Metafounders in ssGBLUP

Daniela Lourenco

BLUPF90 TEAM – 02/2023

UNIVERSITY OF GEORGIA

College of Agricultural & Environmental Sciences

Animal Breeding and Genetics Group

Armidale Animal Breeding Summer Course 2023

Daniela Lourenco
Mehdi Sargolzaei

Missing parents in (ssG)BLUP

- When parents are missing → Average (G)EBV for the animals go back to the base

  - Not ideal because of selection (genetic trends!)

  - Solution: set a base for (G)EBV of animals with unknown parent groups (UPG)

- UPG help to increase accuracy/reliability and reduce inflation when pedigree is missing

UPG in ssGBLUP

- QP-transformation for $\mathbf{A}^{-1}$ (Quaas & Pollack, 1981;Westell et al., 1988)

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix}$$

- QP-transformation for $\mathbf{H}^{-1}$ (Misztal et al., 2013)

$$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

`OPTION exact_upg`

- Altered QP-transformation for $\mathbf{H}^{-1}$ (Tsuruta et al., 2019)

$$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2'(-\mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

G does not depend on the completeness of pedigrees...no UPG for **G**!

`OPTION exact_upg`

`OPTION TauOmegaQ2 0 1`

3

UPG

- UPG definition
  - year of birth, sex, breed, line

- UPG can be inbred (VanRaden, 1992)
  - Average inbreeding for the animals in the years spanned by the UPG

- UPG cannot be related
  - Is it true?
  - Ancestral populations have finite size, so UPG may be related
  - How to deal with that?

Metafounders

- Proxy of base animals

- Inbred and related UPG

- Considered as a random effect

- **A** is modified to be compatible with **G** with 0.5 AF

$$\mathbf{A}_{\Gamma}^{-1} = \begin{bmatrix} \mathbf{A}_{\Gamma}^{11} & \mathbf{A}_{\Gamma}^{12} & \mathbf{A}_{\Gamma}^{1m} \\ \mathbf{A}_{\Gamma}^{21} & \mathbf{A}_{\Gamma}^{22} & \mathbf{A}_{\Gamma}^{2m} \\ \mathbf{A}_{\Gamma}^{m1} & \mathbf{A}_{\Gamma}^{m2} & \mathbf{A}_{\Gamma}^{mm} + \mathbf{\Gamma}^{-1} \end{bmatrix}$$

$$\mathbf{H}_{\Gamma}^{-1} = \mathbf{A}_{\Gamma}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{05}^{-1} - \mathbf{A}_{\Gamma 22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

- Why 0.5 AF?

  - Christensen (2012) on the compatibility between **G** and **A** …

Christensen (2012)

In relation to this problem of compatibility between marker-based and pedigree-based relationship matrices, with data from routine evaluations, selection affects the allele frequencies over time, and in principle both observed marker genotypes and observed phenotypes contain information about allele frequencies in the base population. Therefore, studies on the adjustment of the marker-based relationship matrix to the pedigree-based relationship matrix have overlooked the fact that it should in principle incorporate information on observed phenotypes.

This work explores two possibilities to solve the problem, i.e. 1) in which the pedigree-based relationship matrix is adjusted to the marker-based relationship matrix and 2) in which the single-step genetic evaluation is extended by using a joint likelihood of observed phenotypes and observed marker genotypes. This results in a single-step method in which the marker-based relationship matrix is constructed assuming all allele frequencies are equal to 0.5 and the pedigree-based relationship matrix is constructed using the unusual assumption that animals in the base population are related and inbreed with a relationship coefficient γ and an inbreeding coefficient $\gamma / 2$.

