Can multi-subpopulation reference sets improve the genomic predictive ability for pigs?

Anna Fangmann[1], Sarah Bergfelder-Drüing[2], Ernst Tholen[2], Henner Simianer[1], Malena Erbe[1,3]

[1]Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Goettingen, Germany
[2]Institute of Animal Science, Group of Animal Breeding and Husbandry, University of Bonn, Germany
[3]Institute for Animal Breeding, Bavarian State Research Centre for Agriculture, Poing-Grub, Germany




Introduction

Cattle breeding

- Implementation of genomic prediction successful

Pig breeding

- Possible advantage of genomic prediction: increasing the accuracy of breeding values at the time point of selection

- For decades: separate breeding work of different pig breeding organizations in Germany, Switzerland and Austria

  → stratified subpopulations within breed German Large White

- Limiting factor: size of the training set within a breeding organization

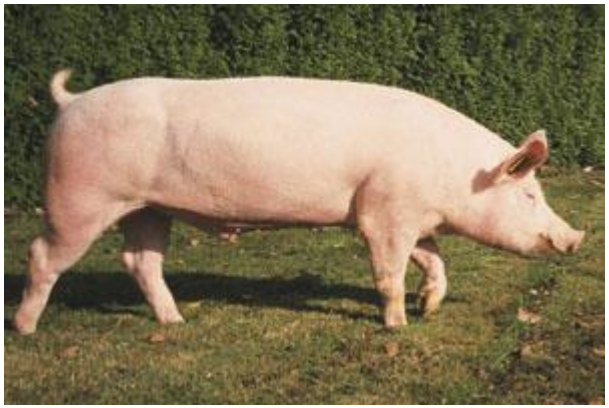
Aim of this study

- Evaluation of a genomic breeding value prediction in the breed German Large White for the trait ‘number of piglets born alive '
- Assessment of the usefulness of multi-subpopulation reference sets based on data from different commercial pig breeding organizations

http://www.bayerfarm.de/static/media/images/upload/2_schwein.jpg

Material and Methods: Data

- Data from individuals of five different commercial pig breeding organizations → different subpopulations
- 2’251 individuals genotyped with Illumina Porcine 60k SNP Chip
- Conventional breeding values for ‚number of piglets born alive‘ (NBA) → deregressed following Garrick et al. (2009)



http://www.gfs-topgenetik.de/images/eber/20080218.jpg

Material and Methods: Genotypes

- Quality control: Callrate per SNP > 97 %
  Callrate per individual > 98 %
  > 10 observations of an allele per marker

