TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Pflanzenzüchtung

Efficiency of different statistical approaches for marker-assisted improvement of European corn borer resistance in maize

Flavio Foiada

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Univ.-Prof. Dr. Donna Pauler Ankerst
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Chris-Carolin Schön
	2.	Univ.-Prof. Dr. Ralph Hückelhoven

Die Dissertation wurde am 23.03.2015 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 19.05.2015 angenommen.

Table of contents

Summary

The European corn borer (ECB), *Ostrinia nubilalis* (Hübner) is a major insect pest of maize (*Zea mays* L.) in temperate climates. Rising average temperatures are expected to increase even more the severity of damage due to ECB in the future. In Central Europe, yield losses are primarily caused by ECB larval feeding within the stalks, resulting in disruption of stalk vascular tissue, secondary infections by fungal pathogens, and direct harvest losses as a consequence of stalk breakage. Native genotypic variation for resistance to ECB stalk damage is available in Central European maize. Hence, exploiting this variation for the development of new varieties with improved ECB resistance would be a significant contribution to securing maize yield productivity. The native resistance to ECB stalk damage has been described as polygenic, making its genetic improvement a challenging task. Given that accurate phenotyping of resistance in the field is cost- and labor intensive, marker-assisted selection (MAS) has been considered as a promising approach for a more efficient resistance improvement.

The main objective of this study was to evaluate the performance of whole genome-based selection, relative to selection based on individual quantitative trait loci (QTL), for improving resistance to ECB stalk damage in Central European elite maize. Three connected biparental populations comprising 590 doubled haploid (DH) lines were genotyped with a high-density single nucleotide polymorphism array and phenotyped at two field locations under artificial ECB infestation and at four locations under natural infestation in 2011. Resistance was evaluated based on stalk damage ratings (SDR), the number of feeding tunnels in the stalk (NT), and tunnel length (TL). A subset of 195 DH lines selected for similar flowering time and diverse SDR was evaluated in the following year as lines *per se* and as testcrosses. Individual- and joint-population QTL analyses were performed, and cross-validated predictive abilities of the QTL models were compared with genomic best linear unbiased prediction (GBLUP) retaining all markers in the model.

For all traits, genome-wide prediction (GP) based on GBLUP models consistently outperformed predictions based on the QTL models, despite the detection of QTL with sizeable effects. Intermediate to high predictive abilities of GBLUP were observed for all resistance traits, with values of up to 0.70 for SDR. These results corroborated the complex polygenic nature of resistance to ECB stalk damage, indicating that a genome-wide approach should be more effective for its genetic improvement.

As a consequence of this outcome, the study evaluated several GP scenarios with the objective of gaining practical insights for implementing GP for ECB resistance in breeding programs. Comparison of GP analyses performed within- and across- biparental populations suggested the existence of different resistance alleles segregating in the three populations. Thus, recombination of material evaluated herein might further increase the overall level of resistance. In a further GP scenario, predictions of SDR performed across environments demonstrated that it should be possible to train GP models using phenotypic data from a relatively limited number of field locations (*i.e.* three to four) without negatively affecting predictive ability. Finally, GP was effective in predicting testcross performance for the trait SDR based on GBLUP models trained at the DH line *per se* level.

Genotypic correlations between SDR and the stalk tunneling traits were intermediate, indicating that SDR is also affected by other factors than ECB feeding. However, at the DH line *per se* level, prediction accuracies of GBLUP for stalk tunneling traits were in the same order as observed for SDR given that TL and NT had lower heritabilities. Thus, a genome-based, multi-trait approach should be considered for the simultaneous improvement of the different components contributing to ECB stalk damage resistance. High negative genetic correlations were observed between resistance traits and flowering time in the unselected populations. However, evaluation in 2012 of the selected subsets of DH lines demonstrated that it is possible to maintain genetic variation for resistance traits while decreasing variation in flowering time.

In conclusion, changing from a QTL-based to a genome-wide approach should increase efficiency of MAS for ECB resistance considerably. With the availability of native resistance to ECB stalk damage in elite maize germplasm adapted to Central European conditions, the obtained results may open up avenues for implementing an integrated genome-based selection approach for the simultaneous genetic improvement of ECB resistance and agronomic traits.

Zusammenfassung

In den gemäßigten Klimazonen ist der Maiszünsler, *Ostrinia nubilalis* (Hübner) einer der bedeutendsten Schädlinge im Maisanbau. Mit steigenden Durchschnittstemperaturen ist zu erwarten, dass die Intensität der durch den Maiszünsler verursachten Schäden weiter zunehmen wird. In Zentraleuropa sind diese Schäden hauptsächlich auf die Fraßtätigkeit der Zünslerlarven innerhalb der Maisstängel zurückzuführen, welche die Zerstörung von Leitungsbündeln, Stängelbruch und Infektionen durch Pilzkrankheiten zur Folge hat. Natürliche genetische Variation für die Resistenz gegen Maiszünsler ist in lokal angepasstem Mais (*Zea mays* L.) vorhanden. Die Nutzung dieser Variation für die Entwicklung neuer Sorten mit verbesserter Resistenz könnte einen signifikanten Beitrag zur langfristigen Sicherung der Maisproduktion leisten. Die züchterische Verbesserung der Maiszünslerresistenz wird aufgrund deren polygener Vererbung erschwert. Zudem ist die phänotypische Charakterisierung der Resistenz teuer und arbeitsaufwendig. Eine Selektion mit Hilfe von molekulargenetischen Markern (markergestützte Selektion, MAS) wird daher als vielversprechender Ansatz für eine effektive Verbesserung der Resistenz gesehen.

Das Hauptziel dieser Arbeit war, die Leistungsfähigkeit der genomweiten Selektion mit der Leistungsfähigkeit einer MAS auf individuellen „quantitative trait loci“ (QTL) im europäischen Elitematerial zu vergleichen. Drei über eine gemeinsame Elternlinie verbundene Populationen von 590 doppelhaploiden (DH) Maislinien wurden mit hochdichten „single nucleotide polymorphism“ (SNP) Markern genotypisiert und in 2011 an zwei Feldstandorten unter künstlicher Infestierung sowie an vier Feldstandorten unter natürlichem Zünslerbefall phänotypisch evaluiert. Die Resistenz wurde anhand von Bonituren des Stängelbruchs (SDR), der Anzahl an Fraßgängen im Stängel (NT) und der Fraßganglänge (TL) erfasst. Anschließend wurden die Populationen einer Selektion auf hohe Variation in der Resistenz bei gleichzeitig ähnlichem Blühzeitpunkt unterzogen. Die selektierten DH Linien wurden im darauf folgenden Jahr als Linien *per se* und als Testkreuzungen evaluiert. QTL Analysen wurden sowohl mit Einzelpopulationen als auch kombiniert über Populationen hinweg durchgeführt. Die kreuzvalidierte Vorhersagefähigkeiten der QTL-Modelle wurden mit denen von genomweiten Vorhersagemodellen („genomic best linear unbiased prediction“, GBLUP) verglichen.

Trotz der Identifizierung von QTL mit relativ großen Effekten war die genomweite Vorhersage der QTL-basierten Vorhersage stets überlegen. Für alle Resistenzmerkmale wurden mittlere bis hohe Vorhersagegenauigkeiten von GBLUP beobachtet, mit Werten von bis zu 0,70 für SDR.

Diese Ergebnisse verdeutlichen die komplexen, polygenen Eigenschaften der natürlichen Zünslerresistenz und zeigen, dass ein genomweiter Ansatz effektiver als ein QTL-basierter Ansatz sein sollte.

In der Folge wurden mehrere genomweite Vorhersageszenarien mit dem Ziel evaluiert, praxisrelevante Erkenntnisse für die Anwendung der genomweiten Selektion in Zuchtprogrammen zu gewinnen. Vergleiche von Analysen innerhalb- und über Populationen hinweg wiesen auf unterschiedliche Resistenzallele in den drei Populationen hin. Die Rekombination von Nachkommen aus diesen Populationen könnte deshalb das Resistenzniveau weiter erhöhen. Ein weiteres Szenario mit Vorhersagen über Umwelten zeigte, dass eine verhältnismäßig überschaubare Anzahl an Versuchsstandorten (zum Beispiel drei bis vier) ausreichend sein sollte, um die genomweiten Vorhersagemodellen ohne nennenswerte Genauigkeitsverluste zu kalibrieren. Zuletzt konnte die Testkreuzungsleistung für das Merkmal SDR effektiv mit GBLUP Modellen vorhergesagt werden, welche mit Daten von DH Linien *per se* kalibriert wurden.

Zwischen SDR und den Fraßgang-Merkmalen wurden mittlere genetische Korrelationen beobachtet. Dies zeigt, dass SDR nicht nur durch Stängelfraß, sondern auch durch andere Faktoren beeinflusst wird. Unter Berücksichtigung der niedrigeren geschätzten Heritabilitäten für TL und NT waren die GBLUP Vorhersagegenauigkeiten für die Fraßgang-Merkmale vergleichbar mit denen für SDR bei den DH Linien *per se*. Für die genetische Verbesserung der verschiedenen Komponenten der Zünslerresistenz sollte deshalb ein genombasierter, multivariater Ansatz gewählt werden. Bei den unselektierten Populationen wurden hohe negative genetische Korrelationen zwischen den Resistenzmerkmalen und dem Blühzeitpunkt beobachtet. Die Prüfung der selektierten Fraktionen von DH Linien in 2012 konnte aber zeigen, dass es möglich ist, genetische Variation für die Zünslerresistenz zu erhalten und gleichzeitig die Variation im Blühzeitpunkt zu reduzieren.

Die Umstellung von einem QTL-basierten hin zu einem genomweiten Ansatz könnte die Effizienz der markergestützten Selektion für die Zünslerresistenz wesentlich erhöhen. Dank der Verfügbarkeit von natürlicher Zünslerresistenz in zentraleuropäischem Elitematerial, könnten die Ergebnisse dieser Studie Wege für eine integrierte genomweite Selektion für die gleichzeitige Verbesserung von Resistenz und agronomischen Merkmalen eröffnen.

List of abbreviations

ADF	Acid detergent fiber
ADL	Acid detergent lignin
ANOVA	Analysis of variance
ANT	Days to anthesis
BBG	Field trial location Bernburg (D)
BLUE	Best linear unbiased estimation
BLUP	Best linear unbiased prediction
bp	DNA base pair
BRIX	Maize stalk sucrose concentration in °Brix degrees
Bt	*Bacillus thuringiensis* (Berliner)
cM	Centimorgan
CIM	Composite interval mapping
CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo (*engl.* International Maize and Wheat Improvement Center)
CV	Cross-validation
DH	Doubled haploid
DIMBOA	2,4-dihidroxy-7-methoxy-1,4-benzoxazin-3-one
DNA	Deoxyribonucleic acid
ECB	European corn borer, *Ostrinia nubilalis* Hübner
EH	Ear height
ES	Estimation set
FER	Field trial location Ferrara (I)
FRS	Field trial location Freising (D)
GBLUP	Genomic best linear unbiased prediction
GP	Genome-wide prediction
h^2	Heritability
HER	Field trial location Herbolzheim (D)
LASSO	Least absolute shrinkage and selection operator
LD	Linkage disequilibrium
LOD	Logarithm of the odds
LSD	Least significant difference

MABC	Marker-assisted backcrossing
MARS	Marker-assisted recurrent selection
MAS	Marker-assisted selection
Mbp	DNA mega base pairs, *i.e.* 10^6 base pairs
MLR	Multiple linear regression
MZH_i	Field trial location Münzesheim (D), artificially infested
MZH_n	Field trial location Münzesheim (D), naturally infested
NAM	Nested association mapping
NT	Number of European corn borer feeding tunnels
PCA	Principal component analysis
PH	Plant height
Pop1	DH population one from the cross R1 × S1
Pop2	DH population two from the cross R2 × S1
Pop3	DH population three from the cross R3 × S1
QTL	Quantitative trait locus / loci
r	Pearson's correlation coefficient
R^2	Coefficient of determination
R1	Resistant parental line one
R2	Resistant parental line two
R3	Resistant parental line three
REML	Restricted maximum likelihood
rep	Repeatability
RFLP	Restriction fragment length polymorphism
RR-BLUP	Ridge-regression best linear unbiased prediction
S1	Common susceptible parental line
SDR	Stalk damage rating
SD	Standard deviation
SE	Standard error
SIL	Days to silking
SMR	Single marker regression
SNP	Single nucleotide polymorphism
TL	Tunnel length, *i.e.* length of European corn borer feeding tunnels
TOS	Field trial location Tournoisis (F)
TS	Test set

List of figures

List of tables

Publications out of this thesis

The following article has been published in advance out of this thesis:

Foiada F, Westermeier P, Kessel B, Ouzunova M, Wimmer V, Mayerhofer W, Presterl T, Dilger M, Kreps R, Eder J, Schön CC (2015) Improving resistance to the European corn borer: a comprehensive study in elite maize using QTL mapping and genome-wide prediction. Theor Appl Genet 128:875-891

The publication is available at:
http://link.springer.com/article/10.1007/s00122-015-2477-1

Candidate`s contribution: construction and randomization of field trial designs, major contribution to experimental field work, statistical data analysis, interpretation and discussion of results, composition of graphs and tables, writing and revisions of the manuscript.

1. INTRODUCTION

1.1. Origin and importance of maize

Maize (*Zea mays* L.) is a crop species of the family Poaceae which was domesticated approximately 9,000 years ago from the wild grass teosinte (*Zea mays ssp. parviglumis*) in current-day Mexico (Beadle 1939). Domesticated maize spread northwards and southwards in the American continent and was particularly popular in the Aztec and Inca empires (Rebourg et al. 2003). The first introduction of maize into Europe dates back to Christopher Columbus who brought it to Spain from his voyage in the West Indies. Additional maize germplasm was brought to Europe from several other North- and South-American regions in the following decades. An important introduction was the so-called Northern Flint maize type from North-Eastern America, which was already adapted to cool climatic conditions in pre-Columbian times and is supposed to have fostered maize cultivation in the colder environments of Central Europe (Dubreuil et al. 2006). Traditional maize cultivation in the past relied on open-pollinated local landraces. With the widespread mechanization of agriculture and the first great successes of hybrid maize breeding in the 1930s (Crow 1998), maize quickly became one of the most important crops worldwide. Landraces soon disappeared from farmers' fields and were replaced with modern hybrid varieties having much higher yield potential (Barcaccia et al. 2003). Maize is nowadays the most widely cultivated agricultural crop together with wheat (*Triticum aestivum* L.) and rice (*Oryza sativa* L.). The amount of maize produced worldwide exceeded 875 million metric tons in 2012 (FAOSTAT 2012) making it the single most important cereal crop by total production. A key advantage of maize lies in its multiple potential uses. In many developing countries, particularly in Eastern and Southern Africa and parts of Latin America, maize is a staple food and represents the most important carbohydrate source in the human diet. In industrialized nations, on the other hand, it is primarily used as feed for livestock (in either grain or whole-plant silage form) or as industrial raw material. In recent years maize also gained importance as energy crop, particularly in the USA where it is mostly employed for the production of fuel ethanol from grain (USDA 2013) and in Europe where whole-plant biomass is used for biogas production. In Germany, for example, the acreage of maize grown for biogas has exponentially increased in the last decade (Grieder et al. 2012). In 2013, the total area under maize cultivation in Germany was almost 2.5 million hectares, out of which approximately 1.2,

0.8 and 0.5 million hectares were used for silage maize (livestock feed), biogas maize and grain maize production, respectively (FNR 2013).

1.2. The European corn borer (ECB)

Stem borers are insects belonging to the order Lepidoptera and are among the most important pests of maize worldwide. Ten different stem borer species have been described which are responsible for significant yield losses across both the tropical and temperate maize growing regions of the world (Malvar et al. 2008). Among these, the European corn borer (ECB) *Ostrinia nubilalis* (Hübner) of the family Crambidae is recognized as a major pest of temperate maize able to cause up to 30 % yield losses (Meissle et al. 2010). As the name implies, the ECB is native to Europe and was introduced in the USA probably between the late 19th and early 20th century (Smith 1920). The species can be subdivided into local races mainly differing with respect to voltinism (*i.e.*, the number of generations per year the insect produces) and to the composition of sexual pheromones secreted by the female moth to attract males for mating (Derron et al. 2009). Two distinct pheromone races of ECB exist which are named the Z and the E race due to the production of different ratios of Z- and E-isomers of the same chemical compound (11-tetradecenyl acetate). Lassance et al. (2010) demonstrated that these subtle differences in sex pheromone composition are the consequence of allelic variation at a single gene responsible for pheromone biosynthesis. Hence, mutations in the coding region of a single enzyme have led to the formation of two sexually isolated races of ECB. In Central Europe both E and Z races can be found in nature. While the E race possesses a broad range of different, mostly dicotyledonous host plant species including mugwort (*Artemisia vulgaris* L.) and hops (*Humulus lupulus* L.) the Z race is adapted to maize and hence primarily responsible for the relevance of ECB as agricultural pest (Saeglitz 2004). Concerning voltinism the ECB is subdivided into univoltine, bivoltine and multivoltine ecotypes in reference to the ability of producing one, two or more generations per year, respectively. Voltinism ecotypes are mainly distinguished by differences in the time point and duration of important phases in their life cycle. For example, univoltine ecotypes must go through a period of (winter) dormancy before pupation whereas dormancy is facultative in bi- and multivoltine ecotypes (Derron et al. 2009). The developmental phase after winter preceding moth emergence is also significantly longer in univoltine compared to bi- and multivoltine ecotypes (Derron et al. 2009). Significant genetic variation does not only exist between sex pheromone races of ECB but between voltinism

ecotypes as well (Pornkulwat et al. 1998) leading to the conclusion that each ecotype can be considered as a different race of ECB also at the genetic level.

The life cycle of ECB in Central Europe is still prevalently univoltine whereas in Southern Europe and in the extensive maize growing areas of the U.S. Corn Belt bi- and multivoltine ecotypes are commonly found. Larvae overwinter in maize stubble until rising temperatures in spring induce pupation, which occurs between May and June in Central Europe. After emergence and mating, the night-active moths will lay clusters of 10 to 40 eggs on the lower side of maize leaves, usually between mid-June and the beginning of July (Papst et al. 2001). Newborn ECB larvae start by feeding on young leaves while showing a preference for pollen if this is available (Brindley et al. 1975). At or before the flowering stage of maize the larvae enter the stalks by burrowing a hole in the most tender stalk rind tissue close to the tassel. For the rest of the season ECB larvae will mostly feed within the stalks and partly within the ear shanks as well (Melchinger et al. 1998a). Thereby the larvae tend to feed from the top towards the bottom of plants while in most cases exiting and re-entering the stalks to avoid the harder tissue at nodes. However, this tendency cannot be generalized as the behavior observed in reality is more complex. Hence, at harvest time several larvae can be found spread across the entire length of the plant and not only concentrated in the portion closer to the soil.

Under current Central European climatic conditions damage to plants is primarily caused by larval feeding within the stalks. The consequences are on the one hand physiological yield reduction due to the disruption of vascular tissue (Klenke et al. 1986), whereby translocation of assimilates and nutrients within the plant is significantly hampered. On the other hand, ECB feeding compromises the stability of stalks leading to stalk breakage and complete harvest loss. Indeed, the most easily identifiable symptoms of ECB infestation are broken stalks that become evident in a maize field as the crop approaches maturity. The ears drop down to the soil and are lost for mechanical harvest. An additional, indirect damage caused by ECB are the infections by fungal pathogens such as *Fusarium ssp.* that find ideal conditions for development with ECB feeding wounds as portals of entry. Fungal infections will often lead to mycotoxin contamination of grains and silage (Magg et al. 2002, Papst et al. 2005).

The spread and severity of the pest is currently favored by global climate change and by the continuously increasing area under maize cultivation in Central Europe. During the last decades the ECB has been steadily spreading northwards. In Germany, for example, occurrence of ECB was reported almost countrywide by 2010 (Figure 1).

Figure 1 Occurrence of ECB in Germany, year 2010 (dark grey area). Source: modified from GMO-Safety (2011)

Moreover, between 2006 and 2010 bivoltine ECB ecotypes were found for the first time in South-Western Germany (Kansy 2010). Rising average temperatures will probably lead to increased ECB pressure in the future: an ecological model by Trnka et al. (2007) prognosticated an area-wide establishment of bivoltine ECB in Central Europe for the period 2025-2050.

Crop management strategies for the control of ECB include crop rotation, cultivation of resistant maize varieties, as well as mechanical, chemical and biological control (Hommel et al. 2006). A diverse crop rotation is one of the most effective options but often not feasible in practice, particularly in areas characterized by intensive maize cultivation. Mechanical control involves thorough mulching and incorporation of harvest residuals into the soil. As a consequence the number of ECB moths emerging after winter is considerably reduced. However, this practice is not desired under conservation agriculture cropping systems where the goal is to reduce the intensity of soil tillage operations. Reduced tillage may even be a necessity for example in areas under severe drought stress, where it is important to limit evaporative water loss from the soil. Chemical control is not straightforward for ECB because of the reduced time window available for an effective insecticide application. This window lies between the time points of larval hatching and larval penetration within the stalks. At this stage the maize canopy has often reached a height that requires special machinery for performing insecticide treatments (*i.e.*, stilt tractors). Biological control is mostly performed using different species of the ichneumon wasp *Trichogramma ssp.*, which is a parasitoid of ECB egg clusters (Liebe 2004, Albert et al. 2008). The extent of ECB-parasitism achieved using *Trichogramma* depends on a number of environmental factors. Hence, this measure of control is not always

successful to the same degree. For both chemical and biological control it is important to follow the prognoses of ECB moth flight carefully in order to determine the optimal time point for application. Moreover, insecticide sprays and treatments with *Trichogramma* are both associated with high costs in terms of materials and labor.

The currently most effective way to control ECB is the cultivation of transgenic *Bt*-maize varieties expressing toxins from the soil bacterium *Bacillus thuringiensis* (Berliner). *Bt*-varieties offer near complete resistance to ECB and are nowadays widely cultivated in the USA. Transgenic varieties constituted 90 % of the maize crop cultivated in the USA in 2013, and more than 70 % of these varieties expressed several stacked genes conferring insect and herbicide resistance (Transgen.de 2013). In Europe, however, cultivation of *Bt*-maize is not a viable alternative at present given the very low public acceptance of green biotechnology. More importantly, *Bt*-induced resistance is monogenic (or a product of a few stacked resistance genes) implying the threat of quick adaptation of the pest to the resistance mechanism. Field-evolved resistance to *Bt*-crops has been reported for insect pests such as the maize stem borer *Busseola fusca* (Fuller) in South Africa (van Rensburg 2007), Western corn rootworm *Diabrotica virgifera virgifera* (LeConte) in the U.S. (Gassmann et al. 2011) and two species of cotton bollworm, *Helicoverpa zea* (Boddie) and *Pectinophora gossypiella* (Saunders) in the U.S., India and China (Tabashnik and Carrière 2010, Dhurua and Gujar 2011, Wan et al. 2012). In the case of ECB a combined strategy of varieties expressing high doses of *Bt*-toxins and the planting of non *Bt*-maize refuges has so far been successful in avoiding the spread of *Bt*-resistant strains in the USA (Siegfried and Hellmich 2012).

1.3. Native host plant resistance to ECB

Resilient maize varieties able to withstand a number of biotic and abiotic stresses will be essential for ensuring yield stability in the context of rapidly changing environmental conditions. Along with efforts for improving tolerance to key abiotic stresses such as drought, a stronger focus on resistance breeding against diseases and insect pests is required (Seifi et al. 2013). Native genotypic variation for ECB resistance is available in the species *Zea mays* (L.) (Hudon and Chiang 1985, Melchinger et al. 1998a). This variation has been described as polygenic (Schön et al. 1993), hence, native ECB resistance has the potential of being more durable than *Bt*-mediated resistance. Considering the future prognoses of an even increased

ECB severity in Central Europe, varieties with improved resistance to ECB would be a significant contribution to secure maize yield productivity.

1.3.1. *Mechanisms of resistance*

Native ECB resistance is highly complex being the product of a number of different morphological and physiological mechanisms. It can be broadly dissected into three main components: non-preference, antibiosis and tolerance (Bohn et al. 2000). Non-preference refers to mechanisms that render plants unattractive for oviposition by ECB moths. Antibiosis is the ability of plants to directly hinder ECB feeding, either through specific chemical compounds or morphological barriers. Antibiosis can include both constitutive and induced mechanisms of defense. Finally, tolerance is the ability to withstand feeding: plants with a high degree of tolerance will not suffer from stalk breakage or relevant yield losses even if leaves and the interior of stalks have been significantly damaged (Kreps et al. 1998a). Mostly constitutive characteristics such as the strength and stability of stalks play an important role in plant tolerance to ECB.

Non-preference

Malvar et al. (2008) reviewed the available knowledge on mechanisms related to oviposition non-preference. At the morphological level, high density of trichomes on the leaf surface was found to reduce oviposition by moths of the tropical stem borer species *Chilo partellus* (Swinhoe) (Kumar 1992), while at the physiological level several metabolites with either repellent or attractant action on stem borer moths have been related to the oviposition behavior. Malvar et al. (2008) subdivided these compounds in contact phytochemicals and volatile phytochemicals. Within the first category, the concentration of carbohydrates and sugars in the leaves was found to correlate with oviposition preference (Derridj et al. 1986, Derridj et al. 1989). Female moths may thus be able to find the plants offering the best energy sources by sensing carbohydrate content. Further chemical compounds such as pentane, methane and hexane extracts were described as elicitors of oviposition (Udayagiri and Mason 1995, Konstantopoulou et al. 2002, Varshney et al. 2003). Among other functions, these chemicals could help the female moths determine which plants are in the optimal developmental stage for larval establishment. Different volatile phytochemicals were also described as either elicitors or deterrents of oviposition, for example in the study by Konstantopoulou et al. (2004) that extracted volatile oils using steam distillation and characterized their effect on oviposition by the Mediterranean stem borer with two-choice bioassays.

Breeding efforts have mostly been focused on antibiosis and tolerance because the screening of hundreds of different maize genotypes for non-preference in the field is practically unfeasible. For example, it is impossible in the open field to discern whether a plant has not been chosen for oviposition as a consequence of non-preference or just as a matter of chance. Studies targeting non-preference were so far only possible on a small scale in closed environments where the oviposition behavior of female insects can be precisely monitored, for example using two-choice bioassays.

Antibiosis and tolerance

In regions where bi- and multivoltine ECB ecotypes occur, such as in the U.S. Corn Belt, ECB resistance has been further subdivided into resistance to leaf feeding by first generation and resistance to stalk tunneling by second generation ECB (Russell et al. 1974). The first generation of ECB attacks plants during the early growth stages leading to extensive foliar damage. The second generation emerges when the maize plants approach flowering and will mostly feed on stalk and ear tissues. Thus, the damage caused by the second generation in the U.S. Corn Belt is comparable to the damage caused by univoltine ECB in Central Europe. Resistance to leaf feeding and resistance to stalk tunneling rely for the most part on different mechanisms (Guthrie and Russell 1989, Cardinal et al. 2001, Jampatong et al. 2002).

The most extensively-studied mechanism of maize antibiosis to ECB leaf feeding is mediated by the secondary metabolite DIMBOA (2,4-dihidroxy-7-methoxy-1,4-benzoxazin-3-one). This hydroxamic acid is constitutively produced by a number of plant species in the Gramineae family and has demonstrated antibiotic properties against several insect pests (Niemeyer 1988). Genetic loci responsible for DIMBOA biosynthesis have been characterized in maize and variation at these loci has been used successfully in breeding varieties expressing higher DIMBOA contents (Grombacher et al. 1989, Barry and Darrah 1991, Jonczyk et al. 2008, Butrón et al. 2010). DIMBOA concentration is especially high in maize seedlings and decreases rapidly with plant age. Latest at the flowering stage DIMBOA is no longer produced by the plant (Barry et al. 1994, Cambier et al. 2000). In the U.S. Corn Belt this mechanism was thus effective against the first but not against the second ECB generation. Hence, so far, DIMBOA-mediated resistance was of limited practical relevance in Central Europe given that the ECB appears not long before flowering time and nearly all economically relevant damages are induced by ECB stalk tunneling after flowering. Yield losses caused by stalk tunneling are the major concern for maize producers also in regions under bi- and multivoltine ECB ecotypes (Jampatong et al. 2002). However, breeding for resistance to stalk tunneling has been more

challenging than breeding for increased DIMBOA content to prevent leaf feeding by the first ECB generation (Krakowsky et al. 2004).

The extensive literature reviews by Malvar et al. (2008) and McMullen et al. (2009a) described a large number of additional mechanisms of maize resistance to stem borers at both morphological and physiological levels. Among these are also several putative mechanisms effective against ECB stalk tunneling. For example, high concentrations of cell wall components such as hemicellulose, cellulose, and particularly lignin and silica have been related to reduced ECB stalk damage in several studies (Coors 1987, Buendgen et al. 1990, Beeghly et al. 1997, Martin et al. 2004, Cardinal and Lee 2005). Plant cell walls are largely indigestible by herbivore insects so that increased amounts of structural components may be deterrent to ECB larval development (Malvar et al. 2008). In addition, the total amount and chemical composition of cell wall components in the outer part of the stalk (the rind) significantly contributes to stalk strength and thus to the stalk breakage tolerance of maize. A problem in utilizing this resistance mechanism relies in a potential conflict with some of the common goals in a maize breeding program. In particular, varieties used as silage feed need to possess a good cell wall digestibility, which is negatively correlated with lignin and fiber concentrations (Lundvall et al. 1994). Hence, breeders have sought for a long time to reduce rather than to increase the content of cell wall components in their maize germplasm. This mechanism could nevertheless be exploited if the goal is to develop varieties destined to be used exclusively for grain maize production.

The total protein content of stalk tissues was also found to positively correlate with the number of ECB feeding tunnels in the stalk and with the number of surviving larvae (Bergvinson et al. 1997), indicating that low protein content in the stalk may hamper larval development as a consequence of nitrogen deficit in the diet. A further putative constitutive mechanism of antibiosis to stalk feeding damage was discovered by Santiago et al. (2003) in a study on maize resistance to the Mediterranean stem borer *Sesamia nonagrioides* (Lefebvre). Following on the observation that larvae enter the stalks by feeding on the meristematic area situated at the base of the internodes, the authors found that this area was significantly larger in susceptible compared to resistant inbred lines. Santiago et al. (2003) suggested that the size and characteristics of this area could be related to the ability of larvae to penetrate the stalks. A follow-up study by Barros et al. (2011) confirmed the potential usefulness of this trait for increasing stem borer resistance but also found negative correlations with important agronomic traits, suggesting that selection for decreasing the length of the internode meristematic area may have adverse consequences on yield.

Even though most of the known mechanisms affecting maize resistance to stem borers are constitutive plant characteristics, some induced mechanisms which are triggered upon larval feeding have been described as well. Both reviews by Malvar et al. (2008) and McMullen et al. (2009a) reported on the identification of a maize protease inhibitor and a cysteine proteinase (Mir1-CP) that are induced by the maize plant in reaction to insect feeding wounds. The maize protease inhibitor was found to inhibit protease enzymes important for the digestive process (specifically the breakdown of proteins) within the larval midgut of the Lepidopteran cotton pest *Spodoptera littoralis* (Boisduval). On the other hand, Mir1-CP, which is itself a protease enzyme, showed a deleterious effect on the larval midgut protective layer of three different Lepidopteran species (*Spodoptera frugiperda* Smith, *Helicoverpa zea* Boddie and *Danaus plexippus* L.) even though no tangible effect on the mid-gut of the ECB was detected (Mohan et al. 2006).

Indirect resistance mechanisms

Additional induced plant reactions after larval feeding include the so called indirect defense mechanisms. Turlings et al. (1990) first discovered that maize plants emitted a complex mixture of volatile compounds (mostly terpenes) after being wounded by Lepidopteran insects and that these volatiles attracted a natural enemy of the pest: the wasp *Cotesia marginiventris* (Cresson) which is a general parasitoid of noctuid Lepidopteran larvae. A similar interaction mediated by terpene compounds was also observed in response to root feeding by the Western corn rootworm in maize (McMullen et al. 2009a). In this case wounded roots were found to emit a chemical signal based on the compound (E)-β-caryophyllene that attracted the insect parasitic nematode *Heterorhabditis megidis* (Poinar). Rasmann et al. (2005) showed that there is variation among maize genotypes in their ability to emit (E)-β-caryophyllene with a clear difference between maize lines of European and North American origin. Hence, genetic variation for potential use in breeding seems to be available also in the case of indirect maize defense mechanisms.

1.3.2. *Phenotypic evaluation of resistance to ECB stalk damage*

From literature it is known that the potential number of different biologic factors affecting native ECB resistance is very high. Especially in the case of stalk damage, the overall level of resistance observed in the field is most probably the result of a complex interplay between several of these factors while it is well possible that other important mechanisms are still unknown. In any case, a clear understanding of the relative contribution of different resistance mechanisms under field conditions is still missing. Therefore, applied resistance breeding

activities have been commonly performed through expensive and time consuming field trials carried out using at least two field locations under artificial ECB infestation sometimes consolidated by additional locations under natural infestation (Russell et al. 1974, Kreps et al. 1998a, Melchinger et al. 1998a, Orsini et al. 2012). The application of laboratory-reared larvae or egg masses upon thousands of maize plants in the field is very costly but also considered as mandatory in order to guarantee a homogeneous level of infestation across all evaluated plants. An evenly distributed inoculum is an important precondition for a sound statistical analysis and interpretation of field trial results. Moreover, not every year offers the climatic conditions necessary for a general level of natural infestation which is sufficient to effectively screen for resistance, even in regions under high ECB pressure.

The evaluation of stalk damage resistance is performed shortly before harvest. The two most common traits used for this purpose are the cumulative length of ECB feeding tunnels (TL) within the stalks (Guthrie 1989) and a visual stalk damage rating (SDR) score (Hudon and Chiang 1991). SDR is an evaluation of the degree of stalk breakage and is mainly performed at field plot level, *i.e.* across several plants of the same genotype planted in the same physical unit (usually a row-plot). While scoring of SDR is relatively straightforward, TL is an extremely labor intensive trait requiring manual splitting of thousands of maize stalks and measuring of feeding tunnels which are at times difficult to distinguish from intact stalk tissue. But despite the amount of labor required, scoring of TL has been considered essential in ECB resistance studies because it is the only trait that allows a direct evaluation of feeding damage within the plant and thus of the antibiosis resistance component. SDR is additionally influenced by the stalk breakage tolerance of plants (Kreps et al. 1998a).

Given that accurate phenotyping of ECB resistance is highly resource intensive, selection aided by molecular DNA markers (marker-assisted selection, MAS) has been considered a promising approach for a faster and cost-efficient resistance improvement in practical breeding.

