MODELLING OF GENOTYPE BY ENVIRONMENT INTERACTION AND PREDICTION OF COMPLEX TRAITS ACROSS MULTIPLE ENVIRONMENTS AS A SYNTHESIS OF CROP GROWTH MODELLING, GENETICS AND STATISTICS

DANIELA BUSTOS-KORTS

Thesis committee

Promotor

Prof. Dr F.A. van Eeuwijk

Professor of Applied Statistics

Wageningen University & Research

Co-promotor

Dr M. Malosetti

Researcher, Applied Statistics

Wageningen University & Research

Other members

Prof. Dr L.F.M Marcelis, Wageningen University & Research

Dr C.D. Messina, DuPont Pioneer, Johnston, USA

Prof. Dr C.C. Schön, Technical University of Munich, Germany

Prof. Dr B.J. Zwaan, Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Production Ecology and Resource Conservation (PE & RC)

Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics

Daniela Bustos-Korts

Thesis

submitted in fulfilment of the requirement for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 15 November 2017
at 4 p.m. in the Aula.

Daniela Bustos-Korts

Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics, 340 pages.

PhD thesis, Wageningen University, Wageningen, The Netherlands (2017)

With references, with summary in English

ISBN: 978-94-6343-669-4

DOI: 10.18174/421321

Abstract

Bustos-Korts, D. (2017). Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics. PhD thesis, Wageningen University, the Netherlands.

The main objective of plant breeders is to create and identify genotypes that are well-adapted to the target population of environments (TPE). The TPE corresponds to the future growing conditions in which the varieties produced by a breeding program will be grown. All possible genotypes that could be considered as selection candidates for a specific TPE are said to belong to the target population of genotypes, TPG. Genotypes commonly show different sensitivities to environmental gradients and then genotype by environment interaction (GxE) is observed. GxE can lead to changes in genotypic ranking, complicating the breeding process. The main aim of this thesis was to investigate statistical models and the combination of statistical and crop growth models to improve phenotype prediction across multiple environments. One aspect that determines the quality of phenotype prediction is the set of genotypes used to train the prediction model, especially when the TPG is structured. We proposed a method that uniformly covers the genetic space of the TPG, leading to a larger prediction accuracy than random sampling. We produced positive results for wheat, maize and rice. A second aspect that influences the accuracy of phenotype predictions is the choice of environments used to train the prediction model, which should capture the heterogeneity in the TPE. When accounting for heterogeneity in environmental quality, it is important to distinguish between repeatable and well predictable elements in the environmental conditions from those that are badly predictable. We proposed statistical methods based on the AMMI model and on mixed models to identify groups of environments that show repeatable GxE, illustrating our ideas with multi-environment wheat data in North-Western Europe. The importance of training set construction strategies and multi-environment genomic prediction models was also demonstrated for barley data. If breeders are interested in identifying the genetic basis of the target traits, it is advantageous to have a higher SNP density. In this thesis, we used exome sequence data of the EU-Whealbi-barley germplasm, which corresponds to a unique set of genotypes with a diverse origin, growth habit and breeding history.

For this diverse data, we assessed the effects of QTLs and haplotypes across multiple environments for awn length, grain weight, heading date and plant height. Our results show that the EU-Whealbi-barley collection possesses a large diversity of promising alleles regulating the four traits we analysed. The last major topic addressed in this thesis is the use of a combination of statistical-genetic models and crop growth models (APSIM) as a strategy to assess the traits and phenotyping schemes to improve the prediction accuracy of a target trait like yield. We assess the potential of the combined modelling approach to characterize a sample of the TPG and TPE, and illustrate how trait correlations are modified by environmental conditions and by the genetic architecture of the sample of the TPE. We discuss the topics mentioned above, from a didactical perspective, proposing a list of subjects that should be covered in a GxE course for plant breeders. Finally, we discuss challenges and opportunities presented by the characterization of the TPE and TPG when using simulations based on statistical and crop growth models.

Contents

Chapter 1

General Introduction

1.1. Introduction

The main goal of plant breeders is to create and identify genotypes that are well-adapted to the future growing conditions defined by the meteorological, soil and management at the growing area of interest (Cooper & Hammer, 1996). These growing conditions will influence the phenotypic response of individual genotypes. The functional form by which environmental inputs are translated into phenotypes is sometimes referred to as the reaction norm (Woltereck, 1909; Dobzhansky & Spassky, 1963; Sarkar, 1999; DeWitt & Scheiner, 2004). Reaction norms depend both on environmental inputs and genetic sensitivities to these environmental inputs. When the reaction norms for different genotypes are not parallel, this indicates the existence of genotype by environment interaction (GEI) (Finlay & Wilkinson, 1963; van Eeuwijk et al., 2005). An extreme form of GEI is cross-over interaction, where the ranking of the genotypes varies with the environmental conditions (Baker, 1988; Crossa et al., 2004). Cross-over interactions complicate the breeding process, making it necessary to recommend specific genotypes for specific environments.

Usually, GxE patterns are analysed for a set of environments that correspond to the future growing conditions of the genotypes created by the breeding programme. This set of environments is also referred to as 'the target population of environments', TPE (Comstock & Moll, 1963; Cooper & Hammer, 1996; Cooper et al., 2014). The TPE will influence the traits and adaptation mechanisms that are necessary for a genotype to perform well, showing a high yield. The traits that are relevant for adaptation to particular growing conditions influence the decision process of breeders, modifying the selection strategy applied within breeding programme. The selection strategy, together with the crossing scheme, shape the target population of genotypes (TPG), which corresponds to all possible genotypes that the breeding programme for the TPE hopes to develop during the coming years. The TPG coincides with the notion of selection candidates (Jannink et al., 2010; Schulz-Streeck et al. 2012; Albrecht et al. 2014).

To select well-adapted genotypes, predictions need to be made for the phenotype as a function of genotype and environment. These predictions can be made with statistical-genetic models, crop growth models or with a combination of statistical-genetic and crop growth

models. Statistical models to characterize GxE range from the conventional ANOVA to factorial regression type of models and mixed models integrating explicit environmental and physiological information (van Eeuwijk et al., 2005; Malosetti et al., 2013). Commonly used fixed models to characterize GxE are the Additive Main effects and Multiplicative Interactions (AMMI) and the Genotype main effects and Genotype by Environment interaction effects model (GGE) models (Gauch, 1992, 2013; Yan et al., 2000; Yan & Hunt, 2001). In fixed model terms, inferences are made with respect to the specific levels of genotypes and environments, whereas in random model terms, there is interest in the underlying distribution of the population (Searle et al., 2009). Both AMMI and GGE models allow examining GxE by means of biplots that provide a visual representation of which genotypes and environments are driving the interaction, allowing to group environments. To structure the network of testing sites, it is important to distinguish between GxE variation due to the consistent differences between locations, from the random year-to year variations (Atlin et al., 2000, 2011; Piepho & Möhring, 2005). If the phenotypic responses at particular locations have a certain degree of repeatability across years, these locations may be classified into 'mega-environments', which correspond to sets of environments of similar quality that show a reduced number of cross-over interactions (Rajaram et al., 1993; Gauch & Zobel, 1997). Within a mega-environment, similar genotypes can be recommended, simplifying the selection and the recommendation processes. Fixed-effects models to characterize GxE are certainly a useful tool for breeders, provided the GxE concerns well defined genotypes under repeatable environmental conditions. However, part of the GxE is commonly due to random trial-to-trial variations and not necessarily to repeatable GxE that would justify splitting the TPE in a number of mega-environments (Atlin et al., 2011). Random year-to-year variations can be handled by an adequate representation and replication of the testing sites across the TPE and by the use of mixed models that consider part of the GxE as a random process. Popular mixed models to characterize GxE are the unstructured model and the factor analytic model (Burgueño et al., 2008; Beeck et al., 2010). When the TPE can be subdivided into homogeneous mega-environments that show a consistent genotypic ranking across years, these mega-environments can also be either integrated into the mixed model as a fixed effect or used to model the variance-covariance structure. Modelling of the mega-environments explicitly allows to increase the response to selection and to obtain genotypes that are better

adapted to the within-mega-environment growing conditions (Atlin et al., 2000, 2011; Piepho & Möhring, 2005).

The availability of molecular markers allows identifying genomic regions underlying the phenotypic differences via QTL models. If the expected genetic similarity is homogeneous across the whole population, as in doubled haploid populations (DHs) or recombinant inbred lines (RILs), the significance of QTL effects is tested using a simple residual structure (Rebai et al., 1995; Lynch & Walsh, 1998). However, a more elaborate structure for the residual genetic variance is needed to model for multi-parent populations or diversity panels (Malosetti et al., 2007, 2011; Korte & Farlow, 2013; Garin et al., 2017). If the interest is not in detecting genomic regions associated with the trait, but in making phenotype predictions, it is in general convenient to move from QTL models to genomic prediction models. In genomic prediction models, no explicit significance threshold is used and all markers are used (Meuwissen et al., 2001). Genomic prediction models range from those accounting only for additive effects, like GBLUP (VanRaden, 2008), up to more elaborate models that allow for nonlinearities associated with non-additive genetic effects. A popular model allowing for non-additive effects is the reproducible Kernel Hilbert Spaces model (RKHS) (Gianola & van Kaam, 2008; de los Campos et al., 2009; Jiang & Reif, 2015). An alternative to mixed models for genomic prediction are the Bayesian models. There is a large number of Bayesian models for genomic prediction and they basically differ in the degree of shrinkage imposed on the markers genome-wide (Meuwissen et al., 2001; Hayes et al., 2009; Perez et al., 2010; Habier et al., 2011; de los Campos et al., 2013).

For a more explicit representation of the genotypic response across environments, it might be convenient to use factorial regression type of models with QTLs modulating the genotypic sensitivity to the environment. Examples for factorial regression-type of models can be seen in (Malosetti et al., 2004; Boer et al., 2007). More elaborate and dynamic characterization of the genotypic response to the environmental conditions can be modelled via functional mapping, which consists of a combination of QTLs and mathematical functions that model the QTL effects over time (Malosetti et al., 2006; Wu & Lin, 2006; van Eeuwijk et al., 2010; Li & Sillanpää, 2015). An increasingly popular approach to explicitly model the functional relationships between plant physiology and the environment is by using crop growth models

(Yin et al., 2000, 2005; Tardieu et al., 2005). Crop growth models usually decompose the target trait (grain yield) into a number of underlying genetically-correlated traits, called 'intermediate traits' (Yin et al., 2004), 'indicator traits' (Calus & Veerkamp, 2011), 'secondary traits' (Rutkoski et al., 2016) or 'components' (Porter & Gawith, 1999) that might have a simpler genetic basis and larger heritability than the target trait (Yin et al., 2004; Tardieu & Tuberosa, 2010; Cabrera-Bosquet et al., 2016). Examples of intermediate traits are biomass, grain number and grain weight. Intermediate traits are calculated indirectly from a set of environmental inputs and genotype-dependent parameters derived from prior experimentation. GxE in the target trait is then a consequence of the interactions between the intermediate phenotypes (Chapman et al., 2003; Tardieu et al., 2005; Chenu et al., 2009; Makumburage et al., 2013). GxE modelling via crop growth models is commonly done using a two-step approach; first, the genotype dependent parameters are estimated from experimental data, via QTL or genomic prediction models. Then, these predicted parameter values are introduced into the crop growth model to generate predictions for the target trait across multiple environments (Yin et al., 2000; Zheng et al., 2013). An alternative-two stage approach had been recently proposed by (Technow et al., 2015; Cooper et al., 2016), where component traits are treated as latent variables for prediction that arise from the propagation of the marker effects on yield through the crop growth model structure.

1.2. Objectives and outline of the thesis

The general objective of this thesis is to propose and evaluate models to characterize GxE and predict complex traits across multiple environments. To achieve this goal, we used statistical models and the crop growth model APSIM-Wheat. The first specific objective was to set the scene about prediction scenarios that are interesting for breeders and to discuss how either statistical models, crop growth models or the combination of both types of models can be used to predict phenotypes across environments (**Chapter 2**). The scenarios that we discussed in **Chapter 2** correspond to the prediction of unobserved genotypes in observed environments, the prediction of observed genotypes in unobserved environments and finally, the prediction of unobserved genotypes in unobserved environments. The second specific objective of this thesis, assessed in **Chapter 3**, was to propose a strategy to construct the training set of genotypes, improving the prediction accuracy of unobserved genotypes in

observed environments. The main underlying hypothesis for our proposed training set construction method was that a homogeneous representation of the genetic diversity (genetic space) in the TPG leads to larger prediction accuracy. We compared our method based on uniform coverage of the genetic space, with commonly used methods like random sampling, stratified sampling and with a method that maximizes the generalized coefficient of determination, proposed by (Rincent et al., 2012).

When the goal is to make predictions in a multi-environment setting, the first task is to characterize the GxE patterns present in the TPE because this will influence the multi-environment prediction model that will have the largest accuracy. Therefore, the strategy to select candidates across a range of environments is highly dependent on the structure of GxE. This structure depends on the importance of year-to-year variability and on whether locations can be classified into more homogeneous groups, also called 'mega-environments'. In **Chapter 4**, we assessed strategies based on the additive main effects and multiplicative interactions model (AMMI, Gauch, 1992, 2013; Gauch & Zobel, 1997)) as applied to repeatable genotype by location interaction and on mixed models to identify regions that are internally more homogeneous. We presented examples for historical multi-environment trials for wheat in Denmark, Germany, The Netherlands and the United Kingdom. In **Chapter 5**, we illustrate the concepts discussed in Chapters 2, 3 and 4, applying them to a multi-environment genomic prediction context. Issues as the similarity between training and prediction environments and the design of training-validation schemes are also discussed and illustrated in Chapter 5. In **Chapter 6**, we apply multi-environment mixed models to characterize the EU-Whealbi germplasm collection as a source of valuable alleles for adaptation to European environments. The EU-Whealbi collection corresponds to 511 barley genotypes with a very wide range of origins and breeding history. EU-Whealbi was genotypically characterized with exome sequence data and phenotypically characterized in six very diverse European environments. Thanks to the genome sequence data, we could explore the effects of haplotypes instead of single SNPs, adding an extra layer the complexity to the statistical analysis.

In **Chapter 7**, we switch from the prediction of single traits observed at a single time point to the modelling of multiple traits simultaneously to improve the prediction accuracy

of the target trait. We also compare strategies to model phenotypes of multiple time-points simultaneously to characterize trait dynamics during the growing season. For intermediate traits to be useful to improve the prediction accuracy of the target trait, they have to be correlated to the target trait and have a large heritability. Trait correlations depend on the environment and heritability depends on the experimental and measurement quality. In Chapter 7, we propose the combination of statistical-genetic models and the APSIM crop growth model as a tool to assess the potential of traits and phenotyping strategies to improve the prediction accuracy of the target trait. Many of the concepts and statistical models discussed in Chapters 2 to 7 are presented in **Chapter 8** from an educational/didactical perspective. In this Chapter, we propose a schedule of topics that should be covered in a GxE course for plant breeders. We also provide an overview of the trends in the usage of different GxE models in the literature over time. Finally, in **Chapter 9**, we discuss the convenience of modelling approaches based on statistical models, crop growth models or a combination of statistical and crop growth modelling for different breeding situations. We also discuss how high throughput genotyping and high throughput phenotyping can be used to increase the chances of selecting better-adapted varieties and about technical considerations when combining statistical-genetic and crop growth models to assess breeding strategies.

Chapter 2

Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics

Daniela Bustos-Korts[1,2], Marcos Malosetti[1], Scott Chapman[3], Fred van Eeuwijk[1*]

1. Biometris, Wageningen University & Research Centre

2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)

3. CSIRO Plant Industry and Climate Adaptation Flagship

This Chapter is published as:

Bustos-Korts D, Malosetti M, Chapman S, van Eeuwijk FA (2016) Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics. In: Yin X, Struik PC (eds) Crop Systems Biology - Narrowing the Gaps between Crop Modelling and Genetics. Springer, pp 55–82

Abstract

Selection processes in plant breeding depend critically on the quality of phenotype predictions. The phenotype is classically predicted as a function of genotypic and environmental information. Models for phenotype prediction contain a mixture of statistical, genetic and physiological elements. In this chapter, we discuss prediction from linear mixed models (LMMs), with an emphasis on statistics, and prediction from crop growth models (CGMs), with an emphasis on physiology. Three modalities of prediction are distinguished: predictions for new genotypes under known environmental conditions, predictions for known genotypes under new environmental conditions, and predictions for new genotypes under new environmental conditions.

For LMMs, the genotypic input information includes molecular marker variation, while the environmental input can consist of meteorological, soil and management variables. However, integrated types of environmental characterizations obtained from crop growth models (CGMs) can also serve as environmental covariable in LMMs. LMMs consist of a fixed part, corresponding to the mean for a particular genotype in a particular environment and a random part defined by genotypic and environmental variances and correlations. For prediction via the fixed part, genotypic and/or environmental covariables are required as in classical regression. For predictions via the random part, correlations need to be estimated between observed and new genotypes, between observed and new environments, or both. These correlations can be based on similarities calculated from genotypic and environmental covariables. A simple type of covariable assigns genotypes to sub-populations and environments to regions. Such groupings can improve phenotype prediction.

For a second type of phenotype prediction, we consider crop growth models. CGMs predict a target phenotype as a non-linear function of underlying intermediate phenotypes. The intermediate phenotypes are outcomes of functions defined on genotype dependent CGM parameters and classical environmental descriptors. While the intermediate phenotypes may still show some genotype by environment interaction, the genotype dependent CGM parameters should be consistent across environmental conditions. The CGM parameters are regressed on molecular marker information to allow phenotype prediction from molecular marker information and standard physiologically relevant environmental information.

Both LMMs and CGMs require extensive characterization of genotypes and environments. High-throughput technologies for genotyping and phenotyping provide new opportunities for upscaling phenotype prediction and increasing the response to selection in the breeding process.

2.1. Introduction

The target production area for most arable crops spans a range of environmental conditions. In the absence of diseases and pests (not considered in this review), local environmental conditions are a function of meteorological and soil variables on the one hand and management interventions on the other hand. These conditions will influence the phenotypic response of individual genotypes, and to some extent genotypes will create their 'own' environment, e.g. depending on how they use soil water across the season. The functional form by which environmental inputs are translated into phenotypes is sometimes referred to as the reaction norm (Woltereck 1909; Dobzhansky and Spassky 1963; Sarkar 1999; DeWitt and Scheiner 2004). Reaction norms depend both on environmental inputs and genetic factors. For a given (multi-locus) genotype, the reaction norm is the functional relationship between the phenotype and an environmental gradient, and is often linearised in some way. Modelling of the reaction norms for a set of genotypes is a central objective in many breeding and genetic studies. The prediction of the phenotypic response as a function of genetic and environmental factors is the basis for decisions that involve selection of superior genotypes for a defined environmental range (Hammer et al. 2006; Chenu et al. 2011; Sadras et al. 2013).

Several important concepts in breeding and genetics have been defined in relation to the behaviour of reaction norms for a population of genotypes. Firstly, when the reaction norms are non-constant, genotypes are said to show 'plasticity' (Bradshaw et al. 1965; DeWitt and Scheiner 2004; Sadras and Lawson 2011). Secondly, when the reaction norms for different genotypes are not parallel, this indicates the existence of genotype by environment interaction (GEI) (Finlay and Wilkinson 1963; van Eeuwijk et al. 2005). An extreme form of GEI is cross-over interaction, where the ranking of the genotypes varies with the environmental conditions (Baker 1988; Muir et al. 1992; Crossa et al. 2004). Another important concept in the context of the comparison of reaction norms is adaptation (Wright 1931, 1932; Finlay and Wilkinson 1963; Romagosa and Fox 1993; Cooper and Hammer 1996; Cooper 1999; Romagosa et al. 2013), i.e., some genotypes do better than other ones in a defined set of environmental conditions, the reaction norms of the adapted genotypes are then always above those of the less adapted. Finally, for a given genotype, 'stability' measures quantify the variation around the reaction norm (Lin and Binns 1988; Piepho 1998). So, while plasticity, GEI and adaptation refer to the expected response curve, which may be most simply thought of as the expectation in a linear regression model, stability refers to the variation around this expected response at a defined set of environmental conditions (Slafer and Kernich 1996; DeWitt and Scheiner 2004; van Eeuwijk et al. 2005; van Eeuwijk et al. 2010).

To select genotypes with superior average performance or a given degree of adaptation, predictions need to be made for the phenotype as a function of genotype and environment. These types of predictions occur at various stages in a breeding programme. In the early stages of breeding programmes, seed is limiting and large numbers of new genotypes produced as offspring from crosses between well-chosen parents are evaluated in one or a few trials, normally in small plots. For the earliest stages of a breeding programme, modelling of reaction norms is not possible and selection takes place on the mean performance. At intermediate stages, offspring populations are tested in a limited number of trials at various locations for one or a few years. In those cases when seed is still limiting, it is attractive to use partially replicated designs (Cullis et al. 2006; Smith et al. 2006) so that genotypes can be tested at a larger sample of environmental conditions. Selection can be done on the mean across trials, but there are also possibilities to select for adaptation. At the later stages, when there is sufficient seed for individual genotypes, a limited number of genotypes can be tested in a large number of trials, with again possibilities for selection on wide adaptation to a wide set of environments or narrow adaptation to a limited set of environments (Cooper et al. 2014). Simultaneously, at this stage selection on stability is possible.

When a population of genotypes is evaluated in multiple trials, reaction norms can be fitted to help in describing the observed data efficiently and to allow some form of selection on properties of the reaction norm. To evaluate the predictive quality of reaction norm models, special cross validation (CV) schemes have been proposed. In CV schemes, the data are subdivided in a training set, used to estimate model parameters, and a test set, used to assess the correlation between predicted values and observed values. Such a correlation is termed prediction ‘accuracy’ (Meuwissen et al. 2001). For multiple environment data, various CV strategies have been proposed (Crossa et al. 2010; Burgueño et al. 2012; Heslot et al. 2012; Zhao et al. 2012; Heslot et al. 2013; Crossa et al. 2014). At this point, it is useful to clarify some nomenclature. When genotypes were tested, evaluated or observed in at least one environment, we indicate this by the letter G. When this was not the case we use nG. For environments the same rule can be defined: E for observed environments, with at least one observed genotype, and nE for environments without observations. Specific combinations of genotype and environment can have been observed, GE, or not, nGE. Following this terminology, the set [G, E, GE] would indicate a genotype that was observed and an environment that was observed, while also the specific combination of genotype and environment was observed. The combination [G, E, nGE] indicates a genotype and environment that have been observed, but the specific combination of genotype and environment was not observed. This latter situation is typical for unbalanced genotype by environment data.

Figure 1 shows four scenarios that are relevant to prediction of phenotypes from genotypes and environments as well as to the calculation of accuracies and CV strategies. Scheme 1 pertains to situations in which both genotypes and environments were observed. Specific combinations of genotypes and environments may be present, [G, E, GE] or absent [G, E, nGE]. Phenotype predictions for Scheme 1 can be made by simple additive models. The Schemes 2, 3 and 4 are more interesting and we will concentrate on those. Potential strategies for assessment of accuracy in genomic prediction are predictions for new genotypes in observed environments [nG, E, nGE] (Scheme 2, Fig. 1); predictions for observed genotypes in new environments [G, nE, nGE] (Scheme 3, Fig. 1); and predictions for new genotypes in new environments [nG, nE, nGE] (Scheme 4, Fig. 1) (Utz et al. 2000; Calus and Veerkamp 2011; Burgueño et al. 2012; Schulz-Streeck et al. 2012; Guo et al. 2013; Crossa et al. 2014). Scheme 4 of CV obviously represents the strictest type of accuracy assessment. (For the notation, whenever nG or nE appears, necessarily nGE needs to appear as well, so for Schemes 2, 3 and 4, we can omit the specification nGE.)

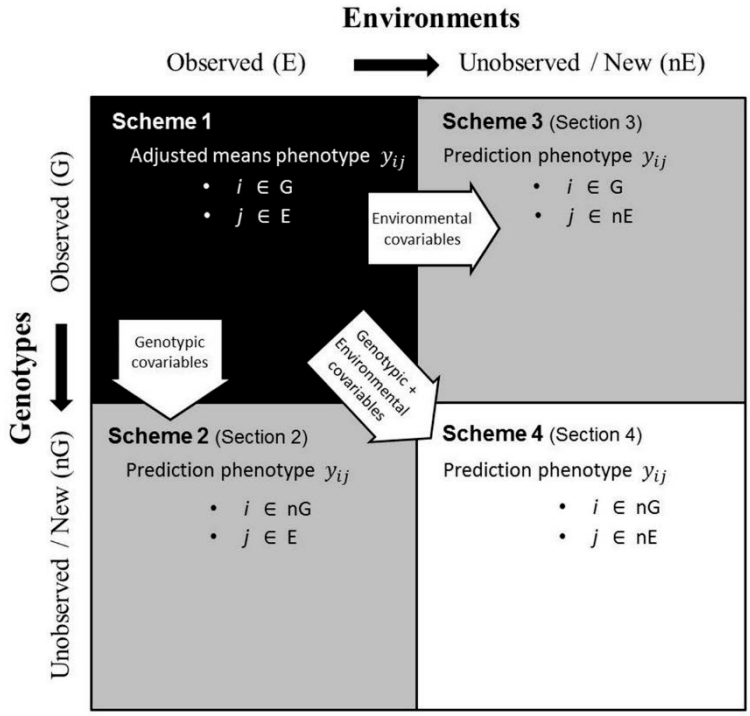


Figure 1. Prediction scenarios, depending on whether genotypes were observed (G) or not observed (nG), and on whether environments were observed (E) or not observed (nE).

To produce phenotype predictions for new genotypes (nG) from observed genotypes (G), it is essential to use statistical models that allow us to connect the new genotypes to the observed genotypes. The connections between nG and G can be achieved by the inclusion of explicit genotypic covariables in the statistical model, and/or by borrowing information via the correlation structure among genotypes, defined by their genetic similarities. Analogously, for predicting new environments, there needs to be a connection between nE and E via explicit environmental covariables and/or the correlation structure among environments. The latter correlation structure is an expression of environmental similarity as estimated from environmental characterizations.

In this chapter, we introduce linear mixed models (LMMs) as our default class of statistical prediction models. LMMs can be described as consisting of two parts: 1) a fixed part, corresponding to the mean; and 2) a random part defined by variances and covariances. Predictions in LMMs can be obtained via the fixed and the random part, although the statistical mechanism for prediction in those two cases is different. As an illustration, we provide an LMM for the phenotype of genotype *i* in environment *j*: $\underline{y_{ij}} = \mu_j + x_i\alpha_j + \beta_i z_j + \underline{GE_{ij}} + \underline{e_{ij}}$ (van Eeuwijk et al. 2010). The fixed part of this model is given by the expectation, or mean, for genotype *i* in environment *j*: $\mu_{ij} = \mu_j + x_i\alpha_j + \beta_i z_j$. Here μ_j is a fixed intercept (mean) for environment *j*, x_i is a genotypic covariable, for example a molecular marker, α_j is an environment specific slope corresponding to x_i. When x_i is a molecular marker, α_j is an environment specific quantitative trait locus (QTL) effect (Malosetti et al. 2004; Boer et al. 2007). For the environments, z_j is an environmental covariable, for example, a drought stress index, and β_i is a corresponding genotype specific slope, for example a genotype-specific sensitivity to drought stress.

For prediction via the fixed part, we use genotypic and/or environmental covariables as in classical regression (van Eeuwijk et al. 1996). Besides values for the covariable, x_i and z_j, prediction requires that we have estimates for the slopes, α_j and β_i. These can be obtained by fitting a model for the mean to training data, where we need to select suitable genotypic and/or environmental covariables. For prediction, we combine the estimated slopes in the training set with the values for genotypic and/or environmental covariables in the test set.

The random part of the model is determined by the terms $\underline{GE_{ij}}$ and $\underline{e_{ij}}$, the first term representing the (residual) genotypic effect of genotype *i* in environment *j*, the second term containing experimental (block) and measurement errors. (Random terms in model formulations are underlined.) The random terms are assumed to have a Gaussian distribution, with expectation zero and proper variance-covariance structures. The important random term for prediction purposes is $\underline{GE_{ij}}$. For this term, the correlations among genotypes on the one

hand and the correlations among environments on the other hand determine the predictive properties of the LMM. Thus, for predictions via the random part of the LMM, correlations need to be estimated between observed and new genotypes (Scheme 2), observed and new environments (Scheme 3), or both (Scheme 4). Correlations among genotypes can be estimated from genotypic covariables, including molecular markers, and pedigree data, or a combination of genotypic covariables and pedigree. Correlations among environments follow from environmental covariables. Although important, we will largely ignore the error term $\underline{e}_{ij}$ in the remainder of this chapter. See Smith et al. (2001a) and Smith et al. (2005) for discussion on models for $\underline{e}_{ij}$.

The realization of the predictive potential of LMMs depends on the selection of genotypic covariables and environmental covariables, for the fixed part as well as for the random part. Physiological knowledge on genotypes and environments can help in the choice of covariables for inclusion in LMMs. For example, knowledge on the structure and use of crop growth models (CGMs) can help in the dissection of complex traits (Chapman et al. 2002b; Edmeades et al. 2004; Reynolds et al. 2009a), thereby suggesting genotypic and environmental covariables for inclusion in predictive LMMs. A CGM can suggest writing a complex target trait as a function of a set of simpler component traits and a set of environmental input variables (Yin et al. 2003, 2004; Chenu et al. 2008; Hammer et al. 2010). These component traits are traditionally related to physiologically parameters in CGMs. The CGM produces GEI as an emerging property of the interaction between the physiological parameters and the environmental information (Chapman et al. 2002a, 2008; Hammer et al. 2002, 2006, 2010). Interpreting the CGM as a function that transforms physiological parameters and environmental inputs into a complex trait, we can understand that when the CGM can be approximated by a linear function, the component traits may be entered as genotypic covariables and the environmental inputs as environmental covariables in an LMM for the complex trait.

In Section 2, we will discuss how statistical LMM models can be used to predict phenotypic responses for new genotypes in observed environments (Scheme 2; [nG, E, nGE]), observed genotypes in unobserved (new) environments (Scheme 3; [G, nE, nGE]), or new genotypes in new environments (Scheme 4; [nG, nE, nGE]). In Section 3, we will discuss the use of CGMs to predict the performance of genotypes for environments in which they were not tested. Section 4 will discuss the contribution of high throughput genotyping and phenotyping to models for phenotype prediction. Strategies to group genotypes and environments will also be discussed in this section. We finish with some concluding remarks in Section 5.

2.2. Statistical models to predict phenotypic performance

Section 2.1 presents statistical models for predicting the phenotype of genotypes that were so far not tested in the environments for which we want to predict, although we do have information about these environments from phenotypic evaluations for other genotypes [nG, E, nGE], Scheme 2 in Fig. 1. The connection between observed genotypes (G) and not observed genotypes (nG) will come from explicit genotypic covariables and/or the genetic correlations among genotypes. Section 2.2 describes statistical models for predicting phenotypes in environments that were not used to test genotypes, although we do have phenotypic information about these genotypes in other environments [G, nE, nGE], Scheme 3 in Fig. 1. The connection between observed environments (E) and unobserved environment (nE), will result from the inclusion of explicit environmental covariables and/or the correlations among environments calculated on the basis of environmental characterizations. Section 2.3 discusses the most challenging prediction scenarios; predicting the phenotype of genotypes that were not tested so far, for environments that neither were tested [nG, nE, nGE], Scheme 4 in Fig. 1. Here, both explicit genotypic and environmental covariables are required for prediction.

2.2.1. Statistical models to predict performance of unobserved genotypes in observed environments [nG, E, nGE]

Quantitative traits are determined by many loci, with allelic effects varying in magnitude. Specific genomic regions significantly associated with phenotypic variation may be identified as quantitative trait loci, or QTLs (see Chapter 1 of this book by Baldazzi et al.). Besides QTLs, or instead thereof, many other loci with small additive effects (polygenic effects) can contribute to phenotypic variation. None of these loci with small effects might by itself have an important phenotypic effect, but these loci together can still make a sizeable contribution to the phenotype. Our first model, Model 1, includes loci with relatively large quantitative effects (QTLs) together with loci that have small effects.

$$\underline{y}_{ij}^{t} = \mu_j + \sum_{q=1}^{Q} x_{iq}\alpha_{jq} + \underline{G}_{ij} + \underline{e}_{ij} \qquad (1)$$

In the multi-environment Model 1, y_{ij}^{t} represents the target trait, *t*, (for example, yield) of genotype *i* in environment *j*, μ_j is a fixed intercept term for each environment, x_{iq} is a genotypic covariable that represents DNA information of genotype *i* at QTL position *q*, and α_{jq} is the additive effect of the fixed QTL q in environment *j*. $\underline{G}_{ij}$ represents the residual genetic effect (polygenic effects) for genotype *i* in environment *j*. The matrix with elements $\underline{G}_{ij}$, $\{\underline{G}_{ij}\}$, has a multivariate normal distribution with zero mean, $\mathbf{0}$, and, as we will see later,

a highly structured variance-covariance matrix $\boldsymbol{\Sigma}$; $\{\underline{G}_{ij}\}\sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$. (For notational simplicity, we will omit the dimensions of the various matrices.) $\boldsymbol{\Sigma}$ defines the genetic variances and covariance for any two pair of observations, $\underline{y}_{ij}^{t}$ and $\underline{y}_{i'j'}^{t}$ and depends on the genetic and environmental similarities of the two genotypes, *i* and *i'*, and the two environments *j* and *j'*. The term $\underline{e}_{ij}$ stands for a non-genetic residual, $\{\underline{e}_{ij}\}\sim MVN(\mathbf{0}, \boldsymbol{R})$, with $\boldsymbol{R}$ often allowing for specific residual variances per environment.

A simplification of Model 1 omits the genetic residual, $\underline{G}_{ij}$, and is appropriate when QTLs account for all of the genetic variation:

$$\underline{y}_{ij}^{t} = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_{jq} + \underline{e}_{ij} \qquad (1.1)$$

When Model 1.1 fits the data well, the performance of the unobserved genotype *i* in environment *j* can be predicted as;

$$\hat{\underline{y}}_{ij}^{t} = \hat{\underline{\mu}}_{j} + \sum_{q=1}^{Q} x_{iq}\hat{\underline{\alpha}}_{jq}$$

Compared with single-environment QTL models, multi-environment QTL models like Model 1 or Model 1.1 are more powerful in picking up QTLs and generally explain a larger amount of the genetic variance (Piepho 2000; Piepho and Möhring 2005; Mathews et al. 2008; Alimi et al. 2013). It has been shown that jointly considering multivariate phenotypes (i.e., the phenotype in multiple environments) allows a substantially greater separation between genotype classes than when considering univariate phenotypes (i.e., phenotype in a single environment) (Stephens 2013).

Another simplification of Model 1 occurs when we assume that there are no large discrete genetic effects in the form of QTLs that drive phenotypic differences, but that genetic effects are exclusively of a polygenic nature. A prediction model that generalizes the single environment genomic best linear unbiased prediction (G-BLUP) approach of (Meuwissen et al. 2001) to multi-environment prediction can be defined as:

$$\underline{y}_{ij}^{t} = \mu_{j} + \underline{G}_{ij} + \underline{e}_{ij} \qquad (1.2)$$

In Model 1.2, the distribution of the polygenic effects $\underline{G}_{ij}$ is $\{\underline{G}_{ij}\}\sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$. Since $\boldsymbol{\Sigma}$ is a function of the genetic and environment similarities, the larger the similarity of unobserved genotypes with observed genotypes, and the larger the similarity of observed environments with unobserved environments, the more information is available for phenotype prediction,

and the higher is the prediction accuracy (Crossa et al. 2006; Albrecht et al. 2014). Analogous to the classical partitioning of genetic and environmental effects, the covariance matrix $\boldsymbol{\Sigma}$ can be partitioned into a 'genotypic' variance-covariance matrix ($\boldsymbol{\Sigma}^G$), and an 'environmental' variance-covariance matrix ($\boldsymbol{\Sigma}^E$), such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^G \otimes \boldsymbol{\Sigma}^E$, i.e., the Kronecker product of the genotypic variance-covariance matrix and the environmental variance-covariance matrix (West et al. 2006; Smith et al. 2005). It is important to realize that although $\boldsymbol{\Sigma}^E$ is called an 'environmental' variance-covariance matrix, $\boldsymbol{\Sigma}^E$ reflects genetic correlations among environments, and so plays a role in forming predictions in the multi-environment context. Examples of commonly used models for these two covariance matrices are given below.

$\boldsymbol{\Sigma}^G$ can be modelled as $\boldsymbol{\Sigma}^G = \boldsymbol{A}$, where $\boldsymbol{A}$ corresponds to the expected additive relationship matrix calculated from the coefficients of coancestry estimated from the pedigree, or to the realized additive relationship matrix estimated from molecular markers (Piepho et al. 2008). If the one step prediction with statistical models uses pedigree information, $\underline{G}_{ij}$ is commonly called "breeding value" (Falconer and Mackay 1996; Piepho et al. 2008). On the other hand, if the prediction uses molecular marker information, it is called "genomic estimated breeding value" (Burgueño et al. 2012; Piepho 2009).

In the multi-environment context, genotypic variances tend to change across environments with consequent changes in genotypic correlations for pairs of these environments. A flexible variance-covariance structure across environments $\boldsymbol{\Sigma}^E$, is required to achieve higher prediction accuracies. One flexible and parsimonious model for variances and covariances/correlations across environments is the factor analytic model (FA) (Table 1) (Smith et al. 2001a, 2005; Mathews et al. 2008).

The decision about when it is convenient to use Models 1, 1.1, or 1.2 depends on the genetic architecture of the target trait. If the trait is regulated by a few QTLs with large effects, a QTL model (Model 1.1) might provide the largest prediction accuracy. On the other hand, traits like grain yield, which are regulated by many genes with small effects might not show any significant QTL that can be included in Model 1.1. In this case, Model 1.2, whose predictions we will call GE-BLUPs because they can account for GEI, should integrate the large number of small additive effects into a multi-environment prediction model. For the intermediate case when traits have a few QTLs with large effects, and many other loci with very small additive effects, Model 1 is adequate. Bernardo (2014) suggested that it is convenient to consider QTLs (or genes) as fixed effects when they account for more than 10% of the genetic variance. The simulations made by Bernardo (2014) show that the most adequate model depends on the genetic architecture of the trait, i.e., on the number of QTLs and the magnitudes of the QTL effects.

Table 1. Variance-covariance models for the environmental covariance (Σ^E), ordered by increasing number of parameters. For simplicity, these examples assume three environments (m=3).

Name	**Number of parameters**	**Structure**
Identity	1	$\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$
Compound symmetry	2	$\begin{bmatrix} \sigma^2+\varphi & \varphi & \varphi \\ \varphi & \sigma^2+\varphi & \varphi \\ \varphi & \varphi & \sigma^2+\varphi \end{bmatrix}$
Factor analytic, order 1	2m	$\begin{bmatrix} \lambda_1^2+\psi_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2+\psi_2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+\psi_3 \end{bmatrix}$
Unstructured	m(m+1)/2	$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$

2.2.2. Statistical models to predict performance of observed genotypes in new environments

After genotypes have been phenotyped in some environments, it can be useful to predict their performance in other environments that were not used for evaluation. New environments could, for example, include future trials at known locations, which implies that none of the genotypes were observed in that environment yet [G, nE, nGE]. Thus, the correlation between observed environments and the predicted environments cannot be estimated from phenotypic data, or direct observations on the complex trait. In this case, we may use environmental covariables, like meteorological, soil or management covariables, as predictors in models for the mean or define correlations between environments in models for the variance-covariance structure.

Models for the mean that can be used to predict phenotypes in unobserved environments usually correspond to factorial regression models that incorporate environmental covariables.

