



Alternative Approaches to handling of missing parents in genetic evaluation of dairy cattle using single-step test-day SNP-BLUP model

**Dawid Słomian, Kacper Żukowski, Monika Skarwecka,
Jan Ten Napel, Jeremie Vandenplas, Joanna Szyda**

Introduction

- Pedigree structure
- Missing data in pedigree
 - Set as missing
 - Genetic Groups
 - Metafounders
- Challenge: How to deal with missing parents?



Purpose of the study



- Compare three approaches to handling missing parents
- Scenario comparison:
 - Validation results
 - GEBV comparison
 - Assessment of average GEBV trend



Scenarios

- Levels of missing parental information:
 - **P_Real**: Routine pedigree (5.6% missing sires, 15.3% missing dams)
 - **P_2010**: Intermediate missingness (~10% sires, ~20% dams missing for animals born before 2019)
 - **P_4020**: High missingness (~20% sires, ~40% dams missing for animals born before 2019)
- Approaches to handling missing parents:
 - **Raw pedigree (RP)**: IDs of missing parents set as missing
 - **Genetic groups (GG)**: Missing parents replaced by unrelated genetic groups
 - **Metafounders (MF)**: Missing parents replaced by metafounders



Material



	Sex	Number of animals	Number of records	
Phenotype data (fat yield) $h^2=0.29$	Cows	3,707,727	Full data set	63,615,019
			Truncated data set	58,446,695
Genotype data	Cows	113,019	181,991	
	Bulls	68,972		
Pedigree data	Cows	4,569,044	4,712,143	
	Bulls	143,099		

- Polish national evaluation April 2024



Single-step test day SNP-BLUP model

$$y = Xh + Wf + Vp + Vu + e$$

y – cow's test day fat yield records in the first 3 lactations

h – fixed effects of herd-test-date-parity-milking-frequency

f – fixed lactation curve effects, modeled as regression on DIM using the Wilmink function

p – permanent environmental effects

u – additive genetic effects

e – residuals



Results

482,810 validation cows:

- Born between 2019 and 2022

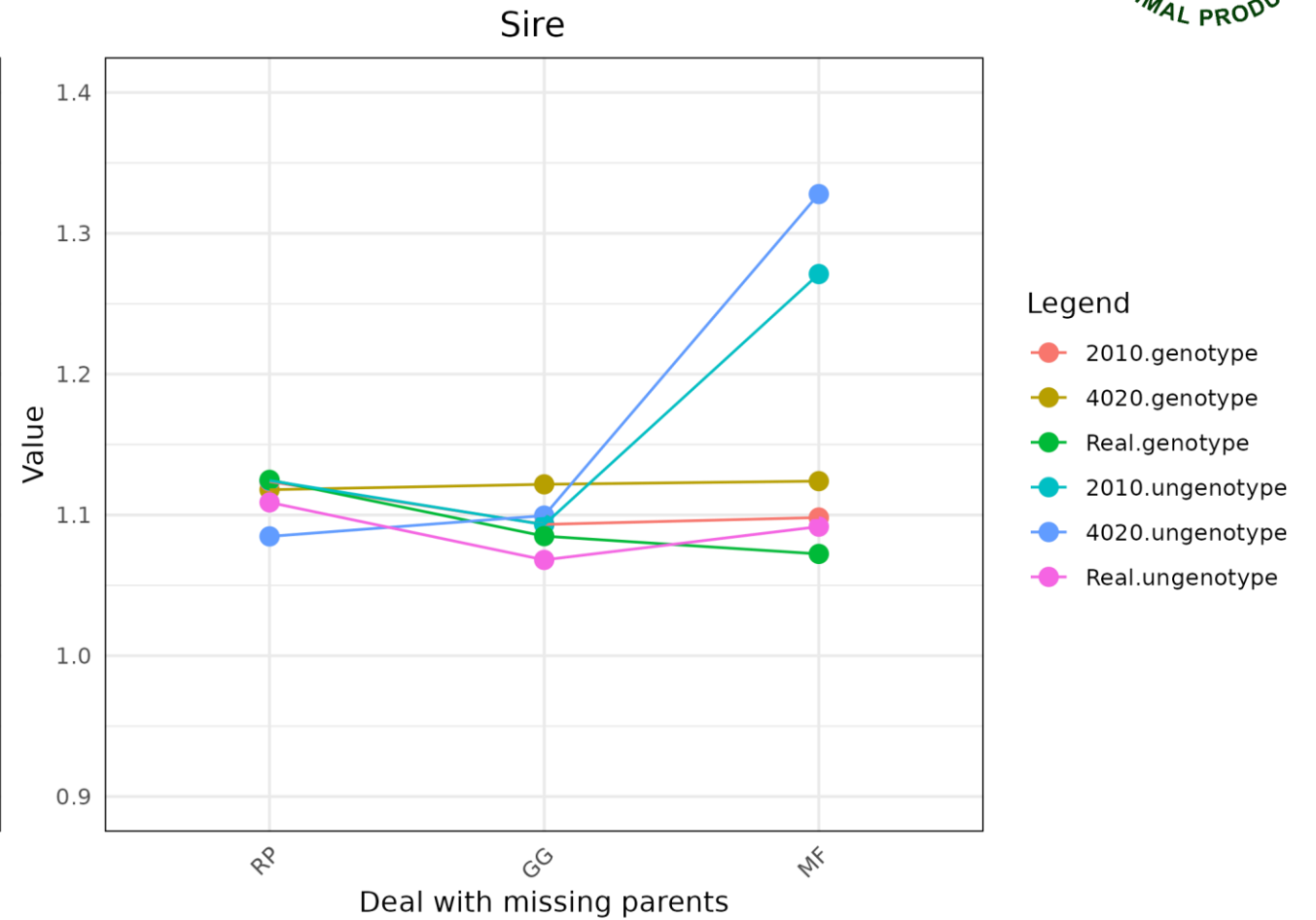
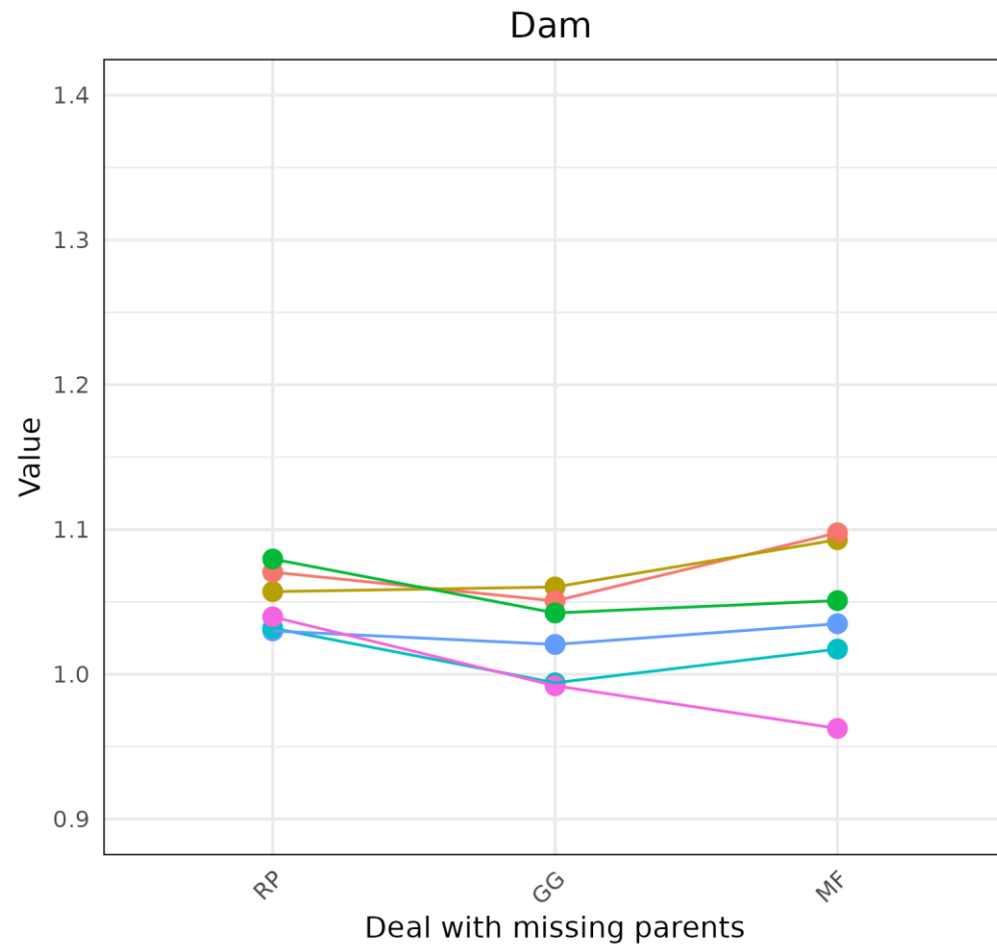
562 validation bulls:

