

Modeling unknown parent groups or metafounders in single step genomic BLUP – results of a simulation study

J. Himmelbauer^{1,2}, H. Schwarzenbacher¹, C. Fuerst¹, B. Fuerst-Waltl²

¹ZuchtData EDV-Dienstleistungen GmbH, Dresdner Straße 89/B1/18, 1200 Vienna, Austria

²BOKU University, Vienna, Gregor-Mendel Str. 33, 1180 Vienna, Austria

Corresponding author: himmelbauer@zuchtdata.at

Abstract

The concepts considering for unknown parents are crucial in improving genetic evaluations in animal breeding by accounting for genetic differences within base populations. This study builds on a previous simulation study for the German-Austrian-Czech Fleckvieh population, presenting results that compare metafounders (MF) and unknown parent groups (UPG) for single-step genomic best linear unbiased prediction, and includes detailed analyses for scaling variance components when using MF. The results show that in both settings with complete and incomplete pedigree, evaluations using MF show the best bias and dispersion results, with minimal impact from incomplete pedigree information. In contrast, evaluations without UPG or MF and evaluations where UPG were incorporated via Quaas-Pollak-transformation in the pedigree-based and genomic relationship matrix (UPG_fullQP) exhibit substantial overestimation and overdispersion, emphasizing the importance of accurate relationship modeling in genetic evaluations. This study found that estimating variance components using MF and scaling variance components lead to the same heritability. However, using adapted variance components results in moderate overestimation and slight overdispersion of GEBV. The validation method based on the linear regression method could not detect the significant overestimation and overdispersion in UPG_fullQP. This means that commonly used validation methods tend to underestimate the advantages of MF in populations with numerous unknown pedigrees, highlighting challenges in model optimization for handling unknown parents.

Key words: ssGBLUP, unknown parents, metafounder, simulation, dairy cattle

Introduction

Thompson (1979) and Quaas (1988) published the concept of unknown parent groups (UPG) to account for genetic differences within subgroups of base populations, incorporating animals with missing parents and diverse genetic backgrounds into genetic evaluations. UPGs can have non-zero means but are assumed to be non-inbred and unrelated, similar to the base population. For single-step genomic best linear unbiased prediction (ssGBLUP) Legarra et al. (2015) extended this concept and introduced metafounder (MF), which can model relationships within and across subpopulations.

ssGBLUP uses an integrated relationship matrix (H), combining the pedigree-based (A) and genomic (G) relationship matrices. Ideally, both matrices should refer to the same base population (Christensen, 2012), though this is often not the case in cattle populations without adjustments. Methods to align G with A include those by VanRaden (2008), Vitezica et al. (2011), and Christensen (2012). MF is addressing this alignment by adapting A to match G.

In the German-Austrian-Czech Fleckvieh population, the first ssGBLUP genomic evaluation was published in April 2021 (Himmelbauer et al., 2021), using 15 UPGs for most fitness traits. MF is considered the gold

