Sequencing for better genomic predictions

Mogens Sandø Lund

Quantitative Genetics and Genomics
Gene 1

ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCTAGCTAGCTAGGTACCACATATAGATAACATC

Gene 2

Gene = a sequence that affect a phenotype
Gene 1

Gene 2

Bull 1
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCTAGCTAGCTAGGTACCACATATAGATACATC

Bull 2
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCTAGCTAGCTAGGTACCACATATAGATACATC

Bull 3
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCTAGCTAGCTAGGTACCACATATAGATACATC

Bull 4
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCTAGCTAGCTAGGTACCACATATAGATACATC

SNP

Breeding value

• Millions of SNPs
• Most without effect
• We don’t know which
Gene 1

ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT...AGCTAGCTAGG........GACCACATATAGATACATC

ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT...AGCTAGCTAGG........TACCACATATAGATACATC

ATGACTAGGTCTCGATCGTAGCTATAGGGCTCGCT..TGCTAGCTAGG........GACCACATATAGATACATC

ATGACTAGGTCTCGATCGTAGCTATAGGGCTCGCT..TGCTAGCTAGG........TACCACATATAGATACATC

Bull 1

Bull 2

Bull 3

Bull 4

SNP

Breeding value
Gene 1

Bull 1
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT...AGCTAGCTAGG......GACCACATATAGATACATC

Bull 2
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT...AGCTAGCTAGG......TACCACATATAGATACATC

Bull 3
ATGACTAGGTCTCGATCGTAGCTATAGGGCTCGCT...TGCTAGCTAGG......GACCACATATAGATACATC

Bull 4
ATGACTAGGTCTCGATCGTAGCTATAGGGCTCGCT...TGCTAGCTAGG......TACCACATATAGATACATC

SNP

Causative SNPs present in data

Breeding value

Higher LD for rare alleles
The challenge

• Lower reliabilities for small breeds

• Increase reference population
  • Collaborate with other populations of same breed
  • Include thousands of genotyped cows

• Across breed predictions
Across breeds predictions
Across breeds predictions

- RDM 3%
- FAY 7%
- SRB 7%
Across breeds predictions

- RDM
- SRB
- FAY

0.65
0.74
0.90

More markers → Higher LD → Higher accuracy
50K → 700K → Higher LD → 1.2% higher accuracy
Small difference in models

- **GBLUP**
  - All SNPs assumed to explain equal amount of variance

- Bayesian variable selection models
  - Some SNP large effects
  - Most SNP small or no effect

- Models fit genomic values equally well

- Thousands of genes affecting mastitis??
Region with large effects
Bayesian variable selection model

Region with large effects

Sequence improves resolution further
Genetic selection across breeds

Genes are IBD

Breed 1 Bred 2

→ estimate effects across breeds
Across breeds predictions
Summary

• Need to use across breed information smarter

• Too many genes to understand traits in full

• Identify most important genes
  • Conserved haplotypes for across breed predictions
  • Prioritise regions in WGS genomic models

• Most efficient with sequence data
What is next?

- WGS of several breeds
- Impute WGS to all bulls
- Concordance study for QTN identification
- GWAS with WGS data
- Identify map of conserved haplotypes between breeds
- Gene prioritised genomic selection within and across breeds
Nordic Red WGS project

- Sequences combined from Sweden, Finland, Denmark
- Bioinformatic platform
  - Sequence alignment
  - SNP calling
  - WGS imputation
- Common research
  - Genetic architecture
  - Prediction models