1.4. Marker-assisted selection

1.4.1. *History and development of marker-assisted selection*

MAS for polygenic resistance traits in plants has been traditionally based on results obtained from quantitative trait loci (QTL) mapping studies (Bernardo 2008). QTL studies identify a limited number of genetic regions that putatively influence a trait of interest and the general goal of QTL-based MAS is to combine favorable alleles at these specific loci in elite germplasm. The history and principles behind QTL mapping trace back to the very beginnings of quantitative genetics. Nilsson-Ehle (1908) and East (1916) realized that the continuous phenotypic variation observed for many traits (including most traits of agronomic importance in crops) is not conflicting with Mendel's laws of inheritance if one assumes the segregation of several different genetic factors jointly affecting the trait's expression. In the wake of these pioneering works, geneticists have sought to better understand quantitative variation by trying to dissect it into its underlying genetic components. A first study applying the basic concepts of QTL mapping was performed by Sax (1923) long before the discovery of DNA and the availability of molecular markers. This study was able to show a statistical association between the quantitative trait of grain size and Mendelian traits such as grain pigmentation and seed-coat pattern in common bean (*Phaseolus vulgaris* L.). Hence, Sax (1923) was the first to "map" a genetic factor influencing a quantitative trait by showing its genetic linkage with loci responsible for the expression of qualitative traits. These loci that resulted in discrete, easy to discern Mendelian traits fulfilled the role of genetic markers in this seminal study. Later, with the advent of the first DNA-based molecular markers in the 1980's (restriction fragment length polymorphism – RFLP markers) it became possible to construct genetic linkage maps containing a much higher number of markers more or less evenly spaced across the whole genome. With a denser and homogeneous genome coverage achieved by RFLP markers, a systematic approach to QTL mapping was possible. Plant researchers recognized the potential of the new DNA technologies for a better understanding of quantitative genetic variation and a more efficient improvement of agricultural crops (Beckmann and Soller 1986). Hence, fueled by the developments in molecular biology, also the statistical framework for QTL analysis rapidly progressed and soon became more sophisticated. Early QTL studies were based on single marker regression (SMR) approaches (e.g. Soller et al. 1976, Edwards et al. 1987). In SMR each marker is tested individually for its effect on the phenotypic trait by linear regression of the (continuous) trait values on the (discrete) genotypic classes at the specific marker locus. The significance of the marker effect on the trait is tested through a one-way analysis of

variance (ANOVA) with the genotypic classes as factor levels. Thereby QTL positions always coincide with marker positions and significant markers are identified without taking into account the effect of the remaining genetic background. Moreover, multiple-tests for statistical significance at each individual marker increase the risk of detecting false positives if the extent of statistical stringency is not adjusted appropriately. A first methodological improvement was the introduction of interval mapping by Lander and Botstein (1989) which allowed a more accurate estimation of QTL positions. By using this approach it became possible to map the most likely QTL position in an interval between two flanking markers. Lander and Botstein (1989) also proposed to substitute the traditional ANOVA-based inference method with the logarithm of the odds (LOD) statistic which went on to become a standard in QTL analysis up to this day. Composite interval mapping (CIM) introduced by Zeng (1994) has been a further landmark development that significantly increased the accuracy of QTL mapping results. This method combined interval mapping with multiple regression and offered the key advantage of being able to account for the effects of other genetic loci in the genome while estimating the effect of a putative QTL.

A traditional QTL analysis relies on the study of a (possibly) large population obtained from a cross between two inbred parents that differ in the mean expression of the quantitative trait of interest. By using designed biparental crosses it is possible to calculate the recombination frequencies between markers and to estimate their positions in the genome by constructing a genetic map. A relatively recent advance in this field is given by multiple-cross (or joint-population) QTL mapping that exploits the advantages of conducting a combined QTL analysis across multiple interconnected biparental populations (Xu 1998). Both simulation (Rebai and Goffinet 2000, Jannink and Jansen 2001) and empirical studies (Blanc et al. 2006, Steinhoff et al. 2011, Bardol et al. 2013) confirmed the greater benefits of multiple-population over individual-population QTL analyses. They include the possibility to explore more diverse plant material with an overall larger genetic basis and thus to identify an increased number of interesting allelic variants in the frame of a single QTL analysis. In such a combined analysis the effects of QTL alleles originating from different parental genotypes can be readily compared making it possible to identify and eventually introgress or pyramid the most interesting ones into a given genetic background. Multiple-cross analyses were also shown to possess a higher statistical power of QTL detection and higher accuracy of position and effect estimates compared to analyses using individual populations (Rebai and Goffinet 2000, Jansen et al. 2003). This is especially true if the biparental populations under study are connected by one or several common parental lines, thereby reducing the number of allelic effects the QTL model

needs to estimate (Bardol et al. 2013). Finally, connected biparental populations represent the standard type of plant material typically found in breeding programs, where a few elite inbred lines may be routinely employed for producing a number of different crosses. Thus, multiple-cross QTL analyses should be well adapted to the needs of breeding programs and could help fostering the application of MAS in practice.

Depending on the number, percentage of explained variance and allelic effects of the QTL identified for a given trait, different MAS strategies are needed as summarized in Figure 2 (Bernardo 2008).

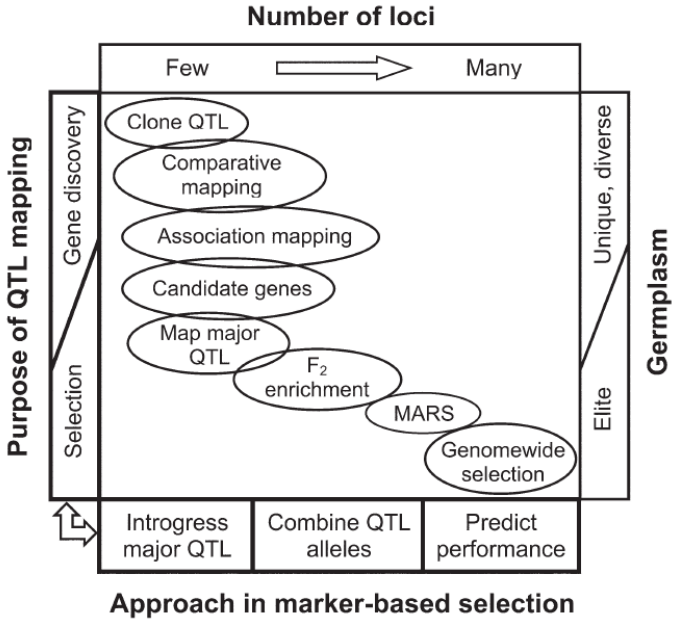

Figure 2 Different MAS strategies in plants depending on the purpose and number of genetic loci affecting the trait of interest. Source: Bernardo (2008)

For traits controlled by a few major QTL with large effects, the most appropriate breeding strategy is to introgress the favorable alleles at these QTL into an elite genetic background, *i.e.* by means of marker-assisted backcrossing (MABC). If, on the opposite, a trait is highly polygenic and several QTL having small to moderate effects are detected, the directed introgression of all favorable alleles into a single elite genotype is nearly impossible, even assuming the absence of common challenges such as (i) the inconsistency of small QTL effects across environments and/or genetic backgrounds and (ii) linkage with alleles that negatively affect other important traits. The recommended procedure for highly polygenic traits is rather to increase the frequency of favorable alleles at multiple loci using a marker-assisted recurrent selection (MARS) breeding strategy (Figure 2, Bernardo 2008). A classical MARS

implementation starts with a QTL detection step performed in a given population (either bi- or multiparental) which is both phenotyped and genotyped. Selected genotypes out of this population are recombined and the resulting progeny is genotyped in order to perform selection based on marker information only. A molecular score is calculated for each of the genotypes using the allelic information at putative QTL positions and the effects of the QTL estimated in the previous generation. Selection is then performed by ranking the genotypes according to their molecular score. Hence, a MARS strategy relies more on *prediction* of total performance based on marker data rather than on the *identification* and *introgression* of specific genomic regions (Figure 2, Bernardo 2008).

This consideration serves as an introduction for the newest development of MAS for quantitative traits in plant breeding: the application of genome-wide selection. This method was first proposed by Meuwissen et al. (2001) fostered by the realization that genotyping individuals with thousands of single nucleotide polymorphism (SNP) markers by means of DNA-arrays has become affordable on a routine basis. Meuwissen et al. (2001) introduced statistical models able to simultaneously exploit the information provided by dense, genome-wide SNP markers for predicting the total genetic merit of individuals (genome-wide prediction, GP). In contrast to MAS based on QTL analysis, GP can be viewed as a form of MAS that does not involve testing for the significance of marker effects (Bernardo and Yu 2007). In this way one of the major drawbacks of QTL analysis can be avoided: the imposition of predetermined significance thresholds. When adopting a stringent significance threshold many loci having small to moderate effects on a quantitative trait will remain undetected. A relaxed threshold, on the other hand, bears a high risk of detecting false positives without ensuring that all loci with a true effect on the trait will be identified. In principle GP is more straightforward than QTL mapping and more oriented towards practical applications in breeding. Its main aim is not to dissect or understand the genetic architecture of quantitative traits but to select the most promising individuals based on genetic values predicted from their DNA profile. The method has already found wide application in animals and particularly in dairy cattle breeding (Hayes et al. 2009a) because in this field the impact of GP on the speed and cost-efficiency of the breeding process is large. Here, the DNA profile of newborn calves is used to predict their future genetic merit as bulls that could be employed for artificial insemination. In recent years the potential of GP has been studied extensively for plant breeding as well (see e.g. reviews by Jannink et al. 2010, Lorenz et al. 2011, Daetwyler et al. 2013, de los Campos et al. 2013, Jonas and de Koning 2013) so that by now its use for a faster and more efficient genetic improvement of crops is considered highly promising and applications in practical breeding programs are under way.

GP is based on a linear regression of phenotypic values on all available SNP markers. The main difference with the statistical models used in QTL analysis (where only a small number of markers or QTL are included as cofactors in the model) is that in GP the number of marker cofactors (p) can be much larger than the number of phenotypic observations (n). Due to this so called "large-p, small-n" problem, marker effects cannot be estimated using ordinary least squares. Procedures involving variable selection or shrinkage estimation are necessary (de los Campos et al. 2013). Several statistical methods have been proposed for GP that mainly differ in the extent of regularization and variable selection (Wimmer et al. 2013). The so far most widely adopted methods rely on the statistical framework of best linear unbiased prediction (BLUP) which has been the standard method for breeding value estimation in cattle breeding for the last three decades (Henderson 1984). This approach was originally developed in order to allow the use of pedigree information in the estimation of animal breeding values. This was achieved by including a so called relationship matrix calculated from pedigree data that provided the variance-covariance structure for the linear mixed model. For GP two statistical models have been developed: ridge-regression best linear unbiased prediction (RR-BLUP) and genomic best linear unbiased prediction (GBLUP). RR-BLUP directly fits individual SNP markers as cofactors in the linear mixed model. The solution provided by RR-BLUP for the "large-p, small-n" problem is to fit SNP markers as random effects while assuming that effects are drawn from an underlying normal distribution. In GBLUP, on the other hand, genome-wide SNP data are used to construct a genomic relationship matrix for the tested genotypes (Habier et al. 2007). This matrix is then used in the same way as the relationship matrix based on pedigree data from traditional BLUP. Hence, in GBLUP the breeding values (and not the markers) are fitted as random effects (Goddard et al. 2009) thus effectively reducing the number of cofactors in the linear mixed model to the number of genotypes. GBLUP was shown to be computationally more efficient than RR-BLUP due to the reduction in dimensionality (Goddard 2009, Hayes et al. 2009b). Both RR-BLUP and GBLUP assume that markers are evenly spread across the genome and therefore their application is considered to be most appropriate for highly polygenic traits controlled by many loci with small effects. In contrast, other statistical approaches used in GP such as the least absolute shrinkage and selection operator – LASSO (Tibshirani 1996) or the elastic net (Zou and Hastie 2005) implement variable selection and are assumed to be more promising for traits characterized by a simpler genetic architecture (Lorenz et al. 2011). The final maximum number of informative SNP markers allowed in the LASSO procedure is equivalent to the number of genotypes. Hence, among GP models the LASSO is known for performing a rather strict form of variable selection. The elastic net can be considered

as intermediate between RR-BLUP and LASSO as it combines the properties of both methods and makes it possible to retain more informative markers in the model than the number of genotypes (Wimmer et al. 2013). Additional statistical models for GP were developed using the framework of Bayesian statistics and include the so called BayesA and BayesB methods (Meuwissen et al. 2001). These models perform variable selection by applying different degrees of shrinkage to marker effects, thus giving more weight to SNPs linked to large-effect QTL. Depending on factors such as genetic architecture, extent of linkage disequilibrium (LD) and trait heritability different models can yield better prediction accuracies. However, according to studies performed on experimental data, RR-BLUP and GBLUP were found to be very robust across different situations (de los Campos et al. 2013).

The general procedure for implementing GP in a plant breeding program is very similar to the above-described QTL-based MARS strategy. First, a population is needed which is genotyped and phenotyped in order to estimate SNP effects, or in a more general way to calibrate the GP models. This population or sample of genotypes is most commonly referred to as training or calibration set. In a second step, genome-based predictions are computed for genotypes of another sample which is only genotyped (e.g. progeny of selected genotypes out of the training set). The sample on which predictions are performed is usually referred to as the prediction or validation set.

1.4.2. *Potential of MAS for ECB resistance improvement*

Traditional QTL-based MAS has been used in resistance breeding against diseases and pests. However, while MAS has been generally successful when applied to simple mono- or oligogenic resistances, the marker-based improvement of many important polygenic resistance traits has posed significant challenges (Miedaner and Korzun 2012). A major difficulty faced by the QTL-based MAS approach has been the instability of the detected resistance QTL across environments and genetic backgrounds. Therefore, despite the very large number of published QTL mapping studies focusing on quantitative resistance traits across different crop species, far fewer reports demonstrated a successful application of MAS in practical breeding (Bernardo 2008, St.Clair 2010). An additional reason for the shortage of reports on MAS is the fact that private breeding companies are generally not publishing detailed results because of intellectual property issues (St.Clair 2010), although they may have been successfully implementing MAS on a large scale and for a number of traits since the beginning of the last decade (Eathington et al. 2007).

Published examples of QTL-based MAS application in breeding for polygenic resistance have been reviewed by Bernardo (2008) and St.Clair (2010). A well-known case is the improvement of *Fusarium* head blight resistance in wheat mainly caused by *Fusarium graminearum* (Schwabe) (Anderson et al. 2008). A few major QTL were identified and the *Fhb1* locus in particular is now used routinely in wheat breeding programs (Miedaner and Korzun 2012). Other examples are the improvement of resistance against (i) the soybean cyst nematode (*Heterodera glycines* Ichinohe) in soybean (*Glycine max* L.) (Cahill and Schmidt 2004), and (ii) three different diseases in common bean: white mold (*Sclerotinia sclerotiorium* Lib. de Bary), bacterial blight (*Xanthomonas axonopodis* pv. *phaseoli* Smith) and bean common mosaic virus (Miklas et al. 2006). A few further cases of successful MAS implementation were described by St.Clair (2010) which may not have been applied on a larger breeding scale yet. These include resistance to (i) late blight (*Phytophthora infestans* Mont. de Bary) in tomato (*Solanum lycopersicum* L.), (ii) leaf rust (*Puccinia hordei* Otth) and stripe rust (*Puccinia striiformis* f. sp. *hordei* Westend) in barley (*Hordeum vulgare* L.), and (iii) root and stem rot (*Phytophthora capsici* Leonian) in pepper (*Capsicum ssp.* L.). The common denominator among these examples is the identification of a few major QTL having effects sufficiently large and stable to be effectively introgressed or pyramided in different genetic backgrounds by means of MABC.

For resistance to ECB stalk damage a number of QTL mapping studies were performed during the last twenty years, especially in the USA (Schön et al. 1993, Cardinal et al. 2001, Jampatong et al. 2002, Krakowsky et al. 2007, Orsini et al. 2012). Genetic variation for ECB resistance was identified in Central European germplasm as well (Melchinger et al. 1998a) allowing the development of QTL mapping populations (Bohn et al. 2000, Papst et al. 2001, Papst et al. 2004). QTL experiments were also conducted for stalk tunneling resistance to the Mediterranean corn borer in Spain (Ordas et al. 2010). All studies revealed a rather complex genetic architecture of the respective resistance traits. Willcox et al. (2002) and Flint-Garcia et al. (2003) performed MAS studies to evaluate the prospects of utilizing QTL mapping results in resistance breeding against maize stem borers. Willcox et al. (2002) focused on resistance to leaf feeding by the first generation of the Southwestern corn borer (*Diatraea grandiosella* Dyar). This study showed that MABC for the introgression of three QTL into an elite line was as effective as phenotypic selection in increasing resistance. Leaf feeding resistance is supposed to be genetically less complex than resistance to stalk damage (see section 1.3.1) and this could explain the success of a MABC strategy. On the other hand, Flint-Garcia et al. (2003) targeted resistance to stalk tunneling by the ECB and chose a recurrent selection strategy for comparing

MARS with phenotypic selection. The positive results obtained in this study confirmed that MAS could be promising also for the improvement of ECB stalk damage resistance. However, to my knowledge, routine implementation of this approach in practical breeding programs is limited.

Empirical studies across different crop species have shown that prediction accuracy of highly polygenic traits based on GP-models was consistently higher than prediction accuracy using traditional MARS approaches based on QTL detection (Lorenzana and Bernardo 2009, Heffner et al. 2011, Massman et al. 2013). Hence, application of GP-based MAS could have profound consequences not only for the improvement of complex agronomic traits such as yield, but also in breeding for quantitative resistances to diseases and pests. First studies of GP applied to polygenic diseases in plants are encouraging. Technow et al. (2013), for example, reported GP accuracies of up to 0.70 for Northern corn leaf blight (*Exserohilum turcicum* Pass.) resistance in maize, while Rutkoski et al. (2012) evaluated the prospects of GP for improving *Fusarium* head blight resistance in North American wheat, concluding that germplasm from different origins could be used to train accurate prediction models. Thus, it seems that GP can be applied successfully also to traits possessing large-effect QTL already exploited in traditional MAS. Indeed, GP showed the ability to adequately capture large effect QTL and additionally cover the remaining genome-wide effects in a single statistical model (Wimmer et al. 2013). Hence, given the quantitative nature of stalk damage resistance it can be hypothesized that a GP approach should be more effective than QTL-based MAS for the improvement of ECB stalk damage resistance in Central European maize.

1.5. Objectives of the study

The main goal of the present study was to evaluate the potential of GP relative to MAS based on QTL mapping for the improvement of resistance to ECB stalk damage in Central European elite maize material.

The detailed study objectives were:

(i) Estimate quantitative genetic parameters of agronomic and ECB stalk damage traits within and across three connected doubled haploid (DH) populations of locally adapted elite maize at both the DH line *per se* and testcross level.

(ii) Perform individual- and joint-population QTL analyses of agronomic and ECB stalk damage traits at both the DH line *per se* and testcross level.

(iii) Estimate and compare cross-validated predictive abilities of genome-wide and QTL-based prediction models for agronomic and ECB stalk damage traits within each individual population and in a joint-population prediction scenario.

(iv) Estimate the performance of GP models in prediction of ECB stalk damage traits *across* biparental populations.

(v) Evaluate the accuracy of GP for predicting ECB stalk damage in testcrosses based on models trained on phenotypes observed at the DH line *per se* level.

(vi) Evaluate the extent of genotype-by-environment interaction for ECB stalk damage resistance in a GP framework by estimating the accuracy of prediction *across* locations (involving both artificially and naturally infested locations). Following this, evaluate the prospects of reducing the number of screening locations and/or the costs of artificial ECB infestation needed for the training of GP models.

The findings obtained in this study should allow the definition of optimized marker-based breeding strategies for the genetic improvement of native ECB resistance in elite maize.

2. MATERIALS AND METHODS

2.1. Plant material

Four elite early maturing inbred lines, originating from the Central European dent heterotic pool, were used as parents of the DH populations evaluated herein. Three inbred lines selected for resistance to ECB stalk damage in observation trials under high ECB pressure (R1, R2, R3) were crossed to the same susceptible line (S1) for the development of three connected populations: Pop1 (R1 × S1), Pop2 (R2 × S1) and Pop3 (R3 × S1) comprising 85, 243 and 262 DH lines, respectively. DH lines were developed by KWS SAAT AG (Einbeck, Germany) using *in vivo* haploid induction technology (Röber et al. 2005). F_1 plants from each of the three crosses were pollinated with pollen of an inducer line. Haploid seeds were identified by means of a dominant anthocyanin marker gene carried by the inducer line. Chromosomes of haploid seedlings were doubled with a colchicine treatment. The success of DH line production varied for the three populations, thereby leading to differences in sample size. In the winter nursery 2011/2012, testcrosses were produced with a subset of DH lines by crossing each DH line with an ECB susceptible tester from the flint heterotic pool. All genetic material is proprietary to and was supplied by KWS SAAT AG.

2.2. Field experiments

Field trials were conducted during 2011 and 2012 at two locations in Southern Germany under artificial ECB infestation conditions and under natural infestation conditions at four locations in Germany, France and Italy characterized by high ECB pressure (Table 1). In 2011 the three populations were evaluated for agronomic and ECB stalk damage traits as DH lines *per se*. Based on these results a bidirectional selection was performed, where 25 % of the most resistant and 10 % of the most susceptible DH lines were selected in each population under the side condition of similar flowering time. The selected 31 (Pop1), 85 (Pop2) and 91 (Pop3) DH lines were evaluated in 2012 as lines *per se* and as testcrosses in separate trials at each location. In both years parental lines and their testcrosses were included as duplicate entries in the respective trials. All experiments were sown in single-row plots of 20 plants in a 40 × 15 alpha-lattice design with two replications in 2011 and a 31 × 7 alpha-lattice design with two replications in 2012. An exception was the trial at Ferrara (Italy) in 2012 which was sown in four replications

in order to exploit the extreme level of natural ECB pressure typical for this site. At location Münzesheim (Germany) both artificially (MZH_i) and naturally infested (MZH_n) trials were conducted and the two experiments were considered as different environments in the statistical analysis. Field trials were managed according to standard maize cultivation practices for the respective growing regions with the exception that no chemical or biological insect control was carried out.

Table 1 Information on field trial environments and on traits scored.

Year	Location	Abbreviation	Replications	ECB treatment	Traits scored[1]
2011	Freising (D)	FRS	2	Artificial	SDR, TL, NT, ANT, SIL, PH
2011	Münzesheim (D)	MZH_i	2	Artificial	SDR, *TL*, *NT*
2011	Münzesheim (D)	MZH_n	2	Natural	SDR, *ANT*, *SIL*
2011	Herbolzheim (D)	HER	2	Natural	SDR
2011	Tournoisis (F)	TOS	2	Natural	SDR, ANT, SIL
2011	Ferrara (I)	FER	2	Natural	SDR
2012	Freising (D)	FRS	2	Artificial	SDR, TL, NT, ANT, SIL, PH, EH, BRIX
2012	Münzesheim (D)	MZH_i	2	Artificial	SDR, TL, NT
2012	Münzesheim (D)	MZH_n	2	Natural	SDR, ANT, SIL, BRIX
2012	Herbolzheim (D)	HER	2	Natural	SDR
2012	Bernburg (D)	BBG	2	Natural	SDR, *ANT*, *SIL*, PH, EH
2012	Ferrara (I)	FER	4	Natural	SDR

[1]Traits are stalk damage rating (SDR), tunnel length (TL), number of tunnels (NT), days to anthesis (ANT), days to silking (SIL), plant height (PH), ear height (EH), and stalk sucrose concentration (BRIX). Traits that were scored in only one replication at a specific field site are marked in italics.

2.3. Artificial ECB infestation

In the two artificially infested locations ECB was applied at three time points in 1-2 week intervals. At each time point approximately 20 neonate larvae were placed in the whorl or leaf collar of 10 adjacent plants per plot starting from the third plant in the respective row plot. The first infestation date was synchronized with the occurrence of natural ECB moth flight which usually coincided with the mid-whorl stage of the plants (end of June to beginning of July). The progression of moth flight was followed using the ECB monitoring services provided by the Bavarian State Research Center for Agriculture (Zellner and Lechermann 2014) for location Freising and the Landwirtschaftliches Technologiezentrum Augustenberg (LTZ 2014) for location Münzesheim. The last infestation was carried out shortly before or during flowering time (end of July).

ECB egg masses were obtained from the entomology unit of Dr. P. Aupinel, INRA Le Magneraud, France. The egg masses were incubated for 2-3 days in growth chambers at 80 %

relative humidity and temperatures between 15 and 30 °C depending on the developmental stage at delivery time. Few hours after hatching the larvae were mixed in corncob grits (EU-GRITS 20, W. Freiherr von Haxthausen, Lichtenau, Germany) and applied to plants using specific volume dispensers (Mihm 1983).

2.4. Evaluation of traits

2.4.1. *Agronomic traits*

Days to anthesis (ANT) and days to silking (SIL) were recorded as the number of days between sowing and the date when 50 % of the plants in a plot had visible anthers or silks, respectively. ANT and SIL were scored in 3 locations in both years (Table 1) at 2-3 day intervals during the flowering period. Plant height (PH) was assessed after flowering at location Freising in 2011 and additionally at location Bernburg in 2012 (Table 1). PH was estimated across the complete field plot using a metric bar with a 5 cm interval scale by measuring the distance between soil level and the average plot level of the lowest tassel branches. Ear height (EH) was assessed simultaneously with PH at locations Freising and Bernburg in 2012. The distance in cm between soil level and the attachment point of the ear shank was measured for three representative plants in each plot and plot averages across the three plants were successively calculated for further analysis.

2.4.2. *ECB stalk damage traits*

Stalk damage rating (SDR) was used to evaluate the tolerance of plants to stalk breakage under ECB pressure, and was scored in all six locations in both years (Table 1). SDR is based on a 1-9 rating scale adapted according to Hudon and Chiang (1991) where the score of 1 is given for intact plants, 2-3 for broken tassels, 4-8 for stalk breakage above the ear whereby each higher score corresponds to breakage within a correspondingly lower internode, up to the highest score of 9 for plants broken at or below the ear (Figure 3). Scoring was performed on a plot basis by estimating a visual mean across all plants in a plot in naturally infested locations and across the ten infested plants in artificially infested locations, respectively. SDR was assessed during the period preceding harvest (from end of August in Italy to mid-October in Germany). After the final SDR scoring, resistance to ECB stalk tunneling was evaluated in the two artificially infested locations by recording the number of feeding tunnels in the stalk (NT) and the cumulative tunnel length in centimeters (TL). To score NT and TL, the stalks of five infested plants per plot were split from the bottom to the eighth internode above ground using a specific

apparatus developed by Kreps et al. (1998b). For both traits, plot averages were calculated across the five plants.

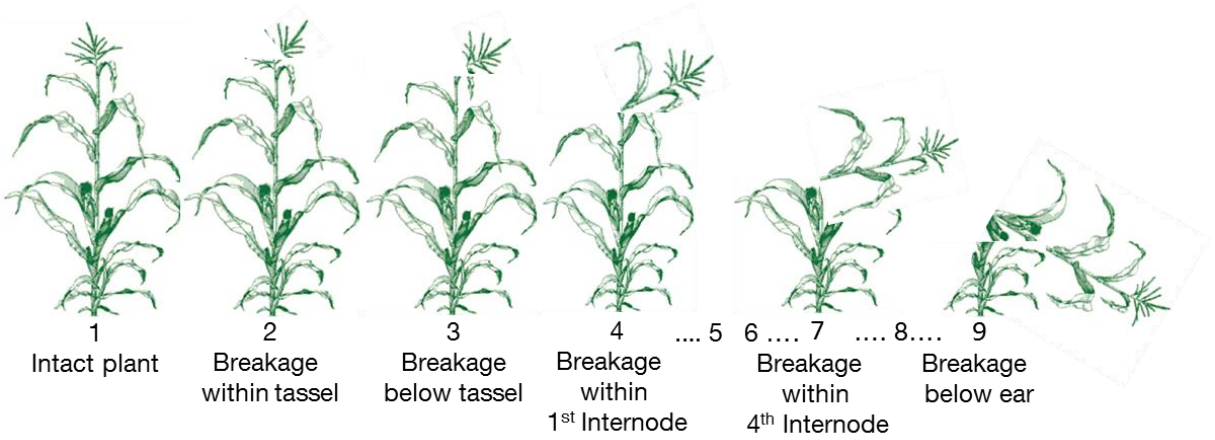

Figure 3 Illustration of the 1-9 scale used in stalk damage rating (SDR).

2.4.3. *Stalk sugar concentration (BRIX)*

To investigate the potential effect of stalk sugar content on ECB resistance in the period of larval establishment, sucrose concentration in the stalk (BRIX) was assessed for the DH lines *per se* at locations Freising and Münzesheim in 2012 (Table 1). During the flowering period the stalks of four adjacent plants per plot were destructively sampled by cutting a stalk portion of approximately 5 cm of length above the attachment point of the ear shank. At each location, all plants were sampled on the same day during a 2-4 hour time period. Immediately after sampling the stalk portions were put in sealed plastic bags and stored in cooling boxes to avoid drying and cellular respiration of sugars prior to analysis. The plant material was successively stored in refrigerators at + 5 °C for the time necessary for measurements which took place during 1-2 days after field sampling. Stalk sucrose concentration was measured in degrees °Brix by squeezing a few drops of stalk juice on an electronic handheld Brix refractometer (Krüss model DR201-95, Krüss Optronic GmbH, Hamburg, Germany). Individual BRIX measurements of the four plants in each field plot were averaged for further analysis.

2.5. Marker analysis and linkage maps

DH populations (N = 590) and their parents were genotyped with 4,790 SNP markers evenly distributed across the genome using a custom Illumina Infinium SNP array (Illumina Inc., San Diego California, USA). The SNPs represented a subsample of the Illumina MaizeSNP50 BeadChip (Ganal et al. 2011). The physical positions of the markers were assigned based on the B73 RefGen_v2 sequence (Schnable et al. 2009). The marker-based genetic distance between the four parental lines was calculated with the full set of markers using the Rogers distance (Rogers 1972). The distribution of polymorphic SNP markers along the genome was investigated for each population through pairwise comparisons between the marker alleles of the common parental line (S1) and the corresponding alleles of each one of the other three parental lines (R1, R2, R3). This analysis was performed based on the physical position of SNP markers using a sliding window approach with the R program (R Development Core Team 2013) with a window size of 30 Mbp and an incrementing step of 3 Mbp.

For further analysis a subset of high quality polymorphic SNPs was selected in each population out of the 4,790 SNP markers according to following criteria: (i) a call rate higher than 0.90, (ii) a minor allele frequency higher than 0.05 and (iii) less than 10 % missing values. After these selection steps 2,411 SNPs were retained. DH lines with more than 20 % missing data in these 2,411 SNPs (4, 29 and 36 DH lines in Pop1, Pop2 and Pop3, respectively) were discarded leaving a total of N = 521 DH lines for further analyses. For each marker, deviations from the expected segregation ratio were tested individually for each population with a chi-square test using the sequentially rejective Holm-Bonferroni method to account for multiple testing (Holm 1979). For the investigation of patterns of variation at the molecular level, among and within populations, a principal component analysis on polymorphic SNP data was performed in R.

Linkage maps were constructed for each population individually using a maximum likelihood mapping approach and Haldane's mapping function (Haldane 1919). To allow for the construction of a genetic consensus map across the three populations, clusters of SNP markers colocalising in each of the three individual genetic maps were identified. For each of these clusters only one marker that was polymorphic in all populations was retained. The consensus map was calculated using 1,034 SNPs, which were polymorphic in at least one of the three populations. All linkage maps were constructed with JoinMap version 4.1 software (Van Ooijen 2006).

2.6. Statistical analyses of phenotypic data

Lattice analyses of variance at individual field environments were performed using PLABSTAT version 3A software (Utz 2011) according to the statistical model:

$$y_{ijk} = \mu + g_i + r_k + b_{km} + e_{ikm} \quad (1)$$

y_{ijk} trait observation

μ overall mean

g_i random effect of genotype i

r_k random effect of replication k

b_{km} random effect of incomplete block m nested within replication k

e_{ikm} random residual error; $e_{ikm}\, iid \sim \mathcal{N}(0, \sigma_e^2)$

Outlying observations were identified by means of residual diagnostic plots in R. Each outlier was considered individually and removed only under evidence of erroneous scoring or obvious problems in the specific field plot. The single-plot repeatability (*rep*) of traits for individual field environments was estimated as following:

$$rep = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2} \quad (2)$$

where $\hat{\sigma}_g^2$ and $\hat{\sigma}_e^2$ are the estimates of the genotypic and error variance components pertaining to model (1), respectively.

The analysis across locations and within years was based on the following linear mixed model implemented in the ASReml-R software package (Butler et al. 2009):

$$y_{ijkm} = \mu + g_i + l_j + gl_{ij} + r_{kj} + b_{mkj} + e_{ijkm} \quad (3)$$

y_{ijkm} trait observation

μ overall mean

g_i random effect of genotype i

l_j fixed effect of location j

gl_{ij} random interaction effect of genotype i with location j

r_{kj} random effect of replication k nested within location j

b_{mkj} random effect of incomplete block *m* nested within replication *k* nested within location *j*

e_{ijkm} random residual error; $e_{ijkm}\, iid \sim \mathcal{N}(0, \sigma_e^2)$

Variance components within the three populations were estimated by introducing a categorical variable in the model assigning DH lines to their respective populations using the at( ) function for conditional factors in ASReml-R. Estimates of heritability ($\hat{h}^2$) on a progeny mean basis were calculated according to Hallauer and Miranda (1981):

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \frac{\hat{\sigma}_{gl}^2}{L} + \frac{\hat{\sigma}_e^2}{L \times R}} \tag{4}$$

$\hat{\sigma}_g^2$ estimate for the genotypic variance component

$\hat{\sigma}_{gl}^2$ estimate for the genotype-by-location variance component

$\hat{\sigma}_e^2$ estimate for the residual error variance component

L number of locations in which a given trait was evaluated

R number of replications per location

In cases where the number of replications for a given trait was different across locations, *R* was calculated as the harmonic mean of the number of replications at each location:

$$R = \frac{L}{\sum_{j=1}^{L} \frac{1}{R_j}} \tag{5}$$

where R_j is the number of replications at location *j*.

To obtain best linear unbiased estimates (BLUEs) for each genotype and trait, model (3) was fitted with genotypes as fixed effects. The BLUEs averaged across locations represented the phenotypic input data for QTL and genome-wide prediction analyses. To estimate the statistical significance of differences between the BLUEs of individual genotypes, the least significant difference (LSD) at the 0.05 significance level was calculated according to Bernardo (2010):

$$LSD_{0.05} = t(0.025)\sqrt{2\left(\frac{\hat{\sigma}_{gl}^2}{L} + \frac{\hat{\sigma}_e^2}{L \times R}\right)} \tag{6}$$

where *t* (0.025) is the 2.5 % quantile of the t-distribution with *(N – 1)(L –1)* degrees of freedom and where *N* is the number of genotypes.

Components of covariance between traits that were evaluated on the same experimental unit were estimated by expanding (3) to a bivariate model. To estimate variance and covariance components between two given traits, only data from the locations where both traits were assessed were fitted in the model. Covariance components between DH lines *per se* and testcrosses were estimated using a two-stage analysis. This approach was necessary since the DH lines *per se* and their corresponding testcrosses were evaluated in separate trials at each location. Here, BLUEs of DH lines *per se* and testcrosses were first calculated for individual locations and were then used as entry data for a bivariate model across locations. Genotypic correlations between traits as well as between DH lines *per se* and testcrosses were extracted from the bivariate variance and covariance components according to Mode and Robinson (1959).

Approximate standard errors of heritabilities and genotypic correlations were calculated based on the standard errors of the underlying components by applying the delta method as suggested by Holland et al. (2003). Pearson's phenotypic correlations between traits and between DH line *per se* and testcross performance were calculated based on BLUEs averaged across locations.

2.7. QTL mapping

For the traits SDR, TL, NT, and ANT three different QTL analyses were carried out using phenotypic data from (i) the unselected populations evaluated as DH lines *per se* in 2011 (*N* = 521), (ii) the selected DH lines *per se* evaluated in 2012 (*N* = 195), and (iii) the corresponding selected testcrosses evaluated in 2012 (*N* = 195). For the unselected DH populations evaluated in 2011, QTL analyses were performed both within the individual biparental populations (hereafter referred to as *within-population* QTL analyses) and combined across all three populations (hereafter referred to as *joint-population* QTL analyses). For the selected DH lines *per se* and their testcrosses evaluated in 2012 only *joint-population* analyses were performed due to the small sample sizes within each selected population.

All analyses were performed based on the genetic consensus map using the software package MCQTL (Jourjon et al. 2005), a program specifically designed for the analysis of multi-parental crosses. Genotypic probabilities of parental alleles were computed at each marker position and, when the distance between two neighboring markers exceeded 2 cM, additionally at intervals

of 2 cM by using information on SNP genotypes at the flanking markers. For each trait and population (including the *joint-population* scenarios) empirical LOD thresholds at the 0.05 genome-wide significance level were assessed from 1,000 permutations according to Churchill and Doerge (1994). QTL detection was carried out using the method iQTLm (Charcosset et al. 2001) which is an iterative extension of composite interval mapping (CIM, Zeng 1994). First, markers associated with the trait were identified and selected as cofactors using a forward stepwise regression model under the predefined condition of a minimal distance of 10 cM between two cofactors. The identified cofactors were then employed for QTL detection using CIM. A chromosome-wise iterative process was applied to refine QTL positions whereby the identified QTL were used as cofactors in a new round of CIM until all QTL positions on a given chromosome reached convergence (Larièpe et al. 2012). LOD support intervals of QTL positions were defined as the map distance in cM spanning a LOD drop of one unit on each side of the LOD peak. QTL for different traits were defined as colocalizing if their respective LOD support intervals overlapped. The sum of allelic effects at each QTL was constrained to be zero in both the *within-* and *joint-population* analyses.

