Low density genotyping and Imputation

Jack Dekkers

**Animal Breeding & Genetics
Department of Animal Science
Iowa State University**

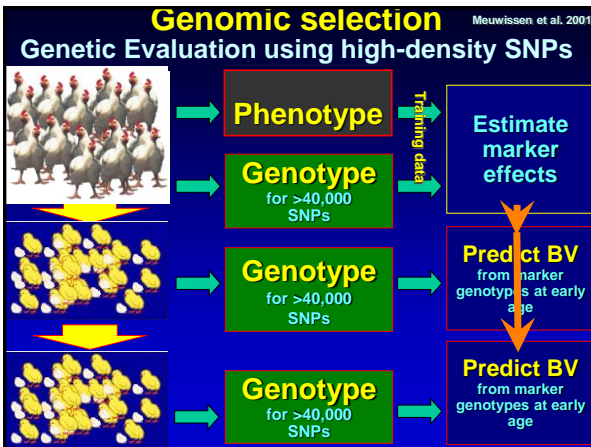IOWA STATE UNIVERSITY

Animal Breeding & Genetics

ANIMAL SCIENCE

Outline

- **Reducing the cost of implementing GS**
  - **Use of low-density panels and imputation**
- **GS for commercial crossbred performance**
  - **Potential benefits**
  - **Training on crossbred/mixed populations**
- **Redesign of breeding programs with GS**
  - **A layer chicken example**
- **Implementation of GS**

Genomic selection
Meuwissen et al. 2001
Genetic Evaluation using high-density SNPs

Phenotype

Genotype for >40,000 SNPs

Training data

Estimate marker effects

Genotype for >40,000 SNPs → Predict BV from marker genotypes at early age

Genotype for >40,000 SNPs → Predict BV from marker genotypes at early age

---

Genomic Selection using Low-Density Panels

Original principle of Genomic Selection (GS)

High-density (HD) SNP genotypes (>40k) used for both
- Estimation of marker effects (training)
- Prediction of G-EBV for selection candidates

Too costly for many species

Need Low- (<400) vs. High-density panel for routine implementation

? <$20   vs. >$150 per animal   ??

'Standard' approach to developing Low-density panels:
- **Select the 'best' SNPs from the HD-panel**
  - ➔ **Panels will be trait and population specific**

Proposed approach for pedigreed populations:
use well-spaced Low-density (<400) SNP genotypes on selection candidates to impute missing HD SNP genotypes
Habier, Dekkers, Fernando. 2009, Genetics

Implementing GS
in Pig/Poultry Programs

Problem

High cost of genotyping ←→value of an individual

Very large numbers of selection candidates

Impossible to implement genomic selection based on high density genotyping in cost efficient manner

Solution

Combination of strategic genotyping and imputation

Information used for imputation

- **LD across the population**
  - To impute from medium density (>10,000 SNP) to high density – up to sequence
- **Linkage within families**
  - To impute from very low density (<1000 SNP) to high density

---

Imputation using population-wide LD

	Haplotypes
HD-genotyped	ACAAGGATTCCGAT
HD-genotyped	GCTATCATGCCTAT
LD-genotyped	--T---A----T--

Imputation using population-wide LD

	Haplotypes
HD-genotyped	ACAAGGATTCCGAT
HD-genotyped	GCTATCATGCCTAT
LD-genotyped	--T---A----T--