→ Finally: 2'053 individuals with 46'064 SNPs

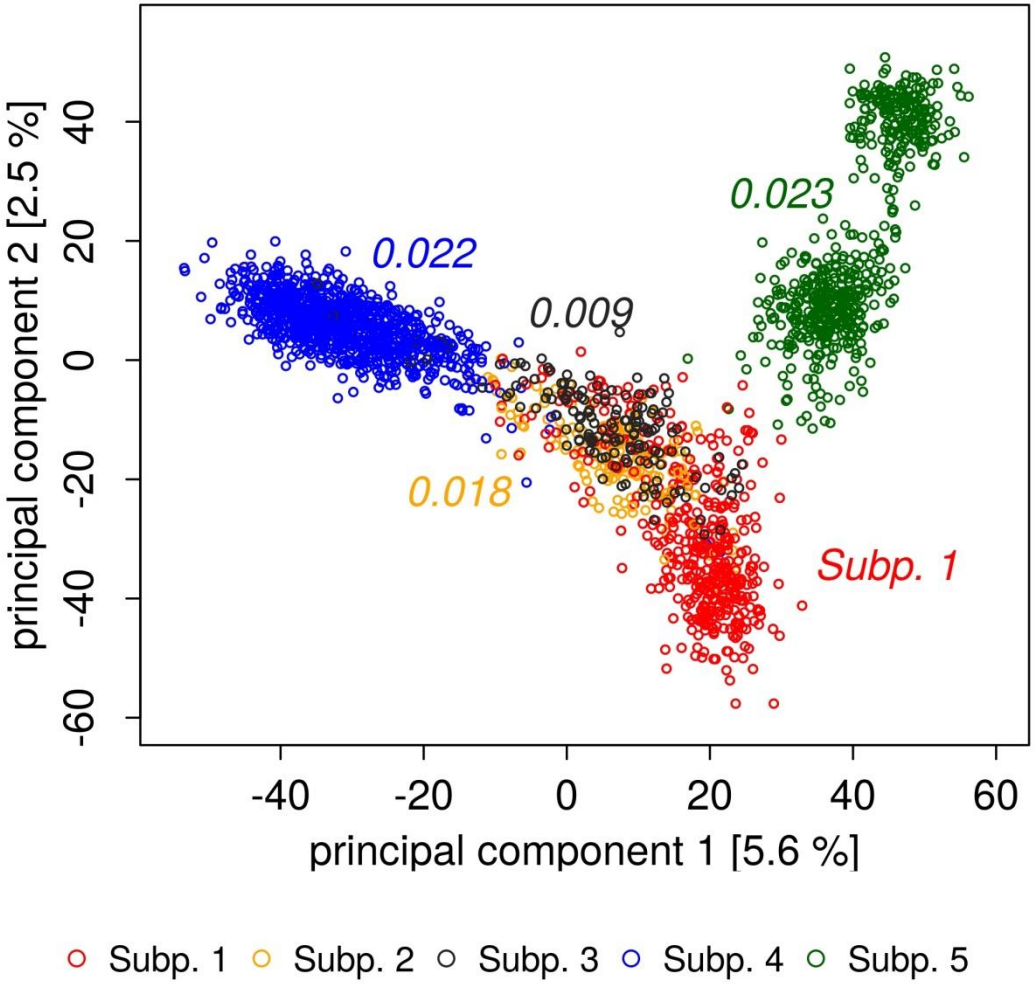
- Genotypic data:

Subpopulation	Total number of animals	Born between
1	187	2002 – 2011
2	140	1997 – 2011
3	155	2001 – 2011
4	821	1993 – 2011
5	540	2002 – 2011

Validation sets

Material and Methods: Subpopulation stratification

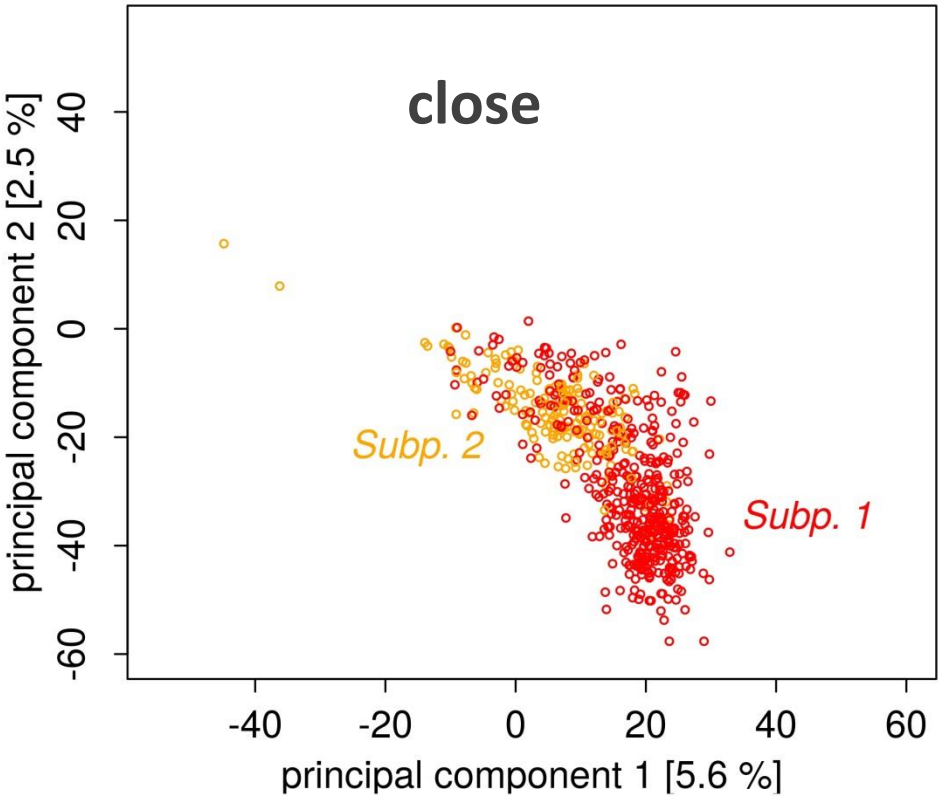
Assessment based on principal component analysis and calculated F_{ST} values between subpopulation1 and another subpopulations

