

Evaluating the benefits and limitations of multiple-trait breeding assisted by genomics in cereal crops



Kumulative Dissertation

zur Erlangung des Doktorgrades der Agrarwissenschaften

Doctor agriculturæ (Dr. agr.)

der Naturwissenschaftlichen Fakultät III Agrar- und Ernährungswissenschaften,

Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Herrn Albert Wilhelm Schulthess Börgel

geboren am 14.02.1987 in Providencia, Santiago de Chile, Chile

1. Gutachter: Prof. Dr. Jochen Reif
2. Gutachter: Prof. Dr. Matthias Frisch

Eini greichvam 12.02.2018

Xgtvkl ki v'co "4; Ø20423: ""

Acknowledgements

Academic: I'm very thankful to all persons and institutions that academically participated during the course of this PhD Thesis, because without their collaboration this work could not be possible. I want to thank Prof. Dr. Jochen Reif, my first supervisor, for his support from the very beginning; even during the DAAD scholarship application process one year before arriving at Germany. I want to gratefully thank Dr. Yusheng Zhao and Dr. Yong Jiang because of their invaluable mathematical support, collaboration, suggestions, patience and constructive criticism respecting my research. I'm thankful to the German Academic Exchange Service (DAAD) for making possible most of my stay in Germany during the course of this PhD Thesis. I also want to thank those scientists that collaborated as co-authors in any of the peer-reviewed research articles that were produced under the frame of this PhD Thesis. In this last respect, I want to specially mention Dr. Yu Wang and Dr. Marion Röder. Last but not least, I want to thank my work colleagues at the Quantitative Genetics Group of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben). Among them, I specially thank Dr. Sang He, Norman Philipp and Guozheng Liu for the scientific discussions and the rich exchange of ideas that we had.

Non-academic: I'm very thankful to all people that somehow played a non-academic or personal role by supporting me during the course of this PhD Thesis, because, as well as those that provided academic support, without them this work could not be possible. First of all I want to thank all my family members, because without their love, values and support I would not be the person I'm today. Secondly, I want to thank all my friends. Among them, I want to specially thank Matze, Karencín, Steffi, Bene, Sofi, Leo, There, P-Illane, Jelena, Sonja, Christoph, Tim, Robert, Yuli, Christin, Macuyis, the Gliège family, Alice, Vivi, Mario-Hugo, Rafa and Seba. I also want to thank Prof. Dr. Jochen Reif because of his empathy by understanding that coming to Germany was a much more complex personal decision for me than simply getting a PhD degree. Last but not least, I want to thank all members of the Quantitative Genetics Group at the IPK Gatersleben for offering a nice working environment during the course of my PhD thesis.

an meine Familie und Freunde... ich habe euch sehr lieb!

„Man soll nur seine Arbeiten so gut und so mannigfaltig machen als man kann, damit sich jeder etwas auslese und auf seine Weise daran Theil nehme“

Johann Wolfgang von Goethe

Table of Contents

1. General introduction	6
1.1 The roles of agriculture and plant breeding.....	6
1.2 The importance of cereal crops and the multiple-trait problem	6
1.3 The genetic basis of trait correlations.....	8
1.4 The many definitions of pleiotropy	8
1.5 Trait correlations in gramineous species: a molecular perspective	9
1.6 Multiple-trait selection in the pre-genomic Era.....	13
1.7 Selection in the genomics Era	16
1.9 Objectives.....	19
Peer-reviewed scientific articles	20
2. The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (<i>Triticum aestivum</i> L.)	20
3. Genomic selection in hybrid breeding	21
4. Advantages and limitations of multiple-trait genomic prediction for <i>Fusarium</i> head blight severity in hybrid wheat (<i>Triticum aestivum</i> L.)	22
5. Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes	23
6. General discussion	24
6.1 How and when do we profit by using multiple-trait instead of single-trait approaches?.....	24
6.1.1 Genetic mapping: it is not only a matter of QTL detection power.....	24
6.1.2 Genomic prediction: improving accuracies by using indicator traits.....	26
6.2 Confronting the tyranny of numbers in multiple-trait complexes.....	27
6.3 Basic considerations for the implementation of a genomic assisted multiple-trait breeding program.....	29
7. Final outlook	34
8. Summary	36
9. Zusammenfassung	38
10. General references	40
11. List of general abbreviations	49
12. Curriculum vitae	50
13. Eidesstattliche Erklärung / Declaration under Oath	54
14. Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren / Declaration concerning Criminal Record and Pending Investigations	55

1. General introduction

1.1 The roles of agriculture and plant breeding

Agriculture is an inherent activity of the human nature. Domestication of plant and animal species, along with the active modification of local ecosystems and the manipulation of biotic communities facilitated the transition from hunter-gathering to a sedentary agriculture-based society during the Neolithic period. In cereal species, for instance, the transition from wild to domesticated forms, in which seeds are firmly held together in ears or panicles, prevented seeds from falling to the ground, thus facilitating their harvest (Doebley et al. 2006; Salamini et al. 2002; Zeder 2011). Agriculture allowed the development of a dense human population in the Fertile Crescent (Salamini et al. 2002), and still plays this role in a continuously growing world population nowadays (Borlaug 1968; Hall and Richards 2013; Saitone and Sexton 2017; Ray et al. 2012). Briefly, world's crops production arises from the combination of the harvested area and the production per soil area unit. While the expansion of harvested area played the major role during the first half of the 20th century, the production per soil area unit gained importance and became the main driving factor of global supply during the following decades (Slafer et al. 1996); an observation that nowadays still persists (FAO 2017a). Retrospectively, improvements in the production per soil area unit have been achieved due to the almost inseparable interaction between plant breeding and technological advances in agronomical practices (Austin et al. 1980; Borlaug 1968; Laidig et al. 2014). Nonetheless, plant breeding is mostly recognized as the driving force of these improvements. Plant breeding is the science, technique, business and art of genetic modification of plants for human benefits (Becker 2011; Bernardo 2010).

1.2 The importance of cereal crops and the multiple-trait problem

Cereal crops of economic significance encompass gramineous plant species (Poaceae or Gramineae botanic family) like barley (*Hordeum vulgare* L.), maize (*Zea mays* L.), pearl millet [*Pennisetum glaucum* (L.) R. Br.], oat (*Avena sativa* L.), rice (*Oryza sativa* L.), rye (*Secale cereale* L.), sorghum [*Sorghum bicolor* (L.) Moench], triticale (\times *Triticosecale* Wittm.), bread wheat (*Triticum aestivum* L.), and durum wheat (*T. turgidum* L. var. *durum*). Grains of cereals consist of an embryo, which contains the genetic material for a new plant, and a starch-rich endosperm, which provides the necessary energy reserves for the process of germination (McKevith

2004; Wrigley et al. 2017). The storability of cereal grains was perhaps the main reason why ancient civilizations were initially interested in cereals as food source, because food surpluses could be stored for their use during less favorable crop production periods (Salamini et al. 2002; Wrigley et al. 2017). This importance is still reflected nowadays, because cereals are the most produced commodity, with an average annual world production of 2.3 billion tons for the period 1994/2014 (FAO 2017a). Maize, rice and wheat correspond to the most produced cereals worldwide. In this sense, it is predicted that 40.6%, 28.7%, and 19.4% of the 2.6 billion tons of cereal production forecasted for the 2017/2018 period would correspond to maize, wheat and rice, respectively (FAO 2017b). Cereal grains are an important source of energy, carbohydrate, protein and fiber, while they also provide some micronutrients like vitamin E, vitamin B, magnesium and zinc (McKevith 2004; Wrigley et al. 2017). Cereal end-uses can be roughly divided in food, animal feed and other purposes. For instance, 43.4%, 35.3%, and 21.3% of the world cereal utilization during the period 2015/2016 corresponded to food, feed and other uses, respectively (FAO 2017b). Therefore, the main contribution of cereals to humans is as food and animal feed sources. Alternative uses of the cereal production include the obtainment of starch, alcoholic beverages and bioethanol (Wrigley et al. 2017).

The ideal grain characteristics would vary according to each end-use and, in most cases, would be also related to local preferences and/or legal restrictions (Sissons 2008; Wrigley et al. 2017). In addition, the distinct stages of the grain production chain, which goes from plant breeding, through crop production, storage and transportation, to processing and final consumption, would also have different needs (Borlaug 1968; Wrigley et al. 2017). The different crop production environments and the consequences of climate change make this picture even more complicated (Hall and Richards 2013; Ray et al. 2012). Moreover, human biology, psychology and their interactions, make human needs continuously evolve (Maslow 1943). For instance, some final consumers of food products demand nowadays healthy and nutritious products with new tastes, textures, colors, sizes and shapes, while also expecting that growers and the industry use social and environmentally friendly production methods (Bandara et al. 2016; Saitone and Sexton 2017). For sure, the final characteristics of end, as well as intermediate, products of the grain production chain would also depend on factors other than genetics (see for example the pasta production chain as reviewed by Sissons 2008). But plant breeding gives crop plant varieties the genetic potential to

achieve a particularly desired performance level (Becker 2011; Bernardo 2010). In this sense, plant breeders spend their life adapting plants and combining different traits in a single variety that could potentially meet certain needs according to the crop production context, end-use, market niche, among other considerations. This simultaneous selection of several traits is termed as multiple-trait selection (Falconer and Mackay 1996). The difficulty of combining different trait levels would be a function of trait correlations and breeding goals. As illustration, high grain yield is important for all production purposes in rye, while grain protein content requirements depend on the end-use of this crop (Miedaner et al. 2012). For example, since rye for feeding purposes must be commercially competitive with alternative crops like wheat, a sufficient protein content level would be required for this particular end-use. In contrast, protein content must be minimized for baking purposes. Thus, a negative correlation between grain yield and protein content would make the breeding of rye cultivars for animal feeding more difficult, while this correlation would be beneficial in the breeding for baking purposes.

1.3 The genetic basis of trait correlations

Trait correlations can be defined either phenotypically or genetically. In principle, a phenotypic correlation corresponds to the correlation among phenotypic values of different traits, whereas a genetic correlation is the correlation between genetic values or breeding values of different traits. In this sense, phenotypic correlations are the result of environmental and genetic variances and covariances (Mode and Robinson 1959). Whereas environmental correlations between traits are due to shared environmental influences on them (Almasy et al. 1997), genetic correlations are basically a function of pleiotropy and close-linkage (Chen and Lübberstedt 2010; Gianola et al. 2015; Hodgkin 1998; Jiang and Zeng 1995; Paaby and Rockman 2013; Solovieff et al. 2013; Stearns 2010).

1.4 The many definitions of pleiotropy

According to Stearns (2010), the term “*Pleiotropie*” was originally coined by the German geneticist Ludwig Plate in a compendium of scientific texts to praise Richard Hertwigs on his 60th birthday in 1910. Brüel (1912) reviewed the original text of Plate (1910) and mentioned “*Pleiotrop heißt dabei eine Einheit, von der mehrere Merkmale abhängen*”, which translates to “*A (genetic) unit, on which many traits rely, is called pleiotropic*“. This definition corresponds to the classical concept of

pleiotropy mentioned in quantitative genetics and breeding textbooks (e.g., Bernardo 2010; Falconer and Mackay 1996). Various authors (Paaby and Rockman 2013; Pavlicev and Wagner 2012; Stearns 2010) mainly attributed the hypothesis of “*universal pleiotropy*” to the works of Fisher (1930) and Wright (1968). According to this hypothesis, most genes, if not all, are capable of simultaneously influencing several, if not all, traits of an organism. Biological evidence points out to a “*modular pleiotropy*”, in which a set of genes tends to have pleiotropic effects on the same set of functionally and/or developmentally related traits (modules) but few and weaker effects on traits belonging to other module(s) (Wagner et al. 2007; Wang et al. 2010). Furthermore, evidence in organisms like yeast (*Saccharomyces cerevisiae*), nematode (*Caenorhabditis elegans*) and house mouse (*Mus musculus*) indicates that the frequency distributions of the number of traits influenced by a single locus are L-shaped (Wang et al. 2010). In other words, most loci tend to influence a small group of traits, while only a few of them tend to influence several traits simultaneously. In addition, advances in molecular genetics have shown that the ways in which one single locus can influence several traits simultaneously are quite diverse and complex, resulting in several, sometimes overlapping or even confounded definitions or subtypes of pleiotropy (e.g., Allison et al. 1998; Chen and Lübberstedt 2010; Hodgkin 1998; Paaby and Rockman 2013; Solovieff et al. 2013; Stearns 2010; Stephens 2013). For instance, Solovieff et al. (2013) used the terms biological and mediated pleiotropy to distinguish two types of genetic causalities. In biological pleiotropy a genetic factor has a direct and simultaneous influence on two traits, whereas a genetic factor has a direct effect on an intermediary phenotype and this phenotype, in turn, influences an outcome phenotype by means of mediated pleiotropy. Similar concepts were presented by other authors (e.g., Allison et al. 1998; Chen and Lübberstedt 2010; Hodgkin 1998; Paaby and Rockman 2013; Stearns 2010; Stephens 2013). Nonetheless, in the present work and following the original concept of Plate (1910), the term “*pleiotropy*” refers to one single locus whose variation is directly or indirectly associated with more than one trait simultaneously.

