**].

Based on these results, we concluded that there was no evidence that RMM.t and GLMM outperformed STM. Thus, in the remainder of this paper, we will only show results from STM.

5.3.2 GWAS results (STM)

Figures 5.1 and 5.2 show the Manhattan plots for the direct and indirect SNP effects for the three crosses, respectively. Several SNPs were associated with direct effects for survival time at $q < 0.3$ in cross W1*WA and W1*WB, but none in cross W1*WC. In none of the crosses, SNPs were associated with indirect effects for survival time at $q < 0.3$.

In cross W1*WA, 17 SNPs were associated with direct effects for survival time at $q < 0.3$. Of these, one SNP was on chromosome 4 at 42 Mb. The remaining 16 SNPs were on chromosome 2, with one SNP at 46 Mb, and 15 SNPs in the region between 87 and 89 Mb. The latter 15 SNPs were in high LD with most pairwise r^2 higher than 0.9. After backwards elimination, one SNP (rs317294317) remained in the model ($q < 0.3$, corresponding to $p < 2.05 \times 10^{-5}$, **Table 5.5**). This SNP is an intron variant at 88 Mb on chromosome 2, had an estimated effect of 22 days, and explained 3% of the phenotypic variance (σ_P^2), and 36% of the total genetic variance (σ_A^2).

In cross W1*WB, seven SNPs were associated with direct effects for survival time at $q < 0.3$. These were located on chromosome 2 (at 9 and 86 Mb), chromosome 5 (two SNPs at 54 Mb), chromosome 7 (at 6 Mb), chromosome 9 (at 17 Mb), and chromosome 20 (at 20 Mb). After backwards elimination, four SNPs remained in the model ($q < 0.3$, corresponding to $p < 2.84 \times 10^{-5}$, **Table 5.5**). Estimated effect sizes ranged from 16 to 35 days, with rs31610924 having the smallest effect size and rs312488612 having the biggest effect size. The SNPs explained 1-6% of σ_P^2 and 9-44% of σ_A^2 .

Biscarini et al. (2010) was the first to investigate the genetic architecture of both direct and indirect genetic effects of plumage condition in laying hens. They found

Table 5.5 Significant ($q < 0.30$) direct and indirect SNP effects, their location, minor allele frequency (MAF), and estimated effect size α , based on model STM

Effect	Cross	SNP	Chr	Position (kbp)	MAF	α (days)	q-value	V (days ²)	% of σ_P^2 explained	% of $\sigma_{A_T}^2$ explained
Direct	W1*WA	rs317294317	2	88 120	0.34	22 \pm 5	0.29	209	2.7	36.3
Direct	W1*WB	rs313098101	2	8799	0.32	20 \pm 5	0.16	167	1.5	11.8
		rs31610924	5	54 321	0.35	16 \pm 4	0.06	122	1.1	8.6
		rs312488612	7	5646	0.47	35 \pm 8	0.05	627	5.8	44.3
		rs14677635	9	17 018	0.22	26 \pm 6	0.28	229	2.1	16.2

11 direct associations and 81 indirect associations between SNPs and plumage condition. Our study did not confirm the large number of indirect SNP associations, although it is one of the first large GWAS that includes both DGE and IGE. We analyzed survival time, which reflects the final stage of cannibalism, while plumage condition is recorded before the end stage. This could partly explain the difference in results with Biscarini et al. (2010).

Candidate genes

Chromosome 2. In cross W1*WA, SNP rs317294317 remained in the model after backwards elimination. This SNP is an intron variant on chromosome 2. The associated gene is *GABBR2* (88.0-88.4 Mb; Fig. 5.1). A clear peak is visible for direct SNP effects in this region for cross W1*WA. Moreover, we observed a clear peak in the same region for indirect SNP effects in cross W1*WB – although not significant at $q < 0.3$ after genomic control (Fig. 5.2). The favorable allele was the same for the direct SNP effect in cross W1*WA and the indirect SNP effect in cross W1*WB. This allele has a positive effect both on the survival time of the individual itself and on the survival time of its group mates. The *GABBR2* gene was found to be associated with both direct phenotypes and behavioral phenotypes (de Almeida et al., 2005; Poshivalov, 1981; Takahashi et al., 2010; Takahashi et al., 2012; Zhang et al., 2012).

The *GABBR2* gene encodes a receptor for gamma-aminobutyric acid (GABA), which plays an important role in the regulation of neurotransmitters in the brain. GABA is an inhibitor of neuronal activity and plays an important role in physiological and behavioral stress response in many species (de Almeida et al., 2005; Poshivalov, 1981; Takahashi et al., 2010; Takahashi et al., 2012; Zhang et al., 2012). Zhang et al. (2012) for example, found that the level of GABA affects the performance and physical condition in Roman laying hens under heat stress. Furthermore, Poshivalov (1981) found that the level of GABA was associated with a change in state of aggressiveness and sociability towards conspecifics in *Mus musculus*.

In addition to GABA, serotonin and dopamine are also important neurotransmitters and are known to be associated with several behavioral disorders in a variety of species. Moreover, several studies have reported a link between the serotonergic, dopaminergic, and GABAergic pathways (Ropert and Guy, 1991; Stutzmann and LeDoux, 1999; Casey et al., 1980). Biscarini et al. (2010) investigated the genetic architecture of direct and indirect genetic effects of plumage condition in laying hens and found a SNP in the *HTR2C* gene, which is involved in the serotonergic system,

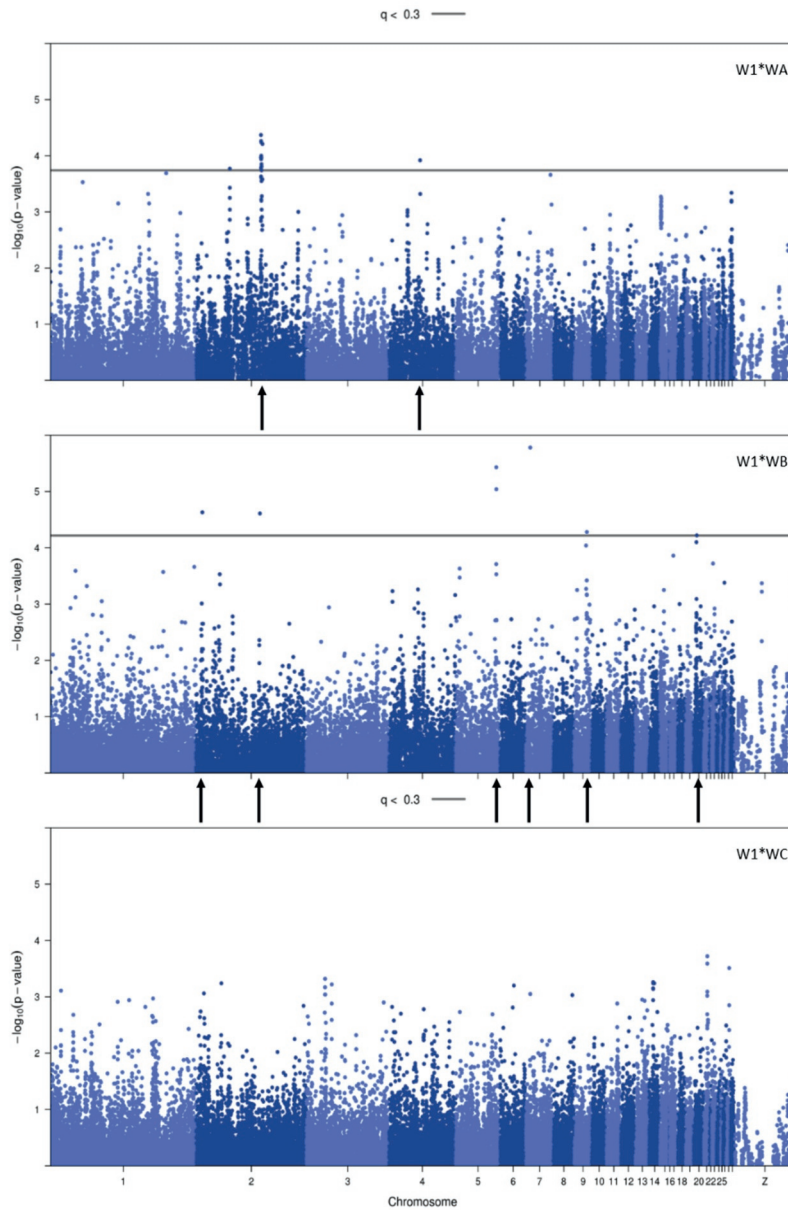


Figure 5.1 Manhattan plots of direct SNP effects for crosses W1*WA, W1*WB, and W1*WC. FDR threshold was 0.30 (solid line). If no SNP reached the FDR-threshold, the threshold could not be estimated (Panel 3). Locations of SNPs with $q < 0.3$ are indicated with an arrow.

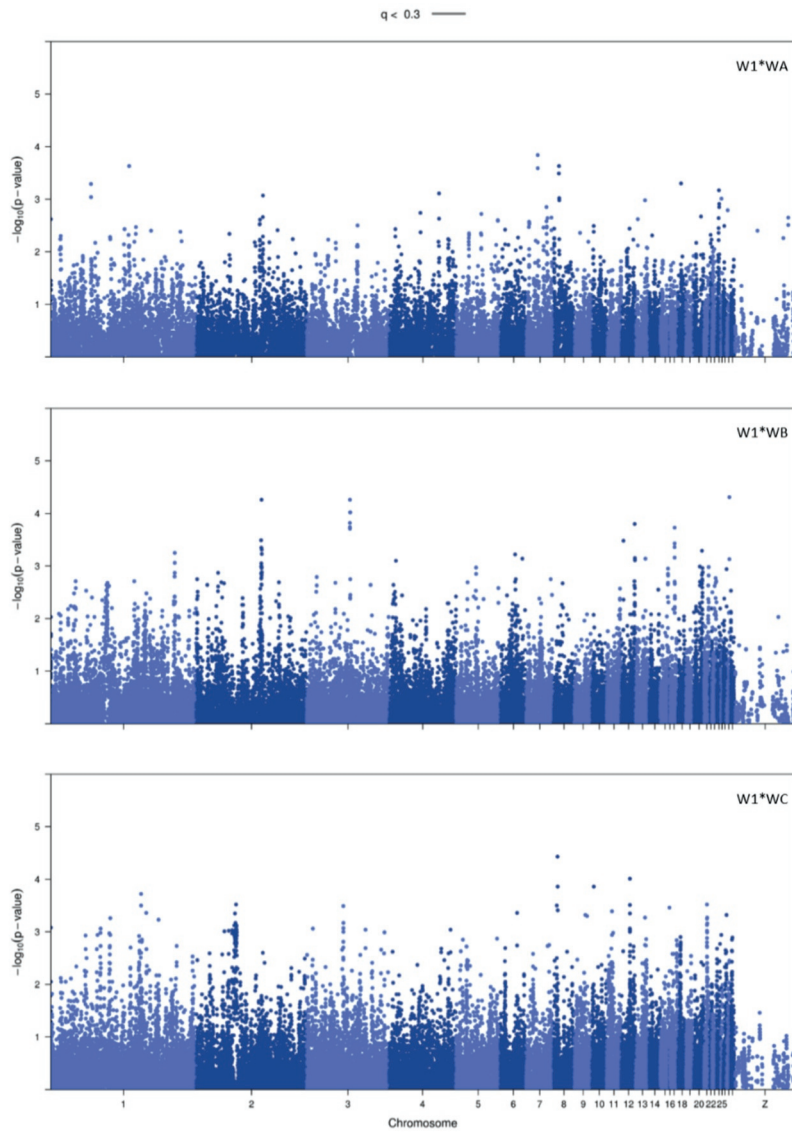


Figure 5.2 Manhattan plots of indirect SNP effects for crosses W1*WA, W1*WB, and W1*WC. FDR threshold was 0.30. If no SNP reached the FDR-threshold, the threshold could not be estimated (Panels 1, 2, and 3).

that was associated with indirect genetic effects. Another study on aggression in chickens revealed a role of the dopaminergic system (Li et al., 2016). Moreover, Bolhuis et al. (2009) investigated the effects of group selection on survival on serotonin levels and suggested that the level of serotonin may be linked to the development of cannibalism. Indeed, the study of Flisikowski et al. (2009) concluded that genomic regions related to the dopaminergic and serotonergic systems were associated with feather pecking behavior in laying hens. Thus, our results support those from two other GWAS that focused on a feather pecking related trait in chickens.

PTPRN2. In cross W1*WB, SNP rs313098101 remained in the model after backwards selection. This SNP is an intron variant on chromosome 2. The associated gene is *PTPRN2* (9 Mb; Fig. 5.1), which encodes a receptor for protein tyrosine phosphatase and is associated with several disease phenotypes (Hale et al., 2017). However, it is likely that the association of SNP rs313098101 with direct effects for survival time is a false positive association, given that no clear peak is visible in this region.

Percentage of genetic variance explained

The contribution of DGE associated SNPs with $q < 0.3$ to the total genetic variance after backwards elimination was large and summer up to 36% for cross W1*WA and to 81% for cross W1*WB (Table 5.5). These genetic variances explained by the SNPs are probably overestimated because of the Beavis effect, i.e., when many effects are tested for significance and only those below the defined significance threshold are considered, SNP estimates tend to be overestimated (Beavis, 1997). This especially occurs when the power of the study is low.

Moreover, the proportion of genetic variance explained by SNPs associated with direct effects was calculated as $\frac{V}{\sigma_A^2}$, where V is the variance explained by the SNPs, and σ_A^2 is an estimate of the total genetic variance. The latter is the sum of variances due to direct and indirect genetic effects, along with a component due to their covariance (Bijma et al., 2007b). Previous research reported negative genetic correlations between direct and indirect effects for survival time in crossbred layers (Peeters et al., 2012). If the correlation between direct and indirect genetic effects is strongly negative, the total σ_A^2 may be smaller than the variance due to the direct effects, which could partly be the reason for the possible over-estimation of the proportion of σ_A^2 explained by direct SNP effects in this study. We are interested in the proportion of direct genetic variance that is explained by DGE associated SNPs.

However, with cages composed of families, an animal model with direct genetic effects only will also pick up many indirect genetic effects (Peeters, 2015). The variance due to direct genetic effects alone was, thus, unknown.

5.3.3 Power

Statistical power of a GWAS depends on several factors such as the number of observations, relatedness among individuals, allele frequency, level of linkage disequilibrium, and the statistical model (Rincent et al., 2014). To get an estimation of the power of our data for GWAS, we calculated power by assuming a number of true direct SNP effects from 1 to 18 days at allele frequencies ranging from 0 to 1, given the population specific parameters of the crosses (N and σ_P^2 ; **Table 5.2**) and significance threshold $q < 0.3$ (~4 standard deviations from the mean; [see **Additional file 5.4**]). We assumed a normal distribution. Theoretical findings as presented in **Additional file 5.4** were supported by the empirical evidence obtained from this study. The results suggest that the statistical power of the GWAS was highest for cross W1*WA and lowest for cross W1*WC. True direct effects, at an allele frequency of 0.5, had to be at least 15 days for W1*WA, 16 days for W1*WB, and 21 days for W1*WC in order to be detected with reasonable probability (power ~ 0.8). True indirect effects, at an allele frequency of 0.5, had to be at least 3.5 days for W1*WA, 4 days for W1*WB, and 5 days for W1*WC in order to be detected with reasonable probability (power ~ 0.8). The number of observations in cross W1*WC was much smaller compared to the other two crosses, which may explain by the power of this cross was lowest.

Theoretical power to detect indirect SNP effects was higher than power to detect direct SNP effects. This is due to multiplication by the number of group mates for indirect SNP effects. It is, therefore, possible that SNP with a lower MAF can be detected at $q < 0.3$ as associated with IGE rather than DGE. In addition, the contribution of IGE to the total heritable variance of survival time in laying hens is often larger than the contribution of DGE (Peeters et al., 2012; Ellen et al., 2008; Brinker et al., 2014). Moreover, in absence of kin selection, IGE are less exposed to natural selection than DGE (Denison et al., 2003; Bijma, 2010a), and some loci may therefore have large effects. Nevertheless, no IGE associated SNPs were found at $q < 0.3$. Perhaps the level of mortality in crosses W1*WA and W1*WB (22 and 25%) was insufficient for detection of IGE associated SNPs, i.e., with lower mortality, fewer individuals will die due to cannibalism and less indirect genetic variance will be available for SNP detection. For cross W1*WC, theoretical power for GWAS was

lower than for W1*WA and W1*WB and more individuals died before blood sampling. These hens were given an average genotype while having an extreme phenotype, which may explain why no direct and indirect SNP associations were found in cross W1*WC.

