

Potentials and limits of genomics-assisted wheat breeding

**Dissertation
zur Erlangung des
Doktorgrades der Agrarwissenschaften (Dr. agr.)**

der

Naturwissenschaftlichen Fakultät III
Agrar- und Ernährungswissenschaften,
Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Herrn He, Sang
Geb. am 25.06.1989 in Shehong, Sichuan, China

1. Gutachter: Prof. Dr. Jochen Reif
2. Gutachter: Prof. Dr. Tobias Würschum

Verteidigung am 22.05.2017

Table of Contents

General Introduction.....	4
Classical wheat breeding methods.....	4
Breeding goals for wheat.....	5
Marker-assisted selection	7
Genomic selection	8
Genetic architecture of grain yield	9
Genetic architecture of resistances to FHB and STB	10
Marker imputation for an economic fingerprinting in marker-assisted and genomic selection	10
Optimum design of a training population in genomic selection.....	11
Prediction accuracy of individual genotypes.....	12
Objectives of this thesis.....	13
Original papers	14
Prospects and limits of marker imputation in quantitative genetic studies in European elite wheat (<i>Triticum aestivum</i> L.).....	14
Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast Central European elite winter wheat population	15
Genomic selection in a commercial winter wheat population.....	16
Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe	17
General discussion and conclusion.....	18
Imputation of missing values paved the way to integrate heterogeneous genomic data across different genotyping platforms	18
Strategy to integrate phenotypic data across different selection cycles	20
Potential and limits of marker-assisted and genomic selection for complex traits.....	22
Choice of biometric model for genomic selection.....	23
Optimum composition of a training population for genome-wide predictions	24
Prediction accuracy is highly variable among genotypes but can be assessed with genomic data ...	25
Implementation of genomic selection in breeding programs using historical data	26
Summary	28
Zusammenfassung	30
Reference.....	32
Abbreviations	39
Acknowledgements	40
Curriculum vitae.....	41
Eidesstattliche Erklärung / <i>Declaration under Oath</i>	42

Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren / *Declaration concerning Criminal Record and Pending Investigations* 43

General Introduction

Wheat (*T. aestivum* L.) is one of the most important crops, providing 20% of the total calories for the world's population (Bushuk 1997; Kumar et al. 2011). In contrast to maize and rice, wheat provides higher protein content and serves as the major ingredient producing a wide variety of food such as bread, biscuit, and cakes (Johnson and Lay 1974). Worldwide wheat production needs to be doubled to feed an estimated world population of 9 billion by 2050 (Godfray et al. 2010; Ray et al. 2013). Nevertheless, it is becoming increasingly difficult to satisfy this rising global demand because the arable land and water are increasingly becoming scarce, average living standards are rising, and investments in increasing agricultural productivity are growing slowly (Fischer et al. 2014; Laidig et al. 2014). Wheat breeding is one viable and sustainable solution to increase grain yield and to improve yield stability.

Classical wheat breeding methods

Wheat is a self-pollinated plant (Wright 1980). Although major efforts have been put in initiating hybrid breeding, most programs are still devoted to inbred line breeding (Kempe and Gils 2011; Mette et al. 2015; Whitford et al. 2013). Becker (2011) suggested classifying the available line breeding approaches into pedigree and bulk methods. Both methods start with the generation of variation by crossing diverse parents. In the pedigree method, selection occurs then within each subsequent generation and inferior candidates are discarded. In this approach the relatedness plays an important role because breeders are able to trace lines from superior families through pedigree. In contrast, only natural selection occurs during early stages using the bulk method until a minimum level of homozygosity is reached. Then, high selection intensity is applied to identify superior varieties. Two popular modifications of the bulk method have been proposed which are often used in wheat breeding. In the single-seed descend method (SSD), the population is not handled as a bulk but F₂ plants are selfed individually, avoiding presence of closely related sister lines in later stages of selection. The SSD method can be implemented in green houses, and allows off-season selection; thus, accelerating the breeding cycle by one year for winter wheat (Knott et al. 2008; Malla et al. 2010; Tee and Qualset 1975). In the double haploid (DH) method, inbred lines are generated in a single step using haploid

tissues. The DH method can shorten the breeding cycle by an additional year in comparison to SSD (Inagaki et al. 1998; Thomas et al. 2003).

All breeding methods are implemented applying multi-stage selection. This includes that breeding starts with traits exhibiting a high plot-based heritability moving forward to more complex traits in later stages of selection (Becker 2011). Moreover, in the course of selection, number of test candidates decreases and phenotyping intensity and consequently heritability increases (Fig. 1; He et al. 2016b; Longin et al. 2015; Würschum 2012).

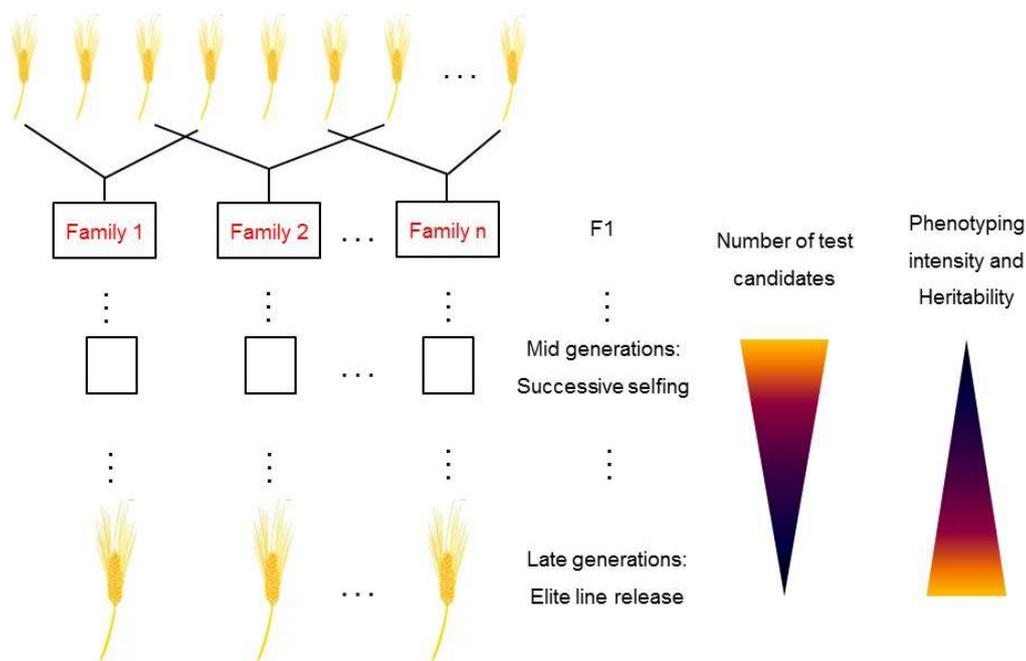


Figure 1 Multi-stage selection of line breeding

Breeding goals for wheat

There are several traits playing an important role for wheat breeding, which can be classified into: (1) yield and yield components, (2) biotic and abiotic stress resistance, and (3) quality traits. Corresponding plot-based heritabilities of relevant traits specifically for European wheat populations are listed in Table 1. Among the long list of relevant traits, grain yield as well as resistance against *Fusarium head blight* (FHB) and *Septoria tritici blotch* (STB) are major issues for wheat breeding in Central Europe. FHB, caused by *Fusarium graminearum*, and STB caused by *Mycosphaerella graminicola* entail a reduction of grain yield and quality (Buerstmayr et al. 2009; Miedaner et al.

2013). In order to control these diseases, fungicides with similar active ingredients or similar mode of action have been widely applied resulting in mutations/selection of pathogens showing resistance against fungicides. In this context, breeding for resistance becomes the most sustainable and efficient way of control (Miedaner et al. 2013; Mirdita et al. 2015).

Table 1 Plot-based heritabilities of relevant traits estimated in wheat populations adapted to Europe

Group	Trait	Heritability	Literature
Yield and yield components	Grain yield	0.33	Reif et al. (2011b)
		0.20	Longin et al. (2013)
		0.20	He et al. (2016b)
	Thousand-kernel weight	0.69	Reif et al. (2011a)
Biotic and abiotic stress resistance	Fusarium	0.62	Liu et al. (2016)
		0.58	Miedaner et al. (2011)
		0.57	Jiang et al. (2015)
		0.30	Mirdita et al. (2015)
		0.37	Miedaner et al. (2001)
	Septoria	0.54	Miedaner et al. (2013)
		0.18	Longin and Würschum (2014)
Quality traits	Frost tolerance	0.14	Mirdita et al. (2015)
		0.72	Zhao et al. (2013a)
		0.72	Longin et al. (2013)
	Protein content	0.58	Reif et al. (2011a)
		0.43	Longin et al. (2013)
		0.47	Liu et al. (2016)
Sedimentation volume	0.64	Longin et al. (2013)	
	0.89	Reif et al. (2011a)	

A recent study based on official variety test in Germany revealed a 0.89% increase in grain yield (Laidig et al. 2014). This relatively small increase in grain yield would hardly cope with the

demand of 100%-110% yield improvement and would not be enough to feed 9 billion people by 2050 (Fischer et al. 2014; Godfray et al. 2010; Tilman et al. 2011). Hence, the advent of more efficient breeding tools is imperative, boosting up the selection gain in wheat.

