

Genomic prediction of heterosis *in White Leghorn crosses*

E. Nancy Amuzu-Aweh, P. Bijma, H. Bovenhuis & D.J. de Koning



- Heterosis is essential in crossbreeding schemes
- Maximize heterosis → which cross to make?
- Can we predict heterosis?
 - Long history with inconsistent results
 - Based on limited number of markers/small data sets

Aim

Develop a predictor of heterosis:

1. At the line level
2. At individual sire level
3. Based on genomic regions affecting the trait

Assumption: Heterosis is solely due to dominance

1. Line-level heterosis

Heterosis is proportional to the **S**quared **D**ifference in **A**llele **F**requency (SDAF) between the parental lines

$$Heterosis_{ij,l} = d_l (p_{i,l} - p_{j,l})^2$$

d : dominance deviation

$p_{i,l}$: allele frequency at locus l in parental line i

$p_{j,l}$: allele frequency at locus l in parental line j

(e.g. Falconer & Mackay, 1996)

Genome-wide SDAF: averaged over all SNP loci

Method

- Predict heterosis

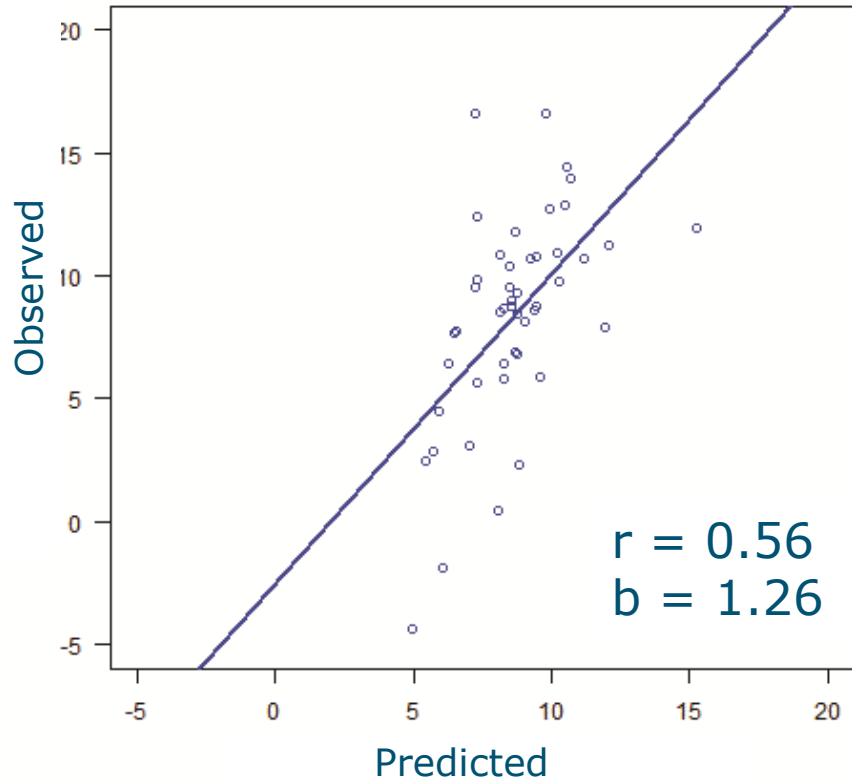
- Regress crossbred phenotypes on SDAF $\rightarrow \hat{\beta}$
- Predict future crosses: $\hat{\beta} \times SDAF$

