**Technische Universität München**

TUM School of Life Sciences

# Genomic prediction in advanced cycle breeding programs

**Hans-Jürgen Auinger**

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation

| | | |
|---|---|---|
| **Vorsitz:** | | Prof. Donna Ankerst, Ph. D. |
| **Prüfende der Dissertation:** | 1. | Prof. Dr. Chris-Carolin Schön |
| | 2. | Prof. Dr. Henner Simianer |

Die Dissertation wurde am 14.04.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 01.06.2022 angenommen.

Ich kam, weiß nicht woher,
ich bin und weiß nicht wer.
I leb‘, weiß nicht wie lang,
ich sterb‘ und weiß nicht wann.
Ich fahr‘, weiß nicht wohin,
mich wundert, dass ich so fröhlich bin.

(unbekannt)

I came, don’t know wherefrom,
I live, don’t know how long.
I am and don’t know who,
I ride, don’t know where to.
I die and don’t know when.
I’m amazed to be a gladsome man.

(unknown)

# Content

# Summary

In plant breeding, genome-based prediction of breeding values is expected to boost selection gain because it allows the reduction of breeding cycle lengths and the application of high selection intensity due to cost-efficient genotyping and savings in phenotyping. Successful application of this novel tool depends heavily on the calibration set with which the model employed for genomic prediction is trained as this largely determines the prediction accuracy regarding new selection candidates. In this thesis, optimum calibration set design was investigated in two commercial breeding programs with maize (*Zea mays*) and rye (*Secale cereale*). For both programs, staggered, interconnected breeding cycles were available.

In the maize study, prediction accuracies were promising for both grain dry matter yield and grain dry matter content but depended strongly on the size of the calibration set and its relatedness to the prediction set. If the calibration or prediction set comprised genetic groups of different performance levels, a significant effect on prediction accuracy was observed with both cross-validation and independent validation. The prediction accuracies of 1,000 calibration sets of size N = 500, sampled at random from initial datasets consisting of thousands of genotypes, were similar to those of the initial sets on average but, due to the large variation among them, some performed better. Thus, searching for optimized calibration sets when many possible candidates from multiple breeding cycles are available seems promising. Therefore, various genomic parameters were investigated for their effect on the prediction accuracy of different calibration sets. A close association among the effective sample size ($N_{eff}$), average maximum kinship ($u_{max}$), and linkage phase similarity (LPS) was observed but was largely attributable to their confounding with sample size. None of these parameters proved useful for identifying highly predictive subsets for calibration. Another measure, the expected reliability calculated from genomic data of the prediction sets, seems to be the only robust predictor of the accuracy of different calibration sets.

In the study with rye, data for grain yield, plant height, and thousand-kernel weight were available. The prediction accuracies of calibration sets sampled from the same and different breeding cycles as the candidates to be predicted by genomic and pedigree-based prediction was investigated. The latter approach was notably inferior to genomic

prediction if the model calibration was performed with aggregated data from multiple cycles. The accuracy of genomic prediction was about 0.70 for all traits when model training was based on aggregated data from all cycles, and exceeded within-cycle accuracy in all cases.

Combining the findings from both studies, aggregate data from several preceding cycles is recommended for compiling the calibration set for new breeding cycles despite its decreasing relatedness over time. Further, it seems promising to concentrate future efforts not only on optimizing the composition of the calibration set but accounting for the population structure of the prediction set as well, especially if the latter displays a clear population structure. Tailor-made calibration sets for unphenotyped families might be useful to maximize the accuracy of genome-based prediction.

# Zusammenfassung

Genom-basierte Zuchtwertvorhersagen von Selektionskandidaten lässt einen Quantensprung beim Selektionsgewinn in der Hybridzüchtung erwarten. Mit genom-basierter Selektion ist es möglich, die Zeit für einen Selektionszyklus deutlich zu senken und die Selektionsintensität zu erhöhen, da ein Teil des Aufwands für die Phänotypisierung wegfällt. Eine erfolgversprechende Anwendung der neuen Methode hängt stark von den Kalibrationsdaten ab, mit denen das Modell für die Vorhersage erstellt wird, da diese die Vorhersagegenauigkeit stark bestimmen. In dieser Arbeit wurde die bestmögliche Zusammensetzung eines Kalibrationsdatensatzes anhand von zwei kommerziellen Zuchtprogrammen bei Mais und Roggen untersucht. Für beide Datensätze lagen Daten aus mehreren Zuchtzyklen vor.

In der Maisstudie waren die Vorhersagegenauigkeiten für die beiden Merkmale Kornertrag und Kornfeuchte vielversprechend, jedoch stark abhängig von der Größe des Kalibrationsdatensatzes und der Verwandtschaft zwischen Kalibrations- und Vorhersagedaten. Unterschiede in den phänotypischen Mittelwerten verschiedener genetischer Gruppen hatten eine starke Auswirkung auf die Vorhersagegenauigkeit, wie mit Kreuzvalidierung und unabhängigen Validierungsmethoden gezeigt werden konnte. Die Vorhersagegenauigkeiten von 1000 zufällig aus den vorhandenen Datensätzen gezogenen Kalibrationsdatensätzen mit N=500 waren im Mittel vergleichbar mit denen der Grundgesamtheiten. Einige der Stichproben erzielten aber deutlich höhere Vorhersagegenauigkeiten, da die Stichproben eine große Spannweite in der Vorhersagegenauigkeit aufwiesen. Deswegen scheint es erfolgversprechend, aus einer großen Grundgesamtheit Individuen für den Kalibrationsdatensatz auszuwählen, die für eine Vorhersage optimal sind. Dieser Umstand war Motivation, den Zusammenhang zwischen genomischen Parametern und der Vorhersagegenauigkeit zu untersuchen. Es zeigte sich ein enger Zusammenhang für die Parameter effektive Populationsgröße $N_{eff}$, die durchschnittliche Verwandtschaft $u_{max}$ und die Ähnlichkeit der Kopplungsphasen LPS (englisch: linkage phase similarity). Leider konnte keiner der Parameter für die Erstellung eines Teildatensatzes mit hoher Vorhersagegenauigkeit herangezogen werden, da auch eine enge Beziehung zur Stichprobengröße des Kalibrationsdatensates bestand. Einzig als robuster Prädiktor für Vorhersagegenauigkeit in Frage zu kommen scheint die aus den genomischen Verwandtschaftsdaten berechnete erwartete Schätzwertsicherheit. Diese wurde in

verschiedenen Kalibrierungsdatensätzen mit den erreichten Vorhersagegenauigkeiten in Beziehung gesetzt.

In der Roggenstudie lagen Daten zu Kornertrag, Tausendkorngewicht und Pflanzenhöhe vor. Es wurden Stichproben für die Kalibration aus den gleichen und anderen Zuchtzyklen gezogen, die auch für die Vorhersagedatensätze verwendet wurden. Für alle Stichproben wurden Zuchtwerte mit genomischen und Pedigree-basierten Zuchtwerten errechnet. Die Zuchtwerte auf Basis des Pedigrees waren deutlich schlechter in der Vorhersagegenauigkeit, vor allem, wenn die Kalibration aus verschiedenen Zuchtzyklen zusammengesetzt war. Die genomischen Vorhersagegenauigkeiten lagen im Bereich von 0.70 für alle Merkmale über aggregierte Datensätze aus Zuchtzyklen hinweg und waren höher als die Vorhersagen innerhalb eines Zuchtzyklus.

Aus den Ergebnissen der beiden Studien kann geschlossen werden, dass Daten aus mehreren vorhergehenden Zuchtzyklen zusammengefasst werden sollten, trotz der über die Zeit abnehmenden Verwandtschaftsnähe. Außerdem scheint es erfolgversprechend nicht nur auf die Zusammensetzung des Kalibrationsdatensatzes zu achten, sondern auch auf die Populationsstruktur innerhalb der Linien, die vorhergesagt werden sollen. Bei deutlichen Populationsstrukturen sollten die Kalibrationsdatensätze speziell für nicht phänotypisierte Familien bzw. Teilpopulationen angepasst werden.

# List of figures

# List of tables

# Publications included in this thesis

## Albrecht et al. (2014)

Albrecht T, Auinger H-J, Wimmer V, Ogutu J O, Knaak C., Ouzunova M., Piepho HP, Schön, C-C (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theoretical and Applied Genetics, 127, 1375–1386. https://doi.org/10.1007/s00122-014-2305-z

**Key message**

The calibration data for genomic prediction should represent the full genetic spectrum of a breeding program. Data heterogeneity is minimized by connecting data sources through highly related test units.

**Abstract**

One of the major challenges of genome-enabled prediction in plant breeding lies in the optimum design of the population employed in model training. With highly interconnected breeding cycles staggered in time the choice of data for model training is not straightforward. We used cross-validation and independent validation to assess the performance of genome-based prediction within and across genetic groups, testers, locations, and years. The study comprised data for 1,073 and 857 doubled haploid lines evaluated as testcrosses in 2 years. Testcrosses were phenotyped for grain dry matter yield and content and genotyped with 56,110 single nucleotide polymorphism markers. Prediction accuracies strongly depended on the relatedness of the doubled haploid lines from the estimation set with those on which prediction accuracy was assessed. For scenarios with strong population heterogeneity it was advantageous to perform predictions within a priori defined genetic groups until higher connectivity through related test units was achieved. Differences between group means had a strong effect on prediction accuracies obtained with both cross-validation and independent validation. Prediction accuracies across subsequent cycles of selection and years were only slightly reduced compared to prediction accuracies obtained with cross-validation within the same year. We conclude that the optimum data set for model training in genome-enabled prediction should represent the full genetic and environmental spectrum of the respective breeding program. Data heterogeneity can be reduced by

experimental designs that maximize the connectivity between data sources by common or highly related test units.

**Candidate's contribution**

The candidate contributed significantly to the data analysis. The candidate contributed significantly to the development and curation of the software package "synbreed", which was used in the data analysis. He discussed and interpreted the results with the coauthors and read, edited, and approved the final manuscript.

## Auinger et al. (2016)

Auinger H-J, Schönleben M, Lehermeier C, Schmidt M, Korzun V., Geiger HH., Piepho HP, Gordillo GA, Wilde P, Bauer E, Schön C-C (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye *(Secale cereale L.)*. Theoretical and Applied Genetics, 129, 2043–2053. https://doi.org/10.1007/s00122-016-2756-5

**Key message**

Genomic prediction accuracy can be significantly increased by model calibration across multiple breeding cycles as long as selection cycles are connected by common ancestors.

**Abstract**

In hybrid rye breeding, application of genome-based prediction is expected to increase selection gain because of long selection cycles in population improvement and development of hybrid components. Essentially two prediction scenarios arise: (1) prediction of the genetic value of lines from the same breeding cycle in which model training is performed and (2) prediction of lines from subsequent cycles. It is the latter from which a reduction in cycle length and consequently the strongest impact on selection gain is expected. We empirically investigated genome-based prediction of grain yield, plant height and thousand kernel weight within and across four selection cycles of a hybrid rye breeding program. Prediction performance was assessed using genomic and pedigree-based best linear unbiased prediction (GBLUP and PBLUP). A total of 1040 S2 lines were genotyped with 16 k SNPs and each year testcrosses of 260 S2 lines were phenotyped in seven or eight locations. The performance gap between GBLUP and PBLUP increased significantly for all traits when model calibration was performed on aggregated data from several cycles. Prediction accuracies obtained from cross-validation were in the order of 0.70 for all traits when data from all cycles (N CS = 832) were used for model training and exceeded within-cycle accuracies in all cases. As long as selection cycles are connected by a sufficient number of common ancestors and prediction accuracy has not reached a plateau when increasing sample size, aggregating data from several preceding cycles is recommended for predicting genetic values in subsequent cycles despite decreasing relatedness over time.

**Candidate's contribution**

The candidate analyzed the data and interpreted the results. The candidate and coauthor Manfred Schönleben contributed equally to writing the manuscript. The candidate and coauthors read, edited, and approved the final version of the manuscript.

## Auinger et al. (2021)

Auinger H-J, Lehermeier C, Gianola D, Mayer M, Melchinger AE, da Silva S, Knaak C, Ouzunova M., Schön C-C (2021) Calibration and validation of predicted genomic breeding values in an advanced cycle maize population. Theoretical and Applied Genetics, 134(9), 3069-3081. https://doi.org/10.1007/s00122-021-03880-5

**Key message**

Model training on data from all selection cycles yielded the highest prediction accuracy by attenuating specific effects of individual cycles. Expected reliability was a robust predictor of accuracies obtained with different calibration sets.

**Abstract**

The transition from phenotypic to genome-based selection requires a profound understanding of factors that determine genomic prediction accuracy. We analysed experimental data from a commercial maize breeding programme to investigate if genomic measures can assist in identifying optimal calibration sets for model training. The data set consisted of six contiguous selection cycles comprising testcrosses of 5968 doubled haploid lines genotyped with a minimum of 12,000 SNP markers. We evaluated genomic prediction accuracies in two independent prediction sets in combination with calibration sets differing in sample size and genomic measures (effective sample size, average maximum kinship, expected reliability, number of common polymorphic SNPs and linkage phase similarity). Our results indicate that across selection cycles prediction accuracies were as high as 0.57 for grain dry matter yield and 0.76 for grain dry matter content. Including data from all selection cycles in model training yielded the best results because interactions between calibration and prediction sets as well as the effects of different testers and specific years were attenuated. Among genomic measures, the expected reliability of genomic breeding values was the best predictor of empirical accuracies obtained with different calibration sets. For grain yield, a large difference between expected and empirical reliability was observed in one prediction set. We propose to use this difference as guidance for determining the weight phenotypic data of a given selection cycle should receive in model retraining and for selection when both genomic breeding values and phenotypes are available.

**Candidate's contribution**

The candidate managed the large dataset, conceived and devised the analysis methods, analyzed and interpreted the data, and wrote the first draft of the manuscript. The candidate and coauthors discussed and interpreted the results and read, edited, and approved the final manuscript.

# 1 Introduction

Food. Everyone needs food almost every day. At the beginning of human history, people lived as hunters and gatherers. Plants and animals occurring in nature were used; sometimes, long distances had to be covered to find them. With the beginning of sedentary lifestyles, plants and animals that exhibited positive properties were propagated in a targeted manner. Animals were fed to compensate for seasonal fluctuations and plants were fertilized or irrigated if necessary. From this point on, plants and animals could be selected and bred according to human requirements. The selection of desired properties was long based purely on visible phenotypes, but are these are subject to environmental influences and, thus, only partially heritable, genetic improvement was a rather slow process. Nevertheless, impressive achievements have been made through selection and breeding over the centuries. Rye, for example, was originally a weed that occurred in fields of common wheat and migrated from the Fertile Crescent to Europe, where it was selected as a cereal for human consumption (Miedaner 2013). Likewise, maize was developed by human selection from its wild ancestor teosinte in Mexico, taking advantage of mutations in a relatively small number of genes. While teosinte has many inflorescences and seeds with hardened fruit cases, maize has few ears and soft kernels and depends on humans for propagation. Besides such changes in characters with relatively simple inheritance, quantitative traits such as yield have also undergone impressive improvements in the millennia after agriculture began.

At the beginning of the 20$^{th}$ century, plant breeding was professionalized and systematized after the rediscovery of Mendel's laws. In many countries, including Germany, breeding companies were founded—often by farmers—and conducted field trials to find the best varieties from genotypes within existing landraces or by generating new genetic variation through intra-specific crosses and subsequent inbreeding. This improved cultivar performance significantly.

The Food and Agriculture Organization of the United Nations (FAO) provides statistics on agricultural production starting in 1961. Over the last six decades, the yield of important field crops has doubled or even tripled (Figure 1). Voss-Fels et al. (2019) showed that these impressive increases in productivity as well as other important traits in wheat are mainly due to the genetic improvement of our crops. The acreage devoted

**Figure 1: Yield development for important field crops worldwide.** Except for sugar beets only the usage of grains is considered. (FAOSTAT 2021).

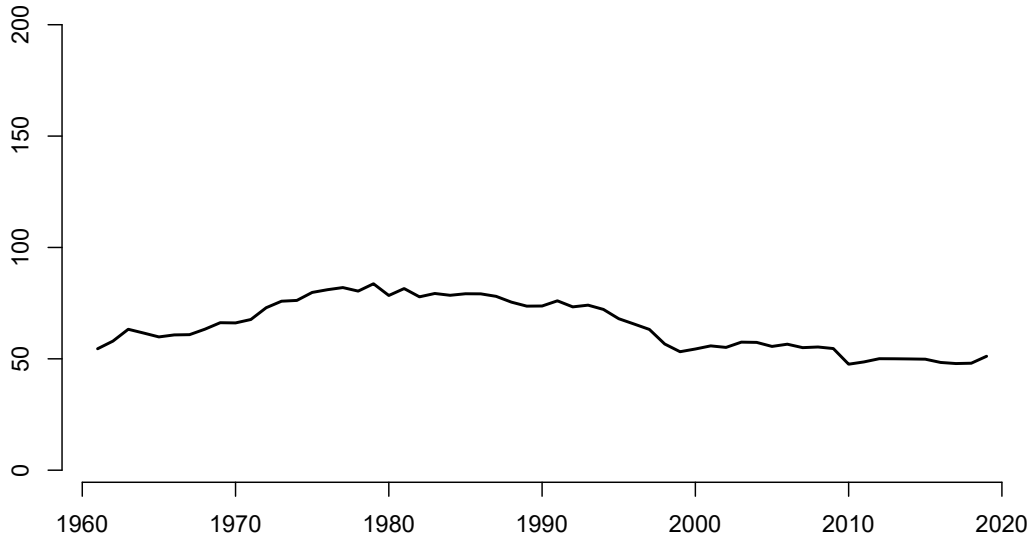


**Figure 2: Acreage of important crops worldwide.** (FAOSTAT 2021).

to maize was expanded tremendously after the introduction of hybrid breeding (Figure 2). Commercialization of hybrid seed allows intensive breeding because farmers buy certified seed every year. Thus, there is a higher return on investment to breeders compared to line breeding. Meanwhile, the rye acreage has slightly decreased despite the introduction of hybrid varieties in the 1980s. In the future, this trend may be reversed with new end-uses, such as fermentation for biogas production.

