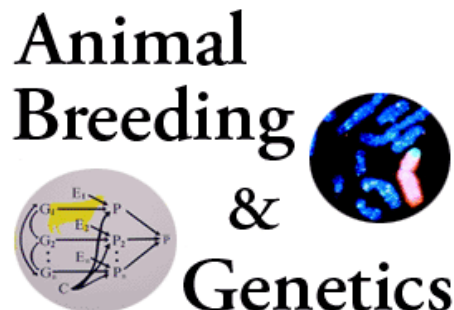


# Low density genotyping and Imputation

Jack Dekkers

Animal Breeding & Genetics  
Department of Animal Science  
Iowa State University

IOWA STATE  
UNIVERSITY



ANIMAL  
SCIENCE

# 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-genotype

ACAAGGATTCCGAT

HD-genotype

GCTATCATGCCCTAT

LD-genotype

--T---A-----T--

# Imputation using population-wide LD

## Haplotypes

HD-genotype

ACAAGGATTCCGAT

HD-genotype

GCTATCATGCCCTAT

LD-genotype

GCTATCATGCCCTAT



# Imputation using population-wide LD

## Haplotypes

HD-genotype

ACAAGGATTCCGAT

HD-genotype

GCTATCATGCCCTAT

LD-genotype

GCTATCATGCCCTAT

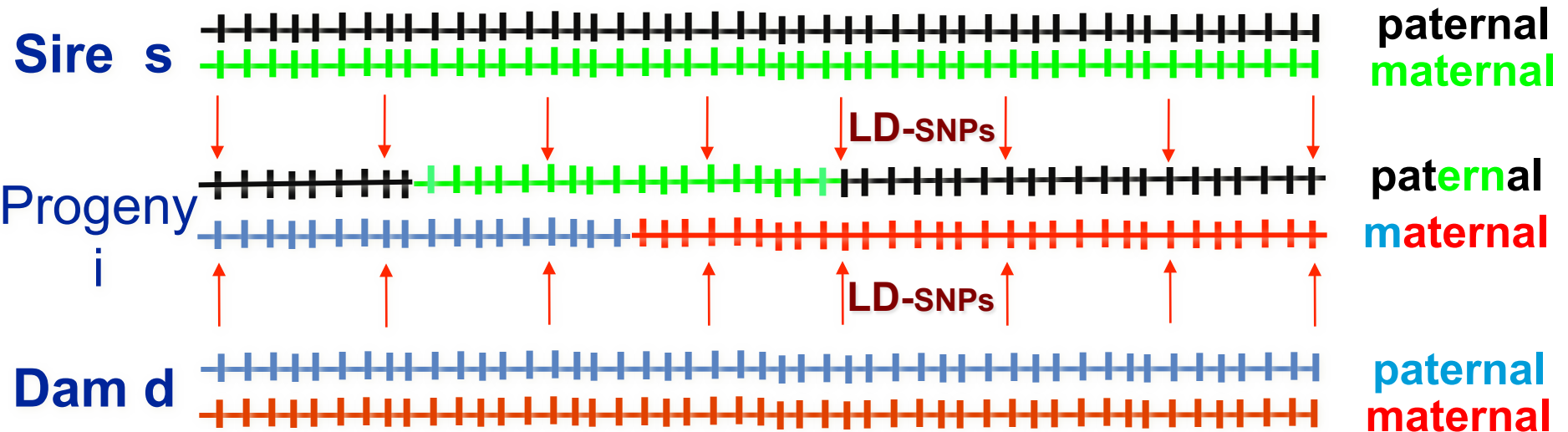
LD-genotype

??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**
  - To impute from very low density ( $<1000$  SNP) to high density

# Imputation based on Linkage Information



**HD-GS**  $\rightarrow$  
$$EBV_i = \sum_{\text{SNP } k} (g_{ik}^m + g_{ik}^p)$$

Sum estimates of effects of maternal and paternal SNP alleles

**LD-GS**  $\rightarrow$  
$$EBV_i = \sum_{\text{SNP } k} (p_{ik}^{md} g_{dk}^m + p_{ik}^{pd} g_{dk}^p + p_{ik}^{ms} g_{sk}^m + p_{ik}^{ps} g_{sk}^p)$$

