Aus dem Institut für
Pflanzenzüchtung, Saatgutforschung und Populationsgenetik
der Universität Hohenheim
Fachgebiet Angewandte Genetik und Pflanzenzüchtung
Prof. Dr. A. E. Melchinger

QTL Mapping and Genomic Prediction of Complex Traits Based on High-density Genotyping in Multiple Crosses of Maize (*Zea mays* L.)

Dissertation
zur Erlangung des Grades eines Doktors
der Agrarwissenschaften
vorgelegt
der Fakultät Agrarwissenschaften

von
M. Sc. agr.
Michael Stange
aus Eschwege

Stuttgart–Hohenheim 2013

Die vorliegende Arbeit wurde am 15. Oktober 2013 von der Fakultät Agrarwissenschaften der Universität Hohenheim als „Dissertation zur Erlangung des Grades eines Doktors der Agrarwissenschaften“ angenommen.

Tag der mündlichen Prüfung: 21. November 2013

1. Prodekan:	Prof. Dr. M. Rodehutscord
Berichterstatter, 1. Prüfer:	Prof. Dr. A. E. Melchinger
Mitberichterstatter, 2. Prüfer:	PD Dr. T. Würschum
3. Prüfer:	Prof. Dr. R. Doluschitz

Contents

---

[1]Stange M, Utz HF, Schrag TA, Melchinger AE, Würschum T (2013) High-density Genotyping: an Overkill for QTL Mapping? Lessons learned from a Case Study in Maize and Simulations. Theor Appl Genet. 126:2563-2574.

[2]Stange M*, Schrag TA*, Utz HF, Riedelsheimer C, Bauer E, Melchinger AE (2013) High-density linkage mapping of yield components and epistatic interactions in maize with doubled haploid lines from four crosses. Mol Breeding. 32:533-546. * Both authors contributed equally.

[3]Riedelsheimer C, Endelman J, Stange M, Sorrells EM, Jannink JL, Melchinger AE (2013) Genomic Predictability of Interconnected Biparental Maize Populations. Genetics 194:493-503.

Abbreviations

cM	centi Morgan
CIM	composite interval mapping
DH	doubled haploid line
FS	full-sib
GER	gibberella ear rot
GS	genomic selection
GY	grain yield
HKW	100-kernel weight
HS	half-sib
IBD	identical-by-descent
ICIM	inclusive composite interval mapping
IM	interval mapping
KN	kernel number
LOD	logarithm of odds
MAS	marker-assisted selection
Mbp	mega base pairs
Pop.	population
QTL	quantitative trait loci
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
SI	support interval
TS	test set
UR	unrelated
VP	validation population

Chapter 1

General Introduction

Maize (*Zea mays* L.) is traditionally used for human and animal consumption and since recent decades for diverse industrial purposes, and for bioenergy. It is grown worldwide on 159 million ha (FAOSTAT 2009) and thus, one of the most important crops besides wheat (*Triticum aestivum* L.) and rice (*Oryza sativa* L.). For this reason, also maize breeding is highly important and the modern era of maize breeding began already about 100 years ago with Shull´s experiments for hybrid breeding (Shull 1909). From this time on, stunning breeding and research progress was made by implementation of several methodological and technological achievements as well as by the application of newly developed selection strategies. Out of all, the most important steps are the hybrid technology, the off-season nursery and the doubled haploid (DH) technology (Seitz 2004). The DH technology, which reduces dramatically the time necessary to obtain fully homozygous inbred lines (Prigge and Melchinger 2012), and which enables the generation of a huge number of inbred lines every year, is meanwhile worldwide routinely applied in maize breeding programs (Schmidt 2003; Seitz 2005; Chen et al. 2009). Typically, DH lines originate from distinct crosses between related or unrelated parents. In practical breeding programs, a parent is often crossed with several other parents in a connected design, which enables the evaluation of the influence of one parent in combination with several others, related or unrelated parents. On a chromosomal level, connected designs enable the evaluation of the contribution of similar or different linkage phases on chromosomal regions, which are contributed by the parents involved.

Detection of quantitative trait loci (QTL)

In natural populations, a remarkable diversity of phenotypic variation for morphology, physiology, and disease susceptibility is present due to a highly complex underlying genetic basis with multiple interacting loci (Mackay et al. 2009). Understanding the

relationship between DNA sequence variation and phenotypic variation for these complex traits can increase the power and speed for breeding of important agronomic traits.

This is difficult because only few traits of agronomic interest are controlled by a single gene (monogenic) or few genes (oligogenic) (Falconer and Mackay 1996; Lynch and Walsh 1998). In contrast, most important agronomic traits like resistance to Gibberella ear rot (GER) or grain yield, both of very high importance in maize breeding, are controlled by tens to thousand quantitative trait loci (QTL). What makes it even more difficult is the fact, that each QTL has only a small to medium effect on the trait of interest (Mackay et al. 2009).

The linkage between these QTL which affect the natural variation of the genetically and physiologically complex traits and polymorphic marker loci with Mendelian segregation has been known since the early twentieth century (Mackay et al. 2009). Since these days, researchers and breeders aim to identify and localize molecular markers linked to QTL which can finally be used in marker-assisted selection of superior lines (Dekkers and Hospital 2002).

Nowadays, QTL detection with genome-wide association mapping in diversity panels and with linkage mapping in biparental populations are well established in genetic studies in plants. Especially for rare alleles, classical linkage mapping approaches offer high QTL detection power due to the balanced allele frequencies in segregating populations (Würschum 2012). Therefore, one focus of this thesis was the development and application of linkage mapping approaches with high-density linkage maps to DH lines of biparental populations to dissect the genetic basis of the complex trait grain yield (Stange et al. 2013a).

Factors influencing QTL mapping

Several factors affecting important parameters of QTL detection were discussed in the literature on the basis of experimental data as well as on the basis of simulation studies (e.g., Darvasi et al. 1993; Utz et al. 2000; Doerge 2002; Li et al. 2010). Thereby, the QTL detection method and factors related to the experimental design like population type, population size, and number of markers which influence power, resolution,

precision of QTL localization, and bias of QTL effect estimates, were important factors, that were considered.

First QTL mapping experiments were performed with single-marker analysis (e.g., t-test or ANOVA) indicating which markers are linked with the quantitative trait of interest and thus, pointing to a putative QTL (Doerge 2002). Afterwards, more sophisticated and statistically more powerful approaches, like interval mapping (IM) and the more precise and effective composite interval mapping (CIM) were developed and applied for QTL mapping (Jansen 1993; Jansen and Stam 1994). CIM relies on the order of markers and uses a multiple regression approach (Haley and Knott 1992) for detection of QTL positions and estimation of their effects. It is extensively used in diverse QTL mapping studies due to implementation in QTL mapping software like PlabMQTL (Utz 2012).

From a practical point of view, the number of markers under limited population size is of special interest due to tremendously fast developments in genotyping technologies. Nowadays, the markers of choice are SNPs, polymorphic insertions or deletions, whereof a huge number was revealed by new developments in sequencing technologies. High-throughput platforms enable a routinely application of this new tool in sequencing (Yan et al. 2009). However, hitherto, most linkage mapping studies for diverse traits in maize were based on low-density linkage maps calculated with only few simple sequence repeat (SSR) markers (e.g., Ma et al. 2007; Guo et al. 2011; Peng et al. 2011; Martin et al. 2012). In these studies, QTL localization was rather imprecise due to the low marker density. Furthermore, the power of QTL detection was insufficient to detect QTL with small genetic effects or to separate closely linked QTL. In contrast, the 50k maize chip provides thousands of markers. However, until now, it was used mostly for genome-wide association mapping and genomic prediction, although it enables the calculation of high-density linkage maps. This raises the question, if high-density linkage maps offer potential to improve important QTL mapping parameters under limited population size. The influence of population size clearly indicated that the statistical power, QTL effect estimates, and precision of QTL localization benefit from larger populations (Darvasi et al. 1993; Beavis 1998; Vales et al. 2005). In contrast, possible advantages of high-density compared to low-density maps are still controversially discussed in the literature.

Genetic variation within a biparental population depends on the genetic similarity between the two parents. Low genetic similarity between the parents might result in large genetic variance and thus, several QTL might segregate. However, under practical considerations, the parents of different populations are related in most breeding programs and hence, show high genetic similarity (Bink et al. 2012). As a consequence, population size should be enlarged to enable sufficient genetic variation. However, population size cannot be arbitrarily enlarged due to increasing costs, e.g., for phenotyping of all lines. Thus, population size is limited and the exploitation of the available high-density SNP arrays for linkage mapping could be a means to gain improvements in QTL mapping. First advantages of high-density linkage maps in QTL mapping were reported in the literature for experimental studies in barley (Hori et al. 2003), rice (Yu et al. 2011), and maize (Shi et al. 2011; Almeida et al. 2012), as well as for simulation studies (Li et al. 2010). In contrast, former simulation studies (Darvasi et al. 1993) and analytical approaches (Piepho 2000) indicated that QTL mapping does not profit from an increase in marker density beyond 10 cM. This raises the question if high-density linkage maps with a marker distance of 1 cM in polymorphic regions would be an overkill for QTL mapping, or if important QTL mapping parameters would benefit, and closely linked QTL could be detected separately with high-density maps (Stange et al. 2013b).