The focus of this thesis lies on the optimization of advanced cycle breeding of two hybrid crops, rye and maize, with the aid of genomic selection. Rye (*Secale cereale* L.) belongs to the family *Poaceae* and genus *Secale*. Rye has a diploid genome of seven chromosomes and a genome size of about 7.9 gigabases with more than 90% repetitive sequences (Bauer et al. 2017). Within the *Poaceae,* rye belongs to the tribus *Triticeae* together with barley and wheat (Jäger et al. 2003). Barley, rye, and wheat presumably evolved from a common *Triticeae* progenitor (Middleton et al. 2014, Murat et al. 2014). The genomes of all three cereals are largely syntenic, except for numerous rearrangements (Devos et al. 1993, Martis et al. 2013). Rye evolved, like barley and wheat, in the Fertile Crescent in the Near East and started its spread after domestication in Anatolia (Hilman 1978). From the Middle Ages on, rye was intensively cultivated and became the main bread cereal in Middle Europe (Behre 1992). Today, Germany, Poland, and Russia are the main producers of winter rye. Rye is mainly grown in dry areas on poor and sandy soils due to its excellent tolerance of unfavorable climatic or soil conditions (FAOSTAT 2021). The acreage of rye for grain production has been decreasing continuously since 1960 (FAOSTAT 2021, Figure 2), but it might gain importance in the context of improving the sustainability of plant production. Rye is an important cover crop to keep fields under cultivation in winter and can contribute to more diverse crop rotation systems. Rye is often planted before maize or sorghum and used as silage for animal feed (Miedaner 2013).

The second species used in this study is maize. Maize (*Zea mays* L.) belongs to the family *Poaceae* and genus *Zea*. The plants are diploid with ten chromosomes and a genome size of about 2.3 gigabases (Schnable et al. 2009). Maize was domesticated about 10,000 to 7,000 years ago from the wild grass teosinte (*Zea mays* ssp. *parviglumis*) in Mesoamerica (Beadle 1939; Goodman and Walton 1988). From there, maize spread quickly over both American continents and became popular even in the cool growing areas of modern Canada. The introduction of maize to Europe is attributed to Christopher Columbus, who brought maize to Europe after his first expedition to the New World in 1493. In the following years, further explorers brought maize germplasm

from other American regions to Europe, mainly from the Caribbean islands but also from Canada (Tenaillon and Charcosset 2011, Mir et al. 2013). During its spread and cultivation in various areas of Southern, Central, and Western Europe, landraces of maize became well adapted to colder climatic conditions (Dubreuil et al. 2006). Today, maize cultivation spans a vast range of latitudes and longitudes and it is one of the most important crops (Crow 1998, Barcaccia et al. 2003) in Germany and worldwide.

The increase in maize yield during the last six decades in Germany is largely the result of hybrid breeding and the introduction of early-maturing cultivars. Maize combines high productivity with relatively low requirements for plant protection. The wide variety of possible end uses also contributes its success as a crop. The whole plant can be used as silage for feeding livestock or for biomass and energy production. Maize grains can be used directly for human consumption, but also as feed, for the production of ethanol, or as raw material for bioplastics.

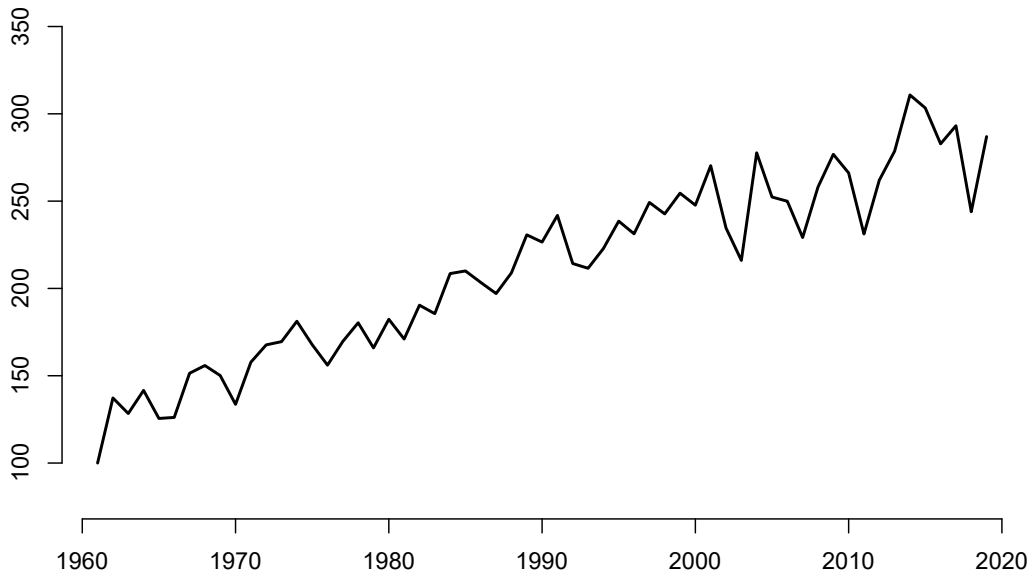


**Figure 3: Grain yield development of the most important crops in Germany relative to 1961 baselines.** (FOASTAT 2021)

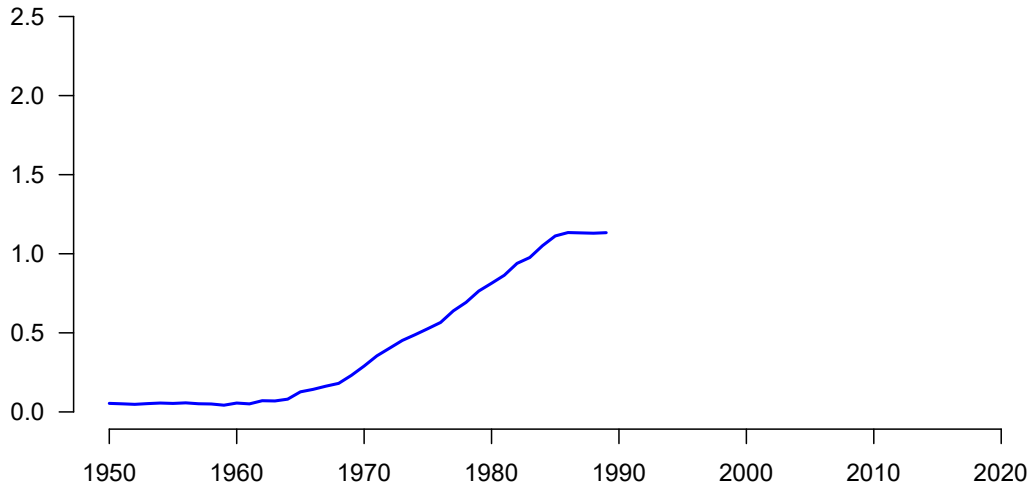


**Figure 4: Acreage of maize in Germany separated by end use.** Before 1990, the values represent only West Germany, omitting the German Democratic Republic. (DESTATIS 2021)

In Germany, the acreage of grain maize has increased only slightly (DESTATIS 2021, Figure 4). It is mainly used for silage production in Germany. To produce silage, the whole plant is chopped, compressed, and preserved through natural lactic acid fermentation. From 1960 on, maize silage replaced other crops, such as beets, as cattle feed. During the 1990s, the maize acreage decreased with the decreasing number of cattle. In the first decade of the 21$^{st}$ century, the acreage of maize for silage production increased impressively again as government programs promoted the generation of renewable energy. To use maize for renewable energy production, bacteria in huge fermenters convert silage, over about 35 days, into methane, which is then converted into electricity by generators.

## 1.1 From phenotypic to genomic selection

### 1.1.1 Quantitative genetics and breeding

With his famous laws derived from experiments with peas, Gregor Mendel (1822–1884) laid the foundation of modern genetics. Mendel was fortunate to work with simply inherited qualitative traits controlled by one or two loci. Even after the rediscovery of Mendel's laws in 1900, the hereditary foundation of quantitative traits was still heavily disputed for almost two decades. Ronald A. Fisher offered a breakthrough in 1918 with his seminal paper on "The correlation between relatives on the supposition of Mendelian inheritance", which is considered the foundation of quantitative genetics. He explained that quantitative traits, such as height in humans, were driven by a genetic component, environmental influence, and possible interactions between these two factors. Most importantly, he explained the hereditary portion as the action of many loci, each having only a small effect on the expression of the trait; nowadays, this is known as the infinitesimal model of quantitative genetics. In his mathematical derivations, Fisher allowed for additive and dominant gene action as well as epistatic interactions between loci, and decomposed the total genetic variance according to these components. Thus, he was able to describe the resemblance between relatives based on these components, which is a fundamental principle of breeding.

A further milestone in quantitative genetics was Malecot's (1948) proposition to quantify relatedness between relatives. He introduced the coefficient of coancestry as the probability that two alleles taken at random from two individuals at a given genetic locus are "identical by descent". The coefficient of coancestry can be calculated from pedigree data. For many genotypes, the values are commonly stored in the pedigree-based numerator relationship matrix, which provides the basis for estimating the components of the phenotypic and genotypic variances and covariances observed in experimental data (Henderson 1974).

In his 1918 paper, Fisher gave an example for the percentage of phenotypic variance that is explained by "essential genotypes," which corresponds to the formal definition of heritability or the ratio of the genetic variance divided by the phenotypic variance corrected for known effects such as nutrition, sex, or controlled environmental factors. Wright (1920) introduced the correlation between genotype and phenotype as a measure of heredity. The square of this correlation is the heritability, denoted as $h^2$, which corresponds to the regression coefficient of the genotypes on the phenotypes. In

breeding, narrow-sense ($h^2$) and broad-sense ($H^2$) heritability are differentiated. To calculate narrow-sense heritability, the variance explained by additive effects is used as the numerator and divided by the phenotypic variance. Broad-sense heritability is not strictly defined, but the genetic variance (including additive, dominance, and epistatic effects) is commonly used as the numerator. In plant breeding, the "operative" heritability suggested by Strube (1967), is commonly used, where the experimental error and genotype x environment interaction variances in the denominator are divided by the number of replications (R) and R x the number of test environments (E). Crucially, this heritability is specific to a given trait, the studied population, and the test environments used for phenotyping. This can be illustrated in the heritabilities estimated in this study (Table 1). In the maize experiments, comprising six different datasets (S1 to S6), two sets (S4 and S6) showed rather low heritabilities for grain yield, whereas set S3 had low heritability for grain moisture. The reasons for these observations are different. In S4, the genetic variance for yield is low; in S6, the experimental error variance is large; and moisture in S3 was subject to large interactions between genotypes and environments. In the rye experiment of this study, cycle C3 showed lower heritabilities for all traits due to rather specific environmental conditions during the study year. Nevertheless, heritabilities for a given trait are often similar among populations and datasets (Visscher et al. 2008) as confirmed, for example, for plant height in the rye experiments. In addition, in plant breeding, broad-sense heritabilities are generally quite high for most traits if experiments are replicated and conducted in several environments.

The genetic improvement of crops depends on efficient selection and breeding progress is commonly measured by the selection response, denoted as $\Delta G$. Selection response is defined as the difference between the mean of the offspring of the selected fraction of genotypes and the mean of the population from which the parental genotypes were selected. The well-known breeder's equation relates $\Delta G$ to its relevant components with the following formula: $\Delta G = h^2 S = h^2 \sigma_p i = h \sigma_g i$ (Falconer and Mackay, 1996). S denotes the selection differential, which is the difference between the mean of the selected genotypes and the population mean, and $h^2$ represents the heritability. The selection differential can be expressed as product of the selection intensity (i) and the phenotypic standard deviation ($\sigma_p$). The selection intensity iscalculated as the expectation of a random variable from a truncated normal distribution with probability $\alpha$ in the tail, and therefore standardized without units. Naturally, $\Delta G$ can

**Table 1: Broad sense heritabilities in the maize and rye data used in this thesis.** The traits are grain dry matter yield (GDY), grain dry matter content (GDC), plant height (PHT) and thousand kernel weight (TKW).

| Cultivar | Dataset | Year | Heritability | | | |
|---|---|---|---|---|---|---|
| | | | GDY | GDC | PHT | TKW |
| Maize | S1 | 2010 | 0.86 | 0.96 | | |
| | S2 | 2011 | 0.78 | 0.97 | | |
| | S3 | 2012 | 0.75 | 0.75 | | |
| | S4 | 2013 | 0.56 | 0.92 | | |
| | S5 | 2014 | 0.74 | 0.88 | | |
| | S6 | 2015 | 0.52 | 0.88 | | |
| Rye | C1 | 2009 | 0.86 | | 0.91 | 0.90 |
| | C2 | 2010 | 0.86 | | 0.94 | 0.80 |
| | C3 | 2011 | 0.77 | | 0.89 | 0.76 |
| | C4 | 2012 | 0.83 | | 0.94 | 0.87 |

also be expressed as the product of the square root of the heritability, the genetic standard deviation ($\sigma_g$), and the selection intensity. In genome-based selection, the square root of the heritability is replaced with the prediction accuracy, which denotes the correlation between the estimated and the true breeding value. When estimated from data, the prediction accuracy is the correlation between the observed phenotype and the predicted breeding value divided by the square root of the heritability (Dekkers 2007).

In many breeding programs, the aim is not only to select the best candidates but to achieve improved performance in their progenies. This led to the introduction of the concept of breeding values in animal breeding. The breeding value of a candidate is defined as twice the deviation of the mean of its offspring from the population mean (Bernardo 2020a). Thus, the breeding value of milk yield for a bull can be estimated by the average milk yield of his daughters. Breeding values depend exclusively on the additive effects of alleles because only individual alleles, and not the allele combinations

of the candidate, are transmitted to the next generation. Thus, dominance effects that depend on the allele combination at the same locus in an individual, as well as epistatic effects that depend on allele combinations at different loci, do not contribute to an individual's breeding value.

A major question in the calculation of breeding values is how to weight the information from different relatives and traits. Henderson (1963, 1975, 1984) solved these problems by applying the mixed linear model to describe the various sources of information and their connection to a candidate's breeding value. He developed the statistical framework for calculating best linear unbiased predictions (BLUPs) to obtain candidates' breeding values from mixed-model equations and developed numerical solutions (Schön and Simianer 2015). The BLUP approach enormously improved selection decisions in animal breeding, especially for traits with low heritability. In plant breeding, BLUPs were adopted much later. One reason was that, in plant breeding, genetically identical plants such as hybrids, clones, or inbred lines can be produced and used in experiments with appropriate field designs for controlling experimental error, while phenotypic records in animal breeding usually come from production farms, where structured designs are typically impossible. Hence, in animal breeding, the collected data are highly unbalanced, numerous confounding effects must be modeled, and heritabilities are much lower than those in plant breeding. One of the first adoptions of BLUP in plant breeding was the prediction of maize single-cross hybrid performance based on molecular marker data and information from related hybrids (Bernardo 1994).

### 1.1.2 Implementation of markers in breeding

The advent of DNA markers more than three decades ago opened new avenues for genetic research and plant breeding. The first milestone was the introduction of restriction fragment length polymorphisms (RFLPs, Botstein et al. 1980), followed by microsatellites (SSRs, Akkaya et al. 1992), and amplified fragment length polymorphisms (AFLPs, Vos et al. 1995). A significant breakthrough was the use of single nucleotide polymorphisms (SNPs, Jordan and Humphries 1994) supported by advances in DNA sequencing techniques. With this advancement, we can now saturate the entire genomes of many plant species with high-density and high-throughput DNA markers. Thus, the genetic architecture of complex traits can be dissected into the contributions of small genomic segments or even individual genes, commonly referred

to as quantitative trait loci (QTL). To detect QTL and map markers, different types of mapping populations can be used, such as biparental families, nested association mapping (NAM) populations, multiparent advanced generation intercross populations (MAGIC), or association panels (Gage et al. 2020). To improve the power of QTL detection and mapping resolution, either recombinant inbred lines (RIL) or doubled-haploid lines from these populations are phenotyped and genotyped.

Figure 5 gives an overview and schematic representation of the different types of mapping populations used in the literature. A major difference among them is the choice of parents and the subsequent mating design. In plant breeding, most approaches advance the genetic material into fully homozygous pure-breeding lines, either by selfing or application of doubled-haploid technology (Molenaar and Melchinger 2019). Due the increased number of meioses, selfing steps lead to smaller haplotype blocks and thus to higher map resolution. In biparental families, two parents are crossed to produce the F1 generation, which serves as starting point for line development, as commonly practiced in many breeding programs. To establish a NAM population, several lines are crossed to a central inbred line. This set of half-sib families serves as a starting point for line development. To produce a MAGIC population, a distinct number of genetically diverse lines are specifically crossed with each other three or more generations before inbreeding is initiated. In an association panel, many very diverse lines, often collected from gene banks, landraces, or breeding programs, are used directly for genetic analyses. All approaches share that genome fragments of different sizes and contents are analyzed for their association with quantitative trait expression. If the appropriate statistical tests are statistically significant for one or several markers in a genomic region, that marker is likely linked to a QTL. Such markers can then be used in marker-assisted selection (MAS), where selection decisions are made based on the marker genotype of the candidate. Successful examples in practical plant breeding include resistance to rust and eyespot, recessive resistance to yellow mosaic viruses and powdery mildew in barley, and resistance to *Fusarium* head blight in wheat (Miedaner and Korzun 2012). Other examples include resistances to bacterial and fungal diseases in rice (Sanchez and Khush 2000) and seedling emergence in maize (Yousef and Juvik 2001). However, for complex traits, such as grain yield in maize, MAS success was rather inconsistent, as the contradictory results reported by Bouchez et al. (2002) and Openshaw and Frascaroli (1997) illustrate. In practical animal breeding, MAS of major genes plays an important role for selection against deleterious mutations that compromise important production traits (Jung et al. 2014; Kipp et al. 2016; Pausch et

al. 2016). Many statistical methods have been developed to integrate MAS directly into the breeding process (Würschum 2012). The systematic mathematical integration of markers in selection indices was proposed by Lande and Thompson (1990).

