MSc Thesis Report

Analysis of the Efficiency of Genomic Selection versus Phenotypic Selection in a Hybrid Rye Breeding Program

WAGENINGEN UNIVERSITY
MSc Plant Sciences
Laboratory of Plant Breeding
Supervision:
Andrés Gordillo
Chris Maliepaard
Roeland Voorrips

Carola López Durán
840726529110
Wageningen, 2015
Acknowledgments

I would like to thank my supervisors Chris Maliepaard, Roeland Voorrips and Andrés Gordillo, for their guidance, patience, useful comments, remarks and engagement through this project.

I am also grateful to KWS SAAT AG, for the financial support throughout my MSc studies at Wageningen University and, especially, to Dorothea Borchardt for introducing me to “the breeders’ world” and for believing in my ability to become part of it.

Finally, I would like to thank my family and friends that have always supported my dreams and aspirations. When I talk about my family, I mainly refer to my mother, my father (although I don’t have his last name) and my grandma, they helped me by keeping me harmonious during this entire process. I will be grateful forever for your love.
Abstract

Genomic selection (GS) based on RR-BLUP and Random Forest (RF) was used to predict three rye traits [(grain dry yield (GDY), plant height (PH), and thousand kernel weight (TKW)] across different selection cycles. Six different training sets displaying a parent-offspring and second-degree relationships with the prediction set were compared. The training sets differed regarding their size and relatedness with the prediction (validation) set. The training populations were composed of elite lines of a hybrid rye breeding program at KWS LOCHOW GmbH tested for their general combining ability between 2006 and 2009 at different selection stages. Elite lines were intermated in 2009 to develop new segregating S_{2}-line (testcross) populations evaluated for their general combining ability in yield trials in 2013 and 2014 (validation set). Genomic prediction abilities (correlation between genomic estimated breeding values and observed phenotypes) were compared with those of mid-parent values (MPV). In all comparisons for GDY and PH, RR-BLUP performed better than RF. Prediction abilities using RR-BLUP for GDY were low to moderate and were in the range of previous publications (0.11-0.35). Prediction abilities using RR-BLUP for PH and TKW were higher (0.38-0.80 and 0.20-0.74, respectively). Excluding all parental lines in the training set so that predictions are based only on second-degree relatives, reduce prediction abilities considerably for GDY (from 0.35 to 0.19) and only slightly for PH (from 0.53 to 0.47) and TKW (from 0.44 to 0.36). When using predicted lines as training set, prediction abilities increased by more than 60% compared to the use of first- and second-degree relatives in the training set. The advantage of the best GS method (RR-BLUP) with respect to the phenotypic prediction by MPV is low (6%, 13%, and 15% for GDY, TKW, and PH, respectively). All results of this study indicate that genetic relatedness is essential in GS for GDY. More research is needed to design optimal training populations, with a sufficient level of relatedness to the validation set to achieve a good accuracy and to compare the gain in selection with the costs of each selection method.
Contents

Acknowledgments ........................................................................................................ ii
Abstract ....................................................................................................................... iii

1. Introduction ........................................................................................................... 1
   1.1. Rye hybrid breeding ....................................................................................... 1
   1.2. Phenotypic Selection .................................................................................... 2
       1.2.1. Heritability and Selection Response ..................................................... 3
   1.3. The use of marker information ..................................................................... 3
   1.4. Genomic Selection ....................................................................................... 4
       1.4.1. Statistical methods in Genomic Selection .............................................. 5
           1.4.1.1. RR-BLUP ................................................................................... 5
           1.4.1.2. Bayesian Methods ...................................................................... 6
           1.4.1.3. Random Forest .......................................................................... 7
   1.5. Aim of the project ......................................................................................... 7

2. Material and Methods .......................................................................................... 9
   2.1. Rye hybrid breeding program ...................................................................... 9
       2.1.1. Description of phenotypic data ........................................................... 9
       2.1.2. Description of genotypic data .............................................................. 10
   2.2. Phenotypic Selection ................................................................................... 10
       2.2.1. Line per se selection .......................................................................... 10
       2.2.2. Mid-parent value selection ................................................................ 10
       2.2.3. Calculation of heritability .................................................................. 10
   2.3. Genomic Selection ....................................................................................... 11
       2.3.1. Description of different scenarios ........................................................ 11
       2.3.2. Statistical methods for genomic prediction .......................................... 12
           2.3.2.1. RR-BLUP .................................................................................. 12
           2.3.2.2. Random Forest .......................................................................... 13
       2.3.3. Correlation among markers. LD decay estimation ............................. 13
       2.3.4. Population structure. Cluster analysis ................................................. 13
   2.4. Comparison of GS and phenotypic selection for three different traits ........ 13

3. Results .................................................................................................................. 15
   3.1. Prediction ability of phenotypic values of lines and parents ....................... 15
   3.2. Prediction ability of genome estimated breeding values .............................. 15
   3.3. Additional results ......................................................................................... 21
3.3.1. Population structure analysis .......................................................... 21
3.3.2. Linkage disequilibrium analysis ...................................................... 22
3.3.3. Comparison of selected fractions from phenotypic selection and different GS sets .......................................................... 23

4. Discussion ........................................................................................................ 24
4.1. Comparison of traits ................................................................................. 24
4.2. Comparison of Phenotypic and Genomic Selection ..................................... 25
4.3. Comparison of different scenarios in Genomic Selection .......................... 26
4.4. Comparison of statistical methods ............................................................. 27
4.5. Further research ....................................................................................... 28

5. Conclusions ..................................................................................................... 30

6. References ....................................................................................................... 31
1. Introduction

1.1. Rye hybrid breeding

Rye (Secale cereale L.) is an important cereal crop grown primarily in regions of Eastern, Central and Northern Europe. The grain is used mainly for bread making and feed, but also for alcohol production and bioenergy. Rye is closely related to wheat (Triticum aestivum L.) and barley (Hordeum vulgare L.), but in contrast with these crops, it is an outbreeding species in which selfing is naturally prevented by a gametophytic self-incompatibility mechanism (Geiger and Miedaner, 2009).

Rye has a highly repetitive diploid genome which is one of the largest among cereal crops (Haseneyer et al., 2011), with a size of more than 8 Gbp and seven pairs of chromosomes (2n = 2x = 14) (Martis et al., 2013). Genetic and genomic resources are limited compared to other Triticeae (Haseneyer et al., 2011). Nevertheless, an advance in the molecular toolbox for rye last years now leads the way to quantitative genetics and genomics research and breeding.

Hybrid rye breeding started in 1970 at the University of Hohenheim in Germany, as an alternative to open-pollinated varieties, based on self-fertile gene pools and cytoplasmatic-genic male sterility (CMS) used as hybridizing system (Geiger and Miedaner, 1999). The development of seed parent lines is a long procedure where the inbred lines have to be converted into CMS analogues by repeating backcrossing. In commercial hybrid breeding programs parent line development is exclusively done by selfing. This is possible because of the overcome of natural self-incompatibility of rye by self-fertility genes that are easily transferred into the parents by backcrossing and certain degree of inbreeding tolerance. However, during the course of inbreeding, trait such as grain yield and yield components suffer from considerable inbreeding depression. This decrease in yield is surpassed by heterosis, the genetic structure of a hybrid usually corresponds to that of a double or triple cross (Geiger and Miedaner, 2009).

Genetically distant pools are used in hybrid breeding to exploit heterosis by crossing them. There are two main heterotic pools described for rye called “Carsten” and “Petkus”. These pools are generally chosen as starting material for the development of pollinator and seed lines, respectively (Geiger and Miedaner, 2009). Carsten pool is characterized by large spikes with excellent seed setting, but low yield and extremely low lodging and sprouting resistance. Petkus pool has high yield, better tillering ability, good kernel development and higher lodging and sprouting resistance.

In 1984, the University of Hohenheim released the first hybrid cultivars, being supported by private breeding companies that initiated soon their own hybrid programs. Grain yield is the economically most important trait in hybrid rye. Other key agronomic traits are plant height, lodging resistance, thousand-kernel weight (TKW), and sprouting resistance and resistances to leaf and stem rust (Puccinia recondita, P. graminis f.sp. secalis), ergot (Claviceps purpurea), and Fusarium diseases (Geiger and Miedaner, 2009). Depending on the use of rye, end user quality traits such as pentosan, protein and starch content may play different roles. For instance, for baking quality, high starch and pentosan content is desirable together with low protein
content. For feeding purposes the opposite is required, protein content is maximized and pentosan content minimized. For ethanol production, breeders focus on high starch content as main trait (Miedaner et al., 2012).

