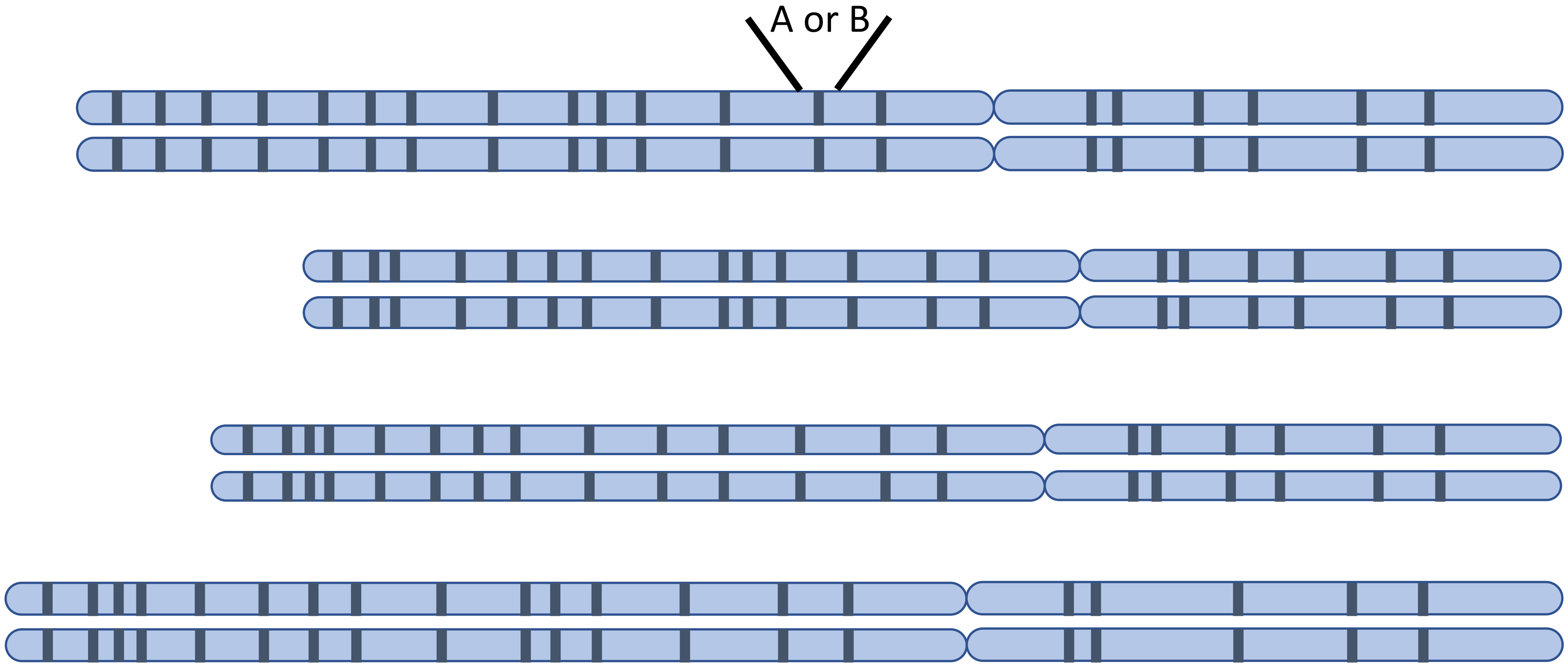
Failed SNP Calls: Useful for Genomic Prediction in Plant Breeding?

Sven Weber

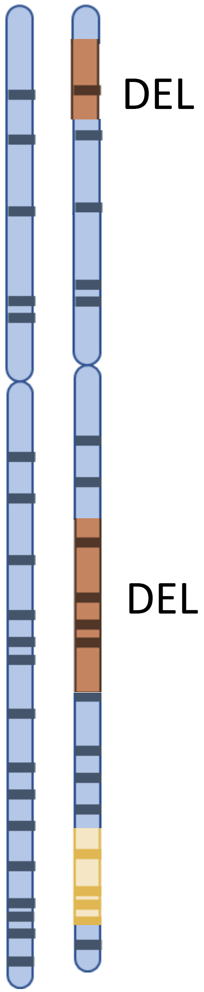
Justus Liebig University Giessen, Germany

Department of Plant Breeding

Single Nucleotide Polymorphism (SNP)



Structural Variations



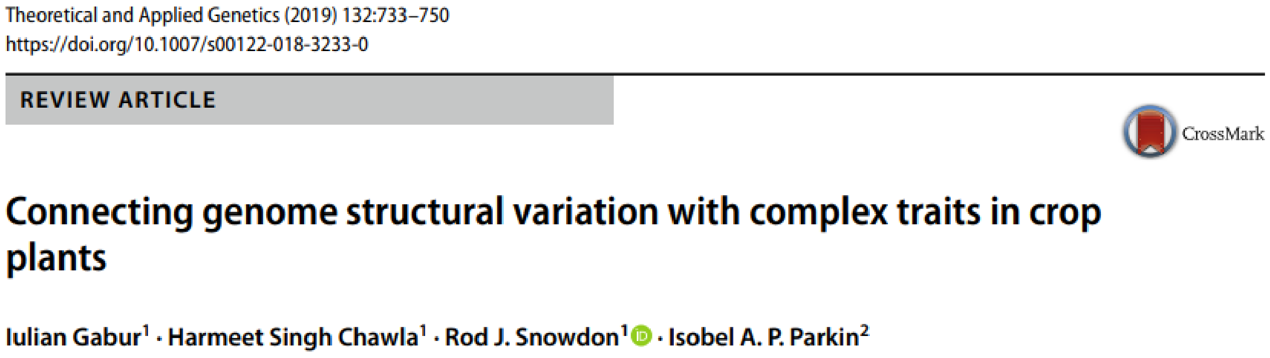
Detectable with UpToDate whole genome sequencing