5.4 Conclusions

This is one of the first large studies that investigates the genetic architecture of a trait by considering both direct and indirect genetic effects. Our results indicate that the same QTL were identified using either a linear mixed model of survival time or models of repeated binomial survival (0/1). Although the power was relatively low, and many SNP associations may have not been detected, our results revealed a link of the GABAergic system with direct and indirect genetic effects for survival time in crossbred layers. The associated gene was *GABBR2*. This supports existing evidence of the involvement of GABA in the development of abnormal behaviours.

5.5 Acknowledgements

We would like to thank the employees of the laying house for taking care of hens and data collection on survival. Furthermore, we would like to thank Jérémie Vandenplas for his help with the quality control of the genomic data. This research was part of a joint project of Hendrix Genetics and Wageningen University on “Genomic solutions for socially-affected traits.” This project was financially supported by the Dutch Technology Foundation TTW, which is part of the Netherlands Organization for Scientific Research (NWO), and is also funded by the Ministry of Economic Affairs.

6

General Discussion

The overarching theme of this thesis is the genetics of socially-affected traits in laying hens, where an individual phenotype depends on the genotype of the individual itself (direct genetic effect; DGE) and on the genotype of its group members (indirect genetic effects; IGE). This thesis focussed on cannibalistic interactions among laying hens which is one of the largest welfare problems in the egg production sector. Previous research clearly showed that genetic solutions need to consider both DGE and IGE in order to reduce mortality due to cannibalism in laying hens (e.g. Ellen et al., 2008; Peeters et al., 2012; Alemu et al., 2016). This thesis is a continuation of research about the relevance of IGE for mortality due to cannibalism, and as such, this thesis contributed to understanding how one of the largest welfare problems in the laying hen sector can be reduced using genetic selection. I believe this is important because I think that we have the responsibility and obligation to have the best practice to our knowledge and to keep on improving animal welfare. Not only for the hen itself (e.g. health, welfare) but also to achieve a high level of food security (e.g. sufficient production) and safety (e.g. reduced antimicrobial use). In this chapter, I will, therefore, not only focus on discussing the results but also reflect on the societal relevance of this thesis.

The outline of this chapter is as follows:

1. Scientific relevance: here I discuss the most important findings of this thesis.
2. Data quality issues: here I discuss the use of phased data in a Genome-Wide-Association Study (GWAS) on crossbred layers.
3. Societal relevance: here I discuss the importance of including DGE and IGE in layer breeding programs to reduce mortality due to cannibalism in laying hens.

6.1 Scientific relevance

6.1.1 Statistical model for survival time

Longevity traits measured in days such as survival time (**this thesis**) are common in any type of animal breeding program. However, such records may still be in progress at the time of genetic evaluation and only a lower bound of the record will be known. This is referred to as censoring. Using models that do not take this into account may reduce accuracies of breeding value estimates.

I contributed a statistical method that improved the accuracy of total breeding values (TBV; the quantity relevant for selection response) for survival time in group-housed laying hens (**Chapter 3**). This model treated survival in consecutive months

as a repeated binomial trait (repeated measures model; RMM) instead of analysing survival time in days. It was shown that the TBV accuracy improved up to 21% when the RMM was used. This is an important result because increased TBV accuracy contributes to higher rates of genetic gain. As mentioned in **Chapter 3**, the benefit of the RMM arises from the fact that this model can better deal with censoring. For survival time, only a lower bound of the true survival time is known for censored records. With the RMM, this problem is circumvented by using a survival indicator, i.e. dead (0) or alive (1) at a given time. **Chapter 3**, therefore, suggests that prediction of TBVs for survival time can be improved using methods that can incorporate censoring.

It is important to realize that censored records were obtained at the same time for all laying hens in this thesis. This is often the case in layer breeding programs, but can be different in other breeding programs. In **Chapter 3**, we investigated the effect on predicted breeding values when 50% of the censored individuals were censored half way during the laying period using a model analysing survival time in days and the RMM. The RMM appeared to be more robust to censoring. In addition, the RMM also produced higher TBV accuracies in genomic predictions (**Chapter 4**). I, therefore, believe that the model presented in **Chapter 3** could also be considered to improve TBV accuracies for censored traits in other species, such as longevity in dairy cattle.

Time-dependent IGE

The composition of the cage changes over time because animals die. The RMM allowed time-dependent IGE in the model. We tested if the prediction of breeding values could be improved by including timing of IGE expression in the model. This was, however, not beneficial, and TBV accuracy reduced substantially compared to any of the other models in **Chapter 3**. The problem probably arises from the fact that modification of the incidence matrices that link observations to IGE directly depends on the observation (Lipschutz-Powell et al. 2012). As a consequence, a considerable amount of the phenotypic variance will be explained by IGE and IGE variance will be overestimated. Recently, Ragab et al. (2018) presented a method where they consider degrees of interaction between each mate pair in the model. Here, the level of social interaction was determined based on feeding behaviour variables. It is, however, not trivial that this will capture all genetic (co-)variance. First, I think that social interactions are not solely based on feeding behaviours. Therefore, using feeding behaviours to shape a social interaction indicator is too narrow and probably misses out on part of the genetic (co-)variances. Second, both absence and presence

may be genetically determined, and it is not clear that all genetic (co)variance will be captured when using a model that focusses on interacting pairs.

The RMM in GWAS

In **Chapter 5**, we investigated whether the RMM would also be beneficial in GWAS analyses. There were, however, no evident differences between GWAS results using the RMM or a model analysing survival time. The same Quantitative Trait Loci (QTL) were appointed with both models. The aims of **Chapter 3** and **Chapter 4** were obviously different: in the GWAS the purpose is to identify QTL, whereas in **Chapter 3** the purpose was to improve TBV accuracy.

6.1.2 Genetic architecture of survival time in laying hens

In **Chapter 5**, we conducted a GWAS to identify SNPs associated with DGE and IGE for survival time. Survival time is a complex trait because it can be representative of a set of other traits. Still, in two of the three crosses, we found SNPs with large DGE (effects ranging from 16 to 35 days). SNPs were detected using a liberal threshold ($q < 0.3$). We expect that some DGE and IGE were left undetected because the power of the study was relatively low.

Our main results include the following:

- GWAS results revealed SNPs with large DGE.
- Within cross, DGE and IGE arose from different SNPs.
- GWAS results revealed a link of DGE (in cross W1*WA) and IGE (in cross W1*WB) between survival time in layers and the GABAergic system.
- DGE associations explained a small proportion of the phenotypic variance (max 6%) and up to 44% of the total genetic variance.
- The power to detect DGE and IGE was relatively low, despite the large number of observations.

6.2 Using phased genotypes in GWAS on crossbreds

In **Chapter 5**, mortality due to feather pecking in laying hens was analysed in a single-SNP GWAS. Survival data of three layer crosses were used of which all originated from the same sire line but different dam lines. Results from this GWAS showed no consensus between the crosses. A possible reason is that SNP effects can differ among crosses due to different levels of linkage disequilibrium (LD) between the SNP and the QTL in the dam lines. Additional evidence that the origin of alleles may be important arises from the fact that detected SNPs (linking to GABBR2) were fixed or nearly fixed for the same allele in the maternal lines (unpublished results). This

suggests that only the contribution of the paternal allele is important. Moreover, if allele-origin matters, the power for detection of SNPs associated with survival time is expected to increase when alleles are mapped specific to their allele origin. Simulation studies indeed show that considering breed-specific effects can be important (Ibáñez-Escriche et al., 2009; Sevillano et al., 2017).

I, therefore, aimed to map DGE and IGE for survival time, while considering the line origin of the alleles. However, accurately phasing crossbred genotypes, i.e. assigning the line origin to alleles, appeared to be challenging. Namely, new Mendel errors were identified after phasing (additional to the identified Mendel errors based on SNP genotypes). In this section, I will first discuss the method used for phasing, followed by a discussion on the background of these errors. The third part of this section presents results of a GWAS on phased data for cross W1*WB, and finally I will discuss the implications of the Mendel errors on the GWAS results.

6.2.1 Data and phasing approach

Genotype data, after quality control from **Chapter 5** were used in the analyses. Quality control encompassed removing SNPs with a missing rate > 0.30 and $MAF < 0.005$. Individuals with a missing rate > 0.10 were also removed. SNPs deviating from Hardy Weinberg Equilibrium ($p < 10^{-5}$) in the parental population were removed across the dataset. Moreover, quality control based on Mendel errors was conducted. SNPs contributing to > 1 Mendel error were removed from the dataset. Finally, individuals contributing to > 5 Mendel errors were removed from the dataset. Remaining genotypes with Mendel errors were set to missing. Ultimately, this means that no detectable Mendel errors were present in the data before phasing.

As a case study, I assigned the line of origin to alleles using data from cross W1*WB. I chose cross W1*WB because the number of initial Mendel errors was limited and, therefore, genotypic data of this cross was considered most reliable. Furthermore, I focussed on chromosome 2 because this chromosome appears to contain a region of interest for mortality due to cannibalism (**Chapter 5**). In total, 1816 genotypes were available for W1*WB crossbreds, with 4717 SNPs (chromosome 2). In total, 1023 ancestral genotypes were available of which 828 of the WB line, and 195 for the W1 line. I assigned the line origin of the alleles using the BOA approach and the corresponding software (Vandenplas et al., 2016). The BOA approach consists of:

- i) phasing genotypes using Alphaphase1.1 (Hickey et al., 2011);
- ii) assigning the line origin to the phased haplotypes

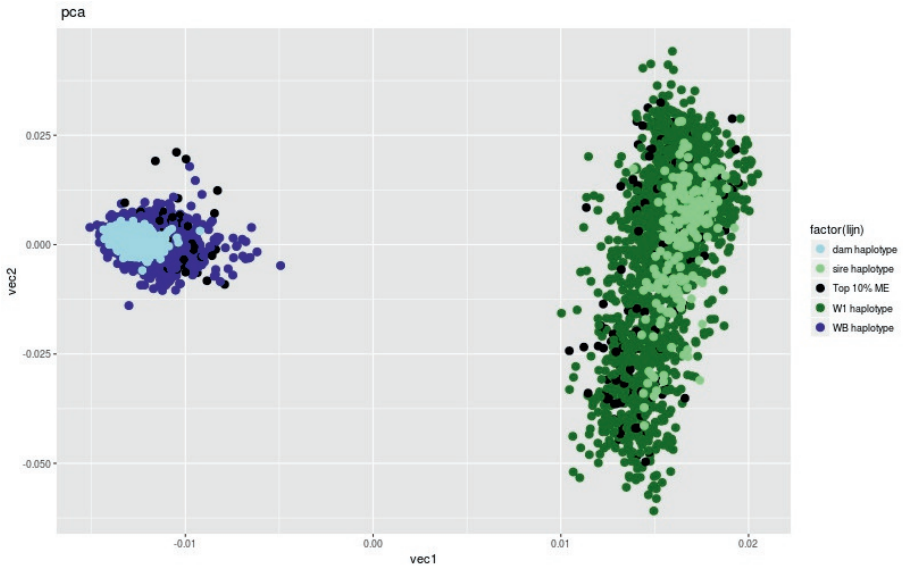


Figure 6.1. Principal component analysis of crossbred haplotypes originating from W1 (W1 haplotype) and from WB (WB haplotype), sire (line W1), and dam (line WB) haplotypes. Black dots indicating the haplotypes part of the top 10% Mendel errors.

- iii)* assigning the line origin to the alleles of the crossbreds based on a library of assigned haplotypes, the line composition of crossbred animals, and their initial SNP genotypes.

In this way, crossbred genotypes (0/1/2) were split into alleles originating from the sire, and alleles originating from the dam (i.e. each individual has two “genotypes” (0/1); referred to as “paternal haplotype” and “maternal haplotype”).

6.2.2 Quality checks of phased data

After assignment, I performed several checks to get an impression of the quality of the phased data. First, I checked whether the input genotype based on count data (0,1,2) equalled the sum of the “paternal haplotype” and “maternal haplotype”. This was true for the full dataset, meaning that the original genotype was maintained throughout all three steps of the BOA approach (see 6.2.1). Second, I conducted a principal component analysis on the haplotypes (PCA; **Figure 6.1**) to investigate if the allele origins were correctly assigned to the specific lines, either W1 or WB. **Figure 6.1** shows that two separate clusters can be identified; one indicating alleles belonging to line W1 and the other indicating alleles belonging to line WB. From this

figure, I conclude that the line origin that was assigned to the haplotypes of the crossbreds matches the parental lines.

Third, I checked for Mendel errors based on the phased data, i.e. I checked whether the “maternal haplotype” and the “paternal haplotype” at a specific locus was possible given the parental genotypes. While no Mendel errors were present based on the SNP count data before phasing, 191 059 Mendel errors (~2%) were found after phasing. Mendel errors are not visible in the data when both parents have heterozygous genotypes, and, therefore, more Mendel errors have been left undetected. Reconstruction of haplotypes, and thus considering multiple markers jointly, may reveal additional Mendel errors.

When using SNP count data to check for Mendel errors, one cannot find all Mendel errors. With (correctly) phased data, it is known which allele comes from the sire and which from the dam, and it is thus possible to reveal additional Mendel errors compared to using count data. **Table 6.1** lists 5 scenarios, each describing a combination of sire and dam genotype and the occurrence (%) of this scenario in the current data. It is also described in which scenarios phased data provides additional information to detect Mendel errors and the % detected in the data for a given scenario. Phased data will not provide any additional information compared to count data for scenario 1 (both parents are homozygous for the same allele) and scenario 3 (both parents are heterozygous). Phased data provides additional information for scenario 2, 4, and 5 for the situation where offspring have heterozygous genotypes. For example, when parents have opposing homozygote genotypes (AA and BB), offspring with heterozygous genotypes (AB) would always pass the parent-offspring trio check based on count data. This is not the case with phased data, where two haplotypes will be identified, the “paternal haplotype” and the “maternal haplotype”. For example, if the sire genotype would be AA, the dam genotype BB, the “paternal haplotype” B, and the “maternal haplotype” A, this would be an error which was undetected when using the count data approach.

I listed the number of errors in **Table 6.1**. No additional errors were revealed based on phased data for situation 1 and 3 and a number of additional errors were found based on phased data that had not been found using count data for situation 2, 4, and 5. Phased data revealed an additional 191 059 of Mendel errors in cross W1*WB on chromosome 2. This comprises ~2 % of the complete dataset. This shows that the number of actual Mendel errors is higher than detected by comparing offspring and

parent genotypes based on count data (0/1/2) and/or that the phasing procedure introduces mistakes.

Mendel errors can occur because of genotyping, pedigree mistakes, or may be introduced by the phasing program that is used. Such errors may have an impact on the power to detect SNP associations. The following was observed:

- In total, 0.2% Mendel Errors were identified where parents had opposing homozygous genotypes (**Table 6.1**)
- The maximum % of Mendel errors detected for an individual was 6%.
- The maximum % of Mendel errors detected for a SNP was 12%

I would argue that the possibility of mistakes in the pedigree is limited because the pedigree was reconstructed by Hendrix Genetics based on genotype data. Parents were set missing if the parent-offspring match was unconvincing. I, therefore, believe that the errors detected in this research are mostly due to genotyping mistakes or phasing mistakes. For the remainder of this section, I will therefore only consider genotyping errors and phasing errors.

To get a feeling about the source of the Mendel errors (crossbred genotype, sire genotype or dam genotype), I looked at the scenario where parent-duos were homozygous for the opposite allele (scenario 4 in **Table 1**). Thus, for each detected Mendel error, I identified whether parent-duos were homozygous for the opposite allele. To make an inference on the source of the error I looked at the genotype of the siblings of the crossbred. First, I identified half-sibs of the crossbred offspring.

Table 6.1. Five situations indicating whether phased data provides additional information compared to count data about Mendel errors.

Scenario	Sire and Dam genotypes	Phased Data	Occurrence (%)	Mendel errors (%)
1	Same homozygote	No additional information	27	0
2	Homozygote and heterozygote	Additional information for heterozygous offspring	30	1.7
3	Both heterozygote	No additional information	4	0
4	Opposing homozygote	Additional information for heterozygous offspring	28	0.2
5	One parent missing	Additional information for heterozygous offspring	10	0.1
	Total		100	2

6. General Discussion

Second, I identified whether parent-duos of the half-sibs also had opposing homozygous genotypes, and finally whether this particular trio was identified as a mismatch. For 4% of the identified Mendel errors in scenario 4, all half-sibs had opposing homozygous parents.