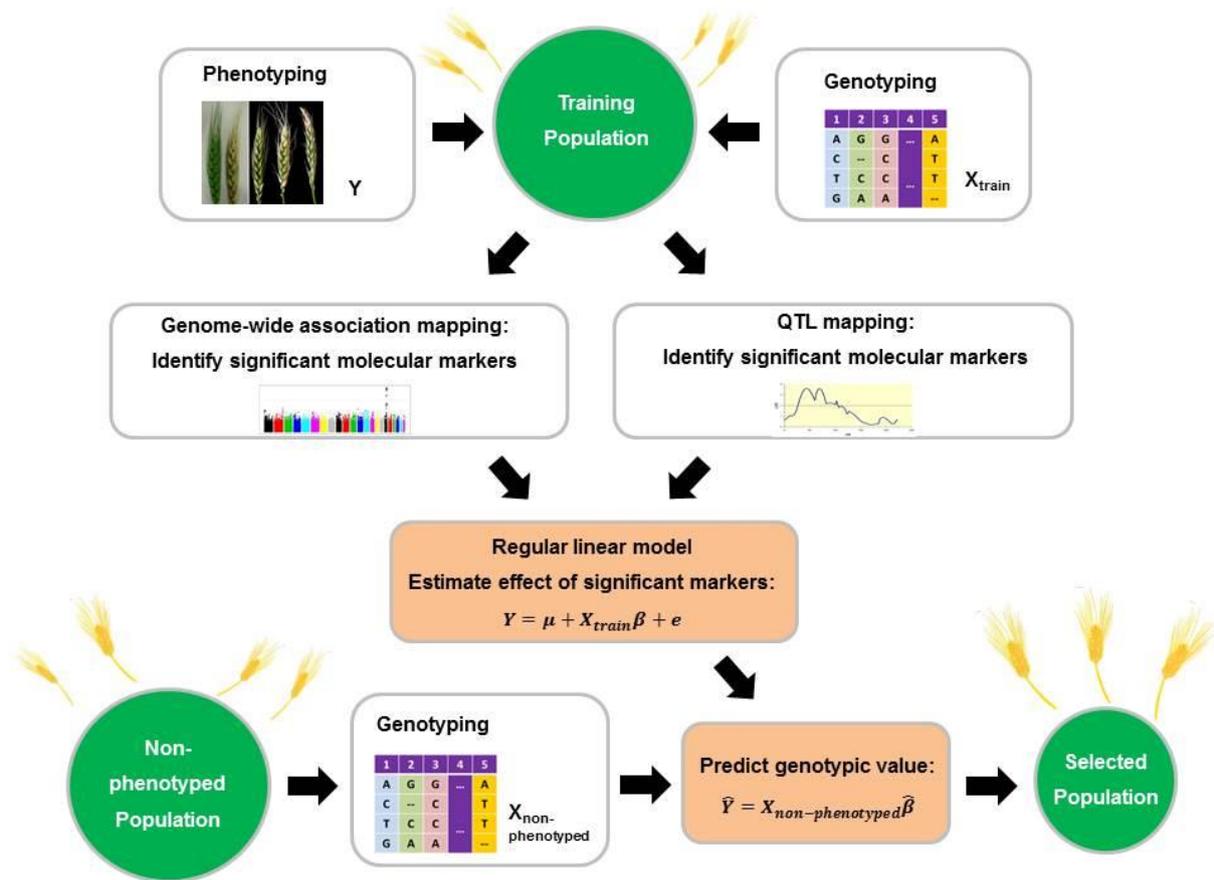


Figure 2 Principle of marker-assisted selection (MAS)

Marker-assisted selection

The rapid advancement in genomics paved the way for marker-assisted selection (MAS). In MAS, selection is based on molecular markers which reflect causal genes underlying traits of interest. MAS reduces the length of breeding cycles, decreases labor for intensive phenotyping, and potentially enhances the selection gain per unit time. Three stages are involved in MAS (Fig. 2): (1) Reliable diagnostic markers are identified in genome-wide mapping studies. (2) The identified markers are validated and their effects are then estimated. (3) The estimated effects of the diagnostic markers are applied using the genotypic profiles of non-phenotyped candidates to estimate their genotypic values.

MAS is successfully implemented in wheat breeding programs for traits that are controlled by large effect quantitative trait loci (QTL) such as leaf or stripe rust resistance (Helguera et al. 2003; Kumar et al. 2010). Nevertheless, the implementation of MAS for traits with a complex genetic architecture such as grain yield failed so far. Hence, there is a necessity to search for alternative approaches which allow exploiting the genomic toolbox for breeding of complex traits.

Genomic selection

Genomic selection (GS) has been proposed to improve the prediction accuracy for complex traits which are controlled by many genes each exhibiting a small effect (Meuwissen et al. 2001). In GS, many markers are used and their effects are estimated in large training populations. The estimated marker effects are then applied to predict the performance of non-phenotyped individuals based on their molecular marker profiles. Genomic selection takes advantage of linear mixed models rather than regular linear models exploited in MAS. In particular, GS models could be divided into marker-based and individual-based models (Gianola and van Kaam 2008; Habier et al. 2011; Hayes et al. 2009a; Jiang and Reif 2015; Meuwissen et al. 2001; Whittaker et al. 2000). The most popular approaches of marker-based models are RRBLUP (Meuwissen et al. 2001) and Bayesian models (Habier et al. 2011; Meuwissen et al. 2001). The unified formula for marker-based models is:

$$\mathbf{y} = \mu + \mathbf{X}\boldsymbol{\alpha} + \mathbf{e},$$

where \mathbf{y} is the vector of phenotypic records, μ is a common intercept, $\boldsymbol{\alpha}$ is the vector of additive effects of markers, \mathbf{X} is the design matrix for $\boldsymbol{\alpha}$, and \mathbf{e} is the residual term. RRBLUP assumes each marker follows an identical normal distribution: $\alpha \sim N(0, I\sigma_{\alpha}^2)$. Bayesian approaches assumes each maker has its own normal distribution: $\alpha_j \sim N(0, \sigma_{\alpha_j}^2)$, being j the j^{th} marker. With respect to individual-based models, GBLUP (Hayes et al. 2009a; VanRaden 2008) and kernel Hilbert space regression (RKHS; Gianola and van Kaam 2008) are two prominent approaches. The unified formula for individual-based models is:

$$\mathbf{y} = \mu + \mathbf{g} + \mathbf{e},$$

where \mathbf{g} is the vector of total genotypic values. The basic assumption for \mathbf{g} is $\mathbf{g} \sim N(0, G\sigma_g^2)$ where G refers to the genomic relationship matrix (VanRaden 2008). Habier et al. (2007) has shown that there

is a underlying homogeneity that RRBLUP is theoretically equivalent to GBLUP inspecting that in RRBLUP $var(X\alpha) = Xvar(\alpha)X^T = XX^T var(\alpha)$ and in GBLUP the G matrix is a scaling pattern of XX^T .

Genomic selection has been evaluated using populations with a maximum of ~400 wheat lines adapted to Europe in the context of line breeding (Bentley et al. 2014). The findings revealed that GS is a valuable tool to boost selection gain in wheat breeding (Heffner et al. 2010; Longin et al. 2015; Zhao et al. 2013b). Evaluating the potential of GS for elite wheat breeding populations including thousands of lines were missing. In this context, it is central so study the optimum implementation of GS in wheat breeding specifically for each trait, because the genetic architecture is expected to strongly impact the prediction accuracy (Daetwyler et al. 2010; Hayes et al. 2010).

Genetic architecture of grain yield

Increasing grain yield is primarily concerned by breeders and regarded as the main goal in wheat breeding. The genetic architecture of grain yield is expected to be complex with absence of large effect QTL (Bordes et al. 2014; Charmet et al. 2014; Snape et al. 2007). Despite this, many studies detected QTL individually explaining more than 10% of the total genetic variation (Bogard et al. 2011; Groos et al. 2003; Maccaferri et al. 2008; Snape et al. 2007) pointing to the potential use of MAS. These findings are astonishing and can be explained most likely by *small n* (individuals) *large p* (markers) scenarios (Beavis effect) in the context of small-scale populations under consideration in genome-wide QTL mapping (Beavis 1994; Beavis 1998; Xu 2003). Nevertheless, presence or absence of reliable QTL underlying grain yield has not been examined based on experimental data of large mapping populations but is of great importance for the optimum design of genomic-assisted breeding programs.

Epistasis refers to interactions between alleles from two or more genetic loci (Carlborg and Haley 2004). A first pioneering study based on populations of elite wheat lines adapted to Europe revealed presence of digenic epistasis but individual effects explained only a small proportion of the genetic variance (Reif et al. 2011b). Nevertheless, the results have to be interpreted with care because mapping epistatic effect is even more vulnerable to a *small n large p* scenario: The number of tests is

not a linear but a quadratic function of the number of markers. Therefore, sophisticated strategies are required to examine the role of epistasis for grain yield in wheat based on large mapping populations.

Genetic architecture of resistances to FHB and STB

FHB infection severely jeopardizes worldwide wheat production (Anderson et al. 2007; Buerstmayr et al. 2009; Mirdita et al. 2015; Yang et al. 2005b). Therefore, major efforts have been spent in unraveling the genetic basis of resistance to FHB (Löffler et al. 2009; Miedaner et al. 2011; Yang et al. 2005b). A few major QTL for FHB resistance have been identified in populations derived from crosses with exotic donor lines (Buerstmayr et al. 2003; Jia et al. 2005; Rawat et al. 2016; Somers et al. 2003; Yang et al. 2005a). These major QTL have so far not been used successfully in wheat breeding in Central Europe mainly because of a grain yield penalty associated with the resistance alleles. Identification of QTL in adapted European wheat germplasm revealed a lack of congruency of QTL results across studies (Buerstmayr et al. 2009). This suggests a complex genetic architecture underlying FHB resistance requiring large populations for QTL mapping.

STB infection caused by *Mycosphaerella graminicola* has become a disastrous leaf disease in Central European winter wheat nowadays (Arraiano and Brown 2006; Ghaffary et al. 2011; Kollers et al. 2013). Twenty major resistance loci have been identified and mapped such as *Stb2*, *Stb3* (Adhikari et al. 2004) and *Stb5* (Arraiano et al. 2001). Most are effective only against virulent genotypes of *Mycosphaerella graminicola* and resistances have been broken through the evolution of pathogen virulence (Krenz et al. 2008). Therefore, it is crucial to gain a clear picture of the presence of resistance loci which are still effective and thus useful for wheat breeding and production in Central Europe.

