Genomic Prediction of Stripe Rust Resistance in Wheat - Opportunities, Limitations and Lessons Learned from a Breeders Population

Hermann Gregor Dallinger, BSc

01540731

Master Thesis

in the field of Plant Sciences (UH 066 455)

submitted April 2020

Supervisor: Hermann Bürstmayr, Univ.Prof. Dipl.-Ing. Dr.nat.techn.
Co-Supervisor: Sebastian Michel, MSc. Dr.
Institute of Biotechnology in Plant Production (H97100)

Eidesstattliche Erklärung / *Affidavit*

Hiermit versichere ich an Eides statt, dass ich die vorliegende Masterarbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch bei keiner anderen Prüferin/ keinem anderen Prüfer als Prüfungsleistung eingereicht.

Mir ist bekannt, dass Zuwiderhandeln geahndet wird („Verwendung unerlaubter Hilfsmittel“) und weitere rechtliche Schritte nach sich ziehen kann.

I hereby swear that I have compiled this Master Thesis without external help and without using sources and aides other than those permitted and that the sources have been cited verbatim or quoted textually in the places indicated.

This work has not been submitted in the same or similar form to any other examiners as a form of examination.

I am aware that offenders may be punished ('use of unauthorized assistance') and that further legal action may ensue.

Date	Unterschrift / *Signature*

Acknowledgements

First and foremost, I want to thank the supervisors of my thesis, Prof. Hermann Bürstmayr for allowing me to work in his department, and Sebastian Michel for providing the idea to this thesis, for co-supervising, and for patiently answering my questions in countless meetings. I also want to thank Barbara Steiner for supporting me in navigating the reference genome browser.

I also want to thank Saatzucht Donau GesmbH & CoKG for providing the phenotypic data, which was gathered in numerous hours of field scoring, and providing the genetic data. Especially I want to thank Anton Neumayer and Franziska Löschenberger for helping me to interpret the phenotyping results.

Further credit goes to Michael Oberforster at AGES for providing literature on stripe rust in Austria. For writing the R-package sommer that aided the statistical analysis in this study, and for answering my questions on it, I want to thank Giovanny Eduardo Covarrubias Pazaran.

I also want to thank my family, for always encouraging me to follow my passions and carry on studying. Special credit goes to my sister Martina for spending time and effort on proofreading this work and Angelika for ensuring my emotional well-being during the making of it.

Abstract

Bread wheat is one of the most important food crops worldwide and stripe rust is one of the most devastating diseases of wheat, especially in temperate climates. Recent epidemics, caused by recombination of the pathogen in the Near-Himalayan center of diversity, are rendering breeding efforts void and demand novel solutions. Resistance breeding relies on phenotypic and marker-assisted selection to combine resistance genes in high yielding genotypes. New genotyping methods, falling costs, and improved statistical methods enable prediction and selection based on a large number of markers in plant breeding. The goal of this thesis was to use these tools and methods, to improve quantitative, non-race-specific resistance against stripe rust using data from wheat breeding trials in the years of 2013, 2014, 2015 and 2016, as well as genotyping by sequencing (GBS) markers. The genetic architecture was analysed and two large effect QTLs were found, one on chromosome 2A, largely fixed in the population and one on 2B, found in low frequency and therefore highly interesting for future breeding efforts. Evidence was found that the QTL on chromosome 2A might be associated with the translocation 2AS-2N. Genomic best linear unbiased predictor (GBLUP) models were used to predict stripe rust within- and across-years. Prediction performance was medium to high within-years, but predictions failed across-years. There was some variability across years (and across trials), possibly due to different races of stripe rust, that impaired genomic prediction. A prediction model can only perform as well as the data that was used to train the model and when a model is trained with conflicting data the predictions will not be reliable. This thesis concluded, that breeder's trial data could show strong variation ultimately impairing the power of genomic prediction models. Data has to be investigated thoroughly, and only the best trials have to be selected to build prediction models.

Keywords: stripe rust resistance, resistance breeding, genomic prediction, genomic selection, wheat breeding, genetic architecture, QTL, GWAS

Zusammenfassung

Weizen ist eine der wichtigsten Nutzpflanzen weltweit, wobei Gelbrost eine der verheerendsten Weizenkrankheiten ist. Epidemien die auf die Rekombination des Erregers im Zentrum der Diversität im Himalaja zurückzuführen sind machen bisherige Erfolge der Züchtung zunichte und erfordern neuartige Lösungen. Resistenzzüchtung nutzt dazu phänotypische und markergestützte Selektion um Resistenzgene in ertragreichen Genotypen zu kombinieren. Neue, kostengünstige Methoden der Genotypisierung mit hoher Dichte und statistische Methoden ermöglichen die Selektion auf der Basis von genetischen Markern. Ziel dieser Arbeit war es die quantitative Resistenz gegen Gelbrost in Weizen anhand von Daten aus Zuchtversuchen in den Jahren 2013, 2014, 2015 und 2016 sowie Genotypisierung durch Sequenzierung (GBS) Markern zu verbessern. Die Analyse der genetischen Architektur lieferte zwei QTL mit großer Wirkung. Jener auf Chromosom 2A ist annähernd fixiert in der Population, ein weiterer auf Chr. 2B dessen resistentes Allel nur in wenigen Genotypen vorhanden ist. Es wurden Hinweise gefunden, dass der QTL auf Chromosom 2A mit einer Translokation assoziiert ist (2AS-2N). GBLUP Modelle wurden verwendet, um Gelbrost innerhalb von Jahren und über Jahre hinweg vorherzusagen. Die Genauigkeit dieser Modell war innerhalb der Jahre mittel bis hoch, über mehrere Jahre hinweg jedoch gering. Die Ursachen dafür sind in der Variabilität der Versuche und Jahre zu suchen, welche zum Teil durch verschiedene Gelbrost-Rassen erklärt werden kann. Die Leistungsfähigkeit eine Modells ist abhängig von der Qualität der Trainingsdaten, trainiert man ein Modell mit widersprüchlichen Daten sind die Vorhersagen nicht zuverlässig. Diese Arbeit kam zu dem Schluss, dass Versuchsdaten starke Unterschiede aufweisen können, die letztendlich die Anwendung genomischer Vorhersagemodelle beeinträchtigen. Um zuverlässige Modelle zu erhalten, sollten nur die besten Versuche zum Trainieren von Vorhersagemodellen auswählt werden.

List of Abbreviations

AGES	Austrian Agency for Health and Food Safety Ltd.
ANOVA	Analysis of variance
APR	Adult plant resistance
ASR	All stage resistance
AU	Austria
BAES	Austrian Authority for Food Safety
BayesR	Bayesian regression model
BLASTN	Basic local alignment search tool (for nucleotides)
BLUP	Best linear unbiased prediction
BOKU	University of Natural Resources and Life Sciences Vienna
CA	Canada
CIMMYT	International Maize and Wheat Improvement Center
CV	Cross validation
DArT	Diversity array technologies
DE	Germany
DH	Doubled haploid
DNA	Deoxyribonucleic acid
DOE	Dörfles
et al.	and others (*et alia*, *et alii*, *et aliae*)
e.g.	example given
FR	France
GBLUP	Genomic best linear unbiased prediction
GBLUPA	Genomic best linear unbiased prediction, fixed QTL effects
GBS	Genotyping by sequencing
GEBV	Genomic estimated breeding values
GRM	Genetic relationship matrix (G)
GS	Genomic selection
GP	Genomic prediction
GWAS	Genome-wide association study
GxE	Genotype by environment
HU	Hungary
i.e.	meaning (*id est*)

INDEL	Insertions and deletions
IWGSC	International Wheat Genome Sequencing Consortium
SR	Seedling resistance
SZD	Saatzucht Donau GesmbH. & CoKG
JKI	German Federal Research Centre for Cultivated Plants
LEO	Leopoldsdorf
LMM	Linear mixed model(s)
MAF	Minor allele frequency
MLK	Melk
MAR	Marchtrenk
MAS	Marker-assisted selection
NA	not available / missing
NGS	Next generation sequencing
PCR	Polymerase chain reaction
PRO	Probstdorf
PRkf	Probstdorf small plots early flowering
PRks	Probstdorf small plots late flowering
PYT	Preliminary yield trials
QTL	Quantitative trait loci
RCD	Replicated control design
RBG	Reichersberg
RI	Recombinant inbred
RO	Romania
rrBLUP	Ridge regression best linear unbiased prediction
RS	Serbia
SK	Slovakia
SNP	Single nucleotide polymorphisms
SSR	Simple sequence repeat (marker)
THE	Thrace region, Turkey
TR	Turkey
WEI	Weikendorf

Index of Figures

Figure 1: Microscopic images of P. striiformis, Uredia with spores emerging. In the corner are images of one uredospore and two teleutospores.....2
Figure 2: Life cycle of *Puccinia striiformis f. sp. tritici* (Jin *et al.*, 2010).....2
Figure 3: Schematic of data preparation, showing the two stage analysis approach.....12
Figure 4: Heritability plot of trials versus years.....21
Figure 5: Stripe rust (YR) adjusted means per year (1-9 score).....22
Figure 6: K-medoids clustering results.....23
Figure 7: Correllogram of all trials used for across-year phenotypic and genetic analysis.....25
Figure 8: Pairwise observations of selected trials to demonstrate variability in trials.....26
Figure 9: Chromosome 2A details of QTLs “2-2014/2-2015” on the left and “3-2014/3-2015” on the right.....28
Figure 10: Cumulative effects of QTL for each year using QTL from 2014 or 2015.....30
Figure 11: Pairwise correlations of QTL found in the GWAS with correlation coefficients in the top right and colour coded circles according to coefficients in the bottom left.....31
Figure 12: Effects of training population size and marker density on within-year prediction ability per year calculated from 160 simulations.....34
Figure 13: Across year prediction abilities of 224 simulations.....35
Figure 14: Allele frequency distribution in the population.....i
Figure 15: A representation of the genetic map.....ii
Figure 16: Heterozygosity distribution histogram among lines.....iii
Figure 17: YR scorings and adjusted means at different stages of mixed modeling.....vii
Figure 18: Correlations of genotypes.....ix
Figure 19: Pairwise correlations of genotypes per year.....ix
Figure 20: Detailed clustering results.....x
Figure 21: Manhattan plot of GWAS results 2013.....xiii
Figure 22: Manhattan plot of GWAS results 2014.....xiii
Figure 23: Manhattan plot of GWAS results 2015.....xv
Figure 24: Manhattan plot of GWAS results 2016.....xv
Figure 25: QTL beta effects vs. phenotypic values.....xvi
Figure 26: GS within years, variations in marker density, and training population size.....xvii

Index of Tables

Table 1: Marker density per chromosome after filtering for MAF (2582 total).....15
Table 2: Software packages used in this thesis.....20
Table 3: Pairwise correlations of scoring across years.....22
Table 4: GWAS results per year, grouped into potential QTL.....29
Table 5: GBS markers, QTL and best-fitting locations on the reference genome of wheat.....32
Table 6: Prediction models ANOVA table, predictions across years, response prediction ability.....36
Table 7: Prediction models ANOVA, table of effects.....36
Table 8: Results from preliminary yield trials simulations using all entries from training year for model training.....37
Table 9: Results from preliminary yield trials simulations using only PYT entries for model training.....37
Table 10: Trials removed due to low number of plots (<55).....iii
Table 11: Trials removed due to low variance (<0.21).....iii
Table 12: Trials removed due to problems in mixed modeling.....iv
Table 13: Trials removed due to low heritability results from mixed modeling.....iv
Table 14: Overview trials usable for mixed modeling.....v
Table 15: Trial heritabilities.....vi
Table 16: Clustering solution for four cluster.....xi

Index of Equations

Equation 1: Linear mixed models.....12
Equation 2: Broad sense heritability.....13
Equation 3: LMM first stage.....13
Equation 4: Rescaling BLUEs.....13
Equation 5: LMM for second stage.....14
Equation 6: LMM for GWAS.....16
Equation 7: Bonferroni correction.....17
Equation 8: LMM for GBLUP.....17

Table of Contents

1 Introduction....1
1.1 Stripe Rust of Wheat....1
1.2 Resistance Breeding....5
1.3 Genetic markers and genome-wide association mapping....6
1.4 Genomic Selection....8
1.5 Aims of this thesis....9
2 Materials and Methods....11
2.1 Plant material and trial locations....11
2.2 Trial designs....11
2.3 Phenotypic Analysis....12
2.4 Genetic Analysis....15
2.5 Statistics software....20
3 Results....21
3.1 Phenotypic analysis....21
3.2 Scoring results across years....22
3.3 Scoring results across trials....24
3.4 Genome-wide association study for stripe rust resistance....27
3.5 Locating QTL in the wheat reference genome....32
3.6 Genomic prediction within years....33
3.7 Genomic prediction across years, for untested and tested years....34
3.8 Genomic prediction and preliminary yield trials....37
4 Discussion....38
4.1 Phenotypic Analysis....38
4.2 Genome-wide association study....40
4.3 Genomic prediction....44
5 Summary, Outlook, Conclusion....49
6 Literature References....50
7 Appendix....i
7.1 Results from marker data....i
7.2 Trial information....iii
7.3 Adjusted means....vii
7.5 Clustering results....x
7.6 GWAS results....xiii

1 Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important crops in Austria, with 280 000 ha planted in 2018 (BMNT, 2019) and has worldwide importance as a food crop, providing, on average, roughly 20% of the total calories and protein of the daily total dietary intake (Shiferaw *et al.*, 2013). Wheat is vulnerable to many diseases, one of which is stripe or yellow rust (*P. striiformis f. sp. tritici Eriks. & Henn*). Stripe rust epidemics do not occur every year, high occurrence depends on autumn infections, mild winters and favourable (cool, moist) weather in spring, which can lead to severe yield reduction of 15 to 40% (Hanson *et al.*, 1982; Oberforster, 2015), making it a serious threat to global food production and food security.

In the developed world, stripe rust is controlled by applying fungicides, but in the developing world, subsistence farmers often do not have access to fungicides (Miedaner, 2017). Stripe rust is also controlled by resistant varieties, but with the introduction of new races in Europe, which happened for example in 2011, formerly resistant varieties might be highly susceptible today (GRRC, 2020). Resistance breeding relies on the application of single or multiple resistance genes, coming from diverse gene pools. In times of rapidly emerging races of stripe rust, the selection of single resistance genes in a breeding scheme do not lead to durable varieties. Therefore another approach, selecting not only for qualitative, single gene resistance, but for quantitative resistance, distributed over the whole genome, may be more favourable (Ellis *et al.*, 2014).

1.1 Stripe Rust of Wheat

Rusts are some of the most devastating fungal diseases since the first cultivation of cereals, and their spores were found on archaeological sites in Israel dating back to 1300 BC. In the 1950s, Stem Rust resistant varieties were developed and disseminated in Africa, helping farmers in this region to control this devastating disease. However, in 1998, a new race of Stem Rust appeared in Uganda, later called Ug99, which had overcome the previously effective resistance genes. This led to heavy infestations in East Africa and West India, threatening global wheat production and food safety with constantly evolving races of Stem Rust up until today. While in warmer regions, Stem Rust is a major problem, in cooler climates newly emerging stripe rust races and epidemics have and are recently and currently threatening wheat production (Miedaner, 2017).

Stripe rust appears on a number of cereal and grass species, but is of high economic importance in barley (*Hordeum vulagere* L.) and wheat. Both its names, stripe rust and yellow rust, are descriptive

of the symptomatic appearance and can be used to distinguish it from other rusts at a certain point in time (X. Chen and Kang, 2017). Stripe rust appears on wheat in the form of yellow to orange urediniospores, emerging from pustules arranged in long narrow stripes along the leaf veins of its host. In seedlings, the infection sites are not constrained by leaf veins and therefore progress outwards from the point of infection in all directions. Pustules are usually oblong, 0.4 to 0.7 mm in length and 0.1 mm in width. Urediniospores are ellipsoidal and broadly obovoid with a mean of 24.5 x 21.6 µm, yellow to orange in colour and have from 6 up to 18 germ pores on their surface (Figure 1). At around 7 to 12 °C, urediniospores germinate rapidly on leaf blades if water is available. At higher temperatures and later development stages of the host, the fungus is producing black telia, which are similar in size to uredia, but contain two-celled brown or black, thick-walled teleutospores (W. Chen *et al.*, 2014). For a long time it was believed that the teleutospores did not serve any purpose in the life cycle of stripe rust, and infections of wheat would only be possible through urediniospores, until Jin *et al.* (2010), found the secondary host to be *Berberis chinensis*, which is infected by the mentioned teleutospores produced on the primary grass host (Figure 2).

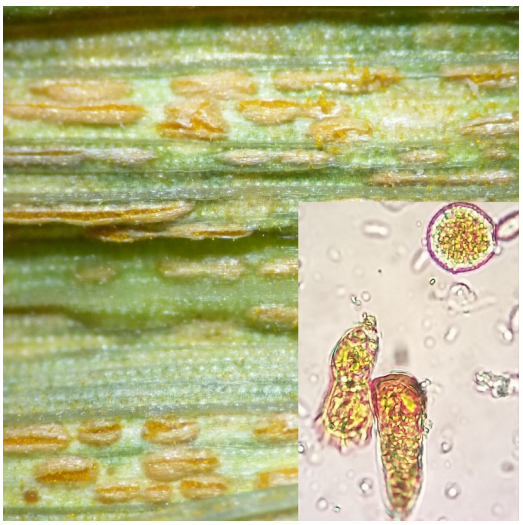


Figure 1: Microscopic images of P. striiformis, Uredia with spores emerging. In the corner are images of one uredospore and two teleutospores.

This does not only show that there is a second host, but while on wheat, stripe rust is reproducing only clonally, on *Berberis chinensis* a sexual stage of reproduction through pycniospores and aeciospores is possible. The details of sexual reproduction are still largely unknown, but this means that genetic diversity in stripe rust is not only possible through mutation but also through genetic recombination (Zheng *et al.*, 2013). Since then, further species were identified to be infectable, with the sexual reproduction stage, under artificial inoculation conditions (M. N. Wang and Chen, 2013).

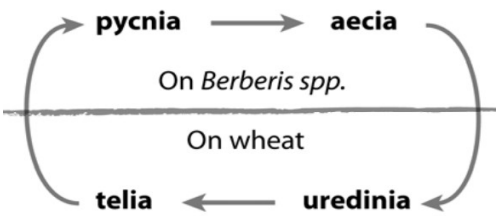


Figure 2: Life cycle of *Puccinia striiformis f. sp. tritici* (Jin *et al.*, 2010).

While in most areas of the world, the clonal reproduction is the main or only way of propagation in stripe rust, Duan *et al.* (2010) showed that Chinese stripe rust populations have a much higher genetic diversity, and annual infections are likely to be originating from the sexual reproduction phase

of the pathogen. The fact that sexual recombination of the pathogen happens only at certain places around the world, allows for the analysis of migration routes across time, showing that European stripe rust isolates usually originate in the Middle-Eastern and East-African areas. Migration of stripe rust has historically not only been documented through natural causes like wind dispersal but also through human activity like travel and commerce (S. Ali *et al.*, 2014). Long-distance travel of stripe rust is documented as far back as 1915, for the first time having been found in the US where it was introduced from Europe, as well as in remote locations like New Zealand in 1980, being introduced from Australia (Hovmøller *et al.*, 2011; Schwessinger, 2017).

A recent epidemic in Europe starting in 2011, was also caused by long-distance dissemination of a novel stripe rust race. The stripe rust race groups "Warrior" and "Kranich" were detected in high frequency at multiple sites firstly in Western Europe, and later all across Europe, overcoming previously durable resistance genes and replacing local rust populations. The origin of the aforementioned races is in the Near-Himalayan region, which is the center of diversity for stripe rust. These newly emerged races differ from previous strains not only in virulence but also in the production of more telia and therefore more spores (Hovmøller *et al.*, 2016).

Historical control of rust diseases included early harvest at signs of infections, as well as pulling nets and cords across the fields to remove morning dew and reduce humidity, which aids spore germination. Other strategies against rust diseases included eradication of the alternate hosts, Berberis sp. as far back as the 1660s in France (Zadoks and Bouwman, 2017). Nowadays, synthetic fungicides are a measure of control in integrated pest management. In Austria, currently 55 different synthetic fungicides are available and registered for the control of stripe rust, containing 13 different active components and combinations thereof (BAES, 2020). While the availability (local registration), effectiveness (pathogen resistance) and societal acceptance (residues, environment) of chemical fungicides are steadily declining, an alternative measure of control is the use of resistant cultivars.

When breeding for resistant cultivars one has to take into account the relationship between host and pathogen. In case of wheat and stripe rust this is on the one hand how the pathogen is able to infect the host, and on the other hand how the host defends itself against infections. Stripe rust is an obligate parasite, meaning it depends on the host for growth, survival and reproduction. If it cannot infect the host, it is not able to complete its life cycle. As previously mentioned, rust fungi have a rather complex life cycle, producing up to five different types of spores on two different hosts. This section will only focus on the uredinial stage, since uredospores can lead to new infections, which

again lead to the production of the same spores. The uredinial stage is the one which is most important for epidemic occurrence of stripe rust on wheat, while other stages are necessary for sexual recombination and the emergence of new races. Once urediniospores are deposited on wheat leaves, heads or awns by rain or wind, they germinate after approximately 3 hours by producing a germ tube. Within 6 to 8 hours post germination, the germ tube penetrates a stoma and enters into the plant tissue. Within the mesophyll cells, in the space between cell walls and plasma membrane, haustoria are produced that extract nutrient from the cells to supplement the formation of more fungal hyphae and further haustoria. Once enough nutrients are gathered, spore buds develop and start producing urediniospores. Sporulation starts around 12 to 14 days after inoculation and newly emerging spores then infect other plants and new leaves (W. Chen *et al.*, 2014; Z. Kang *et al.*, 2017).

Stripe rust cannot infect every plant species, a compatible plant-host interaction is necessary for infection and within compatible host species some genotypes can be resistant to infection. This resistance can be largely categorized into two main classes commonly designated as Adult Plant Resistance (APR) and All Stage Resistance (ASR), or Seedling Resistance (SR) since it is tested on seedlings. The level of ASR is usually complete or at least very high, but also race-specific and therefore prone to the risk of being overcome by new races. APR is usually non-race-specific and therefore more durable, but can be of low level and insufficient if disease pressure is high and conditions are favourable for the pathogen. A series of host responses are involved in seedling resistance, including cell wall apposition, papilla formation, accumulation of antifungal compounds including lignin, plant hydrolases, and reactive oxygen species. Hypersensitive reaction is accompanied by cell apoptosis at the infected location and considered one of the most effective plant defence strategies. Similar reactions can be associated with APR, it is linked to reduced fungal growth in resistant compared to susceptible cultivars (Z. Kang *et al.*, 2017). High temperatures in spring can induce host resistance (Luig, 1985), which leads to a further distinction of High Temperature Adult Plant Resistance (HTAP), while it is believed that most cases of APR are in fact HTAP (M. Wang and Chen, 2017).

Genes responsible for ASR, APR or HTAP are recurrently being described in literature, 78 are permanently and 67 temporarily named, 327 quantitative trait loci (QTL) associated with stripe rust resistance are described in literature (McIntosh *et al.*, 2017). These resistance genes are associated with ASR, APR or HTAP. Another distinction within susceptible varieties can be done by classifying them into “Slow Rusting” or “Fast Rusting” (Caldwell, 1968). They are defined based on the

spread and development under field conditions, whereby slow rusting is associated with partial resistance, similar to APR. While these terms are loosely defined, they are accepted and valuable from a practical viewpoint. Further classifications are made by distinguishing "monogenic" versus "polygenic" resistance, referring to the number of genes involved, and "major-gene" versus minor-gene" or "qualitative" versus "quantitative" resistance, referring to the effect of each gene (M. Wang and Chen, 2017). Monogenic, major-gene, qualitative, ASR is associated with a high level but non-durable resistance, while polygenic, minor-gene, quantitative APR is associated with varying levels of resistance, but is considered more durable. While some argue that breeding for HTAP is the most effective strategy (X. Chen, 2013) and others warn that also APR genes can break down and every genetic resource should be deployed (Ellis *et al.*, 2014), combining all types of resistance might be the most sustainable breeding strategy to obtain a durable stripe rust resistance.

1.2 Resistance Breeding

When a new stripe rust race appears that is virulent on previously resistant cultivars, resistance breeding has to adapt to the situation by looking for new sources of resistance. This is of course not a quick solution, since it takes many years from crossing two varieties to the release of a new variety. Depending on the source of resistance, this involves one or two main activities, pre-breeding and incorporation into the breeding population. Pre-breeding is necessary if the source of resistance is not agronomically adapted. Aside from hexaploid and tetraploid wheat cultivars, sources of resistance are landraces and crossable related or non-crossable wild species (Aktar-Uz-Zaman *et al.*, 2017). To identify resistance genes in cultivars, regional sets of differential lines have been developed that carry single resistance genes. These are only available for selected ASR genes, since for some races, single gene lines are not available or other genes mask the action of a gene (Wan, Wang, *et al.*, 2017). In some cases, embryo-rescue methods, bridge-crossing or genome-doubling might be necessary, if the resistance source is of a different ploidy level, or genetically incompatible. If normal crossing is possible, successive backcrossing is used to transfer resistance genes into genetically adapted material and to get rid of linkage drag. Genetic markers, that are tightly linked with the genes of interest, might be helpful in selecting the desired lines and speeding up the breeding process (Bariana, 2003). Such genetic markers are especially useful for introgression of genes from exotic material rather than selective breeding within elite material (Würschum, 2012). and can facilitate the pyramidization of resistance genes. If this process, which can take many cycles of crossing and back-crossing, is completed, the resulting genotypes can be transferred into elite breeding populations which yield variety candidates (Bariana, 2003).