Figure 2 is a frequency plot of the percentage of half-sibs (excluding self) that have the same identified Mendel errors. My assumptions were the following:

- If none of the individual's half-sibs are implicated in the same Mendel error, i.e. the detected Mendel error for a specific SNP was unique to the offspring-parent trio, the mistake is due to the offspring haplotype. Since we look at heterozygous offspring and opposing homozygote parents, we may conclude the paternal and maternal genotype have been swapped during phasing analysis.
- With an increasing number of half-sib haplotypes implicated in the same Mendel error, chances increase that the mistake is due to errors in the parental genotype. This would be an indication of a genotyping mistake.

The distribution in **Figure 6.2** is skewed to the right, and I, therefore, suspect that most Mendel errors were introduced by phasing.

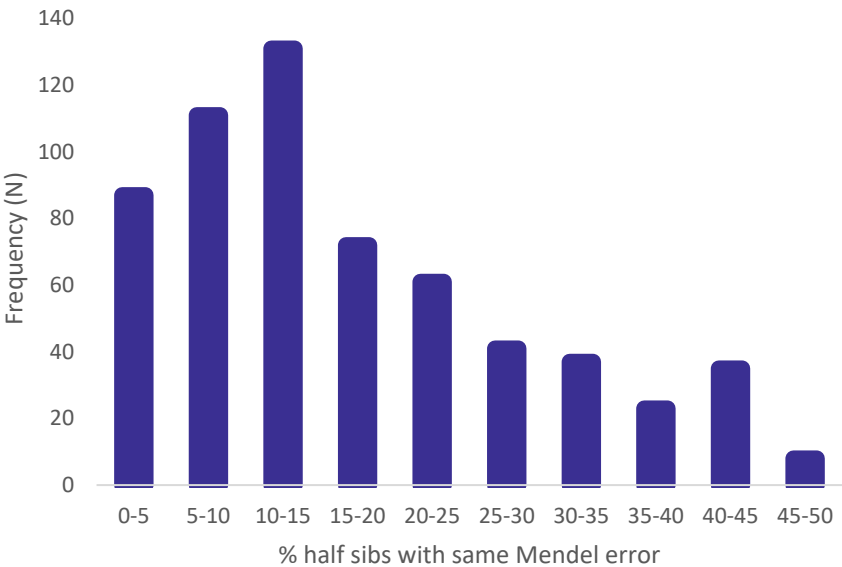


Figure 6.2. Frequency plot of the percentage of half-sibs (excluding self) that have the same identified Mendel errors

Furthermore, I found the following:

- It was expected that genotyping mistakes may have occurred primarily for SNP with low MAF, but this was not the case. There is no correlation between MAF and the number of Mendel errors per SNP.
- There is no correlation between the number of total genotyped relatives and the number of mistakes. It was expected that with a higher number of genotyped relatives, a better phasing quality would be observed, and less Mendel errors.

To summarize, based on the frequency plot of the percentage of half-sibs that have the same identified Mendel errors I conclude that most of the Mendel errors have been introduced by phasing. However, it must be noted that this was real data and I could, therefore, only investigate it for parent-offspring duos with heterozygous offspring and opposing homozygous parents. I would, therefore, recommend a simulation study to investigate the introduction of Mendel errors due to phasing analyses.

6.2.3 GWAS using phased genotypes in W1*WB - Chromosome 2

To understand the implications of the Mendel errors in GWAS analyses, I compare two analyses: GWAS using data including the Mendel errors, and GWAS using data where the known Mendel errors were set missing. Finally, I discuss the results from the GWAS analyses using phased data and compare it to the results in **Chapter 5**.

I used a linear mixed model for survival time for the GWAS analysis using phased genotypes for cross W1*WB on chromosome 2. First, genetic parameters were estimated without SNP effects in the model. Five generations of pedigree information were included in the genetic analysis. Second, SNP effects were estimated one by one. The origin of the alleles were considered in the model. This resulted in four estimated SNP-effects: the paternal and maternal direct allele-effects, and the paternal and maternal indirect allele-effects, which were both modelled as the summed indirect effects of the cage mates. Variance components were fixed to the estimated values from the model without the SNP effect. The following model was used:

$$y_{ik} = fixed + \beta_D^{sire} \cdot SNP_i^{sire} + \beta_D^{dam} \cdot SNP_i^{dam} + \beta_I^{sire} \cdot \sum_{j \neq i}^{n-1} SNP_{j(k)}^{sire} + \beta_I^{dam} \cdot \sum_{j \neq i}^{n-1} SNP_{j(k)}^{dam} + a_i + cage_k + e_{ijk},$$

6. General Discussion

where y_{ik} is the observed survival time (days) for individual i in cage k , with cage mates j , n is the number of cage members ($n = 5$), and **fixed** represents the fixed effect of the combination of wing-row-level; β_D^{sire} , β_D^{dam} , β_I^{sire} , and β_I^{dam} are regression coefficients of y_{ik} on haplotypes for the sire (β_D^{sire} , β_I^{sire}) and dam population (β_D^{dam} , β_I^{dam}), fitted as fixed effects, where β_D represents the fixed direct effect of the SNP of individual i , and β_I is the fixed indirect effect of the same SNP in cage mates j ; SNP_i is the allele count (0, 1) for the individual, $\sum_{j \neq i}^{n-1} SNP_{j(k)}$ is the summed SNP allele counts (0 - 4) of cage mates j , a_i is the random polygenic effect of individual i , $cage_k$ is the random cage effect, and e_{ijk} is the residual. All random effects were assumed to be normally distributed. The covariance structures for the model terms were: $var(a) = A\sigma_a^2$, $var(cage) = I\sigma_{cage}^2$, and $var(e) = I\sigma_e^2$, where A is a pedigree relationship matrix, σ_a^2 is the additive genetic variance, I is an identity matrix, σ_{cage}^2 is the cage variance, and σ_e^2 is the residual variance.

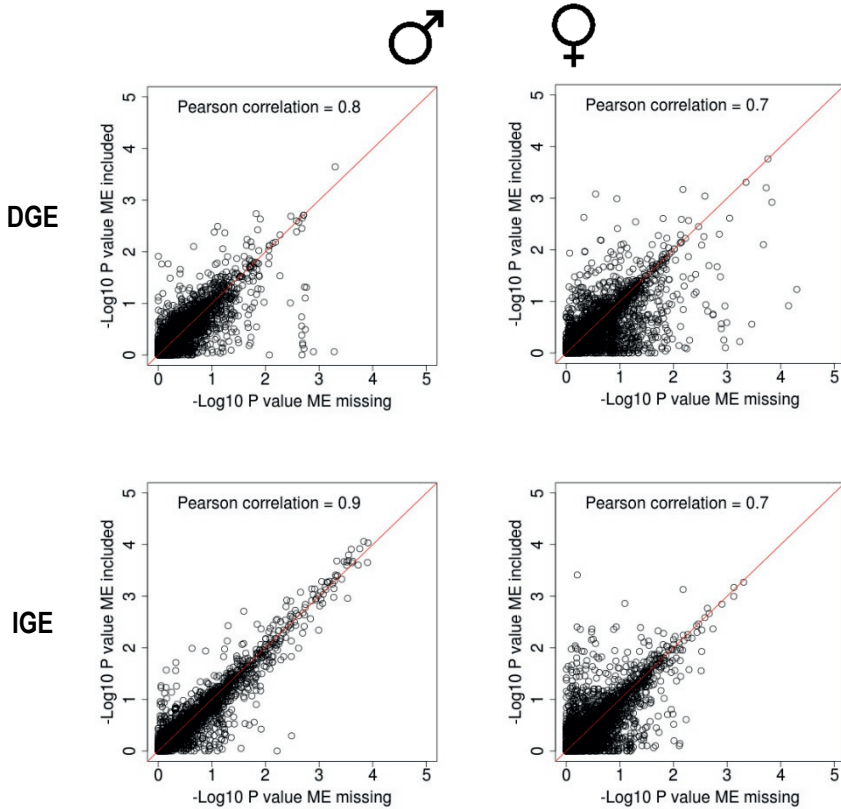


Figure 6.3. Correlation plots of \log_{10} - p values from the analyses including Mendel errors and analyses with Mendel errors set to missing. Correlation plots are for $-\log_{10}$ - p values of the paternal (♂; left) and maternal (♀; right) direct (DGE; top) allele-effects, and the paternal and maternal indirect (IGE; bottom) allele-effects.

Implication of Mendel errors in GWAS.

Figure 6.3 contains four correlation plots of $-\log_{10}p$ values from the analyses including Mendel errors and analyses with Mendel errors set to missing. Correlation plots are for $-\log_{10}p$ values of the paternal and maternal direct allele-effects, and the paternal and maternal indirect allele-effects. The following can be observed:

- Correlations of $-\log_{10}p$ values are rather low (<0.8), except for the correlation between paternal indirect allele-effects (0.9).
- Correlations of $-\log_{10}p$ values for the paternal allele-effects are higher compared to correlations of $-\log_{10}p$ values for the maternal allele-effects.
- Correlations of $-\log_{10}p$ values for the indirect allele-effects are higher compared to correlations of $-\log_{10}p$ values for the direct allele-effects.
- The ranges of the $-\log_{10}p$ values are similar for the paternal allele-effects.
- The ranges of the $-\log_{10}p$ values are similar for the maternal allele-effects.

In conclusion, Mendel errors can have a large impact on the results of the analyses, even when the fraction of identified Mendel errors is small (2% in this dataset). This is supported by findings in other studies, where it was shown that small SNP error rates can have large implications in linkage or association studies (*e.g.* Douglas et al, 2000; Abecasis et al, 2001). Repeating the quality control based on Mendel errors after phasing is, therefore, an important step when conducting GWAS analyses with phased data, i.e. two Mendel error checks are required, one based on count data and one using phased data.

Results of GWAS analyses based on phased data

Figure 6.4 shows the Manhattan plots for paternal and maternal direct allele-effects, and the paternal and maternal indirect allele-effects. In the plots, I added FDR-thresholds for $q<0.1$, $q<0.2$, and $q<0.3$. In the discussion, I will only focus on results based on $q<0.1$. If no SNP reached the FDR-threshold, the threshold could not be estimated. There were no SNPs associated with direct effects for survival time at $q<0.1$. For paternal indirect allele-effects, 34 SNPs were associated with survival time at $q<0.1$, of which 3 SNPs in the region between 27.8 and 28.4 Mb (region 1), 10 SNPs in the region between 62 and 64 Mb (region 2), and the remaining 20 SNPs in the region between 87 and 89 Mb (region 3). In **Chapter 5**, a clear peak for indirect effects was also observed in region 3. However, none of the SNP effects in that peak were significant at $q<0.3$ (and, therefore, also none at $q<0.1$). This peak was associated with a gene, GABBR2 (88.00-88.4 Mb).

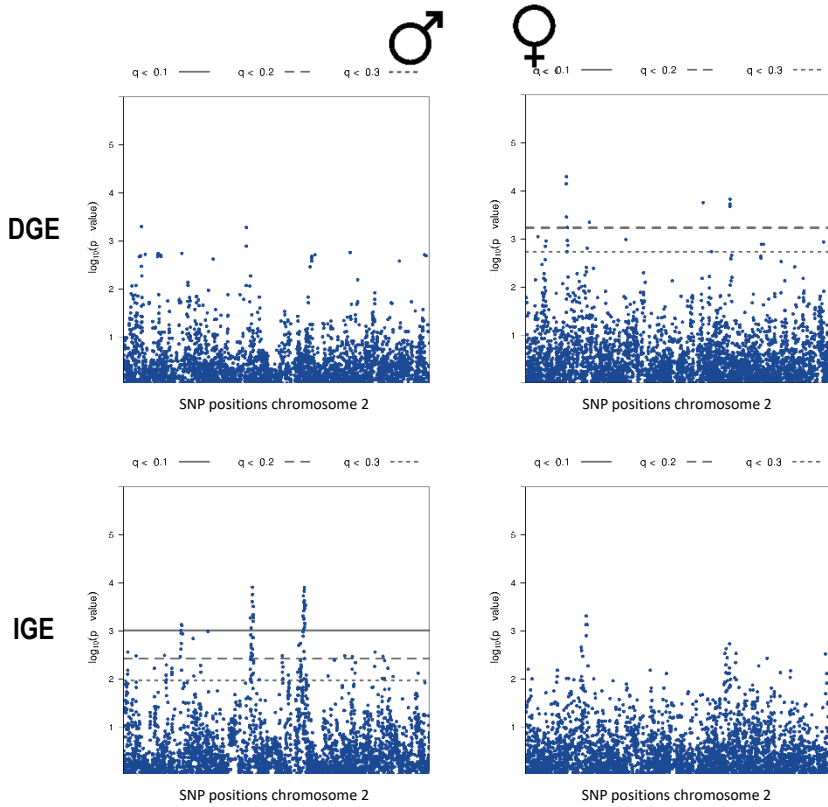


Figure 6.4. Manhattan plots of analyses using phased data where Mendel errors were set to missing. These are Manhattan plots of the paternal (♂; left) and maternal (♀; right) direct (DGE; top) allele-effects, and the paternal and maternal indirect (IGE; bottom) allele-effects. In the plots, FDR-thresholds for $q < 0.1$, $q < 0.2$, and $q < 0.3$ are presented. If no SNP reached the FDR-threshold, the threshold could not be estimated.

Before the analyses based on phased data, I had two hypotheses: *i)* I expected that the contribution of the paternal allele matters because the identified SNP in the GABBR2 region (**Chapter 5**) was fixed in the maternal line (WB), and *ii)* If allele-origin matters, the power for detection of SNPs associated with survival time is expected to increase when alleles are mapped specific to their allele origin. The first hypothesis, that the contribution of the sire allele matters, holds. Namely, two clear peaks were observed for indirect effects based on the “paternal haplotype”, of which one of them was in the GABBR2 region. No associations for indirect effects were found based on the “maternal haplotype”. The second assumption holds as well: the indirect effects based on the “paternal haplotype” were found significant at $q < 0.1$, while based on count data these were not significant at $q < 0.3$.

In conclusion,

- The power of genome-wide association study may increase when alleles are mapped specific to their allele origin.
- In cross W1*WB, the contribution of the paternal alleles are important for indirect effects.

6.3 Animal welfare and sustainable egg production

Sustainable egg production goes beyond the ability to continue to produce enough eggs, but comprises all aspects of egg production, including farm economics, animal welfare, and acceptance of its *modus operandi* by society. A growing world population and increased prosperity in third world countries contribute to an increased demand for eggs. Eggs are very nutritious and are an important ingredient of diets worldwide. Eggs represent ~ 6% of the total animal protein consumption with approximately 9 kg eggs/capita/year consumed globally (Chambers et al., 2017). To meet the demand for eggs, there has been a steady increase in egg production the past decades, with approximately 1400 billion eggs produced in 2016 compared to approximately 1000 billion eggs in 2000 (FAOstat, 2016). Laying hen breeding has contributed to this increase in egg production, and is as such concerned with all aspects of sustainability.

In my opinion, the topic of this thesis is very relevant in the light of sustainable egg production, because *i)* it addresses important societal concerns, and *ii)* it contributes to sustainable development goal “End hunger” (SDG2) of the Food and Agricultural Organization of the United Nations (FAO 2015).

6.1.1 Societal concerns

Animal breeding cannot be sustainable if the *modus operandi* is not accepted by society. Public opinions and attitudes (*e.g.* consumer behaviour) may drive legislation or markets and, as such, affect the way breeding companies operate. The society is increasingly interested in animal production of which animal welfare is one of the larger concerns. Statements about the animal welfare situations in livestock production systems are spreading globally and – perhaps as a result - consumption patterns in many (European) countries have changed, adopting diets containing less animal products or “animal welfare friendly” products (Amos and Sullivan, 2018). Integrating animal welfare in all chains of animal production is, therefore, a growing demand of society. Continued efforts to improve animal welfare using genetic selection is, therefore, crucial.