These models explicitly estimate the sensitivity of the QTL to environmental covariables (Model 2) (Campbell et al. 2004; Boer et al. 2007; Laperche et al. 2007; Malosetti et al. 2013; Romagosa et al. 2013). Hence, model parameters can have biological interpretation.

$$\underline{y}_{ij} = \mu_j + \sum_{q=1}^{Q} x_{iq} \left(\gamma_q + \delta_q z_j\right) + \underline{G}_{ij} + \underline{e}_{ij} \qquad (2)$$

In Model 2, the additive effects (α_{jq}) of the fixed QTL q in environment j of Model 1 are replaced by a regression formulation, $\left(\gamma_q + \delta_q z_j\right)$, in which the effect of QTL q is a function of the environmental covariable z_j, and so changes over environments. When the covariable z_j is centered, the intercept term, γ_q, corresponds to the effect of the QTL in the average environment, while the slope δ_q corresponds to the sensitivity of the QTL q to the environmental covariable z_j. Although Model 2 does not explicitly restrict the environmental covariables to a particular range, it should be considered that crops respond differently to covariables in the environmental extremes (e.g., too cold or too warm). So, the sensitivity of the genotype to the environmental covariables cannot be assumed constant outside the range of environments in which δ_q was estimated. A second issue that needs to be taken into account is that models like Model 2 do not make explicit in which phenological stage the environmental covariable is considered. Since the sensitivity of a crop to the environment varies throughout the development, environmental covariables included in the prediction model need to coincide with the developmental timing used to estimate the sensitivity.

For example, Boer et al. (2007) analysed grain yield and grain moisture for F5 maize testcross progenies evaluated across 12 environments in the U.S. corn belt. Since QTLs did not have a constant effect across environments (QTL by environment interaction), QTL effects were modelled conditional on longitude and year, both consequences of temperature differences during critical stages of the development. This factorial regression model allows prediction of yield and moisture at any location provided that temperatures during specific developmental stages are contained within those of the observed environments.

A second example is shown by Malosetti et al. (2004), who identified QTLs conferring differential sensitivity of grain yield to temperature during heading in a double haploid barley population. In a model like Model 2, the average daily temperature range during heading was the most important environmental covariable explaining differential QTL expression, i.e., the QTL allele from the parental line Steptoe, conferred an extra grain yield of 0.112 ton ha^{-1} for each extra degree Celsius during heading. Hence, yield could be predicted for unobserved environments if the average temperature for such environments was available. In that sense, Model 2 is closer to CGM than Model 1 because Model 2 explicitly represents environments on a continuous scale.

The second way to use environmental information for prediction is using environmental covariables to estimate similarities (covariances) among environments, analogous to the way molecular markers are used to characterize similarity among genotypes. If environmental covariables are considered as an indicator of environmental similarity, they can be used to estimate the environmental variance-covariance matrix in Model 1.2. Hence, $\boldsymbol{\Sigma}^E = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the variance-covariance matrix that accounts for the similarity in environmental conditions. The larger the covariance between observed and unobserved environments, the more information can be shared to make the predictions. The genotypic covariance $\boldsymbol{\Sigma}^G$ can be modelled as explained in Section 2.1 by imposing an additive relationship matrix to define $\boldsymbol{\Sigma}^G = \boldsymbol{A}$, where $\boldsymbol{A}$ can be estimated from the pedigree and/or from molecular markers.

Using multiple climatic variables to model the environmental covariance, as proposed by Jarquín et al. (2013) shows promise as a tool to predict genotypic performance in unobserved environments. However, many environmental covariables are correlated and not all need to be included in the model. Mechanistic CGMs such as APSIM have shown to be a good integrative tool to select subsets of variables that characterize environmental similarity (Chapman 2008).

2.2.3. Statistical models to predict performance of unobserved genotypes in new environments

Section 2.1 presented models that used genotypic covariables to predict the phenotype on unobserved genotypes. Section 2.2 described how environmental covariables can be used in factorial regression models for prediction, and how to estimate the environmental covariance of a random term, necessary for prediction along the random part of an LMM. This Section 2.3 will combine both situations, aiming to predict the phenotype of genotypes that have not been tested yet for environments that have not been used for evaluation.

When predicting unobserved genotypes in new environments, both genotypic and environmental covariables are needed. In factorial regression-type of models, prediction of unobserved genotypes is possible, provided that the additive effects of each QTL-allele can be estimated from the tested genotypes. The phenotypes of unobserved genotypes can also be predicted in new environments, provided that the sensitivity of the QTL effects along an environmental gradient (e.g., temperature), can be estimated from observed environments. In the example of Malosetti et al. (2004) presented in Section 2.2, phenotype prediction is possible for any environment provided the temperature remains within the range used to estimate the QTL sensitivity to temperature.

In models that entirely rely on the use of the variance-covariance structures imposed on genotypes and environments, prediction of unobserved genotypes in new environments is possible via the reconstruction of the full covariance matrix $\boldsymbol{\Sigma}$ from its components, $\boldsymbol{\Sigma}^G$ and $\boldsymbol{\Sigma}^E$. For the genotypic part this runs via explicit pedigree information or information from genotypic covariables (molecular markers), while for the environmental part correlations between environments can be estimated from environmental characterization (temperature, precipitation, soil characteristics, etc.). Note that while in Section 2.1, $\boldsymbol{\Sigma}^G$ was calculated from genotypic covariables, and $\boldsymbol{\Sigma}^E$ was estimated from the phenotypic data on the target trait, here both $\boldsymbol{\Sigma}^G$ and $\boldsymbol{\Sigma}^E$ are estimated from explicit covariables.

2.3. Crop growth models to predict genotypic performance

The algorithms in a CGM predict the target trait (e.g., grain yield) as a non-linear combination of underlying intermediate phenotypes (also commonly called "components", e.g., biomass), which are calculated indirectly from a set of inputs to the CGM that typically comprise environment (soil, weather, and nutrients) data and CGM parameters derived from prior experimentation. GEI in the target trait is then a consequence of the interactions between the intermediate phenotypes (Chapman et al. 2003; Tardieu 2003; Tardieu et al. 2005; Chenu et al. 2009; Makumburage et al. 2013).

Considering the CGM in reverse, we can state that the value of the target trait is able to be 'dissected' into these intermediate phenotypes (See Chapter 7 of this book by Hammer et al.). Although these intermediate phenotypes are likely to show less GEI than the target trait, they still correspond to an integration of genotypic responses to environmental conditions (e.g., they may show GEI). Ideally, a complete dissection of the target trait would comprise of a set of CGM input parameters that depend only on the genotype (for example, a genotypic sensitivity of development rate to the air temperature), and to environmental covariables (Model 2), i.e., CGM parameters that do not show GEI (Slafer 2003; Yin et al. 2003; Bertin et al. 2010; Alam et al. 2014). The target trait for genotype *i* in environment *j* can be written as a function of CGM parameters and environmental inputs as follows:

$$\underline{y}_{ij}^{t} = \int f\left(\underline{\boldsymbol{y}}_{i}^{P};\, \boldsymbol{z}_{j}\right) dt + \underline{e}_{i} \qquad (3)$$

In Model 3, $\underline{y}_{ij}^{t}$ represents the target trait for genotype *i* in environment *j*, which is modelled as a function of multiple CGM parameters $\underline{\boldsymbol{y}}_{i}^{P}$ (*p* for parameter in superscript) and multiple environmental inputs, $\boldsymbol{z}_j$ integrated over time (Fig. 2). The function $f(\,;\,)$ embodies the algorithms that transform CGM parameters into intermediate phenotypes as well as the interactions between the intermediate phenotypes that lead to the target trait.

A commonly-studied CGM is APSIM, which currently has modules for several crops, e.g., wheat, canola, sorghum (Keating et al. 2003; Holzworth et al. 2014). In the case of APSIM-Wheat, growth (biomass accumulation) and development (phenological events, the functionality of plant structures or appearance of new structures) are calculated on a daily basis (Wang et al. 2002). The final phenotype (e.g., grain yield) is calculated as a function of a series of intermediate phenotypes. Examples of intermediate phenotypes are biomass, radiation use efficiency and radiation interception on any given day or accumulated to a given day (Fig. 2). Intermediate phenotypes depend on CGM parameters that are genetically determined, and which modulate the phenotypic response to environmental covariables. An example of CGM parameters is the development rate to flowering in wheat where the parameters are vernalization requirement and sensitivity to photoperiod, which are regulated by the *VRN* and the *PPD* alleles, and regulate the phenotypic response (phenology) to temperature and photoperiod (Zheng et al. 2013).

CGM parameters, $\underline{\boldsymbol{y}}_i^P$, for phenotyped genotypes can be directly observed, estimated or calculated from the phenotypic measurements. However, given that CGM parameters depend on the genotype, they can also be predicted from genotypic covariables, i.e., molecular marker information. When we can identify the genetic basis of physiological parameters in terms of underlying QTLs, or, equivalently, when we can predict the physiological parameters from marker information, we can effectively predict the target trait from marker information and environmental inputs provided the intermediate traits and their interactions have been correctly identified and implemented in the CGM. Hence, predicted CGM parameters enable to predict the phenotype of genotypes that have not been observed yet. The prediction for individual CGM parameters ($\underline{y}_i^P$) would look like Model 4:

$$\underline{y}_i^P = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_q + \underline{G}_i + \underline{e}_i \qquad (4)$$

Like Model 1, Model 4 can be modified to include (i) only the QTLs, in a QTL model (Model 4.1) or (ii) only the polygenic effects ($\underline{G}_i$), in a genomic prediction model with the random effects $\underline{G}_i$ being structured by a genetic relationship matrix (Model 4.2).

$$\underline{y}_i^P = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_q + \underline{e}_i \qquad (4.1)$$

$$\underline{y}_i^P = \mu + \underline{G}_i + \underline{e}_i \qquad (4.2)$$

If more than one CGM parameter is to be predicted from molecular markers and/or pedigree information, Models 4, 4.1, and 4.2 could also be expanded to a multi-trait prediction model that takes into account possible correlations among the CGM parameters,

in a model that is similar to the multi-environment Model 1. Modelling traits simultaneously allows to gain power for QTL detection and to detect QTLs with pleiotropic effects (Alimi et al. 2013; Ste phens 2013).

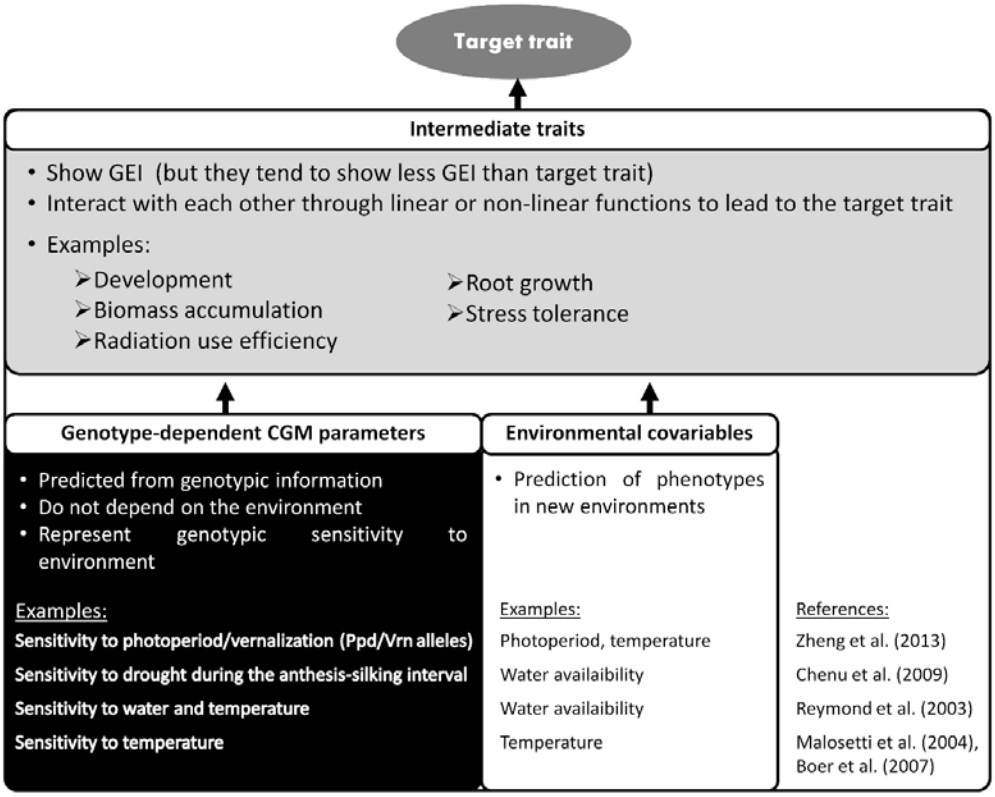


Figure 2. Representation of the main parameters of wheat yield used in the APSIM-CGM. Grey boxes correspond to intermediate phenotypes that are determined by the genotype and the environment (i.e. they show potentially large GEI), black boxes correspond to CGM parameters that are dependent on the genotype, and white boxes represent environmental covariables (http://www.apsim.info/).

Predictions for multiple CGM parameters, $\underline{\hat{\boldsymbol{y}}}_i^P$, can be used as input in Model 3 to calculate intermediate phenotypes, and produce the prediction for the target trait, $\underline{\hat{\mu}}_{ij}^t$, in Model 5.

$$\underline{\hat{\mu}}_{ij}^t = \int f\left(\underline{\hat{\boldsymbol{y}}}_i^P;\ \boldsymbol{z}_j\right) dt \quad (5)$$

In Model 5, the prediction accuracy of the target trait depends on the accuracy of the prediction of each of the components, and on the ability of the functions that transform CGM parameters into intermediate phenotypes to correctly describe the processes leading to the target trait.

CGMs with known/predicted genotypic parameters are a potentially useful tool to understand which traits can be advantageous in a given environment, and also to identify management practices that contribute to improved crop productivity (Yin et al. 2004; Hammer et al. 2006; Reynolds et al. 2009b; Harrison et al. 2014). In the context of adaptation to climate change, Zheng et al. (2012) modelled how phenology of current wheat varieties would influence their adaptation to future environments, which are expected to show different CO_2 and precipitation levels. In their second paper, Zheng et al. (2013) demonstrated that the flowering time of spring wheat genotypes can be modelled using the composition of their *VRN1* and *Ppd-D1* alleles together with responses derived from a single experiment with four environments: +/- treatments for vernalisation and extended photoperiod. Allelic combinations of loci *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, and *Ppd-D1* were used to predict APSIM-wheat parameters of a population of genotypes. From a single experiment (replicated in two years), they validated the model with more than 250 wheat genotypes across the entire Australian wheat belt, and were able to simulate flowering time for any weather records in the wheat belt. These conclusions can be useful to guide breeders in the process of determining which alleles should be considered in the selection process.

Bogard et al. (2014) extended this approach further to model the drivers of flowering time in winter wheat as functions of major genes as well as SNPs derived from association mapping, i.e., allowing prediction of unknown genotypes (but with known genes and SNPs) in new environments. In both Zheng et al. (2013) and Bogard et al. (2014), the predictions for heading date using the gene-based predicted parameters corresponded well with the observed dates to heading. Attributes that contributed to a successful phenotype prediction were (i) a well-defined CGM for heading date (Slafer and Rawson 1994), (ii) a well-defined set of environmental covariables with corresponding CGM parameters, and (iii) a well understood genetic basis of the CGM parameters (Snape et al. 2001).

The same approach has also shown to be successful for other more complex and less heritable traits such as grain yield under drought. For example, Chenu et al. (2009) used APSIM to model the impact of QTLs controlling the intermediate traits leaf and silk elongation on maize grain yield. The intercept and slope of these intermediate traits in response to meristem temperature, evaporative demand and soil water deficit were genotype-dependent (Reymond et al. 2003; Reymond et al. 2004).

Unfortunately, the identification of CGM parameters is sometimes less straightforward for complex traits like grain yield. Yin et al. (2000) showed an example in barley with a successful estimation of QTL effects for the CGM parameters, but with a poor prediction of grain yield. The correlation between the observed CGM parameters, i.e. phenotype of CGM parameters, and the QTL-predictions of the same parameters was high. However, the correlation between yield predictions of the CGM, whether phenotype based or QTL prediction based, and observed yield was not high. The cause of the poor predictions did not reside in the fact that the CGM parameters were replaced by predicted parameters from the QTL model, but in the fact that the CGM was unable to predict yield from its component traits. Similar work has been recently reported by Gu et al. (2014) on grain yield of rice crop, using a new CGM, which gave more promising results. However, efforts to improve CGM for predicting complex traits like grain yield are still strongly needed.

The example from Yin et al. (2000) shows that although the integrated statistical and CGM modelling allows for a larger flexibility, it might result in more complex and fragile models, because the approach can break down at the level of the estimation of the CGM parameters and at the level of the integration of these CGM parameters to calculate the intermediate phenotypes. However, even if the CGM are not fully able to predict the target trait, it is valuable to develop models of intermediate traits as well as for yield *per se*. Breeders can still be interested to recombine lines with high levels of proven intermediate traits with the expectation that these should on average result in better yield when further crossing and selection is done, i.e., because the selection on intermediate traits should already have improved part of the physiological adaptation pathway (Cooper et al. 2014). If breeders select mainly on yield *per se*, then it may be less likely that selections will also have high radiation use efficiency (RUE) or transpiration efficiency (TE) or traits for which genetic variation was not expressed in the given selection environment.

The examples of Zheng et al. (2013), Bogard et al. (2014), and Chenu et al. (2009) show that CGMs are a tool to integrate complex information from the genotypic, organ, and crop level. Dissection of a target trait into component traits at different levels of biological organisation allows phenotype prediction for the target trait in the face of genotype by environment and QTL by environment interactions for that same trait. Hence, the combined approach of statistical QTL-modelling and CGM is an alternative to model complex GEI interactions (Yin et al. 2004).

2.4. Further considerations

2.4.1. Classification of environments

Sections 2.2 and 2.3 presented models to predict the performance of genotypes in new environments ([G, nE, nGE] or [nG, nE, nGE]). However, if there are repeatable patterns that allow to classify environments, these patterns might help to reduce the complexity of $\boldsymbol{\Sigma}^E$ and thereby improve the accuracy of prediction.

One example of repeatable patterns that often justifies to group environments is the presence of regions. Here, we understand 'regions' (or mega-environments) as a group of locations where genotypes perform consistently across years (Bull et al. 1992; Gauch and Zobel 1997; Basford and Cooper 1998; Yan et al. 2000). Environments inside the same region are expected to be more homogeneous in terms of genotypic ranking, i.e., less GEI inside the regions (e.g. Atlin et al. 2000; Burgueño et al. 2008). In dryland production areas, other groupings may relate to characteristics of the soil (shallow/deep, low/high water holding capacity) and the management of the crop (sowing date, row spacing arrangement, etc.). De la Vega and Chapman (2010) showed how multiple component traits related to yield for a complex set of mega-environments in Argentina.

If locations can be grouped into regions, it is generally convenient to breed for specific adaptation to those regions, instead of broad adaptation across regions (Atlin et al. 2000, 2011). In this case, predictions can be produced for the whole of a region, or for new environments within a region. Precision of yield estimates might still benefit from the information of neighbouring regions by means of the covariance structure in a mixed model (Piepho and Möhring 2005; Kleinknecht et al. 2013).

When phenotypes are not available for all the locations of interest, environmental covariables can also be used to classify environments, and reduce the complexity of $\boldsymbol{\Sigma}^E$. Classifying environments into regions on the basis of environmental similarity, potentially allows to (i) predict new environments (as discussed in Sections 2.2 and 2.3, and also (ii) define the target population of environments, where a particular genotype is to be evaluated (Chapman et al. 2000a; Hammer et al. 2002; Chenu et al. 2011). CGMs are a powerful tool to identify relevant environmental factors (Chapman 2008; Messina et al. 2011), and the periods when the crop is most sensitive to those factors (Chenu et al. 2013). For example, considering drought seasonal patterns could give a better indication of the environment types, instead of the total rainfall per year (Chapman et al. 2000a, 2000b).

A further application of explicit environmental characterization is to weight environments based on their expected relevance for future years (Podlich et al. 1999). This means that environmental conditions that are more likely to occur receive more weight when doing the predictions, compared to less likely environmental conditions.

2.4.2. Population structure

Sections 2 and 3 discussed different models to predict phenotypes of unobserved genotypes using molecular marker information. In those sections, $\boldsymbol{\Sigma}^{G}$ had the structure of the genomic relationship matrix, without explicitly specifying sub-populations. However, genetic relatedness between training and validation sets largely influences prediction accuracy (Windhausen et al. 2012; Riedelsheimer et al. 2013). Hence, when there is strong population structure, it is necessary to define whether prediction will be done among or within populations. When predictions are limited to specific sub-populations, accuracy is commonly larger than when predicting across sub-populations, or when correcting for population structure (Daetwyler et al. 2012; Guo et al. 2014).

Methods to consider population structure in the model for genomic prediction can be based on the incorporation of the eigenvectors of the genotype by molecular marker data matrix (Patterson et al. 2006; Janss et al. 2012). Another option is to consider population structure in the design of the cross-validation scheme, for example by a stratified cross-validation design conditional on known population structure to ensure that each sub-population is equally represented in the training and validation sets (Albrecht et al. 2014; Guo et al. 2014).

2.4.3. Next generation sequencing

With the recent development of next generation sequencing technologies, genotyping costs have been largely reduced, allowing improving the genotypic characterization of important crops as barley, wheat and potato (Poland et al. 2012b; Uitdewilligen et al. 2013). In sequence-based genotyping approaches, marker discovery and genotyping are completed at the same time, allowing for faster genotyping processes (Poland and Rife 2012). The shorter time needed is thanks to the combination of restriction enzymes, sequencing, imaging, and genome alignment and assembly methods (Metzker 2010; Elshire et al. 2011).

These technologies permit the genotyping of larger populations of plants with higher marker density and increased mapping resolution (Varshney et al. 2014). Larger marker

density increases the chances of including causal loci that otherwise would not have been considered in models for phenotype prediction (Spindel et al. 2013). More loci in the model means increased genomic prediction accuracy (Poland et al. 2012a). However, models for phenotype prediction have diminishing returns on additional markers once the point of "marker saturation" has been reached, which depends on the genetic diversity of the population (Jannink et al. 2010; Heffner et al. 2011; Poland et al. 2012a).

Other questions regarding larger numbers of markers that remain not fully answered are: (i) how imputation of missing genotype data or haplotype inferences may affect prediction accuracies when genotyping by sequencing is used (Crossa et al. 2013), (ii) how to reduce the computational time needed because of the large number of markers (Verbyla and Cullis 2012), and (iii) how to improve model diagnostics, distinguishing between loci with large effects, and loci with smaller effects (Bernardo 2014).

2.4.4. High-throughput phenotyping to input to models for phenotype prediction

Mixed models and CGM discussed in Sections 2 and 3 are promising tools for phenotype prediction. However, these models require the phenotyping of multiple genotypes, traits and environments. With the reduction of genotyping costs, evaluating the populations phenotypically has become the limiting factor (Cobb et al. 2013).

High-throughput phenotyping platforms can either measure the target trait directly, or measure one or more traits that are correlated with the target trait. The use of CGMs allows estimation of hard-to-measure traits such as seasonal water use, given inputs of leaf area over time and canopy thermal characteristics, for example. Correlated traits measured by high-throughput phenotyping platforms can be used as inputs in models like Model 1. To do so, traits must: (i) have high genetic correlation with the target trait in the target environment, (ii) be less affected by environment (have a larger heritability) than the target trait, and (iii) provide an easy and reliable measurement, which is less expensive than the target trait itself (Bänziger 2000; Araus et al. 2008; Prasanna et al. 2013). When measuring correlated traits, high-throughput phenotyping platforms could be particularly useful for obtaining detailed non-destructive measurements of plant characteristics that collectively provide reliable estimates of trait phenotypes (Cabrera-Bosquet et al. 2012; Prasanna et al. 2013; Cooper et al. 2014).

High-throughput phenotyping platforms are commonly used under two scenarios: (i) precise phenotyping under controlled environments that aim at representing different levels of environmental quality, and (ii) phenotyping in environments that correspond to a sample

of environments in the field. The main advantage of controlled environments is that screening protocols can be more easily standardized, ensuring that plants are exposed to fairly reliable levels of stress. Hence, controlled environments offer the stability to search for attractive phenotypes or genotypes in a specific type of environment, e.g. drought stress (Cobb et al. 2013; Passioura 2012).

Growth under controlled conditions is usually different from that under field conditions. Hence, high-throughput phenotyping platforms in controlled environments might not lead to the identification of important yield-determining processes and promising genotypes in the field (Passioura 2012). This limits the application for phenotyping to specific stages of the crop (e.g., early vigour), or to traits that are correlated with the target trait (e.g., carbon isotope discrimination as an indicator of water use efficiency (Passioura 2012; Prasanna et al. 2013).

Popular high-throughput phenotyping techniques are those based on spectral technologies or remote sensing, as near infrared spectroscopy (NIRS), or image analysis. These techniques are a powerful tool that can provide information about multiple traits from only one or few images, and can be applied in controlled conditions as well as in field trials.

One example of how phenotypes obtained by image analysis can be included in phenotype prediction is shown by van der Heijden et al. (2012). Here, QTLs for leaf area were identified from the 3D representation of the plant canopy reconstructed from stereo images. The QTLs for leaf area from the image analysis agreed with the QTLs detected when using manually measured leaf areas, showing the potential of stereo images to characterize phenotypically breeding populations.

Image analysis introduces potentially larger measurement errors than conventional measurement techniques. For that reason, image information should be first carefully selected with the aid of statistical and physiological knowledge, in an automatized and standardized fashion, before incorporating it in the genetic analysis (Eberius and Lima-Guerra 2009; Hartmann et al. 2011). Hence, models accounting separately for the measurement error and for the experimental (plot) error should be considered (Smith et al. 2001b).

2.4.5. Concluding remarks

This chapter discussed several approaches that aim at predicting the phenotype in a multi-environment context. These approaches ranged from pure statistical models and pure CGMs, to a combination of both types of models. Special attention was given to different prediction scenarios; unobserved genotypes, new environments, and the combination of both. How prediction accuracy can profit from the large availability of environmental and genotypic information was also discussed, aiming at integrating physiological and statistical knowledge. Phenotypic and genomic data start to become abundant. The challenge for better phenotype prediction and more effective selection lies in more sophisticated procedures for selection of genotypic and environmental covariables in models for phenotype prediction, separating the signal from the noise.

Acknowledgements

Marcos Malosetti and Fred van Eeuwijk worked on this chapter as part of a project financed by the Generation Challenge Program - Integrated Breeding Platform (https://www.integratedbreeding.net/).

Chapter 3

Improvement of predictive ability by uniform coverage of the target genetic space

Daniela Bustos-Korts[1,2], Marcos Malosetti[1], Scott Chapman[3], Ben Biddulph[4], and Fred van Eeuwijk[1]

1. Biometris, Wageningen University & Research Centre
2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)
3. CSIRO Plant Industry and Climate Adaptation Flagship
4. Department of Agriculture and Food, Western Australia

This Chapter is published as:

Bustos-Korts, D., M. Malosetti, S. Chapman, B. Biddulph, and F. van Eeuwijk. 2016. Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. G3 Genes|Genomes|Genetics 6(11): 3733–3747.

Abstract

Genome-enabled prediction provides breeders with the means to increase the number of genotypes that can be evaluated for selection. One of the major challenges in genome-enabled prediction is how to construct a training set of genotypes from a calibration set that represents the target population of genotypes, where the calibration set is composed of a training and validation set. A random sampling protocol of genotypes from the calibration set will lead to low quality coverage of the total genetic space by the training set when the calibration set contains population structure. As a consequence, predictive ability will be affected negatively, because some parts of the genotypic diversity in the target population will be under-represented in the training set, whereas other parts will be over-represented. Therefore, we propose a training set construction method that uniformly samples the genetic space spanned by the target population of genotypes, thereby increasing predictive ability. To evaluate our method, we constructed training sets alongside with the identification of corresponding genomic prediction models for four genotype panels that differed in the amount of population structure they contained (maize Flint, maize Dent, wheat, and rice). Training sets were constructed using uniform sampling, stratified-uniform sampling, stratified sampling and random sampling. We compared these methods with a method that maximizes the generalized coefficient of determination (CD). Several training set sizes were considered. We investigated four genomic prediction models: multi-locus QTL models, GBLUP models, combinations of QTL and GBLUPs, and Reproducing Kernel Hilbert Space (RKHS) models. For the maize and wheat panels, construction of the training set under uniform sampling led to a larger predictive ability than under stratified and random sampling. The results of our methods were similar to those of the CD method. For the rice panel, all training set construction methods led to similar predictive ability, a reflection of the very strong population structure in this panel.

3.1. Introduction

The key factor to progress in plant breeding is the number of genotypes that can be evaluated phenotypically (Cooper et al. 2014b). Unfortunately, field testing is slow and costly, forcing breeders to limit the number of genotypes that is phenotyped. Genomic prediction offers the potential to alleviate this limitation, allowing to broaden the pool of genotypes for selection, and thereby increasing selection intensity (Crossa et al. 2013; Windhausen et al. 2012) and efficiency of breeding programs (Heffner et al. 2010; Crossa et al. 2013; Windhausen et al. 2012; Hickey et al. 2014; Longin et al. 2015).

In genomic selection, genome-enabled genotypic or breeding values are calculated from genomic prediction models as sums of effects for large numbers of markers, often without explicitly testing individual marker–trait associations (Meuwissen et al. 2001). Genomic prediction models are developed for a target population of genotypes (TPG). The TPG describes the full collection of existing and future genotypes that is supposed to be suitably adapted to the environmental conditions defined by the target population of environments (Cooper et al. 2014a; Cooper and Hammer 1996; Comstock 1977). Breeders have access to a sample from the TPG, the target sample. This sample of genotypes (or part of it) can be regarded as a calibration set for genomic prediction models when both phenotypic and marker data are available. To estimate the marker effects in prediction models, the calibration set is typically partitioned into a training set and a validation set. Marker effects are estimated on the training set of genotypes, and subsequently, genotypic values are calculated for all genotypes in the training and validation set. For accurate genomic prediction of the genotypic values in the validation set, training and validation sets should have similar genetic diversity, reflected in large kinship coefficients (Saatchi et al. 2011; Auinger et al. 2016). This condition is more likely to be met if the training set covers the whole genotypic, say genetic, space of the calibration set. As the calibration set is assumed to be a representative sample of the TPG, we also hope to cover the genetic space of the TPG. Therefore, a highly diverse TPG requires a larger training set size to capture the whole range of genetic diversity (Hayes et al. 2009).

Conventionally, genomic prediction literature uses random sampling as a strategy to split the calibration set into a training and a validation set (Burgueño et al. 2012; Crossa et al. 2010; Heslot et al. 2013; Schulz-Streeck et al. 2012; Riedelsheimer et al. 2012). In random sampling, genotypes belonging to the calibration set have equal probability to enter the training set. Hence, random sampling reproduces the genotypic frequencies of the calibration set, leading to a more dense coverage of those parts of the genetic space that are represented by a larger number of genotypes (Odong et al. 2013; Jansen and van Hintum 2007). Furthermore, we hypothesize that the heterogeneous coverage of the genetic space produced

by random sampling leads to decreased predictive ability because part of the genetic diversity in the validation set is not well represented in the training set.

One strategy to improve the coverage of the genetic space is to use stratified sampling. In stratified sampling, the calibration set is divided into subpopulations and then a proportion of genotypes is randomly selected from each subpopulation (Guo et al. 2014; Albrecht et al. 2014; Janss et al. 2012; Daetwyler et al. 2012). However, subpopulations are sometimes not clearly defined or they are internally heterogeneous (Crossa et al. 2013). Thus, stratified sampling improves the coverage of the genetic space compared to random sampling, but it does not guarantee that all relevant genotypes are included in the training set.

The importance of an adequate representation of the genetic space for successful genomic prediction has been acknowledged in the recent literature. Rincent et al. (2012) assumed that predictive ability can be improved if genotypes in the training set are chosen in such a way that the precision of the contrasts between each genotype in the validation set and the mean of the calibration set is maximized. This can be achieved by maximizing the generalized coefficient of determination (CD). This method was further adapted by Isidro et al. (2015), who combined the method of Rincent et al. (2012) with stratified sampling. In this method (Isidro et al. 2015), the calibration set is first classified into subpopulations and then the CD mean criterion proposed by Rincent et al. (2012) is applied inside each subpopulation.

The methods proposed by Rincent et al. (2012) and by Isidro et al. (2015) rely on the variance components estimated from phenotypic data to choose genotypes for the training set. Although training set composition is not very sensitive to changes in variance components, some small differences in the genotypes allocated to the training set could be observed from trait to trait due to trait heritability differences (Rincent et al. 2012).

A statistically attractive strategy to increase the genetic similarity between training and validation sets is to uniformly cover the genetic space of the population of genotypes. Uniform coverage of the genetic space as a criterion for choosing members of the training set has the advantage of purely genotypic information being sufficient, without requiring phenotypic information (Jansen and van Hintum 2007; Odong et al. 2011). This principle is well known in the genetic resources literature, where it is used to define germplasm core collections (Odong et al. 2013). Here, we interpret the core collection as a training set because both of them, core collection and training set, are a subset of genotypes that aim at representing the genetic diversity present in a larger population.

Once the training set has been constructed, the next task is to identify a suitable prediction model. A large range of prediction models have been proposed, and they differ in two main aspects. The first aspect is the weight that models assign to specific genomic regions. If large QTL are present, predictive ability might benefit from modifying the common assumption that all marker effects come from a common normal distribution (Hayes et al. 2009). Hence, depending on the trait genetic architecture, it might be convenient to give more importance to genomic regions with large effects (Crossa et al. 2013; Daetwyler et al. 2010; Speed and Balding 2014; Hayes et al. 2009; Bernardo 2010).

The second aspect is whether the model accounts only for additive genetic effects, or also for nonadditive effects (Langer et al. 2014; Reif et al. 2011; Kippes et al. 2014; Stange et al. 2013). The GBLUP model proposed by Meuwissen et al. (2001) can be extended to separately account for nonadditive genetic effects (Oakey et al. 2006). However, the model proposed by Oakey et al. (2006) is computationally demanding. A less demanding model option for various types of nonadditive effects is the class of Reproducing Kernel Hilbert Space (RKHS) models, for example, with a Gaussian Kernel (Gianola and van Kaam 2008; Piepho 2009; Jiang and Reif 2015). The advantage of RKHS models is that they can be used across a spectrum of genetic architectures (de los Campos et al. 2009).

Given the importance of population structure and trait genetic architecture for effective implementation of a genomic prediction strategy, the objectives of this paper were (i) to compare strategies to define the training set, and (ii) to compare the predictive ability for models with explicit QTL with the predictive ability of GBLUP and RKHS models.

3.2. Materials and Methods

3.2.1. Data

To compare the strategies for training set construction and prediction models, we used four genotype panels that differed in the amount of population structure (Flint and Dent maize panels, and a wheat and rice panel).

Maize

The maize data consisted of a Flint panel crossed with a Dent tester (F353) and of a Dent panel crossed with a Flint tester (UH007) to produce hybrid progeny for phenotypic evaluation, published by Rincent et al. (2014b). Both panels were composed of lines aiming at best representing the diversity of Flint and Dent maize in Northern Europe. The panels

included commercially used inbred lines created from open pollinated varieties, and lines recently developed by public institutes or, in the case of the Dent panel, private companies.

The Dent panel consisted of 276 genotypes, whereas the Flint panel had 259 genotypes. Both panels were evaluated infield trials in Germany, France and Spain during 2010 and 2011. In this paper, we used the adjusted means of tasseling date, silking date and dry matter yield for each genotype across all environments (Supplemental Materials 12 and 13 in Rincent et al. (2014b)). Tasseling and silking date were expressed as growing degree days after sowing, considering a base temperature of 6°C, using the mean daily air temperature measured in each environment. Both panels were characterized genotypically with the Illumina maize SNP50 BeadChip described in Ganal et al. (2011). From this set, we used only the markers that were developed by comparing the sequences of nested association mapping founder lines (PANZEA SNPs, Gore et al. (2009); Rincent et al. (2012)). Individuals which had marker missing rate and/or heterozygosity higher than 0.10 and 0.05, respectively, were eliminated. Missing marker genotypes (below 2% in both panels) were imputed with the software BEAGLE. Markers with minor allele frequency lower than 0.05 were eliminated, leading to 28,304 PANZEA markers for the Dent panel, and 25,578 PANZEA markers for the Flint panel (Rincent et al. 2014b).

Wheat

This wheat panel was constructed to represent flowering time variation present in Australian wheat germplasm. Phenotypic data corresponded to the adjusted means across environments for yield and heading date of 149 genotypes observed during 2009. Yield was observed at eight locations, whereas heading date was observed at six locations in the Australian wheat belt. Genotypes were characterized with 4295 SNPs, from which four SNPs were at the position of major genes regulating phenology (Ppd-D1, Vrn-A1, Vrn-B1, Vrn-D1). Missing markers were replaced by imputed genotypic data using the missForest package in R, following the methodology explained in Bogard et al. (2014). One marker was discarded as it showed .25% missing data, 39 markers were removed as they were monomorphic on this panel, and 431 were discarded because they had a minor allele frequency lower than 0.05. This led to 3754 markers for further analysis. Wheat genotypic and phenotypic data are available in Supplemental Material, File S1, File S2, and File S3.

Rice

The rice data consisted of 413 diverse accessions of inbred lines from 82 countries. This data set is publicly available at http://www.ricediversity.org. Phenotypes consisted

of plant height, seed number per panicle and flowering time in Arkansas. Genotypes that were too similar to each other (causing the relationship matrix to be singular) or that had a missing phenotype, were removed, leaving 350 genotypes for the analysis. The panel was genotyped with a 44-K SNP chip. After filtering, 36,091 markers were retained in the published data set. From this set of markers we discarded those that had 5% of missing values. The remaining missing marker scores were imputed with the software BEAGLE. Markers with minor allele frequency lower than 0.05 (considering only the phenotyped lines) were eliminated, leading to 26,259 markers.

3.2.2. Characterization of the population structure

Population diversity was explored by principal component analysis of the identity by state (IBS) matrix among genotypes, calculated from molecular markers (Equation 1). This IBS calculation method indicates the proportion of shared alleles between genotypes.

$$A^{IBS} = \frac{GG' + G_2 G_2'}{K} \qquad (1)$$

The number of subpopulations present in each data set was determined with the Tracy–Widom statistic, following Patterson et al. (2006). Here, the number of subpopulations equals the number of significant principal components, plus one. Genotypes were qualitatively assigned to the subpopulation using the STRUCTURE software (Pritchard et al. 2000) and with the number of groups as determined by the Tracy–Widom statistic. To get an impression about population differentiation, the Fst statistic was calculated following Weir (1996) using a self-coded program in GenStat v.17 (VSN-International 2015).

3.2.3. Training and validation sets

To split the calibration sets into a training and a validation set, we used the following five methods:

Uniform coverage of the genetic space (U)

In U, we used the methodology proposed by Jansen and van Hintum (2007). This method consists of the following steps, which are applied to the list of all genotypes contained in the panel (P_1): (1) Molecular markers are used to calculate identity by state among all genotypes in P_1 (IBS, Equation 1). (2) The first entry of the training set (T_1) is sampled at random from the panel. Genotypes with a distance to T_1; smaller than a sampling radius *r*, are discarded from the training set. The new list of candidate genotypes is called P_2: The genotypes that

are discarded are stored in a list called D_1: (3) The second entry of the training set is sampled at random from P_2 and it is called T_2: Genotypes with a distance to T_2 smaller than the sampling radius *r* are discarded from the list of genotypes. This process is repeated until all the genotypes have been included in the training set T_n, or in the list of discarded genotypes (D_n). U is implemented in the "sampling" method of the GenStat procedure QGSELECT (VSN-International 2015).