Genetic analysis of complex traits by QTL mapping including epistasis

It is well known that direct improvement of genetically complex traits by selection on it is difficult due to their moderate heritability, high sensitivity to environmental conditions, and highly polygenic nature (Holland 2007). However, these genetically complex and physiological multiplicative traits like grain yield (GY) can be decomposed into their underlying component traits. Therefore, an attractive alternative for the improvement of the complex trait can be the investigation of the underlying component traits, which have commonly higher heritabilities, are genetically less complex, and are correlated with the complex trait (Hallauer et al. 2010). Thus, a first promising step for the investigation of the genetic basis of complex traits would be the

application of high-density QTL mapping for the underlying component traits. QTL mapping for GY and its components was performed by several studies in maize (Stuber et al. 1992; Austin and Lee 1996; Ma et al. 2007; Guo et al. 2011). Although these studies mapped QTL for both, the complex and the component traits, they did not account for the multiplicative character of GY by adapting their QTL mapping approaches to the advantages which are obviously provided by the component traits. In contrast, Melchinger et al. (1994), elaborated theoretically the relationship between gene effects of a complex trait and its component traits. Obviously, this approach could also have the potential to improve QTL detection for complex traits.

Epistasis refers to the interaction between a pair of loci in which the genotypic effect of one locus depends on the genotype of the second locus (Carlborg and Haley 2004). Thus, the genotype cannot be predicted by the simple sum of its single locus effects. Although it was reported in QTL mapping studies for maize that epistatic interactions are only of minor importance (Buckler et al. 2009), epistasis should not be ignored beforehand in QTL mapping. Owing to the high number of possible interactions between all detected QTL, epistasis could in sum, still be relevant. High power for detection of epistatic interactions could be gained if QTL mapping approaches consider single QTL of the component traits, which are precisely localized and estimated. Implication of component traits in tests for epistatic interactions could further give insights into the genetic architecture and interaction networks of complex traits and could answer, if meaningful interactions in the analyzed populations exist.

This raises the questions, if mapping of component traits with a well saturated high-density linkage map, could gain in power and resolution of QTL detection as well as in more precise localization of QTL than mapping the complex trait directly, and if the implementation of the component traits in the QTL mapping approach could unravel the genetic architecture.

Genomic prediction

The availability of high-throughput genotyping data in plants paved the way for the application of genomic prediction or selection (GS) in plant breeding (Lorenz et al. 2011), which was originally developed and already successfully applied in animal

breeding (Meuwissen et al. 2001). In classical marker-assisted selection (MAS) (e.g., Dekkers and Hospital 2002), a subset of significant markers, linked to mostly large-effect QTL, is used for selection of superior lines. In contrast, GS predicts breeding values of individual lines, e.g., in a biparental DH population, by incorporating information from all available markers simultaneously. Thus, biases by selection of markers are avoided and more variation caused by small-effect QTL is captured (Heffner et al. 2009). GS consists in the prediction of genotypic performance in a training population or test set (TS) on the basis of both, phenotypic and marker data, and in the prediction of genotypic performance in a validation set or population (VP) on the basis of only marker data (Meuwissen et al. 2001; Heffner et al. 2009). Subsequently, individual lines can be selected on the basis of their predicted performance without ever evaluating them in the field. Therefore, with the application of GS, the length of a breeding cycle can dramatically be reduced due to the shortfall of time and costs consumed by field tests (Lorenz et al. 2011).

Usually, a practical breeding program consists of several related and unrelated small biparental populations. This raises the question for the successful implementation of GS in breeding programs, in particular, how the TS should be constructed from related and unrelated populations to predict individual lines from single populations. Several prediction schemes with changing compositions of the TS are possible which might influence the prediction accuracy $r(g, \hat{g})$ of individual crosses. The latter is measured as the correlation $r(g, \hat{g})$ between the true genotypic value (g) and its estimated value ($\hat{g}$) (Riedelsheimer et al. 2012b).

For the prediction of the genotypic value of complex traits, classical linkage mapping approaches and newly developed approaches, based on component traits, can be directly compared by the calculation of the prediction accuracy ($r(g, \hat{g})$). Thus, the question can be answered if the component trait based approaches result in a meaningful gain in prediction of a complex trait. Additionally, the prediction on the basis of detected QTL and on the basis of GS can be compared by calculating prediction accuracies for both approaches. Thus, the question can be answered if GS outperforms the prediction based on QTL detected with classical linkage mapping.

Objectives

The overall aim of my thesis was the construction of high-density linkage maps, the evaluation of their impact on important QTL mapping parameters, and their application in different QTL linkage mapping approaches with biparental populations of DH lines. In particular, the objectives were to

(1a) investigate the effect of high-density versus low-density linkage maps in QTL mapping with experimental data and a simulation study on the power of QTL detection, the precision of QTL localization, and the bias of QTL effect estimates,

(1b) analyze the resolution of closely linked QTL with varying linkage distances and different linkage phases,

(2) map QTL for yield and yield components with high-density maps in four biparental populations of DH lines and elucidate networks of epistatic interactions,

(3a) compare prediction accuracy of different QTL mapping approaches to predict the complex trait grain yield, i.e., compare direct prediction methods and prediction with the aim of component traits,

(3b) investigate how the training set should be constructed from multiple related and unrelated biparental populations to predict progeny from individual crosses.

Chapter 2

High-density Genotyping: an Overkill for QTL Mapping? Lessons learned from a Case Study in Maize and Simulations[1]

Michael Stange[1], H. Fiedrich Utz[1], Tobias A. Schrag[1], Albrecht E. Melchinger[1], and Tobias Würschum[2]

[1] Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany

[2] State Plant Breeding Institute, University of Hohenheim, 70599 Stuttgart, Germany

Theor Appl Genet (2013) 126:2563-2574

The original publication is available at http://link.springer.com/article/
(DOI: 10.1007/s00122-013-2155-0).

Abstract

High-density genotyping is extensively exploited in genome-wide association mapping studies and genomic selection in maize. In contrast, linkage mapping studies were until recently mostly based on low-density genetic maps and theoretical results suggested this to be sufficient. This raises the question, if an increase in marker density would be an overkill for linkage mapping in biparental populations, or if important QTL mapping parameters would benefit from it. In this study, we addressed this question using experimental data and a simulation based on linkage maps with marker densities of 1, 2, and 5 cM. QTL mapping was performed for six diverse traits in a biparental population with 204 doubled haploid maize lines and in a simulation study with varying QTL effects and closely linked QTL for different population sizes. Our results showed that high-density maps neither improved the QTL detection power nor the predictive power for the proportion of explained genotypic variance. In contrast, the precision of QTL localization, the precision of effect estimates of detected QTL, especially for small and medium sized QTL, as well as the power to resolve closely linked QTL profited from an increase in marker density from 5 to 1 cM. In conclusion, the higher costs for high-density genotyping are compensated for by more precise estimates of parameters relevant for knowledge-based breeding, thus making an increase in marker density for linkage mapping attractive.

Chapter 3

High-density Linkage Mapping of Yield Components and Epistatic Interactions in Maize with doubled Haploid Lines from Four Crosses

Michael Stange[1*], Tobias A. Schrag[1*], H. Fiedrich Utz[1], Christian Riedelsheimer[1], Eva Bauer[2], and Albrecht E. Melchinger[1]

[1] Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

[2] Plant Breeding, Technische Universität München, 85354 Freising, Germany

[*] Both authors contributed equally.

Mol Breeding (2013) 32:533-546

The original publication is available at http://link.springer.com/content/ (DOI: 10.1007/s11032-013-9887-z).

Abstract

Grain yield (GY) is a genetically complex and physiologically multiplicative trait which can be decomposed into the components 100-kernel weight (HKW) and kernel number (KN). Genetic analysis of these less complex yield component traits may give insights into the genetic architecture and predictive ability of complex traits. Here, we investigated how incorporation of component traits and epistasis in QTL mapping approaches influences the accuracy of GY prediction. High-density genetic maps with 7,000 to 10,000 polymorphic SNPs were constructed for four biparental populations. The populations comprised between 99 and 227 doubled haploid (DH) maize lines which were phenotyped in field trials in two environments. Heritability was highest for HKW (88 to 89%), intermediate for KN (72 to 80%), and lowest for GY (64 to 83%). Mapped QTL explained in total between 21 and 55% of the genotypic variance for GY, 22 to 67% for KN, and 24 to 75% for HKW. Support intervals of QTL were short, indicating that QTL were located with high precision. Co-located QTL with same parental origin of favorable alleles were detected between populations for the same traits and within populations for different traits. Using GY predictions based on the detected QTL, prediction accuracies (*r*) determined by cross validation ranged from 0.18 to 0.52. Epistatic models did not outperform the corresponding additive models. In conclusion, models based on QTL positions of component traits support identification of favorable alleles for multiplicative traits and provide a basis to select superior inbred lines by marker-assisted breeding.

Chapter 4

Genomic Predictability of Interconnected Biparental Maize Populations[3]

Christian Riedelsheimer[*], Jeffrey Endelman[§], Michael Stange[*], Mark E. Sorrells[§] Jean-Luc Jannink[§†], Albrecht E. Melchinger[*]

[*] Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

[§] Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA

[†] USDA-ARS, R. W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY 14853, USA

Genetics (2013) 194:493-503

The original publication is available at http://www.genetics.org/content/
(DOI: 10.1534/genetics.113.150227).