One practice in the laying hen industry that receives much critical attention is beak trimming. Beak trimming is a common practice to prevent cannibalism. It has received much critical attention because it compromises the welfare of the laying hens (Riber and Hinrichsen, 2017). For this reason, various European countries have banned the beak trimming practice or plan to ban it in the near future. In **Chapter 4**, we reported mortality to be between 22% and 37% in crossbred hens with intact beaks, while beak-trimmed hens of the same crosses showed a mortality of 2% to 3% at the end of the laying period (personal communication J. Visscher). Thus, cannibalism will increase when hens have intact beaks.

Cannibalism is a multi-factorial problem caused by both environmental and animal related factors. Management solutions could be related to feed (Van Krimpen 2005), light intensity (Kjaer and Vestergaard 2009), provision of floor litter (Blokhus, 1986), and group size (Bilcik and Keeling 2000). However, there are no management solutions that can completely prevent cannibalism. Genetic selection, could contribute to a more permanent solution, that will allow to keep hens with intact beaks. This thesis contributes to understanding how to improve survival time in laying hens with intact beaks. As such, it addresses two, although tightly linked, societal issues: the beak trimming problem and the cannibalism problem.

I believe that conducting research and publishing forthcoming results in scientific journals alone is not sufficient. For ones work to be societal relevant, it is important to communicate results to the right set of stakeholders with the aim to translate the research into practice. This work was established through direct involvement of Hendrix Genetics B.V, a world leader in laying hen breeding. This direct involvement, and the use of their recurrent test data ensured added value of the outcomes of this thesis, especially since directions of this work were fit to the questions from industry, and results were directly shared with the relevant people to be applied in practice.

In my opinion, improving animal welfare using genetic selection is an acceptable approach. However, part of the society will criticize this. Society could question whether it is acceptable to use genetic selection to improve animal welfare given that these animals are bred to be kept in a production environment. They pose the question: should we adapt the animal or change the production environment? Worldwide, the demand for eggs and consequently the production of eggs has increased (FAOstat, 2016). I believe that we have the obligation to produce these eggs in the best way possible, and this includes a high level of welfare. The welfare problem discussed in this thesis, cannibalism among laying hens, is a multifactorial

problem. Here, both environmental and animal related factors play a role. Thus, changing the production environment is not sufficient. In my opinion it is acceptable to use genetic selection to ensure the positive interactions among laying hens because animal breeders make use of the natural variation that occurs in a population. It involves making decisions about the parents for the next generation.

6.1.2 Sustainable development goal: “End hunger”

The Food and Agricultural Organization of the United Nations (FAO) has defined 17 global sustainable development goals (SDGs) in 2015 with the aim to achieve them within 15 years (in 2030). The SDGs were defined within areas of critical importance for both humanity and the planet as a whole. Improved animal welfare contributes to SDG2, “End hunger, achieve food security and improved nutrition and promote sustainable agriculture” (FAO, 2015). As mentioned earlier, animal proteins comprise of an important part of the human diet, and are, therefore, of utmost importance to achieve food security and to improve nutrition. Increased animal welfare by reducing competitive interactions in laying hens may contribute to increased productivity and increased animal health, and consequently increase the availability of animal proteins.

Selection for reduced competitive interactions may, however, be at cost of other traits in the breeding goal, *e.g.* productivity. Ellen and Bijma (2019) aimed to quantify the prospects for genetic improvement of a socially-affected trait in laying hens in a realistic breeding scheme. To this end, authors investigated the response to selection of survival time and “other traits” in the breeding goal for two selection schemes: a multi-trait recurrent testing scheme and a multi-trait genomic selection scheme. **Figure 6.5** displays the gain in survival time (y-axis) relative to the reduction in response to other traits (x-axis). Responses are shown for breeding goals ranging from selection for survival time only to selection for “other traits” only. The following can be observed from **Figure 6.5**:

- The maximum predicted response to selection for survival time was substantially larger in the genomic selection scheme (48 days/year) than for the recurrent testing scheme (25 days/year).
- It is feasible to select for improved survival time with only a small reduction in response to selection in other traits when the genetic correlation between survival and “other traits” is zero. For example, half of the maximum response in survival time can be obtained with only ~13% reduction in response in “other traits” in the recurrent testing scheme, and ~15% in the genomic selection scheme.

6. General Discussion

- With a genetic correlation of -0.3 between survival time and “other traits”, costs in “other traits” were substantially larger. For example, half of the maximum response in survival time can be obtained with ~42% reduction in the recurrent testing scheme and ~52% in the genomic selection scheme.

In conclusion, the weight given to survival time relative to “other traits” will determine the achieved genetic gain in survival time. It is, however, clear that it is feasible to reduce competitive interactions in laying hens and at the same time realize genetic gain in “other traits”. In addition, when more laying hens survive until the end of the laying period, more eggs will be produced. Genetic selection for reduced competitive interactions will, thus, contribute to increased animal welfare,

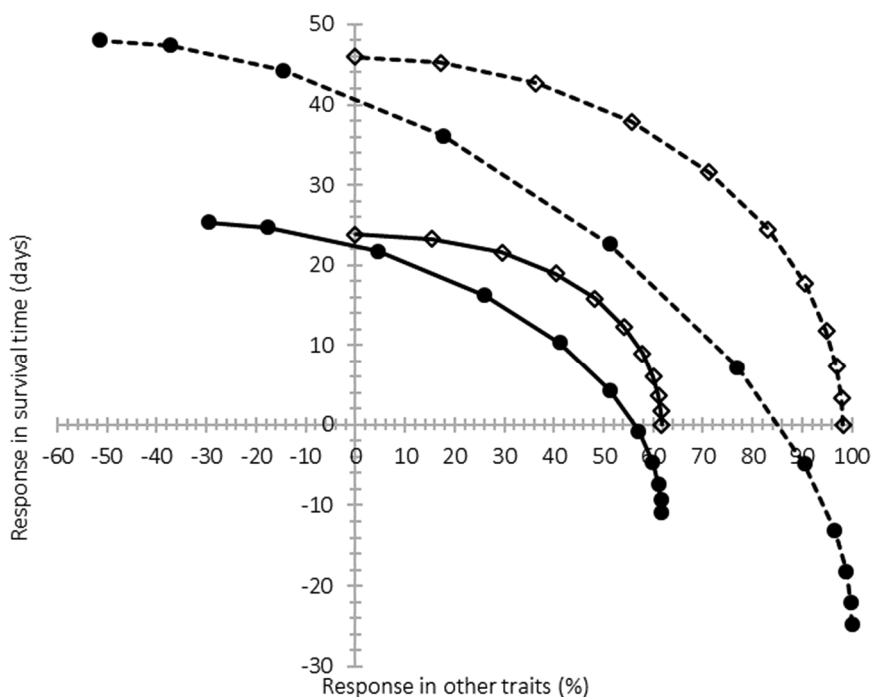


Figure 6.5. Relation between response to selection for survival time and other traits, for a population with 35% mortality. Responses are shown for breeding goals ranging from selection for survival time only to selection for other traits only (when going from left to right in the figure). The genetic correlation between survival time and other traits was either 0 (◊) or -0.3 (●). Solid lines represent the recurrent testing-scheme, while dashed lines represent the genomic selection-scheme. (Figure from E. D. Ellen and P. Bijma, 2019)

and other traits, *e.g.* productivity and health, and consequently increase the availability of animal proteins. As such, this work contributes to SDG2, “End hunger, achieve food security and improved nutrition and promote sustainable agriculture” (FAO, 2015).

Summary

Social interactions are an integral component of life. They occur in both domestic and natural populations. Interacting individuals may affect each other's characteristics. Examples are aggression in mink, cooperation in ants, and spouse effects in human. With social interaction traits, there are two distinct effects: the effect of an individual on its own phenotype (direct effect) and the effects of group mates on an individual's phenotype (indirect effect). Direct and indirect effects can partly be genetically determined, where the phenotype of an individual depends on both the individual's own genes (direct genetic effects; DGE) and the genes from its group mates (indirect genetic effects; IGE). This thesis focussed on cannibalistic interactions among laying hens which is one of the largest welfare problems in the egg production sector. Previous research showed that in order to reduce mortality due to cannibalism in laying hens, genetic solutions clearly need to consider both DGE and IGE.

In this thesis, two traits that show the consequences of cannibalistic interactions are analysed using methods that incorporate both DGE and IGE: plumage condition (**Chapter 2**) and survival time (**Chapter 3-5**). Knowledge of the genetic parameters underlying a trait is required in order to estimate breeding values. The availability of these parameters will, moreover, provide more insight in the potential contribution of IGE to selection response. This thesis shows that up to 94% of the heritable variation in plumage condition (**Chapter 2**) and up to 61% of the heritable variation in survival time relates to IGE (**Chapter 3**). This indicates that methods of genetic selection that include IGE offer perspectives to improve plumage condition and survival time in laying hens.

It is important that breeding values are estimated accurately because this will contribute to higher rates of genetic gain. Studies using DGE-IGE models have mostly focused on the trait survival time to reduce mortality due to cannibalism in laying hens. However, in these models, censored records were considered as exact lengths of life and models assumed that IGE were continuously expressed by all cage members, irrespective of whether they were alive or dead. Neglecting censoring and timing of IGE expression may reduce the accuracy of estimated breeding values. In **Chapter 3**, four models were considered to predict survival time in laying hens. One model was an analysis of survival time and the three other models treated survival in consecutive months as a repeated binomial trait. The latter three models can incorporate censoring and timing of IGE expression. The aim of this chapter was to investigate whether the accuracy of estimated breeding values was improved using the repeated measures models instead of using a model analysing survival time.

Including the timing of IGE expression in the DGE–IGE model reduced accuracy of estimated breeding values compared to analysing survival time. Using repeated measures instead of analysing survival time increased accuracy of estimated breeding values up to 21%. Results in this thesis, therefore, suggest that prediction of breeding values for survival time in laying hens considering methods that can deal with censoring. This is an important result since more accurate estimates of breeding values contribute to higher rates of genetic gain.

Separate estimation of DGE and IGE is not always possible. DGE and IGE can only be estimated based on pedigree information when hens are housed in groups composed of multiple families. Layer breeding organizations, however, often use recurrent testing, where hens are housed in sire-family groups and dam pedigree is unknown. Here, DGE and IGE are fully confounded. A sire model can be used to estimate the total breeding value (TBV) of the sire, which is the linear combination of the sire DGE and the sire IGE. There are, thus, two issues; the dam pedigree is unknown and separate estimation of DGE and IGE is not possible using pedigree information. Genomic information may solve both issues. First, genomic information provides the opportunity to reconstruct the dam pedigree which makes it possible to distinguish full-sibs from half-sibs. In turn, this may allow separate estimation of DGE and IGE, even when a group has the same sire (but not the same dam). Second, genomic information provides information on the actual genetic relationship between individuals. Actual genetic relationships between individuals vary around their expected value based on pedigree information because of linkage and Mendelian sampling. Thus, genomic information may provide new opportunities for estimation of genetic parameters when traits are affected by social interactions. In **Chapter 4** it was, therefore, investigated whether DGE and IGE of survival time can be estimated separately for crossbred laying hens housed in sire-family groups using genomic information. However, the majority of these analyses did not converge and therefore a sire model which directly estimated the TBV was used. Estimates from a sire model using genomic information were compared to those from a sire model using pedigree information. In **Chapter 4**, there was no improvement in the accuracy of breeding value predictions using genomic information rather than using pedigree information. This was probably because of the limited number of sires and the use of a sire model. Moreover, **Chapter 4** showed that, the total estimated heritable variance expressed as a proportion of phenotypic variance ranged from 3-25%. These results suggest that mortality due to cannibalism can be reduced by selection using a sire model, even though underlying DGE and IGE are unknown, because sire models capture the TBV.

Genomic information is also used in Genome-Wide Association Studies (GWAS). GWAS are widely applied to a variety of traits and populations, but have been mainly focussed on DGE. Results from GWAS on DGE show that most quantitative traits in livestock are highly polygenic, and that variants tend to be associated with more than one trait (pleiotropy). The genetic architecture of IGE may, however, differ from the genetic architecture of DGE. With IGE, the phenotype of other individuals are affected rather than the phenotype of the individual itself (*i.e.* DGE). Selection targets such IGE only in the presence of feed-back mechanisms, using *e.g.* group selection or kin selection. IGE may, therefore, be less exposed to natural selection or artificial selection when not accounting for social interactions. In other words, selection has not exhausted variation due to IGE. It is, therefore, expected that some loci may have large IGE. However, only a handful of studies extended GWAS to include IGE. In **Chapter 5**, the aim was, therefore, to identify SNPs associated with direct and indirect effects for survival time in laying hens that show cannibalism. Moreover, GWAS results from analysis of survival time versus using repeated measures models were compared. **Chapter 5** showed that the same quantitative trait loci were identified with all models. Moreover, GWAS results revealed SNPs with large DGE and a link of DGE and IGE for survival time in layers with the GABAergic system (GABBR2 gene), which supports existing evidence for the involvement of GABA in the development of abnormal behaviours.

In **Chapter 5**, mortality due to feather pecking in laying hens was analysed in a single-SNP GWAS. Survival data of three layer crosses were used of which all originated from the same sire line but different dam lines. Results from this GWAS showed no consensus between the crosses. A possible reason is that SNP effects can differ among crosses due to different levels of linkage disequilibrium between the SNP and the QTL in the dam lines. Additional evidence that the origin of alleles may be important arises from the fact that detected SNPs (linking to GABBR2) were fixed or nearly fixed for the same allele in the maternal lines. This suggests that only the contribution of the paternal allele is important. Moreover, if allele-origin matters, the power for detection of SNPs associated with survival time is expected to increase when alleles are mapped specific to their allele origin. **Chapter 6**, therefore, aimed to map DGE and IGE for survival time, while considering the line origin of the alleles. However, accurately assigning the line origin to alleles (phasing), appeared to be challenging. Namely, new Mendel errors were identified after phasing (additional to the identified Mendel errors based on SNP genotypes). In **Chapter 6** it is concluded that Mendel errors can have a large impact on the results of the analyses, even when the fraction of identified Mendel errors is small. Moreover, this thesis shows that

repeating the quality control based on Mendel errors after phasing is an important step when conducting GWAS analyses with phased data, i.e. two Mendel error checks are required, one based on count data and one using phased data.

Finally, **Chapter 6**, shows that the topic of this thesis is very relevant in the light of sustainable egg production, because *i)* it addresses important societal concerns, and *ii)* it contributes to sustainable development goal “End hunger” (SDG2) of the Food and Agricultural Organization of the United Nations.

References

A

- Abecasis, G.R., S.S. Cherny, and L.R. Cardon. 2001. The impact of genotyping error on family-based analysis of quantitative traits. *European Journal of Human Genetics*. 9:130-134.
- Adams, A.W., J.V. Craig, and A.L. Bhagwat. 1978. Effects of flock size, age at housing, and mating experience on two strains of egg-type chickens in colony cages. *Poultry Science*. 57:48-53.
- Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*. 93:743-752.
- Alemu, S.W., M.P.L. Calus, W.M. Muir, K. Peeters, A. Vereijken, and P. Bijma. 2016. Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior. *Genetics Selection Evolution*. 45:68.
- Alemu, S.W., P. Bijma, S.H. Møller, L. Janss, and P. Berg. 2014. Indirect genetic effects contribute substantially to heritable variation in aggression-related traits in group-housed mink. *Genetics Selection Evolution*. 46:30.
- Allen, J. and G.C. Perry. 1975. Feather pecking and cannibalism in a caged layer flock. *British Poultry Science*. 16:441-451.
- Amos, N. and R. Sullivan. 2018. The business on farm animal welfare. *Routledge in association with GSE Research*. 14:159-172.
- Ashbrook, D.G., B. Gini, and R. Hager. 2015. Genetic variation in offspring indirectly influences the quality of maternal behavior in mice. *eLife*. 4:e11814.

B

- Baud, A., F.P. Casale, J. Nicod and O. Stegle. 2018. Genome-wide association study of social genetic effects on 170 phenotypes in laboratory mice. *bioRxiv*. 302349.
- Beavis, W.D. 1997. QTL analyses: power, precision, and accuracy. In: Paterson AH, editor. *Molecular dissection of complex traits*. Boca Raton: CRC Press. 145–62.
- Bergsma, R., E. Kanis, E.F. Knol, and P. Bijma. 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (*Sus scrofa*). *Genetics*. 178:1559-1570.
- Bijma, P. and J.W.M. Bastiaansen. 2014. Standard error of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation?. *Genetics Selection Evolution*. 46:1.
- Bijma, P. 2011. A general definition of the heritable variation that determines the potential of a population to respond to selection. *Genetics*. 189:1347-1359.