Formulas for the QTL models displayed herein are given following the notation of Bardol et al. (2013). *Within-population* analyses were based on the following additive statistical model:

$$\mathbf{y}_p = \mathbf{1}_{N_p}\mu_p + \mathbf{X}_{q_p}\mathbf{a}_{q_p} + \sum_{c \neq q}\left(\mathbf{X}_{c_p}\mathbf{a}_{c_p}\right) + \mathbf{e}_p \qquad (7)$$

$\mathbf{y}_p$ $N_p \times 1$ vector of BLUEs of the N_p genotypes from population p

$\mathbf{1}_{N_p}$ $N_p \times 1$ vector of ones

μ_p mean of population p

$\mathbf{X}_{q_p}, \mathbf{X}_{c_p}$ $N_p \times 2$ matrices assigning genotypic probabilities of the two parental alleles of population p at QTL q and cofactor c, respectively

$\mathbf{a}_{q_p}, \mathbf{a}_{c_p}$ vectors of length 2 containing additive effects of the two parental alleles at QTL q and cofactor c, respectively

$\mathbf{e}_p$ $N_p \times 1$ vector of residuals; $\mathbf{e}_p \sim \mathcal{N}(0, \mathbf{I}\sigma^2_{e_p})$

In the *joint-population* analysis the following connected QTL model was implemented:

$$\mathbf{y} = \mathbf{J}\boldsymbol{\mu} + \mathbf{X}_q\mathbf{a}_q + \sum_{c \neq q}\left(\mathbf{X}_c\mathbf{a}_c\right) + \mathbf{e} \quad (8)$$

$\mathbf{y}$	$N \times 1$ vector of BLUEs of the N genotypes from populations Pop1, Pop2 and Pop3
$\mathbf{J}$	$N \times 3$ design matrix assigning genotypes to their respective population
$\boldsymbol{\mu}$	3×1 vector containing population means of Pop1, Pop2 and Pop3
$\mathbf{X}_q$, $\mathbf{X}_c$	$N \times 4$ matrices assigning genotypic probabilities of the four parental alleles at QTL q and cofactor c, respectively
$\mathbf{a}_q$, $\mathbf{a}_c$	4×1 vectors containing additive effects of the four parental alleles at QTL q and cofactor c, respectively
$\mathbf{e}$	$N \times 1$ vector of residuals; $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$

In our specific case model (8) estimated four allelic effects at each QTL (effects of R1, R2, R3, and S1 parental alleles) in which the effect of the common parent S1 is assumed to be the same in all three populations. The proportion of variance explained by the respective models fitting all detected QTL simultaneously (R^2) as well as the proportion of variance explained by individual QTL was calculated as described in the MCQTL software manual (Mangin et al. 2010).

2.8. Genome-wide prediction (GP)

All polymorphic SNP markers meeting quality criteria (N = 2,411) were used in genome-wide prediction (GP) of the traits SDR, TL, NT, and ANT. As in QTL mapping, GP analyses were performed both within individual biparental populations (hereafter referred to as *within-population prediction*) and combined across populations (hereafter referred to as *joint-population prediction*). Marker genotypes were coded as 0 or 2 according to the number of copies of the minor allele and missing marker genotypes were imputed using the BEAGLE software package (Browning and Browning 2009). A genomic best linear unbiased prediction (GBLUP) model was used to predict genetic values of DH lines. The general specification of

the GBLUP model that served as the basis for all GP scenarios analyzed in this study is given by:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \quad (9)$$

y	$N \times 1$ vector of BLUEs of N genotypes included in a given prediction scenario
X, **Z**	design matrices for the fixed and random effects in the model, respectively
β	$B \times 1$ vector of B fixed effects
u	$N \times 1$ vector of random genotype effects; $\mathbf{u} \sim \mathcal{N}(0, \mathbf{G}\sigma_u^2)$
e	$N \times 1$ vector of residuals; $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$

The vector of genotypic effects **u** follows the normal distribution $\mathbf{u} \sim \mathcal{N}(0, \mathbf{G}\sigma_u^2)$ where σ_u^2 is the genotypic variance component pertaining to model (9) and **G** is the $N \times N$ realized relationship matrix between the N genotypes included in the model. The **G** matrix was computed based on marker data of DH lines according to the method proposed by Habier et al. (2007) and VanRaden (2008):

$$\mathbf{G} = \frac{(\mathbf{W} - \mathbf{P})(\mathbf{W} - \mathbf{P})'}{2\sum_{m=1}^{M} p_m (1 - p_m)} \quad (10)$$

W	$N \times M$ design matrix assigning SNP marker genotypes of M markers to N DH lines
P	$N \times M$ matrix with the m-th column containing the expected genotype score at marker locus m, which is calculated as $2p_m$
p_m	minor allele frequency at marker locus m

The basic GBLUP model was fitted including the overall mean across the N genotypes as fixed effect in **Xβ**. The present study additionally explored whether the inclusion of information obtained from QTL analysis could increase the prediction accuracy of GP. To this purpose an augmented form of GBLUP (hereafter referred to as GBLUP+QTL) was implemented by integrating information on the QTL that were detected using the methods described in section 2.7. The SNP marker genotypes at the nearest flanking markers of each of the detected QTL

were included in the design matrix **X** under (9). Thereby the nearest QTL-flanking markers were fitted as fixed effects in GBLUP in addition to the overall mean. In this way predetermined information on QTL positions (but *not* on QTL effects) was integrated in the GBLUP analysis. Fixed QTL effects were re-estimated in the GBLUP model (9). All GP analyses were performed using the genomic prediction framework provided by the *synbreed* R package (Wimmer et al. 2012).

2.9. Cross-validation for QTL-based and genome-wide prediction models

Figure 4 (A) provides a graphical representation of the cross-validation (CV) schemes used for evaluating and comparing the prediction performance of QTL-based and GP models. CV was performed for *within-* and *joint-population prediction* using the unselected populations evaluated as DH lines *per se* in 2011 (hereafter referred to as 2011 dataset) and additionally in a *joint-population* framework with the selected DH lines *per se* and their testcrosses evaluated in 2012 (hereafter referred to as 2012 datasets). For both QTL and GP analyses fivefold CV was performed to estimate the predictive abilities of the different models. Each of the three datasets was split into five subsets of equal size: four subsets comprising 80 % of the genotypes built the estimation set (ES) and were used for model training while the remaining subset (20 % of genotypes) constituted the test set (TS). Each of the five subsets was used once as TS. The process was replicated 10 times with varying allocations of genotypes to the five CV subsets. In *joint-population prediction* within the 2011 dataset, sampling was performed by taking population information into account. Out of 521 DH lines, 81 (15 %), 214 (41 %) and 226 (43 %) belonged to Pop1, Pop2 and Pop3, respectively. These proportions were maintained in each of the five CV subsets. Each ES comprised DH lines from all three populations, while each TS was subdivided into three separate TSs comprising only DH lines from Pop1, Pop2 and Pop3, respectively. For each CV subset, the predictive performance of the two models was evaluated once for the combined TS across populations and additionally for the three population-specific TSs. In contrast, for *within-population prediction* the ES and TS were constructed through the random sampling of DH lines within biparental populations.

In *joint-population prediction* using 2012 datasets, sampling for CV subsets was implemented to maintain original proportions of genotypes with respect to population and additionally with respect to the two selected fractions (resistant and susceptible) within each population. Here,

both ES and TS always comprised genotypes from all three populations, but in each CV subset two separate TSs were constructed for the two seleced fractions.

For a direct comparison of the prediction performance for each TS between QTL and GP models, an identical allocation scheme of genotypes to CV subsets was used for both approaches across all CV replications. For the QTL model, predictions of genotypes in the TS were based on the sum of additive effects of all significant QTL detected in the ES, while GBLUP predictions were based on the effects of all polymorphic SNP markers in the ES. For the GBLUP+QTL model, predicted values in the TS were calculated as the sum of the fixed effects at the QTL detected in the ES and the remaining random genome-wide marker effects estimated within this GBLUP framework. For the *joint-population prediction* scenario, fixed marker effects were modelled in GBLUP in the same way as they were modelled in the QTL analysis, *i.e.* by assuming four different allelic effects for the four parental lines. The GBLUP+QTL model was evaluated only within the 2011 dataset. Given that no built-in option for CV has so far been available with the MCQTL program (B. Mangin, personal communication, April 5th 2013), a new R package called *cvMCQTL* was developed that performs this task by running a CV-loop on the QTL mapping routine of MCQTL (see Appendix Documentation A1). *cvMCQTL* is executed in combination with the *synbreed* R package.

Predictive abilities of the different models were calculated as Pearson's correlation coefficient between predicted and observed trait values in each TS. Predictive abilities obtained from each of the five TS within a CV replication were averaged to form replication means. The ten replication means were successively averaged to calculate an overall mean predictive ability across all CV subsets and replications. The standard deviation of the replication means was used as a measure of uncertainty of predictive ability estimates following the procedure proposed by Luan et al. (2009). For each TS in the 2012 datasets, correlations between predicted and observed values were calculated separately for the susceptible and resistant fraction to avoid a selection-induced bias.

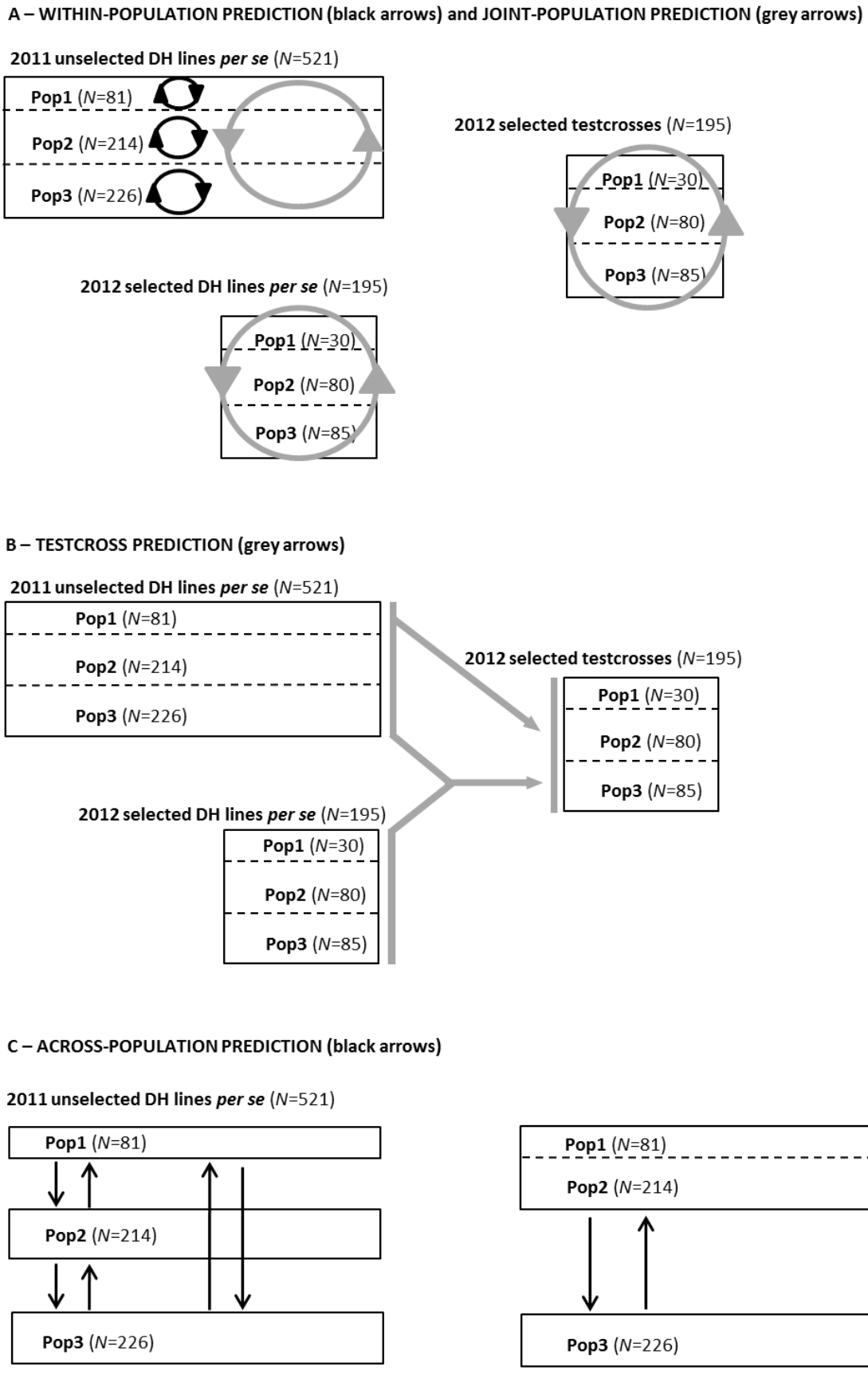

Figure 4 Representation of different cross-validation scenarios applied to the three DH populations Pop1, Pop2 and Pop3: (A) *within-* and *joint-population prediction*, (B) *testcross prediction* and (C) *across-population prediction*.

2.10. Additional cross-validation scenarios for genome-wide prediction

In addition to the comparison of QTL-based and GBLUP models, the prediction performance of GBLUP was evaluated for the prediction scenarios displayed in Figure 4 B-C and Figure 5.

2.10.1. *Testcross prediction*

Testcross prediction refers to the genome-based prediction of ECB stalk damage resistance of the testcrosses using models trained from data observed at the DH line *per se* level (Figure 4 B). The GBLUP model was trained using all DH lines *per se* (N = 521). Following phenotypes were included in the ES: (i) only 2011 DH lines *per se* data, (ii) data from (i) plus 2012 data of the resistant fraction of DH lines *per se* and (iii) data from (i) plus 2012 data of the resistant and susceptible fractions of DH lines *per se*. Two-year observations of the selected DH lines *per se* were fitted as replicated phenotypes in GBLUP including a fixed year effect. Given that only a selected subset of DH lines was evaluated as testcrosses, the CV procedure was slightly modified. In each of the ten CV replications we obtained a predicted genetic value for the N = 195 selected DH lines evaluated as testcrosses. Out of these 195 lines, N = 137 belonged to the resistant and N = 58 to the susceptible fraction, respectively. Predictive abilities were calculated separately for the resistant and susceptible fraction by correlating the N = 137 and N = 58 predicted values with phenotypic observations of their corresponding testcrosses, respectively. The mean predictive ability and its standard deviation were calculated based on the ten CV replications.

2.10.2. *Across-population prediction*

Across-population prediction with the GBLUP model was performed using the 2011 dataset of unselected DH lines *per se* (Figure 4 C). One ES comprised either an individual population or a merged dataset of two populations whereas the remaining population(s) represented the TS. Because of the predefined ES and TS structure, the replication of different CV subsets was not possible, as genome-wide prediction of all DH lines in a given population was based on models trained using a single ES (which contained all DH lines from one or two other populations). Therefore, a bootstrapping approach was implemented in the TS to obtain uncertainty measures of predictive abilities for this scenario. The N DH lines in a given TS were sampled N times with replacement to create one bootstrap sample of size N. This procedure was repeated 1,000 times to generate 1,000 different bootstrap samples of size N. Predictive abilities were calculated as Pearson's correlation coefficient between predicted and observed values of each bootstrap sample. The 1,000 individual predictive abilities were averaged to form an overall

mean predictive ability, and the bootstrap standard error was calculated as the standard deviation of the 1,000 individual values.

2.10.3. *Across-location prediction*

For this GP scenario (hereafter referred to as *across-location prediction*) genome-based predictions were obtained by sampling across both genotypes and field locations for the trait SDR evaluated in 2011. This CV procedure was chosen to maintain comparability of results with the other GP scenarios, which always involved sampling mutually exclusive sets of genotypes for the respective ES and TS. In *across-location prediction*, the fivefold CV scheme described in chapter 2.9 was implemented to include phenotypic data drawn from individual field locations or from different subsets of locations for model training and validation. The *across location* CV-scheme is illustrated with an example in Figure 5. Here, model training is performed based on BLUEs for SDR averaged across the two artificially infested locations (FRS and MZH_i). For the calculation of predictive abilities, predicted values for the genotypes in the TS are correlated with their respective BLUEs averaged across the remaining four naturally infested locations.

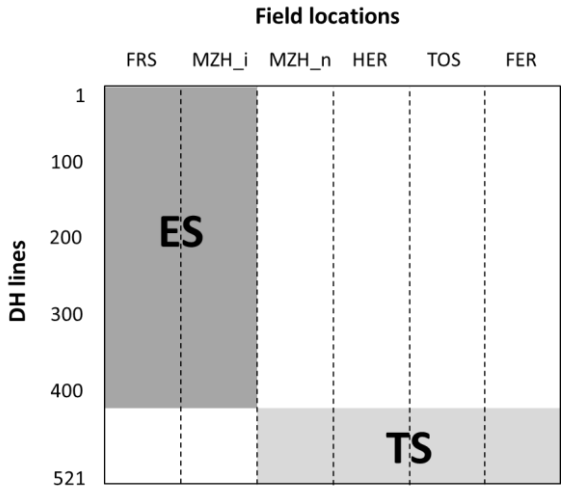

Figure 5 Example of a cross-validation scheme for the *across-location prediction* scenario for SDR evaluated in 2011 in the unselected DH populations (N = 521). The field locations are Freising (FRS), Münzesheim artificially infested (MZH_i), Münzesheim naturally infested (MZH_n), Herbolzheim (HER), Tournoisis (TOS), and Ferrara (FER). ES and TS refer to the estimation set and test set, respectively.

3. RESULTS

3.1. Marker analysis and genetic maps

The overall number of polymorphic SNP markers across populations was 2,411 and 365 of these SNPs were polymorphic in all three populations. Pop2 had the highest number of polymorphic SNPs (1,660), followed by Pop3 (1,340) and Pop1 (977). The polymorphic SNPs of Pop2 were distributed evenly across the genome. Pop1 lacked polymorphisms on the entire chromosome 8, while Pop3 had no polymorphic SNPs on the long arm of chromosome 7 and on the short arm of chromosome 8 (Figure 6). Chromosome 2 showed the highest and most homogeneous coverage with polymorphic SNPs across all three populations whereas chromosomes 9 and 10 displayed polymorphisms only in the distal regions.

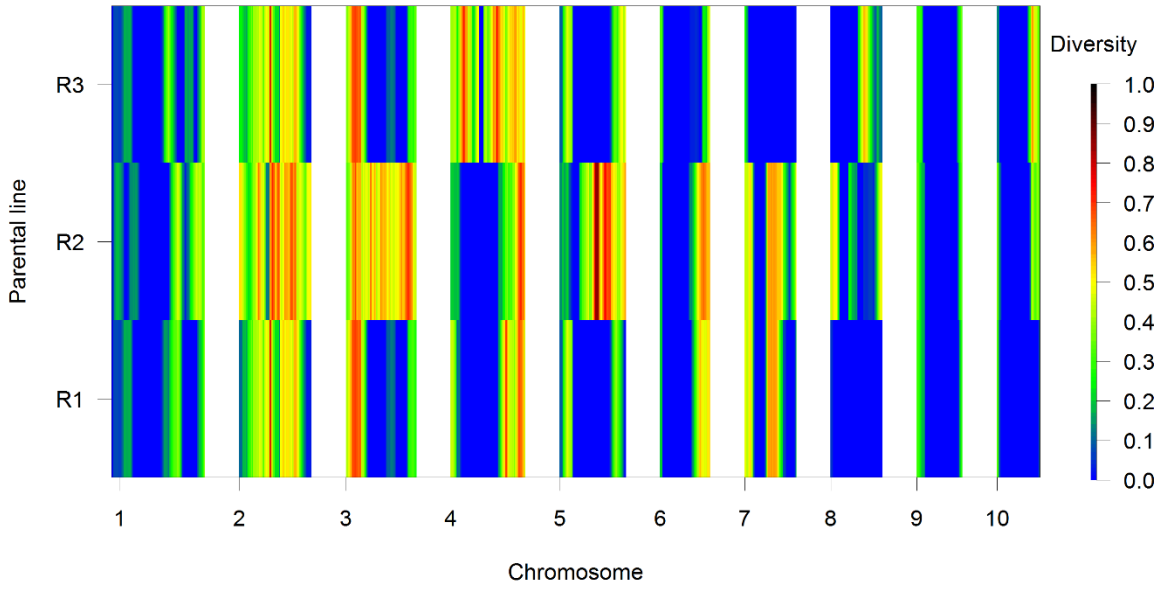

Figure 6 Genome-wide extent of pairwise polymorphism between the three resistant parental lines (R1, R2, R3) and the common susceptible parental line (S1) expressed as relative SNP allelic diversity estimated within a 30 Mbp window at incrementing steps of 3 Mbp along the genome.

The parent R2 was least related to the other three parental lines (Table 2) explaining the higher degree of polymorphism in Pop2 and the broader molecular variation of the DH lines within Pop2, which is also evidenced in the PCA on marker data (Figure 7). Parents R1, R3 and S1 showed a similar degree of relatedness based on their pairwise comparisons (Table 2).

Table 2 Rogers distances calculated based on SNP marker data between the three resistant (R1, R2, R3) and the susceptible (S1) parental lines.

	R1	R2	R3
R1			
R2	0.28		
R3	0.20	0.34	
S1	0.20	0.35	0.22

The PCA performed on polymorphic SNP data clearly separated the populations in three distinct clusters (Figure 7). The marker data did not reveal substructure within the three individual populations.

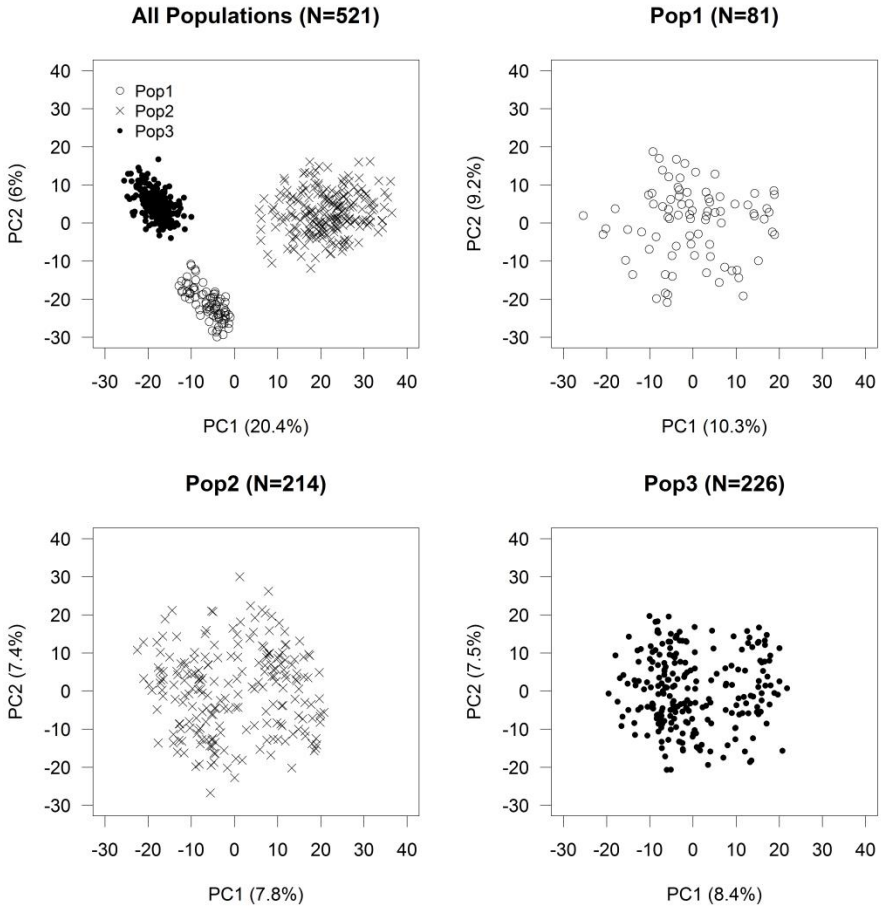

Figure 7 Scatterplots of the loadings of the first two principal components from a PCA on the polymorphic SNP markers computed individually for each of the three DH populations Pop1, Pop2 and Pop3 as well as across populations.

The chi-square test at the Holm-Bonferroni adjusted 0.05 significance level did not reveal any significant segregation distortion of the SNP markers in Pop1. Only 41 markers located on chromosome 7 showed significant distortion in Pop2 (Appendix Figure A 1). The major allele

at all polymorphic markers on chromosome 7 was provided by the R2 parental line. A much larger extent of distorted segregation, in both magnitude and in the number of SNPs involved, was observed in Pop3 across all ten chromosomes (Appendix Figure A 2). 571 out of 1,340 polymorphic SNPs in Pop3 (approximately 40 %) showed significant segregation distortion at the 0.05 Holm-Bonferroni adjusted significance level. Contrarily to the results obtained in Pop2, the segregation distortion of markers in Pop3 was almost exclusively in favor of the allele from the common susceptible parent S1. Moreover, in many genomic regions within Pop3 the intensity of segregation distortion did not follow a continuous gradient along the genome (in contrast to the trend on chromosome 7 observed in Pop2, see Appendix Figure A 1). In Pop3, several genomic regions displayed both SNPs segregating as expected (in a 1:1 ratio) and SNPs with pronounced distorted segregation. From these results it cannot be excluded that during the process of DH line production for Pop3 cross pollination occurred at an unknown rate. Additionally, 277 out of 1,340 SNPs segregating in Pop3 were monomorphic in the parental lines R3 and S1. The vast majority of these 277 SNPs showed a pronounced distorted segregation ratio (black dots in Appendix Figure A 2). These SNPs were removed from linkage mapping and QTL-analyses due to missing information on parental alleles. 133 markers distributed across several chromosomes and displaying a heavily distorted segregation ratio were additionally removed from the linkage analysis of Pop3 leaving 930 mapped markers for QTL analyses. Even though these results hint at some degree of admixture in Pop3, the PCA using all 1,340 polymorphic SNPs did not highlight any apparent substructure within the population nor the presence of outlying genotypes (Figure 7). The order of markers along the genetic map of Pop3 was also consistent with the reference physical map and the genetic maps of Pop1 and Pop2.

The consensus map across all three populations shown in Figure 8 included 1,034 informative SNPs and displayed a total length of 1,787 cM over the 10 chromosomes. The same representation for the three individual genetic maps of Pop1, Pop2 and Pop3 is provided in the Appendix (Figure A 3).

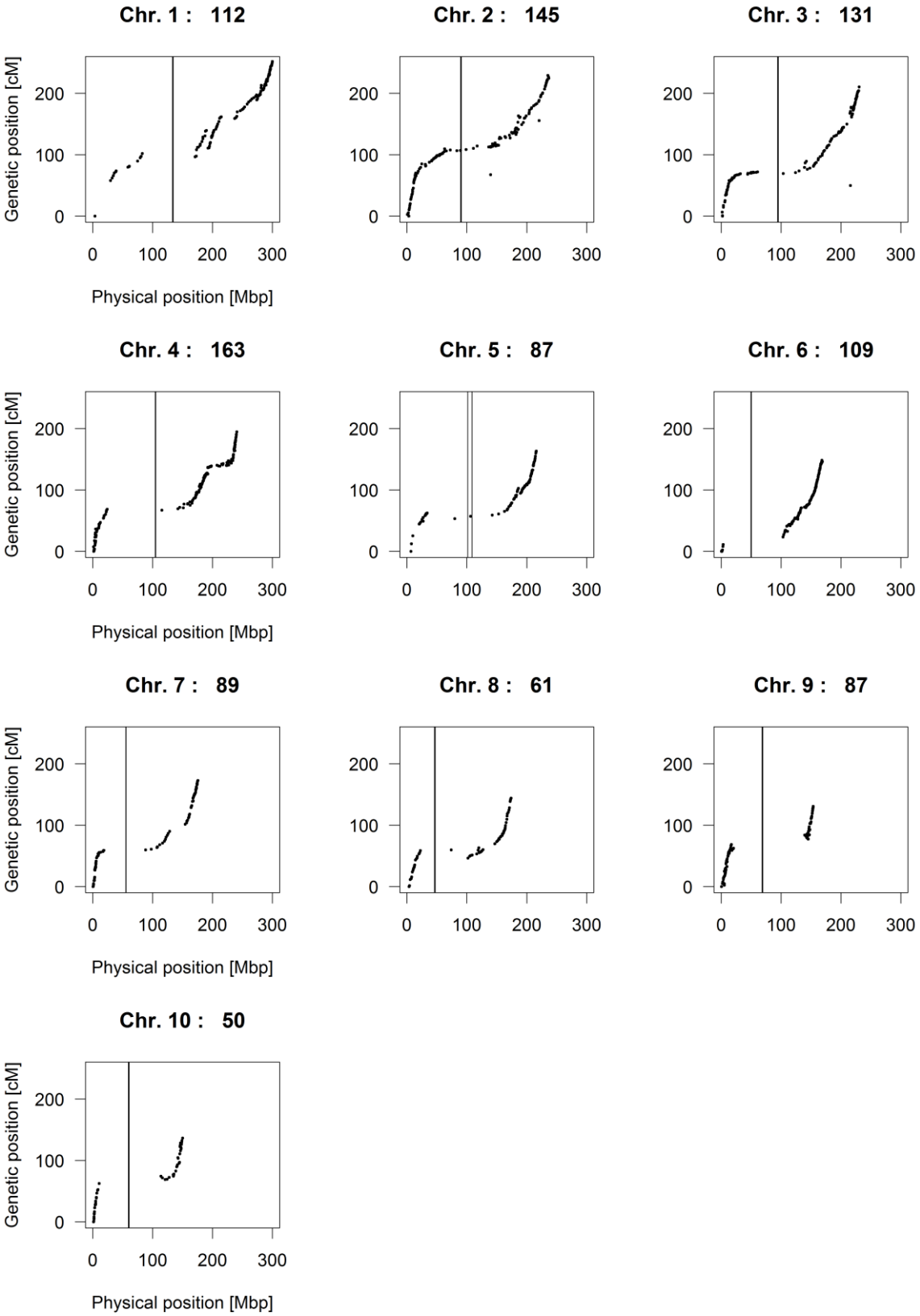

Figure 8 Mapped genetic position [cM] versus corresponding physical position [Mbp] of 1,034 SNP markers on the genetic consensus map based on the three DH populations for all ten chromosomes (Chr. 1 - Chr. 10). Vertical lines within each chromosome denote the estimated position of the centromere according to physical map information. Numbers above each graph refer to the number of SNP markers contained in the consensus map of the respective chromosome.

3.2. Quantitative genetic analysis

3.2.1. *ECB stalk damage traits*

The overall severity of ECB damage was higher in 2012 compared to 2011, due to climatic conditions that were more favorable for the development of ECB larvae. This is shown by the comparison of population trait means between the two years (Table 3). Mean trait values observed in 2012 for DH lines *per se* were considerably higher than in 2011 for all three resistance traits. Because of substantially taller plants and an overall higher center of gravity, testcrosses tend to break more often below the ear compared to inbred lines *per se*, thus leading to higher average SDR scores in the testcrosses compared to the inbred lines *per se*. For TL and NT, testcross means were significantly lower than for their corresponding inbred lines evaluated in the same year. Differences between population means for SDR were not significant in 2011, while significant differences were observed in 2012, particularly between Pop1 and Pop3. Similar results were observed for the other resistance traits but in most cases population mean differences were small.

Significant genetic variance was found for all resistance traits in 2011, with the exception of TL in Pop1 (Table 3). The selected genotypes evaluated in 2012 showed significant genetic variance for most population-trait combinations. Exceptions were observed for the small subset of selected DH lines in Pop1 (N = 31) showing no genetic variation for TL and NT in both DH lines *per se* and testcrosses. Genetic variance for TL and NT was mostly non-significant also for the testcrosses of Pop2 and Pop3.

Bidirectional selection for ECB stalk damage traits was performed based on results from 2011 field trials. Selection was based primarily on SDR under the side condition of similar flowering time, and with TL and NT serving as supporting criteria for DH lines showing more intermediate SDR values. The selection was successful in preserving genetic variance for all three resistance traits, particularly for SDR and TL. Genotypic variance components were generally higher in 2012 than in 2011, which is most probably a consequence of the increased ECB pest pressure observed during the 2012 field trial season. The left-side graphs in Figure 9 show the adjusted means of the selected DH lines *per se* and their testcrosses evaluated in 2012 for SDR. In the DH lines *per se* (Figure 9 A) there was a clear difference in mean SDR of the 147 DH lines selected for low SDR and the 60 DH lines selected for high SDR in 2011. The difference was still significant but less pronounced in testcrosses (Figure 9 B). The mean SDR of resistant parents R1, R2 and R3 were significantly lower than the mean SDR of susceptible parent S1 in both DH lines *per se* and testcrosses, with the sole exception of the testcross of

parent R1. The means of the two selected fractions differed significantly also for TL and NT at the DH line *per se* level, but not at the testcross level. The genetic variance for these two traits was generally not significant in the testcrosses (Table 3). For the DH lines *per se*, differences between both fractions were less pronounced for TL and NT than for SDR (Appendix Figure A 4).

For SDR evaluated across six locations, genotype-by-location interaction was relatively high in DH lines *per se* and in both years when compared to the corresponding genetic variance components. However, genotype-by-location interaction variance was not significant in the testcrosses of Pop2 and Pop3 (Appendix Table A 1). It was significant in the testcrosses of Pop1 but also showed a large SE due to the small number of selected genotypes in this population (N = 31). For TL and NT evaluated in the two artificially infested locations, genotype-by-location interaction was mostly not significant in both years, for both DH lines *per se* and testcrosses, and for all three populations (Appendix Table A 1).

In both years, trait heritabilities were medium to high (h^2 = 0.50 – 0.76) for SDR averaged across six locations and low to medium for TL and NT (h^2 = 0.27 – 0.56) averaged across two locations (Table 3). Heritabilities were highest for SDR in both unselected and selected DH lines *per se* of the two larger DH populations Pop2 and Pop3, with values ranging between 0.70 and 0.76. Heritability estimates for SDR remained relatively high also in testcrosses of Pop2 and Pop3 (h^2 = 0.64). In most cases heritabilities for resistance traits were similar for Pop2 and Pop3 but Pop1 often showed lower values.

Table 3 Estimates of the mean, genetic variance ($\hat{\sigma}_g^2$) and heritability ($\hat{h}^2$) (± standard errors) for the traits SDR, TL, NT, ANT, SIL, PH, EH and BRIX for three DH populations (Pop1, Pop2, Pop3). Estimates are given for the unselected populations evaluated as DH lines *per se* in 2011 as well as for the selected DH lines *per se* and their testcrosses evaluated in 2012.