During the past 20 years, the importance of hybrid rye has progressively increased, reaching larger yield than open-pollinated rye (Geiger and Miedaner, 2009). Currently, the main rye producing countries are Germany, Poland, Russian Federation and Belarus. Highest grain yields are obtained under intensive growing conditions in Germany, where rye grain yield was about 3.5 million tonnes during 2013 with an average of 5.5 tonnes per hectare for the same year (FAO, 2015). In general, hybrid rye yield surpasses population varieties yield by 15-20%, due to major heterotic effects for grain yield (Geiger and Miedaner, 2009). Shorter plant stature, lodging resistance and bread making quality are other traits that show also the superiority of hybrids over population varieties.

The development of alternative breeding strategies to increase the selection gain per year is a continuous challenge for plant breeders (Tomerius and Geiger, 2000). Estimation of GCA of the parents for various traits is used to create promising rye hybrids, GCA being more important than SCA for this crop (Geiger and Miedaner, 2009). Different methods have been proposed to predict and select based on the breeding value of individual lines: (1) the traditional approach based on phenotypic data only and (2) approaches exploiting the same phenotypic data but adding information from molecular markers. Marker assisted selection (MAS) belongs to the second group and it has been implemented for rye improvement. MAS provides a powerful tool to predict genotypic value of material under selection. However, its utility is limited because genotypic values of individuals are predicted based on large effects of a few selected markers. It is known that important traits including yield, are affected by large number of small-effect genes or by a combination of major and minor genes. For such traits, instead of using only a few preselected markers as in MAS, genomic selection (GS) might be a better alternative that considers all marker effects without significance test (Wang et al., 2014).

1.2. Phenotypic selection

In phenotypic selection for hybrids there are three procedures to select promising parental lines: the first one based on general combining ability (GCA), a second one based on the line per se performance, and the last one based on best linear unbiased prediction (BLUP) (Zhao et al., 2014).

GCA is defined as the average phenotypic performance of a particular inbred line in a series of hybrid combinations, therefore is an essential evaluation criterion to select suitable inbred lines that are used as parents in hybrid breeding programs. For the method based on line per se performance, that consists on selection of lines evaluating the behaviour of the inbred parent line itself without prior testcrossing, accuracy of the estimation depends on the association between hybrid performance and its mid-parent value (MPV) (Gowda et al., 2010). MPV is the mean of the values of a quantitative phenotype for two specific parents. In rye breeding, selection per se is not relevant, but the MPV can be used to predict the gain criterion, GCA.
Phenotypic data of individuals and their relatives have been used to calculate the estimated breeding values (EBVs) for quantitative traits. Estimation of genotypic values is usually at the heart of any breeding effort and the analysis of phenotypic data from plant breeding trials can be based on linear mixed models with the following formula (Piepho et al., 2008):

$$ y = X\beta + Zu + e \quad (1) $$

where $y$ is the vector of observations, $\beta$ is a vector of fixed effects, $u$ is a vector of random effects with mean $E(u)=0$ and variance $var(u)=G$, $X$ and $Z$ are the associated design matrices, and $e$ is a random residual vector with mean $E(e)=0$ and variance $var(e)=R$. Best Linear Unbiased Estimation (BLUE) can be used to estimate the fixed effects, while random effects are estimated by Best Linear Unbiased Prediction (BLUP). In practice, the variance components are replaced by their estimates that are preferably obtained by Restricted Maximum Likelihood (REML). BLUE and BLUP may be computed by solving the Mixed Model Equations (MME) (Henderson, 1986):

$$ \begin{bmatrix} X' \cdot R^{-1} \cdot X & X' \cdot R^{-1} \cdot Z \\ Z' \cdot R^{-1} \cdot X & Z' \cdot R^{-1} \cdot Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X' \cdot R^{-1} \\ Z' \cdot R^{-1} \end{bmatrix} \cdot y \quad (2) $$

Often in variety testing and the development of new varieties, genotype effects are considered as fixed and thus become part of $\beta$ in the mixed model. When comparison of genotypes is the main emphasis of the analysis, genotypes should be taken as fixed. BLUEs are used, for instance, in phenotypic data analysis for hybrid rye (Wang et al., 2014). In contrast, when genotypes are regarded as random, genotypic effects become part of $u$ and thus are estimated by BLUP and the fixed part of the model $X\beta$ (environments etc.) is estimated by BLUEs. Shrinkage towards the mean is one major property of BLUP that increases accuracy. BLUP permits accurate estimation of genetic variance components and, furthermore, it can maximize the correlation of true genotypic values and predicted genotypic values (Piepho et al., 2008). Under this approach, genotypes are considered independent, and pedigree information is not exploited. The shrinkage property is responsible for the gain in accuracy compared to BLUE. The use of EBVs via BLUP has been very popular in animal breeding and recently has been also used by plant breeders (Piepho et al., 2008; Heffner et al., 2009). An interesting improvement of BLUP is the use of pedigree information to model and exploit genetic correlation among relatives, called P-BLUP. Exploitation of the pedigree information is based on the relationship matrix computed from the coefficient of coancestry between the different genotypes.

### 1.2.1. Heritability and Selection Response

Phenotypic variance of a population can be divided into genetic and environmental variance, and the genetic variance can be further partitioned into additive, dominance and epistatic variances. The additive variance is the one that breeders can use to predict the performance of the progeny in next generations. Thus, the part of the phenotypic value that is due to additive effects is called breeding value. Usually, the phenotype is used as a selection criterion rather than the breeding value, predicting the response to selection with the heritability.
Wide sense heritability is defined as the proportion of total phenotypic variation among individuals that is due to genetic differences and narrow sense heritability is defined as the proportion of total phenotypic variation among individuals that is due to the genetic variance of additive genetic effects. Thus, low heritability traits present a high variance of dominant and epistatic genetic effects, environmental factors, and or G × E interactions. The higher the heritability, the larger the progress may be in phenotypic selection. This progress is called selection response R (difference between the average phenotypic value measured in the progeny from the selected population and the average phenotypic value of the same trait measured in the original population). One of the main uses of heritability is to predict the response to selection R considered per unit time, defined in the breeders’ equation as:

\[ R = h^2 S \]

where S is the selection differential (difference between the average phenotypic value of the population and the average of the phenotypic values of the selected individuals) (Piepho and Mühling, 2007). Thereby, the selection response R may be predicted from a known heritability \( h^2 \) and the known selection differential S. R depends also on the selection intensity, it is assumed that, in a large population (little or no genetic drift), severe selection will give a better selection response than a mild selection. The higher the heritability the closer R approaches S.

1.3. The use of marker information

Marker-based approaches opened a broad range of possibilities in plant breeding like identification of specific genes or the use of MAS. The use of marker information provides also a logical extension of the BLUP method based only on phenotypic data (Zhao et al., 2014). This new method was called Genomic-Best Linear Unbiased Prediction (G-BLUP) and it estimates the relationship matrix by the molecular markers information (Bernardo, 1998). G-BLUP differs from the traditional P-BLUP in the replacement of the pedigree relationship matrix with a genomic relationship matrix (Habier et al., 2013). In contrast to P-BLUP, molecular markers can, in principle, also capture the random genetic variation in genetic distance between progeny from the same cross combination due to Mendelian Sampling (similarity of different pairs of full sibs may differ based on their molecular profile even when they share the same pedigree). This method may be exploited in genomic selection applications.

1.4. Genomic selection

GS is a selection method that was first introduced by Meuwissen and colleagues in 2001 and it has high potential to improve selection gain per unit time in a hybrid breeding program. GS is based on the prediction of the breeding value from genome-wide marker data and it could improve breeding efficiency by an accurate prediction of the hybrid performance from their inbred parents.

In a GS breeding scheme, a large number of DNA markers in the entire genome are used to calculate genome estimated breeding values (GEBVs) from individuals under selection in a “breeding population”. All marker effects are included in the model, even those with small effects. These GEBVs are estimated using a model that was “trained” from individuals having
both phenotypic and genotypic data and which population is called the “training population” and is related to the population under selection. The basic steps can be summarized in: i) phenotype and genotype a training population, ii) build a GS model estimating regression coefficients for all markers and iii) based on this model, calculate the Genomic Estimated Breeding Values (GEBVs) in a validation population and select without phenotyping in the following generations.