A range of new biotechnology tools is becoming readily available, partly due to advances in genotyping and sequencing technologies. DNA markers are a valuable tool in the pyramidization of resistance genes, since screening for multiple genes is not possible when one gene is masking the effects of others. Some efforts, accelerated by recent advances in genome editing, focus on the development of genetic constructs of multiple resistance genes and types of genes that can be integrated using genetic modification or gene editing techniques (Ellis *et al.*, 2014). While in some parts of the world, genetic modified food is considered safe, public acceptance towards genetic modification is low in Europe (Lucht, 2015). Recent resolutions by the Court of Justice of the European Union, classify new gene editing techniques in the same category as established transgenic techniques, possibly even requiring the recategorizing of century old mutagenesis techniques (Callaway, 2018). Gene editing might promise quick and easy solutions in many areas of research, but the durability of resistance through gene editing is largely unknown, and registration cost are potentially prohibitive to the commercial release of varieties. Therefore the strategy that this work will be focusing on, is the use of statistical tools to enhance breeding progress. Here, stripe rust field scoring data and genetic markers are used to train a model which can predict the performance of untested lines (Juliana *et al.*, 2017).

1.3 Genetic markers and genome-wide association mapping

A major driving force of genetic studies and analyses are advances in high-density genotyping technology. **Genotyping by sequencing (GBS)** uses restriction enzymes and DNA sequencing technology to detect SNP variations in the genome. After cutting the DNA with restriction enzymes, size selection filtering to roughly 100bp is performed, reducing genome complexity drastically, while avoiding repetitive, non coding regions of the genome by using methylation sensitive restriction enzymes. Barcode adapters are then attached to these segments and PCR (Polymerase Chain Reaction) amplification is performed. These fragments are sequenced by next generation sequencing technology and aligned to the specific reference genome (Elshire et al., 2011; He et al., 2014). The advantage of GBS compared to chip-based SNP methods is a lower cost per sample, but with the drawback of a large number of missing data points (Bajgain *et al.*, 2016).

High-density genotyping is needed for powerful genetic studies that link phenotypes to genotypes. With the low cost and advancing technologies in sequencing, GBS is well equipped to be an important tool in the future. Today already many types of GBS exist, providing different options for specific needs. The problem of missing data can only be solved by increasing sequencing depth, which

also drives cost, or by imputation through various methods. However, some analysis can handle non-imputed data, which is preferred if available (Poland and Rife, 2012).

Simple Sequence Repeats (SSR) or microsatellite markers are a PCR based system that relies on non-coding, repeating motifs of variable length in the genome. Flanking site specific primers are used to amplify these regions and variations in length are detected. Advantages of SSR markers include codominance, sub-genome specificity in allopolyploid plants like wheat and good transferability into various genetic backgrounds. This makes them ideal for genetic diversity analysis but also for applied plant breeding (Röder *et al.*, 2005). Markers were developed by groups around the world and aggregated into a consensus map (Somers *et al.*, 2004). What makes them interesting for this thesis, is that SSR markers are widely used in mapping studies, which makes them ideal for the comparison of results across studies. Furthermore, a large number of SSR markers can be found in the wheat reference genome (IWGSC *et al.*, 2018), which nowadays enables precise location in the genome and comparability to other marker systems like GBS.

One of the hypotheses formulated by Mendel (1860) is that plant traits are inherited independently in plant hybrids. This is only true if traits, their genes or markers are on different chromosomes. However, when genes and markers are on the same chromosome, and depending on their distance, they are more likely to be inherited together. **Linkage mapping** is making use of this "linkage disequilibrium" (LD) by correlating phenotypic and genetic data to find markers which are related to plant traits, so-called quantitative trait loci (QTL) (Bernardo, 2002). Linkage mapping is usually performed as a controlled experiment, where two or more parents are crossed and filial generations are analysed for recombination rates of markers and traits. Since recombination events are recent and therefore rare in these populations, genetic maps of coarse resolution can be generated in such studies. **Association mapping** on the other hand uses unrelated or related individuals from any population, exploiting historical recombination, and a high marker density to achieve a higher mapping resolution. Hence, advances in high-density, genome-wide, low cost marker technology have contributed to the rise of Genome-Wide Association Studies (GWAS) as an alternative tool for linkage mapping in genetic studies (Mathew *et al.*, 2019).

In breeding programs, data from a large number of individuals is routinely gathered. When using this data in **Genome-Wide Association Studies (GWAS)**, non-random population structure is introduced into the models, which increases the number of false positive marker-trait association. Appropriate statistical models (like the EMMA algorithm (H. M. Kang *et al.*, 2008)) have to be used, that take into account the non-random structure of such a population. Special multi-parent popula-

tions have been developed, such as NAM in barley and MAGIC in wheat, that combine the advantages of high LD and high population diversity without introducing population structure. GWAS results can be used in breeding programs to perform **Marker-Assisted Selection (MAS)**, which allows the prediction of the performance of untested individuals, by using significant markers (QTL). QTL can usually only explain a small fraction of the variance of quantitative traits, this is supposed to be overcome by genomic prediction (GP) (Dreisigacker, 2019).

1.4 Genomic Selection

Genomic selection (GS) or **Genomic Prediction (GP)** goes one step further than GWAS, by using dense genome-wide marker data to predict the genotypic performance of an individual. This can potentially lead to higher genetic gain compared to marker-assisted selection based only on a few QTL, and was first simulated for cattle breeding (Meuwissen *et al.*, 2001). BLUP (best linear unbiased prediction) allows the estimation of random effects by using LMM (Bernardo, 2002), which was widely used in cattle breeding. Pedigree information and ancestry performance, was used to predict offspring breeding values. Genomic BLUP (GBLUP) uses a genetic relationship matrix (GRM or G-matrix), which is a variance-covariance matrix that is created from the marker data, and provides more precise relatedness information than a pedigree. The G-matrix is used as a random factor in a mixed model that calculates Genomic Estimated Breeding Values (GEBVs) (VanRaden, 2008). GEBVs can then be used instead of real breeding values, to make selection decisions in a breeding program.

Phenotypic and genomic data is used to train a model which is able to predict performance of unphenotyped lines, based on their genomic data. This model has to be validated by splitting the phenotypic data from the training population into two parts. While one part of the population is used for model training, another part is used for validation and optimization of the model. If the model is optimized repeatedly based on the training/validation split, splitting the population into three parts can be useful. The third part can then, at the end of the optimization process, be used to get an unbiased estimate of model performance (Korjus *et al.*, 2016; Roberts *et al.*, 2017).

The true breeding value of an individual will always be unknown, therefore to assess the model performance, predictions are usually compared to an estimated breeding value. The estimated breeding value is calculated from real observations as a BLUP using LMM. One way to do so is to calculate prediction ability, which is the Pearson correlation coefficient of the observed value (BLUP) and the predicted value (GEBV), or the prediction accuracy which is the correlation between the ob-

served value and the true genotypic value. Other, more sophisticated, methods have been proposed for assessing model performance (Ould Estaghvirou *et al.*, 2013), which are promising more robustness of the parameter but are usually more complex and can be computationally demanding.

By using GP in a breeding program untested genotypes can be predicted without or before test data is available. If the genotyping cost is lower than the cost of a field trial plot, more lines can be tested at the same cost, however usually with lower accuracy since the genomic prediction will not explain the whole variation in the population. Heffner *et al.* (2010) demonstrated that the main advantage of GS compared to MAS is not in larger genetic gains per cycle, but in shortening breeding cycle time due to the fact that superior lines can be determined earlier in the breeding cycle and reused as mating partners for the next breeding cycle. In cattle, one of the first agricultural species that were subject to genomic selection, reports of genetic gain per year doubling and generation interval reducing by half are supporting these results (Doublet *et al.*, 2019).

One main challenge of genomic prediction models is, that the number of markers or predictor (p) is usually larger than the number of observations (n). A variety of models have been proposed to deal with this p>n problem in GP. While GBLUP uses the G-matrix to calculate GEBVs, a similar method uses ridge regression best linear unbiased prediction (rrBLUP) and the full marker information to obtain an estimate for each marker effect. Multiplying the marker matrix with the effects vector, generates a GEBV for each line, which has mathematically been shown to be equivalent to the GEBV obtained by GBLUP (Liu *et al.*, 2016). Ridge regression is shrinking the marker effects towards zero to deal with the p>n problem. GBLUP and rrBLUP are based on the assumption of equal contribution of every locus to a quantitative trait. Bayesian models take into account various prior distributions of marker effects and allow distinct variances for each marker. LASSO models are using variable selection rather than shrinkage methods to deal with the p>n problem. Further methods use various machine learning techniques or neural networks models to build increasingly sophisticated models for genomic prediction (Crossa *et al.*, 2017; Poland and Rutkoski, 2016; X. Wang *et al.*, 2018).

1.5 Aims of this thesis

In light of recent epidemic outbreaks of stripe rust around the years 2000 and 2014 in Austria and changes in the susceptibility of registered varieties (Oberforster, 2018), a better understanding of breeding for stripe rust resistance and new stripe rust resistant varieties are in high demand. The main aim of this thesis was therefore to explore statistical and bioinformatic tools to facilitate a

more efficient breeding for stripe rust resistance. To this end, the following research questions (RQ) and goals were stated:

- RQ: Can the susceptibility of wheat to stripe rust be predicted by the use of genetic markers and genomic prediction to replace phenotypic selection in years without natural infections?
- RQ: What are the limitations of genomic prediction for stripe rust in bread wheat?
- RQ: Is it possible to use genomic prediction models to select for quantitative resistance to stripe rust versus race-specific qualitative resistance?
- Goal: Use genome-wide association mapping to dissect the genetic architecture of stripe resistance in a breeding population of wheat in natural and artificial infection scenarios.
- Goal: Use state of the art genomic prediction models to predict stripe rust susceptibility and validate the models to determine their ability to predict future performance.
- Goal: Improve resistance and stability against stripe rust by genomics-assisted wheat breeding.

2 Materials and Methods

The data for this thesis was kindly provided by Saatzucht Donau (SZD) and came from their winter wheat breeding program. It consisted of stripe rust scorings in years of high field infections as well as artificial inoculation sites. This data was supplemented by GBS marker data for part of the breeding lines. Scoring was done on multiple locations in the years 2013, 2014, 2015 and 2016.

2.1 Plant material and trial locations

Plant material used in this thesis were of the following types:

- Mostly F5, F6, F7 and recombinant inbred (RI) as well as doubled haploid (DH) lines from winter wheat breeding programs of SZD.
- Varieties registered in Austria (AGES, 2019) and other countries as well as advanced lines (variety candidates).

In total, 68 trials were analysed, consisting of 11074 lines, of which 3252 were fingerprinted with genome-wide distributed markers. Testing was done at multiple locations in Austria (AU), France (FR), Germany (DE), Hungary (HU), Romania (RO), Serbia (RS), Slovakia (SK), Turkey (TR) and Canada (CA). Locations in Austria also included Probstdorf (PRO, PRks, Prkf), Marchtrenk (MAR), Leopoldsdorf (LEO), Melk (MLK), Weikendorf (WEI), and Dörfles (DOE). Additionally, a total of 6307 lines from the entire set of 11074 breeding lines and varieties were tested in a disease nursery in Reichersberg, Austria (RBG), which is part of the western region of Austria with higher rainfall as compared to the drier continental climate of the other Eastern Austrian trial locations. The disease nursery was artificially spray-inoculated with multiple diseases, as well as an inokulum of stripe rust from the Julius Kühn Institute (JKI) (pathotypes unknown). All other trials were subject to natural infection.

2.2 Trial designs

The disease nursery, as well as most other large trials, are non-replicated designs *i.e.* in which most of the lines are only tested on one plot. One or more lines are replicated across the trials to enable some correction of a field trend along rows or columns, this is also called a “replicated control design” (RCD) or “check plot design”. In some of the smaller trials, lines were replicated up to four

times, with the trials still being modeled like the non-replicated ones, *i.e.* only according to row and column information.

2.3 Phenotypic Analysis

For the genetic and phenotypic analysis the data had to be prepared by using linear mixed modeling (LMM). Linear mixed models allow the definition of fixed and random effects, depending on the type of effect and the goal of the analysis. From these models, effect size and residuals can be extracted and estimates for individuals (BLUEs, BLUPs) can be calculated. Linear mixed models were set up for the phenotypic analysis of the data according to Equation 1, where y is the response vector that is aimed to be modeled, X and Z are incidence matrices for the fixed and random effects. β is a vector with fixed effects, and u is a vector of random effects. The error term, e is used for describing unexplained residuals of the model.

$$y = X\beta + Zu + e$$ Equation 1: Linear mixed models

In Figure 3 the process of the phenotypic analysis is shown in a schematic representation. First, a heritability estimation is done, then estimated marginal means are calculated and finally the data has to be rescaled to the original range. This could have been done in a single process, however, this was regarded to be computationally too demanding with such a large dataset as in the study at hand, since the computation time increases in a non-linear exponential way with the number of datapoints. Therefore the process was split into a two-stage sequence, which reduced the computational load dramatically.

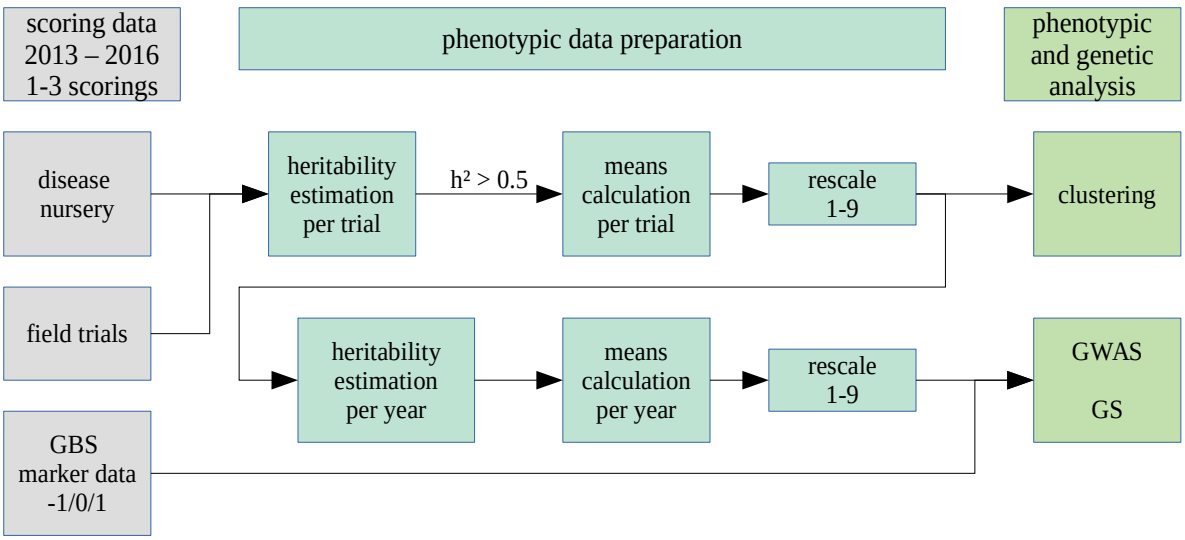


Figure 3: Schematic of data preparation, showing the two stage analysis approach.

The first part of the phenotypic analysis focused on identifying the trials that had a sufficient data quality for further downstream analyses. This was achieved by computing the broad sense heritability (H^2) per trial, using the following formula:

$$H^2 = \frac{V_G}{(V_G + V_e/r)}$$ Equation 2: Broad sense heritability

Heritability (H^2) is calculated as the fraction of genetic variance (V_G), explaining the phenotypic variance ($V_G + V_e/r$) as the sum of the genetic and the error variance, divided by the number of replicates (r). For the purpose of calculating heritability per trial, the following LMM was set up:

$$y = \mu + g_i + r_j + c_k + spline_{jk} + e$$ Equation 3: LMM first stage

The response y represents the value for each plot, and μ is a grand mean. To extract the variance components needed, the genotype (g_i) had to be added as a random factor. Other random factors are row (r_j) and column (c_k) information from the plot layout, as well as a two-dimensional spline model ($spline_{jk}$) across rows and columns, which allows for more precise estimation of variance due to spatial environmental effects (Rodríguez-Álvarez *et al.*, 2018), and a random error term (e). The model did not include any fixed effects. The genetic and error variance was extracted directly from the model for the calculation of broad sense heritability.

To generate adjusted means per genotype, another LMM according to Equation 3 was used. This time, the model contained the genotype as a fixed effect to calculate BLUEs (best linear unbiased estimators). The spatial information was modeled in the same way as in the heritability model as random effects. QQ-plots and residual plots were produced to inspect the residuals concerning the assumptions of a normal distribution and the homogeneity of the residual variance.

Finally, since the scoring data is of an ordinal type but was treated as metric data in the study at hand, the resulting BLUEs were not distributed between 1 and 9 like the original data. This was compensated for by rescaling the data in a way that did not influence their distribution. The following formula was applied to all the resulting BLUEs:

$$y' = 1 + \frac{y - y_{min}}{y_{max} - y_{min}} * 8$$ Equation 4: Rescaling BLUEs

In the second stage of the analysis, heritability estimates were generated, using the trial BLUEs from each trial, additionally for each year. The model contained genotypes (g_i), trials (env_j) and an error term (e) as random factors, and yielded variance components to calculate heritability per year in the same way as per trial:

$$y = \mu + g_i + env_j + e$$ Equation 5: LMM for second stage

Another LMM according to Equation 5 was set up that used the previously generated rescaled BLUEs from each trial to generate BLUEs for each year. Genotype (g_i) was added as a fixed factor and trial (env_j) was added as a random factor. BLUEs per year were extracted from the model and rescaled according to Equation 4.

2.3.1 Scoring results across years

Since the data was recorded from a breeding population, most genotypes were tested only for one or two years, since they proved not to have a high breeding value or failed to meet breeder's selection criteria. However, some genotypes, especially check varieties and other registered varieties were tested across several years. These genotypes can therefore be used to get information about the temporal dynamics of stripe rust, such as changes in race occurrence and virulence.

This was done on the one hand by comparing pairs of years and the correlation of scorings from one year to the other years, as well as by more sophisticated methods like k-means clustering (Hartigan and Wong, 1979) and k-medoids clustering (Kaufmann and Rousseeuw, 1987). The k-means clustering algorithm, as implemented in the stats package for R (R Core Team, 2018), was applied to the scoring data of multiple lines and years, with the aim of grouping similarly behaving lines together into so called "clusters". This algorithm was tested with a range of predefined numbers of clusters that grouped the lines in a way, which minimized the sum of squares of the respective data points to their cluster centers. A local optimum of the cluster number was accordingly defined in this process, as a minimum of total within group sum of squares.

This algorithm was tested with a range of predefined numbers of clusters that grouped the lines in a way which minimized the sum of squares of the respective data points to their cluster centers. A local optimum of the cluster number was accordingly defined in this process, as a minimum of total within group sum of squares.

A similar method for clustering is k-medoids or partitioning around medoids (pam) clustering, which uses data points as cluster centers rather than cluster means. The implementation used in R was from the “fpc” package (Hennig, 2019). The algorithm has to be provided with a range of possible or reasonable cluster centers and returns critical values for each number of cluster centers and the optimal number of clusters. Interaction plots were used to visualise the results. Optimal cluster sizes were selected by finding a maximum in the critical value, which is the average silhouette width of all data points. Silhouette width is a measure of how similar each data-point is to its own cluster versus the other clusters and can range from 0 to 1.

A further analysis was done to assess the similarity of trials within years and across-year. Pairwise correlation coefficients and p-values for the significance of the coefficients were calculated, based on lines present in both years of each pair of trials. In some trials, less than three common lines were tested, which prevented the calculation of a Pearson correlation. A higher number of common observations allowed for a more precise estimation of correlation coefficients.

2.4 Genetic Analysis

2.4.1 Marker data

Using the GBS system, 3188 lines out of the total ~12000 available lines were genotyped with Single Nucleotide Polymorphism (SNP) markers. SNP markers analysis returns only three possible values; the allele is homozygous minor (-1), heterozygous (0) or homozygous major (1). The minor or major allele coding is usually assigned, according to their frequency in the reference population, so that the minor allele is the rare one and the major allele is the more abundant one. However, major or minor can be associated with each one of the alleles in a given population. For further analysis the missing, previously imputed values, which were represented by fractions of (+/-) 1, were rounded to the nearest whole number value. This resulted in a matrix of genotypes in the rows, and markers in the columns coded as “-1”, “0” and “1”.

The markers were filtered for minor allele frequency (MAF), a common threshold for MAF filtering is 5% *i.e.*. Markers with a MAF lower than 5% were not considered for further analysis. This filtering strategy removed 65% of the available 7365 markers, which resulted in 2582 mark-

Table 1: Marker density per chromosome after filtering for MAF (2582 total).

Chr. ID	1	2	3	4	5	6	7
A	104	155	142	149	113	123	149
B	206	338	189	69	218	118	100
D	68	81	62	25	33	52	88

ers that were used for further genomic analyses (Table 1, Appendix 7.1.2). They were distributed roughly equally among all 21 chromosomes.

Heterozygosity per individual was calculated by dividing the number of heterozygous marker allele calls by the total number of markers. The expected heterozygosity is zero for doubled haploid (DH) lines, and in recombinant inbred (RI) lines lower than 12.5% or 6.25% for the F4 or F5 selfing generation. The actual heterozygosity is expected to be lower, since the parental lines are not completely unrelated and usually share some alleles. The mean heterozygosity was thus found to be 4.7% with a maximum of 24.6% and a minimum of 0.7%. Lines with heterozygosity higher than 15% were excluded from the analysis, since they might not produce reliable results, which led to the removal of 45 lines.

2.4.2 Genome-wide association mapping for stripe rust resistance

For the GWAS, each year was seperately analysed using the BLUEs that were generated in the mixed model analysis of 2013, 2014, 2015 and 2016. The available markers were filtered for their MAF in each year, using the "snpQC" function from the package NAM (Xavier *et al.*, 2015), which resulted in 2675, 2717, 2290 and 2212 markers for 2013, 2014, 2015 and 2016 respectively.

The GWAS (R package sommer) analysis was based on the application of an LMM according to Equation 6. The response y represents the phenotypic observations, X is a design incidence matrix and β is a vector with fixed effects, not associated with genetic differences. Z is a covariance matrix of kinship, generated from the marker matrix that represents kinship between individuals and g is a variable of kinship effects. M is the marker matrix and τ a vector of genetic effects not explained by genetic relationship but by the individual SNP, while e is a vector with random error, not explained by the rest of the model (Yu *et al.*, 2006) (Covarrubias-Pazaran, 2016, 2018).

$$y = X\beta + Zg + M\tau + e$$ Equation 6: LMM for GWAS

This model yields effects size, significance ($-\log_{10}(p)$), F-statistic and proportion of explained variation (R^2) for each marker. The threshold for declaring a significant marker trait association was determined by using the Bonferroni method, which takes into account a multiple testing situation. This is necessary because one trait is tested against a lot of markers repeatedly. This leads to some results being significant just by chance, and generates a lot of false positives. Raising this significance

threshold according to the number of markers used, or tests done, leads to the formula in Equation 7,

$$crit = \frac{\alpha}{nm_{maf}}$$ Equation 7: Bonferroni correction

where nm_{maf} designates the number of markers after filtering for MAF. The threshold was calculated for each year seperately, due to a different number of markers in each year after MAF filtering (nm_{maf}). Hence p-values below the $\alpha = 0.05$ threshold (adjusted for number of markers) were declared to be significantly associated with the trait of interest, *i.e.* Stripe rust resistance. Neighbouring significant markers were grouped, while only the marker with the lowest p-value per locus (QTL) was retained for further analysis. Manhattan plots, usually showing the $-\log_{10}(p)$ value, were produced using the "qqman" R package (Turner, 2018) for visualizations of the GWAS results.

QTL were named with a sequential number per year and the year they were found in (number – year / ...). Due to the proprietary nature of the genetic map and missing physical location information, the QTL found in the genome-wide association study were located using BLAST (Altschul, 1997) and compared to the v1.0 reference genome of wheat, using an online genome browser (IWGSC *et al.*, 2018). Known stripe rust resistance genes and QTL were discussed using the online catalogue of gene symbols (McIntosh *et al.*, 2017) as well as other sources.

2.4.3 Genomic selection models for predicting stripe rust resistance

GBLUP, first proposed by VanRaden (2008) and Habier et al. (2007), is chosen in many studies for its high computational efficiency and high predictive ability, and was therefore the method of choice for this thesis. Genetic marker data was used to construct the G-matrix, which is needed to set up a GBLUP model. The data which was fed into the GS model was the same as for the GWAS model. BLUEs from the phenotypic analysis and genotypes were filtered for heterozygosity. The markers used for building the G-matrix were previously filtered for a MAF of 5%. The GBLUP model is based on an LMM, according to:

$$y = X\beta + Zg + e \quad \text{with} \quad var(g) = G * \sigma_g^2$$ Equation 8: LMM for GBLUP

The model uses the G-matrix and the genetic variance (σ^2_g) to construct the vector g, which is used as a random effect in conjunction with the design matrix Z that allocates lines to their genetic values. This model also incorporates fixed effects (β) and a design matrix (X) as well as a random unexplained error term (e) to model the phenotypic performance (y) (Clark and van der Werf, 2013) of

stripe rust resistance. In within-year prediction, no fixed effects are present, but in prediction across multiple years, prediction-year is an additional random effect (u) in the model. Results from the GBLUP model were simply calculated by summing up the intercept plus the random effects plus the fixed effects to obtain predictions for untested genotypes.