Pop.	Entries (no.)	SDR [1-9 scale] 6 locations			TL [cm] 2 locations			NT [counts] 2 locations			ANT [days] 3 locations		
		Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$
2011 unselected populations – DH lines *per se*													
Pop1	85	2.89±0.06	0.10±0.03	0.55±0.08	13.83±0.46	6.84±3.89^{ns}	-	2.20±0.06	0.10±0.06	0.30±0.16	97.00±0.28	5.04±0.92	0.85±0.03
Pop2	243	2.94±0.03	0.28±0.03	0.75±0.03	11.65±0.27	9.79±1.90	0.51±0.07	1.93±0.04	0.16±0.04	0.46±0.07	98.30±0.16	6.09±0.66	0.87±0.02
Pop3	262	2.86±0.03	0.17±0.02	0.68±0.03	12.29±0.26	8.46±1.60	0.47±0.05	2.20±0.04	0.15±0.04	0.41±0.08	96.85±0.16	6.30±0.65	0.86±0.02
All	590	2.90±0.02	0.21±0.02	0.71±0.02	12.25±0.17	9.41±1.23	0.49±0.04	2.09±0.02	0.15±0.02	0.43±0.05	97.46±0.11	6.08±0.41	0.87±0.01
2012 selected populations – DH lines *per se*													
Pop1	31	4.13±0.10	0.20±0.07	0.72±0.08	27.05±0.98	1.97±6.90^{ns}	-	3.52±0.10	0.00±0.00^{ns}	-	90.77±0.26	1.24±0.45	0.73±0.09
Pop2	85	3.94±0.06	0.28±0.06	0.76±0.04	27.09±0.59	16.83±7.32	0.43±0.13	3.28±0.06	0.10±0.06^{ns}	-	91.07±0.16	1.36±0.33	0.66±0.07
Pop3	91	3.66±0.06	0.26±0.05	0.75±0.04	22.73±0.57	16.89±5.26	0.56±0.10	3.40±0.06	0.13±0.05	0.44±0.12	91.07±0.15	1.43±0.32	0.70±0.06
All	207	3.84±0.04	0.27±0.04	0.76±0.03	25.17±0.38	17.09±3.78	0.50±0.07	3.37±0.04	0.11±0.04	0.36±0.09	91.02±0.10	1.44±0.21	0.69±0.04
2012 selected populations – Testcrosses													
Pop1	31	5.30±0.08	0.09±0.05	0.50±0.15	13.28±0.63	4.15±3.10^{ns}	-	1.96±0.06	0.00±0.00^{ns}	-	81.17±0.23	0.55±0.26	0.60±0.16
Pop2	85	5.15±0.05	0.16±0.04	0.64±0.06	14.28±0.38	1.28±2.64^{ns}	-	1.97±0.04	0.00±0.00^{ns}	-	82.57±0.14	3.85±0.71	0.90±0.02
Pop3	91	5.02±0.05	0.15±0.04	0.64±0.06	14.53±0.37	2.66±2.08^{ns}	-	2.10±0.04	0.07±0.03	0.45±0.10	80.66±0.14	1.10±0.27	0.71±0.07
All	207	5.12±0.03	0.16±0.02	0.65±0.04	14.23±0.24	3.16±1.33	0.27±0.09	2.02±0.03	0.04±0.02	0.29±0.08	81.53±0.09	2.07±0.26	0.83±0.02

^{ns}, not significant at the 0.05 probability level.

Table 3 (continued)

Pop.	Entries (no.)	SIL [days] 3 locations			PH [cm] 1-2 locations			EH [cm] 2 locations			BRIX [°Brix] 2 locations		
		Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{h}^2$
2011 unselected populations – DH lines *per se*													
Pop1	85	97.42±0.31	7.28±1.26	0.89±0.02	196.03±1.41	164.30±30.49	-						
Pop2	243	97.47±0.19	6.69±0.70	0.87±0.02	197.62±0.84	134.19±16.30	-						
Pop3	262	97.20±0.18	8.02±0.80	0.89±0.01	178.91±0.80	334.07±35.25	-						
All	590	97.34±0.12	7.43±0.49	0.88±0.01	189.07±0.54	217.92±14.76	-						
2012 selected populations – DH lines *per se*													
Pop1	31	91.05±0.32	2.64±0.85	0.81±0.07	214.53±2.23	179.63±52.01	0.91±0.04	107.38±1.79	96.52±28.80	0.87±0.05	10.33±0.12	0.19±0.12^{ns}	-
Pop2	85	90.60±0.19	2.26±0.51	0.72±0.06	211.46±1.35	116.85±24.15	0.79±0.05	107.19±1.08	83.38±15.59	0.84±0.04	9.71±0.07	0.51±0.13	0.72±0.07
Pop3	91	91.60±0.19	2.62±0.50	0.81±0.04	198.95±1.30	227.86±38.94	0.93±0.02	106.52±1.04	73.37±13.79	0.81±0.04	10.58±0.07	0.32±0.10	0.58±0.10
All	207	91.11±0.12	2.59±0.33	0.78±0.03	206.42±0.86	174.81±19.53	0.89±0.02	106.93±0.69	85.75±10.10	0.84±0.02	10.19±0.05	0.38±0.06	0.64±0.05
2012 selected populations – Testcrosses													
Pop1	31	79.50±0.25	0.79±0.31	0.69±0.08	284.90±1.38	19.90±11.42^{ns}	-	150.84±1.08	25.92±12.08	0.61±0.15			
Pop2	85	80.09±0.15	2.25±0.47	0.80±0.05	285.32±0.83	33.26±10.73	0.55±0.11	155.12±0.65	28.70±7.38	0.66±0.08			
Pop3	91	79.31±0.15	1.23±0.26	0.76±0.04	283.33±0.81	40.95±10.45	0.67±0.08	154.57±0.63	18.10±5.00	0.62±0.09			
All	207	79.66±0.10	1.62±0.22	0.78±0.03	284.40±0.53	32.93±6.38	0.58±0.06	154.21±0.42	24.93±4.06	0.66±0.05			

^{ns}, not significant at the 0.05 probability level.

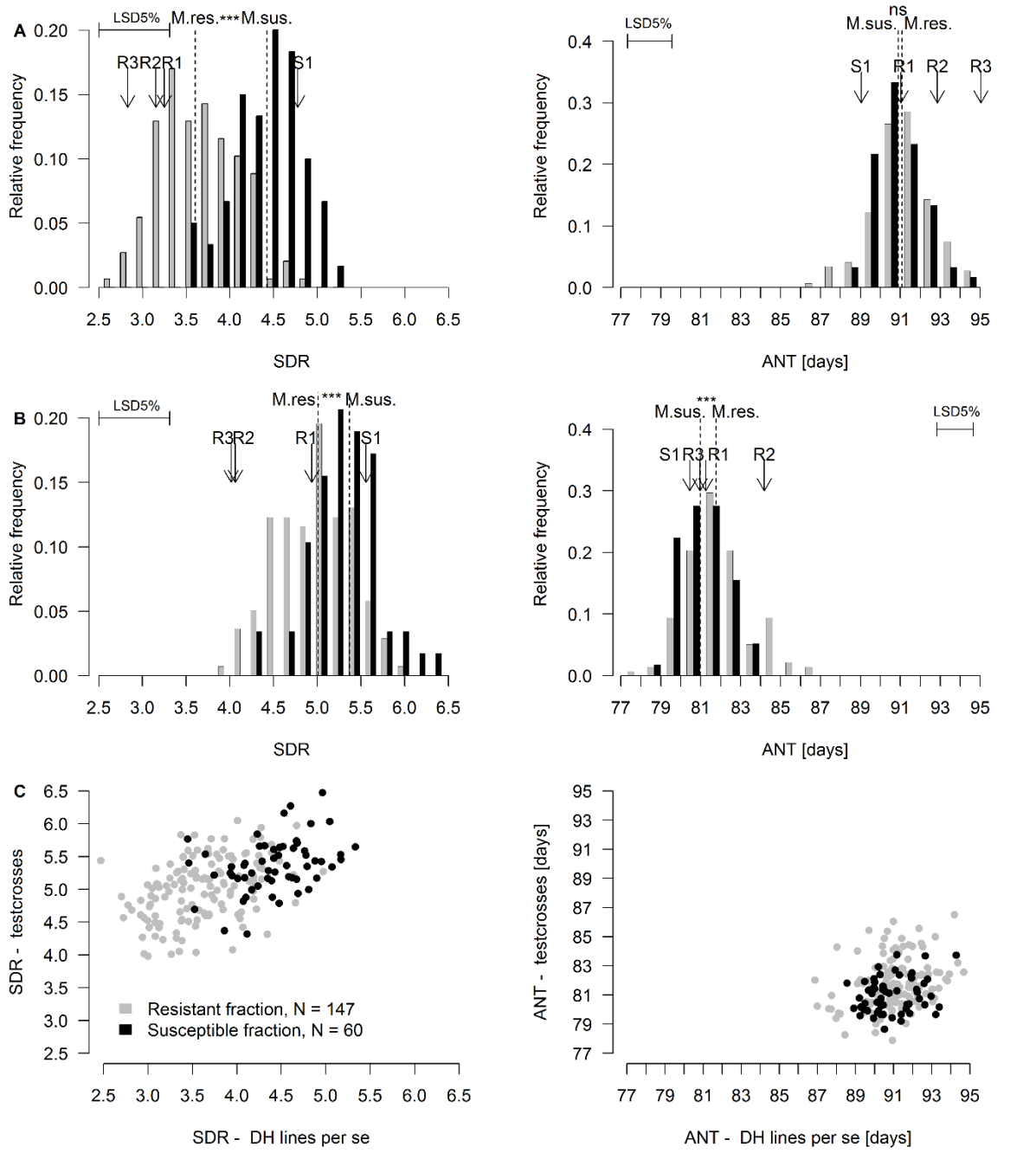

Figure 9 Distributions of the adjusted means for SDR averaged across six locations (left) and for ANT averaged across three locations (right) for the selected DH lines *per se* (A) and their corresponding testcrosses (B) evaluated in 2012, and scatterplots of the phenotypic correlation between DH lines *per se* and testcrosses for SDR and ANT (C). Genotypes from the resistant (N = 147) and susceptible (N = 60) fractions are represented in all graphs by grey and black color, respectively. Arrows show the means of the parental lines R1, R2, R3 and S1. The least significant difference (LSD) at the 5% significance level and the means of the resistant (M.res.) and susceptible (M.sus.) fractions are also shown. Significant differences between M.res. and M.sus. were determined with a t-test (***, ns, difference in fraction means significant at $p<0.001$ and non-significant with $p>0.05$, respectively).

3.2.2. *Agronomic traits and BRIX*

Differences between population means were mostly significant for ANT and SIL but were generally small, similarly to what was observed for resistance traits (Table 3). The maximum difference in flowering time between two population means was 1 to 1.5 days in DH lines *per se* and approximately 2 days for ANT in the testcrosses. These are rather small differences when compared to an overall timespan of the flowering period of 15 to 18 days observed in the unselected DH lines *per se* evaluated in 2011. Population mean differences were small also for EH and BRIX. For EH, only the mean of the testcrosses of Pop1 was significantly lower compared to the other two populations but the difference was only approximately 5 cm. For PH, in contrast, differences between population means were more pronounced in DH lines *per se*, with the mean of Pop3 being 12 to 18 cm lower compared to the means of Pop1 and Pop2 in both years. The common parent S1 was significantly shorter than the three resistant parental lines (R1, R2, R3) while the height of the resistant parent of Pop3 (R3) was not significantly different from the height of R1 and R2 in both years (data not shown). In the testcrosses, on the other hand, population mean differences for PH were minimal given that the testcrosses of Pop3 were on average only 2 cm shorter than the testcrosses of Pop1 and Pop2.

No significant genetic variance was observed for BRIX in the selected DH lines *per se* and for PH in the selected testcrosses of Pop1 (N = 31), respectively. All other population-trait combinations showed significant genetic variance (Table 3). Genetic variance for SIL and ANT decreased strongly in 2012 DH lines *per se* as compared to 2011 as a consequence of selection for reduced variation in flowering time, which was performed simultaneously with the bidirectional selection on ECB resistance traits. As expected the selection also led to a similar range of flowering time for the resistant and susceptible genotypes. The two fractions almost completely overlapped for ANT in both DH lines *per se* and testcrosses, although the means of the two fractions were significantly different in the testcrosses (Figure 9). This difference was mainly driven by the testcrosses of Pop2. Indeed, the testcross of parent R2 flowered significantly later than the testcrosses of the other parents, which was in contrast to the flowering time of the respective parental lines *per se* (Figure 9 A and B). Genetic variance for ANT was also unexpectedly high in the testcrosses of Pop2, exceeding almost three times the genetic variance component of the corresponding (selected) DH lines *per se*. However, it was smaller than the respective variance component observed in the unselected DH lines *per se* (Table 3).

As expected for these highly heritable agronomic traits the magnitude of genotype-by-location interaction was generally low (Appendix Table A 1). For ANT and SIL, genotype-by-location

interaction variance was five to eight times smaller than the corresponding genetic variance component in the unselected DH lines *per se* and not significant in the testcrosses. For PH and EH, genotype-by-location interaction variance was estimated only within the selected DH lines *per se* and testcrosses, given that in 2011 EH was not evaluated and PH was measured only at location Freising. Genotype-by-location interaction was low for EH and even lower for PH, with exception of the testcrosses in Pop2 (Appendix Table A 1).

Heritabilities were very high for ANT and SIL in the unselected populations (always exceeding 0.80 with an average of 0.87, see Table 3), and compared to 2011 decreased as expected in the selected DH lines *per se* evaluated in 2012. Heritabilities of ANT, SIL, PH and EH were generally lower in testcrosses. A striking exception was the very high value for ANT in the testcrosses of Pop2, which was a consequence of the unexpectedly high genetic variance observed within the testcrosses of this population in 2012. Stalk sugar concentration in °Brix degrees (BRIX) was evaluated across two locations in the selected DH lines *per se* in 2012 and showed medium to high heritabilities in Pop2 and Pop3.

3.2.3. *Correlations between traits and between line per se and testcross performance*

Genotypic correlations between the three resistance traits were high for unselected DH lines *per se* (Table 4). In testcrosses correlations between SDR and stalk tunneling traits were not significant or low because of low heritabilities of TL and NT. Medium to high negative genotypic correlations between flowering time and resistance traits were observed in 2011. In general, these correlations were considerably lower in 2012 after selection for reduced flowering time variation. No significant correlations between flowering time and ECB stalk damage traits were observed in testcrosses, except for significant but moderate negative correlations between ANT and SDR and between NT and SIL. PH showed medium positive genotypic correlations with ANT and SIL and low to medium negative correlations were observed between PH and resistance traits in 2011, particularly with NT. A low to moderate negative correlation between PH and NT was confirmed in the 2012 selected DH lines *per se*, while in the testcrosses SDR and TL showed moderate negative correlations with PH. Given the relatively high positive correlation between PH and EH, also the trait EH generally tended to be negatively correlated with the ECB resistance traits. When summarizing the complex landscape of the correlations reported in Table 4, the following general trend is observed: late-flowering genotypes tended to grow taller and to show a reduced level of ECB damage. This observation is also reflected in the parental lines, because all three resistant parents flowered significantly later and were significantly taller than the common susceptible parent S1. Stalk

sugar concentration (BRIX) was not correlated with SDR but displayed low and moderate positive genotypic correlations with TL and NT, respectively. However, BRIX also showed low to moderate negative correlations with the four agronomic traits.

Table 4 Phenotypic (above diagonal) and genotypic (below diagonal) correlation coefficients between the traits SDR, TL, NT, ANT, SIL, PH, EH and BRIX. The table provides estimates across the three unselected populations evaluated as DH lines *per se* in 2011 as well as estimates across the selected DH lines *per se* and their testcrosses evaluated in 2012.

2011 unselected populations – DH lines *per se* (*N* = 590)

	SDR	TL	NT	ANT	SIL	PH
SDR		0.42**	0.41**	-0.38**	-0.40**	-0.07ns.
TL	0.73++		0.85**	-0.40**	-0.30**	-0.19**
NT	0.71++	0.94++		-0.48**	-0.34**	-0.31**
ANT	-0.50++	-0.67++	-0.73++		0.84**	0.45**
SIL	-0.52++	-0.50++	-0.53++	0.86++		0.34**
PH	-0.11+	-0.36++	-0.50++	0.50++	0.40++	

2012 selected populations – DH lines *per se* (*N* = 207)

	SDR	TL	NT	ANT	SIL	PH	EH	BRIX
SDR		0.41**	0.36**	-0.20**	-0.31**	0.08ns.	-0.14*	-0.03ns.
TL	0.63++		0.65**	-0.24**	-0.19**	-0.07ns.	-0.27**	0.09ns.
NT	0.59++	0.69++		-0.23**	-0.08ns.	-0.23**	-0.35**	0.28**
ANT	-0.34++	-0.35++	-0.38++		0.66**	0.25**	0.44**	-0.20**
SIL	-0.43++	-0.26++	-0.12+	0.70++		0.11ns.	0.28**	-0.12ns.
PH	0.12+	-0.09	-0.34++	0.39++	0.17+		0.65**	-0.29**
EH	-0.21+	-0.40++	-0.54++	0.60++	0.39++	0.68++		-0.25**
BRIX	-0.05	0.18+	0.42++	-0.31++	-0.18+	-0.42++	-0.33++	

2012 selected populations – Testcrosses (*N* = 207)

	SDR	TL	NT	ANT	SIL	PH	EH
SDR		0.16*	0.21**	-0.15*	-0.10ns.	-0.21**	-0.18**
TL	-0.17		0.77**	0.03ns.	-0.06ns.	0.01ns.	0.03ns.
NT	0.51+	0.80++		-0.05ns.	-0.19**	0.03ns.	-0.07ns.
ANT	-0.31+	0.03	-0.11		0.75**	0.33**	0.43**
SIL	-0.18	-0.11	-0.35++	0.82++		0.18**	0.44**
PH	-0.37++	-0.23+	-0.16	0.47++	0.37++		0.47**
EH	-0.32+	0.18	-0.10	0.60++	0.58++	0.71++	

*, **, Pearson phenotypic correlation significant at the 0.05 and 0.01 probability levels, respectively. ns. not significant
+, ++, Genotypic correlation exceeds once or twice its standard error, respectively.

Correlations for a given trait between DH lines *per se* and corresponding testcrosses evaluated in 2012 were calculated both across all selected genotypes (*N* = 207) and separately within the resistant (*N* = 147) and susceptible (*N* = 60) fractions to avoid potentially inflated correlation estimates as a consequence of bidirectional selection. Genotypic correlations were high for SDR and surprisingly very high for NT despite rather low phenotypic correlation coefficients for the latter trait (Table 5). The variance component for NT in the testcrosses was slightly smaller than the covariance component between DH lines *per se* and testcrosses for the same trait, thus leading to very high estimates for the genotypic correlation coefficient (> 0.90). However, the

SE of the genotypic correlation coefficients calculated for NT were larger than the SE obtained for TL and much higher than those calculated for SDR, ANT, SIL, PH and EH (data not shown). This is underlined by the genotypic correlation for NT estimated within the susceptible fraction: the very high value of 0.99 exceeded only once a correspondingly large standard error (SE > 0.90). Hence, genotypic correlation coefficients between DH lines *per se* and testcrosses may be upward biased in the case of NT. Genotypic correlation for SDR estimated within the resistant fraction was only slightly lower compared to the corresponding coefficient estimated using all genotypes (Table 5). Hence, inflation of correlations due to selection was limited. For SDR and ANT a graphical display of phenotypic correlations between DH lines *per se* and testcrosses is provided in Figure 9 (C). Correlations were relatively low for ANT and SIL but were also estimated within a sample that was selected for reduced variation in these two traits. For the other two agronomic traits PH and EH, moderately high correlations between DH lines *per se* and testcrosses were observed.

Table 5 Genotypic and phenotypic correlation coefficients between the selected DH lines *per se* (*N* = 207) and their corresponding testcrosses evaluated in 2012 for the traits SDR, TL, NT, ANT, SIL, PH and EH. Correlations are presented across all genotypes and separately for the resistant (*N* = 147) and susceptible (*N* = 60) fractions.

		Genotypic correlation			Phenotypic correlation		
Trait	No. locations	Res. fract. (*N* = 147)	Sus. fract. (*N* = 60)	All (*N* = 207)	Res. fract. (*N* = 147)	Sus. fract. (*N* = 60)	All (*N* = 207)
SDR	6	0.76++	0.66++	0.79++	0.41**	0.36**	0.52**
TL	2	0.36+	0.48	0.42++	0.10ns.	0.13ns.	0.14*
NT	2	0.95++	0.99+	0.90++	0.22**	0.27*	0.26**
ANT	3	0.36++	0.37+	0.37++	0.27**	0.24ns.	0.27**
SIL	3	0.38++	0.66++	0.46++	0.23**	0.48**	0.30**
PH	2	0.61++	0.75++	0.63++	0.54**	0.50**	0.52**
EH	2	0.70++	0.43++	0.65++	0.58**	0.29*	0.52**

*, **, Pearson phenotypic correlation significant at the 0.05 and 0.01 probability levels, respectively. ns. not significant
+, ++, Genotypic correlation exceeds once or twice its standard error, respectively.

3.3. QTL analyses

QTL analyses were performed using the adjusted genotype means (BLUEs) averaged across locations for the traits SDR (six locations), TL and NT (two locations) and ANT (three locations).

3.3.1. *Joint- and within-population analyses in the unselected DH populations*

For the unselected populations evaluated as DH lines *per se* in 2011 (N = 521) results from the *joint-population* QTL analysis are reported in Table 6.

Table 6 Chromosome (Chr.), position (Pos.), LOD support interval (LOD S.I.), LOD score at the QTL position, proportion of variance explained (R^2) and additive effects of QTL alleles derived from parents R1, R2, R3 and S1 detected in the joint analysis across the three unselected populations (N = 521) evaluated as DH lines *per se* in 2011 for the traits SDR, TL, NT and ANT. For each trait, the proportion of phenotypic variance explained by the model fitting all detected QTL simultaneously is given above the R^2 values of the individual QTL.

Chr.	Pos. [cM]	LOD S.I. [cM]	LOD score	R^2	Additive effects			
					R1	R2	R3	S1
SDR [1-9 score]				**0.47**				
1	32	(9 – 85)	4.66	0.04	-0.01	-0.15	0.13	0.03
2	17	(13 – 49)	6.75	0.06	-0.03	-0.08	0.04	0.07
2	98	(97 – 100)	15.90	0.13	0.00	-0.03	-0.08	0.11
3	67	(65 – 68)	27.72	0.21	-0.06	-0.04	-0.05	0.15
5	105	(103 – 110)	9.59	0.08	0.03	-0.02	-0.08	0.07
6	100	(92 – 106)	6.24	0.05	-0.05	-0.02	-0.01	0.08
8	62	(61 – 66)	18.68	0.15	0.00	0.15	-0.04	-0.11
10	103	(95 – 108)	5.38	0.05	0.10	-0.02	-0.10	0.02
TL [cm]				**0.37**				
2	100	(44 – 128)	4.35	0.04	0.00	-0.22	-0.28	0.50
3	69	(68 – 70)	7.71	0.07	-0.48	0.20	-0.46	0.75
3	135	(133 – 146)	5.17	0.05	0.40	-0.81	0.03	0.38
4	14	(1 – 25)	4.34	0.04	0.60	-0.27	0.22	-0.55
4	164	(159 – 166)	6.14	0.05	1.35	-0.35	-1.00	0.00
5	70	(63 – 71)	5.33	0.05	-0.42	-0.13	-0.12	0.67
5	110	(106 – 129)	8.40	0.07	0.87	-0.31	-1.00	0.44
8	36	(32 – 41)	4.64	0.04	0.00	0.53	0.01	-0.54
9	103	(100 – 113)	5.45	0.05	-0.10	-0.67	0.27	0.50
10	126	(123 – 136)	4.83	0.04	-0.05	0.11	-0.57	0.51
NT [count]				**0.25**				
3	70	(68 – 73)	11.71	0.10	-0.05	-0.03	-0.05	0.12
4	120	(105 – 163)	4.97	0.04	0.03	-0.03	-0.07	0.07
5	65	(61 – 67)	6.47	0.06	-0.05	-0.08	0.03	0.09
5	121	(115 – 150)	9.76	0.08	0.13	0.00	-0.19	0.06
ANT [days]				**0.47**				
1	206	(204 – 210)	7.84	0.07	-0.26	-0.25	0.12	0.39
2	17	(13 – 74)	4.76	0.04	0.37	-0.15	0.12	-0.34
2	225	(195 – 229)	5.26	0.05	-0.02	0.64	-0.68	0.06
3	65	(62 – 68)	9.12	0.08	-0.22	0.26	0.44	-0.48
3	91	(89 – 96)	10.91	0.09	1.14	-1.06	0.05	-0.13
5	87	(83 – 97)	6.20	0.05	-0.02	0.10	0.27	-0.35
5	126	(122 – 159)	6.81	0.06	0.21	-0.02	0.24	-0.43
8	58	(57- 64)	9.79	0.08	0.00	-0.69	0.58	0.11
10	69	(64 – 77)	5.14	0.05	-0.17	0.16	0.30	-0.29

Eight, ten, four and nine QTL were identified for SDR, TL, NT and ANT, respectively. R^2 values of individual QTL ranged between 0.04 and 0.21. Total R^2 values representing the proportion of phenotypic variance explained by the model fitting all detected QTL simultaneously ranged between 0.25 for NT and 0.47 for SDR and ANT. Additive allelic effects contributed by the parents R1, R2, R3 and S1 estimated from a connected QTL model were small to moderate for all traits (Table 6). As expected, at most QTL the alleles contributing to lower trait values originated from the resistant parents. Few QTL were detected where the resistance allele was contributed by the common susceptible parent S1. One of these was the QTL for SDR on chromosome 8 which had a negative effect sign for the S1 parental allele. This QTL was found in a similar region as the QTL for TL on chromosome 8 which also showed a negative effect for the S1 allele.

Figure 10 illustrates a comparison between QTL that were detected in the *within-population* analyses of Pop1, Pop2 and Pop3 and the QTL identified in the *joint-population* analysis, as well as comparisons of QTL positions across the four traits. Detailed results of *within-population* analyses are reported in the Appendix (Table A 2). Few QTL were identified in *within-population* analyses that were not detected in the *joint-population* analysis. These were for Pop1 a QTL for ANT on chromosome 4, for Pop2 three QTL for ANT on chromosomes 5 and 7, and for Pop3 two relatively close QTL for SDR on chromosome 4. As seen from the LOD profiles in Figure 10, all these QTL were of rather small magnitude and were also among the QTL showing the smallest R^2 values in the respective *within-population* analyses (Appendix Table A 2). The *joint-population* analysis also identified a few QTL that were not detected in any of the three *within-population* analyses. These included a QTL on chromosome 2 for SDR, on chromosomes 2, 4 and 8 for TL and on chromosomes 2 and 10 for ANT. These QTL explained only a small proportion of the total phenotypic variation. All QTL that displayed larger effects were identified in the *joint*-analysis and additionally in one or more of the *within-population* analyses. Overall, fewer QTL were identified in Pop1 compared to the other two populations. Pop1 (N = 81) was considerably smaller than Pop2 (N = 214) and Pop3 (N = 226) and therefore the power of QTL detection was lower than in the two larger populations. Less than half of the QTL detected for the four traits were shared among two or more populations. The remaining QTL appeared to be expressed specifically in a single population, in most cases in either Pop2 or Pop3. This was particularly true for TL for which the great majority of QTL were population-specific. LOD support intervals of the *joint-population* analysis were in most cases either of similar or smaller width when compared to the smallest support intervals observed at the corresponding QTL in the *within-population* analyses (Figure 10). This result

supports the generally higher accuracy of QTL position estimates that can be achieved under multiple-cross QTL mapping.

The two QTL displaying the largest magnitudes with respect to LOD score and R^2 were located on chromosomes 3 and 8 for SDR, on chromosomes 3 and 5 for TL and NT, and on chromosomes 3 and 8 for ANT. QTL for SDR on chromosomes 3 and 5 colocalized with QTL for both TL and NT. A SDR QTL on chromosome 2 colocalized with a TL QTL. All other QTL for trait TL did not colocalize with SDR. However, the QTL on chromosomes 8 and 10 were detected in similar regions for both traits (Figure 10).

Several of the QTL detected for resistance traits colocalized with ANT QTL. The negative genetic correlation between ECB resistance and flowering time observed in the quantitative genetic analysis was partially reflected in the QTL results as opposite signs of allelic effects at the colocalizing QTL between resistance traits and ANT. This was consistently observed across all four parental alleles for the SDR QTL on chromosome 8 (Table 6).

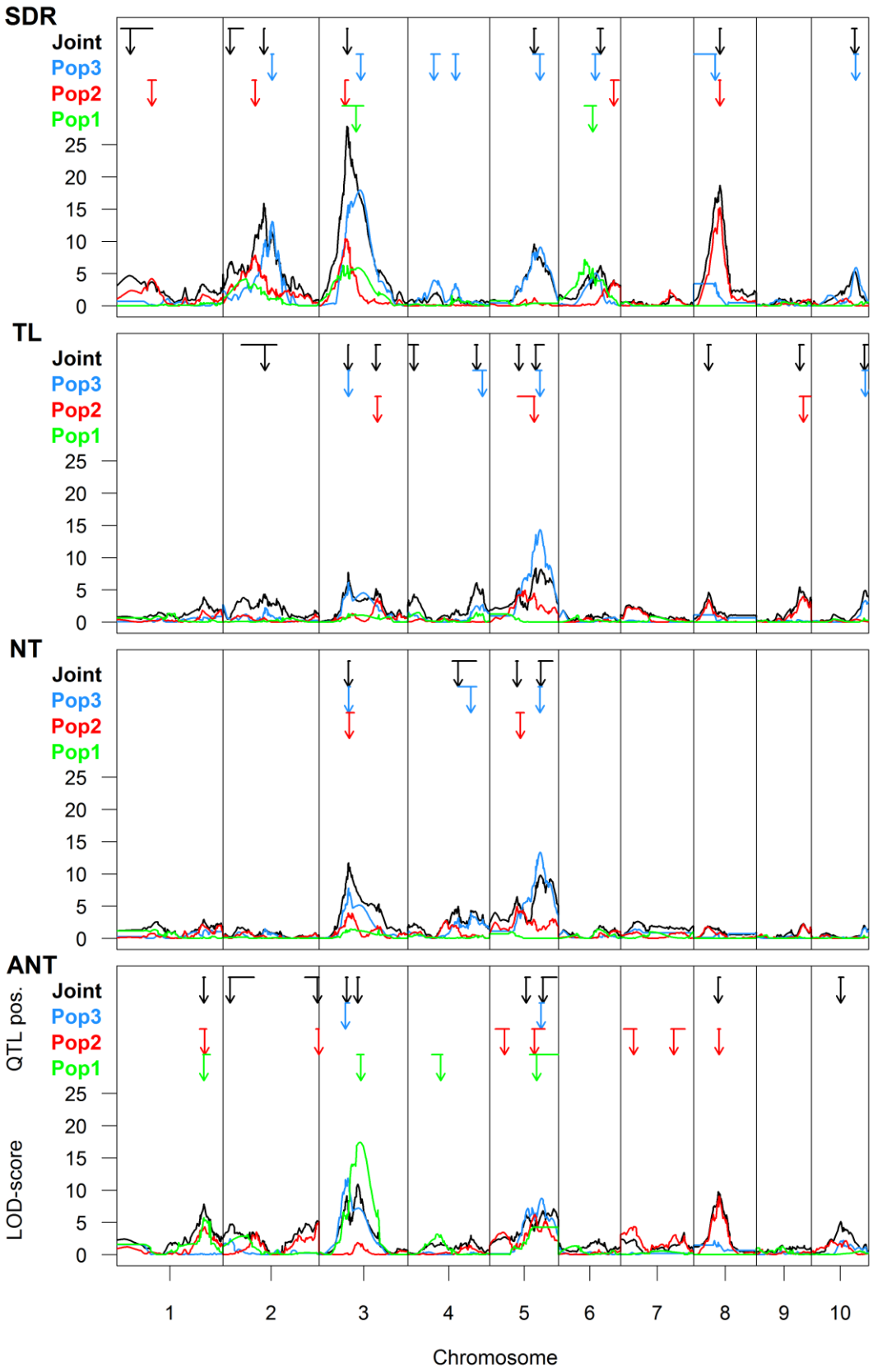

Figure 10 LOD profiles of the *within-population* QTL analyses of Pop1 (*N* = 81, green line), Pop2 (*N* = 214, red line) and Pop3 (*N* = 226, blue line) and of the *joint-population* QTL analysis across the three populations (*N* = 521, black line) based on DH lines *per se* evaluated in 2011 for the traits SDR, TL, NT, and ANT. The arrows indicate the positions of the QTL included in the final model. Horizontal lines attached above each arrow indicate the LOD support interval of the respective QTL position.

3.3.2. *Joint-population analyses in selected DH lines per se and testcrosses*

Compared to 2011, fewer QTL were detected in the selected fractions evaluated in 2012 (*N* = 195) with the only exception of trait NT in the DH lines *per se* (Table 7). However, especially for SDR several of the QTL detected in 2012 were consistent with those identified in 2011: four out of eight QTL detected in 2011 in the unselected populations were identified again in the selected fractions of DH lines *per se* in 2012. For an additional QTL on chromosome 5, the LOD support intervals from the 2011 and 2012 analyses in DH lines *per se* did not overlap only by 1 cM. Three QTL for SDR detected in the DH lines *per se* (on chromosomes 2, 3 and 5) were consistently detected in testcrosses as well (Figure 11). Furthermore, a QTL on chromosome 8 was also detected in testcrosses although its LOD support interval did not overlap with those of the QTL identified for the DH lines *per se*.

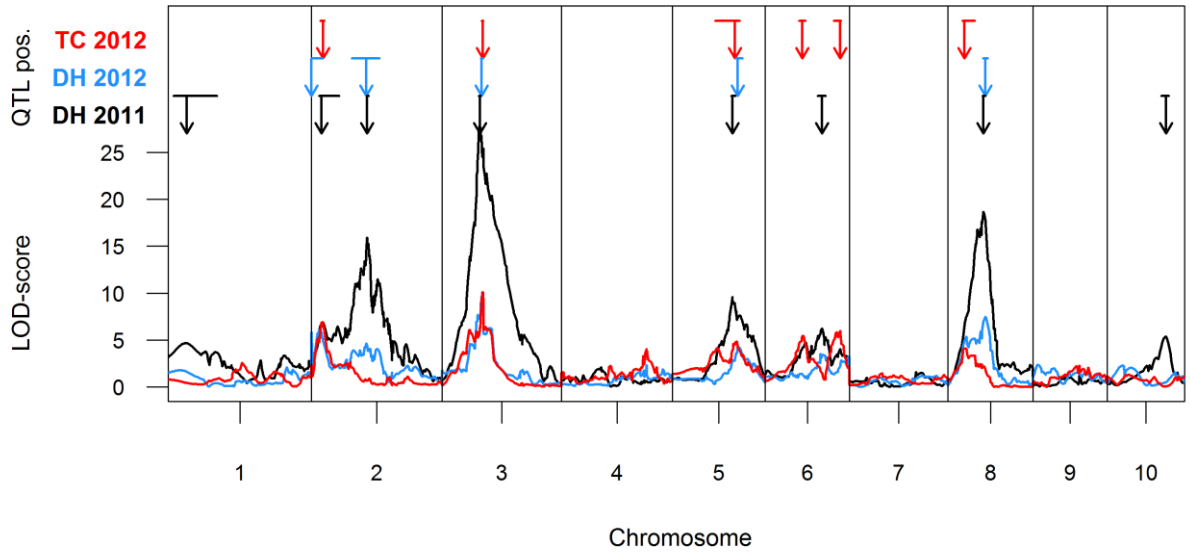

Figure 11 LOD profiles of the *joint-population* QTL analysis for trait SDR across the three unselected populations evaluated as DH lines *per se* in 2011 (DH 2011, black line, *N* = 521), across the selected DH lines *per se* (DH 2012, blue line, *N* = 195) and their corresponding testcrosses (TC 2012, red line, *N* = 195) evaluated in 2012. The arrows indicate the positions of the QTL included in the final model. Horizontal lines attached above each arrow indicate the LOD support interval of the respective QTL position.

The SDR QTL detected in 2012 were those that showed the largest LOD scores and R^2 values in 2011. In contrast, small-effect QTL identified in the unselected populations (*e.g.*, on chromosomes 1 and 10) were not confirmed in the selected fractions in 2012 (Figure 11). For TL, the QTL on chromosome 5 detected in 2012 DH lines *per se* colocalized with a QTL detected in 2011. For the QTL detected on chromosome 8 for TL, the 2011 and 2012 LOD support intervals did not overlap only by 3 cM. With exception of the first QTL on chromosome 5, all QTL for trait NT detected in 2011 were confirmed in the selected DH lines *per se* in 2012.

Contrarily to the results of other traits, two additional QTL that were not identified in 2011 were detected on chromosomes 3 and 6 in 2012 (Table 7). No QTL were detected for TL and NT in testcrosses, a result which is consistent with the low heritability of these two traits in this specific material group.

No QTL were identified for ANT in the selected DH lines *per se* except for a QTL on chromosome 5 (Table 7). This result reflects the consequences of selection for similar flowering time and is consistent with the strongly reduced genetic variance for ANT reported in chapter 3.2.2. The unexpected high heritability for ANT in the testcrosses, particularly for those of Pop2, resulted in the identification of four QTL whereby three of these colocalized with SDR QTL. A *within-population* QTL analysis performed on testcross ANT data revealed that three out of the four QTL found in the *joint*-analysis were consistently identified within Pop2 but not within Pop1 and Pop3 (data not shown), evidencing that Pop2 was indeed the major contributor of ANT QTL in testcrosses.

