University of São Paulo
"Luiz de Queiroz" College of Agriculture

Practical considerations for genotype imputation and multi-trait multi-environment genomic prediction in a tropical maize breeding program

Amanda Avelar de Oliveira

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2019

Amanda Avelar de Oliveira
Agronomist

Practical considerations for genotype imputation and multi-trait multi-environment genomic prediction in a tropical maize breeding program

Advisor:
Prof. Dr. **GABRIEL RODRIGUES ALVES MARGARIDO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2019

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Oliveira, Amanda Avelar de

Practical considerations for genotype imputation and multi-trait multi-environment genomic prediction in a tropical maize breeding program / Amanda Avelar de Oliveira. - - Piracicaba, 2019.

72p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Seleção genômica 2. Dados perdidos 3. Modelos multivariados 4. Genotipagem por sequenciamento I. Título

I dedicate with love to

my parents Carlos Alberto e Maria Lilian,

my sister Ana Luiza

and my husband Ricardo.

ACKNOWLEDGEMENTS

To the “Luiz de Queiroz” College of Agriculture - University of São Paulo (ESALQ-USP) and the Graduate Program in Genetics and Plant Breeding for all the support throughout my academic time.

The financial support from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), are greatfully acknowlegded. Additionally, to Embrapa Milho e Sorgo, who provided the phenotypic and genotypic data.

Appreciation goes to my advisor Dr. Gabriel Rodrigues Alves Margarido for his generosity, guidance and support, which were essential during my graduate studies.

Special acknowledgement goes to Dr. Maria Marta Pastina, who contributed enormously to this work and was always available to assist me, as well for being an example of woman in science.

I am also sincerely grateful to Dr. Marcio Resende Jr. for the opportunity to join his amazing team during six months at Sweet Corn Genomics and Breeding, University of Florida, Gainesville, FL, USA. All his support and guidance during this time were essential for the conducted of this work.

Additionally, I would like to thank all the professors and staff of the Department of Genetics at ESALQ/USP and of the Horticultural Science Department at University of Florida, for their support and extensive knowledge provided.

Thanks to all my colleagues from the Laboratory of Bioinformatic Applied to Bioenergy for all their support, talks, friendship and scientific discussions. I am also grateful to the colleagues from Sweet Corn and Genomic Breeding, it was a pleasure worked with you guys.

My sincerely gratitude goes to Luis Felipe Ferrão e Juliana Benevenuto, who hosted me at Gainesville and were a family to me. I am also thankfull to Rodrigo Amandeu and Maria Fernanda Trientini, I really miss the "Rep UF-ESALQ".

I extend my acknowledgements to all my family. Specially, to my parents, Carlos Alberto and Maria Lilian and my sister Ana Luiza. This work would not be possible without their continuous love, help and support.

To my husband Ricardo for his unconditional support, patience, encouragement and love.

Finally, to God, my guide throughout all the success and frustration.

CONTENTS

RESUMO

Considerações práticas para a imputação de genótipos e predição genômica aplicada a múltiplos caracteres e ambientes em um programa de melhoramento de milho tropical

A disponibilidade de marcadores moleculares cobrindo todo o genoma, como os polimorfismos de nucleotídeos individuais (*single nucleotide polymorphism* - SNP), aliada aos recursos computacionais para o processamento de grande volume de dados, tornou possível o desenvolvimento de uma abordagem de melhoramento assistido para caracteres de herança quantitativa, conhecida como seleção genômica. Na última década a seleção genômica tem sido implementada com sucesso em uma enorme variedade de espécies animais e vegetais, comprovando suas vantagens sobre a seleção assistida por marcadores tradicional e a seleção baseada apenas em informações de parentesco. No entanto, alguns desafios práticos ainda podem limitar a implementação deste método em um programa de melhoramento de plantas. Como exemplos, citam-se o custo da genotipagem de alta densidade de um grande número de indivíduos e a aplicação de modelos mais complexos, que consideram múltiplos caracteres e ambientes. Dessa forma, este estudo teve como objetivos: *i*) investigar estratégias de identificação de SNPs e imputação que possibilitem uma genotipagem de alta densidade economicamente viável; e *ii*) avaliar a aplicação de modelos multivariados de seleção genômica para múltiplos caracteres e ambientes. Este trabalho foi divido em dois capítulos. No primeiro capítulo, comparou-se a acurácia de quatro métodos de imputação: NPUTE, Beagle, KNNI e FILLIN, usando dados de genotipagem por sequenciamento (*genotyping-by-sequencing* – GBS) de 1.060 linhagens de milho, que foram genotipadas usando diferentes profundidades de cobertura. Além disso, duas estratégias de identificação de SNPs e imputação foram avaliadas. Os resultados indicaram que a combinação de estratégias de detecção de polimorfismos e imputação pode possibilitar uma genotipagem economicamente viável, resultando em maiores acurácias de imputação. No segundo capítulo, modelos multivariados de seleção genômica, para múltiplos caracteres e ambientes, foram comparados com suas versões univariadas. Dados de 415 híbridos avaliados na segunda safra em quatro anos (2006-2009) para os caracteres produtividade de grãos, número de espigas e umidade foram utilizados. Os genótipos dos híbridos foram inferidos *in silico* com base nos genótipos das linhagens parentais usando marcadores SNPs obtidos via GBS. No entanto, informações genotípicas estavam disponíveis para apenas 257 híbridos, de modo que foi necessário fazer uso da matriz **H**, a qual combina informações de parentesco genético baseadas em pedigree e marcadores. Os resultados obtidos demonstraram que o uso de modelos de seleção genômica para múltiplos caracteres e ambientes pode aumentar a capacidade preditiva, especialmente para predizer a performance de híbridos nunca avaliados em qualquer ambiente.

Palavras-chave: Seleção genômica; Dados perdidos; Modelos multivariados; Genotipagem por sequenciamento

ABSTRACT

Practical considerations for genotype imputation and multi-trait multi-environment genomic prediction in a tropical maize breeding program

The availability of molecular markers covering the entire genome, such as single nucleotide polymorphism (SNP) markers, allied to the computational resources for processing large amounts of data, enabled the development of an approach for marker assisted selection for quantitative traits, known as genomic selection. In the last decade, genomic selection has been successfully implemented in a wide variety of animal and plant species, showing its benefits over traditional marker assisted selection and selection based only on pedigree information. However, some practical challenges may still limit the wide implementation of this method in a plant breeding program. For example, we cite the cost of high-density genotyping of a large number of individuals and the application of more complex models that take into account multiple traits and environments. Thus, this study aimed to *i*) investigate SNP calling and imputation strategies that allow cost-effective high-density genotyping, as well as *ii*) evaluating the application of multivariate genomic selection models to data from multiple traits and environments. This work was divided into two chapters. In the first chapter, we compared the accuracy of four imputation methods: NPUTE, Beagle, KNNI and FILLIN, using genotyping-by-sequencing (GBS) data from 1060 maize inbred lines, which were genotyped using different depths of coverage. In addition, two SNP calling and imputation strategies were evaluated. Our results indicated that combining SNP-calling and imputation strategies can enhance cost-effective genotyping, resulting in higher imputation accuracies. In the second chapter, multivariate genomic selection models, for multiple traits and environments, were compared with their univariate versions. We used data from 415 hybrids evaluated in the second season in four years (2006-2009) for grain yield, number of ears and grain moisture. Hybrid genotypes were inferred *in silico* based on their parental inbred lines using SNP markers obtained via GBS. However, genotypic information was available only for 257 hybrids, motivating the use of the **H** matrix, which combines genetic information based on pedigree and molecular markers. Our results demonstrated that the use of multi-trait multi-environment models can improve predictive abilities, especially to predict the performance of hybrids that have not yet been evaluated in any environment.

Keywords: Genomic selection; Missing data; Multivariate models; Genotyping-by-sequencing

1 GENERAL INTRODUCTION

The leveraging of heterosis has been extremely successful in affording continuous improvement in commercial maize grain yield. Since the beginning of the hybrid era, maize breeders achieved increases in grain yield that are unmatched among other cereals or oil seeds (Lee and Tracy 2009; Hallauer and Miranda Filho 2010). Classical maize breeding consists of crossing lines from different heterotic groups and measuring phenotypic performance of hybrids in multiple environment trials. However, phenotyping has become one of the most costly and laborious stages in a breeding program. Thus, genomic prediction stands out in virtue of its ability to reduce the time required to complete a breeding cycle, to enable an earlier and more efficient selection of superior genotypes, and to reduce phenotyping costs, representing a promising tool for use in maize breeding programs (Crossa et al. 2017; Wang et al. 2018).

Genomic prediction was first proposed in 2001 by Meuwissen et al. Since then, it has been applied to a variety of crops and routinely practiced in breeding programs of major seed companies, especially for maize and soybean (Bernardo 2016). The key idea of this method is the simultaneous prediction of the effects of a large number of markers spread throughout the genome, in order to ensure that every quantitative trait locus (QTL) affecting a trait be in linkage disequilibrium (LD) with at least one marker. This method remained unexplored for a few years, because the molecular markers available at that time were limited and obtained at high costs. However, the emerging of next-generation sequencing technology presented the possibility of obtaining molecular markers densely distributed across the genome, using high-throughput techniques such as genotyping-by-sequencing (GBS) (Elshire et al. 2011).

Feature of GBS data are the high rates of missingness and heterozygote undercalling, prompting the use of approaches to impute these missing genotypes. In this scenario, several studies have assessed the efficiency of imputing missing data, using different methods and strategies (Howie et al. 2009; Cleveland et al. 2011; Hickey et al. 2012; Swarts et al. 2014; Bouwman et al. 2014; Nazzicari et al. 2016; Gonen et al. 2018). Besides that, the cost of genotyping many samples at high density is still high, representing a barrier to small or public plant breeding programs to routinely implement genomic prediction. Therefore, it is necessary to adopt low cost genotyping strategies to solve this limitation. For species for which genotyping chips are available, combining data from high and low density SNP arrays is a cost effective strategy (Jacobson et al. 2015; Hickey et al. 2015; Gorjanc et al. 2017). When

genotyping chips are not available, the GBS technology allows breeders to adjust the amount of retrieved information and its cost by choosing different restriction enzymes, regulating sequencing depth and the level of multiplexing (Elshire et al. 2011; Deschamps et al. 2012; Poland and Rife 2012).

Currently, the majority of genomic prediction models applied are univariate ones. However, in breeding programs it is common to evaluate several traits simultaneously, because elite genotypes should concentrate favorable alleles for several traits of interest. The existence of genetic correlation between quantitative traits indicates that measures in one trait provide indirect information about other traits, a fact that can be used to improve the predictive ability of genomic selection (Calus and Veerkamp 2011; Jia and Jannink 2012; Guo et al. 2014; Dos Santos et al. 2016; Marchal et al. 2016; Lyra et al. 2017; Covarrubias-Pazaran et al. 2018). Besides the correlation between traits, considering a model that also accommodate the genotype by environment interaction, is also an important issue to plant breeders, since genotypes are evaluated for multiple traits in multiple environments. The types of model that jointly take into account multiple traits and environments are referred to as multi-trait multi-environment (MTME) models. Nonetheless, few studies have simultaneously assessed multiple traits and multiple environments for genomic selection purposes (Montesinos-López et al. 2016; Gomes Torres et al. 2018; Ward et al. 2019).

The complexity of applying genomic prediction in plant breeding programs arises at different levels and is influenced by several factors. In order to investigate some of the challenges faced by breeders, when applying genomic prediction to a maize breeding program, this work is the result of a partnership among: Embrapa Milho e Sorgo (Sete Lagoas, MG, Brazil), the Laboratory of Bioinformatics Applied to Bioenergy at ESALQ/USP ("Luiz de Queiroz" College of Agriculture, University of São Paulo - Piracicaba, SP, Brazil) and the Sweet Corn Genomics and Breeding at University of Florida, Gainesville, FL, USA. In this context, we conducted two studies that are herein organized in two chapters. In the first chapter, we aimed to evaluate different SNP calling and imputation strategies using GBS data of maize lines from the Embrapa maize breeding program. Subsequently, chapter 2 focuses on applications of multi-trait multi-environment genomic prediction models to second season maize hybrids, which also originated from the Embrapa breeding program.

References

Bernardo R (2016) Bandwagons I, too, have known. Theor Appl Genet 129:2323–2332. https://doi: 10.1007/s00122-016-2772-5

Bouwman AC, Hickey JM, Calus MP, Veerkamp RF (2014) Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle. Genet Sel Evol 46:6. https://doi: 10.1186/1297-9686-46-6

Calus MPL, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:26. https://doi: 10.1186/1297-9686-43-26

Cleveland MA, Hickey JM, Kinghorn BP (2011) Genotype imputation for the prediction of genomic breeding values in non-genotyped and low-density genotyped individuals. BMC Proc 5:S6. https://doi: 10.1186/1753-6561-5-S3-S6

Covarrubias-Pazaran G, Schlautman B, Diaz-Garcia L, et al (2018) Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of *Vaccinium macrocarpon* Ait. Front Plant Sci 9:1310. https://doi: 10.3389/fpls.2018.01310

Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22:961–975. https://doi: 10.1016/J.TPLANTS.2017.08.011

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology (Basel) 1:460–483. https://doi: 10.3390/biology1030460

Dos Santos JPR, De Castro Vasconcellos RC, Pires LPM, et al (2016) Inclusion of dominance effects in the multivariate GBLUP model. PLoS One 11:1–21. https://doi: 10.1371/journal.pone.0152045

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1–10. https://doi: 10.1371/journal.pone.0019379

Gomes Torres L, Rodrigues MC, Lima NL, et al (2018) Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. PLoS One 13: 1–15. https://doi: 10.1371/journal.pone.0199492

Gonen S, Wimmer V, Gaynor RC, et al (2018) A heuristic method for fast and accurate phasing and imputation of single nucleotide polymorphism data in bi- parental plant populations. Theor Appl Genet 131:2345–2357. https://doi: https://doi.org/10.1007/s00122-018-3156-9

Gorjanc G, Dumasy J-F, Gonen S, et al (2017) Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. Crop Sci 57:1404. https://doi: 10.2135/cropsci2016.08.0675

Guo G, Zhao F, Wang Y, et al (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:1–7. https://doi: 10.1186/1471-2156-15-30

Hallauer A., Miranda Filho J. (2010) Quantitative genetics in maize breeding., 2.ed. Iowa State University Press, Ames

Hickey JM, Crossa J, Babu R, de los Campos G (2012) Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Sci 52:654. https://doi: 10.2135/cropsci2011.07.0358

Hickey JM, Gorjanc G, Varshney RK, Nettelblad C (2015) Imputation of single nucleotide polymorphism genotypes in biparental, backcross, and topcross populations with a Hidden Markov Model. Crop Sci 55:1934. https://doi: 10.2135/cropsci2014.09.0648

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529. https://doi: 10.1371/journal.pgen.1000529

Jacobson A, Lian L, Zhong S, Bernardo R (2015) Marker imputation before genomewide selection in biparental maize populations. Plant Genome 8:1–9. https://doi: 10.3835/plantgenome2014.10.0078

Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. https://doi: 10.1534/genetics.112.144246

Lee EA, Tracy WF (2009) Modern maize breeding. In: Bennetzen J.L., Hakes S. (eds) Handbook of Maize. Springer, New York, New York, pp 141–160

Lyra DH, Mendonça L F, Galli G, et al (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol Breed 37:80. https://doi: 10.1007/s11032-017-0681-1

Marchal A, Legarra A, Sébastien T, et al (2016) Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq.) progeny tests. Mol Breed 36:2. https://doi: 10.1007/s11032-015-0423-1

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi: 11290733

Montesinos-López OA, Montesinos-López A, Crossa J, et al (2016) A genomic bayesian multi-trait and multi-environment model. G3 Gene Genome Genet 6:2725–2744. https://doi: 10.1534/g3.116.032359

Nazzicari N, Biscarini F, Cozzi P, Brummer EC (2016) Marker imputation efficiency for genotyping-by-sequencing data in rice (*Oryza sativa*) and alfalfa (*Medicago sativa*). Mol Breed 36:1–16. https://doi: 10.1007/s11032-016-0490-y

Poland J a, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome J 5:92–102. https://doi: 10.3835/plantgenome2012.05.0005

Swarts K, Li H, Romero Navarro JA, et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7:1–12. https://doi: 10.3835/plantgenome2014.05.0023

Wang X, Xu Y, Hu Z, Xu C (2018) Genomic selection methods for crop improvement: Current status and prospects. Crop J 6:330–340. https://doi: 10.1016/j.cj.2018.03.001

Ward BP, Brown-Guedira G, Tyagi P, et al (2019) Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. Crop Sci 59:1–17. https://doi: 10.2135/cropsci2018.03.0189

2 SNP CALLING AND IMPUTATION STRATEGIES FOR COST EFFECTIVE GENOTYPING IN A TROPICAL MAIZE BREEDING PROGRAM

Abstract

Genotyping-by-sequencing (GBS) datasets typically feature high rates of missingness and heterozygote undercalling, prompting the use of data imputation. We compared the accuracy of four imputation methods: NPUTE, Beagle, KNNI and FILLIN, using GBS data of maize inbred lines, genotyped using different multiplexing levels. Two strategies for SNP calling and genotype imputation were evaluated. First, only lines genotyped through 96-plex were used for SNP discovery, whereas both 96 and 384-plex were simultaneously used in the second strategy. In the first genotype imputation strategy, only the 96-plex lines were imputed, and next the remaining lines were appended (96-plex-imputed + 384-plex) and then imputed. In the second imputation strategy, we jointly imputed both datasets. We also investigated the impacts of including heterozygous genotypes and distinct rates of missing genotypes per locus. The different SNP-calling strategies and percentage of missing data did not substantially affect the imputation accuracy. However, the different imputation strategies exhibited a sizable effect. Generally, imputations were less accurate for heterozygotes. The scenario 96-plex-imputed + 384-plex showed accuracies similar to the 96-plex scenario. Beagle and NPUTE produced the highest accuracies. Our results indicate that combining SNP-calling and imputation strategies can enhance cost-effective genotyping, resulting in higher imputation accuracies.