Marker imputation for an economic fingerprinting in marker-assisted and genomic selection

Detection of QTL in MAS as well as GS relies on molecular markers that are tightly linked with any relevant casual gene or QTL (Collard and Mackill 2008; Goddard 2009). Nevertheless, fingerprinting many genotypes with high-density marker platforms is costly. In order to counter this dilemma, genotype imputation technology has emerged facilitating to focus on a limited training population for dense genotyping and to predict the marker profiles for the total population (Hickey et al. 2012; Hozé

et al. 2013). Imputation approaches are classified into map-dependent and map-independent algorithms (Browning and Browning 2007; Howie et al. 2009; Sargolzaei et al. 2011; Stekhoven and Bühlmann 2012; Troyanskaya et al. 2001). Map-dependent methods such as Beagle (Browning and Browning 2007), FImpute (Sargolzaei et al. 2011), and IMPUTE2 (Howie et al. 2009) necessitate the availability of physical or genetic maps. Map-dependent methods are precise because they efficiently exploit linkage disequilibrium information among markers. In contrast, map-independent methods are grounded on mathematical algorithms (Breiman 2001; Troyanskaya et al. 2001). Thus, information of linkage disequilibrium is largely ignored which can introduce bias in imputation. As comprehensive physical or genetic maps are for several crops not yet available, map-independent approaches are still demanded.

The potential and limits of marker imputation based on map-dependent and map-independent algorithms for an economic fingerprinting strategy in marker-assisted and genomic selection was not studied in wheat. In particular, it is of interest to examine whether prediction accuracies of genomic selection can be enhanced if missing values are imputed when merging low- and high-density marker data in contrast to a scenario excluding missing marker data. The same holds true for the power of genome-wide association mapping where the question arises whether the power can be boosted if missing values are imputed.

Optimum design of a training population in genomic selection

The size and composition of the training population is an important factor driving the prediction accuracy of genomic selection (Habier et al. 2007; Jannink et al. 2010; Pszczola et al. 2012). The most economic approach to calibrate genomic selection models is to use a training population consisting of individuals that are routinely phenotyped in the course of wheat breeding (Dawson et al. 2013; Rutkoski et al. 2015). Nevertheless, breeding populations are selected which could result in a reduction in accuracy of genomic selection. The impact of selection on the genetic variation is lower at early breeding stages but phenotypic data is often generated in a reduced number of environments, entailing low heritability values. In contrast, phenotypic data of breeding population of late selection stages rely on many environments resulting in high heritabilities. Nevertheless, these populations are highly selected, implying a reduced genetic variation. Thus, there is the need to study sophisticated

approaches to compile optimal training populations based on phenotypic data routinely generated in breeding programs.

Prediction accuracy of individual genotypes

In plant breeding, prediction accuracies of genomic selection are most often estimated for entire breeding populations (Mirdita et al. 2015; Philipp et al. 2016; Poland et al. 2012a; Zhao et al. 2012a; Zhao et al. 2013b). This is in contrast to animal breeding where prediction accuracies are estimated also for individual genotypes, denoted as reliability (Heslot et al. 2015; Pszczola et al. 2012; VanRaden et al. 2009). The concept of reliability is based on the assumption that genomic selection strongly exploits relatedness between the training and test population for prediction (Habier et al. 2007; Habier et al. 2010; Jannink et al. 2010). Thus, it is relevant whether selection candidates are well represented by the training population or not. The reliability is estimated purely based on genotypic data and its application in plant breeding possesses the potential to speed up the implementation of genomic selection. Nevertheless, compared to animal breeding, reliability criterion hitherto has not been examined based on experimental data from plant breeding programs.

Objectives of this thesis

The thesis is based on experimental data generated for two European elite wheat breeding populations. One consists of 372 varieties fingerprinted with 9,926 single nucleotide polymorphism (SNP) markers. Another is composed by up to 3,816 elite lines genotyped with 13,901 SNP markers. The objectives were to:

- (1) investigate the accuracy of genotype imputation in wheat and its influence on prediction accuracy of genomic selection and detection power of genome-wide association mapping (He et al. 2015),
- (2) study the potential and limits of marker-assisted selection and genomic selection for enhancing resistance to FHB and STB (Mirdita et al. 2015),
- (3) explore how to enhance prediction ability of genomic selection by means of optimizing the constitution of training population, extending genomic prediction models, and exploiting relatedness (He et al. 2016b), and
- (4) try to unravel the genetic architecture of grain yield in wheat and dissect the relevance of local and global epistasis in grain yield (He et al. 2016a).

Original papers

Prospects and limits of marker imputation in quantitative genetic studies in European elite wheat (*Triticum aestivum* L.)

Published in:

BMC genomics (2015) 16:168

DOI: 10.1186/s12864-015-1366-y

Authors: Sang He, Yusheng Zhao, M Florian Mette, Reiner Bothe, Erhard Ebmeyer, Timothy F Sharbel, Jochen C Reif and Yong Jiang

The original paper has been published and available online:

<http://link.springer.com/article/10.1186/s12864-015-1366-y>

Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast Central European elite winter wheat population

Published in:

Theoretical and Applied Genetics (2016) 128:2471-2481

DOI: 10.1007/s00122-015-2602-1

Authors: Vilson Mirdita[†], Sang He[†], Yusheng Zhao, Viktor Korzun, Reiner Bothe, Erhard Ebmeyer, Jochen C. Reif, Yong Jiang

[†]: Both authors contributed equally to this work

The original paper has been published and available online:

<http://link.springer.com/article/10.1007/s00122-015-2602-1>

Genomic selection in a commercial winter wheat population

Published in:

Theoretical and Applied Genetics (2016) 129:641-651

DOI: 10.1007/s00122-015-2655-1

Authors: Sang He, Albert Wilhelm Schulthess, Vilson Mirdita, Yusheng Zhao, Viktor Korzun, Reiner Bothe, Erhard Ebmeyer, Jochen C. Reif, Yong Jiang

The original paper has been published and available online:

<http://link.springer.com/article/10.1007/s00122-015-2655-1>

Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe

Published in:

Theoretical and Applied Genetics (2016):1-13

DOI: 10.1007/s00122-016-2840-x

Authors: Sang He, Jochen C. Reif, Viktor Korzun, Reiner Bothe, Erhard Ebmeyer, Yong Jiang

The original paper has been published and available online:

<http://link.springer.com/article/10.1007/s00122-016-2840-x>

General discussion and conclusion

Genomics-assisted breeding is a promising toolbox to boost selection gain (Heffner et al. 2010; Heffner et al. 2009). The research and development phases required to establish genomics-assisted breeding were often considered as separate processes from the routine breeding of improved cultivars. This held true to discover functional markers for marker-assisted selection (Podlich et al. 2004) and to establish diverse training populations for genomic selection (Heffner et al. 2009). Nevertheless, large population sizes are needed to detect minor QTL which segregate in elite germplasms. Moreover, genome-wide prediction models should be established based on representative training populations. Thus, an economic implementation of genomic-assisted breeding would profit from mining phenotypic and genomic data, which have been routinely generated in applied plant breeding programs. Data generated in the course of breeding is highly unbalanced because the intensity of phenotyping strongly varies for the different stages of selection. Moreover, platforms for generating genomic data are often updated leading to a mosaic of genetic fingerprints available. The integration of these unbalanced data sets is challenging but required in order to implement successfully genomics-assisted breeding. In my thesis, I used experimental data from the wheat breeding program of KWS LOCHOW and elaborated strategies to combine and analyze the unbalanced phenotypic and genomic data. Based on the experimental data, I then examined the potential and limits of genomics-assisted wheat improvement of grain yield and resistances against *Fusarium* head blight and *Septoria tritici*.

Imputation of missing values paved the way to integrate heterogeneous genomic data across different genotyping platforms

Several marker platforms are available for high-throughput genotyping in wheat such as SNP arrays (Cavanagh et al. 2013; Wang et al. 2014a), diversity arrays technology (Akbari et al. 2006; Peleg et al. 2008) and genotyping-by-sequencing (Poland et al. 2012a; Poland et al. 2012b). Amongst them, SNP arrays are often used in wheat breeding (Cavanagh et al. 2013; Wang et al. 2014a). The advantages of these SNP arrays in contrast to other marker platforms are a straight-forward quality assessment and a low number of missing values in the genotype-marker matrices (Mammadov et al. 2012). The potential risk of an ascertainment bias resulting from the establishment of the marker platform using

non representative panels of lines is for the narrow genetic basis of the European elite germplasm pool limited.

The progress in developing SNP arrays covering a larger number of markers in wheat is rapid: In the year 2013 a 9,000 SNP array (Cavanagh et al. 2013) was introduced, followed by a 90,000 SNP in 2014 (Wang et al. 2014a) and a 600,000 SNP array in 2016 (Winfield et al. 2015). Based on these platforms, optimized haplotype-based arrays have been developed for cost-efficient use in wheat breeding such as a 15,000 SNP array (He et al. 2016a). As a consequence of the rapid progress in designing optimized genotyping platforms in wheat, breeding lines of different years are often fingerprinted based on altered SNP arrays. This was the case for the wheat lines of KWS LOCHOW which were fingerprinted in the year 2012 by a 9,000 SNP array followed by a 90,000 SNP array in 2013, and a 15,000 SNP array in 2014 (Fig. 3; He et al. 2016a). It is projected that the dynamic in the field of optimizing marker platforms will not slow down in the coming years. Thus, there exists the need to integrate the genomic data across the different genotyping platforms.

