Prediction of Heterosis in White Leghorn Crossbreds using Paternal 60K SNP Genotypes

E. N. Amuzu-Aweh[*,†], H. Bovenhuis[*], D. J. de Koning[†] and P. Bijma[*]

[*]Animal Breeding and Genomics Centre, Wageningen University, the Netherlands, [†]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Sweden

ABSTRACT: Predicting heterosis for crossbred offspring of individual sires would harness variation between sires from the same pure-line, and can increase the utilization of heterosis in crossbreeding schemes. We aimed to derive the theoretical expectation for heterosis in crossbred offspring of individual sires, and then predict heterosis for these offspring. We used 60K SNP-genotypes from 3427 sires, allele frequencies from 9 pure-lines, and egg number records from 16 crosses between those lines, representing ~210,000 crossbred hens. Results show that it is possible to predict heterosis at the sire level, thereby distinguishing between sires within the same pure-line whose offspring will show higher heterosis. However, variation in predicted heterosis between sires within a line was low in our data; most differences were seen between lines. A potential improvement of the method would be to base predictions on a subset of SNPs with identified effects on heterosis.

Keywords: heterosis prediction; laying hens

Introduction

Commercial breeding programs for laying hens make use of crossbreeding schemes to exploit the phenomenon known as heterosis. A major challenge for these breeding schemes is the expensive and time-consuming process of testing many different combinations of pure lines to find which ones produce crossbred offspring with maximum heterosis in the trait of interest. In theory, when heterosis is due to dominance, the heterosis at a single locus is directly proportional to the squared difference in allele frequency (SDAF) between the two parental pure-lines (Falconer and Mackay 1996). Using egg production records from White Leghorn crosses, Amuzu-Aweh et al. (2013) showed that the SDAF between parental pure-lines can predict heterosis at the line level with an accuracy of 0.5, allowing a 50% reduction on the number of pure-line combinations to be field- tested. This concept could be further extended by predicting heterosis for individual sires. This would harness the genetic variation between sires from the same pure-line, and can further increase the utilization of heterosis in crossbred breeding schemes. Our aims were to derive the theoretical expectation for the heterosis of crossbred offspring of individual sires and then predict heterosis for these offspring. We used 60K SNP genotypes from 3427 individual sires, allele frequencies from 9 pure-lines, and records on egg number from 16 crosses between those lines, representing ~210,000 crossbred hens.

Materials and Methods

Population structure. The crossbred hens originated from 9 purebred White Leghorn layer lines. Sire-lines were coded as S1, S4 and S5 and dam-lines were D1, D2, D3, D4, D5 and D6. D1 was also used as a sire-line. A cross produced by an S1 sire and a D1 dam is referred to as S1*D1. There were 1087 S1, 840 S4, 728 S5 and 772 D1 sires. Each of the sires was mated to only one dam-line, but to several dams, on which the specific mating details were not recorded. The mating design produced a total of 16 types of crossbreds. Crossbred hens were housed in cages of full- and half-sibs, with an average of 6 hens per cage. The number of cages per sire ranged from 1- 23, with an average of 11. Crossbred hens had been beak-trimmed.

Phenotypic data. Records on egg number were obtained from routine performance tests for a commercial breeding programme from 2005 through 2010. Egg number is a cage-based record of eggs produced from 100 through 504 days of age, calculated on hen-day basis. There were 34,799 cage-based records of egg number.

Genomic data. Two types of genomic data were used: individual 60K SNP-genotypes of 3427 sires, and allele frequencies of the 9 pure-lines in our data. To obtain allele frequencies for the dam lines, blood from 75 randomly selected males was pooled, and then DNA extracted for genotyping. For the lines which were used as sire lines (S1, S4, S5 and D1), we used the average of the individual sire genotypes as the line allele frequency. The Illumina chicken 60K SNP BeadChip (Groenen et al. 2011), was used for all genotyping. SNPs on the sex chromosome and SNPs that were missing for more than 5% of sires were excluded. The total number of SNPs used in this study was 53,421.

Statistical analyses

Theory. At the population level, heterosis due to dominance is proportional to the SDAF of the two parental lines that produce a crossbred. $\text{Heterosis}_{ij} = \sum_l d_l (p_{i,l} - p_{j,l})^2$,

where d_l is the dominance deviation at locus l, $p_{i,l}$ is the frequency of a particular allele at locus l in parental line i, and $p_{j,l}$ is the frequency of the same allele at locus l in parental line j (Falconer and Mackay 1996). Thus, when the phenotype of crossbred individuals is regressed on the mean SDAF between the two parental lines, the estimated partial regression coefficient is an estimator of the mean dominance deviation across all loci, $\hat{\beta} = \hat{\bar{d}}$. If heterosis is due to dominance, this result holds even when phenotypic data on the pure-lines is not available (Amuzu-Aweh et al. 2013).