An alternative model that was implemented, called GBLUPA (Juliana *et al.*, 2017; Zhao *et al.*, 2014), uses known high-effect loci, in this case QTL from the GWAS, as fixed effects (β) in the GBLUP model to improve the resulting accuracy of GEBVs. If no QTL were found in the GWAS for a particular year, the four highest scored loci were used instead.

2.4.4 Genomic predictions within years

Both of the models for genomic prediction described above were initially tested within each of the years. For this purpose, the population was split randomly into five parts or sets and then one out of five sets was used for validation while the other parts were used for the training of the prediction models. Each set was used as a validation set in combination with the different numbers of training sets to simulate different training population sizes on the one hand, and to generate five replicates for each training population size on the other hand. Different marker densities were simulated by not only using all 7365 markers, but also randomly selecting 100, 1000 and 3000 markers out of the full set. Both models described above, GBLUP and GBLUPA, were used for the simulation. In case of the GBLUPA model, the QTL used as fixed effects were not estimated in the training set each time but taken from the previous GWAS analysis.

All the combinations of 2 models, 4 years, 4 marker densities, 5 validation sets and 4 different training set sizes result in a total of 640 combinations. The models were evaluated based on the Pearson correlation coefficients of the validation set and the predicted values for the same set of lines, which is called the prediction ability. The prediction accuracy was calculated by dividing the prediction ability by the square root of the heritability calculated in the phenotypic analysis, which can be interpreted as the correlation between the true and the predicted breeding values.

2.4.5 Genomic prediction across years

Genomic selection was also tested across years, where one or more years were used for building the model to predict another year's individuals stripe rust resistance. Out of the four years of data, 1 to 3 years were chosen for training and one year for validation. All possible combinations were tested, e.g. using 2013 and 2014 for training and the subsequent year 2015 for validation, as well as more

theoretical combinations with backwards prediction like using 2015 and 2016 for training the model to predict the performance in 2013. A random year effect was added to the model to account for different stripe rust occurrence in the training years.

One thing that has to be taken into account when modeling and predicting across years, is that some genotypes might be present both in the training and in the validation population, which can positively bias the predictive ability. To account for that, two variants were modeled: one where duplicate genotypes were removed from the training population, -and another one where duplicates were removed from the validation population. Using either GBLUP or GBLUPA for building prediction models resulted in the calculation of 48 models of different years and variants when using one or two training years, and 18 models when using three training years.

To find out which factors are relevant for improving across-year prediction, the results from the simulations were further analysed. In total, 224 different variants were calculated for GS across different number of years, untested or tested years, different GBLUP versions, different variants in data splitting and prediction or training years. All these variants were included in a linear model and ANOVA was performed on the model to find out which of factors affect the performance of GS.

2.4.6 Genomic prediction across years, in tested years

In the before mentioned GS approach, the data from tested genotypes was used to predict untested genotypes in untested years. Since the influence of different years can be quite large, the problem could be reformulated in a way that unknown genotypes are tested in a known environment or year. Similar to the previous approach, one two or three years were used to train the model to predict one year, which was then used for validation. Additionally, some lines were moved from the validation set to the prediction set in order to additionally estimate the validation year effect and achieve better predictions for a particular year by reducing the overall model fit. Two variations of this approach were evaluated. In the first variant, genotypes from the validation year that were also present in the training years were also used for training (var 1), these were mostly check varieties and registered varieties. In the second variant, some additional lines (10%) were randomly selected from the validation set to improve the model, and were therefore moved from the validation set to the training set (var 2). These lines are breeding candidates, which should have higher relatedness than for example check varieties.

2.4.7 Genomic prediction and preliminary yield trials

A common feature of a plant breeding program is to use preliminary yield trials (PYT), to test a large number of entries, in a limited number of environments, and preselect lines which are later advanced into extensive multi-location trials (Endelman *et al.*, 2014). In the dataset used in this thesis, a PYT consists of a lot of small plots of F5 or DH breeding lines, named PRkf (Probstdorf small plots early flowering) or PRks (Probstdorf small plots late flowering). Stripe rust data from these trials only exists for 2014, 2015 and 2016.

The goal of this analysis was to simulate a practical approach to genomic prediction in a plant breeding context. The data from one year's PYT was used to predict the breeding line's stripe rust susceptibility in the following year(s). Model variants included GBLUP and GBLUPA, as well as training the model only with the PYT data or with the year's full data. As a reference, the phenotypic correlation of small plots to the following years multi-location trials was calculated.

2.5 Statistics software

All the statistical analysis, as well as the data preparation and graph drawing, was carried out by using R (R Core Team, 2018). Packages for specific purposes that were used, but not delivered with the standard installation, are listed in Table 2.

Table 2: Software packages used in this thesis.

name	version	usage	authors
corrplot	0.84	plotting correlation results	(Wei and Viliam, 2017)
emmeans	1.3.5.1	estimated marginal means	(Lenth, 2019)
fpc	2.2-2	medoids clustering	(Hennig, 2019)
NAM	1.7.2	quality control, MAF filtering	(Xavier *et al.*, 2015)
R	3.5.2	programming language	(R Core Team, 2018)
rrBLUP	4.6	additive relationship matrix	(Endelman, 2011)
sommer	4.0.1	mixed modeling, GWAS, GS	(Covarrubias-Pazaran, 2016, 2018)
qqman	0.1.7	manhattan plots	(Turner, 2018)

3 Results

3.1 Phenotypic analysis

Data from a total of 68 trials was provided for the analysis. After filtering the raw data 30 trials had to be removed, and 38 were passed on to mixed modeling to generate BLUEs for further analysis. Six trials were removed due to their low number of plots, and therefore low number of lines. These trials do not provide reliable information for the analysis of genetic effects, which need a larger number of individuals to be representative. The same is true for seven trials, which have a low variance in yellow rust scorings. Nine trials had to be removed due to mixed modeling errors, which were not investigated further. The final dataset included trials ranging from 60 to 3261 plots, 29 to 2238 lines, 1 to 4 replicates, 15 to 240 rows, 2 to 44 columns, 1, 2 or 3 scorings, mean scorings from 1.21 to 4.56, maximum scorings from 4 to 9 and variance in the scoring from 0.25 to 7.05. The mean number of plots per trial over all trials was 902 with 557 lines per trial, 1.8 replicates, 1.8 scorings, mean scoring of 2.47, mean maximum scoring of 8.08 and mean variance of 2.32.

Before calculating marginal means, the heritability was estimated for each trial and for each year (Figure 4). Some trials yielded heritabilities below 0.5 and were not used for further analysis since they were expected to impair the overall results. The actual limit was increased to 0.55 to exclude one further trial from 2015 which was clustering more with the lower heritability trials, than with the higher heritability trials. The trial heritabilities are listed in section 7.2.3 in the Appendix. Most trials yielded a relatively high heritability, so that after filtering, only two trials were left for 2013, 17 trials for 2014, 10 trials for 2015 and 6 trials for 2016 to be used for the phenotypic analysis across trials within each year. This selection of trials yielded high heritability in the per-year estimation of 0.91 in 2013, 0.91 in 2014, 0.85 in 2015 and merely 0.57 in 2016.

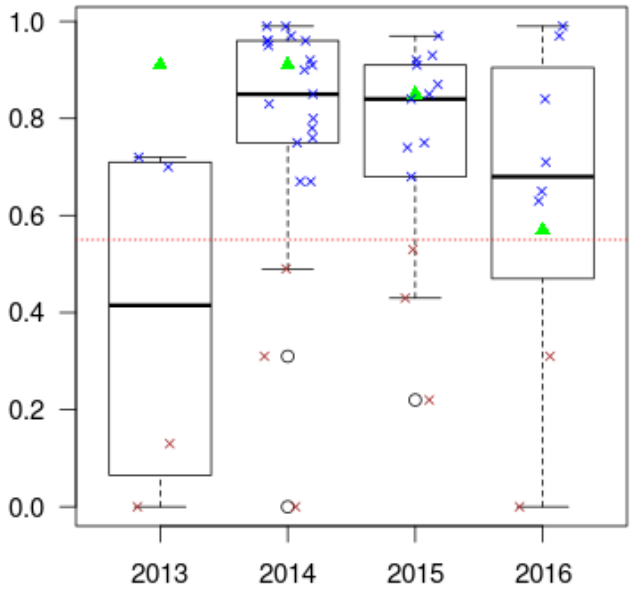


Figure 4: Heritability plot of trials versus years.
Trials above threshold (red line) used for further analysis (blue: per trial, green: h^2 per year, red line: threshold).

The YR scorings show a pattern of relatively high severity in 2013, where only the disease nursery data was available. The average stripe rust severity was lowest in 2014, while it increased in the years 2015 and 2016. The calculation of adjusted means per trial yielded a similar distribution and similar means as in the raw data, with some overcorrection appearing towards the lower scoring. Due to this overcorrection, some of the data had to be rescaled, which led to an increase of the mean scoring values, while conserving the distribution of the individual values. Calculating adjusted means per year did not yield over-adjusted values, but was rescaled to use the full range (1-9) as well. Rescaling per-year adjusted-means did not influence the data distribution, which was inspected using boxplots. The distribution of final, adjusted, rescaled, per-year means can be seen in Figure 5. The distributions of the adjusted means after each step of the whole process can be found in section 7.3 in the Appendix. There, one can see that per year means and distributions of the data were only rescaled but not distorted, which is an important requirement before advancing to further analysis. The final dataset contained 1439 lines for 2013 with an adjusted mean scoring of 4.7, 3132 lines for 2014 with a mean of 4.0, 3308 lines for 2015 with a mean of 4.17 and 3373 lines in 2016 with an adjusted mean scoring value of 4.6.

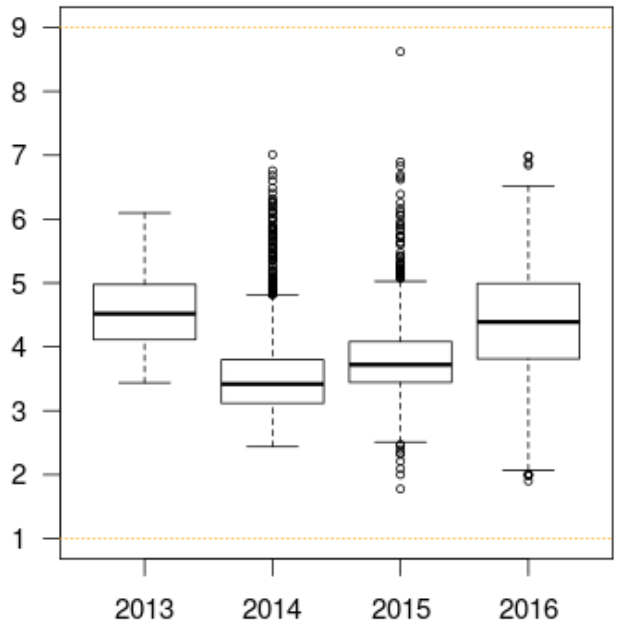


Figure 5: Stripe rust (YR) adjusted means per year (1-9 score).

3.2 Scoring results across years

Comparing different years in terms of overall stripe rust occurrence is one way of investigating the data on hand, while another important aspect is to look at individual lines across multiple years. A total of 143 lines were scored for their stripe rust resistance across all three years, while 39 of those, also had data for 2013 that was assessed in the artificially inoculated disease nursery. The years 2013 and 2016 showed the highest correlation and would therefore be expected to be similar in stripe rust virulence, 2013 and 2014 as well as 2014 and 2016 show

Table 3: Pairwise correlations of scoring across years.

	2014	2015	2016
2013	0.17	0.34	0.58
2014		0.37	0.19
2015			0.49

very low correlations, which would suggest the presence of different races of stripe rust (Table 3). Scorings in the year 2015 show medium to low correlation with all the other years.

The first clustering approach was carried out using the K-means algorithm. Since K-means does not accept missing values, the clustering was done with the data from 2014 to 2016. Cluster sizes from two to 16 clusters were investigated. An optimal number of clusters could not be found, as incrementally increasing the cluster number merely led to a minimal change with respect to minimizing the target criterion until a saturation was reached at around 15 clusters (Appendix 7.5).

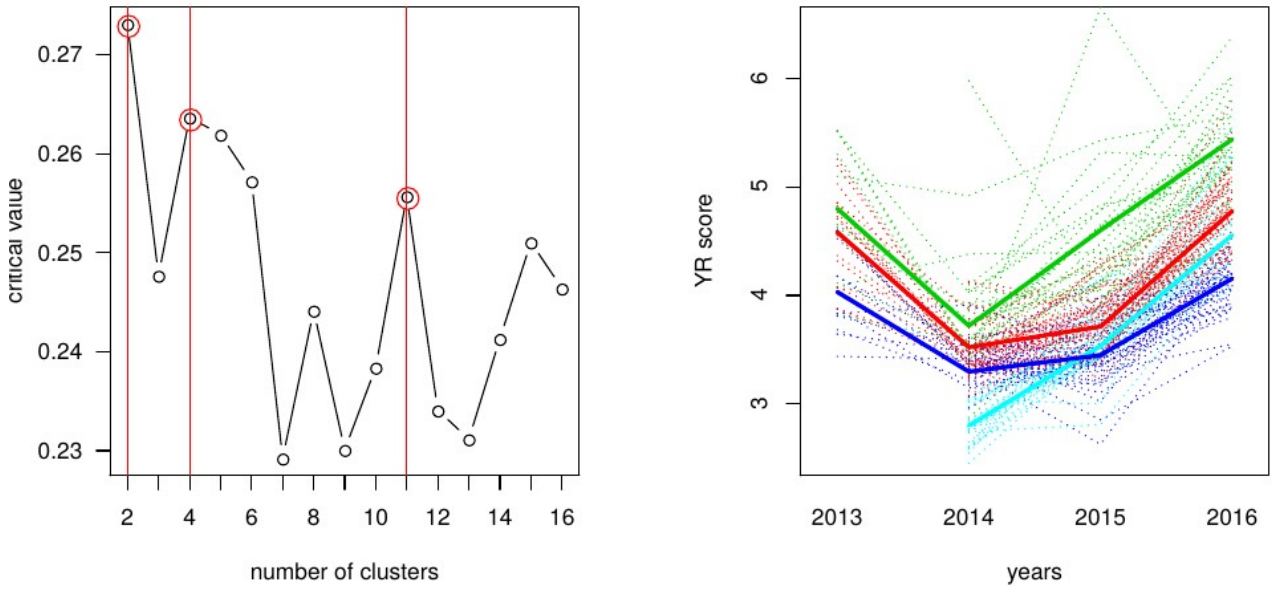


Figure 6: K-medoids clustering results.
On the left, optimal cluster size selection criterion versus the number of clusters, local optima marked in red. On the right, cluster allocation of 149 lines by line for four clusters. Dashed lines represent individual lines and solid lines represent cluster medoids. Clusters are represented by different colours (1 red, 2 green, 3 blue, 4 cyan) (2013 medoids imputed with cluster means)

The pam algorithm used for k-medoids clustering would accept missing values, however, the same data as in the k-means clustering, *i.e.* from 2014 to 2016 without missing values, was used in the analysis. Figure 6 contains a plot of the critical value versus cluster size, where local maxima are marked in red. Optimal cluster-sizes, in order of likelihood, would therefore be two, four and eleven. The solution with two clusters, however, did not provide major insights into the stripe rust dynamics, since it only splits the lines into resistant and highly susceptibility lines with similar relative performance, while the absolute score for each year is higher or lower according to the general occurrence (Figure in chapter 7.5). The clustering result with four clusters showed, similar to the two-cluster solution, partitioning into different levels of susceptibility, but it also identified two clusters in which the average performance crossed over when comparing 2014 and 2016 (Figure 6).

This provides evidence for a shift in susceptibility from 2014 to 2016, although the separation is not quite clear and the cluster are still overlapping in some lines with 25, 28, 40 and 53 lines per cluster (Appendix 7.5). Clustering did not take into account 2013 scoring data due to a lot of missing values. However, available values were plotted and medoids were replaced by cluster means in the plot. The imputed medoids in 2013 support the suggested clustering result, by showing similar grouping to the other years as was generated by clustering. No imputation was possible for cluster four in 2013 because all cluster four lines were missing data for this year. A third clustering solution that was further investigated resulted in eleven clusters. It showed even more crossing over, but cannot be easily interpreted due to a large number of clusters and some clusters consisted only of one line with extreme scorings in 2015.

3.3 Scoring results across trials

Due to the low correlation coefficients across years (Table 3) and low cluster-ability of the adjusted year-wise means, further investigation of the similarity of trials within years and across-year was carried out using a corellogram showing Pearson correlations between trials using pair-wise observations (Figure 7). While some correlation coefficients could not be calculated due to missing pair-wise observations (not enough common genotypes tested), others showed non significant correlations at a confidence level of $\alpha = 0.05$. However, this confidence level has to be decreased further since we have a situation with multiple testing at hand according to the number of tests, which leads to the plot on the right hand side of (Figure 7). Few significant correlations can be found between trials, not only within years, but also across years.

To further demonstrate this, a set of trials was selected, and pairwise observations were plotted (Figure 26). The plots show the numbers of pairwise observations (#), the p-value of the correlation significance test (p) and the correlation coefficient (r). Significant and non-significant correlations are marked (*, n.s.). Some plots, showing significant correlation between trials, demonstrate that trials are agreeing strongly in the recorded stripe rust scores, e.g. “2014 LEO 116” and “2014 LEO 148”. Other combinations, e.g. “2014 AUM-Ob 120” and “2014 AU-RBG_NURS” show medium correlation, but high variance of scores in one trial, and low variance in the other one. A lot of trials however, show non-significant correlation due to a variety of possible reasons.

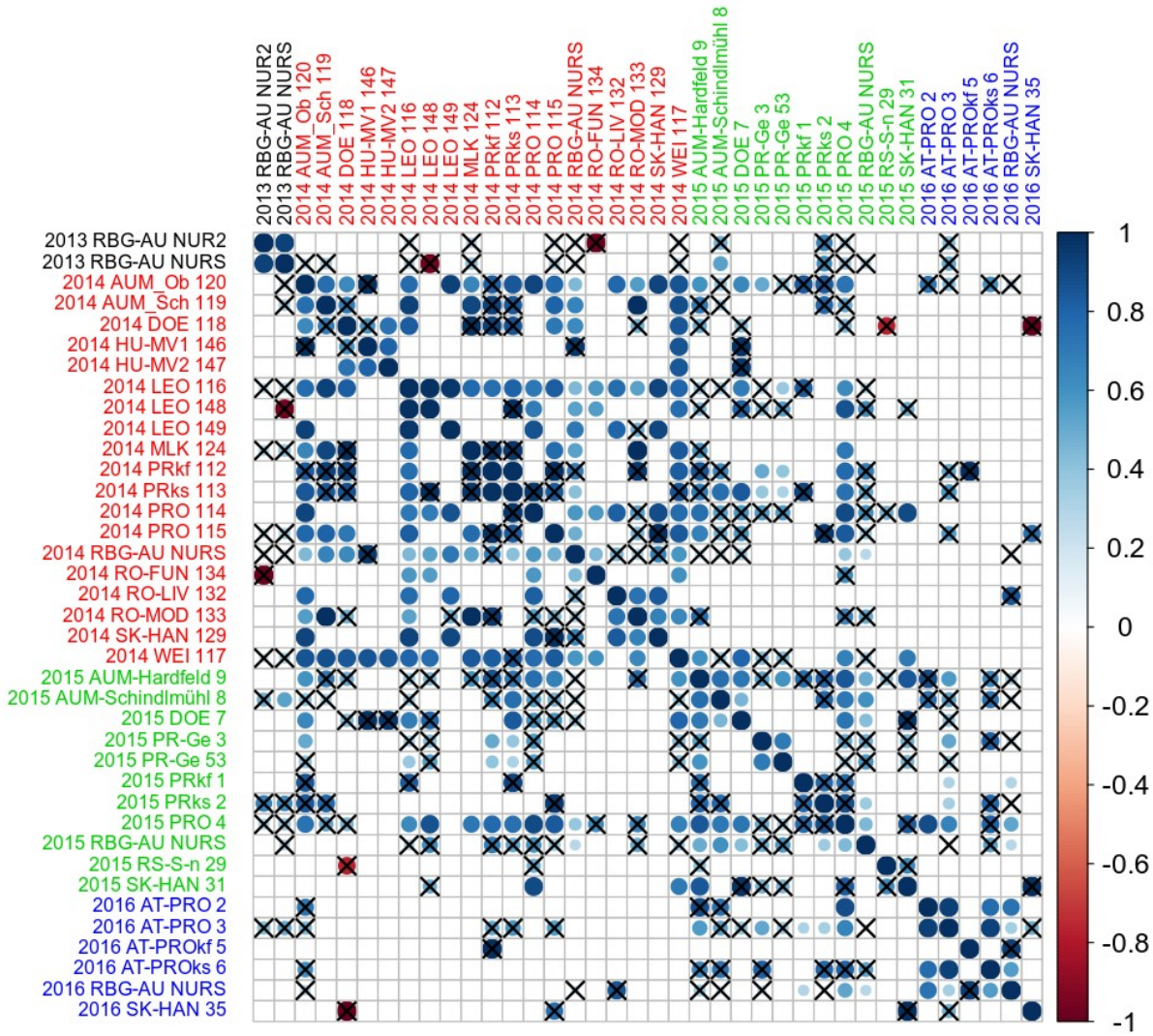


Figure 7: Correllogram of all trials used for across-year phenotypic and genetic analysis.
Non-significant trials $p > 0.05$ are not shown, non-significant trials considering multiple testing situation 0.05/ntest (6.9e-05) $< p < 0.05$ are shown but marked with “x”. Label colours indicate the same years trials.

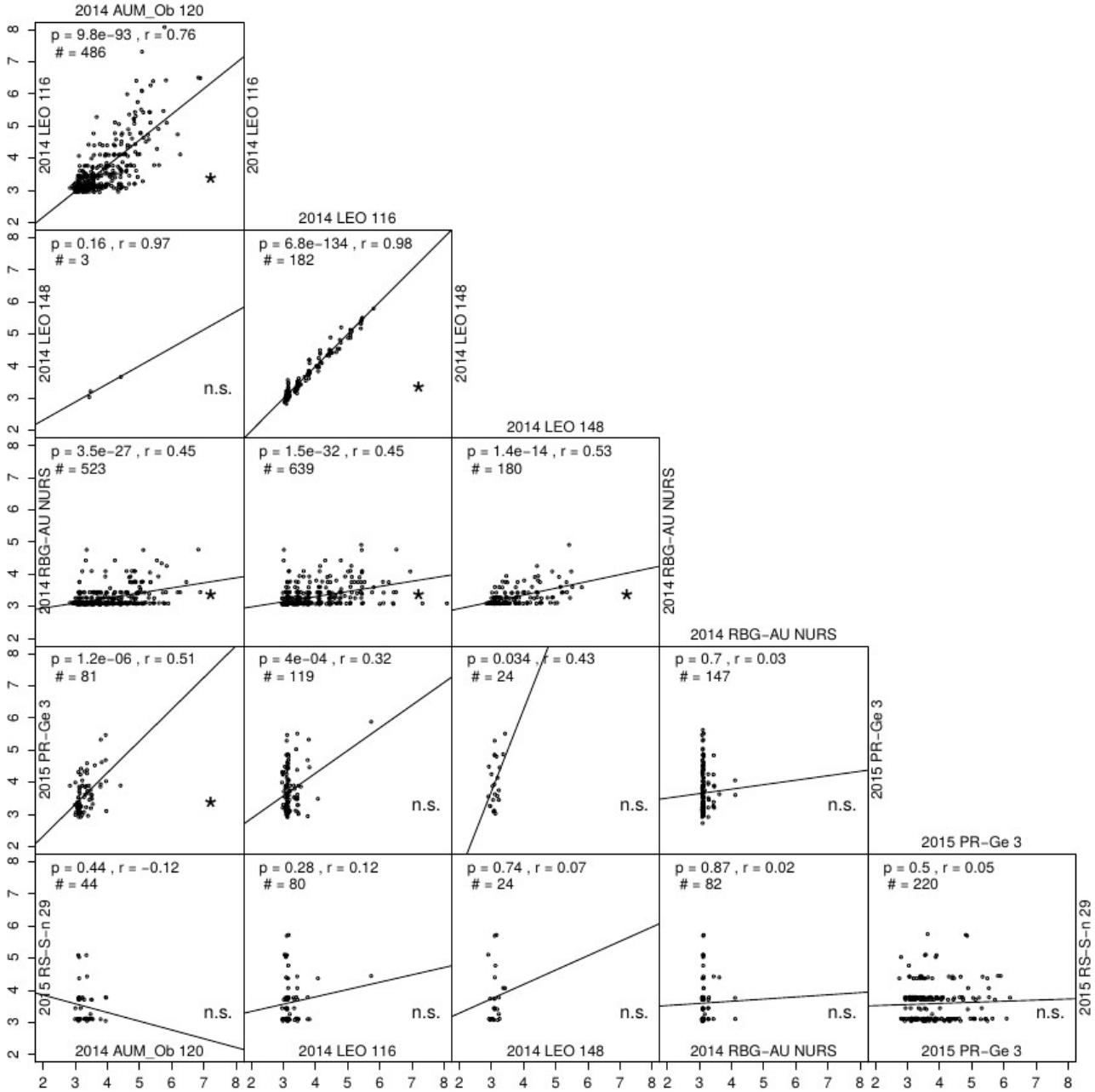


Figure 8: Pairwise observations of selected trials to demonstrate variability in trials.
Linear regression line and number of observations (#), correlation significance (p), correlation coefficient (r)

3.4 Genome-wide association study for stripe rust resistance

Bonferroni correction of the threshold for declaring significance of a marker-trait association in the GWAS was applied for each year separately, which resulted in slightly different thresholds when each year was analysed individually. Due to a different number of individuals, the threshold ranged from a p-value of $-\log_{10} = 4.66$ in 2016 to 4.73 in 2015, and 4.72 for 2013/2014. With these thresholds, no QTL were found for 2013 and 2016, while for 2014 and 2015, 6 potential QTL for stripe rust resistance could be found on chromosomes 1B, 2A, 2B and 5B (Table 4). Close to these QTL, ten markers with p-values below the threshold were found for 2014, and 13 markers for 2015, but generally with lower signal intensity than for 2014.

