GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Schlussbericht gemäß Nr. 8.2 NKBF
Forschungsvorhaben im BMBF-Programm

Pflanzenbiotechnologie – Verbundvorhaben:
RYE-SELECT: Genome-based precision breeding strategies for rye

Teilprojekt E

Förderkennzeichen: 0315946E
Zuwendungsempfänger: Universität Hohenheim, 70593 Stuttgart
Ausführende Stelle: Universität Hohenheim – Fakultät Agrarwissenschaften – Forschungsschwerpunkt Biotechnologie & Pflanzenzüchtung, Fruwirthstr. 21, 70599 Stuttgart
Projektleiter: Prof. Dr. Karl Schmid (Koordinator), Prof. Dr. Hans-Peter Piepho, apl. Prof. Dr. Thomas Miedaner
Projektlaufzeit: 01.10.2011 bis 31.08.2015

„Das diesem Bericht zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 0315946E gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor."

A Work Package 2A: Genomic Selection

I. Kurze Darstellung

1. Aufgabenstellung

The general aim of the work was to develop genome-based breeding strategies to improve selection efficiency for yield and other agronomically important traits in the rye hybrid breeding program. The specific objectives were: (i) to evaluate the effect of phenotypic models into the genomic selection framework, (ii) to revise the current pre-processing steps of trial analysis towards genomic selection, and (iii) to evaluate genomic selection models using forward validation.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Phenotypic datasets were received at the beginning and in the course of the project. As genotypic data, a 15k SNP-chip for entries of the GCA1-2009 and GCA1-2010 was available for the first part of the project. Later on, an 18k SNP-chip was delivered for our use containing entries of candidates of selection for the years 2011 to 2014.

3. Planung und Ablauf des Vorhabens

The first step was to evaluate phenotypic spatial and non-spatial models for multi-location field trials using a set of two sequential years weakly connected. This topic was investigated and published in our first paper. Then, a question regarding the validity and understanding of the pre-processing detecting and deleting outliers was raised by our collaborators in the consortium. The procedure is automatically carried out and its outcome is then routinely employed (and taken for granted) in the following phenotypic analysis. We scrutinized the outlier detection method used by KWS-L (PlabStat-based) and proposed and compared it to easy-to-use methods under a genomic selection frame. This topic was investigated and published in our second paper. The third step covers the joint analysis of the complete dataset using a forward validation approach (prediction of a new cycle) and an evaluation of different scenarios, where the training set and validation set share different parent structures. The investigation is currently under way and a third paper is in preparation and will be submitted soon.

4. Wissenschaftlicher und technischer Stand an den angeknüpft wurde

Every year a new breeding cycle starts with single plants selfing, where line *per se* performance is evaluated and lines are selected. After *per se* selection, testcrosses on S_2 genotypes are evaluated in multi-environment trials (MET) during at least three years. In the first year (GCA1 trials) about 10 to 12% of the genotypes are selected for a further evaluation in the next year (GCA2 trials), where again a subset of 20 to 25% of the entries are selected for a more intense evaluation the following year (FACT trials) in more locations and with different testers than in the previous two years. The datasets available for analysis include the GCAs and FACT trials from 2009 to 2014. These datasets contain genotypes from two programs, i.e., Germany

and Poland, which are tested independently in GCA1 and then, after selection, are merged and evaluated together in GCA2 and FACT.

II. Eingehende Darstellung

1. Verwendung der Zuwendung und die erzielten Ergebnisse im Einzelnen mit Gegenüberstellung der vorgegebenen Ziele

1.1 Phenotypic analysis towards genomic selection (paper #1)

Bernal-Vasquez, A.M., Möhring, J., Schmidt, M., Schönleben, M., Schön, C.C., and Piepho, H.P. (2014). The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics*, 15:646.

Description of the datasets used

The GCA1-2009 and GCA1-2010 consist of subsets of 320 genotypes from S_2 populations tested in several locations within each of the two programs involving two testers. We define a trial as the physical unit within a location, where a subset of genotypes that were testcrossed to the same tester is evaluated. Trials at a location were laid out as α-designs with two replicates. Each trial was randomized independently from the others. Blocks were nested within rows. Row and column coordinates of the plots were available to account for spatial variation. Normally, throughout the program, only a single tester was used per location and year, but in at least one location from each program, some subsets of genotypes were testcrossed with the two available testers. Within years, four common checks were testcrossed with the testers and use to connect the trials. Only one check was in common between the two years.