References

- Bijma, P. 2010a. Estimating indirect genetic effects: Precision of estimates and optimum designs. *Genetics*. 186:1013-1028.
- Bijma, P. 2010b. Multilevel selection 4: modelling the relationship of indirect genetic effects and group size. *Genetics*. 186:1029-1031.
- Bijma, P. and M.J. Wade. 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *Journal of Evolutionary Biology*. 21:1175-1188.
- 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.
- Bilčík, B. and L.J. Keeling. 2000. Relationship between feather pecking and ground pecking in laying hens and the effect of group size. *Applied Animal Behaviour Science*. 68:55-66.
- Bilčík, B. and L.J. Keeling. 1999. Changes in feather condition in relation to feather pecking and aggressive behaviour in laying hens. *British Poultry Science*. 40:444-451.
- Biscarini, F., H. Bovenhuis, J.J. van der Poel, T.B. Rodenburg, A.P. Jungerius, and J.A.M. van Arendonk. 2010. Across-line SNP association study for direct and associative effects on feather damage in laying hens. *Behavior Genetics*. 40:715-727.
- Blokhuys, H.J. 1986. Feather-pecking in poultry: its relation with groundpecking. *Applied Animal Behaviour Science*. 16:63-67.
- Blokhuys, H.J. and J.G. Arkes. 1984. Some observations on the development of feather-pecking in poultry. *Applied Animal Behaviour Science*. 12:145-157.
- Bolhuis, J.E., E.D. Ellen, C.G. van Reenen, J. de Groot, J. ten Napel, R.E. Koopmanschap, G. de Vries Reilingh, K.A. Uitdehaag, B. Kemp, and T.B. Rodenburg. 2009. Effects of genetic group selection against mortality on behavior and peripheral serotonin in domestic laying hens with trimmed and intact beaks. *Physiology & Behavior*. 97:470-5.
- Brinker T., P. Bijma, J. Visscher, T.B. Rodenburg, and E.D. Ellen. 2014. Plumage condition in laying hens: genetic parameters for direct and indirect effects in two purebred layer lines. *Genetics Selection Evolution*. 46:33.
- Brinker, T., B. Raymond, P. Bijma, A. Vereijken, and E.D. Ellen. 2016. Estimation of total genetic effects for survival time in crossbred laying hens showing

- cannibalism, using pedigree or genomic information. *Journal of Animal Breeding and Genetics*. 134:60-8.
- Brinker, T., E.D. Ellen, R.F. Veerkamp, and P. Bijma. 2015. Predicting direct and indirect breeding values for survival time in laying hens using repeated measures. *Genetics Selection Evolution*. 47:1-10.
- Buitenhuis, A.J., T.B. Rodenburg, M. Siwek, S.J. Cornelissen, M.G. Nieuwland, R.P. Crooijmans, M.A. Groenen, P. Koene, H. Bovenhuis, and J.J. van der Poel. 2003. Identification of quantitative trait loci for receiving pecks in young and adult laying hens. *Poultry Science*. 82:1661-1667.
- C**
- Calus, M.P.L. 2015. *calc_grm-a Programme to compute pedigree, genomic, and combined relationship matrices*. Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Wageningen.
- Cantet, R.J.C. and E.P. Cappa. 2008. On identifiability of (co) variance components in animal models with competition effects. *Journal of Animal Breeding and Genetics*. 125:371-381.
- Casey, D.E., J. Gerlach, and E. Christensson. 1980. Behavioral aspects of GABA-dopamine interrelationships in the monkey. *Brain Research Bulletin*. 5:269-73.
- Chang, C.C, C.C. Chow, L.C.A.M. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: rising the challenge of larger and richer datasets. *Gigascience*. 4:7.
- Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*. 42:1-8.
- Costa e Silva, J. B.M. Potts, P. Bijma, R.J. Kerr, and D. Pilbeam. 2013. Genetic control of interactions among individuals: contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. *New Phytologist*. 197:631-641.
- Costa e Silva, J., B.M. Potts, A.R. Gilmour, and R.J. Kerr. 2017. Genetic-based interactions among tree neighbours: identification of the most influential neighbors, estimation of correlations among direct and indirect effects for leaf disease and growth in *Eucalyptus Globulus*. *Heredity*. 119:125-135.
- Craig, J.V. and M.W. Muir. 1996. Group selection for adaptation to multiple-hen cages: beak-related mortality, feathering, and body weight responses. *Poultry Science*. 75:294-302.
- Craig, J.V, and M.W. Muir. 1989. Fearful and associated responses of caged White Leghorn hens: genetic parameter estimates. *Poultry Science*. 68:1040-1046.

D

- Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, 3:e3395.
- Damme, K. and F. Pirchner. Genetic differences of feather-loss in layers and effects on production traits. *Archiv fuer Gefluegelkunde (Germany, FR)*. 48:215-222.
- De Almeida, R.M.M., P.F. Ferrari, S. Parmigiani, and K.A. Miczek. 2005. Escalated aggressive behaviour: dopamine, serotonin and GABA. *European Journal of Pharmacology*. 526:51-64.
- Dempster, E.R., and I.M. Lerner. 1950. Heritability of threshold characters. *Genetics*. 35:212-236.
- Denison, R.F., E.T. Kiers and S.A. West. 2003. Darwinian agriculture: when can humans find solutions beyond the reach of natural selection? *The Quarterly Review of Biology*. 78:145-168.
- Devlin, B., and K. Roeder. 1999. Genomic control for association studies. *Biometrics*. 55:997-1004.
- Douglas, J.A., M. Boehnke, and K. Lange. 2000. A multipoint methods for detecting genotyping errors and mutations in sibling-pair linkage data. *American Journal of Human Genetics*. 66:1287-1297.
- Ducrocq, V., B. Besbes, and M. Protais. 2000. Genetic improvement of laying hens viability using survival analysis. *Genetics Selection Evolution*. 32:23-40.
- Ducrocq, V. and G. Casella. 1996. A Bayesian analysis of mixed survival models. *Genetics Selection Evolution*. 28:505.

E

- Ellen, E.D., P. Bijma. 2019. Can breeders solve mortality due to feather pecking in laying hens?. *Poultry Science*. 98:3431-3442.
- Ellen, E.D., J. Visscher, and P. Bijma. 2014. Comparison of empirical and theoretical responses to selection against mortality due to cannibalism in layers. 10th WCGALP. 17.08-22.08. Vancouver, Canada.
- Ellen, E.D., V. Ducrocq, B.J. Ducro, R.F. Veerkamp, and P. Bijma. 2011. Combining survival analysis and a linear animal model to estimate genetic parameters for social effects on survival in layers. 62nd EAAP. 29.08-02.09. Stavanger, Norway.
- Ellen, E.D., V. Ducrocq, B.J. Ducro, R.F. Veerkamp, and P. Bijma. 2010. Genetic parameters for social effects on survival in cannibalistic layers: combining survival analysis and a linear animal model. *Genetics Selection Evolution*. 42:27.
- 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. *Poultry Science*. 87:233-239.

- Ellen, E.D., W.M. Muir, F. Teuscher, and P. Bijma. 2007. Genetic improvement of traits affected by interactions among individuals: sib selection schemes. *Genetics*. 176:489-499.
- Evans, S.R., D. Waldvogel, N. Vasiljevic, and E. Postma. 2018. Heritable spouse effects increase evolutionary potential of human reproductive timing. *Proceedings Royal Society B*. 285: 20172763.

F

- Falconer, D.S. and T.F.C. Mackay. 1996. *Introduction to quantitative genetics*. Harlow: Pearson Education Limited.
- FAOstat. 2016. Statistical database of the Food and Agricultural Organization of the United Nations.
- FAO. 2015. <https://www.un.org/sustainabledevelopment/>
- Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*. 52:399-433.
- Flisikowski, K. H. Schwarzenbacher, M. Wysocki, S. Weigend, R. Preisinger, J.B. Kjaer, and R. Fries. 2009. Variation in neighbouring genes of the dopaminergic and serotonergic systems affects feather pecking behaviour of laying hens. *Animal Genetics*. 40:192–9.
- Fossum, O., D.S. Jansson, P.E. Etterlin, and I. Vågsholm. 2009. Causes of mortality in laying hens in different housing systems in 2001 to 2004. *Acta Veterinaria Scandinavica*. 51:3.

G

- Gentle, M.J., D. Waddington, L.N. Hunter, and R.B. Jones. 1990. Behavioural evidence for persistent pain following partial beak amputation in chickens. *Applied Animal Behaviour Science*. 27:149-157.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2014. *ASReml User Guide*. Release 4.0. VSN International Ltd: Hemel Hempstead.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. *ASReml User Guide*. Release 3.0. VSN International Ltd: Hemel Hempstead.
- Griffing, B. 1967. Selection in reference to biological groups I. Individual and group selection applied to populations of unordered groups. *Australian Journal of Biological Sciences*. 20:127-139.

H

- Hale A.J., E. ter Steege, and J. den Hertog. 2017. Recent advances in understanding the role of protein-tyrosine phosphatases in development and disease. *Developmental Biology*. 428:283–92
- Han, C.S., C. Tuni, J. Ulcik, and N.J. Dingemanse. 2018. Increased developmental density decreases the magnitude of indirect genetic effects expressed during agonistic interactions in an insect. *Evolution*. 72(11):2435-2448.
- Hickey, J.M., B.P. Kinghorn, B. Tier, J.F. Wilson, N. Dunstan, and J.H.J. van der Werf. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. *Genetics Selection Evolution*. 43:12.
- Hill, W.G., and B.S. Weir. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetic Research*. 93:47-64.
- Hill, W.G., M.E. Goddard, and P.M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*. 4:e1000008.
- Hill, W.G. 2008. Estimation, effectiveness and opportunities of long term genetic improvement in animals and maize. *Lohmann Inform*. 43:3-20.
- Hocking, P.M., C.E. Channing, G.W. Robertson, A. Edmond, and R.B. Jones. 2004. Between breed genetic variation for welfare-related behavioural traits in domestic fowl. *Applied Animal Behaviour Science*. 89:85-105.
- Hoeschele, I. and B. Tier. 1995. Estimation of variance components of threshold characters by marginal posterior modes and means via Gibbs sampling. *Genetics Selection Evolution*. 27:519-540.
- Hughes, B.O. and I.J.H. Duncan. 1972. The influence of strain and environmental factors upon feather pecking and cannibalism in fowls. *British Poultry Science*. 13:525-547.

I

- Ibáñez-Escriche, N., R.L. Fernando, A. Toosi, and J.C.M. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. *Genetics Selection Evolution*. 41:12.

J

- Jairath, L., J.C.M. Dekkers, L.R. Schaeffer, Z. Liu, E.B. Burnside, and B. Kolstad. 1998. Genetic evaluation for herd life in Canada. *Journal of Dairy Science*. 81:550-562.
- Jamrozik, J., J. Fatehi, and L.R. Schaeffer. 2008. Comparison of models for genetic evaluation of survival traits in dairy cattle: a simulation study. *Journal of Animal Breeding and Genetics*. 125:75-83.

K

- Kalbfleish, J.D. and R.L. Prentice. 1980. The statistical analysis of failure time data. Hoboken: John Wiley & Sons.
- Keeling, L., L. Andersson, K.E. Schütz, S. Kerje, R. Fredriksson, Ö. Carlborg, C.K. Cornwallis, T. Pizzari, and P. Jensen. 2004. Feather-pecking and victim pigmentation. *Nature*. 431:645-646.
- Ki Hong, J., Y.M. Kim, K.H. Cho, J.C. Park, and D.H. Lee. 2017. Reproductive performance of sows selected for divergent social genetic effects for growth. *Animal reproduction*, 14:1292-1297.
- Kjaer, J.B., and P. Sørensen, G. Su. 2001. Divergent selection of feather pecking behaviour in laying hens (*Gallus gallus domesticus*). *Applied Animal Behaviour Science*. 71:229-239.
- Kjaer, J.B. and K.S. Vestergaard. 1999. Development of feather pecking in relation to light intensity. *Applied Animal Behaviour Science*. 62:243-254.
- Kjaer, J.B., and P. Sørensen. 1997. Feather pecking behaviour in White Leghorns, a genetic study. *British Poultry Science*. 38:333-341.

L

- Lagakos, S.W. 1979. General right censoring and its impact on the analysis of survival data. *Biometrics*. 35:139-156.
- Li, Z., M. Zheng, B.A. Abdalla, Z. Zhang, Z. Xu, Q. Ye, H. Xu, W. Luo, Q. Nie, and X. Zhang. 2016. Genome-wide association study of aggressive behaviour in chicken. *Scientific Reports*. 6:30981.
- Linksvayer, T. 2006. Direct, maternal, and sibsocial genetic effects on individual and colony traits in an ant. *Evolution*. 60(12):2252-2561.
- Lipschutz-Powell, D., J.A. Wooliams, P. Bijma, R. Pong-Wong, M.L. Bermingham, and A.B. Doeschl-Wilson. 2012. Bias, accuracy, and impact of indirect genetic effects in infectious diseases. *Frontiers in Genetics*. 3:215.

M

- Madgwick, P.A. and M.E. Goddard. 1989. Genetic and phenotypic parameters for longevity in Australian dairy cattle. *Journal of Dairy Science*. 72:2624-2632.
- Meuwissen, T.H.E. 2007. Genomic selection: marker assisted selection on a genome wide scale. *Journal of Animal Breeding and Genetics*. 124:321-322.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157:1819-1829.

References

- Moore, A.J., E.D. Brodie III, and J.B. Wolf. 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution*. 51:1352-1362.
- Muir, W.M., P. Bijma, and A. Schnickel. 2013. Multilevel selection with kin and non-kin groups, experimental results with Japanese quail (*Coturnix japonica*). *Evolution*. 67:1598-1606.
- Muir, W.M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics*. 124:342-355.
- Muir, W.M. 2005. Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics*. 170:1247-1259.
- Muir, W.M. 1996. Group selection for adaptation to multiple-hen cages: selection program and direct responses. 1996. *Poultry Science*. 75:447-458.
- Muranty, H., M. Troggio, I.B. Sadok, M. Al Rifaï, A. Auwerkerken, E. Banchi, R. Velasco, P. Stevanato, W.E. van de Weg, M. di Guardo, S. Kumar, F. Laurens, and M.C.A.M. Bink. 2015. Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research*, 2:15060.
- Mutic, J.J and J.B. Wolf. 2007. Indirect genetic effects from ecological interactions in *Arabidopsis thaliana*. *Molecular Ecology*. 16:2371-2381.

N

- Newberry, R.C. 2004. Cannibalism. In: *Welfare of the laying hen*. Perry GC, editor. Oxfordshire: CAB International. 239-258.
- Nicol, C.J., N.G. Gregory, T.G. Knowles, I.D. Parkman, and L.J. Wilkins. 1999. Differential effects of increased stocking density, mediated by increased flock size, on feather pecking and aggression in laying hens. *Applied Animal Behaviour Science*. 65:137-152

O

- Ødegård, J., I. Olesen, B. Gjerde, and G. Klemetsdal. 2006. Evaluation of statistical models for genetic analysis of challenge test data on furunculosis resistance in Atlantic salmon (*Salmo salar*): prediction of field survival. *Aquaculture*. 256:116-123.
- Olesen, I., M. Perez-Enciso, D. Gianola, and D.L. Thomas. 1994. A comparison of normal and nonnormal mixed models for number of lambs born in Norwegian sheep. *Journal of Animal Science*. 72:1166-1173.

P

- Parmentier, H.K., T.B. Rodenburg, G. de Vries Reilingh, B. Beerda, and B. Kemp. 2009. Does enhancement of specific immune responses predispose laying hens for feather pecking?. *Poultry science*. 88:536-542.
- Peeters, K. 2015. Genetics of social interactions in laying hens: improving survival and productivity. PhD thesis, Animal Breeding and Genomics Centre, Wageningen University, Wageningen.
- Peeters, K., E.D. Ellen, and P. Bijma. 2013. Using pooled data to estimate variance components and breeding values for traits affected by social interactions. *Genetics Selection Evolution*. 45:27.
- Peeters, K., T. Eppink, E.D. Ellen, J. Visscher, and P. Bijma. 2012. Indirect genetic effects for survival in domestic chicken (*Gallus gallus domesticus*) are magnified in crossbred genotypes and show a parent-of-origin effect. *Genetics*. 192:705-713.
- Pollak, E.J., J. van der Werf, and R.L. Quaas. 1984. Selection bias and multiple trait evaluation. *Journal of Dairy Science*. 67:1590-1595.
- Poshivalov, V.P. 1981. GABA-ergic correlates between aggressiveness and sociability in isolated mice. *Bulletin of Experimental Biology and Medicine*. 91:648-50.
- Preisinger, R. 1998. Internationalisation of breeding programmes-breeding egg-type chickens for a global market. 6th WCGALP. 11.01-16.01. Armidale, Australia.
- Purcell, S., C.C. Chang. 2015 PLINK 1.9. <https://www.cog-genomics.org/plink>. Accessed 22 Dec 2017
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, and M.J. Daly. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 81:559-575.

