USE OF GENOMIC INFORMATION TO IMPROVE ACCURACY OF PREDICTION FROM GROUP RECORDS

Thinh T. Chu, John W. M. Bastiaansen, Peer Berg, Hans Komen

AARHUS
UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

EAAP 2019 - GHENT, BELGIUM
26 AUGUST 2019

THINH TUAN CHU
PHD STUDENT

WAGENINGEN
UNIVERSITY & RESEARCH

INTRODUCTION

Individual recording difficult and expensive for important traits such as feed efficiency and egg production

Group recording relatively easier to obtain

Variance components and EBV from group rec. using pedigree-based BLUP

Genotyping becoming cheaper and can be available before phenotyping

How to utilize genotyping info to increase accuracy of selection from group rec.

INTRODUCTION

What we know:

- Genomic info increases accuracy of GEBV from individual rec.
- Accuracy of EBV from group rec. depends relationship group structure:
  - Closer relationship within groups -> better accuracy
- Based on pedigree: full-sibs > paternal half-sibs > half-sibs
- Genomic explains relationships better than pedigree

What we don't know:

- Genomic information available. How to utilize this to improve accuracy of selection in breeding program?

AIMS

Grouping based on genomic information to improve accuracy of selection:

- Propose grouping methods based on genomic information improve accuracy of selection from group records
- Use GBLUP models to estimate variance components from group records

METHODS

Simulation of a breeding scheme (ADAM)

- Genome structure
- 20sire x200dam; 16 full-sibs per dam: all anin. genotyped. One group will have 4 animals.
- Trait: $y_a = a + e$ with $V_a = 0.3$; Ve =0.7

Forming groups by grouping methods (R, STRUCTURE)

- Supervised clustering on GRM
- Unsupervis. clustering STRUCTURE
- Random grouping of full-sibs
- Random grouping of p. half-sibs

Phenotyping: individual records & group records

Estimation of variance components & EBVs from individual & group rec. (DMU)

Evaluation of scenarios: accuracy of prediction, bias, coancestry coef ...

METHODS

Simulation of a breeding scheme
(ADAM)

- Genome structure
- 20sire x200dam; 16 full-sibs per dam: all anin. genotyped. One group will have 4 animals.
- Trait: y_a = a + e with V_a = 0.3; Ve =0.7

Forming groups by grouping methods
(R, STRUCTURE)

Supervised clustering on GRM

Grouping based on genomic info

- Evolutionary algorithm (scripted in R) to <u>maximize genomic relationship</u> between animals <u>within groups</u>.

  Other criteria tried, but not used: minimize relationships between groups. Maximise standard deviation of relationships between groups.

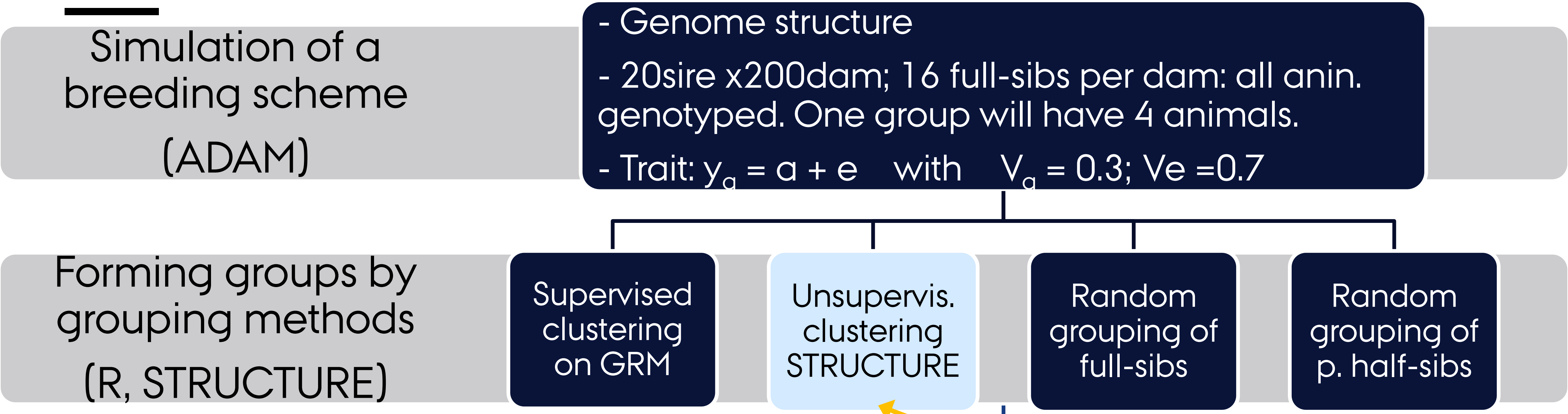
- Do for full-sibs (can do on half-sibs or else, but time-consuming and group members always full-sibs)