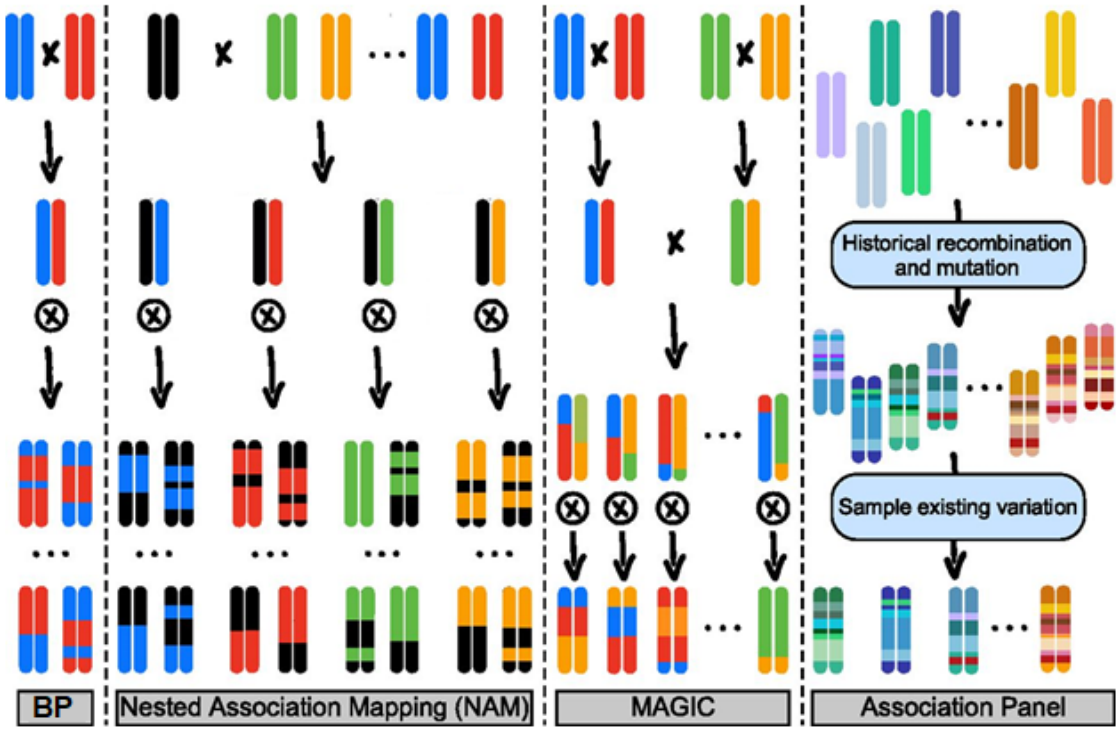


**Figure 5: Comparison of common mapping populations in plants.** BP populations are derived from a single biparental cross, resulting in progeny that are a mosaic of long haplotypes from the two parents. Limited recombination restricts the mapping resolution. NAM populations consist of numerous RIL families that share a common parent (shown in black). The historical recombination accumulated in the parents permits high map resolution and many alleles. MAGIC populations are often derived from 8 or 16 parental lines over several generations of recombination. Similar to NAM populations, MAGIC populations offer increased map resolution and allelic richness compared with BP populations. Association panels are a sample of natural variation from a larger, existing population that has accumulated historical recombination events and mutations. They frequently have more recombination and allelic richness than the other three populations, but are often burdened with an inherent population structure that can be difficult to account for. Black “x”s indicate crosses, and circled “x”s indicate selfing until homozygosity is achieved. Ellipses indicate many other individuals in the population or family (taken from Gage et al. 2020 Figure 1, slightly modified).

A landmark in the selection of complex traits based on DNA information was Meuwissen et al.’s (2001) seminal paper suggesting the use of genome-wide dense marker maps to predict the total genetic value of selection candidates. To establish the association between marker information and phenotypic trait expression, these authors used the

BLUP approach, assuming markers as random effects sampled from a distribution with uniform variance, and two Bayesian methods for estimating the effects of all markers simultaneously. The BLUP approach corresponds to ridge regression (Hoerl and Kennard 1970, Piepho 2009) and Habier et al. (2007) showed its equivalence with the genomic BLUP (GBLUP) approach. In GBLUP, a kinship matrix is calculated from the marker data, leading to a perfect correlation of the resulting predicted breeding values from the two approaches. Genomic selection was rapidly adopted in practical breeding programs and has since been applied to many important traits in animal (Hayes et al. 2009) and plant breeding (e. g., Crossa et al. 2014). In addition to BLUP, other statistical methods have been developed for genome-based prediction, including Bayesian methods (Gianola et al. 2009), reproducing kernel Hilbert spaces regression (de los Campos et al. 2009), and neural networks and machine learning (Maenhout et al. 2007, Montesinos López et al. 2021). However, a comparison of these methods based on various datasets and traits failed to reveal a superior method (Heslot et al. 2012). The relative advantage of a method depends mainly on whether the assumptions associated with it, such as the absence of population structure or simple genetic trait architecture, are met. An improvement of the prediction accuracy estimated as Pearson's correlation between the predicted and observed phenotypic values divided by the square root of heritability could be achieved in some cases by differentially weighting effects and constructing trait-specific kinship matrices (Zhang et al. 2014). Overall, GBLUP has emerged as the gold standard and benchmark for assessing the performance of new prediction methods because it is easy to apply and yields robust results across many selection scenarios (Lorenzana and Bernardo 2009, Bernardo 2020b). Therefore, GBLUP was chosen in this thesis as the statistical method for calculating genomic estimated breeding values.

## 1.2 Hybrid breeding

This thesis is based on data from hybrid breeding programs in rye and maize. A basic prerequisite for hybrid breeding is sufficient heterosis in agronomically important traits, with yield as the first priority (Becker 2019). In rye, C. Kleinert first noticed the heterosis effect, without knowing anything about the phenomenon and its background. He crossed Swedish Snow Rye with Correns Rye and reported more vigorous plants than the Probsteier population (Rimpau 1883). In maize, farmers had already noticed increased vigor when crossing unrelated parents in the 19$^{th}$ century (Kingsbury 2009). This inspired William J. Beal to suggest producing hybrids in maize populations by detasseling one cultivar and harvesting the seed obtained by cross pollination. In practice, the gain in yield was low and, therefore, the effort to produce hybrid seeds was too high for the farmers (Kingsbury 2009).

George H. Shull had a scientific interest in the hybrid technique and noticed the value of inbreeding for developing parental lines, although inbreeding depression in pure-bred lines was high (Shull 1908, 1909). Shull first named the phenomenon of more vigorous hybrid plants “heterozygosis;” later, he coined the term “heterosis” (Shull 1914). In general, heterosis denotes the superiority of the offspring over the mean of the parents. Many explanations exist for the phenomenon of heterosis, but the genetic causes remain elusive. Since the first findings on heterosis, it has been known that unrelated parents are necessary to produce more vigorous hybrid offspring. Hence, for a hybrid breeding program, at least two distinct, genetically distant groups, the heterotic pools, are required. Together they form a heterotic pattern.

For hybrid maize breeding, Figure 6 gives a schematic overview of the selection steps required within each heterotic pool. At the beginning of each selection cycle, new variation is generated by crossing a few elite parental lines to produce many selection candidates. In the first selection step, many candidates are discarded based on unsatisfactory performance per se. The remaining candidates are then crossed to one or several inbred lines as single- or double-cross testers from the opposite pool. The evaluation of the testcrosses provides information about the general combining ability (GCA) of the candidates with the opposite pool. These GCA values reflect the additive genetic values of the candidates and, in the absence of epistasis, correspond to half the breeding values in animal breeding, which also reflect the additive genetic effects with partners from the same population. In a next step, a few of the best candidates

from both pools are crossed and tested as experimental hybrids to find the combinations with the best performance compared with commercial and other experimental hybrids. The genetic values estimated for the hybrids in this stage comprise not only the GCA effects but also the specific combining ability (SCA) of the hybrid, which depends on additive and dominance effects.

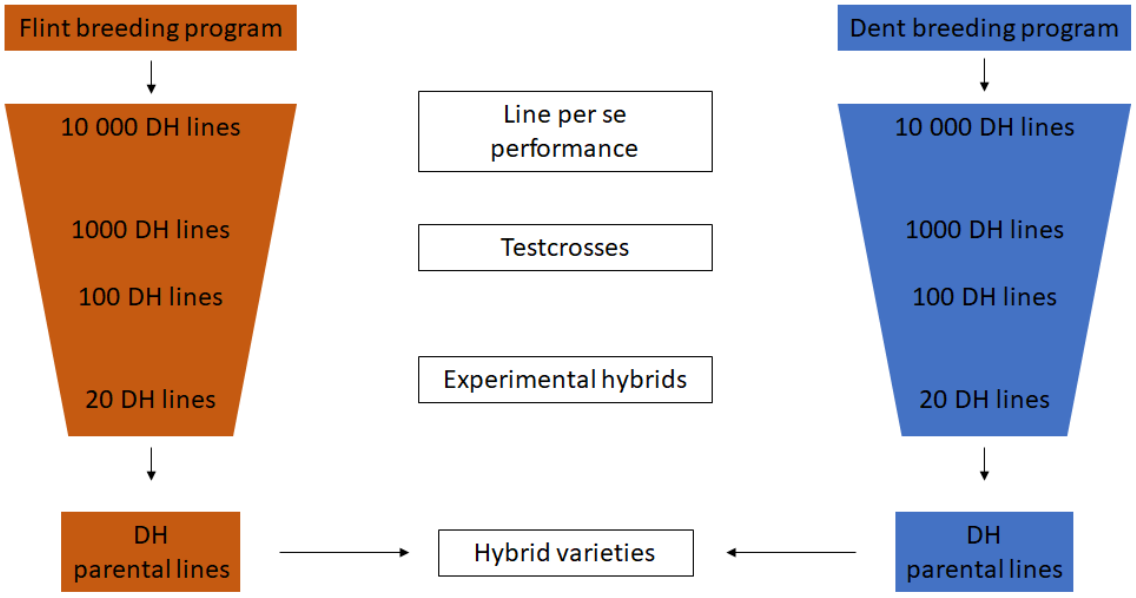


**Figure 6: Principle of one cycle of reciprocal recurrent selection.** (modified after Schmidt 2004)

Different heterotic pools of maize exist. In the US Cornbelt, the main heterotic pools are the Iowa Stiff Stalk Synthetic (BSSS) and non-BSSS pools; the latter is mainly composed of genetic material derived from the Lancaster population. Both pools are dent maize types. In Central Europe, a very important pollen parent (male) heterotic pool is flint material tracing back to landraces cultivated in Europe for centuries after the introduction of maize from the Americas. This material is well adapted to the cool climate in Europe. The seed parent (female) pool is composed of early maturing dent cultivars introduced after World War II at the beginning of hybrid breeding. In the maize experiments of this study (Albrecht et al. 2014, Auinger et al 2021), the implementation of genome-based selection was investigated in the dent seed parent pool and the flint pollen parent pool supplied the testers used to produce the testcross seed. The selection candidates were fully homozygous inbred lines produced with doubled-haploid (DH) technology. The introduction of DH technology was a milestone in maize

breeding. The development of inducer genotypes for in vivo haploid induction permitted the production of DH lines at low costs (Molenaar and Melchinger 2019). Chase (1952) had already described spontaneous haploids in maize, but rates below 1% were inefficient for commercial breeding. In subsequent years, higher rates were achieved and led to the development of widely used inducer lines such as RWS and UH400 (Prigge et al. 2011). Treatment of haploid seedlings with colchicine, chemicals, or nitrous oxide triggers doubled chromosomes at a high rate, and these homozygous diploid plants are generally fertile, allowing selfing. This technique is standard in modern breeding programs because it shortens the cycle length, the most important lever for increasing the selection gain per unit of time.

Compared with maize, rye breeding must accommodate additional technical procedures. Rye is an allogamous species with an effective gametophytic self-incompatibility system (Lundqvist 1956). The development of inbred lines requires self-fertility genes, which occur at very low frequencies in open-pollinated populations. Moreover, rye is recalcitrant to the production of DH lines and the method for producing haploid plants is highly sophisticated (Immonen and Tenhola-Roininen 2003). Furthermore, hybrid seed production on a commercial scale requires an absence of fertile pollen in the seed parent, which cannot be achieved by hand emasculation as in maize. One solution is the use of cytoplasmic male sterility (CMS) (Edwardson 1970) in the seed parent pool. The pollen parent must then possess a dominant restorer gene to restore fertility in the hybrids, if they are cultivated for grain production.

Before the discovery of self-fertility genes (Wricke 1969) and cytoplasmic male sterility in an Argentinean genetic resource (Geiger and Schnell 1970), rye varieties were open-pollinated populations. In a diallel of seven populations (Hepting 1978), the two population varieties Carsten and Petkus exhibited the highest population heterosis and were subsequently used as separate heterotic groups to establish hybrid breeding programs in rye. Traditionally, in hybrid rye breeding, lines of the Petkus pool are used as seed parents and lines of the Carsten pool are used as pollen parents (Geiger and Miedaner 2009). In the rye experiment of this study (Auinger et al. 2016), the implementation of genome-based selection was investigated in the pollen parent pool and the seed parent pool supplied the testers used to produce the testcross seed. As strong inbreeding depression is still an issue in most rye breeding programs, inbred lines are not fully homozygous, Auinger et al. (2016) describe in their study of $S_2$ lines. Because inbreeding depression is still significant in rye, double-cross hybrids are produced for commercialization, unlike maize, where single crosses are the main type

of hybrid. As in other cereals, yield is the primary trait to breed for, but yield stability, yield components, plant height, resistance to fungal diseases, and tolerance of abiotic stresses such as severe frost, drought, and heat are also relevant. Because rye is mainly cultivated as a winter cereal and self-incompatible, the selection cycles are long. Therefore, the development of hybrid components takes many years.

In crops, a hybrid breeding scheme serves two purposes: the development of new hybrid components and the accumulation of favorable alleles in each heterotic pool. Modern breeding programs are called “second cycle” or “advanced cycle” breeding programs because high-performing inbred lines are “recycled” after selection for GCA and SCA to produce the base material for the next cycle. The selection scheme in Figure 6 represents one cycle of such a breeding program. A cycle takes several years and multiple cycles run in parallel. Genome-based selection can be restricted to the prediction of untested selection candidates within the same cycle or across cycles. In this thesis, the advantages and disadvantages of within- and across-cycle predictions were investigated (Auinger et al. 2016, Auinger et al. 2021).

## 1.3 Materials and methods

This thesis comprises two large datasets, one from a rye and the second from a maize breeding program. Both represent mid-sized commercial programs, and in both programs, the data originated from the first testcross selection step, as shown in Figure 6. The programs differ mainly in the progeny’s degree of residual heterozygosity ($S_2$ lines in rye and DH lines in maize) and the number of progenies analyzed in each selection cycle. The main goal of the rye study was to evaluate the prospects of genome-based prediction for grain dry matter yield, plant height, and thousand-kernel weight. The dataset was unique as it allowed us to compare the efficiency of genome-based selection within and across breeding cycles. The main goal of the maize study was to identify useful genomic parameters for choosing datasets from earlier selection cycles to obtain the maximum prediction accuracies in later cycles.

### 1.3.1 Experimental data

The rye dataset comprised testcross data from four subsequent breeding cycles in Poland and Germany from 2009 to 2012. There were 1,416 advanced cycle $S_2$ lines from the pollen parent pool. These lines were progenies of 203 parents used in 430 crosses.

In each cycle, the $S_2$ lines were crossed to two of eight testers from the opposite seed parent pool to obtain testcross seed. The $S_2$ lines were genotyped with a custom Rye 16k Illumina array. The data were filtered for high-quality SNPs (GenTrain score ≥0.7, call rate ≥0.9). SNPs with a minor allele frequency (MAF) of less than 1% and more than 10% missing values were discarded. This resulted in 10,416 useful SNPs, 5,607 of which had available mapping information.

In the field trials, testcrosses of $S_2$ lines with two different but related testers were evaluated for yield, plant height, and thousand-kernel weight. The experimental design used six to nine 16x10 α-designs with two replications, connected by checks. Overall, 14 checks were used with an overlap of one to five checks across cycles. Figure 7 shows how the datasets are connected by common parental inbred lines used to make crosses and generate selection candidates. The sets have 18 to 20 parents in common. With exception of some outliers, the data were balanced.

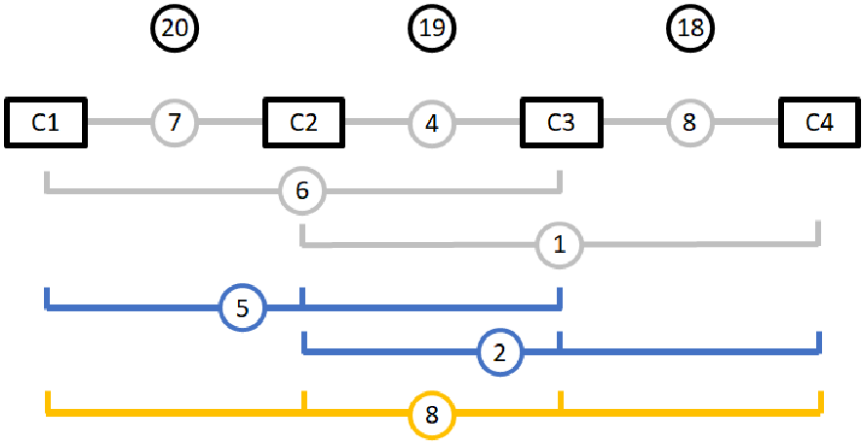


**Figure 7: Parents shared between sets in the rye data.** The black boxes indicate the name of the dataset. The gray circles show the number of parents shared exclusively between two sets. The blue lines and circles refer to the number of parents shared exclusively among three connected sets. The orange lines and circles show the number of parents shared among all four sets. The black circles show the number of parents shared between subsequent cycles.

The maize dataset comprised six datasets (S1, S2…S6) representing the selection candidates from six selection cycles grown from 2010 to 2015 in multiple locations in Germany. The genetic material comprised 5,968 DH lines from the dent heterotic group crossed to one or several flint testers. The publicly available Illumina maize array was used to genotype the DH lines of sets S1 to S4 with 56,110 SNP markers (Ganal et al.

2011). The DH lines of sets S5 and S6 were genotyped with custom-made chips of 12,062 and 22,359 SNPs, respectively. The genotypic data from sets S1 to S6 were merged after quality filtering (GTScore ≥ 0.7, call frequency ≥ 0.9). Monomorphic and fully collinear SNPs, as well as SNPs with the alternative allele appearing only once, were removed. In the full dataset, 9,742 informative SNPs remained. The number of genotyped lines in each dataset varied from 551 to 1,545 DH. Individual datasets were generated by crossing 36 to 148 parents, which resulted in 130 to 607 crosses and 1 to 455 progenies per parent. Figure 8 shows how the sets are genetically connected as the same inbred line might have been used as a parent in multiple datasets. As in the rye data, the connection between the sets varies, e.g., sets S2 and S3 share 49 parental lines, whereas sets S5 and S6 have only three parents in common.