One option to combine heterogeneous genomic data sets is to focus only on the overlapping markers. Nevertheless, this strategy tremendously reduces the marker density. As an alternative, missing values can be imputed in order to integrate the unbalanced genomic data. Our findings have revealed that linkage disequilibrium is the central parameter determining the imputation accuracy (He et al 2015). Moreover, imputation accuracy is hampered for loci exhibiting low minor allele frequencies. Rare alleles exhibiting positive effects are of particular interest for breeding (Jannink 2010; Sun and VanRaden 2014). Thus, the question arises to which extent genome-wide association mapping and prediction profits from imputation of markers.

The accuracy for marker imputation depends on linkage disequilibrium which is also the central parameter driving the power of genome-wide QTL mapping (Hamblin et al. 2011) and the accuracy of genomic selection (Heffner et al. 2009). The extent and distribution of linkage disequilibrium strongly depends on the diversity of the population under consideration (Chao et al. 2010; Maccaferri et al. 2005; Somers et al. 2007). The estimated effective population size in the Central European elite wheat population amounted to around 30 individuals (He et al. 2016a), which leads to a substantial extent of linkage disequilibrium (He et al. 2016a; He et al. 2015). He et al. (2015)

estimated based on the observed decay of linkage disequilibrium with genetic map distance that around 45,000 markers are required to portray the diversity of the gene space guaranteeing high power of QTL detection in the Central European elite wheat population. Expanding the diversity outside of the gene space which reflects around 1% of the wheat genome leads to an up to 100 times larger number of markers required for genome-wide association mapping. In contrast, the required marker density for genomic selection is lower because prediction is driven by a combination of knowledge of the genetic architecture and relatedness (Habier et al. 2007). Using the formula provided by Solberg et al. (2008) in combination with the estimated effective population size for Central European wheat (He et al. 2016a) suggests that around 12,000 markers are required to facilitate high prediction accuracy in genomic selection. In line with these expectations we observed that genome-wide prediction accuracy increased only little when imputing missing marker values from the 9,000 to the 90,000 SNP array (He et al. 2015). In contrast, genome-wide association mapping profited substantially from imputing missing values. Thus, using a mix of high and low-density genotyping of wheat breeding populations is an attractive approach for an economic implementation of association mapping in breeding populations.

Strategy to integrate phenotypic data across different selection cycles

Phenotypic selection is commonly based on data from one specific selection cycle. Thus, once phenotypic data have been applied for selection they were in the pre-genomics era of no further use. In contrast, implementing genomics-assisted breeding profits from large population sizes which adds tremendous value to historic data. This entails substantial challenges on the design of field trials, quality assessment, and analyses of phenotypic data in order to integrate information across different breeding cycles.

Field trials in plant breeding are often designed in a way to maximize the precision of the contrasts among lines within selection cycles considering also important released check varieties. The presence of check varieties served in our studies as important bridge for integrated analyses across selection cycles (Fig. 3; He et al. 2016b). We used a two-step analyses approach with a first step consisting in calculating the Best Linear Unbiased Estimates (BLUEs) of genotypes evaluated in all field trials conducted within a particular environment, i.e. year-by-location combination. This first step

enabled a quality assessment by inspecting the repeatability within each environment. In a second step, BLUEs of genotypes were estimated across environments and used for genome-wide prediction or association mapping. The two-step approach of the analyses of phenotypic data was reported to perform nearly equivalent compared to one-step approaches despite that the latter strategy exploits more information (Möhring and Piepho 2009). We observed also in our study only marginal differences between a one- and two step approaches (data not shown). The major advantage of a two-step analysis is the decrease in computational time (Möhring and Piepho 2009) and the ease to stack continuously data across years. Thus, the implemented two-step strategy is a vital approach to integrate phenotypic data across different selection cycles in wheat breeding programs.

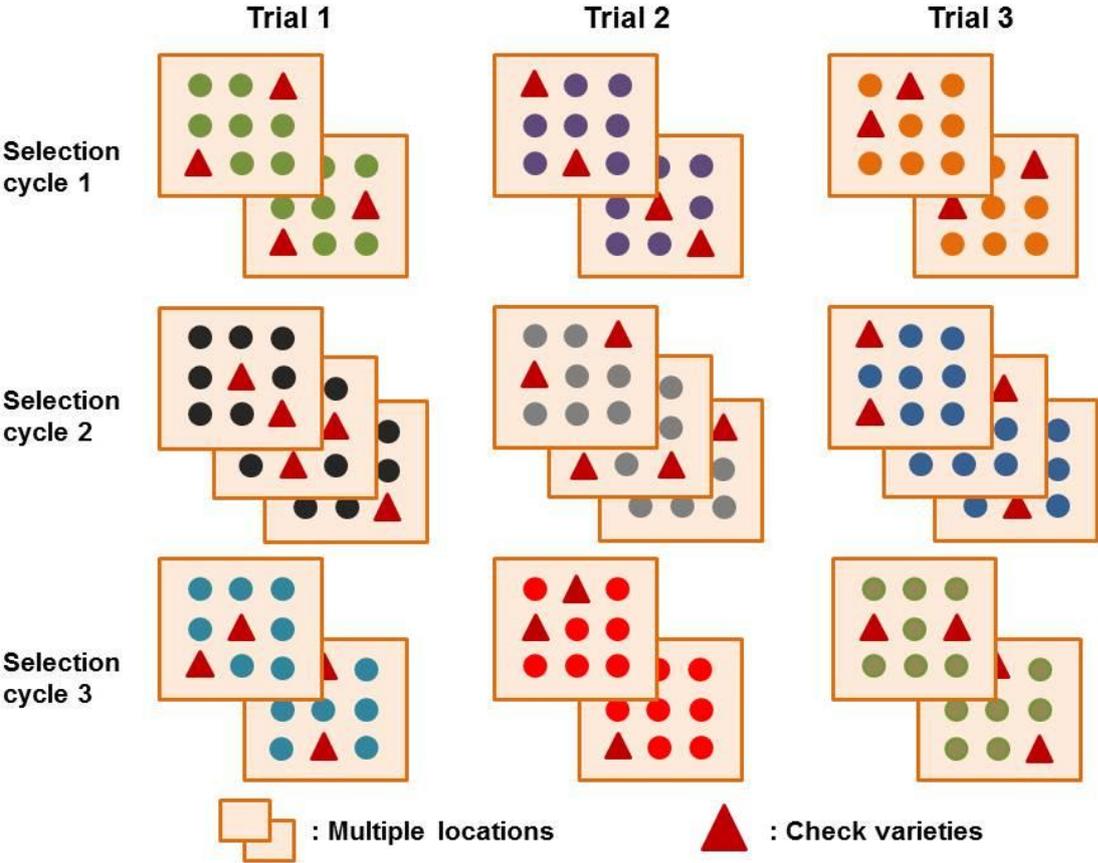


Figure 3 Current field trial design. Check varieties enable to bridge selection cycles and perform integrated data analyses.

Despite the fact that the presented strategy seems to be robust in order to implement genomics-assisted breeding based on phenotypic data routinely generated in the course of breeding,

the question remains whether a complete redesign of field trials is beneficial or not for plant breeding in the genomics era. Increasing the precision of contrasts among genotypes across selection cycles requires more balanced field designs mixing germplasm across cycles. This can increase the prediction accuracy of genome-wide selection. Nevertheless, the comparisons among lines within the same breeding cycles based on phenotypic values will become less precise, which may be partially compensated when combining it with genome-wide predictions. Further research is needed to ultimately answer the question on the optimal design of field trials in the genomic era.

Potential and limits of marker-assisted and genomic selection for complex traits

Marker-assisted selection exploits linkage disequilibrium between diagnostic markers and QTL underlying the trait under consideration and to a lesser extent also the relatedness between the training and test population (Heffner et al. 2009; Jiang et al. 2015; Wang et al. 2014b). Thus, the accuracy of marker-assisted selection depends crucially on the choice of stable diagnostic markers. Our genome-wide association mapping study revealed that grain yield (He et al. 2016a) and resistances against *Fusarium* head blight and *Septoria tritici* (Mirdita et al. 2015) are complex traits approximating an infinitesimal model which ultimately impede the usefulness of marker-assisted selection. In contrast, genomic selection is able to exploit genetic relatedness for predicting grain yield (He et al. 2016b) and resistances against *Fusarium* head blight and *Septoria tritici* (Mirdita et al. 2015) more efficiently compared to marker-assisted selection and is therefore the method of choice. The prediction accuracies for the three traits ranged from 0.74 for resistance against *Fusarium* head blight to 0.87 for *Septoria tritici* corresponding to phenotyping in up to 7 environments for grain yield (He et al. 2016b) and up to 19 environments for resistance against *Septoria tritici* underpinning the large potential of genomic selection.

Relevance of variance due to genotype-by-year interactions determines to a large extent the stability of the developed prediction models across time. The grain yield data in our study from the year 2012 were impacted by frost stress (Zhao et al. 2013a). Nevertheless, prediction accuracies for grain yield were stable across the years 2012 and 2013 (He et al. 2016a). In contrast, prediction accuracies across the years decreased tremendously for resistance against *Fusarium* head blight (Mirdita et al. 2015). This is in line with pronounced genotype-by-year interactions which have been

reported for resistances against Fusarium head blight in wheat (Buerstmayr et al. 2002; Mesterhazy 1995; Steiner et al. 2004) pointing to the need to phenotype training populations across years. Multi-year data for resistances against Fusarium head blight have to be collected outside of the routine breeding process and can be accomplished most efficiently by applying two-tailed selection (Zhao et al. 2012b). In summary, our study revealed that genomic selection is a powerful tool for Central European wheat improvement of grain yield and resistances against Fusarium head blight and *Septoria tritici*.