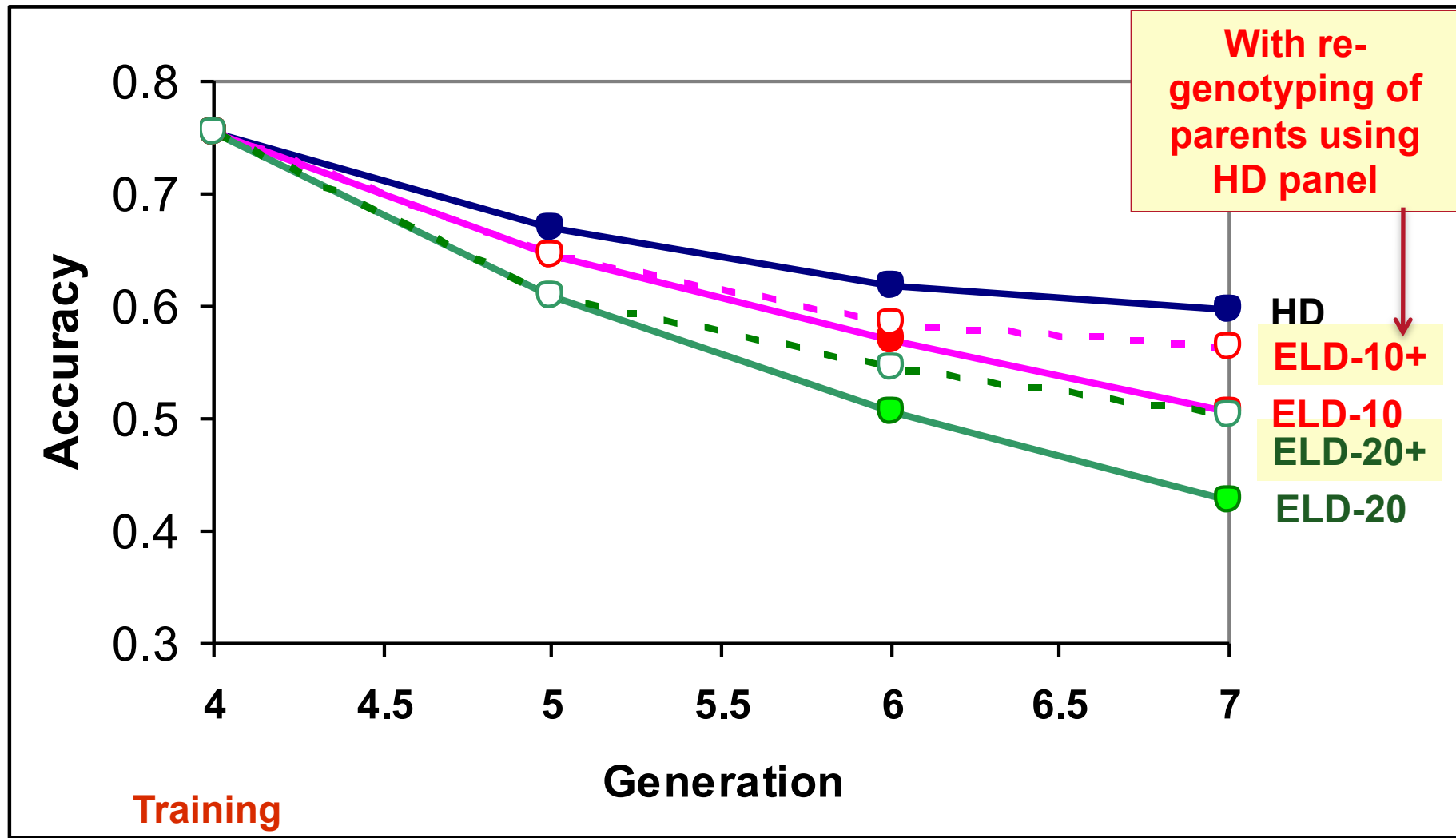
Probability that i received dam's maternal allele at SNP k  
PDM = Probability of Descent of Marker allele