R

- Ragab, M., M. Piles, R. Quintanilla, and J.P. Sánchez. 2018. Indirect genetic effect model using feeding behaviour traits to define the degree of interaction between mates: an implementation in pigs growth rate. *Animal*. 13(2):231-239.
- R Core Team. 2015. R: a language and environment for statistical computing. In: R foundation for statistical computing. Vienna. <http://www.r-project.org/>.
- Riber, A.B., and B. Forkman. A note on the behavior of the chicken that received feather pecks. 2007. *Applied Animal Behaviour Science*. 108:337-341.
- Riber, A.B. and L.K. Hinrichsen. 2017. Welfare consequences of omitting beak trimming in barn layers. *Frontiers in Veterinary Science*. 4:222.

References

- Rincent, R., L. Moreau, H. Monod, E. Kuhn, A.E. Melchinger, R.A. Malvar, J. Moreno-Gonzalez, S. Nicolas, D. Madur, V. Combes, F. Dumas, T. Altmann, D. Brunel, M. Ouzunova, P. Flament, P. Dubreuil, A. Charcosset, and T. Mary-Huard. 2014. Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics*. 197:375–87.
- Rode, N.O., P. Soroye, R. Kassen, and H.D. Rundle. 2017. Air-borne genotype by genotype indirect genetic effects are substantial in the filamentous fungus *Aspergillus nidulans*. *Heredity*. 119:1-7.
- Rodenburg, T.B., M.M. van Krimpen, I.C. de Jong, E.N. de Haas, M.S. Kops, B.J. Riedstra, R.E. Nordquist, J.P. Wagenaar, M. Bestman, and C.J. Nicol. 2013. The prevention and control of feather pecking in laying hens: identifying the underlying principles. *World's Poultry Science Journal*. 69:361-374.
- Rodenburg, T.B., K. de Reu, and F.A.M. Tuytens. 2012. Performance, welfare, health and hygiene of laying hens in non-cage systems in comparison with cage systems. 30th Poultry Science Symposium. 07.09-09.09. Glasgow, Scotland. 210-224.
- Rodenburg, T.B., P. Bijma, E.D. Ellen, R. Bergsma, S. de Vries, J.E. Bolhuis, B. Kemp, and J.A.M. van Arendonk. 2010. Breeding amiable animals? Improving farm animal welfare by including social effects in breeding programmes. *Animal Welfare*. 19:77-82.
- Rönnegård, L, S.E. McFarlane, A. Husby, T. Kawakami, H. Ellegren, and A. Qvarnström. 2016. Increasing the power of genome wide association studies in natural populations using repeated measures-evaluation and implementation. *Methods in Ecology and Evolution*. 7:792-9.
- Roport, N., and N. Guy. 1991. Serotonin facilitates GABAergic transmission in the CA1 region of rat hippocampus in vitro. *Journal of Physiology*. 441:121-36.

S

- SAS. SAS User's Guide. 2004. Release 9.2. Cary: Statistics Institute Inc.
- Savory, C.J. 1995. Feather pecking and cannibalism. *World's Poultry Science Journal*. 51:215-219.
- Sevillano, C.A., J. Vandenplas, J.W.M. Bastiaansen, R. Bergsma, and M.P.L. Calus, 2017. Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles. *Genetics Selection Evolution*. 49:75.
- Schaeffer, L.R. 2004. Application of random regression models in animal breeding. *Livestock Production Science*. 86:35-45.
- Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*. 36:111-147.

- Storey, J.D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences USA*. 100:9440-5.
- Stutzmann G.E., and J.E. LeDoux. 1999. GABAergic antagonists block the inhibitory effects of serotonin in the lateral amygdala: a mechanism for modulation of sensory inputs related to fear conditioning. *Journal of Neuroscience*. 19:RC8.

T

- Takahashi, A., A. Shimamoto, C.O. Boyson, J.F. DeBold, and K.A. Miczek. 2010. GABA(B) receptor modulation of serotonin neurons in the dorsal raphe nucleus and escalation of aggression in mice. *Journal of Neuroscience*. 30:11771-80.
- Takahashi, A., A.N. Schilit, J. Kim, J.F. DeBold, T. Koide, and K.A. Miczek. 2012. Behavioral characterization of escalated aggression induced by GABAB receptor activation in the dorsal raphe nucleus. *Psychopharmacology*. 224:155-66.

U

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

V

- Vandenplas, J., M.P.L. Calus, C.A. Sevillano, J.J. Windig, and J.W.M. Bastiaansen. 2016. Assigning breed origin to alleles in crossbred animal. *Genetics Selection Evolution*. 48:61.
- Van Horne, P.L.M. and T.J. Achterbosch. 2008. Animal welfare in poultry production systems: Impact of EU standards on world trade. *World's Poultry Science Journal*. 64:40-52.
- Van Krimpen, M.M., R.P. Kwakkel, B.F.J. Reuvekamp, C.M.C. van der Peet-Schwering, L.A. den Hartog, and M.W.A. Verstegen. 2005. Impact of feeding management on feather pecking in laying hens. *World's Poultry Science Journal*. 61:663-686.
- Van Pelt, M.L., T.H.E. Meuwissen, G. de Jong, and R.F. Veerkamp. 2014. Genetic analysis of longevity in Dutch dairy cattle using random regression. *Journal of Dairy Science*. 98:4417-4430.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*. 91:4414-4423.

References

- Vazquez, A.I., D. Gianola, D. Bates, K.A. Weigel, and B. Heringstad. 2009. Assessment of Poisson, logit, and linear models for genetic analysis of clinical mastitis in Norwegian Red cows. *Journal of Dairy Science*. 92:739-748.
- Veerkamp, R.F., S. Brotherstone, B. Engel, and T.H.E. Meuwissen. 2001. Analysis of censored survival data using random regression models. *Animal Science*. 72:1-10.

W

- Wolc, A., C. Stricker, J. Arango, P. Settar, J.E. Fulton, N.P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D.J. Garrick, S.J. Lamont, and J.C.M. Dekkers. 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution*. 43:5.
- Wolf, J.B., E.D. Brodie III, J.M. Cheverud, A.J. Moore, and M.J. Wade. 1998. Evolutionary consequences of indirect genetic effects. *Trends in Ecology & Evolution*. 13:64-69.

Y

- Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin and G.W. Montgomery. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. 42:565-569.

Z

- Zhang, M., X.T. Zou, H. Li, X.Y. Dong, and W. Zhao. 2012. Effect of dietary γ -aminobutyric acid on laying performance, egg quality, immune activity and endocrine hormone in heat-stressed Roman hens. *Journal of Animal Science*. 83:141-7.

Additional files

Additional file 2.1

Table S1 Estimates of genetic parameters from the univariate classical animal model and corresponding SE of FCS

Line	Region	σ_A	σ_c^2	σ_p^2
W1	Neck	0.134 ± 0.017	0.099 ± 0.006	0.307 ± 0.007
	Back	0.117 ± 0.015	0.035 ± 0.003	0.186 ± 0.004
	Rump	0.059 ± 0.014	0.105 ± 0.005	0.243 ± 0.006
	Belly	0.180 ± 0.021	0.099 ± 0.007	0.414 ± 0.009
	Total	0.348 ± 0.043	0.805 ± 0.044	2.031 ± 0.048
WB	Neck	0.307 ± 0.029	0.471 ± 0.023	0.956 ± 0.025
	Back	0.221 ± 0.030	0.435 ± 0.022	0.933 ± 0.023
	Rump	0.238 ± 0.030	0.680 ± 0.031	1.184 ± 0.032
	Belly	0.229 ± 0.031	0.235 ± 0.016	0.831 ± 0.019
	Total	0.949 ± 0.091	5.260 ± 0.241	9.479 ± 0.257

Estimates are shown for FCS in each body region and for total FCS in two lines (W1, WB); σ_A is the genetic standard deviation; σ_c^2 is the cage variance; σ_p^2 is the phenotypic variance: $\sigma_p^2 = \sigma_A^2 + \sigma_c^2 + \sigma_e^2$. FCS is feather condition score, for each body region FCS ranges from 1 to 5, and total FCS ranges from 4 to 20.

Table S2 Estimates of genetic parameters from the univariate direct-indirect effects model and corresponding SE of FCS

Line	Region	σ_{A_D}	σ_{A_S}	σ_{TBV}	σ_c^2	σ_p^2	r_A
W1	Neck	0.135 ± 0.019	0.061 ± 0.013	0.244 ± 0.049	0.087 ± 0.007	0.306 ± 0.007	0.16 ± 0.23
	Back	0.118 ± 0.015	0.055 ± 0.010	0.218 ± 0.036	0.026 ± 0.004	0.185 ± 0.004	0.16 ± 0.21
	Rump	0.045 ± 0.019	0.051 ± 0.013	0.154 ± 0.052	0.100 ± 0.006	0.243 ± 0.006	-0.12 ± 0.51
	Belly	0.181 ± 0.021	0.048 ± 0.016	0.243 ± 0.052	0.092 ± 0.009	0.414 ± 0.009	0.11 ± 0.28
	Total	0.335 ± 0.047	0.197 ± 0.033	0.713 ± 0.129	0.701 ± 0.053	2.022 ± 0.049	0.12 ± 0.21
WB	Neck	0.270 ± 0.031	0.193 ± 0.026	0.595 ± 0.097	0.403 ± 0.029	0.954 ± 0.026	-0.17 ± 0.18
	Back	0.157 ± 0.032	0.204 ± 0.025	0.572 ± 0.092	0.367 ± 0.027	0.936 ± 0.025	-0.38 ± 0.24
	Rump	0.185 ± 0.035	0.240 ± 0.030	0.735 ± 0.110	0.559 ± 0.040	1.185 ± 0.034	-0.05 ± 0.22
	Belly	0.241 ± 0.033	0.109 ± 0.022	0.459 ± 0.084	0.197 ± 0.021	0.832 ± 0.019	0.29 ± 0.22
	Total	0.718 ± 0.096	0.775 ± 0.084	2.254 ± 0.309	4.170 ± 0.318	9.475 ± 0.276	-0.25 ± 0.18

Estimates are shown for FCS in each body region and for total FCS in two lines (W1, WB); σ_{A_D} and σ_{A_S} are estimates of the standard deviation of the direct breeding value and the standard deviation of the indirect breeding value. σ_{TBV} is the standard deviation of the total breeding value: $\sigma_{TBV}^2 = \sigma_{A_D}^2 + 2(n-1)\sigma_{A_DS} + (n-1)^2\sigma_{A_S}^2$; σ_c^2 is the cage variance; σ_p^2 is the phenotypic variance: $\sigma_p^2 = \sigma_{A_D}^2 + (n-1)\sigma_{A_S}^2 + \sigma_c^2 + \sigma_e^2$. r_A is the genetic correlation between the direct breeding value and the indirect breeding value. FCS is feather condition score, for each body region FCS ranges from 1 to 5, and total FCS ranges from 4 to 20.

Table S3 Estimates of genetic correlations between the four body regions from the direct-indirect-effects model

Line	Region	Region			
		Neck	Back	Rump	Belly
W1	Neck		0.84 ± 0.13	NC ¹	> 0.99
	Back	0.75 ± 0.11		0.92 ± 0.14	0.75 ± 0.30
	Rump	NC ¹	0.87 ± 0.20		> 0.99
	Belly	0.09 ± 0.18	-0.02 ± 0.18	0.48 ± 0.28	
WB	Neck		NC ¹	0.97 ± 0.03	0.92 ± 0.09
	Back	NC ¹		0.98 ± 0.03	0.85 ± 0.11
	Rump	0.65 ± 0.14	0.74 ± 0.13		0.74 ± 0.13
	Belly	0.47 ± 0.15	0.46 ± 0.20	0.25 ± 0.21	

Estimates are shown for FCS in each body region in two lines (W1, WB); the estimates below the diagonal refer to the direct genetic correlations (r_{12_D}); the estimates above the diagonal refer to the indirect genetic correlations (r_{12_S}); ¹NC = not converged.

Table S4 Means with corresponding SD of the four persons who recorded the FCS

Person	n	W1				WB			
		Neck	Back	Rump	Belly	Neck	Back	Rump	Belly
1	2988	1.30 ± 0.59	1.30 ± 0.41	1.19 ± 0.52	1.36 ± 0.70	2.28 ± 1.18	1.60 ± 1.04	2.17 ± 1.25	1.72 ± 1.01
2	2870	1.22 ± 0.47	1.12 ± 0.40	1.23 ± 0.57	1.32 ± 0.75	1.99 ± 1.01	1.70 ± 1.10	2.26 ± 1.19	1.57 ± 0.99
3	2970	1.48 ± 0.68	1.16 ± 0.46	1.13 ± 0.46	1.36 ± 0.67	2.65 ± 1.21	1.68 ± 1.11	2.13 ± 1.30	1.65 ± 0.96
4	3364	1.46 ± 0.57	1.22 ± 0.48	1.16 ± 0.48	1.17 ± 0.50	2.32 ± 1.11	1.63 ± 0.96	2.11 ± 1.26	1.49 ± 0.82

Estimates with corresponding SD are shown for each body region in two lines (W1, WB); FCS is feather condition score, for each body region FCS ranges from 1 to 5, and total FCS ranges from 4 to 20.

Additional file 3.1 – Predicting phenotypes using STM, RMM.t, RMM.p, and GLMM

In this study, the quality of estimated breeding values was judged using cross validation for survival time. Known phenotypes were set to missing, their values were predicted, and finally the predicted values were correlated to the observed values. Predicted phenotypes were the rank of predicted survival times of individuals. Phenotypes were predicted by combining the estimated DGE (\hat{A}_{D_i}) of the individual itself and the estimated IGE of its cage mates ($n = 3$) that were present at the start of the experiment ($\sum \hat{A}_{I_j}$). Thus the predicted phenotype of individual i was:

$$\hat{P}_i = \hat{A}_{D_i} + \sum_{j=1}^{n-1} \hat{A}_{I_j},$$

where estimated breeding values \hat{A}_{D_i} and $\sum_{j=1}^{n-1} \hat{A}_{I_j}$ refer to survival time. For STM, these estimated breeding values follow directly from the analysis, whereas for RMM.t, RMM.p, and the GLMM the estimated breeding values have to be translated to the survival time scale.

For RMM.t and RMM.p, the sum of all repeated measures of an individual i corresponds to the observed survival time in months of individual i . Hence, for RMM.t, predicted phenotypes were obtained as:

$$\hat{P}_i = \sum_t \left(\left(\hat{A}_{D_i} + \sum_{j=1}^{n-1} \hat{A}_{I_j} \right) t \right),$$

and for RMM.p, predicted phenotypes were obtained as:

$$\hat{P}_i = \sum_t \left(\left(\hat{A}_{D_i} + \sum_{j=1}^{n-1} \hat{A}_{I_j} \right) x_t \right),$$

with x_t being a function of mean survival at time t , $x_t = \sqrt{p_t(1-p_t)}$, p_t denoting mean survival at time t .

The GLMM gives predictions on the underlying logit scale. Before calculating \hat{P}_i , these predictions need to be back transformed to the observed scale. Since $\text{logit}(E(y)) = \log(p/(1-p))$, for any individual i at time t , the predicted survival probability is given by:

$$\hat{p}_{i,t} = \frac{e^{\mu(t) + \hat{A}_{D_i} + \sum_{j=1}^{n-1} \hat{A}_{I_j}}}{1 + e^{\mu(t) + \hat{A}_{D_i} + \sum_{j=1}^{n-1} \hat{A}_{I_j}}},$$

where $\mu(t)$ represents the fixed effect solution at each period coming from, $\eta(E(y)) = \mathbf{Xb}$, being a sixth order polynomial of time. The fixed effect was included because the transformation from the underlying to the observed scale is non-linear. By including the mean, the DGE and IGE predictions are scaled to the mean population survival for each month. Finally, the predicted phenotype for the GLMM is the sum of the predicted survival probabilities at each point in time,

$$\hat{P}_i = \sum_t \hat{p}_{i_t}.$$

Additional file 3.2 – Translating breeding values and genetic parameters of RMM.t to survival days

In this study, survival was modelled as being dead (0) or alive (1) each month. More interest is however in the number of survival days. This appendix shows how the breeding values, estimated genetic parameters, and phenotypic variance for survival can be translated to the survival time scale.