- Sires of cows that were removed for validation
- > 20 daughters
- Born between 2017-2020



Validation results

b_1 (slope)



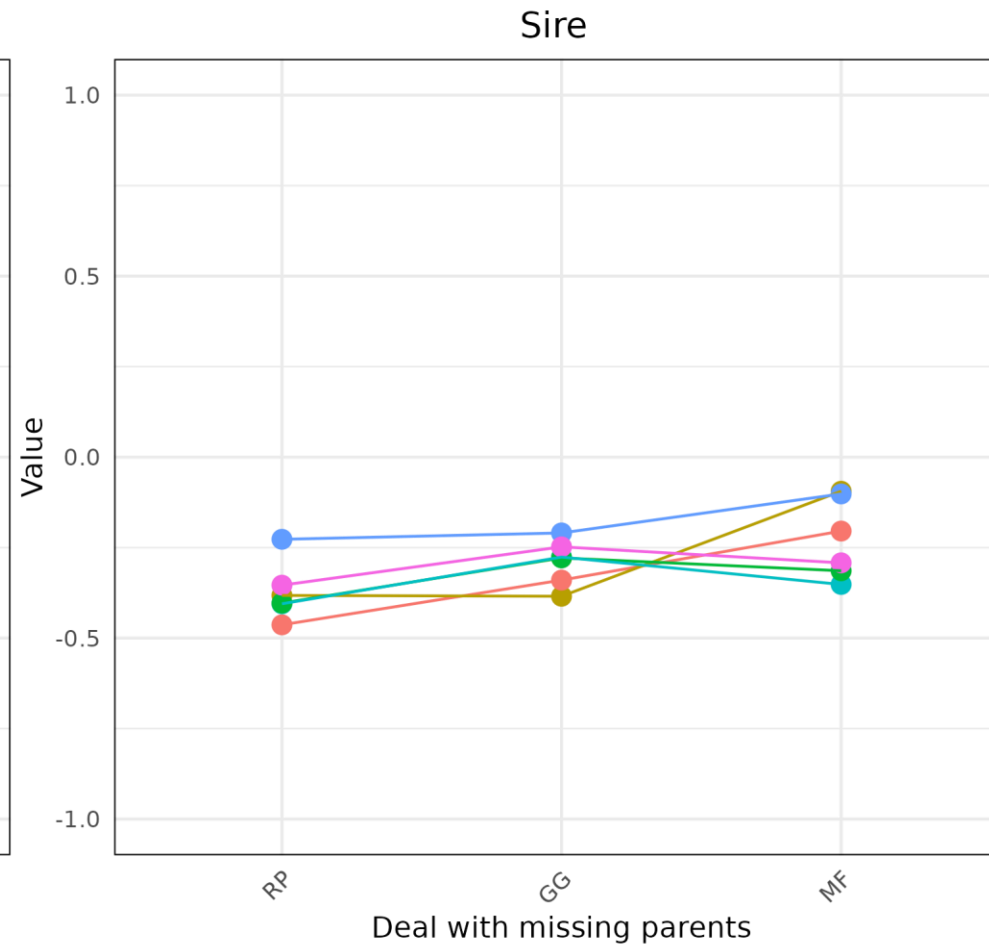
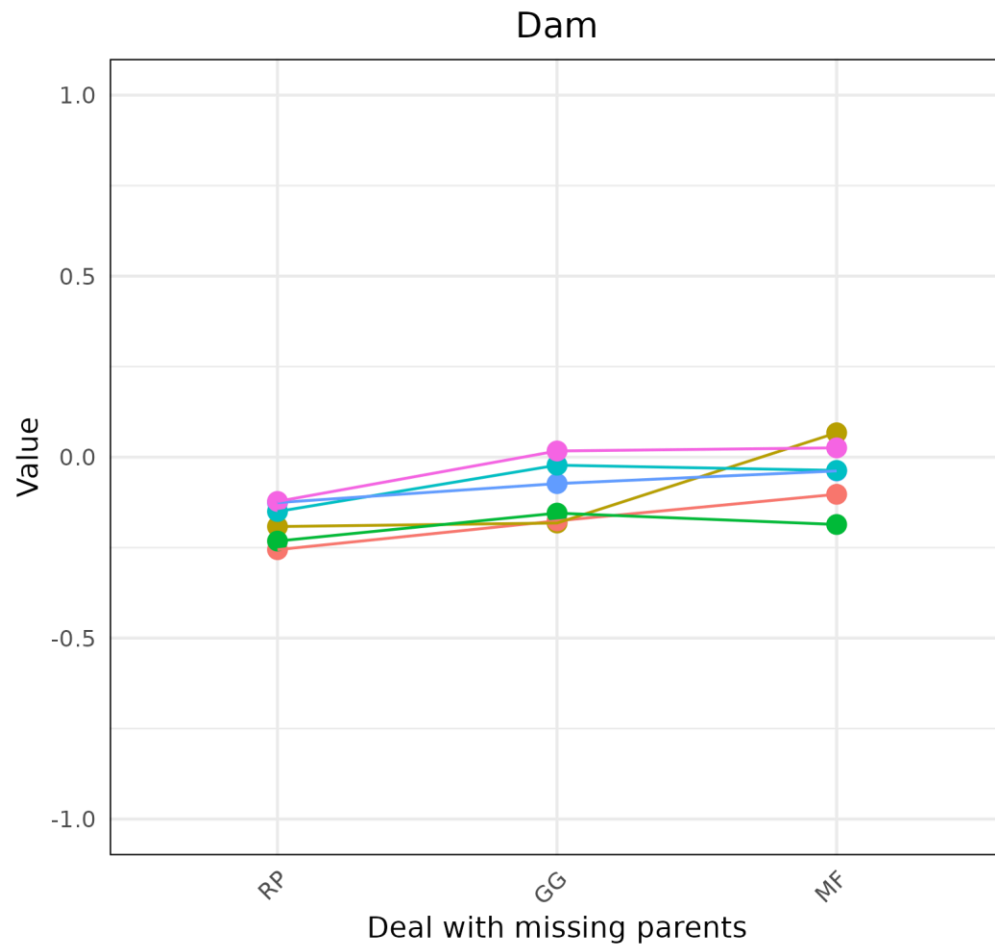
Legend