The sampling radius used in step (2) was obtained empirically. The size of this radius depends on the training set size one aims at. If the desired training set size is larger, the sampling radius becomes smaller. The target *r* is obtained by slowly decreasing its values until the number of sampled genotypes is greater than or equal to the target sample size, following Figure 1 in Jansen and van Hintum (2007).

Stratified sampling with uniform coverage of the genetic space (SU)

In SU, prior information about the grouping of the genotypes was supplied. In this method, an extra restriction was added to the distance restriction. Genotypes are discarded when they are within the sampling radius and they belong to the same group (i.e., they are included in the training set when they are within the sampling radius, but they belong to a different group). This method ensures that each group is represented by at least one genotype.

Generalized coefficient of determination (CD)

The generalized coefficient of determination was used as a criterion to select genotypes for the training set in such a way that the precision of the prediction of the difference between the value of each individual in the validation set and the mean of the total calibration set is maximized (Rincent et al. 2012). Briefly, the precision is maximized when the generalized coefficient of determination (CD, Equation 2) is maximized.

$$CD(c) = diag\left[\frac{\left(c'\left(A^{AB} - \lambda\left(Z'MZ + \lambda\left(A^{AB}\right)^{-1}\right)^{-1}\right)\right)c}{c'A^{AB}c}\right] \quad (2)$$

In Equation 2, *c* is a matrix of the contrasts between each individual in the validation set and the mean of the calibration set, *M*, is an orthogonal projector of the subspace spanned by the columns of the design matrix of the fixed effects, *X*, (in our case, only the

intercept): $M = I - X(X'X)^{-}X'$. λ is the ratio between the residual and the additive genetic variance. For Flint and Dent, we calculated λ from the heritability estimates reported by Rincent et al. (2014b). For wheat heading time and yield, we used an estimate for λ calculated from the phenotypic data (h^2= 0.95 for heading time and h^2= 0.89 for yield). No heritability estimate was available for rice. Thus, we arbitrarily used 0.85 for the three rice traits.

A^{AB} is the realized additive genetic relationship matrix calculated from all molecular markers along the whole genome following the equation proposed by Astle and Balding (2009), with as typical entry for the relationship between genotypes *i* and *j*:

$$A_j^{AB} = \frac{1}{K}\sum_{k=1}^{K}\frac{(G_{ik}-2p_k)(G_{jk}-2p_k)}{2p_k(1-p_k)} \qquad (3)$$

where G_{ik} is a marker score that can take the value 2, 1, or 0 for genotype i at marker *k*, and p_k is the allele frequency of marker *k*. The matrix above was calculated using the "realizedAB" option in the "kin" function of the Synbreed package (Wimmer et al. 2012).

The optimization algorithm used by Rincent et al. (2012) to construct the training set was implemented in R3.2.1. Briefly, at each step, one genotype in the training set is exchanged by one genotype in the validation set. This exchange is accepted if CD is increased and is rejected otherwise. The algorithm was allowed to iterate until the CD did not change anymore (800 times was enough to reach stability in all data sets).

Stratified random sampling (S)

In S, the number of sampled genotypes depended on the logarithm of the subpopulation size, following Franco et al. (2005) and Malosetti and Abadie (2001).

$$n_{t,s} = n_t \frac{\log(n_s)}{\sum_{s=1}^{S}\log(n_s)} \qquad (4)$$

In Equation 4, $n_{t,s}$ is the number of genotypes to be sampled from subpopulation s into the training set, *S* is the number of subpopulations, n_t is the total size of the training set we want to construct, and n_s is the number of individuals belonging to subpopulation *s* in the calibration set. Within the subpopulations, genotypes were sampled at random.

Random sampling (R)

In strategy R, the training set was sampled at random, so each genotype in the calibration set had equal probability of being included in the training set.

One hundred independent realizations of each of the five sampling strategies U, SU, CD, S, and R were generated for each calibration set. Each of the training sets (sampled genotypes) was used for QTL detection and as a training set for the prediction models.

Table 1. Abbreviations and descriptions for training set construction methods

Abbreviation	Description
U	Uniform coverage of the genetic space
SU	Stratified sampling with uniform coverage of the genetic space
CD	Generalized coefficient of determination (Rincent et al. 2012)
S	Stratified random sampling
R	Random sampling

3.2.4. Characterization of the training sets

To characterize the connection between the training and the validation set, we calculated the distance between each genotype in the validation set and the nearest entry in the training set, following the method *Average distance between each accession and the nearest entry (A–NE)* in Odong et al. (2013). Here, we interpret the core collection in that paper, consisting of entries, as a training set. Core collection entries and training set members form a subset of genotypes that aim at representing a larger collection of genotypes. The set of accessions from which a core collection is created, we interpret to represent a calibration set. The distance between a genotype in the validation set and the nearest genotype in the training set (or core collection in Odong et al. (2013)) was calculated as $(1 - IBS)$. The empirical distribution of these distances was plotted for each training set construction method.

To obtain an impression of how each subpopulation is represented in the training set, we calculated the proportion of genotypes from each subpopulation that is included in the training set. The mean IBS in each subpopulation was used to relate the sampling proportion to the genetic diversity in each subpopulation.

3.2.5. QTL detection

Training sets obtained by U, SU, CD, S or R sampling of the genotype panel were used to identify QTL that became part of the prediction model. QTL were identified by a genome-wide association mapping scan (GWAS), following Equation 5:

$$y_i = \mu + x_{ik}\alpha_k + G_i + e_i \qquad (5)$$

In Equation 5, y_i stands for the phenotype of genotype *i*, *m* is the intercept, x_{ik} is a vector that represents information of genotype *i* at marker *k* (0 and 2 for homozygous and 1 for heterozygous), and α_k is the additive QTL effect (fixed) for marker *k*. G_i represents a polygenic effect for genotype *i*, and e_i is the nongenetic residual $(e_i \sim N(0, \sigma_e^2))$. The distribution of G_i is $G_i \sim N(0, A\sigma_g^2)$. A is the additive genetic relationship matrix calculated from the molecular marker information as in Rincent et al. (2014a). In this method, a specific A is calculated for each linkage group by excluding the markers on that particular linkage group. A significance threshold equivalent to a genome-wide significance level of 0.01 was used for the four data sets, following the Li and Ji (2005) multiple-testing correction. We performed the GWAS as implemented in GenStat 17th edition (VSN-International 2015).

3.2.6. Prediction models

The following prediction models were used:

QTL:

$$y_i = \mu + \sum_{q \in Q}(x_{iq}\alpha_q) + e_i \qquad (6)$$

In Equation 6, μ is the intercept, $\sum_{q \in Q}(x_{iq}\alpha_q)$ is for genotype *i* the sum of (random) QTL effects belonging to the QTL set *Q*, where these QTL were identified in a preliminary GWAS scan. Effects for each QTL were allowed to have their own distribution $\left(\alpha_q \sim N(0, \sigma_q^2)\right)$, and e_i is the residual $(e_i \sim N(0, \sigma_e^2))$.

GBLUP:

$$y_i = \mu + G_i + e_i \qquad (7)$$

In Equation 7, μ is the intercept and G_i represents the random genotype effects that follow a distribution $G_i \sim N(0, A^{AB}\sigma_g^2)$. A^{AB} is the additive relationship matrix, following Astle and Balding (2009), Equation 3. The predictions were calculated using GenStat 17th edition (VSN-International 2015).

QGBLUP:

$$y_i = \mu + \sum_{q \in Q}(x_{iq}\alpha_q) + G_i + e_i \qquad (8)$$

The model in Equation 8 combines the QTL and GBLUP model. Again, μ is the intercept, $\sum_{q \in Q}(x_{iq}\alpha_q)$ is the sum of random QTL effects from the QTL set *Q* for genotype *i*, with each

of the QTL effects having proper variance component, $\alpha_q \sim N(0, \sigma_q^2)$. The polygenic effects G_i are assumed to follow a distribution $G_i \sim N(0, A^{ABm}\sigma_g^2)$. e_i is the residual ($e_i \sim N(0, \sigma_e^2)$). A^{ABm} corresponds to a modified additive relationship matrix, calculated from all markers except those that were within a window of ±20cM around QTL. This precaution was taken to avoid accounting for the QTL effects both in the random QTL terms, and in the residual polygenic term. Again, predictions were calculated in GenStat 17th edition (VSN-International 2015).

RKHS:

The RKHS model is as the GBLUP model in Equation 7, but $G_i \sim N(0, A^*)$. $A^* = exp^{\left(\frac{-D}{\theta}\right)}$ represents the genetic relationship matrix, where *D* is a matrix with Euclidean dissimilarities among genotypes calculated from marker scores in the Synbreed package (Wimmer et al. 2012), and Θ is a tuning parameter which determines how the covariance among individuals decays as a function of the genetic distance (Gianola and van Kaam 2008; Piepho 2009). An estimate for Θ was obtained by fitting mixed models along a grid of values between 0.05 and 5. The u value that provided the best predictive ability over a number of validation sets was used as the final Θ value (de los Campos et al. 2010). The final Θ value also showed the lowest AIC across the grid. The RKHS predictions were fitted by the REML procedure in GenStat v.17 (VSN-International 2015).

3.2.7. Training set size

For maize, training sets contained 50, 70, 100, 150, or 200 genotypes (to match sample sizes chosen by Rincent et al. (2012)). The wheat data had a limited panel size, so training set sizes of 50, 75, and 100 genotypes were used. Rice training sets had a size of 50, 100, 150, 200, or 300 genotypes to match the sizes used by Isidro et al. (2015).

3.2.8. Predictive ability

Predictive ability was calculated as the Pearson correlation coefficient between observed and predicted phenotypes (Meuwissen et al. 2001). To evaluate whether predictive ability was driven by population structure, the Pearson correlation was calculated both across subpopulations, so ignoring population structure, and within the subpopulations, where it should be remarked that for smaller sub-populations no reliable estimates of predictive ability may be possible. We wanted to study the influence of training set construction method, prediction model, and training set size on predictive ability. For each combination of these three factors, we calculated mean predictive ability

across 100 training set realizations. We also calculated a standard error (S.E.). To comply with the normality assumption, correlations were analyzed on a transformed scale using Fischer's z transformation, $z = 1/2\ (ln((1+r)/(1-r)))$ and means were back transformed using $r = (\exp(2z)+1)/(\exp(2z)-1)$ before reporting them.

3.2.9. Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

3.3. Results

We first explored population structure for the Flint, Dent, wheat, and rice panels. Subsequently, we investigated the properties of training sets constructed following the training set construction methods U, SU, CD, S, and R. Finally, we present the results of predictive abilities as defined by training set construction method and varying training

3.3.1. Population structure

We present the panels ordered from the least to the most structured. Flint with an Fst statistic of 0.11 was the least structured panel; 5.96% of the total variation was explained by PC1 and 3.84% by PC2 (Figure 1). PC1 separated the Northern Flint genotypes from the other Flint genotypes, coinciding with what was reported by Rincent et al. (2012). The Tracy-Widom statistic indicated that four PCs were significant, suggesting five genetic groups. Although the separation between some groups is not visible in the three dimensions shown in Figure 1, groups were separated in higher dimensions.

The Dent panel with an Fst of 0.19 was slightly more structured than the Flint panel. A larger percentage of variation was explained by the first PCs (5.64% for PC1 and 4.62% for PC2, Figure S1). Five PCs were significant, thus, genotypes were classified into six subpopulations. The first PC separated the IODent from the non-IODent genotypes, the second PC separated the stiff-stalk from the non-stiff-stalk genotypes, and the third PC separated the D06 family from the rest. The remaining subpopulations were separated by PC4 and PC5.

For the wheat panel, Fst was 0.28 and four PCs were significant, indicating the presence of five subpopulations. PC1 (11.41%) tended to separate genotypes by their vernalization requirements, and PC2 (8.48%) tended to separate genotypes by their sensitivity to

photoperiod (Figure S2). Rice was the most structured panel that we analyzed with an Fst of 0.36. Only the first PC was significant (39.41% of the variation), indicating two clearly distinguishable subpopulations (see Figure 2).

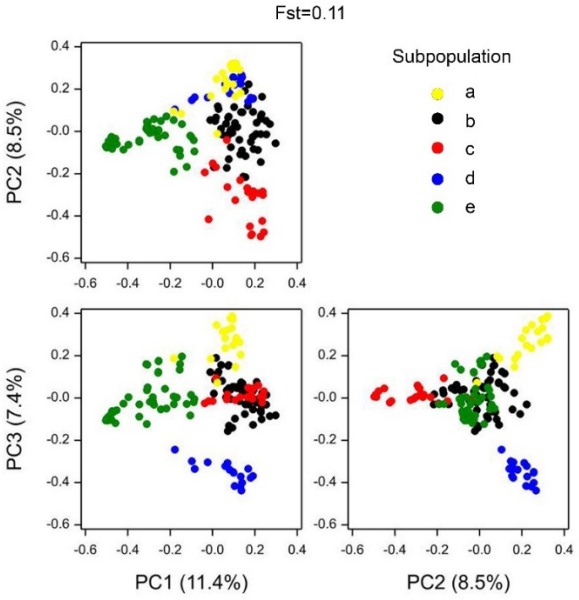


Figure 1. Scatter plots for principal components representing IBS matrix of the Flint panel. Symbol colour represents each of the five subpopulations.

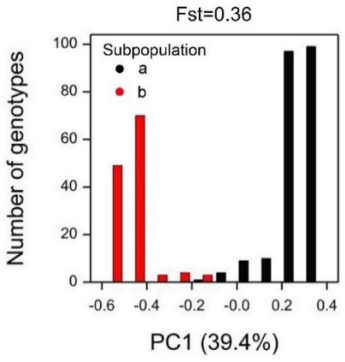


Figure 2. Histogram for the scores of the principal component representing the IBS matrix of the rice panel. Symbol colour represents each of the subpopulations.

3.3.2. Training and validation sets

In this section, we compare five methods to construct training sets from calibration sets (U, SU, CD, S, and R, see Table 1 for a description of the method abbreviations). Each individual calibration set is split into a training set and a validation set. For each combination of training set construction method, training set size, and genomic prediction model, 100 training sets were constructed, or drawn, from a calibration set.

Representation of subpopulations in training sets:

Random sampling of genotypes in the calibration set, i.e., training set construction method R, to create a training set, will lead to a training set with proportional representation of subpopulations. In Table 2, we express the abundance of genotypes coming from a particular subpopulation when using training set construction methods U, SU, CD, and S relative to the abundance for that subpopulation as realized by application of training set construction method R. For all panels it held that large and diverse subpopulations were over-represented in the training sets created by application of U, SU, and CD in comparison to R. The lowest diversity subpopulations were always under-represented when using U, SU, and CD. Subpopulation affected representation in an expected way for the Dent panel and rice panel for the comparison of S vs. R, that is, larger subpopulations were under-represented and smaller subpopulations were over-represented. For the Flint and wheat panel the relationship between representation and subpopulation size was not clear. In conclusion, for U and SU, a relatively larger number of genotypes was allocated to the training set from those parts of the genetic space that were more diverse. CD behaved comparably to U and SU for all panels.

Table 2. Subpopulation size in the calibration set, genetic diversity (*Div =1- median IBS*) and number of calibration set genotypes assigned to the training set, expressed as a percentage of the number realized by random sampling.

	Flint, 200						Dent, 200						Wheat, 200						Rice, 300					
	Size	Div.	U	SU	CD	S	Size	Div.	U	SU	CD	S	Size	Div.	U	SU	CD	S	Size	Div.	U	SU	CD	S
a	50	0.26	-35	-25	-30	13	17	0.26	-61	-53	-42	38	17	0.18	-20	-13	-14	-9	220	0.17	-8	-8	-4	0
b	30	0.30	-5	12	-7	-18	45	0.28	-33	-31	-25	10	19	0.25	-5	-1	-4	-8	129	0.31	14	14	7	0
c	55	0.33	8	6	8	3	13	0.31	-1	-4	-10	35	41	0.31	-28	-24	2	2						
d	30	0.34	1	8	4	-1	38	0.32	-11	-13	-14	22	21	0.35	5	4	-3	6						
e	94	0.39	15	4	13	-3	40	0.36	-3	-4	-5	22	51	0.40	22	14	14	0						
							f	123	0.38	25	24	22	-27											

For the description of the training set construction methods U, SU, CD, S, and R, see Table 1.

Distance between validation set and training set:

Our objective was to evaluate methods for training set construction that provide a more homogeneous coverage of the genetic space and that reduce the genetic distance between genotypes in the validation set and those in the training set. The underlying rationale is that the lower the genetic distance (larger genetic relatedness) between validation and training sets, the better the predictive ability in the validation set is expected to be. Figure 3 shows the distribution of distances of validation set genotypes to the closest training set genotype, with $distance = 1 - IBS$; summed over all 100 realizations of the training set. A broad distribution indicates high heterogeneity of distance, i.e., some validation genotypes are close to the training set, whereas others are distant. Our objective was to construct a training set that is on average close to the validation set with little variation between validation genotypes, reflected in a narrow distribution.

At all training set sizes, SU and U had a narrower distribution than CD, S, and R, showing that training set samples created by SU and U achieve a homogeneous coverage of the genetic space and that these sampling outcomes are consistent from realization to realization. At small training set size, the median and the maximum genetic distance between genotypes in the validation set and those in the training set was similar for U, SU, S, and R. Only CD showed a smaller median distance, compared to the other four methods, especially for wheat and rice (Figure 3 and Table S1). At larger training set sizes, the methods CD, U, and SU showed smaller distances between genotypes in the validation and training sets, compared to S and R (Figure 3 and Table S1). CD coincided with U and SU for the modal genetic distance, but tended to have a broader distance distribution. This broader genetic distance distribution implies that while on average CD, U, and SU are similar, CD tends to achieve a less homogeneous coverage of the genetic space, when compared to U and SU.

Incorporating a priori defined subpopulations into the genetic distance sampling, SU vs. U, had only a small effect for the least structured panels, Flint and Dent. For those panels, U showed a slightly narrower distribution than SU. This difference was most relevant at small sample sizes. In the case of more structured populations (wheat and rice), the incorporation of a priori subpopulation information into the sampling process did not change the distribution of genetic distances between validation and training sets. This means that as a desirable feature of our U method population substructure, whether subtle or not, it will automatically be accounted for in the construction of the training set.

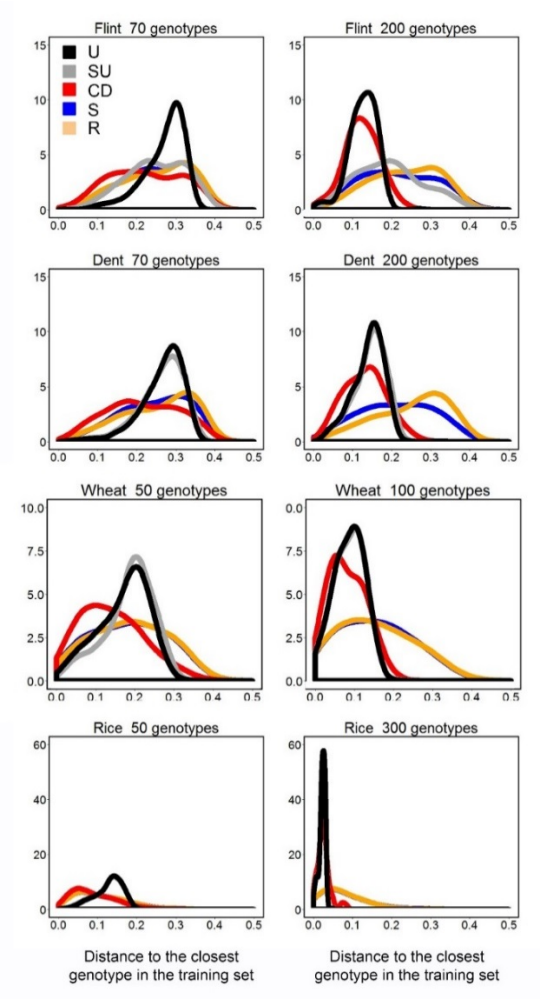


Figure 3. Distribution of genetic distances (distance=1-IBS) between validation set genotypes and the closest genotype in the training set (summed over 100 sampling events). For the description of the training set construction methods U, SU, CD, S, and R see Table 1.

QTL detection in the training set:

The number of detected QTL increased with training set size (Table S2, Table S3, and Table S4). At training set sizes smaller than 100 genotypes, the number of sets in which QTL were detected was very small and their positions changed across training sets. For training set sizes of 100 genotypes or larger, CD, U, and SU produced a larger number of QTL than S and R. In the case of the Flint panel, most consistent QTL were detected on linkage group

1 for tasseling, silking and yield (Table S2). For Dent, QTL were detected most often on linkage groups three and eight for tasseling and silking and in linkage group 5 for yield (Table S3).

Very few QTL for grain yield were detected in the wheat panel. For heading time, large QTL for photoperiod and vernalization requirements appeared only at larger sample sizes, reflecting that the population was too small for QTL detection in the training set. However, given that the population was characterized for loci that are known to be relevant for vernalization and photoperiod sensitivity, we decided to include these four loci in all the QGBLUP and QTL models for heading date. For the rice panel, the most consistent QTL for plant height was detected on linkage group 1 (Table S4). When using the methods U, SU, and CD, an important proportion of the training sets showed a QTL on linkage groups 2 and 6 at larger training set sizes. For seed number, a consistent QTL was detected for a training set size of 300 genotypes on linkage group 12. For flowering date, the most consistent QTL were detected on linkage groups 3 and 5. Again, these QTL were more often detected with U, SU, and CD, than with S and R.

Predictive ability in the validation set, ignoring subpopulations:

First, we present predictive ability as calculated on all genotypes in the validation set, pooling validation genotypes across subpopulations. To investigate the influence of the subpopulations on the accuracy, we have also calculated within subpopulation prediction abilities (see below).

In the Flint, Dent, wheat, and rice panel, as expected, the relative predictive ability of methods depended on the training set size (Figure 4 and Figure 5). While at small training set sizes, differences between all methods were minor, at larger training set sizes, methods that reduced the distances between the validation and the training set (i.e., U, SU, and CD) showed a clear improvement compared to S and R with an absolute increase in predictive ability of between 0.10 and 0.25.

Prediction models differed in predictive ability (Table 3, Table 4, Table 5, and Table 6). For the Flint, Dent, and rice panels, RKHS, GBLUP, and QGBLUP showed a larger predictive ability than the QTL model. This indicates that the evaluated traits were regulated by a large number of loci (Table 3, Table 4, and Table 6). For the same reason, including QTL in a separate model term (QGBLUP) was not advantageous over GBLUP. The comparable results of RKHS and GBLUP indicate that non-additive genetic effects were not so relevant for the analyzed traits in the Flint, Dent, or rice panels.

Model ranking was slightly different for heading date in the wheat panel from that in Flint, Dent, and rice. In the case of heading date, QGBLUP led to larger predictive ability, compared to GBLUP and QTL (Table 5). This indicates that, for heading time in wheat, it is convenient to account separately for loci with large effects. However, RKHS showed a larger predictive ability than QGBLUP, reflecting that non- additive genetic effects contribute to phenotypic variation of heading date. In the case of grain yield, no large QTL were consistently detected and therefore, we only used RKHS and GBLUP to predict this trait in wheat. As for heading, RKHS showed a larger yield predictive ability than GBLUP.

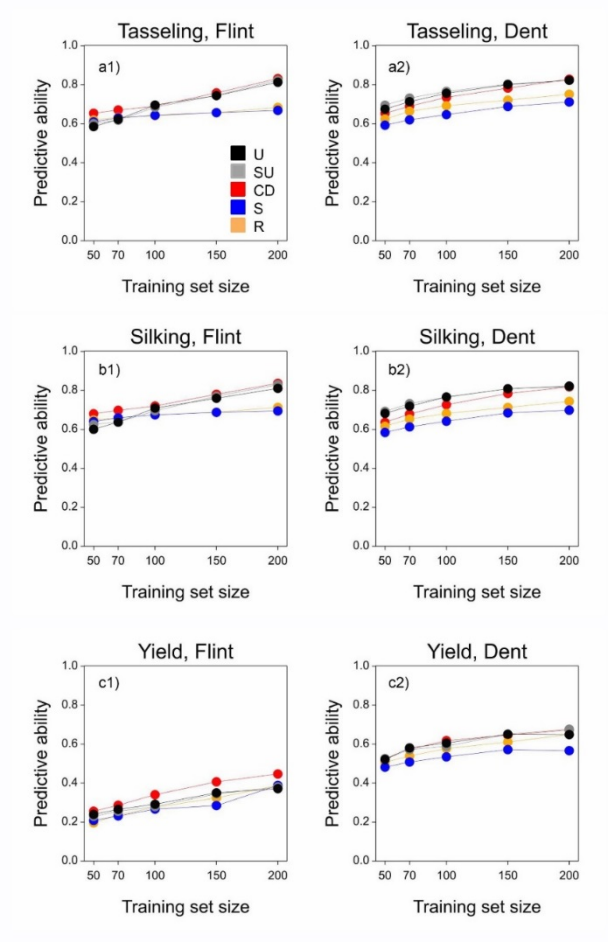


Figure 4. Predictive ability for the Flint and Dent panels as a function of training set size, using the RKHS model. The mean standard error for predictive ability was 0.001. For the description of the training set construction methods U, SU, CD, S, and R see Table 1.

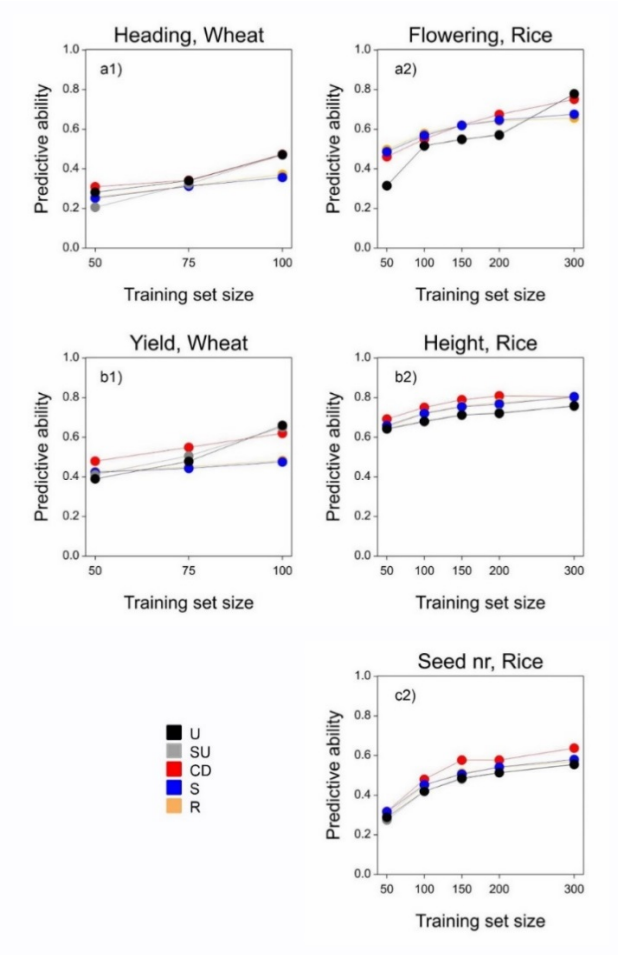


Figure 5. Predictive ability for the wheat and rice panels as a function of training set size, using the RKHS model. The mean standard error for predictive ability was 0.001. For the description of the training set construction methods U, SU, CD, S, and R see Table 1.

Table 3. Predictive ability for the Flint panel, using a training set size of 200 genotypes.

Model	U	SU	CD	S	R	S.E.
Silking, Flint, 200 genotypes						
QTL	0.514	0.531	0.468	0.373	0.378	0.010
GBLUP	0.810	0.830	0.836	0.695	0.713	0.009
QGBLUP	0.806	0.822	0.829	0.680	0.698	0.010
RKHS	0.819	0.832	0.835	0.684	0.706	0.009
Tasseling, Flint, 200 genotypes						
QTL	0.231	0.399	0.328	0.286	0.263	0.017
GBLUP	0.813	0.824	0.832	0.669	0.684	0.009
QGBLUP	0.784	0.800	0.798	0.619	0.635	0.017
RKHS	0.819	0.828	0.834	0.665	0.682	0.009
Tasseling, Flint, 200 genotypes						
QTL	0.287	0.443	0.187	0.067	0.130	0.021
GBLUP	0.372	0.381	0.447	0.388	0.388	0.021
QGBLUP	0.334	0.383	0.373	0.224	0.284	0.021
RKHS	0.380	0.378	0.444	0.373	0.377	0.010

For the description of the training set construction methods U, SU, CD, S, and R see Table 1. S.E. indicates the mean standard error across methods.

Table 4. Predictive ability for Dent, using a training set size of 200 genotypes.

Model	U	SU	CD	S	R	S.E.
Silking, Dent, 200 genotypes						
QTL	0.461	0.471	0.396	0.409	0.367	0.008
GBLUP	0.822	0.820	0.818	0.698	0.744	0.007
QGBLUP	0.842	0.829	0.822	0.696	0.729	0.008
RKHS	0.818	0.814	0.805	0.621	0.678	0.007
Tasseling, Dent, 200 genotypes						
QTL	0.580	0.597	0.530	0.438	0.452	0.009
GBLUP	0.823	0.823	0.829	0.712	0.752	0.009
QGBLUP	0.839	0.832	0.826	0.707	0.741	0.009
RKHS	0.823	0.821	0.817	0.628	0.687	0.009
Tasseling, Dent, 200 genotypes						
QTL	0.416	0.395	0.403	0.241	0.300	0.009
GBLUP	0.649	0.677	0.674	0.567	0.650	0.007
QGBLUP	0.649	0.677	0.678	0.524	0.617	0.009
RKHS	0.621	0.646	0.655	0.523	0.603	0.007

For the description of the training set construction methods U, SU, CD, S, and R see Table 1. SE indicates the mean standard error across methods.

Table 5. Predictive ability for wheat, using a training set size of 100 genotypes

Model	U	SU	CD	S	R	S.E.
Wheat, heading, 100 genotypes						
QTL	0.303	0.301	0.336	0.382	0.351	0.009
GBLUP	0.472	0.472	0.474	0.357	0.371	0.009
QGBLUP	0.512	0.519	0.562	0.517	0.478	0.009
RKHS	0.632	0.611	0.592	0.421	0.419	0.009
Wheat, yield, 100 genotypes						
GBLUP	0.660	0.650	0.620	0.475	0.482	0.009
RKHS	0.699	0.679	0.654	0.538	0.517	0.009

For the description of the training set construction methods U, SU, CD, S, and R see Table 1. SE indicates the mean standard error across methods.

Table 6. Predictive ability for rice, using a training set size of 300 genotypes

Model	U	SU	CD	S	R	S.E.
Silking, Rice, 300 genotypes						
QTL	0.309	0.320	0.303	0.271	0.267	0.013
GBLUP	0.778	0.779	0.751	0.676	0.657	0.013
QGBLUP	0.766	0.770	0.728	0.673	0.653	0.013
RKHS	0.815	0.816	0.787	0.699	0.677	0.013
Tasseling, Rice, 300 genotypes						
QTL	0.379	0.379	0.301	0.361	0.366	0.014
GBLUP	0.759	0.756	0.805	0.804	0.800	0.011
QGBLUP	0.740	0.738	0.801	0.806	0.801	0.011
RKHS	0.785	0.779	0.806	0.790	0.788	0.011
Tasseling, Rice, 300 genotypes						
QTL	0.231	0.223	0.275	0.191	0.191	0.019
GBLUP	0.556	0.554	0.638	0.580	0.571	0.013
QGBLUP	0.479	0.467	0.582	0.515	0.519	0.019
RKHS	0.603	0.599	0.671	0.589	0.579	0.013

For the description of the training set construction methods U, SU, CD, S, and R see Table 1. SE indicates the mean standard error across methods.

Table 7. Predictive ability within groups for Flint silking date, using a training set size of 200 genotypes

Model	U	SU	CD	S	R	S.E.
QTL						
a	0.073	0.104	0.140	0.043	0.232	0.052
b	0.775	0.893	0.663	0.686	0.648	0.061
c	0.803	0.446	0.761	0.347	0.439	0.034
d	0.603	0.797	0.641	0.680	0.622	0.053
e	0.371	0.191	0.258	0.121	0.182	0.023
GBLUP						
a	0.485	0.588	0.485	0.395	0.577	0.031
b	0.625	0.867	0.656	0.579	0.638	0.039
c	0.850	0.331	0.908	0.449	0.575	0.028
d	0.802	0.726	0.860	0.501	0.534	0.043
e	0.666	0.563	0.727	0.634	0.563	0.019
QGBLUP						
a	0.452	0.597	0.509	0.42	0.552	0.038
b	0.611	0.867	0.625	0.654	0.647	0.048
c	0.864	0.402	0.899	0.489	0.561	0.034
d	0.705	0.736	0.803	0.631	0.574	0.053
e	0.737	0.603	0.714	0.523	0.512	0.023
RKHS						
a	0.578	0.576	0.519	0.266	0.554	0.031
b	0.625	0.959	0.627	0.554	0.629	0.039
c	0.807	0.354	0.877	0.523	0.609	0.028
d	0.760	0.753	0.859	0.509	0.582	0.043
e	0.732	0.559	0.732	0.633	0.554	0.019

For the description of the training set construction methods U, SU, CD, S, and R see Table 1. SE indicates the mean standard error across methods.

Predictive ability in the validation set, calculated within subpopulations:

We present predictive ability as calculated within subpopulations for the Flint, Dent, and rice panel. The wheat data were not included in this analysis because the panel was too small, and predictive ability within subpopulations could not be calculated reliably. Within subpopulations, training set construction methods generally maintained their ranking, compared to predictive ability calculated across subpopulations; U, SU, and CD were better than S and R (Table 7, Table S5, Table S6, Table S7, Table S8, and Table S9). This indicates that the improvement in predictive ability observed for U, SU, and CD was not driven by the subpopulations. This result can also be observed in the correlation plot between predicted and observed phenotypes. Figure 6 shows that the relation between predicted and observed trait values was similar within subpopulations and across subpopulations, demonstrating that predictive ability was not driven by population structure. For the rice data, predictive ability within subpopulations was similar for all the training set construction methods, coinciding

with the result observed for the predictive ability across subpopulations (Table S10, Table S11, and Table S12).

For all the panels, the ranking of prediction models with respect to within subpopulation predictive abilities coincided with that for across subpopulations; RKHS, GBLUP, and QGBLUP were similar (with minor differences in the ranking, depending on the panel), whereas the QTL model led to clearly lower predictive ability.

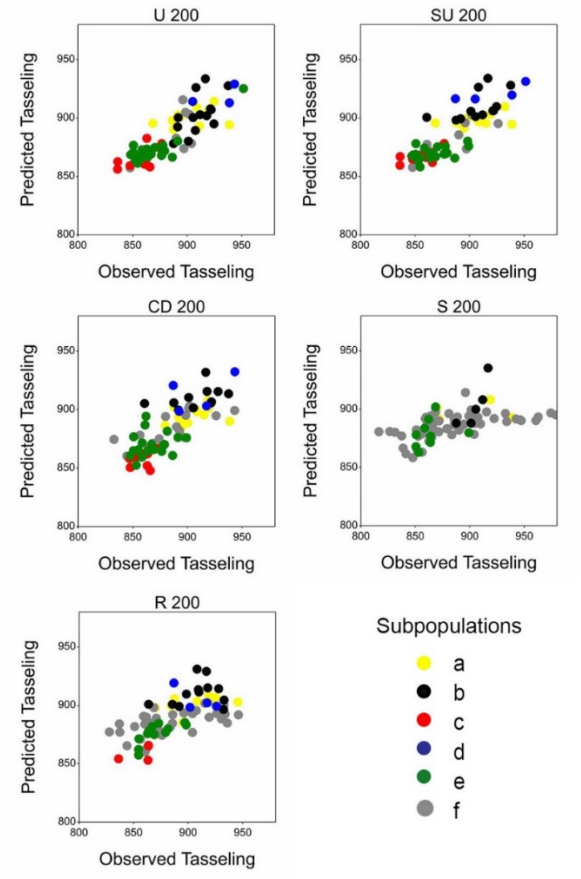


Figure 6. Relation between predicted and observed tasseling date for the Dent panel using the RKHS model and 200 genotypes. A single training set realization is shown for each training set construction method. Symbol colour represents each of the six subpopulations.

3.4. Discussion

The main objective of this study was to assess the impact of five training set construction methods (U, SU, CD, S, and R) on predictive ability in the validation set. A secondary objective was to compare four prediction models that differ in the importance that they assign to specific genomic regions and in the type of genetic effects that they consider (additive/non-additive). The training set construction methods and prediction models were evaluated at different training set sizes in four diversity panels. Predictive ability was calculated for the validation set in all the panels.

3.4.1. Training set construction methods

Prediction of unobserved genotypes is possible provided that genotypes to be predicted are genetically similar to those that have been observed (Habier et al. 2010; Saatchi et al. 2011). Hence, a prerequisite to obtain large predictive ability is that the training set represents well the calibration set and that the calibration set represents well the TPG (Rincent et al. 2012; Crossa et al. 2013; Albrecht et al. 2014; Auinger et al. 2016). Breeding populations are commonly structured. When population structure is present, genetic similarity is heterogeneous, because pairs of genotypes can belong to the same or different subpopulations. Random sampling from the calibration set reproduces its distributional properties without taking into account diversity differences across the genetic space (Jansen and van Hintum 2007). Thus, in structured populations, simple random sampling will not result in training sets that adequately represent the full genetic variation in the calibration set, leading to on average lower similarity between genotypes in the training and the validation set (Pszczola et al. 2012; Albrecht et al. 2011; Wientjes et al. 2016).

We showed that a more homogeneous coverage of the genetic space by applying the methods U and SU leads to smaller distances between genotypes in training and validation sets, and to a higher predictive ability. A uniformly covered genetic space also offers the potential to provide good predictive ability for new genotypes not belonging to the initial calibration set, provided that they are contained within the genetic space spanned by the initial calibration set.

Rincent et al. (2012) proposed to increase predictive ability by maximizing the precision of the contrast between each individual in the validation set and the mean of the calibration set (training and validation sets). This method was also successfully applied to genomic prediction in pea (Tayeh et al. 2015). Here, we show that CD, U, and SU are alternative methods that deliver comparable results because they all provide a training set that has a smaller

genetic distance to the validation set. One of the advantages of U and SU is that no estimate of heritability is required. Thus, it resolves the unavoidable ambiguity when defining a training set for multiple traits with different heritabilities. A second advantage is that U and SU showed more consistency of training set sample properties revealed by a narrower distribution of distances between the validation and the training set, compared to CD, S, and R. The genotypes in the training set are at more constant distances, providing a more uniform coverage of the genetic space and larger predictive ability, even when the distribution of genotypic distances in the validation set is different from that in the training set. Furthermore, U and SU have the advantage that they are computationally easier and faster to apply than CD.

U, SU, and CD are methods that use genetic similarity/distance as a criterion to construct the training set. Thus, the set of markers used for distance calculation influences training set composition. One aspect that could be further explored is the convenience of considering only those genomic regions that influence the trait of interest, especially for traits regulated by a small number of loci. In the same vein, the presence of ascertainment bias in the marker set needs to be evaluated because it might modify the relative distances among genotypes, and, therefore, the training set composition. For that reason, we repeated all calculations for maize, using the full SNP50 BeadChip in place of the PANZEA marker set (results not shown). The relative distances among genotypes were highly comparable between those two marker sets (Frascaroli et al. 2012) and therefore we did not observe changes in the ranking of training set construction methods or prediction models for predictive ability.