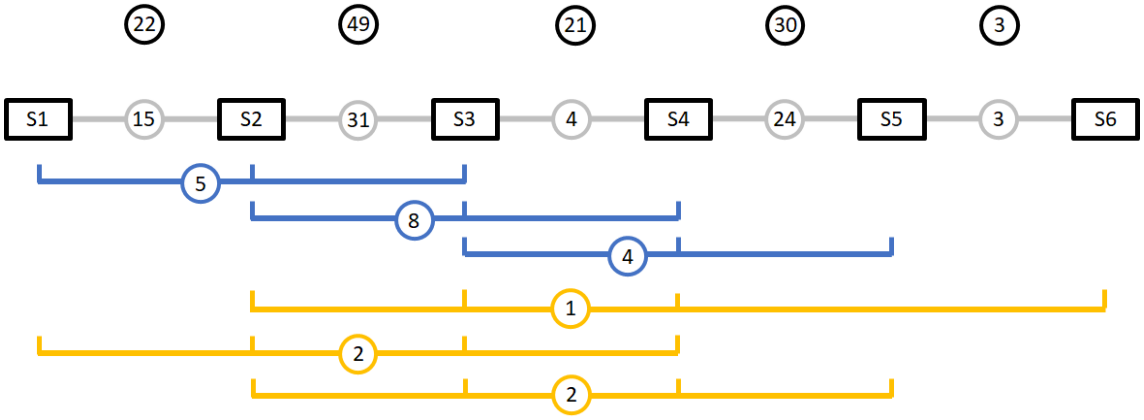


**Figure 8: Parents shared between sets S1 to S6 in the maize data.** The black boxes indicate the name of the dataset. The gray circles show the number of parents shared exclusively between two sets. The blue lines and circles show the number of parents shared exclusively among three sets. The orange lines and circles show the number of parents shared exclusively among four sets. The black circles show the number of parents shared between subsequent cycles.

In the field trials, testcrosses of DH lines were evaluated for yield and dry matter content. The selection candidates from each dataset were grouped in lattice designs with 100 entries each and each lattice featured five to seven checks. Lattice trials were replicated (S1 and S2), partially replicated (S3), or unreplicated (S4, S5, and S6). The six contiguous selection cycles were connected through 11 commercial hybrids that changed over time. In addition, hybrids produced from DH lines related to the selection

candidates and crossed to the same tester as the breeding material were included as checks connecting contiguous selection cycles.

### 1.3.2 Best linear unbiased prediction

Today, the genomic BLUP or the related ridge regression BLUP are considered the gold standards in genomic prediction and are widely used in commercial breeding programs. At the beginning of this study, the implementation and fine-tuning of genomic BLUP were still being discussed, especially for predictions across selection cycles. GBLUP methods can differ in the treatment of the response variable in the model or the calculation of the variance-covariance structure. In the single-stage approach, the response variable is the observation from each plot and all design effects, such as location, years, and replications, are included in the genomic prediction model. This is computationally feasible if the goal is a single prediction from the calibration set to the prediction set. In this study, the single-stage approach was computationally prohibitive because different cross-validation scenarios and many combinations of calibration and prediction sets, as well as subsampling from the calibration sets, were investigated (Auinger et al. 2016, 2021). Hence, phenotypic values from the rye and maize experiments were obtained as Piepho et al. (2012) suggested in a two-stage approach. Adjusted means of environments, corresponding to best linear unbiased estimates (BLUEs) of the genotypes, were calculated and submitted as response variables for a pedigree-based or genomic BLUP model. For data generated in the plant-breeding context, the single- and two-stage approaches have proven to yield highly correlated predictions (Schulz-Streeck et al. 2013). A preliminary analysis confirmed this correlation for both the rye and the maize datasets. The genomic breeding values obtained with the two approaches were highly correlated ($> 0.98$).

Before the era of genomic selection, BLUP was infrequently used in plant breeding. The lack of deep pedigree information, high trait heritabilities as noted above, and the different mating designs compared with animal breeding limited its use. As BLUP relies on relationship matrices constructed from pedigrees, selection among non-phenotyped members of the same family, i.e. progenies derived from the same cross, was impossible because all progenies received the same predicted value. In genomic prediction, the relationship matrix is calculated from the marker profiles of the selection candidates, which allows Mendelian sampling between the members of a family to be

accounted for, and thus, the measurement not only of the expected but also the realized proportion of the genome shared between relatives. In GBLUP, the variance-covariance matrix of the mixed linear model replaces the pedigree-relationship matrix of the conventional BLUP model. The genomic relationship matrix can be calculated in different ways. In Auinger et al. (2016, 2021), Habier et al.'s (2007) method was chosen as it has a defined quantitative genetic interpretation and has been used by many other studies to predict genomic breeding values in plant and animal populations.

For the inbred lines in the rye experiment, up to 10 generations of pedigree information were available. This information was highly useful to evaluate the relative efficiency of genome-based prediction over pedigree-based prediction within and across selection cycles. The pedigree-based relationship matrix had to be adapted to the specific mating scheme employed in the rye experiment. As described in Lynch and Walsh (1998), standard quantitative genetic algorithms exist to calculate the additive genetic relationship matrix A from pedigree information. To increase computational efficiency, the dimensionality of the relationship matrix A was reduced in this study (Auinger et al. 2016). For propagation of $S_2$ lines, a single-seed-descent selfing scheme was assumed. Consequently, the dimensionality of relationship matrix A can be reduced by omitting the selfing steps and modeling the diagonal elements of A for each individual *i* as

$$\bar{} \qquad \bar{}$$

where $A_{ii}$ denotes the diagonal element of A for individual i, $x$ the number of selfing generations, and $A_{gh}$ is twice the kinship coefficient between the parents $g$ and $h$ of individual $i$ in generation $S_0$, i.e., before selfing. For $S_2$ lines derived from $S_1$ plants, $x=1$.

### 1.3.3 Optimal calibration sets for genome-based prediction

In commercial breeding programs, datasets accumulate over time. The datasets used in this thesis allowed a comprehensive analysis of the best scenarios for predicting the phenotypes of the selection candidates. The success criteria for the comparison of different scenarios was the predictive accuracy, calculated as the correlation between predicted and observed phenotypes in the respective prediction sets, standardized for each trait by the square root of its heritability.

In the rye data, fivefold cross-validation (CV), as Utz et al. (2000) suggested, and independent validation were used to evaluate the predictive accuracy of different methods and sample sizes (Auinger et al. 2016). To establish prediction accuracies within cycles using fivefold CV, the dataset was divided into five mutually exclusive subsets. Four subsets (80% of the data) formed the calibration set (CS) for training the prediction model while the fifth subset (20% of the data) represented the prediction set (PS) to determine the prediction accuracy (Figure 9a). CV was performed within each of the four selection cycles. For validation across cycles, lines for the calibration sets were sampled from one, two, or three cycles, and the lines for the prediction set from the fourth cycle were not included in the calibration (Figure 9b). Combined CV sample lines from all four cycles for the calibration set and lines from one of the cycles not included in the calibration were used for prediction (Figure 9c). Each sampling scheme was repeated 50 times. The calibration set sample size was constant to permit direct comparisons between the three scenarios.

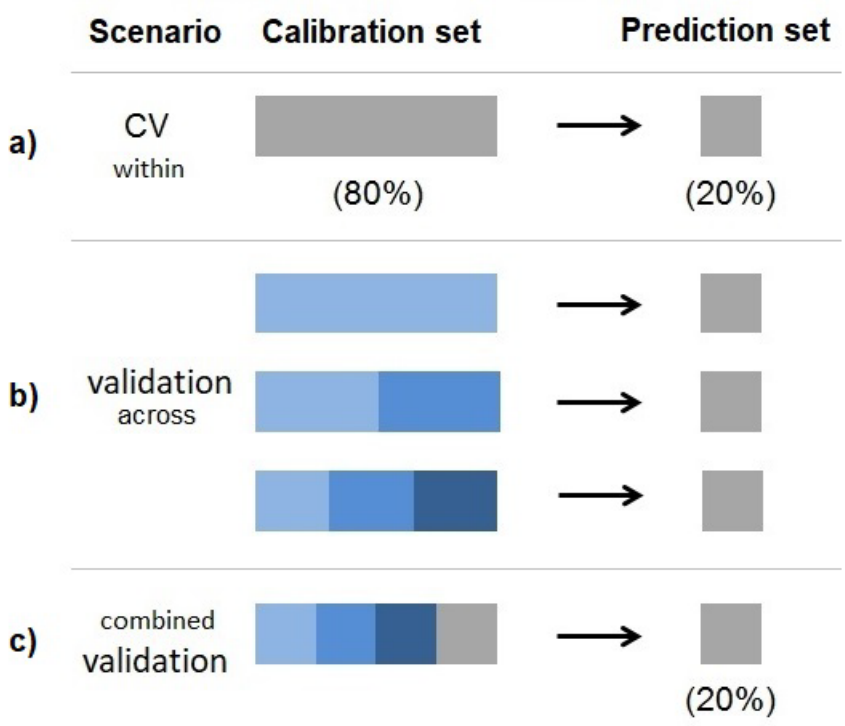


**Figure 9: Cross-validation and validation scenarios.** Cross validation within (a): Lines in calibration and prediction from the same breeding cycle (grey boxes). Validation across (b): Lines in calibration sets from one, two, or three cycles (different shades of blue) with equal numbers of phenotyped lines from each cycle. Combined validation (c): Lines in calibration from all four cycles (blue and grey boxes) and lines in the prediction set from one of the cycles (grey), not represented in the calibration set. Figure modified from Auinger et al. (2016).

In the rye study, increasing the sample size of the calibration set by aggregating data from several cycles increased accuracies in the prediction sets, on average. This was also true for the maize dataset, but some combinations of datasets performed better than others. In maize, data were available for two more selection cycles than for rye. This permitted the construction of 15 forward prediction scenarios to predict the lines tested in dataset S5 and 31 scenarios for S6. The 46 calibration-prediction combinations form a large enough database to investigate the usefulness of genomic parameters for maximizing prediction accuracies in the prediction sets.

Five parameters were investigated in detail for their ability to forecast the success of prediction across selection cycles in the maize data. The parameters described in Auinger et al. (2021) were:

1) The number of common polymorphic SNPs in the 46 combinations of calibration and prediction sets (nPoly).

2) The linkage phase similarities (*LPS*) for all combinations of calibration and prediction sets, according to Schopp et al. (2017):

$$\frac{\quad}{\quad\quad\quad\quad\quad\quad}$$

where $k$ is the index for the $p$ marker pairs. The sign of $r_k^{CS}$ was inferred by calculating the $D = p_{AB} - p_A p_B$ of marker pair $k$, where $p_{AB}$ denotes the frequency of haplotype AB, $p_A$ the frequency of allele A at one marker locus and $p_B$ the frequency of allele B at the other locus in the calibration set.

3) The effective sample size $1 \leq N_{eff} \leq N$ is given by the following formula (Auinger et al. 2021):

$$\frac{\quad}{\quad\quad}$$

where $N$ denotes the size of the calibration set under study and *var*( ) the estimated variance of the eigenvalues of the genomic kinship matrix pertaining to the GBLUP model.

4) The degree of relatedness between the calibration and prediction sets, after Saatchi et al. (2011). For all possible combinations, the average maximum realized kinship coefficients ($u_{max}$) were calculated based on the genomic kinship matrix pertaining to

the GBLUP model. The maximum kinship of line *i* of a given prediction set was derived as the maximum value of the realized kinship coefficients between line *i* and lines *j* of the corresponding calibration set. Averaging the lines in the prediction set produced the $u_{max}$ value for each combination of calibration and prediction sets.

5) The average expected reliability, after Clark et al. (2012). Here, the reliability of line *i* in the prediction set is calculated from its prediction error variance (*PEV* (*i* , derived from the GBLUP model employing a specific calibration set as follows:

where is the genomic variance pertaining to the GBLUP model and $_{ii}$ is the diagonal element of the realized genomic relationship matrix referring to line *i.* Averaging all DH lines in the prediction set gives the expected reliability estimate for each combination of calibration and prediction sets. The expected reliability is trait-specific.

The relative importance of the genomic measures, their correlation with sample size *N*, and their interdependencies were analyzed with principal component analysis (Jolliffe and Cadima, 2016) and multiple regression, with empirical prediction accuracies for grain yield and grain dry matter as response variables (Auinger et al. 2021).

To evaluate variations in the above parameters (nPoly, , and $^2$) for a fixed sample size of the calibration set, 500 lines were sampled randomly without replacement from sets S1 to S4 and all their combinations. Sampling was repeated 1,000 times. Each sample was used as a calibration set to calculate GBVs and the resulting prediction accuracies for S5 and S6. The distribution, mean, and variance were determined for the 1,000 samples of 500 lines for all five parameters.

## 1.4 Outline

Traditional plant breeding methods relied exclusively on estimating the genotypic values of genotypes based on replicated field trials. While the BLUP approach Henderson (1975) introduced revolutionized animal breeding, it was hardly used in plant breeding because it yielded identical predicted values for all members of a given family. This changed dramatically with the introduction of the of the concept of predicting breeding values based on genomic information (Meuwissen et al. 2001). The major difference between breeding values based on pedigree relationships and those based on genomic

relationships is that not only can the expected breeding value be estimated without pedigree information but differences in the true relationships among members of the same family can be measured. These differences are due to the phenomenon of Mendelian sampling, which means, for example, that siblings in a full-sib family might share 10%, 40%, or 80% of their genomes due to Mendelian segregation in meiosis. Thus, the genotypic value of each genotype of a full-sib family can be estimated, whether it was phenotyped or not, and used as a basis for selection. The idea of genomic selection was rapidly accepted because the costs of marker assays decreased dramatically—as Moore's law predicts—over the past two decades, whereas the costs for phenotyping remained almost constant. Thus, the cost for evaluating the breeding potential of a genotype is nowadays typically much lower for genomic than for phenotypic selection.

Additionally, genomic prediction can shorten the generation interval and broaden the basis for selection by assessing many candidates purely based on their predicted genotypic value (Hayes et al., 2009; Elshire et al., 2011; Bhat et al., 2016; Meuwissen et al., 2016; Crossa et al., 2017; Weller et al., 2017). In the last decades, extensive research has been conducted to identify optimal statistical methods to predict breeding values. For traits having a genetic architecture with many small effects and low to medium heritability, the BLUP method remains the method of choice. Various Bayesian and nonparametric methods, as well as machine-learning methods, have been suggested to improve prediction accuracies and account for traits of different genetic architectures (Gianola et al. 2009, Heslot et al. 2012, Morota and Gianola 2014, Zhang et al. 2019). However, the vast number of markers is a computational burden for Bayesian methods because many parameters must be estimated (Zhang et al. 2019). Hence, this thesis uses the GBLUP statistical method.

Albrecht et al. (2014) focus on comparing the traditional pedigree-based BLUP with genome-based BLUP. Special attention was given to how the prediction accuracy is influenced by the genetic structure in the studied material and prediction across locations and years.

In Auinger et al. (2016), the main research question was how the usage of data from additional years and selection cycles influences prediction accuracy. In the four-year data, predictions within and across cycles were compared using PBLUP and GBLUP.

In Auinger et al. (2021), the impact of genetic diversity, the sample size of the calibration set, and the genetic connection between calibration and validation sets were

addressed. To do so, the degree of marker polymorphism, linkage disequilibrium, relatedness and reliability, and their effects on prediction accuracy were investigated.

# 2 Discussion

## 2.1 PBLUP versus GBLUP

Accuracies in genome-based prediction are expected to be higher than in pedigree-based methods because modeling the relatedness based on marker data accounts for Mendelian sampling in crosses sharing one or both parents. Albrecht et al. (2011) convincingly demonstrated this in commercial breeding material in maize and Auinger et al. (2016) in rye. The increased accuracy of GBLUP compared with PBLUP depends on the effective number of segregating loci in the population (Goddard 2008) and the mating design (Habier et al. 2013). In advanced cycle breeding populations, where effective population sizes are small and which are derived from a limited number of crosses as recombinant inbred or doubled haploid lines, genome-based selection exploits long-range LD and the co-segregation of marker-QTL associations.

In the maize study, the variation of realized genetic relationships among selection candidates sharing the same pedigree was high. This leads to a strong increase in prediction accuracy for models using genomic data. For rye, this thesis demonstrated that the advantage of GBLUP over PBLUP was most visible when the model was trained on aggregated data from several selection cycles. While PBLUP accuracies for plant height and thousand-kernel weight were close to zero when predicting across cycles, both methods' models for grain dry matter yield showed intermediate prediction accuracies within and across cycles.

In the rye study, the average family size was rather small for many crosses, and PBLUP and GBLUP achieved similar accuracies for the prediction of grain dry matter yield. This aligned with expectations and suggests that the whole genome, rather than a specific genomic region, is involved in the expression of complex traits. For plant height and thousand-kernel weight, PBLUP and GBLUP prediction accuracies were intermediate to high within cycles. In contrast, across cycles pedigree-based prediction averaged close to zero and genome-based prediction. One explanation is that the close familial relationships within cycles favored pedigree-based prediction, which estimates family means well. Across cycles, the families are more distantly related and family members are rarely included in the calibration set. In contrast, the GBLUP model can capture some persistent effects in the across-cycle scenarios due to linkages between markers and QTL.

Similar to the testcross performance in rye, Michel et al. (2016) compared GBLUP with PBLUP in a commercial wheat breeding program. For grain yield and protein content, GBLUP was superior to PBLUP, especially when information from early generations of the candidates was not available. As a major explanation for their findings, the authors mentioned that PBLUP does not address the Mendelian sampling responsible for the variation within progeny families. If information from early generation trials cannot be exploited, the within-family prediction accuracy is zero for PBLUP. However, combining pedigree with genomic information in two-kernel models substantially increased the prediction accuracy and the selection response y, especially for traits with low heritability. In further studies with autogamous crops (wheat and barley), genome-based prediction accuracy was much lower within families than across a diverse panel of lines and families (Edwards et al. 2019, Sweeney et al. 2020). In simulation studies, prediction accuracies within full-sib families were highest (Schopp et al. 2017), while including some members from the same family significantly improved the prediction accuracy in half-sib families.

Besides the population structure, the genetic architecture of a trait considerably influences the prediction accuracy (Auinger et al. 2016). This hypothesis is supported by a study on genome-based prediction of grain yield and plant height in a biparental rye population derived from two elite parents (Wang et al. 2014), where prediction of plant height based on previously identified QTL performed similarly to genome-wide prediction but the genome-wide prediction of grain yield was clearly superior. This is supported by results from QTL analyses indicating the complex polygenic nature of grain dry matter yield (Miedaner et al. 2012), which can only be accounted for by genomic prediction.

As genomic data will accumulate over time, these data will be useful not only to predict genetic values but also to infer marker effects linked to major QTL. Thus, it seems promising to combine marker-assisted and genomic selection by including the effects of major QTL alleles as fixed effects in the statistical model and a polygenic component capturing the contributions of minor-effect QTL by genomic prediction as random effects with the usual genomic relationship matrix. Zhang et al. (2014) showed that this improved prediction in a cattle dataset for two of three traits, and in a rice diversity panel, including QTL improved the prediction for nine of eleven traits. For other data structures, further research is needed.