- 2010.genotype
- 4020.genotype
- Real.genotype
- 2010.ungenotype
- 4020.ungenotype
- Real.ungenotype



Validation results

b_0 (intercept)



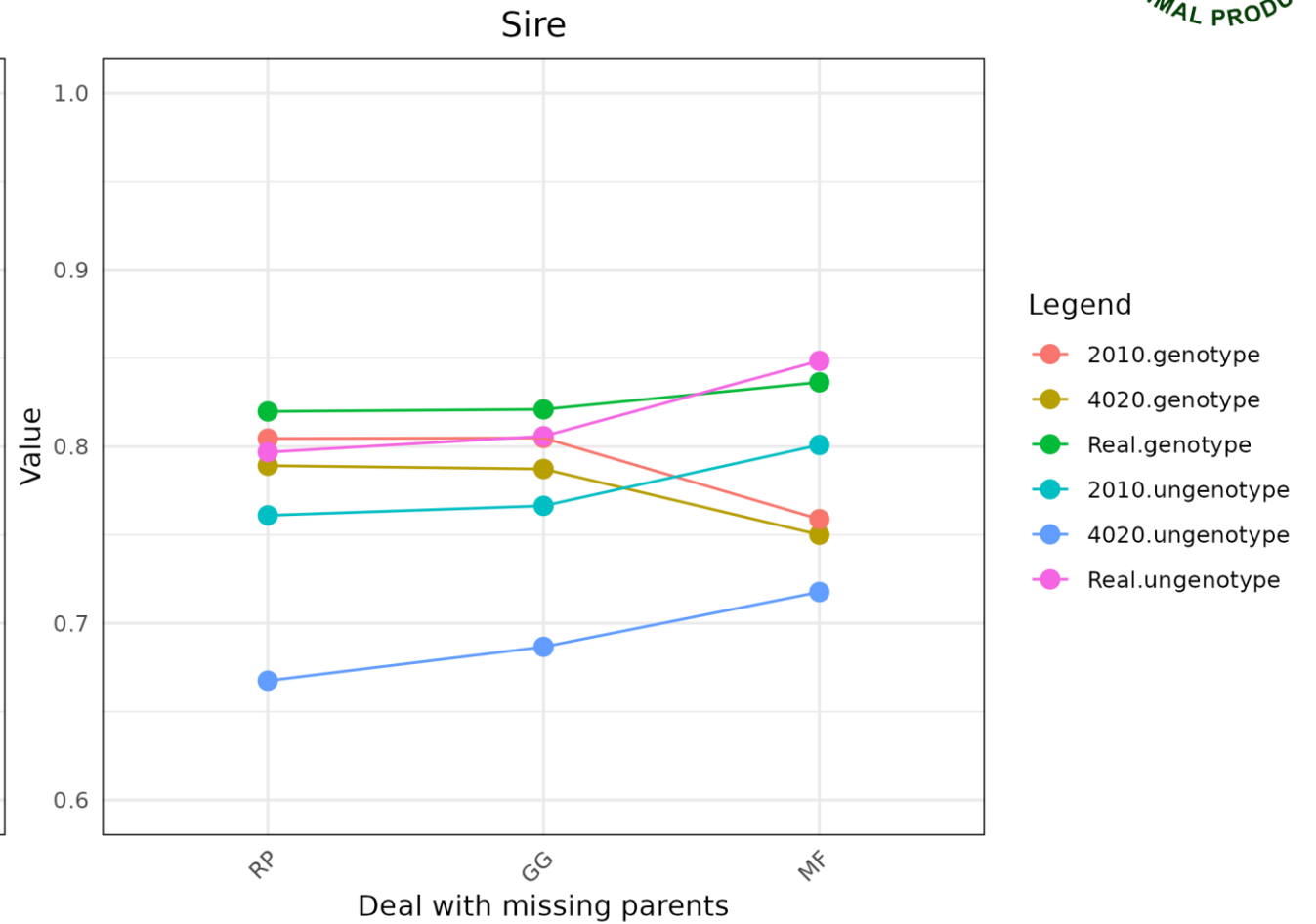
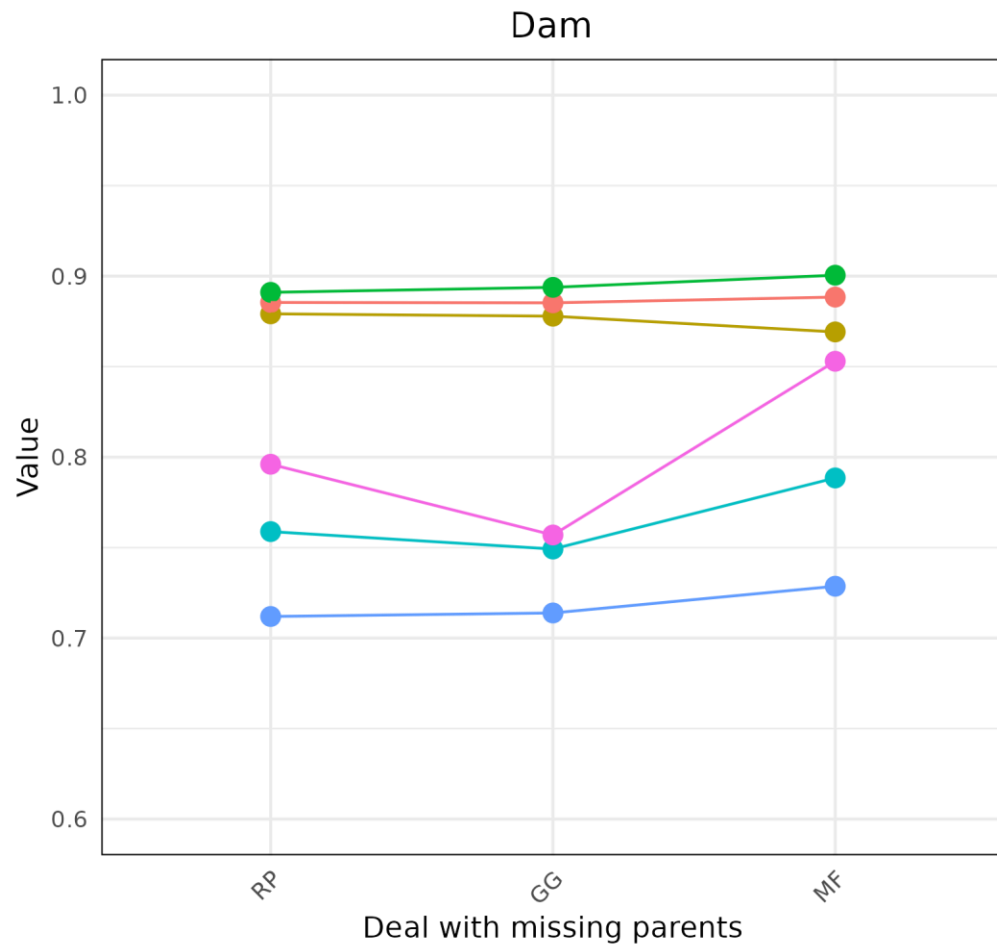
Legend