The advantage of GS over traditional breeding methods may be an increase in prediction accuracy. Gain from selection during GS is proportional to GEBV accuracy. GS can reduce breeding time by increasing the proportion of high-performing offspring in a breeding population, thus accelerating gain from selection (Spindel et al., 2015; Desta and Ortiz, 2015). One of the major limitations to use of GS until now has been the large number of markers required and the cost of genotyping these markers. But nowadays, GS is revolutioning plant breeding due to the decreasing genotyping costs and increasing phenotyping costs, and the ability to select faster individuals in the breeding cycle (Jannink et al., 2010). However, there are many aspects that should be taken under consideration in the use of GS: the need of a proper population design and composition of the TS, a difficult biological interpretation of the level in accuracy, the restricted number of generation cycles between TS and VS, etc. (Desta and Ortiz, 2015)

To assess the potential of GS, researchers have studied the different factors that affect prediction accuracy. These factors include: the LD between markers and QTLs in the training and the validation populations, the size of the training population and relatedness with the breeding population, the heritability of the trait under investigation, the genetic architecture of the trait, and the prediction models (Desta and Ortiz, 2014; Nakaya and Isobe, 2012). Some of the factors cannot be controlled, nevertheless the design of populations and the selection of an adequate statistical method can be achieved (de los Campos et al., 2013).

1.4.1. Statistical methods in GS

High density genotyping technologies made possible the emergence of GS and brought together, as a consequence, the “large p small n” problem that is caused by the use of such a large number of markers. The number of variables p that need to be estimated (marker effects), is larger than the number of available observations n (individuals) (Jannink et al., 2010). Furthermore, there is a high correlation or multicollinearity between the variables. The collinearity between linked markers may be reduced by recombination, suggesting that phenotyping should be done on progeny that experienced a greater number of total recombination events, which is usually not the case in practice (Jannink et al., 2010). These problems result in an over-fitted model with a poor predictive ability. To develop prediction models for GS that overcome these issues, different methods have been proposed. The statistical models can be grouped into shrinkage models, variable selection models, kernel methods, and dimension reduction methods.

1.4.1.1. RR-BLUP

Random regression best linear unbiased prediction (RR-BLUP), also called ridge regression, was first proposed for MAS by Whittaker et al. (2000), becoming later on an important tool for GS.
In this approach the marker effects are modelled as random, and are assumed to follow a normal distribution and to have equal variance. Assuming that all markers effects are drawn from the same distribution does not imply that effects are the same, but they are all equally shrunk towards zero (Jannink et al., 2010). This penalization is regulated by the shrinkage parameter $\lambda$ that is applied on the sum of the squared regression coefficients (de los Campos et al., 2013). In the practice of genomic selection $\lambda$ is estimated as the ratio of the residual variance and the variance of marker effects, under the assumption of equal variance for all marker effects (Meuwissen et al., 2001). The penalization imposed by ridge regression is equivalent to the one imposed by BLUP, for that reason, ridge regression is applied in a mixed model context where marker effects are modelled as random. The capacity of simultaneously estimate effects for all markers avoids marker selection and the biases that go alongside this selection (Whittaker et al., 2000).

Another penalized regression method is the least absolute shrinkage and selection operator (LASSO), in this case $\lambda$ is applied on the sum of the absolute values of the regression coefficients; in using LASSO some regression coefficients are shrunk to zero, which effectively comes down to performing variable selection at the same time as performing shrinkage, so selecting only a subset of the variables. The number of non-zero regression coefficients in LASSO cannot be higher than the number of observations (de los Campos et al., 2013). RR-BLUP captures the small effects (small coefficients receive less shrinkage than in LASSO) while LASSO shrinks the coefficients of small effects to zero, making it less suitable for complex traits. LASSO is less appropriate for GS than RR-BLUP for two reasons: first, because in GS there is no reason to restrict the number of markers with nonzero estimates of regression coefficients to the number of observations (de los Campos et al., 2013); and second, when predictors are correlated (that may occur when LD span over large regions), methods that performs variable selection are outperformed by RR-BLUP. If many markers with small effects determine the phenotype, ridge regression will capture those effects, whereas LASSO will capture large effects with a small number of markers (Nakaya and Isobe, 2012). Thus, RR-BLUP is the most used method in GS and has been proved to achieve high prediction accuracies across crops and traits. Moreover, it is suitable for GS of complex traits and it has been successfully used in recent GS rye studies (Bernal-Vasquez et al., 2014; Wang et al., 2014).

### 1.4.1.2. Bayesian Methods

The potential drawback of the methods introduced above is that the shrinkage parameter is identical for all markers, and that could lead to overshrinking of large effects (Zhao et al., 2014). Therefore, the previous assumption that individual markers have the same variance was unrealistic and not completely satisfactory. Meuwissen et al. (2001) proposed two Bayesian methods to overcome this limitation with two types of prior distribution. Bayes A assumes that each marker effect is drawn from a normal distribution with its own different variance, all markers have non-zero effect. Bayes B assumes also different variances for each marker, but in this case, a proportion of $\pi$ markers have effects equal to zero, and a proportion of $1-\pi$ of markers have effects unequal to zero. Thus, effects are sampled from a mixture distribution (Habier et al., 2011; Zhao et al., 2014). However, the shrinkage of marker effects is affected by $\pi$, and thus it should be treated as unknown and be inferred by the data. An improvement of Bayes A and B is achieved in Bayes C$\pi$ and Bayes D$\pi$ $\pi$ that treat as unknown. The genetic
architecture of the trait is affecting the probability $\pi$. Bayes C$\pi$ and Bayes D$\pi$ considers the prior probability $\pi$ that a marker has zero effect as unknown addressing this drawback. Estimates of $\pi$ from Bayes C$\pi$, in contrast to those from Bayes D$\pi$, are sensitive to training data size and marker density, the shrinkage of SNP effects, increases with SNP density, so the number of SNPs fitted in the model results mainly from larger effects and small effects are not detectable. Habier et al., (2011) compared the Bayesian methods in different traits and training data sizes and they conclude that the best method must be determined for each trait separately because the accuracies were similar and none of the methods outperformed all others and provide information about the genetic architecture of quantitative traits.

1.4.1.3. Random Forest

Machine-learning methods, such as Random Forest (RF), have been proposed and successfully applied to data with the “large $p$ small $n$” problem (Jannink et al., 2010). In this new approach RF builds a non-linear prediction model. RF is an algorithm that uses a collection of tree-structured predictors. Each tree in the ensemble is “grown” on the basis of a sample of the training dataset, which leaves roughly one third of the observations out because some plants will appear more than once and others will not appear at all. The individuals that do not appear are called OOB samples and they act as a validation set at each tree. The target response is individually predicted from each tree and the “forest” predicts the target response as an average of individual tree predictions. This strategy, called bagging, reduces the error prediction by a factor of the number of trees. RF is able to account for correlation and interactions among variables. Selection and ranking of variables is possible by taking advantage of variable importance measures. It is a very robust algorithm for classification and regression when there are thousands of input variables. This makes RF particularly interesting for high-dimensional genomic data analysis, and therefore, for our study, giving a different point of view than the commonly used linear models. As examples in literature of the use of this recent method in plant breeding Spindel et al. (2015) found RF to produce the most consistently accurate GS models for plant height as a trait in rice, and Poland et al. (2012) showed that imputation with RF lead to lower error in wheat.

1.5. Aim of the project

GS, also known as genomic prediction, is becoming a potent and attractive tool for plant breeders (Jannink et al., 2010; Heffner et al., 2009). The potential of GS has been confirmed in several plant species (de los Campos et al., 2013), being in some cases more efficient per unit of time than phenotypic selection or MAS (Heffner et al., 2009). This is also the case of hybrid rye breeding, where GS has been reported as superior to MAS (Bernal-Vasquez et al., 2014; Wang et al., 2014). However, GS is not a perfect method. The superiority of GS has not always been as high as anticipated by simulations (de los Campos et al., 2013). For empirical data, accuracy of prediction varies a lot, it depends on a number of factors and there is a need of further research to validate this method and integrate it optimally in plant breeding programs.

The aim of this study is to validate GS prediction ability across related selection cycles and to quantify the proportion of prediction ability due to differences between and within families in an empirical data set of a hybrid rye breeding program. Furthermore, a comparison of the
efficiency of genomic prediction versus phenotypic prediction will be performed by evaluating selection indices composed of phenotypic and genomic predictions. To accomplish that, the training population (its size and overall relatedness to the validation population), statistical method used to build the GS model, the selection cycles and the trait will be varied to determine their effect on prediction accuracy.
2. Material and Methods

2.1. Rye hybrid breeding program

The data for this study have been kindly provided by KWS LOCHOW. The information available consists of phenotypic and genotypic datasets coming from a German and a Polish rye hybrid breeding program belonging to the pollen-parent pool (Carsten).

