Aus der

Aus der Landessaatzuchtanstalt

der Universität Hohenheim

Apl. Prof. Dr. Thomas Miedaner

Genome-wide prediction of testcross performance and phenotypic stability for important agronomic and quality traits in elite hybrid rye (*Secale cereale* L.)

Dissertation

zur Erlangung des Grades eines Doktors

der Agrarwissenschaften

vorgelegt

der Fakultät Agrarwissenschaften

von

Yu Wang

aus Liaoning

2016

Die vorliegende Arbeit wurde am 12.01.2016 von der Fakultät Agrarwissenschaften der Universität Hohenheim als "Dissertation zur Erlangung des Grades eines Doktors der Agrarwissenschaften" angenommen.

Tag der mündlichen Prüfung: 22.01.2016

1. Prodekan: Prof. Dr. Jörn Bennewitz

Berichterstatter, 1.Prüfer: apl. Prof. Dr. Thomas Miedaner

Mitberichterstatter, 2. Prüfer: Prof. Dr. Jochen. C. Reif

3. Prüfer: Prof. Dr. Ralf T. Vögele

Table of contents

---

[1] 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 15:556

[2] Wang Y, Mette MF, Miedaner T, Wilde P, Reif JC, Zhao Y (2015).
First insights into the genotype-phenotype map of phenotypic stability in rye.
Journal of Experimental Botany 66:3275-3284

[3] Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif CJ, Zhao Y (2015).
Multi-trait- and selection indices genomic prediction for grain yield and protein content in rye with feeding purpose.
Theoretical and Applied Genetics 129(2):273-87

1. General Introduction

Rye (*Secale cereale* L.) is currently used as food, feed, and for bioenergy production and is primarily grown in Northern, Central, and Eastern Europe with importance in Germany, Poland, Scandinavia, Belorussia, and Russia (Geiger and Miedaner 2009; Miedaner et al. 2014). Rye is a member of the *Triticeae* tribe and is closely related to wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) (Pillay 1995). In comparison to other small grain cereals, rye exhibits extraordinary tolerance to frost (Fowler and Limin 1987), drought, salt, and aluminum stress (Geiger and Miedaner 2009). Moreover, rye is very robust to infertile, sandy, or acid soils, as well as poorly prepared land (Geiger and Miedaner 2009). Rye is therefore a promising component towards sustainable, climate-smart agriculture and presents an interesting model crop to study the genetic architecture of phenotypic stability.

Hybrid breeding of rye

Rye was mainly improved as open-pollinating variety in former times (Orellana and Giraldez 1981). The University of Hohenheim is the cradle of hybrid rye breeding with first hybrid varieties resulting out of these activities released in 1984 (Geiger and Miedaner 1999). Hybrids outperform open-pollinating varieties due to a maximum exploitation of heterosis (Geiger and Miedaner 1999), they exhibit a higher uniformity (Geiger and Miedaner 1999), and facilitate a larger return of investment for seed companies due to the built-in plant variety protection (Longin et al. 2012). These advantages led to a market share of hybrid varieties of around 60% of the total rye acreage in Germany (Tomerius et al. 2008).

Hybrid breeding is based on high yielding heterotic groups (Melchinger and Gumber 1998). A heterotic group can be defined "as a group of related or unrelated genotypes from

the same or different populations, which possesses similar combining ability and displays similar heterotic response when crossed with genotypes from other germplasm groups that are genetically distinct" (Melchinger and Gumber 1998). A heterotic pattern refers to a specific pair of two heterotic groups, which express high heterosis and consequently high hybrid performance in their cross (Melchinger and Gumber 1998). In rye, the *Carsten* and *Petkus* populations were identified as the most promising heterotic pattern (Hepting 1978) and hybrid rye breeding is currently based on these two heterotic groups (Geiger and Miedaner 1999). Both populations are highly complementary which may result from their contrasting original target environments and the selection for different ideotypes (Fischer et al. 2010).

Hybrid seed in rye is commonly developed based on cytoplasmatic-male sterility (CMS) systems (Geiger and Miedaner 2009). CMS systems exploit male sterility induced by the mitochondrial genome which can be restored by nuclear genes. According to Łapiński and Stojałowski (2001), sources of CMS for hybrid rye breeding can be classified into two major groups, the "P" (Pampa) and the "V" (Vavilov) types. Almost all rye hybrids are produced by the P cytoplasm, since the V type is difficult to maintain (Geiger and Miedaner 2009). According to Geiger (1982), most hybrids are crosses between a CMS single cross hybrid and a restorer synthetic. Three requirements must be met: (1) the seed parent must be male sterile across a wide range of environments, (2) provide adequate amounts of high-quality seed, and (3) sufficient pollen fertility must be provided by the restorer synthetic.
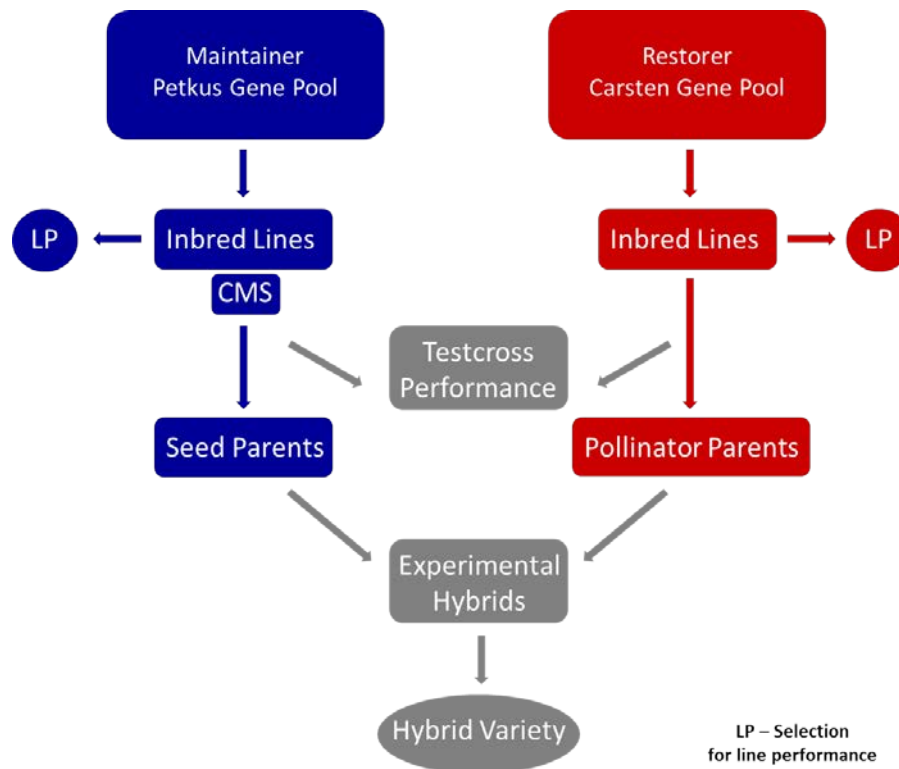
Figure 1 Flow chart for the breeding scheme of hybrid rye (Miedaner et al. 2007)

Hybrid rye breeding is conducted applying multi-stage selection (Tomerius et al. 2008). After generating genetic variation by crossing lines of the same heterotic group, the segregating progenies are fixed through selfing (Fig. 1). Selecting superior inbred lines is implemented on the basis of two successive stages, i.e., selection for line *per se* performance and for combining ability to the respective opposite gene pool (Tomerius et al. 2008). Selection of *per se* performance of lines focus on highly heritable traits such as plant height, thousand-grain weight, falling number, and leaf rust resistance (Tomerius et al. 2008). Evaluation for more complex traits such as grain yield is done at the testcross level mainly due to seed availability and the low covariation between line and testcross performance (Miedaner et al. 2014).

To implement evaluation of testcross performance, the lines of the male pool (*Carsten*) are crossed to a CMS tester from the opposite pool exhibiting high general combining ability.