3.4.2. Prediction models

The main difference among prediction models is the relative importance assigned to specific loci as contrasted with the rest of the genome. It is therefore natural to expect that the degree of success of the different models depends on trait genetic architecture. This study dealt with yield, yield components (regulated by many loci with small effects), and with phenology traits. In the case of wheat, flowering time is regulated mainly by a few loci with large effect. However, despite the apparently simple genetic regulation of heading date favoring a QTL model, it is still beneficial to include a term that accounts for residual genetic variance. This result is in line with Zheng et al. (2013), who showed that flowering time in wheat is not only regulated by major genes for photoperiod and vernalization requirements, but also by a polygenic effect that influences earliness per se. In contrast, in the case of maize and rice, phenology and yield traits are regulated by many QTL (Buckler et al. 2009; Rincent et al. 2014b; Zhao et al. 2011). The more complex

genetic architecture of maize and rice traits is in agreement with our findings of models using genome-wide information showing larger predictive ability than those using information from a few QTL (QTL prediction model).

The importance of considering trait genetic architecture when selecting the prediction model was also discussed by Daetwyler et al. (2010) and by Bernardo (2014), who simulated diverse traits that differed in the number of QTL explaining the genotypic variance. The authors observed that traits regulated by a small number of QTL tend to be predicted better by models that give a larger importance to QTL with large effects, compared to the GBLUP model. This result has also been observed for a set of human diseases regulated by few loci with different effect size, for which it was advantageous to include several random terms (Speed and Balding 2014). We are aware that the number of QTL included in our QGBLUP models contains an element of subjectivity because of the selection of a significance threshold to define when a locus enters the QTL list. Bernardo (2014) gave some guidelines about when to include the QTL in a separate model term.

Previous paragraphs discussed the convenience of separately ac- counting for additive loci, depending on their effect size. However, part of the genetic variance might be non-additive. If the epistasis is simple (interaction between a few loci with large effects), it can be modeled as a QTL-interaction term (Malosetti et al. 2011). Unfortunately, in the case of the traits analyzed here, epistasis has been shown to be largely complex (Reif et al. 2011; Kippes et al. 2014). Langer et al. (2014) showed that epistasis for heading date in wheat can be dissected into at least 30 epistatic interactions, among which many of them did not correspond to interaction between large phenology genes. The results shown by Langer et al. (2014) coincide with the lack of improvement in predictive ability that we observed when we incorporated additional terms accounting for interaction among large phenology genes (results not shown). The RKHS model allows to account for epistatic interactions, without the need of specifying which genomic regions are responsible for this interaction (Crossa et al. 2010, 2013; Gianola and van Kaam 2008; Jiang and Reif 2015).

Traits and crops might also differ in the relative size of epistatic interactions (Langer et al. 2014; Reif et al. 2011; Spindel et al. 2015; Blanc et al. 2006). For example, a larger improvement was observed with the RKHS model for wheat data than for maize and rice. This result coincides with those of Endelman (2011) and Stange et al. (2013), who observed that the advantage of the RKHS model was large in the case of wheat grain yield, but it was small in the case of maize traits. A further issue that needs to be considered in structured populations is the convenience of assuming constant or heterogeneous allele effects across subpopulations (Lehermeier et al. 2015; de los Campos et al. 2015). Models that allow for

subpopulation-specific allele effects range from models that assume fully independent populations (effects estimated in each population separately), to more complex models that allow allele effects to be correlated across subpopulations (Lehermeier et al. 2015; Olson et al. 2012). In this paper, we focused on models that assume homogeneous effects. We also explored the idea of allowing for subpopulation-specific effects by fitting all the models to each subpopulation independently (not shown). However, models that allow for subpopulation-specific effects did not show a clear advantage over models with homogeneous effects, coinciding with Lehermeier et al. (2015), Schulz-Streeck et al. (2012), and Albrecht et al. (2011).

3.4.3. Sample size

Sample size reduction inevitably leads to a larger probability of losing genotypes with extreme values for the trait of interest, thereby narrowing down the phenotypic trait range and the predictive ability. Our results showed a nonlinear decrease in predictive ability as a function of training set size. This nonlinear decrease of the predictive ability was also observed by Heffner et al. (2011), Zhao et al. (2012), and Rincent et al. (2012) and can be explained by the number of individuals, trait heritability, and the effective number of chromosome segments (Daetwyler et al. 2008, 2013). When assessing the sampling methods in relation to sample size,

U produced a more homogeneous representation of the genetic diversity of the original population, compared to S and R, leading to larger predictive ability. The fact that this advantage was maintained only at large sample sizes can be explained by the fact that, at smaller training set sizes, none of the training sets was able to provide enough information for an accurate estimation of genotypic effects.

Conclusions

Training set construction methods that take into account the genetic diversity of the calibration set have higher predictive ability and are not sensitive to population structure in the calibration set: U, SU, and CD vs. S and R.

U and SU and CD produce comparable predictive abilities, but U and SU are simpler to calculate and require less computational cost and no phenotypic information in comparison to CD.

As expected, training sample size reduction led to lower predictive ability, but this reduction was stronger for the wheat and maize panels than for the rice panel.

Acknowledgements

We thank Matthieu Bogard (INRA), who kindly imputed the Australian wheat SNPs, and Renaud Rincent and Hans Jansen for helping us to implement the CD and U methods. We appreciate the comments of the editor and the anonymous reviewers who contributed to improve this paper. We also thank the Department of Agriculture and Food, Western Australia Agronomy and Breeding teams, and Australian Grain Technologies for the Australian wheat data. D.B.K. contributed to this research thanks to a Ph.D. scholarship from Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), Gobierno de Chile.

Supplementary Material

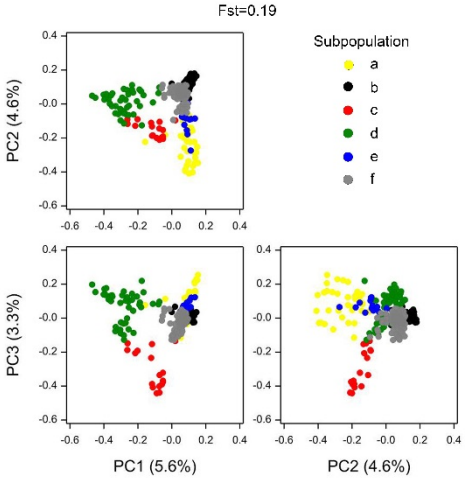


Figure S1. Representation of principal components analysis on the IBS matrix of the Dent panel.

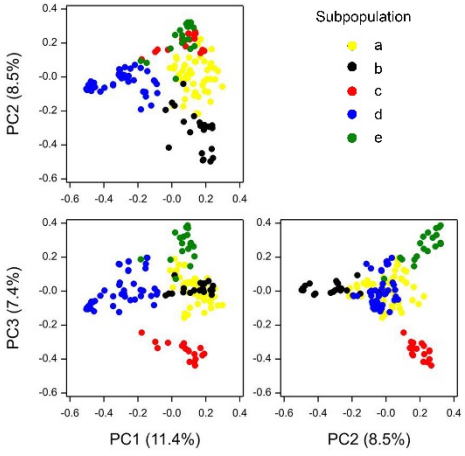


Figure S2. Representation of principal components analysis on the IBS matrix of the wheat panel.

Table S1. Median of the distance (1-IBS) between validation set genotypes and the closest genotype in the training set (over 100 sampling events).

Panel	Size	U	SU	CD	S	R
Flint	70	0.29	0.24	0.22	0.26	0.25
	200	0.13	0.19	0.12	0.23	0.21
Dent	70	0.28	0.27	0.21	0.27	0.25
	200	0.15	0.15	0.14	0.27	0.21
Wheat	50	0.19	0.19	0.13	0.19	0.19
	100	0.09	0.09	0.07	0.14	0.15
Rice	50	0.13	0.13	0.08	0.09	0.09
	300	0.02	0.02	0.02	0.06	0.06

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S2. Flint number of QTLs with a genome-wide significant threshold p<0.01 (Li and Ji (2005)). Columns 1-10 represent the linkage groups where the QTLs were located. Multiple QTLs can occur on a chromosome.

Size	Method	Tasseling										Silking										Yield									
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
50	U			1		2				1			1							1		9									
	SU			1			3	1		4						1		1		2					2					7	
	CD	10			7			2	3	2	2	16	4	1	9			3	3	2	7	2	3			1			1		
	S	9	4	9	10	3	3	4	7	2	2	13	9	8	6	2	2	2	8	3	2	3			5	2					3
	R	12	7	5	12	5	3	8	8	5	2	13	6	7	13	5	2	4	7	4	2	2		1	4					1	1
70	U		3			10			1				8				1					2	1								
	SU			4		4		1	1				6	3		1		1	2	1		13								1	
	CD	11	2	4		1	1	3	6	1	5	5	2	2	5			1	6		8	1	4		1	3	1	1	2		1
	S	19	1	1	6	1	2	2	3	9	1	20	1	14	5	3	2	2	2	7	3	9		2	9				1	1	1
	R	17	2	1	6	2	2	4	3	2	3	15	4	2	7	1	3	2	5	1	2	5			4		2				3
100	U		8			1		3	1	4			17					7	1	9		20									
	SU		1					3					7	2				3		2		19									1
	CD	9		4	7				3		2	11	1	5	5				2			8	3		1			1			3
	S	23	3	2	4		2	3	1		3	28	4	2	3		1	2		1	1	25	1	1	4	2		1		2	
	R	10	4	12	3	3		1	3	9	3	15	4	13	5	1		1	4	5	3	10	2		3					2	
150	U	6		6				1		5		25	1	6				9		15		4		2					1		
	SU	33		2					1			84	2	2				5		2		14								2	
	CD	12		3	5				1			15		4	2				1	1		21		2							
	S	19	2	1	3			2	4	1	1	31	3		2		2		2	1	2	7			5		1		2	1	
	R	22	2	2	3			1		2		17	4	3	2		1	1	2	1		18	1	1	3	2		1			
200	U	7		14								75		2						3		26									
	SU	28		12								100										58									
	CD	9		9								53		9				2		4		10								1	
	S	49	3	2						6		83	8	2	1				1	7		13			8					1	
	R	25	1	2	6			2		2		48	5	3	3		1			4	1	42	3		5					1	

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S3. Dent number of QTLs with a genome-wide significant threshold p<0.01 (Li and Ji (2005)). Columns 1-10 represent the linkage groups where the QTLs were located. Multiple QTLs can occur on a chromosome.

Size	Method	Tasseling										Silking										Yield									
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
50	U		2	4	1				1		1		1	11	1		1	2		1	1				1			1	1			
	SU	1	1	6	2	1	4		3		1		1	14	1	1	3		3				3		4		1			1	1
	CD	1	2		6	5		1		2	9	1	5		3	2			5							9			2	2	
	S	2	1	4	6	7		3	2		1	3		4	5	4	3		4	1	1	6	1	1	6		3	4	2	4	8
	R	7	2	3	3	11		2	5	1	1	2	2	2	2	7		1	5	1		4	3	3	3	14			2	2	1
70	U			8				3	6					5				6	2			1	2		1			1			
	SU	2		11			1		3			3	2	20				1					3		8	4	7	1			
	CD	2		4	2	12	3	3	10		7	2	1	2	6	2	5	4	17		2	1	2	1	3	5	1		3		2
	S	1	1	7	3	4	4		13			2	1	6	3	7	7		13	2		2	2	1	6	8	1	8	4	5	27
	R	5		6	6	10		2	12	2	2	1	3	7	5	6	2	3	12	2	2	4	9	3	5	1	1	2	7		1
100	U		2	14		1			29					3					8			1			4	4	10				
	SU			37			2		45		1			26			1		27		1				15	1	1				
	CD	1	1	26	3	8	4		25		2		3	15		3	3	2	30				8		2	6	1		2		1
	S	5		4	4	11	4	4	12	2	8	3	1	3	1	3	2	5	3	1	2	2	1	1	3	1			2	11	1
	R	3	4	10	5	17	7	3	36		1	1		5	7	9	10	2	37	1	2	1	13	1		5	2		2	2	2
150	U			92		7	2		98					40					77						1	7					
	SU	1		74	1	8	2		53	1			4	30	1	1	1		38				1			20					
	CD			64	4	4	3		71		4	1		17	2				57		2	1			6	33			1		1
	S	1		16	9	24	3	3	64	1	1	5	3	6	3	7	4	1	38		1		2	1	2	11			2	4	2
	R		1	13	7	12	1	15	66	1	2	1	3	4	8	2		10	57		2	1	1		3	15	1		5	6	3
200	U			54					100					1					99							97					
	SU			60					99					5					97							95					
	CD			52	3	5	1	11	94					3	1			8	81						1	69					
	S		2	19	16	8	1	14	96	1		1	10	4	6	1		10	82						5	36			14	1	
	R	1	2	25	10	4	3	11	88			2	6	2	8	3		8	81		1	2	2		3	32			5	6	

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S4. Rice number of QTLs with a genome-wide significant threshold p<0.01 (Li and Ji (2005)). Columns 1-12 represent the linkage groups where the QTLs were located. Multiple QTLs can occur on a chromosome.

Size	Method	Flowering												Height												Seed nr											
		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
50	U					2								29		1			2						2	2	1	1	2	1							1
	SU	1				5			1					26		2							1				5	4			1						1
	CD	2		6	1	5	1							6	4	2	1	1	1		4					2	2	2	2		2	1	1			1	5
	S	2	2	3	2		3	2	2	5		1	1	10	1	1		1	1		1		2			3	3	2	5		3	6		2	2		1
	R	6	3	1	5	1	6	2	5		1	2	3	7			1	2	1		1			1		4	4	1	5	3	3	4			1	1	4
100	U			2		1	1						6	96	17			2			6		2	6	5		13	4	2	1		6	1				
	SU			1		1	1	1				2	10	96	22			1			3		2	7	16		10	3	2	2		9			1		
	CD	4		5	4	2	2		1	2		2		47	1	1		3	2		6				2		1	2		3		1				7
	S	7	8	5	3	3	11	3	6	7	2	2		50	2	2	1	2	4			3	1		1	1	2	1	3	1	5	6	5	2	1	3	4
	R	6	1	4	2	3	7	2	2	7		1		49	5	4		4	4		1	3		4		2		5	4		4	4			1		4
150	U		3	81	1		1			1				106	8				5			6				10				9			1		1	
	SU		4	83				1		2				106	10				5	1		11			2	12				4							
	CD	5	1	37			1		3	6			1	93	3				2			3				8	1			2	1		2			2
	S	7	9	11	2	5	29	2	10	8		3	2	86	9	1		1	3		2				3	5	3	6	9	1	7	4	2		1	2	9
	R	6	5	12	2	4	20	1	13	7	2	4		90	9	1		3	2		1	2		3		3		2	5		14	7	3	2		1	13
200	U			100					4					109					14			9			3	3											
	SU	5		100			1		3					111					16			6			1	1		4									
	CD	1	1	100	1	1			3	3			2	102	11				3	1		5			1	21		1	4		23	1					29
	S	10	17	13	1	5	57	1	17	8		5	2	104	11	4		3	4	3	1	2		1	1	9	2	2	7		7	8		2		6	21
	R	3	2	17	1	1	44	1	10	2		2	1	96	10	1		2	11	2		3			1	15		4	2		6	8				1	21
300	U			100			45		2	1		3		100					40									4	7		2					14
	SU			100			60		5			1		100					46									2	14		6						12
	CD		2	100		1	36		18			4		101	13				22							21		1	4		23	1					29
	S	1	1	69			78		13	1				106	16			2	10					1		8		4	1		16	9					50
	R	1	2	75			78		19	2				114	22			1	9	1		2		3		10		3			18	9					47

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S5. Flint Tasseling date predictive ability within groups using a training set size of 200 genotypes.

			QTL			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.141	0.180	0.185	0.145	0.127	0.091
b	0.627	0.832	0.615	0.663	0.684	0.099
c	0.366	0.368	0.535	0.238	0.271	0.058
d	0.272	0.523	0.424	0.711	0.496	0.098
e	0.214	0.106	0.215	0.008	0.063	0.045
			GBLUP			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.316	0.228	0.357	0.269	0.427	0.034
b	0.615	0.849	0.620	0.594	0.649	0.037
c	0.875	0.290	0.899	0.307	0.484	0.026
d	0.802	0.841	0.879	0.526	0.577	0.047
e	0.821	0.555	0.804	0.607	0.548	0.022
			QGBLUP			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.302	0.309	0.372	0.303	0.324	0.071
b	0.608	0.796	0.559	0.632	0.649	0.077
c	0.881	0.359	0.882	0.331	0.432	0.055
d	0.774	0.824	0.769	0.635	0.495	0.098
e	0.847	0.518	0.743	0.423	0.450	0.045
			RKHS			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.460	0.315	0.426	0.239	0.431	0.034
b	0.654	0.944	0.650	0.593	0.655	0.037
c	0.826	0.308	0.856	0.402	0.529	0.026
d	0.771	0.843	0.887	0.539	0.614	0.047
e	0.847	0.562	0.786	0.608	0.540	0.022

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S6. Flint yield predictive ability within groups using a training set size of 200 genotypes.

			QTL			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	-0.114	-0.003	-0.146	-0.248	-0.048	0.135
b	0.521	0.015	0.328	0.419	0.473	0.194
c	0.030	0.399	0.130	-0.071	0.061	0.136
d	0.628	0.838	0.638	0.666	0.708	0.102
e	0.555	-0.124	0.331	0.043	0.009	0.038
			GBLUP			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.131	0.279	0.238	0.117	0.372	0.040
b	0.564	0.094	0.516	0.401	0.314	0.053
c	0.403	0.195	0.371	0.236	0.270	0.028
d	0.672	0.619	0.783	0.603	0.600	0.039
e	0.702	0.531	0.662	0.457	0.442	0.016
			QGBLUP			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.022	0.158	0.137	-0.127	0.257	0.090
b	0.670	0.044	0.506	0.457	0.450	0.120
c	0.408	0.266	0.402	0.041	0.217	0.064
d	0.674	0.657	0.771	0.708	0.676	0.089
e	0.687	0.361	0.521	0.183	0.170	0.037
			RKHS			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.142	0.337	0.290	0.245	0.392	0.040
b	0.447	0.037	0.408	0.360	0.271	0.053
c	0.382	0.240	0.342	0.259	0.284	0.028
d	0.708	0.680	0.818	0.679	0.688	0.039
e	0.745	0.451	0.666	0.410	0.389	0.016

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S7. Dent silking date predictive ability within groups using a training set size of 150 genotypes.

			QTL			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.049	0.060	-0.057	-0.063	-0.031	0.080
b	0.044	-0.072	0.060	0.040	0.090	0.025
c	0.113	0.020	0.213	0.210	0.243	0.028
d	0.179	0.079	0.124	0.212	0.276	0.116
e	0.387	0.212	0.280	0.112	0.161	0.035
			GBLUP			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.512	0.532	0.364	0.327	0.376	0.051
b	0.610	0.554	0.597	0.706	0.649	0.019
c	0.526	0.458	0.478	0.410	0.410	0.021
d	0.549	0.541	0.517	0.630	0.521	0.036
e	0.772	0.707	0.681	0.683	0.671	0.025
			QGBLUP			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.488	0.570	0.287	0.110	0.108	0.071
b	0.443	0.380	0.431	0.549	0.510	0.025
c	0.532	0.426	0.539	0.462	0.494	0.028
d	0.522	0.503	0.452	0.688	0.547	0.047
e	0.700	0.652	0.609	0.632	0.612	0.032
			RKHS			
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.492	0.465	0.345	0.131	0.158	0.051
b	0.641	0.537	0.602	0.612	0.573	0.019
c	0.477	0.415	0.412	0.356	0.344	0.021
d	0.618	0.610	0.574	0.600	0.531	0.036
e	0.757	0.699	0.664	0.638	0.650	0.025

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S8. Dent tasseling date predictive ability within groups using a training set size of 150 genotypes.

Dent, Tasseling date, 150 genotypes						
QTL						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.308	0.207	0.236	0.164	0.094	0.019
b	0.201	0.140	0.267	0.290	0.337	0.023
c	0.058	-0.075	0.105	0.162	0.204	0.078
d	0.544	0.435	0.443	0.224	0.221	0.028
e	0.385	0.439	0.225	0.094	0.207	0.067
f	0.513	0.510	0.472	-	0.374	0.014
GBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.554	0.543	0.511	0.640	0.581	0.018
b	0.445	0.390	0.444	0.410	0.420	0.021
c	0.547	0.574	0.615	0.632	0.553	0.039
d	0.806	0.743	0.718	0.721	0.703	0.025
e	0.595	0.639	0.593	0.486	0.475	0.046
f	0.812	0.817	0.771	-	0.709	0.013
QGBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.466	0.432	0.414	0.528	0.423	0.019
b	0.530	0.446	0.531	0.518	0.551	0.023
c	0.490	0.508	0.549	0.585	0.536	0.042
d	0.730	0.671	0.658	0.647	0.638	0.028
e	0.654	0.721	0.558	0.328	0.473	0.062
f	0.689	0.729	0.704	-	0.647	0.014
RKHS						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.580	0.558	0.520	0.541	0.501	0.018
b	0.409	0.361	0.368	0.367	0.358	0.021
c	0.618	0.656	0.638	0.598	0.565	0.039
d	0.796	0.735	0.708	0.678	0.680	0.025
e	0.630	0.626	0.614	0.318	0.229	0.046
f	0.759	0.775	0.703	-	0.626	0.013

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S9. Dent yield predictive ability within groups using a training set size of 150 genotypes.

Dent, Yield, 150 genotypes						
QTL						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.055	0.275	0.067	0.394	0.045	0.101
b	0.368	0.305	0.266	0.149	0.215	0.040
c	0.299	-0.06	0.361	0.365	0.286	0.133
d	0.424	0.488	0.399	0.378	0.274	0.047
e	0.391	0.455	0.346	0.121	0.331	0.044
f	0.249	0.189	0.312	-	0.202	0.027
GBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	-0.193	-0.144	-0.026	0.285	0.037	0.042
b	0.380	0.352	0.434	0.359	0.364	0.017
c	0.385	-0.010	0.544	0.608	0.586	0.053
d	0.578	0.501	0.404	0.477	0.441	0.020
e	0.669	0.688	0.676	0.654	0.650	0.019
f	0.574	0.547	0.539	-	0.493	0.012
QGBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	-0.005	0.213	0.072	0.518	0.141	0.096
b	0.478	0.387	0.445	0.321	0.365	0.039
c	0.486	-0.128	0.489	0.516	0.401	0.120
d	0.644	0.604	0.464	0.479	0.462	0.046
e	0.543	0.575	0.609	0.514	0.624	0.044
f	0.604	0.537	0.537	-	0.414	0.027
RKHS						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	-0.103	-0.080	-0.06	0.268	0.032	0.042
b	0.369	0.340	0.385	0.314	0.311	0.017
c	0.393	0.016	0.625	0.572	0.538	0.053
d	0.438	0.423	0.334	0.418	0.370	0.020
e	0.581	0.588	0.575	0.576	0.564	0.019
f	0.557	0.529	0.522	-	0.455	0.012

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S10. Rice height predictive ability within groups using a training set size of 300 genotypes. For the description of the training set construction methods U, SU, CD, S and R see Table 1.

Plant height, rice, 300 genotypes						
QTL						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.216	0.219	0.181	0.160	0.139	0.038
b	-	0.099	-0.356	-	0.148	0.044
GBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.758	0.759	0.717	0.707	0.683	0.015
b	0.973	0.479	0.894	0.974	0.481	0.036
QGBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.741	0.744	0.683	0.691	0.670	0.015
b	0.971	0.501	0.889	0.970	0.518	0.036
RKHS						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.795	0.795	0.753	0.725	0.695	0.015
b	0.963	0.534	0.924	0.963	0.526	0.036

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Table S11. Rice Flowering date predictive ability within groups using a training set size of 300 genotypes.

Flowering, rice, 300 genotypes						
QTL						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.219	0.253	0.396	0.326	0.314	0.019
b	0.894	0.897	0.813	0.690	0.707	0.115
GBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.706	0.702	0.730	0.762	0.741	0.016
b	0.982	0.981	0.939	0.767	0.794	0.028
QGBLUP						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.685	0.680	0.720	0.752	0.728	0.016
b	0.981	0.980	0.944	0.789	0.812	0.028
RKHS						
Subpop.	**U**	**SU**	**CD**	**S**	**R**	**s.e.**
a	0.737	0.730	0.731	0.753	0.736	0.016
b	0.983	0.983	0.943	0.752	0.776	0.028

For the description of the training set construction methods U, SU, CD, S and R see Table 1. SE indicates the mean standard error across methods.

Chapter 4

Identifying regions in multi-environment trials by bilinear and mixed models: a case study of yield in wheat for North-Western Europe

Daniela Bustos-Korts[1,2], Fred A. van Eeuwijk[1], Bertrand E. Schuiling[3], Marcos Malosetti[1]

1. Biometris, Wageningen University & Research Centre

2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)

3. Wiersum Plant Breeding

This Chapter is under review in Crop Science

Abstract

Wheat is adapted to a wide range of environmental conditions and it often shows crossover genotype-by-environment interactions (GxE). To select adapted genotypes, breeders perform multi-environment trials (MET) that aim at representing the target population of environments (TPE). Response to selection can be improved by subdividing the TPE in more homogeneous parts, where ranking of genotypes within subdivisions exhibits higher consistency than across subdivisions. For subdividing the TPE, series of METs can be analysed, concentrating on repeatable genotype-by-location interactions and group locations into regions Various approaches based on fixed linear-bilinear models are popular to group trials in METs (SHMM, AMMI, GGE). Mixed model approaches for the identification of groups of trials with a higher internal homogeneity are potentially powerful alternatives, although in general they require a higher ingenuity in statistical modelling than standard fixed models do. We compare here two strategies for grouping trial locations into regions, one based on a full mixed model analysis, and one based on a relatively simple, yet robust two-step approach based fitting AMMI models to within year genotype by location tables of means. The AMMI predictions are then used to cluster locations within years. Consistent clustering of locations over years is used to assign locations to regions. The mixed model approach uses the parameters of a factor analytic model to classify locations in regions. The approaches are illustrated on yield data from official variety trials from 1995 to 2012, in Denmark, Germany, the Netherlands and the United Kingdom. We identified regions in Denmark, Germany and the United Kingdom that coincided with latitudinal and longitudinal gradients. Regions were most outspoken in Denmark.

Abbreviations

AMMI, Additive main effects and multiplicative interaction model
ANOVA, analysis of variance
FA1, factor analytic model of order 1
FA2, factor analytic model of order 2
GGE, genotype main effects and genotype x environment interaction model
GxE, genotype by environment interaction
MET, multi-environment trials
LSD, least significant difference
PC, principal component
REML, restricted maximum likelihood
SHMM, Shifted Multiplicative Model
SVD; singular value decomposition.

4.1. Introduction

Wheat is a crop with a strong potential for adaptation, or adaptability (Trethowan et al. 2002; van Eeuwijk et al. 2016)). The environmental range in which wheat can grow is wide. For example, wheat is able to deal with the very dry conditions of the Sonora dessert (Fischer and Maurer 1978), and it tolerates extreme temperatures as low as -20°C in the vegetative phase as well as temperatures as high as 45°C during grain filling (Porter and Gawith 1999).

Because of its adaptability to favourable growing conditions wheat can reach yields above 15 Mg ha^{-1}, as observed in Southern Chile (Bustos et al. 2013; García et al. 2013). The strong adaptability of wheat together with the large environmental range makes wheat the largest contributor to the calories supply worldwide (FAO 2013). As a consequence of the same factors, strong adaptability and large environmental range, wheat is susceptible to strong genotype by environment interactions (GxE) including those that lead to rank changes (cross-overs). These genotypic rank changes complicate the selection of new varieties (Crossa et al. 2004; Reynolds et al. 2002).

To select varieties that are well adapted to target growing conditions, breeders perform multi-environment trials (METs). In METs, individual candidate varieties typically are evaluated at a number of locations across a limited set of years. The idea behind METs is that the trials in a MET form a representative sample of the current and near-future growing conditions, the target population of environments, or TPE. The growing conditions that define the TPE result from a combination of soil and meteorological parameters that are determined by location, say geography, latitude and longitude (climate, photoperiod), time (season, year, occurrence of stresses), and cultural practices (management, nutrients, pests, water). In the context of TPEs and METs, it is important to distinguish between repeatable and well predictable elements in the environmental conditions from those that are badly predictable. In general, environmental conditions associated with location and management are better predictable than those associated with year. A relevant consequence for breeders is that genotype by location and genotype by management interactions may be better predictable than genotype by year interactions or genotype by location by year and genotype by management by year interactions. The latter constitute sources of error variation, whereas the earlier represent GxE interactions that can be exploited.

For the analysis of MET data, mixed models are an appropriate class of models METs (Smith et al. 2005). Ignoring the experimental design structure of individual trials for the moment, for classical METs a factorial structure of genotypes, locations and years is defined and frequent model choices are to take all main effects random, or take all genotype related

terms random (Talbot 1984; Piepho and Möhring 2005; Smith et al. 2005; Atlin et al. 2011). Other choices in mixed models for METs aim at choosing repeatable and predictable terms fixed and non-repeatable terms random. For example, genotypes and location effects as well as their interactions may be taken fixed (Atlin et al 2011). The mixed models can easily be extended to include management related terms, in which case the genotype by management interactions will be fixed. As soon as interactions with years are involved, model terms will be chosen random because the idea is that those interactions will be unrepeatable and hard to predict.

When defining a TPE, it makes sense to choose environmental conditions within which relative few rank changes occur. Selection of the best yielding genotypes can then follow a standard best linear unbiased predictor (BLUP) protocol in which a shrinkage estimate is produced for the genotypic performance across the full set of trials taking into account the variance components for the different types of GxE interaction (Piepho and Möhring 2005; Smith et al. 2005; Piepho et al. 2008; Atlin et al. 2011). Although, a TPE may have been devised with the idea of a single homogeneous environmental continuum, in the sample of trials that is included in a MET, the realized genotypic responses may create doubt about the correctness of the sample of trials coming from a single indivisible TPE. If indeed strong heterogeneity occurs between groups of trials and the reason for this heterogeneity can be identified and appear to be predictable in the future, then it may be advisable to split up the original TPE in two or more TPEs and consider the sample of trials to come from various TPEs. Breeding efforts can then be directed at improving the performance for the newly defined TPEs that are a subset of the original undivided TPE.

Part of the analysis of the MET data can consist in checking the homogeneity of the included trials and decide whether to divide the initially targeted TPE into subsets that internally show increased homogeneity and strongly reduced number of crossovers. Preferably, decisions for subdividing the TPE are taken on the basis of the analysis of repeatable form of GxE interaction, like genotype by location and genotype by management interactions. The case that has received a lot of attention and on which we will concentrate in this paper is the grouping of locations into regions (Atlin et al 2000; Piepho and Möhring 2005, Atlin et al 2011). From a breeding point of view, ranking and selection decisions can be performed within-regions, which are supposedly internally more homogeneous increasing the heritability within regions (H^2), and the response to selection (Falconer and Mackay 1996). From a statistical modelling point of view, the original genotype-by-location interaction variation is partitioned into genotype by region and genotype by location within-region variation, where the first interaction is seen as repeatable and fixed, whereas the

second type of interaction stands for a random error term that can only be reduced by taking more locations per region.

Various approaches to group repeatable components of the environment have been proposed, where it needs to be emphasized that in some of the original papers little attention was given to the repeatability of the GxE interactions. The most popular approaches involve bilinear models, as the Shifted Multiplicative Model (SHMM, Cornelius et al., 1992), Genotypic main effects plus Genotype by Environment interaction models (GGE models, Yan et al., 2000) and Additive Main effects and Multiplicative Interaction models (AMMI models, Gauch Jr., 1988; Gauch and Zobel, 1997).

In the mixed model context repeatable GxE interactions can be chosen fixed and non-repeatable GxE interactions will be random. Grouping is then done on fixed interactions, using appropriate weighting schemes for information from individual genotypes and trials by defining a variance-covariance structure based on non-repeatable GxE interactions. Classification of the locations into regions uses the covariance of genotypes between pairs of locations (Atlin et al, 2000; Piepho and Möhring 2005; Burgueño et al 2008; Atlin et al 2011). These covariances can be estimated in various ways. By classical variance components models as in Atlin 2000, Piepho and Möhring 2005; Atlin et al 2011), or more parsimoniously by factor analytic models (FA, Smith et al., 2005). The FA variance covariance model is a multiplicative model, and in that sense it can be considered analogous to the multiplicative terms in the AMMI and GGE models (Piepho 1998; Smith et al. 2001). The magnitude of correlation (covariance) of genotypes between locations can be used to identify regions that are internally homogenous (Crossa et al. 2004; Burgueño et al. 2008; Beeck et al. 2010). Modelling the covariance structure across locations is often not straightforward and it requires a higher ingenuity in statistical modelling than standard fixed models do. A strategy to group locations based on approaches with less computational demands than in mixed models would be useful.

We propose an alternative approach to regionalize trials using an AMMI-fit-based clustering, where we fit AMMI models to genotype by location tables of genotypic means (Best Linear Unbiassed Estimators) to obtain improved genotypic means that serve as the input for a clustering procedure. We define similarity between trials within years as the Jaccard index calculated from performance indicators for genotypes that show whether the genotypes belong to the best ones (value 1) or not (value 0) in that trial. In contrast to current who-won-where proposals our method can use information of a set of highest ranking genotypes, instead of the single best genotype as proposed by (Yan et al. 2000; Gauch 2006).

As we use information of more genotypes than the highest ranking one, our procedure should result in a more robust regionalization.

The focus of our method is on the repeatable part of the GxE, so we group trials annually based on geography (genotype by location interaction), and subsequently assess the consistency of the obtained annual classifications across years. By using the AMMI fitted values instead of the original data we favour the grouping to be driven by patterns of GxE, since we consider only the first principal components (Odong et al. 2013).

Historical data from official wheat variety trials sown in Denmark, Germany, the Netherlands and in the United Kingdom were used in this paper to: (i) quantify GxE interactions on grain yield in North-European wheat trials, (ii) compare strategies to identify regions, AMMI model versus mixed model approach, and (iii) evaluate the resulting classification of locations into regions in terms of response to selection.

4.2. Methods

4.2.1. The study area and datasets

The data corresponded to genotypic yield means obtained from analyses of official trials for the assessment of Value for Cultivation and Use (VCU) in winter wheat for several locations in Denmark, Germany, the Netherlands and the United Kingdom between 1995 and 2012 (Figure1, Table 1). These data were compiled by the 3rd author of this paper. The data were unbalanced as set of trial locations changed from year to year, while, proper to the VCU testing system the set of genotypes changed over the years.

Figure 1. Geographical position of the field trials. Circles are proportional to the number of yield observations accumulated over the period 1995-2012.

Table 1. Institution that generated the data (data origin), years with data, number of locations and number of genotypes per country.

Country	Data origin	Period	Number of locations	Number of genotypes
Denmark	Videncentret for Landbrug	1995-2011	12	426
Germany	Bundessortenamt	1999-2012	17	1103
The Netherlands	Commissie samenstelling aanbevelende rassenlijst landbouwgewassen	2006-2011	10	163
United Kingdom	Agriculture and Horticulture Development Board	2003-2012	24	137

4.2.2. Mixed model analysis

Mixed models were fitted to estimate GxE variance components that will serve to quantify GxE interactions in the various countries. The variance components serve as a basis for establishing to which extent the grouping of locations in regions will lead to a larger response to selection for an evaluation system based on a undivided or a divided TPE following the procedure described by Atlin et al. (2000).

4.2.3. Variance components

The unbalanced genotype by location by year data per country was modelled as in Equation (1);

$$\underline{y}_{ijk} = \mu + L_j + Y_k + LY_{jk} + \underline{G}_i + \underline{GL}_{ij} + \underline{GY}_{ik} + \underline{GLY}_{ijk} \qquad (1)$$

In Equation (1) $\underline{y}_{ijk}$ is the phenotype (yield) of the i^{th} genotype in the j^{th} location and the k^{th} year (i=1,...*ng*; j=1,...*nl*, k=1,... *ny*). μ is the general intercept, L_j and Y_k are the fixed effects of location and year and LY_{jk} is the fixed interaction between location and year. $\underline{G}_i$ is the random main effect of the i^{th} genotype whereas $\underline{GL}_{ij}$ and $\underline{GY}_{ik}$ are the random effects of genotype by location interaction and genotype by year interaction, respectively. $\underline{GLY}_{ijk}$ corresponds to a residual term that contains genotype by location by year interaction and a within trial error. Random effects are assumed to be independent and normal with zero mean and proper variance: σ_g^2, σ_{gl}^2, σ_{gy}^2 and σ_{gly}^2. The residual term $\underline{GLY}_{ijk}$ in the Netherlands and in the United Kingdom was split into the variance of genotype by location by year and the error (σ_e^2). The error was estimated in the Netherlands and in the United Kingdom from the least significant difference (LSD) of the trials. Variance components were estimated by

restricted maximum likelihood (REML) implemented in ASReml-R (VSN-International 2013a).

Broad sense heritability on a plot (H^2_{plot}, Equation 2) and on a line mean basis (H^2_{line}, Equation 3) were calculated from the variance components. The error term σ_e^2 was only included in the Netherlands and in the United Kingdom. In Equation 2 and Equation3, nl, ny and nb represent the median of the number of locations, years and replicates in which genotypes were present, respectively.

$$H^2_{plot} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gl}^2 + \sigma_{gy}^2 + \left(\sigma_{gly}^2 + \sigma_e^2\right)} \qquad (2)$$

$$H^2_{line} = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gl}^2}{nl} + \frac{\sigma_{gy}^2}{ny} + \left(\frac{\sigma_{gly}^2}{nl\ ny} + \frac{\sigma_e^2}{nl\ ny\ nb}\right)} \qquad (3)$$

Variance components were also estimated for intervals of three consecutive years, which reflect the standard testing cycles (cultivars are typically tested for 3 years before release). For example, interval 1=1995, 1996, 1997; interval 2=1996, 1997, 1998 and so on, which resulted in 15 intervals for Denmark, 10 for Germany, 3 for the Netherlands and 8 for the United Kingdom. The median of the number of genotypes that occurred in the three years was calculated across intervals to give an impression of the amount of imbalance (Table 1). The logarithm of the variance of each term was regressed against mean yield of the different series to determine whether variance components were associated to the observed yield level.

Partition of genetic variance between and within regions

Regions were constructed in two ways (see section 2.3). For a given division of the TPE, the contribution of genotype by region interaction to the total variation was estimated to evaluate the efficiency of an evaluation system using regions versus a system with an undivided TPE. In Equation (4), the term genotype by location interaction ($\underline{GL_{ij}}$) of Equation (1) was partitioned into genotype by region interaction ($\underline{GR_{ir}}$) and genotype by location interaction nested within region $GL(R)_{ij(r)}$. The interaction of genotype by location by year ($\underline{GLY_{ijk}}$) was partitioned into genotype by region by year interaction ($\underline{GRY_{irk}}$) and genotype by location within region by year interaction, $GL(R)Y_{ij(r)k}$, following Atlin et al. (2000). A log-likelihood ratio test with two degrees of freedom was used to compare the goodness of fit of Equation (4) and (1) (Welham and Thompson 1997).

$$\underline{y_{ijk}} = \mu + L_j + Y_k + LY_{jk} + \underline{G_i} + \underline{GR_{ir}} + \underline{GL(R)_{ij(r)}} + \underline{GY_{ik}} + \underline{GRY_{irk}} + \underline{GL(R)Y_{ij(r)k}} \quad (4)$$

Modelling genotypic covariance between locations across years

With the random terms in Equation (4) each having their own variance, the variance-covariance structure for the $\underline{GR_{ir}}$ term is compound symmetry, like in Atlin et al. (2000) and in Piepho and Möhring (2005), although in this paper alternative structures are mentioned. To investigate patterns of genotype by location interaction and to classify locations into regions, we chose for the $\underline{GL_{ij}}$ term a factor analytic model of order 1 (FA1), with the genetic variance-covariance between locations $\Sigma_{GL} = \lambda_l \lambda_l' + \psi_l$. λ_l is a vector of location scores of length nl, and ψ_l a diagonal matrix of dimension nl with an location-specific lack of fit parameter. In addition, the genetic variance-covariance for the $\underline{GY_{ik}}$ term was modelled by a FA1 in Denmark, the Netherlands, and the United Kingdom ($\Sigma_{GY} = \lambda_y \lambda_y{}' + \psi_y$). A diagonal model $\Sigma_{GY} = D_y$ for the genetic variance-covariance between years was used in Germany (i.e. different variances for each year, and covariance between years equal to zero).