The training set is composed of 828 S₂ lines from the pollen parent pool. 421 of these S₂ lines belong to the German breeding program and 407 S₂ lines belong to the Polish breeding program. The German and Polish breeding programs are interconnected due to the use of common and related parents and comparable recombination strategies.

In 2009, the training set lines were evaluated in a first general combining ability test (GCA1) to two different testers, across 7 locations and two replicates per location (for the German lines) and 4 different locations and two replicates per location (for the Polish lines). The German program contains 9 parental lines that were evaluated in previous years, referred to as “older lines”, 25 parental lines evaluated in 2009, and a portion of lines that display second-degree relationships with the validation set.

The traits under study were grain dry yield (GDY) measure in dt/ha, thousand kernel weight (TKW) measured in kg, and plant height (PH) measured in cm.

The validation set is composed of 99 recombinant inbred lines (RILs) derived from third generation of selfings from 33 crosses (3 RILs per cross). Parents of 47 RILs correspond to S₂ lines from the German set, evaluated in GCA1 in 2009, whereas 52 RILs are derived from one parent from the German set evaluated in GCA1 in 2009 and another parent from the older lines evaluated in GCA1 in previous years. The validation set was tested across 6 different locations with two testers and two replicates per location in 2013 and 2014.

Testers consist of a single cross of 2 inbred lines from the seed parent pool. The two testers used in the GCA1 test of the training set in 2009 are different than the ones used in the GCA1 tests of the validation set in 2013 and 2014, but they are all representative of the opposite (seed-parent) pool.

2.1.1. Description of Phenotypic Data

The phenotypic data available from the training set corresponds to the raw data and the least squared means (LS means) estimated as BLUEs, for the three traits under study, grain dry yield, thousand kernel weight and plant height in 2009. The LS means are calculated from a combined analysis of 11 different trials. Each trial is composed of a set of lines tested using an alpha design (R replicates and K incomplete blocks with n entries per incomplete block) with two replicates across locations. Adjusted means of “older lines”, parents tested in GCA1 during earlier years (prior to 2009), were used to build a bridge between the “older lines” and “new lines” tested in GCA1 in 2009 to adjust LS means for years effects.
The phenotypic data available from the validation set consists of the LS means for the same three traits in 2013 and 2014.

2.1.2. Description of Genotypic Data

Both training and validation populations were genotyped with 14269 single nucleotide polymorphisms (SNPs) from an Illumina rye 16k-SNP chip. The genotypic data are partially curated, i.e., technical mistakes and SNPs without signals were removed and missing signals for mapped SNPs were imputed via Beagle (Nothnagel et al., 2009). Genetic map position information in cM is available for 5574 markers that are mapped. 8695 markers are not mapped. Pedigree information is also available for every line, describing parents and grandparents. To model the additive effects the marker information was transformed into numbers. Under an additive model and for biallelic SNPs, a score of “0” is given to the heterozygotes (not informative), and “1” to the homozygotes for the allele of the one parent and “-1” for the allele of the other parent \( \{aa, Aa, AA\} = \{-1, 0, 1\} \).

2.2. Phenotypic Selection

2.2.1. Line per se selection

Lines with highest GDY hybrid performance were selected based on the observed phenotypic values from the validation set (LS means) in the years 2013, 2014 and the average phenotype of both years. For that, a truncation selection with a threshold of 10% of the highest yielding RILs was realized on the 99 RILs from the validation set.

2.2.2. Mid-parent value selection

Another method within phenotypic selection is the use of the mid-parent values (MPV) to select before testcross evaluations. The MPV of each line from the validation set was calculated as the mean of the phenotypic values from their two specific parents. These phenotypic values correspond to the LS means from the parents (“old and new lines”) derived from trials in 2009 and earlier years. The top 10% of the 99 RILs from the validation set with highest MPV value for GDY were selected.

2.2.3. Calculation of heritability

The heritability over all locations was calculated for the three traits under study from the raw data. The model considered location as fixed effect and the genotype and the interaction between genotype and location as random effects. The error variance was estimated as the interaction between genotypes and tester and plot replicates. The heritability can be calculated as follows:

\[
 h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{g.env}^2 + \sigma_e^2 (rep \times loc)} \tag{4}
\]

These calculations were made with the software GENSTAT v.10.
2.3. Genomic Selection

2.3.1. Description of different scenarios

The effect of the sample size of the training population and relatedness between training and validation population in hybrid rye was studied by comparing different scenarios (Table 1). The scenarios were built separating the lines from the training set into 6 different training sets that vary in size and relatedness with the validation population.

Table 1. Description of the training sets.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Training set</th>
<th>Origin</th>
<th>Size (number of lines)</th>
<th>Number of parents</th>
<th>Number of older lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1*</td>
<td>All lines from German and Polish program (including all 34 parents)</td>
<td>Germany and Poland</td>
<td>828</td>
<td>34</td>
<td>9</td>
</tr>
<tr>
<td>TS2*</td>
<td>Lines from German program (including “old lines”)</td>
<td>Germany</td>
<td>421</td>
<td>34</td>
<td>9</td>
</tr>
<tr>
<td>TS3*</td>
<td>Lines from German program (excluding all 34 parents)</td>
<td>Germany</td>
<td>387</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TS4*</td>
<td>Lines from German program including “old lines” and excluding 34 random non-parental lines</td>
<td>Germany</td>
<td>387</td>
<td>34</td>
<td>9</td>
</tr>
<tr>
<td>TS5*</td>
<td>Lines from German program (without “old lines”)</td>
<td>Germany</td>
<td>412</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>TS6*</td>
<td>Lines from Polish program</td>
<td>Poland</td>
<td>407</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

TS1*: TS composed of all lines from both programs (German and Polish) and different year of assessment (“new and old lines”), including therefore all 34 parents of the VS.
TS2*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines”).
TS3*: TS composed of lines from the German program, excluding the 34 parents of VS (25 “new lines” and 9 “old lines”).
TS4*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines”) and excluding 34 random non-parental lines.
TS5*: TS composed of lines from the German program, including 25 parents (“new lines”) and excluding 9 parents of the VS (“old lines”).
TS6*: TS composed of lines from the Polish program. No parental lines included.

Table 1 shows the characteristics of each training set. The TS1 is the bigger in size. It is composed of every line available for the training set, no matter the origin of the program (German or Polish) or the year of assessment (“older lines” included). The TS1 is a wider pool than the actual validation set that contains lines from the Polish breeding program and is, therefore, less related to the validation set lines (composed only of lines from the German breeding program) than the TS2, TS3, TS4 and TS5. The TS2 consists of the German lines including the ones evaluated in an earlier year (“older lines”). This training set contains the 34
parents of the VS and relatives. TS3 is composed of the same lines from the TS2 but excluding the 34 parents, so the prediction would be based only on relatives from the VS. TS3 represents a situation where the training set contains fewer lines that are less related to the lines from the validation set. TS4 consists of the 9 “older lines” plus German lines excluding 34 of them that are not parents from the VS. TS4 can be compared to TS3, TS4 is the same size than TS3 but it is more related to the VS than TS3. TS5 contains every German line but excludes the “older lines”. TS6 is composed of the Polish lines being the less related to the VS.

An increase in the accuracy is expected by the larger size of the training population. In contrast, a decrease in relatedness could lead to a substantial decrease in the accuracy of genomic selection model. Therefore, the optimal trade-off between size of the training population and genetic distance between training and test population has to be investigated in detail.

### 2.3.2. Statistical methods for genomic prediction

Two genomic selection methods were chosen based on their demonstrated success in accurately predicting GEBV in variety of crops and because they represent two different types of statistical methodologies used to build GS models: as a linear parametric methods RR-BLUP and as non-linear, non-parametric machine learning methods RF. A comparison of these two methods was done by studying their prediction accuracy for two traits (GDY and PH) and their estimated marker effects for one trait (GDY). Markers effects of both methods were ranked in order of importance for each statistical method and the Pearson correlation was calculated on the rankings of the 50 top markers.

#### 2.3.2.1. RR-BLUP

To estimate the GEBV from the validation set for the three traits under study the “rrBLUP” package from R was used (Endelman, 2011). Phenotypic and genotypic data from training and validation populations were defined and used by RR-BLUP to calculate the markers effects matrix from the following formula:

\[ Y = \mu + Xg + e \] (5)

where \( Y \) is a Nx1 vector of observations, phenotypic means of the training set (N=number of individuals of the training set), \( \mu \) is an overall mean of the training set, \( X \) is an N x number of markers, marker matrix from the training set, \( g \) is the markers effects matrix that will be calculated, and \( e \) is an Nx1 vector of residual effects. Once the markers effects are estimated, they are multiply back to the markers of the validation set to get the predicted phenotypic values of this set (GEBV).