Models and Assumptions

We compared two approaches: year-wise approach vs. across-years approach, both following a stage-wise analysis (Piepho et al., 2012). In the first stage of the year-wise approach, adjusted means of genotypes by location are computed. In the second stage, adjusted means of genotypes across locations by year are calculated, and in the third stage, genomic selection (GS) is performed joining the data of two years and using the mean of all genotypes in a specific year as the year adjustment effect. We can do this analysis assuming that the set of genotypes evaluated in one year is a representative sample of the complete breeding population, which is a realistic assumption in plant breeding scenarios. In the across-years approach, the first stage is the same as the year-wise approach, i.e., adjusted means of genotypes by location are computed. Then, the years are merged and adjusted means of genotypes across locations and years are computed, so that the year effect is already adjusted based on the single check that is shared in 2009 and 2010. In the

third stage, GS is performed (Fig 1 paper#1). Additionally, for the first stage, which is the same for both approaches, we assessed 9 different spatial and non-spatial models (Table 4 paper#1), where variance-covariance structures (e.g. AR(1), LV, AR(1) X AR(1)) were evaluated to model variation along rows, columns and rows and columns within replicates per trial.

Outcome

Rather than relying on the unique check to adjust for the year effect, we showed how by using the year genotype means across all entries in the GS model, we can handle weakly linked datasets and are able to perform more realistic predictions. Further, we showed that fitting row and column effects in the phenotypic analyses captures most of the heterogeneity in the field trials.

1.2 Outlier detection (paper #2)

Bernal-Vasquez, A.M., Utz, H.F., Piepho, H.P. (2016). Outlier detection methods for generalized lattices: A case of study on the transition from ANOVA to REML. *Theoretical and Applied Genetics*. doi:10.1007/s00122-016-2666-6.

Importance of the topic

The PlabStat software (Utz, 2003) for plant trials analysis has been widely used by plant breeders in Germany. The outlier identification method implemented in PlabStat has been trusted and used for decades. PlabStat uses an ANOVA-based approach for the variance parameter estimation. With the computer technology boom, other methods that are meant to have better accuracy estimator properties, i.e., restricted maximum likelihood (REML), are now highly demanded. Thus, a transition from ANOVA-based software to REML-based software is strongly encouraged. Motivated by the popularity and extensively use of PlabStat but also by the lack of detailed information about the outlier detection method and the requirement of finding a smooth transition from ANOVA to REML, we conducted a research work aiming to compare the variance component estimates between ANOVA and REML approaches, to scrutinize the outlier detection method of the ANOVA-based package PlabStat and to propose and evaluate alternative procedures for outlier detection.

Methods evaluated and validation using GS

Using published datasets of generalized lattice designs with and without outliers and with missing (deliberated deleted) observations, we demonstrated that the outputs between ANOVA- and REML-based analyses are very close in terms of variance component estimates and outliers detected (when using the same outlier detection method but different variance estimation approach). The PlabStat outlier detection method is based on the theory described in Anscombe and Tukey (1963), who propose the calculation of a threshold applied to the distribution of the residuals under study. The magnitude of the threshold depends on a constant calculated given:

(i) the ratio between the degrees of freedom of the residuals and the number of observations, (ii) a so-called *premium*, and (iii) an estimate of the scale parameter of the residuals. The *premium* is, in layman's words, the penalty that the researcher is willing to pay as a consequence of having a protection against spurious observations right on the dispersion (variance) allowed for the residuals. The outlier detection method in PlabStat uses a threshold to flag outliers based on a robust estimate of the scale of the distribution of the residuals, i.e., median absolute deviation (MAD) of the residuals, a *premium* at a fixed level of 0.05%, and an additional constant derived from the experience of the PlabStat programmer, Dr. Utz. Another common method among plant breeders is what we named studentized residual razor (SRR). SRR uses studentized residuals (obtained by standardizing raw residuals by the estimate of its standard deviation) and a threshold value derived from the standard normal distribution, normally set by breeders as a fixed value between |2| and |4|.
The other methods we proposed consider: (i) residuals standardized by the MAD, (ii) studentized residuals and (iii) studentized residuals using a robust estimate of the standard deviation. Once residuals are standardized they are subjected to a Bonferroni-Holm significance test (Holm, 1979). The five methods were performed at a trial level of a rye breeding cycle dataset, i.e., GCA1-2009 + GCA2-2010 + FACT-2011. Outliers by trial were identified and deleted and genomic selection - cross validation (GS-CV) implemented. The predictive abilities (correlation between the observations and the predictions of the validation sets) were computed.

Outcome
We reviewed and published the PlabStat outlier detection method step-by-step, supplying with our publication the codes for this and the other methods in SAS and R language together with a thorough study of the properties, advantages and disadvantages of each method. We recommend using an outlier identification procedure in routine analysis as it helps to detect obvious outlying observations that may escape eye scrutiny. Since all the methods showed similar performance in terms of false and true positive rates, we recommend a favorite method, which combines a robust scale estimate to standardize residuals and a test that deals with the multiple testing problem.

1.3 Modelling genotype by year interaction (paper #3)
Bernal-Vasquez, A.M., Gordillo, A., Schmidt, M., Piepho, H.P., Genomic selection using historical data when all or most entries are tested only in a single year: Modelling genotype-by-year interaction is crucial (In preparation).