Keywords: Genotyping-by-sequencing; Unobserved genotype imputation; Beagle; KNNI; FILLIN; NPUTE

2.1 Introduction

The emergence of next-generation sequencing technology presented the possibility of obtaining molecular markers densely distributed across the genome, using high-throughput techniques such as genotyping-by-sequencing (GBS) (Elshire et al. 2011). Making use of these genome-wide genotyping platforms, genomic selection and genome-wide association studies offer great potential to accelerate and enhance selection efficiency of plant breeding programs (Desta and Ortiz 2014; Chang et al. 2018; Dias et al. 2018; Faville et al. 2018; Haile et al. 2018; Gerard et al. 2018; Kayondo et al. 2018). However, the costs of high-density genotyping for large numbers of individuals are still infeasible, representing a barrier to a more widespread adoption of these tools. Because the accuracy of genomic selection and power of association studies usually increase with increasing numbers of individuals and density of markers, low-cost genotyping strategies have to be adopted to address resource limitations (Jacobson et al. 2015; Han et al. 2018; Cericola et al. 2018).

The adoption of genomic selection in a maize breeding program allows breeders to genotype elite lines and to predict the performance of all possible hybrids, even if they are not phenotypically evaluated. This strategy reduces the costs and labor involved in field trials and can increase genetic gains. In any case, cost effective genotyping is crucial. A feature of GBS data are the high rates of missingness and heterozygote undercalling, which vary according to the kind of population and multiplexing level. Several studies have reported the efficiency of imputing missing data, using different methods and strategies (Howie et al. 2009; Cleveland et al. 2011; Hickey et al. 2012; Swarts et al. 2014; Bouwman et al. 2014; Nazzicari et al. 2016; Gonen et al. 2018). An effective strategy involves genotyping some of the individuals at higher marker density, then using these high-density data to impute larger numbers of individuals genotyped at lower marker density. Genomic selection studies adopting this approach reported increases in the predictive accuracy (Jacobson et al. 2015; Gorjanc et al. 2017a, b). For species for which genotyping chips are available, as is the case of economically important animals and some crops, combining data from high and low density SNP arrays is a cost effective strategy (Jacobson et al. 2015; Hickey et al. 2015; Gorjanc et al. 2017b). When genotyping chips are not available or their use is prohibitive, the GBS technology allows breeders to adjust the amount of retrieved information and its cost in different ways (Gorjanc et al. 2017b). For instance, choosing different restriction enzymes, regulating sequencing depth and the level of multiplexing (Elshire et al. 2011; Deschamps et al. 2012; Poland and Rife 2012). However, to the best of our knowledge, no studies have yet empirically investigated the combined use of SNP calling and imputation strategies to improve GBS data quality.

There are several imputation methods available, but most of them were developed for humans (Howie et al. 2009; Browning and Yu 2009; Liu et al. 2013). However, humans are highly heterozygous, obligate outcrossers, show little inbreeding and much less structural variation than that observed in crop plants. These factors make the imputation methods designed for humans not necessarily optimized for use in crop systems. For this reason, it is worthwhile to compare different imputation methods, which may or may not allow for heterozygous genotypes. Situations in breeding programs where there are genotypic datasets with varying levels of multiplexing and heterozygosity are increasingly common. There is therefore scientific and practical interest in gaining knowledge about how to better explore such datasets in order to achieve high imputation accuracies.

In this paper, we compared the imputation accuracy of four imputation methods: NPUTE (Roberts et al. 2007), Beagle (Browning and Browning 2007), K-nearest neighbors imputation (KNNI) (Troyanskaya et al. 2001) and Fast Inbreed Line Library ImputatioN (FILLIN) (Swarts et al. 2014); which are well known algorithms implemented in freely available software libraries. We imputed missing genotypes from GBS data of maize inbred lines, genotyped using different levels of multiplexing per sequencing lane. We evaluated different SNP calling strategies in order to better explore the low and high multiplexing levels of our dataset. Because this dataset represents a panel of maize inbred lines mostly in final stages of the breeding program, we expected that these lines were homozygous for the majority of loci. However, a few of those lines were in initial stages of the breeding program and could thus have higher heterozygosity rates. Hence, we also evaluated the impact of including heterozygous genotypes on imputation accuracy. The main objective of this study was to evaluate different SNP calling and imputation strategies in a real maize breeding program scenario.

2.2 Materials and methods

2.2.1 Experimental data

Data used in this study came from a collection of 1060 maize inbred lines from the Embrapa Maize and Sorghum breeding program. These lines represent dent (34%) and flint (51%) heterotic groups, as well as another group – here called group C (15% of the lines), which is unrelated to both dent and flint sources. We performed DNA extraction from young leaves based on the cetyl trimethylammonium bromide method (Saghai-Maroof et al. 1984). DNA samples were quantified using the Fluorometer Qubit® 2.0, following the manufacturer's instructions (Life Technologies™, USA). Samples were also evaluated on 1% agarose gel in Tris acetate-EDTA buffer, stained with GelRed™ (Biotium, USA) and recorded under UV light in the Imager Gel Doc L-PIX (Loccus Biotecnologia, Brazil). Genotyping-by-sequencing (GBS) was carried out at the Genomic Diversity Facility at Cornell University (Ithaca, NY, USA) using the standard GBS protocol (Elshire et al. 2011) with the restriction enzyme ApeKI. The inbred lines were split into two groups: *i)* 680 lines genotyped using 96 samples per sequencing lane (HiSeq2500 - 1 x 100bp); and *ii)* 380 lines genotyped with 384 samples per lane (NextSeq500 - 1 x 90bp). We thus expected a larger

number of reads per sample in the first group. Tags were aligned to the B73 reference genome (AGPv3) (Law et al. 2015) using the Bowtie2 aligner (Langmead and Salzberg 2012). Then, Single Nucleotide Polymorphisms (SNPs) were called using the GBSv2 Discovery Pipeline, available in the software TASSEL v. 5.2.28 (Glaubitz et al. 2014), using different strategies as shown below.

2.2.2 SNP calling strategies

We evaluated different SNP calling strategies in order to better explore the low and high multiplexing levels of our dataset. In our first strategy, denoted as SNP calling strategy I, we ran the Discovery SNP Caller Plugin using only the 680 lines genotyped with 96 samples per sequencing lane. In this scenario only the lines with higher depth of coverage, which have less missing data and lower genotyping error probability, were used for SNP discovery. We thus expected this to be a set of better quality SNPs, with greater power of detection and less false positives. Next, we ran the Production SNP Caller Plugin with all 1060 lines. By doing so, all lines were effectively genotyped, but only for the loci detected in the first set. For SNP calling strategy II, we ran the Discovery and Production SNP Caller Plugins combining both the high and low multiplexing sets of lines. This strategy was likely to affect the number and quality of discovered SNPs, because we included lines genotyped with 384 samples per sequencing lane throughout the SNP discovery step. Finally, we evaluated the descriptive statistics generated for each discovered marker and applied filters for Minor Allele Frequency (MAF) less than 5% and inbreeding coefficient less than 0.8. Only SNPs that passed both filters were used for further analyses.

2.2.3 Imputation methods

We performed the subsequent imputation analyses using the filtered datasets from the two competing SNP calling strategies. We evaluated four imputation methods: *i*) NPUTE (Roberts et al. 2007), which is based on the similarities between haplotypes of different individuals for the same genomic region. Different window imputation sizes were tested for each chromosome, and the windows with higher accuracies were chosen; *ii*) Beagle (Browning and Browning 2007), which was originally developed for human genetic studies, but also presents a wide application in animal and plant genetics (Law et al. 2015; Nazzicari

et al. 2016). This method infers haplotypes and imputes missing alleles both with known and unknown linkage phase, using a stochastic procedure based on Hidden Markov Models (HMM) to find the most likely haplotype pair for each individual. The method works iteratively using an expectation-maximization (EM) approach; *iii*) K-nearest neighbors imputation – KNNI (Troyanskaya et al. 2001), which is a method based on the weighted average of the k closest markers; *iv*) Fast Inbreed Line Library ImputatioN – FILLIN (Swarts et al. 2014), an imputation method optimized for inbred populations implemented in the software TASSEL v. 5 (Glaubitz et al. 2014). It is based on haplotype reconstruction around recombination break points. Imputation is carried out in two steps: first, the inference of haplotype takes place (FILLINFindHaplotypesPlugin), followed by imputation of missing data based on the resulting haplotypes (FILLINImputationPlugin).

2.2.4 Imputation scenarios

We initially cleaned the two datasets by removing indels and non-biallelic markers. Next, we used two contrasting approaches to compare the influence of heterozygous genotypes, by either keeping or removing any non-homozygous genotype calls. Because our dataset contains a collection of maize inbred lines mostly in final stages of the breeding program (F6 – F7), we expected that these lines were homozygous for the majority of locus. However, a few of those lines were in initial stages of inbreeding (F3 – F4) and could thus have higher heterozygosity rates (up to 25% and 12.5%, respectively).

For each of these scenarios, we then evaluated two different imputation strategies to leverage the varying levels of multiplexing. Towards this end, we first imputed only the 680 lines genotyped with 96 samples per sequencing lane. We expected that the imputation accuracy of this dataset would be higher, because these lines have higher depth of coverage. Later we appended to these imputed data the remaining 380 lines, which were genotyped with 384 samples per sequencing lane, and finally performed the imputation of the remaining missing data. The competing strategy consisted of jointly imputing the high and low multiplexing datasets in a single step.

In addition to the imputation strategies, we aimed to evaluate the impacts of the rate of missing data per marker on the imputation accuracies. Then, we filtered the SNP data to have a maximum of 10, 20, 50 or 80% missing data per marker, generating four sub-datasets. We used these four sub-datasets, in addition to the unfiltered dataset, to perform the imputation

analyses. For each of these datasets, we randomly introduced an additional 10% missing genotypes, based on which imputation accuracy could be measured. All the SNP calling and imputation strategies are summarized in Figure 1.

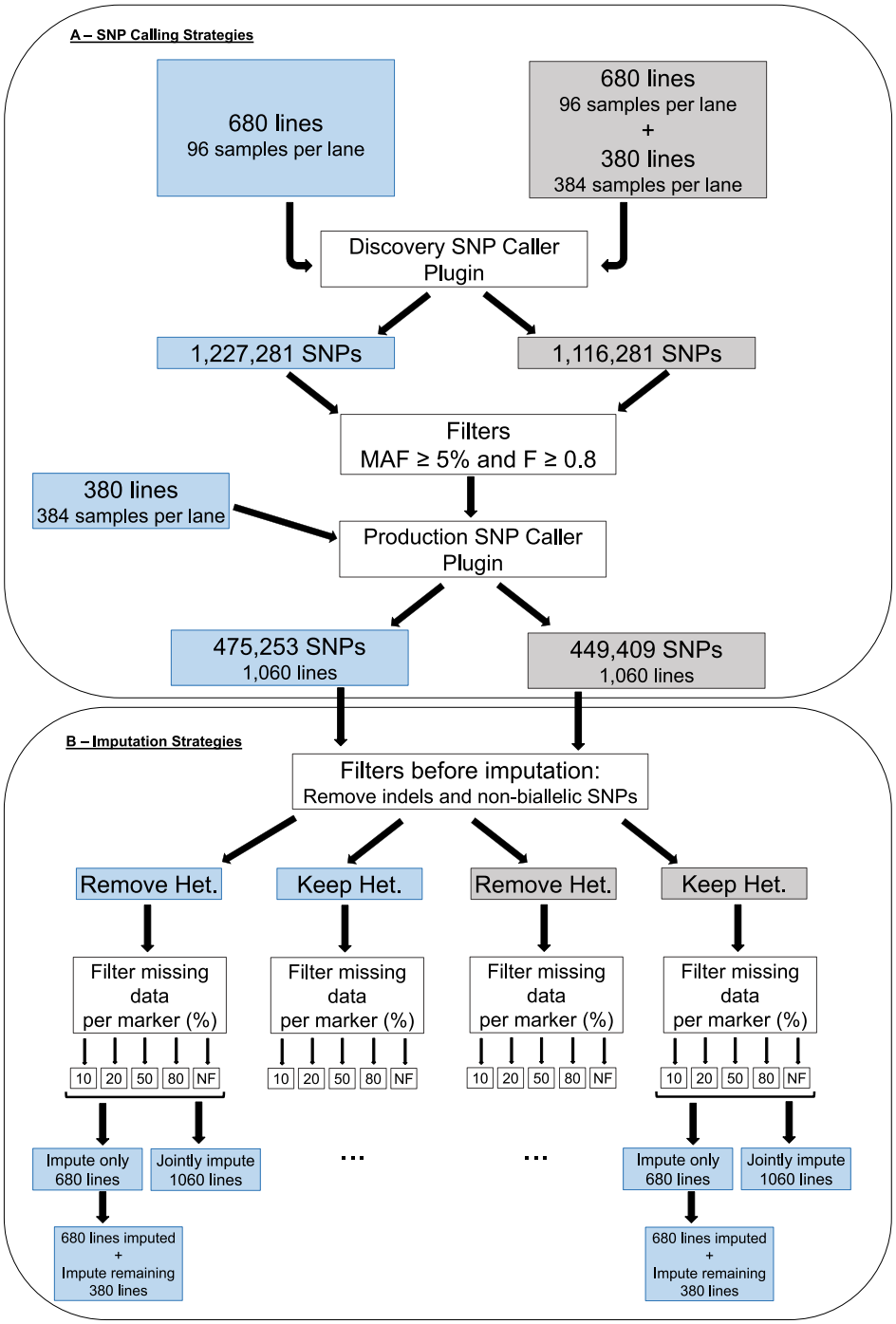


Figure 1 Summary of (A) SNP calling and (B) Imputation strategies. Blue color indicate SNP calling strategy I, whereas gray color indicate SNP calling strategy II. The imputation strategies showed in B were applied to all the four imputation methods evaluated. MAF = minor allele frequency. F = inbreeding coefficient. Het = heterozygous. NF = not filtered

2.2.5 Imputation accuracy and computational time

For each imputation scenario, we used the artificial missing genotypes to measure the overall imputation accuracy and the accuracy for each genotype class. The imputation accuracy was computed as the proportion of correct imputation, measured as the number of correctly imputed missing data divided by the total number of artificially missing data points.

We also measured the computational time required for imputation to be completed in each analysis as an indicator of the software relative performance. To ensure consistency, all jobs were separately submitted to the same computing platform, a multi node server with two Intel® Xeon® E5-2650 v4 @ 2.20 GHz CPUs, with a total of 48 threads, and 256 GB of RAM.

2.2.6 Software

We used the software TASSEL v.5.2.28 (Glaubitz et al. 2014) and the open-source environment for statistical programming R (R Core Team, 2018) for data handling, editing, summarizing results and figure design. We ran the KNNI method in R using the function KNNIcatimpute from the R package Scrime (Schwender and Fritsch 2015). The Beagle imputation method is implemented in the Beagle software version 4.1 (Browning and Browning 2016) and was run using default parameters. We used the TASSEL v.5.2.28 (Glaubitz et al. 2014) plugin FILLINFindHaplotypesPlugin followed by FILLINImputationPlugin to perform the imputation procedure using the FILLIN algorithm (Swarts et al. 2014), considering the options –accuracy and –proSitesMask to calculate the accuracy. For the NPUTE method we used the NPUTE software v.1 (Roberts et al. 2007).

2.3 Results

2.3.1 Genotypic data

Using our SNP calling strategy I, i.e., with only the 680 lines with higher depth of coverage for SNP discovery, we found 1,227,281 SNPs. We initially removed SNPs that did not pass the MAF and inbreeding coefficient filters, generating 475,253 SNPs (Figure 2). Comparatively, we found 1,116,281 SNPs with our SNP calling strategy II, i.e., using all the

1060 lines with high and low depths of coverage for SNP discovery. We again removed SNPs with the same filtering criteria, resulting in 444,409 SNPs (Figure 2). As expected, the number of SNPs found with SNP calling strategy II was slightly smaller, possibly due to the lower detection power with this higher multiplexing dataset. The mean depth of coverage was 3.04 and 0.77 reads per locus per sample, for the lines genotyped using 96 and 384 samples per sequencing lane, respectively. The number of markers found per chromosome and missing data per locus for each SNP calling strategy are in Supplementary Figures 1 and 2, respectively. The number of markers found per chromosome followed similar patterns with both SNP calling strategies (Supplementary Figure 1). In addition to the higher number of markers found with SNP calling strategy I the number of missing data per locus was also higher (Supplementary Figure 2). After removing indels and non-biallelic markers, the different filters of allowed missing data (10%, 20%, 50%, 80% and no filter) generated 12,957, 42,053, 173,328, 368,351, and 474,367 markers, respectively, for SNP calling strategy I; and 17,508, 50,793, 187,440, 380,955, and 443,940 markers, respectively, for SNP calling strategy II.

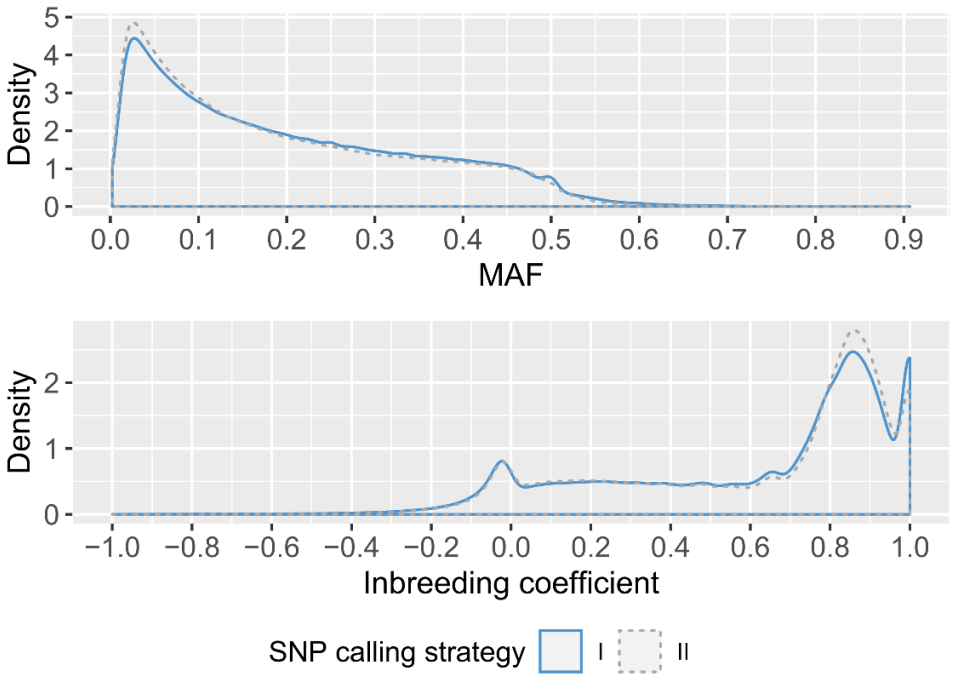


Figure 2 (A) Minor allele frequency (MAF) and (B) Inbreeding coefficient distributions for two alternative SNP calling strategies. Solid blue line corresponds to SNP calling strategy I, i.e., SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines. Dashed gray line corresponds to SNP calling strategy II, i.e., SNP identification and genotype calling using all 1060 samples. A total of 1,227,281 and 1,116,281 markers are represented in SNP calling strategies I and II, respectively

2.3.2 Imputation accuracy

Accuracies are reported for each combination of SNP calling and imputation strategies (Figures 3 – 6). Comparing the imputation accuracies between the SNP calling strategies I and II, we did not observe pronounced differences. When removing heterozygous markers, we observed that Beagle and NPUTE outperformed all other imputation methods in most scenarios evaluated (Figures 3 and 4). The KNNI method presented a computational limitation and in most evaluated scenarios did not run to completion. Interestingly, however, in some cases it outperformed all other methods. For example, in the joint imputation of the whole dataset, the total and major homozygous accuracies of the KNNI method were slightly higher than all other methods when the rate of allowed missing data per locus was 20%. FILLIN resulted in considerably smaller accuracies in all scenarios evaluated.