Phenotyping: individual records & gro

Estimation of variance components &

Evaluation of scenarios: accuracy of p

METHODS



Simulation of a breeding scheme

(ADAM)

- Genome structure
- 20sire x200dam; 16 full-sibs per dam: all anin. genotyped. One group will have 4 animals.
- Trait: $y_a = a + e$ with $V_a = 0.3$; Ve =0.7

Forming groups by grouping methods

(R, STRUCTURE)

Supervised clustering on GRM

Unsupervis. clustering STRUCTURE

Random grouping of full-sibs

Random grouping of p. half-sibs

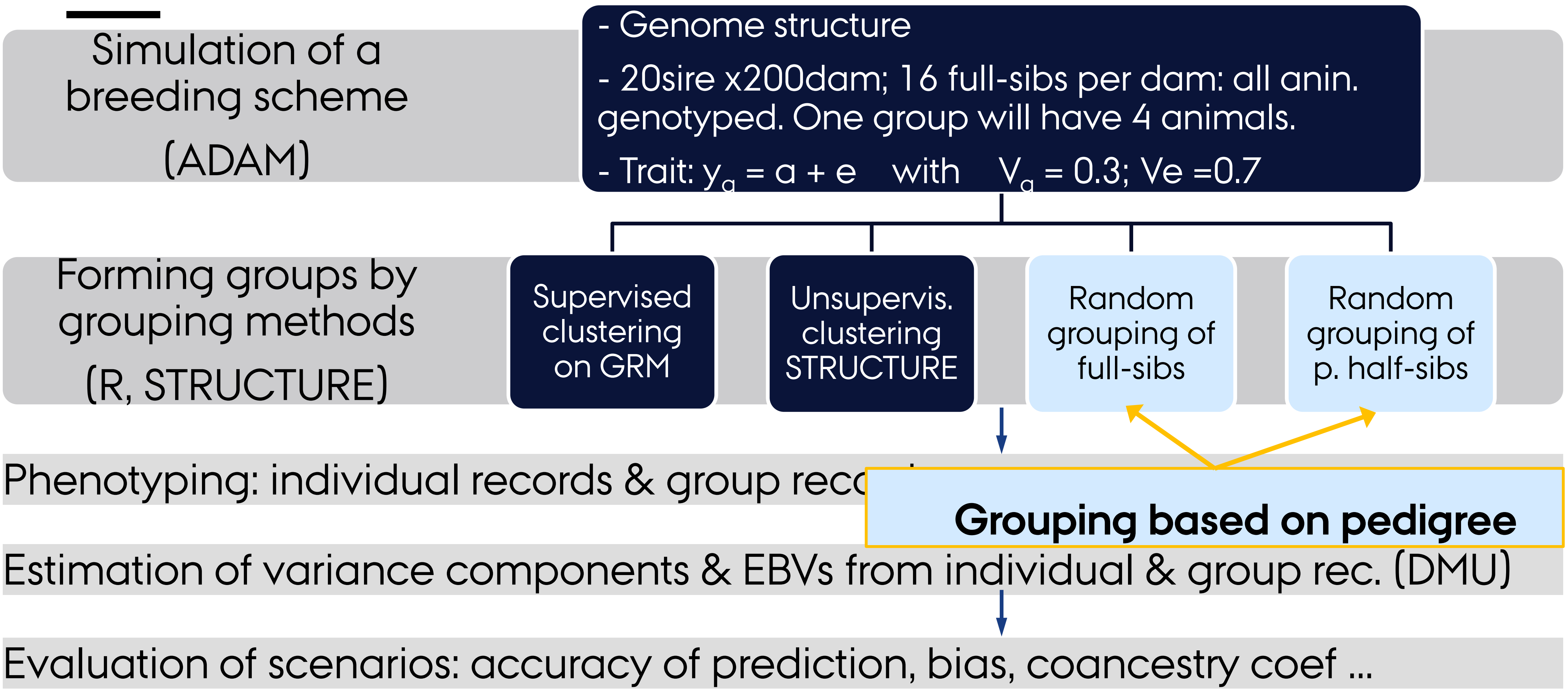
Phenotyping: individual re

Estimation of variance co

Evaluation of scenarios: ac

Grouping based on genomic information:
- STRUCTURE program: unsupervised clustering
- By full-sibs.
- Equal size group approach to equalize number of animals per groups