Aim and motivation
In GS, it is of most interest to be able to predict genetic breeding values of genotypes of a future cycle. We call the validation of this procedure genomic selection – forward validation (GS-FV). Having established routines for single-year phenotypic analysis,

the new aim is to develop an approach to achieve the highest predictive abilities using GS-FV. The main challenge is to properly account for the genotype-by-year interaction effect, since the trials of the breeding program are very unbalanced across cycles. If the genotype-by-year interaction is not properly modelled, the GS model will divert part of the marker information into predicting that interaction instead of the genetic breeding values.

Hypotheses

- Using the kinship matrix to model the genotype-by-year interaction effect will allow us to dissect the genotype effect from the genotype-by-year effect in datasets with none or very few genotypes in common
- Including sufficient common ancestors in the training set may improve the predictive ability since the chance of having segregating parental lines increases.

Scenarios and datasets

We have used two datasets: (i) one complete cycle to evaluate a GCA1 trial 3 and 4 years later, and (ii) two consecutive two-year cycles to evaluate a GCA1 trial 3 and 4 years later.

Within the datasets, predictive abilities for scenarios with different parent composition between the training set and the prediction set have been evaluated.

Outcome (so far)

A preliminary observation is that using the kinship matrix to dissect the genotype-by-year effect from genotype effect is beneficial for scenarios whose training set includes genotype entries from several cycles. As expected, the more related the training and the prediction sets, the better the predictive abilities. Nevertheless, we still need to separate the effect of sample size in our results.

2. Notwendigkeit und Angemessenheit der geleisteten Arbeit

As long as phenotypic and genotypic data become available, it is imperative to efficiently integrate both information sources to support selection decisions. All steps involved are important and should be considered, studied and refined. This is why we put big effort on the pre-processing stage, the establishment of proper phenotypic models and the integration into the genomic selection frame.

3. Nutzung und Verwertung

A genomic selection strategy was described using phenotypic spatial and non-spatial models for weakly connected trials. The statistical codes are published and available. Detailed description and codes in SAS and R for PlabStat outlier detection method as well as for the other methods are published and available.

4. Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Within the consortium

Communication with KWS-L and TUM has flowed smoothly in regard to all decisions and analyses. The models and strategies discussed in the first paper have served as basis for KWS-L to establish a genomic selection pipeline, and for TUM to develop further research within and across rye cycles. The complete understanding of the current outlier detection methods used by KWS-L allows them to modify and improve their methods or implement any other suggested method.

Within the PLANT2030 community

Collaboration of Professor F. Utz from the plant breeding institute in the University of Hohenheim was crucial for the investigation of the outlier methods.

6. Publications

- Bernal-Vasquez, A.M., Möhring, J., Schmidt, M., Schönleben, M., Schön, C.C., and Piepho, H.P. (2014). The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics*, 15:646.
- Bernal-Vasquez, A.M., Utz, H.F., Piepho, H.P. (2016). Outlier detection methods for generalized lattices: A case of study on the transition from ANOVA to REML. *Theoretical and Applied Genetics*. doi:10.1007/s00122-016-2666-6.
- Bernal-Vasquez, A.M., Gordillo, A., Schmidt, M., Piepho, H.P., Genomic selection using historical data when all or most entries are tested only in a single year: Modelling genotype-by-year interaction is crucial (In preparation).

References

- Piepho, H.-P., Möhring, J., Schulz-Streeck, T., & Ogutu, J. O. (2012). A stage-wise approach for the analysis of multi-environment trials. *Biometrical Journal*, 54:844–60.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.
- Anscombe, F. J., & Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics*, 5:141–160.
- Utz, H.F. (2003). PLABSTAT. Manual. http://www.uni-hohenheim.de/ipsp/soft.htlm. Version 3A of 2010-07-19

B Work Package 2B: Population Genomics

I Kurzdarstellung der Ergebnisse

1. Aufgabenstellung

The main goal of the work package was to characterize patterns of genetic variation in the two main breeding pools of rye as well as in the exotic material by analyzing SNP genotyping and resequencing data. Furthermore, tests of selection should identify genomic regions, which were affected by divergent selection between the seed and the pollen parent pools.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Since this work project mainly relied on the data that was provided by the project partners, methods were implemented and adapted for the analysis of SNP and sequencing data until the project data from SNP arrays and Illumina sequencing became available. The key methods for the population genetic analyses were available as software modules implemented in stand-alone programs or R scripts and were immediately available to the project.

3. Planung und Ablauf des Vorhabens

The project was essentially conducted as descripted in the proposal, and progress was discussed at regular project meetings.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

The work package was conducted under the premise that numerous methods for the description of genetic diversity, genetic population structure and selection were published. However, very little was previously known about the genetic diversity and structure of breeding pools of rye and its relationship to their ancestral populations. For this reason, this work package can be considered the first work to characterize genome-wide diversity in rye using state of the art genome-wide genetic diversity, which allowed to conduct powerful tests of selection.

5. Zusammenarbeit mit anderen Stellen

As described in the proposal, the collaboration was conducted with project partners, in particular TUM by exchanging genotyping data and sequencing data, as well as genetic map information, with IPK to discuss the sequence assembly, and with KWS regarding the design and analysis of the SNP array.