Contrary to our expectations, the allowed missing data per locus did not substantially adversely affect the imputation accuracy, with most methods showing a (nearly) flat response to increased missing data (Figure 3 and 4). KNNI showed decreasing imputation accuracy with increasing missing data. On the other hand, FILLIN showed increasing imputation accuracy with increasing missing data in the imputation strategy 96 plex imputed + 384 plex and, particularly in the SNP calling strategy II, for the 96 + 384 plex scenario (Figure 4).

When not removing the heterozygous genotypes, we assessed the imputation accuracy with only the Beagle method, because it accepts all genotype classes, while NPUTE and FILLIN require exclusively homozygous genotypes. Although the KNNI method accepts heterozygous genotypes, its computational limitations precluded further evaluation. The accuracy for heterozygous genotypes was considerably lower than for the two homozygous genotypes (Figures 5 and 6). Again, the allowed missing data per locus did not affect the imputation accuracy.

We observed extensive differences between the two different imputation strategies (Figures 3 – 6). As expected, the 96 plex imputation scenario always showed the best accuracies, while the 96 + 384 plex performed the worst. Interestingly, the imputation scenario 96 plex imputed + 384 plex showed imputation accuracies similar to the 96 plex scenario. This opens up the possibility of combining high and low depth GBS data without compromising the imputation accuracy.

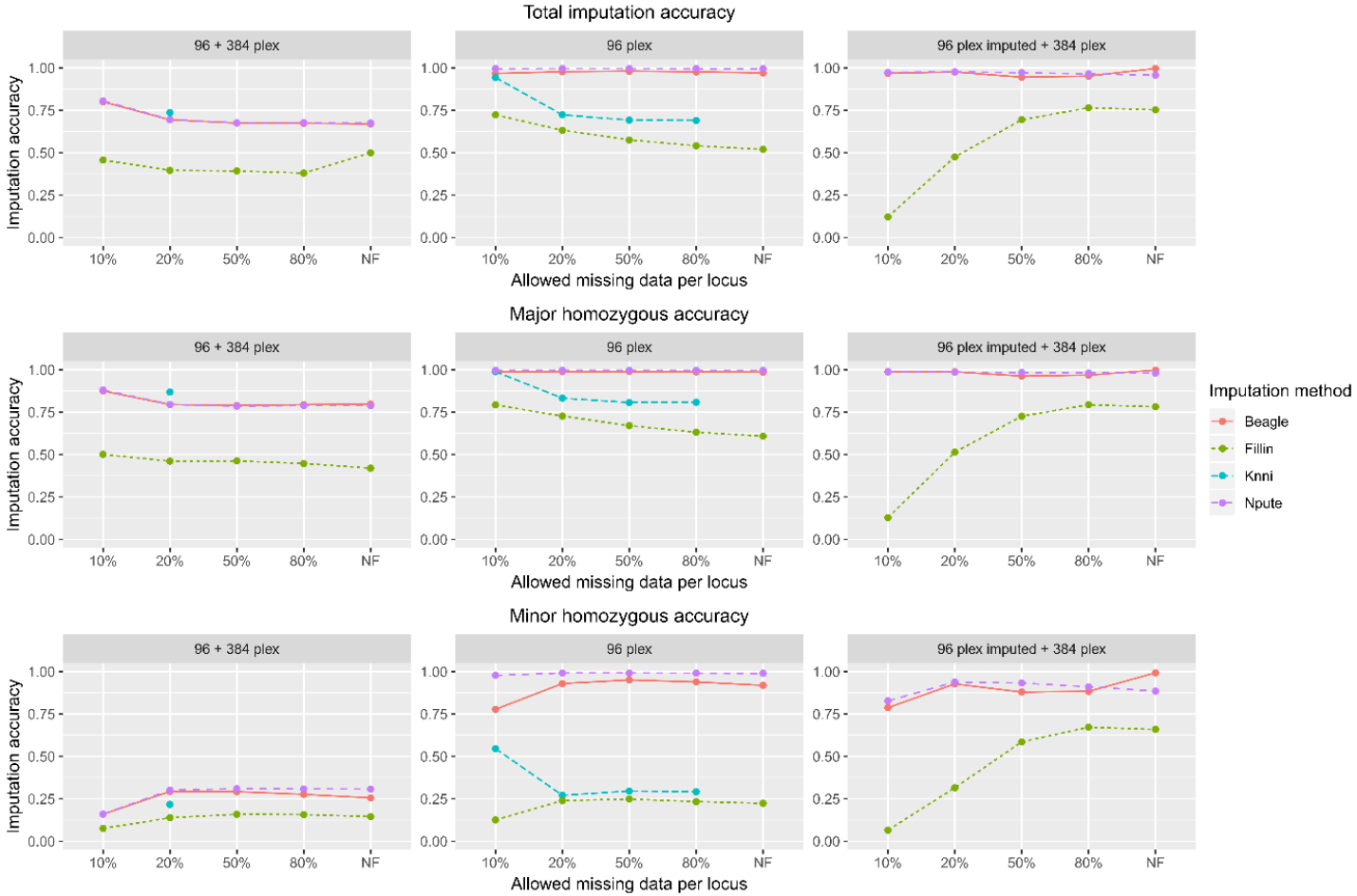


Figure 3 Imputation accuracy using SNP calling strategy I, i.e., SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines, removing heterozygous genotypes. Each row represents imputation accuracy for different genotypic classes: total imputation accuracy, major homozygous accuracy and minor homozygous accuracy. Each column represents an imputation strategy: 96 + 384 plex, 96 plex and 96 plex imputed + 384 plex. The X-axis represents the different filters of allowed missing data per locus (10%, 20%, 50%, 80% and not filtered). Line colors represent the four imputation methods: Beagle (solid red), FILLIN (dotted green), KNNI (dashed blue) and NPUTE (dashed purple)

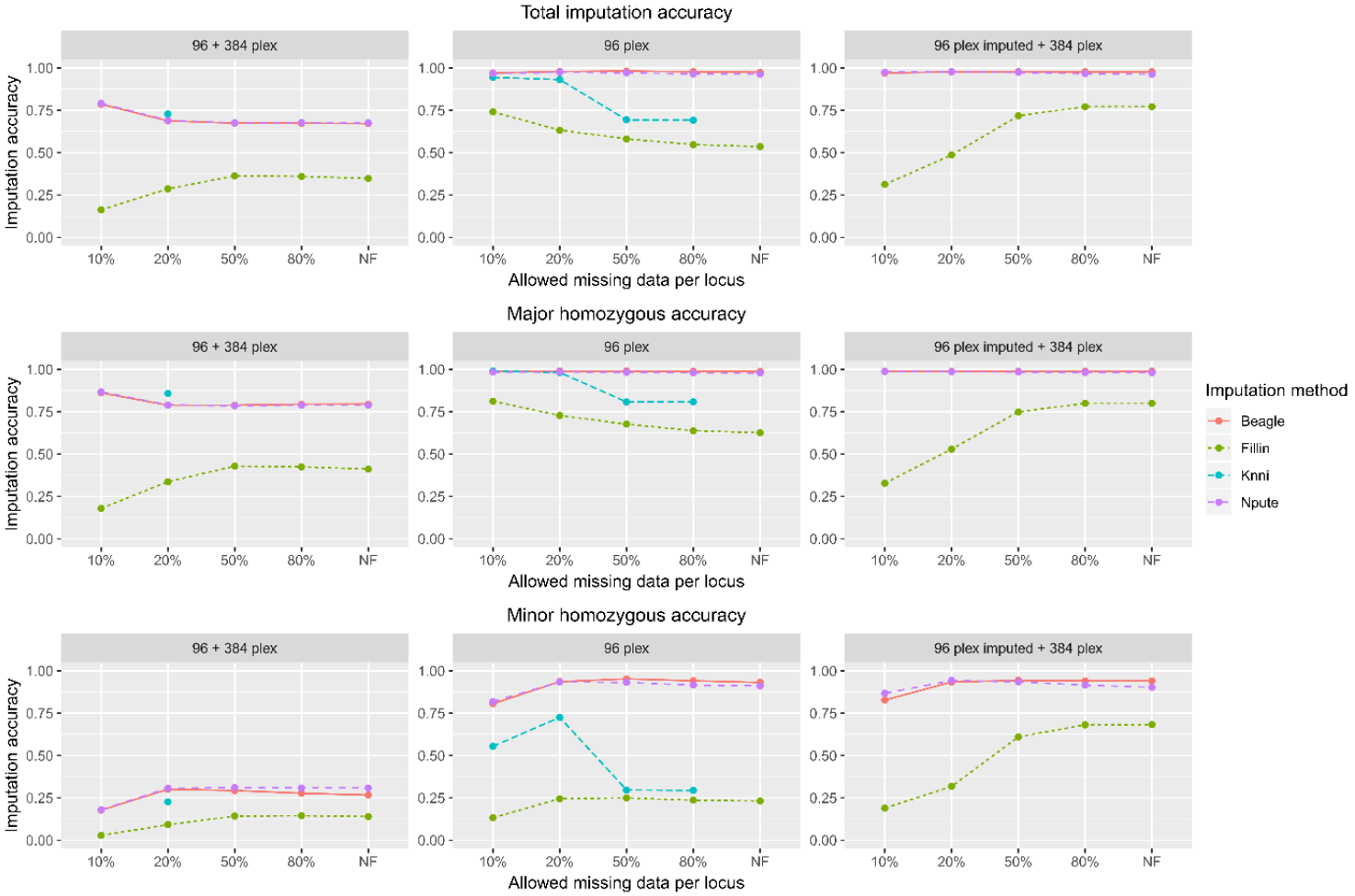


Figure 4 Imputation accuracy using SNP calling strategy II, i.e., SNP identification and genotype calling using all 1060 samples, removing heterozygous markers. Each row represents imputation accuracy for different genotype classes: total imputation accuracy, major homozygous accuracy and minor homozygous accuracy. Each column represents an imputation strategy: 96 + 384 plex, 96 plex and 96 plex imputed + 384 plex. The X-axis represents the different filters of allowed missing data per locus (10%, 20%, 50%, 80% and not filtered). Line colors represent the four imputation methods: Beagle (solid red), FILLIN (dotted green), KNNI (dashed blue) and NPUTE (dashed purple)

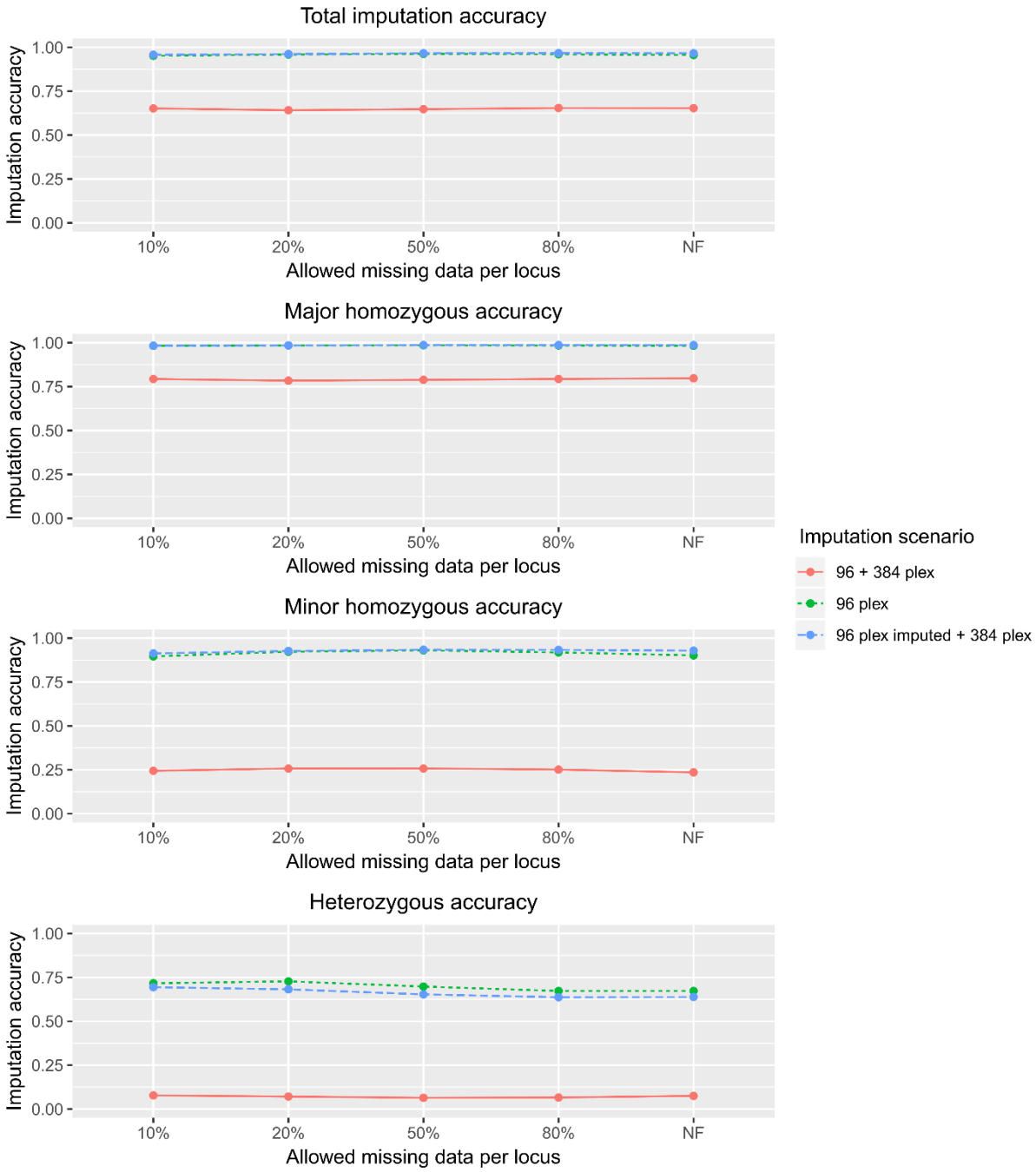


Figure 5 Imputation accuracy using SNP calling strategy I, i.e., SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines, not removing heterozygous markers for the Beagle imputation method. Each row represents imputation accuracy for different genotype classes: total imputation accuracy, major homozygous accuracy, minor homozygous accuracy and heterozygous accuracy. The X-axis represents the different filters of allowed missing data per locus (10%, 20%, 50%, 80% and not filtered). Line colors represent the three imputation scenarios: 96 + 384 plex (solid red), 96 plex (dotted green) and 96 plex imputed + 384 plex (dashed blue)

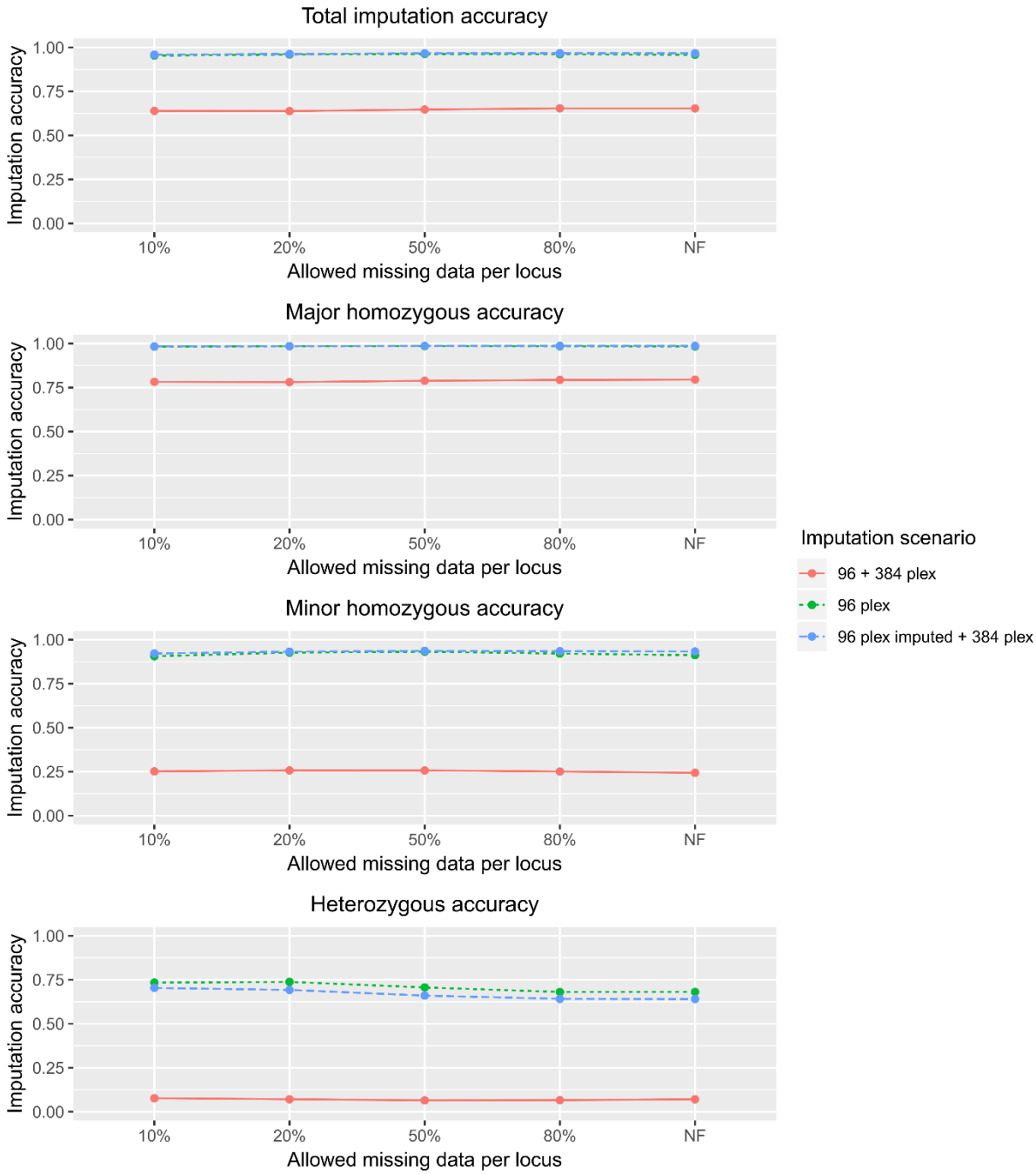


Figure 6 Imputation accuracy using SNP calling strategy II, i.e., SNP identification and genotype calling using all 1060 samples, not removing heterozygous markers for the Beagle imputation method. Each row represents imputation accuracy for different genotype classes: total imputation accuracy, major homozygous accuracy, minor homozygous accuracy and heterozygous accuracy. The X-axis represents the different filters of allowed missing data per locus (10%, 20%, 50%, 80% and not filtered). Line colors represent the three imputation scenarios: 96 + 384 plex (solid red), 96 plex (dotted green) and 96 plex imputed + 384 plex (dashed blue)

2.3.3 Computation time

We tracked the amount of time required to complete the imputation process for each method and imputation scenario, in each SNP calling strategy, with and without removing heterozygotes (Tables 1 and 2). The number of markers for the different filters of allowed missing data varied considerably, which reflected in the computation times. FILLIN required by far the least computational times (Table 1). Even in the most complex scenario, i.e., a higher number of markers and 96 plex imputed + 384 plex, the time required to complete the imputation process was never more than a few minutes. The second fastest method was Beagle, which however required noticeably more time in more complex scenarios (Tables 1 and 2). NPUTE was the slowest algorithm overall, except for some scenarios where KNNI was slower (Table 1).