The first potential QTL "1-2014", containing only one marker, only appeared with a relatively low p-value of 5.2 in 2014. The allele frequency was close to the threshold in 2014 and below the thresholds in 2015 and 2016, therefore no score value is available. This fact and the positive signed effect value indicate that the unfavourable minor allele has been displaced and the favourable major allele has been fixed in the population.

Two markers were associated with the next potential QTL, "1-2015", which were on the same chromosomal location according to the genetic map, and almost indistinguishable by additive effects. Even though they were present in all years except 2016, they only barely passed the threshold in 2015 with a score of 4.8. The positive signed effect value indicates that the major allele increases stripe rust susceptibility, and the decreasing allele frequencies in the populations across years indicate that this unfavourable allele has been selected against, until the allele was so rare that it did not reach the MAF threshold in 2016.

The clearest signal for the presence of a QTL was found for "2-2014/2-2015" on the short arm of chromosome 2A, with four and seven markers being associated with this locus in 2014 and 2015 respectively. All the associated markers have high positive or negative signed additive effects ranging from 0.37 to 0.52 scoring points. Different markers were found to have the highest score for each year, and some markers that were found to be significant in 2015 did not reach the threshold in 2014. Some small effect of the markers seemed to be present in 2013, but in 2016 their effects were not visible any more. From 2014 to 2015 some selection towards the resistant alleles of these markers seemed to have happened, since the respective allele frequencies shifted accordingly. From 2015 to 2016 allele frequencies of these markers in the population did not change. Figure 9 shows detailed score values along the chromosome for each year. Some markers with low score appear in

between markers with high score, which indicates that the genetic map might be inaccurate, or that there are two QTL in close proximity. Manhattan plots can be found in chapter 7.6.1 in the Appendix.

Another potential QTL “3-2014/3-2015” on the short arm of chromosome 2A is supported by two markers, having some of the highest scores in 2014, and significant effect in 2015 as well. The allele frequencies show some selection towards the favourable allele. According to the genetic map, the markers are on the same position, but in the GWAS they show slightly differing results.

On chromosome 2B two positionally separated markers were found to have high scores in the GWAS. One is QTL “4-2014/4-2015” with the associated marker 1030280, which is located very close to the short telomere of chromosome 2B. The GWAS resulted in low p-values in 2014 and 2015, as well as a high effect size of 0.38 scoring points. The allele frequency was reduced by half from 2014 to 2015, indicating selection towards the favourable allele.

The other QTL “5-2014”, which also consisted of only one marker was located on the long arm of chromosome 2B, close to the telomeric end. It was only found to have a significant effect in 2014 with one of the highest negative effects of -0.48. In 2015 the effect was a little smaller and the score was a too small to pass the threshold, but it seemed to be active in 2015 as well. Some selection might have been happening between 2015 and 2016, since the allele frequency doubled from a low value of 0.07 to 0.12.

The only QTL on chromosome 5B is “6-2014/5-2015”. In both years, this one marker had rather low scores, but a relatively high positive effect, which would indicate at least a medium effect. The allele frequency showed a slight selection against the unfavourable allele between 2014 and 2015.

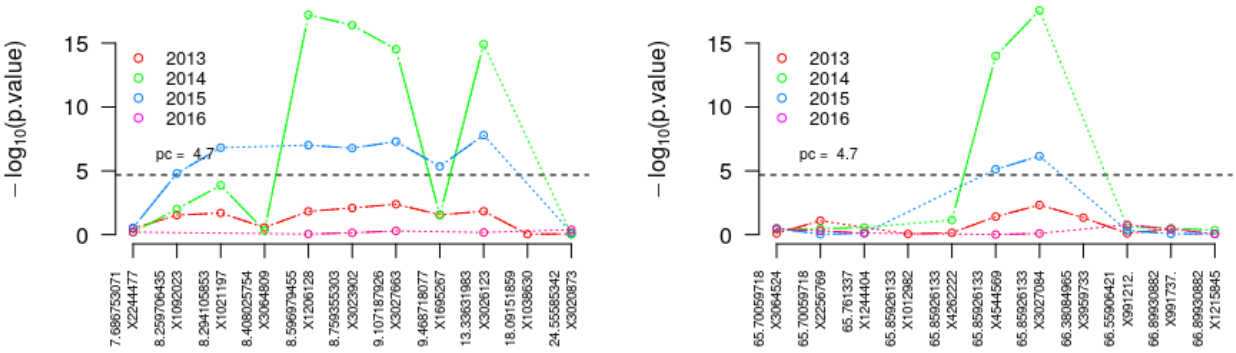


Figure 9: Chromosome 2A details of QTLs “2-2014/2-2015” on the left and “3-2014/3-2015” on the right. P-values versus marker name and marker location in cM shown in consecutive order. Due to strongly varying distances between markers the horizontal axis is not to scale for improved readability.

Table 4: GWAS results per year, grouped into potential QTL.
Peak markers per QTL are in **bold letters**, non-significant results are in *italic letters.*
SC = -10log(p) score, R^2 = explained variance, β = marker effect. Pos. = Position [cM] on the (Chr.) chromosome

QTL #	Chr #	Pos [cM]	Locus name	2013			2014			2015			2016			Allele frequency (Maj.)			
				Sc	R^2	β	Sc	R^2	β	Sc	R^2	β	Sc	R^2	β	2013	2014	2015	2016
1-2014	1B	201.49	1012285	0.5	0.004	0.12	5.2	0.014	0.35		-			-		0.20	0.07	0.02	0.01
1-2015	1B	96.91	1107810	1.8	0.024	0.41	2.4	0.006	0.25	4.8	0.015	0.05		-		0.20	0.12	0.06	0.04
			1114850	1.5	0.020	0.37	2.2	0.005	0.24	4.8	0.015	0.05		-					
2-2014 2-2015	2A	8.26	1092023	1.5	0.020	0.37	2.0	0.005	0.22	4.8	0.015	0.49		-		0.20	0.12	0.06	0.04
		8.29	1021197	1.7	0.022	0.39	3.9	0.010	0.30	6.8	0.56	0.02		-		0.20	0.12	0.06	0.04
		8.60	1206128	1.8	0.024	0.25	17.2	0.052	0.50	7.0	0.022	0.37	0.04	0	0.01	0.41	0.24	0.13	0.12
		8.76	3023902	2.1	0.029	-0.27	6.4	0.049	-0.52	6.8	0.021	-0.38	0.14	0	-0.02	0.59	0.76	0.87	0.88
		9.11	3027663	2.4	0.034	-0.28	14.5	0.043	-0.48	7.3	0.023	-0.40	0.30	0	-0.04	0.61	0.78	0.88	0.89
		9.47	1695267	1.6	0.020	0.38	1.5	0.003	0.17	5.4	0.016	0.49		-		0.20	0.12	0.07	0,04
		13.34	3026123	1.8	0.024	-0.25	14.9	0.045	-0.50	7.8	0.025	-0.41	0.17	0	0.03	0.59	0.77	0.87	0.88
3-2014 3-2015	2A	65.86	3027084	2.3	0.033	-0.28	17.5	0.053	-0.52	6.2	0.019	-0.34	0.09	0	-0.02	0.58	0.75	0.86	0.88
			4544569	1.4	0.018	0.21	13.9	0.042	0.45	5.1	0.31	0.02	0.01	0	0	0.42	0.23	0.13	0.12
4-2014 4-2015	2B	0.655	1030280	1.4	0.017	0.21	9.7	0.028	0.38	7.1	0.022	0.38	0.04	0	0	0.42	0.25	0.13	0.12
5-2014	2B	108.66	3024021	2.4	0.034	-0.54	8.3	0.024	-0.48	4.2	0.012	-0.39	0.70	0	-0.10	0.07	0.07	0.07	0.12
6-2014 5-2015	5B	50.75	7492401	0.1	0	0.03	4.9	0.013	0.31	5.5	0.017	0.51	1.24	0	0.18	0.17	0.10	0.06	0.06

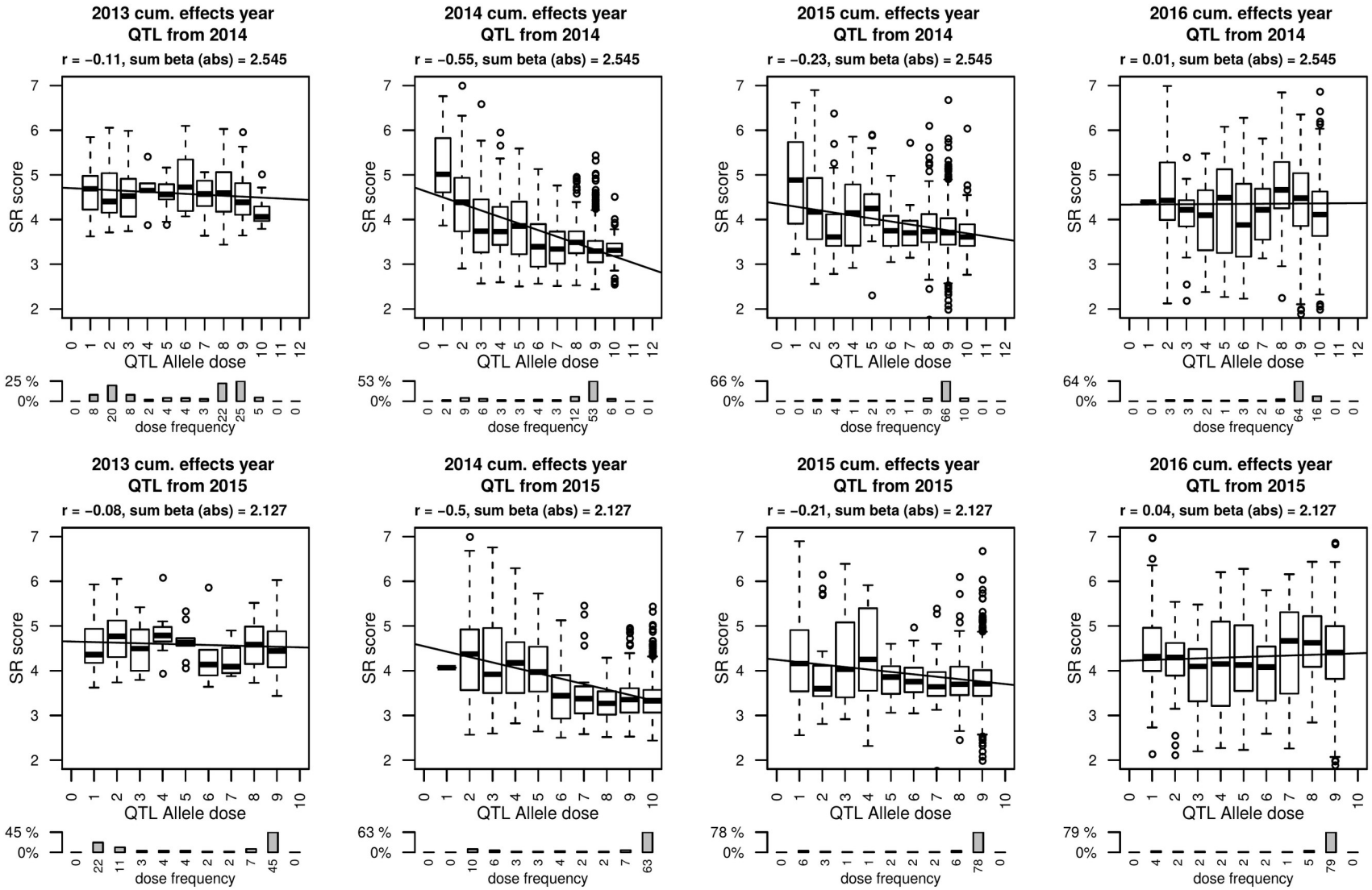


Figure 10: Cumulative effects of QTL for each year using QTL from 2014 or 2015.
Boxplots show the allele dosage effects of QTL, regression lines and effect sums. Bar-plots show: frequency distributions of allele doses in the respective populations. In the top row are QTL found in 2014 and their effects in the phenotypic records from 2013 to 2016, in the bottom row are QTL found in 2015 and their effects in years 2013 to 2016.

In total, there is strong evidence for two QTL on chromosome 2A and one on chromosome 2B. Four other QTL had either low effects or were not detected in multiple years. In the next step, the cumulative effect of the detected QTL on the phenotypic value was investigated more closely. To achieve this, each line was classified according to its total number of favourable allele dosage, with one QTL carrying up to two allele doses. Figure 10 shows these effects for the QTL found in 2014 and 2015 respectively, and the according phenotypes of the lines within each year for 2013-2016. Detailed figures for all years and all QTL, separately, can be found in the Appendix chapter 7.6.2.

It was apparent that the QTL found in 2014 had a major effect on the stripe rust scores for 2014 and 2015, where just a few allele doses significantly decreased the susceptibility, whereas more than three doses did not lead to a major improvement (Figure 10). However, the effect of these QTL seemed to be lower in 2015 than in 2014. On the other hand, the QTL found in 2015 appeared to be even more effective in 2014 than in the year in which they were detected, while most of the QTL were based on the same markers. The QTL detected in 2014 and 2015 did on the other hand not have any positive or negative effect on the phenotypic values for the tested lines in 2013 and 2016. Some dosage classes were not available (e.g. 0, 11, 12), since some alleles were not present homozygously in any genotype, while the frequency of lines carrying multiple dosages of resistance conferring alleles was generally high.

Using the QTL found in the GWAS, a correlation matrix of the markers, as they were present in the whole population of wheat lines under investigation, was produced (Figure 11). It revealed that a lot of the QTL are strongly positively correlated with each other, in the case of a group of four (consisting of “1-2015”, “2-2015/2-2014”, “4-2014/4-2015” and “6-2014/5-2015”) with 70% to 93% correlation, which means that the major alleles are likely to be present together. This group is also strongly negatively correlated to QTL “3-2014/3-2015” which means that the minor allele of one marker is likely to be present with the major allele of the other four. Two QTL, namely “1-2014” and “5-2014” are neither correlated with each other, nor with all the other QTL.

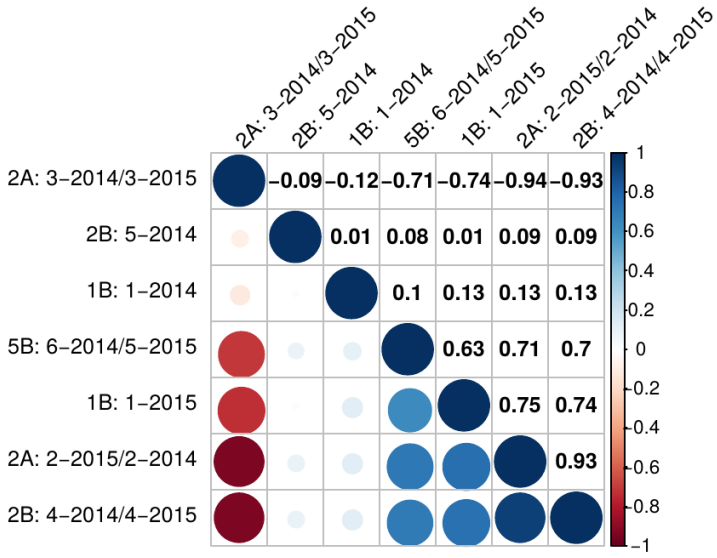


Figure 11: Pairwise correlations of QTL found in the GWAS with correlation coefficients in the top right and colour coded circles according to coefficients in the bottom left.

3.5 Locating QTL in the wheat reference genome

Using the BLASTN tool, the markers found to be significant in the GWAS were located in the wheat reference genome. Some markers were found where they were expected to be according to the genetic map provided by the genotyping service, others were not found at all or on a different chromosome than expected (Table 5).

Table 5: GBS markers, QTL and best-fitting locations on the reference genome of wheat.

Locus name Clone ID	QTL #	TGAC refseq		IWGSC refseq			similarity comment
		Chr	Pos [cM]	Chr	Start pos. [B]	End pos. [B]	
	marker sequence						
1012285	1-2014	1B	201.49	6DL	473516100	473516044	
	TGCAGGAGCTGCGGCTGGAGCAGGAGCTGCGGGTGCTGGTGTGCGTGCACGGCGCCGAGATCGGAAGAG						
1107810	1-2015	1B	96.913	7DL	488123747	488123689	
	TGCAGAGCAGAGCACGCAGACCATGGTGGCCGTCACGGTCGCCGTGACCACTGGCGCCGAGATCGGAAG						
1114850	1-2015	1B	96.91	7AL 6BS	518146417 46351830	518146475 46351772	lower score on 1BL: 540991187 - 540991225
	TGCAGAGCAGAGCACGGAGACCATGGTGGCCGTCACGGTCGCCGTGACCGCTGGCGCCGAGATCGGAAG						
1092023	2-2014/2-2015	2A	8.26	1DL	438585076	438585018	
	TGCAGAGTAGAGCACGGAGACCATGGTGGCCGTCACGGTCGTCGTGACCGCTGGCGCCGAGATCGGAAG						
1021197	2-2014 / 2-2015	2A	8.30	2AS	18495218	18495150	
	TGCAGTTGTGTAAGGGTGCCCTCTGTGAGGTTCACCTGGAACACTGGATTGCACACGGCTAGGTTGCCG						
1206128	2-2014 / 2-2015	2A	8.60	2AS	14215413	14215345	DArT marker 1206128
	TGCAGGTACGTATGTGAATCAAATCAATCACCAGTGCTCTTTTTGTTCCACTCGGTACTATGCTTTCCC						
3023902	2-2014 / 2-2015	2A	8.76	2AS	24002743	24002789	
	TGCAGGGAACCCAAGGACATGGTCGTCTCGTTGTGGCACTTCCTCCGAGATCGGAAGAGCGGTTCAGCA						
3027663	2-2014 / 2-2015	2A	9.11	2AS	15459940	15459898	
	TGCAGCGATGATGGCGCCGAGTTCGTCCACGCCGTGGCGCCCGAGATCGGAAGAGCGGTTCAGCAGGAA						
1695267	2-2014 / 2-2015	2A	9.47	2BS	30430752	30430684	2AS 19447986 - 19447918
	TGCAGGTCTATGATGCATTCTTCCGCCACCTTCGACTGCACCAATTCTTCCTTGACATCTACAAGCTCC						
3026123	2-2014 / 2-2015	2A	13.34	2AS	5928501	5928454	
	TGCAGCCTCAGGAGCACGTCTAGCAGGTCCTCGTTGTCTTTAGTGCCGAGATCGGAAGAGCGGTTCAGC						
3027084	3-2014 / 3-2015	2A	65.86	2AS	24111084	24111040	
	TGCAGCGGCATCAGCCAGAGGTCGTGCAGCCCGTGCATGGCACCGAGATCGGAAGAGCGGTTCAGCAGG						
4544569	3-2014 / 3-2015	2A	65.86	-	not found		
	TGCAGCGCCTTGAGGTTAACGGTTACGCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGAT						
1030280	4-2014 / 4-2015	2B	0.655	2BS	11076245	11076177	also on 2AS/2DS wPt-5960
	TGCAGGATGTTGTCGGTGCTCCACAGGCTGCCGTCGTCGGCGTCGTCGCAGCCGCCGCAGTGGCTGAAA						
3024021	5-2014	2B	108.66	2BL	763798474	763798511	
	TGCAGGCACAACTCTACATCAGCAAGAAGCCTCGGCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCG						
7492401	6-2014 / 5-2015	5B	50.75	2DS	14744721	14744680	
	TGCAGGGTGCTGGTCCAGAAGCCGCCAGGCGGGTTCAGCCCGAGATCGGAAGAGCGGTTCAGCAGGAAT						

3.6 Genomic prediction within years

In addition to conducting a GWAS, genomic prediction for stripe rust resistance was tested in a five-fold cross validation approach, using either the GBLUP or GBLUPA method with different marker densities and training population sizes, resulting in a total of 640 different models combinations.

On the left hand side in Figure 12(a), the effect of the training population size on the within-year prediction ability can be observed. In general, the models using additional information from the previously calculated GWAS (GBLUPA) are performing better than the standard GBLUP models. The best predictions were possible in 2014, while 2016 was second, and 2015 performed worst. In 2013 the number of individuals for training and prediction was rather low, due to the lack of field trial data. Prediction abilities similar to the ones of 2014 and 2016 would be sufficient for effective selection if transferable to across-year prediction. The effect of the training population size is considerable in years with generally lower prediction abilities. While in 2015 an improvement might still be possible by increasing the training population size, in 2014 and 2016 saturation is already noticeable at 50% to 75% of the training data, and little improvement by using the full training population. For 2013 saturation is not yet visible, which means more training data would be beneficial for increasing prediction ability, but similar results as for 2014 or 2016 seem possible. In the GBLUPA models, saturation seems to be reached with a lower number of training data, which means it would be favourable for a breeding program if QTL were established and stable.

On the right hand side in Figure 12(b) within-year prediction ability was calculated with different sets of varying marker density (on a semi-logarithmic scale). The marker set was reduced by randomly selecting a certain number of markers before filtering for MAF. Surprisingly, even a reduction of marker density from the full density of 7365 markers to 100 markers (~35 markers filtered) allowed good prediction, especially in 2014 and 2016, *i.e.* years with generally high prediction ability. A generally lower prediction ability was on the other hand achieved in 2013 and 2015, where the reduction of marker density also led to a stronger decline in prediction ability. Overall, a reduction of marker density from the full set of 7365 to 3000 did not seem to impair the prediction ability, while a reduction to 1000 markers showed a visible reduction. Using additional markers as fixed effects in the GBLUPA models, a stronger reduction of marker density would be possible, as the GBLUPA models generally performed better, leading to similar prediction ability as GBLUP models, even with a lower number of markers.

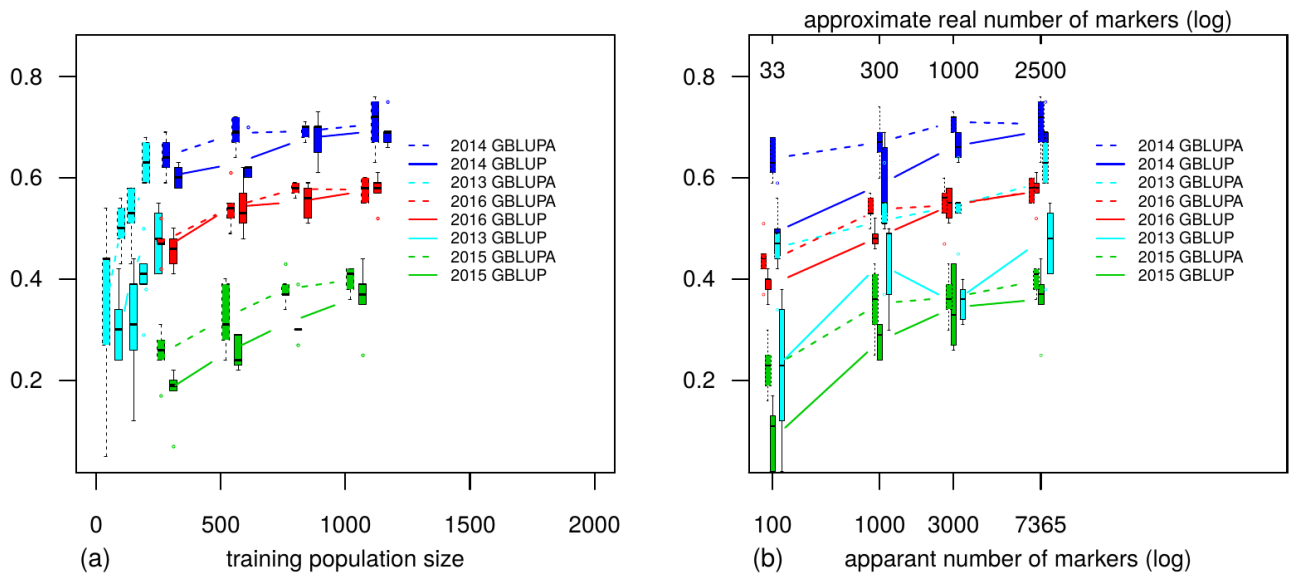


Figure 12: Effects of training population size and marker density on within-year prediction ability per year calculated from 160 simulations.
Figure a on the left hand side shows the effect of training population size on the prediction ability in each year. Figure b on the right hand side shows the effect of marker density size on the prediction ability in each year. Boxplots in the graphs are jittered for better visibility and GBLUPA is a variant of the GBLUP model that uses additional QTL as fixed effects.

In Chapter 7.6.3 in the Appendix, detailed plots for all variations of marker density and the amount of training data are documented.

3.7 Genomic prediction across years, for untested and tested years

When moving from prediction within years to prediction across years, one might expect prediction ability to increase due to a larger number of individuals and a larger number of plots available for model training, or decrease, due to differences in years and environmental effects, which make prediction more difficult. The results obtained in this thesis, are shown in Figure 13.