II Eingehende Darstellung

1.1 Genome-wide analysis of genetic diversity in rye breeding populations

The first goal of the population genomic part of the RYE-SELECT project was to carry out a population genomic analysis of the 10 resequenced rye genomes from the seed and pollen parent pools. The original sequencing of the Lo7 line and resequencing of additional 5 seed and 5 pollen parent pool lines, together with a *S. vavilovii* outgroup was performed by the project partners at TUM. The raw data was then processed at IPK and a dataset containing all 12 individuals and around 90 million unfiltered SNPs was sent to us for further analysis.

Since the goal of this part of our project was to establish the amount of genomic diversity within individual pools and also the differentiation between them, we split the dataset into two based on the pool from which the lines originated. Based on the information from scientific literature we filtered the datasets and have ended up with 1,354,163 SNPs in 186,788 contigs from the seed parent pool and 1,546,363 SNPs in 207,302 contigs from the pollen parent pool. Using a number of custom python and R scripts we measured nucleotide diversity, population mutation rates and standardized difference of nucleotide diversity measures within the two pools (Table 1). Lower values observed in the seed parent pool are most likely explained by a larger number of the selfed generations in comparison to the pollen parent pool (S5/S6 vs S2/S3) . Positive values of Tajima's *D* could be explained by a variety of effects; genome-wide balancing selection, recent population reduction or a within breeding pool population structure. Due to the small number of resequenced samples in each breeding pool it is to conduct further tests the provide an explanation for the observed data.

Table 1: Diversity parameters of the breeding pools

Pool	θ_w	π	Tajima's D
Seed Pool	0.0014	0.0024	0.297
Pollen Pool	0.0016	0.0030	0.498

1.2 Differentiation between the two breeding pools

To look at the differentiation between the breeding pools using the resequenced data we calculated the F_{ST} value for contigs containing 20 or more SNPs. The observed differentiation (F_{ST}=0.089) seems to be low on a genome-wide level, genome-wide but ~10% of the contigs showed high levels of differentiation (F_{ST}>0.25), which would suggest that certain genomic regions have strongly differentiated since the split of the two heterotic breeding pools.

In addition to the resequenced lines project partners from TUM have designed a Rye600k array that contains 600,843 SNPs which they used to genotype 46 exotic accessions, 43 seed and 47 pollen parent pool lines. The larger number of

individuals makes this dataset a much better resource to test for differentiation between the breeding pools.

From the 600,843 SNPs, 78,731 came with a genetic map position in cM. We split the data into three populations; exotic, pollen and seed and used those SNPs to calculate the F_{ST} between them. It was interesting to see that differentiation between the two breeding pools was much higher (F_{ST}= 0.318) for the array than for the resequenced data. Additionally, differentiation between exotic accessions and the breeding pools was only half as high (exotic vs pollen F_{ST}=0.149, exotic vs seed F_{ST}=0.179) which indicates a possibility of pool specific selection processes. The higher values of differentiation between array and resequencing data could be explained by different number of individuals as well as with ascertainment bias of the Rye600k array. We also performed a discriminant analysis of principal components on the Rye600k array and have observed a strong differentiation between the three populations (Figure 1).