The vector λ_l of the FA1 structure for genotypes across locations was used to classify locations in regions, by a k-means clustering procedure (Dillon and Goldstein 1984; Zelterman 2015). Clusters were defined based on Euclidean distances, reducing the variability of individuals within a cluster, while maximizing the variability between clusters. The sum of the squared distances of cluster members to their cluster centroid was plotted as a function of the number of clusters to identify the final number of groups (Hastie et al. 2009).

4.2.4. AMMI-based clustering to identify groups of locations

A two-stage approach was used to identify consistent groups of locations across years. In the first stage, locations were classified within years on the basis of AMMI predictions, and the consistency of these groups over years was evaluated in the second stage.

Grouping within years: AMMI-based clustering of locations

MET data are usually balanced within years for genotypes and locations, but unbalanced across years as the locations can differ from year to year and new genotypes enter the VCU system while other genotypes are discarded and leave the system. Because our data were balanced within years, we could fit standard AMMI models to the within year genotype by location tables. AMMI predictions were obtained by fitting the following model in GenStat 16 (VSN-International 2013b);

$$\underline{y_{ij}} = \mu + G_i + L_j + \sum_{m=1}^{M} b_{im} z_{jm} + \underline{\varepsilon_{ij}} \quad (5)$$

In Equation (5) $\underline{y}_{ij}$ represents the mean yield of the i^{th} genotype in the j^{th} location as obtained from the trial analyses in the VCU system, μ stands for an intercept, G_i is the fixed effect of the i^{th} genotype and L_j is the fixed effect of the j^{th} location within a year. The interaction is explained by M multiplicative terms. Each multiplicative term formed by the product of a genotypic sensitivity b_{im} (genotypic score) and an environmental scores z_{jm}. Finally, $\underline{\varepsilon}_{ij}$ is a residual term. The number of principal components M to be retained was determined according to Gollob's test (1968).

Gauch (2006) suggested to classify locations that have the same highest yielding genotype as belonging to the same group of trials, which were called mega-environment. Grouping trials on the basis of the identity of the best performing genotype according to an AMMI fit in model (5), we will call strategy A. The groups of locations following from the coincidence of the best performing genotypes can be formed for each year. The grouping results across the years need to be considered to establish the consistency of the grouping.

We propose a more robust approach, called strategy B, in which regions are defined using a certain percentage, say 20%, of the best performing genotypes, again using the AMMI fitted values (Figure 2). Effectively, the fitted values from the AMMI model were replaced by binary vectors having the value 1 when a genotype is among the best and 0 otherwise (Figure2, steps 1 and 2). A similarity matrix of locations by locations in a particular year was constructed on these binary vectors using a Jaccard similarity (Giudici 2003) (step 3 in Figure 2). The resulting similarity matrix was used in a hierarchical clustering procedure (average linkage, (Johnson and Wichern 2007)) (step 4 in Figure 2). Locations were considered to be in the same cluster, or region, when their fitted similarity in the dendrogram was above 0.8. This process was repeated for each of the years, delivering a series of clusterings of locations. The consistency of these clusterings then remains to be assessed.

In strategy C, locations within years were clustered on Euclidean distances between locations as calculated from the AMMI fitted values across the full set of included genotypes. As in strategy B, the resulting similarity matrix was used in a hierarchical clustering of locations, by the agglomerative method group average (Johnson and Wichern 2007). A similarity of 0.8 fitted in the dendrogram was used as a threshold to define the groups within years. This cut-off level produced the best fit for a mixed model for the genotype by location means within a year with location fixed and random effects for genotype, genotype by region and genotype by location within region interactions.

Grouping across years

The second step of clustering aimed at identifying locations that tend to consistently belong to the same region across years. The input was an incidence table of region memberships for locations in particular years which had the dimensions of the number of locations by the number of years. This table was constructed by combining the vectors of region memberships obtained within years (step 5 in Figure 2). The entries of the table consist of region levels. Similarities between locations were defined by simple matching coefficients (Dillon and Goldstein 1984), and locations were then clustered by the agglomerative method group average (Steps 6 and 7 in Figure 2). The response to selection using the groups of locations suggested by the branching of the dendrogram to divide the TPE (MET) versus the response to selection of an undivided TPE (MET) was used as a criterion to decide the cut-off value in the clustering. The cut-off value that produced groups with the largest response to selection was used.

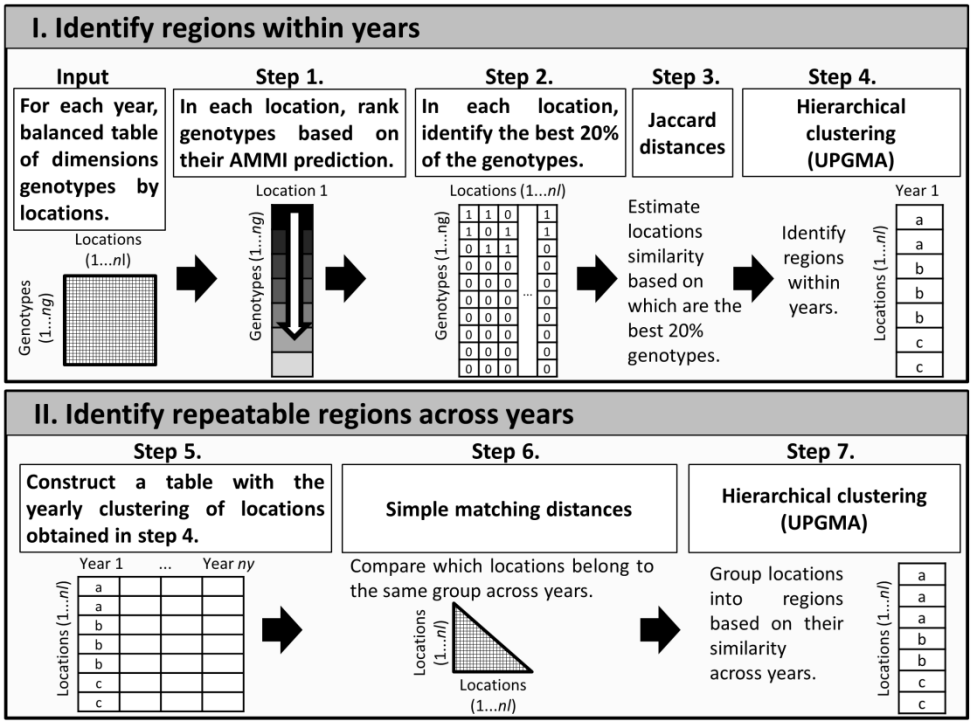


Figure 2. Steps in the strategy to identify regions based on clustering of AMMI predictions.

4.2.5. Evaluation of the regions

As Piepho and Möhring (2005) show, it can always be worthwhile to try to find divisions of the TPE provided that an appropriately chosen weighting regime is chosen to combine the BLUP estimators for the genetic values in the different parts of the subdivision. The utility and efficiency of groupings of trials in METs and thereby subdivisions of the TPE can be evaluated in a number of ways. We describe three of them in this section.

4.2.6. Overlap with agro-climatically attainable wheat yield map

The Global Agro-Ecological Zones (GAEZ) map (http://www.fao.org/nr/gaez/en/) was used as a reference to assess whether the regions resulting from our strategies coincided with those from the agro ecological wheat map. All the locations were georeferenced and superimposed on the GAEZ map for high input level winter wheat, by FAO/IIASA (2011). Different versions of the map are available and the one for a growth cycle length of 45+135 days was chosen because it resembles the cycle length of European wheat genotypes. Groups of locations were visually compared with the geographical pattern of the predicted yield using the software ESRI©Arc Map™ 10.1.

4.2.7. Effects of latitude and longitude

Coordinates of each location (projected using the WGS 1984 Web Mercator coordinate system) were included as covariables in the model of Equation (6). This model aimed at determining whether there is an effect of latitude and longitude on genotypic performance across environments.

$$\underline{y}_{ijk} = \mu + L_j + Y_k + LY_{jk} + G_i + \beta_i x_{1j} + \gamma_i x_{2j} + \underline{GL}_{ij} + \underline{GY}_{ik} + \underline{GLY}_{ijk} \quad (6)$$

In Equation (6), G_i considered as a fixed term, x_{1j} and x_{2j} are covariables representing the latitude and longitude of the j^{th} location respectively. β_i and γ_i represent the genotype-specific regression coefficients. Significance of fixed terms was evaluated with the Wald test (Welham and Thompson 1997). The relevance of latitude and longitude for grouping environments follows from the significance of the slopes and the reduction of the genotype by location variance.

4.2.8. Predicted response to selection in the divided and in the undivided target regions

The grouping of locations in regions was also assessed in terms of the expected response to selection, by comparing the direct response to selection estimated when selecting within a region with the correlated response expected when indirect selection would be performed across all locations (Atlin et al., 2000). This ratio was calculated as in Equation (7), following Falconer and Mackay (1996). Ratios smaller than 1.0 indicate a larger response when selecting within regions, so subdivision into regions is worthwhile, and ratios above 1.0, indicate a larger response when selecting across regions, so better not to subdivide into regions.

$$CR/_{DR} = \rho_g \sqrt{\frac{H^2_{locations}}{H^2_{regions}}} \qquad (7)$$

In Equation (7), ρ_g is the genetic correlation for grain yield between undivided and countries, using all locations, and regions, using only the locations in a region, calculated as follows;

$$\rho_g = \frac{\sigma^2_g}{\sqrt{\sigma^2_g(\sigma^2_g+\sigma^2_{gr})}} \qquad (8)$$

$H^2_{locations}$ and $H^2_{regions}$ are the heritability, or better, the repeatability, of line means in the undivided countries and the regions, respectively, estimated by using equations 9 and 10, in which nl, ny and nr are the median number of locations, years and regions in which genotypes were present, respectively (Atlin et al. 2000).

$$H^2_{locations} = \frac{\sigma^2_g}{\sigma^2_g+\frac{\sigma^2_{gr}}{nr}+\frac{\sigma^2_{gl(r)}}{nl\ nr}+\frac{\sigma^2_{gy}}{ny}+\frac{\sigma^2_{gry}}{ny}+\frac{\sigma^2_{gl(r)y}}{nl\,nr\,ny}} \qquad (9)$$

$$H^2_{regions} = \frac{\sigma^2_g+\sigma^2_{gr}}{\sigma^2_g+\sigma^2_{gr}+\frac{\sigma^2_{gl(r)}}{nl}+\frac{\sigma^2_{gy}}{ny}+\frac{\sigma^2_{gry}}{ny}+\frac{\sigma^2_{gl(r)y}}{nl\,ny}} \qquad (10)$$

4.3. Results

4.3.1. Variance components

Yield in Denmark, Germany, the Netherlands and the United Kingdom showed a complex GxE interaction, with strong σ^2_{gly} (between 43 and 52% of the total variance, Table 2). When estimating variance components without regions (Equation 1), the genotypic variance (σ^2_g) was smaller than σ^2_{gly} in all countries with the exception of Germany where σ^2_g was larger than σ^2_{gly} (σ^2_g=23.6 x 10^{-3} (Mg/ha)^2 and σ^2_{gly}=18.1 x 10^{-3} (Mg/ha)^2). σ^2_{gl} and σ^2_{gy} represented a minor proportion of the total variance (between 3 and 9% for σ^2_{gl} and between 8 and 21% in the case of σ^2_{gy}). σ^2_{gy} was in all countries about twice the size of σ^2_{gl}. The broad sense heritability at a plot level (H^2_{plot}) was similar for Denmark, the Netherlands and the United Kingdom, i.e. 0.29, 0.26 and 0.26 respectively, and was larger in Germany (0.49) mainly because of the larger σ^2_g (Table 2). At a genotype means level, countries showed the same ranking of heritabilities as at a plot level, with Germany with the highest, followed by Denmark, United Kingdom and the Netherlands. The Netherlands, with only 10 locations, had the lowest H^2_{line}.

Table 2. Variance components for each country estimated from the full unbalanced data set. Variance explained by genotype, genotype-by-location, genotype-by-year, genotype-by-location-by-year and the error term are represented by σ^2_g, σ^2_{gl}, σ^2_{gy}, σ^2_{gly} and σ^2_e, respectively, as defined in model (1). Broad sense heritability at the plot (H^2_{plot}) and genotypic mean basis (H^2_{gmean}) are also shown.

Country	Variance (Mg/ha)^2 x10^3					H^2_{plot}	H^2_{gmean}
	σ^2_g	σ^2_{gl}	σ^2_{gy}	σ^2_{gly}	σ^2_e		
Denmark	76.8	19.3	54.9	114.5	-	0.289	0.737
Germany	235.9	20.7	41.6	185.1	-	0.488	0.896
Netherlands	75.5	15.2	51.3	128.3	24.7	0.256	0.675
United Kingdom	78.4	29.0	52.1	107.2	40.6	0.255	0.721

Variance components estimated for the intervals of three years oscillated over time, but there was no general trend to increase or decrease (Figure 3). H^2_{plot} was therefore stable (Figure 4). Average yield across environments did not change much in the period that we analysed and no association of the logarithm of the variance components with grain yield was observed. H^2_{plot} estimated for the original unbalanced dataset was similar to the mean of H^2_{plot} estimated for the collection of data series of three years (Figure 4). Therefore, H^2_{plot} estimates from three years can be considered as representative for what occurred in the longer term.

4.3.2. Using the FA1 model to identify regions

The vector λ_l that contains location loadings was used to classify locations into regions by K-means clustering. Location loadings and the classification of locations into regions are shown in Figure 5. The regions obtained with the K-means clustering were used in Equation (4) to estimate variance components. The CR/DR ratio was calculated using these variance components (upper part of Table 3). In the case of Denmark, the CR/DR ratio was 0.93, suggesting an improvement in the response to selection when selecting directly in regions versus selecting across all locations in a country. More modest results were observed in Germany, the Netherlands and the United Kingdom, where the CR/DR ratio was 0.99, 0.98 and 0.98, respectively. These modest results suggest that response to selection would only marginally improve with regionalization in Germany, the Netherlands and in the United Kingdom.

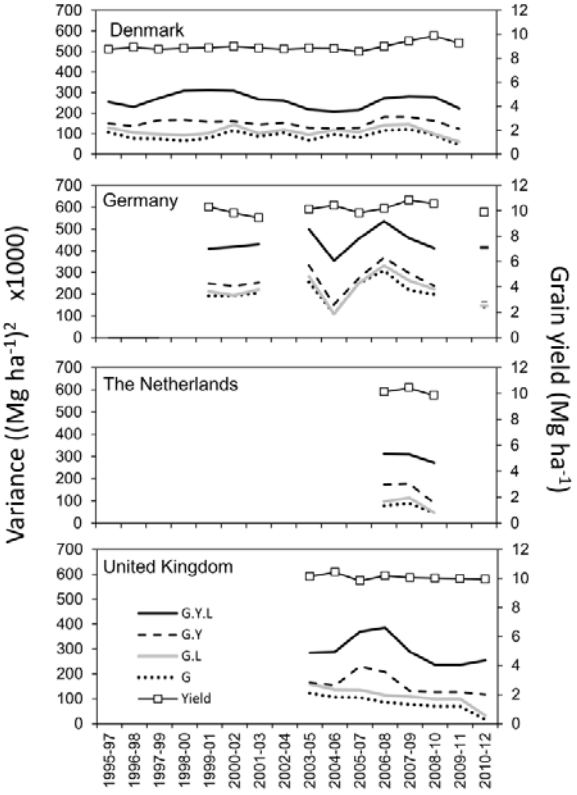


Figure 3. Mean yield and variance components estimated for each of the intervals of 3-years in Denmark, Germany, the Netherlands and the United Kingdom. Components are plotted in chronological order.

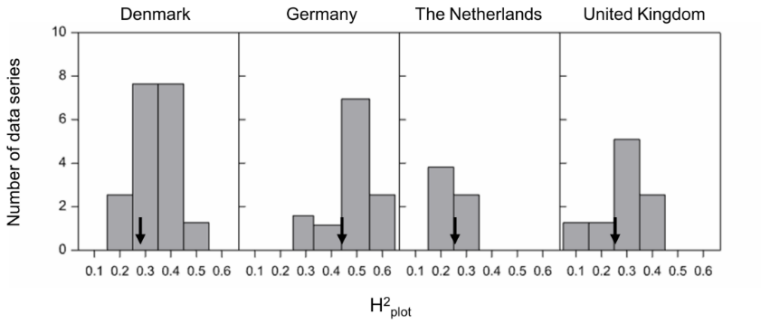


Figure 4. Histogram of broad sense heritability estimated at the plot level for each of the intervals of 3 years of information in Denmark, Germany, the Netherlands and the United Kingdom. Mean heritabilities are 0.34, 0.47, 0.24 and 0.28, for the countries in the same order. Arrows show heritability estimated using form the whole period considered in the analysis (0.28, 0.43, 0.26 and 0.26, respectively).

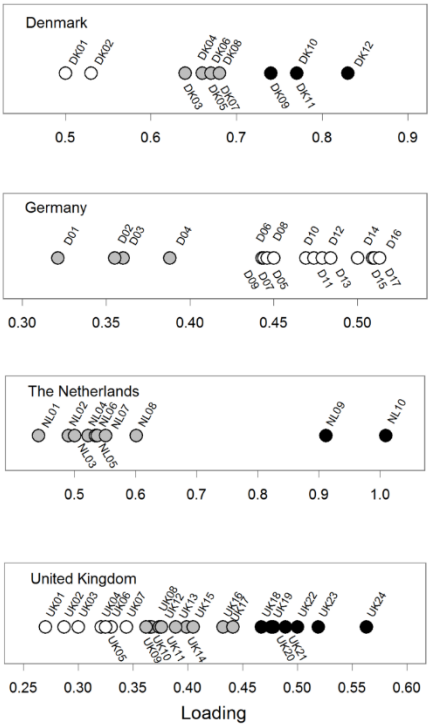


Figure 5. REML estimates of environmental loadings for Denmark, Germany, the Netherlands and the United Kingdom, estimated with a FA1 model. Circles of different colours indicate locations that were classified into different regions.

4.3.3. AMMI-based clustering to identify regions

The AMMI analysis was done separately for the tables of genotypes by locations available for each of the years. These tables of genotypes by locations differed in size (number of genotypes and locations). In consequence, the number of significant components that were retained in the AMMI model differed between years (from 1 to 4 components in Denmark, from 2 to 8 in Germany, from 2 to 3 in The Netherlands and from 1 to 6 in United Kingdom).

The predictions obtained from the AMMI model were used to define the regions though a two-step cluster analysis. In the first step, methods A, B or C grouped locations within year using information of the best genotype, the best 20%, and all of the genotypes, respectively. For strategy B, results obtained by using the top 10% and 30% genotypes were produced as well (Table 4). Depending on the percentage of genotypes that were considered, between 2 and 6 regions were obtained in Denmark, whereas the range in Germany, the Netherlands and the United Kingdom was between 3 and 6, 3 and 8 and 2 and 8 regions, respectively (Table 4, Figure 6). In the four countries, the CR/DR ratio was lowest for regions formed from when considering 20% of the genotypes. If the CR/DR ratio is smaller than 1.0, it suggests that subdivision into regions is worthwhile. When taking the top 20% of genotypes, regionalization was advantageous in Denmark, Germany and in the United Kingdom, with no advantage in the Netherlands. This result was in line with the significant improvement in goodness of fit after regions were included in Denmark, Germany and in the United Kingdom, but not in the Netherlands (Table 3).

4.3.4. Comparison of regions obtained using the FA1 model, the AMMI-based clustering and geographical information

The FA1 and the AMMI-based clustering agreed in that correlated response to selection within regions was largely improved in Denmark, whereas it was only marginally improved in Germany and in the United Kingdom (Table 3). In the Netherlands, the clustering of the FA1 loadings allowed to obtain regions with an improved response to selection, whereas response to selection did not benefit from regions obtained with the AMMI strategy. The lack of improved response to selection in the Netherlands coincides with its absence of significant latitudinal and longitudinal effects and with the homogenous yield predicted by the GAEZ map (Figure 7). In contrast, the latitudinal and longitudinal gradients were significant in Denmark, Germany and in the United Kingdom (Figure7). These gradients were represented by the regions obtained by using the FA1 model and by the regions resulting from the AMMI based-clustering. Main difference between regions obtained with both models is that the FA1 model tended to set the geographical limits to separate locations further South, compared to the AMMI model (Figure 7).

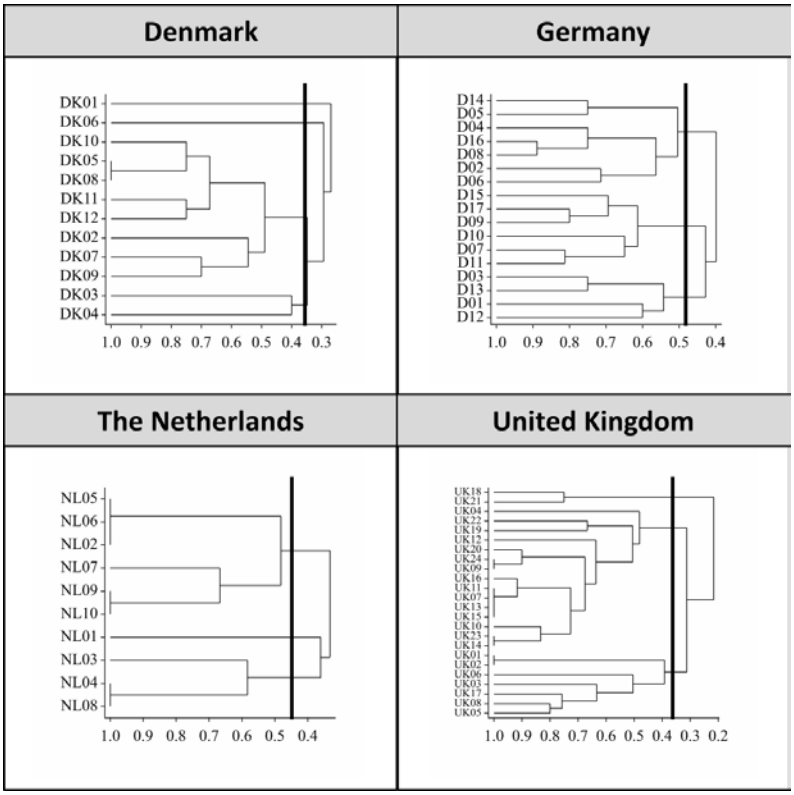


Figure 6. Dendrograms of the locations in Denmark, Germany, the Netherlands and the United Kingdom using the table of yearly clustering of locations of the AMMI-based strategy, based on the 20% best genotypes. The vertical line represents the threshold used to define the regions, which was set using the lowest CR/DR ratio.

Table 3. Variance components with their standard errors for the original unbalanced data estimated with the regions obtained using FA1 model and the top 20% of the genotypes. Variance explained by genotype, genotype-by-region, genotype-by-location-within-region, genotype-by-year, genotype-by-region-by-year and to genotype-by-location-by-year are represented by σ^2_g, σ^2_{gr}, $\sigma^2_{gl(r)}$, σ^2_{gy}, σ^2_{gry} and $\sigma^2_{gl(r)y}$, respectively, as defined in model (4). Difference of deviance was calculated using equation (4) as a full model and equation (1) as a reduced model.

Component	Denmark		Germany		The Netherlands		United Kingdom	
	Variance	s.e.	Variance	s.e.	Variance	s.e.	Variance	s.e.
	$(Mg/ha)^2$ x10^3		$(Mg/ha)^2$ x10^3		$(Mg/ha)^2$ x10^3		$(Mg/ha)^2$ x10^3	
Regions FA1 model								
σ^2_g	55.7	8.6	125.0	20.9	78.6	23.6	57.8	13.1
σ^2_{gr}	7.1	1.3	27.9	6.0	7.2	9.2	3.5	1.7
$\sigma^2_{gl(r)}$	1.8	0.3	6.7	1.0	0.0	0.0	2.5	0.6
σ^2_{gy}	41.7	5.1	0.0	0.0	8.6	14.3	31.1	7.2
σ^2_{gry}	2.7	0.9	13.1	1.4	25.2	8.9	4.9	1.4
$\sigma^2_{gl(r)y}$	118.8	2.4	178.9	4.2	159.1	7.7	160.2	4.4
CR/DR	0.932		0.986		0.975		0.976	
Δ deviance	43.210		94.464		16.410		18.066	
p-value	<0.001		<0.001		<0.001		<0.001	
Regions AMMI clustering (regions identified using the top 20% genotypes)								
σ^2_g	64.7	8.9	232.5	18.5	74.0	21.9	78.5	14.5
σ^2_{gr}	17.2	3.7	9.8	2.3	1.7	6.8	6.5	3.0
$\sigma^2_{gl(r)}$	13.3	1.9	13.9	3.7	14.5	8.3	24.8	3.7
σ^2_{gy}	48.7	4.5	41.6	4.5	44.8	12.9	48.5	6.3
σ^2_{gry}	12.3	3.2	0.0	0.0	11.6	9.6	6.9	3.0
$\sigma^2_{gl(r)y}$	109.8	2.5	185.3	4.4	145.9	10.8	144.0	4.5
CR/DR	0.872		0.982		1.048		0.970	
Δ deviance	92.454		23.024		3.130		28.376	
p-value	<0.001		<0.001		n.s		<0.001	

Table 4. Predicted selection response in a region when doing selection in the undivided country (CR), expressed relative to predicted response to direct selection within a region (DR). The ratio CR/DR and the number of resulting regions is shown for each of the grouping strategies. The groups used in the following analyses are shown in bold.

Grouping strategy	Denmark N. locations=12		Germany N. locations=17		The Netherlands N. locations=10		United Kingdom N. locations=24	
	CR/DR	N. regions	CR/DR	N. regions	CR/DR	N. regions	CR/DR	N. regions
Best genotype	0.980	3	0.997	3	1.142	5	1.006	8
Top 10%	0.973	3	1.014	4	1.071	6	0.979	4
Top 20%	**0.872**	**3**	**0.982**	**3**	**1.048**	**3**	**0.970**	**3**
Top 30%	0.932	6	0.988	6	1.098	8	1.001	7
All genotypes	0.955	2	0.996	2	1.127	5	1.014	2

4.4. Discussion

4.4.1. Variance components

A first objective of this paper was to characterise GxE interactions of wheat grain yield to have an overview of the main sources of variation of multi-environment trials sown in Denmark, Germany, the Netherlands and the United Kingdom. The relatively large size of the variance for the genotypic main effect suggests that the germplasm included in the trials was broadly adapted to the explored environments (in all countries it was above 25% of the total variance). The large σ_g^2, together with the large number of locations and years included in the analysis, resulted in a large H^2_{line}. In addition, the small σ_{gl}^2 compared to σ_g^2 and to the unpredictable components (σ_{gy}^2 and σ_{gly}^2) greatly complicated the identification of homogeneous growing conditions that are stable over time and that can contribute to an improved response to selection for regional evaluations in place of national evaluations. We could not always separate intra trial plot error from the term σ_{gly}^2 because we did not have access to the original plot data. However for the Netherlands and the United Kingdom, the estimation of σ_e^2 was possible from the reported LSD for the trials. The estimated σ_e^2 showed that the variance associated to genotype by location by year interactions corresponded to a large proportion of the residual term (Table 2). Large σ_{gy}^2 and σ_{gly}^2, compared to σ_{gl}^2, are characteristic of environments of North-Western Europe since similar partitioning of the total variance has been observed in other studies in these countries (Cullis et al. 1996; Laidig et al. 2008; Piepho et al. 2014; Weber and Westermann 1994).

The partitioning of the total variation was homogeneous over time and hence, broad sense heritability was also stable. The large heritability indicates that genotypes have a broad adaptation to the environments included in the analysis. Although variance components have also behaved in a stable way in other cases, as in Australia (Cooper et al. 1996), the apparent stability has to be taken with caution because changes might occur in the long term (Laidig et al. 2008). In our data, the magnitude of variance components was not associated to grain yield, which was stable in the time frame of our analysis. If yield potential largely changes in the future, as occurred in the past (Austin 1999; Mackay et al. 2011), the partitioning of the total variation could change.

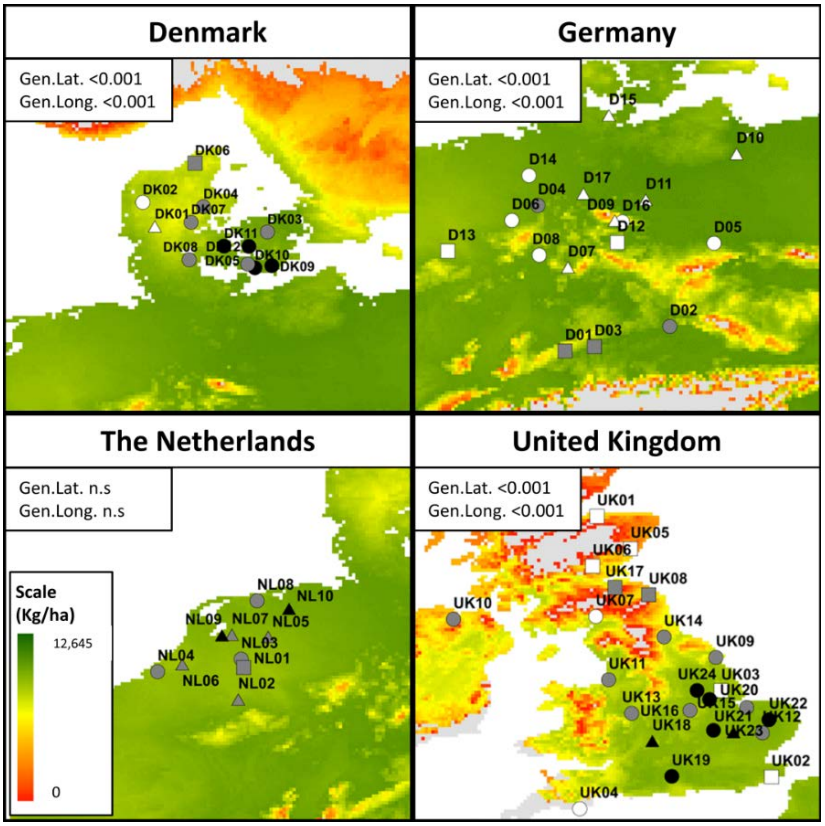


Figure 7. Groups of locations in Denmark, Germany, the Netherlands and the United Kingdom, plotted on the GAEZ map of attainable yield for high input level winter wheat, 45+135 days. Symbols of different shapes represent groups obtained with the strategy that used the top 20% of the genotypes and symbols of different colours represent the groups identified with the FA1 model. Background colours represent yield levels predicted by the GAEZ model, which are indicated in a scale on the bottom left panel. Interaction between genotype and latitude (G.Lat.) and genotype and longitude (G.Long.) is shown in the upper-left part of each map.

4.1.1. Identification of regions

A second and major objective of this paper was to compare strategies based on the FA1 model and on the AMMI model to identify regions. The FA1 covariance structure for genotypes across locations showed to be simple enough for fitting to the data of Denmark, Germany, the Netherlands and of the United Kingdom. Clustering of environmental loadings estimated with a FA model have also shown elsewhere to be a good strategy to group locations (Burgueño et al. 2008). However, fitting more complex covariance structures is not straightforward. Therefore, consideration of a simpler two-step approach to the identification of regions, as in our AMMI-fitted values based clustering, might be worthwhile.

Data tables of genotypes across locations were very unbalanced across years, but they were almost completely balanced within years, which made them suitable for fitting AMMI models by standard approaches (ANOVA and singular value decomposition of interaction residuals (van Eeuwijk 1995; van Eeuwijk et al. 2016). The wheat yield data had the advantageous feature that variance components were stable over time, enabling us to do the analysis across years.

Other applications of the AMMI model, based on stratified rankings or cluster analysis have been proposed elsewhere (Crossa et al. 1991). These improvements have the advantage of allowing a sharper identification of ranking patterns compared to using genotypic means as such. However, they assign equal weights to genotypes in the top as to the ones in the bottom of the ranking. The main change that we proposed for the identification of regions is to include information of more than a single best genotype, while assigning more weight to the best part of the ranking through the use of the Jaccard index, which in our implementation focusses on selected genotypes. Relying on more than a single best genotype is expected to produce a more robust classification of locations into regions, allowing to find regions that are less sensitive to changes of genotypes in the evaluations. The use of the best 20% of the genotypes worked best in our data set (based on the CR/DR ratio). Though, it is recommendable to explore the most suitable percentage of genotypes to be considered, as well as the thresholds used for the within- and across-year clustering when applying the method to other data. Another advantage of our strategy is that the two-step approach allowed us to identify regions across years, which is not possible to be done with a one-step AMMI or GGE model (Gauch and Zobel 1997; Yan and Rajcan 2002).

4.1.2. Correlated response to selection in the regions

This paper used mixed models to identify regions and also as a central criterion to evaluate the defined regionalization. Crossa et al. (2004) used an analogous approach, identifying the

regions with the SHMM model and evaluating these regions by mixed models. Although a model including regions might have a better fit, breeders will only put the regionalization into practice when there is clear evidence showing that response to selection can benefit from the classification of locations into regions. For that reason, it is important to include selection response theory to the evaluation of the regions, as we did.

The correlated response to selection within regions *versus* the response to selection in the undivided region is strongly affected by the size of the ratios $\sigma^2_{gl(r)}/\sigma^2_g$, $\sigma^2_{gr}/\sigma^2_{gl}$ and by the reduction in replication because of a lower number of locations within each region. If the reduction of locations within regions offsets the increase of the genotypic main effect within regions, the response to direct selection in the region will be smaller than for indirect selection in an undivided country (Atlin et al. 2011). If low genotype by location interaction is observed (e.g. $\sigma^2_{gl(r)}/\sigma^2_g$ close to 0.1), the subdivision is not expected to increase response unless σ^2_{gr} accounts for at least 50% of σ^2_{gl}. This was the case in the Netherlands, where response to selection did not improve when classifying the locations into regions. A similar result was observed in work done by Atlin et al. (2000) and by Forkman et al. (2012) who used Canadian and Swedish winter wheat. In those cases, the largest response to selection was observed when pooling all the locations within a single group.

In Germany and in the United Kingdom, the partitioning of the variation was similar for the regions obtained with the FA1 and with the AMMI model. In these countries, σ^2_{gr} explained a larger proportion of the total variance, as compared to the Netherlands. However, increase of the genotypic main effect within regions was offset by the reduction in the number of locations within each region. This offset led to a marginal improvement of the response to selection.

In Denmark, regions obtained with the FA1 and with the AMMI model followed a similar geographical pattern. However, the partitioning of the total variance was different for the regions obtained with the two methods. σ^2_{gr} was larger when using the AMMI strategy and in consequence CR/DR ratio vas also larger. Similar geographical patterns and different CR/DR ratios might appear contradictory. The contradiction can be explained by the fact that in the regions obtained with the AMMI model, two out of the three regions were formed by only one location, resulting in some confounding between σ^2_{gr} and $\sigma^2_{gl(r)}$. In consequence, CR/DR might have been overestimated in the Danish regions that were identified with the AMMI strategy. A possible modification of our grouping strategy may penalize too small groups, as these will be logistically inconvenient and of little practical value while disturbing the evaluation of the efficiency of the grouping.

In this paper we evaluated regions, with a simple compound symmetry model for the covariance between regions, and an assumption of independence for the genotypes. However, genotypic effect estimates will benefit from considering more flexible models for regions and include a relationship matrix between genotypes. Genotypic estimates can be improved by using weighted combinations of regional estimates (Piepho and Möhring 2005). Suggestions for the use of factor analytic models for the variance-covariance structure between regions were given by Piepho and Möhring (2005). Another example of heterogeneous covariances between regions with a FA model can be found in (Kleinknecht et al. 2013).

The relations between genotypes can be modelled by imposing an additive relationship matrix (A matrix) as a covariance structure (Beeck et al. 2010; Burgueño et al. 2012; Crossa et al. 2013). The A matrix could be potentially calculated either from pedigree or from molecular marker information (Legarra et al. 2009; Burgueño et al. 2012; Ashraf et al. 2016; Pérez-Rodríguez et al. 2017). However, comparing models to exploit the correlated response to selection was beyond the scope if this paper and can be covered in further research.

We took a simple approach for evaluating the consequences of regionalization of VCU trials, ignoring relations between genotypes and following Atlin et al. (2000) in comparing direct selection in a region with indirect selection for the same region in an undivided country. Piepho and Möhring (2005) show how a BLUP based estimation procedure will always improve the selection response upon regionalisation, although the logistic costs of such regionalisations may outweigh the gains in selection response. We leave the application of such an alternative BLUP estimator, more complicated models for the correlations between regions and the inclusion of a relationship matrix between genotypes for further research, as the objective of the current paper was to give insight in the sources of variation for yield in VCU trials in North West Europe and to propose alternative ways for the identification of regions with schemes for evaluating the efficiency of different regionalisations.

4.2. Concluding remarks

Wheat sown in Denmark, Germany, The Netherlands and United Kingdom showed large GxE, with variance components for unpredictable GxE variation (σ^2_{gly} and σ^2_{gy}) larger than predictable GxE variation (σ^2_{gl}). These features were stable over the full and extensive evaluation period in this study.

We used two approaches to identify regions; the first one based on modelling the variance covariance structure for the genotype by location interactions by a factor analytic model and then clustering the location scores. The second approach was based on selecting best genotypes using fitted values from AMMI models fitted to within year balanced genotype by location tables of means. We proposed a two-step approach in which first locations within years are grouped based on selected genotypes. The consistency of the location grouping over years is assessed in a second clustering where the similarity between locations is determined by the number of years that locations were assigned to the same group. This process allowed us to identify regions in Denmark, Germany and United Kingdom.

The regions obtained with the FA1 model and with the AMMI strategy were in agreement with the latitudinal and longitudinal gradients. In Denmark, the correlated response to selection was larger when selecting within regions than when selecting across regions. Although still advantageous, the inclusion of regions in Germany and the United Kingdom led to marginal improvements in expected response to selection.

Acknowledgements

We thank Wiersum Plant breeding for kindly providing the data used in this paper. We also thank Roberto Chávez (Laboratory of Remote Sensing, Wageningen) for helping with the use of ArcGIS. Daniela Bustos thanks the programme Becas Chile (CONICYT) for the funding in form of a PhD grant. Marcos Malosetti and Fred van Eeuwijk contributed to this paper in the framework of the project 2.2.1 of the Generation Challenge Program.

Chapter 5

Predicting responses in multiple environments: issues in relation to genotype by environment interactions

Marcos Malosetti[1] , **Daniela Bustos-Korts**[1,2], Martin P. Boer[1], Fred A. van Eeuwijk[1]

1. Biometris, Wageningen University & Research Centre

2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)

This Chapter is published as:

Malosetti, M., D. Bustos-Korts, M.P. Boer, and F.A. van Eeuwijk. 2016. Predicting Responses in Multiple Environments: Issues in Relation to Genotype × Environment Interactions. Crop Sci. 56: 2210–2222.