1.5 Trait correlations in gramineous species: a molecular perspective

Different transcripts and/or proteins can be produced from a single locus by alternative splicing, alternate start/stop codons, among others mechanisms (Chen and Lübberstedt 2010; Hodgkin 1998; Solovieff et al. 2013; Stearns 2010). As illustration, rice plants with mutated *LAGGING GROWTH AND DEVELOPMENT 1* (*lgd1*) locus

showed slower growth, reduced number of tillers, semi-dwarfism, altered panicle architecture and reduced grain yield in comparison to wild-types. The *LGDI* locus encoded at least six transcripts variants which originated from different transcription start sites. These transcripts were predicted to encode at least four different protein isoforms. Moreover, *LGDI* transcripts displayed distinct spatiotemporal expression patterns. For instance, in a spatial sense, the *LGDI.2* transcript were only highly expressed in the panicle and spikelet branches, while *LGDI.4* was expressed in several plant tissues but not in root or leaf blade. In a temporal sense, *LGDI.4* was expressed at many developmental stages, while *LGDI.2* was only expressed during tillering. Furthermore, complementing *lgdl* mutants with complementary deoxyribonucleic acid (cDNA) of *LGDI.1* produced only partially complementary phenotypes, which implies that other *LGDI* transcripts are required to perform the diverse functions associated to *LGDI* during rice growth and development (Thangasamy et al. 2012).

Metabolomic pathways or biochemical processes would be antagonistic to each other if they rely on the same limited metabolite(s)/protein(s) (Chen and Lübberstedt 2010; Paaby and Rockman 2013). For example, the cyclization of lycopene splits the biosynthesis of plant carotenoids into β, β and β, ϵ branches. β -carotene and other compounds are produced by the β, β branch, whereas the β, ϵ branch leads to α -carotene and its derivatives. The enzyme lycopene β -cyclase (β LCY) catalyzes the formation of β -carotene, whereas the co-action of enzymes β LCY and lycopene ϵ -cyclase (ϵ LCY) produces α -carotene (Howitt and Pogson 2006). The *LCYE* locus codifies for ϵ LCY in maize and its genetic variation influences the ratio of α to β branch carotenoids in grains. This last finding opened the door to the breeding of maize with enhanced provitamin A content by selecting *LCYE* alleles which are associated with decreased α to β branch carotenoids ratios (Harjes et al. 2008).

A genetic factor can have a direct influence on an intermediary phenotype and this last one, in turn, affects an outcome phenotype which has no direct association with the former causal genetic factor (e.g., Allison et al. 1998; Chen and Lübberstedt 2010; Hodgkin 1998; Paaby and Rockman 2013; Solovieff et al. 2013; Stearns 2010; Stephens 2013). For instance, traits like grains per spike, grain weight, spikes per plant, among others, are intermediary phenotypes of the more complex outcome phenotype grain yield in cereal species (Slafer et al. 1996). This is also reflected in that some quantitative trait loci (QTL) influencing maize grain yield could be predicted by the multiplicative interactions among QTL of component traits (Balestre et al. 2012).

The same gene product acting in different tissues can also produce correlated phenotypes (Chen and Lübberstedt 2010; Hodgkin 1998; Paaby and Rockman 2013; Solovieff et al. 2013; Stearns 2010). The first step of carotenoid synthesis is the condensation of two geranylgeranyl pyrophosphate molecules, which is catalyzed by the enzyme phytoene synthase (PSY). Carotenoids play tissue-specific roles in plants. For instance, carotenoids are essential components of the light harvesting system and also protect membranes from excess light damage in photosynthetic tissue. In parallel, carotenoids are important for abscisic acid production and dormancy in seeds, while they also play a role in limiting free radical-induced membrane deterioration and aging of seeds (Howitt and Pogson 2006). In maize, the yellow endosperm phenotype has been a strong target of selection since the 1930s because of its former association with increased nutritional value due to an enhanced carotenoid content (Fu et al. 2010). The paralog PSY1 enzyme is codified by the *Y1* locus in maize (Buckner et al. 1996) and a guanine nucleotide insertion at 377 base pairs downstream of the ATG initiator codon produces a new stop codon in the mutant *y1-8549*, which leads to a null *PSY1* allele (Li et al. 2008). This may explain why *y1-8549* mutants are characterized by a white endosperm along with pale green seedlings and mature plants when grown at high temperatures compared to wild-type plants (Li et al. 2008; Robertson and Anderson 1961).

In gene expression QTL (eQTL) analyses, transcript levels of genes are considered as intermediary molecular phenotypes, and statistical associations between genetic polymorphisms and them are investigated (Gilad et al. 2008; Paaby and Rockman 2013; Solovieff et al. 2013). Genomic regions with a high number of distal eQTL might contain so-called “*master regulators*” that influence the expression of many genes (Gilad et al. 2008). In this sense, trait correlations could also arise from interactions of epistatic nature (Chen and Lübberstedt 2010; Hodgkin 1998; Paaby and Rockman 2013; Pavlicev and Wagner 2012; Solovieff et al. 2013; Stearns 2010). For instance, a significant eQTL hotspot was detected on chromosome 6D of wheat. This hotspot influenced the expression of 54 genes related to three different interconnected modules in the mature wheat embryo (Munkvold et al. 2013).

Adaptive modifications of a character are associated with deleterious pleiotropic effects on other characters; an evolutionary process also known as the cost of adaptation or fitness tradeoff (Fisher 1930; Paaby and Rockman 2013; Pavlicev and Wagner 2012; Stearns 2010; Todisco et al. 2010). The fitness penalty due to

pleiotropic effects of some types of weed herbicide resistance is a good illustration of this (Délye 2005; Vila-Aiub et al. 2009). For instance, herbicides inhibiting the activity of the homomeric chloroplastic enzyme acetyl-CoA carboxylase (ACCase) can selectively control grass weed species by blocking the first step in the synthesis of fatty acids. Nonetheless, due to the high effective-control capability of this type of herbicides, their overuse led to the rapid selection of resistant weed plants (Délye 2005). In vitro assays of ACCase activity of monocot species like *Alopecurus myosuroides* (Huds.) (Délye et al. 2005) and *Lolium rigidum* (Yu et al. 2007) have shown that different mutations producing specific amino acid substitutions in the enzyme can decrease the enzyme sensitivity to some ACCase-inhibiting herbicides; thus, these mutations confer herbicide resistance. Interestingly, in the absence of herbicide, some of these mutations conferring resistant presented ACCase in vitro activity levels inferior to those achieved by susceptible plants, which suggests the existence of fitness cost due to the acquired resistance (Délye et al. 2005; Yu et al. 2007). One explanation for this could be that some amino acid substitutions conferring herbicide resistance may also impair enzyme activity and/or reduce substrate affinity (Vila-Aiub et al. 2009). For example, two amino acid substitutions conferring resistance with fitness cost in *Alopecurus myosuroides* (Huds.) may be involved in carboxyl-transferase activity of ACCase, because they are located in the polar area at the bottom of the carboxyl-transferase active-site cavity (Délye et al. 2005).

As previously mentioned, closely linked loci, each of them influencing a single trait, can also induce associations between different traits (Chen and Lübberstedt 2010; Gianola et al. 2015; Hodgkin 1998; Jiang and Zeng 1995; Paaby and Rockman 2013; Solovieff et al. 2013; Stearns 2010). For example, in the development of near isolines or in backcross breeding, closely linked loci associated to other traits different from that of interest can be dragged along with the targeted genome region intended to be introgressed (Brinkman and Frey 1977). This phenomenon is usually referred as linkage drag in the literature (e.g., Chen and Lübberstedt 2010; Häberle 2007; Kowalski et al. 2016; McCartney et al. 2007; Kuchel et al. 2007; Tanksley and Nelson 1996; Voss-Fels et al. 2017). For instance, two adjacent QTL at 137.1 and 143.5 cM on chromosome 5B interact epistatically and were associated to variation in root biomass in a diverse population of wheat. In addition, a QTL associated to heading date was found between these two root biomass QTL. Interestingly, the haplotype associated to increased root biomass was exclusively carried by Chinese accessions,

while this genomic region in European varieties conformed a haplotype block with loci in very strong linkage disequilibrium (Voss-Fels et al. 2017). Since heading date plays an important role in the adaptation of wheat as a crop (Worland 1996), these observations suggested that the haplotype variant associated to increased root mass disappeared from European wheat genotypes as a consequence of linkage drag during the selection for adapted plant material and the fixation of the closely linked heading date QTL (Voss-Fels et al. 2017).

1.6 Multiple-trait selection in the pre-genomic Era

Plant breeders use conscious or unconsciously the metric of selection gain or response to selection in order to measure how effective a trait would be improved by means of selection. Basically, this metric is the difference in average genetic values between the original population before selection and the offspring of the selected fraction. Since one generation interval is needed to observe this difference, the response to selection is usually predicted without the need of true selection by using the “*breeders’ equation*”:

$$\bar{R} = i * h * \sigma_G, \quad (1)$$

with \bar{R} being the expected or predicted response to selection, i denoting the selection intensity, while h and σ_G correspond to the square roots of heritability (h^2) and genetic variance (σ_G^2) of a particular trait, respectively. The selection intensity is a function of the number of individuals selected relative to the phenotypic variation and size of the original population, whereas the heritability is the proportion of the phenotypic variance which can be attributed to the genetic variance (Bernardo 2010; Falconer and Mackay 1996).

Tandem selection and independent culling levels are basic methods of multiple-trait selection. In the tandem method, selection is performed for one trait at a time during successive breeding cycles, until each trait of a set of multiple-traits reaches a desirable level. In other words and assuming only two traits, the first trait is improved during the first breeding cycle(s), while selection is performed for the second trait during the following generation(s) but using the fraction selected from the first generation(s) as base population. In the independent culling levels approach, different thresholds or culling levels are applied to all traits under selection in a single generation and only those individuals that pass all thresholds are selected (Bernardo 2010; Falconer and Mackay 1996). Nonetheless, in certain situations these methods would have some disadvantages. First, if two traits are negatively correlated and are

thought to be selected in the same direction, selection for the second trait during subsequent generation(s) of a tandem selection program could counteract some of the progress that was achieved for the first trait. In addition, applying too strict culling levels when several traits are under selection, i.e. extremely high selection intensities for each trait, could result in no individuals being selected, because even individuals being exceptional for all except one trait would be discarded (Bernardo 2010). Moreover, some traits contribute more than others to the economic value of a cultivar as a whole. This is reflected in the “*aggregated breeding value*” (H , Hazel 1943; Hazel and Lush 1942; Smith 1936), which is basically a linear combination of single trait breeding values based on their relative economic importance:

$$H_i = a_1 * g_{1i} + a_2 * g_{2i} + \dots + a_k * g_{ki}, \quad (2)$$

where H_i is the aggregated breeding value of the i th individual, a_1 to a_k are relative economic values related to each of the k traits under selection, whereas g_{1i} to g_{ki} are the breeding values of the i th individual for each of the k traits. Nevertheless, breeding values are in practice unknown; thus, only linear combinations using phenotypic values can be performed. The Smith-Hazel or net merit selection index (I , Hazel 1943; Hazel and Lush 1942; Smith 1936) was proposed to deal with this limitation:

$$I_i = b_1 * y_{1i} + b_2 * y_{2i} + \dots + b_k * y_{ki}, \quad (3)$$

where I_i is the net merit of the i th individual, b_1 to b_k are index weights related to each of the k traits that maximize the correlation between I and H , while y_{1i} to y_{ki} are the phenotypic values for each of the k traits of the i th individual. Provided that I and H are vectors containing the net merits and aggregated breeding values of all the individuals in the population under selection, respectively, the index weights for k traits are obtained by a multiple-regression procedure of H on I . From a single-trait point of view and assuming i and σ_G in Equation (1) as constants, maximizing h will maximize \bar{R} , i.e. the expected gain of selection (Bernardo 2010; Falconer and Mackay 1996). Similarly, maximizing the correlation between I and H will maximize \bar{H} , i.e. the predicted aggregated gain of selection, which is a measure of the efficiency of multiple-trait improvement (e.g., Hazel 1943; Hazel and Lush 1942; Kempthorne and Nordskog 1959; Smith 1936; Suwantaradon et al. 1975). Moreover, under simplifying assumptions and considering high selection intensities, \bar{H} is expected to be higher using a net merit index than applying independent culling levels and these two methods, in turn, would be more efficient than performing tandem selection (Hazel and Lush 1942). Briefly, the principle underlying net merit indices is that the

excellence of some traits can, to a certain extent, compensate the weaknesses with respect to other traits (Elston 1963). In theory, different independent culling levels could be selected in such a way that \bar{H} is also maximized. Nevertheless, the selection of such culling levels can be quite tedious (Hazel and Lush 1942) and the net merit index solves this problem in a single step. Furthermore, the Smith-Hazel index will result in the maximization of \bar{H} , regardless of whether or not the means of single-trait breeding values in Equation (2) change in a positive or negative direction but, in practice, plant breeders may occasionally want to keep some traits at the mean level or unchanged while improving others. For this reason, Kempthorne and Nordskog (1959) proposed a restricted selection index which allowed the inclusion of such restrictions. A few years later Williams (1962) presented a base index in which the index weights in Equation (3) are simply replaced by the relative economic values of Equation (2). However, economic weights may be difficult to estimate for traits like plant height or abiotic stress resistance, because such traits are not directly handled in the market. In consequence, others modifications to the Smith-Hazel index which do not rely on economic weights like the desired gains index (Pešek and Baker 1969) or the weight-free multiplicative index (Elston 1963) have been proposed. The selection index theory has been used, for instance, for the simultaneous improvement of grain yield, kernel weight and grain protein content in wheat (Gebre-Mariam and Larter 1996), for the simultaneous selection of grain and straw yields in oat (Eagles and Frey 1974) and for the simultaneous breeding of biotic and abiotic stresses resistance along with grain yield and other traits in recurrent selection programs of maize (Suwantaradon et al. 1975).