## 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)





# Imputation results in HyLine data

**Neil O'Sullivan, Janet Fulton, Petek Settar and Jesus Arango**

**HY-LINE INTERNATIONAL**

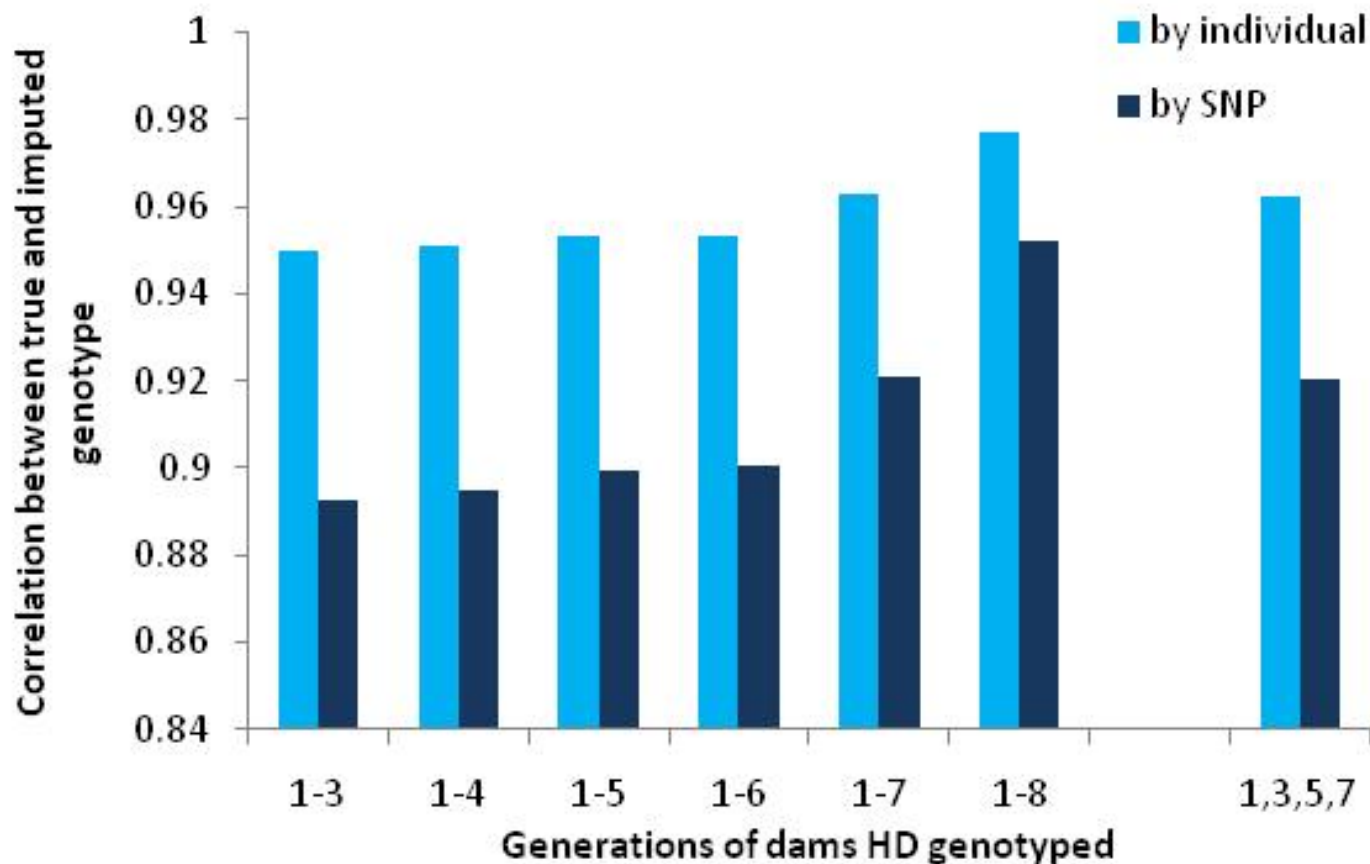
**Anna Wolc, David Habier, John Hickey, Mehdi Sargolzaei,  
Dorian Garrick, Rohan Fernando, Nathan Bowerman,  
Chunkao Wang, Jack Dekkers,**

**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

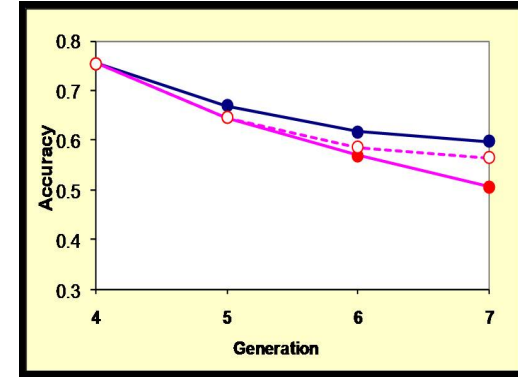


**Accuracy of imputed genotypes in generation 9 with 8 generations of sires and various generations of dams HD genotyped, and the remaining generations of dams Low Density genotyped.**

# 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