Metafounders

- What is $\mathbf{\Gamma}$?
- Matrix with relationships within and across metafounders
  - $\mathbf{\Gamma}_{ii} < 1$: negative inbreeding and higher freq of heterozygous than average population
  - $\mathbf{\Gamma}_{ii} < 1$: base population with large genetic variability
  - $\mathbf{\Gamma}_{ii} > 1$: base population is inbred
  - $\mathbf{\Gamma}_{ij} > 0$: base populations overlap
  - $\mathbf{\Gamma}_{ij} < 0$: divergent base populations
  - $\mathbf{\Gamma}_{ij} = 0$: unrelated base populations

- Estimating $\mathbf{\Gamma}$ is not easy (GLS; Garcia-Bacino et al., 2017)
  - Based on genotypes
  - Enough genotyped animals linked to MF

Metafounders in blupf90+

- Create metafounders based on your definition

- Replace missing parents (0) with –MF in the pedigree

- Run renumf90

```
UPG_TYPE
  in_pedigrees
```

- Make sure UPG are in the renumbered pedigree file: renaddXX.ped

Format: ani/sire/dam/code/......

```
1  12  11  2  0  1  1  1  0  5
2  6   8   1  0  2  1  0  0  3
7  1   11  1  0  2  0  0  3  8
9  10  14  2  0  1  0  0  2  10
3  12  7   2  0  1  1  2  0  1
4  3   9   1  0  2  1  0  0  6
11 13  14  3  0  0  0  0  2  11
5  10  9   1  0  2  1  0  0  4
8  13  7   2  0  1  0  0  1  9
6  3   7   1  0  2  1  1  0  2
10 13  14  3  0  0  0  2  0  7
```

Metafounders in blupf90

- Estimate the gamma matrix
  - `gammaf90`
  - `gammaf90 --help`

```
usage: gammaf90 [options]

--snpfile file          * snp file
--pedfile file          * pedigree file  (renaddXX.ped)
--xrefid file             XrefID file (default: snpfile_XrefID)
--gammafile file          gamma file (default: gamma.txt)
--gammainitfile file      gamma init file with starting values (default: gammainit.txt)
--num-threads             number of threads with OpenMP (default: maximum)
--allele-freqs-base       estimates of base allele frequencies in file freqs
--bounded-allele-freqs  bounds estimates of base allele frequencies to 0-1 interval
--quiet                   output not too verbose
--approximatedLikelihood    computes approximated tr(MM') instead of exact
                        tr(M V^-1 M') likelihood - approximated is faster but not recommended

* Mandatory options
> The pedigree file should be generated with RENUMF90 or compatible programs.
  Metafounders must have the larger integer code than the real animals; and
  they must not appear in the "animal" column (i.e. only in sire/dam columns).
  This coding is compatible with UPG used in BLUPF90 programs.
  If no gammainitfile is given, starting values are Gamma=I*0.01, close to UPG model.
```

Metafounders in blupf90

- Rename the output file "gamma.txt" to "renaddxx.ped_gamma"
  - (xx should be the # of animal effect) which is the default gamma file read by blupf90

- Replace "add_an_animal" (or "add_an_upg" or "add_an_upginb") in the renf90.par with "add_an_meta"

- Run blupf90+
  - Only free software with MF implemented
  - See the output to make sure metafounder subroutines are called

```
Metafounders: effect=2 gamma_file=renadd02.ped_gamma format=matrix

Option changes for metafounders:
* whichfreq      = 1 (allele frequency set to 0.5 for centering)
* whichfreqScale = 1 (allele frequency set to 0.5 for scaling)
* tunedG         = 0 (no tuning on G)
* the number of real animals     66540
* the number of IDs              66542
* the number of metafounders         2
* info in read_gamma: success
    status info from reading gamma           0
* Gamma =
     0.3953       0.2202
     0.2202       0.2820
  eigenvalues
  0.11128       0.56597
* computing inbreeding coefficients
* Scale_k =  0.8899
* Matrix G (genetic covariances) is scaled with k=  0.8899
  Original G
  0.3000
  Scaled G
  0.3371
```

Problems with Gamma estimation

- Gamma values may not make sense
  - Possible range for $\mathbf{\Gamma}_{ii}$: [0, 2f]
  - Possible range for $\mathbf{\Gamma}_{ij}$: [0, 1f]

