First Results on Genome-wide Genetic Evaluation in Swiss Dairy Cattle

C. Stricker[1], J. Moll[2], H. Jörg[3], D.J. Garrick[4] and R.L. Fernando[4]

[1]applied genetics network, Davos, Switzerland; [2]Association of Swiss Cattle Breeders, Zollikofen, Switzerland; [3]Qualitas AG, Zug, Switzerland, [4]Iowa State University, Ames, IA, U.S.A.

Introduction

The Association of Swiss Cattle Breeders has initiated an applied research project to implement genomwide evaluation and genomic selection in the current breeding schemes of Swiss dairy cattle. A brief schedule of the project is below.

Fall 2007: Evaluation of available SNP chip technology of Affymetrix and Illumina. Genotyping of 48 test animals with both Chip technologies and comparing results. Decision for Illumina's Bovine 50k SNP chip.

2008: Simulation study comparing Bayes A, B (Meuwissen et al. 2001) and random-regression BLUP (RRBLUP) for varying effective populations size, linkage disequilibrium and number of SNPs.

Winter 2008/09: Genotyping of 1129 Brown Swiss sires

Spring 2009: Genotyping of 1579 Holstein an Red & White sires

Spring 2009: Analysis of Brown Swiss data (this report)

Summer 2009: Analysis of Holstein and Red & White data

This report considers the analysis of Brown Swiss sires for the trait milk yield only.

Phenotypes

- Data from Swiss Brown Cattle only: 1129 sires, performance tested
- Trait: de-regressed EBVs for milk yield
  - EBV 1^{st} lactation
  - average EBV lactation 1-3
- average reliability of EBVs for milk yield, lactation 1-3 is 0.92

Genotypes

- Illumina Bovine 50K SNP chip: 54001 SNPs
- Exclusion of SNPs with minor allele frequency below 0.025
- Exclusion of SNPs with all heteroygous genotypes
- 42'722 SNPs remaining for analysis
- missing SNPs imputed as average genotype of all 1129 sires

Training and Validation Data

- validation data 1 (maxRelTest): calculating additive genetic relationship coefficients r_{ij} among all sires $i, j = 1, ..., n$. 112 sires with minimum reliability 0.95 and largest average relationsship coefficients $r_{i.}$ assigned to validation data. Average reliability of 0.98 in validation data. 1017 sires remaining for training. Average relationship coefficient $r_{i.}$ in training data 0.025, $r_{i.}$ in validation data 0.068.
- validation data 2 (youngSires): 117 sires from the last two generations. No minimum reliability required, average reliability of 0.88. 1012 sires remaining for training. Average $r_{i.}$ in training data 0.046, average $r_{i.}$ in validation data 0.058.

Models

The general model to analyze the training data is

$$y_i = \sum_j (x_{ij} - \bar{x}_j)\alpha_j\delta_j + e_j$$

where

y_i is the de-regressed EBV of sire i ,

x_{ij} is the genotype of sire i at SNP j,

$\bar{x}_j$ is the mean genotype at SNP j,

α_j is the random effect of SNP j,

δ_j is 0/1- indicator variable, and

e_j is a random residual.

If $\delta_j = 1$, i.e. when SNP j is included in the model, the prior distribution for a_j was assumed to be $N(0, \sigma^2_{\alpha_j})$ and the prior distribution for $\sigma^2_{\alpha_j}$ was scale-inverse chi-square with 4 degrees of freedom. For the error variance σ^2_e, the prior was assumed to be scale-inverse chi-square with 10 degrees of freedom. Markov Chain Monte Carlo was used to sample parameters a_j and σ^2_e via Gibbs Sampling, whereas a Metropolis-Hastings sampler was used to sample $\sigma^2_{\alpha_j}$. The proportion π of SNPs not included in the model was pre-determined as 0, 0.5, 0.75, 0.9 or 0.95. $\pi = 0$ corresponds to a method referred to as Bayes A by Meuwissen et al. (2001). Assuming all a_j to originate from a single distribution with σ^2_α instead of modelling separate distributions for each of the SNP-effects is referred to as Bayes C, whereas all other methods are called Bayes B as in Meuwissen et al. (2001). Setting $\pi = 0$ and assuming a single common variance σ^2_α is random-regression BLUP. 25'000 MCMC samples were taken with 1'000 samples discarded as burn-in. For the animals in the validation data, correlation between the naional EBV for milk yield using a random-regresson test-day model and the genomic EBV for milk yield was calculated.