In contrast, testcrossing of the female lines (*Petkus* pool) is done after the selected S_2 lines are converted to CMS analogues by repeated backcrossing, which is concurrently to the ongoing selfing process (Geiger and Miedaner 2009). Assessment of testcross performance of female lines is often conducted using CMS BC_1-analogues of S_4 lines and CMS BC_2-analogues of S_6 lines crossed with testers from the opposite gene pool. Commercial hybrids derive from the crosses between CMS-seed parents and pollen parents. Restorer synthetics used as pollen parents, consist of two inbred lines, which are crossed by hand and multiplied by random open pollination. Compared with development of seed parent, the line development of pollen parent is less complicated and laborious with inbreeding conducted only till S_2 generation. Finally, selected lines are used for the production of experimental hybrids. Completing one cycle of selection takes 11 years for the seed parent and 9 years for the pollen parent. Consequently, approaches allowing to accelerate selection progress by shortening the lengths of the selection cycles are urgently needed.

Marker-assisted selection

Marker-assisted selection (MAS) has been proposed as one promising approach to enhance the selection gain per time unit. MAS is an indirect selection of individuals, which applies linkage disequilibrium between molecular markers and quantitative trait loci (QTL) (Lande and Thompson 1990). Individuals are selected based on molecular fingerprints without measuring or assessing the phenotype. Consequently, MAS holds the potential to speed up the selection gain per unit time.

MAS involves two stages: (1) Diagnostic markers have to be identified, validated, and their effects have to be estimated; (2) diagnostic markers are used to predict the performance of the population under consideration. MAS is nowadays routinely applied in

many plant breeding programs to improve some specific traits with large effect QTL (Lande and Thompson 1990, Dekkers and Hospital 2002). The success of MAS to improve complex agronomic traits, however, is limited because they are controlled by many genes with rather small effects (Bernardo 2008, Heffner et al. 2009). Consequently, there exists the need to search for alternatives to use genomic data to predict the performance also for complex traits.
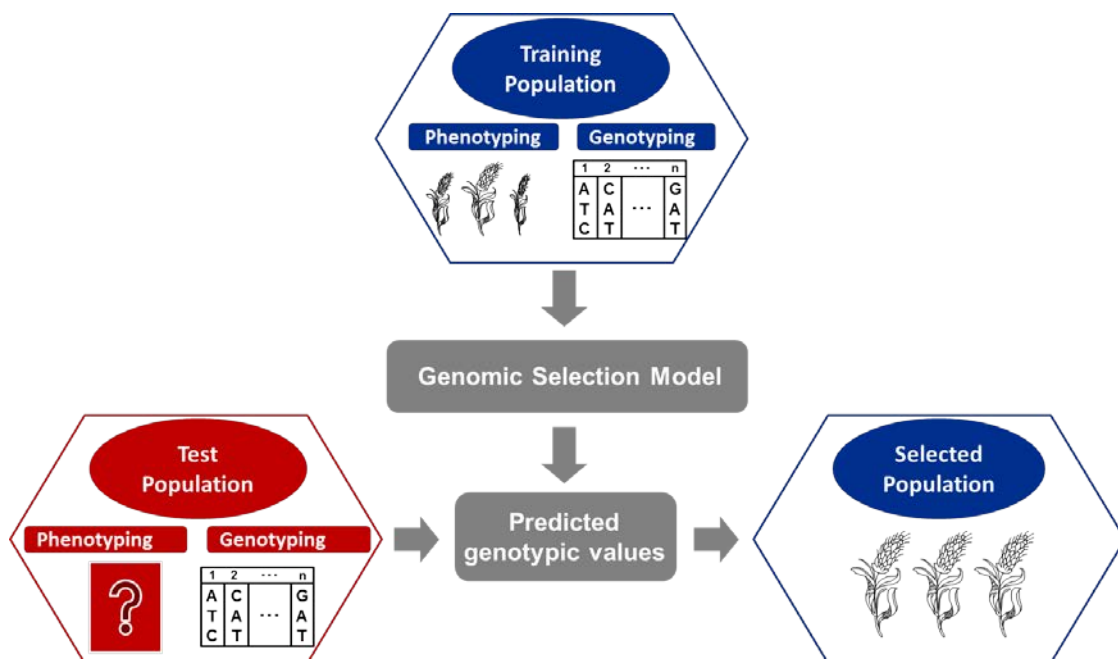
Genomic selection

Genomic selection (GS) is specifically designed to predict complex traits. In GS, effects of all markers across the entire genome are used to estimate the trait performance of genotyped but non-phenotyped individuals (Meuwissen et al. 2001; Fig. 2). The basic principle of GS involves two types of data sets: 1) A training population consisting of individuals, which are genotyped and phenotyped; 2) a test population not being phenotyped. The training population is used to establish a prediction model, which is then applied to predict the trait performance of the test population based on DNA fingerprints.

Before applying GS, assessment of how accurately a predictive model will perform must be obtained. Several studies investigated the potential of GS using experimental data in maize (e.g., Albrecht et al. 2011, Windhausen et al. 2012, Zhao et al. 2012a, Zhao et al. 2012b), wheat (e.g., Heffner et al. 2011a, Zhao et al. 2013), barley (e.g., Heslot et al. 2012, Rutkoski et al. 2012), rice (e.g., Xu et al. 2014, Spindel et al. 2015), and soybean (e.g., Shu et al. 2013, Jarquín et al. 2014). Despite promising magnitudes of accuracies observed for MAS in these crops, potential of GS for rye has not yet been reported.

Accuracy of prediction is affected by the effective population size (Daetwyler et al. 2008), the marker type and density (Chen and Sullivan 2003, Poland and Rife 2012), the

extent and distribution of linkage disequilibrium (Habier et al. 2007), relationship between training and test population (De Roos et al. 2009, Saatchi et al. 2010, Saatchi et al. 2011, Windhausen et al. 2012, Guo et al. 2014, Albrecht et al. 2011, Clark et al. 2011, 2012, Pszczola et al. 2012), choice of GS models (Heslot et al. 2012), and trait heritability (Heffner et al. 2009). Previous studies suggested that one important driving factor of the accuracy for complex traits in population sizes typical for plant breeding programs is the relatedness between training and test population (Albrecht et al. 2011, Gowda et al. 2013). Consequently, when studying the potential of GS this relatedness has to be taken into account.

Figure 2 Principle of genomic selection.

Phenotypic stability

GS focused in many crops mainly on grain yield and its components (e.g., Heslot et al 2012, Poland et al. 2012, Zhao et al. 2013, Xu et al. 2014), biotic and abiotic stress resistances (e.g., Odegard et al. 2009, Rutkoski et al. 2012), and quality traits (e.g., Zhao et al. 2012a, Zhao et

al. 2012b, Spindel et al. 2015). To cope with the consequences due to climate change, however, phenotypic stability including yield stability is becoming of relevance. Phenotypic stability refers to the stability of the performance for a genotype across a range of growing conditions (Becker 1981).

Becker and Léon (1988) distinguished between static and dynamic phenotypic stability. According to the static concept, a stable genotype must show constant trait performance with minimum variance under different environmental conditions. Very often, however, static stability is expected to be associated with low trait performance, which is not desired for instance for grain yield in Central Europe (Mühleisen et al. 2014a). In contrast, in the concept of dynamic phenotypic stability a genotype is considered as stable if it shows low genotype × environment interactions (Lin et al. 1986; Becker and Léon 1988). Thus, the concept of dynamic stability is well suited for grain yield (Mühleisen et al. 2013, Mühleisen et al. 2014a, Mühleisen et al. 2014b).

Only little is known on the genetic basis of phenotypic stability. Recently, heritability estimates have been developed facilitating to examine the quality and the genetic component of phenotypic stability estimates (Mühleisen et al. 2014). The genetic architecture of phenotypic stability has only been examined in one study in barley (Kraakman et al. 2004).

Multi-trait genomic selection

Rye breeders have to improve several traits simultaneously, which is hampered by undesired trait correlations and restricted population sizes due to the budget. For feeding purposes, for instance, grain yield and protein content are negatively correlated, impeding the selection of high-yielding genotypes exhibiting also high protein content (Miedaner et al. 2012). Trait

correlations derive from either pleiotropy due to the same loci controlling different traits (Falconer 1960) or linkage of loci controlling different traits and locating closely on the same chromosome (Mather and Jinks 1971). If the latter is relevant, it is of utmost importance to break undesired trait correlations via fast recombination cycles.