Abstract

Intense structuring of plant breeding populations challenges the design of the training set (TS) in genomic selection (GS). An important open question is how the TS should be constructed from multiple related or unrelated small biparental families to predict progeny from individual crosses. Here, we used a set of five interconnected maize (*Zea mays* L.) populations of doubled-haploid (DH) lines derived from four parents to systematically investigate how the composition of the TS affects the prediction accuracy for lines from individual crosses. A total of 635 DH lines genotyped with 16,741 polymorphic SNPs were evaluated for five traits including three kernel yield component traits and Gibberella ear rot severity. The populations showed a genomic similarity pattern, which reflects the crossing scheme with a clear separation of full sibs, half sibs, and unrelated groups. Prediction accuracies within full-sib families of DH lines followed closely theoretical expectations accounting for the influence of sample size and heritability of the trait. Prediction accuracies declined by 42% if full-sib DH lines were replaced by half-sib DH lines, but statistically significantly better results could be achieved if half-sib DH lines were available from both instead of only one parent of the validation population. Once both parents of the validation population were represented in the TS, including more crosses with a constant TS size did not increase accuracies. Unrelated crosses showing opposite linkage phases with the validation population resulted in reduced or negative prediction accuracies, if used alone or in combination with related families, respectively. We suggest identifying and excluding such crosses from the TS. Moreover, the observed variability among populations and traits suggests that these uncertainties must be taken into account in models optimizing the allocation of resources in GS.

Chapter 5

General Discussion

This thesis was based on five interconnected biparental crosses (Figure 1). Pop. 2, the largest population, was analyzed in combination with a simulation study by Stange et al. (2013b) to evaluate the potential of high-density genotyping for QTL mapping. Pop. 1 to Pop. 4 were investigated by Stange et al. (2013a) with the focus on QTL mapping of grain yield and yield components. Finally, all five populations were analyzed by Riedelsheimer et al. (2013) for genomic prediction. In the following, I will discuss the population design and genetic characteristics of the four parental lines, evaluate the effect of high-density genotyping on QTL mapping parameters, and show possible benefits of multi-population QTL mapping approaches. I will compare prediction accuracies calculated on the basis of genomic prediction with prediction accuracies calculated on the basis of QTL detected with linkage mapping. Finally I will discuss how variation in the recombination rate influences the QTL distribution.

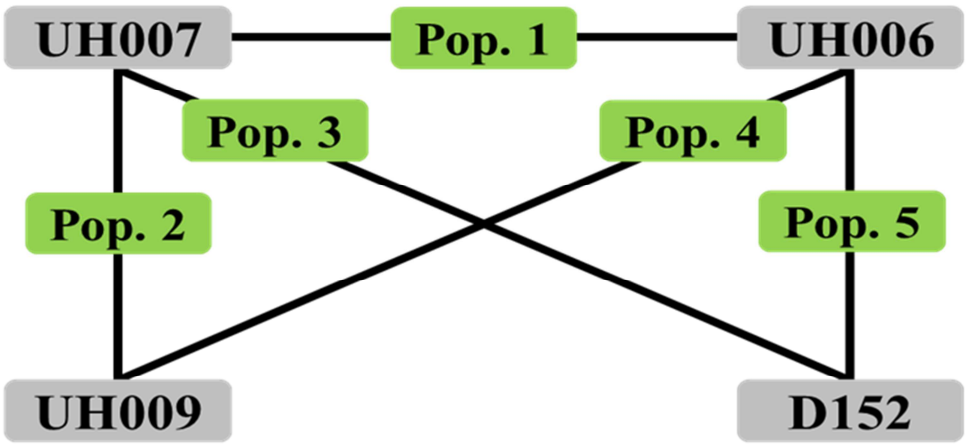

Figure 1 Crossing scheme to obtain the five populations (Pop.). Numbers (green boxes) between two parental lines (grey boxes) indicate the populations.

Population design and genetic characteristics of parental lines

All parental inbred lines are from the flint heterotic pool and genetic similarities, measured as identity-by-state (S_{IBS}) and genomic correlation (S_{GC}), were higher among parents UH006, UH007, and UH009 compared with parent D152 (Riedelsheimer et al. 2013). This structured pattern of relatedness between parents is similar to practical breeding programs, where parental lines are often related (Bink et al. 2012), and when unrelated, not in a population genetic sense. The former can be explained because the development of lines for hybrid breeding is mostly within a certain heterotic pool.

In total, 699 DH lines of all five populations were genotyped with SSR markers (Martin et al. 2011) and SNP markers (Stange et al. 2013a). All markers were used after a quality check for DH lines and SNPs (Stange et al. 2013a) for calculation of high-density genetic linkage maps for Pop. 1 to Pop. 5, containing 8,383, 7,169, 10,351, 8,562, and 10,372 markers, respectively. As expected by the relatively high genetic similarities between all parental lines, linkage maps of all populations showed several monomorphic regions (Figure 2). According to Stange et al. (2013b), these regions were arbitrarily defined as identical-by-descent (IBD) regions if they were larger than 20 cM. With this definition, approximately 9%, 27%, 7%, 10%, and 2% of the genome may be IBD from Pop. 1 to Pop. 5, respectively. The varying proportions of IBD regions of the populations are in agreement with the S_{IBS} and S_{GC} values between the respective population parents. Additionally, the lowest number of monomorphic bins was observed in Pop. 3 compared to Pop. 1, 2, and 4 (Stange et al. 2013a), which is in agreement with a low proportion of IBD regions (7%) in Pop. 3. This indicates that in Pop. 3, a slightly higher level of genetic variation might be exploited for selection of superior lines compared to the other populations. The analysis of co-located QTL within Pop. 3 for different traits, as well as between populations for the same trait, showed different co-located QTL in Pop. 3 compared to the other populations (Stange et al. 2013a). This indicates that in Pop. 3, besides the expected higher level of genetic variation, variation for the analyzed traits exists also in deviating genomic regions. One of the parents of Pop. 3, namely D152, showed lower genetic similarity with the other parents. Including such a parent in a cross might

explain, why in Pop. 3 QTL were detected in different genomic regions compared to the other populations, where parent D152 was not included. Parent D152 was also included in Pop. 5, where deviating results compared to the other populations were observed (Riedelsheimer et al. 2013). Especially in Pop. 5, important QTL were present which was shown by the analysis of the genome partitioning of genetic variance (Riedelsheimer et al. 2013). This was explained by opposite linkage phases in Pop. 5 resulting also in negative prediction signals.

In summary, genetic variance was observed in different genomic regions for the analyzed traits in populations where one parent shares slightly lower genetic similarity with the other parents, although all parents were from the same heterotic pool (Riedelsheimer et al. 2013; Stange et al. 2013a). Thus, knowing genetic patterns of population parents prior to the experiments, might be used to select and cross those parents in which a high level of new genetic variance can be exploited for detection of important genomic regions followed by selection of superior lines.

QTL mapping with high-density maps

High-density genotyping is nowadays routinely applied in maize breeding due to tremendously decreasing costs of high-throughput genotyping platforms. This results in a huge number of available markers which were used for construction of high-density linkage maps as a basis for high-density QTL mapping (Stange et al. 2013a). To evaluate possible advantages of high-density maps compared to low-density maps in QTL mapping, Stange et al. (2013b) constructed linkage maps with three marker densities of 1, 2, and 5 cM. These were used for QTL mapping in the largest experimental population (Pop. 2) with 204 DH lines for diverse phenotypic traits as well as in a simulation study. The simulation study was generated on the basis of QTL mapping results obtained in the experimental population.