In general, running times for the imputation strategy 96 plex were much lower than for situations that included the samples sequenced at lower depth. The second imputation strategy in amount of time required was the 96 + 384 plex. Finally, the 96 plex imputed + 384 plex imputation strategy largely exceeded the others in amount of time required to complete the imputation process (Tables 1 and 2).

Table 1 Running times for the different imputation methods for each allowed missing data per locus (10%, 20%, 50%, 80% and not filtered) and respective number of markers, in each imputation and SNP calling strategy, removing the heterozygous markers

SNP calling strategy*	Imputation strategy	Allowed missing data per locus	Number of markers	Imputation method	Running time (HH:MM:SS)
I	96 plex	10%	12957	Beagle KNNI NPUTE FILLIN	00:02:37 00:07:29 00:42:39 00:00:12
		20%	42053	Beagle KNNI NPUTE FILLIN	00:18:18 00:56:47 04:24:57 00:00:35
		50%	173328	Beagle KNNI NPUTE FILLIN	01:23:00 21:19:08 20:56:10 00:01:51
		80%	368351	Beagle KNNI NPUTE FILLIN	02:58:00 107:06:20 38:34:20 00:03:49
		NF	474367	Beagle KNNI NPUTE FILLIN	03:25:00 - 41:25:49 00:04:30
	96 plex imputed + 384 plex	10%	12957	Beagle KNNI NPUTE FILLIN	04:17:24 - 02:48:06 00:00:13
		20%	42053	Beagle KNNI NPUTE FILLIN	29:28:46 - 15:58:28 00:00:54
		50%	173328	Beagle KNNI NPUTE FILLIN	43:28:34 - 84:54:03 00:02:53
		80%	368351	Beagle KNNI NPUTE FILLIN	61:54:41 - 163:19:23 00:07:03
		NF	474367	Beagle KNNI NPUTE FILLIN	54:25:08 - 189:12:28 00:07:07

(continued)

Table 1 (continued)

SNP calling strategy*	Imputation strategy	Allowed missing data per locus	Number of markers	Imputation method	Running time (HH:MM:SS)
I	96 + 384 plex	10%	12957	Beagle KNNI NPUTE FILLIN	01:11:34 - 01:30:01 00:00:15
		20%	42053	Beagle KNNI NPUTE FILLIN	02:57:11 07:32:04 09:28:58 00:00:45
		50%	173328	Beagle KNNI NPUTE FILLIN	04:17:45 07:32:04 43:46:57 00:02:49
		80%	368351	Beagle KNNI NPUTE FILLIN	06:16:15 - 77:52:46 00:05:28
		NF	474367	Beagle KNNI NPUTE FILLIN	06:21:13 - 83:50:49 00:06:33
II	96 plex	10%	12957	Beagle KNNI NPUTE FILLIN	00:03:41 00:15:52 00:57:22 00:00:18
		20%	42053	Beagle KNNI NPUTE FILLIN	00:27:06 06:53:51 05:31:46 00:00:36
		50%	173328	Beagle KNNI NPUTE FILLIN	01:40:45 10:10:12 24:51:52 00:02:05
		80%	368351	Beagle KNNI NPUTE FILLIN	03:26:19 117:56:04 41:17:26 00:04:05
		NF	474367	Beagle KNNI NPUTE FILLIN	03:33:55 - 42:56:24 00:04:30

(continued)

Table 1 (continued)

SNP calling strategy*	Imputation strategy	Allowed missing data per locus	Number of markers	Imputation method	Running time (HH:MM:SS)
II	96 plex imputed + 384 plex	10%	12957	Beagle KNNI NPUTE FILLIN	03:14:19 - 03:00:25 00:00:20
		20%	42053	Beagle KNNI NPUTE FILLIN	18:47:42 - 16:42:40 00:00:50
		50%	173328	Beagle KNNI NPUTE FILLIN	42:42:46 - 78:20:06 00:03:07
		80%	368351	Beagle KNNI NPUTE FILLIN	69:25:25 - 155:48:01 00:07:41
		NF	474367	Beagle KNNI NPUTE FILLIN	57:18:19 - 179:39:53 00:09:06
	96 + 384 plex	10%	12957	Beagle KNNI NPUTE FILLIN	01:41:32 - 02:52:14 00:00:27
		20%	42053	Beagle KNNI NPUTE FILLIN	03:02:00 13:10:43 15:59:20 00:01:13
		50%	173328	Beagle KNNI NPUTE FILLIN	04:26:05 - 48:18:38 00:04:14
		80%	368351	Beagle KNNI NPUTE FILLIN	06:28:54 - 74:41:06 00:08:41
		NF	474367	Beagle KNNI NPUTE FILLIN	08:09:11 - 81:20:55 00:09:37

*SNP calling strategy I: SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines. SNP calling strategy II: SNP identification and genotype calling using all 1060 samples

Table 2 Running times for Beagle for each allowed missing data per locus (10%, 20%, 50%, 80% and not filtered) and respective number of markers, in each imputation and SNP calling strategy, considering the heterozygous markers

SNP calling strategy*	Imputation strategy	Allowed missing data per locus	Number of markers	Running time (HH:MM:SS)
I	96 plex	10%	12957	00:27:02
		20%	42053	01:08:07
		50%	173328	04:01:21
		80%	368351	04:51:47
		NF	474367	05:20:26
	96 plex imputed + 384 plex	10%	12957	02:28:25
		20%	42053	05:29:07
		50%	173328	11:43:59
		80%	368351	17:24:06
		NF	474367	21:10:02
	96 + 384 plex	10%	12957	01:28:59
		20%	42053	02:56:30
		50%	173328	07:33:40
		80%	368351	08:29:33
		NF	474367	11:16:18
II	96 plex	10%	17508	00:21:33
		20%	50793	01:00:56
		50%	187440	03:12:39
		80%	380955	06:25:55
		NF	443940	07:13:19
	96 plex imputed + 384 plex	10%	17508	03:08:15
		20%	50793	06:55:54
		50%	187440	13:15:22
		80%	380955	20:52:30
		NF	443940	21:48:30
	96 + 384 plex	10%	17508	01:47:01
		20%	50793	03:17:01
		50%	187440	06:44:40
		80%	380955	12:24:20
		NF	443940	12:41:05

*SNP calling strategy I: SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines. SNP calling strategy II: SNP identification and genotype calling using all 1060 samples

2.4 Discussion

Results from our study indicate that combining SNP calling and imputation strategies can enhance cost effective genotyping, resulting in higher imputation accuracies. These approaches thus allow a more widespread adoption of genomic selection and genome-wide association studies in plant breeding programs. The different SNP calling strategies aimed to better explore the high and low multiplexing levels of our dataset and yielded different number of markers. Even after we removed SNPs with MAF less than 5% and inbreeding coefficient less than 0.8, the SNP calling strategy I, i.e., using only the high coverage dataset

to discover SNPs, produced 30 thousand more markers than the alternative scheme. Across the MAF plots, we observed that minor allele frequencies closer to zero were less frequent using only lines genotyped with 96 samples per sequencing lane (Figure 2). The higher coverage dataset likely enabled greater power of detection and less false positives. With regard to the inbreeding coefficient, both SNP calling strategies showed similar patterns, with slightly higher values with SNP calling strategy II (Figure 2). It is more difficult to call heterozygous SNP with the lower depth dataset, such that the homozygous genotypes tend to be called more frequently (Swarts et al. 2014). Overall, the SNP calling strategies did not greatly affect the imputation accuracy, but had influence on the number of markers found (Figures 3 – 6).

In SNP genotype imputation it is important to evaluate not only the total imputation accuracy, but also the per-class accuracy. Data balancedness refers to the ratio that each genotype class (AA, AB, BB) appears. Classification problems are more difficult when the data is unbalanced, that is, the three classes appear at different frequencies in the dataset. Data balancedness is directly related to MAF, because very low MAFs arise when a class is underrepresented (Hickey et al. 2012; Nazzicari et al. 2016). Insofar as the classes of missing genotypes appear at different frequencies, the total imputation accuracy can be dominated by the most frequent class. The imputation accuracy tends to be higher for the more frequent genotype class, and the overall imputation accuracy will predominantly represent the imputation accuracy of that class. Indeed, we found significantly higher error rates in the less frequent class for all imputation methods.

Beagle and NPUTE produced the best imputation results with accuracies close to 100% in the imputation strategies 96 plex and 96 plex imputed + 384 plex, in most scenarios of missing data. The KNNI method did not work in most evaluated scenarios. With large amounts of missing data, the complexity of the imputation problem increases and complicates the identification of k neighbors that are close enough to the data point to be imputed (Nazzicari et al. 2016). Probably as a consequence of the curse of dimensionality (Marimont and Shapiro 1979), scenarios with large amounts of missing data could not be imputed with KNNI. FILLIN performed poorly in all tested scenarios, which may be explained by the fact that this algorithm is optimized for homogeneous inbred populations, while our dataset consists of a collection of lines from different heterotic groups. Similar results using Beagle, KNNI and FILLIN were observed in a study with GBS data from rice and alfalfa (Nazzicari et al. 2016).

The allowed missing data per locus did not reflect on the imputation accuracy for Beagle and NPUTE methods (Figure 3 and 4). Nonetheless, in the 96 plex imputation strategy, with larger quantities of missing genotypes, KNNI showed a decrease in imputation accuracy. In the 96 plex imputation strategy, FILLIN also showed decreasing imputation accuracy with increasing missing rates for total and major homozygous imputation accuracies. However, overall in more stringent scenarios, with only 10 to 20% allowed missing data, imputation accuracy for all classes in the scenarios 96 plex imputed + 384 plex and 96 + 384 plex, as well as the minor homozygous accuracy in the scenario 96 plex were reduced for the FILLIN method.

Despite working with inbred maize lines, some of them were in the initial stages of the breeding program (F3 – F4) and were not yet completely endogamic. Including heterozygous genotypes complicated the imputation problem, because this dataset showed relatively few heterozygotes, which are more susceptible to genotyping errors. As a consequence, the heterozygous accuracy was considerably lower than for both homozygote classes (Figure 3 and 4).

The complexity of the problem directly affected the running time required to complete the imputation process. KNNI and NPUTE were the most demanding methods, with computation times growing both with the number of markers and with the number of missing genotypes to be imputed. The 96 plex imputed + 384 plex imputation strategy exceeded substantially the others in amount of time required. We believe that, despite the smaller amount of missing data to impute, the initial step of identifying the haplotypes is likely more time consuming because there are more informative loci. Considering both imputation accuracy and computational time, the best imputation method was Beagle (Tables 1 and 2). In addition, this method allows for heterozygote genotypes, which is an interesting feature for panels that include individuals with few generations of inbreeding.

Several works have explored imputation strategies combining high and low density genotyping (Hickey et al. 2012, 2015; Huang et al. 2012; Mulder et al. 2012; Gorjanc et al. 2017a; Gonen et al. 2018). These studies, however, do not focus on combining SNP calling and imputation strategies, using real GBS data. In this paper, we investigated the impact on imputation accuracies of combing different SNP calling and imputation strategies, using a real dataset of lines from a maize breeding program genotyped with GBS. We believe that our study is a first stage of what can be done regarding SNP calling and imputation strategies with GBS data. Further research is necessary, for example, to set out the number of high coverage

individuals necessary to ensure high imputation accuracy of low coverage individuals. We suggest that some key individuals could be genotyped using lower multiplexing levels, while others might be included in larger pools. This set of key individuals should be well thought out to represent the entire diversity of heterotic groups in the breeding program.

Our results indicate that designing the SNP calling and imputation strategies in order to better explore the different depths of coverage considerably improves the imputation accuracy, besides reducing costs, since higher multiplexing levels are considerably cheaper. Bringing together SNP calling strategies using only high coverage data to discover variants, followed by genotype calling for all sequenced samples, with the imputation strategy 96 plex imputed + 384 plex produced the larger number of SNPs and higher imputation accuracies. These combined strategies encompass a wide range of applications in breeding programs, representing an opportunity to reduce costs and time by optimizing the genotyping process.

References

Bouwman AC, Hickey JM, Calus MP, Veerkamp RF (2014) Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle. Genet Sel Evol 46:6. https://doi: 10.1186/1297-9686-46-6

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126. https://doi: 10.1016/j.ajhg.2015.11.020

Browning BL, Yu Z (2009) Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. Am J Hum Genet 85:847–861. https://doi: 10.1016/j.ajhg.2009.11.004

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097. https://doi: 10.1086/521987

Cericola F, Lenk I, Fè D, et al (2018) Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (*Lolium perenne* L.). Front Plant Sci 9:369. https://doi: 10.3389/fpls.2018.00369

Chang L-Y, Toghiani S, Ling A, et al (2018) High density marker panels, SNPs prioritizing and accuracy of genomic selection. BMC Genet 19:4. https://doi: 10.1186/s12863-017-0595-2

Cleveland MA, Hickey JM, Kinghorn BP (2011) Genotype imputation for the prediction of genomic breeding values in non-genotyped and low-density genotyped individuals. BMC Proc 5:S6. https://doi: 10.1186/1753-6561-5-S3-S6

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology (Basel) 1:460–483. https://doi: 10.3390/biology1030460

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601. https://doi: 10.1016/j.tplants.2014.05.006

Dias KODG, Alejandro Gezan S, Teixeira Guimarães C, et al (2018) Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity 121:24–37. https://doi: 10.1038/s41437-018-0053-6

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1–10. https://doi: 10.1371/journal.pone.0019379

Faville MJ, Ganesh S, Cao M, et al (2018) Predictive ability of genomic selection models in a multi-population perennial ryegrass training set using genotyping-by-sequencing. Theor Appl Genet 131:703–720. https://doi: 10.1007/s00122-017-3030-1

Gerard GS, Kobiljski B, Lohwasser U, et al (2018) Genetic architecture of adult plant resistance to leaf rust in a wheat association mapping panel. Plant Pathol 67:584–594. https://doi: 10.1111/ppa.12761

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:1–11. https://doi: 10.1371/journal.pone.0090346

Gonen S, Wimmer V, Gaynor RC, et al (2018) A heuristic method for fast and accurate phasing and imputation of single nucleotide polymorphism data in bi- parental plant populations. Theor Appl Genet 131:2345–2357. https://doi.org/10.1007/s00122-018-3156-9

Gorjanc G, Battagin M, Dumasy J-F, et al (2017a) Prospects for cost-effective genomic selection via accurate within-family imputation. Crop Sci 57:216–228. https://doi: 10.2135/cropsci2016.06.0526

Gorjanc G, Dumasy J-F, Gonen S, et al (2017b) Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. Crop Sci 57:1404. https://doi: 10.2135/cropsci2016.08.0675

Haile JK, N'diaye A, Clarke F, et al (2018) Genomic selection for grain yield and quality traits in durum wheat. Mol Breed 38:75. https://doi: 10.1007/s11032-018-0818-x

Han S, Miedaner T, Utz FU, et al (2018) Genomic prediction and GWAS of Gibberella ear rot resistance traits in dent and flint lines of a public maize breeding program. Euphytica 214:1–20. https://doi: 10.1007/s10681-017-2090-2

Hickey J, Crossa J, Babu R, de los Campos G (2012) Factors affecting the accuracy of genotype imputationin populations from several maize breeding programs. Crop Sci 52:654–663. https://doi: 10.2135/cropsci2011.07.0358

Hickey JM, Gorjanc G, Varshney RK, Nettelblad C (2015) Imputation of single nucleotide polymorphism genotypes in biparental, backcross, and topcross populations with a Hidden Markov Model. Crop Sci 55:1934. https://doi: 10.2135/cropsci2014.09.0648

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next neneration of genome-wide association studies. PLoS Genet 5:e1000529. https://doi: 10.1371/journal.pgen.1000529

Huang Y, Hickey JM, Cleveland MA, Maltecca C (2012) Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Genet Sel Evol 44:1. https://doi: 10.1186/1297-9686-44-25

Jacobson A, Lian L, Zhong S, Bernardo R (2015) Marker imputation before genomewide selection in biparental maize populations. Plant Genome. https://doi: 10.3835/plantgenome2014.10.0078

Kayondo SI, Pino Del Carpio D, Lozano R, et al (2018) Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. Sci Rep 8:1549. https://doi: 10.1038/s41598-018-19696-1