- 2010.genotype
- 4020.genotype
- Real.genotype
- 2010.ungenotype
- 4020.ungenotype
- Real.ungenotype



Validation results

R^2



Legend

- 2010.genotype
- 4020.genotype
- Real.genotype
- 2010.ungenotype
- 4020.ungenotype
- Real.ungenotype

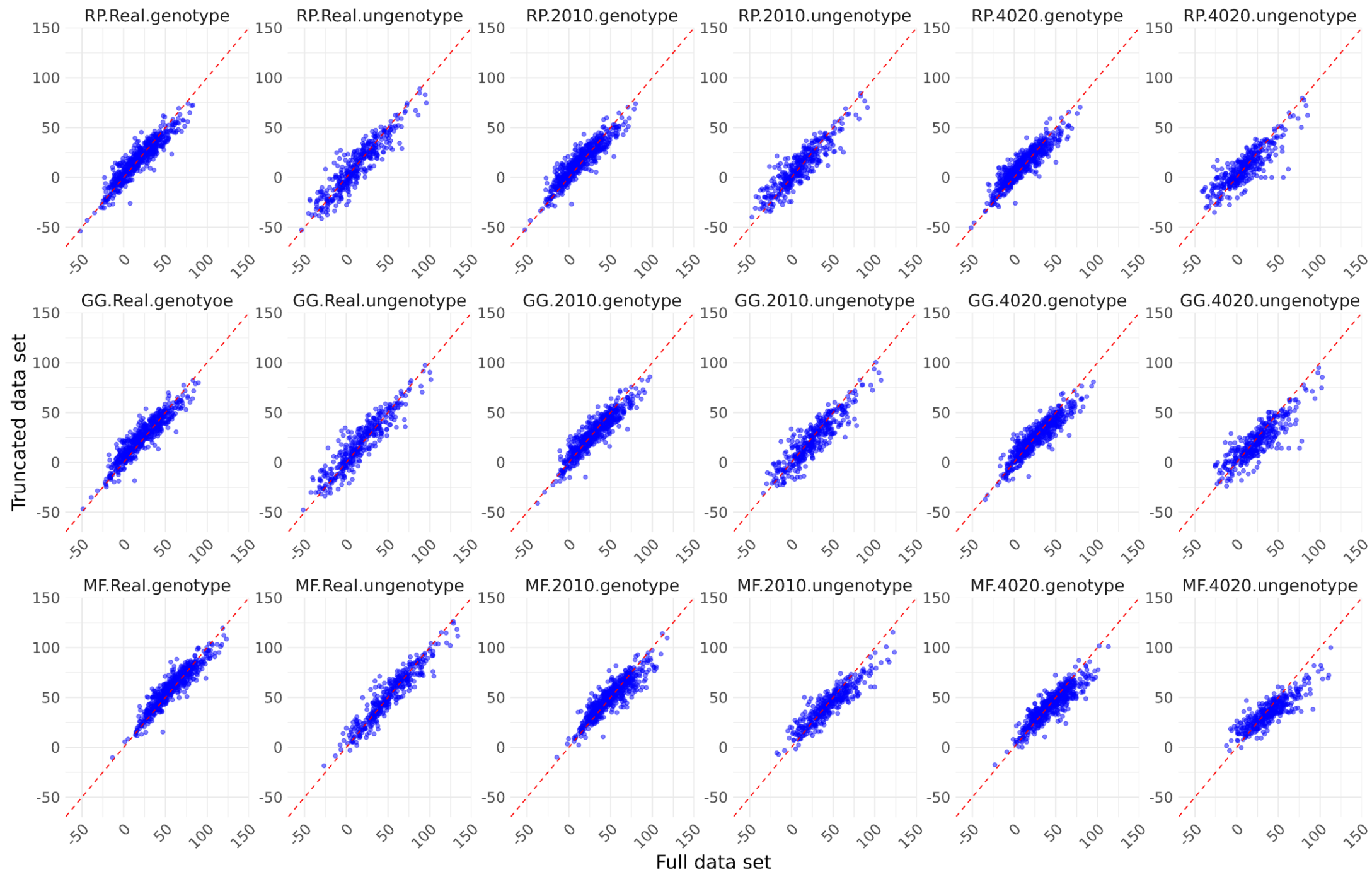


Comparison of GEBV (Full vs Truncated) for sires across scenarios

RP

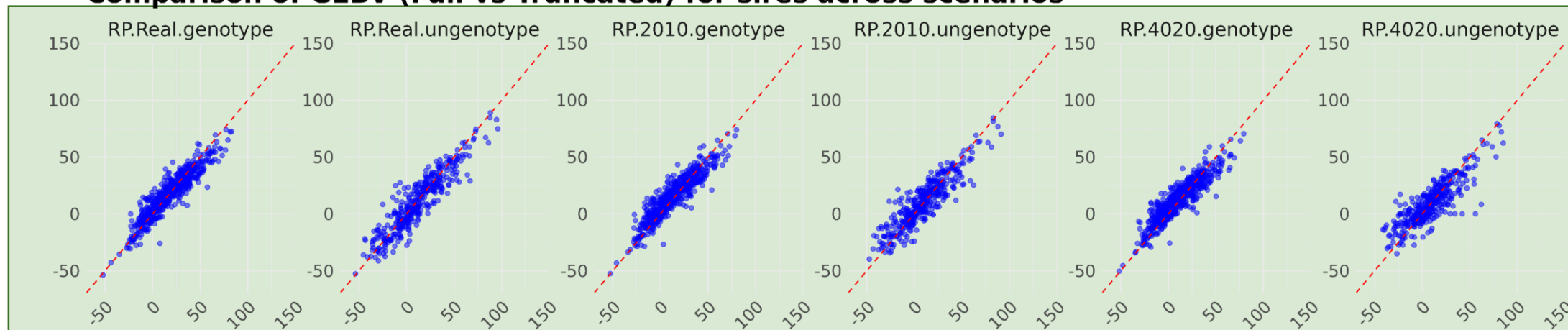
GG

MF

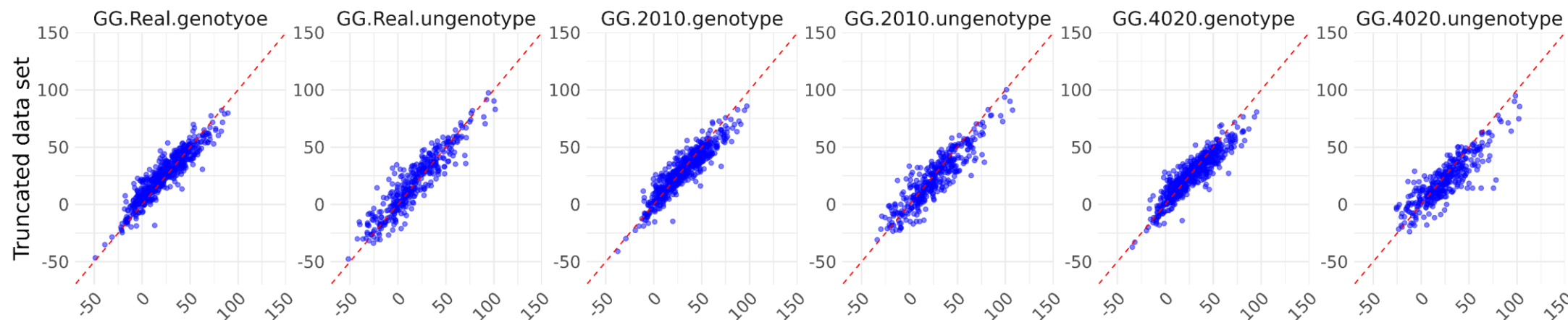


Comparison of GEBV (Full vs Truncated) for sires across scenarios