GS is currently focusing on individual traits ignoring correlations among them. GS may benefit from a shift from uni- to multi-trait analyses by (1) improving accuracy of prediction borrowing precision in phenotyping across traits with varying heritability (Calus and Veerkamp 2011, Jia and Jannink 2012, Guo et al. 2014) and (2) deeper understanding of the genetic architecture of trait correlations (Calus and Veerkamp 2011, Jia and Jannink 2012). Moreover, multi-trait selection is a substantial part of the daily work of each commercial breeder. Despite the potential, multi-trait GS models have not been evaluated based on experimental data in crops.

Objectives of this thesis

This study is based on two segregating $F_{3:4}$ rye testcross populations connected through one common parent, each comprising 220 lines, evaluated for grain yield, thousand kernel weight, test weight, falling number, and protein as well as pentosan content in up to 16 environments. The overall goal of the present thesis was to study the prospects and limits of MAS and GS for important quantitative, quality and agronomic traits in rye.

In particular, the objectives were to:

(1) examine the prediction accuracy of MAS and GS within two related bi-parental rye populations,

(2) investigate the influence of relatedness on the prediction accuracies of MAS and GS,

(3) study the impact of establishing training populations for GS phenotyped in multiple years and locations on the prediction accuracy of GS,

(4) evaluate the heritability of static and dynamic phenotypic stability parameters in rye,

(5) investigate the genetic architecture of phenotypic stability by combining linkage mapping and GS,

(6) examine the potential to predict phenotypic stability parameters applying MAS and GS,

(7) explore the benefits of multi-trait GS over single-trait GS for the prediction of grain yield and protein content in rye using different sizes of balanced and unbalanced training datasets concerning these traits, and

(8) develop a pipeline that allows the combination of selection index with GS approaches and its application for grain yield and protein content improvement.

2. The accuracy of prediction of genomic selection in elite hybrid rye population surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years

Yu Wang[1, 2], Michael Florian Mette[1], Thomas Miedaner[2], Marlen Gottwald[3], Peer Wilde[4], Jochen Christoph Reif[1*], and Yusheng Zhao[1]

[1] Leibniz Institute of Plant Breeding and Crop Plant Research (IPK), 06466 Gatersleben, Germany;
[2] State Plant Breeding Institute, University of Hohenheim 70599 Stuttgart, Germany;
[3] Syngenta Agro GmbH, Am Technologiepark 1-5, 63477 Maintal, Germany;
[4] KWS LOCHOW GMBH, Ferdinand-von-Lochow-Str. 5, 29303 Bergen, Germany;

Abstract. Marker-assisted selection (MAS) and genomic selection (GS) based on genome-wide marker data are becoming powerful tools to predict the genotypic value of selection material in plant breeding. However, case-to-case optimization of these approaches is required to achieve maximum accuracy of prediction with reasonable input. Based on extended field evaluation data for grain yield, plant height, starch content and total pentosan content of elite hybrid rye derived from testcrosses involving two bi-parental populations that were genotyped with 1048 molecular markers, we contrasted the accuracy of prediction of MAS and GS in a cross-validation approach. MAS achieved generally lower and in addition potentially over-estimated accuracies of prediction than GS by ridge regression best linear unbiased prediction (RR-BLUP). The grade of relatedness of the plant material included in the estimation and test sets clearly impacted the accuracy of prediction of GS. Within each of the two bi-parental populations, accuracies differed depending on the relatedness of the respective parental lines. Across populations, accuracy increased when both populations contributed to estimation and test set. In contrast, accuracy of prediction based on an estimation set from one population to a test set from the other population was low despite that the two bi-parental segregating populations under scrutiny shared one parental line. Limiting the number of locations or years in field testing reduced the accuracy of prediction of GS equally, supporting the view that to establish robust GS calibration models a sufficient number of test locations is of similar importance as extended testing for more than one year

3. First insights into the genotype-phenotype map of phenotypic stability in rye

Yu Wang[1, 2], Michael Florian Mette[1], Thomas Miedaner[2], Peer Wilde[3], Jochen Christoph Reif[1*], and Yusheng Zhao[1]

[1] Leibniz Institute of Plant Breeding and Crop Plant Research (IPK), 06466 Gatersleben, Germany;
[2] State Plant Breeding Institute, University of Hohenheim 70599 Stuttgart, Germany;
[3] KWS LOCHOW GMBH, Ferdinand-von-Lochow-Str. 5, 29303 Bergen, Germany;

Abstract. Improving phenotypic stability of crops is becoming pivotal to cope with the detrimental impacts of climate change. The main goal of this study was to gain first insights into the genetic architecture of phenotypic stability in cereals. To this end, we determined grain yield, thousand kernel weight, test weight, falling number, and both protein and soluble pentosan content for two large bi-parental rye populations connected through one common parent and grown in multi-environmental field trials involving more than 15,000 yield plots. Based on these extensive phenotypic data, parameters for static and dynamic phenotypic stability of the different traits were calculated and linkage mapping was applied using whole-genome molecular marker profiles. While we observed no large-effect quantitative trait loci (QTLs) underlying yield stability, but detected large and stable QTLs for phenotypic stability of test weight, soluble pentosan content, and falling number. Applying genome-wide selection, which in contrast to marker-assisted selection also takes into account loci with small-effect sizes, considerably increased the accuracy of prediction of phenotypic stability for all traits by exploiting both genetic relatedness and linkage between single-nucleotide polymorphisms and QTLs. We conclude that breeding for crop phenotypic stability can be improved in related populations using genomic selection approaches established upon extensive phenotypic data.

4. Multi-trait- and selection indices genomic prediction for grain yield and protein content in rye with feeding purposes

Albert Wilhelm Schulthess[1#], Yu Wang[1, 2#], Thomas Miedaner[2], Peer Wilde[3], Jochen Christoph Reif[1], and Yusheng Zhao[1*]

[#] These authors equally contributed to this work;

[1] Leibniz Institute of Plant Breeding and Crop Plant Research (IPK), 06466 Gatersleben, Germany;
[2] State Plant Breeding Institute, University of Hohenheim 70599 Stuttgart, Germany;
[3] KWS LOCHOW GMBH, Ferdinand-von-Lochow-Str. 5, 29303 Bergen, Germany;

Abstract. Multiple-trait genomic selection (MTGS) is specially designed to take advantage of genetic correlation between target and indicator traits, leading to improvement in genomic prediction accuracies. Two segregating $F_{3:4}$ rye testcross populations genotyped using diversity array technology markers and evaluated for grain yield (GY) and protein content (PC) were considered. The goals of our study were to investigate the benefits of MTGS over single trait genomic selection (STGS) for GY and PC prediction and to apply GS to predict different selection indices (SIs) for GY and PC improvement. Our results using a two-trait model (2TGS) empirically confirm that the ideal scenario to exploit the benefits of MTGS would be when the predictions of a low heritable target trait with scarce phenotypic records are supported by an intensively phenotyped genetically correlated indicator trait which has higher heritability. This ideal scenario is expected for PC in practice. According to our GS implementation, MTGS can be performed in order to achieve more cycles of selection by unit of time. If the aim is to exclusively improve the prediction accuracy of a scarcely phenotyped trait, 2TGS will be a more accurate approach than a three-trait model which incorporates an additional correlated indicator trait. In general for balanced phenotypic information, we recommend to perform GS considering SIs as single traits, being this method a simple, direct and efficient way of prediction.

5. General Discussion

Assessment of the experimental design of the study

In the present study, two segregating populations, POP-A ($P_1 \times P_2$) and POP-B ($P_1 \times P_3$), with a total number of 440 individuals were derived from crosses between the three elite winter rye inbred lines P_1, P_2, and P_3, connected via one overlapping parent, P_1 (Figure 3). F_1 plants from crosses of parental lines were self-pollinated under isolation bags during two generations to obtain F_3 plants by single seed descent (SSD). The 440 randomly taken $F_{3:4}$ lines were crossed to an unrelated cytoplasmic-male sterile (CMS) single-cross tester of the Petkus gene pool (seed parent pool) by open pollination between isolation walls.

