The use of sequence SNP in a marker model for an across-breed dairy cattle genomic evaluation

B.L. Harris, R.G. Sherlock, C. Couldrey, M.D. Keehan & K.M. Tiplady

Livestock Improvement Corporation,
Private Bag 3016, Hamilton,
New Zealand

Introduction

- Access to whole-genome sequence data increasing common
- In theory, the sequence data should contain causative mutations associated with most of the genetic variation observed in phenotypic traits
- In theory, the use of sequence data is expected to improve multi-breed genomic evaluation

Introduction

- The aim of this study was to compare the genomic prediction accuracies and validation biases from:
  - subsets of sequence SNP
  - 50k SNP chip
  - New Zealand national across-breed genomic evaluation using a single-step marker effects model

Sequence SNP Selection

- 19.5M SNP >>> biological significance, quality of variant calling and LD trimming >>> 1.6M SNP
- Bayes RC models (70k animals training population)
  - GWAS by chromosomes >>> 200K SNP with non-zero posterior variance per trait
    - Successive genome wide GWAS removing SNP with lowest posterior variance
      - 200k >>> 100k >>> 50K >>> …

Data

- Two Validation Datasets – for genomic breeding values
  - 24.3M animals
    - 1.6M live weight records
    - 11.8M first-lactation protein records
  - Excluded phenotypic records following the completion of the 2013 season and animals born after completion of 2011 season

Genomic Breeding Values

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_g\mathbf{Mm}_g + \mathbf{Z}_n\mathbf{u}_n + \mathbf{Za} + \mathbf{e}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_g & \mathbf{Z}_n \end{bmatrix}$$

$\mathbf{b}$ are the fixed effects

$\mathbf{m}_g$ are the SNP marker effects

$\mathbf{u}_n$ are the marker breeding values for the non-genotyped individuals

$\mathbf{a}$ are the additive polygenic effects

$\mathbf{e}$ is the random residual

- Compared sequence SNP to 50k SNP chip
- Scenario 1: Polygenic variance was set to 20% of total genetic variance
- Scenario 2: Polygenic variance was varied from 10 to 90%

Validation

- Used daughter yield deviations for progeny test sires with daughter phenotypes in the 2014-2016 seasons
- Validation of the genomic breeding values was undertaken using the Interbull validation procedures
- 3 crops of sires
  582 in total: 140 KiwiCross, 282 Holstein Friesian and 160 Jersey

Results

- Live weight – polygenic variance = 20%

Data	Accuracy			Bias		
	HFxJ	HF	J	HFxJ	HF	J
PA	0.593	0.525	0.434	1.08	0.91	0.94
SSNP[1] 23779	0.801	**0.758**	**0.682**	**1.09**	**0.93**	**0.81**
SSNP[1] 11889	0.800	**0.758**	**0.682**	1.07	0.90	0.78
SSNP[1] 743	0.771	0.694	0.584	1.09	0.93	0.74
50k SNP	**0.844**	0.756	0.680	1.23	1.00	0.99

Results

- Protein Yield – polygenic variance = 20%

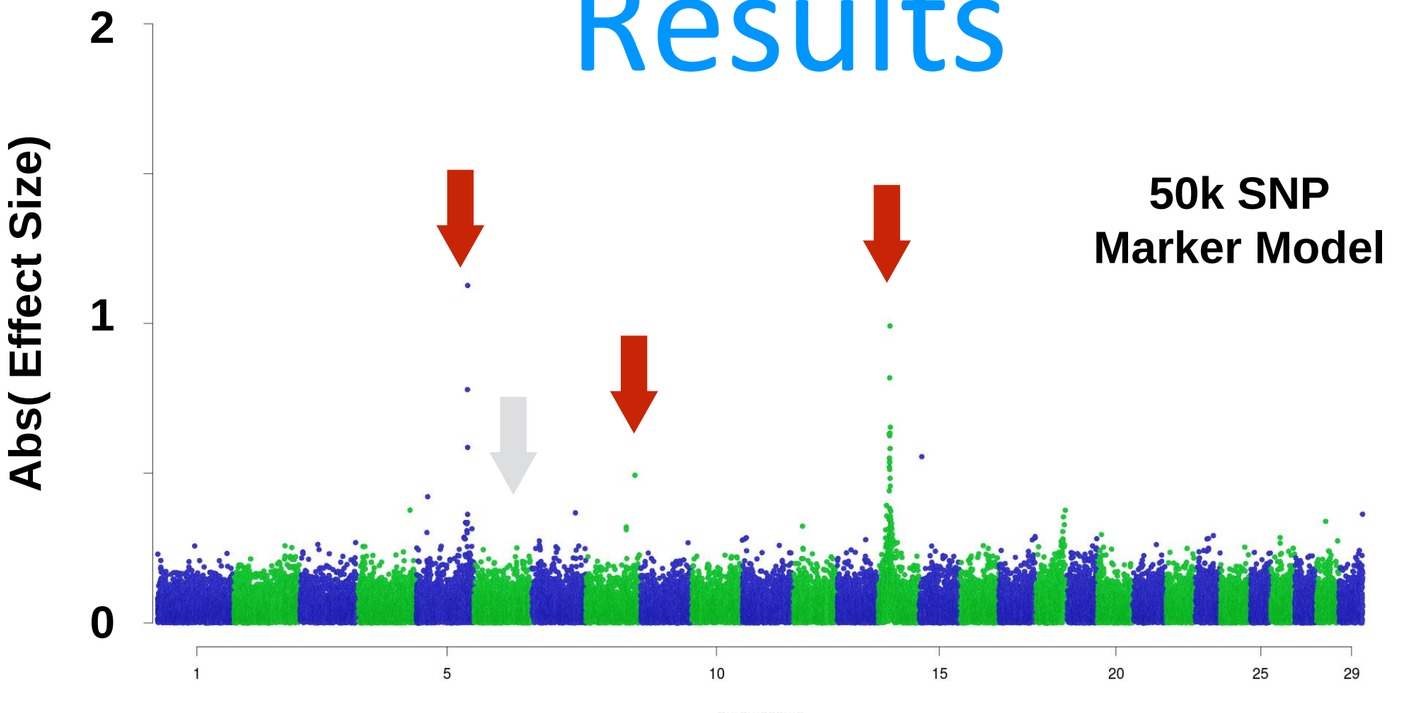
Data	Accuracy			Bias		
	HFxJ	HF	J	HFxJ	HF	J
PA	0.559	0.230	0.300	1.20	0.61	0.80
SSNP[1] 22023	**0.711**	0.602	**0.564**	0.90	0.73	0.66
SSNP[1] 11011	0.691	**0.606**	0.524	0.88	0.71	0.62
SSNP[1] 668	0.663	0.466	0.473	0.98	0.69	0.67
50k SNP	0.677	0.506	0.533	**1.08**	**0.94**	**0.91**