Last but not least, a special form of multiple-trait selection corresponds to indirect selection. Indirect selection is based on the change in average genetic values for a particular trait caused by selection for a correlated trait; a term known as the correlated response to selection. Similarly as in Equation (1), this metric can be predicted without the need of true selection, in the way:

$$\overline{CR}_{T1} = i_{T2} * h_{T2} * r_G * \sigma_{G_{T1}}, \quad (4)$$

where \overline{CR}_{T1} is the expected correlated response to selection for trait T1 when selecting for trait T2, i_{T2} and h_{T2} denote the selection intensity and square root of heritability for T2, correspondingly, r_G represents the genetic correlation between T1 and T2, whereas $\sigma_{G_{T1}}$ is the square root of the genetic variance for T1. In indirect selection, the trait indirectly selected (T1) is usually referred as target trait, while the trait being

directly selected (T2) is denominated indicator trait. Moreover, from Equations (1) and (4) it can be deduced that indirect selection for T1 using T2 as indicator trait would be more effective than direct selection of T1 when $h_{T2} * r_G > h_{T1}$. In other words, a target trait with low heritability would potentially benefit from indirect selection using a genetically correlated indicator trait with higher heritability (Becker 2011; Bernardo 2010; Falconer and Mackay 1996). For instance, these benefits have been evaluated in the indirect selection for grain yield in resource-limited environments by selecting yield in high-yielding environments in barley (Ceccarelli et al. 1992) or by using physiological indicator traits in maize (Ziyomo and Bernardo 2013), or in the indirect selection for grain yield under irrigated conditions by selecting spectral reflectance indices in wheat (Babar et al. 2007). Moreover, the concept of indirect selection is implicit in the crop ideotype theory (Donald 1968), in which selecting for those traits pertaining to the high-yielding plant ideotype may lead to a high-yielding cultivar. Nonetheless, this is not the only reason why indirect selection could be beneficial. Indicator traits which are easier or less expensive to measure than a target trait are also very attractive because of their potential to decrease costs and simplify the phenotyping process in applied breeding (Becker 2011; Bernardo 2010; Falconer and Mackay 1996). In addition, there are target traits like dry matter and biomass yields in rye which are measured by destructive methods. Therefore, if these traits are directly measured during the early stages of selection, this could hamper the seed multiplication process. In this case, using plant height as non-destructive indicator trait has proven to be very effective (Haffke et al. 2014).

1.7 Selection in the genomics Era

Plant breeding is a costly labor in terms of time and inputs. For instance, plant breeders need in average between five and ten years from the development of the base population for selection to the final obtainment of a new crop candidate variety (e.g., Becker 2011; Bernardo 2010; Collard and Mackill 2008; Longin et al. 2015). Therefore, the plant breeding community is continuously searching for more efficient breeding methods that can accelerate breeding programs or reduce their costs. One of these methods is the marker-assisted selection (MAS). According to Lande and Thompson (1990), MAS can be interpreted as an indirect selection method (Section 1.6), in which molecular markers are used as indirect traits for the selection of an associated target trait. In this sense, molecular markers like those based on DNA polymorphisms are less prone to be affected by environmental variation (they are

measured with more precision) and can be in some situations even simpler or cheaper to measure than the target trait (Collard and Mackill 2008; Kumar 1999). Nonetheless, selection purely based on molecular scores would replace phenotypic selection only in those cases when the proportion of additive genetic variance of the target trait that is associated with marker loci equals the heritability of the target trait. Otherwise, selection should be based on a mixed-criteria, either by combining marker and phenotypic information into a selection index or by performing early selection based on marker profiles followed by phenotypic selection in later stages (a sort of tandem selection), among other strategies (Dekkers 2007; Lande and Thompson 1990). In cereal crops, MAS has been mainly applied to assist backcross and pyramiding breeding (Collard and Mackill 2008; Kumar 1999). A subgroup of markers used in MAS is the so-called “*functional markers*”. Functional markers are derived from polymorphic sites within genes that have a biological cause-effect relationship with phenotypic variation (Liu et al. 2012). For example, functional markers linked to dwarfing loci *Rht-B1* and *Rht-D1* allow distinguishing between tall and dwarfing genotypes of wheat (Ellis et al. 2002). It follows from the above that marker-trait associations are the essence of MAS. Basically, the knowledge about these associations arises from genetic mapping studies (Collard and Mackill 2008; Kumar 1999; Lande and Thompson 1990). Among them, linkage and association mapping are the most used ones. Both of these mapping methods use the ability of recombination to break the genome into fragments that can be subsequently correlated with phenotypic variation. However, they mainly differ in the way recombination is being generated. In linkage mapping controlled crosses are performed to generate a mapping population, which sets up a closed genetic system with known levels of relatedness among individuals but few recombination points in genomes. In contrast, association mapping uses diverse populations of genotypes, which configures an open system with a high number of historical recombination events in genomes but uncontrolled relatedness among individuals (Myles et al. 2009). Nonetheless, highly polygenic traits cannot be efficiently targeted by MAS, because minor effect loci remain often undetected by genetic mapping. On the contrary, genomic prediction uses all markers information to predict the breeding values of unobserved traits in selection candidates (Meuwissen et al. 2001; Whittaker et al. 2000). Dekkers (2007) extended the MAS theory of Lande and Thompson (1990) to the context of genomic predictions. Similarly as in MAS, the correlation between predictions and true breeding values would determine if genomic

selection should be purely based on genomic predictions or on a mixture of phenotypic values and genomic predictions in order to maximize selection gains. Genomic prediction of maize testcross grain yield performance under optimal (Endelman et al. 2014; Ziyomo and Bernardo 2013) and drought conditions (Ziyomo and Bernardo 2013), of grain yield in barley (Endelman et al. 2014), of *Fusarium* head blight resistance in wheat populations of inbred lines (Mirdita et al. 2015a; Rutkoski et al. 2012) and hybrids (Mirdita et al. 2015b), of grain yield and other traits in rye (Wang et al. 2014), are only a few of several examples that have evaluated the potential of this technology in cereal crops.

1.8 Multiple-trait mapping and genomic prediction methods

Multiple-trait mixed model approaches were introduced several years ago in the pioneering work of Henderson and Quaas (1976). In principle, the multiple-trait additive genetic model allows the simultaneous modelling of several traits by using the genetic relationships between individuals along with the genetic and residual variances and covariances among traits. Simulation and theoretical studies have revealed some of the advantages of multiple-trait over single-trait analyses in genetic mapping and genomic prediction:

- i) In the context of genetic mapping, multiple-trait methods are expected to provide more precise estimates of QTL effects, an enhanced power of QTL detection, a decreased rate of false positive pleiotropic associations and a formal frame of work to test if the co-location of QTL for more than one trait is because of pleiotropy or due to closely linked single-trait QTL (Allison et al. 1998; Almasy et al. 1997; Banerjee et al. 2008; Caliński et al. 2000; Cheng et al. 2013; Da Costa E Silva et al. 2012; Ferreira and Purcell 2009; Jiang and Zeng 1995; Knott and Haley 2000; Mangin et al. 1998; Thoen et al. 2017; Varona et al. 2004; Williams et al. 1999; Zhou and Stephens 2014; Stephens 2013).
- ii) In genomic prediction, the benefits of incorporating information on genetically correlated indicator trait(s) within a multiple-trait model are similar to those of indirect selection (Section 1.6). In this sense, predictabilities for a low heritable target trait can be improved if a genetically correlated indicator trait with higher heritability is included within the genomic prediction model (Calus and Veerkamp 2011; Dahl et

al. 2016; Guo et al. 2014; Hayashi and Iwata 2013; Jia and Jannink 2012; Jiang et al. 2015).

1.9 Objectives

When the present PhD work was started (June 2014), studies directly dealing with the multiple-trait selection problematic in cereal crops by using state-of-the-art multiple-trait genome-wide association mapping or multiple-trait genomic prediction/selection approaches were scarce. Therefore, the main goal of this work was to explore the potential and limits of multiple-trait genomic methods to assist the improvement of multiple-traits in line and hybrid breeding of cereal crops. The current work covers the following specific objectives:

- (1) To dissect the pleiotropic architecture of the bread wheat yield-syndrome into pleiotropy and close-linkage by means of multiple-trait association mapping in a diverse population of varieties adapted to European environments (Schulthess et al. 2017a).
- (2) To review the state-of-the-art of genomic prediction in hybrid breeding with an emphasis on the implementation of genomic selection in model crops wheat and maize (Schulthess et al. 2017c).
- (3) To study the influence of the relatedness between estimation and prediction sets on the benefits of multiple-trait genomic prediction of the target trait *Fusarium* head blight severity using plant height and heading date as indicator traits in a vast hybrid wheat population (Schulthess et al. 2017b).
- (4) To identify an efficient method for genomic prediction of selection indices for the simultaneous improvement of grain yield and protein content in rye with feeding purposes using two populations of hybrid rye (Schulthess et al. 2016).

Peer-reviewed original research article

2.

The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (*Triticum aestivum* L.)

Published in:

Journal of Experimental Botany, Vol. 68, No. 15 pp. 4089–4101, 2017

DOI: 10.1093/jxb/erx214

Authors: Albert W Schulthess, Jochen C Reif, Jie Ling, Jörg Plieske, Sonja Kollers, Erhard Ebmeyer, Viktor Korzun, Odile Argillier, Gunther Stiewe, Martin W Ganal, Marion S Röder
Yong Jiang

The original paper has been published and available online:

<https://academic.oup.com/jxb/article/68/15/4089/3979722/The-roles-of-pleiotropy-and-close-linkage-as>

Peer-reviewed literature review article

3.

Genomic selection in hybrid breeding

Published as Chapter 7 (pp. 149–183) in the book:

Genomic selection for crop improvement: new molecular breeding strategies for crop improvement. Edited by Rajeev K Varshney, Manish Roorkiwal and Mark E Sorrells.

Published by Springer International Publishing AG, Cham, 258 pages, 2017

DOI: 10.1007/978-3-319-63170-7

Authors: Albert W Schulthess, Yusheng Zhao, Jochen C Reif

The original book chapter has been published and available online:

https://link.springer.com/chapter/10.1007/978-3-319-63170-7_7

Peer-reviewed original research article

4.

Advantages and limitations of multiple-trait genomic prediction for *Fusarium* head blight severity in hybrid wheat (*Triticum aestivum* L.)

Published in:

Theoretical and Applied Genetics, pp. 1–17, 2017

DOI: <https://doi.org/10.1007/s00122-017-3029-7>

Authors: Albert W Schulthess, Yusheng Zhao, C. Friedrich H. Longin, Jochen C Reif

The original paper has been published and available online:

<https://link.springer.com/article/10.1007/s00122-017-3029-7>

Peer-reviewed original research article

5.

Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes

Published in:

Theoretical and Applied Genetics, Volume 129, Issue 2 pp. 273–287, 2016

DOI: 10.1007/s00122-015-2626-6

Authors: Albert W Schulthess[†], Yu Wang[†], Thomas Miedaner, Peer Wilde, Jochen C Reif, Yusheng Zhao

[†]: Both authors contributed equally to this work

The original paper has been published and available online:

<https://link.springer.com/article/10.1007%2Fs00122-015-2626-6>

6. General discussion

6.1 How and when do we profit by using multiple-trait instead of single-trait approaches?

The principle of parsimony is well embodied in the aphorism of George Box (1979): “*All models are wrong but some are useful*”. In this sense, a model with a small number of parameters is considered parsimonious if this can describe a phenomenon as accurate as a more sophisticated model which has more parameters. From the theoretical point of view, a parsimonious model may perhaps be far away from what we conceive as the true mechanistic model describing a phenomenon. Nonetheless, the simplicity and robustness of parsimonious models make them an attractive choice compared to unpractical and unnecessarily complicated models which may only impose difficulties in model fitting, implementation and interpretability. Consequently, the principle of parsimony plays a central role when single- and multiple-trait approaches are being compared. In the current section, the main advantages of multiple-trait methods presented in Chapter 2 (Schulthess et al. 2017a), Chapter 4 (Schulthess et al. 2017b), and Chapter 5 (Schulthess et al. 2016) of the present PhD work are highlighted and discussed.

6.1.1 Genetic mapping: it is not only a matter of QTL detection power

Many authors in the context of genetic mapping in plants recognize the increased power in QTL detection, a decreased rate of false positive pleiotropic associations and the possibility to formally test pleiotropy versus close-linkage as the main potential advantages of multiple-trait over single-trait approaches (e.g., Caliński et al. 2000; Cheng et al. 2013; Jaiswal et al. 2016; Lebreton et al. 1998; Malosetti et al. 2008; Mangin et al. 1998; Schulthess et al. 2017a; Stich et al. 2008; Thoen et al. 2017). For instance, the pleiotropic architecture of grain yield and correlated traits such as plant height, thousand grain weight, grains per ear, among others, was investigated by means of multiple-trait genome-wide association mapping in Chapter 2. This method detected genomic regions which simultaneously influenced grain yield of wheat and, at least, one trait correlated to it. Moreover, a multiple-trait approach implemented in Chapter 2 allowed discerning if pleiotropy per se or, alternatively, two closely linked loci were the most likely cause underlying co-located QTL for grain yield and correlated traits. Here should be emphasized that while both, multiple- and

single-trait methods can detect QTL for one or more traits, only multiple-trait approaches provide the proper statistical frame of work to answer the pleiotropy versus close-linkage question (Jiang and Zeng 1995). Nonetheless, it was also shown in Chapter 2 that decreased allele frequencies and QTL sizes together with increased linkage disequilibrium between simulated QTL located at two different loci hinder the ability to differentiate true close-linkage from presumed pleiotropy. These and other factors limiting or enhancing this ability were already discussed in detail in Chapter 2. However, differences in the abilities of single- and multiple-trait approaches to detect QTL for multiple-trait complexes were barely covered in Chapter 2. Therefore, this topic is extensively discussed in the following paragraphs.

Provided a null correlation between traits, the evidence favoring the association between variation in genetic polymorphisms and correlated-trait complexes found by multiple-trait genetic mapping will be approximately the sum of evidences observed by separately analyzing each trait using single-trait approaches. On the contrary, when traits are correlated, a multiple-trait test statistic may be far different from the sum of separate single-trait test statistics (Jiang and Zeng 1995). Since trait distributions are in this last case no longer independent, the experiment may be more suited to be studied using a multivariate k -dimensional distribution, with k being the number of traits under consideration (Lebreton et al. 1998). For example, variation in a particular genomic region may have a practically negligible effect on certain trait but a small to intermediate effect on a second trait correlated to the first one, and single-trait genetic mapping may fail to detect any QTL in this region. However, the joint analysis of these two traits may increase the power of detection of the QTL for the second trait. In this regard, several theoretical and simulation studies have shown that the joint analysis of traits can potentially increase the power in QTL detection (e.g., Allison et al. 1998; Almasy et al. 1997; Banerjee et al. 2008; Cheng et al. 2013; Da Costa E Silva et al. 2012; Ferreira and Purcell 2009; Jiang and Zeng 1995; Knott and Haley 2000; Mangin et al. 1998; Thoen et al. 2017; Zhou and Stephens 2014; Stephens 2013). Nonetheless, most of these studies have also shown that this advantage of multiple-trait approaches is to a major or minor extent influenced by factors such as the percentage of trait variance explained by QTL, the population size, the genetic distance between markers and QTL, the allele frequencies at a pleiotropic locus, the number of traits, as well as the magnitude and sign of genetic and residual trait covariances or correlations. For instance, Jiang and Zeng (1995) already showed in

their pioneering work that if the residual correlation and the cross product between QTL effects for two simulated traits are different in signs, the power of joint analysis would be always higher than the overall power of two separated single-trait genetic mapping analyses.

Statistical power is not the only criteria of QTL detection to be considered when deciding between single- and multiple-trait approaches for genetic mapping. Meta-analyses of several single-trait genetic mapping analyses run in parallel for different correlated traits may result in declaring an excessive number of loci with significant pleiotropic associations; a type of error which may be better accounted by multiple-trait genetic mapping approaches (Allison et al. 1998; Caliński et al. 2000; Ferreira and Purcell 2009). However, this last advantage lies on the increased critical values due to the additional number of parameters estimated in multiple-trait genetic mapping approaches that could, in turn, decrease QTL detection power. Therefore, the ability of joint mapping to overall increase QTL detection power relies on the inclusion of information on additional traits that can compensate for the more stringent critical values of multiple-trait approaches (Allison et al. 1998; Almasry et al. 1997; Jiang and Zeng 1995; Mangin et al. 1998).

Even though no single genetic mapping approach will be the most powerful in detecting the many different types of genetic effects underlying multiple-trait complexes (for an extensive review refer to Stephens 2013), the overall picture suggests that the balance between decreased false pleiotropic associations and increased QTL detection power of multiple-trait genetic mapping approaches overshadows the losses in power that can result in specific situations where single-trait analyses appear more powerful.

6.1.2 Genomic prediction: improving accuracies by using indicator traits

Over the last couple of years, the crop plant science community became more interested in multiple-trait genomic prediction because of its potential to improve prediction accuracies over those achieved by single-trait approaches (Bao et al. 2015; He et al. 2016; Hori et al. 2016; Jia and Jannink 2012; Lyra et al. 2017; Marchal et al. 2016; Montesinos-López et al. 2016; Qiu et al. 2017; Rutkoski et al. 2012, 2016; Santos et al. 2016; Schulthess et al. 2016, 2017b; Wang et al. 2016). For instance in Chapter 4, the potential of indicator traits plant height and heading date to improve the predictability of *Fusarium* head blight severity in hybrid wheat by means of multiple-trait genomic prediction was evaluated. As well, the benefits of shifting from single-

towards multiple-trait genomic prediction for grain yield and protein content in two populations of hybrid rye were investigated in Chapter 5. Both chapters confirmed what extensive simulation studies have shown so far: provided an indicator trait is genetically correlated to and has higher heritability than a particular target trait, including its information within a multiple-trait genomic prediction model can be more accurate for the target trait than single-trait prediction (Calus and Veerkamp 2011; Dahl et al. 2016; Guo et al. 2014; Hayashi and Iwata 2013; Jia and Jannink 2012; Jiang et al. 2015). Nevertheless, these simulation studies along with Chapter 4 and Chapter 5 also showed that this niche of advantage for multiple-trait genomic prediction may be influenced by several factors other than the gap in heritabilities or the genetic correlation between target and indicator traits. For instance, experimental results of Chapter 4 showed that even though a genetically correlated indicator trait with higher heritability is included in the model, the difference in predictability between multiple- and single-trait genomic prediction for the target trait tends towards zero as the level of relatedness between estimation and prediction set increases. Thus, in those situations where accuracies of single- and multiple-trait approaches are comparable and due to the high computational burden of multiple-trait approaches (e.g., Allison et al. 1998; Calus and Veerkamp 2011; Da Costa E Silva et al. 2012; Hayashi and Iwata 2013; Hori et al. 2016; Malosetti et al. 2008; Zhou and Stephens 2014), the principle of parsimony will finally promote the use of single-trait methods. Nonetheless, it may be anticipated that this current picture would change in the future due to advances in breeding informatics. This last topic is discussed in more detail in the following section.

6.2 Confronting the tyranny of numbers in multiple-trait complexes

Methods that can reduce the computational load but also conserve the advantages of multiple-trait analyses are among the “*Philosopher’s stones*” of many quantitative geneticists and breeding informaticians. In the context of genetic mapping, one of the earliest and simplest methods trying to deal with this issue consisted in obtaining linear combinations of traits by applying, for instance, principal component analysis, followed by or coupled with univariate analyses of each of these new generated variables (e.g., Allison et al. 1998; Cheng et al. 2013; Mangin et al. 1998). A further step in sophistication has been the implementation of multiple-trait approaches with reduced number of parameters as compared to other less parsimonious multiple-trait models (e.g., Hori et al. 2016; Malosetti et al. 2008; Calus

and Veerkamp 2011; Williams et al. 1999). For example, Calus and Veerkamp (2011) proposed two Bayesian models for multiple-trait genomic prediction in which marker effects on traits were expressed as the product between a vector of marker effect sizes and a vector of scaling factors that transform these effects to the scale of each trait. This parameterization implies that the number of additional parameters for marker effect estimation in multiple-trait over single-trait approaches is, at the model level, exactly equal to the number of additional traits included in the multiple-trait model. In contrast, in multiple-trait approaches like those considered in Chapter 2, Chapter 4 and Chapter 5, this number increases multiplicatively, which makes the computational load gap between single- and multiple-trait approaches broader. Moreover, the development or implementation of efficient methods for statistical inference has been a continuous topic in the contexts of multiple-trait genetic mapping (e.g., Banerjee et al. 2008; Caliński et al. 2000; Da Costa E Silva et al. 2012; David et al. 2013; Ferreira and Purcell 2009; Jiang and Zeng 1995; Knott and Haley 2000; Lebreton et al. 1998; Stephens 2013; Varona et al. 2004; Zhou and Stephens 2014) and genomic prediction (e.g., Dahl et al. 2016; Hayashi and Iwata 2013; He et al. 2016; Montesinos-López et al. 2016; Wang et al. 2016). For instance, alternative ways to test pleiotropy versus close-linkage that are potentially more efficient or simpler to implement than the multiple-trait likelihood ratio approach of Jiang and Zeng (1995) have been proposed (e.g., Banerjee et al. 2008; Caliński et al. 2000; David et al. 2013; Knott and Haley 2000; Lebreton et al. 1998; Varona et al. 2004).

In the future and facilitated by advances in modern and automatized –omics technologies besides genomics (e.g., Neumann et al. 2015; Patti et al. 2012; Xiong et al. 2017), the number of expression traits like transcripts, proteins, metabolites, or phenotypes available in multiple-trait datasets would dramatically increase. In this sense, the future bottle neck will not be the generation of expression data anymore but, the search for new methods to wisely and optimally use this extremely large amount of information. Breeding informatics will play a central role to confront this challenge. For instance, Cheng et al. (2013) developed a variable selection algorithm to maximize statistical power in QTL detection when the number of traits is too large and the joint analysis of all traits becomes unpractical. Moreover, in Chapter 4 and Chapter 5 it was shown that when a set of indicator traits is at hand for multiple-trait genomic prediction and the goal was to improve the predictability or prediction accuracy of one particular target trait, a two-trait model using the best indicator trait available for that

target trait would be enough. In contrast, Wang et al. (2016) found that adding more traits into a multiple-trait model can further increase the predictability of a particular target trait in hybrid rice. In this respect, the development of algorithms analogous to the one of Cheng et al. (2013) but considering prediction accuracies instead of QTL detection power will help to optimize the number of indicator traits needed in multiple-trait genomic prediction applications.

6.3 Basic considerations for the implementation of a genomic assisted multiple-trait breeding program

Given a limited budget, the challenge of implementing genomic selection as a tool to assist hybrid plant breeding was dissected in Chapter 3 (Schulthess et al. 2017c) using basic optimization theory. In this problem, the relative efficiency of genomic selection over pure phenotypic selection per unit of time should be maximized. Until the date, simulation and experimental plant data studies evaluating the optimal allocation of resources between estimation and prediction sets have been only assessed from the single-trait point of view (Endelman et al. 2014; Krchov and Bernardo 2015; Longin et al. 2015; Lorenz 2013; Riedelsheimer and Melchinger 2013). Analogously to single-trait cases (as reviewed in Chapter 3), the \bar{H} ratio per unit of time (see Section 1.6 for a detailed explanation of \bar{H}) between multiple-trait selection assisted by genomics and pure phenotypic multiple-trait selection should be the objective function to be maximized (Cerón-Rojas et al 2015). It is anticipated that the current section is far away from solving this non-linear multivariate problem and future studies on this topic are for sure needed. Nevertheless, some principles that should be considered for the implementation of multiple-trait selection assisted by genomics, i.e. using multiple-trait MAS and genomic prediction approaches, are presented here.