Table 7 Chromosome (Chr.), position (Pos.), LOD support interval (LOD S.I.), LOD score at the QTL position, proportion of variance explained (R^2) and additive effects of the QTL alleles derived from parents R1, R2, R3 and S1 in the *joint-population* analysis across the selected *N* = 195 DH lines *per se* (left side) and their testcrosses (right side) evaluated in 2012 for the traits SDR, TL, NT and ANT. For each trait, the proportion of phenotypic variance explained by the model fitting all detected QTL simultaneously is given above the R^2 values of the individual QTL.

DH lines *per se*									**Testcrosses**								
Chr.	Pos. [cM]	LOD S.I. [cM]	LOD score	R^2	Additive effects				Chr.	Pos. [cM]	LOD S.I. [cM]	LOD score	R^2	Additive effects			
					R1	R2	R3	S1						R1	R2	R3	S1
SDR [1-9 score]				**0.49**					SDR [1-9 score]				**0.50**				
2	0	(0 – 20)	5.83	0.13	0.12	-0.19	0.02	0.06	2	20	(15 – 23)	6.88	0.16	0.03	-0.06	-0.07	0.10
2	97	(71 – 119)	4.64	0.11	-0.07	0.08	-0.11	0.10	3	72	(69 – 72)	10.11	0.22	0.00	-0.11	-0.02	0.13
3	70	(68 – 72)	9.68	0.20	-0.04	-0.12	0.02	0.15	5	110	(75 – 118)	4.72	0.12	0.03	-0.05	-0.07	0.08
5	115	(111 – 123)	4.31	0.10	0.03	-0.03	-0.10	0.09	6	65	(59 – 71)	5.42	0.13	0.10	-0.05	-0.12	0.07
8	65	(61 -70)	7.47	0.16	0.00	0.15	-0.02	-0.13	6	132	(120 – 133)	6.00	0.14	0.02	-0.07	-0.05	0.10
									8	29	(25 – 47)	4.11	0.10	0.00	0.16	-0.15	-0.01
TL [cm]				**0.20**					TL [cm]				**0.00**				
5	113	(105 – 123)	6.43	0.14	0.62	-0.70	-1.08	1.16									
8	24	(17 – 29)	2.99	0.07	0.00	1.42	-0.80	-0.61									
NT [count]				**0.45**					NT [count]				**0.00**				
3	32	(29 – 43)	6.41	0.14	0.09	0.06	0.01	-0.15									
3	67	(65 – 70)	9.17	0.19	-0.06	-0.13	0.02	0.16									
4	106	(101 – 120)	5.99	0.13	0.08	-0.09	-0.08	0.09									
5	113	(105 – 118)	9.75	0.20	-0.05	0.01	-0.11	0.15									
6	107	(105 – 110)	3.23	0.08	0.09	0.07	-0.16	0.00									
ANT [days]				**0.12**					ANT [days]				**0.46**				
5	151	(148 – 156)	5.56	0.12	0.39	-0.02	0.12	-0.49	2	125	(116 – 130)	8.13	0.18	0.20	-0.44	0.01	0.23
									3	88	(47 – 108)	9.82	0.21	-0.15	0.40	0.05	-0.31
									6	82	(73 – 88)	4.71	0.11	0.23	0.30	-0.43	-0.10
									6	139	(111 – 144)	5.51	0.13	0.16	0.19	-0.05	-0.29

3.4. Predictive ability of QTL-based and genome-wide prediction models

For the traits SDR, TL, NT and ANT the prediction performance of QTL-based and GBLUP models was compared using fivefold CV within the 2011 and 2012 datasets (Figure 4 A). For the 2011 dataset, Table 8 presents mean predictive abilities of the two approaches from *within-population* and *joint-population prediction* scenarios. Mean predictive abilities of GBLUP were consistently higher than corresponding mean predictive abilities of the QTL-based model. Figure 12 shows this comparison by means of scatterplots of the 50 individual *joint-population* predictive ability values calculated from combined test sets across the three populations. For all traits, the predictive ability of GBLUP was higher in every single test set of the CV scheme. Correlations of $r = 0.40$ to $r = 0.55$ were observed between the predictive abilities obtained from the two different approaches. In general, population-trait combinations that showed higher predictive abilities of GBLUP were characterized by higher predictive abilities of the QTL model, and vice versa (Table 8). The higher heritabilities of SDR and ANT when compared to TL and NT (Table 3) generally resulted in higher predictive abilities of both the GBLUP and QTL models. High GBLUP predictive abilities of approximately 0.70 were observed for SDR and ANT in all populations, and Pop3 showed the highest predictive ability of GBLUP for all traits and in both prediction scenarios. The highest mean predictive ability of the QTL model was observed for the *joint-population prediction* of the ANT trait in Pop1. In this specific case the QTL-based model was able to predict almost as accurately as its GBLUP counterpart. On the other hand, no QTL were found for TL and NT in Pop1 and therefore no predictive ability of the QTL model could be estimated for these traits within this population.

The small Pop1 population benefitted more from *joint-population prediction* than the two larger Pop2 and Pop3 populations (Table 8). *Joint-population* predictive abilities in Pop2 and Pop3 were in several cases similar to or only marginally higher than the corresponding values from *within-population prediction*. For the QTL-based prediction of SDR in Pop2 and NT in Pop3, *within-population* predictive abilities were higher than *joint-population predictive abilities*.

Table 8 Mean predictive abilities (± standard deviation) obtained from cross-validation of the QTL-based and GBLUP models for the traits SDR, TL, NT and ANT evaluated in 2011. For *within-population prediction* (Within) and *joint-population prediction* (Joint), predictive abilities calculated based on separate test sets for each population are reported for Pop1, Pop2 and Pop3.

Population	Entries (no.)	Pred. model	SDR		TL		NT		ANT	
			Within	Joint	Within	Joint	Within	Joint	Within	Joint
Pop1	81	QTL	0.45 ± 0.05	0.56 ± 0.06	-	0.07 ± 0.10	-	0.12 ± 0.10	0.60 ± 0.05	0.68 ± 0.05
		GBLUP	0.64 ± 0.03	0.70 ± 0.03	0.15 ± 0.08	0.31 ± 0.06	0.25 ± 0.08	0.37 ± 0.04	0.67 ± 0.04	0.70 ± 0.05
Pop2	214	QTL	0.52 ± 0.03	0.50 ± 0.05	0.22 ± 0.04	0.26 ± 0.03	0.20 ± 0.05	0.30 ± 0.05	0.42 ± 0.06	0.50 ± 0.03
		GBLUP	0.68 ± 0.01	0.69 ± 0.01	0.43 ± 0.02	0.48 ± 0.02	0.47 ± 0.01	0.52 ± 0.01	0.66 ± 0.01	0.66 ± 0.02
Pop3	226	QTL	0.57 ± 0.02	0.58 ± 0.02	0.39 ± 0.03	0.42 ± 0.03	0.51 ± 0.03	0.49 ± 0.03	0.48 ± 0.02	0.50 ± 0.02
		GBLUP	0.72 ± 0.01	0.73 ± 0.02	0.60 ± 0.03	0.61 ± 0.01	0.65 ± 0.02	0.65 ± 0.02	0.72 ± 0.02	0.71 ± 0.02

-, no QTL identified

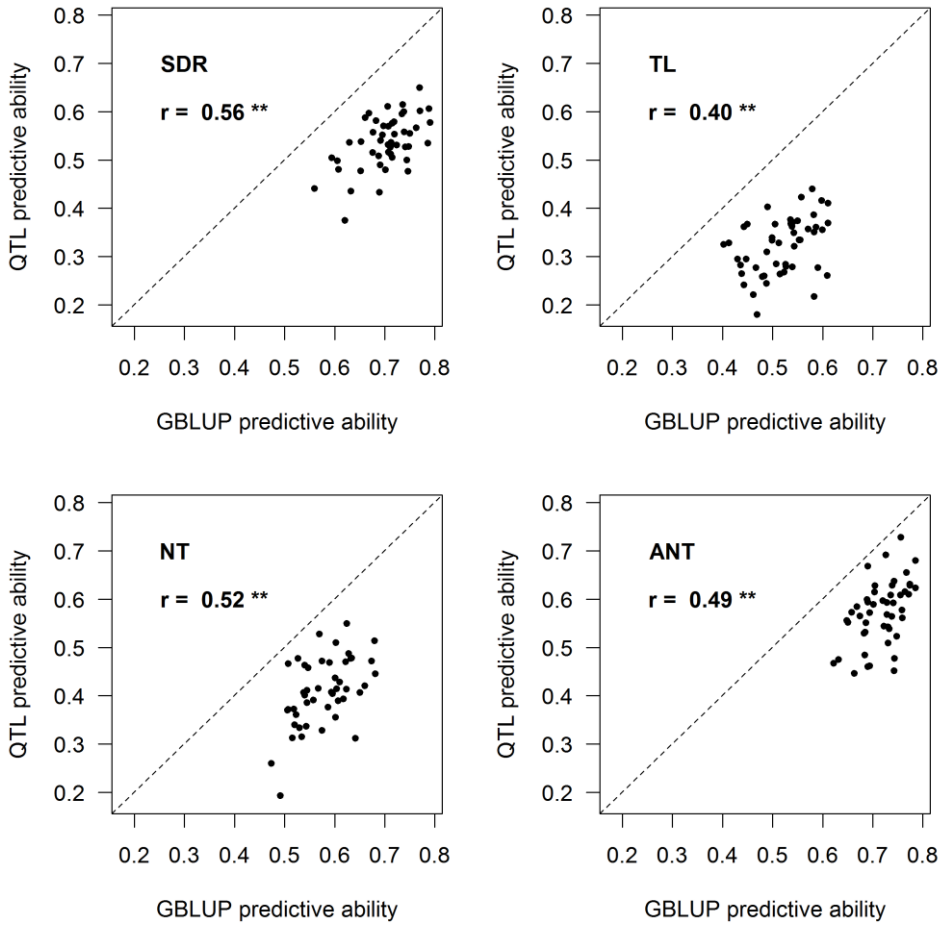

Figure 12 Comparison of the predictive abilities of QTL-based and GBLUP models from the *joint-population prediction* scenario calculated based on combined test sets across the three populations. Predictive abilities from each test set of QTL-based cross-validation are plotted against predictive abilities from the corresponding test set of GBLUP cross-validation. Plots are shown for all four traits based on unselected DH lines *per se* evaluated in 2011. The Pearson's correlation coefficient (r) between the plotted predictive abilities is also given (**, correlation significant at $p<0.01$).

In the 2012 datasets (comprising selected DH lines *per se* and their corresponding testcrosses evaluated in 2012, see Figure 4 A) the prediction performance of GBLUP was also consistently superior to the prediction performance of the respective QTL model (Table 9). In these scenarios the predicted and observed genotypic values were correlated over all populations while separating the resistant and susceptible fractions to avoid a potential selection-induced inflation of predictive abilities. Predictive abilities calculated within the susceptible fraction were generally lower than corresponding values calculated from the resistant fraction and across the complete dataset. These predictive abilities also displayed larger standard deviations due to the reduced TS-size (N = 11-12). In contrast, the magnitude of predictive abilities obtained from the larger resistant fractions were in most cases similar to predictive abilities calculated from the complete dataset. An exception were the values for SDR in the DH lines *per se*, where the

predictive ability was significantly higher when calculated across both fractions than when calculated from the resistant fraction only (Table 9). Given that the selection was based primarily on SDR, selected DH lines showed a clear difference between the means of the resistant and susceptible fractions for SDR in 2012 whereas differences were less pronounced for TL and NT at DH line *per se* level and for SDR, TL and NT at testcross level (Figure 9 and Appendix Figure A 4). At DH line *per se* level, predictive abilities obtained from the resistant fraction were similar for the three resistance traits but considerably lower for ANT. This is a direct consequence of the selection for reduced variation in ANT that was performed simultaneously to the selection on stalk damage traits after the 2011 field trial season. This selection considerably reduced genetic variation for ANT in DH lines *per se* but, surprisingly, not in their testcrosses (Table 3). Hence, in accordance to what was already observed in the phenotypic data and QTL analyses, the unexpected high genetic variance and heritability for ANT resulted in high predictive abilities of both statistical models in the testcrosses. Indeed, the QTL model could predict ANT with almost equal accuracy as the GBLUP model in the testcrosses (Table 9). Conversely, the low heritabilities of tunneling traits TL and NT in testcrosses produced correspondingly low accuracies of prediction.

Table 9 Mean predictive abilities (± standard deviation) obtained with the QTL-based and GBLUP models calculated from 10 replications of fivefold cross-validation for the traits SDR, TL, NT and ANT evaluated in 2012 within the selected DH lines *per se* and their corresponding testcrosses. Predictive abilities were calculated by correlating predicted and observed values across all evaluated genotypes (N = 195) and separately within the resistant (N = 137) and susceptible (N = 58) fractions, respectively.

Fraction	Entries (no.)	Pred. model	SDR	TL	NT	ANT
DH lines *per se* 2012						
Resistant	137	QTL	0.32 ± 0.05	0.42 ± 0.05	0.34 ± 0.07	0.13 ± 0.09
		GBLUP	0.54 ± 0.02	0.58 ± 0.03	0.51 ± 0.03	0.35 ± 0.05
Susceptible	58	QTL	0.26 ± 0.09	0.22 ± 0.06	0.10 ± 0.07	0.24 ± 0.13
		GBLUP	0.26 ± 0.09	0.34 ± 0.03	0.32 ± 0.05	0.23 ± 0.06
Both	195	QTL	0.46 ± 0.05	0.38 ± 0.03	0.33 ± 0.05	0.16 ± 0.07
		GBLUP	0.68 ± 0.02	0.56 ± 0.02	0.53 ± 0.02	0.32 ± 0.03
Testcrosses 2012						
Resistant	137	QTL	0.18 ± 0.07	-	-	0.71 ± 0.03
		GBLUP	0.62 ± 0.03	0.18 ± 0.08	0.34 ± 0.04	0.79 ± 0.01
Susceptible	58	QTL	0.23 ± 0.11	-	-	0.44 ± 0.08
		GBLUP	0.40 ± 0.07	0.17 ± 0.05	0.32 ± 0.05	0.64 ± 0.05
Both	195	QTL	0.30 ± 0.06	-	-	0.67 ± 0.03
		GBLUP	0.63 ± 0.03	0.18 ± 0.06	0.33 ± 0.03	0.78 ± 0.01

-, no QTL identified

A potential integration of the two statistical approaches – QTL-based prediction and genome-wide prediction using GBLUP – was evaluated in the 2011 dataset for the trait SDR. Fitting the QTL detected in each ES as fixed effects in GBLUP during CV slightly decreased prediction performance in both *within* and *joint-population prediction* scenarios (Table 10). In order to evaluate the estimation of QTL effects at a predefined position of the genome within the GBLUP framework, predictive abilities of QTL models were compared with predictive abilities calculated based on fixed QTL effects estimated in the mixed GBLUP model by removing the random effects of all remaining genome-wide SNP markers in the calculation of predicted values in the TS. The obtained mean predictive abilities were similar in magnitude and standard deviation to the predictive abilities of the respective QTL models (see last two columns of Table 10).

Table 10 Mean predictive abilities (±standard deviation) of the following prediction approaches: the standard genomic BLUP model (GBLUP), a variant of GBLUP including detected QTL as fixed effects in the model (GBLUP + QTL), the standard QTL-based model using QTL-effects estimated from the MCQTL software (QTL – MCQTL est.) and a prediction solely based on the fixed QTL effects estimated within the mixed GBLUP model (QTL – GBLUP est.). The trait is stalk damage rating evaluated within the unselected DH populations (Pop1, Pop2, Pop3) and in the *joint-population* scenario (*Joint-pop.*).

Population	Entries (no.)	GBLUP	GBLUP + QTL	QTL - MCQTL est.	QTL - GBLUP est.
Pop1	81	0.64 ± 0.03	0.56 ± 0.04	0.45 ± 0.05	0.40 ± 0.06
Pop2	214	0.68 ± 0.01	0.63 ± 0.02	0.52 ± 0.03	0.51 ± 0.03
Pop3	226	0.72 ± 0.01	0.69 ± 0.02	0.57 ± 0.02	0.54 ± 0.02
Joint-pop.	521	0.70 ± 0.01	0.66 ± 0.01	0.54 ± 0.02	0.51 ± 0.02

3.5. Testcross prediction

Testcross prediction refers to genome-wide prediction of testcross performance based on GBLUP models trained at the DH line *per se* level (Figure 4 B). Moderate predictive abilities between 0.40 and 0.46 were achieved for SDR and ANT, while predictive abilities were low for NT and close to zero for TL (Table 11). The three sub-scenarios (denoted with A-C in Table 11) did not significantly differ in their predictive abilities except for ANT, which was predicted 6 % more accurately when using only data from DH lines *per se* evaluated in 2011. Hence, for all traits, training the model by adding phenotypic information of the selected DH lines *per se* evaluated in 2012 (Table 11 B and C) did not increase prediction performance of testcrosses. The predictive abilities in Table 11 were calculated based on the resistant fraction

of the selected lines. When correlations between predicted genotypic and observed phenotypic values were calculated using all selected lines, *i.e.* the resistant and the susceptible fraction, predictive abilities increased to approximately 0.50 for both SDR and ANT, to approximately 0.10 for TL but did not significantly increase for NT.

Table 11 Mean predictive abilities (±standard deviation) of the GBLUP model for *testcross prediction* of the traits SDR, TL, NT and ANT. The model was trained using the same number of genotypes (N = 521) but varying the number of replicated phenotypic observations across years (A-C). Predictive abilities are given for the resistant fraction of testcrosses.

Estimation set (ES)	Test set (TS)	SDR	TL	NT	ANT
A) DH lines *per se* 2011	Testcrosses 2012	0.43 ± 0.02	0.06 ± 0.02	0.24 ± 0.02	0.46 ± 0.01
B) DH lines *per se* 2011 + res. fract. 2012	Testcrosses 2012	0.44 ± 0.02	0.04 ± 0.02	0.25 ± 0.02	0.40 ± 0.02
C) DH lines *per se* 2011 + both fract. 2012	Testcrosses 2012	0.43 ± 0.02	0.04 ± 0.02	0.24 ± 0.02	0.40 ± 0.02

3.6. Across-population prediction

Results from *across-population prediction* scenarios (Figure 4 C) are summarized in Table 12. Mean predictive abilities were in general significantly lower compared to predictive abilities observed in *within-* and *joint-population prediction* scenarios. A notable exception was observed for TL and NT in the prediction of Pop3 using GBLUP models trained with Pop2 as well as with an ES comprising genotypes of both Pop1 and Pop2. In these particular scenarios *across-population* predictive abilities were in the same order of magnitude as values observed for *joint-population prediction* using combined TSs across the three populations (approximately 0.50 to 0.55).

The variation in *across-population* prediction performance generally reflected the varying degree of relatedness between the three resistant parental lines as seen in Table 2. For all four traits, training the model with DH lines from Pop1 resulted in significantly higher prediction accuracy for DH lines of the more related Pop3 than for DH lines from Pop2. Except for the NT trait, using DH lines from Pop1 in addition to lines from Pop2 for model training significantly increased the accuracy of predicting genotypic values of lines from Pop3. In contrast, when predicting genotypic values of lines from Pop2 the addition of genotypes from Pop1 did not increase or even slightly decreased the predictive ability that was already achieved

with DH lines from Pop3 alone (with the sole exception of a slight increase in predictive ability for ANT– from 0.31 to 0.36). Despite being the least related family among the three, Pop2 was able to predict Pop1 and especially Pop3 with generally good accuracy.

Table 12 Mean predictive abilities (± standard deviation) from *across-population prediction* using the GBLUP model. Predictions were performed across the biparental populations Pop1, Pop2 and Pop3 using phenotypic data evaluated in unselected DH lines *per se* for traits SDR, TL, NT and ANT in 2011. Numbers in brackets represent the number of genotypes included in the respective estimation and test sets.

Estimation set (ES)	Test set (TS)	SDR	TL	NT	ANT
Pop1 (81)	Pop2 (214)	0.36 ± 0.06	0.16 ± 0.07	0.26 ± 0.07	0.23 ± 0.06
Pop1 (81)	Pop3 (226)	0.57 ± 0.04	0.39 ± 0.05	0.39 ± 0.05	0.53 ± 0.04
Pop2 (214)	Pop1 (81)	0.50 ± 0.07	0.11 ± 0.09	0.30 ± 0.09	0.40 ± 0.09
Pop2 (214)	Pop3 (226)	0.53 ± 0.04	0.49 ± 0.05	0.55 ± 0.04	0.56 ± 0.05
Pop3 (226)	Pop1 (81)	0.59 ± 0.08	0.36 ± 0.09	0.34 ± 0.09	0.58 ± 0.06
Pop3 (226)	Pop2 (214)	0.42 ± 0.06	0.39 ± 0.06	0.41 ± 0.06	0.31 ± 0.07
Pop2 + Pop1 (295)	Pop3 (226)	0.61 ± 0.04	0.55 ± 0.04	0.57 ± 0.04	0.63 ± 0.04
Pop3 + Pop1 (307)	Pop2 (214)	0.40 ± 0.06	0.39 ± 0.06	0.41 ± 0.06	0.36 ± 0.06

Across-population prediction was also performed by keeping the sample size of all three populations equal, *i.e* through a random reduction of the sample size of Pop2 and Pop3 to match the size of the smaller Pop1 (N = 81) in the respective ES and TS. Mean predictive abilities were in most cases similar (in both magnitude and standard deviation) to the corresponding values obtained when using Pop2 and Pop3 in their respective original sample size (Appendix Table A 3).

3.7. Across-location prediction

To maintain comparability of results with the other GP scenarios in this study, all CV schemes for the *across-location prediction* scenario involved sampling mutually exclusive subsets of genotypes for the respective ES and TS. In a first step, the ES and TS included phenotypic data drawn from the same locations. This CV scheme corresponds to the one used in *joint-population prediction*, but instead of using adjusted means averaged across all six locations, either individual locations or different combinations of two or more locations were used to form the phenotypic data. Results thereof are presented as bold diagonal values in Tables 13 and 14 and

are hereafter referred to as *within-location* predictive abilities. In a second step, CV was implemented by sampling different genotypes as well as different locations for the respective ES and TS (see example of CV scheme in Figure 5). These results are shown in the off-diagonal values of Tables 13 and 14 and are hereafter referred to as *across-location* predictive abilities. Table 13 displays results from prediction of SDR for the six individual locations using unselected DH lines *per se* evaluated in 2011. All predictive abilities involving single environments were significantly lower compared to the value of 0.70, which was obtained in the corresponding *joint-population prediction* scenario using SDR BLUEs averaged across all six locations. The highest value (0.64) was reported for prediction of SDR within the Italian location FER. This naturally infested field environment displayed the most severe mean level of ECB damage as well as the highest single-plot repeatability for SDR (Table 13). *Within-location* predictive abilities were low for HER (which also showed the overall lowest level of damage) and moderate for the remaining locations. The largest *across-location* predictive abilities ranged between 0.36 and 0.47 and were reported for different combinations of the two artificially infested locations (FRS and MZH_i) and the two locations under higher natural ECB pressure (FER and TOS). All CV scenarios involving the two German environments characterized by low natural pressure (MZH_n and HER) resulted in low *across-location* predictive abilities, except for the prediction of MZH_n using a model trained with HER-data (Table 13). Accuracies of prediction between MZH_i and MZH_n were low even though both trials were planted at the same field site.

Based on the results presented in Table 13, locations were grouped in determined subsets to (i) investigate the effects of a reduction in the number of screening locations on predictive ability and (ii) estimate the mutual accuracies of predicting SDR across trials under natural and under artificial infestation, respectively. Results from these prediction scenarios are given in Table 14. A model training and validation performed with adjusted genotype means averaged across the four locations under highest ECB pressure was sufficient to reach the same level of predictive ability obtained when using the complete dataset with six locations (> 0.70). The heritability of SDR estimated for this subset was almost as high as the heritability observed across the six environments (see Table 3). Remarkably, predictive ability obtained when using only two locations under high natural pressure was just slightly lower compared to values resulting from four or six environments. As expected, predictions across mutually exclusive location subsets were significantly less accurate than predictions across subsets having one or more locations in common (Table 14). Predictive abilities across mutually exclusive subsets reflected the values obtained for prediction across single locations: the highest predictive abilities (approximately

0.47 to 0.50) were achieved for predictions between the artificially infested subset (FRS and MZH_i) and the high natural infestation subset (FER and TOS) whereas lower predictive abilities were observed for predictions between the combination of low infestation environments (HER and MZH_n) and the other subsets.

Table 13 Mean predictive abilities (±standard deviation) of the GBLUP model obtained for prediction of trait SDR across the six individual field locations Freising (FRS), Münzesheim artificially infested (MZH_i), Münzesheim naturally infested (MZH_n), Herbolzheim (HER), Tournoisis (TOS), and Ferrara (FER). This prediction scenario is based on the unselected DH lines *per se* (*N* = 521) evaluated in 2011. The model was trained using adjusted genotype means obtained at the locations displayed vertically on the left side, and validated against adjusted means obtained at the locations displayed horizontally in the fourth row. The first three rows contain the mean and single-plot repeatability for SDR calculated across *N* = 521 DH lines as well as the type of ECB infestation at the respective location.

Mean SDR	3.36	2.91	2.22	1.52	2.90	4.34
Repeatability	0.34	0.24	0.30	0.23	0.42	0.56
ECB infestation	artificial	artificial	natural	natural	natural	natural
	FRS	**MZH_i**	**MZH_n**	**HER**	**TOS**	**FER**
FRS	**0.57 ± 0.01**	0.41 ± 0.01	0.29 ± 0.01	0.21 ± 0.01	0.44 ± 0.01	0.30 ± 0.01
MZH_i	0.47 ± 0.01	**0.45 ± 0.02**	0.29 ± 0.01	0.19 ± 0.01	0.41 ± 0.02	0.44 ± 0.01
MZH_n	0.28 ± 0.01	0.24 ± 0.01	**0.53 ± 0.01**	0.28 ± 0.01	0.25 ± 0.01	0.19 ± 0.01
HER	0.22 ± 0.01	0.19 ± 0.02	0.41 ± 0.01	**0.28 ± 0.02**	0.28 ± 0.02	0.23 ± 0.01
TOS	0.47 ± 0.01	0.39 ± 0.01	0.28 ± 0.01	0.26 ± 0.01	**0.51 ± 0.02**	0.44 ± 0.01
FER	0.27 ± 0.01	0.36 ± 0.01	0.20 ± 0.01	0.18 ± 0.01	0.38 ± 0.01	**0.64 ± 0.01**

Table 14 Mean predictive abilities (±standard deviation) of the GBLUP model obtained for prediction of trait SDR across different combinations of the field locations Freising (FRS), Münzesheim artificially infested (MZH_i), Münzesheim naturally infested (MZH_n), Herbolzheim (HER), Tournoisis (TOS), and Ferrara (FER). This prediction scenario is based on the unselected DH lines *per se* (*N* = 521) evaluated in 2011. The model was trained using adjusted genotype means averaged across the locations displayed vertically on the left side, and validated against adjusted means averaged across the locations displayed horizontally in the fourth row. The first three rows contain the mean and heritability (h^2) for SDR calculated across *N* = 521 DH lines as well as the type and overall level of ECB infestation for the respective combination of locations.

Mean SDR	3.15	3.55	3.40	1.88	3.65	4.34
h^2	0.57	0.57	0.67	0.39	0.50	-
ECB infestation type / level	artificial	artificial+ high natural	artificial+ high natural	low natural	high natural	high natural
	FRS+MZH_i	**FRS+MZH_i +FER**	**FRS+MZH_i +FER+TOS**	**MZH_n+HER**	**FER+TOS**	**FER**
FRS+MZH_i	**0.61 ± 0.02**	0.60 ± 0.01	0.63 ± 0.01	0.33 ± 0.01	0.50 ± 0.01	0.40 ± 0.01
FRS+MZH_i +FER	0.56 ± 0.01	**0.69 ± 0.01**	0.70 ± 0.01	0.32 ± 0.01	0.65 ± 0.01	0.59 ± 0.01
FRS+MZH_i +FER+TOS	0.57 ± 0.01	0.69 ± 0.01	**0.71 ± 0.01**	0.34 ± 0.01	0.66 ± 0.01	0.58 ± 0.01
MZH_n+HER	0.33 ± 0.01	0.34 ± 0.01	0.37 ± 0.01	**0.53 ± 0.01**	0.31 ± 0.01	0.24 ± 0.01
FER+TOS	0.47 ± 0.01	0.66 ± 0.01	0.68 ± 0.01	0.31 ± 0.01	**0.68 ± 0.01**	0.63 ± 0.01
FER	0.38 ± 0.01	0.61 ± 0.01	0.62 ± 0.01	0.24 ± 0.01	0.65 ± 0.01	**0.64 ± 0.01**

4. DISCUSSION

4.1. The potential of genome-wide prediction for ECB resistance improvement

Improving native resistance to ECB stalk damage in maize is challenging and resource-intensive (Cardinal et al. 2001, Papst et al. 2001, Krakowsky et al. 2004). Resistance breeding activities require expensive artificial infestation programs to guarantee a sufficiently high and homogeneous level of ECB pest pressure in at least a subset of screening locations, because natural ECB infestation is often unreliable, especially under Central European climatic conditions. In addition, the phenotypic evaluation of ECB stalk tunneling is extremely laborious, requiring manual harvest, dissection and evaluation of thousands of individual maize plants.

This study evaluated the potential of genome-wide prediction for the improvement of resistance to ECB stalk damage in Central European elite maize. Intermediate to high prediction accuracies were obtained for all resistance traits at the DH line *per se* level, and for SDR at the testcross level. In general, accuracies of GP were comparable to or even higher than those reported by Technow et al. (2013) for Northern corn leaf blight resistance in maize and by Rutkoski et al. (2012) for *Fusarium* head blight resistance in wheat. Considering the costs and difficulties of phenotyping, the obtained results demonstrate that GP can be an effective strategy for increasing genetic gain per unit time by predicting the resistance level of unphenotyped individuals based on their DNA profile. For ECB stalk damage resistance, GP might be highly cost-efficient already when applied within breeding cycles, for example by training GP models using reduced subsets of DH lines in each cross and predicting the resistance level of the progenies without phenotypes. Implementation of GP in practical breeding should be facilitated if sources of resistance are recombined with elite material in earlier selection cycles, as it was carried out in this study, to allow for the simultaneous genome-based prediction of resistance and other agronomically important traits.

4.2. Comparison of the QTL-based and genome-wide approaches

In resistance breeding, strategies based on QTL mapping results have been implemented successfully for traits influenced by at least one or a few major QTL, like for the trait *Fusarium* head blight resistance in wheat (Miedaner and Korzun 2012). For such traits, QTL-based

approaches allowed a sufficiently accurate identification and characterization of specific resistance alleles and their directed introgression and/or pyramidization into elite genetic backgrounds. Targeted combinations of alleles influencing different, even negatively correlated traits may also be feasible with a QTL approach. Especially when resistance alleles originate from non-adapted genetic sources the introgression of individual genomic regions into elite material might be advantageous over genome-wide approaches. In GP, emphasis is not placed on specific regions but on whole-genome marker profiles which are used to predict the total genetic merit of individuals. This shift of focus needs the implementation of a different breeding strategy, *i.e.*, marker-assisted recurrent selection in elite breeding populations instead of marker-assisted backcrossing involving non-adapted material (Bernardo 2008). Hence, depending on the genetic architecture of the trait and on the type of plant material either a QTL- or a genome-based approach can be more appropriate for MAS. The results obtained in this work confirmed the availability of significant genetic variation for resistance to ECB stalk damage in elite maize germplasm adapted to Central European climatic conditions. Therefore, the plant material used should be suitable for a potential direct implementation of a marker-based approach in practical breeding.

A major goal of the present study was to compare the two methods with respect to prediction of genotypic values. A comprehensive empirical dataset was used, which included data on ECB stalk damage and agronomic traits from two seasons of field trials conducted in up to six environments per year. The large sample sizes of Pop2 and Pop3 (more than 200 DH lines each) and the dense genome coverage with SNP markers were a good precondition for efficient QTL and GP analyses. Using three interconnected DH populations allowed the analysis of a *joint-population* scenario with a considerable overall sample size of $N = 521$ DH lines tested during the 2011 field season. In addition, this study addressed the relationship between line *per se* and testcross performance in both the QTL mapping and GP frameworks, thanks to the parallel evaluation as lines *per se* and testcrosses of the selected subsets of DH lines in 2012.

In the unselected populations, some of the identified QTL explained a significant proportion of the total phenotypic variance, in particular for the SDR trait (*e.g.*, QTL on chromosomes 2, 3 and 8 with R^2 values of 0.13, 0.21 and 0.15, respectively). The positions of these three loci, along with that of an additional QTL on chromosome 2, have been confirmed by overlapping LOD support intervals in the selected DH lines *per se* evaluated in 2012. Furthermore, three QTL identified in DH lines *per se* were consistently detected in testcrosses as well (Figure 11). Thereby the QTL on chromosome 3 always displayed the highest magnitude among SDR QTL in terms of both R^2 value and LOD score. A further locus that might be of particular interest

was the QTL on chromosome 5, which colocalized with a QTL for the TL trait. The position of this locus showed good congruency with positions of QTL detected for TL in several previous studies (Bohn et al. 2000, Cardinal et al. 2001, Jampatong et al. 2002, Krakowsky et al. 2004, Papst et al. 2004). For the three resistance traits, the largest individual allelic substitution effects extrapolated from the *within-population* QTL analyses of the unselected Pop1, Pop2 and Pop3 ranged between approximately 0.20 and 0.30 stalk damage rating scores (SDR), between 1.20 and 2.00 cm of tunnel length (TL), and between 0.15 and 0.30 ECB tunnels (NT) (Appendix Table A 2). These values are comparable to the largest effects reported in most earlier Central European and North American QTL studies (Bohn et al. 2000, Cardinal et al. 2001, Jampatong et al. 2002, Krakowsky et al. 2004, Papst et al. 2004) with the exception of studies by Schön et al. (1993) and Krakowsky et al. (2002) which reported effects of up to 5.5 cm for TL.

In summary, QTL with sizeable effects could be identified for all resistance traits, and some of these QTL also displayed good consistency across years and at both the line *per se* and testcross level. However, as has been shown previously, QTL effects are optimistically biased due to model selection unless they are derived from an independent validation sample (Schön et al. 2004). Therefore, a cross-validation R routine for MCQTL was developed to obtain unbiased estimates of the proportion of variance explained by the *within-* and *joint-population* QTL mapping approaches. After bias correction, statistical models fitting putative QTL identified for the three resistance traits explained only a small proportion of the respective total phenotypic variance (*i.e.*, 0.29, 0.10, and 0.17 for SDR, TL, and NT in the unselected DH populations, respectively). These results corroborate the polygenic nature of resistance to ECB stalk damage reported in previous studies (Schön et al. 1993, Melchinger et al. 1998a, Krakowsky et al. 2002, Papst et al. 2004) and suggest that for the marker-based improvement of ECB resistance a genome-wide approach might be more promising.

This conclusion was confirmed by results achieved with the GBLUP analysis. In the unselected populations, predictive abilities obtained with GBLUP for the *joint-population* scenario were on average 31% higher for SDR and even 62% for TL than those obtained by QTL-based prediction. In a comparison across 50 CV test sets predictive abilities obtained with GBLUP surpassed those obtained with the QTL-based approach without exception (Figure 12). Results obtained from *joint-population prediction* using the selected DH lines *per se* and their testcrosses evaluated in 2012 further consolidated the superior predictive power of the GBLUP over the QTL model.

A similar comparison of GBLUP and QTL-based prediction was carried out by Peiffer et al. (2013) for stalk strength in maize using nearly 5,000 inbred lines including the U.S. nested

association mapping panel (McMullen et al. 2009b). Like in this study, the authors reported cross-validated predictive abilities of GBLUP that were consistently higher compared to those of QTL models.