Merging data across selection cycles strongly affects prediction accuracy. As Auinger et al. (2016) demonstrated, the prediction accuracy for thousand-kernel weight in rye increased with GBLUP but remained almost constant with PBLUP. Aggregating data from several selection cycles increased the sample size of the calibration set and allowed marker effects to be modeled over several testers and years, which increased the mean prediction accuracy and slightly reduced the prediction error variances, improving prediction reliability. This is very important to optimize breeding schemes but is often overlooked when discussing the potential of genome-based selection. Thus, the reduced prediction error variance argues for model training across several selection cycles.

Results with testcrosses accord with those reported for a hybrid breeding program in maize, for which data from factorials evaluated over a period of 15 years were analyzed (Schrag et al. 2019). Like aggregating results across selection cycles, GBLUP benefited more than PBLUP from including data from early years in the extended calibration set, especially for grain dry matter content. The calibration sets tested in the early years were less related to the prediction set, so the pedigree relationships were rather weak, but genome-based prediction could use linkage disequilibrium between QTL and markers. According to the theory presented by Habier et al. (2013), this is the component of prediction accuracy that persists over generations. When years in the calibration set, which are closest to the prediction set, were skipped and the gap between the calibration set and the prediction set widened, the advantage of genome-based over pedigree-based prediction was largest. This was especially true for the calibration sets comprising all previous years. Thus shows, pedigree-based prediction is heavily dependent on calibration set data from years closest to the prediction set, while genome-based prediction benefits from adding past years to extend the calibration set and improves the prediction of the genotypes of future breeding cycles.

## 2.2 Predictions across locations

In most plant-breeding programs, lines are tested usually only in a single year. Only hich performing candidates are tested in the field in subsequent years again. Selected lines are grown in locations, which are chosen to represent environments the breeding program is destined for, and at best capture genotype × environment interactions, to receive representative results. Using data from Albrecht et al. (2014), their analysis was extended to across-location cross-validation (CV) to assess whether across-location accuracies can predict accuracies across years. CV of environmental sampling was performed for a subset of 928 DH lines evaluated in four locations in 2010. The accuracies were compared to results from validating a constant set of genotypes tested in 2010 and 2011. Predictions for grain dry matter yield ranged from 0.36 to 0.53 and were quite similar to the results of validation across years. Thus, for grain dry matter yield, the effect of across-location prediction was well captured by across-location CV. For grain dry matter content, the accuracies were slightly overestimated. This could have resulted from the different growing conditions in 2010 and 2011. While 2011 was optimal for maize production, 2010 was a cold and wet year. Some of the late DH lines might have struggled to mature properly. Burgueño et al. (2012) suggested including genotype x environment interactions in the prediction model. However, as only two years are compared in Albrecht et al. (2014), little prediction improvement could be expected by including interactions.

## 2.3 Predictions across testers

For selection in hybrid breeding genomic prediction of testcross performance was found to be a promising tool. Restrictions are the usage of the same tester and the same environments in the calibration as in the prediction set (Albrecht et al. 2011, Riedelsheimer et al. 2012, Lehermeier et al. 2014, Lian et al. 2014, Galán et al. 2021). In practice, however, prediction accuracy is reduced due to change of testers, or target environments, of selection targets, or traits. If DH lines are evaluated for their testcross performance with only one inbred or single-cross tester, the effects of their general and specific combining ability cannot be separated with GBLUP (Albrecht et al. 2011). Nevertheless, it would be desirable to estimate the expected decline in the prediction accuracy of genomic prediction for given calibration data in such conditions.

A quantitative genetic analysis of the genetic correlation of testcross performance with different testers shows that the correlation among testers depends directly on the relative size of the variances of GCA and SCA effects, and , respectively, which can be modeled with:

$$\frac{}{\sqrt{}} \qquad \frac{}{}$$

where is the general combining ability of the candidate and ( ) is its specific combining ability with the tester *T1* (*T2*). Thus, if is much smaller than , the genetic correlation between testcrosses is close to 1. If there are distinct heterotic groups, as in hybrid maize breeding in Europe and the USA, then the relative importance of relative to is expected to be small (Reif et al. 2007). The genetic correlations between homozygous lines crossed to different testers of the opposite should be medium to high. Schopp et al. (2015) showed that under certain assumptions about the stochastic independence of the error terms of genomic prediction with different testers, the prediction accuracy for testcross performance with a new tester can be described as the product of the prediction accuracy for testcross performance with the original tester and the genetic correlation between the two testers. In the maize data, genetic correlations between testers could not be estimated because each DH line was only crossed to one of several testers, so there was no correlation among testers. However, if a lower bound of is assumed based on results from the literature (e.g., Melchinger et al. 1998), the prediction accuracies obtained with a different tester in Albrecht et al. (2014) were rather low. This is especially unsatisfactory, because the two Flint single-cross testers T1 and T3 shared a common parent. DH lines were assigned to the tester according to maturity rather than at random. This could explain the low prediction accuracies across testers. Population structure due to subsets or large effects of specific combining ability may also have affected the results. The choice of tester and the size of specific combining ability effects needs further research in the context of genomic prediction.

## 2.4 Predictions across selection cycles

In both studies conducted for this thesis (Auinger et al. 2016 and Auinger et al. 2021), prediction across selection cycles was possible. However, prediction accuracies were reduced compared to within-cycle prediction unless model training was performed with data from several selection cycles, which naturally requires substantially larger calibration sets. The prediction accuracy is known to depend heavily on the relationship between the calibration set and prediction set (Clark et al. 2012, Scutari et al. 2016), so the relationships among selection cycles were analyzed.

In the rye study, the average maximum kinship of calibration and validation sets in across-cycle scenarios was about half that of within-cycle scenarios. As the breeding program advances, new unrelated parents are used to counterbalance inbreeding, so more distant selection cycles share fewer common ancestors (Figure 7 and Figure 8). This decreased connection through common parents over time was reflected in reduced kinship, which was accompanied by significantly reduced prediction accuracy when predicting lines from cycle 4 with a model trained in cycle 1. However, in the rye study, the association of prediction accuracy and average maximum kinship was only intermediately strong for grain yield and not significant for the other traits. This contrasted with other studies where the average maximum kinship and the prediction accuracy showed a tight linear relationship (Habier et al. 2010, Windhausen et al. 2012, Albrecht et al. 2014, Schrag et al. 2019). In plant populations are often introduced unrelated material. Therefore, the average maximum kinship must be interpreted with caution as a predictor for prediction accuracy, even for complex traits such as grain yield. If the calibration set comprises entries that are closely related to many entries of the validation set, e. g. one line with a resistance gene introduced as parental line in many crosses, the average maximum kinship will be high, but the prediction accuracy may not be. The latter depends on the genomic relationships of every genotype in the prediction set to all genotypes in the calibration set, as follows from the general formula of the BLUPs obtained as solutions to the mixed model equations (Henderson 1975). Therefore, the role of parameters like the average maximum kinship on the prediction accuracies was further investigated in Auinger et al.'s (2021) maize study.

While adding related genotypes should always improve the prediction accuracy, the benefit of including unrelated individuals in the calibration set remains unclear. No effect is expected because unrelated or distantly related lines contribute only little to prediction performance (de los Campos et al. 2013), but a few experimental studies

reported a negative effect. A closer examination revealed that these studies generally involved structured populations such as different animal breeds (e.g., Lund et al. 2014), different plant breeding programs (e.g., Lorenz and Smith 2015), or large biparental families (e.g., Riedelsheimer et al. 2013). In Auinger et al. (2021), no decrease in the prediction accuracy was observed when aggregating data from several cycles. Thus, as long as selection cycles share sufficient common ancestors and prediction accuracy has not plateaued to further increases in the sample size, aggregating data from several selection cycles to predict the phenotypes of subsequent selection candidates is recommended despite the decreasing relatedness over time.

In the across-cycle prediction scenarios of the rye study, whether the increased prediction accuracy was mainly attributable to the increased calibration set size or the combination of data from different years and testers was investigated. Therefore, CV was performed with fixed calibration set sample sizes of N = 208, 416, and 624 but different compositions in that the genotypes originated from one, two, or three cycles. The prediction accuracies were essentially the same when, for a given N, the lines were sampled from one, two, or three cycles. This shows that the main driver for the increased prediction accuracy observed with aggregated data over cycles was the increase in the sample size of the calibration set, whereas estimating marker effects from data collected with multiple testers over several years was only of minor importance. These findings align with the experimental results Auinger et al. (2021) reported in maize, where sample size was also a major driver of prediction accuracies.

Lorenz (2013) investigated the effect of sample size and replication on the prediction accuracy of GBLUP in a simulation study. He observed that the allocation of the two factors is very flexible and therefore a second order problem. However, when experimental plant populations have small effective population sizes, prediction accuracy seemingly cannot be increased beyond a certain upper limit, even with greater sample sizes (Albrecht et al. 2011; Jan et al. 2016). Compared with our results for the genome-based prediction of complex traits in rye and maize, studies on genome-based prediction in self-pollinating crops showed few advantages of merging datasets from subsequent progeny sets (e.g., Sallam et al. 2015). A major difference is that, in autogamous crops, the calibration set is often composed of highly unbalanced historical datasets with many lines phenotypically evaluated at low intensity, which only marginally enhances the prediction accuracy. In the rye study (Auinger et al. 2016), the predictive ability of GBLUP for grain yield increased steadily up to a sample size of about 800 S2 lines (Figure 10). The variation of predictive ability increased with

increasing calibration set sample sizes because the size of the prediction set decreased simultaneously. Erbe et al. (2010) observed this as well. For a given constant prediction set, as in the maize study, the findings are inverted. The range decreases dramatically with increasing calibration set size (Auinger et al. 2021). Considering the costs of phenotyping and genotyping, the results suggest how to optimize the allocation of resources in a rye breeding program to maximize the selection gain from genome-based selection per unit of time and budget.

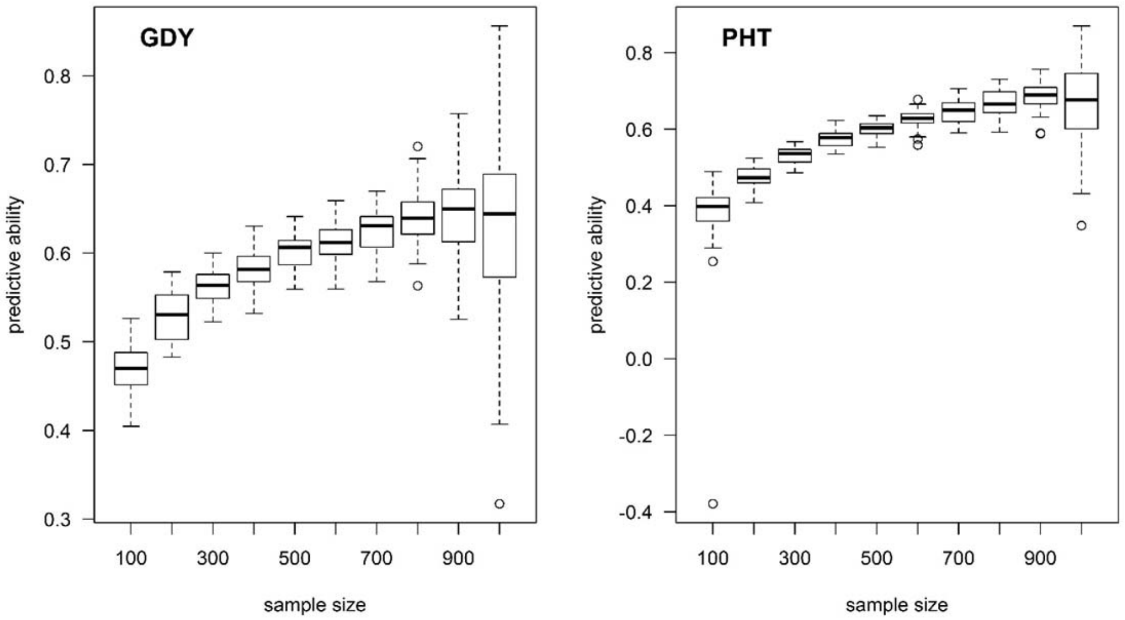


**Figure 10: Predictive abilities for grain dry matter yield (GDY) and plant height (PHT) as a function of increasing sample size in the calibration set (CS) and decreasing sample size in the prediction set.** The calibration and prediction sets from the 1,040 S2 lines were sampled in steps of 100 and repeated 50 times for a given CS. Here, predictive abilities instead of accuracies are given because accuracies could not be estimated properly with the prediction set formed of lines from several cycles with different heritabilities (Supplemental Figure S7 from Auinger et al. 2016).

The data analyzed for maize (Auinger et al. 2021) also originated from a medium-sized commercial hybrid breeding program. However, when the experiments were planned, they were designed for phenotypic selection, because practical implementations of genomic selection were in their infancy. Thus, there was considerable heterogeneity within and between datasets S1 to S6 for sample size, mating design, the number of crosses, progenies per cross, and the relatedness of DH lines. The six datasets all

comprised germplasm from the same heterotic group that was adapted to the same maturity zone, but the testcrosses from each set were evaluated in different years and thus experienced different environmental conditions. As the breeding cycles were at least three years, direct progenies produced from crosses with selected DH lines of a previous cycle were tested several years later (e.g., when the selected parents were members of set S1, some of their progenies were tested in S4). To capture this connectedness among cycles, analyzing at least four cycles was of utmost importance.

By merging datasets across years, prediction accuracy could be significantly increased for both prediction sets and both grain yield and grain dry matter, but some calibration sets comprising only three and four datasets yielded slightly higher accuracies than those observed for the full set (see Fig. 2 and Suppl. Table S3 from Auinger et al. [2021]). Thus, there seems to be room to create optimized calibration sets using existing genotypic and phenotypic data. In the literature, different optimization criteria have been suggested, including the prediction error variance and the $CD_{mean}$ criterion (Rincent et al. 2012; Isidro et al. 2015), which require only genotypic data, the EthAcc criterion (Mangin et al. 2019), which requires both genotypic and phenotypic data, and the sparse selection index (Lopez-Cruz and de los Campos 2021), which integrates selection index methodology with sparsity-inducing techniques commonly used for high-dimensional regression to identify the best calibration set for each individual in the prediction set. Despite the availability of these tools, identifying the best calibration sets is difficult and, for larger datasets, a computational challenge as well. To model systemic effects such as testers, genetic groups, years, locations, and trials correctly, in-debth knowledge of the genetic composition of the population is necessary (Albrecht et al. 2014). Consequently, the calibration set must be specifically optimized for each trait. In general, optimization is most rewarding for improving prediction accuracy for small sample sizes but shows diminishing returns for larger calibration sets (e. g., Daetwyler et al. 2008, Zhong et al. 2009, Isidro et al. 2015, Michel et al. 2017, Norman et al. 2018, de Bern Oliviera et al. 2020, Auinger et al. 2021). As the range of prediction accuracy for random samples decreases for larger N values (Figure 3 of Auinger et al. [2021]), the application of any of the above trait-specific optimization criteria to choose DH lines from the entire dataset for better prediction accuracy is not convincing. Brandariz and Bernardo (2019) proposed the general combining ability (GCA) model for genomic prediction in hybrid breeding, where only crosses sharing a common parent with the prediction set are used for model training. However, this approach is recommended only if both the biparental families to be predicted and the calibration set are fairly large,

and thus, did not apply to the medium-sized breeding populations in this study, where fewer progenies per cross and fewer crosses per parent were available. Thus, other genomic measures were investigated for their association with the prediction accuracy of the entire dataset comprising different selection cycles and different years, as detailed in section 2.5.

Michel et al. (2016) pointed out that prediction accuracies estimated by cross-validation within cycles could be biased. To examine the potential bias, they treated breeding cycles as folds in cross-validation and sampled nonoverlapping sets of genotypes from multiple breeding cycles as training populations to predict an independent breeding cycle. Their prediction yielded accuracies on the upper bound of previous reports from studies of admixed wheat populations (Heffner et al. 2011; Poland et al. 2012; Combs and Bernardo 2013; Storlie and Charmet 2013; Charmet et al. 2014; Isidro et al. 2015, Dreisigacker at al. 2021), but lower as in Haile et al. (2020) with a special composition of the material under study. Longin et al. (2015) found prediction accuracies of about 0.3. Combining these findings, Michel et al. (2016) considered prediction accuracies between 0.3 and 0.4 realistic for predicting wheat grain yield across breeding cycles. Interestingly, Piepho et al. (2014) analyzed German official variety trials of bread wheat and calculated variance components of long-term trends. They found, that the values for prediction accuracy correspond to the broad-sense heritability. In a barley breeding population, Sallam et al. 2015 reported prediction accuracies from 0.03 to 0.99 for grain yield and three other traits and found connections to genetic architecture and heritability.

Hofheinz et al. (2012) also investigated the genome-based prediction of testcross performance in sugar beets using genotypes from the preceding breeding cycle for training and compared it with the prediction accuracies obtained by cross-validation within cycles. In the latter case, the estimates were high regardless of the trait, whereas independent validation using materials from the subsequent breeding cycle were high for sugar content but low for molasses loss, similar to the heritabilities for both traits. Thus, cross-validation in breeding pools of related material from the same breeding cycle seems to over-fit the prediction model. Consequently, high correlations between predicted and observed performance in cross-validation are no certainty that the prediction model can be applied successfully to new breeding materials. This brings the validity of comparisons of different statistical approaches for genomic prediction reported in the literature, such as RR-BLUP or various Bayesian methods that were based on comparisons of prediction accuracies determined by cross-validation within

materials from the same breeding cycle, into question for their prediction performance across cycles.

In contrast to the above studies, He et al. (2016) reported only a marginal loss in accuracy for a large commercial population comprising 2,325 European winter wheat lines when data collected from one year were used to predict genotypes in the following year. They concluded that using robust genomic selection models calibrated with a huge dataset can predict the performance of materials in subsequent breeding stages with decent accuracy. Interestingly, excluding genotypes evaluated only in a single location from the calibration set maximized the prediction ability, but the accuracy decreased again when phenotyping was performed in more than two locations. Consequently, an updated training population should include all selected genotypes except those evaluated in a single location. However, the findings from this study must be interpreted with caution because, although no clear grouping of the materials was detected by principal component analysis, the proportion of the variation explained by the first two principal components far exceeded the values reported in other comparable studies.

For rye, Bernal-Vasquez et al. (2017) evaluated different genomic prediction approaches to simultaneously model genotype x year interactions and breeding values. They compared different scenarios in practical breeding programs, including different degrees of relatedness of the calibration and prediction sets and selected genotypes from the previous generation. They concluded that using two or three years in the calibration set is always advantageous because it allows genotype x year effects to be separated from genomic estimated breeding values. Moreover, the authors stressed the critical importance of the predicted candidates being sufficiently related to the genotypes in the calibration set.