In order to evaluate the prediction accuracy of the method, the GEBV from the training set itself were estimated. 5 random samples of 99 lines from the training set were used as validation set and the prediction accuracy with itself was calculated as the mean of the accuracies of these 5 predictions.
2.3.2.2. Random Forest

Random Forest was used as a second method to estimate the GEBV from the validation set for one trait, GDY. For that, Breiman and Clueter’s RF approach was implemented by the use of the package “Random Forest” (Charmet et al., 2014). The number of predictors (SNPs) per decision per tree was one third of the SNPs, and the number of trees was 5, both parameters were set as default. To evaluate the method, 5 random samples of 99 lines from the training set were used as validation set and the prediction accuracy of the TS with itself was calculated as the mean of the accuracies of these 5 predictions.

2.3.3. Correlation among markers. LD decay estimation

Linkage disequilibrium (LD) between markers of the same chromosome was calculated as the squared correlation \( r^2 \) between alleles at two loci (Hill and Robertson, 1968). In order to have a graphical representation of the correlation between markers an LD decay plot was created, the genetic distance of the markers in centimorgans (cM) was plotted against the squared correlation coefficient \( r^2 \). To verify that the marker density was high enough for a GS study, a boundary of LD was chosen visually for each chromosome to set the genetic distance necessary between markers or markers and QTLs.

2.3.4. Population structure. Cluster analysis

In order to visualize the population structure of the training and validation sets a cluster analysis was performed in R using the 14269 SNPs. The “Euclidean” distances were calculated between markers and a hierarchical cluster algorithm was applied to create the clusters.

2.4. Comparison of GS and phenotypic selection for three different traits

The study was done separately for the three traits, by the application of two genomic selection methods to estimate the regression coefficients in the six training populations described in Table 1. The training population, statistical method used to build the GS model, the selection cycles and the trait were varied to determine their effect on prediction accuracy. The statistical software used was R with different packages available for the different prediction models. In order to do a general comparison of the predictive ability of: (i) mid-parent predictions (MPV), (ii) GS (GEBV), and (iii) phenotypic selection (LS means), the following calculations were done during this study:

Prediction ability of phenotypic values for genetically identical material in different years was estimated by the correlation between LS means of the VS in 2013 and 2014 for three traits. This prediction ability of phenotypic values was used as a general reference to study the potential reduction of prediction ability due to genotype x year interaction.

GS across different cycles was the main experiment using individuals evaluated in 2009 and earlier to create 6 different TS, and individuals evaluated in years 2013 and 2014 as VS. GS across different cycles was compared with GS within the same selection cycle, in a separate experiment using as training set the individuals evaluated in year 2013, and as validation set, genetically identical individuals evaluated in year 2014. The latter allow to study the effect of

Heritability of GDY, PH and TKW was also calculated as explained in section 2.2.3.

Since the true breeding value was unknown, prediction ability was estimated for each trait as: the Pearson correlation coefficient between the different GEBVs and the observed phenotypic values from the VS (LS means from the VS from 2013, 2014 and the mean of the two years) for GS; and the Pearson correlation coefficient between the MPV estimated from the phenotypic GCA1 data in 2009 and the observed values from the VS. The average phenotypic values from two years may be considered the closer estimate of the true breeding value.

The effect of the genetic complexity of the trait was studied by comparing the predictive abilities described above in three different traits: grain dry yield (GDY) as a more genetic complex trait affected by a large number of small effect QTLs; and thousand kernel weight (TKW) and plant height (PH) as genetically more simple traits.
3. Results

3.1. Prediction ability of phenotypic values of lines and parents

Reproducibility of phenotypic values and prediction ability of MPV for genetically identical material tested in two different years (VS 2013, VS 2014) is shown as the Pearson correlation for three traits in table 2. The MPV correlation is stable among the two years of study for GDY, being higher when is correlated to the averaged phenotype from those years and always lower than the correlation between the observed phenotypic values in different years. For PH and TKW, the prediction ability of the MPV is higher for 2013 than for 2014. As expected, the correlation between the observed phenotypic values for PH and TKW is considerably higher than for GDY. PH is the trait with the highest predictability and GDY is the trait with the lowest predictability.

Table 2. Pearson correlations for phenotypic predictors

<table>
<thead>
<tr>
<th>Pearson correlation</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDY</td>
</tr>
<tr>
<td>Phenotypic values 2013/Phenotypic values 2014</td>
<td>0.57</td>
</tr>
<tr>
<td>Phenotypic values 2013/MPV</td>
<td>0.24</td>
</tr>
<tr>
<td>Phenotypic values 2014/MPV</td>
<td>0.24</td>
</tr>
<tr>
<td>Average of phenotypic values 2013-2014/MPV</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Similar results are obtained when studying the heritability of the traits that can be interpreted as a second estimator of prediction ability of phenotypic data (table 3). In this case, TKW is the trait with a higher heritability, followed by PH and GDY.

Table 3. Heritability estimates of GDY, PHT, TKW calculated over locations.

<table>
<thead>
<tr>
<th>Traits</th>
<th>GDY</th>
<th>PH</th>
<th>TKW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heritability</td>
<td>0.56</td>
<td>0.71</td>
<td>0.83</td>
</tr>
</tbody>
</table>

3.2. Prediction ability of genome estimated breeding values

Table 4 shows the results obtained from correlating the GEBV calculated with RR-BLUP, of three traits and six scenarios described in table 1, to the observed phenotypic values of the VS for two different years and the average for those years. Estimates of prediction abilities of GEBVs ranged from -0,042 [TS6 (TS composed of lines from the Polish program, no parental lines included), VS 2013] to 0,364 [TS4 (TS composed of lines from the German program, including all 34 parents of VS and excluding 34 random non-parental lines), VS 2013] for GDY, from 0,337
(TS6, VS 2014) to 0.540 (TS1, VS 2013) for PH, and from 0.002 (TS6, VS 2014) to 0.483 (TS5, VS 2013) for TKW.

Table 4. Prediction ability calculated as Pearson correlation between GEBV and phenotypic values of VS for GDY, PHT and TKW.

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Trait</th>
<th>Year of Validation Set</th>
<th>2013</th>
<th>2014</th>
<th>Average of 2013 and 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1*</td>
<td>GDY</td>
<td>0.239</td>
<td>0.252</td>
<td>0.295</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>0.540</td>
<td>0.467</td>
<td>0.530</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TKW</td>
<td>0.351</td>
<td>0.261</td>
<td>0.293</td>
<td></td>
</tr>
<tr>
<td>TS2*</td>
<td>GDY</td>
<td>0.363</td>
<td>0.268</td>
<td>0.346</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>0.526</td>
<td>0.466</td>
<td>0.521</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TKW</td>
<td>0.472</td>
<td>0.426</td>
<td>0.456</td>
<td></td>
</tr>
<tr>
<td>TS3*</td>
<td>GDY</td>
<td>0.165</td>
<td>0.165</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>0.461</td>
<td>0.432</td>
<td>0.468</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TKW</td>
<td>0.433</td>
<td>0.318</td>
<td>0.358</td>
<td></td>
</tr>
<tr>
<td>TS4*</td>
<td>GDY</td>
<td>0.364</td>
<td>0.274</td>
<td>0.351</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>0.536</td>
<td>0.463</td>
<td>0.525</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TKW</td>
<td>0.461</td>
<td>0.412</td>
<td>0.442</td>
<td></td>
</tr>
<tr>
<td>TS5*</td>
<td>GDY</td>
<td>0.305</td>
<td>0.284</td>
<td>0.344</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>0.511</td>
<td>0.437</td>
<td>0.499</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TKW</td>
<td>0.483</td>
<td>0.412</td>
<td>0.447</td>
<td></td>
</tr>
<tr>
<td>TS6*</td>
<td>GDY</td>
<td>-0.042</td>
<td>0.187</td>
<td>0.106</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>0.379</td>
<td>0.337</td>
<td>0.376</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TKW</td>
<td>0.080</td>
<td>0.002</td>
<td>0.018</td>
<td></td>
</tr>
</tbody>
</table>

TS1*: TS composed of all lines from both programs (German and Polish) and different year of assessment (“new and old lines”), including therefore all 34 parents of the VS.

TS2*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines”).

TS3*: TS composed of lines from the German program, excluding the 34 parents of VS (25 “new lines” and 9 “old lines”).

TS4*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines) and excluding 34 random non-parental lines.

TS5*: TS composed of lines from the German program, including 25 parents (“new lines”) and excluding 9 parents of the VS (“old lines”).

TS6*: TS composed of lines from the Polish program. No parental lines included.