Choice of biometric model for genomic selection

We contrasted the prediction accuracies of four biometric models for genomic selection. Two of the four models, RRBLUP and Bayes-C π , exploit additive effects and vary with respect to the distribution from which marker effects are sampled. RRBLUP assumes that markers are independently and identically distributed but BayesC π allows marker-specific and independent variance (Habier et al. 2011; Meuwissen et al. 2001). Thus, it is expected that RRBLUP fits better the infinitesimal model and BayesC π is supposed to outperform RRBLUP in tackling traits controlled also by major genes (Meuwissen et al. 2001; Sun et al. 2011; Wolc et al. 2016). In our study, we observed only marginal differences between both approaches (He et al. 2016b; Mirdita et al. 2015). This finding is in line with previous studies (Haws et al. 2015; Heslot et al. 2012; Jiang et al. 2015). RRBLUP is less computational demanding (Heslot et al. 2012). Therefore, RRBLUP seems to be the method of choice for genomic selection based on additive effects for the traits grain yield and resistances against Fusarium head blight and *Septoria tritici*.

Another two popular biometric models, EGBLUP and RKHS, directly estimate genotypic values by means of constructing relationship matrices on the basis of SNP profiles. The main difference between EGBLUP and RKHS versus RRBLUP and BayesC π is that the first two approaches are both capable to exploit besides main also epistatic effects (de los Campos et al. 2010; Heslot et al. 2012; Jiang and Reif 2015). Our study revealed higher prediction accuracies for EGBLUP and RKHS versus RRBLUP and BayesC π in wheat for all three examined traits grain yield (He et al. 2016b) and resistances against Fusarium head blight and *Septoria tritici* (Mirdita et al. 2015). Consequently, genomic selection based on EGBLUP and RKHS are the methods of choice for

selecting superior lines. Using EGBLUP additionally enables to decompose the genotypic value into the breeding value and a non-additive component. The former is in particular of interest because it reflects an interesting source of information for parental selection to initiate the next cycle of selection.

Global epistasis cannot be exploited in a recurrent fashion, because independent assortment and recombination would break the established linking phase of pairs of QTL (Falconer and Mackay 1996; Akdemir and Jannink 2015). In contrast, local epistasis originating from regional linkage blocks could be conserved across breeding generations and exploited in a recurrent fashion (Akdemir and Jannink 2015). Local epistasis can be portrayed using haplotype-based genomic selection approaches (He et al. 2016a; Wei et al. 2014). Our study indicated presence of local epistasis in the elite winter wheat population with improved prediction abilities when shifting from marker- to haplotype-based genome-wide prediction approaches. Therefore, haplotype-based genomic selection offers an opportunity to consider local epistasis when selecting parents to initiate the next cycle of selection.

Optimum composition of a training population for genome-wide predictions

The composition of training populations is crucial to establish genome-wide prediction models. On one hand, marker effects are estimated from training populations and subsequently used for predicting genotypic values of selection candidates. On the other hand, relatedness between training populations and selection candidates is an important factor influencing the prediction accuracy (Habier et al. 2007; Hayes et al. 2009b; Rutkoski et al. 2015). A vast amount of historical phenotypic data generated from different selection cycles provide a promising repository to build up a diverse training population. Focusing on relatedness, Rincent et al. (2012), Akdemir et al. (2015), and Isidro et al. (2015) proposed strategies to optimize the composition of training populations by specifically selecting training subsets for each selection candidate. The merit of relatedness is fully exploited when closely related individuals of a given selection candidate have been incorporated in the training population. Nevertheless, genetically more distant individuals in the training population are not decreasing the prediction accuracy. Thus, there is no need to compose training populations by sampling only individuals related to selection candidates, when historical data of breeding programs are considered (Heslot et al. 2015). Consequently, optimizing the composition of training populations based on historical data should focus on other factors than relatedness.

The quality of the phenotypic data and the diversity of the training population are possible alternative criteria to sample training populations based on historical data. Best linear unbiased estimates (BLUEs) of genotypic values are used commonly as response variable in genome-wide prediction models. The precision of BLUEs depends on the underlying phenotyping density which varies among selection candidates (Lorenz 2013). In the early selection cycles, phenotyping intensity is low but a large amount of selection candidates is evaluated. Thus, BLUEs of selection candidates which are rejected in early selection cycles are imprecise impairing the response variable of genomic prediction models. By contrast, selection candidates of late selection cycles have been intensively evaluated in field trials but the population size is small. We inspected the best compromise between population size and phenotyping intensity/phenotypic data quality (He et al. 2016a) and observed that selection candidates solely evaluated in one location should be discarded in order to optimize the prediction accuracy of genomic selection.

Prediction accuracy is highly variable among genotypes but can be assessed with genomic data

Prediction accuracy of genomic selection in plant breeding is commonly estimated for the entire breeding population (Mirdita et al. 2015; Philipp et al. 2016; Poland et al. 2012a; Zhao et al. 2012a; Zhao et al. 2013b). Nevertheless, the degree of relatedness between test candidates and the training set is highly variable in the population of test candidates. Therefore, prediction accuracy is treated in animal breeding separately for each individual which is denoted as reliability (Heslot et al. 2015; Pszczola et al. 2012; VanRaden et al. 2009). The reliability criterion was so far not been implemented in the context of plant breeding. Our findings points to the great merit of using the reliability criterion in plant breeding (He et al. 2016b): The prediction accuracy of the total population amounted to 0.71 but was approaching one for the subset of the most reliable genotypes. These differences in prediction accuracies reflect differences in phenotyping intensities ranging from field evaluation in 5 to 50 environments (Fig. 4). Thus, fast-track genomics-based breeding strategies can be implemented and require genomic selection strategies specifically designed for every genotype. This genomic selection specifically tailored for each genotype entails a paradigm shift and strongly depends on massive adjustments in the logistic chains of wheat breeding programs. These adjustments are for instance relevant for selection candidates with low reliabilities being genetically distant from training

population. For these selection candidates phenotypic evaluations cannot be replaced by genomic selection.

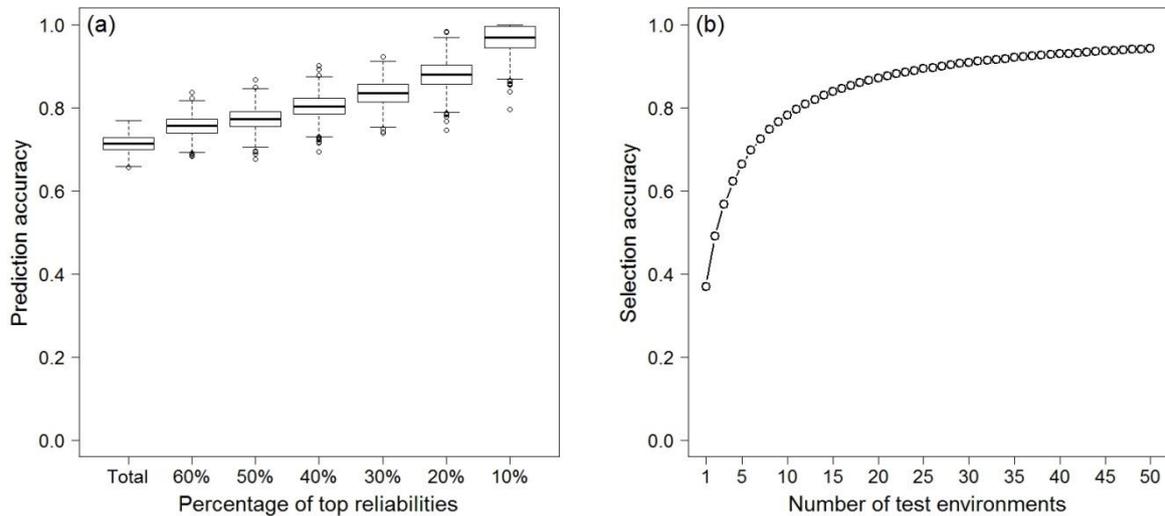


Figure 4 (a) Prediction accuracy of genomic selection for grain yield inspecting different subsets composed by genotypes with top 60% to 10% reliabilities; (b) The association between selection accuracy and number of test environments

Implementation of genomic selection in breeding programs using historical data

Genomic selection is expected to shift the paradigm of breeding schemes in the genomics era. Towards a conservative attitude utilizing genomic selection, it is suggested to treat it as an auxiliary means assisting traditional phenotypic selection. Focusing on phenotypic selection, in the early stages phenotyping intensity is limited due to a vast amount of selection candidates needed to be tested. In consequence, heritabilities in early stages are unambiguously low entailing a disappointing selection gain (Endelman et al. 2014; Heslot et al. 2015). Nevertheless, having a powerful training population available, genomic selection could markedly increase the selection accuracy in preliminary stages boosting selection gain. Endelman et al. (2014) and Lorenz (2013) even proposed to entirely ignore phenotyping at preliminary stages and focusing exclusively on genomic predicted values. The main constraint for applying genomic prediction is the extra expenditure on genotyping in contrast to phenotypic selection. Nevertheless, this additional cost would become more and more trivial as the advent of cost-effective platforms such as genotyping-by-sequencing (Riedelsheimer and Melchinger

2013). By contrast, in the medium or late stages of breeding programs genomic selection would not arouse remarkable profit because intensive field evaluations implemented in phenotypic selection has guaranteed a high accuracy level (He et al. 2016b).

Once genomic selection has demonstrated its merits, breeders have to gain the knowledge on how to allocate resources between genotyping and phenotyping (Longin et al. 2015; Lorenz 2013); how to interpret genotype-by-environment interaction in the genomics era (Lopez-Cruz et al. 2015; Schulz - Streeck et al. 2013); and how to renew the design of field trials to better connect different selection cycles and tune the tradeoff between investment on number of tested location and replicates per location (Lorenz 2013). More empirical studies with regard to the effectiveness of genomic selection are needed to provide more evidence and reference on the way to optimize the application of genomic selection in practice.