## 2.5 Genomic parameters to maximize prediction accuracy

Compared with calibration sets chosen at random, the application of criteria such as CDmean proved useful for identifying more optimal calibration sets for the genomic prediction of individuals or sets of genotypes with the aid of genetic or simulated annealing algorithms or a combination of these (Akdemir et al. 2021). Nevertheless, to redesign breeding schemes and take full advantage of the new possibilities offered by genomic prediction, knowledge about the association of prediction accuracy with various genomic parameters could be enlightening. All of the genomic parameters

(effective sample size [$N_{eff}$], average maximum kinship [$u_{max}$], expected reliability [ $^2$], number of common polymorphic SNPs [nPoly], and linkage phase similarity [LPS]) analyzed by Auinger et al. (2021) in the maize dataset were significantly correlated with prediction accuracy.

The sample size (N) was tightly correlated to the prediction accuracy for both grain dry matter yield and grain dry matter content (Figure 11). The importance of the calibration set sample size to the prediction accuracy was quantified under simplifying assumptions in the formula derived by Daetwyler et al. (2008). A larger sample size improves the precision of the estimated effect of markers associated with trait expression and, therefore, directly affects the prediction accuracy provided the markers are good proxies for the quantitative trait loci affecting the trait. As the parameter $N_{eff}$ depends on the linkage disequilibrium and N, it was unsurprising that prediction accuracy and $N_{eff}$ were closely associated.

The average maximum kinship ($u_{max}$) was also closely correlated with prediction accuracy, which translated into specific relations between the calibration and prediction sets. For example, if one of the datasets (S3) was included in calibration sets to predict set S6, the $u_{max}$ was much higher than in calibration sets without S3, illustrating the shortcomings of the $u_{max}$ as a parameter to choose and optimize calibration sets. Therefore, quantiles of kinship coefficients, or the mean and median of the kinship coefficients, were also evaluated as alternatives to reduce the exclusive weight given to the closest related individual and get a more robust measure. However, these alternatives yielded no improvement supporting the use of the $u_{max}$ to characterize the relationship between the calibration and prediction sets.

The tight correlation of prediction accuracy with expected trait-specific reliability ( $^2$) aligned with Ould Esthaghvirou et al.'s (2013) theoretical results. They suggested this parameter as an alternative to determining the prediction accuracy by cross-validation. He et al. (2016) also suggested reliability as an optimization criterion and found it superior to taking random samples for calibration. While these authors recommended including the genotypes with the highest $^2$ in the calibration set, based on Ould Esthagvirou et al.'s (2013) theoretical results and the experimental findings in this thesis, I recommend choosing lines for calibration that maximize the mean reliability of the genotypes in the prediction set.

As the number of polymorphisms (nPoly) shared between the calibration and prediction sets depends strongly on N, its correlation to the prediction accuracy for grain dry

matter yield was unsurprising. The magnitude of nPoly also depends on the composition of the prediction sets and varied by about 20%, exceeding the variation among the calibration sets. This difference could be due to the SNPs selected for the custom-made chips used to genotype sets S5 and S6, which had only about one-third of the polymorphic markers on the commercial 50k Illumina chip.

Schopp et al. (2017) examined the association of linkage phase similarity (LPS) with prediction accuracy in detail. With high LPS, the associations between markers and QTL present in the calibration set are more likely to carry over to the prediction set. In synthetics, this association was tight if (i) the populations were generated from a small number of parents and (ii) the calibration and prediction sets were closely related and separated by no more than one generation of intermating (Müller et al. 2017). Weir and Hill's (1980) analytical results provide a link between the LD parameter $r^2$ and the sample size. Accordingly, the expectation for $r^2$ in populations of DH lines is given by:

$$\_\_\_\_\_ \quad \_$$

where c is the recombination frequency between the loci, is the effective population size and $n$ is the sample size. As the measured LD depends on the size of the underlying set, the LPS is also influenced by the sample size via the LD in the compared sets. This may explain why, in this thesis, LPS values between calibration and prediction sets differed substantially by whether the prediction set was S5 or S6 because S5 is nearly three times bigger than S6. However, this formula does not explain the differences in LPS between calibration sets of equal size. Potentially, equal-sized sets with different LPS have different effective population sizes. Alternatively, Weir and Hill (1980) mention that deviations from random mating in the population might cause deviations from their theoretical results; therefore, their theoretical results might cause the differences between the equal-sized sets in this thesis.

To determine which of the genomic parameters and N had the strongest influence on the prediction accuracy, multiple regression models were fitted with a prediction set coded as a categorical covariable and calibration set size and the genomic parameters as regressors. The variance explained by the different models was very similar. Reliability was consistently included in the best 10 models with the best fit in model selection, together with one or two additional factors. nPoly was also always included in the best models for grain dry matter yield, with other parameters. Further, the size N

of the calibration set was always included in the best models, although it showed some degree of collinearity with all genomic parameters.

In summary, the size of the calibration set is indisputably the main driver of prediction accuracy. However, the genomic parameters have additional features that contribute to differentiation between the calibration sets. Their utility for improving the prediction accuracy seems largely trait-dependent. Hence, further research is warranted to investigate how additional knowledge about genetic trait architecture could help to identify calibration sets that are optimized for high prediction accuracy for each trait.

## 2.6 Predictions with constant sampling size

Several studies in the literature reported substantial variation in the prediction accuracy among biparental progenies (Marulanda et al. 2015; Lian et al. 2014) and other types of populations (Albrecht et al. 2011, Würschum et al. 2013, Lehermeier et al. 2014) for a given size of calibration set, without finding convincing explanations. This motivated me to more closely investigate the prediction accuracies obtained in sets S5 and S6 with different calibration set compositions. First, all genotypes in sets S1 to S4 and their combinations were used as “full” calibration sets with a large sample size ($842 \leq N \leq 3,872$) range. Second, random subsets of $N = 500$ genotypes from each of these full calibration sets were sampled as new calibration sets and the mean, minimum, and maximum of the prediction accuracies obtained from 1,000 subsets derived with this sampling process were calculated. Third, for a given full calibration set, the 1,000 subsets with $N = 500$ were ranked in decreasing order of their prediction accuracy and for each genotype, the average of the rank of the subsets in which it was included was calculated. Finally, two new calibration sets with $N = 500$ were assembled, one containing the genotypes with the highest average rank, termed highly predictive (+), and one containing those with the lowest average rank, termed poorly predictive (-), and the prediction accuracies for these calibration sets were determined.

Tables 2 and 3 show the prediction accuracies obtained with this huge number of calibration subsets for prediction sets S5 and S6 and grain dry matter yield and grain dry matter content in the maize study presented by Auinger et al. (2021). The prediction accuracies were similar in both sets for grain dry matter content but much lower in set S6 than S5 for grain dry matter yield. Interestingly, the prediction accuracies for the full calibration sets, with N ranging from 842 to 3,872, were almost the same as the mean

of the 1,000 subsamples with N = 500. This indicates that the average prediction accuracy seems to have plateaued for N > 500. However, the variation in the prediction accuracy among the 1,000 replications with N = 500 was quite substantial, as the range shows. The range was extremely large for grain dry matter yield in set S6, where the minimum values were often negative. Consequently, if many genotyped and phenotyped candidates are available from previous breeding cycles, it seems rewarding to find subsets for calibration that surpass the prediction accuracy of the original set. The potential to identify such calibration sets is clearly demonstrated by the higher predictive power of the samples composed of the 500 highly predictive genotypes and the lower prediction accuracy of the sample with the 500 poorly predictive genotypes compared with the entire sets from which they were sampled.

This raises the question of how to identify highly predictive genotypes from a large set of potential candidates. The criterion of the average rank of genotypes described above can only be applied after enough genotypes from the prediction set have been genotyped and phenotyped. However, this is quite laborious and expensive, especially if genomic selection is to be used for testcross performance in hybrid breeding. Furthermore, it would considerably increase the duration of breeding cycles and product development.

**Table 2: Prediction accuracies for grain dry matter yield (GDY) and grain dry matter content (GDC)** for genotypes in set S5 obtained with different compositions of the calibration sets (CS).

| | | | | | | | predictive § | |
|---|---|---|---|---|---|---|---|---|
| Trait | CS | N† | all | mean‡ | min‡ | max‡ | - | + |
| | S1 | 928 | 0.47 | 0.48 | 0.30 | 0.57 | 0.17 | 0.57 |
| | S2 | 842 | 0.46 | 0.47 | 0.34 | 0.57 | 0.11 | 0.58 |
| | S3 | 1085 | 0.49 | 0.37 | 0.22 | 0.51 | 0.09 | 0.56 |
| | S4 | 1017 | 0.41 | 0.44 | 0.33 | 0.56 | 0.01 | 0.58 |
| | S1_2 | 1770 | 0.53 | 0.55 | 0.48 | 0.63 | 0.18 | 0.65 |
| | S1_3 | 2013 | 0.53 | 0.55 | 0.45 | 0.62 | 0.09 | 0.67 |
| | S1_4 | 1945 | 0.49 | 0.52 | 0.40 | 0.59 | 0.10 | 0.65 |
| **GDY** | S2_3 | 1927 | 0.53 | 0.47 | 0.33 | 0.58 | 0.19 | 0.60 |
| | S2_4 | 1859 | 0.45 | 0.50 | 0.39 | 0.59 | 0.00 | 0.67 |
| | S3_4 | 2102 | 0.48 | 0.49 | 0.40 | 0.57 | 0.03 | 0.62 |
| | S1_2_3 | 2855 | 0.57 | 0.59 | 0.50 | 0.66 | 0.18 | 0.68 |
| | S1_2_4 | 2787 | 0.52 | 0.56 | 0.49 | 0.64 | 0.15 | 0.68 |
| | S1_3_4 | 3030 | 0.53 | 0.55 | 0.46 | 0.64 | 0.11 | 0.69 |
| | S2_3_4 | 2944 | 0.51 | 0.54 | 0.44 | 0.61 | 0.06 | 0.67 |
| | S1_2_3_4 | 3872 | 0.56 | 0.60 | 0.53 | 0.66 | 0.15 | 0.71 |
| Average | | | 0.50 | 0.51 | 0.40 | 0.60 | 0.32 | 0.45 |
| | S1 | 928 | 0.64 | 0.50 | 0.38 | 0.61 | 0.35 | 0.70 |
| | S2 | 842 | 0.56 | 0.48 | 0.36 | 0.56 | 0.15 | 0.65 |
| | S3 | 1085 | 0.57 | 0.50 | 0.33 | 0.61 | 0.29 | 0.63 |
| | S4 | 1017 | 0.67 | 0.50 | 0.33 | 0.62 | 0.33 | 0.68 |
| | S1_2 | 1770 | 0.67 | 0.59 | 0.52 | 0.67 | 0.43 | 0.74 |
| | S1_3 | 2013 | 0.65 | 0.62 | 0.53 | 0.68 | 0.37 | 0.74 |
| | S1_4 | 1945 | 0.71 | 0.65 | 0.57 | 0.74 | 0.48 | 0.78 |
| **GDC** | S2_3 | 1927 | 0.60 | 0.49 | 0.35 | 0.60 | 0.32 | 0.68 |
| | S2_4 | 1859 | 0.69 | 0.67 | 0.60 | 0.73 | 0.38 | 0.78 |
| | S3_4 | 2102 | 0.68 | 0.64 | 0.54 | 0.70 | 0.38 | 0.75 |
| | S1_2_3 | 2855 | 0.67 | 0.63 | 0.56 | 0.69 | 0.43 | 0.75 |
| | S1_2_4 | 2787 | 0.73 | 0.70 | 0.64 | 0.76 | 0.51 | 0.78 |
| | S1_3_4 | 3030 | 0.70 | 0.71 | 0.65 | 0.75 | 0.44 | 0.80 |
| | S2_3_4 | 2944 | 0.70 | 0.69 | 0.62 | 0.75 | 0.44 | 0.77 |
| | S1_2_3_4 | 3872 | 0.72 | 0.72 | 0.67 | 0.77 | 0.50 | 0.80 |
| Average | | | 0.67 | 0.61 | 0.51 | 0.68 | 0.49 | 0.59 |

† N number of genotypes in the full calibration set (CS)

‡ Mean, maximum and minimum of the prediction accuracies for calibration subsets with samples of size N= 500 taken from the CS, calculated from 1000 repetitions

§ Prediction accuracies of calibration subsets with N = 500 composed of highly predictive (+) or poorly predictive (-) genotypes

**Table 3: Prediction accuracies for grain dry matter yield (GDY) and grain dry matter content (GDC)** for genotypes in set S6 obtained with different compositions of the calibration sets (CS).

| | | | | | | | predictive § | |
|---|---|---|---|---|---|---|---|---|
| Trait | CS | N† | all | mean‡ | min‡ | max‡ | - | + |
| | S1 | 928 | 0.19 | 0.32 | 0.02 | 0.57 | -0.34 | 0.55 |
| | S2 | 842 | 0.03 | 0.15 | -0.09 | 0.40 | -0.38 | 0.43 |
| | S3 | 1085 | 0.35 | 0.23 | -0.03 | 0.55 | -0.10 | 0.57 |
| | S4 | 1017 | 0.21 | 0.17 | -0.13 | 0.48 | -0.38 | 0.61 |
| | S1_2 | 1770 | 0.14 | 0.18 | -0.10 | 0.43 | -0.42 | 0.56 |
| | S1_3 | 2013 | 0.33 | 0.35 | 0.11 | 0.58 | -0.26 | 0.69 |
| | S1_4 | 1945 | 0.23 | 0.35 | 0.10 | 0.57 | -0.38 | 0.65 |
| **GDY** | S2_3 | 1927 | 0.24 | 0.21 | -0.01 | 0.50 | -0.19 | 0.61 |
| | S2_4 | 1859 | 0.14 | 0.16 | -0.09 | 0.36 | -0.45 | 0.60 |
| | S3_4 | 2102 | 0.32 | 0.39 | 0.11 | 0.60 | -0.30 | 0.66 |
| | S1_2_3 | 2855 | 0.26 | 0.28 | 0.05 | 0.48 | -0.29 | 0.68 |
| | S1_2_4 | 2787 | 0.20 | 0.23 | -0.02 | 0.44 | -0.44 | 0.64 |
| | S1_3_4 | 3030 | 0.33 | 0.39 | 0.13 | 0.57 | -0.28 | 0.68 |
| | S2_3_4 | 2944 | 0.25 | 0.31 | 0.11 | 0.51 | -0.31 | 0.64 |
| | S1_2_3_4 | 3872 | 0.28 | 0.30 | 0.09 | 0.49 | -0.35 | 0.69 |
| Average | | | 0.23 | 0.27 | 0.02 | 0.50 | 0.04 | 0.28 |
| | S1 | 928 | 0.57 | 0.53 | 0.35 | 0.69 | 0.14 | 0.74 |
| | S2 | 842 | 0.56 | 0.50 | 0.28 | 0.63 | 0.12 | 0.70 |
| | S3 | 1085 | 0.68 | 0.41 | 0.23 | 0.60 | 0.33 | 0.72 |
| | S4 | 1017 | 0.65 | 0.58 | 0.44 | 0.71 | 0.16 | 0.77 |
| | S1_2 | 1770 | 0.64 | 0.59 | 0.44 | 0.72 | 0.20 | 0.77 |
| | S1_3 | 2013 | 0.70 | 0.63 | 0.49 | 0.74 | 0.37 | 0.79 |
| | S1_4 | 1945 | 0.69 | 0.67 | 0.54 | 0.77 | 0.25 | 0.82 |
| **GDC** | S2_3 | 1927 | 0.68 | 0.40 | 0.23 | 0.54 | 0.40 | 0.76 |
| | S2_4 | 1859 | 0.73 | 0.68 | 0.55 | 0.78 | 0.22 | 0.81 |
| | S3_4 | 2102 | 0.74 | 0.69 | 0.59 | 0.77 | 0.37 | 0.82 |
| | S1_2_3 | 2855 | 0.70 | 0.64 | 0.52 | 0.72 | 0.44 | 0.80 |
| | S1_2_4 | 2787 | 0.74 | 0.72 | 0.61 | 0.79 | 0.34 | 0.83 |
| | S1_3_4 | 3030 | 0.74 | 0.75 | 0.67 | 0.81 | 0.38 | 0.85 |
| | S2_3_4 | 2944 | 0.76 | 0.73 | 0.60 | 0.81 | 0.44 | 0.83 |
| | S1_2_3_4 | 3872 | 0.76 | 0.74 | 0.67 | 0.81 | 0.44 | 0.84 |
| Average | | | 0.69 | 0.62 | 0.48 | 0.73 | 0.50 | 0.63 |

† N number of genotypes in the full calibration set (CS)

‡ Mean, maximum and minimum of the prediction accuracies for calibration subsets with samples of size N= 500 taken from the CS, calculated from 1000 repetitions

§ Prediction accuracies of calibration subsets with N = 500 composed of highly predictive (+) or poorly predictive (-) genotypes

**Table 4: Pairwise correlations between genomic measures** effective sample size ($N_{eff}$), number of polymorphic SNPs shared by the calibration and the prediction set (nPoly), average maximum kinship ($u_{max}$), linkage phase similarity (*LPS*), trait specific reliability ( $^2$), and trait specific prediction accuracy ($r$) for calibration subsets with N=500. In the upper triangle, the values are based on all calibration subsets with S5 as prediction set, in the lower triangle calibration subsets with S6 as prediction set.

| | $N_{eff}$ | nPoly | $u_{max}$ | *LPS* | (GDY) | (GDC) | $r$(GDY) | $r$(GDC) |
|---|---|---|---|---|---|---|---|---|
| $N_{eff}$ | | 0.08 | -0.06 | 0.04 | -0.01 | 0.01 | 0.00 | -0.02 |
| nPoly | -0.01 | | 0.00 | -0.01 | -0.05 | 0.00 | -0.04 | -0.04 |
| $u_{max}$ | 0.03 | 0.02 | | 0.29 | 0.08 | 0.13 | 0.04 | 0.11 |
| *LPS* | 0.37 | -0.12 | 0.31 | | 0.07 | 0.10 | 0.05 | 0.03 |
| (GDY) | 0.01 | 0.08 | 0.10 | 0.04 | | 0.12 | 0.06 | |
| (GDC) | -0.02 | -0.01 | 0.11 | 0.07 | 0.14 | | | -0.01 |
| $r$(GDY) | 0.03 | 0.06 | 0.03 | 0.06 | -0.08 | | | |
| $r$(GDC) | 0.01 | 0.03 | 0.08 | 0.06 | | 0.09 | | |

The genomic parameters $N_{eff}$, nPoly, $u_{max}$, and discussed in the previous paragraph were fairly tightly correlated with the prediction accuracy when values were based on combinations of all 15 calibration sets for S5 and all 31 calibration sets for S6 (see Table 4 in Auinger et al. [2021]). For the sampling procedure, from all calibration sets were sampled. Correlations of previously described variables were calculated for individual samples of size N = 500 as well and averaged over all samples. However, in the combination of calibration set S4 and prediction set S6 with the largest range between the maximum and minimum prediction accuracy for grain dry matter yield (Table 3), no association of the prediction accuracy with a genomic parameter was found, which seems promising as a tool to identify the most predictive training dataset. Averaging these correlations across all 1,000 replications for all combinations of calibration data sets (Table 4) confirmed this conclusion.