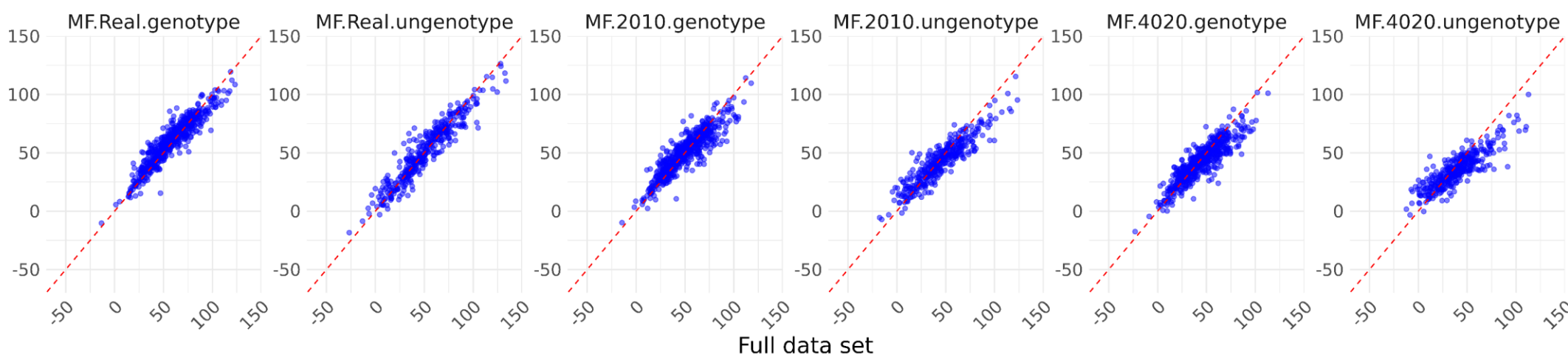
RP



GG

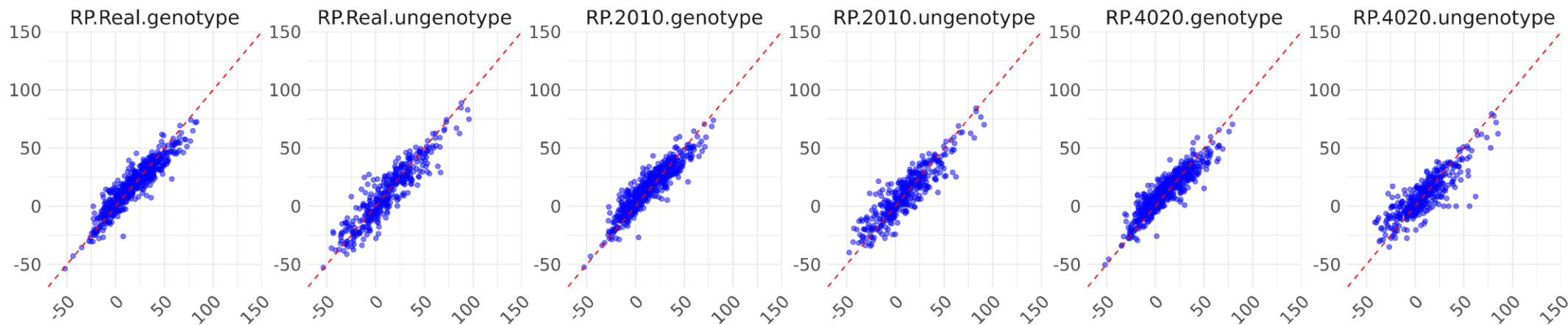


MF

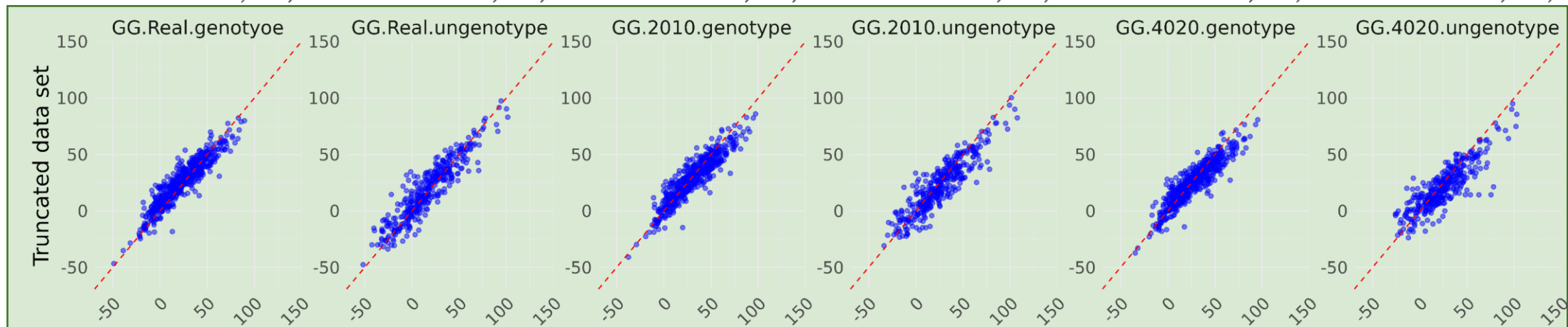


Comparison of GEBV (Full vs Truncated) for sires across scenarios

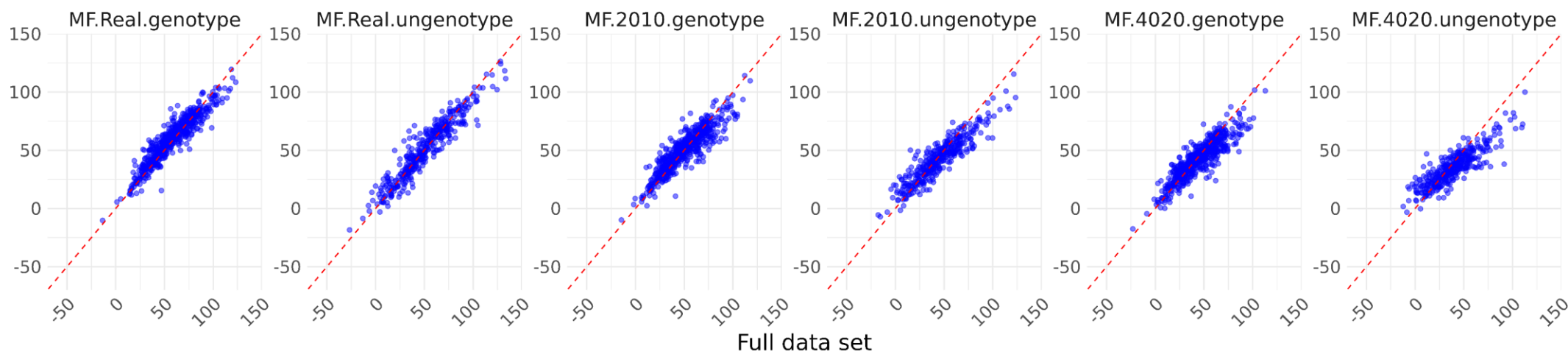
RP



GG



MF

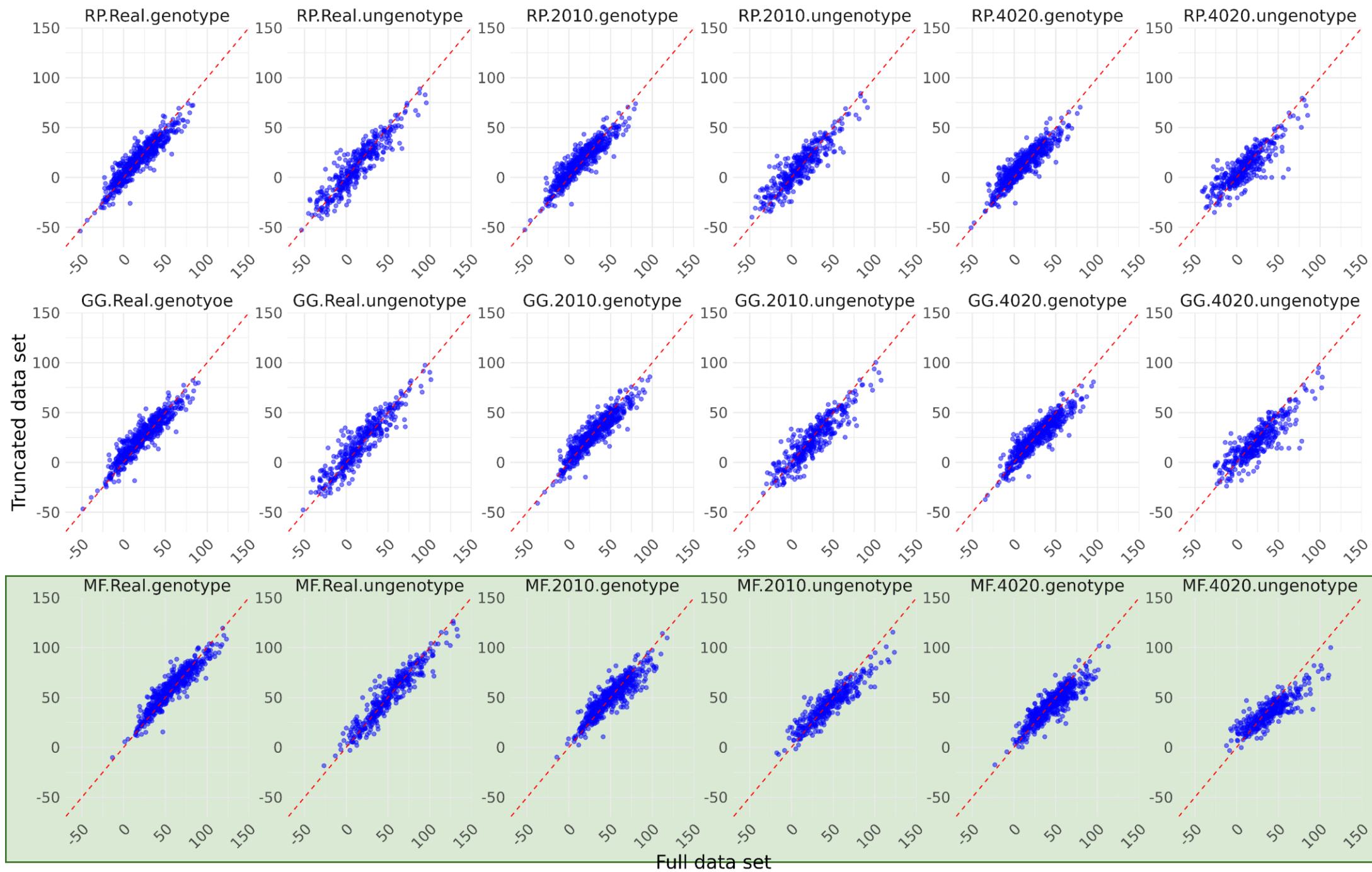


Comparison of GEBV (Full vs Truncated) for sires across scenarios

RP

GG

MF

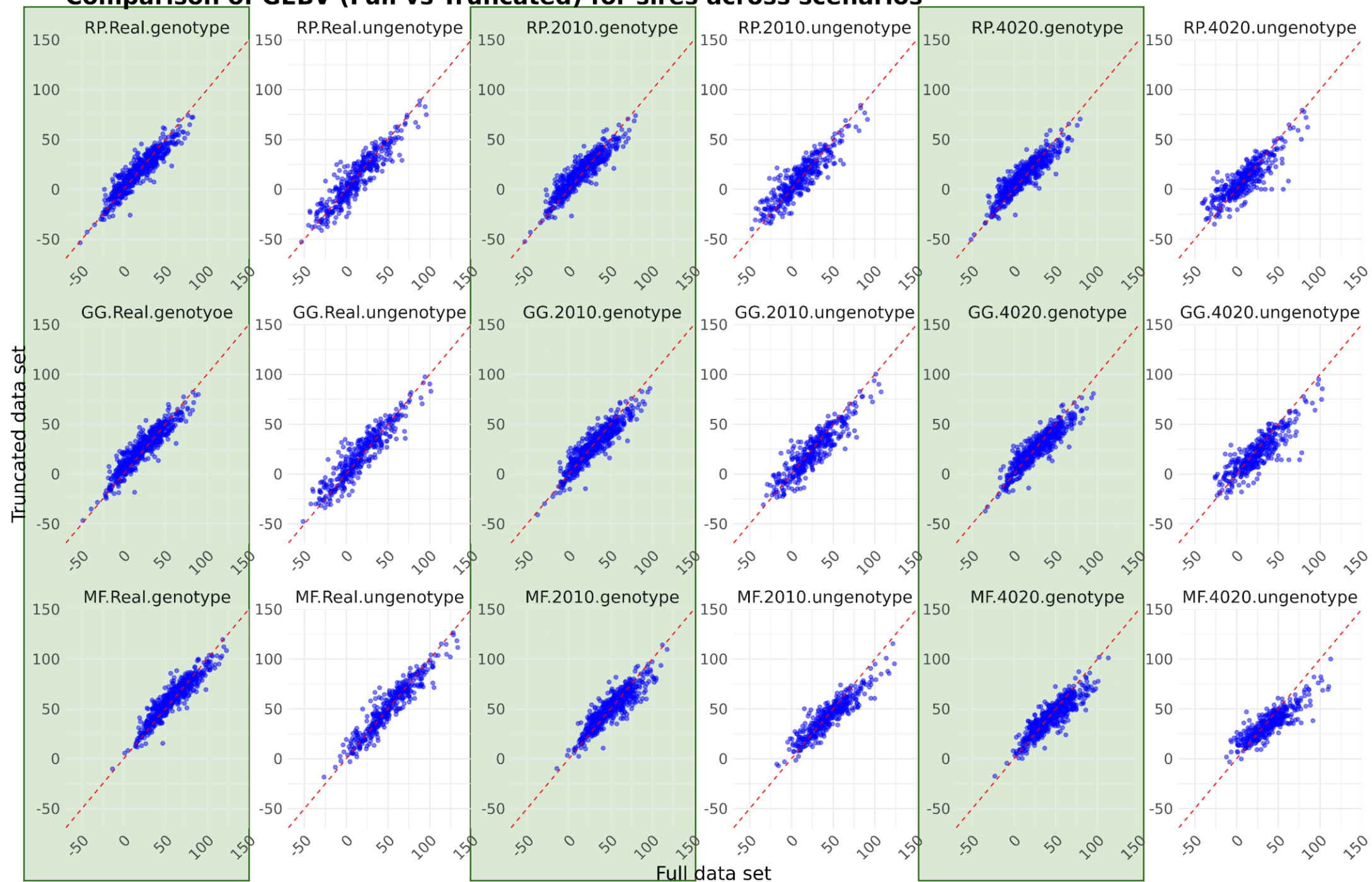


Comparison of GEBV (Full vs Truncated) for sires across scenarios

RP

