Slide 1

Metafounders in ssGBLUP

Daniela Lourenco

BLUPF90 Team – 02/2023

1

Slide 2

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

2

Slide 3

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

Slide 4

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?

4

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

$$A_\Gamma^{-1} = \begin{bmatrix} A_\Gamma^{11} & A_\Gamma^{12} & A_\Gamma^{1m} \\ A_\Gamma^{21} & A_\Gamma^{22} & A_\Gamma^{2m} \\ A_\Gamma^{m1} & A_\Gamma^{m2} & A_\Gamma^{mm} + \Gamma^{-1} \end{bmatrix} \qquad H_\Gamma^{-1} = A_\Gamma^{-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & G_{05}^{-1} - A_{\Gamma 22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- 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 inbred with a relationship coefficient γ and an inbreeding coefficient $\gamma / 2$.

Metafounders

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

- Estimating Γ 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

---

Problems with Gamma estimation

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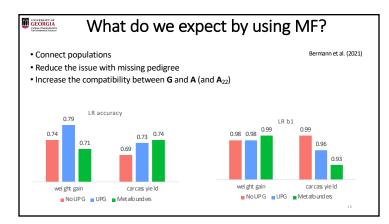 - Possible range for Γ_{ij} : [0, 1f]

- 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+ → $(\sigma_u^2 \, (1 - \frac{\gamma}{2}))$

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
```

Slide 13

What do we expect by using MF?

Bermann et al. (2021)

- Connect populations
- Reduce the issue with missing pedigree
- Increase the compatibility between **G** and **A** (and A_{22})

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

13

Slide 14

What do we expect by using MF?

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

14