To extend this concept to the sire level, we derived a term that is proportional to the expected heterosis due to dominance for crossbred offspring of a particular sire, *s*, from sire-line *i* that is mated to one or more dams from dam-line *j*.

In the following model:

$$y_{s_i j} = \text{sire-line}_i + \text{dam-line}_j + \beta x + e_{s_i j} \quad , \qquad [1]$$

$y_{s_i j}$ is a phenotypic record of an offspring of sire *s* from pure-line *i* mated to dams from pure-line *j*, and β is a regression coefficient. Following the notations for a 1-locus model, we derived that the *x* term in model [1] is (detailed derivation Amuzu-Aweh et al., in preparation):

$$x = (p_i - p_j)^2 + (p_{s_i} - p_i)(1 - 2p_j).$$

The first component, $(p_i - p_j)^2$, is the squared difference in allele frequency (SDAF) between the sire-line and dam-line. The second term measures how much the expected performance of the offspring of this sire deviates from the mean of the cross, due to dominance. This is a combination of the deviation of the sire allele frequency from its line allele frequency, $(p_{s_i} - p_i)$, and of the dam-line allele frequency, $(1 - 2p_j)$. The biggest contribution of the individual sire information is seen when the genetic distance between the sire-line and dam-line is small, and the sire diverges from the average of its sire-line. In general, the value of *x* increases as the genetic distance between the sire and dam-line increases. For a sire that has the same allele frequency as its sire-line, the $(p_{s_i} - p_i)$ term becomes zero, and *x* reduces to SDAF.

We calculated *x*, per SNP, between the 3427 sires in our dataset and all dam-lines that they had been mated to. Sire frequencies (p_s) were based on their genotypes and equal to 0, 0.5 or 1. Missing SNP genotypes for the sires were replaced by the sire-line allele frequency at that SNP. We then calculated $\bar{x}$ as:

$$\bar{x}_{s_i j} = \frac{\sum_{n=1}^{N}\left[(p_i - p_j)^2 + (p_s - p_i)(1 - 2p_j)\right]}{N},$$

where N was the total number of SNP loci (53,421).

Model: We predicted the heterosis per sire by fitting a linear mixed model where we regressed phenotypes of crossbreds on $\bar{x}$. We used the following statistical model:

$$y_{s_i jklm} = \mu + \text{sire-line}_i + \text{dam-line}_j + \beta \cdot \bar{x}_{s_i j} + \text{test}_k + \text{hen density}_{l:k} + HRT_m + e_{s_i jklm} \quad ,$$

where $y_{s_i jklm}$ was the average egg record of hens in cage *k*, offspring of sire s_i; *sire-line*$_i$ and *dam-line*$_j$ were the fixed effects of the i^{th} sire-line and j^{th} dam-line of each cross (*i*,=1 - 4, *j* = 1 - 7), β was the regression coefficient of *y* on $\bar{x}$, *test*$_k$ was the fixed effect of each performance test (k=1 – 33 year-farm classes). *hen density*$_l$ was a fixed effect accounting for the initial number of hens within a cage. It had 128 levels, and was nested within *test* because the physical size of cages differed across some performance tests. The combined effect of the *h*en-house, *r*ow and *t*ier of the cage was accounted for by including the term "HRT_m" as a random effect (m = 1 - 767) and $e_{s_i jklm}$ was the random residual error term. Data were analysed using the MIXED procedure in SAS version 9.2.

For the crossbred offspring of each sire, heterosis was calculated by multiplying the estimated regression coefficient, $\hat{\beta}$, by the $\bar{x}$ value between the sire and the dam-line:

$$\text{Estimated heterosis}_{s_i j} = \hat{\beta} \cdot \bar{x}_{s_i j} \ .$$

Next, to determine the relative importance of adding individual sire genotypes to predict heterosis at the sire level versus estimating heterosis only at the population level, we calculated the proportion of variance in *x* that is contributed by SDAF versus that which is contributed by $(p_{s_i} - p_i)(1 - 2p_j)$.

Results and Discussion

Descriptive statistics: Cage-based egg numbers ranged from 163.9 - 375.3. The S5*D3 cross had the highest mean of 345.2 eggs, whilst the S4*D6 had the lowest mean egg number of 326.1.

For the genomic data, values of $\bar{x}$ ranged from 0.08 to 0.18, with an average of 0.12 and a standard deviation of 0.018. Values of $\bar{x}$ increased as the genetic distance between the sire and the dam-line increased (note that this is not the same as the genetic distance between the sire-line and dam-line).

