Metabolomic phenotypic prediction of growth in pigs

Pernille Sarup
J. Jensen, T. Ostersen, O.F. Christensen, P. Sørensen
How to increase EBV accuracies by 20-50%

Pernille Sarup
J. Jensen, T. Ostersen, O.F. Christensen, P. Sørensen
Background

Accurate EBVs need lots of phenotypic measurements

But

Some phenotypes are expensive or difficult to measure

Current solution:

Breeding values using information on correlated phenotypes and additive genetic relationships with tested animals
Background

New option

Use large scale molecular phenotypes (e.g. H\(^1\) NMR metabolomics) for prediction of relevant phenotypes in large cohorts.
Background

NMR spectra are affected by both the genotype and the environment, but only the genetic component is useful for prediction of breeding values.
Aim

Can we use these NMR predicted phenotypes to increase accuracy of EBVs?
Data

2320 finishing Duroc x(Yorkshire x Landrace)
Genotypes (6K)
NMR – metabolomics on blood serum
Total Feed Intake
Average Daily Gain
Back Fat
Adjusted phenotypes

Growth phenotypes were adjusted for fixed effects of stable-year-month and sex

\[ y = X_{sec} \cdot sec + X_{sex} \cdot sex + e \]

\[ y_{adj} = \hat{e} \]
Models

**GBLUP**

\[ y = Z_l l + Z_g g + e \]

\[ g \sim N(0, G \sigma_g^2) \]

\[ G = (W W' / n) \]
Models

MBLUP

\[ y = Z_l l + Z_m m + e \]

\[ m \sim N(0, M\sigma_m^2) \]

\[ M = (XX'/n) \]

\[ X = \text{scaled and centered NMR intensities} \]
Cross validation schemes

Prediction across environments and family:
   Oldest half of stable-year-month in the training population (across)

Prediction within environments and family:
   Alternate stable-year-month in the training population (within)
Cross validation schemes

GBLUP

\[ y = Z_l l + Z_g g + e \]

Cross-validated:

Univariate \( \hat{g}_{cv} \)
Cross validation schemes

**MBLUP**

\[ y = Z_l l + Z_m m + e \]

\[ \hat{y}_m = \hat{m} \]
Cross validation schemes

Bivariate GBLUP

\[ y = Z_l l + Z_g g + e \]
\[ \hat{y}_m = Z_l l + Z_g g + e \]

Cross-validated:

Bivariate \( \hat{g}_{cv} \)
Proxies for "True" genetic values

The adjusted observations themselves:

\[ y \]

accuracy = \( \text{cor}(y, \hat{g}_{\text{cv}})/h \)

Bivariate GBLUP with all information:

Bivariate \( \hat{g} \)

accuracy = \( \text{cor}(\text{Bivariate } \hat{g}, \hat{g}_{\text{cv}}) \)}
Accuracy ADG

ADG across

- Univariate $\hat{\theta}_{cv}$
- Bivariate $\hat{\theta}_{cv}$

ADG within

- Univariate $\hat{\theta}_{cv}$
- Bivariate $\hat{\theta}_{cv}$
Accuracy

BF across

BF within

TFI across

TFI within
Conclusions

20-50% increased accuracy of EBVs for all three traits with added information from NMR metabolomics.

Increased accuracy of EBVs for all three traits when predicting within environment and family (within) – Benefit from NMR information remains!

Questions?
Only NMR data in training population

Bivariate GBLUP

\[ y_{adj} = Z_l l + Z_g g + e \]
\[ \hat{y}_m = Z_l l + Z_g g + e \]

Cross-validated:

\[ B_{\hat{g}_{adj, cv, cv}} \]