The highest prediction ability was achieved when predicting the year 2014, while there was some variation of the prediction ability in terms of training years with all training year combinations containing 2015 performing best. The prediction ability for 2015 was, in contrast, a lot lower, although being highest when 2014 was used for training compared to other training years. Having prior information about the year's performance in the model ("tested" variant) surprisingly performed worse than having no prior information when predicting 2013, while for 2014 and 2015, the models with prior information performed better than predicting untested genotypes in untested years. Contrary to expectations, the number of training years did not improve prediction ability by a large extent, while in the within-year predictions the GBLUPA model generally performed better than the GBLUP model, which could not be observed in across-year prediction.

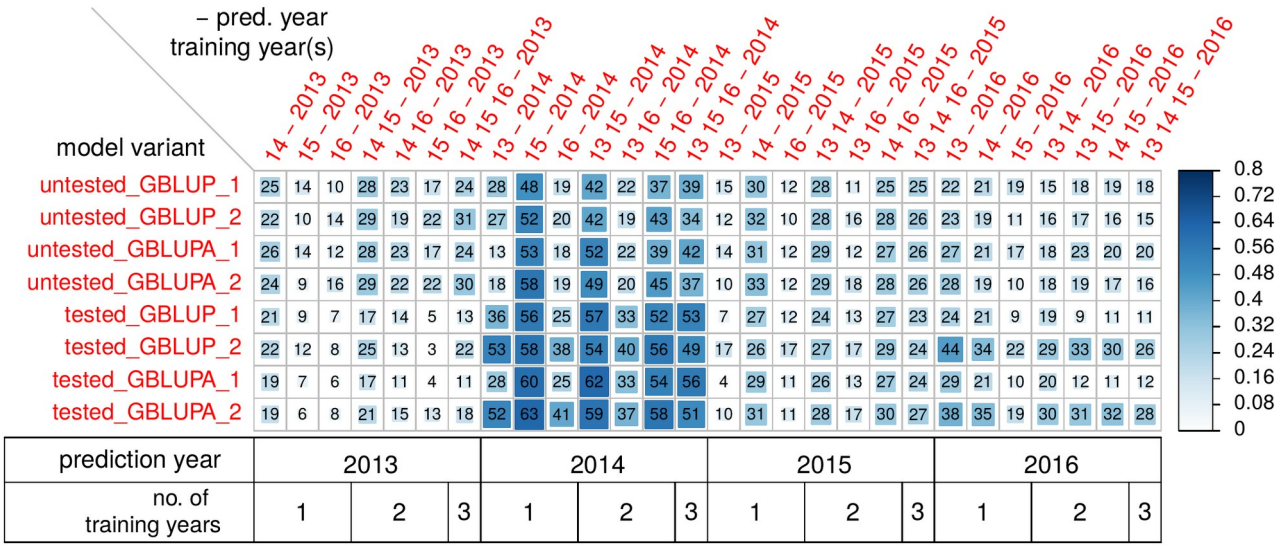


Figure 13: Across year prediction abilities of 224 simulations.
GS model variants on the vertical axis (GBLUP vs. GBLUPA, untested vs. tested prediction years and variants "1" and "2" of data partitioning) and years (grouped into validation, training and number of training years) on the horizontal axis. Variants for untested years: 1 = lines present in training and validation years eliminated from validation set (more lines in training set), 2 = eliminated duplicates from training set (more lines in validation year set). Variants for tested years: 1 = lines present in training and validation years used for estimation of year effect, 2 = additionally, 10% of the validation set lines used for year effect estimation (more training data).

The results of the ANOVA, using the prediction ability of across-year results to find significant effects influencing prediction ability, can be found in Table 6, factor levels and effects are found in Table 7. While a lot of factors can be identified as having significant influence in prediction ability, the ANOVA and effects table show that prediction year had the largest effect, with 2014 performing best and 2013 performing worst. A similar picture is visible by inspecting the effect of the training years, as including 2014 and to a lesser extent 2015, increased prediction ability. Whereas training models with 2013 data did not improve prediction ability and including 2016 into the training population reduced the prediction ability significantly. The positive effect of including QTL in the within-year prediction model (factor GS model) is not visible in across-year predictions (Table 6 and Figure 13). As one would expect, the factor "Tested", which indicates whether there is prior knowledge of the year that is being predicted, has a significant positive effect on prediction ability. However, having prior knowledge or not depends on the strategy of the breeding program and the specific strategy in implementing GS. A similar effect size was found in the training data partitioning variants (factor "tested:variant"), which are mostly an indicator of the effect of training population size which was not investigated separately in across-year prediction.

Table 6: Prediction models ANOVA table, predictions across years, response prediction ability.

Factor	Df	Sum Sq.	Mean Sq.	F value	p value	sig.
Tested	1	0.0559	0.0559	7.79	0.0057	**
GS model	1	0.0016	0.0016	0.22	0.6366	n.s
Prediction Year	3	1.987	0.6624	92.29	<2.2e-16	***
Training 2013	1	0.0005	0.0005	0.072	0.7888	n.s.
Training 2014	1	0.2693	0.2693	37.52	4.33e-09	***
Training 2015	1	0.1308	0.1308	18.22	2.96e-05	***
Training 2016	1	0.0879	0.0879	12.25	0.00057	***
Tested:variant	2	0.1353	0.0676	9.427	0.00012	***
Residuals	212	1.522	0.0072			
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						

Table 7: Prediction models ANOVA, table of effects.

Factor / levels	Level 1		Level 2		Level 3		Level 4	
Tested	untested [c]	-0.0158	tested [c]	0.0158				
GS model	GBLUP[c]	-0.0027	GBLUPA[c]	0.0027				
Prediction Year	2013[a]	-0.073	2014[a]	0.162	2015[a]	-0.048	2016[a]	-0.041
Training 2013	no[d]	0.0011	yes[b]	-0.0015				
Training 2014	no[d]	0.0345	yes[b]	-0.026				
Training 2015	no[d]	-0.018	yes[b]	0.024				
Training 2016	no[d]	0.0145	yes[b]	-0.193				
Tested:variant	untested:1[a]	0.01	tested:1[a]	-0.03	untested:2[a]	-0.01	tested:2[a]	0.03

Factor level description, see Figure 13 description. Number of observations: 56[a], 96[b], 112[c], 128[d]

Testing for the number of years used for model training as well as for the effect of each training year at the same time is not possible in the dataset at hand since the factors are collinear.

3.8 Genomic prediction and preliminary yield trials

Only a limited number of year combinations can be tested in this case, since only forward prediction might be regarded as being reasonable. Results were similar, or better, compared to the previous across-year prediction ability simulation. Additionally, prediction accuracy was listed in the results as well as the number of entries. Prediction ability results differed only slightly between using all available data for training (Table 8) and using only the PYT entries for model training (Table 9). Depending on the year only one half or one third of the entries were advanced from PYT to yield trials, as can be seen in the number of entries in Table 9. In all cases, GBLUPA is performing slightly worse than the basic GBLUP model. Phenotypic correlation is the Pearson correlation coefficient of entries from training year with entries from the validation year, similar to the prediction ability, which is the correlation of validation data and the predicted data. Prediction abilities are generally in the range of phenotypic correlations.

Table 8: Results from preliminary yield trials simulations using all entries from training year for model training.

training year	validation year	n. training entries	n valid. entries	phenotypic correlation	prediction ability		prediction accuracy	
					GBLUP	GBLUPA	GBLUP	GBLUPA
2014	2015	1407	346	0.46	0.34	0.29	0.40	0.34
2015	2016	1276	379	0.36	0.44	0.42	0.67	0.64
2014	2016	1407	62	0.38	0.21	0.20	0.33	0.31

Table 9: Results from preliminary yield trials simulations using only PYT entries for model training.

training year	validation year	n. PYT entries	n valid. entries	phenotypic correlation	prediction ability		prediction accuracy	
					GBLUP	GBLUPA	GBLUP	GBLUPA
2014	2015	1004	346	0.46	0.35	0.32	0.41	0.37
2015	2016	705	379	0.36	0.41	0.39	0.62	0.60
2014	2016	1004	62	0.38	0.28	0.19	0.43	0.29

4 Discussion

Breeding and growing wheat varieties with resistance to stripe rust is part of a sustainable plant production strategy to maintain the status of wheat as a profitable crop for conventional, organic and low-input farming. Using genomic tools is promising to aid the breeders in achieving this goal. The goal of this thesis was to find out if state of the art genomic prediction can be successfully used to predict stripe rust resistance of breeding lines, without having to conduct extensive field testing, especially in years, when no natural disease pressure is present. For this purpose, stripe rust recordings from plant breeding trials in years with a relative high disease pressure and genetic marker data was used. The phenotypic recordings have been preprocessed to assess the data quality and to accommodate for field variability, and clustering was used to find groups of similarly behaving lines across years. Further analysis aimed at using the genetic marker data to map QTL or possibly active resistance genes, using a genome-wide association study. Finally, the phenotypic and the whole genotypic data was used to simulate different situations in which stripe rust resistance was to be predicted.

4.1 Phenotypic Analysis

Heritability is a measure of trial quality, and the results in this investigation revealed a large variation in heritabilities across all years. In 2013, heritabilities were generally low, *i.e.* after filtering for low heritability trials only the disease nursery was left. Therefore the data from 2013 was not particularly useful, since disease nursery results seemed to differ from naturally infected trials in the subsequent years of 2014 to 2016. In the other years, a lot of high heritability trials were left after filtering leading to relatively high heritabilities in the across trial analysis of each year. Only 2016 resulted in a medium heritability, probably due to high variability across trials. One explanation for the low heritability of 2016 could be that due to different stripe rust races being present at different sites, varying susceptibilities led to contrasting scorings across sites.

The two-stage phenotypic analysis resulted in adjusted means for stripe rust resistance per line and per year, with the 2013 data being only partially comparable due to including only the disease nursery data. While, again, large variability was present in each year, the mean stripe rust susceptibility increased slightly from one year to the next (2014 to 2016). This is, on the one hand, surprising, since in a breeding population selection for resistance takes place. On the other hand it is to be expected, since new lines are added each year to determine their susceptibility. According to Oberforster (2018), who reported stripe rust incidence in the official variety trials in Austria, a stripe rust

epidemic happened between 1998 and 2001, indicated by high incidence. Starting from 2002 up to 2012, stripe rust incidence was very low, with a further increase in incidence starting from 2013 and lasting up to 2016, marking a new epidemic outbreak happening in the years analysed in this thesis. This epidemic is attributable not only to favourable environmental conditions for stripe rust infection and proliferation, but also to a change in stripe rust pathotypes which occurred in the preceding years. From 2011 onwards the new race “Warrior” (PstS7 grouped with PstS8 “Kranich”) was the most prevalent in Europe, while from 2013 onwards, a pathotype called “Warrior -” (PstS10) was starting to gain importance (GRRC, 2020; Hovmøller *et al.*, 2016).

Not only newly emerging races are a problem for breeders, due to the breakdown of resistance genes, but also changing prevalence of stripe rust races are. The correlations of scoring results across years, were used as one way to find out if the analysed years were behaving similarly or differently. Medium correlations were found between 2013/2016, as well as between 2015/2016. 2013/2015 and 2014/2015 showed medium low correlation, while 2013/2014 as well as 2014/2016 showed a very low correlation. This indicates that the breeding lines behaved rather heterogeneously in the different years, which will also make stripe rust resistance prediction challenging. One attempt to add some structure to the data was made by clustering the genotypes present in all years into groups which behave similar across years. This resulted in a solution with 4 clusters, two of which showed a crossover between 2014 and 2016. This means that previously (2014) resistant varieties were more susceptible in 2016. The rest of the groups only showed differentiation into levels of susceptibility, rather than changes across lines. At this stage, a variety of different reasons could be pointed out for the low comparability across years. One reason could be inconsistently or poorly scored trials. This was not only eliminated by using only trials with high variance in the scoring and a minimum size of 60 plots, but also by looking at the heritability in the first stage of the analysis as a data quality measure, and by only using trials with at least medium heritabilities. Another source of error could be the two-stage design of the mixed model preprocessing step, since the variance information of each line which is associated with the adjusted means was not carried on to the second stage of the analysis. This could be done by using variances as weights in the second stage. However, this was computationally too demanding and since it is not strictly necessary (Möhring and Piepho, 2009), this was deemed not to be a major issue. It could also be possible that different stripe rust races were present not only in different years, but also on different locations within years.

In an attempt to find out how well trials agree with one another within years, a correlation matrix was produced, based on pairwise observations, to visualise differences and similarities of trials across years and within years. Few trials showed statistically significant correlation in their scoring results. Some highly correlated trials were found within years, but not all showed statistically significant correlation. Across years, statistically significant correlations were even rarer. Missing correlations across trials or low significance of correlations is sometimes caused by a low number of common genotypes. Especially yield trials do usually test a small number of genotypes and therefore only two or three "standard" genotypes allow comparability. Low correlation coefficients are however also caused by differing scoring results, with some genotypes performing well in one trial and not so well in other trials.

4.2 Genome-wide association study

During the genetic analysis markers were selected according to allele frequencies, excluding markers with an allele frequency below 5%. From the 7365 markers only one third, around 2500, were left after this filtering. An explanation for such a large number of markers with low allele frequencies could be that this is a breeding population which is heavily selected towards specific traits and attributes. The genome-wide association study found some single markers and groups of markers to be significant in explaining stripe rust susceptibility in 2014 and 2015 respectively. None were found in 2013 and 2016. Stripe rust incidence in 2013 and 2016 was found to be quite high in general. Possible reasons for missing QTL are the following:

- For 2013, only the disease nursery data was available: Lines did not properly differentiate in terms of stripe rust susceptibility due to mixed infections with other diseases. In the disease nursery, other pathotypes of stripe rust may be present than are prevalent in natural infected trials, which makes the data from the disease nursery only partly comparable with other trials.

- Presence of different stripe rust races: Looking at the 2014 and 2015 QTL allele dosage effects on the stripe rust scores of 2016, one can clearly see that these QTL were not useful in 2016. Maybe some effect might be visible in 2013. This would explain why the QTL found in 2014 and 2015 were not found in 2016, but would also suggest that races present in 2014 and 2015 were already part of the inoculum used in the disease nursery of 2013. According to information from the Global Rust Research Center (GRRC, 2020), from 2013 onwards "Warrior" (PstS7) was slowly replaced by "Warrior-" (PstS10). However, these races are

quite similar in virulence, only differentiating on phenotype “Amb”. In lower prevalence stripe rust race “Kranich” (PstS8) was present, mostly in northern Europe, which differs from the above-mentioned in a few additional virulence phenotypes, but none of the putative Yr genes considered in the next section.

In the process of locating the markers in the reference genome of wheat, QTLs “1-2014” and “1-2015”, which were supposed to be on chromosome 1BL, were located with high confidence on 6D and 7D. Due to this result, and having just barely passed the significance threshold in the GWAS, they will not be discussed further. QTL “4-2014/4-2015” was supposed to be close to the telomeric end of the short arm of chromosome 2B but was located to be on chromosome 2D. Despite having a strong signal according to the p-value, possible candidates resistance genes for this QTL will not be discussed further due to the uncertainty in its chromosomal location.

Another dislocated QTL is “6-2014/5-2014”, which was located on chromosome 2D, while it was supposed to be on 5BS, is therefore also excluded from resistance gene discussion. However, even if QTL cannot be located precisely, compared to resistance genes they show signs of selection. Allele frequencies are shifted towards the resistant allele through the years, up to the point of being fixed in the population (“1-2014”, 1% in 2016). Additionally, none of the above is present in the susceptible homozygous state, except “1-2015” in 2014. Effect size indicates, compared to the other QTL in this thesis, medium to high quantitative influence on stripe rust resistance.

The QTL “2-2014/2-2015” is located close to the telomeric end of chromosome 2AS. It was defined by significant effects on up to 8 markers, all of which show comparably high effects, and allele frequencies indicate selection towards the resistant allele. Resistance genes on chromosome 2A(S) are Yr17, Yr56 (Qyr.sun-2A), Yr69 (YrCH86, linkage with Yr17 (Hou *et al.*, 2016)) as well as temporarily designated resistance gene YrR61 (McIntosh *et al.*, 2017). Ledesma-Ramírez *et al*. (2019) also found the marker 1206128, which is part of the QTL “2-2014/2-2015”, to be associated with stripe rust resistance in a study with 419 pre-breeding lines developed at the International Maize and Wheat Improvement Center (CIMMYT). They argue that Yr17 and YR56 are in close proximity of this marker. According to this information, it is likely that QTL “2-2014/2-2015” is associated with the linkage group also containing Yr17, Yr56 and Yr69, and not distinguishable with the resolution of this map. In 2014, marker 1206128 showed up as the peak marker on this QTL, while in 2015, marker 3026123 had the strongest signal. Due to this, 3026123 could be considered an additional QTL being located 5cM from marker 1206128 and 4cM from the closest other marker of this QTL. However, it was assigned to this QTL due to unclear separation by interspaced markers. According

to literature (Sajid Ali *et al.*, 2017; Hovmøller *et al.*, 2016) recently prevalent races “Warrior” and “Warrior-” are virulent on Yr17, and in fact epidemic occurrence of stripe rust in the late 1990s and early 2000s in Europe was due to the breakdown of this resistance gene (Bayles *et al.*, 2000). For Yr56, Yr61 and YrR69 no data on virulence is available.

The other QTL “3-2014/3-2015” on chromosome 2AS is closer to the centromere. According to the Catalogue of Gene Symbols for Wheat (McIntosh *et al.*, 2017), and previous considerations, only temporarily designated resistance gene YrR61 is left as a candidate for this locus or the marker 3026123 of QTL “2-2014/2-2015”, which might be a separate QTL. No information on virulence of stripe rust on this resistance gene is available. According to allele frequencies across the years, this locus was heavily selected towards the resistant allele. While in 2013 the allele frequencies were roughly equal on both alleles (50%), in the last year 2016, the frequency of the resistant allele is 88%, which means 78% of lines are homozygous for the resistant QTL allele and 1.5% are homozygous susceptible, while the rest is heterozygous at this locus. The additive effect was found to be comparably large at this locus in the GWAS.

Even though being on different chromosomes (2A, 2B and 5B), QTL “2-2014/2-2015”, QTL “4-2014/4-2015” and QTL “6-2014/5-2015” are highly correlated in occurrence. Surprisingly, the marker associated with QTL “6-2014/5-2015” which is supposed to be on chromosome 5B was located on chromosome 2D using BLASTN, as well as with lower scores on chromosome 2B and 2AS. Additionally, the markers associated with QTL “1-2015” and “4-2014/4-2015”, were also found with a lower score on chromosome 2AS. This would indicate that these three QTL, while supposedly being dislocated from one another, are really one QTL which could not be detected due to low quality of the marker map. Further investigation of the correlations between markers showed that QTL “3-2014/3-2015”, which is also located on chromosome 2AS, is highly negatively correlated (which means that the opposite allele is favourable) with the previously mentioned QTL. This is surprising, since according to the genetic map and BLASTN results, these QTL should be on separate locations, with one being closer to the telomeric end, and one being close to the centromere. This unexpected linkage can only be explained by the markers and QTL being together on a translocated part of the chromosome, which has a low recombination rate and therefore high linkage disequilibrium, which was already described in the literature (Bremenkamp-Barrett *et al.*, 2008; Helguera *et al.*, 2003; Xue *et al.*, 2018). The 2AS-NS translocation supposedly confers resistance to leaf rust, which is not associated with yield penalties but increased protein content and reduced mixing strength (Dyck and Lukow, 1988), as well as powdery mildew resistance (Bariana and McIn-

tosh, 1993). A large study, using large-scale phenotypic and genomic data from CIMMYT, revealed that this translocation carries not only resistance to all three rusts (Stem Rust, stripe rust, Leaf Rust) but is also positively influencing grain yield and stability (Juliana *et al.*, 2019). Further QTL "1-2014" and "5-2014" are not correlated to each other or any other QTL.

Another important QTL in this thesis is "5-2014", which is located on chromosome 2BL. Associated with this chromosomal region are a number of stripe rust resistance genes and resistance gene candidates: Yr5 and Yr7 (Zhang *et al.* (2009) reported Yr5, Yr7 and YrSp to be allelic), Yr43, Yr44, Yr53, Yr72, and temporarily designated resistance genes YrV23, YrS2199, YrSte, YrSp (McIntosh *et al.*, 2017). According to Ali *et al.* (2017) and Hovmøller *et al.* (2016), "Warrior" and "Warrior-" races are virulent for Yr7 and YrSp from the above-mentioned candidates, but not for Yr5. For the other proposed resistance genes no virulence information is available for Europe, while for example in China, races virulent on Yr43 and Yr44 have been recorded (Wan, Muleta, *et al.*, 2017). For Yr5 no virulent race has been recorded so far, neither in Europe nor China. Muleta *et al.* (2017) located the group containing Yr43, Yr44 and Yr53 close to the telomeric end of chromosome 2BL and since the marker was found to be close to the telomeric end of 2B by using BLASTN, it is more likely that this QTL is associated with one of these resistance genes, rather than the group containing Yr5/Yr7, or in fact is a novel QTL. Xu *et al.* (2013) used deletion lines and SSR markers to map the interval between Yr5, Yr44, Yr53 and Yr43. According to this mapping, Yr53 is between the physical locations of Xwmc441 and Xwmc149 SSR markers, and according to the physical location of these markers on the reference genome, the marker found for QTL "5-2014" (Marker 3024021) is in the same interval. This means that it is highly likely that QTL "5-2014" found in this thesis, is in fact Yr53 introduced from Ethiopian durum wheat, which was first described by Xu *et al.* (2013). This QTL is quite interesting for breeding, since it was found to have a high additive affect, while not even being present in the homozygous resistant state. In 2013, the allele frequency of the resistant allele was merely 7% in the breeding population, rising to only 12% in 2016 with no homozygous lines being present. This makes it highly promising for future breeding activities, but it could also mean that it is in strong linkage with a non-wanted locus. However, even if it was not found by the GWAS in 2016, it still shows some effect on stripe rust susceptibility in the comparative box-plot in Appendix 7.6.2.

This leaves us with two QTL, 5-2014 and 2-2014/2-2015, one being still available for selection, and the other locus being mostly fixed in the population. Since it was not the main goal of this thesis to find marker-trait associations, the GWAS was done with the year-wise means, so only QTL per

year could be detected. A more sophisticated approach would have been to look for QTL in each trial of each year. The results from such an approach would not be robust since line recordings in trial are mostly not replicated, but it would certainly have resulted in a larger number and maybe some interesting QTL. However, the aim of this thesis was to employ genomic prediction based on the whole genome rather than single markers, which will be discussed in the next section.

4.3 Genomic prediction

The main goal of this thesis was the prediction of stripe rust susceptibility in wheat breeding lines by using genetic marker data as a way of reducing the need for field testing, especially in non-virulent years. In the previous sections of this chapter, the phenotypic analysis and QTL analysis was outlined, in this section an attempt to answer the main research question will be made. Is it possible to use genomic prediction models to reliably predict stripe rust susceptibility not only within but also across years?

Prediction of performance within years is only partly useful. It allows a breeder to reduce the number of entries tested in a year, since other individual's performance can be inferred by markers, but it is unknown if the model is useful in future years. However, it is very much suited to simulate and optimize model parameters. The model of choice in this thesis was GBLUP, since it is computationally efficient and the most widely used in plant breeding. Different marker densities, training population sizes and a model with additional QTL from the GWAS as fixed effects were simulated. In the models that were aimed at predicting performance across years, variations included: the number of years for training, all combinations of training and validation years available, and an approach which included some data from the validation year into the training data to improve the model performance.

In most genomic prediction studies the main focus is on (grain) yield and quality traits (e.g. protein content) and less on disease traits, which of course influence yield and quality but are not necessarily correlated. In the following section, results from this thesis will be compared to results from literature, but for the specific topic of predicting stripe rust in wheat only a few studies were found to be comparable.

- Ornella *et al. (2012)* compared different prediction models for stem rust and stripe rust in biparental wheat populations, not applying a GBLUP model but a ridge regression BLUP model (rrBLUP) which usually performs equally well. 1400 DArT markers were used to

predict performance across and within populations in two locations. They found a lower predictability of stripe rust than stem rust.

- Daetwyler *et al.* (2014) investigated stem, stripe and leaf rust in four years on a collection of Australian wheat landraces. Marker data was extracted using a 9k SNP marker chip. Prediction models were GBLUP, GBLUP with fixed QTL effects (GBLUPA) and BayesR.

- Rutkoski *et al.* (2014) worked with a large CIMMYT dataset from 6 years and 12 environments to predict stem rust resistance in wheat, comparing different models, applying GBLUP and GBLUPA models.

- Juliana *et al.* (2017) used data from 4 years of replicated CIMMYT trials and GBS markers to assess the prediction accuracy for leaf, stem and stripe rust. Different model types were compared, applying GBLUP and GBLUPA models.

- Muleta *et al.* (2017) investigated a germplasm collection of spring wheat in two seasons for stripe rust resistance and performed genomic prediction. Different marker densities of a 9k SNP chip and training population sizes were simulated to find optimal values for these parameters.

- Juliana *et al.* (2019) conducted a large genomic study (44624 lines, 78606 GBS markers, 4 years, evaluated in America, Africa and South-East Asia), looking at 35 traits in bread wheat, one of which was stripe rust. GWAS was performed and Genomic prediction models with different marker densities and prediction within as well as across panels were evaluated. The 2AS-NS translocation described earlier in our study was also found in this thesis as effective against rust diseases.

- Miedaner *et al.* (2019) did a study about genomic prediction of stem rust and stripe rust. 12000 GBS markers and scoring data from trials in 2 years were used to compare three different models (marker-assisted selection, rrBLUP, rrBLUP with fixed QTL effects). Three QTL for stripe rust were found and prediction ability of stripe rust was found to be similar to prediction ability of stem rust.

