EFFECT OF IMPUTED MARKER GENOTYPES ON ACCURACY OF GENOMIC BLUP IN AQUACULTURE-LIKE POPULATIONS

Sergio Vela Avitúa *‡•

Jørgen Ødegård*†

* Norwegian University of Life Sciences
† Akvaforsk Genetics Center
• Nofima
GENOMIC BLUP, MARKER DENSITY AND THE GENOMIC MATRICES

Genomic BLUP (GBLUP) as the simplest form of Genomic Selection.

For the implementation of GBLUP (and Genomic Selection in General), it is required a large number of markers distributed across the genome.

In order to improve genomic estimates, methods to include phenotypes from ungenotyped individuals through pedigree have been developed.
WHY TO IMPUTE?

Economical reasons
- Sparse genotypes cheaper than dense genotypes
- Increase the number of genotyped individuals

Technological reasons
- Change of platform
- New available products

Dense genotypes VS Sparse genotypes

prices:

€

€
IMPUTATION: POPULATION-BASED METHODS

Haplotype library

Target samples

A C G O C A G T C
A T G G C G A T C
A C G T T A A T A
G T G G C G T C C
G C T G C C G C A
T T G G C A G T C
T T C G C A C A C

A C G O C A G T C
T T G O C A G T C
IMPUTATION FAMILY-BASED METHODS
IMPUTATION FAMILY-BASED METHODS
IMPUTATION FAMILY-BASED METHODS
OBJECTIVE

To assess the effect of imputed genotypes in the accuracy of Genomic Estimated Breeding Values (GEBVs) under different scenarios by varying:

- Number of target (dense) markers
- Number of reference (sparse) markers
- Reference (dense genotyped) population

After imputation, to assess the effect of including phenotypes of ungenotyped relatives (full and half sibs of parents and grandparents) in the accuracy of GEBVs.
METHODS
SIMULATED DATA

QMSim (Sargolzaei, M and Schenkel, E.S., 2009)

Historical Population

Ne=1,000
Random Mating

Generation 0
Mean segregating SNPs/Chr = 2900
Mean segregating QTLs/Chr = 280

Generation 3
5 selection candidates/family
training data 10 full sibs/family

Base Population

15 Males 30 Females

30 offspring 30 offspring

10,000 Generations

Genome Structure

10 Chromosome pairs/individual
1 Morgan/chromosome
30,001 Markers $\mu=1e^{-5}$
1,000 QTLs $\mu=1e^{-5}$
Normal trait $h^2=0.3$
QTLs with gamma distribution ($\beta=0.4$)
10,000 Generations

30 offspring
IMPUTATION: STRATEGIES

Imputing down the pedigree

- Dense genotyped (reference individuals)
- Sparse genotyped (target individuals)
IMPUTATION: STRATEGIES

Imputing up the pedigree

Training data  Candidates  Candidates  Training data

Dense genotyped (reference individuals)
Sparse genotyped (target individuals)
DENSITIES

All the imputations were made using one of the densities combinations as follow:

<table>
<thead>
<tr>
<th>Density</th>
<th>Sparse</th>
<th>Target (Dense)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-950</td>
<td>10 SNPs/Morgan</td>
<td>~950 SNPs/Morgan</td>
</tr>
<tr>
<td>10-2.9k</td>
<td>10 SNPs/Morgan</td>
<td>All Markers (~2900 SNPs/Morgan)</td>
</tr>
<tr>
<td>100-950</td>
<td>100 SNPs/Morgan</td>
<td>~950 SNPs/Morgan</td>
</tr>
<tr>
<td>100-2.9k</td>
<td>100 SNPs/Morgan</td>
<td>All Markers (~2900 SNPs/Morgan)</td>
</tr>
</tbody>
</table>
GENOMIC AND PHENOTYPIC DATA

Scenario 1

- Genotype + phenotype
- Genotype
- Phenotype
- No genotype or phenotype

Legend:
- Green: Phenotype
- Orange: Genotype
- Black: No genotype or phenotype
- Blue: Genotype + phenotype
GENOMIC AND PHENOTYPIC DATA

Scenario 2

Legend:
- Green: Phenotype
- Yellow: Genotype
- Black: No genotype or phenotype
- Blue: Genotype + phenotype
GENOMIC AND PHENOTYPIC DATA

Scenario 3

[Genetic tree diagram showing generations with symbols for phenotype, genotype, and genotype + phenotype.]

Legend:
- Green: Phenotype
- Black: No genotype or phenotype
- Yellow: Genotype
- Blue: Genotype + phenotype
IMPUTATION: METHOD

Population-based method:
- Beagle V4 (S. R. Browning and B. L. Browning., 2007)

Family-based method:
- Fimpute (Sargolzaei et al., 2014)

Combined family-and population-based
- Alphalmpute (Hickey et al., 2011)
GENOMIC RELATIONSHIP MATRICES AND GENOMIC BLUP

Genomic Relationships after imputation were constructed for each method using the software Gmatrix (version 2) (Guosheng Su and Per Madsen).

Genomic BLUP was carried out using DMU (version 6 release 5.2) (Per Madsen & Just Jensen). Genomic and Pedigree relationship matrices were merged using the PGMIX option in DMU in order to include phenotypic information from ungenotyped relatives from previous generations.

Accuracies of EBVs and G-EBVs were estimated as correlation between true (TBVs) and estimated breeding values (EBVs or G-EBVs).
RESULTS
RESULTS IMPUTING DOWN THE PEDIGREE

G-EBVs accuracy

Imputation accuracy

All methods better than traditional pedigree method
RESULTS IMPUTING DOWN THE PEDIGREE

![Graph showing correlation with sparse-target density (SNPs/Chr) for different scenarios and methods.](image)
RESULTS IMPUTING UP THE PEDIGREE

G-EBVs accuracy

![Graph showing G-EBVs accuracy with different methods (Fam-Pop-based, All known, Pop-based, Family-based, Pedigree) across different Sparse-Target density (SNPs/Chr).](image)

Imputation accuracy

- **Pop-based UpPed**
  - Accuracy (% of correctly imputed genotypes)
  - Graph showing accuracy across Sparse-Target density (SNPs/Chr)

- **Family-based UpPed**
  - Accuracy (% of correctly imputed genotypes)
  - Graph showing accuracy across Sparse-Target density (SNPs/Chr)
RESULTS IMPUTING UP THE PEDIGREE

![Graph showing correlation between sparse-target density and scenario results.](image)

**Method:**
- Fam-Pop-based
- All known
- Pop-based
- Family-based
- Pedigree