- Accuracy: correlation observed and predicted heterosis

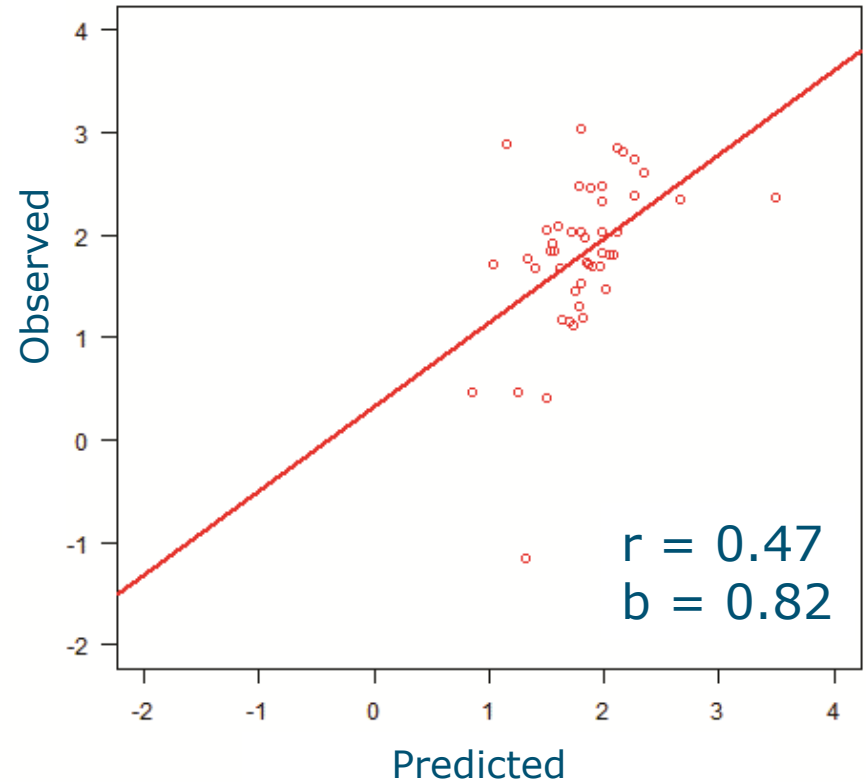
- 47 crosses
- Leave-one-out cross-validation

Results for egg number and egg weight

Egg number



Egg weight



Benefits of predicting line-level heterosis

Preselect crosses before field testing

Two-stage selection

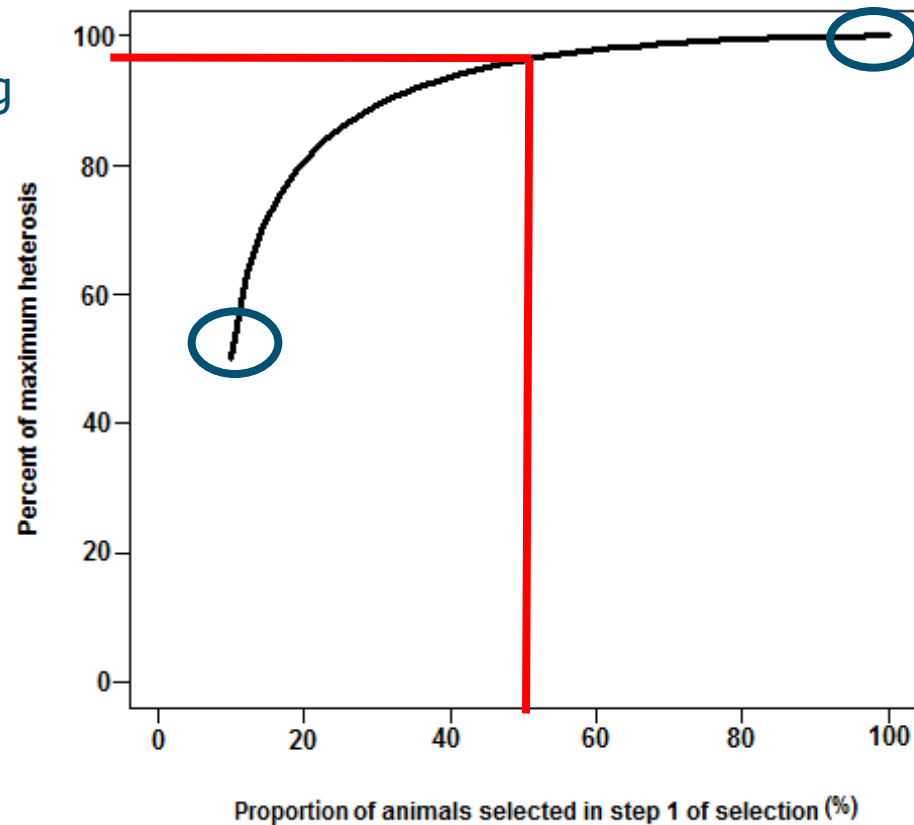
- Stage 1: predicted heterosis
- Stage 2: field testing