Survival time (ST_i) of an individual i is the sum of its survival records (S_i) at each day,

$$ST_i = c \cdot \sum_{t_1}^{t_2} S_{i_t},$$

with c being a multiplication factor to translate monthly survival into days; $c = 30.4$, t_1 is the start of the laying period (in months), and t_2 the end of the laying period. The continuous time equivalent of this expression is:

$$ST_i = c \cdot \int_{t_1}^{t_2} S_{i_t} dt.$$

Substituting ST_i by Equation 1 and S_{i_t} by Equation 2, ignoring the fixed effects, and using an animal model for survival time yields:

$$\begin{aligned} & A_{ST,D_i} + \sum_{j \neq i}^{n-1} A_{ST,I_j} + cage_{ST,k} + e_{ST,ijk} \\ &= c \int_{t_1}^{t_2} \left(A_{S,D_i} t_l + \sum_{j \neq i}^{n-1} A_{S,I_j} t_l + cage_{S_{kl}} + cage_{S_k} t_l + PE_{S_i} t_l + e_{S_{ijkl}} \right) dt, \end{aligned}$$

where subscript ST denotes survival time, and subscript S denotes survival. This expression identifies the following correspondence between the direct and indirect breeding values for survival time and survival,

$$A_{ST,D_i} = c \int_{t_1}^{t_2} A_{S,D_i} t_l dt$$

and

$$\sum_{j \neq i}^{n-1} A_{ST,I_j} = c \int_{t_1}^{t_2} \sum_{j \neq i}^{n-1} A_{S,I_j} t_l dt$$

Solving the integrals shows how direct and indirect breeding values for survival are related to survival time,

$$A_{ST,D_i} = \frac{1}{2}c(t_2^2 - t_1^2)A_{S,D_i}$$

and

$$A_{ST,I_i} = \frac{1}{2}c(t_2^2 - t_1^2)A_{S,I_i}.$$

These two expressions can also be applied to the direct and indirect EBV. Taking the (co)variance shows that the relationship between the genetic parameters for survival time and survival is given by:

$$\sigma_{A_{ST,D}}^2 = \frac{1}{4}c^2(t_2^2 - t_1^2)^2 \sigma_{A_{S,D}}^2$$

$$\sigma_{A_{ST,I}}^2 = \frac{1}{4}c^2(t_2^2 - t_1^2)^2 \sigma_{A_{S,I}}^2$$

$$\sigma_{A_{ST,DI}} = \frac{1}{4}c^2(t_2^2 - t_1^2)^2 \sigma_{A_{S,DI}}.$$

These equations are used in Table 3.3 to translate estimated genetic parameters for survival to the survival time scale.

The relationship between the phenotype for survival time and the model for survival is given by (ignoring fixed effects):

$$y_{ST,ijk} = c \int_{t_1}^{t_2} \left(A_{S,D_i} t_l + \sum_{j \neq i}^{n-1} A_{S,I_j} t_l + cage_{S_{kl}} + cage_{S_k} t_l + PE_{S_i} t_l + e_{S_{ijkl}} \right) dt,$$

Solving the integral yields:

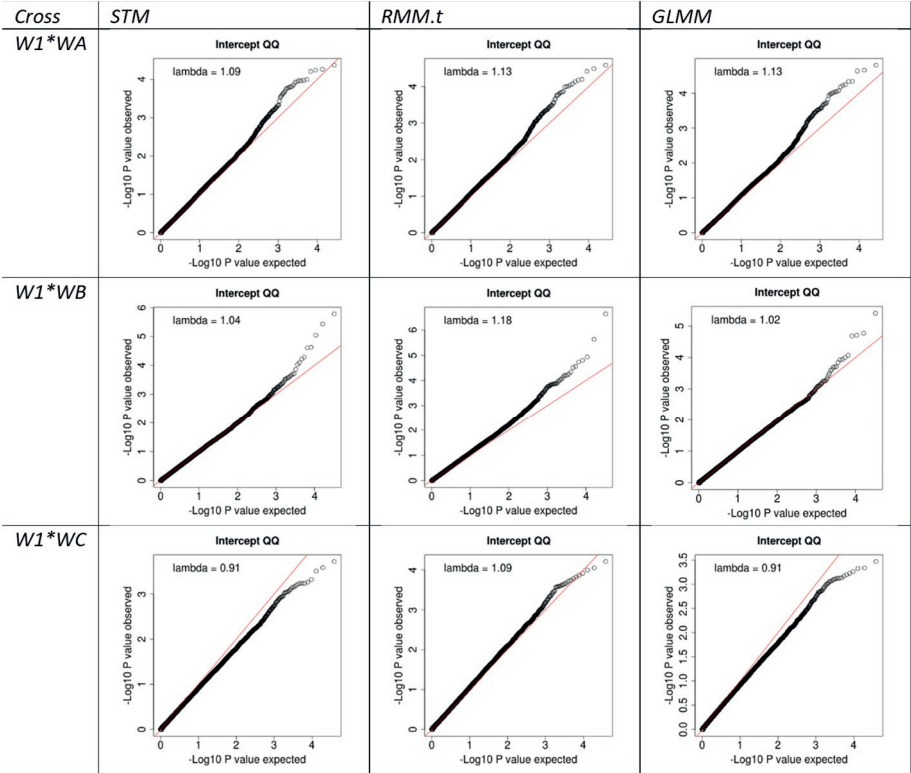
$$y_{ST,ijk} = c \left(\frac{1}{2} (t_2^2 - t_1^2) \left(A_{S,D_i} + \sum_{j \neq i}^{n-1} A_{S,I_j} + cage_{S_k} + PE_{S_i} \right) + \sum_{t_1}^{t_2} (cage_{S_{kl}} + e_{S_{ijkl}}) \right).$$

The last two terms remain summations because a different effect is fitted at each time point. Taking the variance on both sides shows that phenotypic variance for survival time is given by:

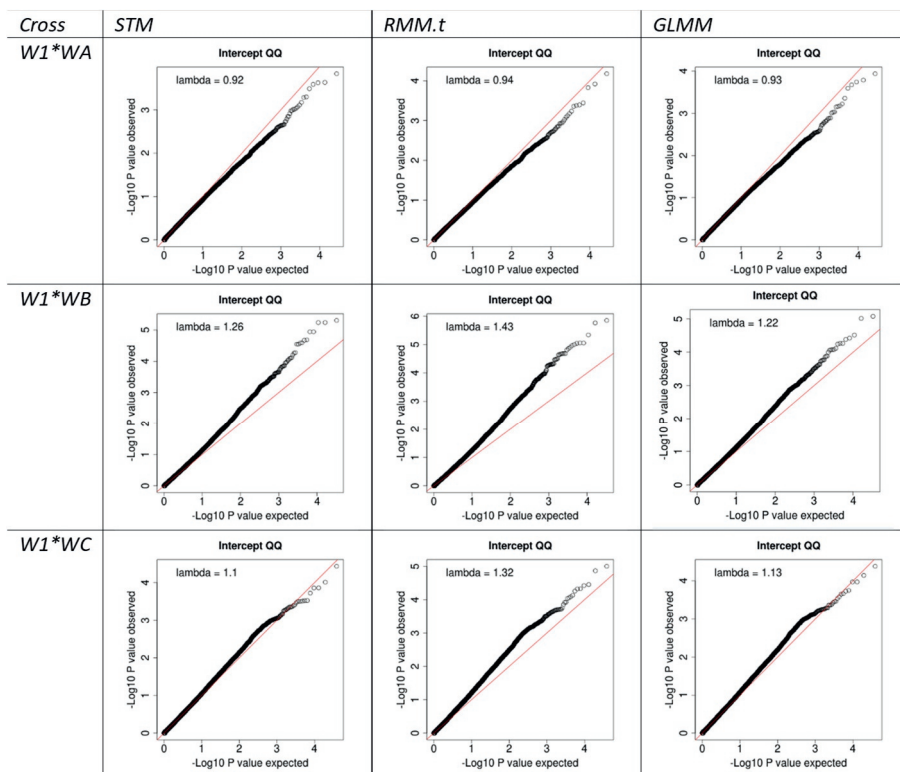
$$\sigma_{P,ST}^2 = c^2 \left[\frac{1}{4} (t_{end}^2 - t_{start}^2) (\sigma_{A_D}^2 + (n-1)\sigma_{A_I}^2 + \sigma_{cage_k}^2 + \sigma_{PE}^2) + (t_{end} - t_{start}) \cdot \sigma_{cage_{km}}^2 + \sum_{t=end}^{t=start} \sigma_e^2 \right].$$

The last term is kept as a sum because, in the survival model, a separate residual variance was fitted for each time point (whereas a variance common to all time points was fitted for the cage effect). This expression is used in Table 3 to calculate phenotypic variance on the survival time scale from the parameter estimates of the survival model.

Additional file 5.1 – QQ-plots of direct SNPs for STM, RMM.t, and GLMM



Additional file 5.2 – QQ-plots of indirect SNPs for STM, RMM.t, and GLMM



Additional file 5.3 – Number of direct and indirect SNPs with $p < 0.001$ for all crosses and models after genomic control

Effect	Model	W1*WA	W1*WB	W1*WC
Direct	STM	236	278	345
	RMM.t	226	283	332
	GLMM	237	245	330
Indirect	STM	226	356	374
	RMM.t	236	355	376
	GLMM	213	413	361

Additional file 5.4 – Power of direct and indirect SNP effects

To get an impression of the power in our data we calculated the power assuming a true SNP effect in days at allele frequencies ranging from 0-1, given the population specific parameters of the crosses (**Table 1**), and with an acceptance threshold of ~4 standard deviations from the mean.

Assuming a simple model for direct SNP effects: $y_i = \mu + b_D \cdot SNP_i + e_i$,

The SE of direct SNP effect b_D is given by:

$$SE(b_D) = \frac{\sigma_e}{\sqrt{2Np(1-p)}}$$

With $SE(b_D)$ the standard error of direct SNP effect b_D , σ_e being the residual standard deviation, N the sample size, and p the allele frequency.

The power was obtained from:

$$PWR(b_D) = 1 - \text{pnorm}\left(4 - \frac{|b_D|}{SE(b_D)}\right)$$

Assuming a simple model for indirect SNP effects: $y_i = \mu + b_I \cdot \sum_{j \neq i}^{n-1} SNP_j + e_i$,

The SE of indirect SNP effect b_I is given by:

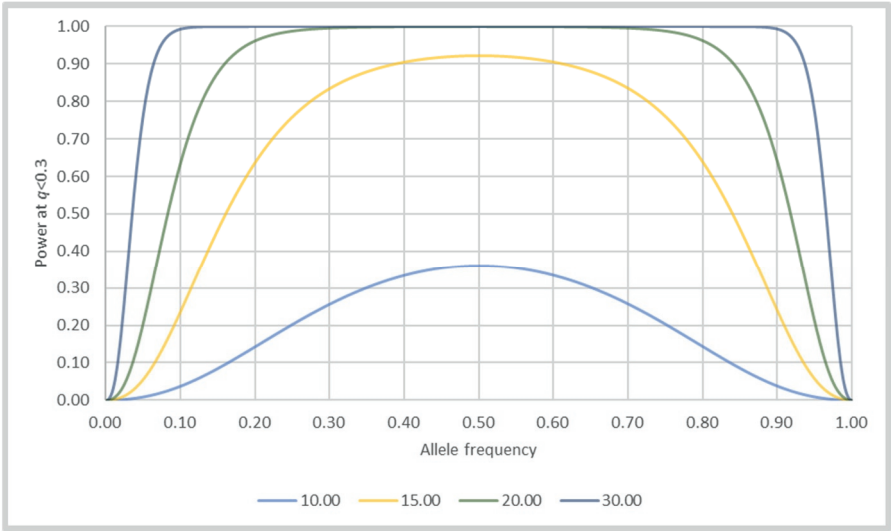
$$SE(b_I) = \frac{\sigma_e}{\sqrt{(n-1)^2 2Np(1-p)}}$$

With $SE(b_I)$ the standard error of indirect SNP effect b_I , and n the number of group members.

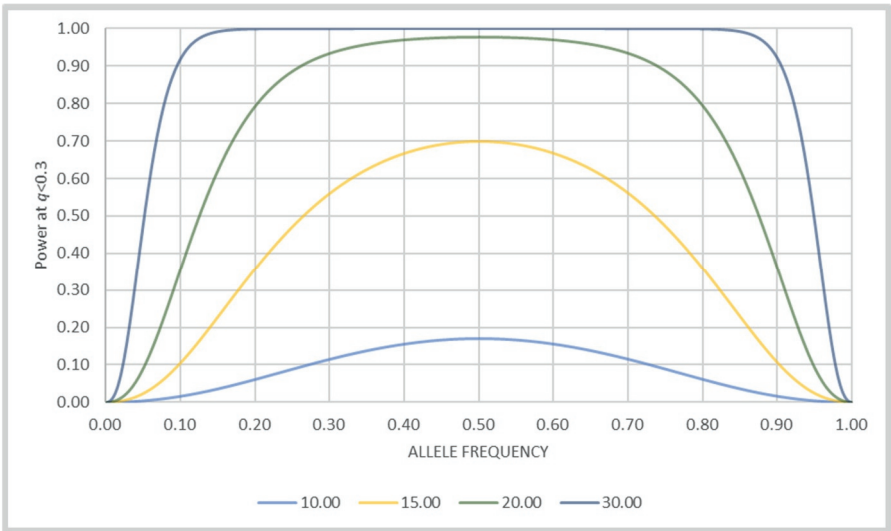
The power was obtained from:

$$PWR(b_I) = 1 - \text{pnorm}\left(4 - \frac{|b_I|}{SE(b_I)}\right)$$

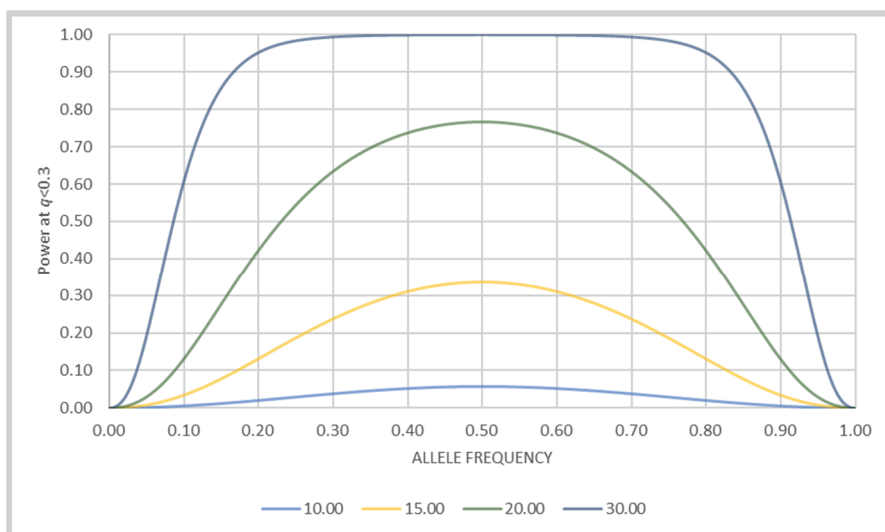
Graphs of power at allele frequencies 0-1.00 for direct SNP effects *b*



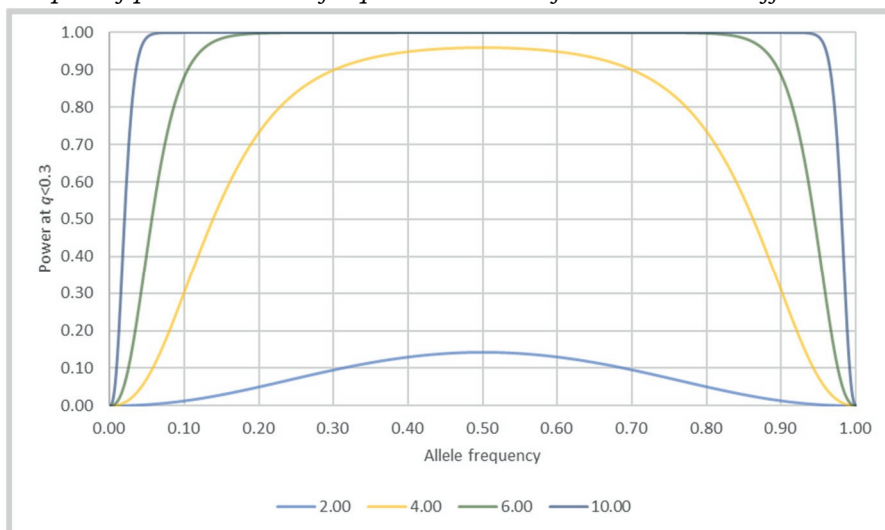
Power of detecting direct SNP effect *b* (ranging from 0-30 days) for a range of allele frequencies (0-1.00) for cross W1*WA.