A theoretical decision tree that may maximize \bar{H} by means of genomic tools is proposed in Figure 6.1. The root node in Figure 6.1 is the estimation set for marker effects on traits, which corresponds to a multivariate box of k traits, n genotypes and p DNA-based molecular markers. If there is previous knowledge about useful indicator traits (for example from the literature or from own breeder's experience), these should be included in the estimation set along with those traits directly involved in H . Nonetheless, certain assumptions and clarifications must be first introduced before genomic-based breeding strategies for multiple-trait improvement are explained:

- i) Although it is very likely that different traits have a mixture of genetic architectures; it is assumed that either only a few major or only several minor effect loci underlie variation for all traits.
- ii) Individuals within the estimation set and selection candidates are assumed to be inbred lines. Thus, traits are assumed to be under the control of additive genetic effects and their epistatic interactions.
- iii) It was already discussed in Chapter 2 that traits appearing uncorrelated at the phenotypic level may be correlated at the genetic level and vice versa. Likewise, it was discussed in Chapter 2 that even if traits have a null net genetic correlation, they could still have a partially shared genetic control. Furthermore, it is clear from Section 6.1 that multiple-trait approaches exploit the information contained in correlated-trait complexes in a better manner than their single-trait counterparts, either because trait covariation is of genetic, environmental or phenotypic nature. Therefore, a scenario in which no benefits at all are expected from using multiple-trait approaches needs the very strong assumption that there are absolutely no genetic or environmental factors inducing covariation among traits. However, this case is simply included for illustration purposes.
- iv) n and p have been already optimized according to certain restrictions and parameters; although this is not a trivial issue (see Chapter 3). In this regard, markers are assumed to be in very strong linkage-disequilibrium with functional loci underlying trait variation, which is of uppermost relevance for the obtainment of reliable genetic correlations estimates in the multiple-trait context (Gianola et al. 2015).
- v) The $k \times n \times p$ multidimensional system is assumed to be balanced and if this is not the case, marker (He et al. 2015) and phenotype imputation (e.g., Dahl et al. 2016; Hori et al. 2016; Jia and Jannink 2012) strategies should be considered. In this sense, relying on a balanced instead of an unbalanced estimation set may, for instance, increase power in QTL detection (He et al. 2015; Dahl et al. 2016).
- vi) In Chapter 4 it was concluded that a low level of relatedness between estimation and prediction sets is needed to benefit from multiple-trait

genomic prediction. Therefore, estimation set and predicted selection candidates are assumed as not or, at most, barely related.

- vii) Phenotype imputation is the method of choice for multiple-trait genomic prediction. Thus, indicator traits not only must be measured within the estimation set but also on selection candidates. Consequently, this additional phenotyping cost should be considered in the implementation problem.

The first node of the tree in Figure 6.1 corresponds to the phenotypic analyses, in which first and second degree statistics such as trait correlations and heritabilities are studied. Information generated at this stage will be useful to evaluate, for example, the potential of benefits from indirect selection (See Section 1.6 for more details). The second level of nodes is marked by genome-wide association mapping within the estimation set, which may reveal if major effect loci underlie trait variation. In the case that traits are uncorrelated and major effect loci are detected, haplotype effects will be calculated using estimated single-trait marker effects of major loci and marker profiles of selection candidates. If traits have a truly polygenic architecture and traits are uncorrelated, H predictions of selection candidates by obtained by means of a direct genomic prediction approach (Chapter 5).

When traits are correlated, multiple-trait genome-wide association and prediction approaches will have a central role. In this case, if multiple-trait genetic mapping reveals no major effect loci, multiple-trait genomic prediction can be applied to obtain more accurate predictions for traits with low heritability by using information on indicator traits of selection candidates. Afterwards, H of selection candidates may be predicted using a reverse genomic prediction approach (Chapter 5) or other alternative methods (Cerón-Rojas et al. 2008, 2015; Dekkers 2007). In case some detected QTL with major effects on different traits co-locate, differentiating true pleiotropy from close-linkage will define different breeding strategies depending on the sign of the locus-induced covariation and the multiple-trait breeding goals (Chapter 2). In this regard, trait covariation which fits perfectly to multiple-trait breeding goals is termed as desired trait covariation while covariation which does not, is considered undesired. Thus, when a test of pleiotropy versus close-linkage cannot discard pleiotropy in the estimation set, performing MAS using pleiotropic loci that induce desired trait covariation may be a very efficient way to simultaneously improve multiple-traits (pleiotropic multiple-trait-MAS in Figure 6.1). However, if co-located

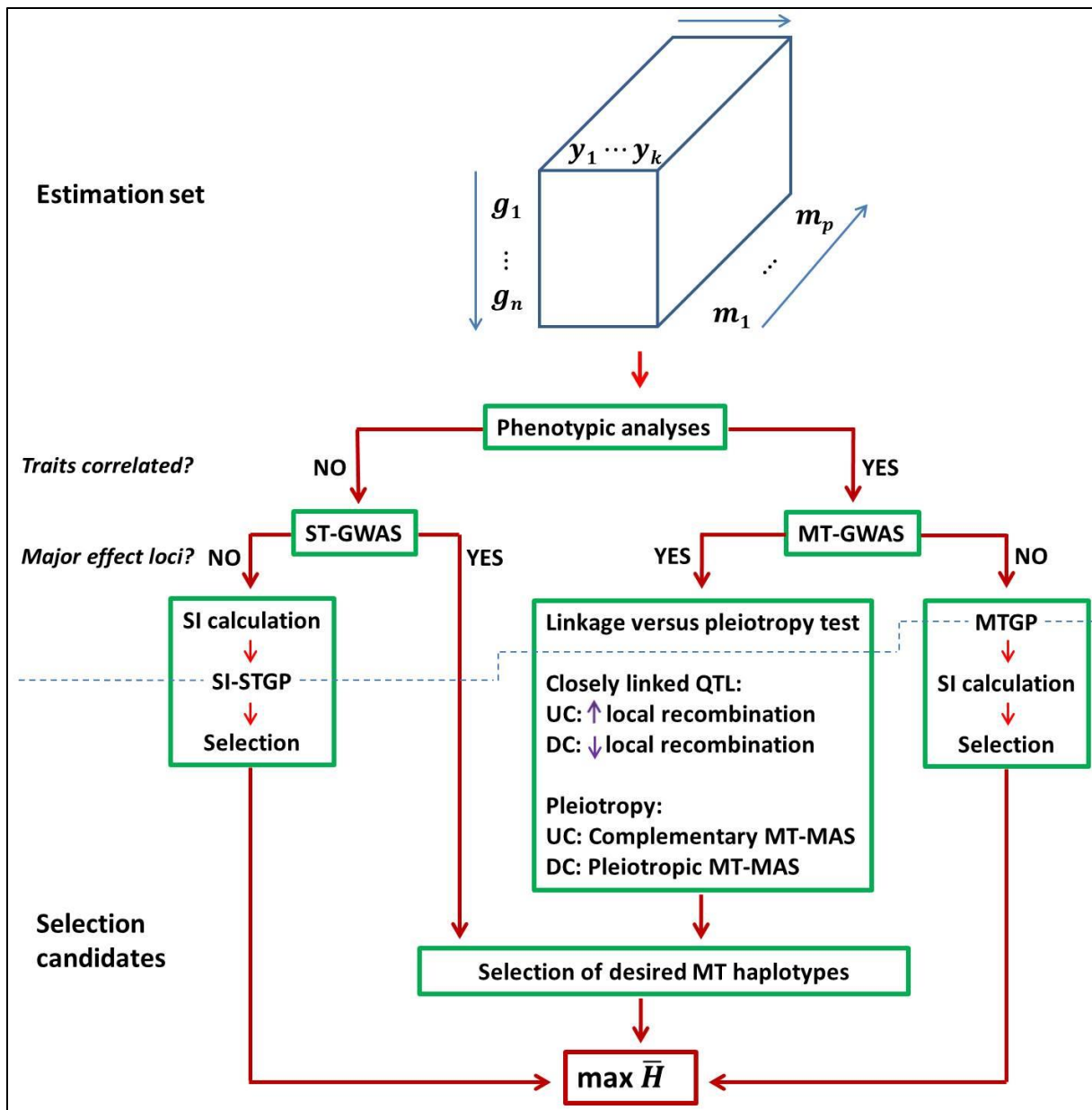


Figure 6.1. Decision tree in a multiple-trait (MT) breeding program assisted by genomics for the maximization of the aggregated gain of selection (\bar{H}). The estimation set is represented by a three-dimensional system involving phenotypic values (y) for k traits of n individuals genotyped with p DNA-based molecular markers. Depending on the strategy, single-trait (ST) and MT genome-wide association analyses (-GWAS) as well as ST and MT genomic prediction (-GP) models are considered. In the absence of major effect loci, selection index (SI) prediction methods are contemplated. When major loci are present, the type of trait covariation induced by loci, i.e. undesired (UC) or desired covariation (DC) is taken into account for marker-assisted selection (MAS). A dashed line separates the tree into two halves, with each of them involving those tasks related to the estimation set or to the group of selection candidates, respectively. See text for a more detailed description.

QTL induce undesired trait covariation, it should be evaluated which of the traits controlled by a pleiotropic locus will be improved relying on that particular locus, while effects on the rest of the traits are assumed as “*improvement costs*” and should be complemented using other major effect loci detected by multiple-trait genetic mapping (complementary multiple-trait-MAS in Figure 6.1). From an evolutionary perspective, this last case may basically correspond to a selection-pleiotropy-compensation model (Pavlicev and Wagner 2012). Nonetheless, the cost-benefit evaluation of such multiple-trait haplotypes is not a trivial issue and further applied breeding studies are needed on this topic. The amount of trait variation explained by QTL as well as the economic relevance of traits should be among the basic factors considered by these future studies.

In case traits are correlated and the co-location of some detected QTL with major effects on different traits is due to closely linked loci, strategies of multiple-trait MAS will basically depend on the frequency of multiple-trait haplotypes within the group of selection candidates (Chen and Lübberstedt 2010). If alleles in line with breeding goals are in coupling phase in some selection candidates, their selection is straight forward and would lead to haplotype fixation (a decrease in local recombination) for the selected material (Bernardo 2010). When alleles that fit to breeding goals are in repulsion phase in all selection candidates, strategies to increase recombination rate along with genetic variation must be evaluated. For instance, one strategy would be to cross suboptimal selection candidates, which already have desired haplotypes at all other major loci, with individuals less related to them. Subsequently new inbred line populations could be developed from these crosses. In this sense, less related individuals could be, for example, lines in the estimation set that already carry the desired haplotypes for those region(s) with undesired linkage phases among selection candidates. Alternatively, breeders could rely on the decreased linkage disequilibrium and increased genetic variation existent in unadapted material to increase the local recombination rate at genomic region(s) targeted by multiple-trait MAS (Chen and Lübberstedt 2010; Longin and Reif 2014; Tanksley and Nelson 1996). In this regard, coupling backcross breeding (Tanksley and Nelson 1996) and doubled haploid techniques (Melchinger et al. 2013) may accelerate the development of new selection candidates with desired multiple-trait haplotypes. Other strategies to increase the frequency of rare events of recombination in multiple-trait MAS were mentioned by Chen and Lübberstedt (2010). Nonetheless, since recombination rates

are not evenly distributed in genomes, these strategies may be not effective in all cases. The most common observation is that in many species like human (*Homo sapiens*), rat (*Rattus norvegicus*), house mouse (Jensen-Seaman et al. 2004), and *Arabidopsis thaliana* (Fernandes et al. 2017) the recombination rate tends to increase towards telomeric regions of chromosomes, whereas this rate is reduced near centromeres. As illustration in cereals, this may explain why the haplotype structure of the *laxatum-a* locus of barley, which is located in the pericentromere area of chromosome 5HL, is significantly less diverse than that of the *uniculme4* locus (Jost et al. 2016), which maps to the distal end of chromosome 3HL (Tavakol et al. 2015). Moreover, a recent study in *Arabidopsis thaliana* revealed that combined effects of mutated *figl1*, *recq4*, and *fancm* loci can increase the recombination rate up to 7.8-fold (genome-wide average) as compared to recombination in wild-type plants (Fernandes et al. 2017). Nevertheless, no significant changes in recombination rate were observed between these mutants and wild-type plants for intervals encompassing or immediately flanking centromeres. Further studies are needed to evaluate the prospects and limits to enhance recombination rate relying on homologues of these loci in cereal species. However, here should be emphasized that a limited recombination rate and genetic variation among selection candidates is not exclusively problematic when the undesired trait covariation is caused by closely linked loci (Figure 6.1). Actually, these features are also limited in all traditional breeding programs that rely on elite plant material (e.g., Becker 2011; Longin and Reif 2014; Tanksley and Nelson 1996). Thus, these and further strategies to overcome these problems [see for instance the strategy to exploit genetic resources for the improvement of quantitative traits in Longin and Reif (2014)] should be evaluated in a case-by-case fashion and are beyond the scope of this PhD thesis.