METHODS



Simulation of a breeding scheme (ADAM)

- Genome structure
- 20sire x200dam; 16 full-sibs per dam: all anin. genotyped. One group will have 4 animals.
- Trait: $y_a = a + e$ with $V_a = 0.3$; Ve =0.7

Forming groups by grouping methods (R, STRUCTURE)

Supervised clustering on GRM

Unsupervis. clustering STRUCTURE

Random grouping of full-sibs

Random grouping of p. half-sibs

Phenotyping: individual records & group reco

Grouping based on pedigree

Estimation of variance components & EBVs from individual & group rec. (DMU)

Evaluation of scenarios: accuracy of prediction, bias, coancestry coef ...

METHODS

Simulation of a breeding scheme

(ADAM)

- Genome structure
- 20sire x200dam; 16 full-sibs per dam: all anin. genotyped. One group will have 4 animals.
- Trait: $y_a = a + e$ with $V_a = 0.3$; Ve =0.7

Forming groups by grouping methods

(R, STRUCTURE)

Supervised clustering on GRM

Unsupervis. clustering STRUCTURE

Random grouping of full-sibs

Random grouping of p. half-sibs

Phenotyping: individual records & group records

Estimation of variance components & EBVs from individual & group rec. (DMU)

Evaluation of scenarios: accuracy of prediction, bias, coancestry coef ...

METHODS-MODELS

Simulation of a breeding scheme (ADAM)

Forming groups by grouping methods (R, STRUCTURE)

Individual records: y= 1μ + Zg + e

Model for group records

Similar to Olson *et al.* 2006 (Exact method) & Su *et al.* 2018

Use genomic relationship matrix G, applied for equal group size.

Eg: Two groups of 8 animals (3-10) that from 2 parents (1-2):

$$\begin{bmatrix}2.6\\3.5\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}\mu + \begin{bmatrix}0&0&1&1&1&1&0&0&0&0\\0&0&0&0&0&0&1&1&1&1\end{bmatrix}\begin{bmatrix}g_1\\g_2\\g_3\\\dots\\g_{10}\end{bmatrix} + \begin{bmatrix}\sum_{i=3}^{6} e_i\\\sum_{i=7}^{10} e_i\end{bmatrix}$$

Phenotyping: individual records & group records

Estimation of variance components & EBVs from individual & group rec. (DMU)

Evaluation of scenarios: accuracy of prediction, bias, coancestry coef ...

METHODS - OUTPUT

Accuracy of prediction: Correlation between TBV and GEBV

Bias: Regression TBV on GEBV

Coancestry coef. - Inbreeding measures: relationship between 20 top ranking males and 200 top rank females.

Relationship distributions:

Realized genomic relationship matrix **G** from VanRaden [11] method 1:

$$G = \frac{MM'}{2\sum p_j^*(1-p_j^*)}$$

Simu
breed
(

Formir
groupi
(R, ST

andom
uping of
alf-sibs

Phenotyp

Estimatio

c. (DMU)

Evaluation of scenarios: accuracy of prediction, bias, coancestry coef ...

METHODS

Simulation of a breeding scheme

(ADAM)

- Genome structure

- 20sire x200dam; 16 full-sibs per dam: all anin. genotyped. One group will have 4 animals.

- Trait: $y_a = a + e$ with $V_a = 0.3$; Ve =0.7

Forming groups by grouping methods

(R, STRUCTURE)