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi: 10.1038/nmeth.1923

Law M, Childs KL, Campbell MS, et al (2015) Automated update, revision, and quality control of the Maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. Plant Physiol 167:25–39. https://doi: 10.1104/pp.114.245027

Liu EY, Li M, Wang W, Li Y (2013) MaCH-Admix: Genotype imputation for admixed populations. Genet Epidemiol 37:25–37. https://doi: 10.1002/gepi.21690

Marimont RB, Shapiro MB (1979) Nearest Neighbour searches and the curse of dimensionality. IMA J Appl Math 24:59–70. https://doi: 10.1093/imamat/24.1.59

Mulder HA, Calus MPL, Druet T, Schrooten C (2012) Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. J Dairy Sci 95:876–889. https://doi: 10.3168/jds.2011-4490

Nazzicari N, Biscarini F, Cozzi P, Brummer EC (2016) Marker imputation efficiency for genotyping-by-sequencing data in rice (*Oryza sativa*) and alfalfa (*Medicago sativa*). Mol Breed 36:1–16. https://doi: 10.1007/s11032-016-0490-y

Poland J a, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome J 5:92–102. https://doi: 10.3835/plantgenome2012.05.0005

R Development Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Roberts A, McMillan L, Wang W, et al (2007) Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. Bioinformatics 23:401–407. https://doi: 10.1093/bioinformatics/btm220

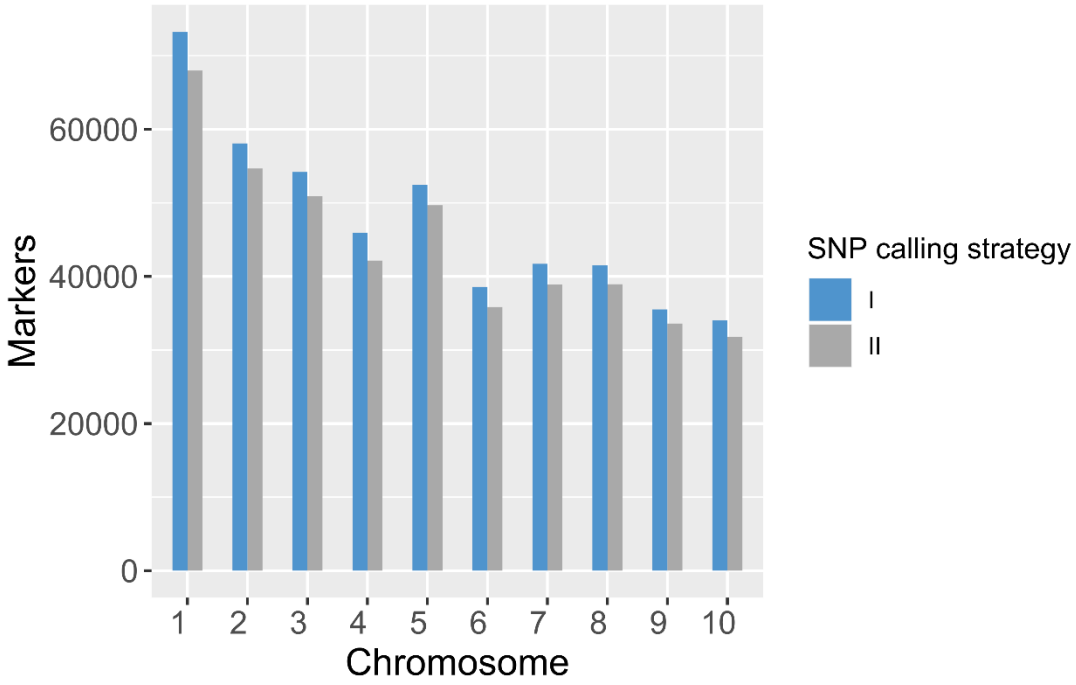
Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Population Biology Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics (ribosomal DNA spacer-length variation/restriction fragment-length polymorphisms/Rrnl/Rrn2). Proc Natl Acad Sci 81:8014–8018.

Schwender H, Fritsch A (2015) Scrime: Analysis of high-dimensional categorical data such as SNP data.

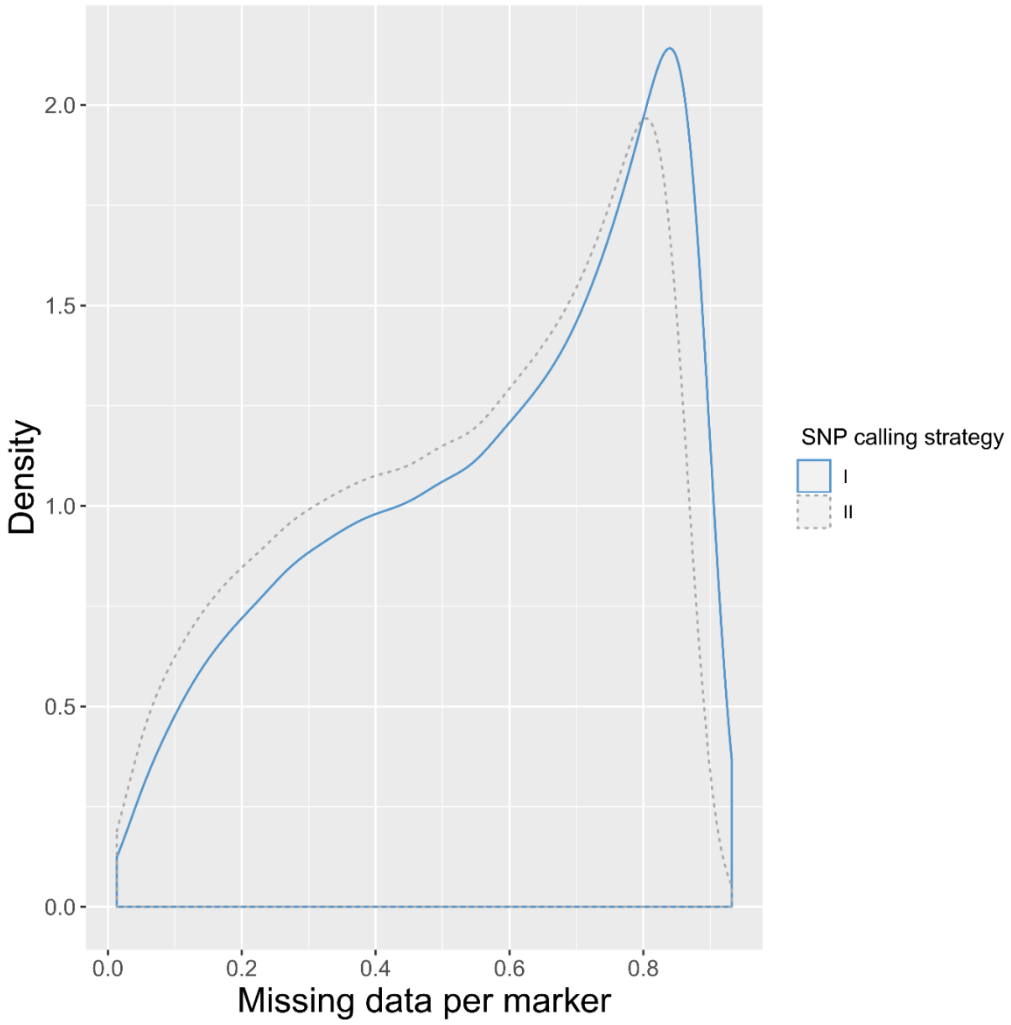
Swarts K, Li H, Romero Navarro JA, et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome. https://doi: 10.3835/plantgenome2014.05.0023

Troyanskaya O, Cantor M, Sherlock G, et al (2001) Missing value estimation methods for DNA microarrays. Bioinformatics 17:520–525.

Supporting information



Supplementary Figure 1 Number of markers found per chromosome for two alternative SNP calling strategies. Blue bars correspond to SNP calling strategy I, i.e., SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines. Gray bars correspond to SNP calling strategy II, i.e., SNP identification and genotype calling using all 1060 samples



Supplementary Figure 2 Missing data per discovered marker for two alternative SNP calling strategies. Solid blue line corresponds to SNP calling strategy I, i.e., SNP identification using only the 680 lines genotyped with 96 samples per sequencing lane, followed by genotype calling with all 1060 lines. Dashed gray line corresponds to SNP calling strategy II, i.e., SNP identification and genotype calling using all 1060 samples

3 GENOMIC PREDICTION APPLIED TO MULTIPLE TRAITS AND ENVIRONMENTS IN SECOND SEASON MAIZE HYBRIDS

Abstract

Genomic selection has become a reality with the reduction in genotyping costs. Especially in maize breeding programs, it emerges as a promising tool for predicting hybrid performance. The dynamics of a commercial breeding program involve the evaluation of several traits simultaneously in a large set of environments. Therefore, multi-trait multi-environment (MTME) genomic prediction models can leverage these data sets by exploring the correlation between traits and GxE interaction. Herein, we assess predictive abilities of univariate and multivariate genomic prediction models in a maize breeding program. To this end, we used data from 415 maize hybrids evaluated in four years of second season field trials for the traits grain yield, number of ears and grain moisture. Genotypes of these hybrids were inferred *in silico* based on their parental inbred lines using single nucleotide polymorphism markers obtained via genotyping-by-sequencing. Because genotypic information was available for only 257 hybrids, we used the single-step procedure to obtain the **H** matrix for all 415 hybrids, combining pedigree and genomic relationship information. Our results demonstrated that the use of MTME models improved the predictive abilities, especially to predict the performance of hybrids that have not been evaluated in any environment. However, the computational requirements of this kind of model could represent a limitation to its practical implementation and further investigation to resolve this bottleneck is necessary.

Keywords: Genotype-by-environment interaction; Genetic correlation; GBLUP; Multivariate models

3.1 Introduction

The Brazilian maize production is currently concentrated in the second season, from February to June, representing more than 66% of the production in the 2017/2018 season (Conab, 2019). The second season is an alternative crop rotation system in the Center-South region, with maize grown mostly after soybean, contributing to a greater profitability of the Brazilian agribusiness. However, this ensuing season poses some challenges, such as diseases and, especially, water deficiency stress. Much of the progress made in growing maize in the second season is due to genetic improvement in drought tolerance. Due to climate changes and the limitation of water resources, yield stability even under water stress is a highly desirable feature in agriculture nowadays (Cooper et al. 2014).

Maize breeding programs for the second season target genotypes that are highly productive under normal growing conditions, but which are able to maintain good performance even under conditions of water scarcity. The biggest challenge faced by breeders remains in the fact that grain yield is a quantitative trait, strongly influenced by environmental

effects and showing low heritability under stress conditions (Comstock 1978; Hallauer and Miranda Filho 2010). Therefore, to increase the experimental precision of phenotypic evaluations under water deficiency, a large number of replicates and adequate plot sizes are required (Edmeades et al. 1999; Bãnziger et al. 2000). However, phenotyping accounts for a large part of the cost of a plant breeding program, limiting progress by restricting the number of evaluated genotypes and the sizes of experiments.

A large number of hybrids can be obtained from the cross of a relatively small number of lines in a maize breeding program (Technow et al. 2014). Due the financial unfeasibility of evaluate all these possible hybrids in field trials, predicting hybrid performance through genomic selection is an attractive alternative to maize breeders. Since proposed by Meuwissen et al. 2001, genomic selection models have been applied to a variety of crops and become an important tool in maize hybrid breeding (Bernardo 2009; Massman et al. 2013; Dias et al. 2018; Fristche-Neto et al. 2018; Han et al. 2018). Besides the opportunity to reduce costs and labor involved in field trials, this approach allows an early and more efficient selection, increasing genetic gains. These models were initially applied in a univariate context, by using a separate model for a single environment and a single trait. However, breeders commonly evaluate several traits simultaneously in a large set of environments, because elite genotypes should concentrate favorable alleles for various traits of interest and perform well in different target environments. As a prime consequence, the use of univariate approaches does not match the reality of many programs that seek to estimate the magnitude of the Genotype-by-Environment (GxE) interaction and explore the genetic correlation between important agronomic traits.

The presence of genetic correlation between quantitative traits implies that measures in one trait indirectly provide information about other traits, which can be used to improve the predictive ability of genomic selection. However, univariate genomic selection models do not take advantage of this shared information. Multivariate genomic selection models, known as multi-trait models, allow the information between secondary traits to be explored through modeling of the covariance between them. The main factors that have been reported to contribute to increasing predictive ability of multi-trait models are: traits highly correlated with the trait of interest, low heritability coefficients for the target trait, but high for the correlated trait (Calus and Veerkamp 2011; Jia and Jannink 2012; Guo et al. 2014; Dos Santos et al. 2016; Marchal et al. 2016; Lyra et al. 2017; Covarrubias-Pazaran et al. 2018). Grain yield is the trait of major interest in a maize breeding program and it is a direct function of its

components: number of ears per plant, number of rows of grain in the ear, number of grains per row, ear length, ear diameter, average grain weight and grain depth (Jugenheimer 1976). Because these traits are less complex than the grain yield, with higher heritability coefficients, and are highly correlated to it, they are feasible to perform indirect selection for grain yield. Several studies reported higher predictive abilities using models that consider high-heritability secondary trait information, in addition to grain yield (Henderson and Quaas 1976; Mrode and Thompson 2005; Malosetti et al. 2008; Piepho et al. 2008).

In addition to the correlation between traits, the relationship among environments, in terms of the GxE interaction patterns, is also a relevant issue to plant breeders. Burgueño et al. 2012 were the first to accommodate the GxE interaction in the context of genomic selection. Following this study, other also examined the possibility of increasing the predictive ability in several crops by incorporating the GxE interaction (Lopez-Cruz et al. 2015; Cuevas et al. 2016; Ferrão et al. 2017; Sousa et al. 2017; Roorkiwal et al. 2018). Proper understanding of GxE interaction provides valuable information and can help breeders to predict completely untested combinations of hybrids and environments through the use of cross validation schemes as proposed by (Burgueño et al. 2012).

The large amount of phenotypic data collected in breeding programs across years is a valuable source of information, of which genomic selection is recently taking advantage. Nonetheless, the quality and unbalanced nature of these historical data raise a new challenge to plant breeders - how to optimally exploit this kind of data (Gapare et al. 2018). Few studies have simultaneously assessed multi-trait and multi-environment (MTME) models for genomic selection (Montesinos-López et al. 2016; Gomes Torres et al. 2018; Ward et al. 2019). Therefore, our objectives were to: *i*) evaluate the applicability of a MTME model, *ii*) compare the results of using this model with its univariate counterparts, *iii*) predict completely new and untested hybrids and years. Historical data from three traits from second season maize hybrids of the Embrapa breeding program were used to this end.

3.2 Materials and methods

3.2.1 Plant material

The genetic material consisted of 415 hybrids evaluated in field trials for four years (2006-2009) in Campo Mourão, Paraná, Brazil. Of the 415 hybrids, 304 are single cross

hybrids, 76 are triple cross hybrids, 19 are double cross hybrids and 16 are commercial checks. The experimental design of the phenotypic trials from 2006 to 2008 was a 10 × 10 squared lattice design with two replicates, where 100 hybrids were evaluated. In 2009, 125 hybrids were evaluated side-by-side in two trials. In each trial, 60 hybrids and four common checks were evaluated using an 8 × 8 lattice design with two replicates. The connection across years was based in a few common checks (Table S1).

The evaluated traits were grain yield (GY), determined by weighing all the grains in each plot, adjusted to 13% of grain moisture and converted to tons per hectare (t/ha); number of ears (NE), consisted of counting all ears in each plot; and percentage of grain moisture (GM), assessed with the Wintersteiger Classic Plot Combine automatic harvester (Wintersteiger AG, Austria), which automatically weighs each parcel and infers the moisture via NIRS (near-infrared spectroscopy)..

3.2.2 Phenotypic analysis

We computed the best linear unbiased estimation (BLUE) for each trial and trait, using the following mixed model:

$$y_{ijk} = \mu + r_k + g_i + b_{j(k)} + \varepsilon_{ijk} \qquad [1]$$

where y_{ijk} is the phenotype of the ith genotype in block j, replicate k; μ is the common intercept; r_k is the fixed effect of replicate k; g_i is the fixed effect of the ith genotype; $b_{j(k)}$is the random effect of block j, in replicate k; and ε_{ijk} is a random non-genetic effect. Outliers were removed by deleting observations with residuals that deviated more than four times the standard deviation.

The broad-sense and narrow-sense heritability were computed based on model [1], but considering the genotype effects as random. Also, in order to compute the narrow-sense heritability we assumed that $g_i \sim MVN(0, \boldsymbol{H})$, where $\boldsymbol{H}$ represents the relationship matrix of additive effects. We then estimated the heritability based on the following equation:

$$h^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)} \qquad [2]$$

where σ_g^2 is the total genetic variance component and the additive variance component for the broad-sense and narrow-sense heritability, respectively, and σ_e^2 is the residual variance component. The BLUE and heritability coefficients were estimated using the ASREML-R package version 3.0 (Butler et al. 2009) in the R environment v.3.5.1 (R Core Team 2018).

3.2.3 Genotypic data

A collection of 1060 maize inbred lines from the Embrapa Maize and Sorghum breeding program were genotyped, of which 228 are parents of hybrids used in this study. We performed DNA extraction from young leaves based on the cetyltrimethylammonium bromide method (Saghai-Maroof et al. 1984). DNA samples were quantified using the Fluorometer Qubit® 2.0, following the manufacturer's instructions (Life TechnologiesTM, USA). Samples were also evaluated on 1% agarose gel in Tris-acetate-EDTA buffer, stained with GelRed™ (Biotium, USA) and recorded under UV light in the Imager Gel Doc L-PIX (Loccus Biotecnologia, Brazil). Genotyping-by-sequencing (GBS) was carried out at the Genomic Diversity Facility at Cornell University (Ithaca, NY, USA) using the standard GBS protocol (Elshire et al. 2011) with the ApeKI restriction enzyme. The inbred lines were genotyped in two different batches: first, we genotyped eight libraries of 96 samples each, with one HiSeq 2500 sequencing lane per library; next, we genotyped one library of 384 samples with NextSeq500 in a single lane. Tags were aligned to the B73 reference genome (AGPv3) (Law et al. 2015) using the Bowtie2 aligner (Langmead and Salzberg 2012). Then, single nucleotide polymorphisms (SNPs) were called using the GBSv2 Discovery Pipeline, available in the software TASSEL v. 5.2.28 (Glaubitz et al. 2014), using SNP calling strategy I described in Chapter 2. We applied filters for Minor Allele Frequency (MAF) less than 5% and inbreeding coefficient less than 0.8. Subsequently, we performed imputation of missing data using Beagle software version 4.1 (Browning and Browning 2016). Before imputation, we removed indels, non-biallelic SNPs and considered heterozygous loci as missing data. Because Beagle can introduce heterozygous genotypes, after imputation we again removed heterozygous loci. Finally, from the genotypes of the 228 parental lines we inferred the genotypes of 257 single cross hybrids.