In summary, the decision tree (Figure 6.1) would lead to the maximization of \bar{H} , either by selecting those individuals with desired multiple-trait haplotypes when all traits are under the control of major loci or by choosing those selection candidates that maximize genomically predicted H values when all traits are influenced by minor loci.

7. Final outlook

It is clear that more benefits than disadvantages are expected when shifting from single- towards multiple-trait genome-wide association and prediction approaches. Advances in breeding informatics may further facilitate this shift in the future. In addition, the selection index theory for the improvement of multiple-traits

was introduced almost eight decades ago and, along with the indirect selection theory, have been a theoretical backbone for marker-assisted and genome-wide selection approaches. Nonetheless, it results somehow curious that during the genomic era the selection index theory has been barely considered by plant molecular breeders for what it was originally conceived for: multiple-trait improvement. Thus, we expect that the current work can illustrate annual crop breeders on how to embark into the improvement of multiple-traits assisted by genomics.

8. Summary

Plant breeders spend most of their time and efforts adapting plants and combining different traits in a single variety that can potentially meet certain production, end-use, and market niche needs. This task is known as multiple-trait breeding. In this respect, breeders dealing with multiple-traits can take advantage of some trait correlations or in the contrary, they must try to break undesired correlations by means of recombination. When this PhD work was started (June 2014), studies focused on multiple-trait selection by using multiple-trait genome-wide association mapping or multiple-trait genomic prediction/selection approaches were scarce in cereal crops. Therefore, the main goal of this work was to explore the potential and limits of multiple-trait genomic methods to assist the improvement of multiple-traits in line and hybrid breeding of cereal crops.

Basically, traits are genetically correlated because of the net effect of co-located quantitative trait loci (QTL) for different traits. A multiple-trait genome-wide association mapping study in a population of 372 bread wheat (*Triticum aestivum* L.) varieties adapted to Europe and genotyped with 18,856 polymorphic markers revealed that several QTL for grain yield and its components co-locate. In the Era of molecular breeding, distinguishing if this QTL co-location is due to pleiotropy or caused by closely linked genes, would determine the optimum strategy for multiple-trait breeding. In most genomic regions with co-located QTL for grain yield and its components there was not enough statistical evidence to strongly rule out pleiotropy as the cause of QTL co-location. A simulation study in this wheat population showed that the ability to differentiate QTL-colocation due to close-linkage from that caused by pleiotropy was potentially limited by low minor allele frequencies, along with small QTL sizes and high rates of linkage disequilibrium between QTL. On this respect, increasing population size and marker density may improve the ability to dissect the pleiotropic architecture of grain yield and correlated traits.

An extensive literature survey on genomic prediction in hybrid crops showed that the relatedness between estimation and prediction sets is a major factor influencing genomic predictability. In parallel, past simulation studies revealed that provided an indicator trait is genetically correlated to and has higher heritability than a particular target trait, including its information within a multiple-trait genomic prediction model can be more accurate for the target trait than single-trait prediction. Nonetheless, a study on multiple-trait genomic prediction of the target trait *Fusarium*

head blight severity of hybrid wheat showed for the very first time in crop plants that the benefits of multiple-trait prediction are also impacted by relatedness. In this sense, even though a genetically correlated indicator trait with higher heritability is included in the model, the difference in predictability between multiple- and single-trait genomic prediction for the target trait tends towards zero as the level of relatedness between estimation and prediction set increases.

Genetic correlations are beneficial in multiple-trait genomic prediction regardless of whether or not these correlations complicate multiple-trait improvement. Thus, multiple-trait genomic prediction cannot be directly used to break undesirable trait correlations. Grain yield and protein content were negatively correlated in two testcross populations of rye (*Secale cereale* L.). Different selection indices potentially allow the simultaneous improvement of these traits when rye is used for feeding purposes. These selection indices can be accurately and efficiently predicted by single-trait genomic prediction considering selection indices as if they were single-traits.

9. Zusammenfassung

Pflanzenzüchter beschäftigen sich mit der Anpassung von Kulturpflanzen an ihre Zielumwelten und der Kombination von verschiedenen Eigenschaften in einer einzigen Sorte, um die Bedürfnisse der Pflanzenproduktion, Verbraucher und der jeweiligen Marktnische zu erfüllen. Diese Aufgabe wird als Züchtung von mehreren Merkmalen bezeichnet. Während der Selektion auf mehrere Merkmale können Züchter von Merkmalskorrelationen profitieren oder müssen unerwünschte Assoziationen durch genetische Rekombination reduzieren. Zu Beginn dieser Doktorarbeit (Juni 2014) gab es nur wenige Studien zur Anwendung multivariater Assoziationskartierung und genomischer Selektion in Getreiden. Deswegen war es das Hauptziel dieser Doktorarbeit, die Chancen und Grenzen multivariater genomischer Methoden für die Selektion auf mehrere Merkmale in der Linien- und Hybridzüchtung von Getreidearten zu untersuchen.

Merkmale können aufgrund von überlappenden *Quantitative Trait Locus* (QTL)-Regionen genetisch korreliert sein. In einer Population von 372 europäischen Brotweizensorten (*Triticum Aestivum* L.), die mit 18,856 polymorphen molekularen Markers genotypisiert worden sind, wurden durch multivariate Assoziationskartierung mehrere überlappende QTL-Regionen für den Kornertrag und verschiedene Ertragskomponenten detektiert. In der Ära der molekularen Züchtung sind detaillierte Kenntnisse zu solch überlappenden QTL-Regionen entscheidend, um eine optimale Selektionsstrategie auf mehrere Merkmale zu wählen. Hauptursachen von überlappenden QTL-Regionen sind Pleiotropie oder enge Kopplung zwischen Genen. Für die Mehrheit der überlappenden QTL-Regionen konnte Pleiotropie als genetische Ursache nicht ausgeschlossen werden. In einer Simulationsstudie basierend auf dieser Weizenpopulation konnte gezeigt werden, dass die Möglichkeiten zwischen Pleiotropie und enger Kopplung zu unterscheiden von einer niedrigen Frequenz seltener Allele, von QTL mit kleinen Effekten und einem hohen Kopplungsungleichgewicht zwischen Genen begrenzt werden. In diesem Sinne würde eine Erhöhung der Populationsgröße und der Markerdichte die Chancen erhöhen, um die pleiotrope genetische Architektur vom Kornertrag und seinen korrelierten Merkmale besser zu verstehen.

Eine umfassende Literaturarbeit über die genomische Vorhersage in Hybridkulturarten belegt, dass die Verwandtschaftsbeziehung zwischen den Trainings- und Validierungspopulationen einen wesentlich Einfluss auf die Genauigkeit der

genomischen Vorhersagen hat. Weiterhin zeigen bisherige Simulationsstudien, dass die Merkmalsvorhersage genauer werden kann, wenn man anstatt eines univariaten ein multivariates genomisches Vorhersagemodell nutzt. Letzteres ermöglicht es, die Informationen von genetisch korrelierten Hilfsmerkmalen zu berücksichtigen. Dies ist vorteilhaft sofern die Heritabilität des Hilfsmerkmals höher als die vom Zielmerkmal ist. Eine experimentelle Studie über multivariate genomische Vorhersagen der Anfälligkeit gegenüber der Ährenfusariose vom Hybridweizen belegt weiterhin, dass der Zugewinn an Vorhersagegenauigkeit durch multivariate versus univariate Modelle von Verwandtschaftsbeziehung zwischen Trainings- und Validierungspopulationen abhängt: je enger der Verwandtschaftsgrad desto geringer ist der Genauigkeitsunterschied.

Genetische Korrelationen sind von Vorteil in der multivariaten genomischen Vorhersage unabhängig davon, ob sie die Selektion auf mehreren Merkmale komplizierter oder leichter machen. Deswegen können multivariate Vorhersagen nicht direkt für das Aufbrechen ungewünschter Merkmalskorrelationen verwendet werden. In zwei Testkreuzungspopulationen vom Roggen (*Secale cereale* L.) waren der Korntrag und der Proteingehalt negativ korreliert. Verschiedene Selektionsindizes erlauben die kombinierte Verbesserung beider Merkmale was wichtig ist, wenn Roggen als Futtermittel genutzt wird. Die Selektionsindizes können genau und effizient durch die Anwendung von univariaten Modellen vorhergesagt werden, indem jeder Selektionsindex als ein eigenes Merkmal berücksichtigt wird.

10. General references

- Allison DB, Thiel B, St Jean P, Elston RC, Infante MC, Schork NJ (1998) Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *Am J Hum Genet* 63:1190–1201
- Almasy L, Dyer TD, Blangero J (1997) Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet Epidemiol* 14:953–958
- Austin RB, Bingham J, Blackwell RD, Evans LT, Ford MA, Morgan CL, Taylor M (1980) Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *J Agric Sci* 94:675–688
- Babar MA, van Ginkel M, Reynolds MP, Prasad B, Klatt AR (2007) Heritability, correlated response, and indirect selection involving spectral reflectance indices and grain yield in wheat. *Aust J Agric Res* 58:432–442
- Balestre M, Von Pinho RG, Souza CL Jr, Bueno Filho JSS (2012) Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. *Theor Appl Genet* 125:479–493
- Bandara BES, De Silva DAM, Maduwanthi BCH, Warunasinghe WAAI (2016) Impact of food labeling information on consumer purchasing decision: with special reference to faculty of agricultural sciences. *Procedia Food Science* 6:309–313
- Banerjee S, Yandell BS, Yi N (2008) Bayesian quantitative trait loci mapping for multiple traits. *Genetics* 179:2275–2289
- Bao Y, Kurle JE, Anderson G, Yong ND (2015) Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Mol Breed* 35:128. doi:10.1007/s11032-015-0324-3
- Becker H (2011) *Pflanzenzüchtung* (in German). Auflagenr. 2. Verlag Eugen Ulmer, Stuttgart
- Bernardo R (2010) *Breeding for quantitative traits in plants*. Stemma Press, Woodbury
- Borlaug NE (1968) Wheat breeding and its impact on world food supply. *Australian Academy of Science, Australia*, pp 1–36
- Brinkman FN, Frey KJ (1977) Yield component analysis of oat isolines that produce different grain yields. *Crop Sci* 17:165–168
- Brüel L (1912) Referat: Vererbungslehre und Deszendenztheorie (Plate, L). In: Baur E (ed), *Zeitschrift für Induktive Abstammungs- und Vererbungslehre*. Band VIII, Heft 3. Verlag von Gebrüder Borntraeger, Berlin, pp 302-304
- Buckner B, San Miguel P, Janick-Buckner D, Bennetzen J (1996) The *yl* gene of maize codes for phytoene synthase. *Genetics* 143:479–488
- Caliński T, Kaczmarek Z, Krajewski P, Frova C, Sari-Gorla M (2000) A multivariate approach to the problem of QTL localization. *Heredity* 84:303–310
- Calus MPL, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol* 43:26. doi:10.1186/1297-9686-43-26
- Ceccarelli S, Grando S, Hamblin J (1992) Relationship between barley grain yield measured in low- and high-yielding environments. *Euphytica* 64:49–58
- Cerón-Rojas JJ, Crossa J, Arief VN, Basford K, Rutkoski J, Jarquín D, Alvarado G, Beyene Y, Semagn K, DeLacy I (2015) A genomic selection index applied to simulated and real data. *G3* 5:2155-2164

- Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008) A restricted selection index method based on eigenanalysis. *J Agric Biol Environ Stat* 13:440–457
- Chen Y, Lübberstedt T (2010) Molecular basis of trait correlations. *Trends Plant Sci* 15:454–461
- Cheng R, Borevitz J, Doerge RW (2013) Selecting informative traits for multivariate quantitative trait locus mapping helps to gain optimal power. *Genetics* 195:683–691
- Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos T Roy Soc B* 363:557–572
- Da Costa E Silva L, Wang S, Zeng ZB (2012) Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genet* 13:67. doi:10.1186/1471-2156-13-67
- Dahl A, Iotchkova V, Baud A, Johansson A, Gyllensten U, Soranzo N, Marchini J (2016) A multiple-phenotype imputation method for genetic studies. *Nat Genet* 48:466–472
- David I, Elsen JM, Concorde D (2013) CLIP test: a new fast, simple and powerful method to distinguish between linked or pleiotropic quantitative trait loci in linkage disequilibria analysis. *Heredity* 110:232–238
- Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. *J Anim Breed Genet* 124:331–341
- Délye C (2005) Weed resistance to acetyl coenzyme A carboxylase inhibitors: an update. *Weed Sci* 53:728–746
- Délye C, Zhang XQ, Michel S, Matejicek A, Powles SB (2005) Molecular bases for sensitivity to acetyl-coenzyme-A-carboxylase inhibitors in black-grass. *Plant Physiol* 137:794–806
- Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321
- Donald CM (1968) The breeding of crop ideotypes. *Euphytica* 17: 385-403
- Eagles HA, Frey KJ (1974) Expected and actual gains in economic value of oat lines from five selection methods. *Crop Sci* 14:861–864
- Ellis MH, Spielmeier W, Gale KR, Rebetzke GJ, Richards RA (2002) Perfect markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. *Theor Appl Genet* 105:1038–1042
- Elston RC (1963) A weight-free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics* 19:85–97
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, Jannink JL (2014) Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci* 54:48-59
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Ronald Press Company, New York
- FAO (2017a) Food and Agriculture Organization of the United Nations, FAOSTAT, <http://www.fao.org/faostat/en/#data>. Accessed 02 July 2017
- FAO (2017b) Food and Agriculture Organization of the United Nations, Food outlook June 2017. <http://www.fao.org/3/a-i7343e.pdf>. Accessed 02 July 2017