INV

BreedPatH

Giessen Plant Breeding

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN



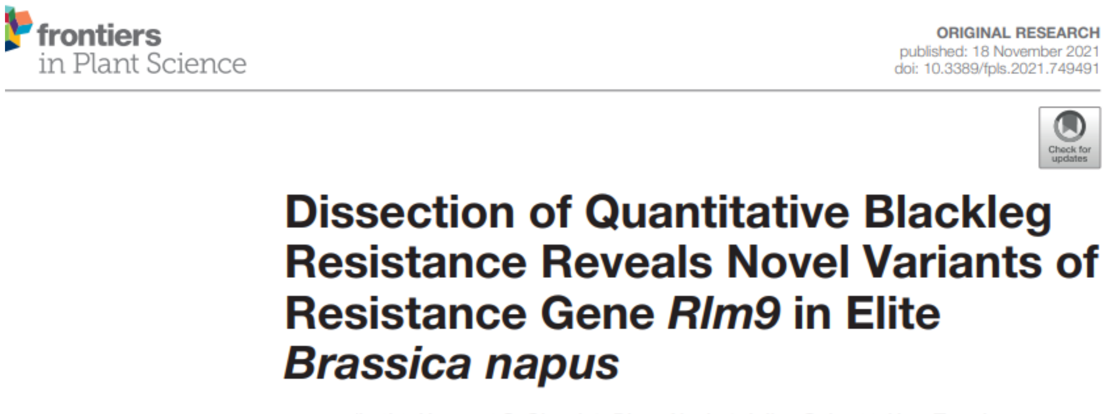
Theoretical and Applied Genetics (2019) 132:733–750
https://doi.org/10.1007/s00122-018-3233-0

REVIEW ARTICLE

CrossMark

Connecting genome structural variation with complex traits in crop plants

Iulian Gabur[1] · Harmeet Singh Chawla[1] · Rod J. Snowdon[1] · Isobel A. P. Parkin[2]



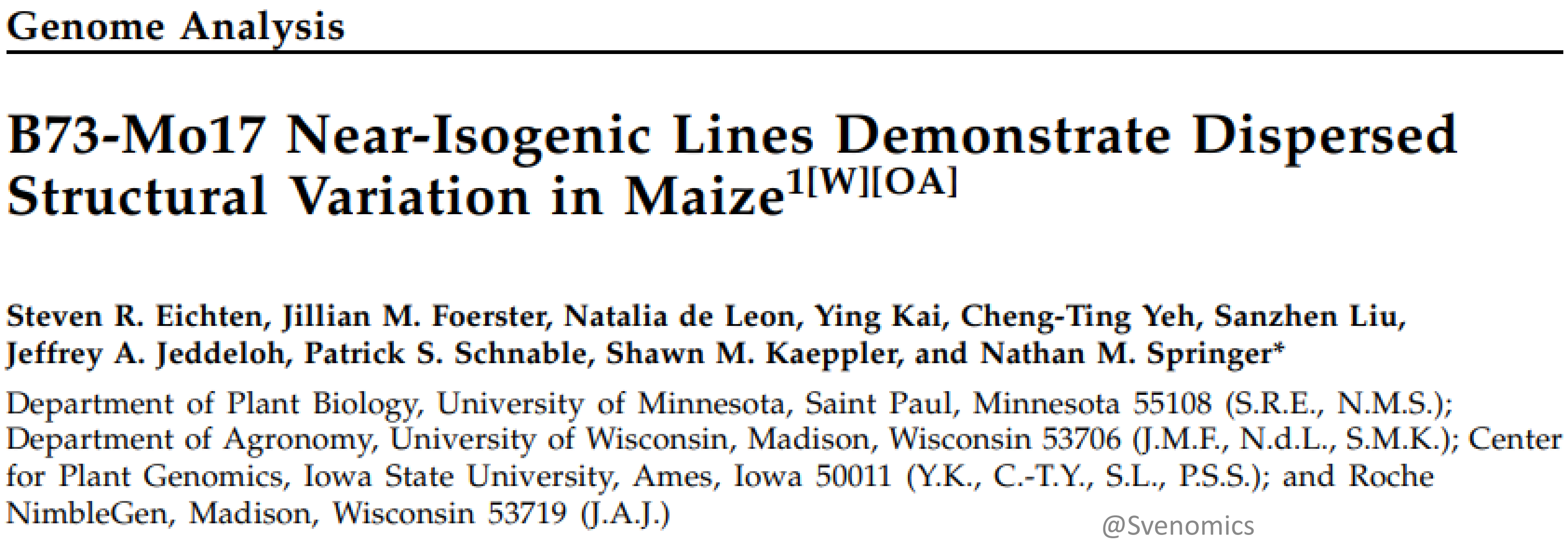
frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fpls.2021.749491

Check for updates

Dissection of Quantitative Blackleg Resistance Reveals Novel Variants of Resistance Gene *Rlm9* in Elite *Brassica napus*

lrath[1], Harmeet S. Chawla[2], Dima Alnajar[3], Iulian Gabur[1,4], HueyTyng Lee[1],
eber[1], Lennard Ehrig[1], Birger Koopmann[3], Rod J. Snowdon[1] and

