Computation of many relationships between metafounders replacing phantom parents

Mario Calus, Claudia Sevillano, Rob Bergsma &
Jeremie Vandenplas

29 August 2019, EAAP, Ghent

BREED4FOOD

Breed4Food is dedicated to be the leading research consortium in animal breeding, genetics and genomics enabling the Breed4Food partners to breed better products to benefit society's needs.

WAGENINGEN UR
For quality of life

CRV

Cobb

HENDRIX GENETICS

Topigs Norsvin
PROGRESS IN PIGS

Background

- Pedigrees contain many phantom parent groups
  - Defined by year(class), sex, line, etc.

- How to consider all of those in genomic evaluations?

- Include as:
  - Unrelated genetic groups
  - Related metafounders
    - Relationships may benefit estimation

Objective

- Treat many genetic groups as metafounders (MF)
- Compute relationships between MF
- Use those MF in ssGBLUP

Pig data (Topigs Norsvin)

- 38,488 animals with genotypes from 3 lines
  - 34,293 from 3 purebred lines
  - 4195 3-way crossbred
- 835,773 animals in pedigree; 211 genetic groups

- Bivariate ssGBLUP model:
  - Purebred & crossbred Average Daily Gain (ADG)
  - MF: (1) none, (2) 1 per line, or (3) all 211

Computation of MF relationships I

- MF relationship = 8*covariance(allele frequencies)[1]
- *Base* allele frequencies; GLS estimator
- *Current* allele frequencies, computed:
  - Per breed
  - By regression on breed composition
- *Randomize* allele coding to avoid biased allele frequencies

[1] Garcia-Baccino, C.A., A. Legarra, O.F. Christensen, I. Misztal, I. Pocrnic, Z.G. Vitezica, and R.J.C. Cantet. 2017. Metafounders are related to F_{st} fixation indices and reduce bias in single-step genomic evaluations. Genet. Sel. Evol. 49:34.

Computation of MF relationships II

Initial idea:

- Estimate allele frequencies for 5 MF with the highest average relationships with the genotyped population
- Interpolate to obtain allele frequencies by birth-year
- BUT: unrealistic MF relationships obtained (not shown)

Computation of MF relationships III

Alternative idea:

- Define 1 MF per line; compute:
    - Allele frequencies per line
    - MF relationships per & between lines
- Expand those to all 211 genetic groups
    - Within line: all relationships assumed the same

Validation ssGBLUP

- Validation records of 83-139 (recent) boar line sires
  - Using average offspring performance (AOP)

- Evaluated measures:
  - Accuracy
  - Regression of AOP on GEBV; $E(\beta)=0.5$

RESULTS

(1) 1 MF per line – MF relationships

Allele frequency	**Base**	**Current**
Original allele coding	0.48 0.07 0.18 0.07 0.25 0.11 0.18 0.11 0.31	0.49 0.05 0.17 0.05 0.32 0.07 0.17 0.07 0.35

(1) 1 MF per line – MF relationships

Allele frequency	Base	Current
Original allele coding	0.48 0.07 0.18 0.07 0.25 0.11 0.18 0.11 0.31	0.49 0.05 0.17 0.05 0.32 0.07 0.17 0.07 0.35
Randomized allele coding	0.54 0.10 0.24 0.10 0.27 0.15 0.24 0.15 0.37	0.55 0.08 0.23 0.08 0.33 0.10 0.23 0.10 0.41

(2) Expanding relationships to all 211 MF

	Line A		
Line A	**0.54**	**…**	**0.54**
	⋮	**0.54**	⋮
	0.54	**…**	**0.54**

(2) Expanding relationships to all 211 MF

	Line A		
Line A	**0.54**	**…**	**0.54**
	⋮	**0.54**	⋮
	0.54	**…**	**0.54**
Line B	0.10	…	0.10
	⋮	0.10	⋮
	0.10	…	0.10
Line C	0.24	…	0.24
	⋮	0.24	⋮
	0.24	…	0.24

(2) Expanding relationships to all 211 MF

	Line A			Line B			Line C		
Line A	**0.54**	**…**	**0.54**	0.10	…	0.10	0.24	…	0.24
	⋮	**0.54**	⋮	⋮	0.10	⋮	⋮	0.24	⋮
	0.54	**…**	**0.54**	0.10	…	0.10	0.24	…	0.24
Line B	0.10	…	0.10	**0.27**	…	**0.27**	0.15	…	0.15
	⋮	0.10	⋮	⋮	**0.27**	⋮	⋮	0.15	⋮
	0.10	…	0.10	**0.27**	…	**0.27**	0.15	…	0.15
Line C	0.24	…	0.24	0.15	…	0.15	**0.37**	…	**0.37**
	⋮	0.24	⋮	⋮	0.15	⋮	⋮	**0.37**	⋮
	0.24	…	0.24	0.15	…	0.15	**0.37**	…	**0.37**

=> Not positive definite; added 0.001 to diagonal

Results ssGBLUP

Scen	#MF	MF rel. ships	#iterations	Acc.	Bias
1	None	No	2194	0.574	0.631
2	1 per line (3)	Yes (Base)	3067	0.562	0.548
3	All (211)	Yes (Base)	4814	0.140	0.050
4	1 per line (3)	Yes (Curr)	3099	0.561	0.544
5	All (211)	Yes (Curr)	6109	0.237	0.071

Conclusions

- MF relationships:
  - Similar with current or base allele frequencies
  - Important to randomize allele coding

- Impact on ssGBLUP results:
  - 1 MF per line:
    - Comparable accuracy & considerably less bias
  - All 211 MF:
    - No meaningful results
- => Fitting few MF may have favourable impact

Scaling genetic variances when using MF

- Compute average (across lines) MF self-relationship: $\bar{\gamma}$
- Divide all (co)variances by: $\left(1-\frac{\bar{\gamma}}{2}\right)$
- E.g. using current allele frequencies:
  - $\bar{\gamma}$=0.43
  - $\left(1-\frac{\bar{\gamma}}{2}\right)$=0.78
  - Effectively: (co)variances are multiplied by 1.28