Comparing the **absolute values** of model performance between studies is challenging for several reasons. Some studies report prediction accuracy while others report prediction ability, and the associated heritabilities to link these parameters are not always given. A further problem is that differ-

ent marker densities, training populations sizes and other design factors can influence these values. However, minimum and maximum values for optimal design parameters will be provided in the following, as an attempt to compare the results. In this thesis, the maximum within-year prediction abilities per year ranged from 0.3 to 0.7. In prediction across-year, the maximum (optimal) prediction abilities per year were found to be around 0.3 in 2013 up to 0.65 for 2014, with some combinations being almost zero, and the mean value being 0.26. It was found that prediction ability depends not only but mostly on the validation year, which means that the model fits better for some years than for others. Ornella *et al.* (2012) reported prediction ability within populations between 0.15 and 0.6, while across populations the predictive ability decreased to a value of 0.05 to 0.16. Compared to this result, our study achieved somewhat better results. Daetwyler et al (2014) reported prediction accuracy and heritability, prediction ability was calculated to compare the results with our study. They found a prediction ability of 0.31 for GBLUP and a prediction ability of 0.33 for GBLUPA. Compared to the results in this thesis, this is slightly better than the mean prediction ability. Miedaner *et al.* (2019) found across-year and trial prediction ability to be 0.5 for rrBLUP and 0.55 for rrBLUP with fixed QTL effects. The following results were reported as prediction accuracy, the values for prediction accuracy in our study ranged, for within-year prediction from 0.55 to 0.75, and for across-year prediction from 0.32 to 0.68. Juliana *et al.* (2017) reported prediction accuracies for across trials, within populations, and also highly dependent on the year used for model validation, ranging from 0.39 to 0.73 for GBLUP and 0.41 to 0.78 for GBLUPA, which are slightly higher than in our study. Muleta *et al.* (2017) reported peak within trial prediction accuracy to be between 0.6 and 0.65. In another extensive study, Juliana *et al.* (2019) found prediction accuracies for within panel predictions to be 0.52 and 0.55 and across panel predictions between 0.23 and 0.37. All these results show a few things and while some consensus about the methods to report results is established, details about populations and models make it hard to directly compare the results. However, the results found in our study are well within the results reported in other studies.

In within-year prediction, including **QTL as fixed effects** (GBLUPA) was found to increase predictive ability and always outperformed the simpler model without fixed effects. Especially in situations with a low number of markers and a small training population size, as opposed to a large number of markers and a large training population, the difference in predictive ability was negligible. In across-year prediction, this effect disappeared almost immediately and prediction abilities can be substantially lower or higher, but are mostly equal. Some studies find that including QTL as fixed effects does consistently increase predictability (Daetwyler *et al.*, 2014; Miedaner *et al.*, 2019), while Rutkoski *et al.* (2014) also found that GBLUPA does sometimes, but not always out-

perform GBLUP. Juliana *et al.* (2017) found GBLUPA to be equal or better than GBLUP for stripe rust, however for leaf rust, GBLUPA performed worse than GBLUP. In this thesis, the GBLUPA models might be performing worse due to overfitting caused by autocorrelation in the QTL used as fixed effects, which was outlined in the discussion of QTL.

In this thesis, the effect of **marker density** on prediction ability was investigated in within-year prediction simulations. While a large improvement in prediction ability was possible by increasing the number of markers from 100 to 1000, further increasing the number of markers to the full set of 7365 only had a marginal benefit. In all years, except in 2015, only including known QTL (GBLUPA) and 100 randomly selected markers yields prediction abilities close to the best performing models with the full marker set, which is probably due to the QTL explaining a large part of the variation. Muleta *et al.* (2017) compared genomic prediction models, using a variety of marker densities from 1 SNP per 1.3cM (~6000 markers total) to 1 SNP per 14.8cM, and found saturation at 1 SNP per 3.2cM, which is ~1850 markers total, noting that this is higher than in other GP studies on different traits and species. The simulation results from this thesis also show similar results with saturation being reached around 1000 markers. More simulations with a higher resolution of marker densities would be needed to really find an optimum value, but based on these results, it could be advised that obtaining 1000 markers for genomic prediction of stripe rust is feasible within years. However, simulations of marker densities across years were not performed and might lead to different conclusions.

Using a five-fold CV approach, it was possible to simulate four different **training population sizes**, in within-year prediction. The results from this simulation are highly dependent on the year of simulation. In some years (2013, 2015), no saturation of prediction ability can be observed using the available data of ~250 and ~1000 lines, in the years 2014 and 2016, however, saturation is reached around 500 training individuals. Again, these results can be compared to the study done by Muleta *et al.* (2017), who found saturation only at the largest training population size of 959 lines in within-year prediction, or did not find signs of saturation, suggesting that a larger population size would be needed to optimize prediction ability.

In the genomic prediction simulations within years, the main factor influencing the variation in prediction ability is the **year of training and prediction**. When simulating prediction across years, the effect of training and prediction year might be distinguished. However, most combinations of training and prediction years yielded low prediction accuracies, only a few combinations showed results that indicate useful models. Most combinations using 2015 data to train the model and predicting

2014 performed reasonably well, with prediction accuracy in the range of 0.37 to 0.63 (prediction accuracy 0.39 to 0.67). Even though showing up as significant effect in the ANOVA, including data from more than one training year did not consistently improve model performance, neither did including some data from the validation year into the prediction model. The breakdown of model performance when moving to prediction across years can also be found in other studies, Ornella *et al.* (2012) reported prediction ability up to 0.56 within environments and prediction ability of 0.05 to 0.15 across two environments (years) for rrBLUP, similarly Juliana *et al.* (2019) report prediction accuracy of 0.52 and 0.59 within panels, and 0.23 and 0.37 across panels using GBLUP.

Simulating the situation of a breeder, using PYT showed similar or even better results than using all the available data. Using more data for model training and validation, does not necessarily lead to to higher prediction ability. Using less validation data narrows the variability in the data, and therefore increases prediction ability, but not necessarily the performance of the model. Using less training data, and therefore less genetic variability, might improve the accuracy of the GRM, since entries might be more related.

All of these factors lead to the conclusion that variability across and within years is a major problem in predicting stripe rust. Investigation of the correlations of stripe rust across all trials showed that scoring results differ between pairs of trials. If prediction models are therefore fed with data that is conflicting, it is not surprising that the predictions obtained from such models are not performing very well when validated. The conclusion of this is, that this thesis should be repeated, with a very rigorous preselection of trials. Only trials with high stripe rust infestation, showing clear separation of lines, and correlating well with each other should be used to train the prediction models. If trials show high stripe rust infections but do not correlate well with other trials, different stripe rust races might be present, these should not be included in the prediction model.

Due to the race-specific nature of stripe rust, there is no “true” breeding value for stripe rust, the breeding value would be different for each race. If one wanted to breed **only** for quantitative resistance, the selection process would be impossible, since e.g. varieties with a low quantitative resistance, could be highly resistant due to one ASR resistance gene. Breeding for real quantitative resistance would therefore only be possible in a set of lines with absolutely no major resistance genes present, and such a set is hard to impossible to create in the real world.

5 Summary, Outlook, Conclusion

In this thesis, a large number of field trials across multiple years were used to dissect the genetic architecture of stripe rust in wheat and develop genetic prediction models using a breeding population. Solid evidence for two QTL was found, while one of them might be part of a large translocation from *Aegilops ventricosa* and being almost fixed in the population, the other one is highly interesting due to its large effect and still being available for selection. Genomic prediction results were promising when validated within years, but failed to perform across years.

Advanced methods of GWAS are focused on using variable selection prior to GWAS to increase its statistical power, and to potentially find more QTL (Mieth *et al.*, 2016). Further improvements in prediction accuracy of genomic prediction might also be possible by using more sophisticated approaches. Incorporating genotype by environment interaction (GxE) into prediction models can potentially increase prediction performance by 10 – 40% (Crossa *et al.*, 2017). The models used in this thesis do only capture additive genetic effects, while Random Forrest for GWAS or GP could also capture epistatic effects (Brieuc *et al.*, 2018). Dominance effects are not relevant for inbreeding crops like wheat, at least when aiming to develop line varieties. Some argue, that machine learning methods (like support vector classification) are useful in predicting ordinal scored traits like stripe rust (González-Camacho *et al.*, 2018).

The number of methods proposed for genomic prediction is almost unfathomable, but the best prediction model can only be as good as the training data that is fed into it and the genetic variability available in the breeding population. For a disease like stripe rust, which is highly destructive in the field and appears in new races, due to spontaneous recombination a non-race-specific resistance is highly desirable. Race-specific resistance genes will still be important in the short term, in years of high disease pressure. However, slow rusting varieties with quantitative resistance should be built up in the long term, since these varieties do not only reduce disease severity but also the selection pressure on the pathogen.

The poor performance of genomic prediction across years might be improved, by doing a more rigorous preselection of trials before further analysis, using only trials for model training that show strong differentiation and correlate well with other trials.

6 Literature References

AGES. (2019). Österreichische Beschreibende Sortenliste 2019, Landwirtschaftliche Pflanzenarten. Schriftenreihe 21/2019, ISSN 1560-635X. (Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH, Ed.). Retrieved from https://bsl.baes.gv.at

Aktar-Uz-Zaman, M., Tuhina-Khatun, M., Hanafi, M. M., Sahebi, M. (2017). Genetic analysis of rust resistance genes in global wheat cultivars: an overview. *Biotechnology & Biotechnological Equipment*, *31*(3), 431–445. https://doi.org/10/ggkw75

Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., … de Vallavieille-Pope, C. (2014). Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. *PLoS Pathogens*, *10*(1), e1003903. https://doi.org/10/ggkvc6

Ali, Sajid, Rodriguez-Algaba, J., Thach, T., Sørensen, C. K., Hansen, J. G., Lassen, P., … Hovmøller, M. S. (2017). Yellow Rust Epidemics Worldwide Were Caused by Pathogen Races from Divergent Genetic Lineages. *Frontiers in Plant Science*, *8*, 1057. https://doi.org/10/ggkw8s

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10/d8t8f7

BAES. (2020). Pflanzenschutzmittelregister - Verzeichnis der in Österreich zugelassenen Pflanzenschutzmittel. Retrieved from https://psmregister.baes.gv.at/psmregister/

Bajgain, P., Rouse, M. N., Anderson, J. A. (2016). Comparing Genotyping-by-Sequencing and Single Nucleotide Polymorphism Chip Genotyping for Quantitative Trait Loci Mapping in Wheat. *Crop Science*, *56*(1), 232–248. https://doi.org/10/f76nr7

Bariana, H. S. (2003). DISEASES | Breeding for Disease Resistance. In *Encyclopedia of Applied Plant Sciences* (pp. 244–253). Elsevier. https://doi.org/10.1016/B0-12-227050-9/00150-2

Bariana, H. S., McIntosh, R. A. (1993). Cytogenetic studies in wheat. XV. Location of rust resistance genes in VPM1 and their genetic linkage with other disease resistance genes in chromosome 2A. *Genome*, *36*(3), 476–482. https://doi.org/10/bpnxjs

Bayles, R. A., Flath, K., Hovmøller, M. S., de Vallavieille-Pope, C. (2000). Breakdown of the Yr17 resistance to yellow rust of wheat in northern Europe. *Agronomie*, *20*(7), 805–811. https://doi.org/10/bzd5w3

Bernardo, R. (2002). *Breeding for Quantitative Traits in Plants* (Second Edition). Woodburry (MN): Stemma Press.

BMNT. (2019). Getreideanbau und Getreidearten in Österreich. (Bundesministerium für Nachhaltigkeit und Tourismus und Getreidearten in Österreich, Ed.). Retrieved from https://www.bmnt.gv.at/land/produktion-maerkte/pflanzliche-produktion/getreide/Getreide.html

Bremenkamp-Barrett, B., Faris, J. D., Fellers, J. P. (2008). Molecular Mapping of the Leaf Rust Resistance Gene in Wheat. *Crop Science*, *48*(3), 1124. https://doi.org/10/bxqwjm

Brieuc, M. S. O., Waters, C. D., Drinan, D. P., Naish, K. A. (2018). A practical introduction to Random Forest for genetic association studies in ecology and evolution. *Molecular Ecology Resources*, *18*(4), 755–766. https://doi.org/10/gf4d72

Callaway, E. (2018). CRISPR plants now subject to tough GM laws in European Union. *Nature*, *560*(7716), 16–16. https://doi.org/10/gdvnj7

Chen, W., Wellings, C., Chen, X., Kang, Z., Liu, T. (2014). Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*: *Puccinia striiformis* , yellow rust. *Molecular Plant Pathology*, *15*(5), 433–446. https://doi.org/10.1111/mpp.12116

Chen, X. (2013). Review Article: High-Temperature Adult-Plant Resistance, Key for Sustainable Control of Stripe Rust. *American Journal of Plant Sciences*, *04*(03), 608–627. https://doi.org/10/ggkw8m

Chen, X., Kang, Z. (2017). Introduction: History of Research, Symptoms, Taxonomy of the Pathogen, Host Range, Distribution, and Impact of Stripe Rust. In X. Chen & Z. Kang (Eds.), *Stripe Rust* (pp. 1–33). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-024-1111-9_1

Clark, S. A., van der Werf, J. (2013). Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. In C. Gondro, J. van der Werf, & B. Hayes (Eds.), *Genome-Wide Association Studies and Genomic Prediction* (Vol. 1019, pp. 321–330). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-447-0_13

Covarrubias-Pazaran, G. (2016). Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *PLOS ONE*, *11*(6), e0156744. https://doi.org/10/ggjp6v

Covarrubias-Pazaran, G. (2018). Software update: Moving the R package *sommer* to multivariate mixed models for genome-assisted prediction. https://doi.org/10.1101/354639

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., … Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10/gch6f7

Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., Hayes, B. J. (2014). Genomic prediction for rust resistance in diverse wheat landraces. *Theoretical and Applied Genetics*, *127*(8), 1795–1803. https://doi.org/10/ggkw79

Doublet, A.-C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., … Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetics Selection Evolution*, *51*(1), 52. https://doi.org/10/ggkw8n

Dreisigacker, S. (2019). Genome-wide association studies (GWAS) in wheat. In Julius Kühn Institute, Germany & W. Freidt, F. Ordon (Ed.), *Burleigh Dodds Series in Agricultural Science* (pp. 477–502). Burleigh Dodds Science Publishing. https://doi.org/10.19103/AS.2019.0051.23

Duan, X., Tellier, A., Wan, A., Leconte, M., de Vallavieille-Pope, C., Enjalbert, J. (2010). *Puccinia striiformis* f.sp. *tritici* presents high diversity and recombination in the over-summering zone of Gansu, China. *Mycologia*, *102*(1), 44–53. https://doi.org/10/dvmt5r

Dyck, P. L., Lukow, O. M. (1988). THE GENETIC ANALYSIS OF TWO INTERSPECIFIC SOURCES OF LEAF RUST RESISTANCE AND THEIR EFFECT ON THE QUALITY OF COMMON WHEAT. *Canadian Journal of Plant Science*, *68*(3), 633–639. https://doi.org/10/bkxzvj

Ellis, J. G., Lagudah, E. S., Spielmeyer, W., Dodds, P. N. (2014). The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science*, *5*. https://doi.org/10/ggkw8p

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, *6*(5), e19379. https://doi.org/10/d9782d

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal*, *4*(3), 250. https://doi.org/10/cpsdzs

Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., Jannink, J.-L. (2014). Optimal Design of Preliminary Yield Trials with Genome-Wide Markers. *Crop Science*, *54*(1), 48–59. https://doi.org/10/f558wb

González-Camacho, J. M., Ornella, L., Pérez-Rodríguez, P., Gianola, D., Dreisigacker, S., Crossa, J. (2018). Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance. *The Plant Genome*, *11*(2), 170104. https://doi.org/10.3835/plantgenome2017.11.0104

GRRC. (2020). Wheat Rust Toolbox. (Global Rust Reference Center, Ed.). Retrieved from http://agro.au.dk/forskning/internationale-platforme/wheatrust/yellow-rust-tools-maps-and-charts/

Habier, D., Fernando, R. L., Dekkers, J. C. M. (2007). The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics*, *177*(4), 2389–2397. https://doi.org/10/d75fc6

Hanson, H., Borlaug, N. E., Anderson, R. G. (1982). *Wheat in the Third World*. Boulder, Colo: Westview Press.

Hartigan, J. A., Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1), 100. https://doi.org/10/fc6g7t

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, *5*. https://doi.org/10/gfznw9

Heffner, E. L., Lorenz, A. J., Jannink, J.-L., Sorrells, M. E. (2010). Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science*, *50*(5), 1681–1690. https://doi.org/10/bsx26v

Helguera, M., Khan, I. A., Kolmer, J., Lijavetzky, D., Zhong-qi, L., Dubcovsky, J. (2003). PCR Assays for the *Lr37-Yr17-Sr38* Cluster of Rust Resistance Genes and Their Use to Develop Isogenic Hard Red Spring Wheat Lines. *Crop Science*, *43*(5), 1839–1847. https://doi.org/10/bnhqgn

Hennig, C. (2019). fpc: Flexible Procedures for Clustering. R package version 2.2-2. Retrieved from https://CRAN.R-project.org/package=fpc

Hou, L., Jia, J., Zhang, X., Li, X., Yang, Z., Ma, J., … Chang, Z. (2016). Molecular Mapping of the Stripe Rust Resistance Gene *Yr69* on Wheat Chromosome 2AS. *Plant Disease*, *100*(8), 1717–1724. https://doi.org/10/f8w3mr

Hovmøller, M. S., Sørensen, C. K., Walter, S., Justesen, A. F. (2011). Diversity of *Puccinia striiformis* on Cereals and Grasses. *Annual Review of Phytopathology*, *49*(1), 197–217. https://doi.org/10/fmm62w

Hovmøller, M. S., Walter, S., Bayles, R. A., Hubbard, A., Flath, K., Sommerfeldt, N., … de Vallavieille-Pope, C. (2016). Replacement of the European wheat yellow rust population by new races from the centre of diversity in the near-Himalayan region. *Plant Pathology*, *65*(3), 402–411. https://doi.org/10/ggkw8j

IWGSC, Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., … Quesneville, H. (2018). Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*, *19*(1), 111. https://doi.org/10/gf23xv

Jin, Y., Szabo, L. J., Carson, M. (2010). Century-Old Mystery of *Puccinia striiformis* Life History Solved with the Identification of *Berberis* as an Alternate Host. *Phytopathology*, *100*(5), 432–435. https://doi.org/10/fnm3pv

Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., … Singh, R. P. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nature Genetics*, *51*(10), 1530–1539. https://doi.org/10/ggkw8b

Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Huerta-Espino, J., Lan, C., … Sorrells, M. E. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theoretical and Applied Genetics*, *130*(7), 1415–1430. https://doi.org/10/gbj276

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., Eskin, E. (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics*, *178*(3), 1709–1723. https://doi.org/10/fgtkcz

Kang, Z., Tang, C., Zhao, J., Cheng, Y., Liu, J., Guo, J., … Chen, X. (2017). Wheat-Puccinia striiformis Interactions. In X. Chen & Z. Kang (Eds.), *Stripe Rust* (pp. 155–282). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-024-1111-9_3

Kaufmann, L., Rousseeuw, P. (1987). Clustering by Means of Medoids. *Data Analysis Based on the L1-Norm and Related Methods. 405-416*.

Korjus, K., Hebart, M. N., Vicente, R. (2016). An Efficient Data Partitioning to Improve Classification Performance While Keeping Parameters Interpretable. *PLOS ONE*, *11*(8), e0161788. https://doi.org/10/f9q94k

Ledesma-Ramírez, L., Solís-Moya, E., Iturriaga, G., Sehgal, D., Reyes-Valdes, M. H., Montero-Tavera, V., … Singh, S. (2019). GWAS to Identify Genetic Loci for Resistance to Yellow Rust in Wheat Pre-Breeding Lines Derived From Diverse Exotic Crosses. *Frontiers in Plant Science*, *10*, 1390. https://doi.org/10/ggkw8c

Lenth, R. (2019). emmeans. Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.5.1. Retrieved from https://CRAN.R-project.org/package=emmans

Liu, Z., Goddard, M. E., Hayes, B. J., Reinhardt, F., Reents, R. (2016). Technical note: Equivalent genomic models with a residual polygenic effect. *Journal of Dairy Science*, *99*(3), 2016–2025. https://doi.org/10/f8bhhr

Lucht, J. (2015). Public Acceptance of Plant Biotechnology and GM Crops. *Viruses*, *7*(8), 4254–4281. https://doi.org/10/ggnnhq

Luig, N. H. (1985). Epidemiology in Australia and New Zealand. In *Diseases, Distribution, Epidemiology, and Control* (pp. 301–328). Elsevier. https://doi.org/10.1016/B978-0-12-148402-6.50018-3

Mathew, B., Sillanpää, M. J., Léon, J. (2019). Advances in statistical methods to handle large data sets for genome-wide association mapping in crop breeding. In Julius Kühn Institute, Germany, W. Freidt, & Justus Liebig University Giessen, Germany, F. Ordon (Ed.), *Burleigh Dodds Series in Agricultural Science* (pp. 437–450). Burleigh Dodds Science Publishing. https://doi.org/10.19103/AS.2019.0051.20

McIntosh, R. A., Dubcovsky, J., Appels, R., Xia, X. C. (2017). Catalogue of Gene Symbols for Wheat - Current Edition at KOMUGI, Originally published in 2013 plus supplements through 2017. In: KOMUGI – Wheat Genetic Resource Database. Retrieved from https://shigen.nig.ac.jp/wheat/komugi/genes

Mendel, G. (1860). Versuche über Pflanzen-Hybriden. *Verhandlungen Des Naturforschenden Vereins Zu Brünn*, *4*, 3–47.

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps, 11.

Miedaner, T. (2017). *Pflanzenkrankheiten, die die Welt beweg(t)en*. Berlin: Springer.

Miedaner, T., Rapp, M., Flath, K., Longin, C. F. H., Würschum, T. (2019). Genetic architecture of yellow and stem rust resistance in a durum wheat diversity panel. *Euphytica*, *215*(4), 71. https://doi.org/10/ggkw76

Mieth, B., Kloft, M., Rodríguez, J. A., Sonnenburg, S., Vobruba, R., Morcillo-Suárez, C., … Müller, K.-R. (2016). Combining Multiple Hypothesis Testing with Machine Learning Increases the Statistical Power of Genome-wide Association Studies. *Scientific Reports*, *6*(1), 36671. https://doi.org/10/f9c9xn

Möhring, J., Piepho, H.-P. (2009). Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. *Crop Science*, *49*(6), 1977–1988. https://doi.org/10/fvmbkr

Muleta, K. T., Bulli, P., Rynearson, S., Chen, X., Pumphrey, M. (2017). Loci associated with resistance to stripe rust (Puccinia striiformis f. sp. tritici) in a core collection of spring wheat (Triticum aestivum). *PLOS ONE*, *12*(6), e0179087. https://doi.org/10/gbgwjn

Muleta, K. T., Bulli, P., Zhang, Z., Chen, X., Pumphrey, M. (2017). Unlocking Diversity in Germplasm Collections via Genomic Selection: A Case Study Based on Quantitative Adult Plant Resistance to Stripe Rust in Spring Wheat. *The Plant Genome*, *10*(3), plantgenome2016.12.0124. https://doi.org/10/ggkw8q

Oberforster, M. (2015). 2015 droht erneut eine Gelbrostepidemie! *Der Pflanzenarzt*, *4*(68.Jg 2015), 4–7.

Oberforster, M. (2018). Im Kampf gegen Rostpilze bei Weizen und Gerste. *Der Pflanzenarzt*, *9-10/2018*(71.Jg 2018), 13–15.