nsupervis.
ring

Random grouping of full-sibs

Random grouping of p. half-sibs

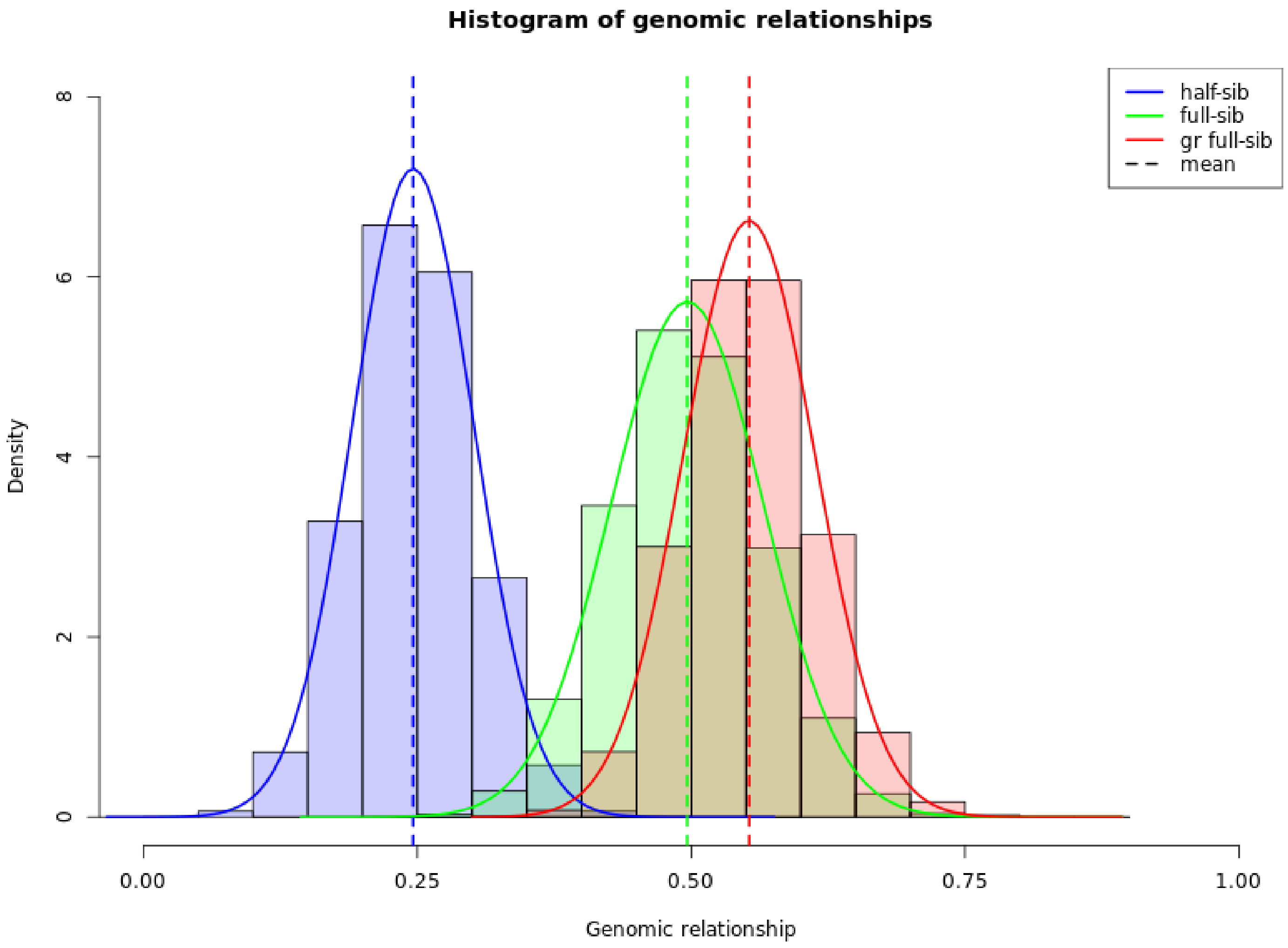
100 replicates

Phenotyping: individual records & grou

Estimation of variance components & EBVs fro group rec. (DMU)

Evaluation of scenarios: accuracy of prediction, bias, co cestry coef ...

RESULTS



Relationships	Mean	SD
Half-sibs	0.246	0.055
Paternal half-sibs	0.270	0.093
Full-sibs	0.496	0.070
Genomic-close full-sibs grouped by supervised clustering based on genomic relationships	0.553	0.061
Genomic-close full-sibs grouped by unsupervised clustering based on genotypes	0.538	0.067

RESULTS

Records	σ^2_a (SD)	σ^2_e (SD)	Accuracy (SD)	Bias (SD)	Coancestry coef. (SD)
Simulated parameters	0.30	0.70			
Individual records	0.300 (0.030)	0.698 (0.022)	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group rec. from supervised clustering based on genomic relationships	0.302 (0.042)	0.691 (0.048)	0.762 (0.029)	1.007 (0.054)	0.041 (0.010)
Group rec. from unsupervised clustering based on genotypes	0.301 (0.043)	0.693 (0.050)	0.758 (0.030)	1.009 (0.054)	0.041 (0.009)
Group rec. from random full-sibs	0.298 (0.045)	0.695 (0.052)	0.749 (0.031)	1.015 (0.060)	0.043 (0.010)
Group rec. from random paternal half-sibs	0.301 (0.062)	0.695 (0.050)	0.682 (0.040)	1.017 (0.092)	0.048 (0.010)