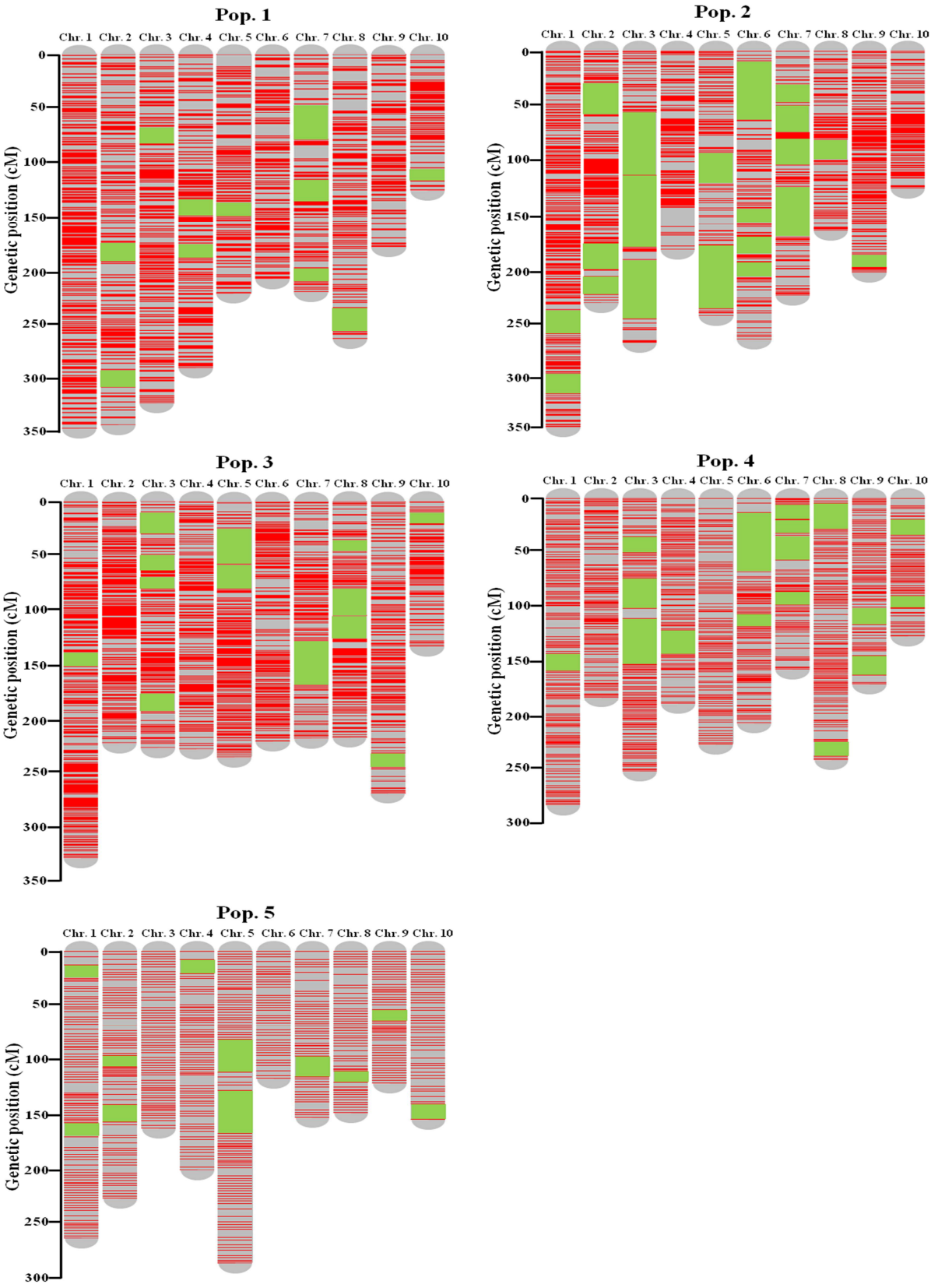

Figure 2 Mapped polymorphic markers with their genetic map positions in cM for chromosomes (Chr.) 1 to 10 based on doubled haploid lines of population 1 (Pop. 1) to Pop. 5. Regions, which are identical-by-descent (IBD; arbitrarily defined as monomorphic regions >20 cM) are indicated in light green.

Support interval (SI) length of detected QTL decreased with increasing marker density, indicating increasing precision of QTL localization (Stange et al. 2013b). Also Stange et al. (2013a) performed QTL mapping with a high-density map for grain yield and yield components. Across all detected QTL, the SI length was as narrow as observed by Stange et al. (2013b) with a marker density of 1 cM. Thus, both studies indicate that the precision of QTL localization is improved with high-density maps. However, across all detected QTL, two QTL detected for grain yield showed very wide SIs with a length of 40 and 58 cM, respectively (Stange et al. 2013a). Both intervals stretched over IBD regions, which explains the large difference compared to the average SI length across all detected QTL of only 12 cM (Figure 3) (Stange et al. 2013a).

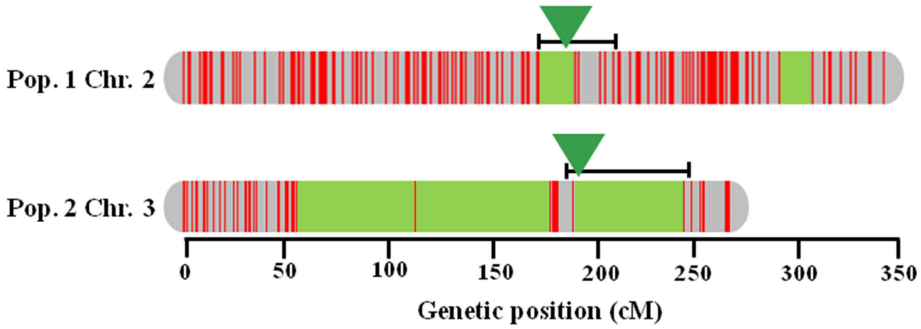

Figure 3 Mapped polymorphic markers with their genetic map positions in cM based on doubled haploid lines of population (Pop.) 1 and chromosome (Chr.) 2, and of population 2 and chromosome 3. Regions, which are identical-by-descent (IBD; arbitrarily defined as monomorphic regions >20 cM) are indicated in light green. QTL detected for grain yield are indicated as dark green triangles and their respective support intervals in cM are indicated as black horizontal bars.

To evaluate the precision of QTL effect estimates, a simulation study was designed in a way to answer why only medium to large effect QTL were detected in the experimental population (Stange et al. 2013b). Thus, independent QTL with reference additive effects from 0.10 to 0.75 were simulated and the precision of their estimated genetic effects was measured as deviation between the reference genetic effect and the estimated genetic effect (Stange et al. 2013b). In addition to high precision of QTL localization as observed in both studies, also the relative distribution of effect sizes of detected QTL for each trait indicated that only QTL with medium to large effects were identified, but no QTL with small effects (Stange et al. 2013a, b). Thus, the design and the question of the simulation study were confirmed by the distribution of QTL effects observed in all experimental populations. Moreover, conclusions drawn by Stange et

al. (2013b) from the simulation study onto the experimental population regarding the increasing precision of QTL effect estimates with increasing marker density, might be also valuable for the other populations analyzed by Stange et al. (2013a).

Simulation results showed that the resolution of closely linked QTL in coupling phase with linkage distances of 5 and 10 cM, respectively, was by far higher with the highest marker density as compared to the lower marker densities (Stange et al. 2013b). The high resolution of linked QTL only with the highest marker density might explain why Stange et al. (2013a) found besides individual QTL several pairs of co-located QTL within individual populations. This indicates that the separate detection of co-located QTL, i.e., separate detection of two QTL of different traits located closely together in narrow chromosomal regions, profited most from the high marker density. In contrast to the separate detection, low-density maps with lower resolution might detect only one of both QTL. This is probably the larger effect QTL although besides this QTL, a second smaller effect QTL is located nearby. Nevertheless, lower resolution might be sufficient if the main interest is the detection of chromosomal sections where only one large QTL is detected, which might be a QTL of a complex trait. However, if the main interest is the detection of interaction networks as well as to unravel the genetic architecture of complex and component traits (Stange et al. 2013a), high-density maps with high resolution and precise localization of co-located QTL are obligatory.

In conclusion, advantages of high-density maps for important QTL mapping parameters, relevant for knowledge-based breeding and analysis of genetic networks, were observed in the experimental populations and in the simulation study (Stange et al. 2013a, b). This broad basis of meaningful results allows a generalization of the advantages of high-density maps for QTL mapping. Thus, higher costs for high-density genotyping are compensated for by these gains.

Single population and multi-population QTL mapping

The analysis of large individual biparental populations is a general practice in QTL detection in order to guarantee a high power for QTL detection (Blanc et al. 2006). It is advisable to have mapping populations of around 100 individuals where a QTL that explains 10% of phenotypic variance, can be detected with a power of 85% (Charcosset and Gallais 1996). Therefore, this population size maybe considered roughly as lowest limit for QTL mapping experiments. The smallest population of this thesis, Pop. 5, was with only 45 DH lines by far below this size. The small population size explains why QTL mapping for Gibberella ear rot (GER) resistance was not informative (Martin 2012). Thus, Pop. 5 was not included in this thesis for QTL mapping, whereas the other populations were close (Pop. 4; N=99), or clearly above this size (Pop. 1, 2, and 3). Consequently, Pop. 1 to Pop. 4 were used for mapping of QTL for grain yield (GY), 100-kernel weight (HKW), and kernel number (KN) (Stange et al. 2013a).

The main focus of QTL mapping experiments in the four largest populations (Pop. 1 to Pop. 4) (Stange et al. 2013a) was the dissection of the complex trait GY into its underlying component traits HKW and KN. In contrast to the reported inconsistency of QTL localization when analyzing several populations for the same trait (Mihaljevic et al. 2004), Stange et al. (2013a) found in total four pairs of co-located QTL. These pairs were co-located for the same trait in two of the four analyzed populations, respectively, and each pair showed the same parental contribution of favorable or unfavorable alleles. This indicates that consistent genomic regions between the connected populations were found for individual traits, having similar effects on phenotypic trait expression. However, Pop. 3 was not involved in any co-location. Thus, to finally detect common QTL segregating in several populations, multi-population QTL mapping approaches might be a solution. These approaches provide also high QTL detection power and precision. If a common QTL is detected across several populations and estimated with similar genetic effects, this might indicate high reliability for this QTL. Using this QTL for marker-assisted selection (MAS) should result in high power for selection of superior lines.

For multi-population QTL detection, different approaches are available. One approach consists in the combination of logarithm of odd (LOD) curves from single population QTL mapping to fit meta-QTL positions and effects (Sosnowski et al. 2012; Utz 2012). Another approach is linking families by assuming that the QTL locations are the same in all analyzed populations (Jourjon et al. 2005; Bink et al. 2012). Both approaches were integrated in different software packages and were already successfully applied (e.g., MCQTL: Jourjon et al. 2005; BioMercator: Sosnowski et al. 2012; PlabMQTL: Utz 2012).

Besides main-effect QTL, several additive by additive epistatic interactions of QTL detected for individual traits, were found in all populations (Stange et al. 2013a). However, the total contribution of epistasis to the explained genotypic variance was rather small in all populations. This might be explained by the low power to detect epistatic interactions due to limited population size, which indicates that even 204 DH lines as in Pop. 2, are not sufficient. Therefore, to profit from large populations, Stange et al. (2013a) concluded that the application of multi-population approaches could increase the power for detection of epistatic interactions. Another benefit of multi-population approaches, especially when applied in connected populations, might be the test for QTL by genetic background interactions (Blanc et al. 2006). With connected populations, epistasis can be tested by comparison between a model assuming identical QTL effects in different populations and a model assuming nested QTL effects within populations. Thereby, the latter model accounts for possible interactions with the genetic background (Blanc et al. 2006).
Another multi-population approach was suggested by Bink et al. (2012). This Bayesian approach considers that parents of multiple connected populations are not unrelated. Bink et al. (2012) found that inclusion of parental IBD data in the Bayesian QTL mapping approach resulted in an increase of power and precision of mapped QTL. Owing to different levels of IBD (Figure 2) and genetic similarities between the five populations (Riedelsheimer et al. 2013), this approach seems promising for multi-population QTL detection for the five populations analyzed in this thesis.