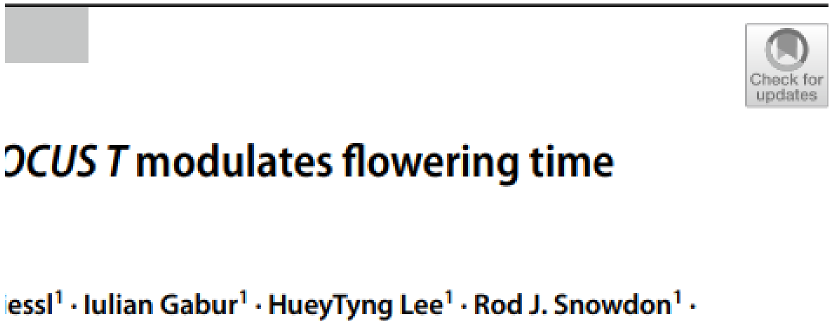


Genome Analysis

B73-Mo17 Near-Isogenic Lines Demonstrate Dispersed Structural Variation in Maize[1[W][OA]]

Steven R. Eichten, Jillian M. Foerster, Natalia de Leon, Ying Kai, Cheng-Ting Yeh, Sanzhen Liu, Jeffrey A. Jeddeloh, Patrick S. Schnable, Shawn M. Kaeppler, and Nathan M. Springer*

Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108 (S.R.E., N.M.S.); Department of Agronomy, University of Wisconsin, Madison, Wisconsin 53706 (J.M.F., N.d.L., S.M.K.); Center for Plant Genomics, Iowa State University, Ames, Iowa 50011 (Y.K., C.-T.Y., S.L., P.S.S.); and Roche NimbleGen, Madison, Wisconsin 53719 (J.A.J.)

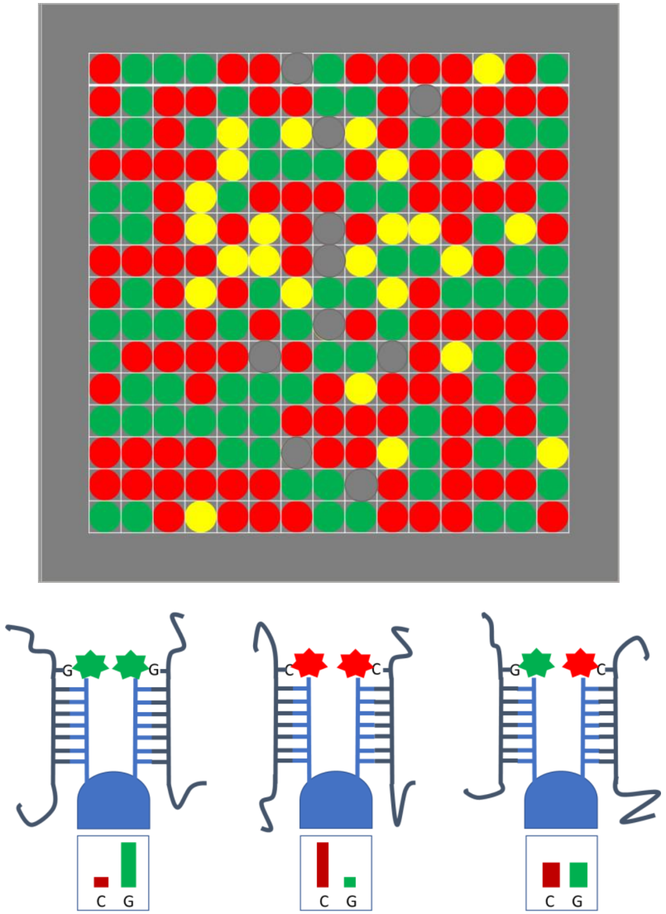


Check for updates

OCUS T modulates flowering time

essl[1] · Iulian Gabur[1] · HueyTyng Lee[1] · Rod J. Snowdon[1] ·

SNP array



$$
\begin{bmatrix}
0 & 2 & 0 & 2 & 1 & NA & 0 & 2 & 2 & 1 \\
2 & 0 & NA & 2 & NA & 0 & 2 & 0 & 2 & NA \\
NA & 0 & 1 & NA & 2 & NA & 1 & 0 & 2 & 2 \\
2 & 0 & 1 & 0 & 2 & 1 & 0 & 2 & 2 & 0 \\
0 & NA & 0 & 2 & 0 & 0 & 1 & 1 & 2 & 0 \\
2 & 1 & 1 & 0 & 2 & NA & NA & NA & 0 & 2 \\
1 & NA & 2 & 0 & 1 & NA & 0 & 2 & NA & 1 \\
1 & 0 & 2 & 2 & 0 & 2 & 2 & 0 & 1 & 2 \\
2 & 0 & 2 & 2 & 2 & NA & 1 & 0 & 0 & 1 \\
2 & 0 & 2 & 2 & 1 & 2 & 0 & 2 & 2 & 0
\end{bmatrix}
$$

SNP array

How predictive are failed marker calls in genomic prediction?