Imputation using population-wide LD

Haplotypes

HD-genotyped   ACAAGGATTCCGAT

HD-genotyped   GCTATCATGCCTAT

LD-genotyped   GCTATCATGCCTAT

Imputation using population-wide LD

Haplotypes

HD-genotyped   ACAAGGATTCCGAT

HD-genotyped   GCTATCATGCCTAT

LD-genotyped   GCTATCATGCCTAT

LD-genotyped   --T---A----G--

Imputation using population-wide LD

Haplotypes

HD-genotyped   ACAAGGATTCCGAT

HD-genotyped   GCTATCATGCCTAT

LD-genotyped   GCTATCATGCCTAT

LD-genotyped   ??T???A????G??
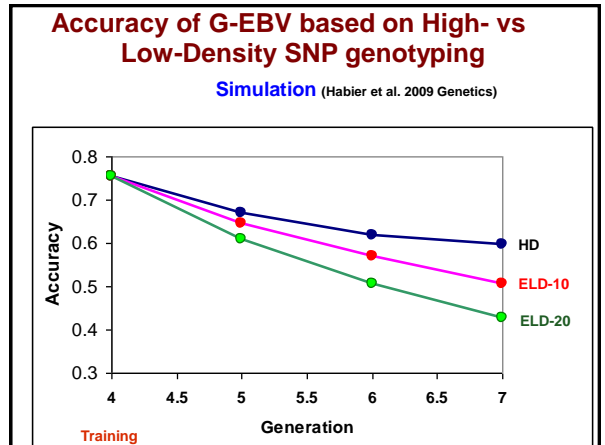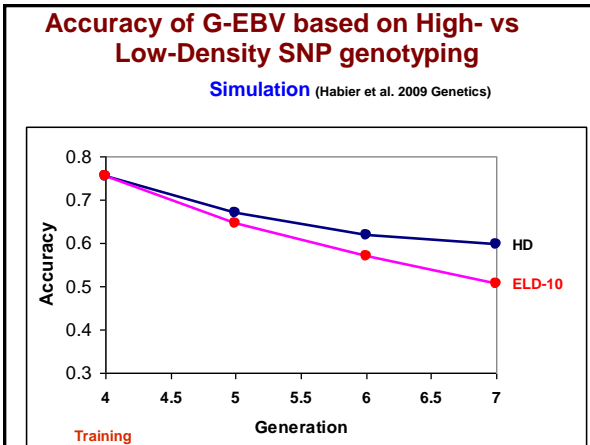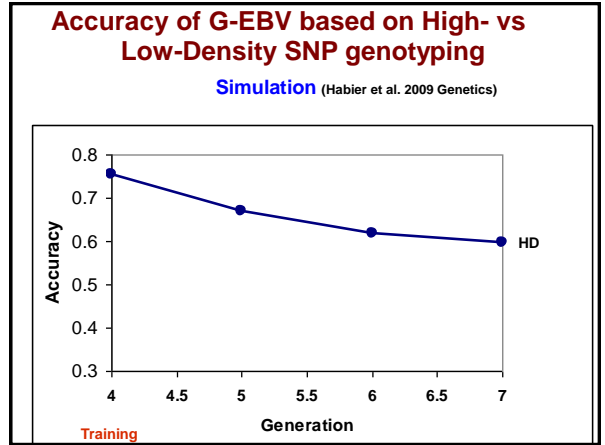

Information used for imputation

- **LD across the population**
  - To impute from medium density (>10,000 SNP) to high density – up to sequence
- **Linkage within families** – as explained before
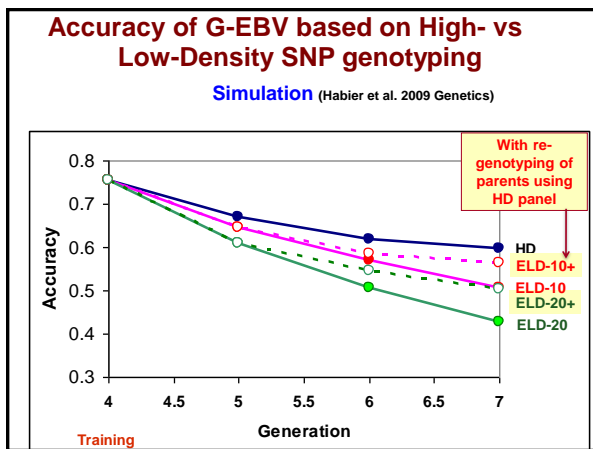 - To impute from very low density (<1000 SNP) to high density