Summary

Wheat (*Triticum aestivum* L.) is one of the most important crops providing 20% of the total calorie consumption of the entire human population. Worldwide wheat production needs to be doubled to feed an estimated human population of 9 billion by 2050. Wheat breeding is considered as an environmentally sound approach to boost grain yield and to improve yield stability. The advent of high-throughput molecular markers offers new perspectives to realize these goals in the genomics era. Genomic selection jointly exploits pedigree information, co-segregation, and linkage disequilibrium between markers and QTL to predict complex traits and has been widely applied in animal breeding. Nevertheless, this was not the case for wheat improvement. The main goal of this research was to examine the potential and limits of genomics-assisted wheat improvement of grain yield and resistances against *Fusarium* head blight and *Septoria tritici*. The comprehensive phenotypic and genomic data of the study comprising up to 3,816 elite wheat lines adapted to Central Europe were generated in a commercial wheat breeding program.

As a consequence of a rapid progress in designing optimized genotyping technologies in wheat, breeding lines of different years have been often fingerprinted based on altered marker platforms. Thus, there exists the need to integrate these unbalanced genomic data. Genotype imputation is a promising approach to merge data of different marker platforms. The objective of the presented simulation study was to investigate the accuracy of genotype imputation in wheat. Imputation accuracy was highly dependent on linkage disequilibrium between markers in reference panels and the marker to be imputed. The detection power of genome-wide association mapping can be increased substantially by imputing missing values. Genomic selection profited marginally from genotype imputation.

The complexity of the genetic architecture plays a crucial role for the selection of the optimal genomics-assisted breeding strategy. Genomic selection is expected to outperform marker-assisted selection if the trait under consideration approximates an infinitesimal model with presence of many genes contributing to the phenotypic variation. The goal of the presented experimental studies were to use genome-wide association mapping in combination with cross-validations or validations in

independent data sets to investigate the genetic architecture of grain yield and resistances against *Fusarium* head blight and *Septoria tritici*. The genome-wide association mapping analyses revealed that the three traits are complex approximating the infinitesimal model. Consequently, genomic selection outperformed marker-assisted selection. Contrasting different genomic selection models suggested that interaction effects among genes, i.e. epistasis, contribute substantially to the phenotypic variation. Nevertheless, only 5% of the epistatic effects occurred within conserved local linkage blocks and can therefore be exploited across generations.

Prediction accuracy of genomic selection is ordinarily expressed as an average value of the total population. Nevertheless, the degree of relatedness between training set and selection candidates which serves as the most essential factor influencing genomic prediction is highly divergent among predicted candidates. Reliability criterion offers a prospect to inspect the degree of relatedness specifically for each individual candidate. The findings of an experimental study demonstrated that a subset of genotypes with high reliability values were prone to possess a high prediction accuracy whilst low reliable ones could not be accurately predicted. This finding implies a paradigm shift in genomic selection which should be designed specifically for each genotype and entails massive adjustments in the logistic chains of wheat breeding programs.

Zusammenfassung

Weizen (*Triticum aestivum* L.) ist eine der wichtigsten Kulturpflanzen und trägt 20 % zum gesamten Kalorienbedarf der menschlichen Bevölkerung bei. Die Weizenproduktion muss bis zum Jahr 2050 verdoppelt werden, um die geschätzte Weltbevölkerung von 9 Milliarden Menschen zu ernähren. Die Weizenzüchtung ist ein nachhaltiger Ansatz zur Steigerung des Ertrags und der Ertragsstabilität. Die Entwicklung von hochdurchsatzfähigen molekularen Markertechnologien eröffnet im Zeitalter der Genomik neue Ansätze, um diese Ziele in der Züchtung zu erreichen. Die genomische Selektion kombiniert Informationen über die Herkunft, die Kosegregation und das Kopplungsphasenungleichgewicht zwischen Markern und QTL, um komplexe Merkmale vorherzusagen und wird bereits in größerem Umfang in der Tierzucht verwendet. Zur züchterischen Verbesserung von Weizen fand diese Methode jedoch bislang keine Anwendung. Das Hauptziel dieser Forschungsarbeit war es, das Potenzial und die Grenzen von Genomik-basierter Verbesserung des Kornertrags, sowie der Resistenz gegen Ährenfusariosen und *Septoria tritici* zu untersuchen. Die phänotypischen und genomischen Daten dieser Studie umfassen bis zu 3,816 mitteleuropäische Weizenelitelinien und wurden im Rahmen eines kommerziellen Weizenzüchtungsprogramms generiert.

Als eine Konsequenz aus dem schnellen Fortschritt im Design von optimierten Genotypisierungstechnologien bei Weizen, wurden Zuchtlinien in den verschiedenen Jahren oft mit unterschiedlichen Markerplattformen charakterisiert. Es ergibt sich daraus die Notwendigkeit, diese nicht balancierten, genomischen Daten zu integrieren. Das Schätzen von fehlenden Markerdaten ist eine vielversprechende Herangehensweise, um die Daten verschiedener Markerplattformen zusammenzuführen. Das Ziel der Simulationsstudie war es, die Präzision der Schätzung von fehlenden Markerdaten zu untersuchen. Die Genauigkeit der Schätzung wurde stark vom Kopplungsphasenungleichgewicht zwischen den Markern im Referenzpanel und den einzelnen betroffenen Markern beeinflusst. Die Güte der genomweiten Assoziationskartierung kann deutlich durch die Schätzung fehlender Werte erhöht werden. Die genomische Selektion hat geringfügig von der Schätzung von Fehlwerten profitiert.

Die Komplexität der genetischen Architektur spielt eine entscheidende Rolle bei der Wahl der optimalen Genomik-basierten Zuchtstrategie. Die genomische Selektion wird gegenüber der markergestützten Selektion als überlegen angesehen, wenn die zu selektierenden Merkmale annähernd einem infinitesimalen Modell mit einer Vielzahl von Genen entsprechen, die zur phänotypischen Variation beitragen. Das Ziel der experimentellen Studien war es, genomweite Assoziationskartierung zusammen mit Kreuzvalidierung oder Validierungen an unabhängigen Datensätzen zur Untersuchung der genetischen Architektur von Kornertrag und Resistenz gegen Ährenfusariosen und *Septoria tritici* zu nutzen. Die Analyse der genomweiten Assoziationskartierung ergab, dass diese drei Merkmale komplex vererbt werden und näherungsweise dem infinitesimalen Modell entsprechen. Die genomische Selektion war folgerichtig der markergestützten Selektion überlegen. Der Vergleich verschiedener biometrischer Modelle zur genomischen Selektion deutet darauf hin, dass die Auswirkungen der Interaktion zwischen Genen, d. h. Epistase, deutlich zur phänotypischen Variation beitragen. Allerdings traten nur 5 % der epistatischen Effekte innerhalb konservierter lokaler Kopplungsgruppen auf und können über Generationen hinaus genutzt werden.

Die Vorhersagegenauigkeit der genomischen Selektion wird gewöhnlich als ein Mittelwert für die gesamte Population ausgedrückt. Der Grad der Verwandtschaft zwischen dem Training-Set und den Selektionskandidaten, als wichtigster Faktor der genomischen Vorhersage, ist dessen ungeachtet hochgradig divergent zwischen den vorhergesagten Kandidaten. Die Resultate der experimentellen Studie demonstrieren, dass eine Untergruppe von Genotypen mit Werten hoher Verlässlichkeit zu einer hohen Vorhersagegenauigkeit tendierte, während gering verlässliche nicht korrekt vorhergesagt werden konnten. Dies bedingt einen Paradigmenwechsel für die genomische Selektion, die daher spezifisch für jeden einzelnen Genotyp geplant werden sollte, und verursacht substantielle Anpassungen in der Logistikkette von Weizenzuchtprogrammen.

Reference

- Adhikari TB, Wallwork H, Goodwin SB (2004) Microsatellite Markers Linked to the and Genes for Resistance to Septoria Tritici Blotch in Wheat. *Crop Science* 44:1403-1411
- Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and applied genetics* 113:1409-1420
- Akdemir D, Jannink J-L (2015) Locally epistatic genomic relationship matrices for genomic association and prediction. *Genetics* 199:857-871
- Akdemir D, Sanchez JI, Jannink J-L (2015) Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selection Evolution* 47:1
- Anderson JA, Chao S, Liu S (2007) Molecular breeding using a major QTL for Fusarium head blight resistance in wheat. *Crop Science* 47:S-112-S-119
- Arraiano L, Brown J (2006) Identification of isolate-specific and partial resistance to septoria tritici blotch in 238 European wheat cultivars and breeding lines. *Plant Pathology* 55:726-738
- Arraiano L, Worland A, Ellerbrook C, Brown J (2001) Chromosomal location of a gene for resistance to septoria tritici blotch (*Mycosphaerella graminicola*) in the hexaploid wheat 'Synthetic 6x'. *Theoretical and Applied Genetics* 103:758-764
- Beavis W (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proceedings of the forty-ninth annual corn and sorghum industry research conference*. Chicago, IL, pp 250-266
- Beavis WD (1998) QTL analyses: power, precision, and accuracy. *Molecular dissection of complex traits* 1998:145-162
- Becker H (2011) *Pflanzenzüchtung*. UTB
- Bentley AR, Scutari M, Gosman N, Faure S, Bedford F, Howell P, Cockram J, Rose GA, Barber T, Irigoyen J (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. *Theoretical and applied genetics* 127:2619-2633
- Bogard M, Jourdan M, Allard V, Martre P, Perretant MR, Ravel C, Heumez E, Orford S, Snape J, Griffiths S (2011) Anthesis date mainly explained correlations between post-anthesis leaf senescence, grain yield, and grain protein concentration in a winter wheat population segregating for flowering time QTLs. *Journal of experimental botany* 62:3621-3636
- Bordes J, Goudemand E, Duchalais L, Chevarin L, Oury FX, Heumez E, Lapierre A, Perretant MR, Rolland B, Beghin D (2014) Genome-wide association mapping of three important traits using bread wheat elite breeding populations. *Molecular breeding* 33:755-768
- Breiman L (2001) Random forests. *Machine learning* 45:5-32
- Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* 81:1084-1097
- Buerstmayr H, Ban T, Anderson JA (2009) QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. *Plant breeding* 128:1-26
- Buerstmayr H, Lemmens M, Hartl L, Doldi L, Steiner B, Stierschneider M, Ruckenbauer P (2002) Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). *Theoretical and Applied Genetics* 104:84-91
- Buerstmayr H, Steiner B, Hartl L, Griesser M, Angerer N, Lengauer D, Miedaner T, Schneider B, Lemmens M (2003) Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. *Theoretical and Applied Genetics* 107:503-508
- Bushuk W (1997) Wheat breeding for end-product use. *Wheat: Prospects for global improvement*. Springer, pp 203-211
- Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* 5:618-625
- Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A (2013) Genome-wide comparative diversity uncovers multiple

- targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the national academy of sciences* 110:8057-8062
- Chao S, Dubcovsky J, Dvorak J, Luo M-C, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC genomics* 11:1
- Charmet G, Storlie E, Oury FX, Laurent V, Beghin D, Chevarin L, Lapierre A, Perretant MR, Rolland B, Heumez E (2014) Genome-wide prediction of three important traits in bread wheat. *Molecular Breeding* 34:1843-1852
- Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:557-572
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021-1031
- Dawson JC, Endelman JB, Heslot N, Crossa J, Poland J, Dreisigacker S, Manès Y, Sorrells ME, Jannink J-L (2013) The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Research* 154:12-22
- de los Campos G, Gianola D, Rosa GJ, Weigel KA, Crossa J (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research* 92:295-308
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, Jannink J-L (2014) Optimal design of preliminary yield trials with genome-wide markers. *Crop Science* 54:48-59
- Fischer R, Byerlee D, Edmeades G (2014) Crop yields and global food security: will yield increase continue to feed the world? *ACIAR*:158:634
- Ghaffary SMT, Robert O, Laurent V, Lonnet P, Margalé E, van der Lee TA, Visser RG, Kema GH (2011) Genetic analysis of resistance to septoria tritici blotch in the French winter wheat cultivars Balance and Apache. *Theoretical and Applied Genetics* 123:741-754
- Gianola D, van Kaam JB (2008) Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178:2289-2303
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245-257
- Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010) Food security: the challenge of feeding 9 billion people. *Science* 327:812-818
- Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theoretical and Applied Genetics* 106:1032-1040
- Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389-2397
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics* 12:1
- Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* 42:1
- Hamblin MT, Buckler ES, Jannink J-L (2011) Population genetics of genomics-based crop improvement methods. *Trends in Genetics* 27:98-106
- Haws DC, Rish I, Teyssedre S, He D, Lozano AC, Kambadur P, Karaman Z, Parida L (2015) Variable-Selection Emerges on Top in Empirical Comparison of Whole-Genome Complex-Trait Prediction Methods. *PLoS One* 10:e0138903
- Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME (2009a) Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution* 41:1
- Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet* 6:e1001139
- Hayes BJ, Visscher PM, Goddard ME (2009b) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research* 91:47-60

- He S, Reif JC, Korzun V, Bothe R, Ebmeyer E, Jiang Y (2016a) Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. *Theoretical and Applied Genetics*:1-13
- He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y (2016b) Genomic selection in a commercial winter wheat population. *Theoretical and Applied Genetics* 129:641-651
- He S, Zhao Y, Mette MF, Bothe R, Ebmeyer E, Sharbel TF, Reif JC, Jiang Y (2015) Prospects and limits of marker imputation in quantitative genetic studies in European elite wheat (*Triticum aestivum* L.). *BMC genomics* 16:168
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. *Crop science* 50:1681-1690
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Science* 49:1-12
- Helguera M, Khan I, Kolmer J, Lijavetzky D, Zhong-Qi L, Dubcovsky J (2003) PCR assays for the cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. *Crop Science* 43:1839-1847
- Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for genomic selection applications and research in plants. *Crop Science* 55:1-12
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. *Crop Science* 52:146-160
- Hickey JM, Crossa J, Babu R, de los Campos G (2012) Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Science* 52:654-663
- Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529
- Hozé C, Fouilloux M-N, Venot E, Guillaume F, Dassonneville R, Fritz S, Ducrocq V, Phocas F, Boichard D, Croiseau P (2013) High-density marker imputation accuracy in sixteen French cattle breeds. *Genetics Selection Evolution* 45:1
- Inagaki M, Varughese G, Rajaram S, Van Ginkel M, Mujeeb-Kazi A (1998) Comparison of bread wheat lines selected by doubled haploid, single-seed descent and pedigree selection methods. *Theoretical and applied genetics* 97:550-556
- Isidro J, Jannink J-L, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. *Theoretical and applied genetics* 128:145-158
- Jannink J-L (2010) Dynamics of long-term genomic selection. *Genetics Selection Evolution* 42:1
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics* 9:166-177
- Jia G, Chen P, Qin G, Bai G, Wang X, Wang S, Zhou B, Zhang S, Liu D (2005) QTLs for Fusarium head blight response in a wheat DH population of Wangshuibai/Alondra's'. *Euphytica* 146:183-191
- Jiang Y, Reif JC (2015) Modeling epistasis in genomic selection. *Genetics* 201:759-768
- Jiang Y, Zhao Y, Rodemann B, Plieske J, Kollers S, Korzun V, Ebmeyer E, Argillier O, Hinze M, Ling J (2015) Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). *Heredity* 114:318-326
- Johnson VA, Lay CL (1974) Genetic improvement of plant protein. *Journal of agricultural and food chemistry* 22:558-566
- Kempe K, Gils M (2011) Pollination control technologies for hybrid breeding. *Molecular breeding* 27:417-437
- Knott CA, Van Sanford DA, Souza EJ (2008) Comparison of selection methods for the development of white-seeded lines from red x white soft winter wheat crosses. *Crop science* 48:1807-1816
- Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, Hinze M, Plieske J, Kulosa D, Ganai MW (2013) Genetic architecture of resistance to *Septoria tritici blotch* (*Mycosphaerella graminicola*) in European winter wheat. *Molecular breeding* 32:411-423
- Krenz J, Sackett K, Mundt C (2008) Specificity of incomplete resistance to *Mycosphaerella graminicola* in wheat. *Phytopathology* 98:555-561

- Kumar J, Mir R, Kumar N, Kumar A, Mohan A, Prabhu K, Balyan H, Gupta P (2010) Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. *Plant breeding* 129:617-621
- Kumar P, Yadava R, Gollen B, Kumar S, Verma R, Yadav S (2011) Nutritional contents and medicinal properties of wheat: a review. *Life Sci Med Res* 22:1-10
- Laidig F, Piepho H-P, Drobek T, Meyer U (2014) Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theoretical and Applied Genetics* 127:2599-2617
- Liu G, Zhao Y, Gowda M, Longin CFH, Reif JC, Mette MF (2016) Predicting Hybrid Performances for Quality Traits through Genomic-Assisted Approaches in Central European Wheat. *PLoS One* 11:e0158635
- Löffler M, Schön C-C, Miedaner T (2009) Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. *Molecular Breeding* 23:473-488
- 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. *Theoretical and Applied Genetics* 128:1297-1306
- Longin CFH, Sieber AN, Reif JC (2013) Combining frost tolerance, high grain yield and good pasta quality in durum wheat. *Plant Breeding* 132:353-358
- Longin CFH, Würschum T (2014) Genetic variability, heritability and correlation among agronomic and disease resistance traits in a diversity panel and elite breeding material of spelt wheat. *Plant Breeding* 133:459-464
- Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink J-L, Singh RP, Autrique E, de los Campos G (2015) Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker× Environment Interaction Genomic Selection Model. *G3: Genes| Genomes| Genetics* 5:569
- 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* 3:481-491
- Maccaferri M, Sanguineti MC, Corneti S, Ortega JLA, Salem MB, Bort J, DeAmbrogio E, del Moral LFG, Demontis A, El-Ahmed A (2008) Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics* 178:489-511
- Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding* 15:271-290
- Malla S, Ibrahim AM, Yen Y, Berzonsky W, Glover KD, Stein J (2010) QTL analysis of a putative novel source of resistance to *Fusarium* head blight in hard winter wheat. *Int J Plant Breed* 4:47-54
- Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *International journal of plant genomics* 2012: 728398
- Mesterhazy A (1995) Types and components of resistance to *Fusarium* head blight of wheat. *Plant Breeding* 114:377-386
- Mette MF, Gils M, Longin CFH, Reif JC (2015) Hybrid breeding in wheat. *Advances in Wheat Genetics: From Genome to Field*. Springer, pp 225-232
- Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829
- Miedaner T, Reinbrecht C, Lauber U, Schollenberger M, Geiger H (2001) Effects of genotype and genotype—environment interaction on deoxynivalenol accumulation and resistance to *Fusarium* head blight in rye, triticale, and wheat. *Plant Breeding* 120:97-105
- Miedaner T, Würschum T, Maurer HP, Korzun V, Ebmeyer E, Reif JC (2011) Association mapping for *Fusarium* head blight resistance in European soft winter wheat. *Molecular breeding* 28:647-655
- Miedaner T, Zhao Y, Gowda M, Longin CFH, Korzun V, Ebmeyer E, Kazman E, Reif JC (2013) Genetic architecture of resistance to *Septoria tritici* blotch in European wheat. *BMC genomics* 14:1