Results

Note that gEBV refers to the genomic breeding value, whereas EBV is the breeding value from the national evaluation. π as explained above is the a-priori proportion of SNPs not included in the model and h^2 is the ratio of the variance explained by the SNPs to the total variance. As a main result, $corr(gEBV, EBV)$ gives the correlation between genomic and 'traditional' breeding values from the national evaluation.

De-regressed EBVs 1^{st} lactation

$\sigma^2_{\alpha_j}$	validation data	π	corr(gEBV,EBV)	h^2
$\sigma^2_{\alpha_j} = \sigma^2_\alpha$ for all j	maxRelTest	0	0.65	0.88
		0.5		
		0.75		
		0.9		
		0.95		
	youngSires	0	0.57	0.86
		0.5		
		0.75		
		0.9		
		0.95		
		0.99	0.55	0.82
		0.999	0.48	0.67
		0.9999	0.41	0.52
$\sigma^2_{\alpha_j}$ for each j	maxRelTest	0	0.65	0.94
		0.5		
		0.75		
		0.9		
		0.95		
	youngSires	0	0.57	0.91
		0.5		
		0.75		
		0.9		
		0.95		
		0.99	0.53	0.82
		0.999	0.47	0.67
		0.9999	0.40	0.52

De-regressed average EBVs lactation 1-3

$\sigma^2_{\alpha_j}$	validation data	π	corr(gEBV,EBV)	h^2
$\sigma^2_{\alpha_j} = \sigma^2_\alpha$ for all j	maxRelTest	0	0.72	0.84
		0.5		
		0.75		
		0.9		
		0.95		
		0.99		
		0.999	0.64	0.7
		0.9999	0.59	0.57
	youngSires	0	0.62	0.92
		0.5		
		0.75		
		0.9		
		0.95		
		0.99	0.60	0.85
		0.999	0.54	0.72
		0.9999	0.46	0.59
$\sigma^2_{\alpha_j}$ for each j	maxRelTest	0	0.71	0.89
		0.5		
		0.75		
		0.9		
		0.95	0.72	0.9
		0.99	0.73	0.83
		0.999	0.62	0.71
		0.9999	0.59	0.57
	youngSires	0	0.62	0.94
		0.5		
		0.75		
		0.9		
		0.95	0.61	0.9
		0.99	0.60	0.86
		0.999	0.54	0.70
		0.9999	0.46	0.57

Conclusions

- Correlation of 0.72 between gEBVs and 'traditional' BLUP EBVs from the national evaluation for milk yield lactations 1-3 and 0.65 for 1^{st} lactation, when around 1000 genotypes were used to estimate SNP effects. These results were obtained when sires with high average relationship to all other sires were separated into the validation data. When the two last generation of performance tested sires were selected as validation data, correlations dropped by about 8-10%. This is mostly due to the fact, that the average reliability of those young sires is considerably lower (0.88 vs. 0.98).
- Correlation between gEBVs and 'traditional' BLUP EBVs is not improved when each SNP effects is assumend to originate from a separate distribution with variance $\sigma^2_{\alpha_j}$. Fitting a common variance σ^2_α for all SNP effects α_j is sufficient.
- Results are not sensitve to π, the a-priori proportion of SNPs not included in the model. Correlations between gEBVs and 'traditional' BLUP EBVs drop only for extremely high values of π.

References

Meuwissen, T.H.E., Hayes, B.J. and M. E. Goddard (2001): Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157, 1819-1829.