Ornella, L., Singh, S., Perez, P., Burgueño, J., Singh, R., Tapia, E., … Crossa, J. (2012). Genomic Prediction of Genetic Values for Resistance to Wheat Rusts. *The Plant Genome*, *5*(3), 136–148. https://doi.org/10/ggkw78

Ould Estaghvirou, S., Ogutu, J. O., Schulz-Streeck, T., Knaak, C., Ouzunova, M., Gordillo, A., Piepho, H.-P. (2013). Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. *BMC Genomics*, *14*(1), 860. https://doi.org/10/gb3f2j

Poland, J., Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*, *5*(3), 92–102. https://doi.org/10/gfzn68

Poland, J., Rutkoski, J. (2016). Advances and Challenges in Genomic Selection for Disease Resistance. *Annual Review of Phytopathology*, *54*(1), 79–98. https://doi.org/10/ggkw7x

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., … Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10/gc4h8p

Röder, M. S., Huang, X.-Q., Ganal, M. W. (2005). Wheat Microsatellites: Potential and Implications. In H. Lörz & G. Wenzel (Eds.), *Molecular Marker Systems in Plant Breeding and Crop Improvement* (Vol. 55, pp. 255–266). Berlin/Heidelberg: Springer-Verlag. https://doi.org/10.1007/3-540-26538-4_15

Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*, *23*, 52–71. https://doi.org/10/ggkw72

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., … Sorrells, M. E. (2014). Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *The Plant Genome*, *7*(3), plantgenome2014.02.0006. https://doi.org/10.3835/plantgenome2014.02.0006

Schwessinger, B. (2017). Fundamental wheat stripe rust research in the 21 st century. *New Phytologist*, *213*(4), 1625–1631. https://doi.org/10/f9xbzq

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, *5*(3), 291–317. https://doi.org/10/ggkw73

Somers, D. J., Isaac, P., Edwards, K. (2004). A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, *109*(6), 1105–1114. https://doi.org/10/ftv7mq

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, *3(25)*, *731*, 2. https://doi.org/10/ggkw8k

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10/c85k62

Wan, A., Muleta, K. T., Zegeye, H., Hundie, B., Pumphrey, M. O., Chen, X. (2017). Virulence Characterization of Wheat Stripe Rust Fungus *Puccinia striiformis* f. sp. *tritici* in Ethiopia and Evaluation of Ethiopian Wheat Germplasm for Resistance to Races of the Pathogen from Ethiopia and the United States. *Plant Disease*, *101*(1), 73–80. https://doi.org/10/f9j6b5

Wan, A., Wang, X., Kang, Z., Chen, X. (2017). Variability of the Stripe Rust Pathogen. In X. Chen & Z. Kang (Eds.), *Stripe Rust* (pp. 35–154). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-024-1111-9_2

Wang, M., Chen, X. (2017). Stripe Rust Resistance. In X. Chen & Z. Kang (Eds.), *Stripe Rust* (pp. 353–558). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-024-1111-9_5

Wang, M. N., Chen, X. M. (2013). First Report of Oregon Grape (*Mahonia aquifolium*) as an Alternate Host for the Wheat Stripe Rust Pathogen (*Puccinia striiformis* f. sp. *tritici*) Under Artificial Inoculation. *Plant Disease*, *97*(6), 839–839. https://doi.org/10/ggkw74

Wang, X., Xu, Y., Hu, Z., Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, *6*(4), 330–340. https://doi.org/10/gdz9sd

Wei, T., Viliam, S. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Retrieved from https://github.com/taiyun/corrplot

Würschum, T. (2012). Mapping QTL for agronomic traits in breeding populations. *Theoretical and Applied Genetics*, *125*(2), 201–210. https://doi.org/10/ggkw8g

Xavier, A., Xu, S., Muir, W. M., Rainey, K. M. (2015). NAM: association studies in multiple populations. *Bioinformatics*, btv448. https://doi.org/10/ggkw8h

Xu, L. S., Wang, M. N., Cheng, P., Kang, Z. S., Hulbert, S. H., Chen, X. M. (2013). Molecular mapping of Yr53, a new gene for stripe rust resistance in durum wheat accession PI 480148 and its transfer to common wheat. *Theoretical and Applied Genetics*, *126*(2), 523–533. https://doi.org/10/f4j8tc

Xue, S., Kolmer, J. A., Wang, S., Yan, L. (2018). Mapping of Leaf Rust Resistance Genes and Molecular Characterization of the 2NS/2AS Translocation in the Wheat Cultivar Jagger. *G3: Genes|Genomes|Genetics*, *8*(6), 2059–2065. https://doi.org/10/ggkw8f

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., … Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10/d8npfr

Zadoks, J. C., Bouwman, J. J. (2017). Stripe Rust Epidemiology. In X. Chen & Z. Kang (Eds.), *Stripe Rust* (pp. 283–352). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-024-1111-9_4

Zhang, P., McIntosh, R. A., Hoxha, S., Dong, C. (2009). Wheat stripe rust resistance genes Yr5 and Yr7 are allelic. *Theoretical and Applied Genetics*, *120*(1), 25–29. https://doi.org/10/bd8j62

Zhao, Y., Mette, M., Gowda, M., Longin, C. F. H., Reif, J. C. (2014). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity*, *112*(6), 638–645. https://doi.org/10/f536kj

Zheng, W., Huang, L., Huang, J., Wang, X., Chen, X., Zhao, J., … Kang, Z. (2013). High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications*, *4*(1), 2673. https://doi.org/10/gbd4zw

7 Appendix

7.1 Results from marker data

7.1.1 Allele frequencies

Histogram of allele frequencies found in the genotyped population, markers with allele frequencies outside the area between the red lines were removed.

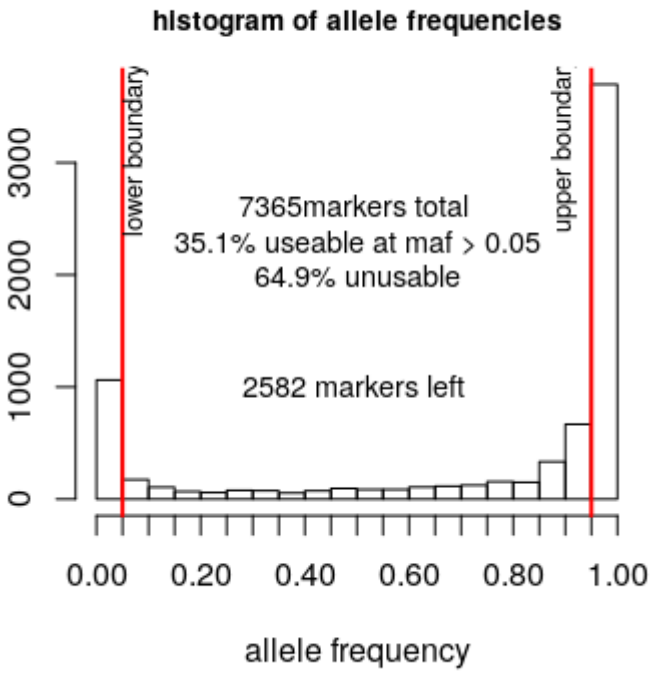


Figure 14: Allele frequency distribution in the population

7.1.2 Genetic Map

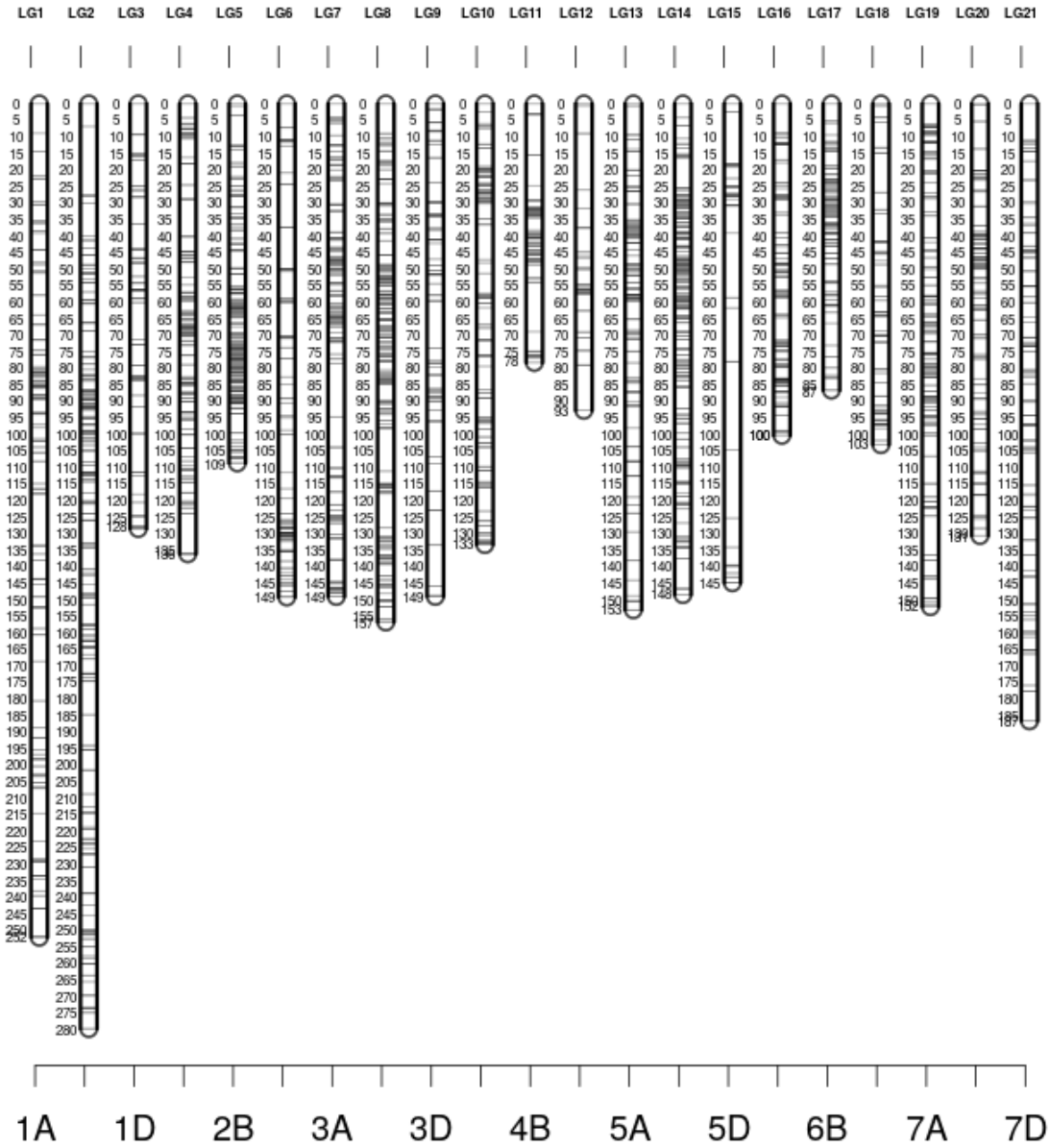


Figure 15: A representation of the genetic map
filtered for 5% MAF, showing the density of markers.

7.1.3 Heterozygosity of lines

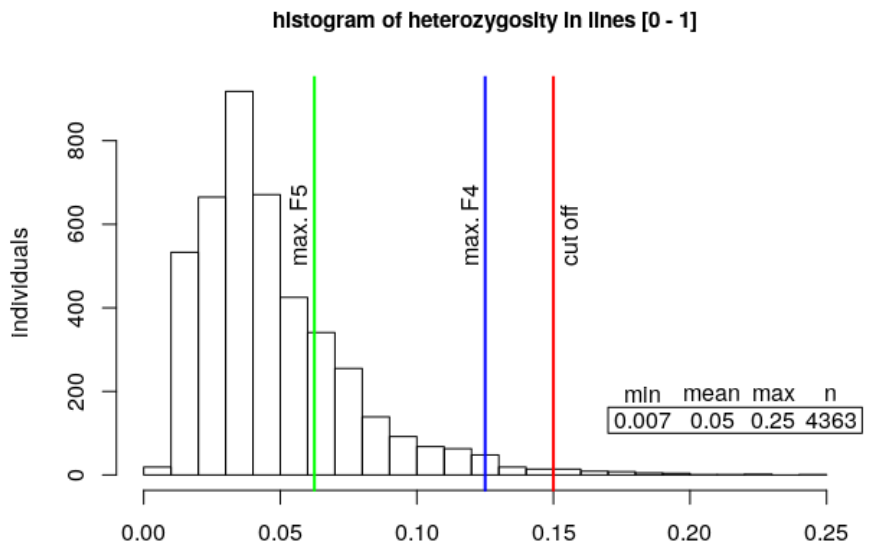


Figure 16: Heterozygosity distribution histogram among lines

7.2 Trial information

7.2.1 Trials removed

Table 10: Trials removed due to low number of plots (<55)

Year and location	mean score	var. score	max. score	plots	lines	reps	rows	cols	nscore
2013 DOE 103	1	0	1	70	64	1.1	70	1	1
2013 TR-BAB 97	1	0	1	25	25	1	25	1	1
2013 TR-KON 99	4	1	5	3	2	1.5	3	2	1
2013 TR-PIN 100	1	0	1	25	25	1	25	1	1
2014 RBG 121	1.2	0.21	5	25	25	1	25	1	2
2016 TR-ALT 43	1.2	0.37	4	30	22	1.4	30	2	1

Table 11: Trials removed due to low variance (<0.21)

Year and location	mean score	var. score	max. score	plots	lines	reps	rows	cols	nscore
2013 TR-BAB 91	1	0	1	200	174	1.1	50	4	1
2013 TR-KAR 111	1	0	1	100	25	4	25	4	1
2016 AT-AUM_Jetzing 11	1.17	0.16	7	585	315	1.9	105	8	2
2016 AT-AUM_Pöding 10	1.48	0.18	5	90	44	2	45	2	2

2016 AT-PRO 4	1.17	0.2	4	528	436	1.2	132	4	1
2016 RO-L-n 31	1.04	0.02	2	268	223	1.2	67	4	2
2016 RO-LIV 30	1.12	0.05	2	268	223	1.2	67	4	2

Table 12: Trials removed due to problems in mixed modeling

Year and location	mean score	var. score	max. score	plots	lines	reps	rows	cols	nscore
2014 TR-Bab 141	2.37	2.58	7	200	182	1.1	50	4	1
2014 TR-Bab 142	2.99	5.53	9	100	25	4	25	4	1
2014 TR-Pin 143	1.29	0.89	6	200	182	1.1	50	4	1
2014 TR-Pin 144	1.72	2.04	8	100	25	4	25	4	1
2015 DE-BIE 16	1.25	0.78	7	200	185	1.1	25	8	2
2015 DE-BIE-n 17	1.33	0.9	6	199	184	1.1	25	8	2
2015 MAR 12	1.25	0.25	4	200	185	1.1	100	2	1
2015 MAR-n 13	1.32	0.27	3	200	185	1.1	100	2	1
2016 FR-AUC 22	3.21	3.15	8	230	222	1	65	4	1
2014 HU-MV2 147	3.45	3.45	8	99	33	3	35	3	1

Table 13: Trials removed due to low heritability results from mixed modeling

Year and location	mean score	var. score	max. score	plots	lines	reps	rows	cols	nscore
2013 TR-KAI 98	1.07	0.49	8	100	25	4	25	4	1
2013 TR-KON 92	1.32	0.74	5	200	174	1.1	50	4	1
2014 RO-DRA 131	1.92	2.31	6	60	29	2.1	30	2	1
2014 RS-SOM 135	3.38	4.54	8	260	209	1.2	50	6	1
2015 LEO 5	1.14	0.62	8	610	276	2.2	102	6	1
2015 RS-SOM 29	1.51	0.65	5	300	239	1.3	65	6	1
2016 AT-PRO 1	1.32	0.38	8	456	406	1.1	76	6	1
2016 CA-RID 17	1.77	0.39	4	260	222	1.2	13	20	1

7.2.2 Trials used

Table 14: Overview trials usable for mixed modeling

Year and location	mean score	var. score	max. score	plots	lines	reps	rows	cols	nscore
2013 RBG-AU NURS	2.84	0.99	8	3122	1439	2.2	240	14	2
2013 RBG-AU NUR2	3.79	2.45	8	2823	1360	2.1	240	14	1
2014 AUM_Ob 120	1.93	1.27	9	865	623	1.4	173	5	3
2014 AUM_Sch 119	1.68	1.62	8	540	260	2.1	90	6	2
2014 DOE 118	2.13	2.39	8	618	346	1.8	125	5	1
2014 HU-MV1 146	3.26	2.54	7	99	33	3	35	3	1
2014 LEO 116	1.73	1.46	9	1040	774	1.3	175	6	2
2014 LEO 148	1.77	1	7	200	182	1.1	50	4	2
2014 LEO 149	1.51	0.82	7	60	29	2.1	30	2	2
2014 MLK 124	1.51	0.75	7	220	146	1.5	40	8	2
2014 PRkf 112	2.95	3.42	9	1224	1086	1.1	102	12	3
2014 PRks 113	2.32	2.36	9	1284	1146	1.1	107	12	3
2014 PRO 114	3.68	3.56	9	260	209	1.2	65	4	3
2014 PRO 115	2.75	2.54	9	838	193	4.3	210	4	3
2014 RBG-AU NURS	1.21	0.25	8	2312	1385	1.7	200	13	2
2014 RO-FUN 134	2.08	3.24	8	200	182	1.1	25	8	2
2014 RO-LIV 132	2.25	3.02	8	60	29	2.1	20	3	2
2014 RO-MOD 133	1.37	1.5	8	146	63	2.3	45	4	1
2014 SK-HAN 129	3.33	7.05	9	60	29	2.1	30	2	2
2014 WEI 117	2.39	2.76	9	838	467	1.8	140	6	2.1
2015 AUM-Hardfeld 9	2.57	2.4	9	410	173	2.4	82	5	3
2015 AUM-Schindlmühl 8	2.02	1.79	9	868	339	2.6	177	5	2.4
2015 DOE 7	2.05	3.68	8	402	134	3	67	6	1
2015 PR-Ge 3	1.91	1.33	7	468	404	1.2	117	4	2
2015 PR-Ge 53	1.65	0.82	8	468	404	1.2	117	4	2
2015 PRkf 1	3.05	3.86	9	1248	1059	1.2	104	12	2
2015 PRks 2	2.17	1.9	9	1248	977	1.3	104	12	2
2015 PRO 4	3.23	1.82	9	1668	680	2.5	139	12	2
2015 RBG-AU NURS	1.42	0.79	7	2470	1528	1.6	90	30	1
2015 RS-S-n 29	1.69	0.78	5	260	220	1.2	65	4	1
2015 SK-HAN 31	2.42	3.37	8	60	29	2.1	15	4	1
2016 AT-PRO 2	3.99	3.93	8	122	78	1.6	42	4	1
2016 AT-PRO 3	3.09	2.92	9	1870	816	2.3	172	16	1.6

2016 AT-PROkf 5	4.56	4.5	9	1248	1042	1.2	104	12	2
2016 AT-PROks 6	4.32	4.79	9	1248	1009	1.2	104	12	1
2016 RBG-AU NURS	2.03	0.75	8	3261	2238	1.5	80	44	2
2016 SK-HAN 35	1.6	0.41	4	60	29	2.1	15	4	2

7.2.3 trial heritabilities

Table 15: Trial heritabilities

trials above threshold (>0.55)				**below threshold (< 0.55)**	
Trial	h^2	Trial	h^2	Trial	h^2
2013 RBG-AU NUR2	0.72	2015 RS-S-n 29	0.68	2013 TR-KAI 98	0
2013 RBG-AU NURS	0.7	2015 PRks 2	0.74	2013 TR-KON 92	0.13
2014 LEO 148	0.67	2015 PR-Ge 53	0.75	2014 RO-DRA 131	0
2014 LEO 116	0.78	2015 SK-HAN 31	0.85	2014 RS-SOM 135	0.31
2014 DOE 118	0.83	2015 AUM-Schindlmühl 8	0.92	2014 RBG-AU NURS	0.49
2014 AUM_Ob 120	0.80	2015 AUM-Hardfeld 9	0.91	2015 RS-SOM 29	0.22
2014 AUM_Sch 119	0.91	2015 PRkf 1	0.84	2015 LEO 5	0.43
2014 RO-FUN 134	0.85	2015 PRO 4	0.93	2015 RBG-AU NURS	0.53
2014 MLK 124	0.90	2015 PR-Ge 3	0.87	2016 AT-PRO 1	0
2014 WEI 117	0.92	2015 DOE 7	0.97	2016 CA-RID 17	0.31
2014 PRO 115	0.97	2016 AT-PROkf 5	0.63		
2014 RO-LIV 132	0.96	2016 AT-PROks 6	0.65		
2014 SK-HAN 129	0.96	2016 RBG-AU NURS	0.71		
2014 PRO 114	0.96	2016 AT-PRO 3	0.84		
2014 PRks 113	0.95	2016 AT-PRO 2	0.97		
2014 LEO 149	0.99	2016 SK-HAN 35	0.99		
2014 PRkf 112	0.99				

7.3 Adjusted means

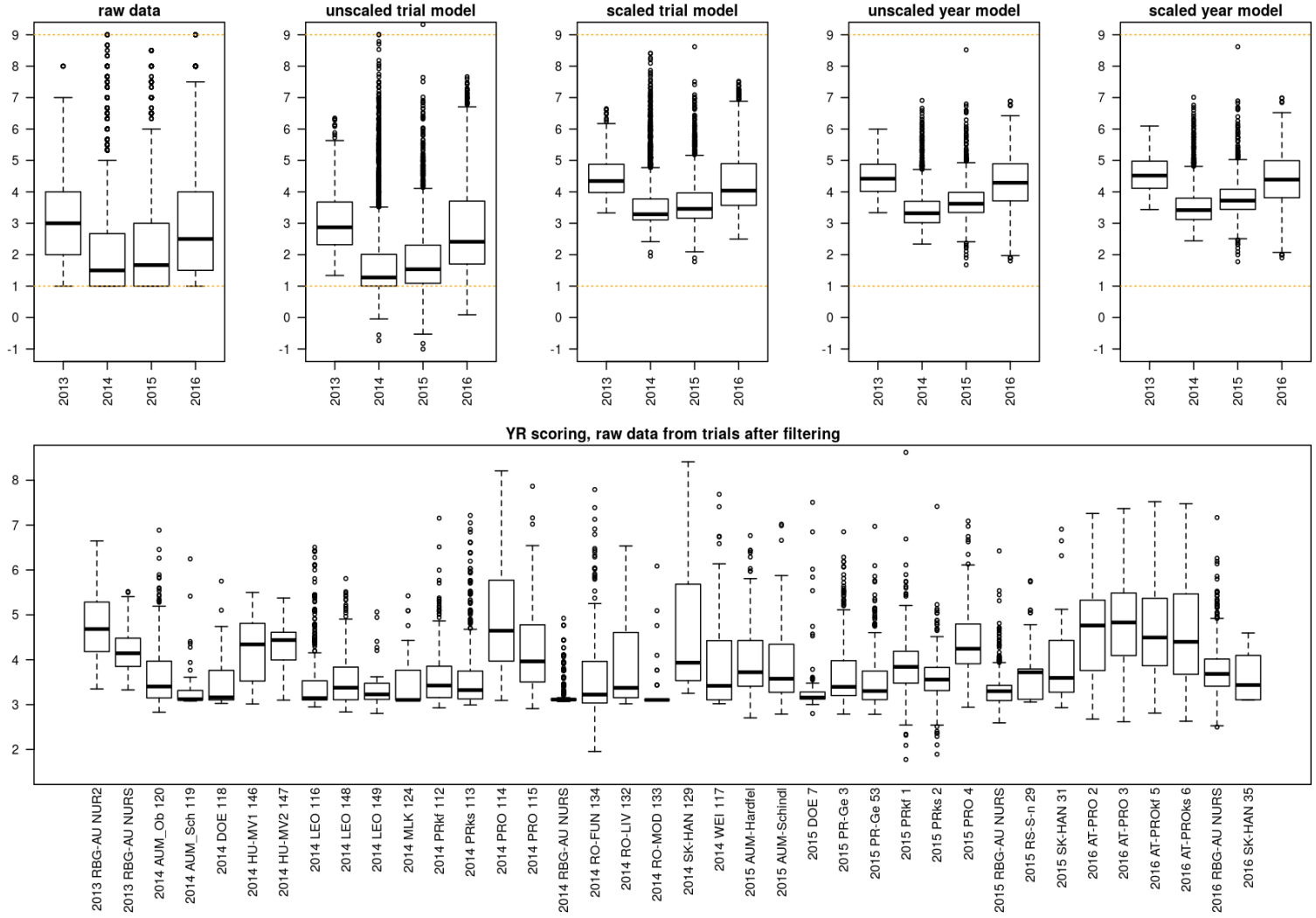


Figure 17: YR scorings and adjusted means at different stages of mixed modeling

7.4 Correlations across years

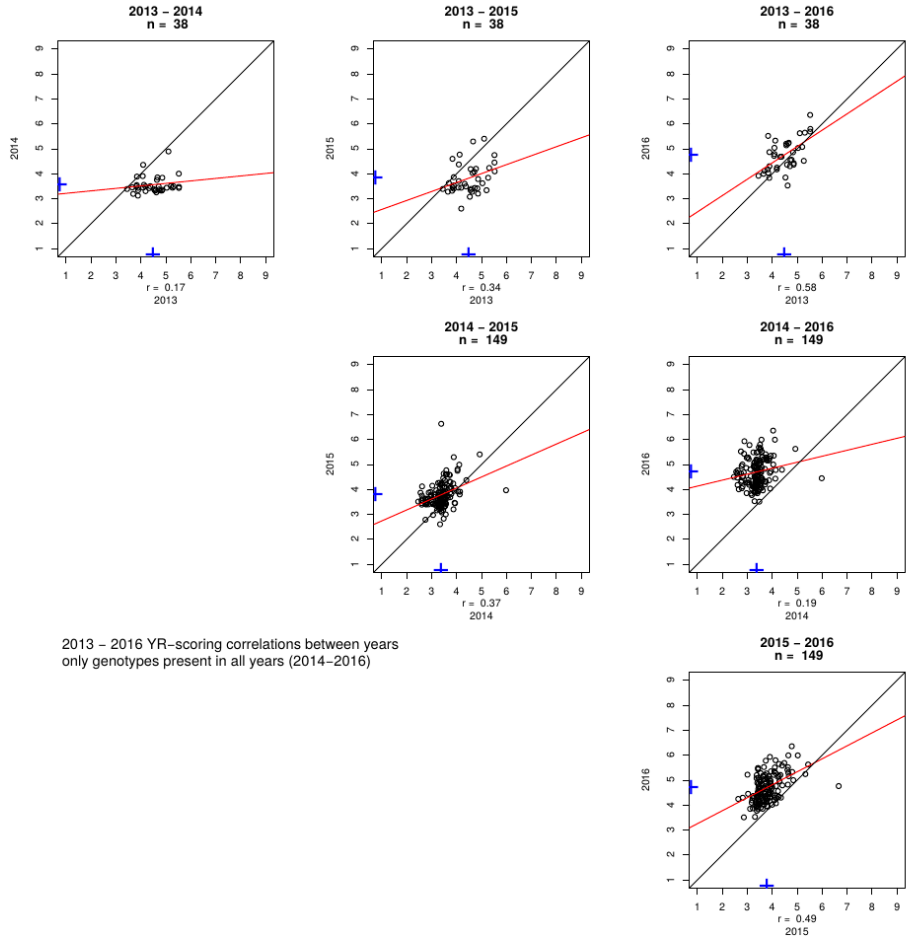


Figure 18: Correlations of genotypes present in all years of 2014, 2015 and 2016, not 2013