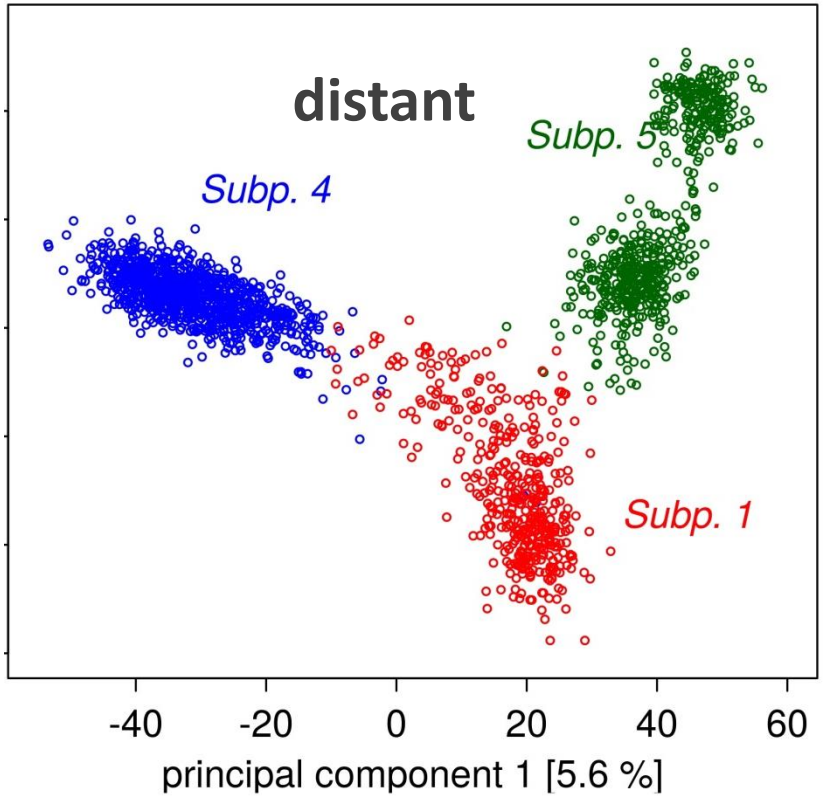


Material and Methods: Subpopulation stratification

Multi-subpopulation reference sets for validation set:

Subpopulation 1





Material and Methods: GBLUP model

Genomic Predictions with ASReml (Gilmour et al., 2009):

$$y = Xb + Wg + e$$

y = vector of DRPs for NBA

X = design matrix for fixed effects

b = vector containing the fixed effects

a) within subpopulation: overall mean

b) multi-subpopulation: general mean and subpopulation

W = design matrix for the random genomic effects

g = vector of random genomic effects (DGV)