Stange et al. (2013b) reported that important QTL mapping parameters, like increased precision of QTL localization, reflected by narrow support intervals, profited from

high-density linkage maps. These benefits might be observed also when analyzing several populations with multi-population approaches on the basis of high-density consensus maps. Both, Blanc et al. (2006) and Bink et al. (2012) concluded that their approaches would gain from the application of haplotype maps instead of marker maps and that the assignment of haplotypes to the parental inbreds would profit from dense marker data. High-density consensus maps, calculated on the basis of single population high-density maps as calculated in this thesis, would be the perfect basis for unambiguously assigning haplotypes to all parents of the mapping populations.

Prediction accuracies calculated with GS and with QTL detected by linkage mapping

The main difference between marker-assisted selection (MAS) and genomic prediction or selection (GS) is, that the latter predicts breeding values of lines by incorporating information of all available markers (Heffner et al. 2009). In contrast, MAS is based only on medium to large selected effect QTL that were detected with linkage mapping. If in a population mainly medium to large effect QTL segregate, which are also precisely estimated with high-density maps (Stange et al. 2013b), prediction accuracies on the basis of detected QTL by linkage mapping and prediction accuracies obtained with GS might be (i) similar and (ii) prediction accuracies calculated for different validation populations (VP) and the same training set (TS), respectively, might show similar trends. To evaluate both hypotheses, we calculated prediction accuracies exemplarily on the basis of QTL detected in Pop. 3 for Gibberella ear rot (GER) severity under different scenarios depending on the VP, with Pop. 3 as TS, respectively.

Prediction accuracies obtained for full-sibs (FS) from the same cross (Pop. 3; scenario 1A-CV) were highest, in between for the two half-sib (HS) families (Pop. 1 and Pop. 2; scenarios 1B-1 and 1B-2), and lowest if the TS comprised an unrelated (UR) population (Pop. 4; scenario 1C), independent of the number of QTL or cofactors used for calculation of prediction accuracies (Table 1). This decline from FS over HS to UR families is in agreement with prediction accuracies for GER severity reported by

Riedelsheimer et al. (2013) using whole genome prediction. It indicates that both methods, i.e., GS and prediction on the basis of QTL, depend on the level of relatedness between the genotypes in the TS and VP, which is in agreement with empirical studies (Clark et al. 2012; Habier et al. 2010).

Table 1 Prediction accuracies for Gibberella ear rot (GER) severity with population 3 (Pop. 3) as training set (TS), respectively, under four scenarios depending on different individual validation populations (VP) with different relatedness to the TS. Prediction accuracies were calculated on the basis of 21 cofactors, 14 QTL, and 3 QTL, respectively.

Scenario	VP	Relatedness	21 cofactors[†]	14 QTL[†]	3 QTL[‡]
1A-CV[¶]	Pop. 3	full sib	-	0.464	0.144
1B-1	Pop. 1	half sib	0.331	0.317	0.029
1B-2	Pop. 2	half sib	0.184	0.133	0.013
1C	Pop. 4	unrelated	0.131	0.085	0.005

[†] Cofactor selection and QTL selection with the Akaike´s Information Criterion (AIC) (Akaike 1974).
[‡] Cofactor selection with the modified Bayes Information Citerion (mBIC; Baierl et al. 2006) and QTL selection with the Bayes Information Criterion (BIC; Schwarz 1978).
[¶] Prediction accuracies for Pop. 3 as VP were calculated on the basis of five-fold cross validation (CV).

Interestingly, prediction accuracies on the basis of QTL increased strongly with increasing number of used QTL, which depends on the applied criteria for cofactor selection and QTL detection (Table 1). Applying the modified Bayes Information Criterion (mBIC, Baierl et al. 2006), which is described as a more stringent criterion for cofactor selection (Bogdan et al. 2008) and the Bayes Information Criterion (BIC; Schwarz 1978) for QTL selection, only one cofactor and three reliable QTL were detected in Pop. 3, and thus, used for prediction. This resulted in lowest prediction accuracies for all scenarios. In contrast, 21 cofactors were chosen and 14 QTL were detected with the Akaike´s Information Criterion (AIC; Akaike 1974). Prediction accuracies calculated with these cofactors and QTL, respectively, were by far higher for all scenarios compared to those on the basis of the low number of QTL. This indicates that the criterion for cofactor selection should be chosen according to the focus of the study. If the focus is the detection of reliable QTL, more stringent criteria like mBIC should be applied. In contrast, if the focus is to obtain high prediction accuracy, it is advisable to use less stringent criteria like AIC for QTL detection. Although the gap between highest prediction accuracies with GS (Riedelsheimer et al.

2013) and lowest values with prediction on the basis of only three QTL was substantially reduced when prediction was performed with the high number of QTL or cofactors, there was still a considerable gap to GS. This might be explained by possible small-effect QTL segregating in Pop. 3. These might remain undetected with linkage mapping, even on the basis of high-density maps probably due to the low power caused by limited population size (Stange et al. 2013b). Taken together, we could show that GS should be applied due to its superiority over MAS if the focus is the selection of superior lines. However, if the focus is the dissection of complex traits and to unravel the genetic architecture, linkage mapping with high-density maps offer high power and precision for detection of QTL (Stange et al. 2013a, b).

Influence of recombination rate on QTL distribution along the chromosome

Several studies detected QTL for yield and yield components (Guo et al. 2011; Li et al. 2007; Li et al. 2009; Lu et al. 2010; Ma et al. 2007; Peng et al. 2011; Yang et al. 2011). They found that chromosome 1, which is the largest chromosome, harbors the highest number of QTL in agreement with results reported by Stange et al. (2013a). However, Stange et al. (2013a) observed an accumulation of QTL in centromeric regions. Interestingly, recombination rate measured in cM/Mbp was lowest in centromeric regions and by far higher in teleomeric regions (Figure 4).

This shape of recombination cold and hot spots is in agreement with results reported by Nachman (2002) for several species, and by Schnable et al. (2009) and Farkhari et al. (2011) for maize. The latter authors observed that the recombination rate was around 100 fold lower in centromeric regions compared to teleomeric regions, a feature that might be explained by retrotransposon clusters. These are one of the factors, that account for most of the repetitive DNA in maize and that can enhance or suppress the recombination rate (Dooner and He 2008). Retrotransposon clusters vary in composition and location relative to genes (Wang and Dooner 2006), which might explain the existence of recombination cold and hot spots.

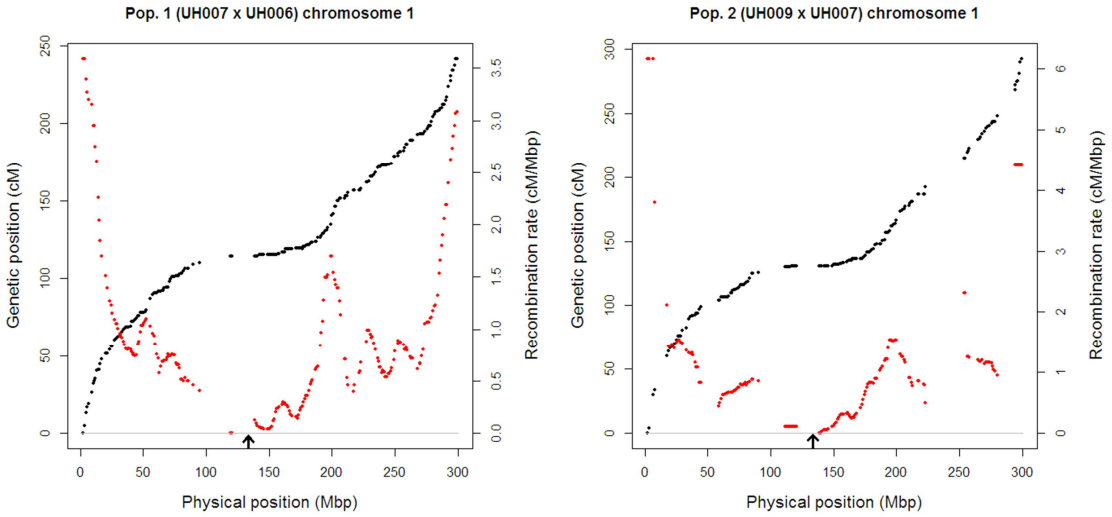

Figure 4 Relationship between physical (Mbp) and genetic (cM) map positions (black dots) and the corresponding recombination rates (cM/Mbp) (red dots) exemplarily for chromosome 1 of populations (Pop.) 1 and 2, respectively. The arrow indicates the approximated position of the centromere.

The effect of variation in recombination rates along the chromosome on QTL distribution was analyzed by Noor et al. (2001) in a simulation study based on the *Drosophila melanogaster* genome. These authors observed a clustering of QTL in regions of low recombination rates, which were primarily centromeric regions. In contrast, in regions of high recombination rates, only single QTL were detected. They concluded that this trend does not result from the QTL mapping algorithms and that large effect QTL, detected in regions with high recombination rates are more likely single genes of large effect. In contrast, QTL detected in regions of low recombination, are more likely QTL of several genes with small effects. To analyze these conclusions in the populations of this thesis, where the phenomenon of QTL clustering in regions of low recombination was observed (Stange et al. 2013a), a detailed analysis of the genome sequence could be a starting point.