- How to guarantee better estimates

  --allele-freqs-base   estimates of base allele frequencies in file freqs

  --bounded-allele-freqs bounds estimates of base allele frequencies to 0-1 interval

- Should not estimate VC with metafounders
  - Variances are scaled within blupf90+ $\rightarrow$ $(\sigma_u^2 \, (1 - \frac{\gamma}{2}))$

Problems with Gamma estimation

- Gamma vs. Robust Gamma

  - Use Robust Gamma when Gamma has a problem (e.g., very complex populations) and the estimates are not good

  - The robust is a "hack" to get some estimates (median vs. mean)

```
Estimated gamma: 2*sigma to gamma.txt
  0.39527       0.22018
  0.22018       0.28198
Statistics of estimated gamma
diagonal min, median, mean, max
        0.2819782094        0.3386243623        0.3386243623        0.3952705153
offdiagonal min, median, mean, max
        0.2201822526        0.2201822526        0.2201822526        0.2201822526
robust gamma in gammarobust.txt using medians:
 robust gamma
  0.33862       0.22018
  0.22018       0.33862
```

What do we expect by using MF?

- Connect populations

- Reduce the issue with missing pedigree

- Increase the compatibility between **G** and **A** (and $\mathbf{A}_{22}$)

Bermann et al. (2021)

LR accuracy

weight gain: No UPG 0.74, UPG 0.79, Metafounders 0.71
carcass yield: No UPG 0.69, UPG 0.73, Metafounders 0.74

LR b1

weight gain: No UPG 0.98, UPG 0.98, Metafounders 0.99
carcass yield: No UPG 0.99, UPG 0.96, Metafounders 0.93

No UPG · UPG · Metafounders
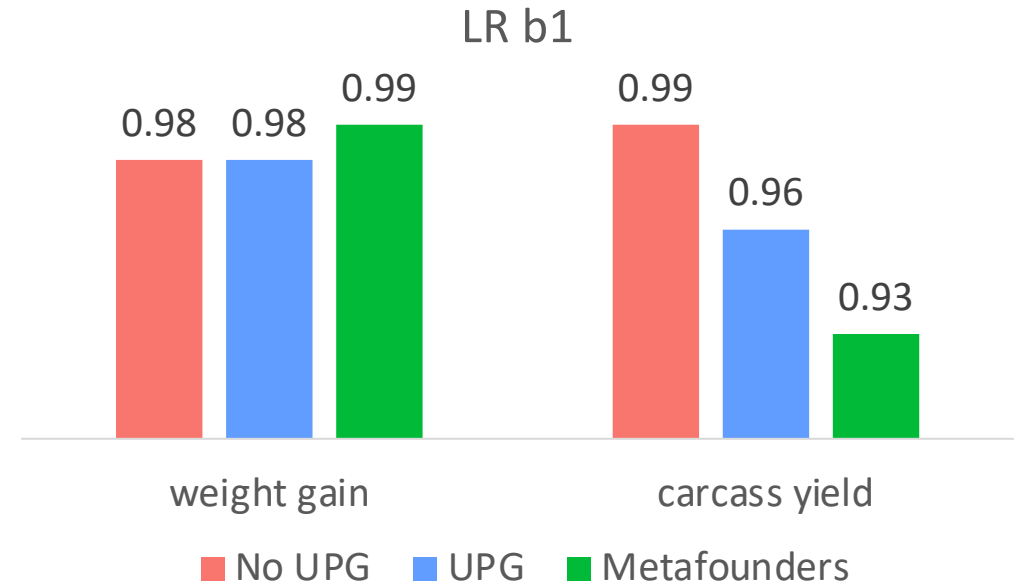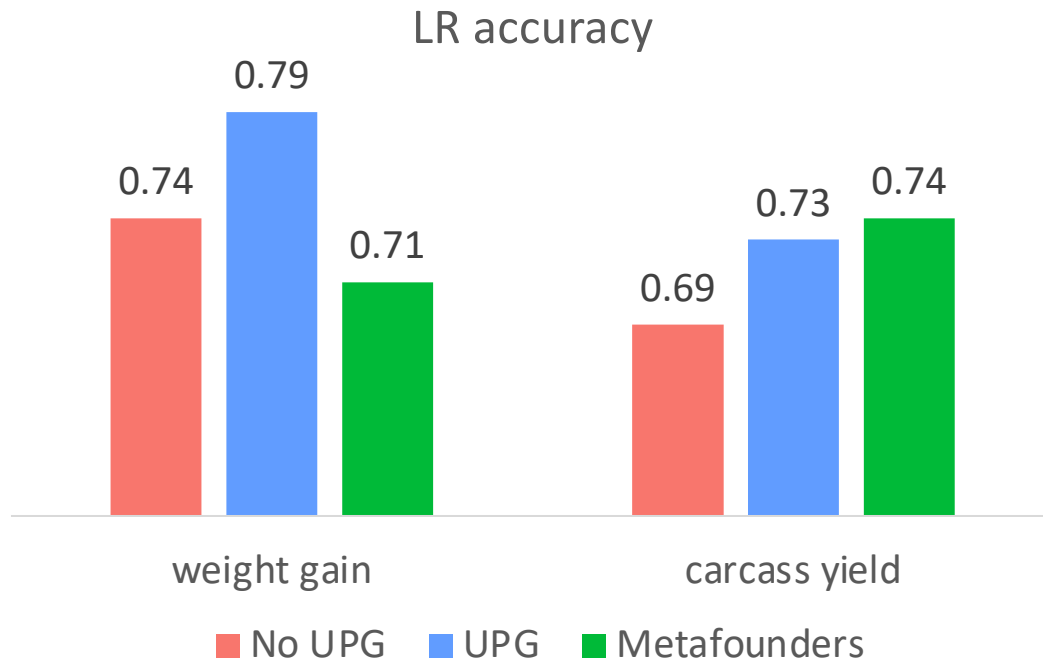
What do we expect by using MF?

