

Optimizing the construction of haplotype blocks to increase genomic prediction accuracy across maize landraces

Daniel Valle Torres¹, Manfred Mayer¹, Torsten Pook², Milena Ouzunova³, Thomas Presterl³, Chris-Carolin Schön¹

¹Plant Breeding, Technical University of Munich, Freising, Germany; ²Animal Breeding and Genetics, Georg-August-University, Göttingen, Germany; ³Maize Breeding, KWS SAAT SE, Einbeck, Germany

Introduction

Landraces are a rich source of novel genetic variation, but landrace derived lines have to be selected and improved before they can be integrated into elite breeding pools [1]. Genomic prediction can be an appropriate tool for rapid pre-breeding of landraces. In genomic prediction, haplotype blocks might capture ancestral information better than SNPs and thus increase prediction accuracy across landraces.

Materials & Methods

873 doubled haploid lines were derived from two landraces (Table 1), genotyped with 600k SNPs and phenotyped in up to 11 environments [1,2]. The following traits were measured: early vigor (EV), plant height (PH), female flowering (FF) and root lodging (RL).

Table 1. Sample size, abbreviation and origin of the doubled haploid libraries derived from maize landraces.

| Landrace population | Abbr. | Origin | Size |
|------------------------|-------|---------|------|
| Kemater Landmais Gelb | KE | Austria | 471 |
| Petkuser Ferdinand Rot | PE | Germany | 402 |

Haplotype blocks, defined as allelic sequences of arbitrary length, were constructed using the HaploBlocker package in R [3], representing genomic segments present in both landraces (Figure 1).

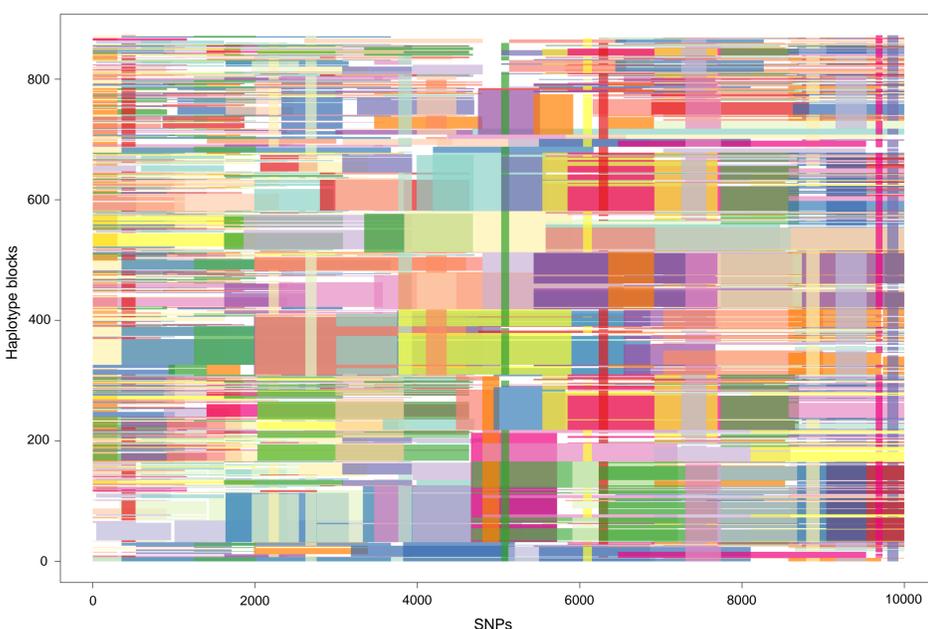


Figure 1. Graphical representation of the haplotype blocks shared by at least five individuals of landraces KE and PE aligned to ten thousand SNPs in chromosome 1.

Genomic prediction was conducted based on the quality filtered set of SNPs and on different sets of haplotype blocks following the GBLUP model.

Predicting across landraces

By optimizing the settings for the construction of haplotype blocks, genomic prediction accuracy across landraces was increased compared to SNPs (Figure 2).

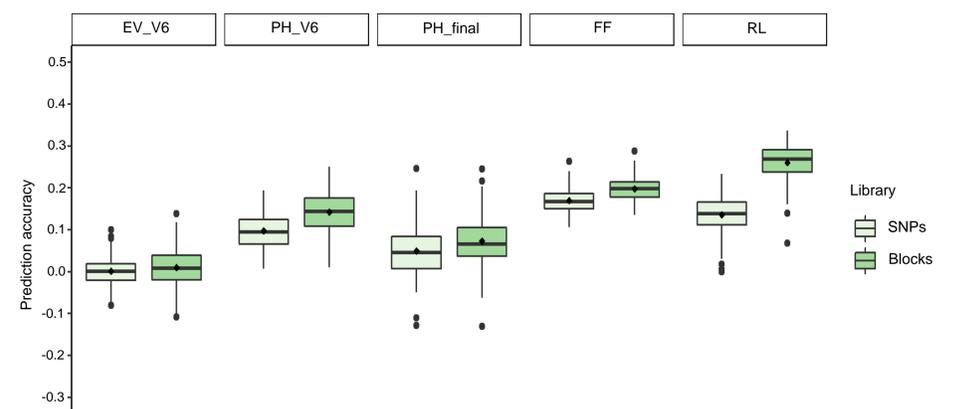


Figure 2. Prediction accuracy for five traits using SNPs or haplotype blocks for predicting across landraces, from PE (training set, n = 300) to KE (prediction set, n = 471). Diamonds indicate the mean derived from 100 repetitions.

Dimensionality reduction was achieved from 501,124 SNPs to 89,305 haplotype blocks, with an average size of 68 SNPs and on average 603 individuals. The haplotype block dataset led to a coverage of 88.7% of the initial SNP matrix.

The increase in prediction accuracy achieved with haplotype blocks compared to SNPs could be due to stronger LD between haplotypes and QTL compared to the LD between SNPs and QTL. Moreover, haplotypes naturally capture short-range epistatic interactions, which further could be accountable for an increase in prediction accuracy.

Conclusion & Outlook

In this study, it has been shown that using haplotypes can increase genomic prediction accuracy across maize landraces. Further research is required comparing additional methods of haplotype block construction and statistical models in the context of genomic prediction across plant populations.

References

- [1] Mayer M, Unterseer S, Bauer E, et al (2017) Is there an optimum level of diversity in utilization of genetic resources? *Theor Appl Genet* 130:2283–2295
- [2] Hölker AC, Mayer M, Presterl T, et al (2019) European maize landraces made accessible for plant breeding and genome-based studies. *Theor Appl Genet* 132:3333–3345
- [3] Pook T, Schlather M, de los Campos G, et al (2019) HaploBlocker: Creation of subgroup-specific haplotype blocks and libraries. *Genetics* 212:1045–1061