Conclusions

This thesis was to the author´s best knowledge the first work calculating high-density linkage maps with the Illumina 50k maize chip resulting in marker densities of 1 cM in polymorphic regions. These linkage maps were used to investigate limitations and benefits of high-density QTL mapping. On the basis of the high-density maps, different QTL mapping models were applied for the dissection of the complex trait GY into its components HKW and KN in connected maize populations of DH lines.
We used a connected design of five populations which enabled evaluating the influence of single parents in crosses with up to three different parents. Genetic similarities, measured as identity-by-state (S_{IBS}) and genomic correlation (S_{GC}) between the parents indicated a structured pattern: three of the four parents showed slightly higher similarities among each other than each to the fourth parent. Consequently, this small difference yielded deviating results in Pop. 3, where this parent was crossed with another parent, compared to the crosses where it was not involved. Thus, knowing the structure of genetic similarities between the parents in advance, might give first hints for possible common or different QTL. This information should be considered when comparing several populations analyzed individually, and more importantly, when combining these populations in a multi-population QTL mapping approach.

Relevant QTL mapping parameters such as the precision of QTL localization and effect estimates, as well as the resolution of closely linked QTL are improved with high-density maps, as demonstrated in a large experimental population and a simulation study. Results of the experimental population, especially the high precision of QTL localization, were confirmed by QTL mapping in further DH populations. This allows a generalization of these findings and shows that the higher costs for high-density genotyping are by far outweighed.

QTL mapping with high-density maps for yield and yield components revealed several pairs of co-located QTL between the analyzed populations. Nevertheless, multi-population QTL mapping approaches could confirm these common QTL and, due to the higher QTL detection power, could detect more common QTL. Additionally,

multi-population QTL mapping facilitates the test for QTL by genetic background interactions. Thus, the genotypic variance explained by the detected QTL can be separated into its three components: main-effect QTL, QTL by QTL epistatic interactions, and QTL by genetic background interactions.

Prediction of genotypic values for GER severity on the basis of QTL detected with linkage mapping did not reach accuracies obtained with GS. This indicates a superiority of GS over MAS. However, the decline of prediction accuracies with GS from FS over HS to UR families was confirmed by the prediction on the basis of QTL detected with linkage mapping. Consequently, predicting genotypic values of additional new DH lines, developed from the same cross as used for the development of the TS, would result in highest prediction accuracies. In contrast, predicting genotypic values of additional new DH lines, developed from a cross where only one parent is common with the TS, would result in a strong decrease in prediction accuracies. Thus, both parents of the VP should be represented in an optimal TS. Further, it was clearly shown that prediction accuracies on the basis of detected QTL profited from an increased number of QTL or cofactors used for calculation of prediction accuracies. This indicates that also small-effect QTL segregate in the analyzed plant material and that these should be included in the prediction. In contrast to linkage mapping, GS incorporates information from all available markers simultaneously and thus, uses also variation from small-effect QTL which might remain undetected with linkage mapping. Consequently, selection of superior lines should be conducted with the aid of GS.

Chapter 6

Summary

Most important agronomic traits like disease resistance or grain yield (GY) in maize show a quantitative trait variation and, therefore, are controlled by dozens to thousands of quantitative trait loci (QTL). Mapping of these QTL is well established in plant genetics to elucidate the genetic architecture of quantitative traits and to detect QTL for knowledge-based breeding. Nowadays, high-density genotyping is routinely applied in maize breeding and offers a huge number of SNP markers used in association mapping and genomic selection (GS). This enables also the construction of high-density linkage maps with marker densities of 1 cM or even higher. Nevertheless, QTL mapping studies were until recently mostly based on low-density maps. This raises the question if high-density maps are an overkill for QTL mapping, or in contrast, if important QTL mapping parameters would profit from them. High-density maps could also be beneficial for dissection of the complex trait GY into its components 100-kernel weight (HKW) and kernel number (KN). Analysis of these less complex traits may help to unravel the genetic architecture and improve the predictive ability for complex traits. However, an open question is whether consideration of component traits and epistatic interactions in QTL mapping models are beneficial for predicting the performance of untested genotypes for the complex trait GY.

In this thesis, high-density linkage maps were constructed for biparental maize populations of doubled haploid (DH) lines and applied in different QTL linkage mapping approaches. In detail, the objectives of this study were to (1) investigate the effect of high-density versus low-density linkage maps in QTL mapping of important QTL mapping parameters and to analyze the resolution of closely linked QTL with experimental data and computer simulations, (2) map QTL for HKW, KN, and GY with high-density maps and to analyze epistatic interactions, (3) compare the prediction accuracy for GY with different QTL mapping models, and (4) answer the question how the composition of the test set (TS) influences the accuracy in genomic prediction of progenies from individual crosses.

This thesis was based on five interconnected biparental populations with a total of 699 DH lines evaluated in field experiments for GER resistance related traits as well as for HKW, KN, and GY. All DH lines were genotyped with the Illumina MaizeSNP50 Bead Chip and high-density linkage maps were constructed separately for each population.

For evaluation of high-density versus low-density maps on QTL mapping parameters, three linkage maps with marker densities of 1, 2, and 5 cM were constructed, starting from the full linkage map with 7,169 markers mapped in the largest population (*N*=204). QTL mapping was performed with all three marker densities in the experimental population for GER resistance related traits and for yield related traits, as well as in a simulation study with different population sizes. In the simulation study, independent QTL with additive effects explaining 0.14 to 7.70% of the expected phenotypic variance, as well as linked QTL with map distances of 5 and 10 cM, were simulated. Results showed that high-density maps had only minor effects on the QTL detection power and the proportion of genotypic variance explained. In contrast, support interval length decreased with increasing marker density, indicating an increasing precision of QTL localization. The precision of QTL effect estimates was measured as deviation between the reference additive effects and the estimated QTL effects. It gained from an increase in marker density, especially for small and medium effect QTL. Increasing the marker density from 5 to 1 cM was advantageous for separately detecting linked QTL in coupling phase with both linkage distances. In conclusion, this study showed that QTL mapping parameters relevant for knowledge-based breeding profited from an increase in marker density.

For QTL mapping of the complex trait GY and the components HKW and KN, three QTL mapping models were applied to the four largest populations, of which two models were based on the component traits HKW and KN. All models included tests for epistatic interactions. The results showed that heritability was slightly higher for the component traits compared to the complex trait. The average length of support intervals of detected QTL was short with 12 cM, indicating high precision of QTL localization. Co-located QTL with same parental origin of favorable alleles were detected within populations for different traits and between populations for same traits,

reflecting common QTL across populations. However, to finally confirm these common QTL, multi-population QTL mapping should be conducted. Based on the detected QTL, predictions for GY showed that epistatic models did not outperform the respective additive models. Nevertheless, component trait based models can be advantageous for identification of favorable allele combinations for multiplicative traits.

For all five populations, the comparison of genetic similarities reflected the crossing scheme with full-sib families, half-sib families and unrelated families. The evaluation of prediction accuracies for different scenarios depended on the composition of the TS. Highest prediction accuracies were observed for DH lines within full-sib families, medium values if full-sib DH lines were replaced by half-sib DH lines, and lowest values if the TS comprised of DH lines from unrelated crosses.

In conclusion, I found high-density linkage maps to be advantageous for linkage mapping in biparental DH populations by improving important QTL mapping parameters. Higher costs for high-density genotyping are by far compensated by these advantages. Dissecting the complex trait GY into its component traits HKW and KN by component trait based QTL mapping models revealed a complex genetic network of GY. Future research should focus on high-density consensus maps applied in multi-population QTL mapping to take advantage of the improved QTL detection power and to confirm common QTL across populations.

Chapter 7

Zusammenfassung

Viele agronomisch bedeutende Eigenschaften von Kulturpflanzen zeigen eine quantitative Merkmalsvariation. Die der Ausprägung solcher Merkmale zugrunde liegenden Genomregionen (sog. *quantitative trait loci* (QTL)) können mittels molekularer Marker und statistischer Verfahren kartiert werden. Die Kartierung dieser QTL ist in der Pflanzengenetik weit verbreitet, um die genetische Architektur von wichtigen Merkmalen wie Kornertrag oder Krankheitsresistenzen zu erforschen und um gezielter und effizienter züchten zu können. Mittlerweile sind bei Mais mehrere tausend sog. *single nucleotide polymorpishm* (SNP)-Marker bekannt, die auf Unterschieden in der Basenabfolge in der Mais-DNA beruhen. Diese SNP-Marker lassen sich routinemäßig durch Hoch-Durchsatz-Genotypisierungsverfahren ermitteln und bieten daher ein enormes Potential für die Maiszüchtung. Bisher wird dieses Potential an SNP-Markern jedoch lediglich in der Assoziationskartierung und in der Genomischen Selektion (GS) ausgeschöpft, obwohl auch die Möglichkeit besteht hochdichte genetische Karten zu erstellen, die in der QTL-Kartierung eingesetzt werden können. Allerdings wurde die QTL-Kartierung bisher meistens mit genetischen Karten mit geringer Markerdichte durchgeführt. Somit stellt sich die Frage, ob hochdichte genetische Karten eine genauere QTL-Kartierung ermöglichen. Hochdichte genetische Karten könnten ferner die Möglichkeit bieten, das Komplexmerkmal Kornertrag (GY) in seine Komponentenmerkmale 100-Korngewicht (HKW) und Kornanzahl (KN) zu zerlegen. Die Analyse von einfach vererbten Komponentenmerkmalen verspricht tiefere Einblicke in die genetische Architektur des Komplexmerkmals. Allerdings stellt sich die Frage, ob durch das Einbeziehen von Komponentenmerkmalen und epistatischen Interaktionen zwischen QTL auch die Vorhersage des Komplexmerkmals GY genauer wird.