The 440 testcross progenies of two segregating rye populations were evaluated in up to 16 diverse location × year × irrigation level combinations, denoted as environments. High heritability estimates on an entry-mean basis (h^2) in the range from 0.65 to 0.96 were obtained for the nine traits under consideration, resulting from the broad variation observed for genotypic values of the 440 $F_{3:4}$ testcross progenies. Four quality traits were determined using near-infrared reflectance spectroscopy (NIRS), including protein content, starch content, soluble pentosan content, and total pentosan content. Stable and accurate prediction models for two quality traits have been constructed (Table 1). Therefore, indirect measurement of these traits should not impair the efficiency of QTL and GS analyses, leading to solid data for studying the potential of MAS and GS.
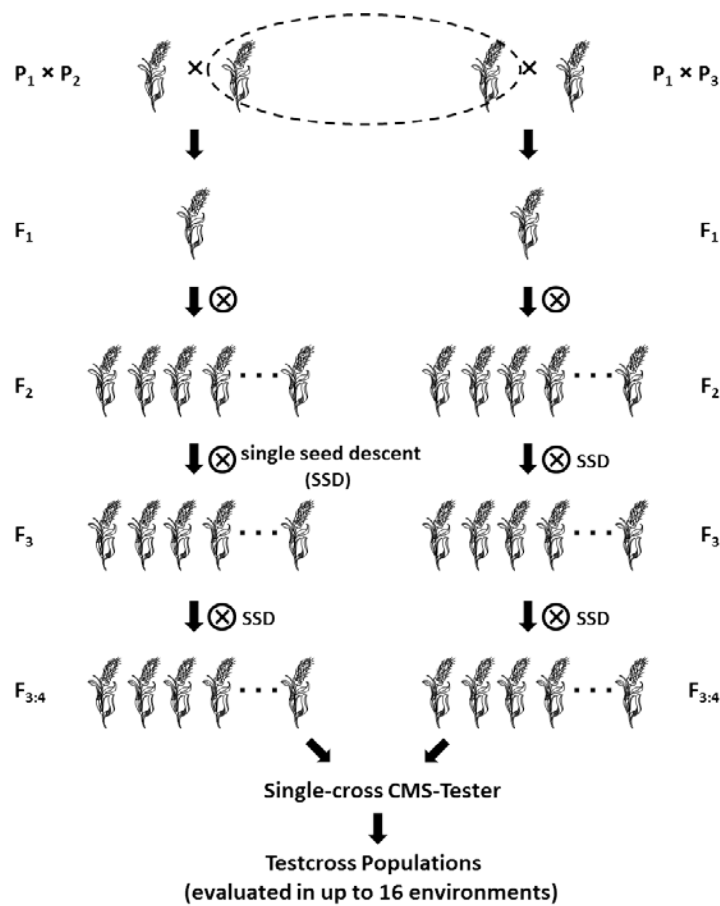
Figure 3 Plant material used in the study.

Table 1. Accuracy of prediction for developed near-infrared reflectance spectroscopy (NIRS) calibrations of protein content (PC, %) and soluble pentosan content (SPC, %)

Model	Calibration				Validation					
	N_C	R_C	R_C^2	SE	N_V	Bias	R_V	R_V^2	SE	SD
PC	330	0.99	0.98	0.23	108	0.01	0.99	0.98	0.31	2.07
SPC	321	0.91	0.82	0.18	107	-0.02	0.86	0.74	0.22	0.43

N_C and N_V denote sample size for calibration and validation, respectively. R_C and R_V refer to correlation coefficient of calibration and validation, respectively. R_C^2 and R_V^2 represent coefficient of determination of calibration and validation, respectively. SE is standard error of calibration of calibration and validation, respectively, and SD denotes the standard deviation within the validation set.

Two hundred-twenty $F_{3:4}$ lines in each segregating population had been genotyped with diversity array technology (DArT) markers. After quality checks to these marker data, 394 and 584 DArT markers were included in analysis for POP-A and POP-B, respectively. The genetic linkage maps possess a total length of 980 cM for POP-A and 2,349 cM for POP-B.

For QTL mapping, we have not implemented approaches allowing to detect QTL across both populations (Wang et al. 2014, Wang et al. 2015) and focused on each biparental population individually. This can lead to a lower QTL detection power (Martínez and Curnow 1992) but reduces problems arising due to ghost QTL (Lander and Botstein 1989, Knapp 1991). The accuracy of the estimates for QTL effect within biparental populations cannot be substantially improved with a decrease in marker density below 10 cM (Darvasi et al. 1993, Piepho 2000). Therefore, marker saturation guarantees a solid basis for the further analysis in MAS and GS within biparental populations with the average genetic distance between adjacent markers of less than 2 cM (Figure 4).
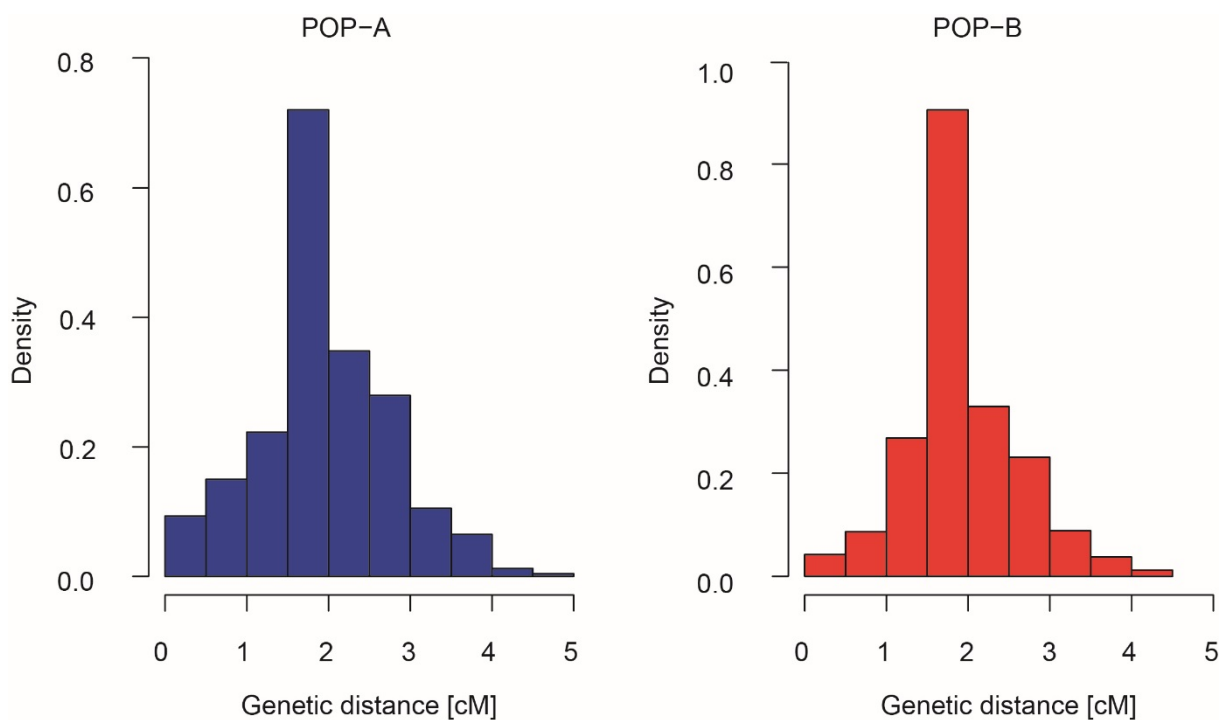
Figure 4 Histogram of the distribution of genetic distance between adjacent markers in population POP-A and POP-B, respectively.