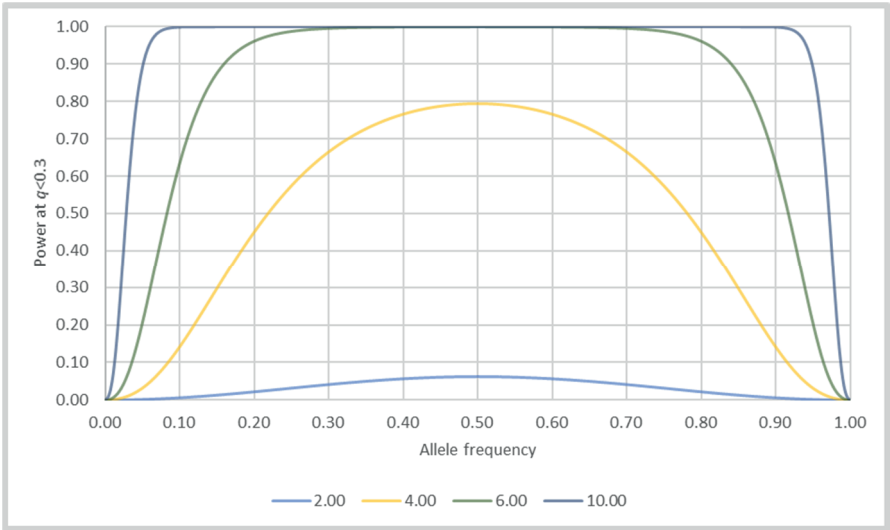


Power of detecting direct SNP effect *b* (ranging from 0-30 days) for a range of allele frequencies (0-1.00) for cross W1*WB.

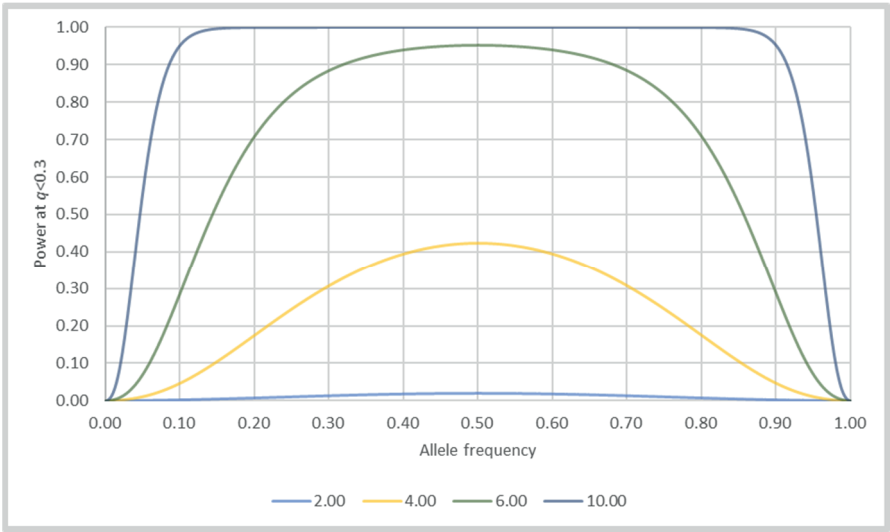


Graphs of power at allele frequencies 0-1.00 for indirect SNP effects b





Power of detecting indirect SNP effect b (ranging from 0-10 days) for a range of allele frequencies (0-1.00) for cross W1*WB.



Power of detecting indirect SNP effect b (ranging from 0-30 days) for a range of allele frequencies (0-1.00) for cross W1*WC.

Curriculum Vitae

About the author

Tessa Brinker was born on October 24th 1989 in Zevenaar, the Netherlands. She grew up in Zevenaar where she obtained her high school degree in 2007. After one year of studying Archeology at Leiden University, she made up her mind and enrolled for the BSc Animal Sciences at Wageningen University. During her BSc, she was a student assistant for the course Introduction of Animal Sciences and also for the course Advanced Statistics. Her BSc thesis was in the direction of Animal Production Systems. Afterwards, Tessa enrolled for the MSc Animal Sciences at Wageningen University. She pursued her job as student assistant for the course Advanced Statistics. She specialized in Animal Breeding and Genetics, doing both her internship and MSc thesis in this direction. For her internship, she went to CRV (Arnhem, NL) where she estimated genetic parameters for ketosis in dairy cows. For her MSc thesis, she estimated genetic parameters of plumage condition in laying hens. In her MSc thesis, Tessa performed analyses on a socially-affected trait, and her interest in this topic grew. This led to her decision to apply for a PhD project on the same topic: "The genetic architecture of socially-affected traits". She started her project in 2013. As part of her job as PhD candidate, she spent a three month period as a visiting PhD at the Quantitative Genetics department of Aarhus University (DK). Moreover, she was PhD representative in the WUR council for the period of one year. Tessa was also interested to share knowledge and she, therefore, supervised several BSc and MSc students, and also participated as course assistant in the course Genetic Improvement of Livestock. Finally, she was involved in the organization of several courses and workshops. From September 2017 until September 2019, she worked at the European Forum of Farm Animal Breeders. Among many other tasks, she was involved in the communication and dissemination package of several European projects focusing on animal breeding and reproduction. On 1st of May 2020, Tessa started working at a plant breeding company, Rijk Zwaan, in Fijnaart (NL). She works as biostatistician at the department of Quantitative Genetics.

Peer reviewed publications

- Brinker, T., B. Raymond, P. Bijma, A. Vereijken, and E.D. Ellen. 2016. Estimation of total genetic effects for survival time in crossbred laying hens showing cannibalism, using pedigree or genomic information. *Journal of Animal Breeding and Genetics*. 134.
- Brinker, T., E.D. Ellen, R.F. Veerkamp, and P. Bijma. 2015. Predicting direct and indirect breeding values for survival time in laying hens using repeated measures. *Genetics Selection Evolution*. 47:75.
- Brinker, T., P. Bijma, J. Visscher, T.B. Rodenburg, and E.D. Ellen. 2014. Plumage condition in laying hens: genetic parameters for direct and indirect effects in two purebred layer lines. 46:33.
- Brinker, T., P. Bijma, A. Vereijken, and E.D. Ellen. 2018. A GWAS for direct and indirect genetic effects on survival time in laying hens showing cannibalism. *Genetics Selection Evolution*. 50:38

Conference proceedings

- Ellen, E.D., T. Brinker, and P. Bijma. Genome-wide association study for direct and indirect genetic effects on survival time in layers. 68th Annual meeting of the EAAP. 28.08.2017-31.08.2017.
- Vandenplas, J., M.P.L. Calus, T. Brinker, E.D. Ellen, M.C.A.M. Bink, J. ten Napel. Single-step GBLUP using metafounders to predict crossbred performance of laying hens. 68th Annual meeting of the EAAP. 28.08.2017-31.08.2017.
- Brinker, T., P. Bijma, E.D. Ellen. Genome-wide association study for direct and indirect genetic effects on survival in crossbred layer hens. Gordon Research Conference on Quantitative Genetics & Genomics. 26.02.2017-03.03.2017. Galveston, USA.
- Brinker, T., E.D. Ellen, R. Veerkamp, and P. Bijma. Predicting direct and indirect genetic effects on survival time in laying hens using repeated observations. 5th International Conference on Quantitative Genetics. 12.06.2016-17.06.2016. Madison, USA.
- Brinker, T., E.D. Ellen, and P. Bijma. Breeding value predictions for survival in layers showing cannibalism. 66th Annual meeting of the EAAP. 31.08.2015-04.09.2015. Warsaw, Poland.
- Brinker, T., E.D. Ellen, J. Visscher, and P. Bijma. Improving analysis of direct and indirect genetic effects on survival data in laying hens. 10th WCGALP. 17.08.2014-22.08.2014. Vancouver, Canada.

Other presentations

Brinker, T., “Did you ever keep birds? Chickens, perhaps?”, FAMELAB competition, 12.03.2017. Groningen, the Netherlands.

Brinker, T., E.D. Ellen, P. Bijma. Genome-wide association study for direct and indirect genetic effects on survival in crossbred laying hens. WIAS Science Day 2017. 06.02.2017. Wageningen, the Netherlands.

Brinker, T., E.D. Ellen, R. Veerkamp, P. Bijma. Predicting direct and indirect genetic effects on survival time in laying hens using repeated observations. WIAS Science Day 2016. 04.02.2016. Wageningen, the Netherlands.

Brinker, T., E.D. Ellen, R. Veerkamp, and P. Bijma. Predicting direct and indirect genetic effects on survival time in laying hens using repeated observations. Hendrix Genetic Academy. 19.05.2015-21.05.2015. Eindhoven, The Netherlands.

Media

Brinker, T. “Heb je ooit vogels gehouden? Kippen, misschien?”. 12.04.2017. Weblog Wageningen University.

Brinker, T. “Using DNA information to reduce feather pecking”. 19.12.2016. news article Wageningen University.

Training and supervision plan

Training and supervision plan



The Basic Package (3 ECTS)

WIAS Introduction Course	2013
Ethics and philosophy in life sciences	2014

Scientific Exposure (14 ECTS)

International conferences

EPGS - European Symposium on Poultry Genetics	2013
WCGALP - World congress on genetics applied to livestock production	2014
EAAP - European association of animal production	2015
ICQG - International conference of quantitative genetics	2016
GRC - Gordon Research Congress	2017

Seminars and workshops

WPSA seminar day	2013
WIAS science day	2014
NVG workshop, Societal communication and societal relevance	2014
F&G connectedagen	2014, 2017
KNAW - Genome Editing	2016

Presentations

WCGALP - World congress on genetics applied to livestock production	2014
EAAP - European association of animal production	2015
Hendrix Genetics Academy	2015
ICQG - International conference of quantitative genetics	2016
F&G connectiedagen	2017
GRC - Gordon Research Congress	2017

In-Depth Studies (7 ECTS)

Disciplinary and interdisciplinary courses

Social genetic effects: theory and genetic analysis	2013
Introduction to theory and implementation of genomic selection	2014
Societal impact	2015

Advanced statistics courses

Advanced statistical and genetical analysis of complex data using ASReml 4	2014
--	------

Training and supervision plan

Statistical models for genomic prediction in animals and plants 2015

PhD students' discussion groups

Quantitative genetics discussion group (QDG) 2013-2017

Professional Skills Support Courses (3 ECTS)

Data Management Planning Course 2014

Techniques for writing and presenting a scientific paper 2016

The Final Touch: Writing the General Introduction and Discussion 2016

Career Assessment 2017

Famelab cursus & pitch 2017

Research Skills Training (3 ECTS)

Introduction to R for statistical analysis 2013

External training period, Aarhus University 2015

Didactic Skills Training (10 ECTS)

Supervising practicals and excursion

Genetic improvement of livestock 2014

Supervising theses

MSc major 1 2015

MSc major 2 2015

BSc thesis 2015

MSc minor 2017

Management Skills Training (7 ECTS)

Organisation of seminars and courses

Organisation societal impact course 2015

Organisation seminar breeding against infectious diseases 2016

Membership of boards and committees

PhD representative WUR Council 2014-2015

Education and training total 46 ECTS

Acknowledgments

My PhD has almost come to an end. This is the final chapter I am writing before I can send it to the press. It has been an interesting ride, with ups and downs – both professionally and personally, and I could not have done it without the support of many. I have experienced the PhD trajectory as way more than conducting science and growing to become an independent researcher. It has also been a journey of self-discovery.

Esther, je was de eerste die mij ooit vroeg of ik interesse had om een PhD te gaan doen. Dit was tijdens mijn MSc thesis, waar ik met jou als begeleider werkte aan een thesis over sociale interacties. Door deze ervaring met jou, raakte ik geïnteresseerd in het doen van een PhD. Wat een mazzel dat daar een project kwam over sociale interacties! Precies in mijn straatje, omdat mijn interesse toch vooral ligt in dierenwelzijnsonderwerpen. Tijdens de PhD werd je wederom mijn begeleider. Ik ben je enorm dankbaar voor al je begeleiding, samenwerking en steun tijdens dit traject.

Piter, ik herinner mij nog goed dat je tijdens het vak “Genetic Improvement of Livestock” een presentatie kwam geven over sociale interacties. Op dit moment raakte ik in de ban van het onderwerp en wilde daarom persé een MSc thesis in die richting gaan doen. Eigenlijk heeft deze presentatie een grote invloed gehad op het vervolg van mijn carrière, waar ik enorm dankbaar voor ben. Samen met Esther heb je mij begeleid en ik had geen beter koppel kunnen indenken hiervoor, met Esther haar inzicht in de praktijk en jouw inzichten in de theorie. Bedankt! Daarnaast wil ik je bedanken voor mijn eerste MTB-tocht, waar jij me mee hebt genomen om te trainen voor de Alpe d’HuZes. En natuurlijk bedankt voor de vele MTB en wielrentochten die daarna volgden.

Johan, destijds was jij nog leerstoelgroep houder van ABGC. Ik wil je bedanken voor al jouw input en motivatie in de eerste helft van mijn PhD. Ook enorm bedankt dat ik deel mocht uitmaken van het Alpe d’HuZes team. Ik vond het een bijzonder avontuur.

Henk, je hebt Johan zijn rol als promotor in mijn project overgenomen. Ik wil je bedanken voor al jouw input en motivatie in de tweede helft van mijn PhD. En daarnaast, ook jij bedankt voor de fietstochten!

Acknowledgments

Thanks to the secretariat, with special thanks to Lisette. Lisette, bedankt voor alle zorgen. Ik waardeer enorm wat je afgelopen jaren allemaal voor mij geregeld hebt en dat je met mij meedacht. En natuurlijk voor de persoonlijke gesprekken.

Luc, bedankt voor de mogelijkheid om naar Aarhus University te komen. Ik heb het als een enorm leerzame periode ervaren. Ik vond het erg leuk om bij een andere universiteit in de keuken te kijken. Daarnaast voelde ik me meteen thuis in de groep. Ik zal altijd met een warm gevoel terug denken aan mijn periode in Denemarken. Mange tak! Also many thanks to my colleagues from Aarhus University. You contributed greatly to a wonderful time.

During my PhD, I used data from Hendrix Genetics. I want to thank you for providing the data and also for the fruitful discussions and input during our “chicken meetings”. With special thanks to Addie who also did a lot of preparatory work on the genomic data. Moreover, thanks to Hendrix Genetics for having me present in their office from time to time. In this way, I could experience company life and also directly ask questions about the data. It was very helpful.

To all my colleagues and friends, thank you for the welcome distractions and scientific discussions. Special thanks to my two paranympths, Hamed and Vahid. I would also like to give special thanks to Setegn. Unfortunately you could not make it to be my paranymp but I know you will be thinking of me. You are like family and I miss you my brother. I hope we meet soon.

I would, furthermore, like to thank Lauren Scheuer for allowing me to use her chicken drawings.

Als laatste wil ik graag mijn familie en vrienden bedanken voor hun steun en vertrouwen in mij. Met een speciale dank aan mijn ouders. Mama en papa, jullie betekenen de wereld voor mij. Jullie steun is van onschatbare waarde, love you!

Colophon

This research was part of a joint project of Hendrix Genetics and Wageningen University and Research on “Genomic solutions for socially-affected traits”. This project was financially supported by the Dutch Technology Foundation TTW, which is part of the Netherlands Organization for Scientific Research (NOW), and is also funded by the Ministry of Economic Affairs (grant number 12654).

The cover of this thesis was created by Tessa Brinker

This thesis was printed by Wageningen Academic Publishers, Wageningen, the Netherlands.