- Under tests in purebred and multibreed populations (small and large)

Research Article | Open Access | Published: 12 August 2020

Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups

Fernando L. Macedo ✉, Ole F. Christensen, Jean-Michel Astruc, Ignacio Aguilar, Yutaka Masuda & Andrés Legarra

Genetics Selection Evolution **52**, Article number: 47 (2020) | Cite this article

2473 Accesses | **15** Citations | **7** Altmetric | Metrics

Research

Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions

F.L. Macedo [1, 2, 3], J.M. Astruc [4], T.H.E. Meuwissen [5], A. Legarra [1] ✉

Show more ∨

+ Add to Mendeley   ✑ Share   ❝❞ Cite

https://doi.org/10.3168/jds.2021-20860
Under a Creative Commons license   Get rights and content
                                                    open access

ORIGINAL RESEARCH article

Front. Genet., 03 December 2020 | https://doi.org/10.3389/fgene.2020.556399

Applying the Metafounders Approach for Genomic Evaluation in a Multibreed Beef Cattle Population

Vinícius Silva Junqueira[1,2*], Paulo Sávio Lopes[1], Daniela Lourenco[3], Fabyano Fonseca e Silva[1] and Fernando Flores Cardoso[4]

ORIGINAL RESEARCH article

Front. Genet., 20 August 2021 | https://doi.org/10.3389/fgene.2021.678587

Metafounders May Reduce Bias in Composite Cattle Genomic Predictions

Sabrina Kluska[1,2*], Yutaka Masuda[2], José Bento Sterman Ferraz[3], Shogo Tsuruta[2], Joanir Pereira Eler[3], Fernando Baldi[1] and Daniela Lourenco[2]

G3, 2022, 12(9), jkac180
https://doi.org/10.1093/g3journal/jkac180
Advance Access Publication Date: 3 August 2022
Multiparental Populations

Accounting for population structure in genomic predictions of *Eucalyptus globulus*

Andrew N. Callister,[1,*] Matias Bermann,[2] Stephen Elms,[3] Ben P. Bradshaw,[4] Daniela Lourenco,[2] Jeremy T. Brawner[5]

14