Abstract

Prediction of the phenotypes for a set of genotypes across multiple environments is a fundamental task in any plant breeding program. Genomic prediction (GP) can assist selection decisions by combining incomplete phenotypic information over multiple environments (ME) with dense sets of markers. We compared a range of ME-GP models differing in the way environment-specific genetic effects were modeled. Information among environments was shared either implicitly via the response variable, or by the introduction of explicit environmental covariables. We discuss the models not only in the light of their accuracy, but also in their ability to predict the different parts of the incomplete G×E table: $(G^t; E^t), (G^u; E^t), (G^t; E^u)$, and $(G^u; E^u)$, where $G =$ genotype, $E =$ environment, t= tested (in one or more instances), and u= untested. Using the Steptoe × Morex barley population (*Hordeum vulgare* L.) as an example, we show the advantage of ME-GP models that account for genotype by environment interactions. In addition, for our example data set, we show that for prediction in the most challenging scenario of untested environments (E^u), the use of explicit environmental information is preferable over the simpler approach of predicting from a main effects model. Besides producing the most general ME-GP model, the use of environmental covariables naturally links with eco-physiological and crop growth models (CGMs) for G×E. We conclude with a list of future research topics in ME-GP, where we see CGMs playing a central role.

Abbreviations:

G×E	Genotype by Environment Interaction
MNV	Multivariate Normal Distribution
FA	Factor Analytic
G BLUP /GE BLUP	Main/Environment Specific Genomic Best Linear Unbiased Prediction
CTD / DTD	Connected / Disconnected Training Design
DH	Double Haploid
CGM	Crop Growth Model
CV	Cross Validation
QTL	Quantitative Trait Locus
QTL×E	QTL by Environment Interaction

5.1. Introduction

Introductory courses in biology and genetics commonly start with the model $P = G + E$, that is, the phenotype is the result of the effect of the DNA constitution of an organism (the genotype), and of the effects of external inputs (the environment). However, for most complex traits, that simple link between DNA and phenotype is an oversimplification. Complex traits are the outcome of many small effects induced by DNA polymorphisms. These effects are subject to additional variation caused by environmental changes. The integration of the small genetic effects varying across environments may lead to genotype by environment interaction (GxE): genotypic differences vary in relation to the environmental conditions.

G×E is ubiquitous in plant breeding, where varieties should be selected for environments that are almost by definition heterogeneous, both in space and time. For example, heterogeneity can be caused by soil and climate conditions (structural factors), or by plot to plot variability and weather fluctuations (non-structural factors). While dealing with an essentially biological phenomenon, the term G×E is highly statistical as it implies a departure from the simple additive model, $P = G + E$. A large inventory of models related with the classical two-way ANOVA and regression have been used to describe G×E, including linear-bilinear models (Finlay and Wilkinson, 1963; Gauch, 1992; Crossa and Cornelius, 1997; Crossa et al., 2002; Yan and Kang, 2003) and mixed model versions of it (Piepho, 1997, 1998; Smith et al., 2001, 2005). Factorial regression models (Denis, 1988; van Eeuwijk et al., 1996; Denis et al., 1997) are a particularly interesting type of models because by using explicit environmental information, they can predict phenotypic responses in conditions not necessarily observed. In addition, they offer the opportunity to enhance models by biological knowledge, as a bonus. Classical reviews on G×E models include (Cooper and Hammer, 1996; Kang and Gauch, 1996; van Eeuwijk, 2006), more recent ones are (Crossa, 2012; Malosetti et al., 2013).

Plant breeding is going through a turning point in history thanks to: 1) an increase of scale via massive data generation (phenotypic and genotypic), and 2) a greater use of modelling and prediction methods via increased data management and processing capacity (Cooper et al., 2014). Clearly, the use of large numbers of markers to predict phenotypic responses (genomic prediction) fits within this context, and has a role to play in modern plant breeding. The idea of estimating breeding values by a large set of markers was introduced by Meuwissen et al. (2001), and can be regarded as an extension of earlier ideas by Lande and Thompson (1990). While initiated within the animal breeding context, genomic prediction quickly showed its potential in plant breeding (Bernardo and Yu, 2007). A comprehensive

overview of the role of genomic prediction in plant breeding can be found in Heffner et al. (2009). Genomic prediction can play a role especially at early and intermediate stages of the breeding cycle, when abundant DNA information on thousands of new genotypes can be used to predict candidate line performance even when the phenotypic information is scarce or absent. These predictions can then be used to select the best genotypes, promoting only those genotypes that are more likely to perform well to the later stages of expensive multi-environment phenotyping (advanced testing). Simulations have shown the increase in expected genetic gain per unit of time in both under high- and low-investment breeding schemes (Heffner et al., 2010).

Initial examples of genomic prediction in plant breeding dealt with within environment predictions (Piepho, 2009; Crossa et al., 2010; Heslot et al., 2012). However, the need to account for genotype by environment interaction effects when predicting for multiple environments was quickly recognized (Burgueño et al., 2011; Burgueño et al., 2012; Schulz-Streeck et al., 2013). More recently, factorial regression type of models, which include explicit environmental covariables to form predictions have been proposed (Heslot et al., 2013a; Jarquín et al., 2014).

In this paper, we discuss the performance in terms of prediction accuracy of different types of prediction models for incomplete genotype by environment data sets. The compared models differ in whether they allow environment-specific effects or not, in whether they allow borrowing information between environments or not, and whether they allow forming predictions for fully unobserved environments or not. Different layers of the prediction problem are treated, that is, predictions in observed environments of genotypes that were either tested in other environments or never tested, and the most challenging scenario, prediction in new environments. The effect of the training set design on prediction accuracy is also discussed. As illustration example we use a double haploid barley population that is simple but well known and for which multi-environment yield data and explicit environmental covariables are available. We analyzed the data, draw conclusions, and underline future research areas regarding the difficult task of predicting for multiple environments.

5.2. Materials and Methods

5.2.1. G×E Table with Empty Cells

The general structure of a genotype by environment data can be summarized by a two-way table ($I \times J$) with genotypes in the rows ($i = 1 \ldots I$) and environments in the

columns ($j = 1 \ldots J$). In a balanced data set all $I \times J$ cells in the table are present. However, in most of the cases, the G×E table has empty cells for different reasons. For example, a particular combination of genotype and environment might not have been tested, or the information of a genotype is available in some but not all of the tested environments (e.g. because of the limited amount of seeds or available plots). In some cases, the information might be absent either because a particular genotype was never tested in the field or because an environment was never observed (e.g. a new testing site or a future planting season). To accommodate these situations, the $i = 1 \ldots I$ genotypes can be grouped $i \in (G^t, G^u)$, and the $j = 1 \ldots J$ environments $j \in (E^t, E^u)$ where the superscript t and u stand for *tested* and *untested* respectively. A schematic representation of the entire G×E table with filled and empty cells is shown in Figure 1, where the white areas represent absence of phenotypic information. Note that in the case of *tested* genotypes and environments, $(G^t; E^t)$, it is not implied that all genotype-environment combinations are observed. Also note that the empty cells in the table can have regular or irregular patterns (the one displayed in Figure 1 is just an arbitrary example for easy visualization that shows no particular pattern).

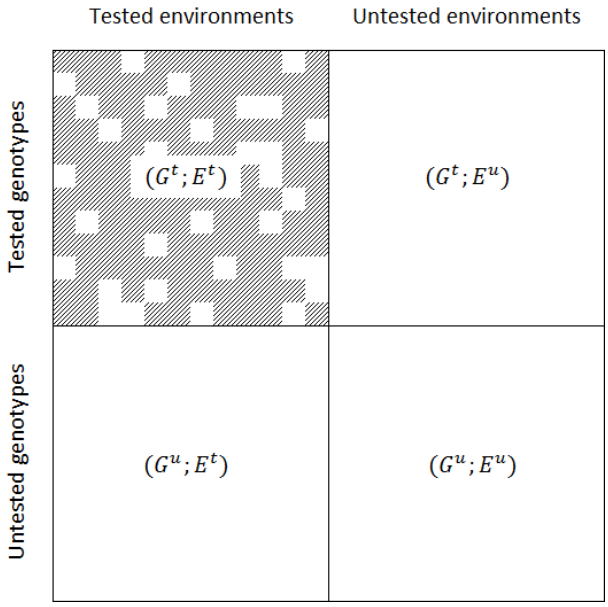


Figure 1. Representation of an incomplete G×E table with genotypes in the rows and environments in the columns. Dashed areas represent available phenotypic observations, and white areas representing unavailable phenotypic data.

5.2.2. The General Model for the G×E Table

The general model for the G×E table is given by

$$y_{ij} = \mu_j + G_{ij} + \epsilon_{ij} \quad (1)$$

where y_{ij} is the trait value of the i^{th} genotype $i \in (G^t, G^u)$ in the j^{th} environment $j \in (E^t, E^u)$, μ_j is an intercept term for environment j, G_{ij} the environment-specific genotypic random effect of genotype i in environment j that follows $G_{ij} \sim MVN(0, \Sigma)$, where Σ is a covariance matrix, and ϵ_{ij} a residual term that includes non-additive genetic effects and error variance (both environment-specific), $\epsilon_{ij} \sim MVN(0, R)$. For the residual variation, heteroscedasticity is typically required, so R takes a diagonal form $D = Diag\{\sigma_j^2\}$, where each environment has its own residual variance. Although not treated here, further modelling of the residual term is possible to account for autocorrelations due to non-additive genetic variation and/or local environmental variation (Oakey et al., 2006; Burgueño et al., 2007).

Model (1) can be used to predict empty cells in the G×E table. The prediction ability for empty cells by the model comes from the fact that via the term G_{ij} information can be borrowed from the filled cells in the table. In particular, the recovery of information is possible via the covariance matrix Σ, which approximates the multi-dimensional genotypic space defined by the genotype-environment combinations. In this paper we discuss different models for Σ and their implications in terms of the predictions that the corresponding models would allow.

As the white areas in Figure 1 suggest, there are different types of predictions to be made, namely: 1) prediction of tested genotypes in tested environments but for which the particular combination was not observed $(G^t; E^t)$, 2) prediction of new genotypes in environments that were observed $(G^u; E^t)$, and 3) prediction of genotypes (previously tested or not) in new (future) environments $(\{G^t, G^u\}; E^u)$. The complexity of the prediction problem increases from situation 1 to 3, and as we will argue later in this paper, will require the use of different types of information to model the covariance matrix Σ.

5.2.3. Models for Σ

The matrix Σ is a genotypic covariance matrix that can be decomposed into a genotypic-related and an environment-related component, say $\Sigma = \Sigma^G \otimes \Sigma^E$, with $\otimes$ the Kronecker product of the two matrices. The matrix Σ^G reflects similarities among genotypes in terms of the DNA sequence, whereas Σ^E reflects similarities among environments induced by the growing conditions, and both jointly determine the similarities among genetic effects.

The genetic similarity Σ^G can be derived from pedigree records (i.e., expected kinship) or from molecular marker information (i.e., realized kinship). With a biallelic marker system as SNPs with say, alleles A and B, a standard expression for Σ^G is $\Sigma^G = WW'$, with W an $(I \times M)$ marker scores matrix, I is the population size and M is the total number of markers. Each element in the matrix, w_{im}, contains the number of B alleles in genotype i for marker m, $w_{im} = \{0, 1, 2\}$, or after centering and standardization $w_{im} = \frac{W_i - 2p_m}{\sqrt{2p_m(1-p_m)}}$, with p_m the frequency of the B allele at SNP m (Speed and Balding, 2014).

Here we give special attention to matrix Σ^E, which conditions the genetic variation on implicit or explicit properties of the environments. The model for Σ^E determines the prediction scenario for which a particular model can be useful, i.e. $(G^t; E^t)$, $(G^u; E^t)$, or $(\{G^t, G^u\}; E^u)$. We start with predictions in tested environments $(G^t; E^t) \cup (G^u; E^t)$, which have in common that predictions are made in environments in which at least part of the genotypes were observed. Therefore, we can model Σ^E in relation to the implicit environmental information that is available in the phenotypic data. In the basic G-BLUP model, all environment-specific effects are assumed to come from the same distribution with variance σ_G^2 (Table 1). However, if genetic effects are conditional on the environment (i.e. there is G×E), the genetic variance should be allowed to vary from environment to environment. A simple extension of the model for Σ^E is to fit a separate variance component per environment, with a diagonal structure, with the genetic variance component $\sigma_{G_j}^2$ attached to each environment (Table 1, GE-Diag model). The GE-Diag model assumes no genetic correlations between environments, which is not realistic if two environments have properties in common, and so, one would expect genetic correlations to occur between those environments. Genetic correlations between environments can be modeled by allowing non-zero covariances among environments. The most general option to model environmental covariance is the unstructured model that has a separate variance component for each environment, $\sigma_{G_j}^2$, and a separate covariance parameter for each pair of environments $\sigma_{G(j,j^*)}$ (Table 1, GE-Unstr). Depending on the number of environments involved, parsimonious models such as the Factor Analytic model can be an interesting alternative for Σ^E (Denis et al., 1997; Piepho, 1997; Smith et al., 2001). The models discussed above assume that it is possible to estimate the covariance between the environments, implying that environments have been tested, i.e. E^t. However, the covariance between tested and untested environments, i.e. $cov(E^t; E^u)$, or between untested environments, i.e. $cov(E^u; E^u)$ cannot be estimated from the phenotypic data, since by definition observations are not available. Without implicit environmental information (phenotypes), prediction for new environments will necessarily require the use of explicit environmental information to model Σ^E. Analogous to molecular markers (which are explicit genotypic covariables), explicit environmental covariables can be used to define similarities between environments (i.e. form an "environmental kinship",

Table 1, model GE-KE). Each entry in the environmental similarity matrix is a function of K covariables ($k = 1 \ldots K$), that are chosen to characterize the environments, with the covariables having the environment-specific values z_{kj}. At a finer resolution, the environmental characterization can be made at the individual genotype level (z_{kji}) by considering the individual genotype phenology, in which case $\Sigma = Z_g \Sigma^G Z_g' \circ \Sigma^{E^*}$, with Z_g an incidence matrix of genotypes, and $\circ$ the Hadamard product of the two matrices (Jarquín et al., 2014). Note that the environmental matrix Σ^{E^*} is defined at the individual genotype level, so of dimension ($IJ \times IJ$). An important difference of the inclusion of explicit environmental covariables in the model is that modeling is not restricted to the tested environments (E^t), but covers the whole environmental range $j \in (E^t, E^u)$. Therefore, this last type of prediction model is the most general one as it will be useful to predict any environment (either tested or not tested).

Table 1. Models for the genetic variation in a multi-environment space, $\Sigma = \Sigma^G \otimes \Sigma^E$, with Σ^G a genotypic-related component, and Σ^E an environment-related component.

Model name	$\boldsymbol{\Sigma^G}$	$\boldsymbol{\Sigma^E}$
G-BLUP	WW'	$\Sigma^{E^t}_{(j,j^*)} = \sigma^2_G$ if $j = j^*$, zero otherwise
GE-Diag	WW'	$\Sigma^{E^t}_{(j,j^*)} = \sigma^2_{G_j}$ if $j = j^*$, zero otherwise
GE-Unstr	WW'	$\Sigma^{E^t}_{(j,j^*)} = \sigma^2_{G_j}$ if $j = j^*$, $\Sigma^{E^t}_{(j,j^*)} = \sigma_{G(j,j^*)}$ otherwise
GE-KE	WW'	$\Sigma^{E}_{(j,j^*)} = 1 - \sum_{k=1}^{K} \frac{z_{kj} - z_{kj^*}}{\max(z_k) - \min(z_k)}$

The superscript *t* stands for *tested*, *j* and *j** are indexes for any two environments. The similarity coefficient in model GE-KE is defined based on *K* covariables (z_k).

5.2.4. Example Data Set

As illustration of a G×E data set, we use the well-known Steptoe × Morex double haploid (DH) population (Hayes et al., 1993). More information about the phenotypic and meteorological data can be found in Malosetti et al. (2004). The phenotypic information consisted of grain yield (Mg ha^{-1}) of 148 DH lines observed in 10 trials, covering sites in US and Canada in two consecutive years (1991-1992). Meteorological information consisted of daily records of temperature, available water (rainfall and irrigation when applicable), and photoperiod. The climatic information was summarized per crop stage: vegetative (from sowing to visible awns), heading time (from visible awns to end of anthesis), and grain filling (from end of anthesis to maturity). The original markers for this population consisted of a low density set of 116 markers, but a much denser set of about 3000 SNPs was published later for this population (Close et al., 2009). Here we used a subset of 794 SNPs evenly spaced over the seven linkage groups, such that the largest gaps between consecutive SNPs were about 5 cM.

We considered two major prediction scenarios: 1) a *time structured prediction* scenario, which emulates the situation where selections are performed for a future environment using past phenotypic data, and 2) a *physically structured prediction* scenario, which emulates the situation where the trials used to train the prediction model have been selected based on their explicit physical properties defined by a number of environmental covariables.

In the *time structured prediction* scenario, we used a fraction of the phenotypic data of 1991 to train the prediction model, and predicted the untested genotypes in 1991 and all genotypes in 1992. In addition, we compared two alternative training designs for the phenotyping in 1991. Under the so-called *connected training design* (CTD), the entire population was split into 5 sets, and the sets randomized over the four environments, with one set observed in all the environments, and the remaining sets observed in two out of the four environments (Figure 2a). So, the CTD design was unbalanced with respect to environments. Therefore, predictions within 1991 were for genotypes and environments that were tested $(G^t; E^t)$ and predictions for 1992 were for tested genotypes but in new (untested) environments $(G^t; E^u)$. Under the *disconnected training design* (DTD), the DH lines were split into a *training set* (60% of the lines), and a *test set* (40% of the lines), implying that DH lines were either fully observed in 1991 or not observed (Figure 2b). The DTD design was balanced with respect to environments. Therefore, predictions within 1991 were for untested new genotypes $(G^u; E^t)$, and predictions for 1992 were for both tested and untested genotypes $(\{G^t, G^u\}; E^u)$. Both training designs (CTD and DTD) are equivalent in terms of the amount of phenotypic data that was used (60% of the 1991 data), and in both designs, the training population represents as a whole 24% of the total data and the remaining 76% is used as testing/validation set (since all the 1992 information was assumed absent).

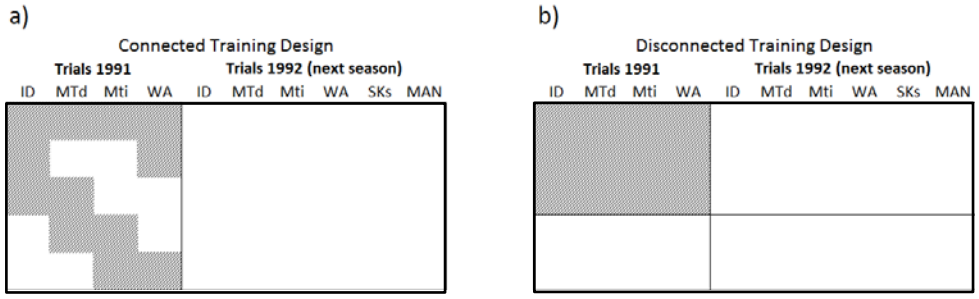


Figure 2. Diagram of the *time structured prediction* scenario, under two designs of the training phenotyping (a) Connected Training Design, and (b) Disconnected Training Design applied to the Steptoe x Morex barley DH population. The rows represent genotypes, and the columns represent environments, which are grouped by year (1991 and 1992 trials). Shaded areas represent available phenotypic information, white areas represents unavailable information.

In the *physically structured prediction* scenario, we took a different approach regarding the 10 trials as a sample of possible environments, and from which a *training* set of environments should be selected. The physical properties of the environments (and not yield information) were used to assign environments to the training set, via the explicit environmental information. First a similarity matrix was estimated using the environmental covariables, and a selection of the trials in the training set was made after inspection of a Principal Coordinate plot (Gower, 1971). Five trials were designated as training set, chosen such that they covered the entire environmental space in the plot. The two training designs, CTD and DTD, were also applied in this prediction scenario (Figure 3).

The different scenarios were repeated 25 times (independent realizations) and the models fitted by Residual Maximum Likelihood (REML) in GenStat 17th Edition (VSNi, 2014). The accuracy of predictions was assessed by the Pearson correlation coefficient between predictions and their actual observations. A linear mixed model including prediction model, training design, and environment (and their interactions) as fixed effects and replicates as random effects (blocks) was fitted to assess the influence on the observed accuracies (r). For the analysis, and to comply with the normality assumption, the Fisher's z transformation was used, $z = 1/2\ (\ln((1+r)/(1-r)))$. Final results (tables of means following the findings from the Wald/F test statistic) were presented on the original scale by reporting the values after back-transformation: $r = \frac{\exp(2z)-1}{\exp(2z)+1}$.

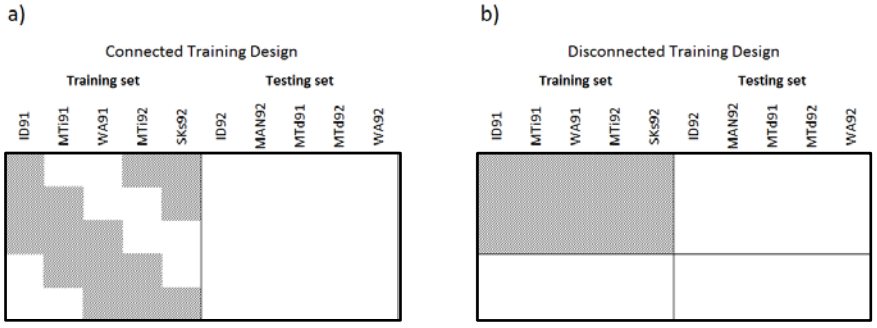


Figure 3. Diagram of the *physically structured prediction* scenario, with five environments selected as training environments and the remaining five used as testing/validation environments. The information within the training environments is incomplete following a connected (a), and a disconnected design (b), similar as the one described in Figure 2. The rows represent genotypes, and the columns represent environments. Shaded areas represent available phenotypic information, white areas represents unavailable information.

5.3. Results

5.3.1. Prediction accuracies in tested environments (1991)

Table 2 summarizes the results of the simplest prediction situation, that is, the prediction of unobserved genotypes in tested environments (E^t) under the *time structured prediction* scenario. Note that while here we report only the results of unobserved DH lines (thus focusing on prediction accuracies), both, BLUP of observed and unobserved DH lines are relevant, because essentially all genotypes are of interest for genomic selection. In the case of the observed lines, predictions can be regarded as improved estimates of their breeding values over the ones estimated from phenotypic records alone (Heffner et al., 2009). The value of the F-test statistics shows a strong effect of prediction model by environment interaction (Table 2). Figure 4A shows that in ID91 and WA91 the GE-BLUP models were substantially better than the G-BLUP model and not substantially different among each other. However, in MTd91 and MTi91 no major differences were found between all models (the GE-Diag somewhat lower in MTi91). In addition, and although relatively less important, the model by design interaction also showed some effect (Table 2). From Figure 4B can be observed that most of the effect is associated with the relative differences among GE-BLUP models. Although the total number of phenotypic observations is the same in the CTD and in the DTD designs, the total number of genotypes in the training population is larger in the former than in the latter. That explains why the accuracies under the CTD are in general higher than under the DTD for the same model (Figure 4B). The same observation was made by Endelman et al. (2014), who compared the performance of the G-BLUP model using the Harrington × T306 barley double haploid population under a wide range of CTD and DTD designs and training population sizes. However, here an extra advantage of the CTD was uncovered, and that is the one of exploiting the information from correlated environments. Only the CTD allowed the factor analytic or the environmental kinship model to express their superiority over the diagonal model by exploiting the correlation structure in the data. Under the DTD design, the advantage of the former two over the latter vanished due to the much lower connectivity in the data set.

Table 2. Wald test (and corresponding approximate F-test stastistic) associated with the effects of prediction model, design of training population, environment, and their corresponding interactions on the prediction accuracy ($\bar{r}$) of unobserved DH lines in four North American environments in 1991. The probability (F pr) corresponds to the F-test statistic.

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Model	381.56	3	127.19	576	<0.001
Design	6.49	1	6.49	48	0.014
Environment	342.90	3	114.30	144	<0.001
Model × Environment	691.36	9	76.82	576	<0.001
Model × Design	36.30	3	12.10	576	<0.001
Environment × Design	4.43	3	1.48	144	0.224
Model × Environment × Design	8.02	9	0.89	576	0.533

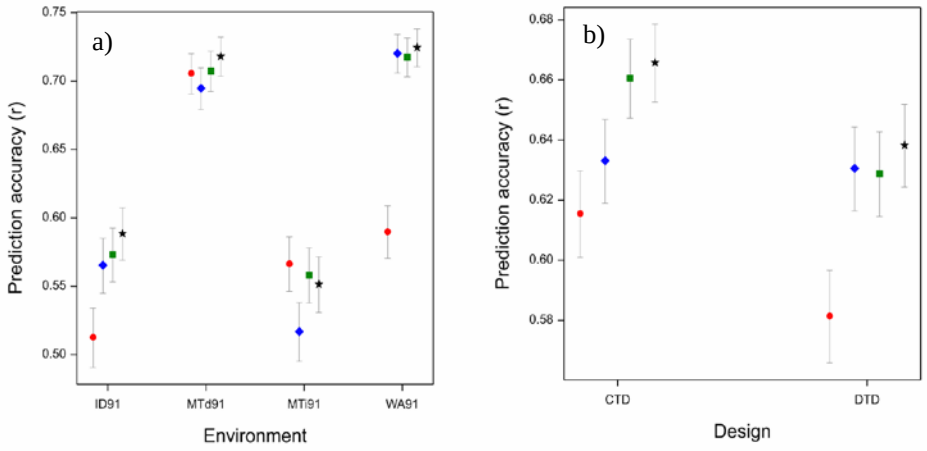


Figure 4. Mean prediction accuracies in tested environments in 1991 (E^t) of four prediction models: a main effect model (G BLUP, circle), a diagonal model (GE Diag, diamond), a factor analytic model (GE-FA, square), and an environmental kinship model (GE-KE, star). A) The four tested environments and B) the connected (CTD) and disconnected (DTD) training designs. The bars give the upper and lower bounds of the 95% confidence interval.

5.3.2. Prediction accuracies in new environments (1992)

The results of the performance of the models under the most demanding prediction scenario, i.e. the prediction for new environments (E^u), are shown in Table 3 and Figure 5. As one would expect given the far more difficult task at hand, the results were less conclusive than the within year prediction. Only in two environments (ID92 and MTi92) the accuracies

were comparable (or slightly lower) to those observed in the within year prediction scenario (compare Figure 4A with Figure 5). The ranking of the models depended on the environment, which was reflected in the large F test statistic associated with model by environment interaction (Table 3). For example, in ID92 all models differ from each other, with G-BLUP the best model followed by GE-KE, GE-FA, and GE-Diag. However, in MTi92 no differences were found among these models. For the other three environments accuracies were either moderate or low (and in some cases even negative). In 1992 two new testing sites were included (MAN and SKs). As was mentioned in the model description section, and reflected in Figure 5, not all GE-BLUP models were able to produce predictions for the unobserved environments. Model G-BLUP predicts under the strong assumption that genetic effects are not conditional on the environment (main effects prediction model), and so predictions are always possible. The diagonal model and the factor analytic model do condition genetic effect on the environment, but in return, they require implicit information for those environments to form predictions, which were not available. In contrast, even when no information was available from 1991 in MAN and SKs, the environmental kinship model did allow predicting for these new environments, and in the case of MAN92 it showed to be superior over the G-BLUP model.

Table 3. Wald test (and corresponding approximate F-test stastistic) associated with the effects of prediction model, design of training population, environment, and their corresponding interactions on the prediction accuracy ($\bar{r}$) of the performance of DH lines in four North American environments in 1992 (E^u) using past data (1991). The probability (F pr) corresponds to the F-test statistic.

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Model	759.28	3	253.09	679.1	<0.001
Design	1.76	1	1.76	48.0	0.191
Environment	14438.44	5	2887.58	249.5	<0.001
Model × Environment	4249.79	11	386.34	679.1	<0.001
Model × Design	1.06	3	0.35	679.1	0.786
Design × Environment	1.35	5	0.27	249.5	0.929
Model × Environment × Design	13.54	11	1.23	679.1	0.262

5.3.3. Using environmental covariables in G×E prediction

The prediction accuracies for new environments under a *time structured prediction* (i.e. prediction of 1992 using 1991 data) were very low. Obviously, the prediction of an untested environment (E^u) is a more difficult case than prediction within a tested environment (E^t). However, the low level of accuracy, suggested the need of an additional explanation. As in any prediction exercise, the prediction ability of models depends not only on the choice of

the model, but also on the suitability of the training data used to estimate the model parameters. A detached training data set from the target population, will lead to poor prediction ability. Lack of consistency of performance across years as the result of complex genotype by environment interaction patterns is not an uncommon situation in plant breeding data, and the Steptoe x Morex population is not an exception of that (Romagosa et al., 1996). The imperfect overlapping between the 1991 and 1992 environments as training and testing set respectively is clearly shown in Figure 6. The plot shows a two-dimensional representation of the growing conditions in the two years based on the physical properties used to characterize these environments (temperature, water availability and photoperiod in the different crop stages). Clearly, the growing conditions were quite different between the two years, and so, they help to explain the poor prediction of the 1992 performance by models trained with 1991 data.

When environmental information is available for potential testing sites, a selection of the training set of environments can be made based on that information. That was the rationale behind the so-called *physically structured prediction* scenario. As an example, after inspection of the plot in Figure 6, five environments (ID91, MTi91, MTi92, SKs92, and WA91) that roughly covered the environmental spectrum were selected as training set, and the remaining environments left as testing/validation set. Under this scenario, the two models that can produce predictions in untested environments (E^u) were compared: the G-BLUP model and the GE-KE model. The results in Table 4 show that model by environment interaction had a dominant effect, reflected by the largest F test statistic among all interaction terms. One first observation in Figure 7 is that accuracies were in general above 0.3 for all environments except in one, WA92 and in MTd91 for the G-BLUP model. Therefore, there was a clear improvement over the accuracies in the untested environments in Figure 5. In addition, no large differences in accuracies were found between tested (E^t) and untested (E^u) environments (left and and right of the gray dotted line in Figure 7). If WA92 is not considered, the mean prediction accuracies for the G-BLUP and GE-KE were in the range of 0.4-0.5. What was consistently observed in all environments was that the GE-KE model showed higher accuracies than the G-BLUP model, with a few exceptions where differences were not large (MTi91 and WA92). Note that in WA92 none of the models had predictive power. Regarding the training population design, and in agreement with the conclusions in Endelman et al. (2014), for the same model, the CTD design showed to be superior over the DTD design, although the magnitude of the effect was comparatively lower than that of the prediction model (Figure 7B).

Table 4. Wald test (and corresponding approximate F-test stastistic) associated with the effects of prediction model, design of training population, environment, and their corresponding interactions on the prediction accuracy ($\bar{r}$) of the performance of DH lines in the training and testing environments in North America. The probability (F pr) corresponds to the F-test statistic.

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Model	1316.95	1	1316.95	480	<0.001
Design	6.43	1	6.43	48	0.015
Environment	4176.65	9	464.07	432	<0.001
Model × Environment	1774.81	9	197.20	480	<0.001
Model × Design	14.54	1	14.54	480	<0.001
Design × Environment	27.94	9	3.10	432	0.001
Model × Environment × Design	20.34	9	2.26	480	0.017

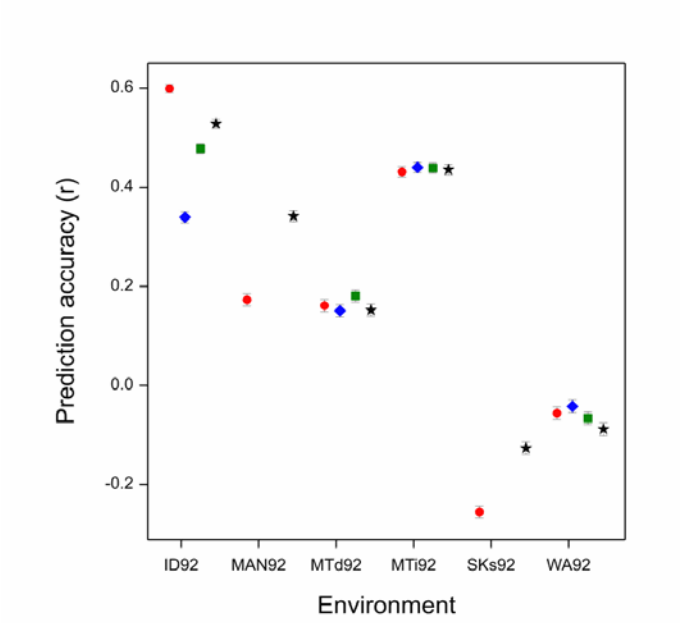


Figure 5. Mean prediction accuracies in untested environments in 1992 (E^u) of four prediction models: a main effect model (G-BLUP, circle), a diagonal model (GE-Diag, diamond), a factor analytic model (GE-FA, square), and an environmental kinship model (GE-KE, star). In MAN92 and SKs92 predictions from GE-Diag and GE-FA were not available since those environments were not present in the 1991 data. The bars give the upper and lower bounds of the 95% confidence interval.

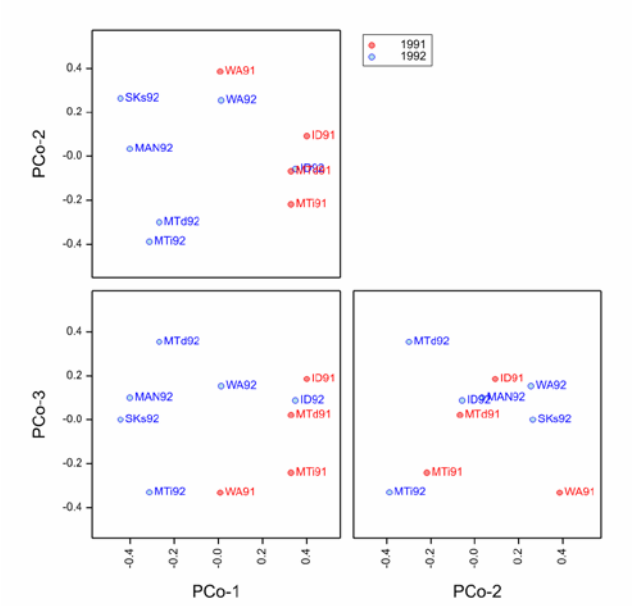


Figure 6. Principal Coordinate (PCo) plot of the environmental kinship estimated from explicit environmental covariable (temperature, water availability and photoperiod) measured in different crop stages (vegetative, heading time, and grain filling). Explained variances by the first three axes were 31.5%, 18.0%, and 14.6%.

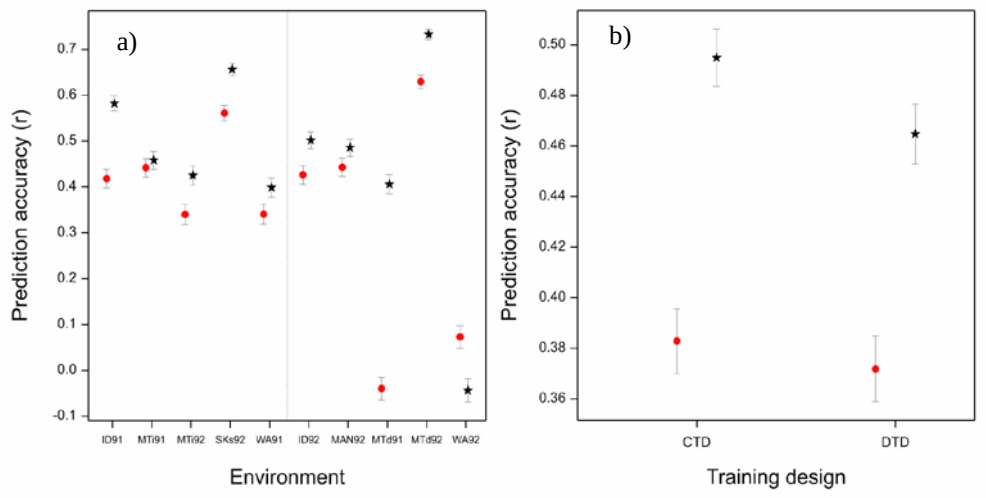


Figure 7. Mean prediction accuracies of the main effect model (G BLUP, circle), and the environmental kinship model (GE-KE, star) in: A) tested (E^t) and untested environments (E^u), the gray dotted line separates tested environments (to the left), and untested environments (to the right); and B) under two training population designs: connected design (CTD), and disconnected design (DTD). The bars give the upper and lower bounds of the 95% confidence interval.

5.4. Discussion

We have compared the performance of different prediction models in the context of multi-environment data. The prediction models that we compared can be classified as either a main effects model (G-BLUP) or models that allow environment-specific genetic effects, GE-BLUP models. The GE-BLUP models differ among each other in two major features: a) whether they allow to borrow information from other environments or not, and b) whether they use implicit or explicit environmental information. In addition, we have compared the effect of the design of the phenotyping of the training population set, that is, either assuring that all genotypes have some phenotyping information (although not in all environments), or having some genotypes fully phenotyped (training set), while others are left without any phenotypic information (test set). Finally, the performance of the models have been compared at different layers of the prediction exercise, that is, prediction in tested environments (E^t) of partially tested genotypes ($G^t; E^t$), or fully untested genotypes ($G^u; E^t$), and the most challenging prediction scenario, prediction in new environments (E^u) of tested or untested genotypes ($\{G^t, G^u\}; E^u$). The major findings were:

- In our data set, models that conditioned genetic effects on the environment (GE-BLUP models) were in general superior to the main effect model (G-BLUP).

- In prediction scenarios for tested environments ($G^t; E^t$) and ($G^u; E^t$), GE-BLUP models that borrow information between environments were superior, especially when the design of the training set included all genotypes (connected design).

- In the most challenging prediction scenario ($\{G^t, G^u\}; E^u$) the overlap of the training and test set becomes a crucial issue (when predicting for tested environments, the overlap is guaranteed since the predictions are restricted to the same set of environments, E^t).

- The use of external environmental information in the prediction model allows formulating the most general GE-BLUP prediction model, as it can predict for any situation, i.e. ($G^t; E^t$), ($G^u; E^t$), ($G^t; E^u$), and ($G^u; E^u$).

We now elaborate our major conclusions, and give some final remarks regarding future research in this area.

We found that GE-BLUP models were superior to or not different from G-BLUP. This result implies that it was advantageous (or at least had no penalty) to allow environment-specific genetic (marker) effects when dealing with multi-environment

prediction. A similar conclusion was found by Burgueño et al. (2012), where for example their models 14 and 15 (equivalent to our GE-Diag, and GE-FA) were as good as or better than model 13 (equivalent to our G-BLUP model). In a similar work, but within a Bayesian framework, Lopez-Cruz et al. (2015) showed an improvement of accuracy or no difference with the across environment model when marker-specific effects were allowed. Naturally, the superiority of the GE-BLUP models over main effect models depends on the importance of G×E. In a large international wheat data set no differences were observed between main effects and (mega)environment-specific effects models (Dawson et al., 2013). The result was explained by the relatively high correlations between environments, and it was shown by simulations that GE-BLUP models have no advantage when correlations between environments are high (above 0.7). Finally, one should note the parallelism of multi-environment prediction with that of predicting multiple populations, where the prediction model can be equally allowed to have population-specific marker effects (Schulz-Streeck et al., 2012; Technow and Tutoir, 2015).

We also found that the gains of more sophisticated models to borrow information between environments (e.g. FA models) depend on the design of the training set. As one would expect, a strategy of partial phenotyping of the population within environments as our connected design, offers more scope for borrowing information among environments than a full phenotyping of part of the population as in the disconnected design. Similarly, Burgueño et al. (2012) and Lopez-Cruz et al. (2015), observed a difference in favor of models allowing correlations between environments under the cross-validation scheme CV2 (that corresponds to predictions for $(G^t; E^t)$ in our connected design), and not under CV1 (the equivalent of our predictions for $(G^u; E^t)$ under the disconnected design). Therefore, and for a fixed amount of resources for example in terms of available evaluation plots, one should consider training the model using as many different genotypes as possible and spreading them over the testing environments, instead of exclusively fully phenotyping part of the population.