Final selected prop = 10%

$$r_{\text{pred heterosis}} = 0.5$$

Pre-select 50% of crosses

→ 96% of total heterosis



Conclusions: line-level heterosis

- Genome-wide SDAF predicts heterosis in egg traits with an accuracy of ~ 0.5
- Implement 2-step selection and save up to 50% on field tests

2. Sire-level heterosis

Can we exploit heterosis at the sire level?

- Utilise the variation between sires from the same pure-line
 - Allocate sires to dam lines

Sire-level heterosis

- Heterosis depends on heterozygosity excess due to between- and within-line differences in allele frequencies

$$Heterosis_{ij,l} = \left[(p_i - p_j)^2 + (p_{s_i} - p_i)(1 - 2p_j) \right] \cdot d_l$$

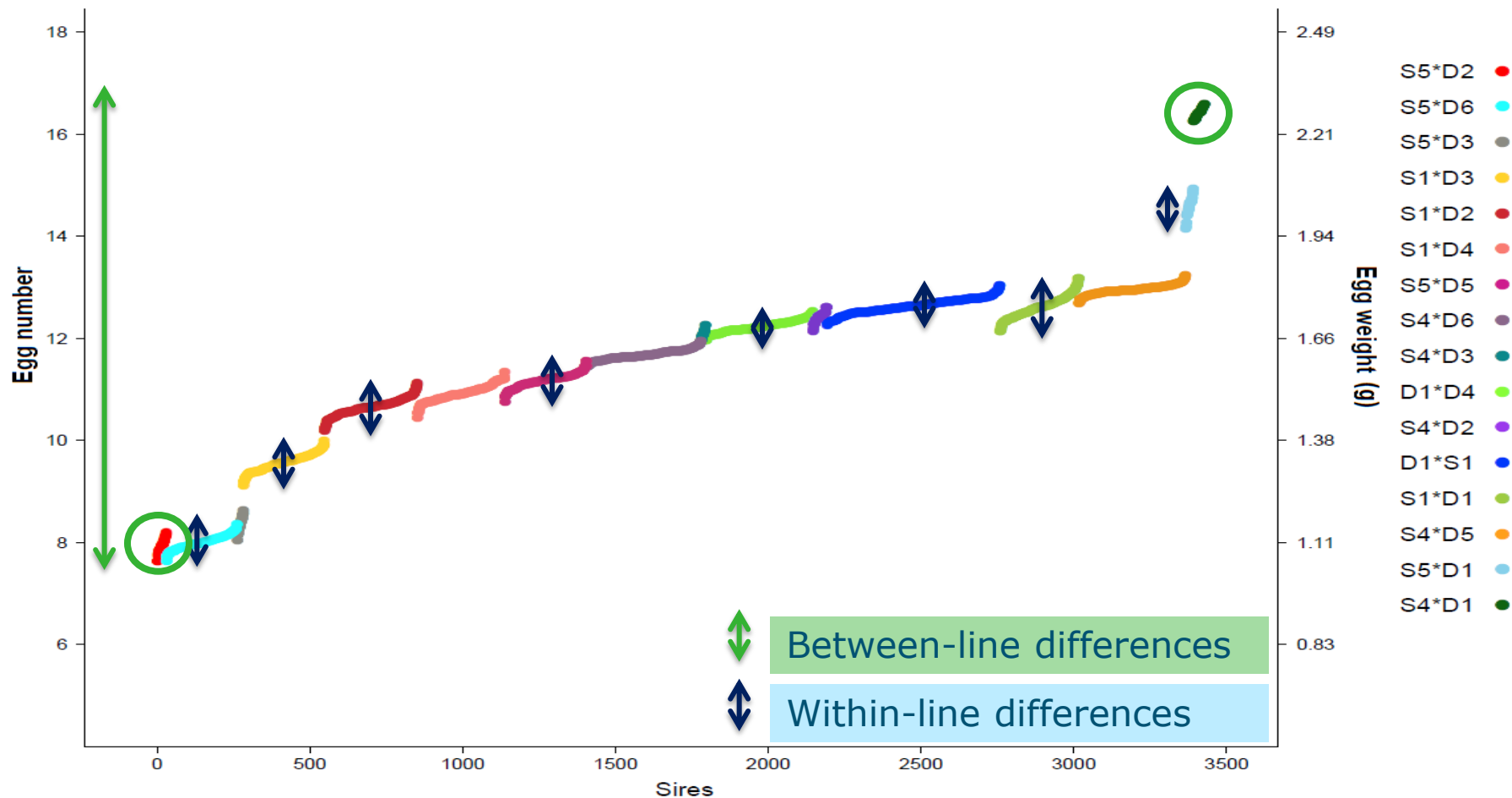
p_i = allele frequency in sire line , p_{s_i} = allele frequency in the sire