Predicted heterosis per sire: The estimated regression coefficient of egg number on $\bar{x}$ was $\hat{\beta}$ = 93.5 (s.e = 18.3) ($p < 0.0001$), showing a positive and highly significant association between $\bar{x}$ and heterosis in egg number. This implies that sires that diverge more from the dam-line to which they are mated produce offspring with higher levels of heterosis for egg number. This is in agreement with a study by Haberfeld et al. (1996), who estimated correlations between heterosis and genetic distance between mating-pairs and concluded that offspring were superior when they were from parents with a relatively distant genetic relationship.

Figure 1 shows heterosis for egg number for the 3427 sires in our study. Between lines, the lowest predicted heterosis for egg number was 7.6 eggs for an S5 sire mated to the D6 dam-line, and the highest was 16.7 eggs for an S4 sire mated to the D1 dam-line. The biggest within-line difference was 1 egg, seen among the 318 S1 sires that were mated to the D2 dam-line.

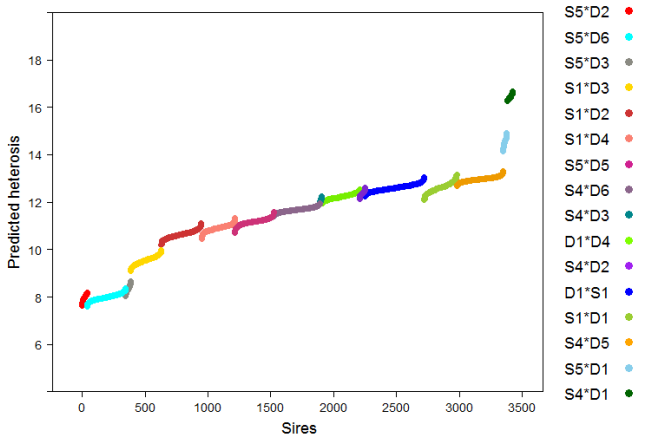

Figure 1. Estimated heterosis in egg number for the 3427 sires in this study. On the x axis, the sires are numbered from 1 to 3427. The y-axis shows estimated heterosis. Each point on the graph represents the average heterosis in the offspring of a particular sire. Each sire was mated to one dam-line, but to several dams. The colours represent the 16 crosses in this study. The key is ordered from the cross with the lowest to the highest predicted heterosis.

Proportion of heterosis explained by the within-line sire difference: Next we quantified the added value of individual sire genotypes to the estimation of heterosis. We did this by comparing variances. The first component of *x* is the squared difference in allele frequency (SDAF) between the two parental pure-lines, and represents the between-line variance in allele frequency. The second component of *x* represents the within-line variance (in allele frequency) of the sires and a term related to the dam-line each sire is mated to. The variance of the between-line component, $\sum(p_i - p_j)^2$, was 3.08E-4 and that of the within-line component, $\sum(p_{s_i} - p_i)(1-2p_j)$, was 2.23E-6. The proportion of variance in $\overline{x}$ that was explained by the between-line component was 99%, and that explained by the within-line component was 0.72%. This shows that the extra genomic information from individual sires only explained a small proportion of the variance in $\overline{x}$, implying that the differences between sires for predicted heterosis is small as compared to differences between crosses. This could be because the genetic variation between sires within a line was indeed low, or that the genetic variation between sires was reduced because we averaged *x* across the genome.

The use of an average *x* across the entire genome follows the assumption that all SNPs have an equal effect on heterosis in egg number. In a preliminary analysis where we estimated $\overline{x}$ using only a subset of 12,097 SNPs with a putative effect on heterosis, we observed a bigger range and variance of $\overline{x}$ and some re-ranking of individual sires as well as among crosses.

Conclusion

We conclude that based on a dominance model, it is possible to estimate heterosis using allele frequency differences between individual sire-genotypes and the dam-lines that they are mated to. This indicates that within one pure-line, one can identify specific sires whose offspring are expected to have relatively higher levels of heterosis than others. In our data however, the within-line differences between sires was relatively low compared to the between-line differences. A potential improvement of the method would be to base predictions on a subset of SNPs with identified effects on heterosis.

Acknowledgements

The first author benefited from a joint grant from the European Commission and ISA- Hendrix Genetics within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG".

References

Amuzu-Aweh E. N., Bijma P., Kinghorn B. P. et al. (2013). Heredity. 111: 530–8

Falconer, D. S. and Mackay T. F. C. (1996). Introduction to quantitative genetics, 4th edn. Longman Harlow.

Groenen M. A., Megens H. J., Zare Y., et al. (2011). BMC Genomics 12: 274.

Haberfeld A., Dunnington E. A., Siegel P. B. et al. (1996). Poult Sci 75: 951-953.