GG

MF

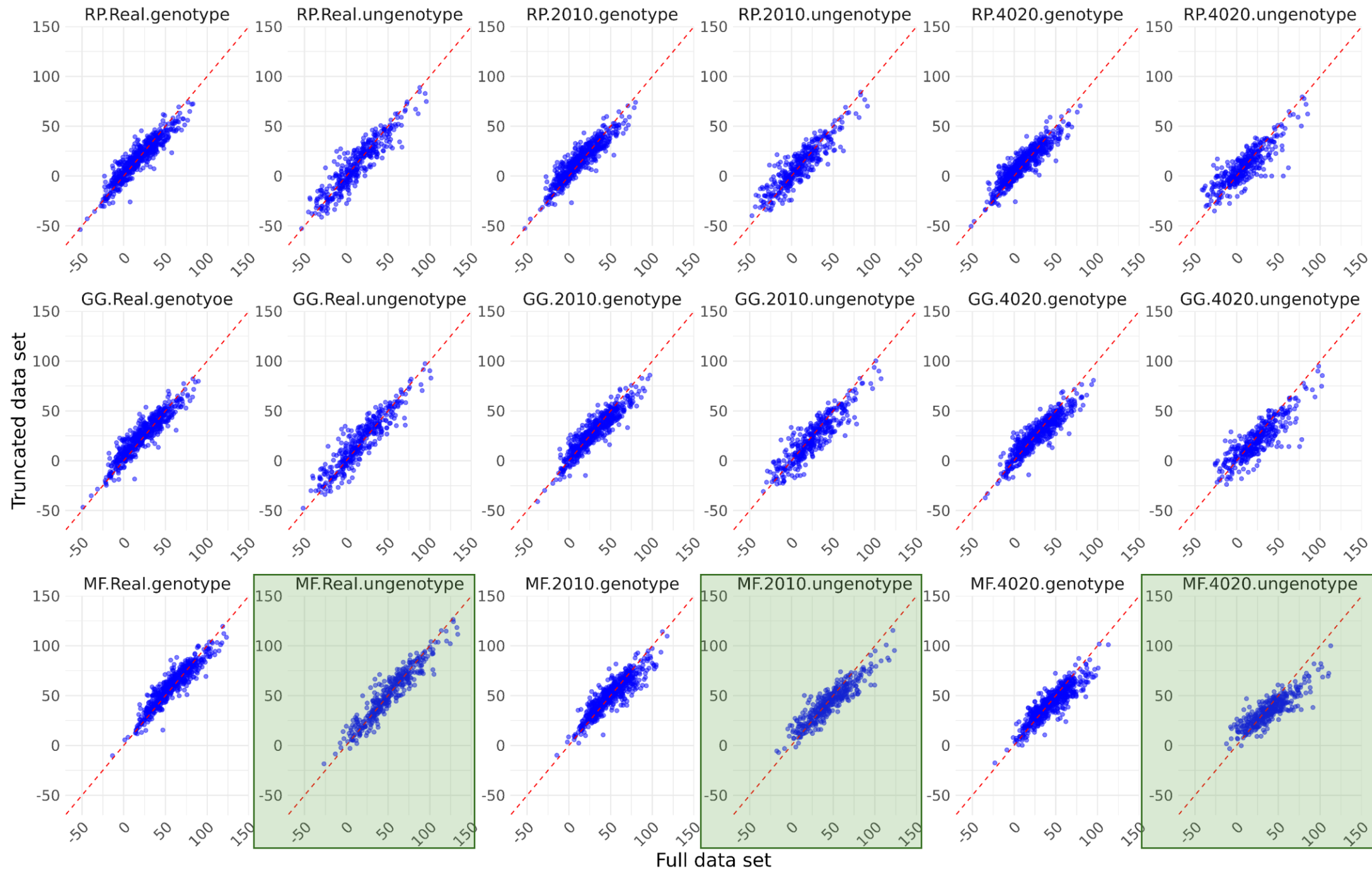


Comparison of GEBV (Full vs Truncated) for sires across scenarios

RP

GG

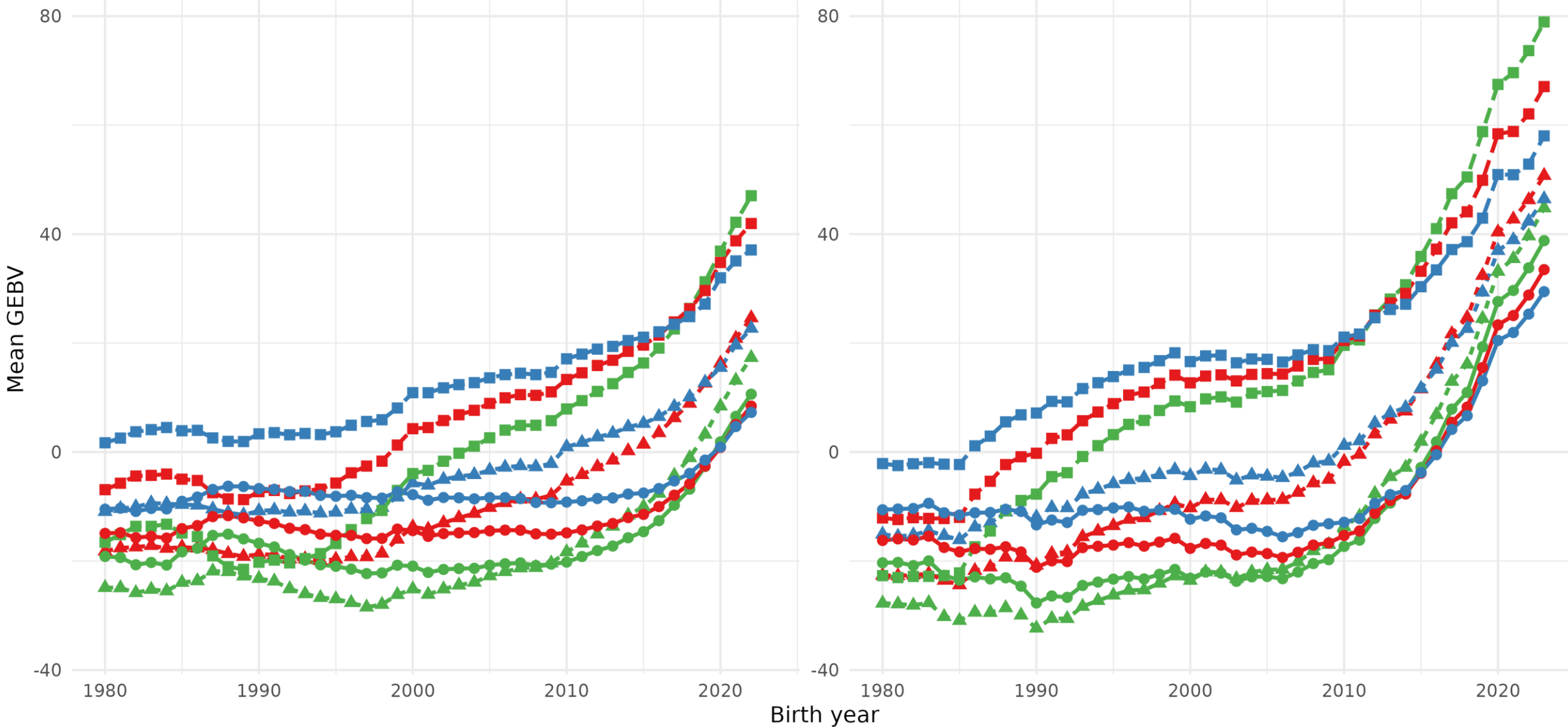
MF



Average GEBV trend

Dam

Sire



Deal with missing parents RP GG MF Pedigree scenario Real 2010 4020

Conclusions

- GEBV accuracy is impacted by the handling of missing parents (**GG, MF**)
- Missing parental information may result in overdispersion of GEBV
- Evaluation reliability is increased with more genotyped animals and more pedigree completeness
