Use of GWAS results for a custom LD chip to exploit sequence data for prediction

Rasmus Froberg Brøndum, Guosheng Su, Luc Janss, Goutam Sahana, Bernt Guldbrandtsen, Didier Boichard, Mogens Sandø Lund







Background

- Customizable Illumina Bovine-LD SNP-chip
- Add QTL markers from sequence data
- LD-chip used for large scale genotyping of cows
- Markers can be accurately imputed to bulls
- Higher accuracy of genomic prediction from adding QTL markers?



QTL marker selection

Breedwise GWAS in Nordic Holstein, Jersey and Red

CENTER FOR QUANTITATIVE

GENETICS AND GENOMICS

- Index traits categorized, Top QTLs selected
- 3-5 markers per QTL/ trait
- Selected on e.g. P-value, functional annotation
- Breedwise LD pruning: (pairwise $R^2 > 0.95$)
- **Duplicates removed**
- Checked for assay performance (Illumina assay tool)
- ~1,500 markers in total

Category-I (15 QTL)	Category-II (10 QTL)	Category-III (5 QTL)	
Mastitis	Longevity	Nordic Total Merit	
Fertility	Other-diseases	Growth	
Leg	Birth	Udder	
Milk	Calving	Body-conformation	
Fat		Milking-speed	
Protein		Yield	
		Temperament	



Validation data

Breed	Cut-off	Reference (#)	Validation (#)	50k markers
Holstein	2005/01/01	3,953	1,046	~42k
Nordic Red	2005/01/01	3,605	874	~43k
French Holstein	2005/01/01	3,804	1369	~42k



- 1,443 QTL markers extracted from imputed sequence data (1000 bull genomes run4 reference data)
- De-regressed proofs



Methods

Prediction done using either:

- 1. 50k data
- 2. 50k data pooled with QTL markers
- 3. Separate variance components for 50k and QTL data

Milk Yield, Fat Yield, Protein Yield, Fertility and Mastitis

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS

GGG.

Reliability of prediction:
$$\frac{Cor^2(DGV, DRP)}{R_{DRP}^2}$$



GBLUP

> Model used for 50k and pooled data: $y = 1\mu + Za + e$

Model used for separate 50k and QTL data $y = 1\mu + Z_{50k}a_{50k} + Z_{QTL}a_{QTL} + e$

$$\mathbf{a}_i \sim N(\mathbf{0}, \mathbf{G}_i \sigma_{gi}^2)$$
$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{W}^{-1} \sigma_e^2)$$



Bayesian

>> Model used for 50k and pooled data $\mathbf{y} = \mathbf{1}\mathbf{\mu} + \mathbf{Xg} + \mathbf{e}$

Model used for separate 50k and QTL data $y = 1\mu + X_{50k}g_{50k} + X_{QTL}g_{QTL} + e$

Priors

- $\mathbf{g}_{ij} \sim \pi_i N(0, 0.0001) + (1 \pi_i) N(0, \sigma_{ij}^2)$
- $\pi_j \sim \beta(10,1)$
- $e \sim N(0, W^{-1}\sigma_e^2)$
- $\sigma_e^2 \sigma_{ij}^2 \sim Uni$



Nordic Holstein Reliabilities



GenSAP 2nd annual meeting

UNIVERSITY



Nordic Red Reliabilities

UNIVERSITY







French Holstein Reliabilities



GenSAP 2nd annual meeting

UNIVERSITY

Conclusion

- Increase in accuracy from QTL markers
- Easy to implement in practice Impute from cow data



- More trait-specific markers
- Bayesian methods for mapping



Danish Council for Strategic Research

CC

CENTER FOR QUANTITATIVE

The

GENETICS AND GENOMICS