One problem in assessing the usefulness of genomic parameters to discover the optimal calibration set is their dependency on the sample size (N) of the calibration set. This dependency is examined in Figure 11, showing the association of various genomic parameters for sets S5 and S6 and the sample size of the 46 combinations of calibration and prediction sets. The association was fairly weak for parameters $N_{eff}$, $u_{max}$, and LPS but quite tight for nPoly and the reliability $^2$ for both traits. For a deeper analysis, I

calculated the genomic parameters for the samples of size N = 500 and averaged the 1,000 repetitions for each of the combinations of full calibration sets. Figure 12 shows scatter plots for the genomic parameters in the full calibration sets versus the average of the samples with N = 500. For $N_{eff}$, $u_{max}$, and LPS, the association was very tight, indicating that these genomic parameters are mainly properties of the population

**Figure 11: Association of sample size N of the calibration sets with various genomic measures** for 46 combinations of sets S1 to S6.

**Figure 12: Genomic parameters in the entire calibration set vs. their mean** across 1,000 samples of size N = 500 taken from the respective calibration sets for all 46 combinations of sets S1 to S6.

**Figure 13: Genomic parameters $u_{max}$, $N_{eff}$, and LPS calculated for the sample sets.** The highly predictive sets consisting of 500 highly predictive genotypes and the poorly predictive sets consisting of 500 poorly predictive genotypes identified amongst 1000 repetitions. As calibration sets were the 15 calibration sets used, which were used for predicting S5. The optimization was done for grain yield and grain dry matter content and for both prediction sets.

and minimally influenced by the size of the calibration set once N is sufficiently large. By comparison, no clear pattern emerged for nPoly and the reliabilities for both grain dry matter yield and grain dry matter content, demonstrating their dependency on N. The dependency of the reliabilities follows directly from the BLUP equations.

Whether $N_{eff}$, $u_{max}$, and LPS could be used to discriminate the calibration sets yielding high or low prediction accuracy for a given N was also investigated. The values of these genomic parameters for each full calibration were calculated in the two highly predictive and poorly predictive calibration subsets with N = 500. For prediction sets S5 and S6 and both grain dry matter yield and grain dry matter content, the genomic parameters obtained for the high and poor samples were plotted. As shown in Figure 13, the values were closely distributed along the bisector, indicating that none of these genomic parameters is promising for identifying calibration subsets.

Hence, further research is needed to identify highly predictive genotypes based on genomic and phenotypic data of earlier breeding cycles and the genomic data of the prediction dataset. I suggest applying a reverse prediction approach in which the original prediction set serves as the calibration set and is used to calculate the reliability for each genotype in the full calibration set using variance components estimated from the latter. The candidates with the highest reliability can be selected to form the subset of highly predictive genotypes. For a proof-of-concept, one must investigate the degree of variation in the prediction reliability and how to choose the reliability criterion threshold to include a genotype in the calibration set.

In summary, this thesis confirmed that, in principle, a set of about 500 carefully chosen genotypes can predict the breeding values of unphenotyped candidates with satisfactory accuracy. This aligns with numerous studies showing diminishing returns from adding further candidates past a certain calibration set size (Albrecht et al. 2014; Pembleton et al. 2018; Brandariz and Bernardo 2019). However, the magnitude of N to reach a plateau seems to depend on the germplasm and the genetic architecture of the trait. In rye and maize (Auinger et al. 2016, 2021), no robust criteria could be found by which to identify, from several successive cycles of commercial breeding programs, smaller subsets of genotypes yielding a consistently higher prediction accuracy across traits and new breeding cycles. Thus, it must be acknowledged that the idea of finding an optimal calibration set for one prediction set and applying it to an independent set has failed. The creation of a calibration set with specific characteristics seems crucial to the quality of the genomic prediction and further research in this direction is of questionable utility.

This thesis focused on the composition of the calibration set. However, the composition of the prediction set is equally important. When estimating the prediction accuracy with the Pearson correlation of the BLUEs and the genomic-predicted values of candidates

in cross-validation or independent validation, the prediction set is implicitly assumed to be a homogeneous population. However, unlike in animal breeding, where all individuals usually originate from the same breed and prediction across breeds is not mixed with prediction within breeds, in plant breeding, the composition of the germplasm is more complex. It is usually composed of groups of related and unrelated full- and half-sib families produced from selected genotypes from previous breeding cycles. Moreover, new genotypes derived from foreign sources such as genetic resources or exotic material are commonly introgressed into elite germplasm to counterbalance the effect of narrowing the genetic base. In such cases, the prediction accuracy can break down from the effect of a small amount of new, unrelated germplasm being added to the material from previous breeding cycles. Identifying such outliers in prediction by calculating their reliability, which is expected to be much lower than that of genotypes descended from recycling breeding with the best lines from previous cycles, is promising. If the germplasm has a strong population structure due to large biparental full-sib or back-cross families, genomic selection might be applied primarily within families to identify the best family members. This is because the genetic variance within families due to Mendelian sampling can only be successfully exploited by genomic selection, not by PBLUP, as all genotypes within families are related to the same degree. In contrast, under an additive genetic model, as applies to *per se* and testcross performance in hybrid breeding, the differences between family means can be predicted fairly reliably from the performance of the parents. Hence, selection among families is mainly a question of the choice of parents, whereas the segregation variance within families can only be exploited with genomic selection.

## 2.7 Conclusions

The results from the studies on rye and maize included in this thesis underline the huge potential of genomic selection to improve the selection progress in breeding in general and in the hybrid breeding of these and other crops in particular. Once genotypic and phenotypic data from multiple breeding cycles have been gathered after implementing genomic selection, these large-scale calibration sets generally yield satisfactory prediction accuracies for new, unphenotyped germplasm. Hence, breeders can shorten breeding cycles and apply a much greater selection intensity by monitoring many candidates and saving on phenotyping. This research also found substantial variation in the prediction accuracy among smaller subsets of about 500 candidates chosen from

many genotypes. The best of these subsets surpassed the prediction accuracy of the full set, composed of several thousands of genotypes. However, the study data did not reveal robust criteria to identify the most predictive genotypes *a priori*, although various genomic parameters were tightly correlated to the prediction accuracy. Hence, based on theoretical results about the relationship between prediction accuracy and sample size (Daetwyler et al. 2008) and the experimental findings of numerous studies, including this one, I recommend including the entire set of genotypes for which genotypic and phenotypic data are available in the calibration set. Unexplored possibilities remain for further research to identify optimized calibration sets, such as the reverse prediction approach I suggested. Moreover, I recommend focusing not only on the composition of the calibration set but also on the composition of the prediction set. There seems to be particular room for improvement by identifying tailor-made calibrations for subsets, such as finding a specific calibration set of close relatives for each larger family in the prediction set, because close kinship is a major driver of prediction accuracy for full-sib and half-sib families (Riedelsheimer et al. 2013, Lehermeier et al. 2014, Jacobson et al. 2014, Brauner et al. 2019). Last but not least, to take full advantage of the tremendous potential of genomic selection, the breeding schemes in hybrid breeding should be redesigned with a systematic interweaving of breeding cycles so that each genotype in the prediction set has a balanced number of close relatives in the calibration set.

# 3 References

Akdemir D, Sanchez JI, Jannink, J-L (2015) Optimization of genomic selection training populations with a genetic algorithm. Genet. Sel. Evol., 47, 38. https://doi.org/10.1186/s12711-015-0116-6

Akdemir D, Rio S, Sánchez IY (2021) Trainsel: an r package for selection of training populations. Front Gene, 12, 607.

Akkaya MA, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics, 132, 1131–1139.

Albrecht T, Wimmer V, Auinger H, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön C-C (2011) Genome-based prediction of testcross values in maize. Theor. Appl. Genet., 123, 339–350. https://doi.org/10.1007/s00122-011-1587-7

Albrecht T, Auinger HJ, Wimmer V, Ogutu JO, Knaak C, Ouzunova M, Piepho HP, Schön CC (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor. Appl. Genet., 127(6), 1375–1386.

Andreescu C, Avendano S, Brown SR, Hassen A, Lamont SJ, Dekkers JC (2007) Linkage disequilibrium in related breeding lines of chickens. Genetics, 177, 2161–2169. https://doi.org/10.1534/genetics.107.082206

Atanda SA, Olsen M, Burgueño J, Crossa J, Dzidzienyo D, Beyene Y, Gowda M, Dreher K, Zhang X, Prasanna BM, Tongoona P, Danquah EX, Olaoye G, Robbins KR (2021) Maximizing efficiency of genomic selection in cimmyt's tropical maize breeding program. Theor. Appl. Genet., 134, 279–294. https://doi.org/10.1007/s00122-020-03696-9

Auinger H-J, Schönleben M, Lehermeier C, Schmidt M, Korzun V, Geiger HH, Piepho HP, Gordillo A, Wilde P, Bauer E, Schön C-C (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). Theor. Appl. Genet., 129, 2043–2053. https://doi.org/10.1007/s00122-016-2756-5

Auinger H-J, Lehermeier C, Gianola D, Mayer M, Melchinger AE, da Silva S, Knaak C, Ouzunova M, Schön C-C (2021) Calibration and validation of predicted genomic breeding values in an advanced cycle maize population. Theor. Appl. Genet., 134(9), 3069–3081.

Barcaccia G, Lucchin M, Parrini P (2003) Characterization of a flint maize (*Zea mays* var. *indurata*) Italian landrace, II. Genetic diversity and relatedness assessed by SSR and Inter-SSR molecular markers. Genet. Resour. Crop Evol., 50, 253–271.

Bauer E, Schmutzer T, Barilar I, Mascher M, Gundlach H, Martis MM, Twardziok, SO, Hackauf B, Gordillo GA, Wilde P, Schmidt M, Korzun V, Mayer KFX, Schmid K, Schön C-C, Scholz U (2017) Towards a whole-genome sequence for rye (*Secale cereale* L.). The Plant Journal, 89 (5), 853–869.

Beadle GW (1939) Teosinte and the origin of maize. J Hered 30: 245–247.

Becker H (2019) Pflanzenzüchtung. utb GmbH.

Behre K-E (1992) The history of rye cultivation in Europe. Veget Hist Archaeobot, 1(3), 141–56. http://www.jstor.org/stable/23416112.

Bernardo R (1994) Prediction of maize single cross performance using RFLPs and information from related hybrids. Crop Science, 34(1), 20–25. https://doi.org/10.2135/cropsci1994.0011183X003400010003x

de Bern Oliveira I, Amadeu RR, Ferrão, LFV, Muñoz, PR (2020) Optimizing whole-genomic prediction for autotetraploid blueberry breeding. Heredity, 125, 437–448. https://doi.org/10.1038/s41437-020-00357-x

Bernal-Vasquez AM, Gordillo A, Schmidt M, Piepho HP (2017) Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. BMC Genetics, 18(1), 1–17.

Bernardo R (2020a) Breeding for quantitative traits in plants. 3rd edition, Stemma Press, Woodbury.

Bernardo R (2020b) Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. Heredity, 125(6), 375–385.

Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jainj N, Singh PK, Singh GP, Prabhu KV (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Front Gene, 7, 221.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet., 32, 314–331.

Bouchez A, Hospita, F, Causse M, Gallais A, Charcosset A (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics, 162(4), 1945-1959.

Brandariz SP, Bernardo R (2018) Maintaining the accuracy of genomewide predictions when selection has occurred in the training population. Crop Science, 58, 1226–1231. https://doi.org/10.2135/cropsci2017.11.0682

Brandariz SP, Bernardo R (2019) Small ad hoc versus large general training populations for genomewide selection in maize biparental crosses. Theor. Appl. Genet., 132(2), 347–353.

Brauner PC, Schipprack W, Utz HF, Bauer E, Mayer M, Schön C-C, Melchinger AE (2019) Testcross performance of doubled haploid lines from European flint maize landraces is promising for broadening the genetic base of elite germplasm. Theor. Appl. Genet., 132(6), 1897–1908.

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 Science, 52(2), 707–719.

de los Campos G, Gianola D, Rosa GJ (2009) Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. Journal of Animal Science, 87(6), 1883–1887.

de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, 193(2), 327–345.

Dreisigacker S, Crossa J, Pérez-Rodríguez P, Montesinos-López OA, Rosyara U, Juliana P, Mondal S, Crespo-Herrera L, Govindan V, Singh PR, Braun, H. J. (2021). Implementation of genomic selection in the CIMMYT global wheat program, findings from the past 10 years. Crop Breeding, Genetics and Genomics, *3*(2)

Charmet G, Tran LG, Auzanneau J, Rincent R, Bouchet S (2020) BWGS: AR package for genomic selection and its application to a wheat breeding programme. PLoS One, 15(4), e0222733.

Clark SA, Hickey JM, Daetwyler HD, and van der Werf JH (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Gen Sel Evol, 44,4. https://doi.org/10.1186/1297-9686-44-4

Combs E, Bernardo R (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. The Plant Genome, 6(1), 1-7. https://doi.org/10.3835/plantgenome2012.11.0030

Crossa J, Perez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang S, Dreisigacker S, Babu R, Li Y, Bonnet D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112(1): 48–60.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci, 22(11), 961–975.

Crow JF (1998) 90 years ago: the beginning of hybrid maize. Genetics, 148, 923–928.

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PloS One, 3(10), e3395.

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed Genet., 124(6), 331–341. https://10.1111/j.1439-0388.2007.00701.x.PMID: 18076470.

Destatis (2021) Statistisches Bundesamt. Retrieved November 2, 2021, from https://www-genesis.destatis.de/genesis/online?operation=abruftabelleBearbeiten&levelindex=1&levelid=1649104958871&auswahloperation=abruftabelleAuspraegungAuswaehlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&code=41141-0001&auswahltext=&werteabruf=Werteabruf#abreadcrumb

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theor. Appl. Genet., 85, 673–680. https://doi.org/10.1007/BF00225004

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica, 51, 281–291.

Edwards SM, Buntjer JB, Jackson R, Bentley AR, Lage J, Byrne E, Burt C, Jack P, Berry S, Flatman E, Poupard B, Smith S, Hayes C, Gaynor RC, Gorjanc G, Howell P, Ober E, Mackay IJ, Hickey JM (2019) The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Gen 132(7):1943–1952.

Edwardson JR (1970) Cytoplasmic male sterility. The Botanical Review, 36(4), 341–420.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One, 6(5), e19379.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. 4th Edition, Addison Wesley Longman, Harlow.

FAOSTAT (2021) Statistics database of the Food and Agriculture Organisation of the United Nations (FAO), Data, Crops and livestock products. Retrieved September 21, 2021, from https://www.fao.org/faostat/en/#data/QCL

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edin., 53, 399–433.

Gage JL, Monier B, Giri A, Buckler ES (2020) Ten years of the maize nested association mapping population: impact, limitations, and future directions. The Plant Cell, 32(7), 2083–2093.

Galán RJ., Bernal-Vasquez AM, Jebsen C, Piepho HP, Thorwarth P, Steffan P, Gordillo A, Miedaner T (2021) Early prediction of biomass in hybrid rye based on hyperspectral data surpasses genomic predictability in less-related breeding material. Theor Appl Genet 134, 1409–1422. https://doi.org/10.1007/s00122-021-03779-1

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, et al. (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PloS One, 6(12), e28334.

Geiger HH, Miedaner T (2009) Rye breeding. Cereals, 3, 157–181.

Geiger HH, Schnell FW (1970) Cytoplasmic male sterility in rye (*Secale cereale* L.) Crop Science, 10(5), 590–593.

Gianola D, de Los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. Genetics, 183(1), 347–363.

Goddard ME (2008) The use of high density genotyping in animal health. In Animal Genomics for Animal Health (Vol. 132, pp. 383–389). Karger Publishers.

Goodman MM, Walton CG (1988) The history and evolution of maize. Critical Reviews in Plant Sciences, 7(3), 197–220. 10.1080/07352688809382264

Guo T, Yu X, Li X, Zhang H, Zhu C, Flint-Garcia S, McMullen MD, Holland JB, Szalma SJ, Wisser RJ, Yu J (2019) Optimal designs for genomic selection in hybrid crops. Mol. Plant, 12, 390–401. https://doi.org/10.1016/j.molp.2018.12.022

Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics, 177, 2389–2397. https://doi.org/10.1534/genetics.107.081190

Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution, 42(1), 1–12.

Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics, 194(3), 597–607.

Haile TA, Walkowiak S, N'Diaye A, Clarke JM, Hucl PJ, Cuthbert RD, Knox RE, Pozniak CJ (2021) Genomic prediction of agronomic traits in wheat using different models and cross-validation designs. Theor Appl Genet 134**,** 381–398. https://doi.org/10.1007/s00122-020-03703-z

Hayes BJ, Goddard ME (2008) Prediction of breeding values using marker-derived relationship matrices. Journal of Animal Science, 86(9), 2089–2092.

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research, 91(1), 47–60.

He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y (2016) Genomic selection in a commercial winter wheat population. Theor. Appl. Genet., 129, 641–651. https://doi.org/10.1007/s00122-015-2655-1

Heffner EL, Jannink JL, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. The Plant Genome 4(1), 65–75.

Hepting L (1978) Analyse eines 7 X 7 Sortendiallels zur Ermittlung geeigneten Ausgangsmaterials für die Hybridzüchtung bei Roggen. Z Pflanzenzucht.

Henderson CR (1963) Selection index and expected genetic advance. In: Statistical genetics and plant breeding. Nat Acad Sci — Nat Res Counc Publ No 982, Washington DC, pp 141–163

Henderson CR (1974) General flexibility of linear model techniques for sire evaluation. J. Dairy Sci., 57 (8), 963–972. https://doi.org/10.3168/jds.S0022-0302(74)84993-3

Henderson, CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics, 31, 423–447.

Heslot N, Yang HP, Sorrells ME, Jannink JL (2012) Genomic selection in plant breeding: a comparison of models. Crop Science, 52(1), 146–160.

Hillman G (1978) On the origins of domestic rye—*Secale cereale*: the finds from aceramic Can Hasan III in Turkey. Anatolian Studies, 28, 157–174.