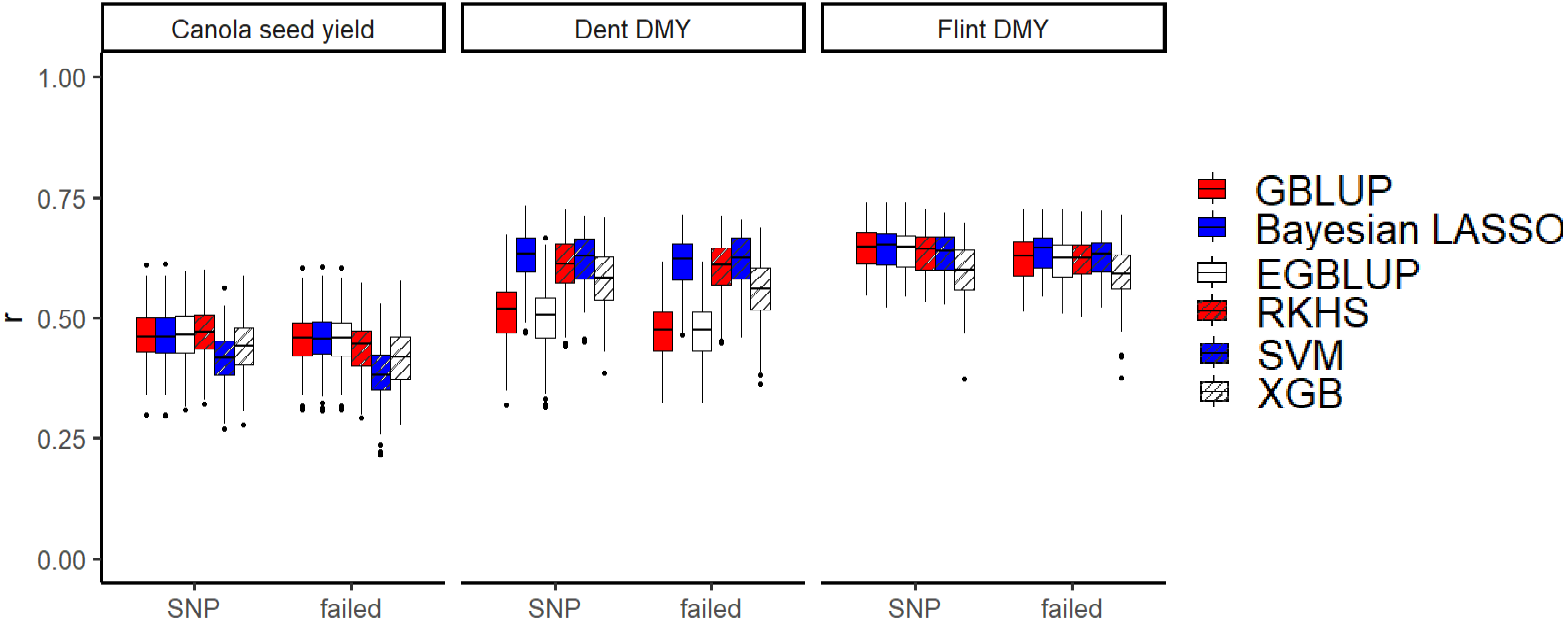
Can the addition of failed marker information to standard SNPs improve genomic prediction accuracy?

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Datasets

- Canola Jan HU, Abbadi A, Lücke S, et al (2016) PLOS ONE 11:e0147769. https://doi.org/10.1371/journal.pone.0147769
    - 457 lines x 2 MS Lines
    - 910 hybrids
    - Illumina *Brassica* 60 k SNP array
- Maize Lehermeier C, Krämer N, Bauer E, et al (2014) Genetics 198:3–16. https://doi.org/10.1534/genetics.114.161943
    - 847 Dent lines
    - 918 Flint lines
    - Illumina MaizeSNP50 SNP array