The prediction using the average phenotypic data from two different years (2013 and 2014) is frequently higher than the prediction for a single year, especially when compared to the year 2014, being at the same range or lower than the prediction for year 2013. As expected, higher prediction abilities are achieved when PH is analysed and lower prediction abilities when GDY is under study. Prediction ability varies depending on the size and relatedness of the training set for GDY and TKW but stays quite stable for PH across different training sets (Figure 1). Training sets containing the parents from the VS (TS1, TS2, TS4 and TS5) have higher prediction ability than the ones without parents (TS3 and TS6), decreasing the prediction when the relatedness between TS and VS is lower. The same effect is magnified when predicting the
exact same individuals in different years, using as TS the 99 RILs evaluated in year 2013 and as VS the same 99 RILs evaluated in year 2014 (table 5).

Table 5. Prediction ability calculated as Pearson correlation between GEBV and phenotypic values of VS evaluated in 2014 for GDY, PHT and TKW.

<table>
<thead>
<tr>
<th>Training / Validation Set</th>
<th>99 RILs evaluated in 2014</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 RILs evaluated in 2013</td>
<td>Method</td>
<td>GDY</td>
</tr>
<tr>
<td>RR-BLUP</td>
<td></td>
<td>0.488</td>
</tr>
</tbody>
</table>

GS prediction ability increases when using a maximum genetic relationship between TS and VS (TS=2013; VS=2014) from 0.295 to 0.488 (TS1 for GDY), from 0.530 to 0.800 (TS1 for PH) and from 0.293 to 0.741 (TS1 for TKW).

![Figure 1. Prediction ability calculated as Pearson correlation between the GEVB for the different traits and training sets using RR-BLUP and the average phenotypic values across 2013 y 2014.](image)

In order to compare RR-BLUP and Random Forest, GEBV were calculated for GDY and the prediction ability of each TS with itself and with the VS for both methods was estimated. The results are presented on the table 6. For every training set and different years of evaluation RR-BLUP has superior prediction accuracy than Random Forest.
Table 6. Pearson correlation between the GEBV of GDY (calculated by RR-BLUP and Random Forest) and prediction ability of the training sets with themselves.

<table>
<thead>
<tr>
<th>Training / Validation Set</th>
<th>RR-BLUP</th>
<th>2013</th>
<th>2014</th>
<th>13/14</th>
<th>Itself</th>
<th>Methods</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.295</td>
<td>0.919</td>
<td></td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.622</td>
</tr>
<tr>
<td>TS2*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.363</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.346</td>
<td>0.957</td>
<td></td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.631</td>
</tr>
<tr>
<td>TS3*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.188</td>
<td>0.964</td>
<td></td>
<td>-0.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.667</td>
</tr>
<tr>
<td>TS4*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.364</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.351</td>
<td>0.959</td>
<td></td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.649</td>
</tr>
<tr>
<td>TS5*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.344</td>
<td>0.950</td>
<td></td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.641</td>
</tr>
<tr>
<td>TS6*</td>
<td>-0.042</td>
<td></td>
<td>0.187</td>
<td>0.106</td>
<td>0.934</td>
<td></td>
<td>-0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.084</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.573</td>
</tr>
</tbody>
</table>

TS1*: TS composed of all lines from both programs (German and Polish) and different year of assessment (“new and old lines”), including therefore all 34 parents of the VS.

TS2*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines”).

TS3*: TS composed of lines from the German program, excluding the 34 parents of VS (25 “new lines” and 9 “old lines”).

TS4*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines”) and excluding 34 random non-parental lines.

TS5*: TS composed of lines from the German program, including 25 parents (“new lines”) and excluding 9 parents of the VS (“old lines”).

TS6*: TS composed of lines from the Polish program. No parental lines included.

Random Forest leads to lower prediction abilities within training sets than RR-BLUP. However, prediction ability when using part of the TS as VS shows acceptable values (around 0.6).

The rank correlation of the 50 largest SNP effect estimates obtained with RR-BLUP and RF was close to zero (0.018).

Fig. 2 shows the marker effect estimates across the genome obtained by RR-BLUP for GDY for TS1, whereas unmapped SNPs where placed on a virtual “eighth chromosome”.

Most SNP effects have very low values in a range between -0.03 and 0.03, as can be expected from this model that shrinks the marker effects towards zero. However, there are a few data points that have a superior value close to 0.05 or -0.05. In chromosome 4 higher values for marker effects are found. To investigate the existence of a possible QTL for yield in the fourth chromosome of rye, the plot of marker effects of that chromosome against their relative position was drawn also for the statistical method RF (figure 3). The marker effects of RF show also two high values at the end of the chromosome.
The comparison of RR-BLUP and Random Forest was also performed for the trait PH (Table 7). Again, RR-BLUP has superior prediction accuracy than Random Forest for each training set and different years of evaluation except for the Polish TS. For PH the prediction accuracy of RF was acceptable and did not show a collapse as it did in GDY when predicting the VS. The prediction abilities within training sets were similar to those for GDY.
3.3. Additional results

3.3.1. Population structure analysis

A marker-based cluster analysis was performed to characterize the genetic structure of the training set(s). The cluster analysis indicated that there is no differentiation in the structure of the TS6 (TS composed of lines from the Polish program, no parental lines included) and TS5 (TS composed of lines from the German program, including 25 parents “new lines” and excluding 9 parents “old lines”) (Fig. 4). RILs from the VS and parents included in the TS formed clusters that overlap with the Polish and German TS (TS6 and TS5), showing a lack of population structure (Fig. 5). This is in agreement with the breeder’s information about interconnection between these two breeding programs.
3.3.2. Linkage disequilibrium analysis

Linkage disequilibrium (LD), measured as $r^2$, between pairs of polymorphic SNP marker loci was plotted against the genetic distance (cM) for 7 chromosomes (Fig. 6).

Setting a visual boundary at $r^2=0.4$, LD was, on average, 10 cM for chromosome 1, 15 cM for chromosome 2, 20 cM for chromosome 3, 5 cM for chromosome 4, 20 cM for chromosome 5, 10 cM for chromosome 6, 5 cM for chromosome 7, and across chromosomes was about 10 cM. The average available marker density was 0.1 cM.
3.3.3. Comparison of selected fractions from phenotypic selection and different GS selection sets

The 10% highest yielding lines was selected for each phenotypic and genomic selection set. Table 8 shows the proportion of common lines for individual selection-set pairs, when using a selection fraction of 10%.
Table 8. Proportions of lines selected in common by different selection methods.

<table>
<thead>
<tr>
<th>Lines commonly selected</th>
<th>PS 2013</th>
<th>PS 2014</th>
<th>PS 13/14</th>
<th>MPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR-BLUP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS1*</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>TS2*</td>
<td>0.10</td>
<td>0.10</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>TS3*</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>TS4*</td>
<td>0.10</td>
<td>0.10</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>TS5*</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>TS6*</td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
</tr>
</tbody>
</table>

TS1*: TS composed of all lines from both programs (German and Polish) and different year of assessment ("new and old lines"), including therefore all 34 parents of the VS.

TS2*: TS composed of lines from the German program, including all 34 parents of VS (25 "new lines" and 9 "old lines").

TS3*: TS composed of lines from the German program, excluding the 34 parents of VS (25 “new lines” and 9 “old lines”).

TS4*: TS composed of lines from the German program, including all 34 parents of VS (25 “new lines” and 9 “old lines) and excluding 34 random non-parental lines.

TS5*: TS composed of lines from the German program, including 25 parents (“new lines”) and excluding 9 parents of the VS ("old lines").

TS6*: TS composed of lines from the Polish program. No parental lines included.

The comparison of these 10 RILs selected by each method, shows that in most cases only one tenth of lines are selected in common in the top 10% fraction (i.e. 1 RIL out of 10). The methods that showed the highest proportion of common lines in the selected fraction (0.3) were the TS2 and TS4 with the phenotypic selection done on the average phenotype over the two years (which is the closer value to the true breeding value). The selection considering the average phenotype of both years is the one with a higher number of RILs selected in common with the other selection methods.
4. Discussion

4.1. Comparison of traits

The number and distribution of genes that affect a trait, the magnitude of their effects, and the relative contributions of additive, dominant, and epistatic gene effects is referred to as genetic architecture. In our study comparison has been done between three traits with different complexity in their genetic architecture.