Imputation based on Linkage Information

Sire s — paternal / maternal

LD-SNPs

Progeny i — paternal / maternal

LD-SNPs

Dam d — paternal / maternal

HD-GS → $EBV_i = \sum_{SNP\ k} \left(g^m_{ik} + g^p_{ik} \right)$ — Sum estimates of effects of maternal and paternal SNP alleles

LD-GS → $EBV_i = \sum_{SNP\ k} \left(p^{md}_{ik} g^m_{dk} + p^{pd}_{ik} g^p_{dk} + p^{ms}_{ik} g^m_{sk} + p^{ps}_{ik} g^p_{sk} \right)$

Probability that i received dam's maternal allele at SNP k
PDM = Probability of Descent of Marker allele

Requirements:
- Ordered/phased HD SNP genotypes of parents
- Imputation of HD SNP genotypes on progeny

Accuracy of G-EBV based on High- vs Low-Density SNP genotyping

Simulation (Habier et al. 2009 Genetics)

Accuracy of G-EBV based on High- vs Low-Density SNP genotyping

Simulation (Habier et al. 2009 Genetics)

Accuracy of G-EBV based on High- vs Low-Density SNP genotyping

Simulation (Habier et al. 2009 Genetics)

Accuracy of G-EBV based on High- vs Low-Density SNP genotyping

Simulation (Habier et al. 2009 Genetics)

Imputation results in HyLine data

Neil O'Sullivan, Janet Fulton, Petek Settar and Jesus Arango
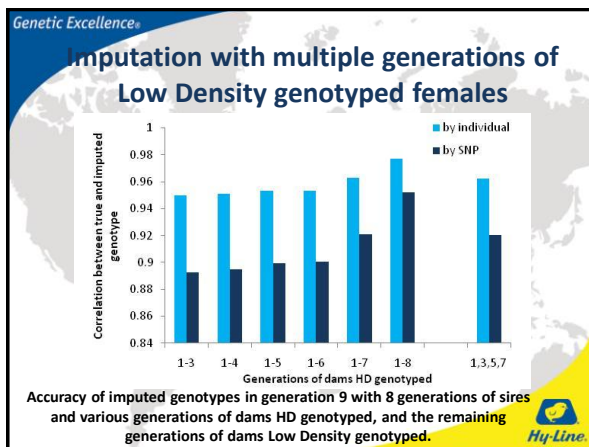
HY-LINE INTERNATIONAL

Anna Wolc, David Habier, John Hickey, Mehdi Sargolzaei,
Dorian Garrick, Rohan Fernando, Nathan Bowerman,
Chunkao Wang, Jack Dekkers,

IOWA STATE UNIVERSITY, USA
POZNAN UNIVERSITY OF LIFE SCIENCES, POLAND
UNIVERSITY OF NEW ENGLAND, AUSTRALIA
**UNIVERSITY OF GUELPH, CANADA**

Hy-Line data

- 8 generations of HD sires and dams
- Selection candidates : 544 individuals from generation 9

- **High Density genotypes** - 4,893 segregating SNPs on chromosome 1

- **Low Density genotypes** – Simulated panel of 73 ~equally spaced SNPs (equivalent to ~400 SNP across the genome)

Imputation with multiple generations of Low Density genotyped females

Accuracy of imputed genotypes in generation 9 with 8 generations of sires and various generations of dams HD genotyped, and the remaining generations of dams Low Density genotyped.

5

Genomic Selection using Low-Density SNPs

Conclusions

GS can be implemented by genotyping selection candidates for <400 SNPs spread across the genome
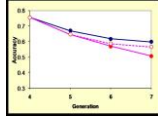
• Loss in accuracy limited: < 5 % - if parents re-genotyped HD
         - sufficient to genotype only sires

• Cost effectiveness depends on cost
          of Low-   vs.   High-density genotyping
           $20 ←??→ $150

• Loss in accuracy ~ independent of # QTL and # traits

• LD-genotyped individuals can also be used for training

• Allows imputing to higher densities / sequence from founders