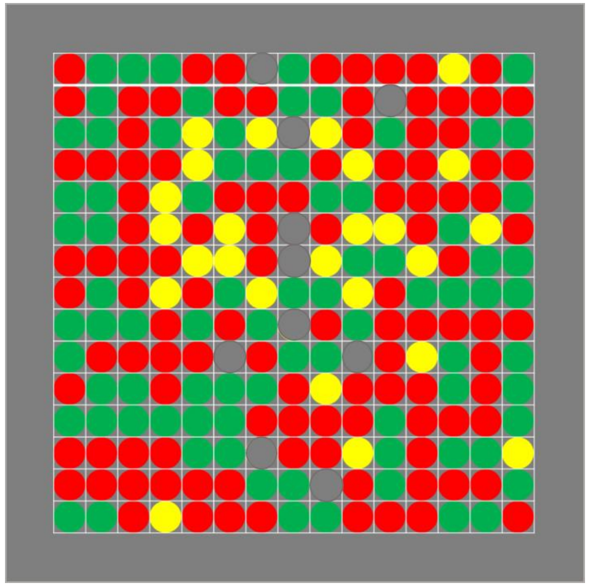
Genomic Prediction

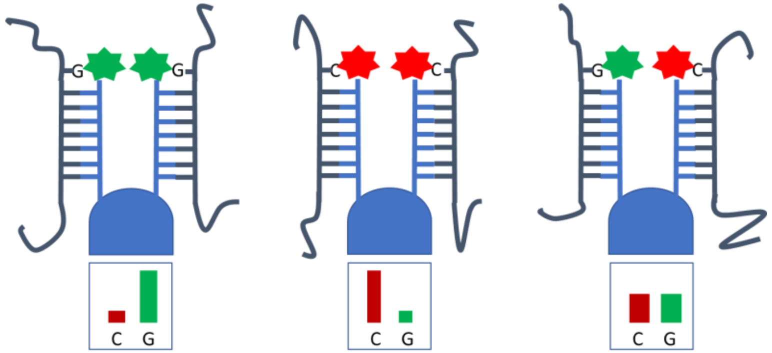


Random CV 80%/20%, 150 repetitions

Technical errors

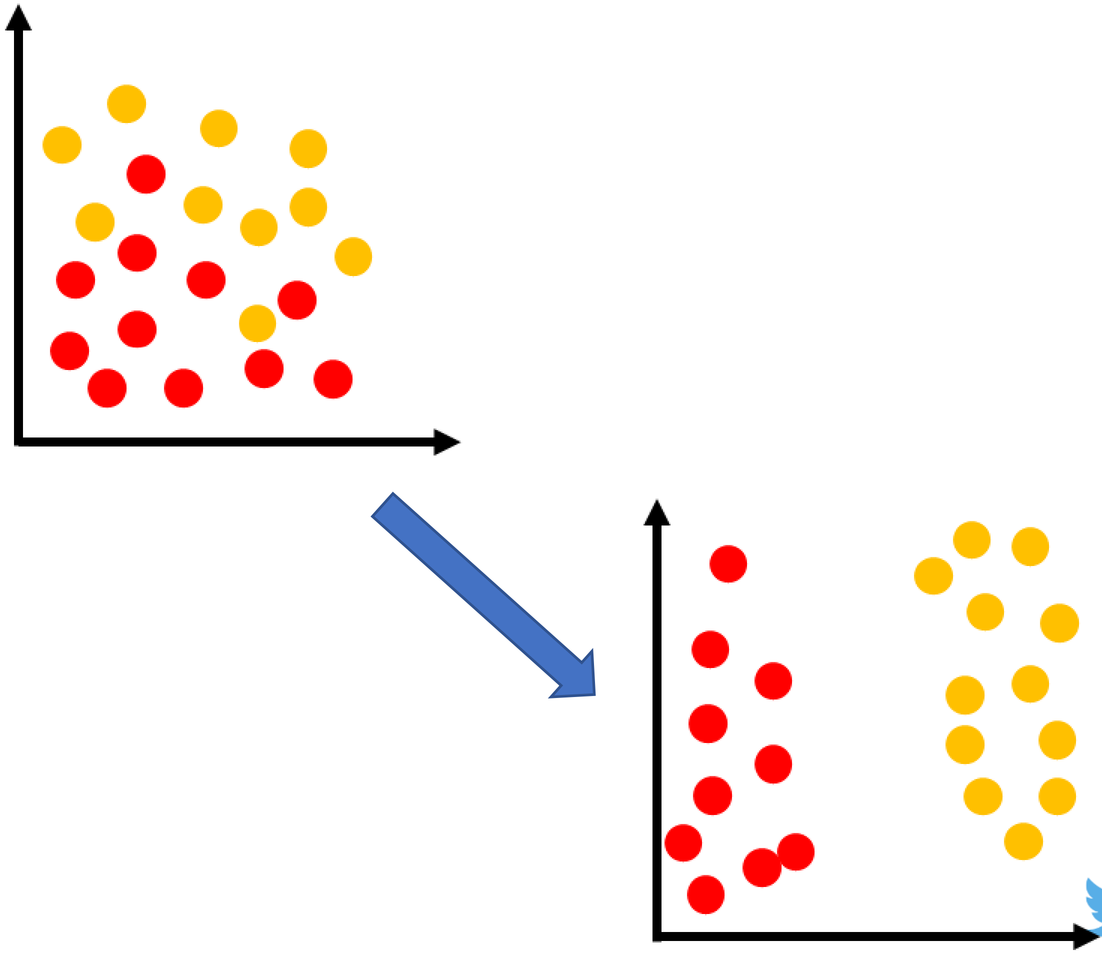
- This may mask valuable information
- How to distinguish them?