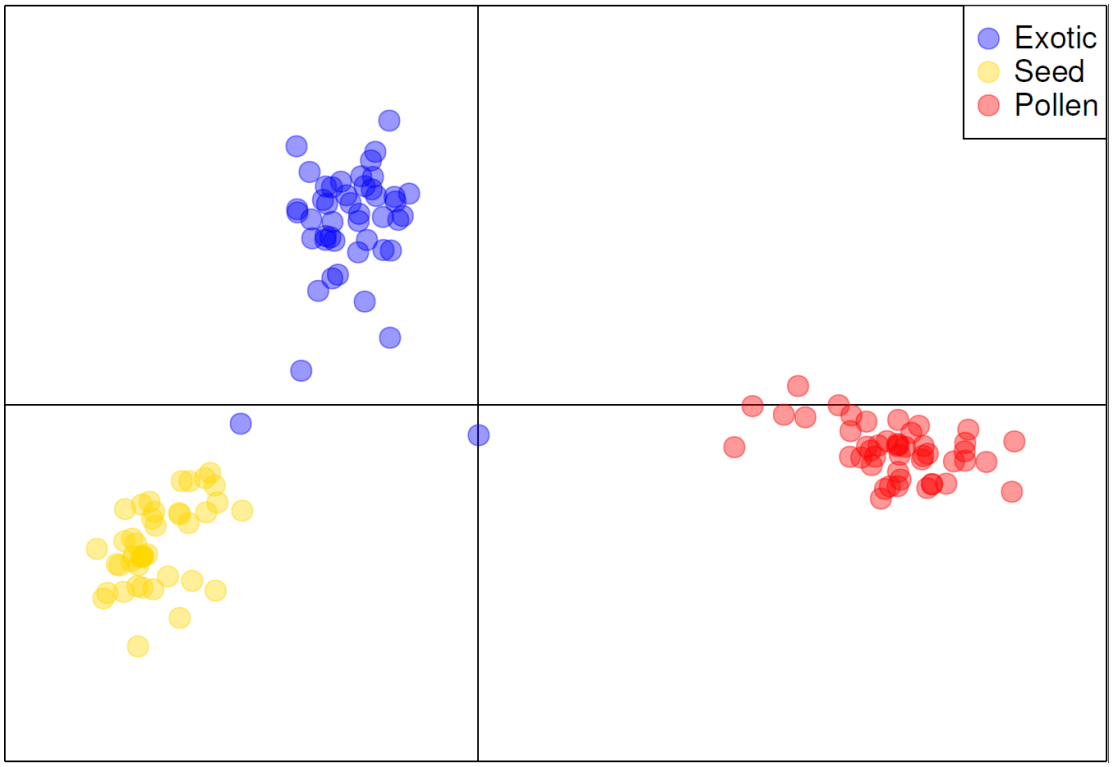

Figure 1: DAPC of the exotic accessions, pollen and seed pool lines based on the 78,731 SNPs from the Rye600k array

The results of this analysis suggest that the diversity levels of the two breeding pools are comparable to similar crops and that there exists a significant level of differentiation between the two breeding pools and the exotic accessions. Using this information we could in the future find interesting candidate genes or regions to use in the improvement of modern breeding varieties.

1.3 Tests of selection

The third and final goal we set was utilizing existing tools and methods to detect selection in seed and pollen breeding pools. Using the 11 resequenced lines from the breeding pools and the Rye600k array we performed multiple tests of selection. First, we used OmegaPlus and SweeD to detect signals of selection in the resequenced lines. With OmegaPlus we looked at the linkage disequilibrium over a certain length of DNA. To find a pattern of a selective sweep (a fixed mutation flanked by regions of excessive LD), it calculates the ω statistics. High values of the ω suggest a possible selective sweep. In similar fashion SweeD implements a composite likelihood ratio test for detecting selective sweeps. It looks at the site frequency spectrum pattern of the SNPs and looks for a region where around a fixed mutation the SFS has been shifted to low- and high-frequency derived variants giving it a high score.

Based on the description of the methods it is easy to discern that the quality of results will depend on the number of individuals and on the length of DNA sequences used. Since we have 6 lines from the seed and 5 lines from the pollen parent pool we needed to be careful with the interpretation of the results to avoid losing to much information as well as gaining false positive values. To counteract this we performed both of the aforementioned tests on contings in our two datasets and combined the results to get only those values that appear in the top 1% in both methods (Figure 2). With this approach we identified 3,240 contigs in the pollen and 3,149 contigs in the seed parent pool as best selection candidates.

We then looked into 78,731 genetically mapped SNPs from the Rye600k array in order to detect selection between the breeding pools as well as between the pools and exotic accessions. For this part of the analysis we used the Lositan and Bayenv2.0 tools. In both of the methods we look at the population differentiation and use it to identify sites under selection. Lositan uses an F_{ST} outlier detection method. It calculates the average neutral F_{ST} of the dataset and detects SNPs that are candidates for positive selection as outliers.

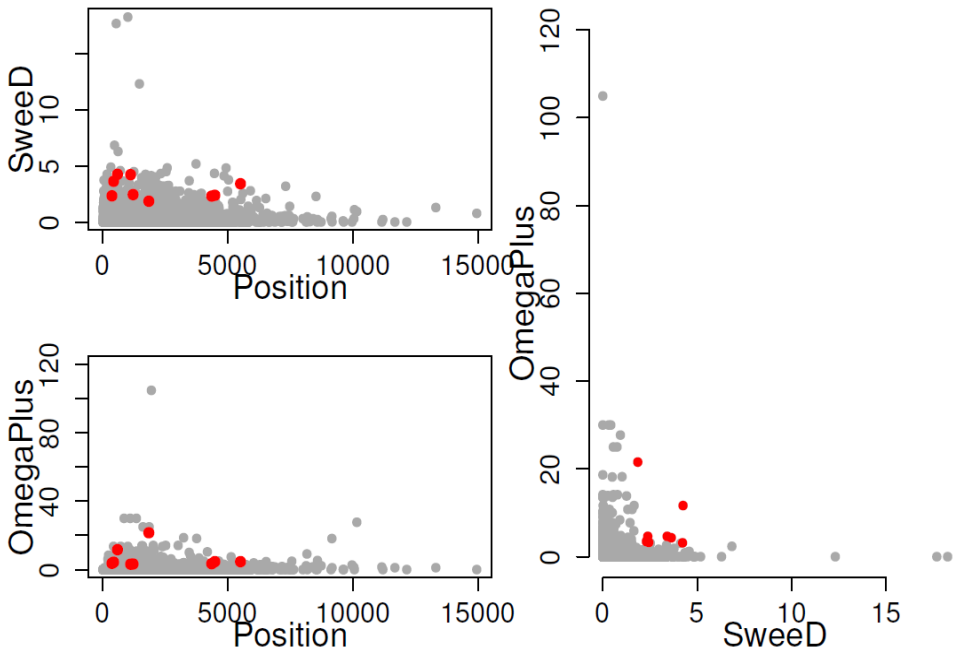

Figure 2: SweeD and OmegaPlus over a set of contigs with the intersect of top 1% values in both methods represented by red dots. Position indicates the position of SNPs within assembled contigs.