Hoerl AE, Kennard RW (1970) Ridge regression: applications to nonorthogonal problems. Technometrics, 12(1), 69–82.

Hofheinz N, Borchardt D, Weissleder K, Frisch M (2012) Genome-based prediction of test cross performance in two subsequent breeding cycles. Theor. Appl. Genet., 125(8), 1639–1645.

Immonen S, Tenhola-Roininen T. (2003) Protocol for rye anther culture. In: Maluszynski M, Kasha KJ, Forster BP, Szarejko I (eds) Doubled haploid production in crop plants: a manual. Kluwer Academic Publishers, Dordrecht, pp 141–150.

Isidro J, Jannink J-L, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. Theor. Appl. Genet., 128, 145–158. https://doi.org/10.1007/s00122-014-2418-4

Jacobson A, Lian L, Zhong S, Bernardo R (2014) General combining ability model for genomewide selection in a biparental cross. Crop Sci., 54, 895–905.

Jäger EJ, Neumann S, Ohmann E (2003) Botanik. Spektrum Akademischer Verlag, Heidelberg.

Jan HU, Abbadi A, Lücke S, Nichols RA, Snowdon RJ (2016) Genomic prediction of testcross performance in canola (*Brassica napus*). PLoS One, 11(1), e0147769.

Jordan SA, Humphries P (1994) Single nucleotide polymorphism in exon 2 of the BCP gene on 7q31-q35. Human Molecular Genetics, 3(10), 1915.

Jung S, Pausch H, Langenmayer MC, Schwarzenbacher H Majzoub-Altweck M, Gollnick NS, Fries R (2014) A nonsense mutation in PLD4 is associated with a zinc deficiency-like syndrome in Fleckvieh cattle. BMC Genomics, 15, 623. https://doi.org/10.1186/1471-2164-15-623

Kingsbury N (2011) Hybrid: the history and science of plant breeding. University of Chicago Press.

Kipp S, Segelke D, Schierenbeck S, Reinhardt F, Reents R, Wurmser C, Fries R, Thaller G, Tetens J, Pott J, Haas D, Raddatz BB, Hewicker-Trautwein M, Proios I, Schmicke M, Grünberg W (2016) Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. J. Dairy Sci., 99(11), 8915–8931.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124(3), 743–756.

Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N, Moreau L, Moreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack W, Schönleben M, Walter H, Charcosset A, Schön C-C (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. Genetics, 198(1), 3–16.

Lian L, Jacobson A, Zhong S, Bernardo R (2014) Genomewide prediction accuracy within 969 maize biparental populations. Crop Science, 54(4), 1514–1522.

Longin CFH, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor. Appl. Genet., 128(7), 1297–1306.

Lopez-Cruz M, de los Campos G (2021) Optimal breeding-value prediction using a sparse selection index. Genetics, 218(1), https://doi.org/10.1093/genetics/iyab030.

Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3: Genes, Genomes, Genetics, 3(3), 481–491.

Lorenz A, Smith KP (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55(6):2657. https://doi.org/10.2135/cropsci2014.12.0827

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet., 120(1), 151–161.

Lund MS, Su G, Janss L, Guldbrandtsen B, Brøndum RF (2014) Genomic evaluation of cattle in a multi-breed context. Livestock Science, 166, 101–110.

Lundqvist A (1956) Self incompatibility in rye: I. Genetic control in the diploid. Hereditas, 42(3 4), 293–348.

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits, Sinauer Associates, Sunderland, Massachusetts.

Maenhout S, De Baets B, Haesaert G, van Bockstaele E (2007) Support vector machine regression for the prediction of maize hybrid performance. Theor. Appl. Genet., 115(7), 1003–1013.

Malécot G (1948) The mathematics of heredity. Masson, Paris.

Mangin B, Rincent R, Rabier CE, Moreau L, Goudemand-Dugue E (2019) Training set optimization of genomic prediction by means of EthAcc. PLoS One, 14(2), e0205629.

Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrána J, Kubaláková M, König S, Kugler KG, Scholz U, Hackauf B, Korzun V, Schön C-C, Doležel J, Bauer E, Mayer KFX, Stein N (2013) Reticulate evolution of the rye genome. Plant Cell, 25, 3685–3698.

Marulanda JJ, Melchinger AE, Würschum T (2015) Genomic selection in biparental populations: assessment of parameters for optimum estimation set design. Plant Breeding, 134(6), 623–630.

Melchinger AE, Utz HF, Schön C-C (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics, 149(1), 383–403.

Meuwissen T, Hayes B, and Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157, 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Meuwissen T, Hayes B, Goddard M (2016) Genomic selection: A paradigm shift in animal breeding. Animal Frontiers, 6(1), 6–14.

Michel S, Ametz C, Gungor H, Epure D, Grausgruber H, Löschenberger F, Buerstmayr H (2016) Genomic selection across multiple breeding cycles in applied bread wheat breeding. Theor. Appl. Genet., 129(6), 1179–1189.

Michel S, Ametz C, Gungor H, Akgöl B, Epure D, Grausgruber H, Löschenberger F, Buerstmayr H (2017) Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. Theor. Appl. Genet., 130(2), 363–376.

Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, Kilian B (2014) Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. PLoS One, 9, e85761

Miedaner T, Hübner M, Korzun V, Schmiedchen B, Bauer E, Haseneyer G, Wilde P, Reif JC (2012) Genetic architecture of complex agronomic traits examined in two testcross populations of rye (Secale cereale L.). BMC Genomics, 13(1), 1–13.

Miedaner T (2013) Roggenanbau: Eine erfolgreiche Alternative. DLG-Verlag GmbH, AgrarPraxis kompakt.

Mir C, Zerjal T, Combes V, Dumas F, Madur D, Bedoya C, Dreisigacker S, Franco J, Grudloyma P, Hao PX, Hearne S, Jampatong C, Laloë D, Muthamia Z, Nguyen T, Prasanna BM, Taba S, Xie CX, Yunus M, Zhang S, Warburton ML, Charcosset A (2013) Out of America: tracing the genetic footprints of the global diffusion of maize. Theor. Appl. Genet., 126, 2671–2682.

Molenaar WS, Melchinger AE, Production of doubled haploid lines for hybrid breeding in maize (2019) Frank Ordon and Wolfgang Friedt (eds), Advances in breeding techniques for cereal crops, Burleigh Dodds Science Publishing, Cambridge, UK. ISBN: 978-1-78676-244-3

Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, Barrón-López JA, Martini JW, Fajardo-Flores SB, Gaytan-Lugo LS, Santana-Manchilla PC, Crossa J (2021). A review of deep learning applications for genomic selection. BMC Genomics, 22(1), 1–23.

Morota G, Gianola D (2014) Kernel-based whole-genome prediction of complex traits: a review. Front Gene, 5, 363.

Müller D, Schopp P, Melchinger AE (2017) Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. G3: Genes, Genomes, Genetics, 7(3), 801–811.

Murat F, Pont C, Salse J (2014) Paleogenomics in *Triticeae* for translational research. Curr. Plant Biol., 1, 34–39.

Norman A, Taylor J, Edwards J, Kuchel H (2018) Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. G3: Genes, Genomes, Genetics, 8(9), 2889–2899. https://doi.org/10.1534/g3.118.200311

Olatoye MO, Clark LV, Labonte NR, Dong H, Dwiyanti MS, Anzoua KG, Brummer JE, Ghimire BK, Dzyubenko E, Dzyubenko N, Bagmet L, Sabitov A, Chebukin P, Głowacka K, Heo K, Jin X, Nagano H, Peng J, Yu CY, Yoo JH, Zhao H, Long SP, Yamada T, Sacks EJ, Lipka AE (2020) Training population optimization for genomic selection in miscanthus. G3: Genes, Genomes, Genetics, 10, 2465–2476. https://doi.org/10.1534/g3.120.401402

Ould Estaghvirou SB, Ogutu JO, Schulz-Streeck T, Knaak C, Ouzunova M, Gordillo A, Piepho HP (2013) Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. BMC Genomics, 14(1), 1–21.

Openshaw S, Frascaroli E (1997) QTL detection and marker-assisted selection for complex traits in maize. In: ASTA (ed) Proc 52nd Annu Corn Sorghum Res Conf. ASTA (American Seed Trade Association), Washington D.C., pp 44–53.

Pausch H, Venhoranta H, Wurmser C. Hakala K, Iso-Touru T, Sironen A, Vingborg RK, Lohi H, Söderquist L, Fries R, Andersson M (2016) A frameshift mutation in ARMC3 is associated with a tail stump sperm defect in Swedish Red (*Bos taurus*) cattle. BMC Genetics, 17, 49. https://doi.org/10.1186/s12863-016-0356-7

Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. Crop Science, 49(4), 1165–1176.

Piepho HP, Möhring J, Schulz Streeck T, Ogutu JO (2012) A stage wise approach for the analysis of multi environment trials. Biometrical Journal, 54(6), 844–860.

Piepho HP, Laidig F, Drobek T, Meyer U (2014) Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials. Theor. Appl. Genet., 127(5), 1009–1018.

Poland JA, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S; Crossa J; Sánchez-Villeda H; Sorrells M; Jannink J-L (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome, 5(3), 103–113.

Pembleton LW, Inch C, Baillie RC, Drayton MC, Thakur P, Ogaji YO, Spangenberg GC, Forster JW, Daetwyler HD, Cogan NOI (2018) Exploitation of data from breeding programs supports rapid implementation of genomic selection for key agronomic traits in perennial ryegrass. Theor. Appl. Genet., 131(9), 1891–1902.

Prigge V, Sánchez C, Dhillon BS, Schipprack W, Araus JL, Bänziger M, Melchinger AE (2011) Doubled haploids in tropical maize: I. Effects of inducers and source germplasm on in vivo haploid induction rates. Crop Science, 51(4), 1498–1506.

Reif JC, Gumpert FM, Fischer S, Melchinger AE (2007) Impact of interpopulation divergence on additive and dominance variance in hybrid populations. Genetics, 176(3), 1931–1934.

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics, 44, 217–220. https://doi.org/10.1038/ng.1033

Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor. Appl. Genet., 126(11), 2835–2848.

Rimpau W (1883) Züchtung auf dem Gebiete der landwirtschaftlichen Kulturpflanzen Mentzel & von Lengerke's Landwirtschaftl. Kalender, 2, 33–92.

Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodríguez VM, Moreno-Gonzalez J, Melchinger AE, Bauer E, Schön C-C, Meyer N, Giauffret C, Bauland C, Jamin P, Laborde J, Monod H, Flament P, Charcosset A, Moreau L (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics, 192, 715–728. https://doi.org/10.1534/genetics.112.141473

Saatchi M, McClure MC, McKay SD, Rolf MM, Kim JW, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel RG, Garrick DJ, Taylor JF. (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genetics Selection Evolution, 43, 1–16. https://doi.org/10.1186/1297-9686-43-40

Sallam AH, Endelman JB, Jannink JL, Smith KP (2015) Assessing genomic selection prediction accuracy in a dynamic barley breeding population. Plant Genome 8:1. https://10.3835/plantgenome2014.05.0020.

Sanchez AC, Khush GS (2000) Chromosomal localization of five mutant genes in rice, Oryza sativa, using primary trisomics. Plant Breeding, 119(1), 84–86.

Schmidt W (2004) Hybrid maize breeding at KWS Saat AG. Tagungsband der 54. Jahrestagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs: Hybridmais, Züchtung und Verwertung, 25–27 November 2003, Gumpenstein, Irdning, Austria, 1–6.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock S.M, Belter E, Du F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science, 326, 1112–1115.

Schön C-C, Simianer H (2015) Resemblance between two relatives–animal and plant breeding. Journal of Animal Breeding and Genetics, 132(1), 1–2.

Schopp P, Riedelsheimer C, Utz HF, Schön CC, Melchinger AE (2015) Forecasting the accuracy of genomic prediction with different selection targets in the training and prediction set as well as truncation selection. Theor. Appl. Genet., 128(11), 2189–2201.

Schopp P, Müller D, Technow F, Melchinger AE (2017) Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. Genetics, 205(1), 441–454.

Schrag TA, Frish M, Dhillon B, Melchinger AE (2009) Marker-based prediction of hybrid performance in maize single-crosses involving doubled haploids. Maydica. 54, 353. https://doi.org/10.1007/s00122-008-0934-9

Schrag TA, Schipprack W, Melchinger AE (2019) Across-years prediction of hybrid performance in maize using genomics. Theor. Appl. Genet., 132(4), 933–946.

Schulz-Streeck T, Ogutu JO, Piepho HP (2013) Comparisons of single-stage and two-stage approaches to genomic selection. Theor. Appl. Genet., 126(1), 69–82.

Scutari M, Mackay I, Balding D (2016) Using genetic distance to infer the accuracy of genomic prediction. PLoS Genetics, 12(9), e1006288.

Seye A, Bauland C, Charcosset A, Moreau L (2020) Revisiting hybrid breeding designs using genomic predictions: simulations highlight the superiority of incomplete factorials

between segregating families over topcross designs. Theor. Appl. Genet., 133, 1995–2010. https://doi.org/10.1007/s00122-020-03573-5

Shull GH (1908) The composition of a field of maize. Journal of Heredity, (1), 296–301.

Shull GH (1909) A pure-line method in corn breeding. Journal of Heredity, (1), 51–58.

Shull GH (1914) Duplicate genes for capsule-form in *Bursa bursa-pastoris*. Z. Vererbungslehre, 12, 97–149. https://doi.org/10.1007/BF01837282

Storlie E, Charmet G (2013) Genomic selection accuracy using historical data generated in a wheat breeding program. Plant Genome 6:1–9. https://10.3835/plantgenome2013.01.0001

Strube HG (1967): Merkmalskorrelationen bei Hybridmais und Ihre Bedeutung für die Selektion. Diss. Univ. Hohenheim.

Sweeney DW, Rutkoski J, Bergstrom GC, Sorrells ME (2020) A connected half sib family training population for genomic prediction in barley. Crop Science, 60(1), 262–281.

Tenaillon MI, Charcosset AA (2011) European perspective on maize history. Comptes Rendus Biologies, 334(3), 221–228.

Utz HF, Melchinger AE, Schön C-C (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics, 154(4), 1839–1849.

Visscher P, Hill W, Wray N (2008) Heritability in the genomics era—concepts and misconceptions. Nat. Rev. Genet., 9, 255–266. https://doi.org/10.1038/nrg2322

Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M, Friters A, Pot J, Paleman J Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 23(21), 4407–4414.

Voss-Fels KP, Stahl A, Wittkop B. Lichthardt C, Nagler S, Rose T. Chen T. W, Zetzsche H, Seddig S, Baig MM, Ballvora A, Frisch M. Ross E, Hayes BJ, Hayden MJ, Leon J, Kage H, Friedt W, Stützel H, Snowdon RJ (2019) Breeding improves wheat productivity under contrasting agrochemical input levels. Nature Plants, 5, 706–714. https://doi.org/10.1038/s41477-019-0445-5

Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics, 15(1), 1–12.

Weir BS, Hill WG (1980) Effect of mating structure on variation in linkage disequilibrium. Genetics, 95(2), 477–488.

Weller JI, Ezra E, Ron M (2017) Invited review: A perspective on the future of genomic selection in dairy cattle. J. Dairy Sci., 100(11), 8633–8644.

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3: Genes, Genomes, Genetics, 2, 1427–1436. https://doi.org/10.1534/g3.112.003699

Wricke G (1969) Untersuchungen zur Vererbung der Selbstfertilität beim Roggen (*Secale cereale*). Theor. Appl. Genet., 39(8), 371–378.

Wright S (1920) The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. Proc Natl Acad Sci USA 6(6):320–332.

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor. Appl. Genet., 125, 201–210. https://doi.org/10.1007/s00122-012-1887-6

Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. BMC Genetics, 14(1), 1–8.

Yousef GG, Juvik JA (2001) Comparison of phenotypic and marker assisted selection for quantitative traits in sweet corn. Crop Science, 41(3), 645–655.

Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, Li J, Simianer H (2014) Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PloS One, 9(3), e93017.

Zhang H, Yin L, Wang M, Yuan X, Liu X (2019) Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. Front Gene, 10, 189.

Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics, 182(1), 355–364.

# 4 Publications

This section includes the three publications underlying this thesis. The three publications, including supplementary material, can be accessed via the following links:

**Albrecht at al. 2014**

https://doi.org/10.1007/s00122-014-2305-z

**Auinger et al. 2016**

https://doi.org/10.1007/s00122-016-2756-5

**Auinger et al. 2021**

https://doi.org/10.1007/s00122-021-03880-5

# 5 Acknowledgements

I would like to express the greatest gratitude to my supervisor, Prof. Chris Schön. As I started at her side, I thought I was merely doing a job. She convinced me to seek a PhD, and without her support and encouragement, I might not have finished this thesis. She also provided the opportunity to work on my own ideas. So, altogether, the suggestion to achieve a PhD was right. Prof. Chris Schön, for all of your time, assistance, and patience: Again, thanks a lot!

I would also like to thank Prof. Albrecht Melchinger for his support in the last phase of my PhD. It was a great help, and I hope we can successfully finish our current projects and maybe start new ones. Among my senior colleagues, I thank Prof. Leo Dempfle, who inspired me to take up the scientific field of breeding and who is always available for questions. Prof. Daniel Gianola also must be mentioned for sharing his idea of calculating a genomic measure based on the variation in the relatedness of candidates of interest, which is included among the papers of this thesis.

Thanks also go to Prof. Donna Ankerst and Prof. Henner Simianer for serving on my graduate committee. I would like to thank all of the coauthors of the papers for their good collaboration. Namely, the team of KWS for providing much data and helping me understand the constraints and limitations of a real-life breeding program. I also acknowledge everyone who was involved in organizing the projects and participated in various meetings and conferences where I had many fruitful discussions. Of course, I would like to thank the German Federal Ministry of Education and Research (BMBF) for funding the AgroClustEr Synbreed—Synergistic plant and animal breeding (Grant 0315528) and RYESELECT (Grant ID 0315946) projects.

Special thanks to all of the colleagues of the plant breeding chair, present and past, whether in the office, the laboratory, or the field. It was a pleasure to meet you, and I think we had good exchanges. I cannot think of anyone I didn’t enjoy meeting during my time in the scientific world. Thank you all for that.

Last but not least: Ein herzlicher Dank an meine Familie und meine Freunde, die meist Verständnis aufgebracht haben, wenn in der Zeit etwas liegen geblieben ist oder nicht möglich war. Ein spezieller Dank geht auch an meine Eltern, die mir meine Ausbildung ermöglicht haben.

I hope no one is disappointed not to be mentioned by name, but you know me and that I would certainly have forgotten someone if I had started listing people individually.