The correlation between observed VS values is indicative for the predictability. GDY showed a lower prediction ability (table 2) than PH and TKW. This result was expected, considering the complexity of the trait GDY, regulated by a larger number of genes and therefore subject to genotype x environment interactions to a larger extent (Charmet et al., 2014). The prediction ability was moderate for GDY (0.57) and high for PH and TKW (0.84 and 0.73, respectively). Heritability estimates (table 3) are in the same range as the correlations, showing the potential of each trait to be selected in phenotypic or genomic selection. The high values for the phenotypic correlation over years and for heritability show that the quality of the phenotypic data is acceptable. Other studies have similar or higher results in heritability estimates for GDY, TKW and PH (0.7, 0.85 and 0.87) (Miedaner et al., 2012) and GDY and PH (0.75 and 0.94) (Wang et al., 2014).

The phenotypic average of years 2013 and 2014 is a value closer to the true breeding value of the genotypes. Thus, the study of correlations with the average phenotype will give values closer to the prediction accuracy than the correlations with one year phenotype. In this study, prediction ability is obtained by calculating the correlation between GEBV and observed phenotypic data. In several studies (Spindel et al., 2015; Schopp et al., 2015), researchers find interesting to obtain the prediction accuracy that is the correlation between the GEBV and the true genetic values. In simulation studies the true genetic value can be estimated. In empirical studies, prediction accuracy could be estimated analytically as prediction ability divided by h, where h is the root of heritability (Dekkers, 2007). However, this analytical estimate of prediction accuracy is associated with estimation errors as well, so that it is preferred to use prediction abilities as references to assess the relative merit of different statistical methods and training sets. Moreover, the use of the average phenotype is giving a closer estimate of the true genetic value than the use of phenotypic data from an individual year.

Prediction ability of every scenario and for both statistical methods was consistently higher for PH than for GDY. The lower accuracies obtained for GDY, compared to the other traits, may be due to a lower heritability or to a more complex genetic architecture (many small QTLs). G x E interactions seem to be a major confounding factor in other studies (Charmet et al., 2014). In our case it could be playing a role as well, the more genes controlling the trait, the more gene x environment interaction, and therefore less prediction ability for phenotypic selection and GS.
4.2. Comparison of Phenotypic and Genomic Selection

The main objective of our study was to compare the prediction ability of phenotypic selection and genomic selection under different scenarios for three different traits in rye.

Accuracy of phenotypic selection can be studied through the correlations between MPV values and observed phenotypes. These correlations show the same trend as the correlation between phenotype of different years, being lower for the trait with lower heritability, GDY, and higher for PH and TKW (table 2). As expected, the prediction ability of MPV is higher with respect to the average of observed values in 2013 and 2014 than for individual years for all traits.

In GS higher prediction abilities are obtained by RR-BLUP using TS that contain parental lines. Prediction ability from the GEBV using RR-BLUP and the TS3 (TS composed of lines from the German program, excluding the 34 parents of VS) has different behaviour depending on the trait. When excluding all parental lines from the TS (TS3), the prediction ability from the GEBV (0.19) is lower than the prediction ability from the MPV using the average phenotype of GDY (0.29), but is similar in the case of TKW (0.36 and 0.33 respectively) and higher for PH (0.47 and 0.37). In the case of GDY, if GS is applied through independent growing cycles, it is essential that the parents are present in the TS, without their presence the prediction is not accurate and phenotypic selection would be a better selection method. In addition, the prediction ability for the TS6 (TS composed of lines from the Polish program), a TS without parents, less related to the VS and evaluated in a different set of locations, is really low for GDY (0.11) and for TKW (0.02), staying at the same level of prediction with MPV for the PH (0.38). PH is the only trait that can be predicted with GS when the VS is not highly related to the TS and it is evaluated in a different environment. A direct comparison between phenotypic selection and GS is possible by using as TS the observations of the 99 RILs in 2013 and as VS the same 99 RILs observed in 2014. The correlation between observed phenotypes from the two years and the prediction ability of GS is 0.57 and 0.49 for GDY, 0.84 and 0.80 for PH, and 0.73 and 0.74 for TKW respectively. When the use of predicted lines as training set is compared to the use of first- and second-degree relatives in the training set, prediction abilities increased by from 0.35 to 0.49 (for GDY), from 0.52 to 0.80 (for PH) and from 0.46 to 0.74 (for TKW), indicating that genetic relatedness is essential in GS. These results are in accordance with Zhao et al. (2014) who stated that relatedness between training and validation sets is the driving factor of accuracy of prediction for complex traits.

Prediction ability within families of each method can be compared by the prediction ability of MPV and the prediction ability of the TS2 (TS composed of lines from the German program, including all 34 parents of VS). The gain of using markers obtaining a different prediction for each individual is 6% for GDY, 13% for TKW and 15% for PH. The capture of random genetic variation between progeny from the same cross due to Mendelian sampling is exploited in GS, but for our study it is not giving a great advantage over phenotypic selection. The MPV prediction does not need molecular marker data, is faster and simpler, so there is a need of a deeper discussion on the profitability of the methods in comparison with the prediction ability. Increasing the number of RILs per family could improve the efficiency of GS, because prediction ability of GS for hybrid breeding varies with the number of lines per cross within the
families, small families (few lines per cross) give a lower prediction accuracy than larger families (Schopp et al., 2015).

Looking at the results of table 8, GS with TS2 (TS composed of lines from the German program, including all 34 parents of VS) and TS4 (TS composed of lines from the German program, including all 34 parents of VS) are the methods that have a higher percentage of lines selected also with the average phenotype of two years (closer value to the true breeding value). This suggests that there is some information given by GS (including first- and second-degree relatives) that is lost in phenotypic selection. But in order to investigate the real gain of GS versus phenotypic selection, further research is needed: the selection should continue one more cycle and the response to selection should be evaluated.

### 4.3. Comparison of different scenarios in Genomic Selection

Focussing on genomic selection via RR-BLUP, we explored abilities of prediction between different training sets. The performance of RR-BLUP for GDY, PH and TKW, the traits under study in 6 different scenarios generated by varying the training sets size and relatedness to the VS, has been characterized by the prediction ability (table 4). In general, differences between training sets in their predictive ability (defined by the correlation between the GEBV and the phenotypic value), were only modest, with the exception of the significant decrease in accuracy when parents or relatives are excluded from the TS, stressed in the complex trait GDY.

Prediction ability increases with increasing training set size (Technow et al., 2014). In our study, the opposite is observed, prediction ability decreased with increasing training set size (table 4). However, the decrease was relatively small. Training sets TS1 (TS composed of all lines from German and Polish programs and different years of assessment, including all 34 parents), TS2 (TS composed of lines from the German program, including all 34 parents) and TS4 (TS composed of lines from the German program, including all 34 parents and excluding 34 random non-parental lines) contain the same 34 parental lines each one, and their size is 828, 421, and 387 lines respectively. The average prediction ability for the larger TS with RR-BLUP is 0.373 (TS1), 0.441 for the medium size set (TS2), and 0.440 for the smaller size set (TS4). This is in contrast to studies on genomic prediction in plant breeding, where tripling the size of TS could double the accuracy (Technow et al., 2014). Albrecht et al. (2011) observed that the prediction accuracy was smaller when the training set size was halved. The decrease in prediction ability with decreasing size of TS is most likely a consequence of the increase in the relative percentage of parents in the TS. Training sets TS1, TS2 and TS4 contain the same in number of parents (34) but not in relation to the number of total lines, so the larger TS included more unrelated, less informative material. Thus, a possible explanation could be that the increase in relatedness is raising the prediction ability. If the TS5 (TS composed of lines from the German program, including 25 parents “new lines” and excluding 9 parents “old lines”) and TS6 (TS composed of lines from the Polish program, no parental lines included) are compared, their size is practically the same but the relatedness to the VS is different. The TS5 has a higher prediction ability for all traits, due to the presence of the parents in the TS but also the importance of closest relatives has to be taken into account.
In order to study the importance of second degree relatives TS2 (TS composed of lines from the German program, including all 34 parents) and TS3 (TS composed of lines from the German program, excluding the 34 parents) were compared (table 4). For all traits, TS2 showed a higher prediction ability than TS3, with an remarkable decrease in GDY when excluding the parental lines and including second degree relatives (from 0.35 to 0.19). Technow et al. (2014) demonstrated the disproportional importance of the closest relatives for prediction accuracy. The rather small differences in prediction ability between the various scenarios with German background investigated seem to reflect the importance of the presence of close relatives that determined prediction ability. One explanation for the reduced accuracies when predicting unrelated populations is the presence of different alleles (Charmet et al., 2014). Then, the drop in prediction ability when the TS6 is used may be explained by of the lack of parents and relatives despite the known relation with VS. Under the suspicion of a hidden population structure, the analysis of the population structure was done. In order to confirm the lower relatedness of Polish lines (TS6) with the VS, a cluster analysis was performed (Figures 3 and 4). Surprisingly, the genetic distance was not obvious so that another explanation of the decrease in accuracy may be considered: the different set of locations for evaluation of this TS and the VS in different countries. In the cluster analysis it can be seen that the Polish individuals are not grouping in a closely related genetic subgroup apart from the rest of individuals. Thus, the lower prediction ability of the TS6 can only be explained by the direct absence from parents and second degree relatives in the TS and the different trial location for the VS and the TS.