To find our candidate SNPs we made three population pairs: pollen *vs*. seed parent pool, pollen parent pool *vs*. exotic material and seed parent pool *vs*. exotic material. According to the method the F_{ST} outliers found are considered to be positively selected. To be sure to eliminate false results we performed three independent Lositan runs and have kept only those SNPs that appear as outliers in all of them. With this method we found 3,846 SNPs as outliers between the seed and pollen, 1,172 SNPs between the seed and exotic and 963 SNPs between the pollen and exotic pools. The higher number of SNP outliers found between the two breeding pools suggests that selection, probably artificial, has differentiated them more in the short amount of time since their split than it had done in regard to the exotic lines. We also identified 19 SNPs as being shared between the exotic vs. pollen and exotic vs. seed parent pool comparisons. These SNPs could possibly reflect the adaptation to Central European environment.

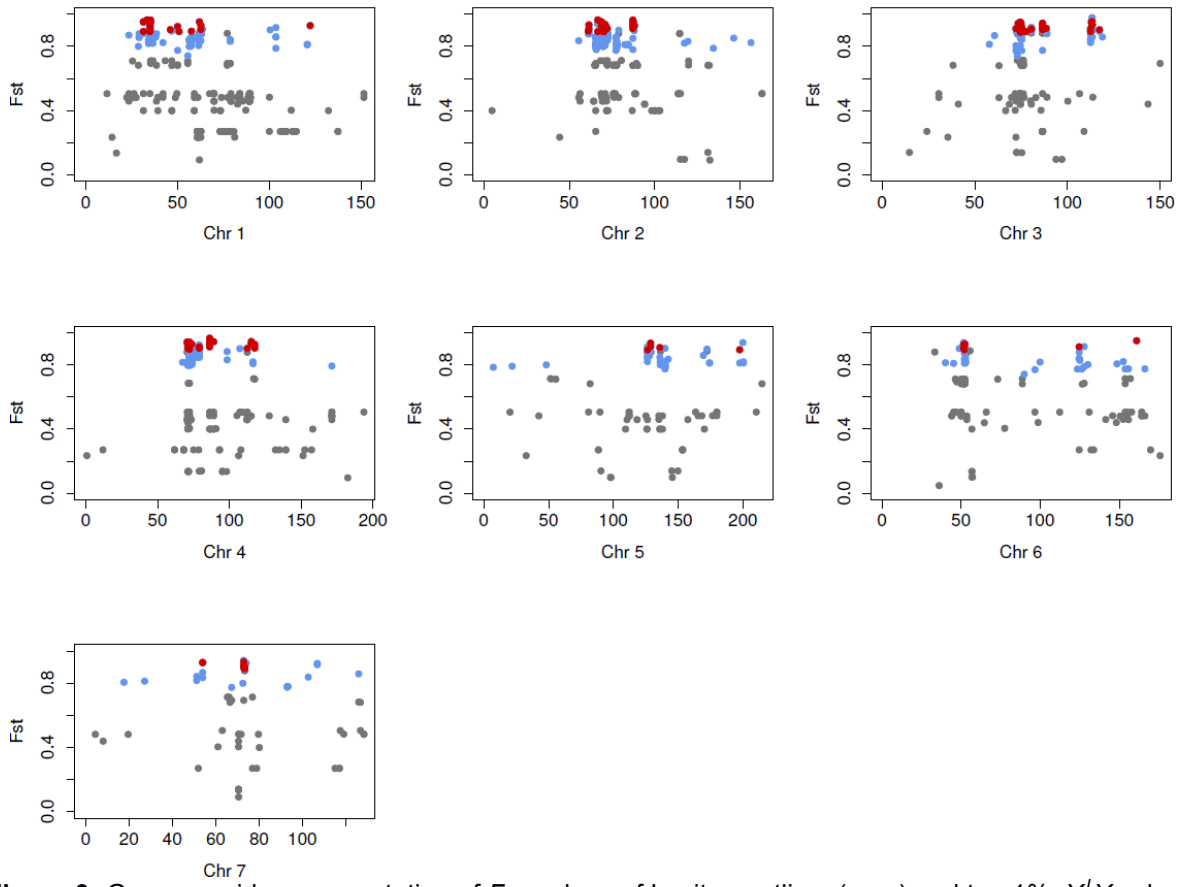

Figure 3: Genome-wide representation of F_{ST} values of Lositan outliers (grey) and top 1% $X^{t}X$ values (blue) together with their intersect (red).