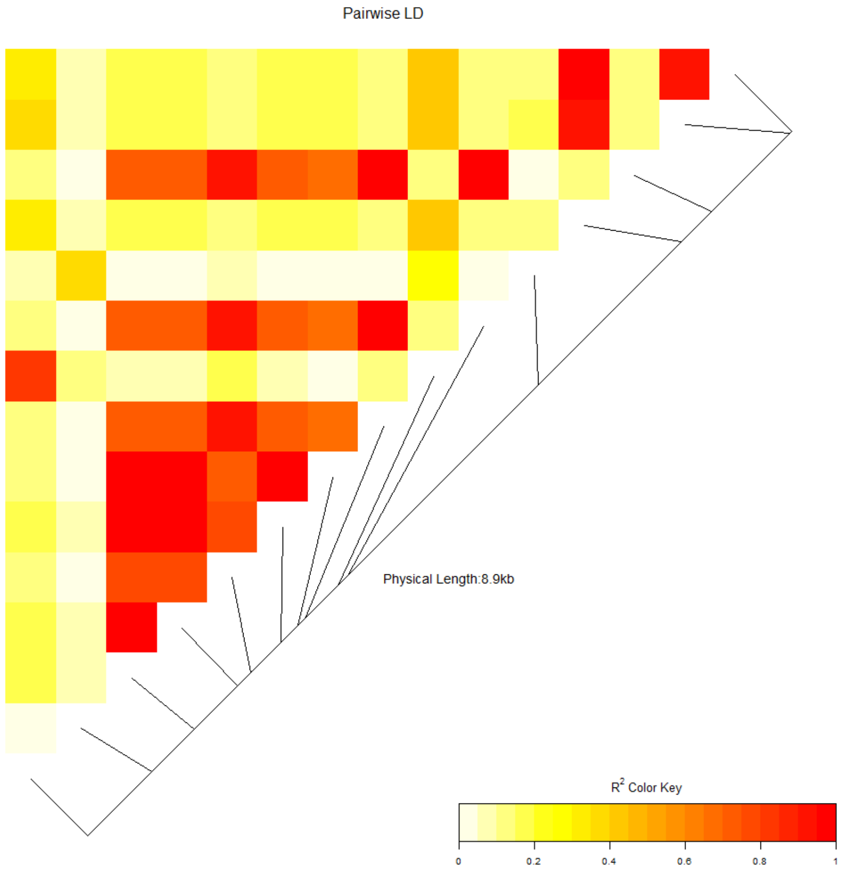


Recovery

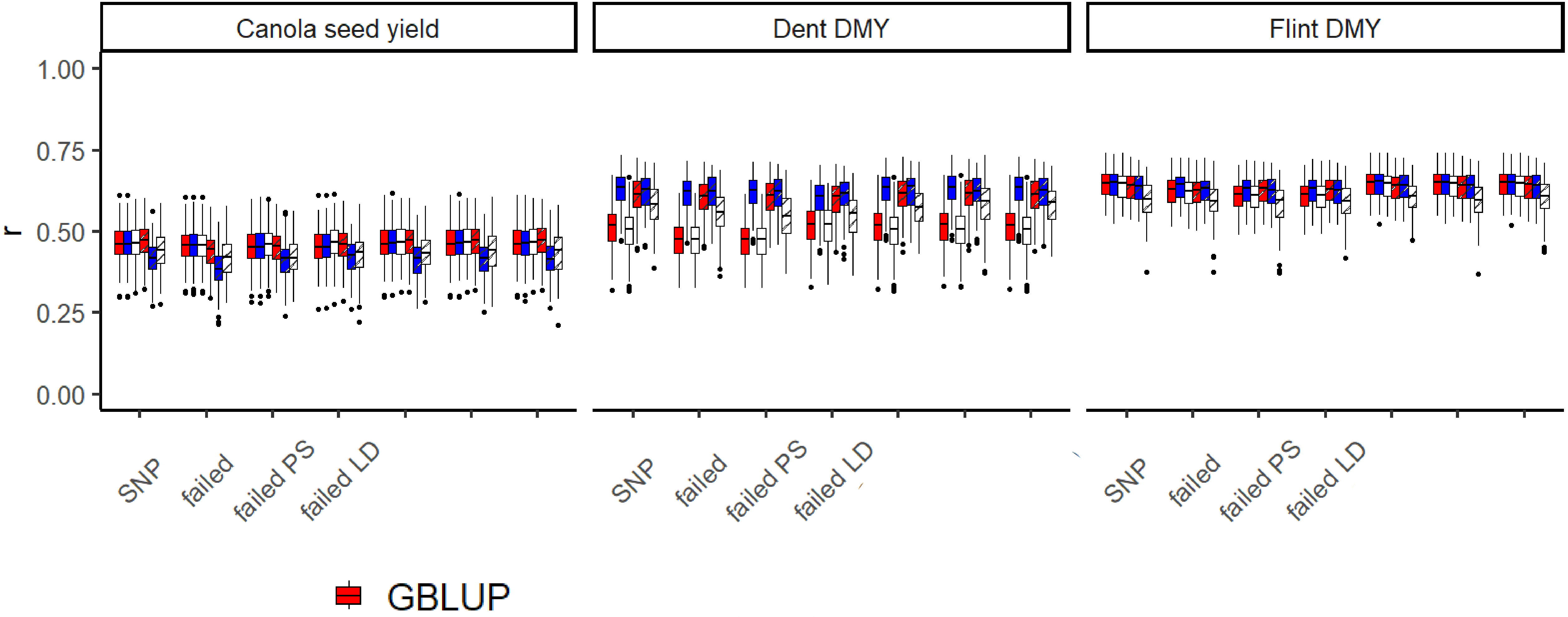
Population structure



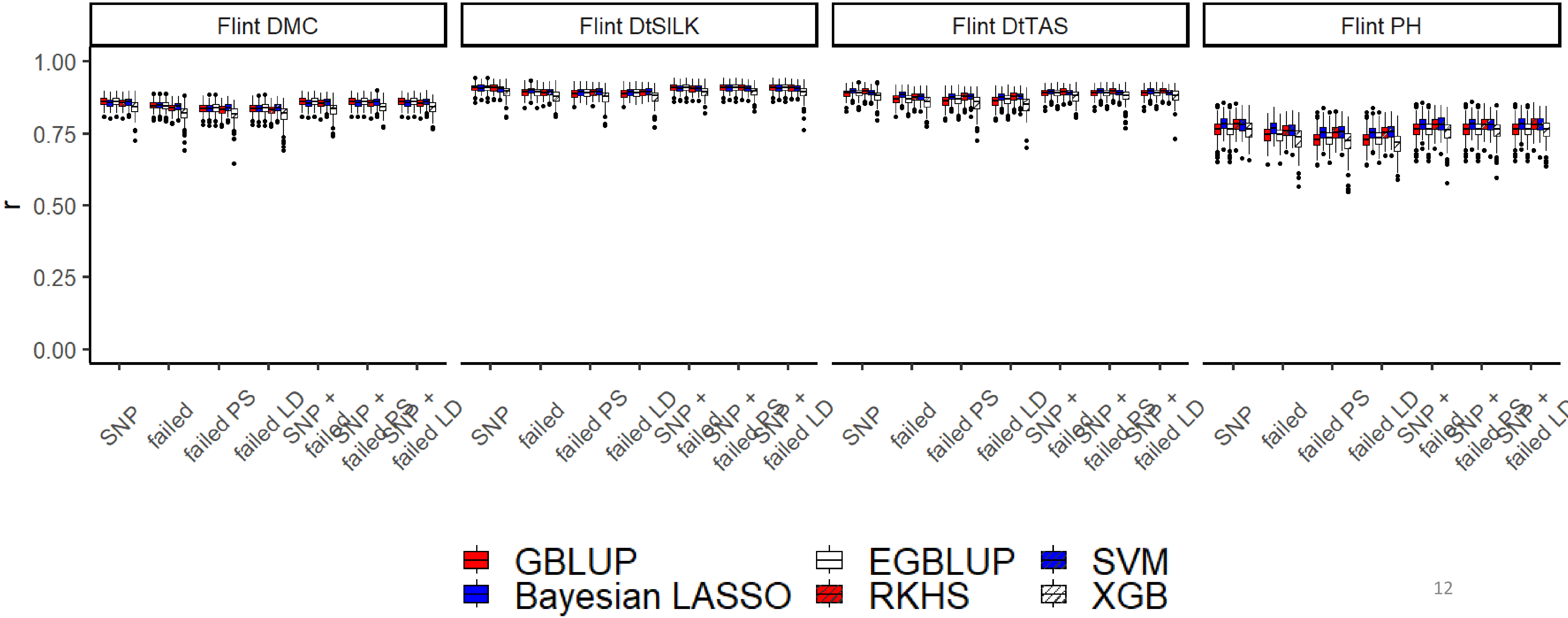
Linkage Disequilibrium