RESULTS

Records	σ^2_a (SD)	σ^2_e (SD)	Accuracy (SD)	Bias (SD)	Coancestry coef. (SD)
Simulated parameters	0.30	0.70			
Individual records	0.300 (0.030)	0.698 (0.022)	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group rec. from supervised clustering based on genomic relationships	0.302 (0.042)	0.691 (0.048)	0.762 (0.029)	1.007 (0.054)	0.041 (0.010)
Group rec. from unsupervised clustering based on genotypes	0.301 (0.043)	0.693 (0.050)	0.758 (0.030)	1.009 (0.054)	0.041 (0.009)
Group rec. from random full-sibs	0.298 (0.045)	0.695 (0.052)	0.749 (0.031)	1.015 (0.060)	0.043 (0.010)
Group rec. from random paternal half-sibs	0.301 (0.062)	0.695 (0.050)	0.682 (0.040)	1.017 (0.092)	0.048 (0.010)

RESULTS

Records	σ^2_a (SD)	σ^2_e (SD)	Accuracy (SD)	Bias (SD)	Coancestry coef. (SD)
Simulated parameters	0.30	0.70			
Individual records	0.300 (0.030)	0.698 (0.022)	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group rec. from supervised clustering based on genomic relationships	0.302 (0.042)	0.691 (0.048)	0.762 (0.029)	1.007 (0.054)	0.041 (0.010)
Group rec. from unsupervised clustering based on genotypes	0.301 (0.043)	0.693 (0.050)	0.758 (0.030)	1.009 (0.054)	0.041 (0.009)
Group rec. from random full-sibs	0.298 (0.045)	0.695 (0.052)	0.749 (0.031)	1.015 (0.060)	0.043 (0.010)
Group rec. from random paternal half-sibs	0.301 (0.062)	0.695 (0.050)	0.682 (0.040)	1.017 (0.092)	0.048 (0.010)

RESULTS

Records	σ^2_a (SD)	σ^2_e (SD)	Accuracy (SD)	Bias (SD)	Coancestry coef. (SD)
Simulated parameters	0.30	0.70			
Individual records	0.300 (0.030)	0.698 (0.022)	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group rec. from supervised clustering based on genomic relationships	0.302 (0.042)	0.691 (0.048)	0.762 (0.029)	1.007 (0.054)	0.041 (0.010)
Group rec. from unsupervised clustering based on genotypes	0.301 (0.043)	0.693 (0.050)	0.758 (0.030)	1.009 (0.054)	0.041 (0.009)
Group rec. from random full-sibs	0.298 (0.045)	0.695 (0.052)	0.749 (0.031)	1.015 (0.060)	0.043 (0.010)
Group rec. from random paternal half-sibs	0.301 (0.062)	0.695 (0.050)	0.682 (0.040)	1.017 (0.092)	0.048 (0.010)

RESULTS

Records	σ^2_a (SD)	σ^2_e (SD)	Accuracy (SD)	Bias (SD)	Coancestry coef. (SD)
Simulated parameters	0.30	0.70			
Individual records	0.300 (0.030)	0.698 (0.022)	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group rec. from supervised clustering based on genomic relationships	0.302 (0.042)	0.691 (0.048)	0.762 (0.029)	1.007 (0.054)	0.041 (0.010)
Group rec. from unsupervised clustering based on genotypes	0.301 (0.043)	0.693 (0.050)	0.758 (0.030)	1.00 [illegible] 54)	0.041 (0.009)
Group rec. from random full-sibs	0.298 (0.045)	0.695 (0.052)	0.749 (0.031)	1.01 [illegible] 60)	0.043 (0.010)
Group rec. from random paternal half-sibs	0.301 (0.062)	0.695 (0.050)	0.682 (0.040)	1.01 [illegible] 92)	0.048 (0.010)