Finally we used the Bayenv2.0 to detect strongly differentiated SNPs between our three population pairs. Bayenv2.0 tests for unusual allele frequency patterns and correlations between those frequencies and environmental variables. It calculates a set of standardized allele frequencies that can be used to detect SNPs deviating strongly from the neutral model. Its X^TX statistic is closely related to F_{ST} and we used it to do a comparison with the Lositan results and validate the selection candidate SNPs. As previously with SweeD and OmegaPlus we compared the top 1% of Bayenv2.0 with Lositan outlier results. We found a strong overlap of highly differentiated SNPs with both methods. We made a genome-wide map of signals of genetic differentiation between the breeding pools where we can see that outliers cluster in few regions that could contain targets of divergent selection (Figure 3).

By doing genome-wide and Rye600k array based selection screens we identified regions of potential selective sweeps in the rye genome. Additionally we discovered SNPS that may be associated with the adaptation to Central European environment, information which could prove useful to breeders working on future elite rye breeding material. Finally we discovered a multitude of SNPs that differentiate strongly between the two breeding pools. These SNPs suggest fixation and selection of different trains within the breeding pools and could therefore be able to contribute to the heterosis effect. As a proof of concept we also successfully applied the Bayenv2 approach to characterize genetic differentiation in whole genome resequencing data (Günther et al., Mol. Ecol., in Revision).

2. Notwendigkeit und Angemessenheit der geleisteten Arbeit

This work package was conducted by the University of Hohenheim in collaboration with the project partners and was aimed at characterizing genome-wide patterns of genetic variation and to evaluate, which genomic regions were strongly affected by the selection process in breeding programs. The work was successfully conducted. Funding of this project by federal funds was required because a funding by own funds of the University of Hohenheim was not possible.

3. Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

The most important result of this work package is that the genetic diversity of rye populations is strongly influenced by the breeding process, and genomic regions affected by this could be identified. This information is of great use for the management of genetic diversity of breeding populations. Furthermore, the methods and tools applied and further developed in this project are available for ongoing breeding efforts and were communicated to the industrial partner. They are therefore available for utiliziation in ongoing breeding programmes.

4. Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

To our knowledge, there are currently no other projects or publications aimed at characterizing genetic diversity in rye populations.

5. Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 11:

Eingereichte Publikationen:

Günther, Lampei, Barilar and Schmid. Phenotypic and genomic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian Alps. Molecular Ecology, in Revision.

Geplante Publikationen:
Bauer et al. Structure, diversity and selection signatures of the rye (*Secale cereal*e L.) genome.

Erfolgskontrollbericht

1. Beitrag des Ergebnisses zu den förderpolitischen Zielen

Ein wichtiges Ziel der Ausschreibung des Plant 2030-Programms war es, die aktuellen Entwicklungen in der Pflanzenforschung weiterzuführen und in die praktische Anwendung einzubringen. Die Genomische Selektion als Zuchtmethode hat in den vergangenen Jahren die Tierzüchtung revolutioniert und ein ähnlicher Prozess vollzieht sich derzeit in der Pflanzenzüchtung. Dieser ist aber durch die Vielzahl an Kulturarten, Fortpflanzungsmethoden und auch Wirtschaftsunternehmen etwas aufwendiger. Durch ein Gemeinschaftsprojekt von Universitäten, ausseruniversitären Forschungsinstituten, der Ressortforschung und einem Züchtungsunternehmen konnte modellhaft die Implementierung und Weiterentwicklung von statistischen Verfahren zur genomischen Zuchtwertschätzung in die Untersuchung von aktuellen Zuchtmaterial und der Fortführung im vorwettbewerblichen Bereich erzielt werden. Der Beitrag des von der Universität Hohenheim durchgeführten Teilprojekts bestand insbesondere in der Verbesserung der Modellierung der Zuchtwertschätzung, der Charakterisierung der genetischen Diversität im Zuchtmaterial sowie der Charakterisierung von genomischen Regionen mit funktioneller genetischer Variation für Qualität, Fruchtbarkeitsrestoration und Ertrag. Dadurch wurde zum einem ein Netzwerk gebildet, welches die förderpolitischen Ziele des beschleunigten Wissens- und Technologietransfers erfüllt und eine Grundlage für weitere Kooperationen mit dem Ziel die Wettbewerbsfähigkeit der deutschen und europäischen Saatgutindustrie zu erhöhen.

2. Wissenschaftlich-technisches Ergebnis, erreichte Nebenergebnisse und gesammelte wesentliche Erfahrungen

Das Vorhaben zeigte, eine signifikante Verbesserung bei der Zuchtwertschätzung erreicht werden kann, wenn räumliche Modelle und eine Identifizierung von Ausreissern bei der Analyse von Feldversuchen verwendet werden sowie verbesserte Validation beim Vergleich von Trainings- und Validierungspopualationen verwendet werden. Des weiteren konnte das Ausmass genetischer Variation in den beiden heterotischen Pools sowie in exotischem Material beschrieben werden. Es wurde gezeigt, dass verschiedene genomische Regionen stark von der Selektion im Rahmen des Züchtungsprogramms beeinflusst werden und die genetische Vielfalt dadurch signifikant verändert wird. Schliesslich konnte durch die Charakterisierung einer Rückkreuzungspopulation die Effekte von funktioneller genetischer Variation mit Restorergenen aus exotischen Quellen die Effekte auf Pollenfertilität, agronomischer und Qualitätsmerkmale evaluiert werden.