TS that contain parental lines are the ones providing the highest prediction abilities: the accuracy is not highly influenced when 34 random lines are removed from the TS, but if the 34 parents are removed the accuracy drops dramatically for GDY from 0.35 to 0.19. This shows that the effect of the parents is really high for a complex trait. PH and TKW show the same effect but less pronounced. Some questions arise: is it necessary to have all the parental lines in the training set? Is it enough to have only the parents evaluated in 2009? Or is it important to have the old lines included? A mild reduction of the prediction ability is observed when comparing TS2 (TS composed of lines from the German program, including all 34 parents) (0.36) and TS5 (TS composed of lines from the German program, including 25 parents “new lines” and excluding 9 parents “old lines”) (0.30) that is still a high prediction for GDY. The same occurs to PH (0.52 to 0.50) and TKW (0.46 to 0.45) suggesting that it is enough to have the parents that were evaluated in 2009, and it would not be necessary to have the complete set of parental lines included in the TS. The TS5 contains both parental lines of 47 RILs from the VS and one parental line of the other 52 RILs from the VS, excluding only the second parental of 52 RILs. Including in the TS at least one parent from each line of the VS seems to be necessary to obtain acceptable prediction ability. In another investigation, Zhao et al. (2013) found that the prediction accuracy was reduced from 0.55 to 0.37 for hybrid wheat when the training and validation sets were not related versus having at least one common parent. Looking at the prediction ability of RR-BLUP (Table 4) of the TS5 (TS composed of lines from the German program, including 25 parents “new lines” and excluding 9 parents “old lines”) and the TS2 (TS composed of lines from the German program, including all 34 parents) obtained from the correlation with the average year phenotype, they were almost the same: 0.34 for GDY. However, prediction ability obtained from the individual years for GDY is not the same for TS5 and TS2 (TS composed of lines from the German program, including all 34 parents). Prediction
ability shows an increase from 0.305 to 0.363 when adding the old lines for year 2013 (Table 4). This difference is not there in the prediction ability with the average phenotype, meaning that the principal effect comes from the German parental lines evaluated in 2009.

4.4. Comparison of statistical methods

In our study, the parametric method (RR-BLUP) outperformed the non-parametric method (RF), the former being the best method under every scenario and for the two traits analysed (tables 6 and 7). The same results were obtained by other researchers (Spindel et al., 2015).

In order to evaluate the prediction accuracy of the methods, the GEBV from random samples of individuals of the training set were estimated. RR-BLUP showed an average of 0.947 prediction ability for the prediction using part of the training set as validation set for GDY and 0.975 for PH. This high correlation indicates that the phenotypic data of GDY and PH are correctly associated with the molecular markers and there are no inconsistencies, no systematic error is found. In the case of Random Forest, average prediction ability using part of the training set as validation set is a bit lower (0.631 for GDY and 0.629 for PH) but still acceptable and there is no signal of a systematic error.

The prediction ability for GDY using RF is acceptable in the case that the predicted individuals (99 lines from the TS that are used as VS) are at the same time part of the TS, i.e., the training set is predicted against itself. But the prediction ability collapses when the actual VS is predicted for GDY and the prediction accuracy drops for the PH. That could suggest that the VS is not so related to the training set, but our cluster analysis and the results from RR-BLUP indicate the contrary. Moreover, prediction ability for PH with RF is lower than using RR-BLUP but there is no a prominent drop of it.

RR-BLUP differs from the other approach used in this study in that the unconditional variance of marker effects is normally distributed, with the same variance for all markers, all markers are penalised equally. The suitability of different methods of developing predictive models to estimate GEBV is expected to be trait dependent, conditional on the genetic architecture of the trait. This can be seen in our results, where the simpler trait (PH) has better prediction ability in both methods, but especially in RF, where GDY was not predictable at all. The methods may capture different aspects of the relationship between phenotype and SNPs. Spindel et al., 2015 concluded that RR-BLUP compared to RF was the best performing method for grain yield in rice, where no large effect QTL was detected. Random Forest is able to capture the higher marker effects for a simpler trait but the lower number of samples to construct the prediction could lead to a reduction of power which would affect mostly the smaller effects of a complex trait. Nevertheless, other researchers found completely opposite results, where RR-BLUP prediction was outperformed by RF for grain dry yield (Jannink et al., 2010), suggesting that our results may be incorrect. They explain that a non-linear prediction model may be especially useful when the relationships between predictors and responses are nonlinear (as would occur if epistatic effects were responsible of great amount of genetic variation of the trait).

The different behaviour of the two statistic models when predicting GDY is reflected also in the correlation close to zero of the ranking of marker effects. The comparison of importance in use
of marker effects in RF and RR-BLUP, shows a clear difference in the use of marker effects. Probably, the use of wrong predictors is the reason of the low prediction ability of RF for GDY.

4.5. Further research

Complex traits, like grain yield, are probably controlled by a large number of QTLs with small effects (Charmet et al., 2014). Unfortunately, not much is known about genetic architecture of most traits for rye (Miedaner et al., 2012). Recently, different researchers found several QTLs for GDY in rye which explained from 5 to 24% of the genotypic variation (Miedaner et al., 2012). Our marker effects estimated by RR-BLUP for GDY with the TS1 (TS composed of all lines from both programs (German and Polish) and different year of assessment (“new and old lines”), including therefore all 34 parents of the VS) (Fig. 2) show a possible QTL on the last part of chromosome 4. This position corresponds approximately with one of the QTLs found by Miedaner et al. (2012). Nevertheless, a better approach would be to plot the marker effects from RF which have not been shrunk and the effects are not diluted (divided by the number of adjacent markers), so the exact position cannot be seen anymore, especially with such a high density of markers. This approach has been done for chromosome 4 (Fig. 3) using RF marker effects, and again, the plot suggests a possible QTL at the end of chromosome 4. However, any statistical method that models simultaneously all or a large number of marker effects may not be appropriate to identify QTL for quantitative (complex) traits, so this statement cannot be concluded in this study.

The main use of markers for this study was to perform GS. Marker density in this study seems enough to realize a GS prediction even with the fast LD decay that shows rye. Zhao et al., (2013) suggest that a SNP marker density resulting in average r values among adjacent loci above 0.2 can be used for genome wide approaches to predict phenotypic performance. LD was very common for distances <10cM. Occasionally, LD occurred between loci further apart. However, an extended study on LD is needed, in order to find gaps in the linkage map where associations with the trait may have been missed. The genomewide LD in this population, i.e. the LD decay as a function of genetic distance, is relatively high compared to that in other studies (Schopp et al., 2015). This is a consequence of the small effective population size of this experiment (around 34 parents), which leads to a strong population structure.

Genomic selection is a marker based selection method that promises to improve and accelerate the breeding process in plants and animals. Numerous studies have investigated the gain per unit time; however only a few have directly compared gains from genomic and phenotypic selection. In this study phenotypic selection and GS are compared as two different selection methods, but there is a need for further research. The best performing 10% RILs were selected using the different methods, these lines should be evaluated together again in a following breeding cycle in order to compare the gain from selection of each method.
5. Conclusions

Several factors can affect the prediction ability of genomic selection for rye: trait heritability, genetic architecture, the statistical model used, and size and relatedness of the training set, being the latter of great importance in this study.

PH was the trait with a higher predictability, followed by TKW and GDY.

RR-BLUP proved to be an interesting statistical method to be used in GS of important traits in rye outperforming in all cases RF. RF performed good for a simpler trait, PH, but no prediction ability was obtained with this method for a more complex trait, GDY.

GS presents slightly higher prediction ability than phenotypic selection for three traits under study. The gain of GS by capturing random genetic variation between progeny from the same cross due to Mendelian sampling for our study is not giving a great advantage over phenotypic selection. There is a need of a deeper discussion on the profitability of the methods in comparison with the prediction ability. More research is needed to design optimal training populations, with a sufficient level of relatedness to the validation set to achieve a good accuracy and to compare the gain in selection with the costs of each selection method.
6. References


Wang, Y., Mette, M.F., Miedaner, T., Gottwald, M., Wilde, P., Reif, J.C. and 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, 1-12.