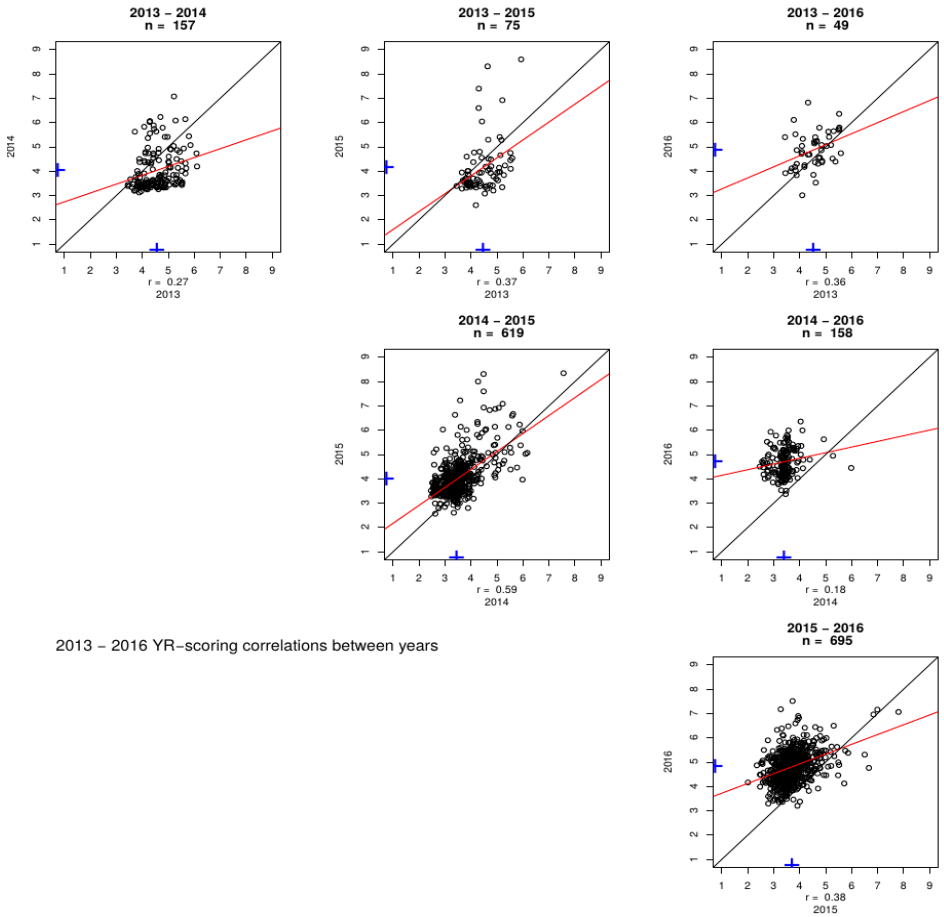


Figure 19: Pairwise correlations of genotypes per year

7.5 Clustering results

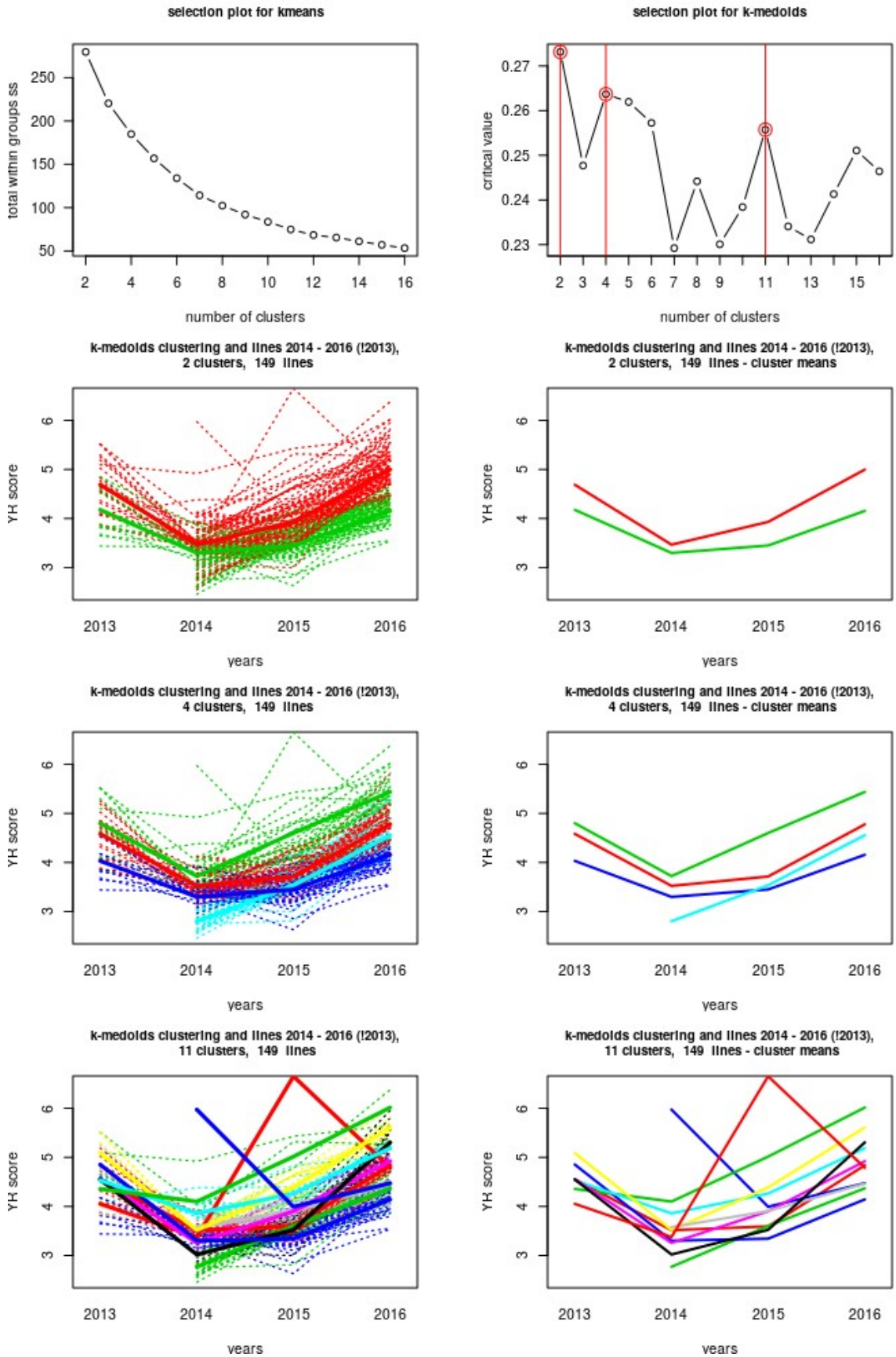


Figure 20: Detailed clustering results

7.6 GWAS results

7.6.1 Manhattan plots per year

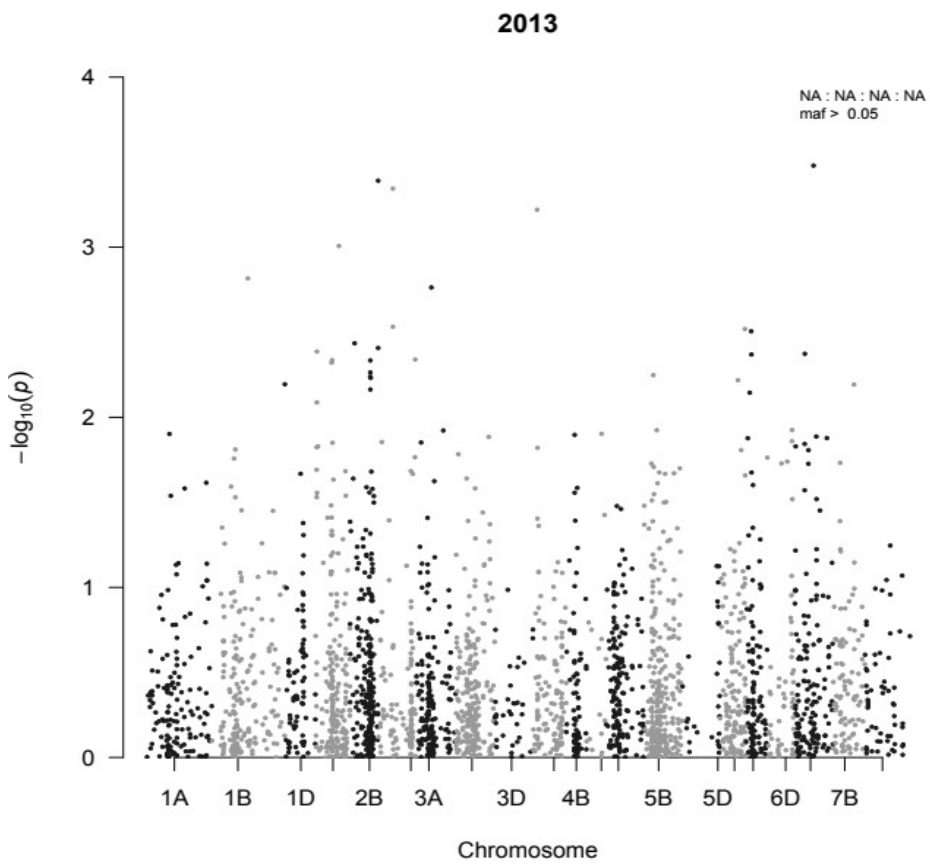


Figure 21: Manhattan plot of GWAS results 2013

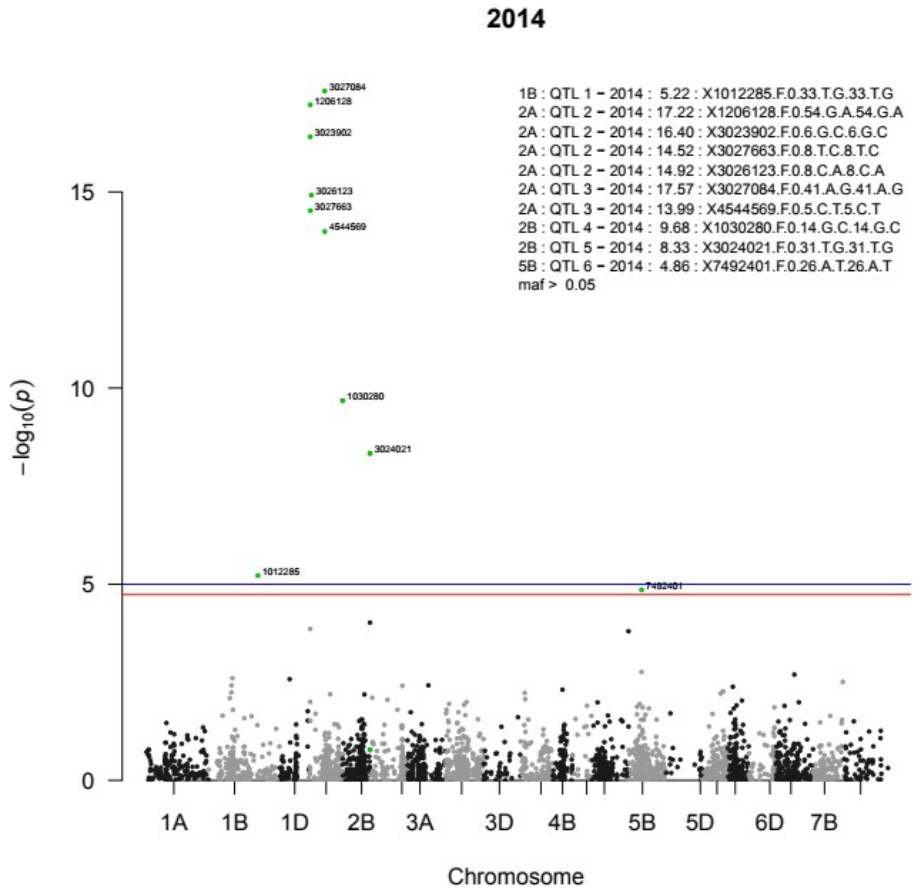


Figure 22: Manhattan plot of GWAS results 2014

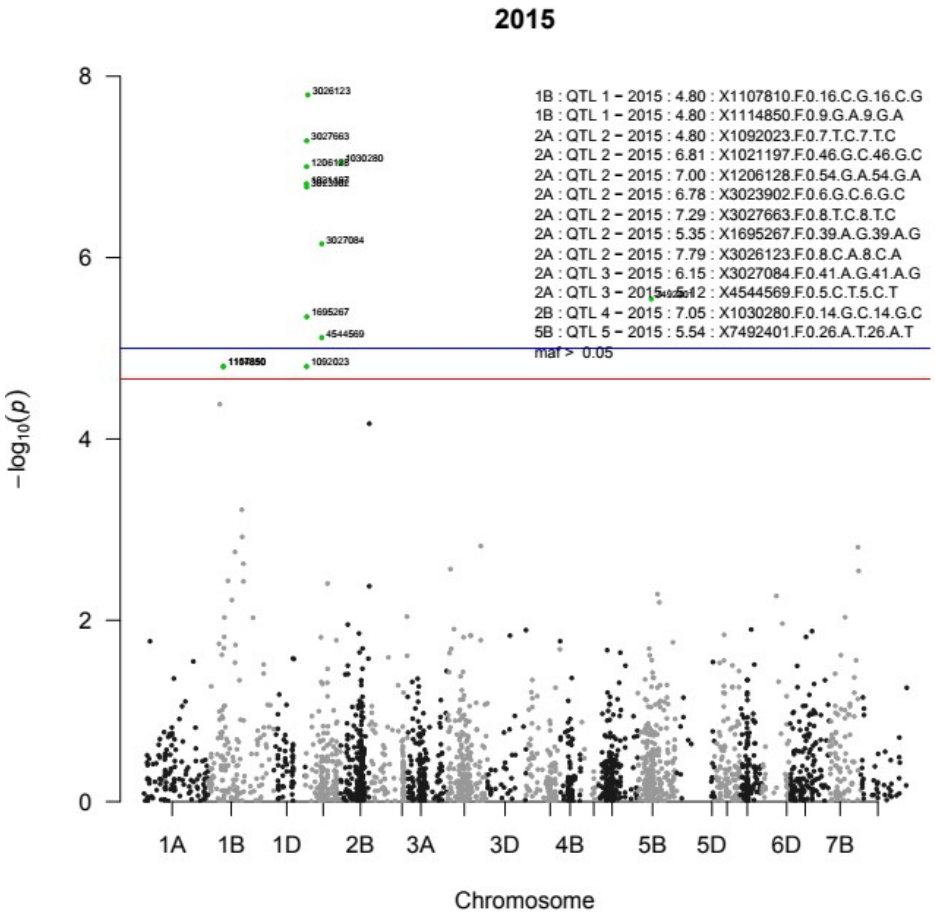
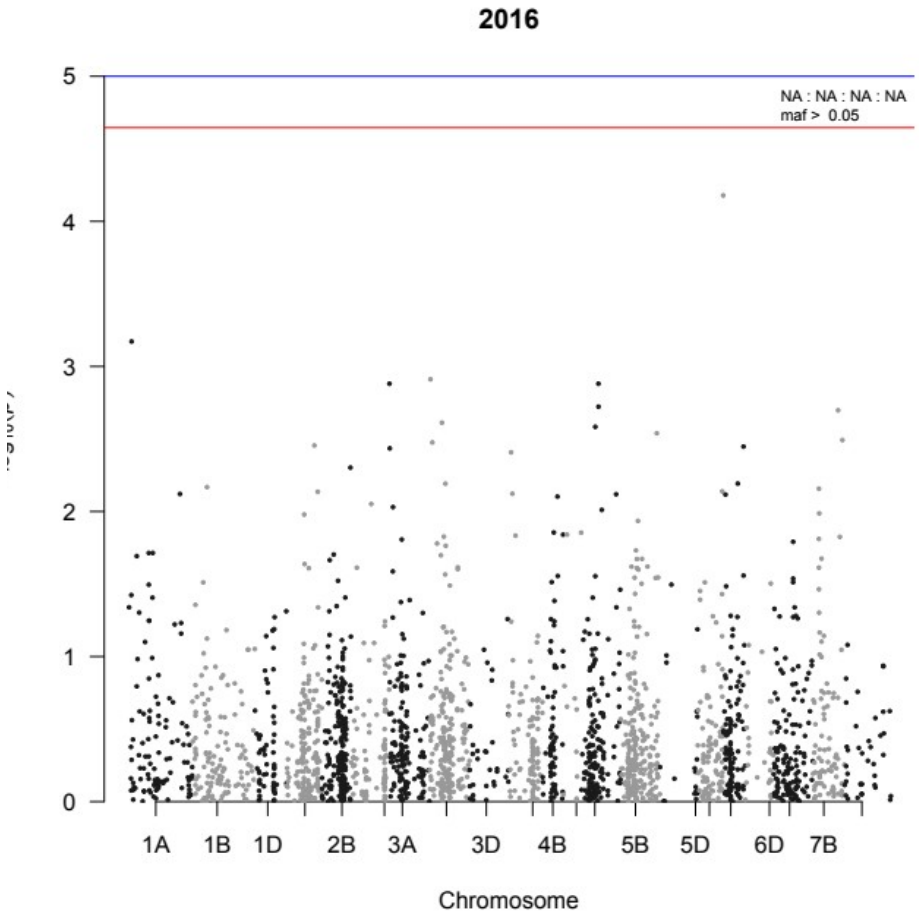


Figure 23: Manhattan plot of GWAS results 2015



igure 24: Manhattan plot of GWAS results 2016

7.6.2 QTL effects and additivity

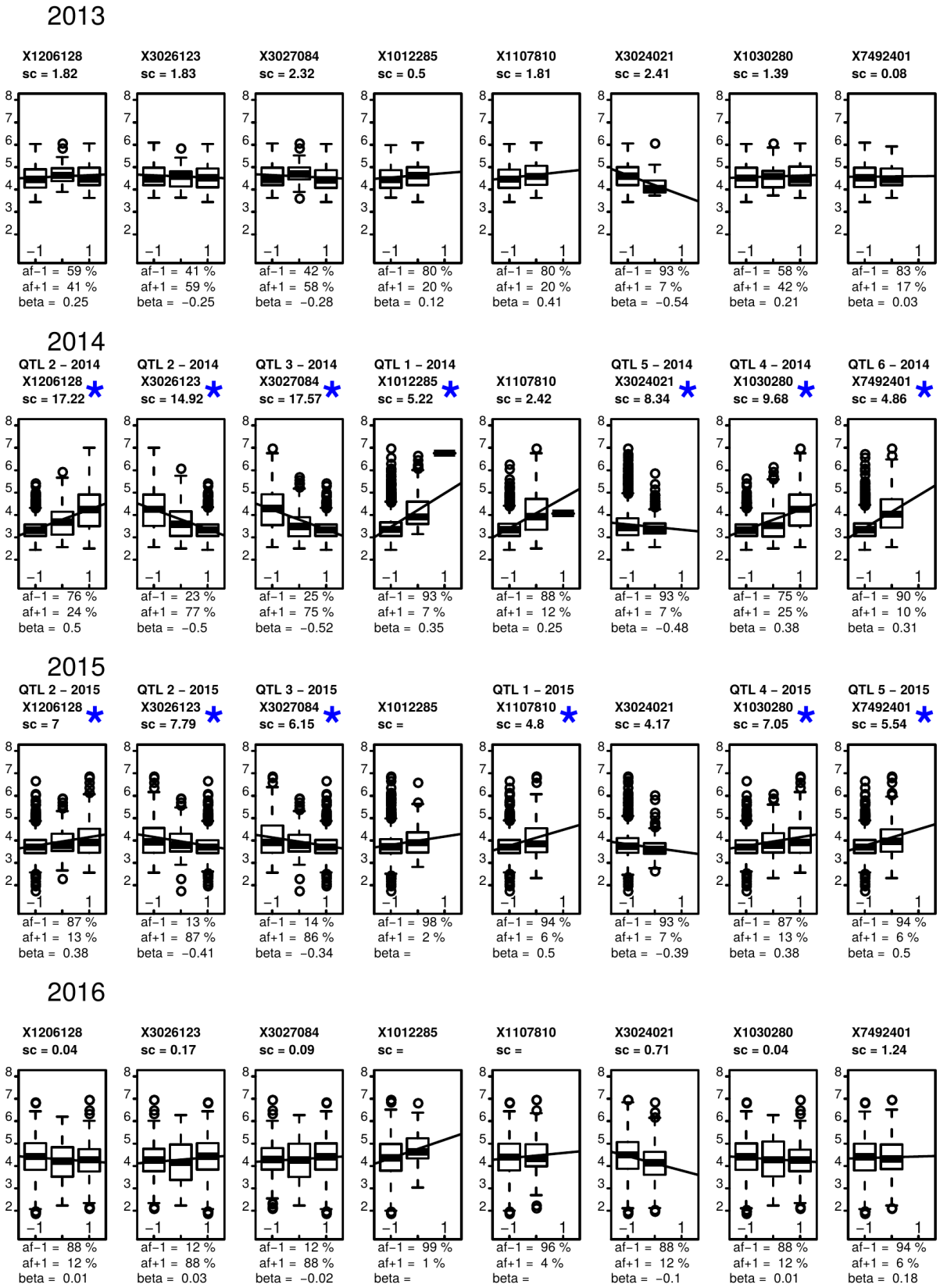


Figure 25: QTL beta effects vs. phenotypic values

7.6.3 GS in years detailed results

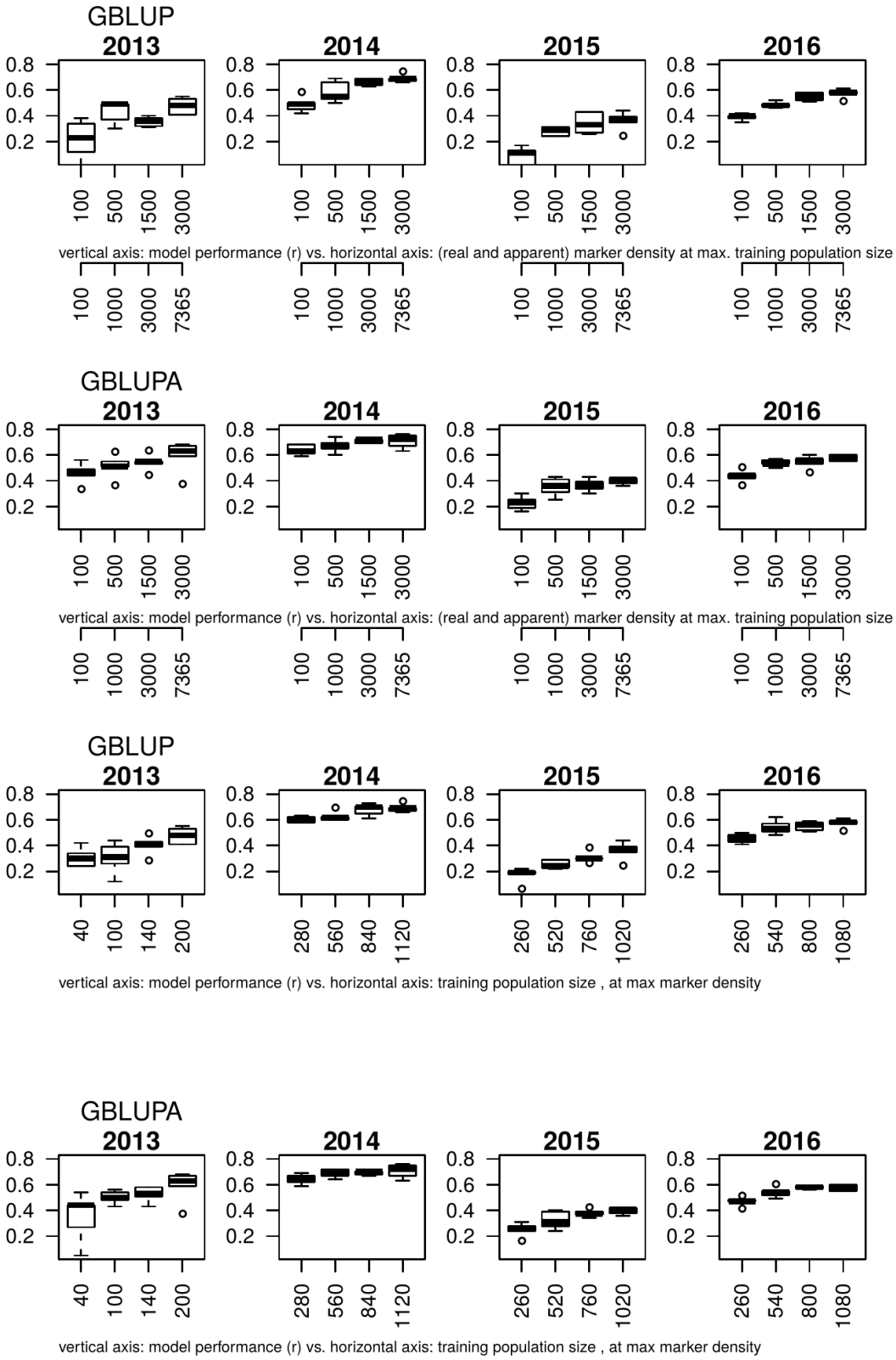


Figure 26: GS within years, variations in marker density, and training population size