p_j = allele frequency in dam line

- Genome-wide average

Results

Predicted heterosis for egg number and egg weight



Conclusions: sire-level heterosis

- Heterosis can be predicted at the sire level
- But within-line sire differences contribute very little

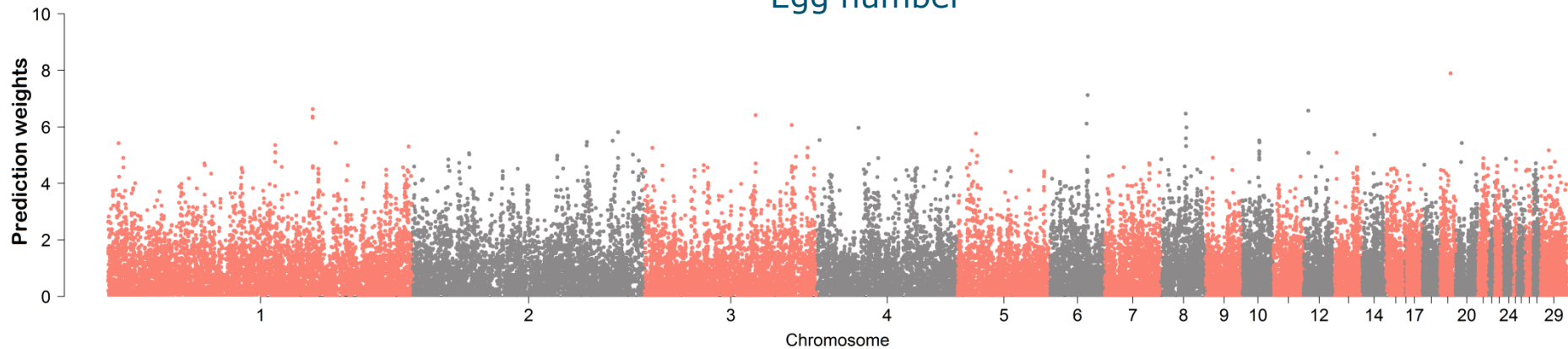
3. Heterosis focussing on genomic regions

- GWAS to identify SNPs with effects on egg number and weight
 - Average effect
- Weighted SNPs by average effect (α) and $SE(\alpha)$