3. Fortschreibung des Verwertungsplans

3.1 Erfindungen und Schutzrechte

Innerhalb des Projects RYESELECT wurden vom Berichterstatter keine Erfindungen und Schutzrechte angemeldet.

3. 2. Wirtschaftliche Erfolgsaussichten

Die wirtschaftlichen Erfolgsaussichten der genomischen Zuchtwertschätzung und der Charakterisierung von genomweiter und funktioneller Variation können aufgrund der Ergebnisse dieses Projekts positiv beurteilt werden. Deswegen gehen wir davon aus, dass die geleisteten Arbeiten und die dabei entstandenen Ressourcen zum Erfolg der genomischen Selektion und Hybridzüchtung in der kommerziellen Roggenzüchtung beitragen werden.

3. 3. Wissenschaftliche Erfolgsaussichten

Die wissenschaftlichen Erfolgsaussichten sind gut, was durch die geplanten Publikationen dokumentiert wird. Die im Projekt entstandenen Netzwerke sind für weitere Anschlussprojekte genutzt worden.

3. 4. Wissenschaftliche und wirtschaftliche Anschlussfähigkeit

Die wissenschaftliche und wirtschaftliche Anschlussfähigkeit ist sehr gut, weil die im Projekt entstandenen statistischen Verfahren, Softwareprogramme und genetischen Ressourcen für weitere Arbeiten verwendet werden können.

4. Arbeiten, die zu keiner Lösung geführt haben

Im Wesentlichen haben alle Arbeiten zu einer Lösung geführt.

5. Präsentationsmöglichkeiten für mögliche Nutzer

Die Ergebnisse wurden bei Projekttreffen mit allen Partnern, bei den jährlichen Plant 2030 Statusmeetings und bei nationalen und internationalen Tagungen (z. B. EUCARPIA Biometry Workshop) vorgestellt.

6. Einhaltung der Kosten- und Zeitplanung

Die Ausgabenplanung wurde wie im Antrag beschrieben eingehalten. Die Zeitplanung wurde ebenso im Wesentlichen eingehalten.

Berichtsblatt

<table>
<tr><td>1. ISBN oder ISSN</td><td colspan="2">2. Berichtsart (Schlussbericht oder Veröffentlichung)
Schlussbericht</td></tr>
<tr><td colspan="3">3. Titel
Genom-basierte Präzisionszüchtungsstrategie für Roggen (RYE SELECT) – Teilprojekt E</td></tr>
<tr><td rowspan="3">4. Autor(en) [Name(n), Vorname(n)]
Prof. Schmid, Karl
Prof. Piepho, Hans-Peter
Apl. Prof. Miedaner, Thomas</td><td colspan="2">5. Abschlussdatum des Vorhabens
31.08.2015</td></tr>
<tr><td colspan="2">6. Veröffentlichungsdatum</td></tr>
<tr><td colspan="2">7. Form der Publikation
Report</td></tr>
<tr><td rowspan="3">8. Durchführende Institution(en) (Name, Adresse)
Universität Hohenheim
Forschungsschwerpunkt Biotechnologie & Pflanzenzüchtung (762)
Fruwirthstrasse 21
70599 Stuttgart</td><td colspan="2">9. Ber. Nr. Durchführende Institution</td></tr>
<tr><td colspan="2">10. Förderkennzeichen
0315946E</td></tr>
<tr><td colspan="2">11. Seitenzahl
31</td></tr>
<tr><td rowspan="3">12. Fördernde Institution (Name, Adresse)

Bundesministerium für Bildung und Forschung (BMBF)
53170 Bonn</td><td colspan="2">13. Literaturangaben
4</td></tr>
<tr><td colspan="2">14. Tabellen
-</td></tr>
<tr><td colspan="2">15. Abbildungen
-</td></tr>
<tr><td colspan="3">16. Zusätzliche Angaben</td></tr>
<tr><td colspan="3">17. Vorgelegt bei (Titel, Ort, Datum)</td></tr>
<tr><td colspan="3">18. Kurzfassung
zur Auswertung von Feldersuchen für Zuchtwertschätzung beim Roggen verglichen und optimiert, und als Softwareskripte implementiert. Die genom-weite genetische Variation von Züchtungspopulationen und exotischem Material wurde charakterisiert. Funktionelle genetische Variation von Restorergenen wurde in Feldversuchen im Hinblick auf Pollenfertilität, Qualitäts- und Ertragsmerkmale evaluiert.</td></tr>
</table>

19. Schlagwörter Populationsgenetik, Zuchtwertschätzung, Roggen, QTL, Selektion	
20. Verlag	21. Preis