Genomic selection outperformed marker-assisted selection

MAS is a process of indirect selection of individuals, where molecular markers in linkage disequilibrium with quantitative traits locus are applied as genetic determinant (Lande and Thompson 1990). Thus, a genotype is selected based on its marker profile, not on the trait itself. GS differs to MAS by taking a large number of markers (most often all) into account to establish a prediction model. GS is expected to outperform MAS especially for complex traits such as grain yield. We observed in our study an increase in prediction accuracy when applying GS instead of MAS (Wang et al. 2014, Wang et al. 2015). Interestingly, the superiority of GS versus MAS was not always associated with the complexity of the trait, which was estimated by the plot-based heritability. Consequently, there are further factors besides the knowledge of the genetic architecture expected, which drive the prediction accuracy of MAS in biparental populations.

Relatedness severely impacts prediction accuracy not only in genomic but also in marker-assisted selection

GS is heavily promoted as a tool revolutionizing plant breeding (Meuwissen et al. 2001, Jannink et al. 2010). While MAS is expected to profit mainly from linkage disequilibrium between markers and QTL, GS exploits additionally pedigree information in order to predict the trait performance. Nevertheless, recent studies based on factorial crosses (Gowda et al. 2014) and diversity panels (Jiang et al. 2014) pointed to a prominent role of relatedness driving the prediction accuracy also in MAS. We tested the hypothesis that relatedness also influences the prediction accuracy in biparental populations applying a simulation study (Wang et al. 2014). Our results clearly underlined that prediction accuracy of MAS in biparental populations is not only driven by linkage disequilibrium between QTL and functional markers but exploits also relatedness between training and test populations.

Nevertheless, our results also underline that relatedness is more efficiently exploited in GS than in MAS. Consequently, GS is the tool of choice for trait prediction in rye.

Selection for yield stability is challenging but phenotypic stability of quality traits can be improved through genomic selection

So far yield stability and phenotypic stability for quality traits are not directly considered in applied hybrid rye breeding programs. This is mainly due to the large number of environments required to reach a moderate heritability for the stability parameters (Mühleisen et al. 2014b). We explored the potential of GS to predict yield stability and phenotypic stability for quality traits (Wang et al. 2015). We found that it is challenging to predict yield stability but observed moderate to high predicting accuracies for phenotypic stability of quality traits. Consequently, if training populations are available which have been evaluated in extensive field trials, robust GS calibration can be established for phenotypic stability of quality traits in rye. Official variety testing data meets the required phenotyping intensity and is therefore a promising data source to build up GS models for phenotypic stability. Further research is needed to investigate this option in more detail.

Predicting ability of genomic selection across biparental populations

In hybrid breeding programs, genomic selection is conducted not only within but also across biparental populations with the final goal to identify the best performing hybrid in the total population. Windhausen et al. (2012) reported for maize that the prediction ability for performance from one to the other biparental populations was low (e.g., 0.12 for grain yield). We observed in our study also a drop in prediction accuracy when shifting from within population into across population predictions. This is surprising because both populations

were linked by pedigree link through the common parent P_1. In contrast, Riedelsheimer et al. (2013) have shown that using partially related biparental populations allows calibrating GS models also across biparental maize populations. Our study is based on only two populations. Therefore, further research is needed in rye expanding the number of biparental populations to study the potential to predict the performance across different biparental populations.

Prediction accuracy did not increase by applying multi-trait genomic selection

In plant breeding programs, rye breeders have to consider not only one but several traits in parallel, such as yield components, grain quality, and resistance to biotic or abiotic stress factors. These traits are often correlated, either positively or negatively. Therefore, it was hypothesized that prediction accuracy of GS can be improved by using them simultaneously in multi-trait GS models. We tested this hypothesis using the experimental data in rye and concluded that the increase was only marginal in contrast to single-trait GS, indicating that multi-trait GS models were only of interest if traits under consideration are correlated and differ largely in their heritability. For rye, no such a situation could be identified based on the available experimental data. Consequently, multi-trait GS models are only of limited interest for applied rye breeding programs.

Implementing genomic selection in applied hybrid rye breeding programs

In our study, we observed prediction accuracies for grain yield across biparental rye populations ranging from 0.15 to 0.22 (Wang et al. 2014), which failed to achieve a precision of unreplicated field trials at one location (Figure 5). Thus, these prediction accuracies lead to a pessimistic prognosis of the potential to successfully expand the current hybrid rye

breeding schemes considering GS. Changing the training populations towards diversity panels or more diverse partially connected biparental populations can increase the prediction accuracy across different biparental populations (Würschum et al. 2013, Riedelsheimer et al. 2013). Alternatively, more efforts could be concentrated on the design of training sets to improve the accuracy of GS calibration models (Rincent et al. 2012, Akdemir et al. 2015). Akdemir et al. (2015) proposed a dynamic GS calibration model where genotypic information of test set was incorporated when determining training set, indicating that training set by optimized samples leading to improvement of accuracy in contrast to the normally random samples of the same size. However, this dynamic GS model is beneficial only when a subset of genotyped individuals needs to be phenotyped in a breeding program.
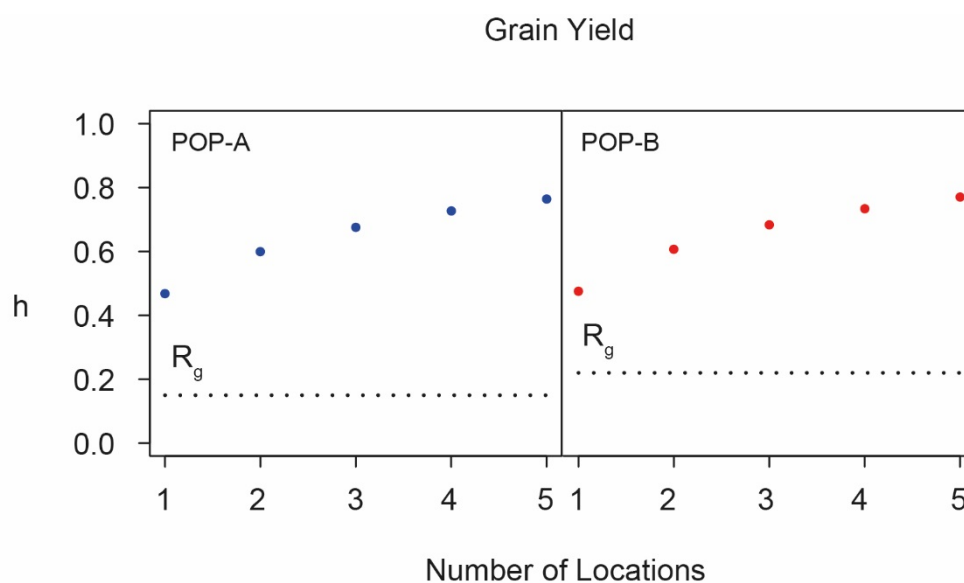
Figure 5 Square root of heritability (h, blue and red points) based on experimental data at different number of locations for grain yield in population A (POP-A) and population B (POP-B), respectively. Five-fold cross-validated standardized accuracies of prediction for grain yield across biparental populations (R_g, black dashed lines) for genomic selection using RR-BLUP across genotypes based on data at five locations over one year.
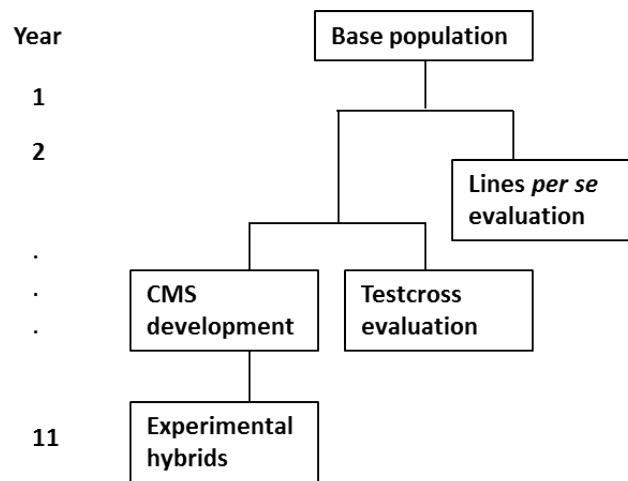
Figure 6 Flow chart breeding scheme for hybrid rye (Tomerius et al. 2008)

If prediction accuracies are increased by using improved training populations, hybrid rye breeding schemes have to be expanded including genomic selection. This could be accomplished by revising the currently devised breeding schemes (e.g., Geiger and Miedaner 1999, 2009). For doing so, optimal allocation of test resources is pivotal. Tomerius et al. (2008) optimized non-GS multi-stage selection schemes in hybrid breeding of rye and suggested that inbred lines are evaluated in single row plot for non-yield traits in unreplicated trails at no more than three locations (Figure 6). Grain yield is evaluated then at the testcross level in multi-location replicated trials using drilled plots in further two stages of selection. The number of proposed locations for first yield test was four. If the corresponding accuracy of GS across biparental populations reaches the level observed for other crops such as maize (Riedelsheimer et al. 2013; Zhao et al. 2012) which is equivalent to field trials at 3-4 locations, then it seems tempting to apply GS immediately after the per se field evaluation.