Quite frequently, genomic prediction procedures are compared using cross-validation (CV) schemes. Within the context of multi-environment data, CV almost by definition restricts the prediction problem to only part of the scenarios of interest, i.e. $(G^t; E^t)$ and $(G^u; E^t)$. In addition, CV results might overestimate the ability of the prediction models, which may do differently outside the boundaries of the training/testing space (Albrecht et al., 2014). Prediction in new environments (E^u) is a more difficult task but works as a validation procedure, that is, data that is not used to estimate the model parameters but that is used to independently assess the performance of the prediction models. In addition, $(\{G^t, G^u\}; E^u)$ is clearly a more relevant type of prediction for a plant breeding program that needs to forecast which lines are likely to do well in future testing and eventually in production.

However, $(\{G^t, G^u\}; E^u)$ introduces an extra complication that is the need to establish a link between tested (E^t) and untested (E^u) environments. A good understanding of the G×E patterns by classical analytical approaches using historical data can be instrumental in finding those links (van Eeuwijk et al. 2015, this issue). Alternatively, prediction models can be generalized by the addition of explicit environmental information, as we shown here with our GE-KE model.

The idea of modeling genetic variation conditional on explicit environmental data has been applied in the context of QTL×E (Crossa et al., 1999; Malosetti et al., 2004; Boer et al., 2007). The extensions within the framework of multi-environment prediction models were simultaneously proposed by Heslot et al. (2013) and Jarquín et al. (2014), although they differ in that the former was oriented towards prediction of new environments (E^u) while the latter still focused on prediction of tested environments (E^t). Our results showed that the inclusion of environmental covariables allowed predicting reasonably well in the more challenging situation of an unobserved environment $(\{G^t, G^u\}; E^u)$. In tested environments, predictions for $(G^t; E^t)$ and $(G^u; E^t)$ from the GE-KE model did equally well (or even better) than other GE-BLUP prediction models. The distinctive feature of the use of explicit environmental covariables is that it shifts from predicting at a discrete set of environments, to prediction in a continuous genotypic space defined by the configuration of genotype and environmental combinations, previously observed or not. Of course, forming groups of environments (mega-environments) is an alternative way of dealing with prediction of future environments. Effectively it implies defining a representative set of landmarks within that continuous genotypic space that are used as reference for future predictions. The use of genomic predictions for classifying and selecting representative environments has been proposed (Heslot et al., 2013b). However, the approach will work as long as future conditions are indeed reasonably well represented by those reference environments.

Dealing with a $(\{G^t, G^u\}; E^u)$ scenario raises another important issue that might be overlooked when focusing on predicting $(G^t; E^t)$ or $(G^u; E^t)$, and that is the overlap between the training and testing/validation environments. Predicting for tested environments (E^t), automatically assures the overlap along the environmental dimension. Of course, the same overlapping requirement holds true along the genotype dimension, especially when dealing with structured genetic material or multiple populations (Albrecht et al., 2011, 2014; Rincent et al., 2012; Guo et al., 2014). In our rather limited 2-year data set, the simple temporal partitioning of the environments (1991 versus 1992) resulted in an insufficient overlap between training and testing set, and caused the prediction models to fail. However, a re-structuring of the training and validation set using external environmental information to select representative training environments reverted to relatively high accuracy levels for

a complex trait as yield. The sharp increase in accuracy observed, even when dealing with a short series of data (two years) was remarkable. It is to be expected that the addition of more information should improve the results of the prediction ability of the GE-KE model. For example, we still failed to predict environment WA92, which is not surprising, since that environment has shown to be quite distinct from the rest. In a previous study, WA92 was the only environment where neither of the two major yield QTL on chromosomes 2H and 3H showed a significant effect (Malosetti et al., 2004). Again, a good understanding of the major drivers of the G×E provides insights on how to enhance the power of the prediction models by identifying the most relevant information to add to the existing prediction machinery. In other words, while developing the prediction models, it is important to keep in mind the basic classical questions regarding G×E and its causes (van Eeuwijk et al., 2015, this issue).

The inclusion of environmental information in the prediction model brings in as an additional important feature the opportunity to enhance the statistical models with biological knowledge, moving from purely statistical to eco-physiological prediction models (van Eeuwijk et al., 2005). The selection of key environmental covariables at specific crop stages as predictors in the statistical model is a first form of incorporating biology into the statistical model. That was what our GE-KE model implicitly did. However, that is just the tip of the iceberg. A much stronger interaction with biology is possible by integrating crop growth modeling in the prediction process of G×E (Cooper et al., 2009; van Eeuwijk et al., 2010). Crop growth models (CGM) predict responses via functions of properties of the genotypes and the environments, and so, they offer a promising platform for prediction of complex non-linear responses over environmental gradients. Examples of the integration of molecular markers and CGM to predict complex phenotypic responses have been published (Reymond et al., 2003; Yin et al., 2005; Chenu et al., 2009). In all cases, the basic idea was to use QTLs to shape CGM parameters that in turn, and in combination with the environmental inputs, produced the final predicted phenotype. This can be regarded as a simplified genomic prediction model integrated with CGM (the difference being that only few genotypic regressors enter the model). Heslot et al. (2013a) did use CGM but only to define environmental stress covariables to feed in the prediction model. However, they did not use the CGM to obtain predictions directly from it. A general discussion of the integration of CGM in genomic prediction can be found in Bustos-Korts et al. (2015), and a first proof of the concept in Technow et al. (2015). In the latter, DH lines were simulated in a set up analogous to our disconnected design, where 50 DH lines were used to estimate model parameters, and 1500 DH lines were predicted in an observed environment $(G^u; E^t)$ and in a new environment $(G^u; E^u)$. Mean prediction accuracies by their best method, an approximate Bayesian computation method (Tavaré et al., 1997), were around 0.8 and 0.6 for $(G^u; E^t)$ and $(G^u; E^u)$ respectively, showing the potential of the approach.

5.5. Future work

We regard the integration of CGM as an exciting area of research for prediction of complex phenotypes, and foresee it will gain momentum. As a matter of fact, the inherent property of CGM of regarding traits as a dynamic system of interconnected components that mediates between inputs and the final output broadens even further the possible models to explore. Complex traits are indeed a dynamic system that results from the integration of processes at different level of biological organization, and those can be described by sets of interconnected differential equations (Sun and Wu, 2015). Therefore, prediction models do not necessarily need to rely exclusively on covariables related with the DNA sequence, but other levels of information could potentially be integrated. The challenges ahead are computational on the one side, and more importantly, on making choices of the right level of detail of the models to tailor the specific prediction problem (van Eeuwijk, 2015). This is an area we are currently starting to explore. Highly related with it, is the research in multi-trait and multi-population prediction problems, including the investigation of design issues of training sets and the development of computationally efficient alternatives to the standard mixed model approach used here. We also see the need of investigating the prediction not only of the mean performance, but also of the expected consistency of that performance (i.e. stability). That is an important target in plant breeding, especially when targeting difficult/harsh environments, where is important to select genotypes that guarantee a minimum return. Finally, methods should be moved from the lab to the commercial level (both private and public), so an infrastructure to make these methods widely accessible is needed. The challenge of that infrastructure is that it should carry the latest developments and be user-friendly. We would like to recommend the Integrated Breeding Platform (https://www.integratedbreeding.net/) as an interesting option to explore in the search for such an infrastructure.

Acknowledgements

The work of MM and FvE was conducted as part of the Integrated Breeding Platform phase II project (Bill & Melinda Gates Foundation). We thank the two anonymous reviewers for their constructive comments that improved the presentation of this manuscript.

.

Chapter 6

Modelling the genetic basis of adaptation and stability of the EU-WHEALBI-barley collection using exome capture and common garden experiments in contrasting environments

D. Bustos-Korts[1,2], J. Russell[6], A. Tondelli[3], L., Cattivelli[3], I. Dawson[6], L. Rossini[5], N. Trabanco[5], C. Ferrandi[5], D. Guerra[3], S. Kyriakidis[6], A. Booth[6], F. Strozzi[5], E. Nicolazzi[5], H. Ozkan, E. Çakır[4], E. Yakışır[4], M. Molnar[4], M. Megyeri[4], P. Miko[3], S. Delbono[3], R. Waugh[6], F. van Eeuwijk[1]

1. Biometris, Wageningen University & Research Centre

2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)

3 Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

4 Magyar Tudomanyos Akademia Agrartudomanyi Kutatokozpont

5 Parco Tecnologico Padano

6 The James Hutton Institute

Abbreviations:

AMMI: additive main effects and multiplicative interaction model

BAM files: binary version of a tab-delimited text file that contains sequence alignment data

DP: the number of reads covering each position

WHEALBI: WHEAt and barley Legacy for Breeding Improvement

GbS: genotyping by sequencing

GWAS: genome-wide association study

GxE: genotype by environment interaction

HTL: haplotype locus

Fst: Fixation index

FP7: The Seventh Framework Programme of the European Commission

LD: linkage disequilibrium

MQ: mapping quality

QD: quality depth

qPCR: quantitative polymerase chain reaction

QTL: quantitative trait locus

QTLxE: interaction between quantitative trait locus and the environment

TPE: target population of environments

SNP: single nucleotide polymorphism

SKAT: sequence kernel association test

6.1. Background

Barley is the fourth largest cereal crop after maize, rice and wheat (http://faostat.fao.org). It is cultivated across a wide range of environments, making it a relevant crop in the context of climate change (Dawson *et al.*, 2015; Khoury & Achicanoy, 2016). Besides its importance for food security, barley has a historical relevance because, together with wheat, it was a founder crop for agriculture in the Fertile Crescent (Mascher *et al.*, 2016). Barley is regarded as a model species and therefore, knowledge about barley genetics is relevant to other crops (Mayer *et al.*, 2012). Unfortunately, total genetic diversity currently available to breeding programmes has been limited by the bottleneck of the selection process of modern varieties (Dawson *et al.*, 2015).

Broadening the genetic basis of field crops is key to increase the chances of developing varieties that are adapted to a wide range of production environments (Ellis *et al.*, 2000; Tester & Langridge, 2010). An important part of the genetic diversity for barley has been preserved in germplasm banks worldwide (Igartua *et al.*, 1998; Knüpffer, 2009; Muñoz-Amatriaín *et al.*, 2014). However, accessions in germplasm banks are seldom well characterised genotypically and phenotypically, hampering their use in modern breeding programmes (Muñoz-Amatriaín *et al.*, 2014).

One strategy for genotypic characterization of large diversity collections is Exome sequencing, which was first described by (Mascher *et al.*, 2013) and used recently to examine genome wide diversity across a collection of landraces and wild barley accessions. Exome sequencing has the advantage of reducing genomic complexity more than 50-fold, thus dramatically reducing the heavy sequencing and analysis load (Mascher *et al.*, 2013; Uauy, 2017). One of the attempts of intensive genetic characterisation of barley germplasm is shown in (Russell *et al.*, 2016), who used exome sequencing to characterise 267 geographically diverse barley landraces and wild relatives. In the EU-Whealbi-barley collection presented in this paper, the range of diversity that has been genetically characterised with exome sequencing by (Russell *et al.*, 2016) has been expanded to more relevant cultivated germplasm. The EU-Whealbi-barley collection includes 511 genotypes/accessions ranging from wild barley species (*H. agriocriton*, *H. spontaneum*), *H. vulgare* landraces to formally-bred genotypes chosen by consultation with breeders, gene bank curators and researchers. These genotypes originated from a wide range of sites in Europe, Africa, Middle-East and Asia (http://www.whealbi.eu/project/strategy/). The unique EU-Whealbi-barley collection is the first collection with both detailed genotypic and phenotypic data for a number of basic traits like awn length, grain weight, heading date and plant height to address a wide range of

important biological questions that will be addressed by techniques like genome wide association analyses and allele mining specific candidate genes.

Connecting the genotypic and phenotypic information across a range of environments is crucial to orient breeders with respect to the identification of adaptive alleles. In this paper, we propose to use a top-down approach, starting with a phenotype of interest observed across a range of contrasting environments and using genetic analysis to identify candidate genes (Ross-Ibarra *et al.*, 2007). Characterising the genetic basis underlying GxE is not straightforward, especially not for structured populations like the EU-Whealbi-barley collection. Population structure poses a number of statistical methodological challenges to identify genomic regions that are potentially regulating the target traits (Hoffman, 2013; Korte & Farlow, 2013; Vilhjálmsson & Nordborg, 2013). Furthermore, the genotype to phenotype modelling has to deal with the levels of organisation of the barley genome, revealed by the availability of exome sequence, and with complex gene interactions and context-dependencies (Cooper *et al.*, 2009). One approach to integrating genotypic information across genome organisation levels is to assess the impact of gene haplotypes instead of single SNPs (Cockram *et al.*, 2007; Maurer *et al.*, 2016). Such an approach is also used in the supervised test for the joint effects of multiple variants on a phenotype (SKAT method in human genetics, Wu *et al.*, (2011)). The relevance of haplotypes for breeding has to be interpreted in the context of the environmental conditions in which barley is expected to grow (the target population of environments, TPE). Here, we characterise alleles of the EU-Whealbi-barley collection in relation to contrasting environments in Europe and Eurasia. The combined genotypic and phenotypic analysis would allow prioritising genomic regions for further inspection in crosses, relating the diversity of genomic regions to the geographical information of the sites of origin and genotype breeding history. Main objectives In this paper are i) characterise adaptation patterns of the EU-Whealbi-barley collection, containing formally-bred genotypes and landraces using a range of GxE models, and ii) identify genomic regions and candidate genes that drive genotypic differences for awn length, grain weight, plant height and heading date across a number of environments.

6.2. Materials and Methods

6.2.1. Germplasm

Germplasm belonged to the EU-Whealbi-barley collection, which represents an important range of world-wide barley genetic diversity (http://www.whealbi.eu/). Barley genotypes included in the EU-Whealbi barley collection consist of wild barley species (*H. agriocriton*, *H. spontaneum*) and cultivated barley (*H. vulgare).* The set of cultivated barley genotypes

was highly diverse in terms of flowering time habit, row type and country of origin. Phenotypic, passport data and geographical information about the sites of origin can be found at www.whealbi.eu/science/whealbidatabase/. From the 511 genotypes belonging to the EU-Whealbi-barley collection, 403 were successfully characterized with high-quality exome sequence data. From the 403 genotypes, we restricted ourselves to the *H. vulgare* genotypes, leading to 381 genotypes (wild genotypes were not considered in the statistical analysis presented in this paper). From the 381 *H. vulgare* genotypes with sequence data, we removed those that did not have high-quality phenotypic information, arriving at 371 genotypes with high quality genotypic and phenotypic data that were used for the statistical analysis.

6.2.2. Exome sequencing

Library preparation and sequencing

Genomic DNA (gDNA) was extracted from barley leaf material from a single plant for each genotype. DNA samples were checked with a Genomic DNA ScreenTape on Agilent 2200 Tape Station System (Santa Clara, CA, USA) in order to verify gDNA integrity. Samples were quantified by Picogreen assay (Thermo Fisher, CA, USA) and normalised to 20 ng/ul in 10 nM Tris-Hcl (pH 8.0) as suggested in the NimbleGen SeqCap EZ Library SR protocol. The gDNA was fragmented to a size range of 180-200 bp using Covaris microTUBES and a Covaris S220 Instrument (Covaris, MA, USA) and whole genome libraries were prepared according to the Kapa Library Preparation protocol. Libraries were quantified using a Nanodrop (Thermo Fisher, CA, USA) and analysed electrophoretically with an Agilent 2200 Tape Station System using a D1000 ScreenTape. Libraries were pooled in 8-plex and used for the hybridization with the barley SeqCap Ez oligo pool (Design Name: 120426_Barley_BEC_D04, (Mascher *et al.*, 2013) in a thermocycler at 47°C for 48 h. Capture beads were used to pull down the complex of capture oligos and genomic DNA fragments and unbound fragments were removed by washing. Enriched fragments were amplified by PCR and the final library was quantified by qPCR and visualised by Agilent Tape Station. Sequencing libraries were normalised to 2nM, NaOH denatured and used for cluster amplification on the cBot. The clustered flow cells were sequenced on Illumina HiSeq2000 with an 8-plex strategy (i.e. 8 samples per HiSeq lane) with a 100 bp paired-end run module.

Sequence processing and alignment

Sequence quality control was assessed with FastQC (Babraham Institute: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw Illumina reads were then quality trimmed to a base quality of 20 from both ends with Trimmomatic version 0.30 (Bolger *et al.*, 2014). Only correctly paired reads longer than 50 bp were used for further processing. Trimmed reads were then mapped to the reference genome (provided by Martin Mascher on September 4, 2015: file 150831_barley_pseudomolecules_parts.fasta) with BWA version 0.7.5a using the mem algorithm with default parameters (Li & Durbin, 2009). In total, about 24 million reads were mapped to the reference genome. The resulting BAM files were sorted with Samtools (http://samtools.sourceforge.net/) and duplicate reads marked using Samblaster (http://www.htslib.org/).

SNP calling and validation

Variant calling and realignment around indels were performed with GATK version 2.7.4 (https://www.broadinstitute.org/gatk/). All the final BAM files were processed together using GATK UnifiedGenotyper with default parameters and minimum base quality of 30. The raw variant calls produced were initially hard filtered by requiring QD > 30.0, MQ > 40.0 and sample DP >= 10.

Phenotypic data

Phenotypic data consisted of measurements of awn length, grain weight, heading date and plant height. These traits were chosen because they represent important adaptive traits that are commonly considered during the breeding process. The phenotypic data are available in https://wheat-urgi.versailles.inra.fr/Projects/Whealbi. Field trials were sown in Dundee (Scotland, winter and spring), Martonvasar, (Hungary, winter and spring), Fiorenzuola d'Arda (Italy, winter) and Adana (Turkey). The Turkish trial was irrigated, all other trials were rain fed. For more details about environments explored with the field trials, see Table 1.

Field trials were set up as augmented partially-replicated designs. In each field trial, a subset of the genotypes (about 20%) was replicated twice. This subset was a random draw from all genotypes and therefore differed from one experiment to the next. To link the experiments and account for spatial effects, two check cultivars were included, the winter cultivar Meridian and the spring cultivar Irina. These cultivars were chosen because they represent recently released elite cultivars grown across Europe. The check cultivars were assigned to eight plots per experiment. To distribute the checks uniformly over the field, the

full replicate was subdivided into eight incomplete blocks, following the principles of an alpha design. The 20% genotypes that were replicated twice were grouped together in another two incomplete blocks. Plots consisted of 2 rows of 1.5 m long, with 30 seeds per row. In the field, experimental plots were separated from each other by a buffer row with a wheat variety.

Environmental characterizations for field trials were collected from nearby meteorological stations. Photoperiod was calculated for each site of origin and field trial using latitude and the day of the year, as proposed by (Sellers W.D., 1965) and implemented in the 'Daylength' procedure in GenStat 18 (VSN-International, 2016). Environmental variables were summarised both for the whole growing season and for the first half of the growing season, as described in Table 1.

Table 1. Environmental characterization of the field trials. Rainfall was calculated as the sum over the growing season, Tmean is the mean temperature during the growing season, Thermal time is the sum of degree days during the growing season, photop. min and photop max are the minimum and maximum day length during the growing season and vern_days is the number of days with vernalizing temperatures (≥4ºC and ≤9ºC).

Planting	Env	Planting Date	Location	Lat.	Long.	Rain (mm)	Tmean (ºC)	Thermal time (ºCd)	Photop. Min (h)	Photop. Max (h)	Vern days
Spring	BS2JHI	03-03-15	Scotland	56.5 ºN	3.1 ºW	239	10.70	1213.34	12.95	17.18	4
	BS3ATK	11-03-15	Hungary	47.3 ºN	18.8 ºE	105	13.34	1293.55	11.60	15.60	3
Winter	BW1CRA	27-10-14	Italy	44.9 ºN	9.9 ºE	651	8.96	1952.52	8.70	15.12	59
	BW2JHI	29-10-14	Scotland	56.5 ºN	3.1 ºW	450	7.14	1741.29	6.82	17.18	75
	BW3ATK	20-10-14	Hungary	47.3 ºN	18.8 ºE	315	8.31	2150.25	8.39	15.61	59
	BW4TUR	12-11-14	Turkey	37.1 ºN	35.8 ºE	215	6.32	1420.20	9.52	14.45	39

6.2.3. Phenotype adjusted means

Adjusted means were calculated after fitting a model using two-dimensional (2D) penalised splines (P-splines) for correction of spatial trends and heterogeneity, following the methodology described by (Rodríguez-Álvarez *et al.*, 2016) and in (Velazco *et al.*, 2017) and implemented in the SpATS package (https://cran.r-project.org/package=SpATS). The following model was used;

$$y = f(u, v) + Z_r c_r + Z_c c_c + X_g c_g + \varepsilon \quad (1)$$

In model (1), $f(u,v)$ is a smooth two-dimensional trend jointly defined over the row and the column directions. The bivariate surface can be decomposed in three different spatial components (Lee & Durbán, 2011): (a) a component that contains the smooth main effect (smooth trend) along one of rows, (b) a component that contains the smooth main effect (smooth trend) along the columns and (c) a smooth interaction component (sum of the linear-by-smooth interaction components and the smooth-by-smooth interaction component). $Z_r c_r$ represents a design matrix and independent random effects for the rows, with $c_r \sim N(0, \sigma_r^2)$, while $Z_c c_c$ does the same for the columns, with $c_c \sim N(0, \sigma_c^2)$. $X_g c_g$ represents the design matrix and the fixed effects for the 513 genotypes (511 Whealbi genotypes, plus the two elite cultivars used as checks) and ε is a residual, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. To quantify the importance of the genotypic variance with respect to the total variance, we calculated so-called generalised heritabilities, following again (Rodríguez-Álvarez *et al.*, 2016)**.**

6.2.4. Population structure

To characterise population structure, a relationship matrix *A* was calculated from the SNPs following (Patterson *et al.*, 2006), after removing SNPs with minor allele frequency below 0.05.

$$A = \frac{XX'}{n_m} \qquad (2)$$

The *A* matrix is of dimensions number of genotypes by number of genotypes and is proportional to the genetic covariance among individuals. *X* is a matrix of dimensions number of genotypes by number of SNPs, whose entries are marker scores, coded 0, 1 or 2, representing the number of copies of the minor allele. The marker scores were standardised as in (Patterson *et al.*, 2006). In (2), n_m is the number of markers. To infer the number of subpopulations present in the EU-Whealbi-barley collection, we calculated the number of significant principal components after applying a spectral decomposition to the matrix *A*, where the significance was assessed following (Patterson *et al.*, 2006). The number of subpopulations is then the number of significant components plus one. The calculations were done in Genstat 18 (VSN-International, 2016).

Genotypes were grouped and assigned to subpopulations using a hierarchical clustering procedure applied to the significant principal components, following (Odong *et al.*, 2013). The cut-off for the dendrogram was chosen such that the number of subpopulations was equal to the number of principal components plus one.

6.2.5. GxE interaction analysis by AMMI analysis

To characterize GxE interaction patterns and explore possible adaptations, we fitted the Additive Main Effects and Multiplicative Interaction (AMMI) model (Gauch & Zobel, 1997; Gauch, 2013; van Eeuwijk *et al.*, 2016). The adjusted means, after spatial correction, were organised in a genotype by environment table of means, and the following AMMI model fitted, (Equation 3).

$$y_{ij} = \mu + E_j + G_i + \sum_{m=1}^{M} b_{im} z_{jm} + \epsilon_{ij} \qquad (3)$$

In Equation (3), $\underline{y}_{ij}$ represents the phenotype of the i^{th} genotype in the j^{th} environment, μ stands for the intercept, G_i is the fixed effect of the ith genotype and E_j is the fixed effect of the j^{th} environment. The interaction in an AMMI model is described by M multiplicative terms. There are many ways for deciding how to choose M (van Eeuwijk *et al.*, 2016), but we chose the relatively straightforward method by (Gollob, 1968). Each multiplicative term in (3) is formed by the product of a genotypic sensitivity b_{im} (genotypic score) and an environmental score z_{jm}. Finally, ϵ_{ij} is a residual term, that contains the part of the two-way analysis of variance interaction that is not explained by the AMMI interaction terms and a contribution of the plot error, the latter will be ignored in this paper. GxE interactions, using the AMMI model, were visualised in the form of biplots, which are just scatter plots for genotypic and environmental scores. These scores define the coordinates for genotypic and environmental vectors, and provide approximations to genetic variances (length environmental vectors), genetic correlations (angle between environmental vectors), genotypic stabilities (length genotypic vectors), and genotypic adaptations (projections of genotypic vectors on environmental vectors).

6.2.6. GxE interaction analysis introducing genotypic groups

To study the relevance of groupings of genotypes or identified subpopulations for describing genotypic differences and GxE interaction patterns, we fitted various two-way mixed models with and without groupings of genotypes to see to which extent the genotypic and GxE interaction variance, V_G and V_{GE}, was reduced by the introduction of genotype groupings. As a kind of null model in which no genotype groups are distinguished we take the following model for the phenotype of genotype *i* in environment *j*:

$$y_{ij} = \mu + E_j + G_i + GE_{ij} + \varepsilon_{ij} \quad (3.1)$$

where μ is an intercept term, E_j the fixed environmental main effect, G_i a random genotypic main effect, GE_{ij} a random genotype by environment interaction, and ε_{ij} a

residual term that is confounded with the GE_{ij} term when model (3.1) is fitted to a two-way genotype by environment table of adjusted means. Model (3) provides estimates for the genotypic and GxE variance, V_G and V_{GE}, when no grouping of genotypes is introduced to describe GxE interaction. When such terms are introduced, these variances will be reduced.

Upon introduction of groupings, clusters, or subpopulations for genotypes, the genetic variance of G_i and GE_{ij} as occurring in model (3.1) will be partitioned into a part expressing the variance between groups for the genotypic main effect and the GxE interaction and another part that describes the variation not accounted for by the grouping. Two groupings of genotypes that we investigated were one due to subpopulations as obtained from the clustering of the kinship principal components and another one due to the contrast between two- and six-row barleys, (models 3.2 and 3.3).

$$y_{ijk} = \mu + E_j + S_k + G_{ik} + SE_{jk} + GE_{ijk} + \varepsilon_{ijk} \quad (3.2)$$

$$y_{ijl} = \mu + E_j + R_l + G_{il} + RE_{jl} + GE_{ijl} + \varepsilon_{ijl} \quad (3.3)$$

In model (3.2), S_k is the fixed effect of subpopulations and SE_{jk} is the subpopulation by environment interaction. Equation (3.3) considers row type to represent population sub-structuring, with R_l representing the fixed effect of row type and RE_{jl} the row type by environment interaction.

Model (3.2) can be modified to become a model for a genome wide association scan by the introduction of a QTL and a QTLxE term (see section 6.2.9). Effectively, the genotypic main effect is then split up in a QTL main effect and a residual genotypic main effect, while the GxE interaction term is split in a QTLxE term and residual GxE interaction. Model (3.2) contains a correction for population (sub)structure, S_k, with respect to the main effect of a QTL and a correction for subpopulation by environment interaction, SE_{jk}, with respect to QTLxE. Examples of applications of model (3.2) are given in (Millet *et al.*, 2016; Thoen *et al.*, 2017) and will also be discussed below.

In place of the subpopulation factor to correct for population structure in model (3.2), we can also use the principal components approximating the kinship structure to partition the genotypic main effect and the GxE interaction:

$$y_{ij} = \mu + E_j + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_p^G) + G_i + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_{jp}^{GE}) + GE_{ij} + \varepsilon_{ij} \qquad (3.4)$$

Equation (3.4) is as Equation (3.2), with S_k replaced by $\sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_p^G)$ and SE_{jk} replaced by $\sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_{jp}^{GE})$, where x_{ip}^{PC} stands for the genotype specific scores on the *p*-th

kinship principal component, with $p=1...P$, and β_p^G and β_{jp}^{GE} the corresponding fixed regression coefficients for these principal components correcting for population structure with respect to the genotype main effect and the GxE interaction, respectively. Model (3.4) provides another way of estimating the effect of population structure on V_G and V_{GE}.

Models (3.2) and (3.4) can be combined to study the contribution of subpopulations and row type, after accounting for population structure by kinship principal components, as presented in Equations (3.5) and (3.6).

$$y_{ij} = \mu + E_j + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_p^G) + S_k + G_{ik} + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_{jp}^{GE}) + SE_{jk} + GE_{ij} + \varepsilon_{ij} \quad (3.5)$$

$$y_{ij} = \mu + E_j + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_p^G) + R_l + G_i + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_{jp}^{GE}) + RE_{jl} + GE_{ij} + \varepsilon_{ij} \quad (3.6)$$

When the kinship principal components capture the population structure as contained in the subpopulation factor well, expressed as S_k and SE_{jk} in model (3.5), we don't expect any additional variation to be explained by those latter terms.

6.2.7. GxE interaction and adaptation analysis for flowering time

In addition to an AMMI analysis (section 2.6) and the mixed model analyses (section 2.7), we investigated adaptation to a given environment by defining it for individual genotypes by the difference between their flowering time and a target flowering time in that environment (flowering time ideotype/reference). This ideotype was defined as the average heading time of a number of modern cultivars considered to be well adapted to the trial conditions (Table S1). Adaptation to each of the field trials was compared between the following groups using analysis of variance; i) spring-formally-bred cultivars (97 genotypes), ii) winter-formally-bred cultivars (26 genotypes) and iii) spring landraces (47 genotypes, from which 39 had a known site of origin).

6.2.8. Genome Wide Association Scans (GWAS)

Single locus GWAS model

Models (3.2) and (3.4) in the previous section were modified to become a model for a genome wide association scan by the introduction of a QTL and a QTLxE term. Effectively, the genotypic main effect is then split up in a QTL main effect and a residual genotypic main effect, while the GxE interaction term is split in a QTLxE term and residual GxE interaction.

Model (3.2) contains a correction for population (sub)structure, S_k, with respect to the main effect of a QTL and a correction for subpopulation by environment interaction, SE_{jk}, with respect to QTLxE. Examples of applications of model (3.2) are given in (Millet *et al.*, 2016; Thoen *et al.*, 2017) and will also be discussed below.

For our GWAS analyses, we preferred to depart from model (3.4). To Identify genomic regions and candidate genes that drive differences between genotypes for awn length, grain weight, plant height and heading date across a number of environments, we used the following single locus mixed model to scan the genome:

$$y_{ij} = \mu + E_j + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_p^G) + G_i + \sum_{p=1}^{P}(x_{ip}^{PC}\ \beta_{jp}^{GE}) + x_i^{SNP}\beta_j^{SNP} + GE_{ij} + \varepsilon_{ij} \qquad (4)$$

In comparison to model (3.4), model (4) contains a term for the fixed SNP effect β_j^{SNP}, while x_i^{SNP} contains the marker information. This means that we fit a QTL that is allowed to have an environment specific effect, or, we fitted at each marker position a QTL main effect and a QTLxE term simultaneously. The test for β_j^{SNP} being zero in all environments or being non zero in at least one environment was a Wald test (Welham & Thompson, 1997; Boer *et al.*, 2007). The fixed structure correction terms were not tested for inclusion are tested for, as they are supposed to be in the model for an unbiased estimate of the QTL effect. The random terms for G_i and GE_{ij} have variances V_G and V_{GE} that were restricted to be positive. The error term ε_{ij} was confounded with the GE_{ij} term, a fact that we will ignore. Model (4) was fitted in Genstat18 (VSN-International, 2016).

A significance threshold for testing for marker – trait associations was established using a Bonferroni correction based on the number of independently segregating chromosome segments (Li & Ji, 2005), where we calculated this number from the local linkage disequilibrium (LD) decay extension.

Local Linkage Disequilibrium (LD)

To characterise local LD, markers were thinned to one every 25 SNPs. A sliding window of 500 thinned SNPs was then used to calculate the local LD corrected by population structure for each SNP position, following the method proposed by (Mangin *et al.*, 2012), implemented in the LDcorSV R package (https://cran.r-project.org/package=LDcorSV). Consecutive sliding windows had an overlap of 475 out of the 500 SNPs.

Subsequently, SNPs within a distance of 20000 bp were binned and the 0.95 quantile for the population-structure corrected LD was calculated for each bin. To these 0.95 quantiles, a

cubic spline was fitted that describes the LD decay as a function of the distance between markers. The maximum of the spline second derivative was used as an indicator of the distance at which LD starts to stabilise. For each sliding window along the genome, the LD decay distance was calculated and the median of these distances was used as an estimate for genome wide LD decay. From the latter, the number of independently segregating segments in the genome was estimated. In addition to this Bonferroni correction, we applied a genomic control correction (Devlin & Roeder, 1999; Kang *et al.*, 2008).

QTL identification

After the single locus GWAS scan by the model (3.7) using the above-described testing procedure, we identified a number of significant SNPs. All significant SNPs were first collated in a long list of QTL candidates. This long list was pruned as follows. First, we looked at the highest significant SNP in the total long list. For this SNP, the local LD determined a QTL region to the left and right. Next, all SNPs in that region were removed from the long list as they were seen as belonging to the same QTL region. Subsequently, in the reduced list, once again we looked for the highest significant SNP, a QTL region was defined from the local LD and the SNPs in this region were removed from the long list. The procedure was repeated until no further SNPs were significant.

6.2.9. Allele mining

After identification of QTL regions, using the new genome assembly and annotation (Beier *et al.*, 2017; Mascher *et al.*, 2017) the genes inside the QTL position interval were registered and potential candidate genes were identified using a range of different approaches depending on the trait. Published known candidate genes were used, as well as examining homologies to genes in other cereal crops such as rice, maize and wheat. Additionally for grain weight, where no candidate genes have been published in barley, searches using the recent version of the barley genome accessible through the IPK Barley Blast Server (http://webblast.ipk-gatersleben.de/barley_ibsc/) and available gene expression data (http://apex.ipk-gatersleben.de/apex/f?p=284:10::::::) were used.

For those traits that do have published candidate genes (awn length, heading date and plant height), we restricted ourselves to the list of genes contained in those publications. Thus, a more directed approach was used for these three traits, compared to grain weight, where all the candidate genes within the QTL region were assessed. In the case of awn length, the following studies were used: (Tanaka *et al.*, 2012; Luo *et al.*, 2013; Toriba & Hirano, 2014; Gu *et al.*, 2015; Hua *et al.*, 2015; Bessho-Uehara *et al.*, 2016; Jin *et al.*, 2016;

Magwa *et al.*, 2016; Liller *et al.*, 2017), gene function and expression profile, focusing mainly on transcription factors involved in plant development and expressed in inflorescence. In the case of flowering time, we used (Drosse *et al.*, 2014; Hill & Li, 2016) as a reference. To select candidate genes we based ourselves on (Chen *et al.*, 2009; Jia *et al.*, 2009; Dockter *et al.*, 2014; Alqudah *et al.*, 2016; Gruszka *et al.*, 2016; Mikołajczak *et al.*, 2017). We selected those annotated genes having a role in hormones signaling/transport/metabolism, plant architecture regulation or belonging to gene families already known to regulate mechanism related to plant development and growth.

Inside candidate gene regions, we constructed haplotypes combining all SNPs with MAF>0.03 by the method of (Gabriel *et al.*, 2002), as implemented in Haploview (Barrett *et al.*, 2005). Briefly, candidate genes are scanned for historical recombination events, calculated from the local LD patterns. Then, haplotype blocks are defined as genomic regions within genes, over which most of the informative SNP pairs show little evidence of recombination (high LD). We will refer to the genomic regions containing those haplotypes within genes as haplotype loci (HTLs). As some of the haplotype loci had multiple alleles at very low frequencies, we pooled those low-frequency haplotype alleles to achieve a reduced number of alleles. In the pooling step, we first calculated the simple matching distances among genotypes, based on all the SNPs considered in the haplotype blocks (Sokal, 1958). If two haplotypes share exactly the same profile for the underlying SNP alleles, their simple matching coefficient would be 1.0 and if all SNPs have the opposite SNP allele, their coefficient would be 0.0. For convenience, we applied hierarchical clustering (Ward's method) to pool low-frequency haplotypes with the haplotype that is most similar to it. A dendrogram cut-off threshold was defined for forming groups such that each class contained at least 10 genotypes.

6.2.10. Multi-locus GWAS model based on haplotype loci

To evaluate the contribution of haplotype loci (HTLs) to the genotypic main effect and the GxE we fitted single and multi-locus models containing terms for the main effect of haplotype loci, $x_{ih}^{HTL}\beta_h^{HTL}$ and for their interaction with the environment, $x_{ih}^{HTL}\beta_{jh}^{HTL}$, where x_{ih}^{HTL} stands for the ith row (genotype) of a genotypes by alleles matrix containing the haplotype allelic information for all genotypes at the h-th HTL and β_h^{HTL} is a vector containing the corresponding HTL main effects and β_{jh}^{HTL} is the j^{th} column (environment) in an alleles by environments matrix containing the interaction effects with the environment for the h-th HTL. The sets H^G and H^{GE} indicate the collection of HTLs with a main effect and those with an interaction with the environment, respectively, where the latter set is a subset of the first. For the single locus model, those sets contain one and the same HTL.

The multi HTL GWAS model is then:

$$y_{ij} = \mu + E_j + \sum_{p=1}^{P} (x_{ip}^{PC}\ \beta_p^G) + \sum_{h \in H^G} (x_{ih}^{HTL}\ \beta_h^{HTL}) + G_i +$$

$$\sum_{p=1}^{P} (x_{ip}^{PC}\ \beta_{jp}^{GE}) + \sum_{h \in H^{GE}} (x_{ih}^{HTL}\ \beta_{jh}^{HTL}) + GE_{ij} + \varepsilon_{ij} \quad (5)$$

The significance of HTLs was established by Wald tests without correction for multiple testing. To decide which HTL would be included in model (5), we used a forward selection procedure. Here, we started with the most significant HTL from the single-locus HTL model. We allowed only one HTL per QTL region to enter the multi HTL model. Thus, all HTLs that belonged to the same QTL region as the HTL that was already included in model (5) were excluded from the list of candidates to enter the next model term. We continued including HTLs until additional haplotypes were no longer significant ($p<0.05$).

From the fitted multi-HTL model we calculated the HTL contributions to the genotype effects per environment by adding the estimated HTL effects for each environment:

$$\hat{y}_{ij}^{HTL} = \sum_{h \in H^G} (x_{ih}^{HTL}\ \hat{\beta}_h^{HTL}) + \sum_{h \in H^{GE}} (x_{ih}^{HTL}\ \hat{\beta}_{jh}^{HTL}) \quad (6)$$

The HTL fitted values $\hat{y}_{ij}^{HTL}$ were used to construct groups of genotypes that show similar adaptation patterns across the six field trials. Euclidean distances were calculated and a hierarchical clustering (Ward's method) was used to assign genotypes to adaptation classes.

To assess the amount of GxE that was explained by those adaptation classes, we fitted a mixed model similar to model (3.5), where the subpopulations are now replaced by adaptation classes to give:

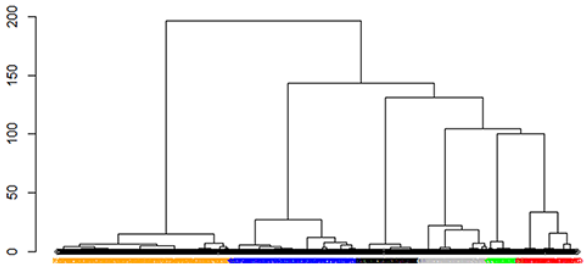
$$y_{ijk} = \mu + E_j + \sum_{p=1}^{P} (x_{ip}^{PC}\ \beta_p^G) + A_k + G_{ik} + \sum_{p=1}^{P} (x_{ip}^{PC}\ \beta_{jp}^{GE}) + AE_{jk} + GE_{ijk} + \varepsilon_{ijk} \quad (3.7)$$

Akaike's Information Criterion (Akaike, 1973) and the reduction of the magnitude of V_G and V_{GE} were used as criteria to quantify the efficiency of HTLs to describe adaptation. We also linked these groups to the geographical information from the sites of origin, assessing per site of origin, the proportion of genotypes that belonged to a certain adaptation group.