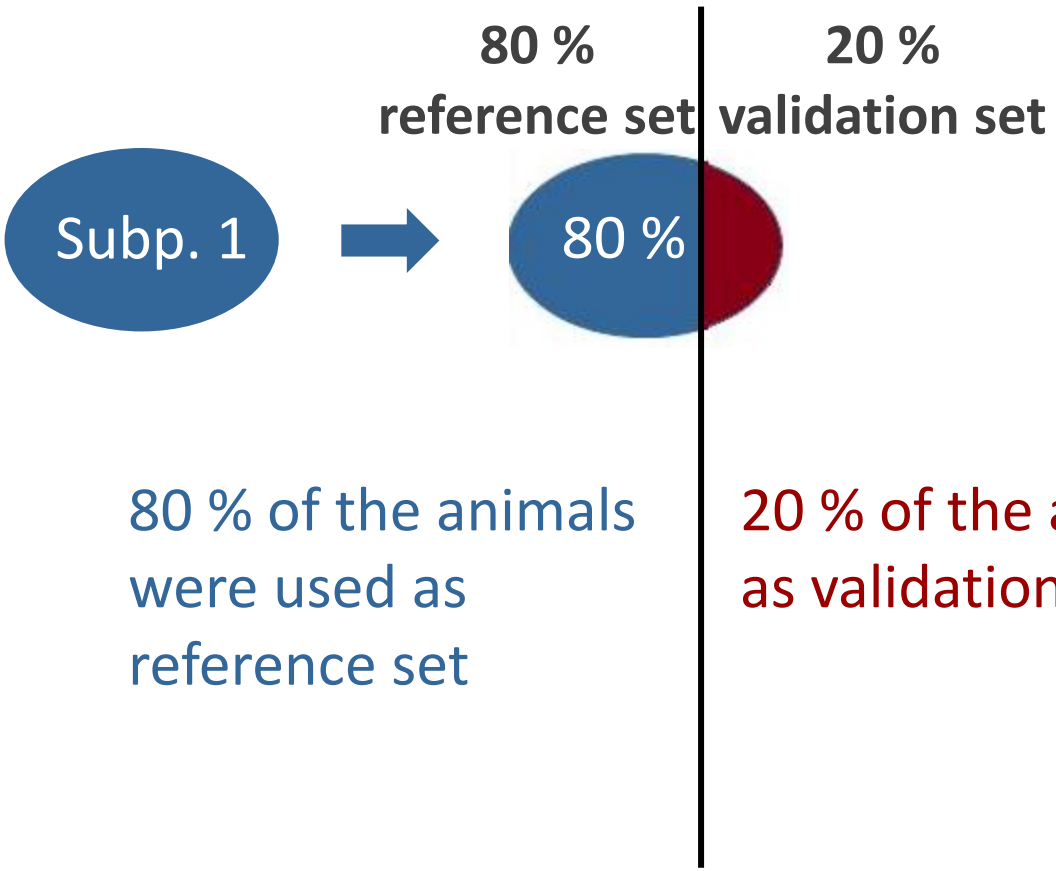
e = vector of random residual effects

with $g \sim N(0, G_x \sigma^2_g)$ and G_x = Genomic relationship matrix according to different approaches