6. References

Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Akdemir D, Sanchez J, Jannink JL (2015) Optimization of genomic selection training populations with a genetic algorithm. Genetics Selection Evolution (In Press) doi 10.1186/s12711-015-0116-6

Becker HC (1981) Correlations among some statistical measures of phenotypic stability. Euphytica 30:835–840

Becker HC, Leon J (1988) Stability analysis in plant breeding. Plant Breeding 101:1–23

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Science 48:1649–1664

Calus MP, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43:26

Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenomics J 3:77–96

Clark SA, Hickey JM, Daetwyler HD, Van der Werf JHJ (2012) The importance of information on relatives for the prediction of genomic breeding values and implications for the makeup of reference populations in livestock breeding schemes. Genet Sel Evol 44:4

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLos ONE 3:e3395. doi:10.1371/journal.pone.0003395

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Dekkers, JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet 3:22–32

De Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic breeding values across multiple populations. Genetics 183:1545–1553

Falconer DS (1960) An Introduction to Quantitative Genetics, Oliver & Boyd, Edinburgh/London Ronald Press.

Fischer S, Melchinger AE, Korzun V, Wilde P, Schmiedchen B, Möhring J, Piepho HP , Dhillon BS, Würschum T, Reif JC (2010) Molecular marker assisted broadening of the Central aeuropean heterotic groups in rye with Eastern European germplasm. Theor Appl Genet 120:291–299. Doi: 10.1007/s00122-009-1124-0

Fowler DB, Limin AE (1987) Exploitable genetic variability for cold tolerance in commercially grown cereals. Canadian Journal of Plant Science 67:278–278

Geiger HH (1982) Züchtung synthetischer Sorten. III. Einfluß der Vermehrungsgeneration und des Selbstungsanteils. Vortr. Pflanzenzüchtung 1:41–72

Geiger HH, Miedaner T (1999) Hybrid rye and Heterosis. *In* Coors JG and Pandey S (ed.), Genetics and Exploitation of Heterosis in Crops, Crop Science Society, America Madison Wisconsin USA. p. 439–450

Geiger HH, Miedaner T (2009) Rye (Secale cereale L.). *In* Carena MJ (ed.) Cereals. Handbook of Plant Breeding, Volume 3, Springer, US Verlag New York. p. 157–181

Gowda M, Zhao Y, Würschum T, Longin CFH, Miedaner T, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Martinant JP, Mette MF, Reif JC (2013) Relatedness severely impacts accuracy of marker-assisted selection for disease resistance in hybrid wheat. Heredity 112:552–561

Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G (2014) The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127:749–762

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397

Hayashi T, Iwata H (2013) A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC bioinformatics 14:34

Heffner EL, Sorrells ME, Jannink JL (2009) genomic selection for crop improvement. Crop Science 49:1–12

Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME (2011) genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Science 51:2597–2606

Hepting L (1978) Analyse eines 7 x 7-Sortendialles zur Ermittlung geeigneten Ausgangsmaterials für die Hybridzüchtung bei Roggen. Z. Pflanzenzüchtg. 80:188–197

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

Jannink JL, Lorenz AJ, Iwata H (2010) genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Jarquín D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. GMC Genomics 15:740

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

Jiang Y, Zhao Y, Rodemann B, Plieske J, Kollers V, Ebmeyer E, Argillier O, Hinze M, Ling J, Röder MS, Ganal MW, Mette MF and Reif JC (2014) Potential and limits unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.) Heredity doi: 10.1038/hyd.2014.104

Knapp SL (1991) Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny. Theor. Appl. Genet. 81:333–338

Kraakman ATW, Niks RE, Van den Berg PMMM, Piet S, Van Eeuwijk FA. (2004) Linkage isequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168: 435–446

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

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

Łapiński M, Stojałowski S (2001) Occurrence of male sterility-inducing cytoplasm in rye (Secale spp.) populations. Pl. Breed. Seed Sci. 48:7–23

Lin CS, Binns MR, Lefkovitch LP (1986) Stability analysis: where do we stand? Crop science 26:894–900

Longin CFH, Mühleisen J, Maurer HP, Zhang H, Gowda M, Reif JC (2012) Hybrid breeding in autogamous cereals. Theor. Appl. Genet. 125:1087–1096

Longin CFH, Mi X, Mühleisen J (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor. Appl. Genet. DOI 10.1007/s00122-015-2505-1

Martínez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85:480–488

Mather K, Jinks JL (1971) Biometrical Genetics (2nd edn). Chapman and Hall

Melchinger AE, RK Gumber (1998) Overview of heterosis and heterotic groups in agronomic crops. *In* Lamkey KR and Staub JE (ed.) Concepts and breeding of heterosis in crop plants. CSSA Madison WI. p. 29–44

Meuwissen THE, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Miedaner T, Daume C, Mirdita V, Schmiedchen B, Wilde P, Geiger HH (2007) Verminderung von Alkaloiden in der Nahrungskette durch die züchterische Verbesserung der Mutterkornresistenz von Winterroggen. *In* Zikeli S et al. (eds.) Zwischen Tradition und Globalisierung. Beiträge zur 9. Wissenschaftstagung Ökologischer Landbau. Volume 1, Köster, Berlin. P. 225–228

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, 706

Miedaner T, Schwegler DD, Wilde P, Reif JC (2014) Association between line per se and testcross performance for eight agronomic and quality traits in winter rye. Theor. Appl. Genet. 127: 33–41

Mühleisen J, Reif JC, Maurer HP, Möhring J, Piepho HP (2013) Visual scorings of drought stress intensity as covariates for improved variety trial analysis. Journal of Agronomy and Crop Science 199:321–330

Mühleisen J, Piepho HP, Maurer HP, Longin CFH, Reif JC (2014a) Yield stability of hybrids versus lines in wheat, barley, and triticale. Theoretical and Applied Genetics 127:309–316

Mühleisen J, Piepho HP, Maurer HP, Zhao Y, Reif JC (2014b) Exploitation of yield stability in barley. Theoretical and Applied Genetics 127:1949–1962

Odegard J, Yazdi MH, Sonesson AK, Meuwissen TH (2009) Incorporating desirable genetic characteristics from an inferior into a superior population using genomic selection. Genetics 181:737–745

Orellana J, Giraldez R (1981) Metaphase I bound arms and crossing over frequency in rye. I. Open pollinated varieties. Chromosoma (*Berl*.) 84:439–449

Piepho HP (2000) Optimal marker density for interval mapping in a backcross population. Heredity 84:437–440

Pillay M (1995) Chloroplast DNA similarity of smooth bromegrass with other pooid cereals: Implications for Plant Breeding. Crop Sci. 35:869–975

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

Poland J, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Gen 5:92–102

Pszczola M, Strabel T, Mulder HA, Calus PL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci 95:389–400

Riedelsheimer C, Endelmann JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503

Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodrígues VM, Moreno-Gonzalez J, Melchinger A, Bauer E, Schön CC, Meyer N, Giauffret C, Bauland C, Jamin P, Laborde J, Monod H, Flament O, 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 (Z*ea mays* L.) Genetics 192:715–28

Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells M (2012) Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. The Plant Genome 5:51–61

Saatchi M, Miraei-Ashtiani SR, Nejati-Javaremi A, Moradi-Shahrebabak M, Mehrabani-Yeganeh H (2010) The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. Afr J Biotechnol 9:438–442

Saatchi M, Mcclure MC, Mckay SD, Rolf MM, Kim JW et al (2011) Accuracies of genomic breeding values in American Angus beef cattle using k-means clustering for cross-validation. Genet Sel Evol 43:40

Shu Y, Yu D, Wang D, Bai X, Zhu Y, Guo CH (2013) Genomic selection of seed weight based on low-density SCAR marker in soybean. Genet. Mol. Res. 12:2178–2188

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink JL, McCouch SR (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLOS Genetics. DOI:10.137/journal.pgen.1004982

Tomerius AM, Miedaner T, Geiger HH (2008) A model calculation approach towards the optimization of a standard scheme of seed-parent line development in hybrid rye breeding. Plant Breeding 127: 433–440

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 15:556

Wang Y, Mette MF, Miedaner T, Wilde P, Reif JC, Zhao Y (2015) First insights into the genotype–phenotype map of phenotypic stability in rye. Journal of Experimental Botany doi: 10.1093/jxb/erv145

Windhausen VS, Atlin GN, Crossa J, Hickey JM, Grudloyma P, Terekegne A et al (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. Genes Genomes Genet 2:1427–1436 Wricke G (1962) Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. Zeitsshrift für Pflanzenzuechtung 47:92–96

Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. BMC Genetics 14:85

Xu S, Dan Zhu, Zhang Q (2014) Prediction hybrid performance in rice using genomic best linear unbiased prediction. PNAS 111:12456–12461

Zhao Y, Gowda M, Longin CFH, Würschum T, Ranc N, Reif JC (2012a) Impact of selective genotyping in the training population on accuracy and bias of genomic selection. Theor Appl Genet 125:707–13

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012b) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776

Zhao, Y, Zeng J, Fernando R, Reif JC (2013) Genomic prediction of hybrid wheat performance. Crop Science 3: 802–810

7. Summary

Plant breeding is based on intensive phenotyping and subsequent selection. In future, knowledge-based breeding approaches should complement phenotypic selection increasing efficiency and speed. Marker-assisted selection (MAS) exploits linkage disequilibrium between individual molecular markers and quantitative trait loci (QTL). Genomic selection (GS) was suggested for selection of complex, quantitative traits. Here, effects of a large number of markers are estimated in comprehensive training populations which are genotyped and phenotyped. Predictions are then based on the marker profiles of non-phenotyped individuals exploiting marker information across the whole genome. On the potential of genomic selection in winter rye (*Secale cereale* L.) nothing has been published yet. Rye is primarily grown in Northern, Central, and Eastern Europe for use as food, feed and bioenergy production. Rye is robust and tolerant to biotic and abiotic stress factors and can be successfully cultivated on infertile, sandy, or acid soils, as well as poorly prepared land. Consequently, rye is a promising model crop to study phenotypic stability.

The objectives of the present thesis were to (1) examine the prediction accuracy of marker-assisted and genomic selection within two related bi-parental rye populations, (2) investigate the influence of relatedness on the prediction accuracies of marker-assisted and genomic selection, (3) study the impact of establishing training populations for genomic selection phenotyped in multiple years and locations on the prediction accuracy of genomic selection, (4) evaluate the heritability of static and dynamic phenotypic stability parameters in rye, (5) investigate the genetic architecture of phenotypic stability by combining linkage mapping and genomic selection, (6) examine the potential to predict phenotypic stability parameters applying marker-assisted selection and genomic selection, (7) explore the

benefits of multi-trait genomic selection over single-trait genomic selection for the prediction of grain yield and protein content in rye using different sizes of balanced and unbalanced training datasets concerning these traits, and (8) develop a pipeline that allows the combination of selection index with genomic selection approaches and its application for grain yield and protein content improvement.

In this study, 440 testcross families of two biparental rye populations were evaluated in comprehensive field trials for important agronomic traits including grain yield and quality traits. The corresponding $F_{3:4}$ lines were additionally genotyped with a total of 1,048 molecular markers. Genomic selection outperformed marker-assisted selection for all investigated traits. Cross-validated accuracies of prediction of marker-assisted selection, however, were even for complex traits such as grain yield substantial. Applying a simulation study clearly revealed that this can be explained by relatedness of genotypes in both training and test sets, which is not only exploited in genomic but also in marker-assisted selection. Accordingly, we observed a severe drop in accuracies when predicting the performance across unrelated populations. This led to a rather pessimistic prognosis for the potential genomic selection in rye but further research is required on genetically more diverse training populations.

Phenotypic stability has been selected in rye breeding so far indirectly by testing progenies on ecologically divergent locations. In our study, substantially lower estimates of heritability for yield stability were observed compared with grain yield per se: To reach a heritability for yield stability of 0.5 requires large field trials conducted at more than 40 environments. Thus, the low precision of the phenotypic data for yield stability prevented clear conclusions on its genetic architecture. In contrast, heritabilities for phenotypic stability of quality traits were larger than those for yield stability unravelling large and stable

QTL for test weight, soluble pentosan content, and falling number as a measure for pre-harvest sprouting. Applying genomic selection in contrast to marker-assisted selection increased again the accuracies of prediction of phenotypic stability for all studied quality traits.

Breeders have to consider not only one but several traits in parallel. These traits are often correlated, either positively or negatively. Exploiting this information in multi-trait genomic selection models may result in increased accuracies of genomic selection. The analyses of the experimental rye data sets of grain yield and protein content revealed, however, that the increase in prediction accuracy of multi-trait genomic selection models was only marginal in contrast to single-trait approaches. This is mainly due to the lack of correlated traits which differ largely in their heritability. Consequently, rye breeders can stick to the simple to handle and computational fast single-trait genomic selection models.

In conclusion, genomic selection offers a greater potential for improving complex, quantitative traits in winter rye than marker-assisted selection. Prediction accuracies for grain yield for unrelated test populations have, however, to be improved. Nevertheless, they are already favorable for selecting phenotypic stability of quality traits.

8. Zusammenfassung

Die Pflanzenzüchtung beruht bisher auf der umfangreichen Phänotypisierung von Merkmalen und der nachfolgenden Selektion. Durch wissensbasierte Techniken sollten in Zukunft Effizienz und Geschwindigkeit der Zuchtverfahren gesteigert werden. Die markergestützte Selektion (MAS, marker-assisted selection) nutzt dazu das Kopplungsungleichgewicht zwischen einzelnen molekularen Markern und komplexen Merkmalsloci (QTL, quantitative trait loci), um die Leistung eines Genotyps vorherzusagen. Für hochkomplexe Merkmale wurde die genomische Selektion (GS) vorgeschlagen, bei der die Effekte einer großen Zahl von Markern und ihre phänotypischen Werte in umfangreichen Trainingspopulationen ermittelt werden. Der genomische Zuchtwert wird dann auf der Basis biometrischer Modelle für nicht-phänotypisierte Genotypen in einer Testpopulation vorhergesagt, indem die Markerinformation des gesamten Genoms gleichzeitig genutzt wird. Entscheidend für den Erfolg von GS ist ihre Vorhersagegenauigkeit (prediction accuracy), also der Zusammenhang zwischen dem genomischen Zuchtwert und dem wahren Wert der Testpopulation, der anhand einer umfangreichen Phänotypisierung anschließend ermittelt wird. Das Potenzial der GS wurde bisher bei Winterroggen (*Secale cereale* L.) nicht erforscht. Winterroggen ist vor allem in Nord-, Mittel- und Osteuropa verbreitet und dient hier als Nahrungsmittel, Tierfutter und zu Herstellung von Bioenergie. Er ist robust, tolerant für biotische und abiotische Stressfaktoren und kann deshalb auch auf leichten, sauren oder trockenen Böden erfolgreich angebaut werden. Daher ist Roggen ein gutes Modell, um phänotypische Stabilität zu erforschen.