- Mirdita V, He S, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y (2015) Potential and limits of whole genome prediction of resistance to *Fusarium* head blight and *Septoria tritici* blotch in a vast Central European elite winter wheat population. *Theoretical and Applied Genetics* 128:2471-2481
- Möhring J, Piepho H-P (2009) Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Science* 49:1977-1988
- Peleg Z, Saranga Y, Suprunova T, Ronin Y, Röder MS, Kilian A, Korol AB, Fahima T (2008) High-density genetic map of durum wheat × wild emmer wheat based on SSR and DArT markers. *Theoretical and Applied Genetics* 117:103-115
- Philipp N, Liu G, Zhao Y, He S, Spiller M, Stiewe G, Pillen K, Reif JC, Li Z (2016) Genomic Prediction of Barley Hybrid Performance. *The Plant Genome* 9
- Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go. *Crop Science* 44:1560-1571
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome* 5:103-113
- Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7:e32253
- Pszczola M, Strabel T, Mulder H, Calus M (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of dairy science* 95:389-400
- Rawat N, Pumphrey MO, Liu S, Zhang X, Tiwari VK, Ando K, Trick HN, Bockus WW, Akhunov E, Anderson JA (2016) Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to *Fusarium* head blight. *Nature Genetics* 48: 1576-1580
- Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8:e66428
- Reif JC, Gowda M, Maurer HP, Longin C, Korzun V, Ebmeyer E, Bothe R, Pietsch C, Würschum T (2011a) Association mapping for quality traits in soft winter wheat. *Theoretical and Applied Genetics* 122:961-970
- Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T, Würschum T (2011b) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. *Theoretical and applied genetics* 123:283-292
- Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theoretical and applied genetics* 126:2835-2848
- Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodriguez VM, Moreno-Gonzalez J, Melchinger A, Bauer E (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
- Rutkoski J, Singh R, Huerta-Espino J, Bhavani S, Poland J, Jannink J-L, Sorrells M (2015) Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. *The Plant Genome* 8
- Sargolzaei M, Chesnais J, Schenkel F (2011) FImpute-An efficient imputation algorithm for dairy cattle populations. *J Dairy Sci* 94:421
- Schulz-Streck T, Ogutu JO, Gordillo A, Karaman Z, Knaak C, Piepho HP (2013) Genomic selection allowing for marker-by-environment interaction. *Plant Breeding* 132:532-538
- Snape JW, Foulkes MJ, Simmonds J, Leverington M, Fish LJ, Wang Y, Ciavarrella M (2007) Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. *Euphytica* 154:401-408
- Solberg T, Sonesson A, Woolliams J (2008) Genomic selection using different marker types and densities. *Journal of animal science* 86:2447-2454
- Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome* 50:557-567
- Somers DJ, Fedak G, Savard M (2003) Molecular mapping of novel genes controlling *Fusarium* head blight resistance and deoxynivalenol accumulation in spring wheat. *Genome* 46:555-564

- Steiner B, Lemmens M, Griesser M, Scholz U, Schondelmaier J, Buerstmayr H (2004) Molecular mapping of resistance to Fusarium head blight in the spring wheat cultivar Frontana. *Theoretical and Applied Genetics* 109:215-224
- Stekhoven DJ, Bühlmann P (2012) MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28:112-118
- Sun C, VanRaden PM (2014) Increasing long-term response by selecting for favorable minor alleles. *PLoS One* 9:e88510
- Sun X, Habier D, Fernando RL, Garrick DJ, Dekkers JC (2011) Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. *BMC proceedings. BioMed Central* 5:1
- Tee TS, Qualset C (1975) Bulk populations in wheat breeding: comparison of single-seed descent and random bulk methods. *Euphytica* 24:393-405
- Thomas W, Forster B, Gertsson B (2003) Doubled haploids in breeding. *Doubled Haploid Production in Crop Plants*. Springer, pp 337-349
- Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* 108:20260-20264
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17:520-525
- VanRaden P (2008) Efficient methods to compute genomic predictions. *Journal of dairy science* 91:4414-4423
- VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of dairy science* 92:16-24
- Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L (2014a) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* 12:787-796
- Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014b) 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* 15:556
- Wei W-H, Hemani G, Haley CS (2014) Detecting epistasis in human complex traits. *Nature Reviews Genetics* 15:722-733
- Whitford R, Fleury D, Reif JC, Garcia M, Okada T, Korzun V, Langridge P (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *Journal of experimental botany* 64:5411-5428
- Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. *Genetical research* 75:249-252
- Winfield MO, Allen AM, BurrIDGE AJ, Barker GL, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G (2015) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant biotechnology journal* 14: 1195-1206
- Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Dekkers JC, Fernando R, Garrick DJ (2016) Mixture models detect large effect QTL better than GBLUP and result in more accurate and persistent predictions. *Journal of animal science and biotechnology* 7:1
- Wright H (1980) Commercial hybrid seed production. *Hybridization of Crop Plants*:161-176
- Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. *Theoretical and Applied Genetics* 125:201-210
- Xu S (2003) Theoretical basis of the Beavis effect. *Genetics* 165:2259-2268
- Yang J, Bai G, Shaner GE (2005a) Novel quantitative trait loci (QTL) for Fusarium head blight resistance in wheat cultivar Chokwang. *Theoretical and Applied Genetics* 111:1571-1579
- Yang Z, Gilbert J, Fedak G, Somers DJ (2005b) Genetic characterization of QTL associated with resistance to Fusarium head blight in a doubled-haploid spring wheat population. *Genome* 48:187-196
- Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012a) Accuracy of genomic selection in European maize elite breeding populations. *Theoretical and Applied Genetics* 124:769-776

- Zhao Y, Gowda M, Longin FH, Würschum T, Ranc N, Reif JC (2012b) Impact of selective genotyping in the training population on accuracy and bias of genomic selection. *Theoretical and Applied Genetics* 125:707-713
- Zhao Y, Gowda M, Würschum T, Longin CFH, Korzun V, Kollers S, Schachschneider R, Zeng J, Fernando R, Dubcovsky J (2013a) Dissecting the genetic architecture of frost tolerance in Central European winter wheat. *Journal of experimental botany* 64:4453-4460
- Zhao Y, Zeng J, Fernando R, Reif JC (2013b) Genomic prediction of hybrid wheat performance. *Crop Science* 53:802-810

Abbreviations

Abbreviation	Explanation
SSD	Single-seed descend
DH	Double haploid
FHB	Fusarium head blight
STB	Septoria tritici blotch
MAS	Marker-assisted selection
QTL	Quantitative trait loci
GS	Genomic selection
RRBLUP	Ridge Regression best linear unbiased prediction
GBLUP	Genomic best linear unbiased prediction
SNP	Single nucleotide polymorphism
BLUE	Best Linear Unbiased Estimate
EGBLUP	Extended genomic best linear unbiased prediction
RKHS	Reproducing kernel Hilbert space

Acknowledgements

Life, to some extent, cannot be planned. You could never fit a linear model to predict the upcoming based on what you learned and experienced before. Since I studied animal breeding for six years, my brain has been solidified within a certain paradigm of an animal breeder. In the regard, I want to immensely express my thanks and appreciations to my supervisor Prof. Dr. Jochen C. Reif for his talented enlightenment and patient guidance, which leads my ways of thinking step-by-step shift from an animal to a plant pattern. Towards the view of research, he always considers issues from a prospect of breeding and emphasizes the importance that research should consistently serve the practice, not just be restricted within the cycle of publishing papers, which makes me distinct of what scientific research really targets. Apart from that, his passion on work and optimistic attitude on life also impress and impact me a lot.

Then I want to express my particular thanks to my outstanding Chinese colleague Dr. Yong Jiang. His incomparable mathematical skill impresses me so much and confirms my realization that every technical issue is in essence a mathematical issue, and how deep you dive into mathematical principles and formulas entailing how distinct regarding issues you could be. He helped me to solve plenty of mathematic- or computation-related issues while contributed a lot of ideas to the papers that we cooperated to publish. He performs as the super brain behind all the work I have done within my PhD period. Next I want to thank another Chinese colleague Dr. Yusheng Zhao. He helped me to form a systematic paradigm on quantitative genetics used in plant breeding in the genomics era and enlightened me how to simplify issues through an associative ways of thinking.

I am also thankful for the helps from my office buddies, Albert Schulthess, Norman Phillip, Carl Beuchel and other colleagues Lars Otto, Guozheng Liu, Fang Liu, and Yu Wang. Without your consistent helps on both work and life, I could not finish my PhD study smoothly. The sponsorship and empirical data support from KWS LOCHOW GmbH are likewise substantial. I am really appreciated that constant economic support and the fantastic job in data preparation from Mr. Reiner Bothe and his colleagues. I really want to give many thanks for that. What is more, family always plays an important role in our life. I appreciate the encouragement and support from my parents and all my relatives living in China. Although we are distant from each other, I could feel the warm of family from you as well. Thank you all!

Curriculum vitae

Personal information

Name	Sang He
Gender	Male
Date of birth	25.06.1989
Place of birth	Shehong, Sichuan, China
Nationality	China
Marital status	Single
Present address	Hans-Stubbe-Str. 18, 06466 Gatersleben, Germany

Education

09/1995 – 06/2001	Jingzhongjie Primary School, Mianyang, Sichuan, China
09/2001 – 06/2004	Shiyan High School, Mianyang, Sichuan, China
09/2004 – 06/2007	Nanshan High School, Mianyang, Sichuan, China
09/2007 – 06/2011	South China Agriculture University, Guangzhou, China
09/2011 – 06/2013	China Agriculture University, Beijing, China
Since 11/2013	Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Working experience

06/2013-11/2013	Beijing dairy cattle center, Beijing, China
-----------------	---

Eidesstattliche Erklärung / *Declaration under Oath*

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

Datum / Date

Unterschrift des Antragstellers / *Signature of the applicant*

Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren /

Declaration concerning Criminal Record and Pending Investigations

Hiermit erkläre ich, dass ich weder vorbestraft bin noch dass gegen mich Ermittlungsverfahren anhängig sind. / *I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.*

Datum / Date

Unterschrift des Antragstellers / *Signature of the applicant*