Several factors contribute to the difference in prediction performance between the QTL-based and the GBLUP approach. For one, the genetic architecture of the trait under study has a strong influence on the relative efficiency of the two methods. For highly polygenic traits it is difficult to achieve accurate estimates of both QTL positions and effects as model selection becomes a challenge when the contribution of individual genes to quantitative trait variation is small. Estimates of effects for the significant QTL are usually inflated (Melchinger et al. 1998b) because effects pertaining to undetected QTL in other genomic regions are incorrectly absorbed by the few QTL that remain in the model (Lorenz et al. 2012). However, even with large-effect QTL it can be difficult to select a set of markers that maximizes prediction accuracy, if the population exhibits long-range linkage disequilibrium and low trait heritability (Wimmer et al. 2013). These characteristics are commonly found in elite plant breeding populations like the ones used in the present study. Here, the relatively large LOD support intervals of QTL positions and the more pronounced superiority of GBLUP over the QTL-based approach for the two traits with lower trait heritability (TL and NT) point in this direction.

Recent studies investigated the potential for a better utilization of major QTL in the GP framework, in order to further increase predictive ability of GP models (Bernardo 2014, Zhang et al. 2014, Zhao et al. 2014). In general, these studies showed that accounting for specific QTL in a GP model is advantageous, provided that one or a few well-characterized QTL with major and stable effects are available for the trait under consideration. In the GP study by Rutkoski et al. (2012) on *Fusarium* head blight resistance in wheat, the resistance trait deoxynivalenol (DON) content could even be predicted more accurately when including only a few QTL-targeted markers in the model in contrast to retaining all genome-wide markers in the analysis. Therefore, the potential of an integrated QTL-GP approach was also investigated in this study, but the results showed that specifically accounting for QTL information in the GBLUP model did not increase predictive ability (Table 10). In such a prediction scenario involving sufficient marker coverage and related genotypes, the standard GBLUP model seems to adequately capture the effects of important QTL. Zhao et al. (2014) showed that the integrated approach can be effective when the two sources of information complement each other. This study performed genome-based prediction of heading time and plant height across *unrelated* wheat genotypes and with an overall lower marker coverage (as is typical for wheat), thereby additionally exploiting information provided by three well established functional loci (*Ppd-D1*,

Rht-B1 and *Rht-D1*). Another reason for the different results obtained in the present study could be the genetic architecture of ECB stalk damage resistance in maize, which may be more complex than the respective genetic architectures of *Fusarium* head blight resistance, plant height and heading time in wheat. In addition, no previous knowledge about major QTL was available for ECB stalk damage traits in the plant material under study. QTL first needed to be mapped and therefore the most realistic prediction scenario for a combined QTL-GP approach was to integrate the QTL detection step in the cross-validation procedure. Thereby the SNP markers closest to each of the QTL detected in a given ES were included as fixed effects in a GBLUP model that was trained based on the same ES used for QTL detection. Zhang et al. (2014) proposed a different method for systematically exploiting the wealth of published QTL information (as SNP markers detected in either QTL mapping or genome-wide association studies) through integration into the genomic relationship matrix of GBLUP. It remains to be seen if this approach could also be valid for the prediction of ECB resistance in maize. From a broader perspective, the development and refinement of statistical methods allowing an efficient utilization of pre-existing QTL-information in GP is a subject that deserves further research.

4.3. Genome-wide prediction scenarios

The obtained results demonstrated the superior predictive performance of GP over a QTL-based approach for native resistance to ECB stalk damage in Central European elite maize. As a consequence of this outcome, a further goal of this study was to compute predictive abilities of GP models for different prediction scenarios in order to evaluate potential strategies for implementation of GP in ECB resistance breeding.

4.3.1. *Within-, joint-, across-population prediction*

Similarly to the increased statistical power of QTL detection that can be expected from *joint-population* QTL analyses in multi-parental designs compared to analyses within individual biparental populations (Rebai and Goffinet 2000, Jansen et al. 2003), an increase in predictive ability of GP models is expected when several populations are combined, primarily due to the larger ES size (Asoro et al. 2011, Technow et al. 2013). However, for Pop2 and Pop3 the predictive abilities of *within-population prediction* were almost as high as or even higher than those obtained from *joint-population prediction* despite the substantially smaller sample sizes of the respective ES. Only the small Pop1 population benefitted significantly from *joint-*

population prediction. A simulation study by Lorenz (2013) showed that above a population size of N = 125 DH lines only marginal gains in predictive ability could be obtained, both for traits with high (0.6) and low (0.2) trait heritability. This could at least in part explain the generally lower predictive abilities estimated within Pop1 (N = 81) with a population size well below N = 125 compared to those estimated within Pop2 (N = 214) and Pop3 (N = 226). Especially the two traits with lower trait heritability (TL and NT) showed substantially lower predictive abilities within Pop1 relative to predictive abilities obtained within the two larger populations. This result was expected, since with decreasing trait heritability an increased minimum ES size is needed in order to accurately estimate marker effects (De Roos et al. 2009). Another explanation for the similar *joint-* and *within-population* predictive abilities could be that at least partially different resistance alleles are segregating in the three populations, a notion which is supported by results from the QTL analysis. For example, among the larger QTL identified for SDR, one of the QTL on chromosome 2 and the QTL on chromosome 8 were mainly segregating in Pop2, while Pop3 provided the major allelic effects at the second QTL on chromosome 2 and the QTL on chromosome 5 (Table 6). Only the two QTL on chromosomes 3 and 6 were detected in all three populations and effects were congruent in sign for the three resistant parents. The fact that several QTL appeared to be population-specific might also explain why no clear advantage in terms of number of detected QTL and magnitude of QTL effects was found when applying a *joint-population* QTL analysis. Indeed, all QTL possessing sizeable effects could be detected also in either one or more of the *within-population* analyses (Figure 10). Nevertheless, other advantages of the *joint-population* analysis could be observed, such as more precise estimates of the positions for some of the detected QTL, as well as the possibility to simultaneously compare the magnitudes and signs of additive effects between all four parental lines.

The *across-population prediction* scenario also supported the hypothesis of different resistance alleles contributed by the three resistant parents. Prediction accuracies that could be obtained across populations were in good agreement with the level of relatedness among parental lines. Resistant parents R1 and R3 showed higher pairwise molecular genetic similarity with each other than each of them with R2 (Table 2) and consequently the mutual predictive abilities of Pop1 and Pop3 were in general substantially higher than their respective predictive abilities for DH lines derived from Pop2. Interestingly, predictive abilities in the *across-population prediction* involving Pop2 differed significantly for reciprocal scenarios. Substantially higher predictive abilities could be obtained when Pop2 constituted the ES and Pop1 and Pop3 the TS as compared to when either Pop1 or Pop3 - or both - constituted the ES and Pop2 the TS. The

effect remained significant also when accounting for differences in sample size of the estimation sets (Appendix Table A 3). These results were observed consistently across all traits and mutual prediction scenarios involving Pop2, with the sole exception of trait TL for predictions with Pop1, probably due to the very low mean predictive abilities for TL in these specific scenarios. Because of the higher genetic distance of R2 to the susceptible line S1, it can be hypothesized that a higher number of predictive haplotypes may have formed in Pop2 compared to Pop1 and Pop3 thus leading to a higher effective sample size and higher predictive power of Pop2.

These results indicate that the choice of population employed in model training can significantly affect predictive abilities. Parental lines should be chosen to maximize the effective sample size of a training population especially when breeding schemes allow for only a few recombination events. If different sources of resistance are available for a given trait, model training can be performed in multi-parental populations, thus ensuring predictive power across a diverse spectrum of resistance alleles and genetic backgrounds.

4.3.2. *Testcross prediction*

The genotypic correlation between line *per se* and testcross performance is an important parameter for the optimization of hybrid breeding programs (Mihaljevic et al. 2005). If this correlation is sufficiently large, a selection based on line *per se* performance may be advantageous due to the larger additive genetic variance that can be exploited in inbred lines, the saving of costs related to the production of testcross seed, and the accelerated phenotypic evaluation (Miedaner et al. 2014). Moreover, both artificial ECB infestation and field scoring of ECB resistance traits are technically more challenging to perform on testcrosses than on inbred lines *per se,* due to the higher plants and overall denser canopies of testcross trials. Hence, focusing on DH line *per se* evaluation may also provide more accurate ECB resistance data for the training of GP models. The reduced heritabilities of the ECB stalk damage traits observed in testcrosses when compared to DH lines *per se*, especially for tunneling traits TL and NT, might be at least to some extent a consequence of the lower phenotyping accuracy.

So far, empirical GP studies conducted in maize have either performed both model training and validation at the testcross level (Lorenzana and Bernardo 2009, Albrecht et al. 2011, Riedelsheimer et al. 2012, Rincent et al. 2012, Windhausen et al. 2012, Zhao et al. 2013, Crossa et al. 2014) or at the line *per se* level (Guo et al. 2012, Peiffer et al. 2013, Riedelsheimer et al. 2013, Technow et al. 2013). To the best of my knowledge, this is the first study that explores the relationship between line *per se* and testcross performance in a GP framework, *i.e.* by

validating models trained at the line *per se* level against phenotypic data observed in the corresponding testcrosses.

Given that the correlation between predicted and observed values was calculated for the selected fraction of resistant DH lines only, and prediction was performed across different years, predictive abilities were sufficiently high for SDR, exceeding 0.43 in all scenarios. These results can be explained by the high genotypic correlation between DH lines *per se* and testcrosses evaluated for SDR in 2012 (Table 5) and the absence of significant genotype-by-year interaction variance for resistance traits (data not shown). The three *testcross prediction* scenarios yielded almost identical results for ECB stalk damage traits, *i.e.*, including additional data from selected DH lines *per se* evaluated in 2012 in model training did not lead to a significant increase in predictive abilities for testcrosses. This result can also be explained by the non-significant genotype-by-year interaction variance observed for SDR, TL and NT. Hence, phenotypic data collected in 2012 did not add new information relevant to improving model training.

For tunneling traits, accuracies of *testcross prediction* were low due to the low heritabilities of TL and NT in the testcrosses. These results are in good agreement with Kreps et al. (1998a), where genotypic correlations between DH line *per se* and testcross performance were also high for SDR and non-significant to moderate for stalk tunneling. In our study, higher correlations between line *per se* and testcross performance were observed for NT as compared to TL and as a consequence also the accuracy of *testcross prediction* was higher for NT. Although no significant genetic variance was found for NT in the testcrosses of Pop1 and Pop2, significant genetic variance and an intermediate heritability value of 0.45 could be observed in the testcrosses of Pop3 (Table 3). This was in contrast to results obtained for TL where none of the three populations displayed significant genetic variance at the testcross level. For TL, significant but low genetic variance was detected only in a combined analysis of testcross data across the three populations. Mean stalk tunnelling damage was significantly lower in testcrosses as compared to their corresponding DH lines *per se* evaluated in 2012 (Table 3). Artificial ECB infestation was carried out for the two field environments in which TL and NT were scored, and it might have been less effective for the testcrosses than for the inbred lines *per se*. The success of artificial infestation could be related significantly to the developmental stage of the plants during the first days of larval establishment. Differences in plant development were accounted for by conducting artificial infestation for testcrosses one week earlier than for the inbred lines. Even this measure may not have been sufficient for reaching a

level of ECB pressure to enable an effective evaluation of stalk tunnelling traits at the testcross level.

Testcross predictive abilities for ANT were surprisingly high and comparable in magnitude to those obtained for SDR. Low prediction performance was expected when predicting ANT in material selected for similar flowering time, given the reduced variability for ANT among the genotypes included in the TS. However, at the testcross level genetic variance for ANT remained high in spite of selection, particularly within testcrosses of Pop2, which explains the good *testcross prediction* accuracy. Interestingly, predictive ability for ANT even decreased when the model was trained using additional phenotypic data collected on selected DH lines *per se* in 2012. Including more data from DH lines selected for similar flowering behavior may have added noise to the phenotypes used for model calibration thus resulting in lower GP accuracies for ANT. These results highlight the importance of good quality phenotypic records and broad genetic variability as a prerequisite for GP accuracy. Adding more data not always translates to better accuracies. Hence, data to be used for model training should be selected carefully in order to avoid the inclusion of redundant phenotypic observations which are potentially detrimental to prediction accuracy.

In summary, the obtained results demonstrate that, for improvements in ECB stalk breakage tolerance in testcrosses, it is promising to use GP models trained on DH lines *per se* phenotyped in one season at multiple locations. Despite the promising results obtained for TL and NT with DH lines *per se*, the potential of GP for improving ECB stalk tunnelling resistance at the testcross level needs further investigation.

4.3.3. *Across-location prediction*

As stated before, phenotypic evaluation of resistance to ECB stalk damage is resource intensive in terms of both materials and labor. Artificial infestation of field trials is by far the most cost-intensive procedure among ECB resistance breeding activities. The main goal of the *across-location prediction* scenario was to evaluate the prospects of reducing the number of screening locations and/or the costs of artificial ECB infestation without significant losses in the quality of phenotypic data which are needed for an accurate training of GP models.

Significant genotype-by-location interaction variance was found within years for SDR in both 2011 and 2012 datasets (Appendix Table A 1). Therefore, genome-wide prediction accuracies of SDR across individual locations or across mutually exclusive location subsets were substantially lower compared to scenarios in which the ES and TS contained data from at least

one shared location. These results confirm that multi-environmental trials are required for a sound evaluation of SDR.

So far, ECB resistance studies have never questioned the need for artificial infestation applied to at least a subset of field locations (Hudon and Chiang 1985, Kreps et al. 1998a, Bohn et al. 2000, Flint-Garcia et al. 2003, Papst et al. 2004, Ordas et al. 2010). Here, single-plot repeatabilities for SDR estimated for the two artificially infested German locations were not higher than single-plot repeatabilities estimated from MZH_n and HER, the other two German locations which in 2011 were characterized by low natural ECB damage. Additionally, predictive ability obtained in a prediction scenario *within* MZH_n was comparable to the values obtained for prediction *within* the two artificially infested locations. Hence, at a first glance, artificial infestation does not seem to have significantly increased the quality of the ECB stalk damage phenotypic data. However, *across-location* predictive abilities were particularly low for predictions between MZH_n and HER and the remaining locations under higher ECB pressure (the artificially infested locations and the naturally infested locations FER and TOS). Higher *across-location* predictive abilities could be achieved for the artificially infested trial MZH_i and the geographically and climatically distant location FER than for MZH_i and MZH_n, which were planted at the same field site. The latter results point to considerable differences between the stalk breakage behavior of the same genotypes when placed under low and under high ECB pest pressure. In 2011, ECB stalk damage in MZH_n and HER was almost exclusively restricted to the uppermost part of the stalk (*i.e.* the tassel), which is the usual first entry point for ECB larvae (Liebe 2004). I hypothesize that a given genotype may have fragile tassels but show increased tolerance to breakage in lower stalk portions, or vice versa. Assuming that this condition applies to a relevant proportion of the tested genotypes, it could explain the low *across-location* predictive abilities for field locations contrasting in mean level of ECB stalk damage. When restricted to the tassel, stalk damage will only marginally affect yield or harvest quality and is therefore of reduced practical relevance. It follows that a sufficiently high level of ECB pest pressure is required in order to effectively screen for resistance. In Germany and especially during the 2011 field trial season this could only be achieved by means of artificial ECB infestation.

The next question to address was whether it might be possible to avoid artificial infestation by concentrating on a few locations with high and reliable natural pest pressure. Especially the Italian location FER displayed the highest single-plot repeatability for SDR, the highest *within-location* predictive ability coupled with the by far highest mean level of ECB damage. FER was thus confirmed as a highly suitable field location for ECB resistance screening, due to optimal

climatic conditions for ECB development (in this region the ECB is able to produce two to three generations per season) plus several subsequent years of maize monoculture that favored the establishment of an extremely high and evenly distributed level of natural inoculum. However, it is increasingly difficult to find suitable naturally infested locations and to exploit them in the long term, because of mandatory crop rotation in many regions affected by ECB. It is also important to stress that the material evaluated in this study is not adapted to cultivation under Southern European climatic conditions. Given the genetic complexity of native resistance to ECB stalk damage, solely relying on screening in a few Southern European environments will most probably not lead to long-term resistance improvement under Central European conditions. *Across-location* predictive abilities between the German artificially infested environments and FER were surely not high enough as to suggest screening in Central Europe is superfluous.

Nevertheless, results clearly showed that an optimization of resources is possible by reducing the number of screening locations at least through elimination of field sites characterized by low and/or unreliable natural ECB pressure. Indeed, predictive ability achieved using the four locations under higher ECB pressure was as high as the value obtained from all six environments. Although artificial ECB infestation remains essential for an effective resistance screening in Central Europe, the number of artificially infested locations might be reduced to one without substantial losses in GP accuracy if two to three additional locations under high and reliable natural infestation are available.

4.4. Relationships between traits and putative resistance mechanisms

4.4.1. *Suitability of different traits for ECB resistance screening*

In contrast to the tight genetic correlation of 0.94 between TL and NT, genetic correlations between tunneling traits and SDR in the DH lines *per se* were intermediate, thus confirming that SDR is also affected by factors unrelated to host plant resistance to ECB stalk feeding (Kreps et al. 1998a). Some of the QTL for SDR and stalk tunneling traits were detected at common genetic positions, while others did not colocalize. This suggests that SDR, TL and NT share only in part a common genetic basis. For instance, mechanisms leading to increased strength and stability of stalks may significantly benefit SDR while having no or limited effect on ECB stalk tunneling. Moreover, SDR might also be influenced by environmental factors completely unrelated to ECB damage, although their impact should be low in lodging-tolerant

elite maize materials screened under severe ECB pressure. Additionally, breakage due to ECB can generally be distinguished in the field from breakage due to other factors. The SDR trait should thus be suitable for assessing the tolerance of genotypes to stalk breakage caused by ECB. However, it remains to be seen if training of GP models based on SDR will be effective for a long-term improvement of stalk damage resistance or if a multi-trait approach including SDR and tunneling traits needs to be implemented. Prediction accuracies obtained for TL and NT at the DH line *per se* level were of a similar order as observed for SDR given that both traits had substantially lower heritabilities. The higher heritability of SDR was at least to some extent a consequence of the fact that SDR could be scored in all six field locations in both years. Hence, a genome-based, multi-trait approach should be considered for the simultaneous genetic improvement of the different components contributing to resistance to ECB stalk damage in maize.

To my knowledge, nearly all North American and European ECB resistance studies published so far focused on TL as evaluation criterion for stalk tunneling. Tropical stem borer resistance studies conducted in East Africa by the International Maize and Wheat Improvement Center (CIMMYT) scored the number of larval entry/exit holes as an additional criterion for evaluating stalk tunneling damage (e.g. in Beyene et al. 2011, Tefera et al. 2011). This criterion is very similar to the trait NT used in the present study. If not identical, both traits should at least be highly correlated given that they both provide an assessment of the number of damaged spots along the stalk. ECB larvae exit and re-enter the stalks frequently, particularly when avoiding the harder tissue at nodes. This leaves holes which are visible from the exterior. Thus, scoring the number of entry/exit holes should not require as much labor as an assessment of TL because it does not necessarily imply manual harvest and dissection of maize stalks. One might even consider scoring this trait directly on erect plants in the field by only removing the foliage to visualize larval holes. Heritabilities and predictive abilities of GP obtained for the NT trait were in a similar order of magnitude as the values observed for TL, and NT was the only stalk tunneling trait to show significant genetic variance at the testcross level (for Pop3 testcrosses). Hence, if resource constraints do not allow screening for TL in a resistance breeding program, the results obtained here suggest that scoring the number of entry/exit holes might provide a viable and more economical alternative for evaluating ECB stalk tunneling.

4.4.2. *Putative mechanisms of resistance in the evaluated plant material*

In addition to ECB resistance traits, this study evaluated common agronomic traits potentially related to stalk damage caused by the ECB. Plant height (PH) was measured because taller

plants could represent the first targets for oviposition by ECB moths under conditions of natural infestation. Ear height (EH) was assessed based on the hypothesis that a higher attachment point of the ear shank might increase stalk breakage. Ears being by far the heaviest parts in a maize plant, significant variability in ear height could potentially affect stalk stability, especially in testcrosses which are characterized by overall higher plants and heavier ears. However, correlation analyses with the ECB stalk damage traits did not confirm either hypothesis. Moderate correlations between PH, EH and the stalk damage traits were observed, but pointed in the opposite direction to what was expected. An explanation for this finding could be that both PH and EH were positively correlated with flowering time, and that flowering time was in turn negatively correlated with the ECB stalk damage traits.

Studies on stem borer resistance in rice (Padhi 2004, Sarwar 2012), sorghum (*Sorghum bicolor* Moench) (Kumar et al. 2006) and tropical maize (Munyiri et al. 2013) reported significant associations between the concentration of sugars in the stalk pith and insect resistance. The trait BRIX was measured in this study to investigate whether such correlations might also exist in Central European maize. The hypothesis is that lower sugar concentrations may decrease the quality of the diet available to ECB. This could hamper the development of larvae during a period which in Central Europe is crucial for larval establishment. No significant correlation was found between BRIX and SDR, but moderate positive genetic correlations were observed between BRIX and stalk tunneling, especially with NT. However, BRIX was also negatively correlated with ANT, PH and EH. It is not straightforward to identify potential causal relationships in this complex landscape of correlations between agronomic and ECB stalk damage traits. Although it is possible that BRIX had a moderate effect on ECB larval feeding, the variation in this trait seems also related to the developmental stage of the maize plants during sampling, which was conducted in the flowering period. An increased flow of photosynthetic assimilates is directed towards the ear after flowering. This could explain why early-flowering genotypes tended to display a higher concentration of stalk sugars compared to late-flowering genotypes. Considering that this analysis was performed in 2012 among DH lines selected for similar flowering time, an even stronger relationship between BRIX and ANT might have been expected in the unselected populations. In addition, the negative correlations with PH may even point to a dilution effect of stalk sugars caused by differences in plant size. While a potential influence of stalk sugar concentration on ECB feeding remains unclear, the rather low correlations of BRIX with the stalk damage traits, and the significant associations with plant development and morphology make BRIX unsuitable as an indirect selection criterion for ECB resistance breeding in Central European maize.

From literature it is also known that higher concentrations of cell-wall compounds such as lignin and cellulose can contribute to the fortification of cell-walls and lead to an increase in ECB resistance (Coors 1987, Buendgen et al. 1990, Beeghly et al. 1997, Martin et al. 2004). Cardinal and Lee (2005) and Krakowsky et al. (2007) found associations between QTL for ECB stalk tunneling and QTL for concentration of cell-wall components. The authors hypothesized that these regions may contain genes involved in the biosynthesis of cellulose, hemicellulose and lignin. In the present study, a QTL identified for the TL trait on chromosome 5 showed good congruency with an ECB stalk tunneling QTL which Cardinal and Lee (2005) detected in the same region as QTL for cell-wall component traits, *i.e.* for acid detergent fiber (ADF) and acid detergent lignin (ADL). Following on this observation, a preliminary analysis of the ADF and ADL traits was performed based on stalk samples collected from the four parental lines R1, R2, R3 and S1 at two field sites in 2012. The four lines did not differ significantly for ADF, while the parent R2 displayed a significantly lower stalk ADL concentration compared to the R1 and S1 parental lines (Appendix Figure A 5). Thus, none of the three ECB resistant parental lines showed higher ADF or ADL values as compared to the susceptible parental line S1. This result does not exclude the presence of segregating QTL for ADF and ADL in the progenies. However, as a consequence thereof this topic was not investigated in more depth, also because of the elite, commercial nature of the tested germplasm, which had probably undergone systematic selection for reduced concentrations in cell-wall compounds in order to improve the digestibility of the plant tissues. If this holds true, it should be possible to combine resistance to ECB stalk damage with a good cell-wall digestibility, and thus to develop varieties with increased ECB resistance not only for use as grain or energy maize, but also for use as silage maize.

In summary, several putative mechanisms of resistance were investigated in this study, but none of these could be elucidated as playing a unique role within the tested germplasm. However, results offered insights that the underlying mechanisms should be exploitable in breeding for increased ECB resistance without necessarily affecting important agronomic traits or product quality.

4.4.3. *Relationship between ECB resistance and flowering time*

High negative genetic correlations were observed between ECB stalk damage traits and ANT in the unselected populations. These correlations were reflected at the genetic level by common positions of QTL showing in some cases opposite signs of additive effects. This was most evident for a SDR/ANT colocalizing QTL on chromosome 8 (bin 8.05) with a physical LOD

support interval that included the well-characterized flowering time *vgt1* locus (Salvi et al. 2007, Ducrocq et al. 2008). It is possible that this and other loci affecting ANT have an influence on ECB stalk damage traits. Negative correlations between stalk damage resistance and flowering time have been repeatedly observed since early ECB resistance phenotyping efforts (Russell et al. 1974, Jarvis and Guthrie 1980, Hudon and Chiang 1991) as well as in QTL mapping studies (Bohn et al. 2000, Krakowsky et al. 2004). Gardner et al. (2001) observed that plants at the V10-V12 stage (*i.e.* approaching flowering) provided better conditions for larval establishment compared to less developed plants at the V6-V8 stage. Possible biological reasons include (i) the preference of young larvae toward feeding on pollen accumulating on leaf axils and sheaths (Guthrie et al. 1969, Brindley et al. 1975), (ii) larvae may be better protected from predators by entering sooner into the stalks of early flowering plants, and (iii) less developed plants may still contain significant levels of DIMBOA acting against larval leaf feeding.

In breeding it is important to identify genotypes possessing resistances based on mechanisms independent from the plant developmental stage. In this study, findings obtained from the phenotypic evaluation of the selected fractions in 2012 demonstrated that it was possible to maintain genetic variation for resistance traits while considerably reducing variation in flowering time. Also the results from the QTL and GP analyses performed on data from the selected fractions supported this notion: at the DH line *per se* level, predictive abilities remained intermediate to high for the three ECB stalk damage traits but decreased substantially for ANT. In contrast, at the testcross level predictive abilities for ANT were high, for both the QTL-based and the GBLUP models (Table 9), as a consequence of the unexpectedly high genetic variance observed for ANT in the testcrosses, most notably in those of Pop2. It is difficult to provide a conclusive explanation for this surprising finding, given that testcrosses were produced from lines selected for similar flowering time. It might be that interactions with the genome of the flint tester at important flowering time loci have disclosed genetic variability that remained unexpressed at the inbred line *per se* level. However, at the testcross level, correlations between flowering time and the ECB stalk damage traits were non-significant to low, indicating that differences in flowering time exerted only a marginal influence on ECB stalk damage of testcrosses. This could be explained by the reduced timespan of testcross flowering (approximately eight to nine days), as compared to the considerably longer flowering timespan observed in the unselected DH lines *per se* in 2011 (approximately 17 to 19 days).

The results obtained here indicated that it should be possible to improve resistance to ECB stalk damage independently of flowering time. When evaluating the potential application of genome-

wide prediction in resistance breeding, a genome-based selection index should be considered as an option for the simultaneous genetic improvement of ECB resistance and maturity.

4.5. Conclusions

Marker-assisted selection can be effective for improving resistance to ECB stalk damage in maize. The results presented in this study demonstrate that the efficiency of MAS can be increased considerably when progressing from a QTL-based towards a genome-wide prediction approach. Because of the cost- and time-consuming phenotyping of ECB stalk damage traits, GP can be a meaningful strategy for increasing genetic gain per unit time, by predicting performance of unphenotyped individuals based on their DNA profile. The potential of GP for improving stalk breakage tolerance (SDR) was demonstrated at both the DH line *per se* and testcross levels. Improvements of stalk tunneling resistance (TL and NT) were shown at the DH line *per se* level. Recombination of progeny from the three presently evaluated populations may further increase the overall level of resistance due to different resistance alleles segregating. With the availability of native resistance to ECB stalk damage in elite maize germplasm adapted to Central European conditions, the obtained results may open up avenues for implementing an integrated genome-based selection approach for the simultaneous improvement of yield, maturity and resistance. Training of GP models for ECB stalk damage should be performed based on phenotypic records obtained from one or two artificially infested field locations in the target environment consolidated by additional two to three naturally infested locations situated in regions characterized by high and reliable natural ECB pest pressure.

References

Albert R, Maier G, Dannemann K (2008) Maiszünslerbekämpfung – Bekämpfung und neue Entwicklungen beim *Trichogramma brassicae*-Einsatz. Gesunde Pflanzen 60:41-54

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

Anderson JA, Chao S, Liu S (2008) Molecular breeding using a major QTL for *Fusarium* head blight resistance in wheat. Crop Sci 47:112-119

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Gen 4:132-144

Barcaccia G, Lucchin M, Parrini P (2003) Characterization of a flint maize (*Zea mays* var. *indurata*) Italian landrace, II. Genetic diversity and relatedness assessed by SSR and Inter-SSR molecular markers. Genet Resour Crop Evol 50:253-271

Bardol N, Ventelon M, Mangin B, Jasson S, Loywick V, Couton F, Derue C, Blanchard P, Charcosset A, Moreau L (2013) Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. Theor Appl Genet 126:2717-2736

Barros J, Malvar RA, Butrón A, Santiago R (2011) Combining abilities in maize for the length of the internode basal ring, the entry point of the Mediterranean corn borer larvae. Plant Breed 130:268-270

Barry D, Alfaro D, Darrah LL (1994) Relation of European corn borer (Lepidoptera: Pyralidae) leaf feeding resistance and DIMBOA content in maize. Environ Entomol 23:177–182

Barry D, Darrah LL (1991) Effect of research on commercial hybrid maize resistance to European corn borer (Lepidoptera: Pyralidae). J Econ Entomol 84:1053-1059

Beadle GW (1939) Teosinte and the origin of maize. J Hered 30:245-247

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111-124

Beeghly H, Coors J, Lee M (1997) Plant fiber composition and resistance to European corn borer in four maize populations. Maydica 42:297-303

Bergvinson DJ, Arnason JT, Hamilton RI (1997) Phytochemical changes during recurrent selection for resistance to the European corn borer. Crop Sci 37:1567-1572

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

Bernardo R (2010) Breeding for quantitative traits in plants, 2nd Edition edn. Stemma Press, Woodbury, MN, USA

Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54:68-75

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082-1090

Beyene Y, Mugo S, Gakunga J, Karaya H, Mutinda C, Tefera T, Njoka S, Chepkesis D, Shuma JM, Tende R (2011) Combining ability of maize (*Zea mays* L.) inbred lines resistant to stem borers. Afr J Biotechnol 10:4759-4766

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

Bohn M, Schulz B, Kreps R, Klein D, Melchinger AE (2000) QTL mapping for resistance against the European corn borer (*Ostrinia nubilalis* H.) in early maturing European dent germplasm. Theor Appl Genet 101:907-917

Brindley T, Sparks A, Showers W, Guthrie W (1975) Recent research advances on the European corn borer in North America. Annu Rev Entomol 20:221-239

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210-223

Buendgen MR, Coors JG, Grombacher AW, Russell WA (1990) European corn borer resistance and cell wall composition of three maize populations. Crop Sci 30:505-510

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual. Queensland Department of Primary Industries and Fisheries, Toowoomba, Australia

Butrón A, Chen YC, Rottinghaus GE, McMullen MD (2010) Genetic variation at *bx1* controls DIMBOA content in maize. Theor Appl Genet 120:721-734

Cahill DJ, Schmidt DH (2004) Use of marker assisted selection in a product development breeding program. Proc 4th Int Crop Sci Cong, 26 Sep–1 Oct 2004, Brisbane, Australia

Cambier V, Hance T, de Hoffmann E (2000) Variation of DIMBOA and related compounds content in relation to the age and plant organ in maize. Phytochemistry 53:223-229

Cardinal AJ, Lee M (2005) Genetic relationships between resistance to stalk-tunneling by the European corn borer and cell-wall components in maize population B73×B52. Theor Appl Genet 111:1-7

Cardinal AJ, Lee M, Sharopova N, Woodman-Clikeman WL, Long MJ (2001) Genetic mapping and analysis of quantitative trait loci for resistance to stalk tunneling by the European corn borer in maize. Crop Sci 41:835-845

Charcosset A, Mangin B, Moreau L, Combes L, Jourjon M-F, Gallais A (2001) Heterosis in maize investigated using connected RIL populations. Quantitative genetics and breeding methods: the way ahead. Les colloques 96, INRA Editions, Paris, France

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-71

Coors JG (1987) Resistance to the European corn borer, *Ostrinia nubilalis* (Hubner), in maize, *Zea mays* L., as affected by soil silica, plant silica, structural carbohydrates, and lignin. In: Gabelman WH, Loughman BC (eds) Genetic aspects of plant mineral nutrition. Springer Netherlands, pp 445-456

Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Ceron-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48-60

Crow JF (1998) 90 years ago: the beginning of hybrid maize. Genetics 148:923-928

Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193:347-365

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327-345

De Roos A, Hayes B, Goddard M (2009) Reliability of genomic predictions across multiple populations. Genetics 183:1545-1553

Derridj S, Fiala V, Jolivet E (1986) Increase of European corn borer (*Ostrinia nubilalis*) oviposition induced by a treatment of maize plants with maleic hydrazide: Role of leaf carbohydrate content. Entomol Exp Appl 41:305-310

Derridj S, Gregoire V, Boutin JP, Fiala V (1989) Plant growth stages in the interspecific oviposition preference of the european corn borer and relations with chemicals present on the leaf surfaces. Entomol Exp Appl 53:267-276

Derron JO, Goy G, Breitenmoser S (2009) Caractérisation biologique de la race de la pyrale du maïs (*Ostrinia nubilalis*) à deux générations présente dans le Bassin lémanique. Revue Suisse Agric 41:179-184

Dhurua S, Gujar GT (2011) Field-evolved resistance to *Bt* toxin *Cry1Ac* in the pink bollworm, *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae), from India. Pest Manag Sci 67:898-903

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica 51:281-291

Ducrocq S, Madur D, Veyrieras J-B, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of *vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178:2433-2437

East EM (1916) Studies on size inheritance in *Nicotiana*. Genetics 1:164-176

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S154-S163

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113-125

FAOSTAT (2012) Global production of cereal crops for 2012. Statistics Division of the Food and Agriculture Organization of the United Nations. http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E (accessed on February 3, 2014)

Flint-Garcia SA, Darrah LL, McMullen MD, Hibbard BE (2003) Phenotypic versus marker-assisted selection for stalk strength and second-generation European corn borer resistance in maize. Theor Appl Genet 107:1331-1336

FNR (2013) Entwicklung der Maisanbaufläche in Deutschland. Fachagentur Nachwachsende Rohstoffe e.V. (FNR). http://mediathek.fnr.de/grafiken/daten-und-fakten/bioenergie/biogas/entwicklung-der-maisanbauflache-in-deutschland.html (accessed on February 12, 2014)

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier M-C, McMullen MD, Montalent P, Rose M, Schön C-C, et al (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334

Gardner J, Hoffmann M, Smith M, Wright M (2001) Influence of plant age and genotype on resistance to European corn borer in sweet corn. Maydica 46:111-116

Gassmann AJ, Petzold-Maxwell JL, Keweshan RS, Dunbar MW (2011) Field-evolved resistance to *Bt* maize by Western corn rootworm. PLoS ONE 6:e22629

GMO-Safety (2011) European corn borer: an ingenious pest. *GMO Safety - Genetic engineering - Plants - Environment*. http://www.gmo-safety.eu/basic-info/126.ingenious-pest.html (accessed on January 9, 2014)

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257

Goddard ME, Wray NR, Verbyla K, Visscher PM (2009) Estimating effects and making predictions from genome-wide marker data. Stat Sci 24:517-529

Grieder C, Dhillon BS, Schipprack W, Melchinger AE (2012) Breeding maize as biogas substrate in Central Europe: I. Quantitative-genetic parameters for testcross performance. Theor Appl Genet 124:971-980

Grombacher AW, Russel WA, Guthrie WD (1989) Resistance to first-generation European corn borer (*Lepidoptera: Pyralidae*) and DIMBOA concentration in midwhorl leaves of the BS9 maize synthetic. J Kans Entomol Soc 62:103-107

Guo Z, Tucker DM, Lu J, Kishore V, Gay G (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet 124:261-275

Guthrie WD (1989) Methodologies used for screening and determining resistance in maize to the European corn borer. In: CIMMYT (ed) Toward Insect Resistant Maize for the

Third World: Proceedings of the International Symposium on Methodologies for Developing Host Plant Resistance to Maize Insects, Mexico, D.F.: CIMMYT

Guthrie WD, Huggans JL, Chatterji SM (1969) Influence of corn pollen on the survival and development of second-brood larvae of the European corn borer. Iowa State J Sci 44:185-192