Material and Methods: random five-fold cross validation

Assessment of predictive ability of DGV prediction

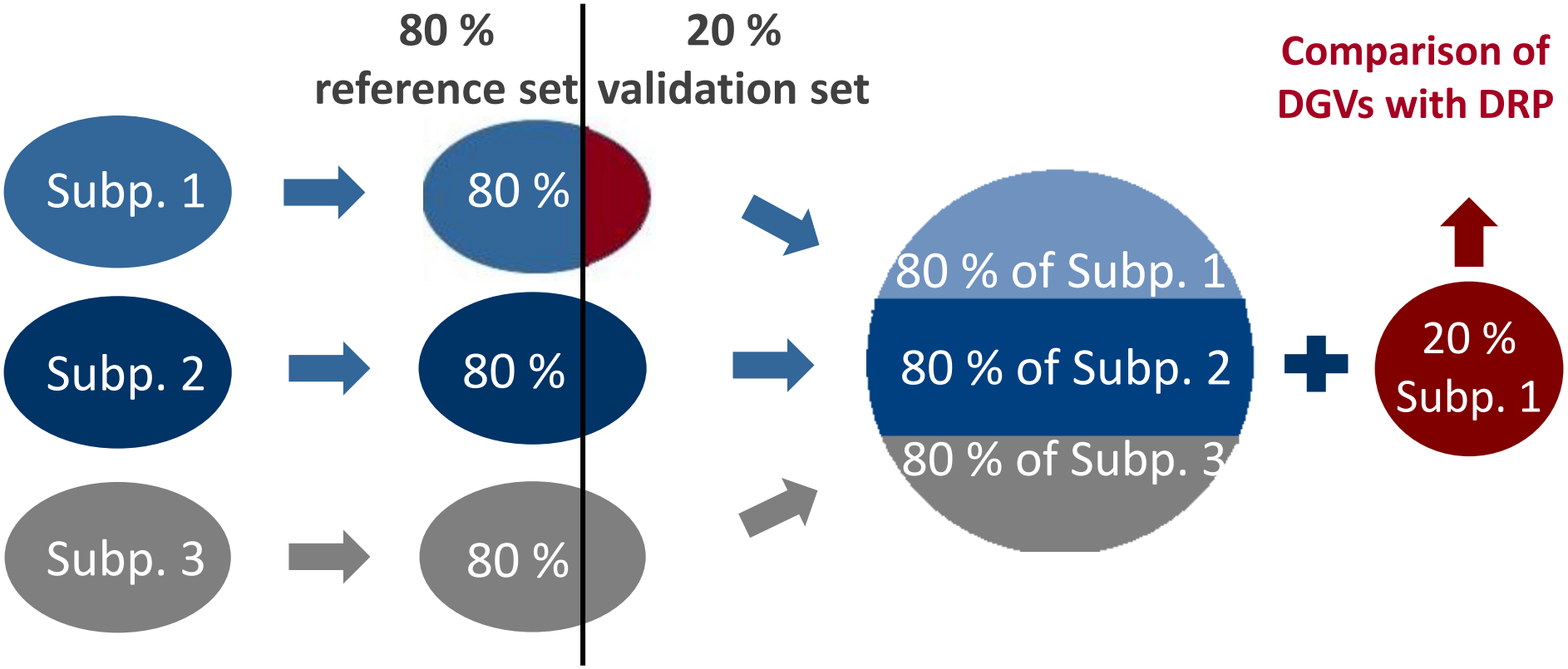
a) Within subpopulation:



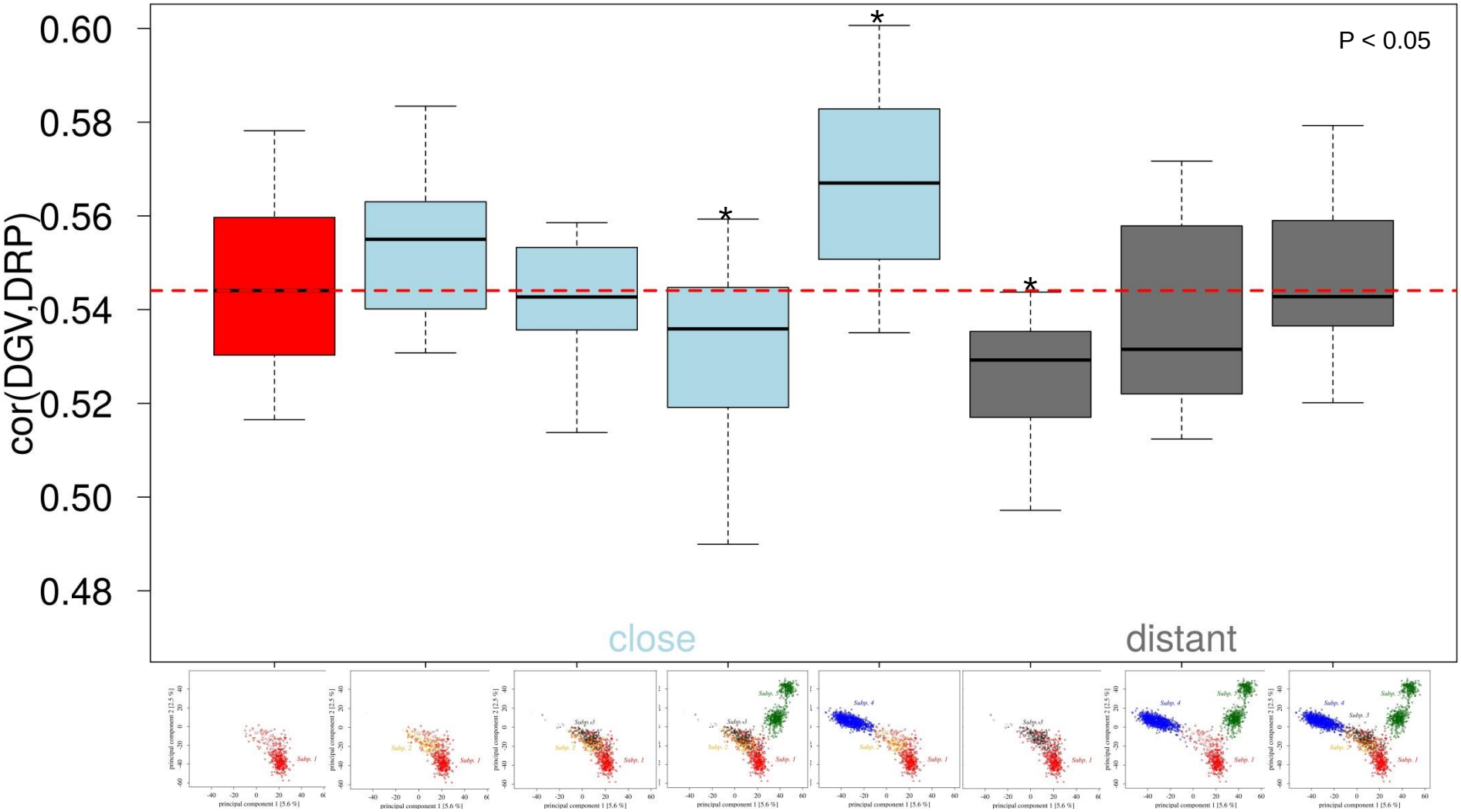
Material and Methods: random five-fold cross validation

Assessment of predictive ability of DGV prediction

b) Multi-subpopulation:



Results: random five-fold cross validation

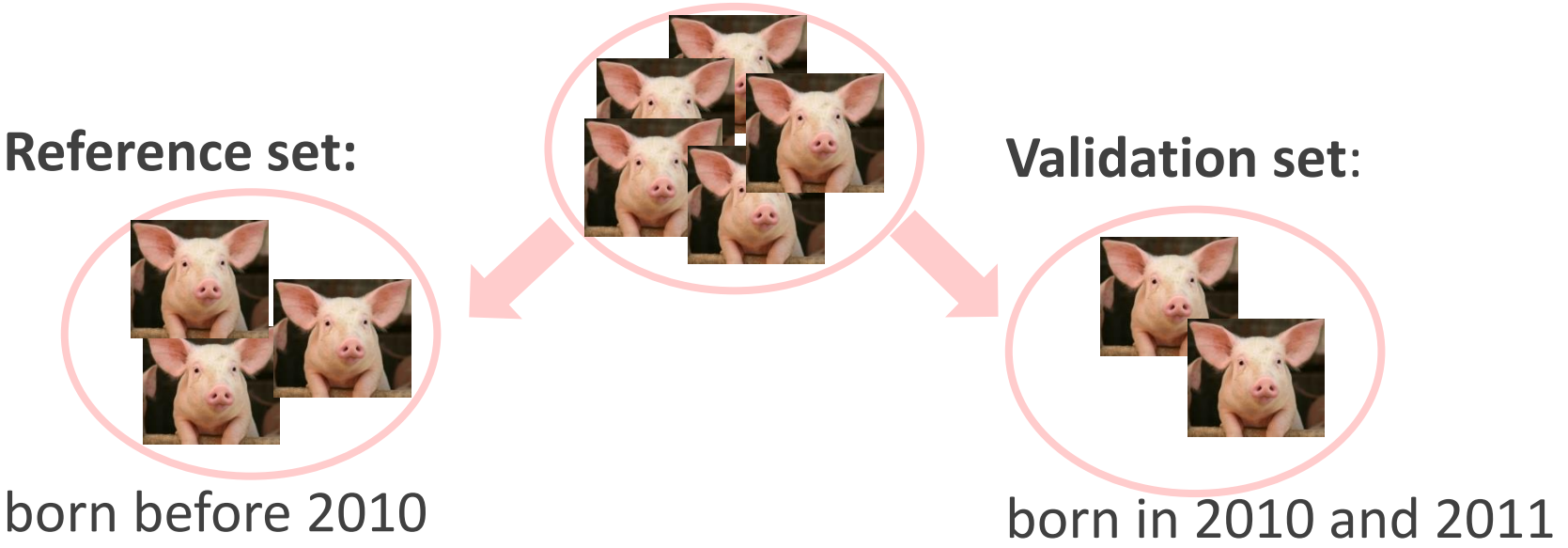


Predictive ability with **DRP** and $\mathbf{G_{VanRaden}}$ exemplary for **subpopulation 1**