3.2.4 H matrix

The hybrids used in this study were originated from 296 inbred lines that belong to three different heterotic groups: Dent (116 lines), Flint (126 lines) and another group herein denominated group C (54 lines). Pedigree information was available for all 415 hybrids, but only 257 of those were (indirectly) genotyped. In this situation, the use of the single-step approach, where the pedigree relationship matrix **A** and the genomic relationship matrix **G** are combined into one matrix called **H**, is a practical way to combine these two sources of information (Legarra et al. 2009; Misztal et al. 2009; Aguilar et al. 2010; Christensen and Lund 2010).

The genomic relationship matrix **G** was computed following the method described by Yang et al. (2010). The pedigree relationship matrix **A** was computed based on Henderson's recursive method described in Mrode (2005). Both **G** and **A** matrices were obtained using the R package AGHMATRIX (Amadeu et al. 2016). We implemented the **H** matrix using the two scaling factors, τ and ω, as proposed by Misztal et al. (2010) and Tsuruta et al. (2011):

$$\mathbf{H}_{\tau,\omega}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & (\tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}) \end{pmatrix} \qquad [3]$$

We further evaluated the effect of these factors on the accuracies of genomic prediction models. Using the approach presented by Martini et al. (2018), we searched for the optimal values of τ and ω by evaluating 420 combinations, varying both parameters on grids defined by the intervals [−1, 1] for ω and [0.1, 2] for τ, in steps of size 0.10 in both cases. To evaluate the performance of each parameter combination, we constructed 420 different **H** matrices in R, one for each combination of the scaling factors, and used these to estimate the breeding values using the single-step procedure for each single-trait single-environment (STSE) model fitted.

3.2.5 Genomic prediction models

The genomic prediction model used in this study was the GBLUP (Genomic Best Linear Unbiased Prediction) (VanRaden 2008). We fitted univariate and multivariate models via a Bayesian approach, as detailed below.

3.2.5.1 Single-trait single-environment model

Using the STSE model the genomic estimated breeding values (GEBV) were obtained for each of the three traits evaluated, separately for each of the four environments, as follows:

$$y_i = \mu + G_i + \varepsilon_i \qquad [4]$$

where y_i is the previously obtained BLUE of the ith genotype ($i = 1, \dots, n$), where n indicates the number of hybrids evaluated in the environment at hand; μ is the intercept; G_{i} is the random effect of the ith genotype, such that $G_i \sim MVN(0, \boldsymbol{H})$; and ε_i is a random non-genetic effect, with $\varepsilon_i \sim MVN(0, \boldsymbol{I})$. $\boldsymbol{H}$ represents the relationship matrix of additive effects and $\boldsymbol{I}$ is an identity matrix for the residual effects.

3.2.5.2 Multi-trait single-environment model

Combining information of the three evaluated traits, separately in each of the four environments, we obtained the GEBV using the multi-trait single-environment (MTSE) model:

$$y_{ic} = \mu + G_{ic} + \varepsilon_{ic} \qquad [5]$$

where y_{ic} is the BLUE of the ith genotype for trait c ($c = 1, \dots, 3$); μ is the intercept; G_{ic} is the random effect of the ith genotype for trait c, $G_{ic} \sim MVN(0, \boldsymbol{H} \otimes \boldsymbol{\Sigma_c})$; and ε_i is a random non-genetic effect, $\varepsilon_{ic} \sim MVN(0, \boldsymbol{I} \otimes \boldsymbol{R_c})$. In this model, $\boldsymbol{\Sigma_c}$ is the variance-covariance (VCOV) matrix for the additive genetic effects of the three traits, with dimension 3×3. $\boldsymbol{R_c}$ represents the VCOV matrix for the residual effects of the three traits, also with dimension 3×3. We assumed an unstructured form for the genetic $\boldsymbol{\Sigma_c}$ and residual $\boldsymbol{R_c}$ VCOV matrices, which allows the assumption of heterogeneity of variance and presence of a specific genetic correlation for each combination of trait and environment.

3.2.5.3 Single-trait multi-environment model

Through the single-trait multi-environment (STME) model, we obtained the GEBV separately for each of the three traits, but jointly modeling the four environments (years), as follows:

$$y_{ij} = \mu + G_{ij} + \varepsilon_{ij} \quad [6]$$

where y_{ij} is the BLUE of the ith genotype, in the jth environment ($j = 1, \ldots, 4$); μ is the intercept; G_{ij} is the random effect of the ith genotype, in the jth environment, with $G_{\text{ij}} \sim MVN(0, \boldsymbol{H} \otimes \boldsymbol{\Sigma_j})$; and ε_{ij} is a random non-genetic effect, such that $\varepsilon_{ij} \sim MVN(0, \boldsymbol{I} \otimes \boldsymbol{R_j})$. $\boldsymbol{\Sigma_j}$ is the VCOV matrix for the additive genetic effects in the j environments, with dimension 4×4. $\boldsymbol{R_j}$ represents the VCOV matrix for the residual effects in the j environments, with dimension 4×4. We again assumed an unstructured form for the genetic $\boldsymbol{\Sigma_j}$ and residual $\boldsymbol{R_j}$ VCOV matrices.

3.2.5.4 Multi-trait multi-environment model

In our most complex model, for MTME genomic selection, we jointly modeled all traits and environments, in order to obtain the GEBV for each trait in each environment:

$$y_{ijc} = \mu + G_{ijc} + \varepsilon_{ijc} \quad [7]$$

where y_{ijc} is the previously obtained BLUE of the ith genotype, in the jth environment, for trait c; μ is the common intercept; G_{ijc} is the random effect of the ith genotype, in the jth environment, for the trait c, $G_{ijc} \sim MVN(0, \boldsymbol{H} \otimes \boldsymbol{\Sigma_{jc}})$; and ε_{ijc} is a random non-genetic effect, $\varepsilon_{ijc} \sim MVN(0, \boldsymbol{I} \otimes \boldsymbol{R_{jc}})$. $\boldsymbol{\Sigma_{jc}}$ represents the VCOV matrix for the additive genetic effects in the four environments for the three traits, with dimension 12×12. In this case, this matrix models variances and covariances for all combinations of traits and environments. Similarly, $\boldsymbol{R_{jc}}$ represents the VCOV matrix for the residual effects in each trait × environment combination, with dimension 12×12. We assumed an unstructured form for the genetic and residual VCOV matrices.

3.2.6 Cross validation schemes

To assess the performance of each model we used the predictive ability as measured by cross-validation. To this end, we implemented three different schemes. In our first cross-validation scheme (hereinafter denoted as CV_R), the complete pool of individuals was randomly split in five folds, such that four of them were used as a training set, while the remaining group was used as a testing set. This procedure was repeated five times, using a different set of individuals as the testing set each time. Therefore, at the end of the process all individuals had their GEBV. This CV_R scheme was applied to models 4, 5, 6 and 7 as described above. In order to compare the models applied to single-environments with those applied to multi-environments, we also evaluated the CV_R scheme using the same training/testing partition of single-environment models to the multi-environment ones.

In the multi-environment context, as described above for models 6 and 7, we evaluated two different cross-validation schemes in order to take advantage of the correlated information between environments, according to the ideas presented by Burgueño et al. (2012). First, we aimed to measure the ability of the model to predict the performance of hybrids that have not been evaluated in any environment (hereinafter denoted as CV_1). In CV_1 we randomly assigned the hybrids to a 5-folds scheme, but in this case ensuring that hybrids in the testing set had not been evaluated in any environment. Alternatively, to assess the ability of the model to predict performance based on data from different years we assigned years to folds (such scheme is hereinafter denoted as CV_2). In CV_2 we had as many folds as years, thus when analyzing the ith fold, hybrids from the ith year were assigned to the testing set and all the hybrids from other years were used as the training set. In all cross-validation schemes, the predictive abilities were estimated by Pearson's correlation coefficient between the GEBV and the corresponding BLUE.

3.2.7 Computational implementation

All genomic prediction models were implemented in the MCMCGLMM R-package (Hadfield 2010). A total of 30000 MCMC samples were generated, assuming a burn-in period and sampling interval (thin) of 6000 and 5 iterations, respectively. To check the convergence of the models we used the Geweke criteria (Geweke 1992) implemented in the coda R-package, as well as the visual inspection of trace plots of the chains.

3.3 Results

3.3.1 H matrix

Based on the evaluation of the 420 combinations of τ and ω scaling factors for each STSE model fitted, we found common optimal values of τ = 0.1 and ω = -0.8. It is important to note that the maximum predictive ability did not substantially differ between the parameter combinations tested; for example, they ranged from 0.33 to 0.42 for GY in 2006.

The heatmap of the **H** matrix across hybrids grown in different years and common checks showed that the patterns between years differed considerably, with hybrids grown in 2008 being less correlated with the others (Figure 1). We stress that 2008 was the year with less hybrids (34) for which genotypic information was available (Supplementary Table 1). Among the 11 checks, nine were exclusive to a single environment, and only four were genotyped (Supplementary Table 1). We can also observe a pattern of lower relatedness between the checks (Figure 1).

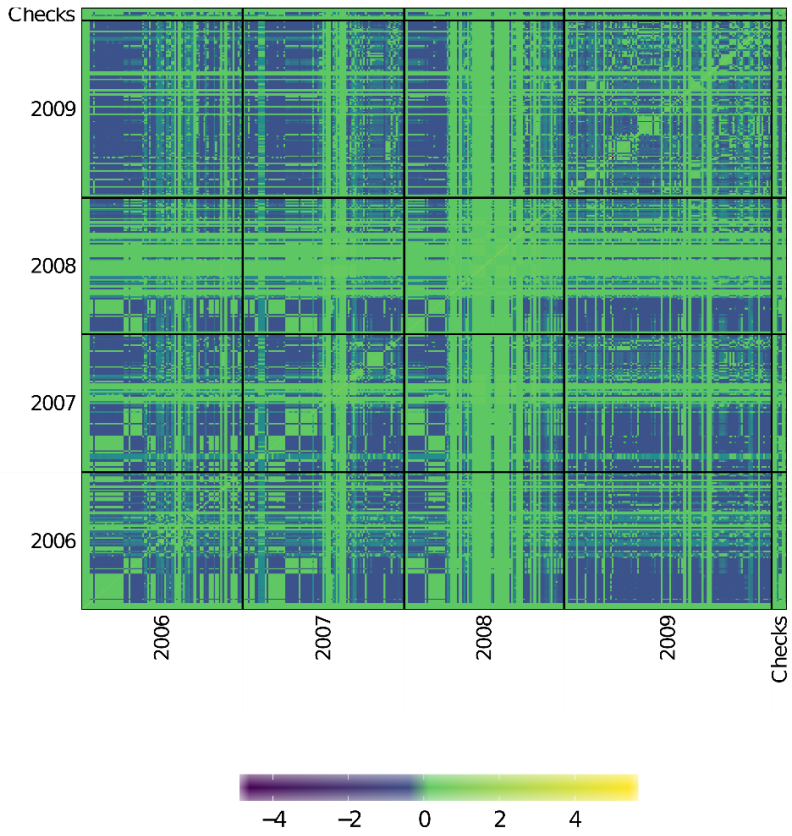


Figure 1 Heatmap of the **H** matrix across hybrids cultivated in different years and common checks. A total of 415 hybrids were evaluated

3.3.2 Genetic parameters

Broad-sense and narrow-sense heritability coefficients varied considerably between traits (Table 1). Overall, the heritabilities were higher for GM and lower for NE, ranging from 0.70 to 0.91 and 0.13 to 0.55, respectively. For GY the broad-sense heritability did not show large variation among years, ranging from 0.51 in 2008 to 0.59 in 2007. However, we observed larger variation between years for the narrow-sense heritability, from 0.18 in 2006 to 0.41 in 2008. The lowest narrow-sense heritability for NE was seen in 2006, with a coefficient of 0.03. The differences between broad- and narrow-sense heritabilities were less pronounced for GM, with no observed difference in 2007 (h^2 = 0.91). The phenotypic means also varied considerably between years, ranging from 3.80 to 8.04, 34.02 to 40.19 and 16.43 to 31.88, for GY, NE and GM, respectively.

Table 1 Phenotypic mean, broad-sense and narrow-sense heritability coefficients for each trait and year evaluated. Grain yield (GY) in t/ha, number of ears (NE) per plot and grain moisture (GM) in percentage

Trait	Year	Phenotypic mean	Broad-sense h^2	Narrow-sense h^2
GY	2006	5.62	0.51	0.18
	2007	3.80	0.59	0.26
	2008	5.58	0.56	0.41
	2009	8.04	0.52	0.28
NE	2006	40.19	0.36	0.03
	2007	34.02	0.13	0.07
	2008	34.6	0.32	0.11
	2009	39.45	0.55	0.21
GM	2006	31.88	0.77	0.67
	2007	25.63	0.91	0.91
	2008	29.05	0.70	0.50
	2009	16.43	0.85	0.80

Genetic correlations estimated based on the full MTME model showed that correlations varied noticeably across years, for each of the three traits (blocks closer to the diagonal in Figure 2). For GY, 2009 showed lower correlation coefficients, presenting negative correlations with years 2007 (-0.13) and 2008 (-0.38). Genotype effects in 2006 were

negatively correlated with all others for NE. In general, the GM was the trait with lowest correlation between years, particularly for 2006 and 2009. However, the years 2007 and 2008 showed the highest correlation (0.80) compared to the other traits.

We observed a large number of small and negative values of genetic correlation between traits. Overall, the correlations between traits varied considerably across years. GY and NE were more positively correlated, with a peak correlation of 0.86 between GY in 2008 and NE in 2007. On the other hand, GY and NE presented the highest negative correlation, -0.67 between GY in 2007 and NE in 2006.

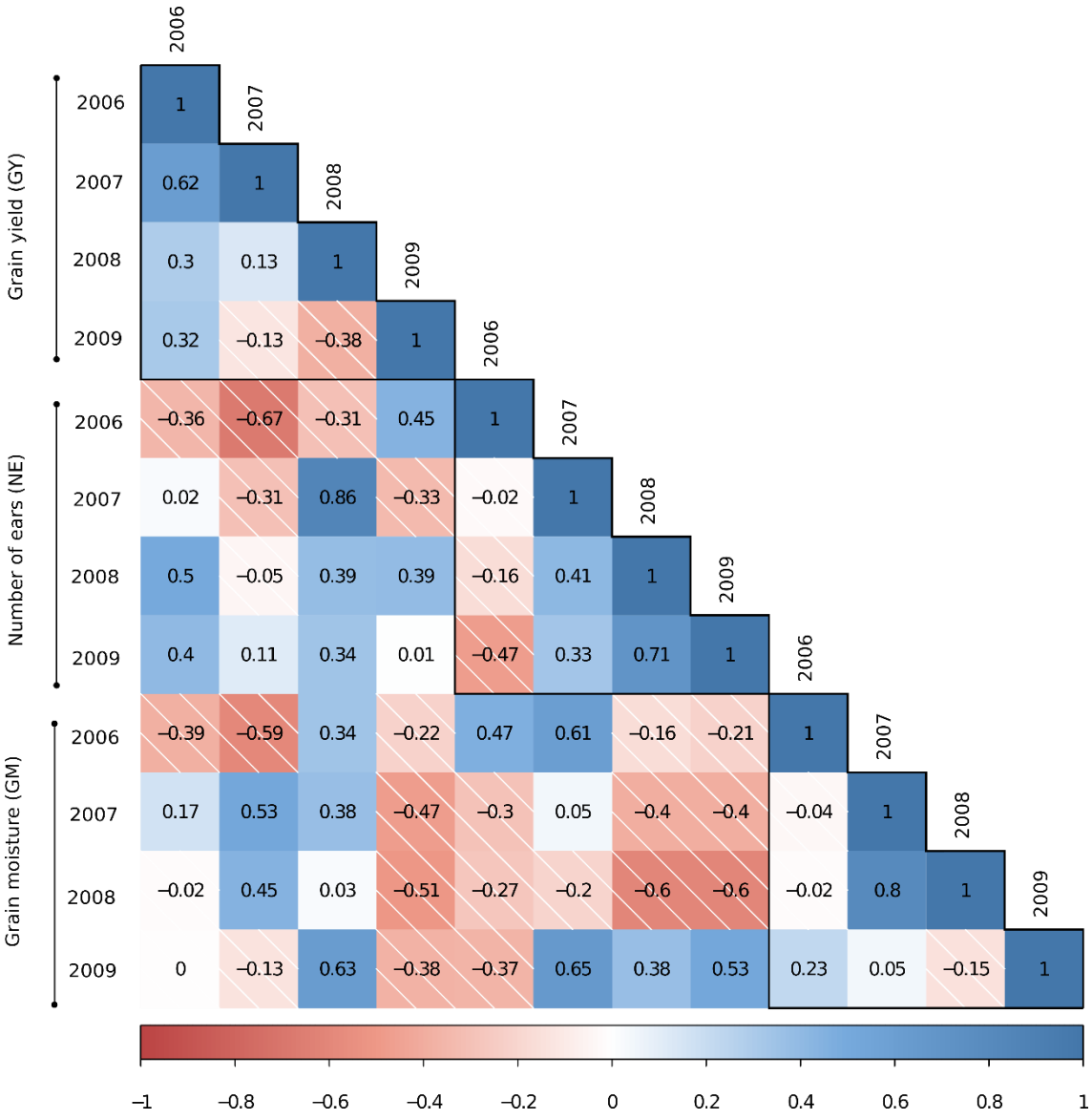


Figure 2 Genetic correlations between combinations of traits and years, estimated based on the multi-trait multi-environment (MTME) model

3.3.3 Genomic prediction

Predictive abilities varied considerably when comparing STSE and MTSE models (Figure 3). The MTSE models were superior to the STSE only in 2007 for GM (0.68 for STSE and 0.70 for MTSE) and in 2009 for NE (0.47 for STSE and 0.49 for MTSE). However, even in these cases, we note that the differences were minimal. Within each trait, the predictive abilities differed greatly across years. For example, values for GY ranged from 0.30 in 2009 to 0.54 in 2007 and 0.21 in 2009 to 0.53 in 2007 for STSE and MTSE models, respectively.