Genomic Prediction

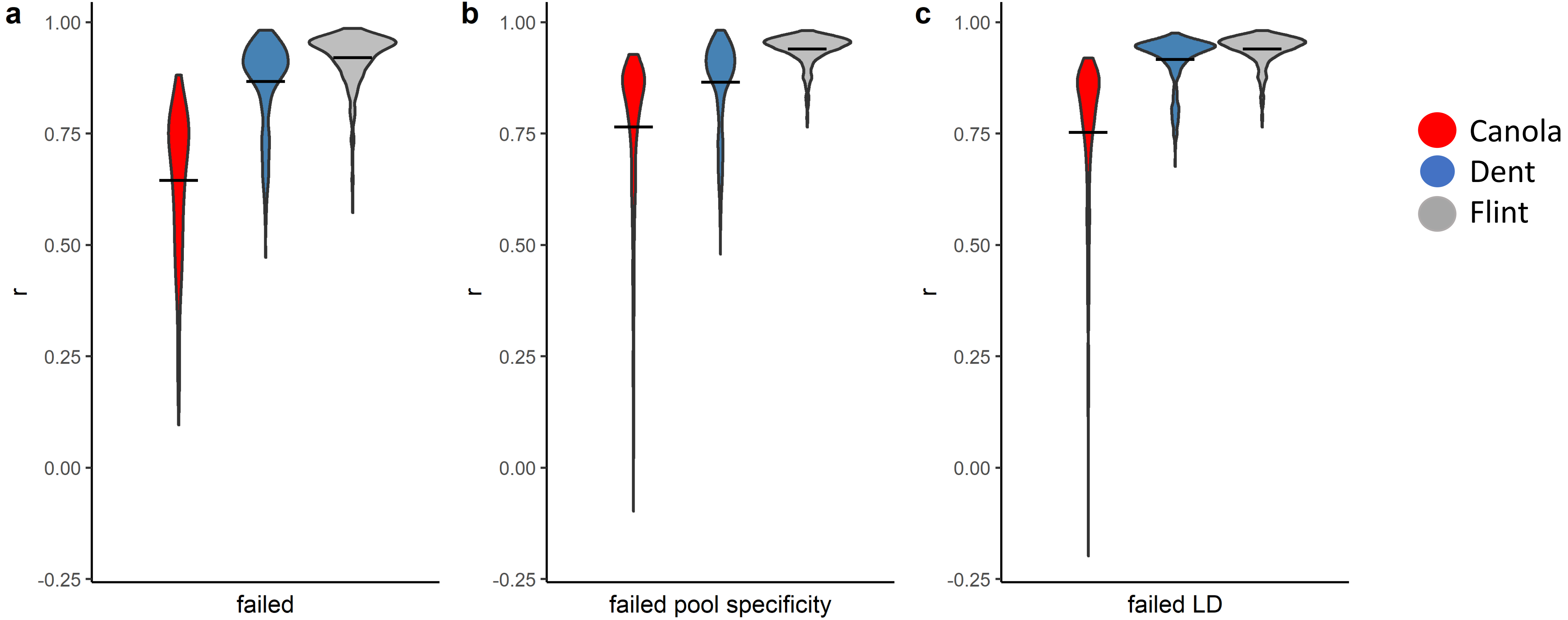


Genomic Prediction



Relationship

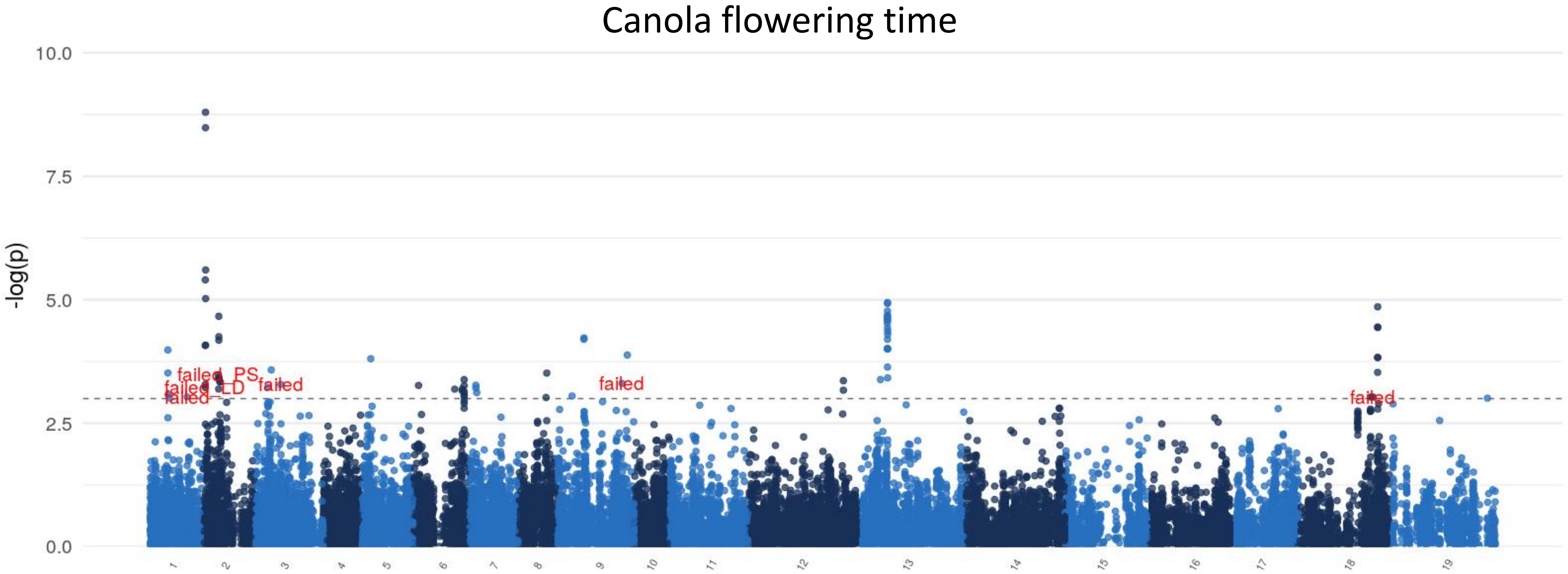
Correlation between realtionships (SNP vs failed)