Die Ziele der vorliegenden Arbeit waren, (1) die Vorhersagegenauigkeit von MAS und GS anhand zweier verwandter, biparentaler Roggenpopulationen zu vergleichen, (2) die

Effekte der Verwandtschaft auf die Vorhersagegenauigkeit beider Verfahren zu ermitteln, (3)

den Einfluss einer Phänotypisierung der Trainingspopulation über mehrere Orte und Jahre

auf die Vorhersagegenauigkeit zu prüfen, (4) die Heritabilität der statischen und

dynamischen phänotypischen Stabilität bei Roggen zu erforschen, (5) die genetische

Architektur der phänotypischen Stabilität durch die Kombination von QTL-Kartierung und GS

zu untersuchen, (6) die phänotypische Stabilität mittels MAS bzw. GS vorherzusagen, (7) die

Chancen einer GS aufgrund von Merkmalskombinationen mit derjenigen von nur einem

Merkmal zu vergleichen und schließlich (8) ein Verfahren für eine GS auf der Basis von

Selektionsindices zu entwickeln und anhand der beiden korrelierten Merkmale Kornertrag

und Proteingehalt zu erproben.

Zu Erreichung dieser Ziele dienten 440 Testkreuzungen von zwei biparentalen

Roggenpopulationen, die in mehrortigen und mehrjährigen Feldversuchen für acht wichtige

Merkmale, einschließlich Kornertrag und Qualitätsmerkmalen, phänotypisiert wurden. Die

korrespondierenden $F_{3:4}$-Linien wurden mit insgesamt 1.048 molekularen Markern

genotypisiert. GS war bei allen untersuchten Merkmalen der MAS überlegen. Die

kreuzvalidierten Vorhersagegenauigkeiten der MAS waren allerdings auch für komplexe

Merkmale wie Kornertrag bedeutend. Eine Simulationsstudie zeigte, dass dieses Ergebnis

durch die Verwandtschaft der Genotypen von Trainings- und Testpopulation bedingt war,

was nicht nur die GS, sondern auch die MAS positiv beeinflusst. Dazu passend fanden wir

einen erheblichen Abfall in der Genauigkeit, wenn die Leistung von mit der

Trainingspopulation nicht-verwandten Genotypen anhand ihres genomischen Zuchtwerts

vorhergesagt wurde. Dies lässt für die Anwendung der GS in der praktischen

Roggenzüchtung derzeit keine optimistische Prognose zu. Allerdings sollte in Zukunft der

Einfluss von weiteren, genetisch diverseren Trainingspopulationen untersucht werden.

Die phänotypische Stabilität wurde bisher in der Roggenzüchtung nur indirekt durch die Phänotypisierung in einer möglichst großen Zahl von ökologisch diversen Umwelten selektiert. Die Heritabilitätsschätzwerte für Ertragsstabilität waren jedoch deutlich geringer als diejenigen für den Kornertrag selbst. Modellrechnungen zeigten, dass mehr als 40 Umwelten nötig sind, um auch nur eine moderate Heritabilität für Ertragsstabilität von 0,5 zu erreichen. Dagegen fanden sich höhere Heritabilitäten für die phänotypische Stabilität von Qualitätsmerkmalen. So wurden für die Stabilität von Hektolitergewicht, löslichem Pentosangehalt und Fallzahl als Maß für die Auswuchsfestigkeit große und stabile QTL detektiert. Die Anwendung von GS führte auch bei der phänotypischen Stabilität der Qualitätsmerkmale zu einer höheren Vorhersagegenauigkeit als die MAS.

In der Züchtung werden üblicherweise zahlreiche Merkmale gleichzeitig selektiert. Diese sind oft korreliert, sei es positiv oder negativ. Die Nutzung solcher Korrelationen durch genomische Selektionsmodelle, die mehrere Merkmale enthalten, könnte die Vorhersagegenauigkeit der GS erhöhen. Unsere Untersuchung von Kornertrag und Porteingehalt zeigte jedoch, dass der Unterschied zu GS-Modellen mit nur einem Merkmal marginal war. Ursache war das Nichtvorhandensein von korrelierten Merkmalen, die stark unterschiedliche Heritabilitäten aufwiesen. Somit können Roggenzüchter weiterhin die GS-Modelle basierend auf einem Merkmal erfolgreich nutzen.

Zusammenfassend zeigte die GS ein größeres Potential für die Verbesserung von komplexen, quantitativen Merkmalen in Winterroggen als die MAS. Allerdings müssen die Vorhersagegenauigkeiten für nicht-verwandte Genotypen noch deutlich verbessert werden. Sie sind jetzt schon vorteilhaft für die Vorhersage der phänotypischen Stabilität von Qualitätsmerkmalen.

9. Publications and presentations

Publications (with peer review)

Wang Y, Li H, Zhang L, Lü W, Wang J (2012) On the use of mathematically derived traits in QTL mapping. Mol. Breeding 29: 661–673

Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of GS 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

Wang Y, Mette MF, Miedaner T, Wilde P, Reif C, Zhao Y (2014) First insights into the genotype-phenotype map of phenotypic stability in rye. Journal of Experimental Botany 66:3275-3284

Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif CJ, Zhao Y (2015) Multi-trait- and selection indices genomic prediction for grain yield and protein content in rye with feeding purposes. Theoretical and Applied Genetics 129(2):273-87

Presentations

Wang Y (2015) Genome-wide prediction of testcross performance and phenotypic stability for important agronomic and quality traits in elite hybrid rye. International Conference on Rye Breeding and Genetics, Wrocław, Poland, 6.24.2015

10. Acknowledgments

I would like to thank Prof. Dr. Thomas Miedaner and Prof. Dr. Jochen C. Reif that they provided me with lots of supervision and support me during the period of my writing all publications and doctoral thesis, including, statistical explanations, linguistic corrections and improvements, structuring contents and improving comprehensibility as well as references to appropriate literature.

Prof. Dr. Jochen C. Reif is more than an excellent superior academically, who providing me the opportunity to complete my PhD thesis at University of Hohenheim and the experience in working in IPK, and he is a perfect superior mentally, a friend, too, and I was my greatest honor to work with him.

As a foreign student, I would like to show my greatest respect to my advisor Prof. Dr. Thomas Miedaner for his intellectual guidance, and for his warm and constant encouragement during the process of my PhD.

Many thanks for Prof. Dr. Hans Peter Piepho and Prof. Dr. Karl Schmid. I benefited greatly from their lecturers and studying materials of statistical courses. Also their staffs' statistical consultancy helped me to understand and apply statistical procedures.

Sincere thanks to Prof Dr. Gerd Weber for his patient support in Biology and Biotechnology learning. Furthermore, I thank Prof. Dr. H. Friedrich Utz for his support in using the statistical software PLABQTL as well as for statistical consultancy.

Further thanks to Dr. Yusheng Zhao for his support in using the statistical software R as well as for discussions about statistics. Many thanks go to Dr. Michael Florian Mette for precise and perfect linguistic corrections and improvements of my publications and my presentations.

Last but not least, I would like to express my great thanks to numerous members of the University of Hohenheim and IPK who supported me in my work and afforded me a feeling like home. It would go beyond the scope of these acknowledgements to name all persons and their specific contribution.

11. Curriculum vitae

Name	Yu Wang	
Birth	03 May 1984 in Liaoning, PR China	
Marital status	Single	
Employment record	09/2012 – 10/2015	Doctorate candidate University of Hohenheim, Stuttgart
	02/2011 – 07/2012	Research assistant, Chinese Academy of Agricultural Science, Being, PR China
University education	09/2007 – 07/2010	Studies of Plant Genetics and Breeding, Shenyang Agricultural University, Shenyang, Master of Science
	09/2003 – 07/2007	Studies of Agronomy, Shenyang Agricultural University, Shenyang, Bachelor of Science
School education	09/2000 – 07/2003	High school Shenyang Foreign Language School
	09/1997 – 07/2000	Secondary school Shenyang Foreign Language School
	09/1990 – 07/1997	Elementary school The No. 1 Primary School of Sijing Street