Guthrie WD, Russell WA (1989) Breeding methodologies and genetic basis of resistance in maize to the European corn borer. In: CIMMYT (ed) Toward Insect Resistant Maize for the Third World: Proceedings of the International Symposium on Methodologies for Developing Host Plant Resistance to Maize Insects, Mexico, D.F.: CIMMYT

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

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. J Genet 8:299-309

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames IA, USA

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433-443

Hayes BJ, Visscher PM, Goddard ME (2009b) Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res 91:47-60

Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597-2606

Henderson CR (1984) Applications of linear models in animal breeding. Univ. of Guelph, Ontario, Canada

Holland JB, Nyquist WE, Cervantes-Martínez CT (2003) Estimating and interpreting heritability for plant breeding: An update. Plant Breed Rev 22:9-112

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat:65-70

Hommel B, Schorling M, Langenbruch G-A (2006) Wie den Maiszünsler bekämpfen? Mais 3:117-119

Hudon M, Chiang MS (1985) Resistance and tolerance of maize germplasm to the European corn borer *Ostrinia nubilalis* (Hübner) and its maturity in Quebec. Maydica 30:329-337

Hudon M, Chiang MS (1991) Evaluation of resistance of maize germplasm to the univoltine European corn borer *Ostrinia naubilalis* (Hünber) and relatioship with maize maturity in Quebec. Maydica 36:69-74

Jampatong C, McMullen MD, Barry BD, Darrah LL, Byrne PF, Kross H (2002) Quantitative trait loci for first- and second-generation European corn borer resistance derived from the maize inbred Mo47. Crop Sci 42:584-593

Jannink J-L, Jansen R (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 157:445-454

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166-177

Jansen RC, Jannink J-L, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations. Crop Sci 43:829-834

Jarvis JL, Guthrie WD (1980) Resistance of maize plant introductions to sheath-collar feeding by 2nd-generation European corn borers. Maydica 25:25-32

Jonas E, de Koning D-J (2013) Does genomic selection have a future in plant breeding? Trends Biotechnol 31:497-504

Jonczyk R, Schmidt H, Osterrieder A, Fiesselmann A, Schullehner K, Haslbeck M, Sicker D, Hofmann D, Yalpani N, Simmons C, Frey M, Gierl A (2008) Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of *Bx6* and *Bx7*. Plant Physiol 146:1053-1063

Jourjon M-F, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128-130

Kansy FJ (2010) Maiszünsler-Monitoring im südlichen Oberrheingraben – Ergebnisse einer neuen Bekämpfungsstrategie 2009. Presentation maize conference Emmendingen-Hochburg, February 8th, 2010

Klenke JR, Russell WA, Guthrie WD (1986) Grain yield reduction caused by second generation European corn borer in BS9 corn synthetic. Crop Sci 26:859-863

Konstantopoulou MA, Krokos FD, Mazomenos BE (2002) Chemical stimuli from corn plants affect host selection and oviposition behavior of *Sesamia nonagrioides* (Lepidoptera: Noctuidae). J Econ Entomol 95:1289-1293

Konstantopoulou MA, Krokos FD, Mazomenos BE (2004) Chemical composition of corn leaf essential oils and their role in the oviposition behavior of *Sesamia nonagrioides* females. J Chem Ecol 30:2243-2256

Krakowsky MD, Brinkman MJ, Woodman-Clikeman WL, Lee M (2002) Genetic components of resistance to stalk tunneling by the European corn borer in maize. Crop Sci 42:1309-1315

Krakowsky MD, Lee M, Holland JB (2007) Genotypic correlation and multivariate QTL analyses for cell wall components and resistance to stalk tunneling by the European corn borer in maize. Crop Sci 47:485-488

Krakowsky MD, Lee M, Woodman-Clikeman WL, Long MJ, Sharopova N (2004) QTL mapping of resistance to stalk tunneling by the European corn borer in RILs of maize population B73 x De811. Crop Sci 44:274-282

Kreps RC, Gumber RK, Schulz B, Klein D, Melchinger AE (1998a) Genetic variation in testcrosses of European maize inbreds for resistance to the European corn borer and relations to line per se performance. Plant Breed 117:319-327

Kreps RC, Klein D, Melchinger AE (1998b) Apparatus for dissecting stalks to evaluate stem borer insect resistance in maize. Agron J 90:233-234

Kumar H (1992) Inhibition of ovipositional responses of *Chilo partellus* (Swinhoe) (Lepidoptera: Pyralidae) by the trichomes on the lower leaf surface of a maize cultivar. J Econ Entomol 85:1736-1739

Kumar VK, Sharma HC, Reddy KD (2006) Antibiosis mechanism of resistance to spotted stem borer, *Chilo partellus* in sorghum, *Sorghum bicolor*. Crop Prot 25:66-72

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

Larièpe A, Mangin B, Jasson S, Combes V, Dumas F, Jamin P, Lariagon C, Jolivot D, Madur D, Fiévet J, Gallais A, Dubreuil P, Charcosset A, Moreau L (2012) The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). Genetics 190:795-811

Lassance J-M, Groot AT, Lienard MA, Antony B, Borgwardt C, Andersson F, Hedenstrom E, Heckel DG, Lofstedt C (2010) Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature 466:486-489

Liebe D (2004) Molekulargenetische Untersuchungen zur Abgrenzung von Populationen des Maiszünslers *Ostrinia nubilalis* Hübner als eine Voraussetzung für das Insektenresistenzmanagement (IRM) von B*acillus thuringiensis*-Mais (*Bt*-Mais). Dissertation, Justus-Liebig-Universität Gießen, Deutschland

Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3 3:481-491

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink J-L (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77-123

Lorenz AJ, Smith KP, Jannink J-L (2012) Potential and optimization of genomic selection for *Fusarium* head blight resistance in six-row barley. Crop Sci 52:1609-1621

Lorenzana R, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151-161

LTZ (2014) Maiszünsler Monitoring Baden-Württemberg. Pflanzenschutz-Warndienst. Landwirtschaftliches Technologiezentrum Augustenberg (LTZ). http://www.wetter-bw.de/schaderreger/maiszuensler/index.php (accessed on January 24, 2014)

Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen THE (2009) The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics 183:1119-1126

Lundvall JP, Buxton DR, Hallauer AR, George JR (1994) Forage quality variation among maize inbreds: in vitro digestibility and cell-wall components. Crop Sci 34:1672-1678

Magg T, Melchinger AE, Klein D, Bohn M (2002) Relationship between European corn borer resistance and concentration of mycotoxins produced by *Fusarium spp.* in grains of transgenic *Bt* maize hybrids, their isogenic counterparts, and commercial varieties. Plant Breed 121:146-154

Malvar R, Butrón A, Ordás B, Santiago R (2008) Causes of natural resistance to stem borers in maize. In: Burton EN, Williams PV (eds) Crop protection research advances. Nova Science Publishers, Inc., New York, USA, pp 51-96

Mangin B, Cathelin R, Delannoy D, Escalière B, Lambert S, Marcel J, Ngom B, Jourjon M-F, Rahmani A, Jasson S (2010) MCQTL: a reference manual. Rapport Ubia Toulouse N° 2010/1 - February 2010. Département de Mathématiques et Informatique Appliquées. Institut National de la Recherche Agronomique (INRA), France

Martin SA, Darrah LL, Hibbard BE (2004) Divergent selection for rind penetrometer resistance and its effects on European corn borer damage and stalk traits in corn. Crop Sci 44:711-717

Massman JM, Jung H-JG, Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci 53:58-66

McMullen MD, Frey M, Degenhardt J (2009a) Genetics and biochemistry of insect resistance in maize. In: Bennetzen JL, Hake SC (eds) Handbook of maize: its biology. Springer, New York, USA, pp 271-289

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, et al (2009b) Genetic properties of the maize nested association mapping population. Science 325:737-740

Meissle M, Mouron P, Musa T, Bigler F, Pons X, Vasileiadis VP, Otto S, Antichi D, Kiss J, Pálinkás Z, Dorner Z, Van Der Weide R, Groten J, Czembor E, Adamczyk J, et al (2010) Pests, pesticide use and alternative options in European maize production: current status and future prospects. J Appl Entomol 134:357-375

Melchinger AE, Kreps R, Späth R, Klein D, Schulz B (1998a) Evaluation of early-maturing European maize inbreds for resistance to the European corn borer. Euphytica 99:115-125

Melchinger AE, Utz HF, Schön CC (1998b) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403

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

Miedaner T, Korzun V (2012) Marker-assisted selection for disease resistance in wheat and barley breeding. Phytopathology 102:560-566

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

Mihaljevic R, Schön CC, Utz HF, Melchinger AE (2005) Correlations and QTL correspondence between line per se and testcross performance for agronomic traits in four populations of european maize. Crop Sci 45:114-122

Mihm JA (1983) Efficient mass rearing and infestation techniques to screen for host plant resistance to maize stem borers, *Diatratea spp.* CIMMYT, Mexico

Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:105-131

Mode CJ, Robinson HF (1959) Pleiotropism and the genetic variance and covariance. Biometrics 15:518-537

Mohan S, Ma PWK, Pechan T, Bassford ER, Williams WP, Luthe DS (2006) Degradation of the *S. frugiperda* peritrophic matrix by an inducible maize cysteine protease. J Insect Physiol 52:21-28

Munyiri SW, Mugo SN, Otim M, Mwololo JK, Okori P (2013) Mechanisms and sources of resistance in tropical maize inbred lines to *Chilo partellus* stem borers. J Agric Sci 5:51-60

Niemeyer HM (1988) Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones), defence chemicals in the gramineae. Phytochemistry 27:3349-3358

Nilsson-Ehle NH (1908) Einige Ergebnisse von Kreuzungen bei Hafer und Weizen. Bot Not:257–294

Ordas B, Malvar R, Santiago R, Butron A (2010) QTL mapping for Mediterranean corn borer resistance in European flint germplasm using recombinant inbred lines. BMC Genomics 11:174

Orsini E, Krchov LM, Uphaus J, Melchinger AE (2012) Mapping of QTL for resistance to first and second generation of European corn borer using an integrated SNP and SSR linkage map. Euphytica 183:197-206

Padhi G (2004) Biochemical basis of resistance in rice to yellow stem borer, *Scirpophaga incertulas* Wlk. Madras Agric J 91:253-256

Papst C, Bohn M, Utz HF, Melchinger AE, Klein D, Eder J (2004) QTL mapping for European corn borer resistance (*Ostrinia nubilalis* Hb.), agronomic and forage quality traits of testcross progenies in early-maturing European maize (*Zea mays* L.) germplasm. Theor Appl Genet 108:1545-1554

Papst C, Melchinger AE, Eder J, Schulz B, Klein D, Bohn M (2001) QTL mapping for resistance to European corn borer (*Ostrinia nubilalis* Hb.) in early maturing European dent maize (*Zea mays* L.) germplasm and comparison of genomic regions for resistance across two populations of F3 families. Maydica 46:195-205

Papst C, Utz HF, Melchinger AE, Eder J, Magg T, Klein D, Bohn M (2005) Mycotoxins produced by *Fusarium spp.* in isogenic *Bt* vs. non-*Bt* maize hybrids under European corn borer pressure. Agron J 97:219-224

Peiffer JA, Flint-Garcia SA, De Leon N, McMullen MD, Kaeppler SM, Buckler ES (2013) The genetic architecture of maize stalk strength. PLoS ONE 8:e67066

Pornkulwat S, Skoda SR, Thomas GD, Foster JE (1998) Random amplified polymorphic DNA used to identify genetic variation in ecotypes of the European corn borer (Lepidoptera: Pyralidae). Ann Entomol Soc Am 91:719-725

R Development Core Team (2013) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rasmann S, Kollner TG, Degenhardt J, Hiltpold I, Toepfer S, Kuhlmann U, Gershenzon J, Turlings TCJ (2005) Recruitment of entomopathogenic nematodes by insect-damaged maize roots. Nature 434:732-737

Rebai A, Goffinet B (2000) More about quantitative trait locus mapping with diallel designs. Genet Res 75:243-247

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895-903

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217-220

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

Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodríguez VM, Moreno-Gonzalez J, Melchinger A, Bauer E, Schoen C-C, Meyer N, Giauffret C, Bauland C, Jamin P, et al (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics 192:715-728

Röber FK, Gordillo GA, Geiger HH (2005) In vivo haploid induction in maize - performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50:275-283

Rogers JS (1972) Measures of genetic similarity and genetic distances. Studies in Genetics, Univ Texas Publ 7213:145-153

Russell WA, Guthrie WD, Grindeland RL (1974) Breeding for resistance in maize to first and second broods of the European corn borer. Crop Sci 14:725-727

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

Saeglitz C (2004) Untersuchungen der genetischen Diversität von Maiszünsler-Populationen (*Ostrinia nubilalis*, Hbn.) und ihrer Suszeptibilität gegenüber dem *Bacillus thuringiensis* (*Bt*)-Toxin als Grundlage für ein Resistenzmanagement in *Bt*-Maiskulturen. Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, Deutschland

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski J-A, et al

(2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. PNAS 104:11376-11381

Santiago R, Souto XC, Sotelo J, Butrón A, Malvar RA (2003) Relationship between maize stem structural characteristics and resistance to pink stem borer (Lepidoptera: Noctuidae) attack. J Econ Entomol 96:1563-1570

Sarwar M (2012) Management of rice stem borers (Lepidoptera: Pyralidae) through host plant resistance in early, medium and late plantings of rice (*Oryza sativa* L.). J Cereals Oilseeds 3:10-14

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115

Schön CC, Lee M, Melchinger AE, Guthrie WD, Woodman WL (1993) Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. Heredity 70:648-659

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485-498

Seifi A, Visser RGF, Bai Y (2013) How to effectively deploy plant resistances to pests and pathogens in crop breeding. Euphytica 190:321-334

Siegfried BD, Hellmich RL (2012) Understanding successful resistance management: The European corn borer and *Bt* corn in the United States. GM Crops Food 3:184-193

Smith HE (1920) Broom corn, the probable host in which *Pyrausa nubilalis* reached America. J Econ Entom 13:425-430

Soller M, Brody T, Genizi A (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47:35-39

St.Clair DA (2010) Quantitative disease resistance and quantitative resistance loci in breeding. Annu Rev Phytopathol 48:247-268

Steinhoff J, Liu W, Maurer HP, Würschum T, Friedrich C, Longin H, Ranc N, Reif JC (2011) Multiple-line cross quantitative trait locus mapping in European elite maize. Crop Sci 51:2505-2516

Tabashnik BE, Carrière Y (2010) Field-evolved resistance to *Bt* cotton bollworm in the US and pink bollworm in India. Southwest Entomol 35:417-424

Technow F, Bürger A, Melchinger AE (2013) Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3 3:197-203

Tefera T, Mugo S, Tende R, Likhayo P (2011) Methods of screening maize for resistance to stem borers and post-harvest insect pests. CIMMYT, Nairobi, Kenya

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc B 58:267-288

Transgen.de (2013) USA 2013: Keine Trendwende - Farmer bleiben bei Gentechnik-Sorten. *Transparenz Gentechnik*. http://www.transgen.de/anbau/flaechen_international/189.doku.html (accessed on February 3, 2014)

Trnka M, Muška F, Semerádová D, Dubrovský M, Kocmánková E, Žalud Z (2007) European corn borer life stage model: regional estimates of pest development and spatial distribution under present and future climate. Ecol Model 207:61-84

Turlings TCJ, Tumlinson JH, Lewis WJ (1990) Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. Science 250:1251-1253

Udayagiri S, Mason CE (1995) Host plant constituents as oviposition stimulants for a generalist herbivore: European corn borer. Entomol Exp Appl 76:59-65

USDA (2013) Corn: background. Economic Research Service. United States Department of Agriculture (USDA). http://www.ers.usda.gov/topics/crops/corn/background.aspx#.Uu_VvLSin3E (accessed on February 3, 2014)

Utz HF (2011) PLABSTAT v3A. A computer program for statistical analysis of plant breeding experiments. Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany

Van Ooijen JW (2006) JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands

van Rensburg JBJ (2007) First report of field resistance by the stem borer, *Busseola fusca* (Fuller) to *Bt*-transgenic maize. S Afr J Plant Soil 24:147-151

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414-4423

Varshney AK, Babu BR, Singh AK, Agarwal HC, Jain SC (2003) Ovipositional responses of *Chilo partellus* (Swinhoe) (Lepidoptera: Pyralidae) to natural products from leaves of two maize (*Zea mays* L.) cultivars. J Agric Food Chem 51:4008-4012

Wan P, Huang Y, Wu H, Huang M, Cong S, Tabashnik BE, Wu K (2012) Increased frequency of pink bollworm resistance to *Bt* toxin *Cry1Ac* in China. PLoS ONE 7:e29975

Willcox MC, Khairallah MM, Bergvinson D, Crossa J, Deutsch JA, Edmeades GO, González-de-León D, Jiang C, Jewell DC, Mihm JA, Williams WP, Hoisington D (2002) Selection for resistance to southwestern corn borer using marker-assisted and conventional backcrossing. Crop Sci 42:1516-1528

Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics 28:2086-2087

Wimmer V, Lehermeier C, Albrecht T, Auinger H-J, Wang Y, Schön C-C (2013) Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics 195:573-587

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 2:1427-1436

Xu S (1998) Mapping quantitative trait loci using multiple families of line crosses. Genetics 148:517-524

Zellner M, Lechermann T (2014) Maiszünsler Prognose Flugbeginn und Befallserhebung. Bayerische Landesanstalt für Landwirtschaft (LfL), Institut für Pflanzenschutz. http://www.lfl-design3.bayern.de/ips/pflanzenschutzhinweise/25944/index.php (accessed on January 24, 2014)

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-68

Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, Li J, Simianer H (2014) Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PLoS ONE 9:e93017

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Piepho HP, Reif JC (2013) Choice of shrinkage parameter and prediction of genomic breeding values in elite maize breeding populations. Plant Breed 132:99-106

Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity 112:638-645

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J Roy Stat Soc B 67:301-320

Appendix

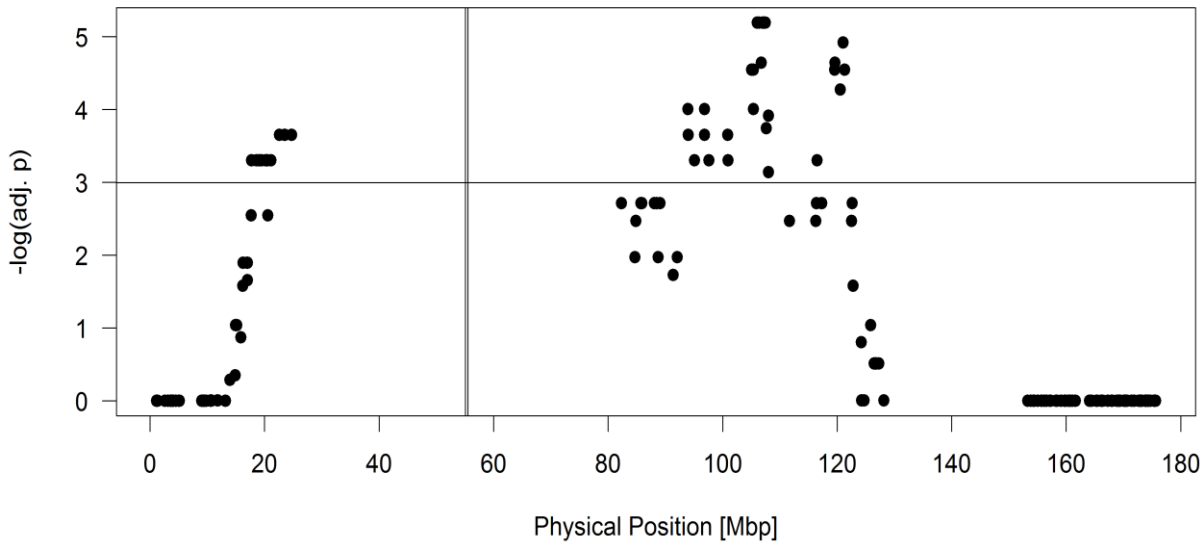

Figure A 1 Extent of segregation distortion of the SNP markers on chromosome 7 observed in the DH lines of Pop2. For each marker, the negative logarithmus of the Holm-Bonferroni adjusted p-value obtained from the Chi-Square test is plotted against the physical position on the chromosome. The horizontal line represents the 5% significance level at –log(0.05). Vertical lines denote the estimated position of the centromere according to physical map information.

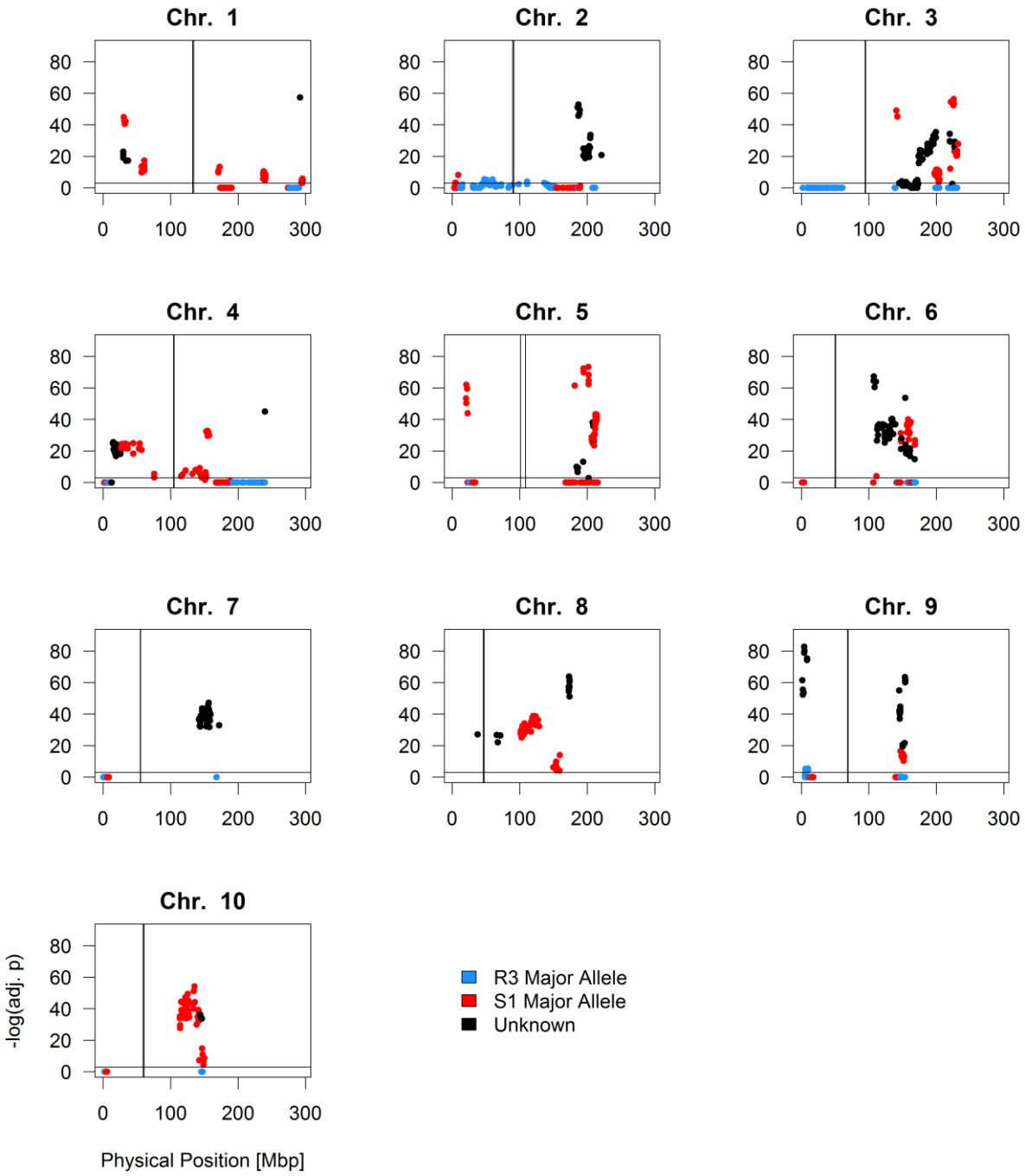

Figure A 2 Extent of segregation distortion of the 1,340 polymorphic SNP markers in the DH lines of Pop3 across ten chromosomes (Chr. 1 – Chr. 10). For each marker, the negative logarithmus of the Holm-Bonferroni adjusted p-value obtained from the Chi-Square test is plotted against the physical position on the respective chromosome. The horizontal line represents the 5% significance level at –log(0.05). Vertical lines denote the estimated position of the centromere according to physical map information.

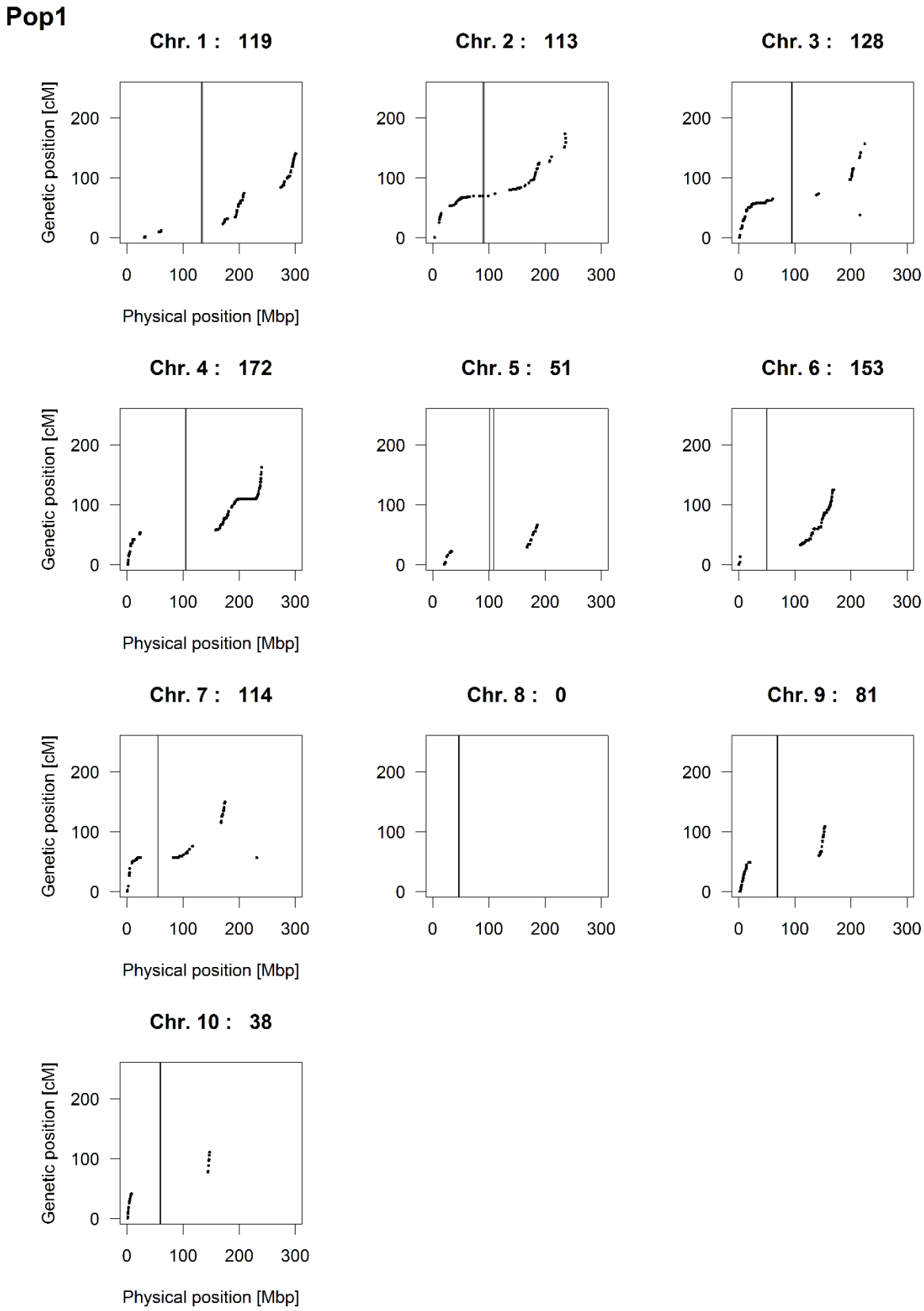

Figure A 3 Mapped genetic position [cM] versus corresponding physical position [Mbp] of SNP markers on the genetic maps of Pop1, Pop2 and Pop3 for all ten chromosomes (Chr. 1 – Chr. 10). Vertical lines within each chromosome denote the estimated position of the centromere according to physical map information. Numbers above each graph refer to the number of SNP markers contained in the genetic map of the respective chromosome.

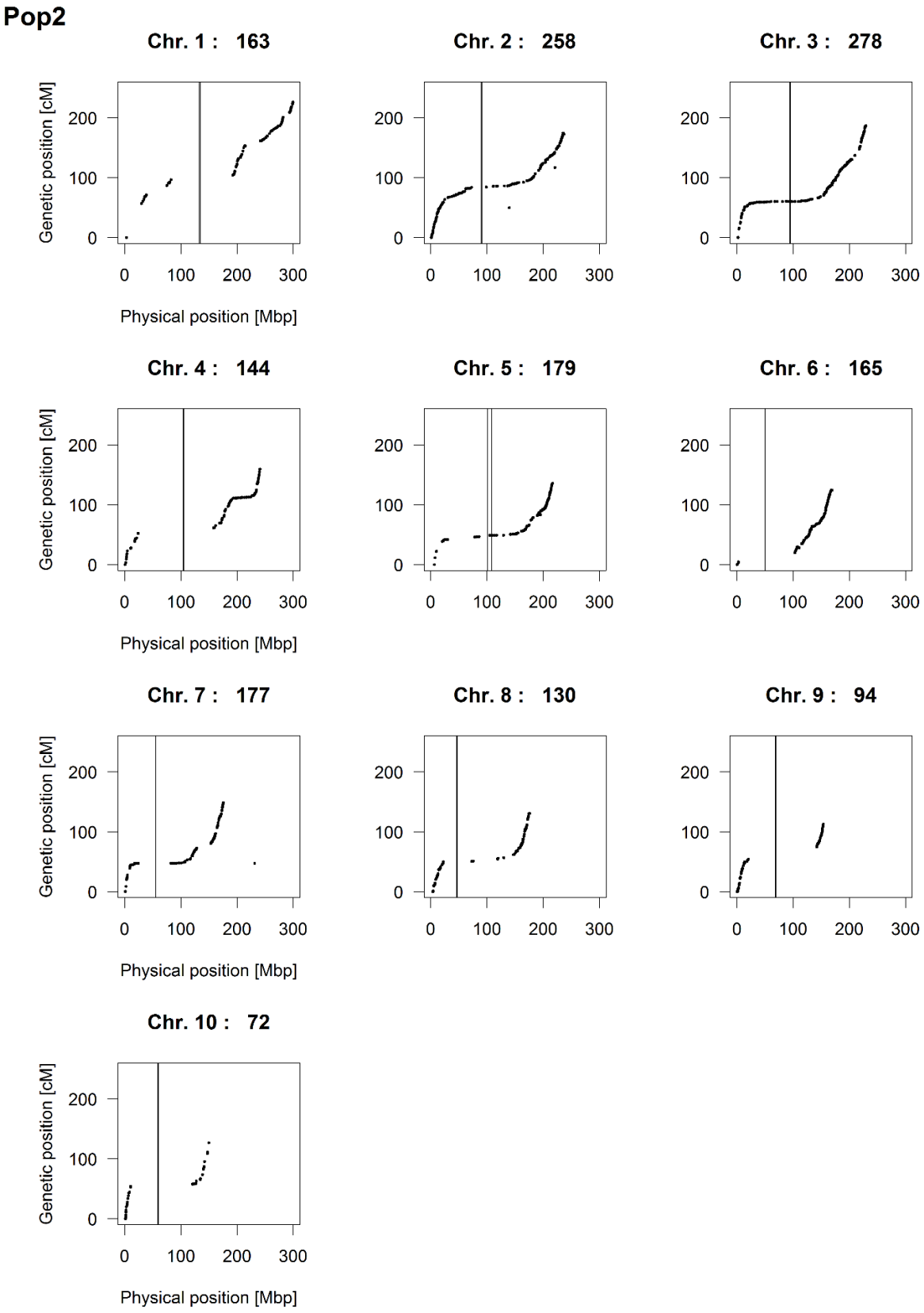

Figure A 3 (continued)

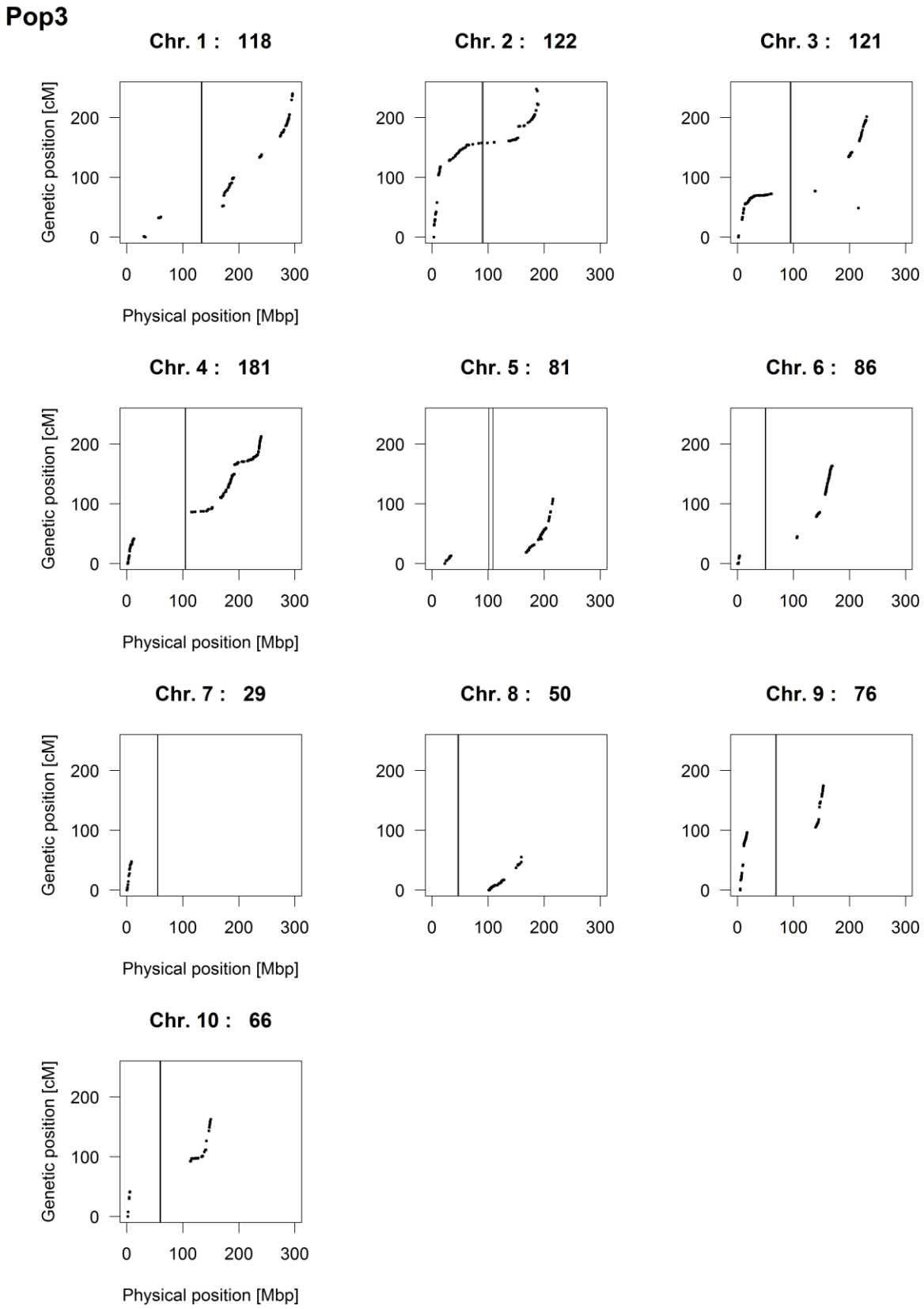

Figure A 3 (continued)

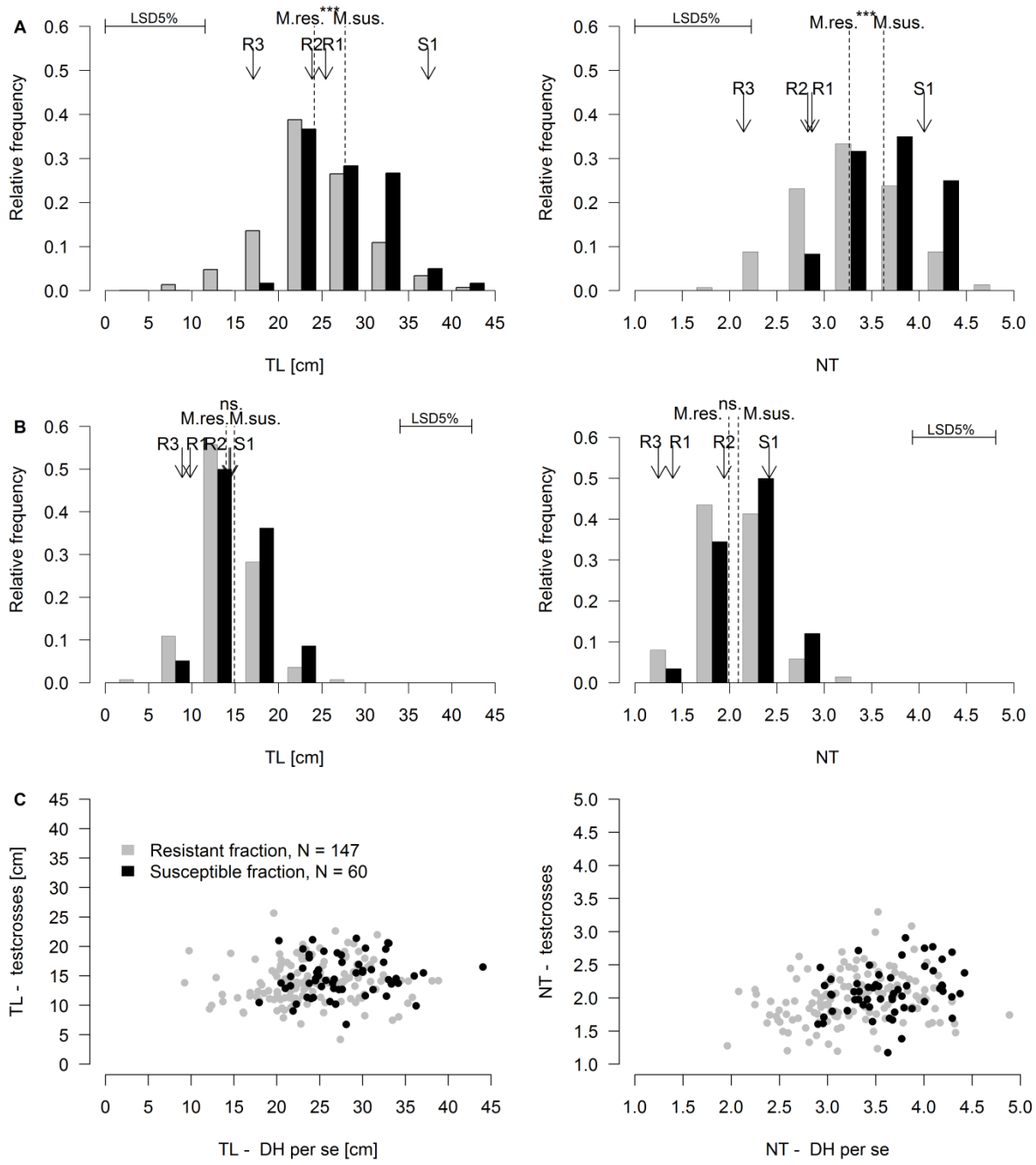

Figure A 4 Distributions of the adjusted means of TL (left) and NT (right) across two locations for the selected DH lines *per se* (A) and their corresponding testcrosses (B) evaluated in 2012, and scatterplot of the phenotypic correlation between DH lines *per se* and testcrosses for TL and NT (C). Genotypes from the resistant (N = 147) and susceptible (N = 60) fractions are represented in all graphs by grey and black color, respectively. Arrows show the means of the parental lines R1, R2, R3 and S1. The least significant difference (LSD) at the 5% significance level and the means of the resistant (M.res.) and susceptible (M.sus.) fractions are also shown in the graphs. Significant differences between M.res. and M.sus. were determined with a t-test (***, ns, difference in fraction means significant at $p<0.001$ and non-significant with $p>0.05$, respectively).