Geno. relat. increases

RESULTS

Records	σ^2_a (SD)	σ^2_e (SD)	Accuracy (SD)	Bias (SD)	Coancestry coef. (SD)
Simulated parameters	0.30	0.70			
Individual records	0.300 (0.030)	0.698 (0.022)	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group rec. from supervised clustering based on genomic relationships	0.302 (0.042)	0.691 (0.048)	0.762 (0.029)	1.007 (0.054)	0.041 (0.010)
Group rec. from unsupervised clustering based on genotypes	0.301 (0.043)	0.693 (0.050)	0.758 (0.030)	1.009 (0.054)	0.041 (0.009)
Group rec. from random full-sibs	0.298 (0.045)	0.695 (0.052)	0.749 (0.031)	1.015 (0.060)	0.043 (0.010)
Group rec. from random paternal half-sibs	0.301 (0.062)	0.695 (0.050)	0.682 (0.040)	1.017 (0.092)	0.048 (0.010)

RESULTS

Individual records vs group records:

- VC from group & individual rec. consistent; and same as simulated values.
- Accuracy of EBV from individual rec. higher than from group rec.
- No different for bias
- SD for VC estimates and accuracy lower from individual than group rec.

Among prediction from group records:

Higher genomic relationships between animals within groups

- -> higher accuracy
- -> tend to lower SD of accuracy & lower coancestry coef.

Grouping based on genomic information better than pedigree information

DISCUSSIONS

Genomic BLUP vs pedigree-based BLUP

- Pedigree-based BLUP:
  - Full-sibs that were from the same group would have identical EBV
- GBLUP:
  - Full-sibs that were from the same group would have different GEBV

- GBLUP > pedigree-based BLUP for accuracy of prediction

DISCUSSIONS

Supervised clustering based on genomic relationships vs Unsupervised clustering based on genotyping

- Same concept that aims to increase genomic similarity between group members

- But because we assume fixed facilities for rearing (number of groups and group sizes:
  - Accuracy, coancestry coef., SD better for supervised clustering

- Computation demanding less for supervised clustering

- Good for designing breeding programs where facilities fixed (number of groups, group sizes)

DISCUSSIONS

Factors affecting supervised clustering based on genomic relationships

- Redundant offspring for grouping, in which some animals not assigned into any groups

- Group sizes, number of groups, family sizes, genome sizes

CONCLUSIONS

Use of genomic info. to improve accuracy of prediction from group rec.:

- Through use of G matrix in prediction
- Through grouping based on genomic info.

Of 2 grouping methods based on genomic info. proposed:

- Supervised clustering method better: computation time, fixed facility, clear distinction of which groups animals belong to.

To further improve benefit of grouping based on genomic info

- Redundant offspring available for grouping (some animals not assigned into any groups)

AARHUS
UNIVERSITY

WAGENINGEN
UNIVERSITY & RESEARCH

SUPERVISED CLUSTERING BASED ON GRM

- Round 1:
- Initial: Animals randomly assigned into different groups/clusters
- Iterated steps:
  - Calc. relationships betw. animals within groups (WG_{old})
  - Exchange of 2 randomly animals between 2 randomly groups
  - Calc. relationships betw. animals within groups (WG_{new}) after the exchange
  - Update group membership if $Wg_{new}>WG_{old}$.
  - Converged when the exchange of two animals between two groups did not increase genomic relationships for a certain number of iterations.
- Round 2: repeated ….
- Need inputs for group sizes, number of groups.

UNSUPERVISED CLUSTERING IN STRUCTURE

STRUCTURE program: unsupervised clustering

Notes:

- Apply for whole population: half-sibs clustered into groups
- Apply for half-sibs: full-sibs clustered into groups
- Apply for full-sibs: genomic close full-sibs clustered into groups
- Group sizes or number of groups are not fixed. Need to re-arrange animals for breeding programs
- For breeds/species/origins of animals, STRUCTURE is optimum, but for designing breeding program, STRUTURE is not so good.

EG. UNSUPERVISED CLUSTERING - STRUCTURE

	G1	G2	G3	G4
221	0.6630	0.1381	0.1949	0.0040
222	0.8463	0.1154	0.0071	0.0313
223	0.5744	0.1555	0.1659	0.1042
224	0.0058	0.1793	0.7307	0.0842
225	0.1817	0.2500	0.4134	0.1549
226	0.4563	0.0692	0.3816	0.0930
227	0.9723	0.0102	0.0043	0.0132
228	0.7769	0.1124	0.0051	0.1057
229	0.5111	0.0091	0.4746	0.0052
230	0.7902	0.0042	0.1959	0.0097
231	0.9968	0.0019	0.0002	0.0011
232	0.0010	0.0010	0.9965	0.0015
233	0.0016	0.0148	0.9743	0.0093
234	0.9982	0.0006	0.0006	0.0006
235	0.0006	0.0010	0.9970	0.0015
236	0.6882	0.0032	0.2962	0.0124