Material and Methods: Forward Prediction

Assessment of predictive ability of DGV prediction

1. Within and multi-subpopulation:



Material and Methods: Forward Prediction

Assessment of predictive ability of DGV prediction

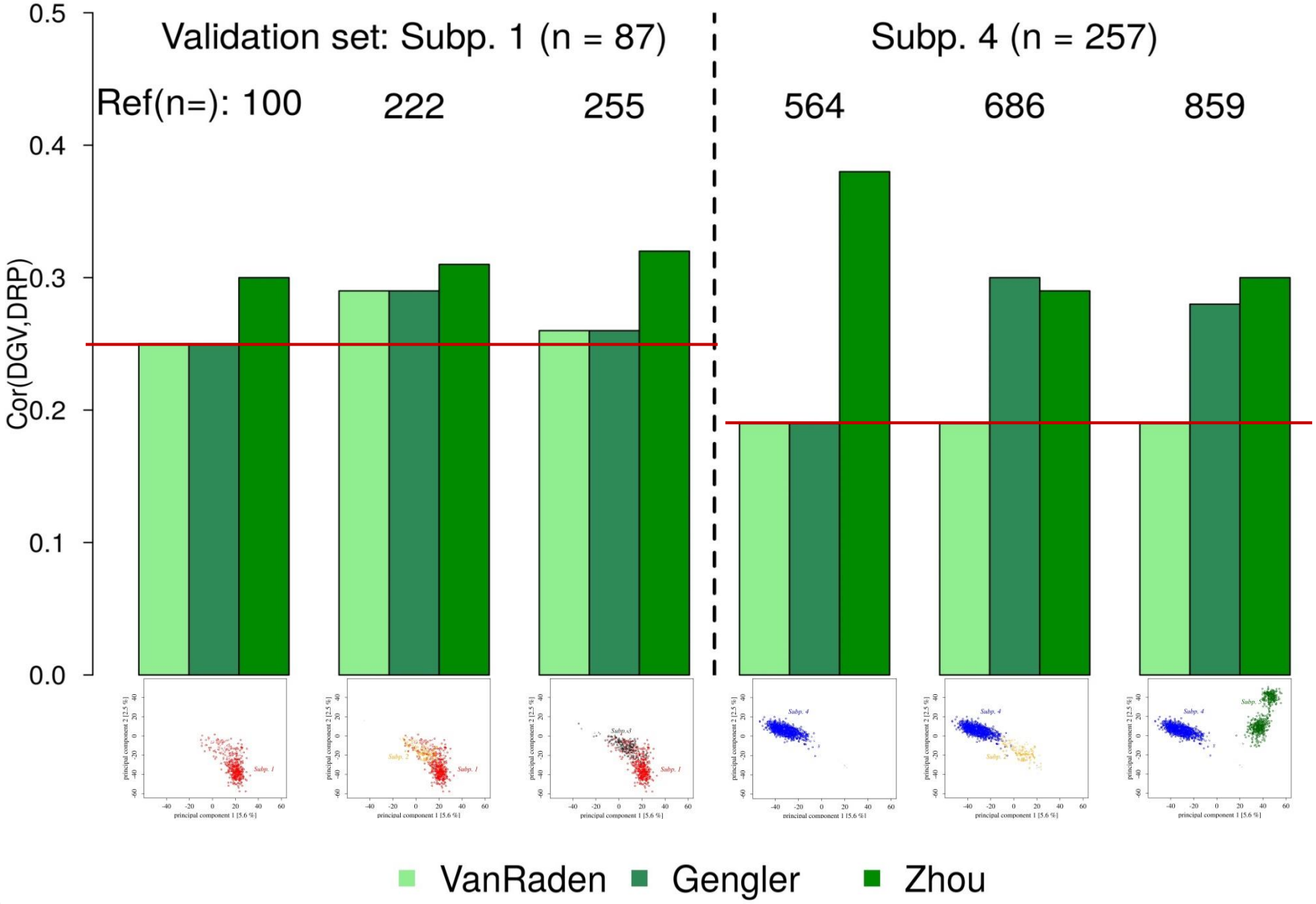
1. Within and multi-subpopulation

2. Effect of different G matrices:

   - G introduced by **VanRaden** (2007)
     - with actual allele frequencies over total set of individuals
     - with founder allele frequencies (Gengler et al., 2007) per subpopulation

   - G introduced by **Zhou et al.** (2014)
     - accounting for substructure by including information of marker effects (estimated from reference set) and linkage disequilibrium

Results: Forward Prediction



Conclusions

- ✓ 5-fold CV: Decrease (slight decrease) in predictive ability for distantly (closely) related multi-subpopulation reference sets
- ✓ Forward prediction:
  - ✓ Slight increase in predictive ability, especially by adding subpopulation 2 to the reference set
  - ✓ Slight increase in predictive ability when using different G matrices, especially when accounting for substructures
- ✓ **Forming** a multi-subpopulation reference population **generally did not lead** to a better predictive ability for individuals within a specific subpopulation
- ✓ Necessity to **creating more concurrent links** between subpopulations, e.g. by using the same boars across populations

Thank you for your Attention!

Acknowledgment:
The author gratefully acknowledge the financial support of the project ‘pigGS’ by the Europäischen Fonds für regionale Entwicklung (EFRE),
the state of North Rhine-Westphalia,
the project management Jülich and
the pig breeding organizations for providing the data.

We especially thank the EAAP for granting a scholarship for the EAAP Meeting 2015 in Warsaw.




References

Garrick, D.J., Taylor, J.F. und Fernando, R.L. (2009): Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41:55.

Gengler, N., P. Mayeres, and M. Szydlowski (2007): A Simple Method to Approximate Gene Content in Large Pedigree Populations: Application to the Myostatin Gene in Dual-Purpose Belgian Blue Cattle. Animal 1.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., und Thompson, R. (2009): ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

VanRaden, P.M. (2007): Genomic Measures of Relationship and Inbreeding. Interbull Bull 37: 33.

Wright , S. (1943): Isolation by distance. Genetics 28: 139-156.

Zhou, L, M.S. Lun, Y. Wang und G. Su (2014): Genomic predictions across Nordic Holstein and Nordic Red using the genomic best linear unbiased prediction model with different genomic relationship matrices. J. Anim. Breed. Genet. 131: 249–257.