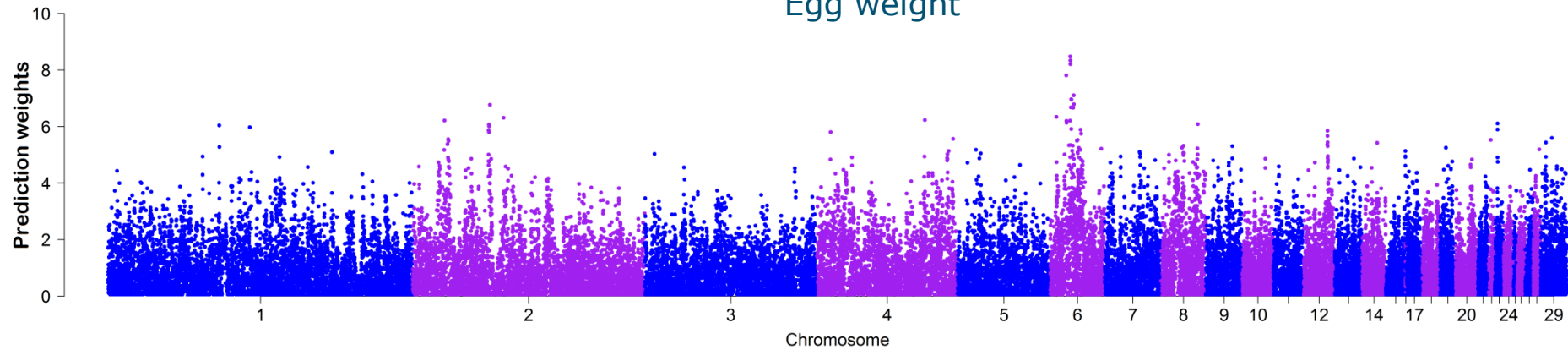
$$\text{Weighted SDAF}_{ijk} = \text{SDAF}_{ijk} * |\hat{\alpha}_k| * \frac{1}{\text{se}_k^2}$$

Resulting Weights

Egg number



Egg weight



Results: weighting genomic regions

- SDAF and $\text{SDAF}_{\text{weighted}}$ have a correlation of 0.99
- Little benefit from using $\text{SDAF}_{\text{weighted}}$ based on average effects

Conclusions

😊 Line-level prediction based on a genome-wide average SDAF can save 50% on field-testing

☹️ Prediction at the sire level or based on genomic regions not beneficial

esinam.amuzu@wur.nl



Thank you!

1. Line-level heterosis: data and models

Genotypes: Illumina chicken 60K SNP chip → 5 sirelines, 6 damlines

Phenotypes:

- 47 crosses
- ~76 000 cage-based records

Model: $y_{ijk} = \mu + \text{sireline}_i + \text{damline}_j + \beta \cdot \overline{SDAF}_{ij} + \dots + e_{ikj}$

2. Sire-level heterosis: Data and models

Genotypes: Illumina chicken 60K SNP chip →
3427 sires (4 lines), 6 damlines

Phenotypes:

- 16 crosses
- ~35 000 cage-based records

Model: $y_{s_i j k} = \mu + \text{sireline}_i + \text{damline}_j + \beta \cdot \bar{x}_{s_i j} + \dots + e_{s_i j k}$

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

3. Heterosis considering genomic regions: data and models

Genotypes: Illumina chicken 60K SNP chip →
3427 sires (4 lines), 6 damlines

Phenotypes: 16 crosses, ~35 000 cage-based records

GWAS:

$$y_{s_i j k} = \mu + \text{sireline}_i + \text{damline}_j + \text{cross}_{ij} + \alpha \cdot (p_{s_i} + p_j - 1) + \dots + e_{s_i j k}$$

$\hat{\alpha}$ = average effect per SNP



3. Heterosis considering genomic regions: data and models

Heterosis prediction:

$$y_{ijk} = \mu + \text{sireline}_i + \text{damline}_j + \beta \cdot \overline{\text{SDAF}}_{wt,ij} + \dots + e_{ikj}$$

where $\text{SDAF}_{wt,ij} =$

$$\frac{\sum_1^N (\text{SDAF}_{ijk} * \left(\frac{|\alpha_k| * \frac{1}{\text{se}_k^2}}{\sum_1^N \frac{|\alpha_k| * \frac{1}{\text{se}_k^2}}{N}} \right))}{N}$$