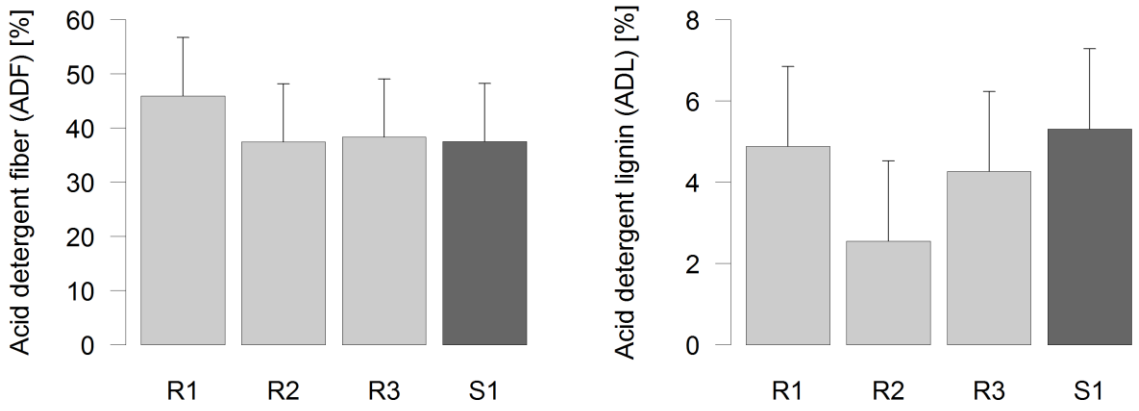

Figure A 5 Genotype means of the four parental lines R1, R2, R3 and S1 for the two cell-wall component traits "acid detergent fiber" (ADF, left graph) and "acid detergent lignin" (ADL, right graph). Stalk samples were collected from three plants per field plot at the two field sites FRS and MZH_n during July 2012. ADF and ADL were measured for each individual field plot based on a mixed stalk sample of the three plants. The data were analysed by means of an ANOVA across the two locations. The error bars indicate the least significant difference at the 5% probability level.

Table A 1 Estimates of the mean, genetic variance ($\hat{\sigma}_g^2$), genotype-by-location interaction variance ($\hat{\sigma}_{gl}^2$), residual error variance ($\hat{\sigma}_e^2$), and heritability ($\hat{h}^2$) (± standard errors) for the traits SDR, TL, NT, ANT, SIL, PH, EH and BRIX for the three DH populations (Pop1, Pop2, Pop3). Estimates are given for the unselected populations evaluated as DH lines *per se* in 2011 as well as for the selected DH lines *per se* and their testcrosses evaluated in 2012.

Population	Entries (no.)	SDR [1-9 scale] 6 locations					TL [cm] 2 locations				
		Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$
2011 unselected populations – DH lines *per se*											
Pop1	85	2.89±0.06	0.10±0.03	0.22±0.04	0.57±0.04	0.55±0.08	13.83±0.46	6.84±3.89^{ns}	6.12±4.94^{ns}	28.55±4.20	-
Pop2	243	2.94±0.03	0.28±0.03	0.21±0.03	0.67±0.03	0.75±0.03	11.65±0.27	9.79±1.90	1.42±2.24^{ns}	22.80±2.14	0.51±0.07
Pop3	262	2.86±0.03	0.17±0.02	0.17±0.02	0.62±0.02	0.68±0.03	12.29±0.26	8.46±1.60	0.00±0.00^{ns}	25.38±1.63	0.47±0.05
All	590	2.90±0.02	0.21±0.02	0.20±0.02	0.63±0.02	0.71±0.02	12.25±0.17	9.41±1.23	0.78±1.48^{ns}	25.14±1.44	0.49±0.04
2012 selected populations – DH lines *per se*											
Pop1	31	4.13±0.10	0.20±0.07	0.11±0.06	0.77±0.07	0.72±0.08	27.05±0.98	1.97±6.90^{ns}	0.00±0.00^{ns}	82.79±12.79	-
Pop2	85	3.94±0.06	0.28±0.06	0.21±0.04	0.72±0.04	0.76±0.04	27.09±0.59	16.83±7.32	12.58±8.20^{ns}	65.51±7.57	0.43±0.13
Pop3	91	3.66±0.06	0.26±0.05	0.21±0.04	0.67±0.04	0.75±0.04	22.73±0.57	16.89±5.26	7.40±4.78^{ns}	38.22±4.46	0.56±0.10
All	207	3.84±0.04	0.27±0.04	0.19±0.02	0.70±0.03	0.76±0.03	25.17±0.38	17.09±3.78	5.27±3.94^{ns}	57.69±4.18	0.50±0.07
2012 selected populations – Testcrosses											
Pop1	31	5.30±0.08	0.09±0.05	0.16±0.07	0.76±0.08	0.50±0.15	13.28±0.63	4.15±3.10^{ns}	0.00±0.00^{ns}	25.64±4.26	-
Pop2	85	5.15±0.05	0.16±0.04	0.04±0.05^{ns}	1.06±0.06	0.64±0.06	14.28±0.38	1.28±2.64^{ns}	0.15±4.31^{ns}	41.30±4.94	-
Pop3	91	5.02±0.05	0.15±0.04	0.06±0.04^{ns}	0.95±0.06	0.64±0.06	14.53±0.37	2.66±2.08^{ns}	0.00±0.00^{ns}	35.21±3.41	-
All	207	5.12±0.03	0.16±0.02	0.07±0.03	0.95±0.04	0.65±0.04	14.23±0.24	3.16±1.33	0.00±0.00^{ns}	34.97±2.16	0.27±0.09

^{ns}, genetic ($\hat{\sigma}_g^2$) or genotype-by-location ($\hat{\sigma}_{gl}^2$) variance components not significant at the 0.05 probability level

Table A 1 (continued)

Population	Entries (no.)	NT [counts] 2 locations					ANT [days] 3 locations				
		Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$
2011 unselected populations – DH lines *per se*											
Pop1	85	2.20±0.06	0.10±0.06	0.10±0.08^{ns}	0.48±0.07	0.30±0.16	97.00±0.28	5.04±0.92	1.04±0.38	2.49±0.31	0.85±0.03
Pop2	243	1.93±0.04	0.16±0.04	0.06±0.04^{ns}	0.41±0.04	0.46±0.07	98.30±0.16	6.09±0.66	1.23±0.23	2.34±0.17	0.87±0.02
Pop3	262	2.20±0.04	0.15±0.04	0.04±0.05^{ns}	0.53±0.05	0.41±0.08	96.85±0.16	6.30±0.65	0.94±0.24	3.04±0.21	0.86±0.02
All	590	2.09±0.02	0.15±0.02	0.06±0.03	0.47±0.03	0.43±0.05	97.46±0.11	6.08±0.41	1.07±0.15	2.65±0.12	0.87±0.01
2012 selected populations – DH lines *per se*											
Pop1	31	3.52±0.10	0.00±0.00^{ns}	0.28±0.11	0.53±0.10	-	90.77±0.26	1.24±0.45	0.80±0.30	0.89±0.20	0.73±0.09
Pop2	85	3.28±0.06	0.10±0.06^{ns}	0.11±0.07^{ns}	0.61±0.07	-	91.07±0.16	1.36±0.33	0.95±0.28	1.78±0.22	0.66±0.07
Pop3	91	3.40±0.06	0.13±0.05	0.09±0.06^{ns}	0.49±0.06	0.44±0.12	91.07±0.15	1.43±0.32	0.81±0.23	1.52±0.18	0.70±0.06
All	207	3.37±0.04	0.11±0.04	0.12±0.04	0.54±0.04	0.36±0.09	91.02±0.10	1.44±0.21	0.93±0.16	1.55±0.12	0.69±0.04
2012 selected populations – Testcrosses											
Pop1	31	1.96±0.06	0.00±0.00^{ns}	0.00±0.00^{ns}	0.36±0.05	-	81.17±0.23	0.55±0.26	0.20±0.23^{ns}	1.05±0.22	0.60±0.16
Pop2	85	1.97±0.04	0.00±0.00^{ns}	0.00±0.00^{ns}	0.45±0.04	-	82.57±0.14	3.85±0.71	0.00±0.00^{ns}	1.72±0.17	0.90±0.02
Pop3	91	2.10±0.04	0.07±0.03	0.00±0.00^{ns}	0.36±0.04	0.45±0.10	80.66±0.14	1.10±0.27	0.15±0.18^{ns}	1.53±0.19	0.71±0.07
All	207	2.02±0.03	0.04±0.02	0.00±0.00^{ns}	0.40±0.02	0.29±0.08	81.53±0.09	2.07±0.26	0.05±0.11^{ns}	1.58±0.12	0.83±0.02

Table A 1 (continued)

Population	Entries (no.)	SIL [days] 3 locations					PH [cm] 1-2 locations				
		Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$
2011 unselected populations – DH lines *per se*											
Pop1	85	97.42±0.31	7.28±1.26	0.62±0.37^{ns}	2.99±0.36	0.89±0.02	196.03±1.41	164.30±30.49	-	54.41±9.27	-
Pop2	243	97.47±0.19	6.69±0.70	1.05±0.24	2.77±0.20	0.87±0.02	197.62±0.84	134.19±16.30	-	57.18±5.73	-
Pop3	262	97.20±0.18	8.02±0.80	0.84±0.25	3.40±0.23	0.89±0.01	178.91±0.80	334.07±35.25	-	68.71±6.49	-
All	590	97.34±0.12	7.43±0.49	0.90±0.15	3.03±0.13	0.88±0.01	189.07±0.54	217.92±14.76	-	61.40±3.77	-
2012 selected populations – DH lines *per se*											
Pop1	31	91.05±0.32	2.64±0.85	0.95±0.43	1.45±0.31	0.81±0.07	214.53±2.23	179.63±52.01	11.16±11.61^{ns}	50.68±10.47	0.91±0.04
Pop2	85	90.60±0.19	2.26±0.51	1.28±0.35	2.11±0.26	0.72±0.06	211.46±1.35	116.85±24.15	33.22±10.64	55.92±6.73	0.79±0.05
Pop3	91	91.60±0.19	2.62±0.50	0.91±0.25	1.46±0.18	0.81±0.04	198.95±1.30	227.86±38.94	7.43±6.20^{ns}	49.44±6.01	0.93±0.02
All	207	91.11±0.12	2.59±0.33	1.06±0.18	1.74±0.13	0.78±0.03	206.42±0.86	174.81±19.53	19.23±5.11	51.96±3.92	0.89±0.02
2012 selected populations – Testcrosses											
Pop1	31	79.50±0.25	0.79±0.31	0.00±0.00^{ns}	1.43±0.24	0.69±0.08	284.90±1.38	19.90±11.42^{ns}	0.00±0.00^{ns}	83.13±13.29	-
Pop2	85	80.09±0.15	2.25±0.47	0.44±0.22^{ns}	1.40±0.17	0.80±0.05	285.32±0.83	33.26±10.73	24.30±9.74	58.15±7.28	0.55±0.11
Pop3	91	79.31±0.15	1.23±0.26	0.00±0.00^{ns}	1.55±0.15	0.76±0.04	283.33±0.81	40.95±10.45	9.37±8.03^{ns}	62.71±7.74	0.67±0.08
All	207	79.66±0.10	1.62±0.22	0.17±0.11^{ns}	1.45±0.11	0.78±0.03	284.40±0.53	32.93±6.38	15.72±5.61	64.52±4.95	0.58±0.06

Table A 1 (continued)

Population	Entries (no.)	EH [cm] 2 locations					BRIX [°Brix] 2 locations				
		Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	Mean	$\hat{\sigma}_g^2$	$\hat{\sigma}_{gl}^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$
2011 unselected populations – DH lines *per se*											
Pop1	85										
Pop2	243										
Pop3	262										
All	590										
2012 selected populations – DH lines *per se*											
Pop1	31	107.38±1.79	96.52±28.80	13.45±8.57^{ns}	31.92±6.34	0.87±0.05	10.33±0.12	0.19±0.12^{ns}	0.12±0.13^{ns}	0.53±0.11	-
Pop2	85	107.19±1.08	83.38±15.59	13.20±5.67	36.17±4.36	0.84±0.04	9.71±0.07	0.51±0.13	0.11±0.08^{ns}	0.55±0.07	0.72±0.07
Pop3	91	106.52±1.04	73.37±13.79	14.30±5.87	38.27±4.45	0.81±0.04	10.58±0.07	0.32±0.10	0.17±0.09	0.61±0.08	0.58±0.10
All	207	106.93±0.69	85.75±10.10	13.64±3.51	36.28±2.69	0.84±0.02	10.19±0.05	0.38±0.06	0.15±0.05	0.56±0.04	0.64±0.05
2012 selected populations – Testcrosses											
Pop1	31	150.84±1.08	25.92±12.08	3.87±10.48^{ns}	59.03±11.19	0.61±0.15					
Pop2	85	155.12±0.65	28.70±7.38	4.20±5.72^{ns}	49.86±5.92	0.66±0.08					
Pop3	91	154.57±0.63	18.10±5.00	6.28±4.15^{ns}	31.80±3.82	0.62±0.09					
All	207	154.21±0.42	24.93±4.06	3.96±3.12^{ns}	43.38±3.28	0.66±0.05					

Table A 2 Chromosome (Chr.), position (Pos.), LOD support interval (LOD S.I.), LOD score at the QTL position, proportion of variance explained (R^2) and additive effects of QTL alleles derived from the respective parental lines R1, R2, R3 and S1 detected in the QTL analyses within the individual biparental populations Pop1, Pop2 and Pop3 evaluated as DH lines *per se* in 2011 for the traits SDR, TL, NT and ANT. For each trait, the total proportion of phenotypic variance explained by all detected QTL is given above the R^2 values of the individual QTL.

Pop1 (*N* = 81)

Chr.	Pos. [cM]	LOD S.I. [cM]	LOD score	R^2	Additive effects	
					R1	S1
SDR [1-9 score]				**0.39**		
3	88	(54 – 106)	5.79	0.26	-0.10	0.10
6	82	(61 – 90)	5.81	0.26	-0.10	0.10
TL [cm]						
NT [count]						
ANT [days]				**0.64**		
1	206	(204 – 222)	5.74	0.26	-0.41	0.41
3	99	(89 – 106)	17.38	0.51	0.82	-0.82
4	79	(57 – 83)	3.09	0.16	-0.30	0.30
5	111	(94 – 163)	4.23	0.20	0.41	-0.41

Pop2 (*N* = 214)

Chr.	Pos. [cM]	LOD S.I. [cM]	LOD score	R^2	Additive effects	
					R2	S1
SDR [1-9 score]				**0.44**		
1	83	(73 – 94)	4.22	0.09	-0.07	0.07
2	77	(69 – 80)	7.85	0.15	-0.10	0.10
3	61	(61 – 69)	10.39	0.19	-0.11	0.11
6	132	(124 – 144)	3.96	0.08	-0.07	0.07
8	62	(59 – 65)	15.24	0.25	0.13	-0.13
TL [cm]				**0.20**		
3	138	(133 – 147)	3.75	0.08	-0.57	0.57
5	105	(65 – 107)	4.48	0.09	-0.61	0.61
9	112	(102 – 129)	4.00	0.08	-0.58	0.58
NT [count]				**0.15**		
3	71	(64 – 83)	3.66	0.07	-0.07	0.07
5	73	(62 – 82)	4.24	0.08	-0.08	0.08
ANT [days]				**0.44**		
1	209	(196 – 213)	4.34	0.09	-0.30	0.30
2	229	(221 – 229)	5.17	0.10	0.33	-0.33
5	35	(13 – 46)	3.37	0.07	0.30	-0.30
5	106	(100 – 131)	6.20	0.12	0.36	-0.36
7	30	(6 – 37)	4.33	0.09	-0.30	0.30
7	126	(119 – 153)	3.10	0.06	0.26	-0.26
8	60	(57 – 64)	9.06	0.17	-0.44	0.44

Table A 2 (continued)

Pop3 (*N* = 226)

Chr.	Pos. [cM]	LOD S.I. [cM]	LOD score	R^2	Additive effects	
					R3	S1
SDR [1-9 score]				**0.57**		
2	118	(114 – 121)	13.11	0.22	-0.10	0.10
3	99	(88 – 105)	17.90	0.28	-0.14	0.14
4	62	(55 – 77)	3.97	0.08	0.06	-0.06
4	114	(106 – 121)	3.46	0.07	-0.05	0.05
5	119	(103 – 128)	9.11	0.16	-0.08	0.08
6	87	(80 – 98)	5.08	0.10	-0.06	0.06
8	51	(0 – 54)	3.61	0.07	0.06	-0.06
10	105	(100 – 111)	5.90	0.11	-0.07	0.07
TL [cm]				**0.33**		
3	70	(68 – 72)	6.19	0.11	-0.62	0.62
4	178	(154 – 187)	2.77	0.05	-0.42	0.42
5	119	(109 – 124)	14.34	0.23	-1.03	1.03
10	128	(120 – 136)	3.39	0.07	-0.49	0.49
NT [count]				**0.35**		
3	70	(68 – 74)	7.78	0.14	-0.10	0.10
4	150	(119 – 164)	3.97	0.08	-0.07	0.07
5	119	(115 – 125)	13.37	0.22	-0.14	0.14
ANT [days]				**0.31**		
3	62	(62 – 72)	11.67	0.19	0.54	-0.54
5	122	(117 – 129)	8.70	0.15	0.49	-0.49

Table A 3 Mean predictive abilities (± standard deviation) from *across-population prediction* with the three biparental populations Pop1, Pop2 and Pop3 for the traits SDR, TL, NT and ANT evaluated in unselected DH lines *per se* in 2011. Numbers in brackets represent the number of genotypes included in the respective estimation and test sets for prediction across original, complete populations. Predictive abilities thereof are shown for each trait in the respective left-side columns (complete). For each trait, predictive abilities shown in the right-side columns (reduced) were obtained by reducing the sizes of Pop2 and Pop3 to the size of Pop1 (N = 81) in both estimation and test set.

Estimation set (ES)	Test set (TS)	SDR complete	SDR reduced	TL complete	TL reduced	NT complete	NT reduced	ANT complete	ANT reduced
Pop1 (81)	Pop2 (214)	0.36 ± 0.06	0.38 ± 0.10	0.16 ± 0.07	0.17 ± 0.11	0.26 ± 0.07	0.25 ± 0.11	0.23 ± 0.06	0.23 ± 0.10
Pop1 (81)	Pop3 (226)	0.57 ± 0.04	0.56 ± 0.07	0.39 ± 0.05	0.38 ± 0.08	0.39 ± 0.05	0.38 ± 0.08	0.53 ± 0.04	0.52 ± 0.07
Pop2 (214)	Pop1 (81)	0.50 ± 0.07	0.51 ± 0.08	0.11 ± 0.09	0.10 ± 0.10	0.30 ± 0.09	0.21 ± 0.10	0.40 ± 0.09	0.35 ± 0.09
Pop2 (214)	Pop3 (226)	0.53 ± 0.04	0.51 ± 0.08	0.49 ± 0.05	0.46 ± 0.08	0.55 ± 0.04	0.47 ± 0.08	0.56 ± 0.05	0.51 ± 0.09
Pop3 (226)	Pop1 (81)	0.59 ± 0.08	0.55 ± 0.08	0.36 ± 0.09	0.31 ± 0.10	0.34 ± 0.09	0.29 ± 0.09	0.58 ± 0.06	0.54 ± 0.06
Pop3 (226)	Pop2 (214)	0.42 ± 0.06	0.39 ± 0.10	0.39 ± 0.06	0.37 ± 0.10	0.41 ± 0.06	0.39 ± 0.09	0.31 ± 0.07	0.30 ± 0.10

Documentation A 1 cvMCQTL R package documentation

Package ‘cvMCQTL’

February 4, 2015

Type Package

Title Cross-validation for MCQTL software

Version 1.09

Date 2015-02-04

Author Valentin Wimmer and Flavio Foiada

Maintainer Valentin Wimmer <Valentin.Wimmer@tum.de>

Description This package provides a framework to cross-validate the results of MCQTL analyses and exploit QTL effects for prediction. This includes the identification of QTL using pre-defined estimation sets, estimation of genetic values based on these QTL effects, and estimation of predictive ability in test sets. This research was funded by the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr Synbreed - Synergistic plant and animal breeding (FKZ 0315528A). We thank the MCQTL software development team, in particular Ms. Brigitte Mangin and Mr. Sylvain Jasson who provided us with insightful technical advice on how to implement parts of our cross-validation procedure.

Depends R (>= 2.14), synbreed

License GPL-2

R topics documented:

cv.MCQTL *Main function to perform cross-validation with MCQTL software*

Description

Identification of QTL using pre-defined estimation sets, estimation of genetic values based on these QTL effects, and estimation of predictive ability in test sets.

Usage

```
cv.MCQTL(pop.list, pop.names, parent.lines, TS.pop, pop.type, pre = NULL, trait = 1,
    progression.step = 2, dir = getwd())
```

Arguments

pop.list	list of gpData objects for each population
pop.names	character vector with names for each population
parent.lines	list with two characters in each element specifying the parent lines for each population. The first is the parent that transmitted the A allele whereas the second is the parent that transmitted the B allele in the respective population. For the special case of a backcross population, a third character should be specified identifying the recurrent parent.
TS.pop	a list for each population containing the names of individuals in each TS
pop.type	character vector specifying the population type for each population (either "bc", "dh", "ril", "f2", "f3", "f4", "f5", "f6", or "f7", see Mangin et al. (2010))
pre	prefix for the files generated for the MCQTL software
trait	name or number of the trait in the gpData object to analyze
progression.step	numeric giving the progression step supplied to function ProbaPop in MCQTL. Default is 2.
dir	directory with consensus map named Consensus.map and QTL-detection parameter file named detection.xml. Chromosomes in Consensus.map must be named as chrom1, chrom2, …

Details

To run this function, the software MCQTL must be installed. In addition, put the following line once in your .bashrc file

export MCQTL_DUMP_FINAL_MODEL=1

Alternatively, the same command can be executed before each analysis session.

Data in the estimation set (ES) will be used to identify QTL and estimate QTL effects using MCQTL. The resulting QTL effects will be used to predict the genetic values for the individuals in the test set (TS).

A consensus map named Consensus.map must be provided in the working directory. A file named detection.xml must be provided which describes the parameters for the MCQTL analysis. See the manual for more information on how to specify this file (Mangin et al, 2010). In principle, the same file can be used which is used for the non cross-validated analysis.

If no QTL are identified within one fold, no genetic values and no predictive ability will be com-puted.

Input data can be prepared using the function create.gpData within the package synbreed (Wim-mer et al. 2012).

Value

A list of length number of replications, each including a list of length number of folds. Each list element contains the results for one fold within one replication

QTL.pos	A data.frame with chromosome of the QTL, position within chromosome, and explained variance (R2) identified in the respective ES
R2_global	global R2 value in the respective ES
predAbi	predictive ability defined as Pearsons correlation coefficients between predicted and observed genetic values in the respective TS
MSE	Mean-squared error of the predicted genotypic values in the respective TS
id.TS	genotypes included in the respective TS
reg.slope	regression slope of observed versus predicted genotypic values in the respectiveTS
yhat.TS	vector of predicted genotypic values in the respective TS

Author(s)

Valentin Wimmer and Flavio Foiada

References

Jourjon, M. F., S. Jasson, J. Marcel, B. Ngom, and B. Mangin (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21: 128-130.

Mangin B, Cathelin R, Delannoy D, Escalihre B, Lambert S, Marcel J, Ngom B, Jourjon M-F, Rahmani A, Jasson S (2010) MCQTL: a reference manual. Dipartement de Mathimatiques et In-formatique Appliquies. Institut National de la Recherche Agronomique (INRA), France

Wimmer V, Albrecht T, Auinger HJ, Schoen CC (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics, 28: 2086-2087

See Also

write.estimation, TS.generator

Examples

```
## Not run:
#  This example demonstrates the data preparation steps with
#  the maize data set from the synbreed package

#  Load maize data from synbreed package
library(synbreed)
library(synbreedData)
data(maize)
ids <- maize$covar$id
fam <- maize$covar$family
```

```
# generate one gpData object for one family
gp1 <- discard.individuals(maize,ids[is.na(fam) | fam!=24]) gp2 <-
discard.individuals(maize,ids[is.na(fam) | fam!=25])

# prepare input data pop.names <-
c(24,25)
# prepare test set allocations using random sampling TS.24 <-
TS.generator(ids[!is.na(fam) & fam==24],2,5) TS.25 <-
TS.generator(ids[!is.na(fam) & fam==25],2,5) TS.pop <- list(TS.24,
TS.25)
pop.list = list(gp1, gp2) pop.type =
c("dh", "dh")
parent.lines <- list(c(11331,11323),c(11323,11313))

# run cross-validation with MCQTL for trait 'Trait'
CV <- cv.MCQTL(pop.list=pop.list, pop.names=pop.names, parent.lines=parent.lines, TS.pop=TS.pop,
pop.type = pop.type, trait="Trait", dir=getwd())

## End(Not run)
```

TS.generator *Generation of cross-validation fold allocations*

Description

Set up a cross-validation scheme for a number of individual names using a defined number of replications and folds. The output equals the scheme of the crossVal function in the synbreed package.

Usage

TS.generator(id.names, Rep = 1, fold = 2)

Arguments

id.names	character vector of individual names
Rep	number of replications
fold	number of folds

Details

Output can be used to generate new cross-validation folds for the cv-MCQTL analysis.

Value

A list containing the individual names in the respective replications and folds

Author(s)

Valentin Wimmer

write.estimation 5

References

Wimmer V, Albrecht T, Auinger HJ, Schoen CC (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics, 28: 2086-2087

See Also

cv.MCQTL

Examples

```
TS.generator(letters,2,5)
```

---

write.estimation *Helper function to prepare input for MCQTL software.*

Description

Only used internally in cv.MCQTL

Usage

```
write.estimation(QTL.pos, QTL.chr, trait, parent.lines, nChr, dir = getwd())
```

Arguments

QTL.pos	QTL position within chromosome
QTL.chr	chromosome of QTL
trait	trait which is analyzed
parent.lines	parent lines used for the analysis
nChr	number of chromosomes
dir	working directory

Value

none

Author(s)

Valentin Wimmer

See Also

cv.MCQTL

Acknowledgements

In the first place I would like to express my deepest gratitude to Prof. Dr. Chris-Carolin Schön for giving me the opportunity to pursue my PhD studies at the Chair of Plant Breeding of TUM, in the frame of a most interesting research project combining experimental field work with current "hot topics" such as genome-wide prediction. Thanks to her dedicated supervision, I was able to substantially expand my knowledge in various subjects such as plant breeding methodology, quantitative genetics, statistical analysis and scientific writing.

Many thanks as well to Prof. Dr. Donna Pauler Ankerst and Prof. Dr. Ralph Hückelhoven for their engagement in the frame of my PhD thesis committee.

To the daily supervisor of my research at the Chair, Dr. Peter Westermeier, my most heartfelt thanks for his constant support and encouragement during all project's phases. From the introduction to the project and to experimental work in 2010 up to the writing of the paper and thesis manuscripts in 2013/2014, I highly appreciated his willingness to help as well as his constructive feedbacks.

My sincerest gratitude goes to our project partners at KWS SAAT AG for the pleasant and enriching collaboration, in particular to Dr. Bettina Kessel, Dr. Milena Ouzunova, Dr. Thomas Presterl, Dr. Michael Dilger and Dr. Ralph Kreps, all of whom were significantly involved in this project. Many thanks as well to Dorothee Seyfang and her field team at the KWS breeding station in Gondelsheim for their support during the corn borer artificial infestation and the maize stalk dissection activities at the Münzesheim field location.

From the Maize Breeding Group at the Bavarian State Research Center for Agriculture (LfL), I would like to kindly acknowledge Dr. Joachim Eder, Albert Ziegltrum and Monika Mesjasz for the excellent technical assistance at the Freising field location.

Dr. Rainer Messmer, former researcher at the Group of Crop Science, ETH Zurich (Switzerland), has been my mentor in the frame of the TUM graduate school program. I would like to thank him heartily for his commitment: his encouragements and numerous advices were deeply appreciated.

Special thanks go to Wolfgang Mayerhofer for his extensive engagement in this project, first in the frame of his internship at KWS SAAT AG and second with his master thesis that set the base for the genome-wide prediction analyses.

I also wish to express my gratitude to several colleagues from the technical staff at the Chair and to a large number of students for their engagement in the experimental field work. Their contribution was crucial to the amount and quality of the data produced. Many thanks in particular to Sylwia Schepella, Iris Leineweber, Felicitas Dittrich, Stefan Schwertfirm, Kurt Walter, Georg Meier, Amalie Fiedler, Lisa Schimanski, Fabian Hetterich as well as to Florian Steinbacher from the Gewächshauslaborzentrum Dürnast for providing access to the climate chambers for cultivating the corn borer larvae.

In 2012 I had the chance to spend two unforgettable months as a visiting scientist at CIMMYT in Nairobi, Kenya, in the frame of the "Insect resistant maize for Africa" (IRMA) project. My heartfelt thanks to Dr. Stephen Mugo and his team for this opportunity, for the numerous scientific inputs, and not least for their kindness and hospitality. Many thanks as well to the TUM graduate school for financial support during this stay.

I gratefully acknowledge the MCQTL software development team at the Applied Mathematics and Informatics Division of INRA in Toulouse (France) with thanks in particular to Dr. Brigitte Mangin and Sylvain Jasson, who provided technical advice on how to implement parts of our cross-validation procedure for the MCQTL software.

I also wish to thank Prof. Dr. Leo Dempfle and Prof. Dr. Daniel Gianola for a number of helpful discussions in the frame of the genome-wide prediction meetings held at our Chair.

I would like to acknowledge the German Federal Ministry of Food and Agriculture (BMEL) for financial support, with special thanks to Dr. Renate-Kaiser Alexnat who was in charge of this project. The research was funded by the "Innovationsförderung Programme" (FKZ: 2814504610).

To my fellow PhD students and all colleagues at the Chair: thank you for your friendship, for the enriching exchanges and of course for the various leisure activities. Time was great under Weihenstephan Hill! Many thanks in particular to Sebastian Gresset for his numerous scientific

inputs; to Valentin Wimmer who took the lead in developing the cvMCQTL cross-validation routine; to Christina Lehermeier for statistical counsel and feedback on this thesis; to Wiltrud Erath, Sandra Unterseer and Sebastian Steinemann for feedback on the paper and the cvMCQTL routine; to Manfred Schönleben, Theresa Albrecht, Hans-Jürgen Auinger and Christos Dadousis for sharing R-scripts and helping with R-analyses.

My gratitude also goes to my current work colleagues at Delley Seeds and Plants Ltd. (Switzerland) for showing flexibility and understanding whenever time was needed for executing urgent tasks related to the completion of this thesis.

I wish to express my deepest feelings for my parents who always supported me along my way, even if it sometimes took me far away from home.

One person supported me without reservation during the four years I was working on this thesis at Freising, and even more during the last intense year back in Switzerland. Estelle, I'm in lack of words, but thank you!

Curriculum vitae

Personal information

Flavio Foiada
Date of birth: July 3rd, 1983
Place of origin: Lavertezzo, Switzerland
Nationality: Swiss

Education

03 / 2010 – 03 / 2014
PhD thesis in plant breeding, Chair of Plant Breeding, Technische Universität München, Freising, Germany

10 / 2006 – 12 / 2008
M.Sc. in Agricultural Sciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

10 / 2003 – 10 / 2006
B. Sc. in Agricultural Sciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

09 / 1999 – 06 / 2003
High school at Liceo Cantonale di Locarno, Locarno, Switzerland

Professional experience

04 / 2014 –
Breeding, testing and maintenance breeding in cereals, Delley Seeds and Plants Ltd., Delley, Switzerland

03 / 2010 – 03 / 2014
Scientific collaborator at Chair of Plant Breeding, Technische Universität München, Freising, Germany

02 / 2009 – 09 / 2009
Civil service in rural development at Helvetas Swiss Intercooperation, Rajshahi, Bangladesh