Ziele der vorliegenden Arbeit waren, (1) potentielle Vorteile von hochdichten Karten im Vergleich zu Karten mit geringer Markerdichte auf wichtige QTL-Kartierungsparameter und die Auflösung eng gekoppelter QTL zu untersuchen, (2)

QTL für HKW, KN und GY mit hochdichten Karten zu kartieren und epistatische Interaktionen zu analysieren, (3) die Vorhersagegenauigkeit für GY mit verschiedenen QTL-Kartierungsmodellen zu vergleichen und (4) die Genauigkeit der genomischen Vorhersage von Nachkommen aus Kreuzungen in Abhängigkeit von der Zusammensetzung des Trainingsets (TS) zu untersuchen.

Die hier vorgestellte Arbeit basierte auf fünf verbundenen biparentalen Maispopulationen mit insgesamt 699 doppelt-haploider (DH) Linien, für die Merkmale der *Fusarium graminearum*-Resistenz und HKW, KN sowie GY erfasst wurden. Alle DH-Linien wurden mit mehr als 50.000 SNP-Markern genotypisiert und hochdichte genetische Karten für jede Population erstellt.

Ausgehend von der genetischen Karte der größten experimentellen Population (N=204) mit 7.169 Markern wurden genetische Karten mit Markerdichten von 1, 2 und 5 cM erzeugt. Die QTL-Kartierung wurde in dieser experimentellen Population für verschiedene Merkmale der *Fusarium graminearum*-Resistenz und des Kornertrags sowie in einer Computersimulation durchgeführt. In der Simulationsstudie wurden unabhängige QTL mit additiven Effekten angenommen, welche 0.14 bis 7.70% der phänotypischen Varianz erklärten und gekoppelte QTL mit 5 und 10 cM Abstand simuliert. Die Ergebnisse zeigten, dass hochdichte Karten nur einen geringen Effekt auf die Anzahl der detektierten QTL und den Anteil der erklärten genotypischen Varianz haben. Im Gegensatz dazu stieg die Präzision der QTL-Lokalisation mit steigender Markerdichte beträchtlich an. Die Genauigkeit der Schätzung der QTL-Effekte, insbesondere für QTL mit kleinen und mittleren Effektgrößen profitierte von ansteigender Markerdichte. Auch für die Auflösung eng gekoppelter QTL war ein Anstieg der Markerdichte vorteilhaft, da es nur mit der höchsten Markerdichte möglich war, die eng gekoppelten QTL separat zu detektieren. Das aus dieser Studie gezogene Fazit ist, dass QTL-Kartierungsparameter mit hoher Relevanz für die wissensbasierte Züchtung von einem Anstieg der Markerdichte profitieren.

Die QTL-Kartierung des multiplikativen Komplexmerkmals GY und der Komponentenmerkmale HKW und KN wurde mit drei QTL Kartierungsmodellen in den vier größten Populationen durchgeführt. Zwei Modelle basierten auf den

Komponentenmerkmalen und alle Modelle wurden ferner um epistatische Interaktionen erweitert. Die hochdichte Karte führte auch in dieser Studie zu einer exakteren Lokalisierung der detektierten QTL. Ko-lokalisierte QTL wurden innerhalb von Populationen für verschiedene Merkmale und zwischen Populationen für die gleichen Merkmale detektiert, so dass gemeinsame QTL über die Populationen hinweg vorliegen dürften. Die Vorhersage des GY von DH-Linien, die auf den detektierten QTL basierte, zeigte, dass die epistatischen QTL Modelle den entsprechenden rein additiven Modellen nicht überlegen waren. Dagegen trugen die beiden komponentenbasierten Modelle zur Aufdeckung von vorteilhaften Allelkombinationen für multiplikative Merkmale bei.

Die genetischen Ähnlichkeiten der fünf Populationen reflektierten das Kreuzungsschemata mit Vollgeschwister-, Halbgeschwister- und nicht-verwandten Familien. Die Zusammensetzung des TS beeinflusste die Genauigkeit der genomischen Vorhersage von Nachkommen erheblich. Höchste Vorhersagegenauigkeit wurde für DH-Linien innerhalb von Vollgeschwisterfamilien beobachtet, mittlere Werte, wenn Vollgeschwister-DH-Linien durch Halbgeschwister-DH-Linien ersetzt wurden, und geringste Werte wurden gefunden, wenn das TS aus DH-Linien von nicht-verwandten Kreuzungen bestand.

Die experimentellen Ergebnisse dieser Arbeit zeigten eindrucksvoll, dass hochdichte genetische Karten ein enormes Potential bieten wichtige QTL-Kartierungsparameter genauer zu schätzen. Somit werden die höheren Kosten der hochdichten Genotypisierung bei weitem durch die aufgezeigten Vorteile kompensiert. Die Zerlegung des Komplexmerkmals GY in die Komponentenmerkmale HKW und KN deckte ein komplexes genetisches Netzwerk für GY auf. Zukünftige Forschungsarbeiten sollten sich auf hochdichte Consensuskarten und auf QTL-Kartierung in multiplen Populationen fokussieren, um gemeinsame QTL über Populationen hinweg aufzufinden und damit die Züchtung effizienter zu gestalten.

Chapter 8

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Contr 19:716–723

Almeida GD, Makumbi D, Magorokosho C, Nair S, Borém A, Ribaut J-M, Bänziger M, Prasanna BM, Crossa J, Babu R (2012) QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. Theor Appl Genet. doi:10.1007/s00122-01202003-7

Austin D, Lee M (1996) Comparative mapping in $F_{2:3}$ and $F_{6:7}$ generations of quantitative trait loci for grain yield and yield components in maize. Theor Appl Genet 92:817–826

Baierl A, Bogdan M, Frommlet F, Futschik A (2006) On locating multiple interacting quantitative trait loci in intercross designs. Genetics 173:1693–1703

Beavis W (1998) QTL analyses: Power, precision and accuracy. In: Paterson AH (ed) Molecular analysis of complex traits. CRC Press, Boca Raton, FL, pp 145–161

Bink MCAM, Totir LR, Ter Braak CJF, Winkler CR, Boer MP, Smith OS (2012) QTL linkage analysis of connected populations using ancestral marker and pedigree information. Theor Appl Genet 124:1097–1113. doi:10.1007/s00122-011-1772-8

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224

Bogdan M, Frommlet F, Biecek P, Cheng R, Ghosh JK, Doerge RW (2008) Extending the modified bayesian information criterion (mBIC) to dense markers and multiple interval mapping. Biometrics 64:1162–9

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer J a, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, Da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718

Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Genet 5:618–625

Charcosset A, Gallais A (1996) Estimation of the contribution of quantitative trait loci (QTL) to the variance of a quantitative trait by means of genetic markers. Theor Appl Genet 93:1193–1201

Chen S, Li L, Li H (2009) Maize Doubled Haploid Breeding (in Chinese). China Agricultural University Press, Beijing

Clark SA, Hickey JM, Daetwyler HD, Van der Werf JHJ (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genetics, selection, evolution: GSE 44. doi:10.1186/1297-9686-44-4

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

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics 3:22–32. doi:10.1038/nrg701

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics 3:43–52

Dooner HK, He L (2008) Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. The Plant cell 20:249–58. doi:10.1105/tpc.107.057596

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Addison Wesley Longman, Harlow

Farkhari M, Lu Y, Shah T, Zhang S, Naghavi MR, Rong T, Xu Y (2011) Recombination frequency variation in maize as revealed by genomewide single-nucleotide polymorphisms. Plant Breeding 130:533–539. doi:10.1111/j.1439-0523.2011.01866.x

Food and Agriculture Organization of the United Nations, FAOSTAT (2009)

Guo J, Chen Z, Liu Z, Wang B, Song W, Li W, Chen J, Dai J, Lai J (2011) Identification of genetic factors affecting plant density response through QTL mapping of yield component traits in maize (*Zea mays* L.). Euphytica 182:409–422

Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. GSE 42. doi:10.1186/1297-9686-42-5

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hallauer AR, Carena MJ, Miranda JB (2010) Quantitative genetics in Maize Breeding. Iowa State Univ. Press, Ames, IA

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic Selection for Crop Improvement. Crop Sci 49:1–12. doi:10.2135/cropsci2008.08.0512

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161

Hori K, Kobayashi T, Shimizu A, Sato K, Takeda K, Kawasaki S (2003) Efficient construction of high-density linkage map and its application to QTL analysis in barley. Theor Appl Genet 107:806–813

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Jourjon M-F, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128–30. doi:10.1093/bioinformatics/bth481

Li H, Hearne S, Bänziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity 105:257–267

Li YL, Niu SZ, Dong YB, Cui DQ, Wang YZ, Liu YY, Wei MG (2007) Identification of trait-improving quantitative trait loci for grain yield components from a dent corn inbred line in an advanced backcross BC_2F_2 population and comparison with its $F_{2:3}$ population in popcorn. Theor Appl Genet 115:129–140