For the ME models using the CV_R cross-validation scheme (Figure 4), we observed that the predictive abilities were lower than those observed with the SE models (Figure 3). Only for GY did the MTME model outperform the STME model in terms of predictive ability, but we note that both values were low (0.07 for STME and 0.15 for MTME). For GM the difference between STME and MTME was low, but the variance was higher in the MTME scenario. On the other hand, NE showed higher variance and predictive ability with the STME model (0.36) compared to the MTME model (0.28).

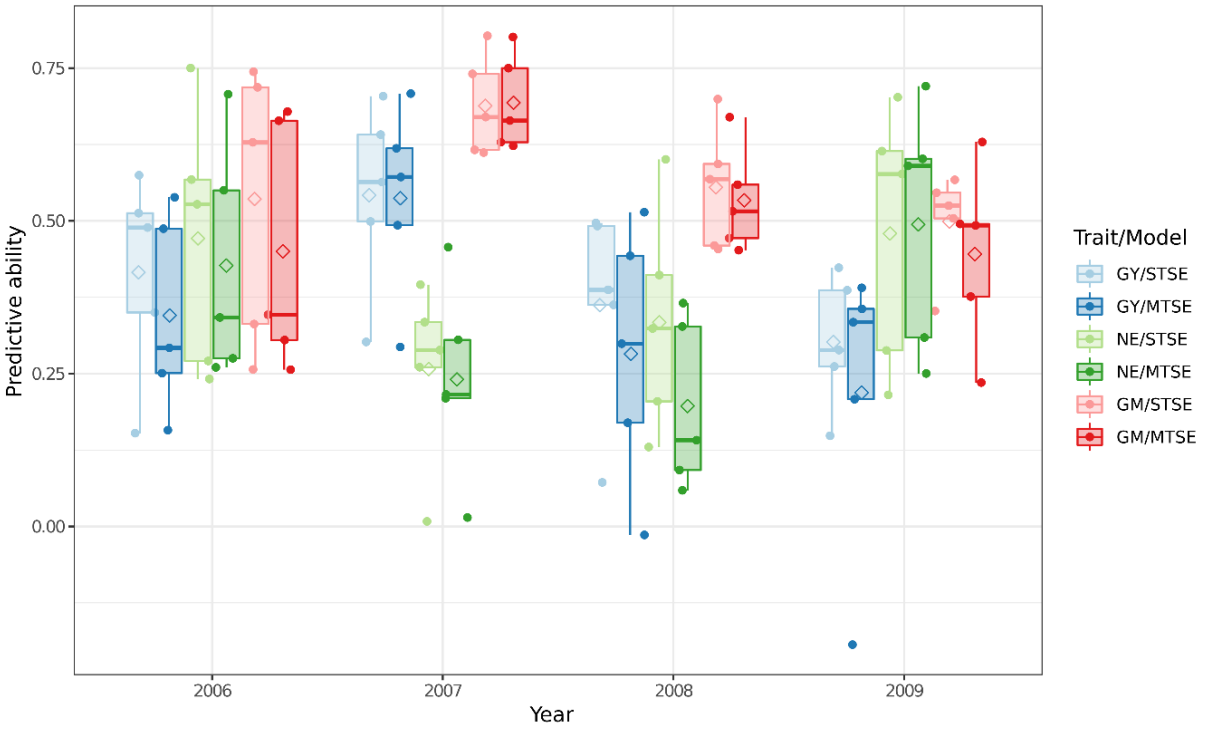


Figure 3 Predictive abilities of single-trait single-environment (STSE) and multi-trait single-environment (MTSE) models obtained with CV_R cross validation scheme, for the traits grain yield (GY), number of ears (NE) and grain moisture (GM). Diamonds correspond to the mean predictive abilities

When we used the same training/testing partition of single-environment models to fit multi-environment models, we observed that the most complete model MTME outperformed all the others, except for NE in 2006 (Figure 5).

The prediction of non-evaluated hybrids using the cross-validation scheme CV_1 showed similar predictive abilities for GY when compared to the cross-validation scheme CV_R. However, for NE and GM the multi-trait models also outperformed the single-trait ones when using the cross-validation scheme CV_1 (Figure 6). This is contrast to the results found using the cross-validation scheme CV_R (Figure 4).

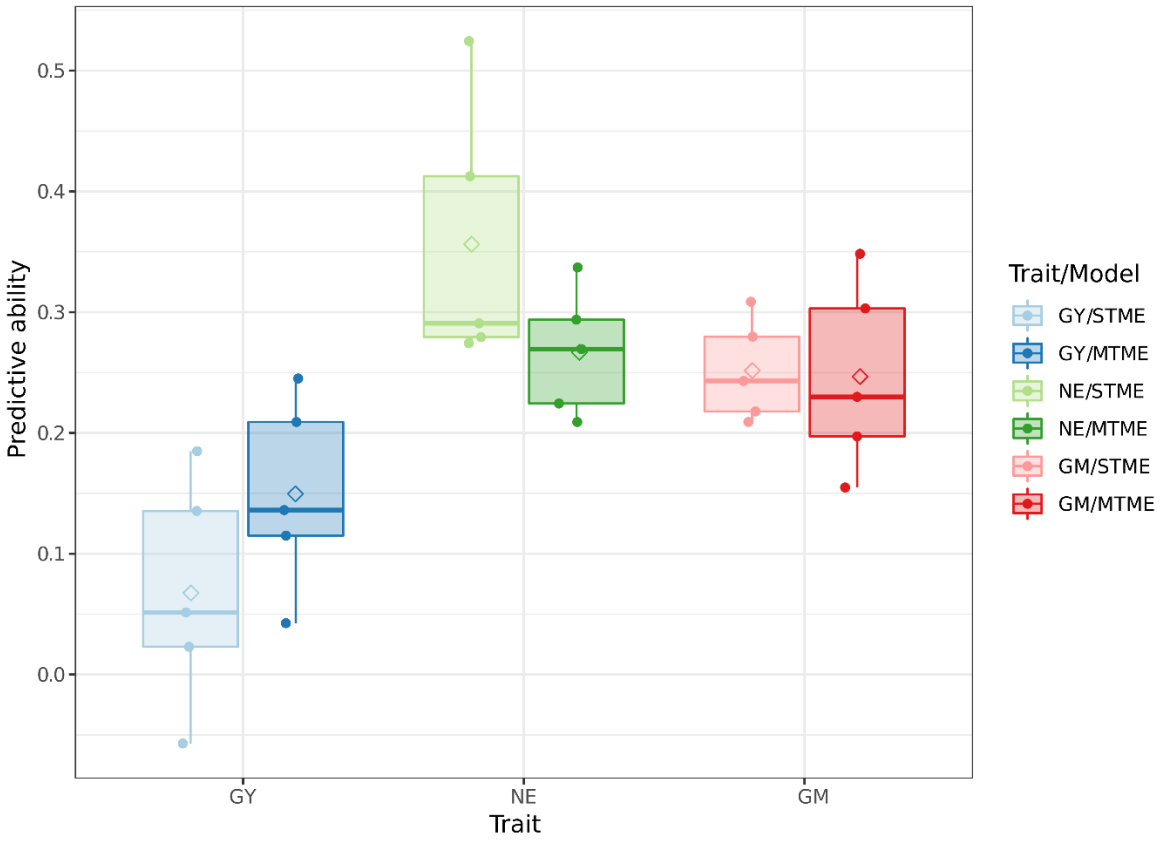


Figure 4 Predictive abilities of single-trait multi-environment (STME) and multi-trait multi-environment (MTME) models obtained with the CV_R cross validation scheme, for the traits grain yield (GY), number of ears (NE) and grain moisture (GM). Diamonds correspond to the mean predictive abilities

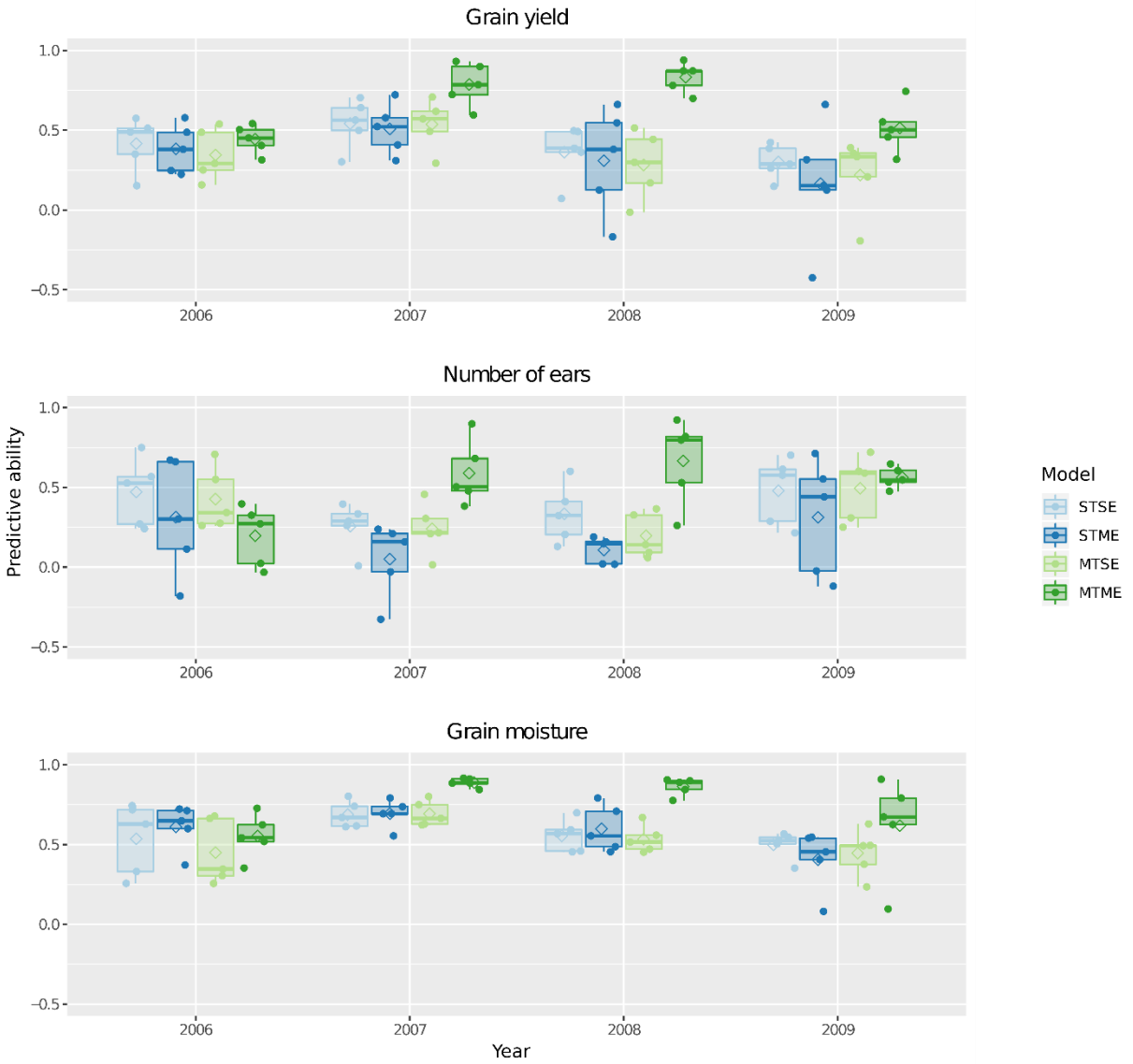


Figure 5 Predictive abilities of single-trait single-environment (STSE), single-trait multi-environment (STME), multi-trait single-environment (MTSE) and multi-trait multi-environment (MTME) models obtained with the CVR cross validation using the same training/testing partitioning of single-environment models, for the traits grain yield (GY), number of ears (NE) and grain moisture (GM). Diamonds correspond to the mean predictive abilities

The results from the cross-validation scheme CV_2 even revealed negative predictive abilities (Figure 7). This reflects difficulties in accurately predicting hybrid performance in different years. However, it is important to note that for some specific years and traits the predictive abilities achieved with CV_2 were comparable to those obtained with SE models. For example, in 2006 the predictive ability of the STME model for GY was 0.40, similar to the value of 0.42 found with STSE in the same year (Figures 3 and 7). The comparison

between MT and ST models when using the CV_2 scheme did not show any clear pattern, with substantial variation across traits and years.

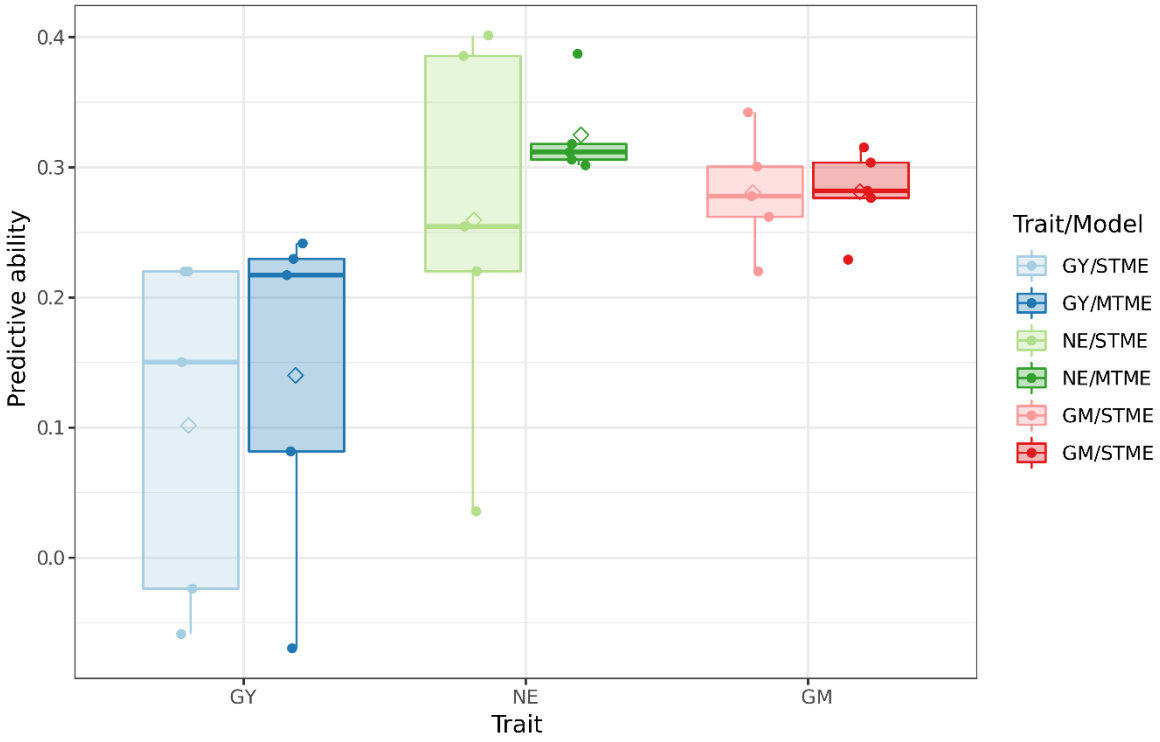


Figure 6 Predictive abilities of single-trait multi-environment (STME) and multi-trait multi-environment (MTME) models obtained with the CV_1 cross validation scheme, for the traits grain yield (GY), number of ears (NE) and grain moisture (GM). Diamonds correspond to the mean predictive abilities

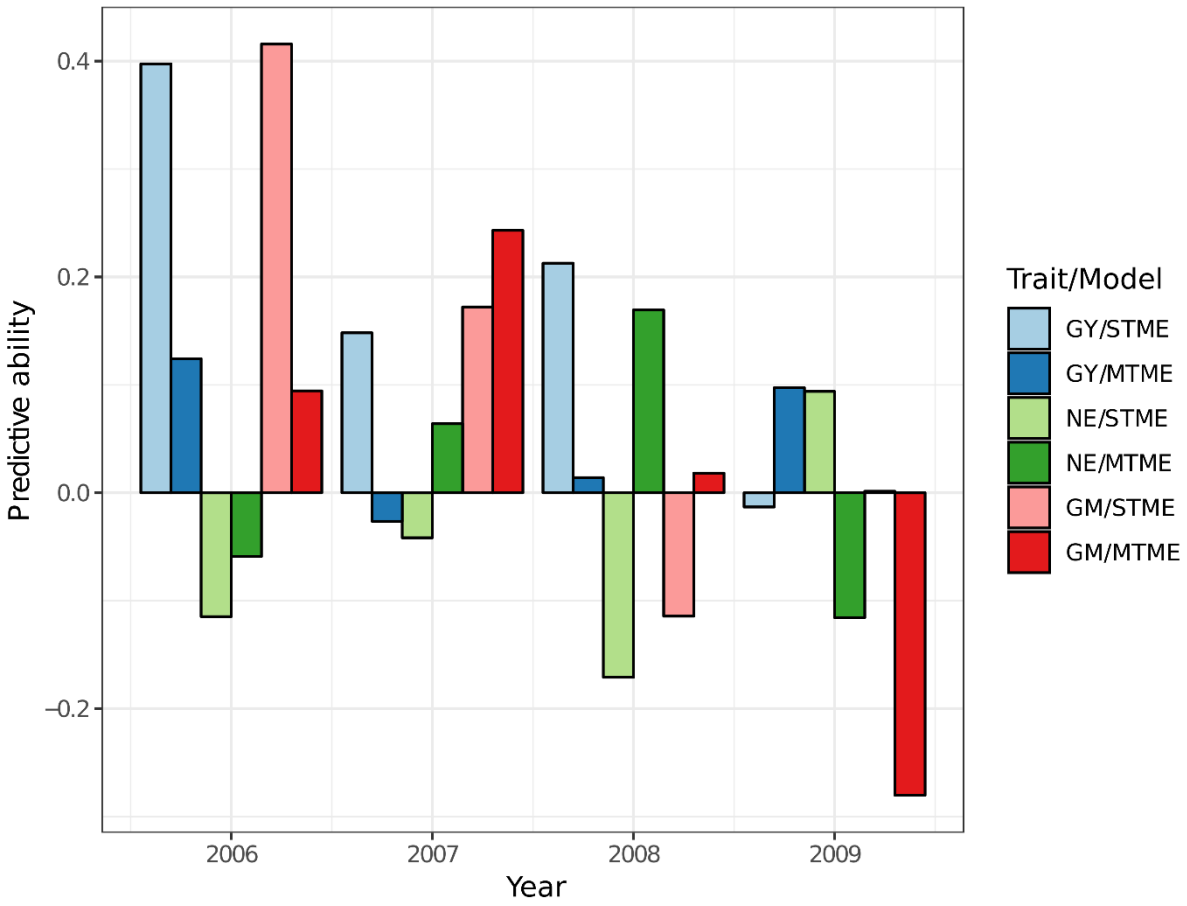


Figure 7 Predictive abilities of single-trait multi-environment (STME) and multi-trait multi-environment (MTME) models obtained with the CV_2 cross validation scheme, for the traits grain yield (GY), number of ears (NE) and grain moisture (GM)

3.4 Discussion

Genomic prediction models have been widely adopted in plant breeding of a variety of species, especially in maize (Bernardo 2009; Massman et al. 2013; Dias et al. 2018; Fristche-Neto et al. 2018; Han et al. 2018). However, the adoption of models that simultaneously take into account multiple traits and multiple environments has been much more limited (Montesinos-López et al. 2016; Gomes Torres et al. 2018; Ward et al. 2019). In this study, we applied genomic prediction to MTME trials of second season maize hybrids and compared their predictive abilities with univariate models.