Conclusion

- Failed allele calls are represantative for genomic relationship
- Competitive in terms of genomic prediction accuracy
- Naïve methods to recover “deletions” help to better distinguish between errors and “deletions”
- No improvement when combining SNPs and failed calls

GWAS


Canola flowering time

Gabur I, Chawla HS, Liu X, et al (2018) Finding invisible quantitative trait loci with missing data. Plant Biotechnology Journal 16:2102–2112. https://doi.org/10.1111/pbi.12942

Thank you!

✉ Sven.E.Weber@agrar.uni-giessen.de

@Svenomics

Lennard Ehrig

Rod Snowdon

Matthias Frisch

Andreas Stahl

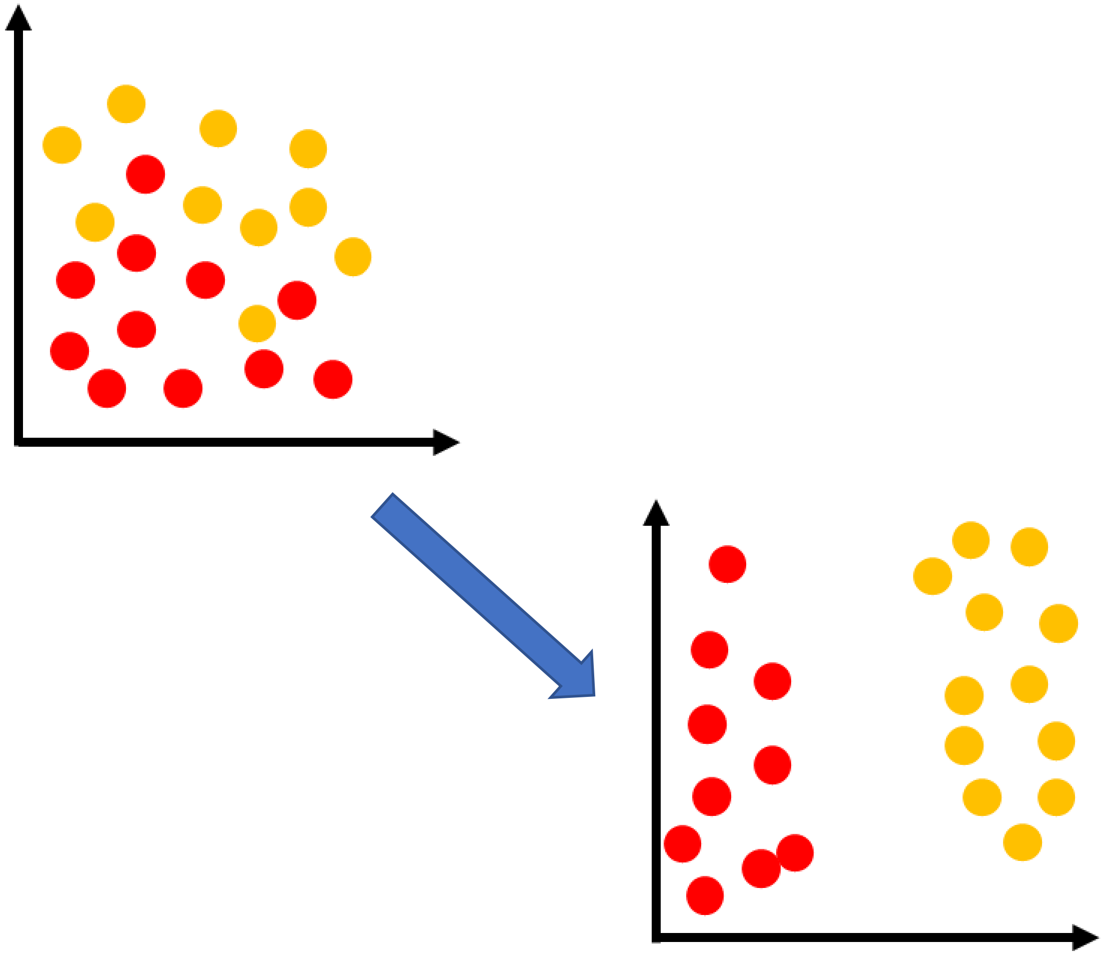
Phillipp Heilmann

SPONSORED BY THE

Federal Ministry
of Education
and Research

Population structure

- Knowledge about pool assignment
- Are the failed allele calls independent of the pool assignment
- X^2- test of independence
  - If significant dependencies are found, we consider the failed allele as "recovered"



Linkage Disequilibrium

- LD calculation between all available markers
- Compare the average LD of a failed allele (treated as independent marker) with it´s SNP counterpart
- One sided t-test
    - If LD is significantly not lower with the failed allele, we consider the allele as “recovered”