Li YL, Li XH, Li JZ, Fu JF, Wang YZ, Wei MG (2009) Dent corn genetic background influences QTL detection for grain yield and yield components in high-oil maize. Euphytica 169:273–284

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink J (2011) Genomic Selection in Plant Breeding: Knowledge and Prospects. Advances in Agronomy 110:77–123. doi:10.1016/B978-0-12-385531-2.00002-5

Lu M, Xie C-X, Li X-H, Hao Z-F, Li M-S, Weng J-F, Zhang D-G, Bai L, Zhang S-H (2010) Mapping of quantitative trait loci for kernel row number in maize across seven environments. Mol Breeding 28:143–152

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associtates, Sunderland, Massachusetts

Ma XQ, Tang JH, Teng WT, Yan JB, Meng YJ, Li JS (2007) Epistatic interaction is an important genetic basis of grain yield and its components in maize. Mol Breeding 20:41–51

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10:565–577

Martin M (2012) Genetic analysis of resistance to ear rot and mycotoxin contamination caused by Fusarium graminearum in European maize. Dissertation. University of Hohenheim

Martin M, Miedaner T, Dhillon BS, Ufermann U, Kessel B, Ouzunova M, Schipprack W, Melchinger AE (2011) Colocalization of QTL for gibberella ear rot resistance and low mycotoxin contamination in early european maize. Crop Sci 51:1935–1945

Martin M, Miedaner T, Schwegler DD, Kessel B, Ouzunova M, Dhillon BS, Schipprack W, Utz HF, Melchinger AE (2012) Comparative Quantitative Trait Loci Mapping for Gibberella Ear Rot Resistance and Reduced Deoxynivalenol Contamination across Connected Maize Populations. Crop Sci 52:32–43

Melchinger AE, Singh M, Link W, Utz HF, Kittlitz E (1994) Heterosis and gene effects of multiplicative characters: theoretical relationships and experimental results from *Vicia faba* L. Theor Appl Genet 88:343–348

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

Mihaljevic R, Utz H, Melchinger A (2004) Congruency of quantitative trait loci detected for agronomic traits in testcrosses of five populations of European maize. Crop science 124:114–124

Nachman M (2002) Variation in recombination rate across the genome: evidence and implications. Current Opinion in Genetics & Development 12:657–663. doi:10.1016/S0959-437X(02)00358-1

Noor M a, Cunningham a L, Larkin JC (2001) Consequences of recombination rate variation on quantitative trait locus mapping studies. Simulations based on the Drosophila melanogaster genome. Genetics 159:581–588

Peng B, Li Y, Wang Y, Liu C, Liu Z, Tan W, Zhang Y, Wang D, Shi Y, Sun B, Song Y, Wang T, Li Y (2011) QTL analysis for yield components and kernel-related traits in maize across multi-environments. Theor Appl Genet 122:1305–1320

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

Prigge V, Melchinger AE (2012) Production of haploids and doubled haploids in maize. Methods Mol Biol 877:161–172

Riedelsheimer C, Technow F, Melchinger AE (2012b) Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics 13:452

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink J-L, Melchinger AE (2013) Genomic Predictability of Interconnected Biparental Maize Populations. Genetics 194:493–503. doi:10.1534/genetics.113.150227

Schmidt W (2003) Hybridmaiszüchtung bei der KWS SAAT AG. Bericht über die Arbeitstagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreich, Gumpenstein, Österreich, 25-27 November:1–6

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. doi:10.1126/science.1178534

Schwarz G (1978) Estimating the dimension of a model. Ann Statist 6:461–464

Seitz G (2004) Interview published in Farmjournal. Link provided at: http: //www.greatlakeshybrids.com/performace/research-information/double-haploid-breeding-technology/ (accessed July 2011)

Seitz G (2005) The use of doubled haploids in corn breeding. Proc. of the 41st Annual Illionois Corn Breeders' Scholl, Urbana-Champaign, Illinois, USA

Shi L, Hao Z, Weng J, Xie C, Liu C, Zhang D, Li M, Bai L, Li X, Zhang S (2011) Identification of a major quantitative trait locus for resistance to maize rough dwarf virus in a Chinese maize inbred line X178 using a linkage map based on 514 gene-derived single nucleotide polymorphisms. Mol Breeding 30:615–625

Shull GH (1909) A pure line method of corn breeding. Am Breeders Assoc Rep 5:51–59

Sosnowski O, Charcosset A, Joets J (2012) BioMercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. Bioinformatics 28:2082–2083. doi:10.1093/bioinformatics/bts313

Stange M, Schrag TA, Utz HF, Riedelsheimer C, Bauer E, Melchinger AE (2013a) High-density Linkage Mapping of Yield Components and Epistatic Interactions in Maize with Doubled Haploid Lines from Four Crosses. Mol Breeding. doi:10.1007/s11032-013-9887-z

Stange M, Utz HF, Schrag TA, Melchinger AE, Würschum T (2013b) High-density Genotyping: an Overkill for QTL Mapping? Lessons learned from a Case Study in Maize and Simulations. Theor Appl Genet

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Utz HF (2012) PlabMQTL - Software for meta-QTL analysis with composite interval mapping. Version 0.5s. Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim.

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849

Vales MI, Schön CC, Capettini F, Chen XM, Corey a E, Mather DE, Mundt CC, Richardson KL, Sandoval-Islas JS, Utz HF, Hayes PM (2005) Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor Appl Genet 111:1260–1270. doi:10.1007/s00122-005-0043-y

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. PNAS 103:17644–17649. doi:10.1073/pnas.0603080103

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210

Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch JH, Xu Y (2009) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breeding 25:441–451

Yang G, Li Y, Wang Q, Zhou Y, Zhou Q, Shen B, Zhang F, Liang X (2011) Detection and integration of quantitative trait loci for grain yield components and oil content in two connected recombinant inbred line populations of high-oil maize. Mol Breeding 29:313–333

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PloS one 6

Acknowledgments

First and foremost I want to express my sincere thanks to my academic supervisor Prof. Dr. A. E. Melchinger for giving me the opportunity to do my PhD in his group, and for his advice, suggestions and continuous support during my work.

I am very indebted to Prof. Dr. H. F. Utz for his advice in QTL mapping and statistical issues, and that he always had an open door and mind for enlightening discussions.

I am very grateful to Dr. Tobias Schrag for all his suggestions, very fruitful discussions on marker data and linkage maps, his helpful ideas and for his advice on learning programming, and for his guidance in teaching.

Sincere thanks to PD Dr. T. Würschum for his guidance in scientific publishing and for his patience in proofreading my manuscript.

Many thanks to Dr. W. Schipprack for his work in coordinating the field trials and the sample processing. Many thanks to J. Jesse, F. Mauch, H. Pöschel, T. Schmid, and R. Volkhausen, and all other people at the research stations who helped to successfully conduct the field experiments.

I thank all colleagues at the Institute of Plant Breeding, Seed Science and Population Genetics and at the State Plant Breeding Institute, Hohenheim, for creating a pleasant working environment. Special thanks to my PhD colleagues in my working group for the nice and pleasant atmosphere and the fruitful discussions in all scientific fields.
I am indebted to H. Kösling, B. Devezi-Savula, M. Lieb, and S. Meyer for resolving many organizational issues during my work.

I highly appreciate the financial support of the Sino-German IRTG research project and all involved members from Chinese and German side.

I very much appreciate the support of Prof. Dr. S. Chen in China. I want to deeply thank all my Chinese counterparts from the IRTG project for their help and guidance

in China. Thanks to my German IRTG colleagues for the time we shared in China and Germany.

Finally, my heartfelt thanks go to my parents for enabling everything and for supporting me in as many fields as possible, and to my sister for encouraging me throughout my work and for her very helpful suggestions in science.

Curriculum vitae

Name:	Michael Stange
Date and place of birth:	20. April 1985 in Eschwege

Education

May 2010 - June 2013	Doctorate candidate in Applied Genetics and Plant Breeding (Prof. Dr. A. E. Melchinger), University of Hohenheim
October 2005 - April 2010	M. Sc. and B. Sc. studies and graduations, Agricultural Sciences, University of Hohenheim, Germany
September 2004 - June 2005	Civilian service, Jugendwaldheim Hoher Meißner, Werra-Meißner-Kreis, Germany
July 2004	High school degree (Abitur), Oberstufen-gymnasium, Eschwege, Germany

Professional Experience

September 2010 - June 2013	Research Assistant, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Germany
July - September 2009	Intern, Strube Research GmbH, Söllingen, Germany
June - September 2008	Intern, State Plant Breeding Institute, Hohenheim, Germany
July - September 2007	Intern, Plant breeding, CPB Twyford, Cambridge, Great Britain
July - September 2006	Intern, Dairy and pig breeding, Zucht- und Besamungsunion Hessen eG, Alsfeld, Germany
July - October 2005	Intern, Crop production and livestock breeding farm, Christian Stange, Waldkappel, Germany
July - August 2004	Intern, Dairy farm, Robert Henning GbR, Weitnau, Germany

_____

Michael Stange

Hohenheim, den 24.06.2013

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und lediglich unter Zuhilfenahme der angegebenen Hilfsmittel und Quellen angefertigt habe. Wörtlich oder inhaltlich übernommene Stellen wurden von mir als solche gekennzeichnet.

Die vorliegende Arbeit wurde in gleicher oder ähnlicher Form keiner anderen Institution oder Prüfungsbehörde vorgelegt.

_____

Michael Stange

Hohenheim, den 24.06.2013