It is well known that the genetic correlation between traits, as well as the fact that the trait of interest be of low heritability and the correlated trait be of high heritability, are key factors for the success of multi-trait models (Calus and Veerkamp 2011; Jia and Jannink 2012; Guo et al. 2014). However, when major quantitative trait loci (QTL) are not present, that is, for complex polygenic traits, the benefits of multi-trait models are limited even with

heritability differences among highly correlated traits (Jia and Jannink 2012). Moreover, studies based on real maize data sets using multi-trait genomic prediction models have reported little benefit of applying multivariate models (Dos Santos et al. 2016; Lyra et al. 2017; Lado et al. 2018). Our results indicated a similar pattern, where in general the multi-trait models did not show higher predictive abilities than the single-trait counterparts. One hypothesis for this similarity is the complexity of the data set considered in this study. Overall, we observed low to moderate values of correlation and heritabilities across the three traits evaluated. For example, for NE the narrow-sense heritability values were close to zero in 2006, 2007 and 2008. Similarly, low values were also observed for GY trait. It is also worthwhile to highlight that the correlations and heritabilities varied considerably across years, showing the challenges of dealing with the quality and unbalanced nature of our historical data.

Besides the correlation between traits, a model that also accommodates the GxE interaction mimics in a more realistic way the type of data generated in plant breeding programs, where genotypes are evaluated for multiple traits in multiple environments. Because in single-environment models the training and testing sets are exposed to the same environmental effects, it is biologically reasonable to expect higher predictive abilities than in scenarios that take multiple environments into account. When we compared STSE with MTSE models, we found that univariate models often outperformed the multivariate ones, in terms of predictive abilities. However, when we compared SE models with ME models we observed the opposite behavior. In this situation, the most complex MTME models outperformed all the others. This shows that multivariate models can improve the predictive ability of genomic prediction by appropriately taking into account the GxE interaction and the correlation between traits to.

We assessed the predictive ability for different combination of models using cross-validation schemes that mimic real scenarios in the maize breeding program. In the CV_R scheme the complete pool of individuals was randomly split in five folds. For CV_1 we randomly assigned the hybrids to a 5-folds scheme, ensuring that hybrids in the testing set had not been evaluated in any environment. Finally, in CV_2 we assigned years to folds. For GY, we found similar results when comparing the CV_R with CV_1 schemes. However, for NE and GM the CV_1 scheme showed higher predictive abilities using multi-trait models. It is noteworthy that, for this particular data set, the CV_R and CV_1 schemes are in fact very similar, because few hybrids are common between years. In any case, for two traits we did observe an

advantage of multivariate models when hybrids had not been evaluated in any environment, despite this limited connection between trials. The most challenging scenario was the prediction of hybrid performance in different years, as we observed with our cross-validation scheme CV_2.

It is important to note that the lack of all hybrids genotyped made necessary the use of the single-step procedure, through the **H** matrix, representing a complicating factor in our study. The impact of this matrix in the predictive abilities of genomic prediction models has been widely discussed in the animal breeding context (Pszczola et al. 2011; Christensen et al. 2012; Legarra et al. 2014; Martini et al. 2018; Teissier et al. 2019). However, in a plant breeding context the single step-procedure is less widespread and its impact in the prediction ability should be further investigated. Motivated by the possibility of combining information from the kinship and genomic relationship, here we are proposing to use the **H** matrix for genomic prediction of maize hybrids. To this end, we estimated the scaling factors τ and ω, – both important parameters to define how the **A** and **G** matrixes will be combined – as proposed by Misztal et al. (2010) and Tsuruta et al. (2011). In any case, we note that this blending is just one of several possibilities to approach the problem.

The use of a Bayesian inference in this study emerged as an alternative to the commonly used Restricted Maximum Likelihood (REML) estimation method. We had previously attempted to fit the same models used here with procedures based on REML, but could not achieve convergence. A similar problem was also reported in a study using MTME genomic prediction models in maize (Gomes Torres et al. 2018). Another study with MTME genomic prediction models applied to unbalanced wheat trials, despite reporting the use of REML, also documented convergence problems for several traits, highlighting the limitation of this technique when using a multivariate approach (Ward et al. 2019). As an alternative, we used algorithms based on Markov Chain Monte Carlo (MCMC) methods and implemented in the MCMCGLMM R-package. As a limitation, the computational requirements of the Bayesian method presented here may be challenging for practical applications. We evaluated different variance-covariance structures for the genetic and residual terms, and noticed that the computational time required to fit the more complex structure (unstructured) was not different from the simplest one (identity). Because using an unstructured matrix to model (co)variances reflects assumptions that are biologically more realistic, we chose to model the genetic and residual terms using this kind of structure.

The work presented here is an initial investigation of what could be done with MTME

models for prediction of hybrid maize, and we believe that the results are promising to justify further research. One important extension would be to incorporate dominance effects in the genomic prediction models, by leveraging the heterosis phenomenon. Several studies including non-additive effects have been conducted and reported the benefits of taking these effects into account (Dos Santos et al. 2016; Resende et al. 2017; Dias et al. 2018). On the other hand, other studies reported little advantage in terms of predictive ability when considering non-additive effects in the model (Muñoz et al. 2014; Nishio and Satoh 2014; Kumar et al. 2015; Minamikawa et al. 2017; Enciso-Rodriguez et al. 2018). The additive effects capture a large part of dominant and higher-order interaction effects, being difficult to properly separate the additive and dominance effects in genetic analyses (Varona et al. 2018). Besides that, the genetic architecture of a trait influences the proportion of each variance component (Huang and Mackay 2016). The non-additive genetic variance is expected to be low for most traits (Hill 2010; Crow 2010). For these reasons, the fact that the inclusion of non-additive effects in genomic prediction models provides little or no improvement in predictive abilities is somewhat reasonable. In this work, we also tried to investigate the consequences of including dominance effects into the models. However, due to the computational limitations we at first modeled dominance in the STSE models. We found little or no advantage of including this effect and thus decided that the steep computational demands for this kind of analysis were not justifiable in our more complex models.

Finally, we believe that our work also helps to better understand the practical challenges to successfully applying MTME genomic prediction models to a second season maize breeding program. We demonstrated that the use of MTME models can increase predictive ability when compared to univariate ones. However, in some cases we did not observe any improvement, which can at least partly be explained by the low correlation between traits and small heritability differences that we found. Besides that, the low levels of connection between trials in different environments and the necessity of using the single-step procedure highlight the complexity of the historical data we used. We believe that further research is needed to explore ways of dealing with these limitations, which represent the reality of a commercial maize breeding program. This study additionally suggests that there is room for further work in optimizing multivariate genomic prediction models in Bayesian and frequentist frameworks, allowing the practical application of these complex models.

References

Aguilar I, Misztal I, Johnson DL, et al (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score 1. J Dairy Sci 93:743–752. https://doi: 10.3168/jds.2009-2730

Amadeu RR, Cellon C, Olmstead JW, et al (2016) AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. Plant Genome 9:3. https://doi: 10.3835/plantgenome2016.01.0009

Bãnziger M, Edmeades G, Beck D, Bellon M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to pratice. CIMMITY

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419. https://doi: 10.2135/cropsci2008.08.0452

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126. https://doi: 10.1016/j.ajhg.2015.11.020

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719. https://doi: 10.2135/cropsci2011.06.0299

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual. 1–145. https://doi: citeulike-article-id:10128936

Calus MPL, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:26. https://doi: 10.1186/1297-9686-43-26

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. Genet Sel Evol 42:2. https://doi.org/10.1186/1297-9686-42-2

Christensen OF, Madsen P, Nielsen B, et al (2012) Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571. https://doi: 10.1017/S1751731112000742

Comstock RE (1978) Quantitative genetics in maize breeding. In: Walden D.B. (ed) Maize breeding and genetics. Wiley, New York, pp 191-206.

Conab (2019) Companhia Nacional de Abastecimento. Séries históricas. Available at: http://www.conab.gov.br. Accessed 07 Jan 2019.

Cooper M, Gho C, Leafgren R, et al (2014) Breeding drought-tolerant maize hybrids for the US corn-belt: discovery to product. J Exp Bot 65:6191–6204. https://doi: 10.1093/jxb/eru064

Covarrubias-Pazaran G, Schlautman B, Diaz-Garcia L, et al (2018) Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of *Vaccinium macrocarpon* Ait. Front Plant Sci 9:1310. https://doi: 10.3389/fpls.2018.01310

Crow JF (2010) On epistasis: why it is unimportant in polygenic directional selection. Philos Trans R Soc B 365:1241–1244. https://doi: 10.1098/rstb.2009.0275

Cuevas J, Crossa J, Soberanis V, et al (2016) Bayesian genomic prediction of genotype x environment interaction kernel regression models. G3 Gene Genome Genet 7:41–53. https://doi: 10.1534/g3.116.035584

Dias KODG, Gezan SA, Guimarães CT, et al (2018) Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity 121:24–37. https://doi: 10.1038/s41437-018-0053-6

Dos Santos JPR, De Castro Vasconcellos RC, Pires LPM, et al (2016) Inclusion of dominance effects in the multivariate GBLUP model. PLoS One 11:1–21. https://doi: 10.1371/journal.pone.0152045

Edmeades GO, Bolaños J, Chapman SC, et al (1999) Selection improves drought tolerance in tropical maize populations: I. Gains in biomass, grain yield, harvest index. Crop Sci 39:1306–1315.

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1–10. https://doi: 10.1371/journal.pone.0019379

Enciso-Rodriguez F, Douches D, Lopez-Cruz M, et al (2018) Genomic Selection for late blight and common scab resistance in tetraploid potato (*Solanum tuberosum*). G3 Gene Genome Genet 8:2471–2481. https://doi: 10.1534/g3.118.200273

Ferrão LFV, Ferrão RG, Ferrão MAG, et al (2017) A mixed model to multiple harvest-location trials applied to genomic prediction in *Coffea canephora*. Tree Genet Genomes. https://doi: 10.1007/s11295-017-1171-7

Fristche-Neto R, Akdemir D, Jannink J-L (2018) Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. Theor Appl Genet 131:1153–1162. https://doi: 10.1007/s00122-018-3068-8

Gapare W, Liu S, Conaty W, et al (2018) Historical datasets support genomic selection models for the prediction of cotton fiber quality phenotypes across multiple environments. G3 Gene Genome Genet 8:1721–1732. https://doi: 10.1534/g3.118.200140

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo J, Berger J, Dawid A, Smith AF. (eds) Bayesian Statistics 4. Oxford Uni. Oxford, pp 625–631

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One. https://doi: 10.1371/journal.pone.0090346

Gomes Torres L, Rodrigues MC, Lima NL, et al (2018) Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. PLoS One 13: 1–15. https://doi: 10.1371/journal.pone.0199492

Guo G, Zhao F, Wang Y, et al (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:1–7. https://doi: 10.1186/1471-2156-15-30

Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. J Stat Softw 33:1–22. https://doi: 10.18637/jss.v033.i02

Hallauer A., Miranda Filho J. (2010) Quantitative genetics in maize breeding., 2.ed. Iowa State University Press, Ames

Han S, Miedaner T, Utz FU, et al (2018) Genomic prediction and GWAS of Gibberella ear rot resistance traits in dent and flint lines of a public maize breeding program. Euphytica 214:1–20. https://doi: 10.1007/s10681-017-2090-2

Henderson CR, Quaas RL (1976) Multiple trait evaluation using relatives records. J Anim Sci 43:1188–1197. https://doi: 10.2527/jas1976.4361188x

Hill WG (2010) Understanding and using quantitative genetic variation. Philos Trans R Soc B 365:73–85. https://doi: 10.1098/rstb.2009.0203

Huang W, Mackay TFC (2016) The genetic architecture of quantitative traits cannot be inferred from variance component analysis. PLOS Genet 12:e1006421. https://doi: 10.1371/JOURNAL.PGEN.1006421

Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. https://doi: 10.1534/genetics.112.144246

Jugenheimer RW (1976) Corn improvement, seed production and uses. Wiley-Interscience, New York

Kumar S, Molloy C, Muñoz P, et al (2015) Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. G3 Gene Genome Genet 5:2711–2718. https://doi: 10.1534/g3.115.021105

Lado B, Vázquez D, Quincke M, et al (2018) Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. Theor Appl Genet 131:2719–2731. https://doi: 10.1007/s00122-018-3186-3

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi: 10.1038/nmeth.1923

Law M, Childs KL, Campbell MS, et al (2015) Automated update, revision, and quality control of the Maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. Plant Physiol 167:25–39. https://doi: 10.1104/pp.114.245027

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. J Dairy Sci 92:4656–4663. https://doi: 10.3168/jds.2009-2061

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single Step, a general approach for genomic selection. Livest Sci 166:54–65. https://doi: 10.1016/j.livsci.2014.04.029

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 Gene Genome Genet 5:569–82. https://doi: 10.1534/g3.114.016097

Lyra DH, Mendonça L de F, Galli G, et al (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol Breed 37:80. https://doi: 10.1007/s11032-017-0681-1

Malosetti M, Ribaut JM, Vargas M, et al (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica 161:241–257. https://doi: 10.1007/s10681-007-9594-0

Marchal A, Legarra A, Sébastien T, et al (2016) Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq.) progeny tests. Mol Breed. https://doi: 10.1007/s11032-015-0423-1

Martini JWR, Schrauf MF, Garcia-Baccino CA, et al (2018) The effect of the H −1 scaling factors τ and ω on the structure of H in the single-step procedure. Genet Sel Evol 50:16. https://doi: 10.1186/s12711-018-0386-x

Massman JM, Jung HJG, Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci 53:58–66. https://doi: 10.2135/cropsci2012.02.0112

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi: 11290733

Minamikawa MF, Nonaka K, Kaminuma E, et al (2017) Genome-wide association study and genomic prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. Sci Rep 7:4721. https://doi: 10.1038/s41598-017-05100-x

Misztal I, Aguilar I, Legarra A, Lawlor TJ (2010) Choice of parameters for single-step genomic evaluation for type I. In: Proceedings of the 61st annual meeting of the European association for animal production. Heraklion, pp 23–27

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655. https://doi: 10.3168/jds.2009-2064

Montesinos-López OA, Montesinos-López A, Crossa J, et al (2016) A Genomic Bayesian Multi-trait and Multi-environment Model. G3 Gene Genome Genet 6:2725–2744. https://doi: 10.1534/g3.116.032359

Mrode RA, Thompson R (2005) Linear models for the prediction of animal breeding values, 2ed.

Muñoz PR, Resende MFR, Gezan SA, et al (2014) Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics 198:1759–1768. https://doi: 10.1534/genetics.114.171322

Nishio M, Satoh M (2014) Including dominance effects in the genomic BLUP method for genomic evaluation. PLoS One 9:e85792. https://doi: 10.1371/journal.pone.0085792

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. https://doi: 10.1007/s10681-007-9449-8

Pszczola M, Mulder HA, Calus MPL (2011) Effect of enlarging the reference population with (un)genotyped animals on the accuracy of genomic selection in dairy cattle. J Dairy Sci 94:431–441. https://doi: 10.3168/jds.2009-2840

R Development Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Resende RT, Resende MDV, Silva FF, et al (2017) Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity 119:245–255. https://doi: 10.1038/hdy.2017.37

Roorkiwal M, Jarquin D, Singh MK, et al (2018) Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. Sci Rep 8:11701. https://doi: 10.1038/s41598-018-30027-2

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Population biology ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics (ribosomal DNA spacer-length variation/restriction fragment-length polymorphisms/Rrnl/Rrn2). Proc Natl Acad Sci 81:8014–8018.

Sousa MB, Cuevas J, Couto EG de O, et al (2017) Genomic-enabled prediction in maize using kernel models with genotype × environment interaction. G3 Gene Genome Genet 7:1995–2014. https://doi: 10.1534/g3.117.042341

Technow F, Schrag TA, Schipprack W, et al (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197:1343–1355. https://doi: 10.1534/genetics.114.165860

Teissier M, Larroque H, Robert-Granie C (2019) Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. J Dairy Sci 102:3142–3154. https://doi: 10.3168/jds.2018-15650

Tsuruta S, Misztal I, Aguilar I, Lawlor TJ (2011) Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J Dairy Sci 94:4198–4204. https://doi: 10.3168/JDS.2011-4256

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423. https://doi: 10.3168/jds.2007-0980

Varona L, Legarra A, Toro MA, Vitezica ZG (2018) Non-additive effects in genomic selection. Front Genet 9:78. https://doi: 10.3389/FGENE.2018.00078

Ward BP, Brown-Guedira G, Tyagi P, et al (2019) Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. Crop Sci 59:491. https://doi: 10.2135/cropsci2018.03.0189

Yang J, Benyamin B, Mcevoy BP, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569. https://doi: 10.1038/ng.608

Supporting information

Supplementary Table 1 – Connections between experiments across years. The diagonal numbers represent the number of hybrids evaluated in each year. The off-diagonal numbers represent the number of common checks between different years. Numbers between parentheses represent the number of hybrids for which genotypic data were collected

Year	**2006**	**2007**	**2008**	**2009**
2006	100 (65)	-	-	-
2007	3 (1)	100 (68)	-	-
2008	2 (0)	2 (1)	100 (34)	-
2009	0	1 (1)	3 (1)	125 (93)