- Handling of missing parents is increasingly important with increasing incomplete pedigree
- **Metafounders > Genetic Groups > Raw Pedigree**



Thank You!



dawid.slomian@iz.edu.pl



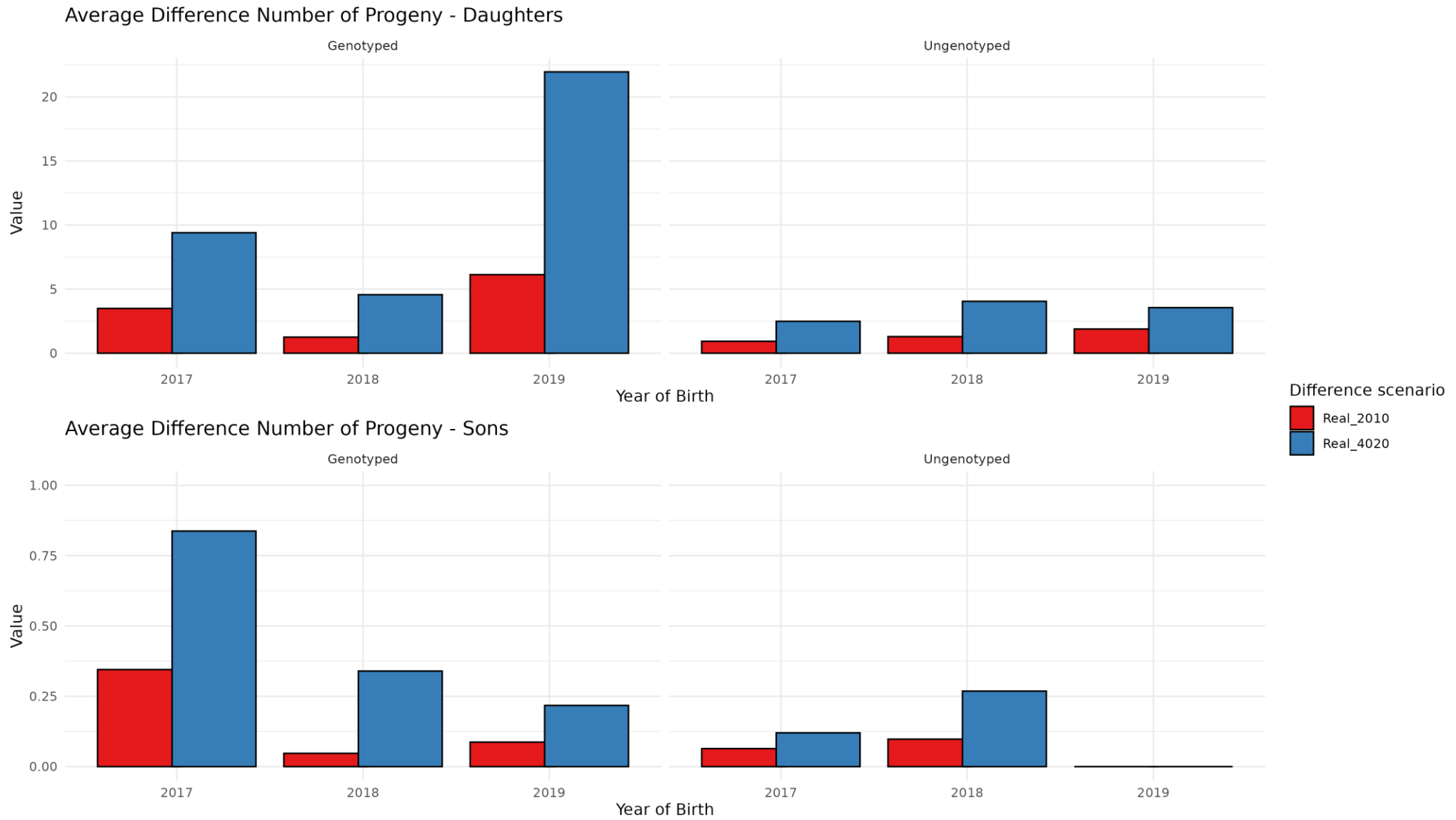
Unrelated genetic groups

- Created based on country of origin, birth of year and sex

Country	Year of birth	Male	Female
	< 1960	-99	-99
POL	1960-1969	-1	-2
USA/CAN	1960-1969	-3	-4
OTHERS	1960-1969	-5	-6
POL	1970-1979	-7	-8
USA/CAN	1970-1979	-9	-10
OTHERS	1970-1979	-11	-12
POL	1980-1989	-13	-14
USA/CAN	1980-1989	-15	-16
OTHERS	1980-1989	-17	-18
POL	1990-1999	-19	-20
USA/CAN	1990-1999	-21	-22
OTHERS	1990-1999	-23	-24
POL	2000-2009	-25	-26
USA/CAN	2000-2009	-27	-28
OTHERS	2000-2009	-29	-30
POL	2010-2019	-31	-32
USA/CAN	2010-2019	-33	-34
OTHERS	2010-2019	-35	-36
POL	2020-now	-37	-38
USA/CAN	2020-now	-39	-40
OTHERS	2020-now	-41	-42



Difference in number of progeny for validation bulls



Percent of missing parents per year divided by pedigree scenario and sex