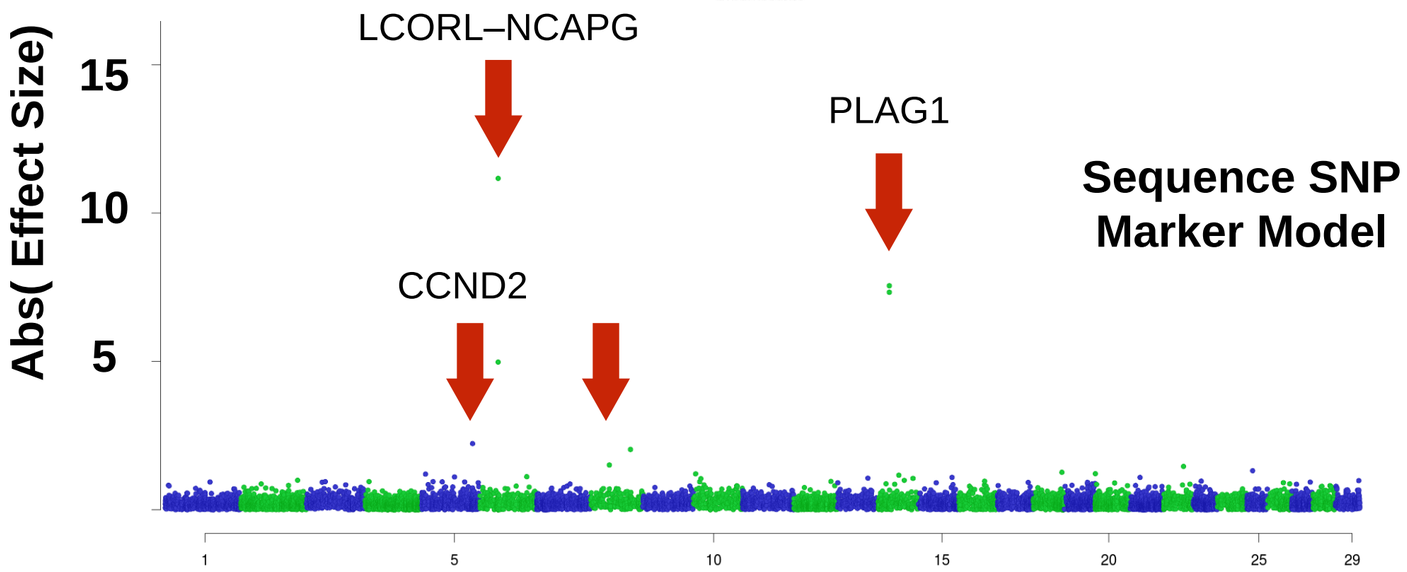
Results

- Protein Yield – vary the polygenic variance

	Accuracy			Bias		
Proportion[1]	HFxJ	HF	J	HFxJ	HF	J
0.2	0.736	0.608	0.528	1.11	0.96	0.82
0.3	0.731	0.615	0.541	1.05	0.89	0.76
0.4	0.724	0.615	0.542	1.00	0.84	0.72
0.5	0.717	0.614	0.547	0.96	0.80	0.70
0.6	0.709	0.612	0.537	0.93	0.76	0.67
0.7	0.701	0.609	0.532	0.91	0.73	0.64
0.8	0.691	0.606	0.524	0.88	0.70	0.62
0.9	0.679	0.602	0.514	0.86	0.68	0.60

Results


Abs(Effect Size)
2
1
0
1
5
10
15
20
25
29
50k SNP
Marker Model


Abs(Effect Size)
15
10
5
LCORL–NCAPG
PLAG1
CCND2
1
5
10
15
20
25
29
Sequence SNP
Marker Model

Conclusions

- The results of this study are in line with several studies that reported little or no increase the accuracy of genomic evaluation when using sequence data compared to SNP chip data

- The increase in bias when using the sequence SNP compared to the 50K SNP chip genotypes was also reported previously

Conclusions

- The sequence data used in this analysis was based on imputed sequence SNP values which were derived from several hundred of sequenced individuals
- Phasing and imputation errors are introduced in the imputation process could impact on the validation results
- Imputation errors in the validation animals could lead to reduced genomic prediction accuracy estimates

Conclusions

- Breed-by-QTL interactions may be present in real data, or causative variants may be present in only one breed, or have effects that are in different directions in different breeds

- Exploring the use of breed of origin haplotypes may enable improved genomic prediction accuracy if breed-by-QTL interactions are present in our multi-breed population

Post-Doc Position available at Livestock Improvement NZ in genomic evaluation

See the App
In the positions tab for more details