- Fernandes JB, Seguela-Arnaud M, Larcheveque C, Lloyd AH, Mercier R (2017) Unleashing meiotic crossovers in hybrid plants. bioRxiv. doi:<https://doi.org/10.1101/159640>
- Ferreira MA, Purcell SM (2009) A multivariate test of association. *Bioinformatics* 25:132–133
- Fisher RA (1930) *The genetical theory of natural selection*. Clarendon, Oxford
- Fu Z, Yan J, Zheng Y, Warburton ML, Crouch JH, Li JS (2010) Nucleotide diversity and molecular evolution of the *PSYI* gene in *Zea mays* compared to some other grass species. *Theor Appl Genet* 120:709–720
- Gebre-Mariam H, Larter EN (1996) Genetic response to index selection for grain yield, kernel weight and per cent protein in four wheat crosses. *Plant Breed* 115:459–464
- Box G (1979) Robustness in the strategy of scientific model building. In: Launer RL and Wilkinson GN (eds), *Robustness in Statistics*, Academic Press New York, pp 201–236
- Gianola D, de los Campos G, Toro MA, Naya H, Schön CC, Sorensen D (2015) Do molecular markers inform about pleiotropy? *Genetics* 201:23–29
- Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet* 24:408–41
- Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G (2014) Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet* 15:30. doi:10.1186/1471-2156-15-30
- Häberle J, Schmolke M, Schweizer G, Korzun V, Ebmeyer E, Zimmermann G, Hartl L (2007) Effects of two major *Fusarium* head blight resistance QTL verified in a winter wheat backcross population. *Crop Sci* 47:1823–1831
- Haffke S, Kusterer B, Fromme FJ, Roux S, Hackauf B, Miedaner T (2014) Analysis of covariation of grain yield and dry matter yield for breeding dual use hybrid rye. *Bioenergy Res.* 7:424–429
- Hall AJ, Richards RA (2013) Prognosis for genetic improvement of yield potential and water-limited yield of major grain crops. *Field Crops Res* 143:18–33
- Harjes C, Rocheford T, Bai L, Brutnell T, Kandianis C, Sowinski S (2008) Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science* 319:330–333
- Hayashi T, Iwata H (2013) A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics* 14:34. doi: 10.1186/1471-2105-14-34
- Hazel LN (1943) The genetic basis for constructing selection indexes. *Genetics* 28:476–490
- Hazel LN, Lush JL (1942) The efficiency of three methods of selection. *J Hered* 33:393–399
- He D, Kuhn D, Parida L (2016) Novel applications of multitask learning and multiple output regression to multiple genetic trait prediction. *Bioinformatics* 32:i37–i43
- He S, Zhao Y, Mette MF, Bothe R, Ebmeyer E, Sharbel TF, Reif JC, Jiang Y (2015) Prospects and limits of marker imputation in quantitative genetic studies in European elite wheat (*Triticum aestivum* L.). *BMC Genomics* 16:1. doi:10.1186/1471-2164-16-1

- Henderson CR, Quass RL (1976) Multiple trait evaluation using relatives' records. *J Anim Sci* 43:1188–1197
- Hodgkin J (1998) Seven types of pleiotropy. *Int J Dev Biol* 42:501–505
- Hori T, Montcho D, Agbangla C, Ebana K, Futakuchi K, Iwata H (2016) Multi-task Gaussian process for imputing missing data in multi-trait and multi-environment trials. *Theor Appl Genet* 129:2101–2115
- Howitt CA, Pogson BJ (2006) Carotenoid accumulation and function in seeds and non-green tissues. *Plant Cell Environ* 29:435–445
- Jaiswal V, Gahlaut V, Meher PK, Mir RR, Jaiswal JP, Rao AR, Balyan HS, Gupta PK (2016) Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T. aestivum* L.). *PLoS One* 11:e0159343
- Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Jacob HJ (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* 14:528-538
- Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192:1513–1522
- Jiang C, Zeng Z (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140:1111–1127
- Jiang J, Zhang Q, Ma L, Li J, Wang Z, Liu JF (2015) Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity* 115:29-36
- Jost M, Taketa S, Mascher M, Himmelbach A, Yuo T, Shahinnia F, Rutten T, Druka A, Schmutzer T, Steuernagel B, Beier S, Taudien S, Scholz U, Morgante M, Waugh R, Stein N (2016) A homolog of *Blade-On-Petiole 1* and 2 (*BOP1/2*) controls internode length and homeotic changes of the barley inflorescence. *Plant Physiol* 171:1113–1127
- Kempthorne O, Nordskog AW (1959) Restricted selection indices. *Biometrics* 15:10–19
- Knott SA, Haley CS (2000) Multitrait least squares for quantitative trait loci detection. *Genetics* 156:899–911
- Kowalski AM, Gooding M, Ferrante A, Slafer GA, Orford S, Gasperini D, Griffiths S (2016) Agronomic assessment of the wheat semi-dwarfing gene *Rht8* in contrasting nitrogen treatments and water regimes. *Field Crops Res* 191:150–160
- Krchov and Bernardo 2015
- Kuchel H, Williams KJ, Langridge P, Eagles HA, Jefferies SP (2007) Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor Appl Genet* 115:1029–1041
- Kumar LS (1999) DNA markers in plant improvement: an overview. *Biotechnol Adv* 17:143–182
- Laidig F, Piepho HP, Drobek T, Meyer U (2014) Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theor Appl Genet* 127:2599–2617
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756

- Lebreton CH, Visscher PM, Haley CS, Semikhodskii A, Quarrie SA (1998) A nonparametric bootstrap method for testing close linkage versus pleiotropy of coincident quantitative trait loci. *Genetics* 150:931–943
- Li F, Vallabhaneni R, Yu J, Rocheford T, Wurtzel ET (2008) The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress tolerance. *Plant Physiol* 147:1334–1346
- Liu YN, He ZH, Appels R, Xia XC (2012) Functional markers in wheat: current status and future prospects. *Theor Appl Genet* 125:1–10
- Longin CFH, Reif JC (2014) Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci* 19:631–636
- Longin CFH, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theor Appl Genet* 128:1297–1306
- 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
- Lyra DH, de Freitas Mendonça L, Galli G, Couto Alves F, Correia Granato IS, Fritsche-Neto R (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol Breed* 37:80. doi:10.1007/s11032-017-0681-1
- Malosetti M, Ribaut JM, Vargas M et al (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* 161:241–257
- Mangin B, Thoquet P, Grimsley N (1998) Pleiotropic QTL analysis. *Biometrics* 54:88–99
- Marchal A, Legarra A, Tisné S, Carasco-Lacombe C, Manez A, Suryana E, Omoré A, Nouy B, Durand-Gasselín T, Sánchez L, Bouvet J-M, Cros D (2016) Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq.) progeny tests. *Mol Breed* 36:2. doi:10.1007/s11032-015-0423-1
- Maslow AH (1943) A theory of human motivation. *Psychol Rev* 50:370–396
- McCartney CA, Somers DJ, Fedak G, DePauw RM, Thomas J, Fox SL, Humphreys DG, Lukow O, Savard ME, McCallum BD, Gilbert J, Cao W (2007) The evaluation of FHB resistance QTLs introgressed into elite Canadian spring wheat germplasm. *Mol Breed* 20:209–221
- McKevith B (2004) Nutritional aspects of cereals. *Nutr Bull* 29:111–142
- Melchinger AE, Schipprack W, Würschum T, Chen S, Technow F (2013) Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. *Sci Rep* 3:2129. doi:10.1038/srep02129
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Miedaner T, Hübner M, Korzun V, Schmiedchen B, Bauer E, Haseneyer G, Wilde P, Reif JC (2012) Genetic architecture of complex agronomic traits examined in two testcross populations of rye (*Secale cereale* L.). *BMC Genomics* 13:706. doi:10.1186/1471-2164-13-706
- Mirdita V, He S, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y (2015a) Potential and limits of whole genome prediction of resistance to *Fusarium* head blight

and *Septoria tritici* blotch in a vast Central European elite winter wheat population. *Theor Appl Genet* 128:2471–2481

- Mirdita V, Liu G, Zhao Y, Miedaner T, Longin CFH, Gowda M, Mette MF, Reif JC (2015b) Genetic architecture is more complex for resistance to *Septoria tritici* blotch than to *Fusarium* head blight in Central European winter wheat. *BMC Genomics* 16:430. doi:10.1186/s12864-015-1628-8
- Mode CJ, Robinson HF (1959) Pleiotropism and the genetic variance and covariance. *Biometrics* 15:518–537
- Montesinos-López OA, Montesinos-López A, Crossa J, Toledo FH, Pérez-Hernández O, Eskridge KM, Rutkoski J (2016) A genomic Bayesian multi-trait and multi-environment model. *G3* 6:2725-2744
- Munkvold JD, Laudencia-Chingcuanco D, Sorrells ME (2013) Systems genetics of environmental response in the mature wheat embryo. *Genetics* 194:265–277
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–2202
- Neumann K, Klukas C, Friedel S, Rischbeck P, Chen D, Entzian A, Stein N, Graner A, Kilian B (2015) Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis. *Plant Cell Environ* 38:1980-1996
- Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends Genet* 29:66–73
- Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 13:263–269
- Pavlicev M, Wagner GP (2012) A model of developmental evolution: selection, pleiotropy and compensation. *Trends Ecol Evol* 27:316–322
- Pešek J, Baker RJ (1969) Desired improvement in relation to selection indices. *Canadian J Plant Sci* 49:803–804
- Plate L (1910) Vererbungslehre und Deszendenztheorie (in German). *Festschr f Hertwig R II*. Fischer, Jena, pp 537-610
- Qiu Z, Tang Y, Ma C (2017) An effective strategy for trait combinations in multiple-trait genomic selection. In: Huang DS, Jo KH, Figueroa-García J (eds), *Intelligent Computing Theories and Application*. ICIC 2017. Lecture Notes in Computer Science, vol 10362. Springer, Cham, pp 230-239
- Ray DK, Ramankutty N, Mueller ND et al (2012) Recent patterns of crop yield growth and stagnation. *Nat Commun* 3:1293. doi:10.1038/ncomms2296
- Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor Appl Genet* 126:2835–2848
- Robertson DS, Anderson IC (1961) Temperature-sensitive alleles of the *y1* locus in maize. *J. Heredity* 52:53–60
- Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells ME (2012) Evaluation of genomic prediction methods for *Fusarium* head blight resistance in wheat. *Plant Genome* 5:51–61
- Rutkoski J, Poland J, Mondal S, Autrique E, González Pérez L, Crossa J, Reynolds M, Singh R (2016) Canopy temperature and vegetation indices from high-throughput

phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3* 6:2799-2808

- Saitone TL, Sexton RJ (2017) Agri-food supply chain: evolution and performance with conflicting consumer and societal demands. *Eur Rev Agric Econ*: 1–24
- Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the Near East. *Nat Rev Genet* 3:429–441
- Santos JPR, Vasconcellos RCC, Pires LPM, Balestre M, Von Pinho RG (2016) Inclusion of dominance effects in the multivariate GBLUP model. *PLoS One* 11:e0152045
- Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y (2016) Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor Appl Genet* 129:273–287
- Schulthess AW, Reif JC, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Ganal MW, Röder MS, Jiang Y (2017a) The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (*Triticum aestivum* L.). *J Exp Bot* 68:4089-4101
- Schulthess AW, Zhao Y, Longin CFH, Reif JC (2017b) Advantages and limitations of multiple-trait genomic prediction for *Fusarium* head blight severity in hybrid wheat (*Triticum aestivum* L.). *Theor Appl Genet*:1–17. doi:<https://doi.org/10.1007/s00122-017-3029-7>
- Schulthess AW, Zhao Y, Reif JC (2017c) Genomic selection in hybrid breeding. In: Varshney R, Roorkiwal M and Sorrells ME (eds) *Genomic Selection for Crop Improvement*. Springer, Cham
- Sissons M (2008) Role of durum wheat composition on the quality of pasta and bread. *Food*. 2:75–90
- Slafer GA, Calderini DF, Miralles DJ (1996) Yield components and compensation in wheat: Opportunities for further increasing yield potential. In: Reynolds MP, Rajaram S, McNab A (eds), *Increasing Yield Potential in Wheat - Breaking the Barriers*. CIMMYT, Mexico, pp 101–133
- Smith HF (1936) A discriminant function for plant selection. *Ann Eugen* 7:240–250
- Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 14:483–495
- Stearns FW (2010) One hundred years of pleiotropy: a retrospective. *Genetics* 186:767–773
- Stephens M (2013) A unified framework for association analysis with multiple related phenotypes. *PLoS One* 8:e65245
- Stich B, Piepho HP, Schulz B, Melchinger AE (2008) Multi-traits association mapping in sugar beet (*Beta vulgaris* L.). *Theor Appl Genet* 117:947–954
- Suwantaradon K, Eberhart SA, Mock JJ, Owens JC, Guthrie WD (1975) Index selection for several agronomic traits in the BSSS2 maize population. *Crop Sci* 15:827–833
- Tanksley S, Nelson J (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191–203

- Tavakol E, Okagaki R, Verderio G et al (2015) The barley *Uniculme4* gene encodes a BLADE-ON-PETIOLE-like protein that controls tillering and leaf patterning. *Plant Physiol* 168:164–174
- Thangasamy S, Chen PW, Lai MH, Chen J, Jauh GY (2012) Rice *LGDI* containing RNA binding activity affects growth and development through alternative promoters. *Plant J* 71:288–302
- Thoen MPM, Davila Olivas NH, Kloth KJ, Coolen S, Huang PP, Aarts MGM, Bac-Molenaar JA, Bakker J, Bouwmeester HJ, Broekgaarden C, Johan Bucher, Busscher-Lange J, Cheng X, Fradin EF, Jongsma MA, Julkowska MM, Keurentjes JJB, Ligterink W, Pieterse CMJ, Carolien Ruyter-Spira, Smant G, Testerink C, Usadel B, van Loon JJA, van Pelt JA, van Schaik CC, van Wees SCM, Visser RGF, Voorrips R, Vosman B, Vreugdenhil D, Warmerdam S, Wiegers GL, van Heerwaarden J, Kruijer W, van Eeuwijk FA, Dicke M (2017) Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *New Phytol* 213:1346-1362
- Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, Laitinen RAE, Huang Y, Chory J, Lipka V, Borevitz JO, Dangl JL, Bergelson J, Nordborg M, Weigel D (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465:632–636
- Varona L, Gomez-Raya L, Rauw WM, Clop A, Ovilo C, Noguera JL (2004) Derivation of a Bayes factor to distinguish between linked or pleiotropic quantitative trait loci. *Genetics* 166:1025–1035
- Vila-Aiub MM, Neve P, Powles SB (2009). Fitness costs associated with evolved herbicide resistance alleles in plants. *New Phytol* 184:751–767
- Voss-Fels KP, Qian L, Parra-Londono S, Uptmoor R, Frisch M, Keeble-Gagnère G, Appels R, Snowdon RJ (2017) Linkage drag constrains the roots of modern wheat. *Plant Cell Environ* 40:717–725
- Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. *Nat Rev Genet* 8:921–931
- Wang Z, Liao BY, Zhang J (2010) Genomic patterns of pleiotropy and the evolution of complexity. *Proc Natl Acad Sci USA* 107:18034–18039
- Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genomics* 15:556
- Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C, Hu Z (2016) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity* 118:302–310
- Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. *Genet Res* 75:249–252
- Williams JS (1962) The evaluation of a selection index. *Biometrics* 18:375–393
- Williams JT, Van Eerdewegh P, Almasy L, Blangero J (1999) Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. *Am J Hum Genet* 65:1134–1147

- Worland AJ (1996) The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica* 89:49–57
- Wright S (1968) *Evolution and the genetics of populations*. Vol. 1. Genetic and biometric foundations. University of Chicago Press, Chicago
- Wrigley C, Batey I, Miskelly D (2017) *Cereal grains: assessing and managing quality*, 2nd edn. Woodhead Publishing, Cambridge
- Xiong H, Guo H, Xie Y, Zhao L, Gu J, Zhao S, Li J, Liu L (2017) RNAseq analysis reveals pathways and candidate genes associated with salinity tolerance in a spaceflight-induced wheat mutant. *Sci Rep* 7:2731. doi: 10.1038/s41598-017-03024-0
- Yu Q, Collavo A, Zheng MQ, Owen M, Sattin M, Powles SB (2007) Diversity of acetyl-coenzyme A carboxylase mutations in resistant *Lolium* populations: evaluation using clethodim. *Plant Physiol* 145:547–558
- Zeder MA (2011) The origins of agriculture in the near east. *Curr Anthropol* 52:S221–S235
- Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods* 11:407–409
- Ziyomo C, Bernardo R (2013) Drought tolerance in maize: indirect selection through secondary traits versus genomewide selection. *Crop Sci* 53:1269–1275

11. List of general abbreviations

cDNA – complementary deoxyribonucleic acid

DNA – deoxyribonucleic acid

eQTL – expression quantitative trait loci

FAO – Food and Agriculture Organization of the United Nations

MAS – marker-assisted selection

QTL – quantitative trait loci

12. Curriculum vitae

(Last update: 14.01.2018)

Personal information

Surname, given name Schulthess Börgel, Albert Wilhelm
Birthdate February 14th, 1987
Place of birth Santiago de Chile, Chile
Nationality Chilean
Citizenship Chilean
Current home address Südstraße 8
06110 Halle (Saale)
Sachsen-Anhalt
Deutschland

Education

- 03.2010 – 03.2013 **Master degree in Plant sciences**
Graduated with maximum distinction and ranked 1st among 18 students graduated from the same master program in 2013.
Master's thesis title: "*Genotypic, environmental and agronomical factors that influence the semolina yellowness of durum wheat in Chile*"
Pontifical Catholic University of Chile
Santiago, Chile
- 03.2005 – 07.2010 **Professional title of Agricultural Engineer**
Graduated with two votes of distinction and ranked 4th among 209 students graduated from the same Agronomy program in 2010.
Pontifical Catholic University of Chile
Santiago, Chile
- 03.2005 – 12.2009 **Bachelor of science in Agriculture and Natural Resources**
Pontifical Catholic University of Chile
Santiago, Chile
- 03.2001 – 12.2004 **High school**
Deutsche Schule Sankt Thomas Morus, Santiago, Chile
- 03.1991 – 12.2000 **Primary school**
Liceo Alemán de Santiago, Santiago, Chile

Work experience

- 06.2014 – **Leibniz Institut of Plant Genetics and Crop Plant Research (IPK-Gatersleben), PhD student**
Research group: Quantitative Genetics (headed by Prof. Dr. Jochen C. Reif)
- 04.2013 – 12.2013 **Faculty of Agronomy and Forestry of the Pontifical Catholic University of Chile, Research assistant**
Working for the project "*Genetic improvement of the yellowness and the nutritional value of the grains of Durum Wheat (Triticum Turgidum L. Ssp. Durum)*" (grant Fondecyt de Iniciación N°11110066).
- 12.2008 – 02.2009 **SEM Chile, Summer professional practice**

Field inspections of flowering time and detasseling labor in hybrid maize seed production.
SEM's Curacaví and Polpaico field stations, Chile

Academic experience

- 03.2008 – 12.2013 **Teaching assistant**
Faculty of Agronomy and Forestry
Pontifical Catholic University of Chile
- Undergraduate level courses:
2011 – 2013 Field Crops Production
2012 Principles of Plant Breeding
2010 – 2011 Crop Production
2008 – 2011 Applied Statistics
- Postgraduate level courses:
2011 Biometrics and Experimental Design
- 08.2013 – 10.2013 **Private teacher**
Private lessons of applied statistics to PhD students.

Scholarships and awards

- 06.2014 – 05.2017 Forschungsstipendium für Doktoranden und Nachwuchswissenschaftler, Deutscher Akademischer Austausch Dienst (DAAD).
- 03.2010 – 03.2013 Master's Full Scholarship, granted by the Faculty of Agronomy and Forestry of the Pontifical Catholic University of Chile.
- 10.2010 Integral Agriculture Engineer, awarded by the Faculty of Agronomy and Forestry of the Pontifical Catholic University of Chile.
- 03.2009 – 12.2009 Academic Honor Scholarship ("Matrícula de Honor"), granted by the Pontifical Catholic University of Chile's rector.

Language certificates

- 03.2013 TOEFL iBT score: 98/120
- 11.2004 *Deutsches Sprachdiplom der Kultusministerkonferenz (DSD) Stufe II*

Skills related to scientific fields of interest

- Molecular biology Experience with DNA extraction and PCR optimization.
- Statistics Microsoft Excel: Advanced knowledge
R-Project: Advanced knowledge
Statistical Analysis System (SAS): Basic knowledge
- Bioinformatics At the user level, experience with ab-initio gene annotation, sequence alignments (complete and local) and three-dimensional protein modeling (mainly homology modeling).
- Plant breeding Experience with design and cultivation of field experiments in wheat. Experience with detasseling and selfing in maize and with manual emasculation and crossing in wheat.

Language skills (not certified)

Spanish	Native language
German	Fluent in spoken and written
English	Fluent in spoken and written

Other activities of interest

2001 –	Amateur musician (bass guitar). Backpacking, trekking and hiking.
--------	--

List of scientific publications

Research articles

- **Schulthess AW**, Zhao Y, Longin CFH, Reif JC (2017) Advantages and limitations of multiple-trait genomic prediction for *Fusarium* head blight severity in hybrid wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*:1–17. doi:<https://doi.org/10.1007/s00122-017-3029-7>
- **Schulthess AW**, Zhao Y, Reif JC (2017) Genomic selection in hybrid breeding. In: Varshney R, Roorkiwal M and Sorrells ME (eds) *Genomic Selection for Crop Improvement*. Springer, Cham.
- **Schulthess AW**, Reif JC, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Ganal MW, Röder MS, Jiang Y (2017) The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (*Triticum aestivum* L.). *Journal of Experimental Botany* 68:4089-4101.
- Jiang Y, **Schulthess AW**, Rodemann B, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Ganal M W, Röder MS, Reif JC. 2017. Validating the prediction accuracies of marker-assisted and genomic selection of *Fusarium* head blight resistance in wheat using an independent sample. *Theoretical and Applied Genetics* 130:471-482.
- Miedaner T, **Schulthess AW**, Gowda M, Reif JC, Longin CFH. 2017. High accuracy of predicting hybrid performance of *Fusarium* head blight resistance by mid-parent values in wheat. *Theoretical and Applied Genetics* 130:461-470.
- Vargas VH, **Schulthess AW**, Royo C, Matus I, Schwember AR. 2016. Transcripts levels of *Phytoene synthase 1 (Psy-1)* are associated to semolina yellowness variation in durum wheat (*Triticum turgidum* L. ssp. *durum*). *Journal of Cereal Science* 68:155-163.
- He S, **Schulthess AW**, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y. 2016. Genomic selection in a commercial winter wheat population. *Theoretical and Applied Genetics* 129:641-651.
- Campos KM, Royo C, **Schulthess AW**, Villegas D, Matus I, Ammar K, Schwember AR. 2016. Association of phytoene synthase *Psy1-A1* and *Psy1-B1* allelic variants with semolina yellowness in durum wheat (*Triticum turgidum* L. var. *durum*). *Euphytica* 207:109-117.
- **Schulthess AW**, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y. 2016. Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for

feeding purposes. Theoretical and Applied Genetics 129:273-287.

- **Schulthess AW**, Matus I, Schwember AR. 2013. Genotypic and environmental factors and their interactions determine semolina color of elite genotypes of durum wheat (*Triticum turgidum* L. var. *durum*) grown in different environments of Chile. Field Crops Research 149:234-244.

- **Schulthess AW**, Schwember AR. 2013. Improving durum wheat (*Triticum turgidum* L. var *durum*) grain yellow pigment content color through plant breeding. Ciencia e Investigación Agraria 40:475-490.

13. Eidesstattliche Erklärung / Declaration under Oath

Hiermit erkläre ich, Albert Wilhelm Schulthess Börgel, an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. / I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

Unterschrift, Ort, Datum / Signature, Place, Date

**14. Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren /
Declaration concerning Criminal Record and Pending Investigations**

Hiermit erkläre ich, Albert Wilhelm Schulthess Börgel, dass ich weder vorbestraft bin noch dass gegen mich Ermittlungsverfahren anhängig sind. / I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.

Unterschrift, Ort, Datum / Signature, Place, Date