6.3. Results

6.3.1. Population structure

The collection of 371 *H. vulgare* genotypes was highly structured. A spectral decomposition of the kinship matrix showed five eigenvalues to be significant. The first eigenvector, say principal component, explained 42% of the variation and it separated genotypes according to their row type (Figures S1 and S2). The six subpopulations identified by cluster analysis on the first five principal components were related to row type, growth habit and country of origin for the genotypes (accessions) (Figures 1 and S1). The first subpopulation, number 1, contained mainly 6-rowed winter formally-bred genotypes, while subpopulation 2 was determined by 2-rowed spring formally-bred genotypes (Figure 1). Subpopulation 3 corresponded to 2-rowed landraces from the Near-East and North Africa. Subpopulation 4 was dominated by 6-rowed Asian genotypes. Subpopulation 5 was clearly separated from the others and contained 2 and 6-rowed genotypes, with almost all Ethiopian landraces included in this group (Figure S3, coinciding with (Muñoz-Amatriaín *et al.*, 2014)). Subpopulation 6 had mainly 6-rowed landraces from the Near-East and North Africa.



Data	Category	1	2	3	4	5	6	Total
Breeding history	Formally bred spring	11	85		1			97
	Formally bred winter	26						26
	Landrace Spring	20	11	4	2	10		47
	Other/Unknown	35	28	38	45	33	22	201
Row type	2	20	122	36	4		9	191
	6	72	2	6	42	43	10	175
	deficiens						3	3
	intermedium				2			2
	Total	**92**	**124**	**42**	**48**	**43**	**22**	**371**

Figure 1. Subpopulations identified with the cluster analysis of the significant Kinship principal components.

6.3.2. Environmental conditions at the field trials

Field trial conditions were clearly different between spring-planted and winter-planted trials. Most winter-planted trials (BW1CRA, BW2JHI, BW3ATK) received considerable lower rainfall than spring trials (BS2JHI and BS3ATK) with the exception of BW4TUR, which had only 215 mm, comparable to spring-planted trials (Table 1). The most humid environment was BW1CRA (651 mm) and the driest environment was BS3ATK (105 mm). Mean photoperiod was shorter in winter-planted than in spring-planted trials, but the photoperiodic range was larger for winter-planted trials (maximum photoperiod was similar between spring and winter trials). Winter-planted trials had a much larger number of days with vernalizing temperatures when compared to spring-planted trials. (Table 1).

6.3.3. Phenotypic analysis

Spring and winter trials showed similar population mean for awn length, grain weight and plant height (the arithmetic means were; 9.9 cm for awn length in spring trials, 10.6 cm for awn length in winter trials, 52.3 mg g^{-1} for grain weight in spring trials, 50.0 mg g^{-1} for grain weight in winter trials, 79.7 cm for height in spring trials and 78.2 cm for height in winter trials. For heading date, spring trials were clearly earlier than winter trials (77.9 days for spring trials and 156.7 days for winter trials, Figure S4, lower left plot). Subpopulations also differed in their mean performance across environments. Subpopulation 4 (6-rowed landraces and cultivars from Asia, coloured grey on Figure S4) showed shorter awn length and grain weight than the other subpopulations (the arithmetic mean for subpopulation 4 was 8.5 whereas awn length in other subpopulations was between 9.3 and 11.3 cm). Subpopulations 4 and 5, with a high proportion of 6-rowed genotypes, had a smaller grain weight (grey and green on Figure S4). However, subpopulation 1, also with a high proportion of 6-rowed genotypes, had a similar grain weight to subpopulations 2, 3 and 6.

The importance of GxE was quantified as the percentage of GxE with respect to the total genetic variance (G and GxE, model 3.1). None of the four traits analysed showed a large GxE; for awn length, grain weight, heading date and plant height, GxE was 45, 39, 37 and 34% of the total genetic variance, respectively (Tables 3 and 4, model 3.1). We quantified the importance of population structure by the difference in genotypic variance, V_G, between models (3.1) and (3.4). Population structure, represented by the first five kinship principal components, reduced the genotypic variance by 37, 31, 38 and 17% for awn length, grain weight, heading date and plant height. The GxE variance, V_{GE}, was reduced by 14, 4, 17 and 22%, respectively by the introduction of kinship principal components that interacted with the environment (Tables 3 and 4, models (3.1) and (3.4)). Adding row type as an additional

fixed term after the kinship principal components hardly changed V_G, except for grain weight, where V_G was reduced by 34% (Table 4, models 3.4 and 3.6). Similarly, adding a factor for subpopulations after the kinship principal components did not change V_G and V_{GE}, as expected because the kinship principal components are assumed to coincide with the subpopulations. V_G was reduced by 3, 5, 1 and 4%, while V_{GE} was reduced by 3, 2, 1 and 1%. (Tables 3 and 4, models 3.4 and 3.5).

In general, subpopulations didn't explain too much of the GxE interaction. This finding is clearly reflected in the AMMI biplots (Figures 2, S5, S6 and S7), where a labelling of the genotypes by subpopulation membership doesn't provide visual separation of subpopulations. Genotypes that appear close together in biplots share adaptation mechanisms, environments that group together share induced stress reactions. For the rules to interpret biplots, see (Kempton, 1984; Malosetti *et al.*, 2013; van Eeuwijk *et al.*, 2016). For awn length, the subpopulation by environment interaction was somehow associated with a positive interaction between subpopulation 4 (6-rowed Asian landraces and cultivars) and BW4TUR (Figure S5). The main environmental contrast for heading date was given by the spring trials (BS2JHI and BS3ATK) versus the winter trials (Figure 2).

Subpopulation 2 (2-rowed spring formally-bred) was almost parallel to the AMMI1 axis, showing some genotypes that had a positive interaction with the spring trials and other genotypes that had a negative interaction with spring trials. Subpopulation 1 (6-rowed winter formally-bred genotypes) tended to align parallel to the AMMI2 axis, with some genotypes showing a positive interaction with BW1CRA. However, this group was very heterogeneous in their distance from the origin in the biplot, indicating different adaptation patterns. When analysing heading date from the perspective of adaptation (adaptation understood as deviations from the ideotypic flowering time in each environment), spring landraces showed an earlier flowering time in spring trials (about four days), compared to the ideotype and winter cultivars flowered later (about five days, Figure 3). Spring cultivars showed to be well adapted to spring trials, coinciding with the ideotype. For the winter trial BW1CRA, genotypes showed considerable variation and all groups achieved heading about four days later than the ideotype. Spring landraces were slightly delayed in BW2JHI (about four days). In all other winter trials, heading date for the three groups coincided with the ideotype. In the case of plant height, a more clear interaction pattern was observed, with subpopulation 3 (2-rowed landraces from the near-east/North Africa) showing a positive interaction with BW1CRA and BW4TUR (Figure S7).

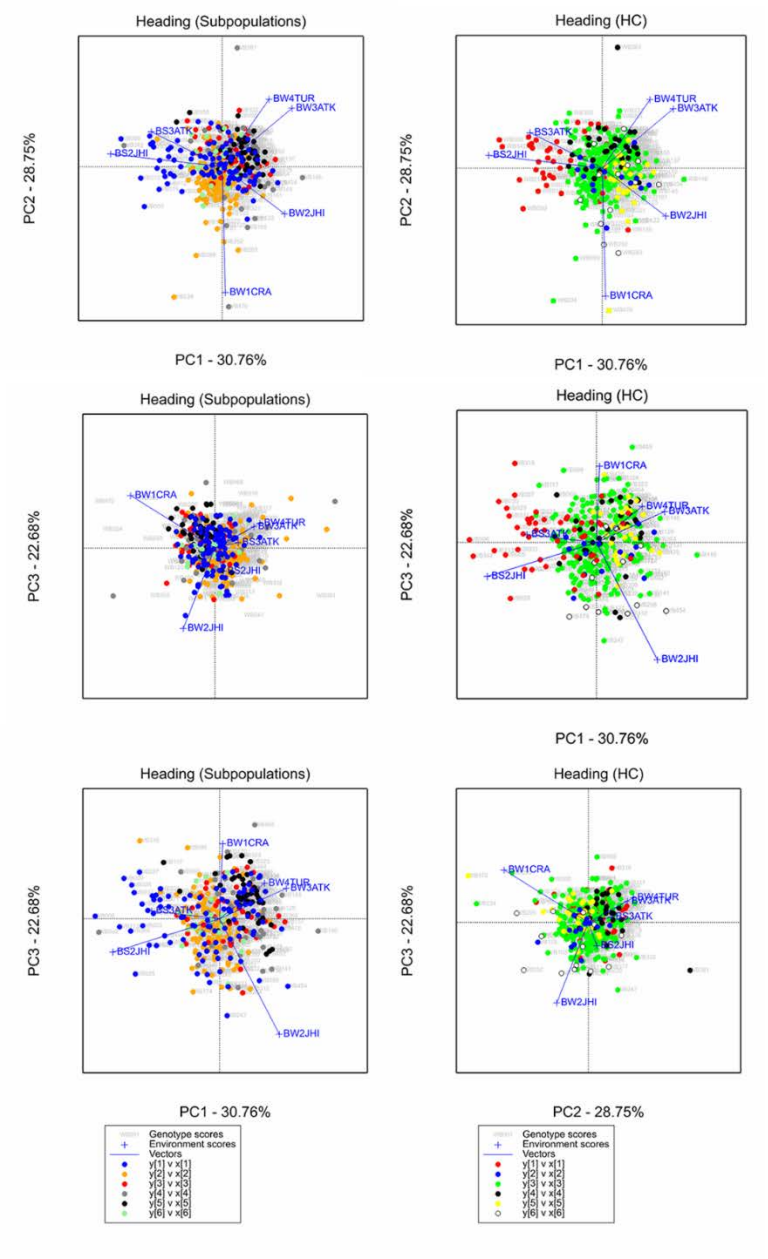


Figure 2. AMMI biplots for heading date. In the left panels, colours show the six subpopulations identified with the cluster analysis. In the right panels, colours show the genotype classes constructed with the haplotypes for heading date genes (same colours as in Figure 9).

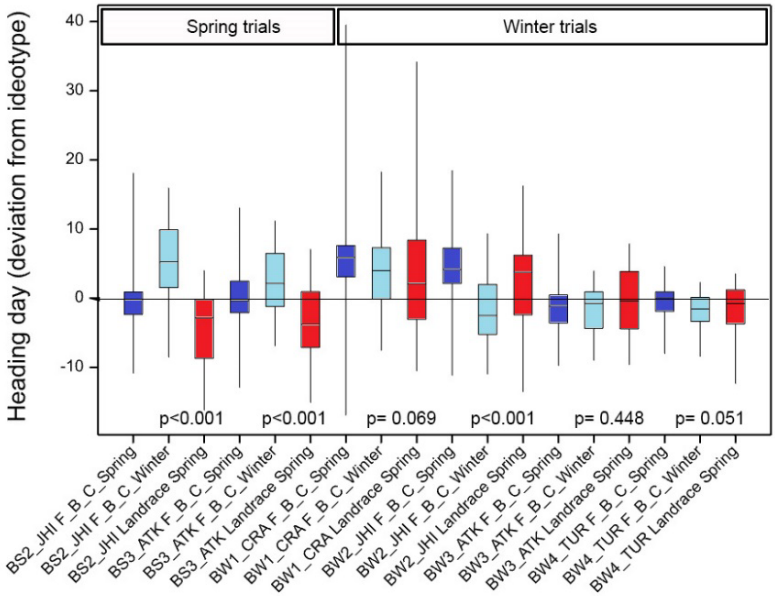


Figure 3. Boxplot of the heading date deviations from the ideotype (target heading date). Dark blue boxes show spring formally-bred cultivars, light blue boxes show winter formally-bred cultivars and red boxes show spring landraces. P-values indicate significance of the ANOVA within environments.

6.3.4. GWAS

LD stabilised at about 20 Mbp, but it was largely variable between chromosomes and also within chromosomes (Figure 4). LD was larger closer to the centromere and it was smaller towards the extremes of the chromosome, where LD decayed very rapidly. The only chromosome that clearly had a different LD pattern was chromosome 5H, which showed a very large LD block covering the first half of the chromosome.

For awn length, 15 QTLs were detected by a single-locus GWAS scan; 1 at the end of chromosome 1H, 3 at the second half of chromosome 2H, 2 at the end of chromosome 3H, 3 at the end of chromosome 4H, 1 at the beginning of chromosome 5H and 6 in chromosome 6H, (Figure 5, Figure S8). For most QTLs, their associated markers showed moderate to high correlations with each other, even when they belonged to different chromosomes (centre of Figure 5, Figure S9). The only exception was the QTL at the end of chromosome 1H, where the marker showed a very low correlation with the QTL detected on the other chromosomes. QTL additive effects differed across environments, ranging from -2.17 to 1.66 cm (21.6 and 16.5% of the population mean). In most cases, QTLxE was related to changes in the

magnitude of the additive effects, without cross-overs. The only exception was the QTL at the end of chromosome 1H, which had a positive effect in BW4TUR and a negative effect in the other environments (except in BW3ATK and BS3ATK, where grain weight and awn length were not observed).

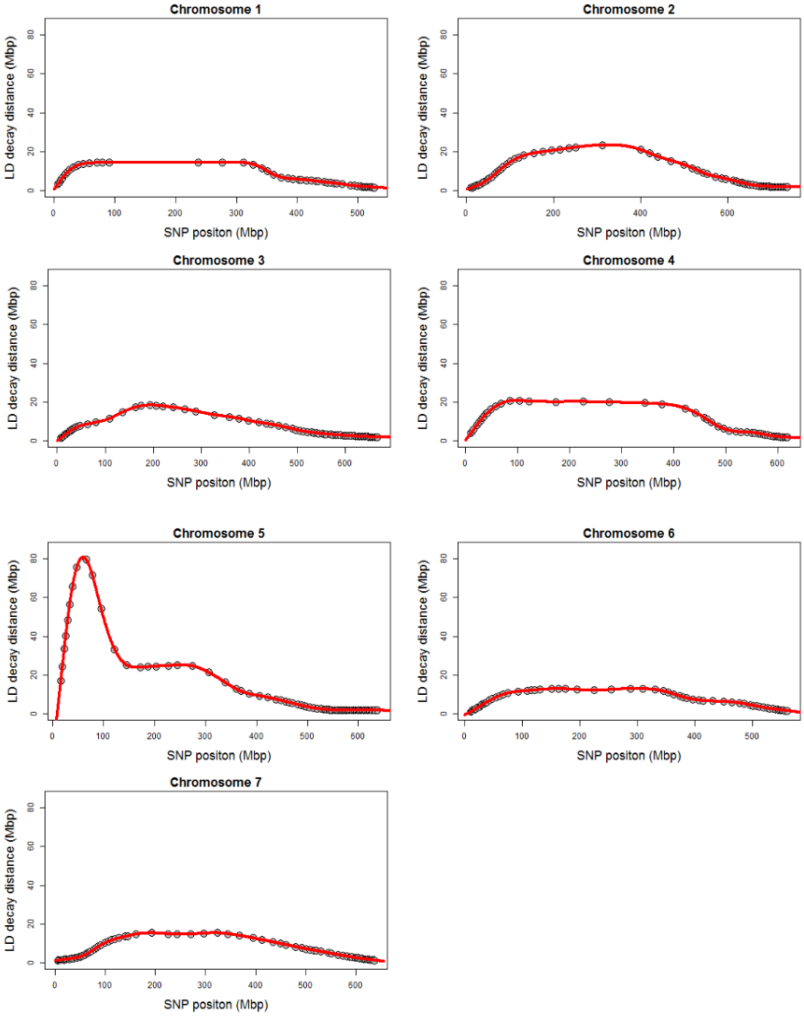


Figure 4. LD decay along the chromosome. Grey dots show the LD decay in each of the sliding windows.

In the case of grain weight, 11 QTLs were detected (3 on chromosome 1H, one on chromosome 2H, four on chromosomes 1H and 4H). Additive effects ranged from -0.82 to 5.13 mg g^{-2} (1.6 to 10.2% of the population mean). In contrast to awn length, QTLs for grain weight showed crossover type of QTLxE, with all QTLs showing a negative effect in BW4TUR and a positive effect in the other environments (except in BS3ATK, where grain weight was not observed).

For heading date, 24 QTLs were detected (2 on chromosome 1H, 13 on chromosome 2H, 2 on chromosome 3H, 1 on chromosome 4H, 4 on chromosome 5H and 2 on chromosome 7H, Figure S8). The QTL at the second half of chromosome 1H (SNP at position 515.2 Mbp) was only 1.13 Mbp from the causal gene Protein FLOWERING LOCUS T (HvPPD-H2). The large QTL close to the HvPPD-H2 locus is in agreement with the contrast between winter (short photoperiod) and spring trials (longer photoperiod) that drives most of the GxE interactions. Most of the QTLs in chromosome 2H were in large LD with each other, suggesting that many of them might be significant just because they were correlated to the causal loci, instead of having a direct effect on the phenotype.

Three of the QTLs on chromosome 2H (SNPs at positions 516.1, 528.8 and 543.4 Mbp) were within a large LD block of around 50 Mb that contains HvCen gene, that is the gene controlling the well-known early maturity locus Eam6 (Comadran *et al.*, 2012). The large number of adjacent QTLs might be the result of high-LD blocks that might have been caused by genetic drift (in the case of landraces) or the selection process (in the case of formally-bred genotypes). In Chromosome 7H (69.5 Mbp), a QTL at about 2 Mb from the FLOWERING LOCUS T 1 (Vrn-H3) gene was detected (Faure *et al.*, 2007; Cockram *et al.*, 2015). An important effect of Vrn-H3 was to be expected because of the difference in the number of vernalizing days between winter and spring trials (Table 1), as previously reported by (Cuesta-Marcos *et al.*, 2008). However, as the Vrn-H3 gene was not covered by the exome sequencing, the QTL detected at 69.5 Mbp might be the result of the LD between the adjacent SNPs and the Vrn-H3 gene. Additive QTLxE effects ranged from -3.55 to 4.30 days (2.2 and 2.7% of the population mean). For heading date, a larger QTLxE was observed compared to the other traits, with most of the QTLs showing a cross-over type of interaction, commonly driven by the contrast between BS2JHI and the other environments, coinciding with the GxE described by the AMMI biplot (Figure 2).

In the case of plant height, 27 QTLs were detected (1 on chromosome 1H, 4 on chromosome 2H, 8 on chromosome 3H, 1 on chromosome 4H, 10 on chromosome 5H, 2 on chromosome 6H and 1 on chromosome 7H. Although a large number of significant loci were observed in chromosome 5H, the peaks in this region were not very clear and for that reason,

it is likely that many of them correspond to loci that are in LD with the actual QTL (Figure S5). This was also reflected in a large block of highly correlated QTLs on chromosome 5H (Figures 9 and S10) and coincides with the wide LD block in this chromosome (Figure 4). Additive effects ranged from -6.55 to 7.01 days (8.3 to 8.9% of the population mean). Most of the QTLs showed little QTLxE interaction, reflected in moderate changes in additive effects across environments, without at crossover type of interaction.

6.3.5. Allele mining

QTL location and its confidence interval determined by the local LD pattern were used to identify candidate genes. A total of 24, 30 and 51 candidate genes were identified for awn length, heading date and height. For grain weight, because there have been no reported candidate genes, all of the genes identified (1439) within the QTL boundaries were subjected to analyses. Only some of the candidate genes were partially or fully covered by the exome sequence data and had a good gene model available (24 for awn length, 968 for grain weight, 23 for heading date and 51 for plant height). As some of the genes that were covered by the exome sequence suggested the presence of more than one haplotype block, 26 HTLs were tested in Equation (8) for awn length, 1473 for grain weight, 29 for heading date and 50 for plant height. From these HTLs, 11 were significant ($p<0.05$) for awn length, 708 for grain weight, 23 for heading date and 35 for plant height (Tables S2-S5, Figures 5-8). Additive effects for the HTLs were in general larger than those of the QTLs, reflecting the advantage of the haplotype analysis with respect to the single SNPs (Figures 5, 6, 7 and 8).

When combining haplotypes in a multi-locus HTLs model like model (3.7), four HTLs were included for awn length, 12 for grain weight, nine for heading date and 11 for plant height (Tables S2, S3, S4, S5). The selected HTLs were more effective in explaining genotype main effect than genotype by environment interaction. The estimate for the genotypic variance was reduced by 18, 43, 30 and 53% for awn length, grain weight, heading date and plant height, whereas the GxE interaction variance was reduced by 6, 14, 16 and 16%, for the traits in the same order (Tables 3 and 4, models 3.4 and 3.8).

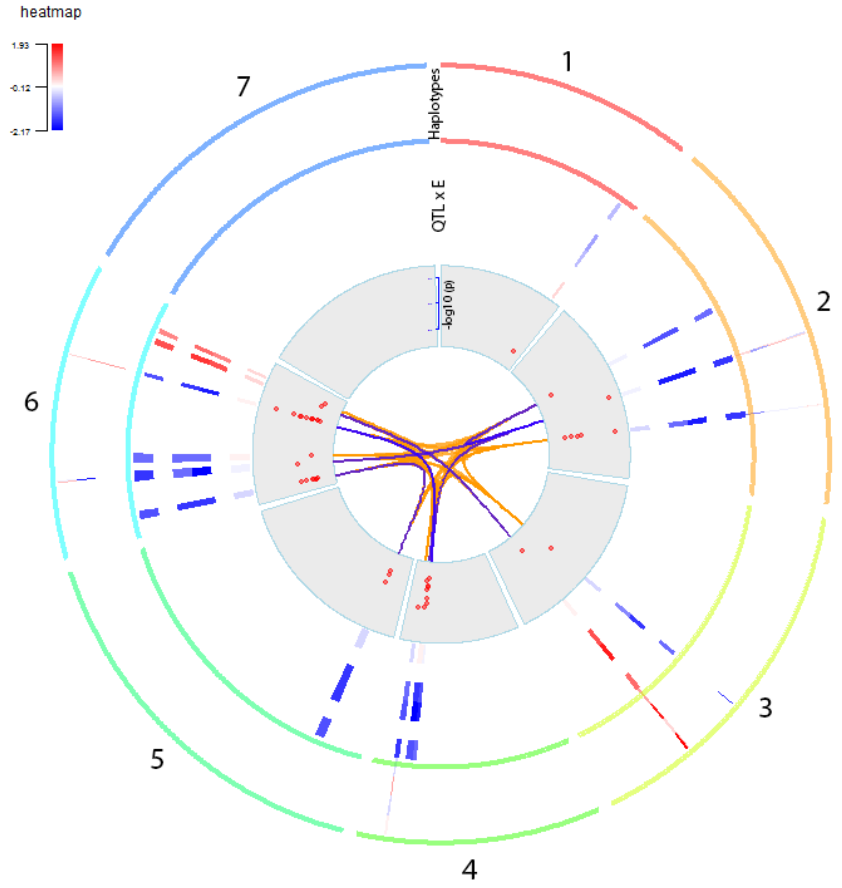


Figure 5. Circos plot for **awn length**, with chromosome 1 shown in red and chromosome 7 shown in blue (clockwise). In the plot centre, orange lines connecting SNPs show absolute SNP correlations between 0.4 and 0.7 and dark blue lines connecting SNPs show absolute SNP correlations larger than 0.7. Dots in the first track (grey ring) show the –log10(p) values for the SNPs above the threshold of 4.19. In the second track, blue and red bars at the QTL positions show additive effects in each environment estimated with a single-locus-multi-environment model (starting from inside; BW4TUR, BW3ATK, BW2JHI, BW1CRA, BS3ATK and BS2JHI). Bar size is proportional to the confidence interval around the QTL. In the third track, blue and red bars show the additive main effects of each of the haplotype alleles from the candidate genes. Labels indicate the name of the candidate gene. Scale for the additive effects shown in the upper-left corner, with effects ranging from -2.17 to +1.93 cm. Mean awn length was 10.45 cm.

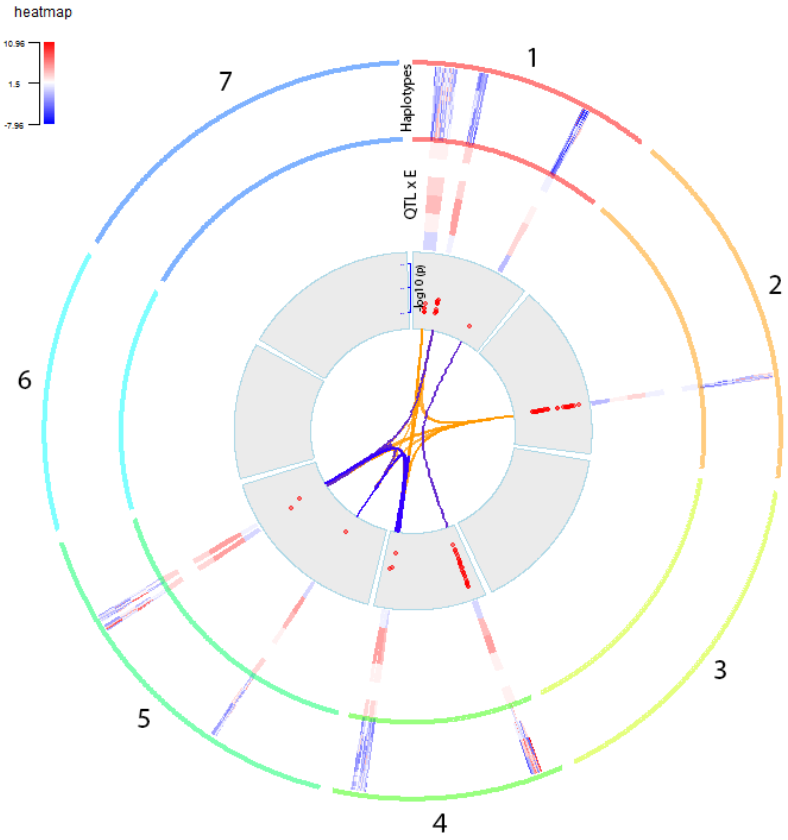


Figure 6. Circos plot for **grain weight,** with chromosome 1 shown in red and chromosome 7 shown in blue (clockwise). In the plot centre, orange lines connecting SNPs show absolute SNP correlations between 0.4 and 0.7 and dark blue lines connecting SNPs show absolute SNP correlations larger than 0.7. Dots in the first track (grey ring) show the –log10(p) values for the SNPs above the threshold of 4.19. In the second track, blue and red bars at the QTL positions show additive effects in each environment estimated with a single-locus-multi-environment model (starting from inside; BW4TUR, BW3ATK, BW2JHI, BW1CRA, BS3ATK and BS2JHI). Bar size is proportional to the confidence interval around the QTL. In the third track, blue and red bars show the additive main effects of each of the haplotype alleles from the candidate genes. Labels indicate the name of the candidate gene. Scale for the additive effects is shown in the upper-left corner, with effects ranging from -7.96 to +1.96 mg. Mean grain weight was 49.79 mg.

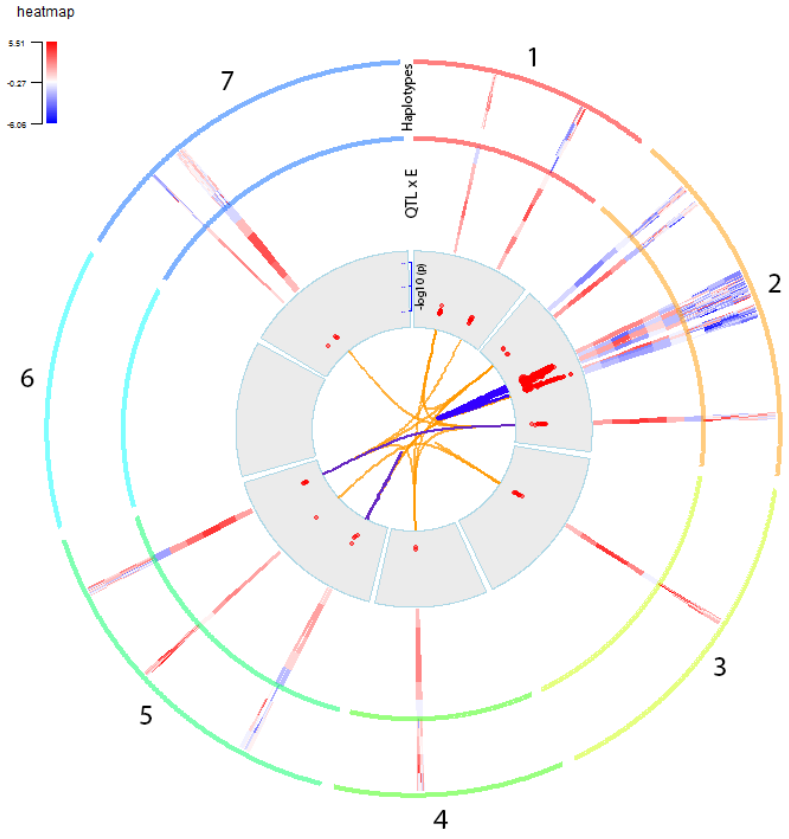


Figure 7. Circos plot for **heading date** with chromosome 1 shown in red and chromosome 7 shown in blue (clockwise). In the plot centre, orange lines connecting SNPs show absolute SNP correlations between 0.4 and 0.7 and dark blue lines connecting SNPs show absolute SNP correlations larger than 0.7. Dots in the first track (grey ring) show the –log10(p) values for the SNPs above the threshold of 4.19. In the second track, blue and red bars at the QTL positions show additive effects in each environment estimated with a single-locus-multi-environment model (starting from inside; BW4TUR, BW3ATK, BW2JHI, BW1CRA, BS3ATK and BS2JHI). Bar size is proportional to the confidence interval around the QTL. In the third track, blue and red bars show the additive main effects of the first five haplotype alleles from the candidate genes. Labels indicate the name of the candidate gene. Scale for the additive effects is shown in the upper-left corner, with effects ranging from -6.06 to +5.51 days. Mean heading date was 155.37 days.

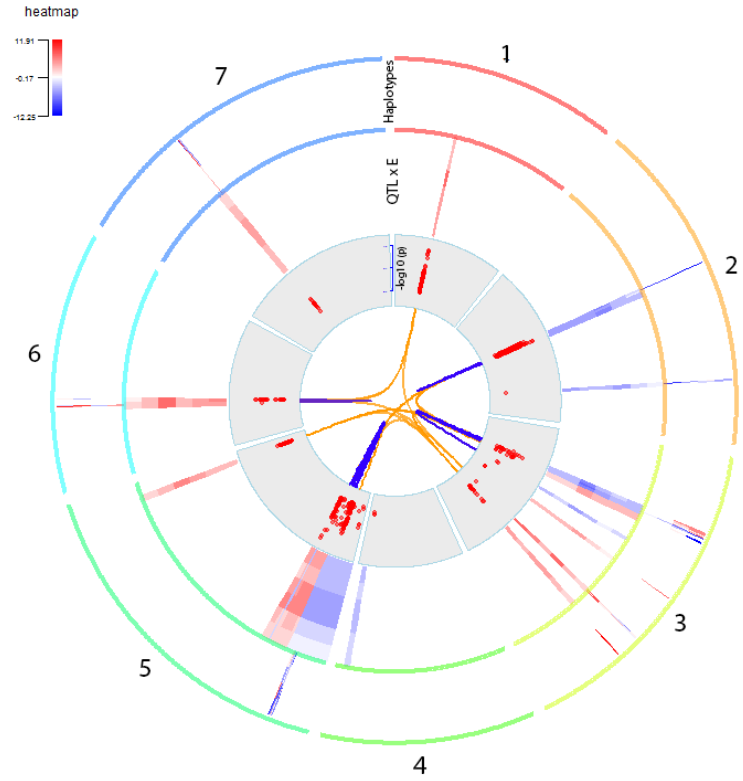


Figure 9. Circos plot for **plant height**, with chromosome 1 shown in red and chromosome 7 shown in blue (clockwise). In the plot centre, orange lines connecting SNPs show absolute SNP correlations between 0.4 and 0.7 and dark blue lines connecting SNPs show absolute SNP correlations larger than 0.7. Dots in the first track (grey ring) show the –log10(p) values for the SNPs above the threshold of 4.19. In the second track, blue and red bars at the QTL positions show additive effects in each environment estimated with a single-locus-multi-environment model (starting from inside; BW4TUR, BW3ATK, BW2JHI, BW1CRA, BS3ATK and BS2JHI). Bar size is proportional to the confidence interval around the QTL. In the third track, blue and red bars show the additive main effects of each of the haplotype alleles from the candidate genes. Labels indicate the name of the candidate gene. Scale for the additive effects is shown in the upper-left corner, with effects ranging from -12.25 to +11.91 cm. Mean plant height was 79.46 cm.

All the alleles of the haplotypes included in the multi-locus HTLs model for awn length showed a geographical gradient from east to west, where the most distinctive genotypes originate from Japan, Korea and China (Figure S10). Part of the genotypes from Ethiopia, Yemen and Fertile Crescent also showed a different type of alleles, especially for HORVU4Hr1G085590 in chromosome 4H (Subtilisin-like protease). A gene of the subtilisin-protease family has been previously involved in awn development in rice (*OsSLP*

Bessho-Uehara *et al.*, 2016). In Table 2, we show the additive effects across environments of two example genes that were included in the multi-locus HTLs model; HORVU3Hr1G093310 (the transcription factor bHLH62) and HORVU6Hr1G019700 (Squamosa promoter-binding-like protein 3). The transcription factor bHLH62 has three haplotypes. However, only haplotype b showed a different additive effect, compared to the reference haplotype a and the other minor haplotype c in awn length.. The high-value haplotype b was most commonly present in subpopulation 4, composed by Asian genotypes (Table 2, Figure S10). Similarity searches indicated high similarity to rice gene Os01g0915600, identified as OsbHLH90 transcription factor (Li *et al.*, 2006). This gene is phylogenetically close to OsbHLH93, also known as *An1*, that was described to control awn length in rice (Luo *et al.*, 2013). HORVU6Hr1G019700 also contained 3 alleles. In this case, the effect of allele c was 2.1 cm shorter than allele a (20% of the population mean). Alleles a and b had similar effects. HORVU6Hr1G019700 encodes a protein sharing high similarity with the one encoded by rice gene *OsSPL12*. Transcription factors of the SQUAMOSA promoter binding protein-like (SPL) family play roles in flower development and other plant developmental processes (Chen *et al.*, 2010). Similar to the situation of the bHLH transcription factor, the trait-increasing allele for Squamosa promoter-binding-like protein 3 was mostly present in subpopulation 4, reinforcing the idea that Asian genotypes constitute a source of different alleles for awn length.

The most important gene for grain weight was HORVU4Hr1G007040.1 in chromosome 4H (Transcription factor HvINT-C). This gene did not follow a clear geographical pattern (Figure S10), but it matched closely to row type (row type could be predicted correctly from HORVU4Hr1G007040.1 for 97% of the genotypes, coinciding with (Ramsay *et al.*, 2011; Muñoz-Amatriaín *et al.*, 2014)). Three haplotypes at this HTL were identified, with haplotypes b and c mainly 6 rows and haplotype a for 2 rows (Table 2). The fact that HTL alleles b and c have similar additive effects also suggests that HTL alleles b and c are functionally equivalent.

The multi-locus HTLs model for heading date included nine HTLs, mainly known flowering time genes, such as the photoperiod-response genes HvPPD-H1 (Pseudo-response regulator 7) and HvPPD-H2 (Protein FLOWERING LOCUS T) and the earliness per se locus *HvCEN* (also called Eam6/Protein TERMINAL FLOWER 1, (Comadran *et al.*, 2012)). The HvPPD-H2 locus was not significant in the single-locus HTL model (Table S4). However, it was significant ($p<0.001$) after including the other flowering time genes. HvPPD-H1 haplotypes showed some geographical distribution patterns, with haplotype a present mostly in European genotypes, haplotype f present mostly in genotypes from the Middle-East and haplotype g present most commonly in Japan, Korea, Ethiopia and the Fertile Crescent (Table

2). In Figure 9, we show the geographical distribution for a limited number of example genes. All the genes retained in the multi-HTL model were used to construct the adaptation classes. Haplotypes c, e and g had similar effects and reached heading earlier than haplotype a (Table 2). However, this difference was greater in spring than in winter trials, coinciding with the GxE driven by the contrast between spring and winter trials and the more inductive photoperiod in spring trials (Table 1, Figure 2, coinciding with (Russell *et al.*, 2016)). The haplotype for the *HvCEN* locus also showed some geographical patterns, with haplotype a dominating in higher latitudes, haplotype c mostly in central Asian genotypes and haplotype d mostly in African genotypes. Haplotype b was distributed across the whole range of sites of origin, but with a lower frequency in higher latitudes (Figure 9). Haplotype b and c were similar in their effects and were in general earlier than haplotype a and d, but this difference was larger in winter trials than in spring trials (Table 2). HvPPD-H2 had several alleles, that were clustered into three major haplotypes (Table S4). These alleles did not show a clear geographical distribution, except for the fact that allele c was not present in Ethiopian genotypes. HvPPD-H2 alleles had a smaller effect than HvPPD-H1; allele b was in general later than the reference and allele c was in general earlier than the reference, except in BW2JHI, where both alleles were earlier than the reference (allele b=-0.8 and allele c=-2.02 days). The Vrn-H3 gene on chromosome 7H was not covered by the sequence data and for that reason, it could not be tested. However, GWAS shows a QTL at about 2 Mb from this gene. In the same region of the QTL at chromosome 7H, we tested the MADS-box transcription factor 25, which was included in the multi-locus HTLs model. In this case, the significance of the MADS-box transcription factor 25 could be due to a direct effect or because of its LD with the Vrn-H3 gene.

In the case of plant height, 11 loci were included in the multi-locus HTLs model (Table S5). All of them showed some geographical patterns, with a trend from East to West. The sdw1/denso (HORVU3Hr1G090980) and ABSCISIC ACID-INSENSITIVE 5-like protein 2 are shown as examples in Figure 10. sdw1/denso is known to be responsible of the dwarf phenotype in barley (Jia *et al.*, 2009). Gibberellin 20-oxidase 3 showed 3 major haplotypes, but 2 out of them had similar effects, associated with taller plants than the reference haplotype a. The additive effects of haplotypes b and c were around 10cm, similar across environments, except for BS2JHI, where they had a smaller effect (Table 2). The gene ABSCISIC ACID-INSENSITIVE 5-like protein 2 had 6 haplotypes, from which haplotype b and c were similar and had a negative effect across environments (Table 2). Haplotypes e and f were also similar to each other and resulted in taller plants than the reference haplotype a, except in BS2JHI, where they had a similar effect than the reference haplotype.

Table 2. Examples of gene haplotypes, with the proportion of each subpopulation carrying a particular haplotype, additive effects across environments and mean standard error.

Trait	chr	gene	Annotation	pvalue	Allele	Haplotypes	Subpopulations prop. 1	2	3	4	5	6	Freq	BS2 JHI	BS3 ATK	BW1 CRA	BW2 JHI	BW3 ATK	BW4 TUR	mean se
Awn length	3H	HORVU3Hr1G093310 Start: 641851932 End: 641855001	Transcription factor bHLH62	1E-09	a		0.9	1.0	1.0	0.4	1.0	1.0	**332**	0.00		0.00	0.00		0.00	0.53
					b		0.0	0.0	0.0	0.5	0.0	0.0	**27**	1.91		3.00	2.10		0.82	0.53
					c		0.1	0.0	0.0	0.1	0.0	0.0	**12**	-0.09		0.08	0.28		0.72	0.53
	6H	HORVU6Hr1G019700 Start: 53909817 End: 53916886	Squamosa promoter-binding-like protein 3	2E-07	a		1.0	1.0	0.7	0.2	0.9	0.9	**314**	0.00	0.00	0.00	0.00			0.57
					b		0.0	0.0	0.3	0.4	0.1	0.1	**33**	-0.30	-0.49	-0.23	-0.36			0.57
					c		0.0	0.0	0.0	0.5	0.0	0.0	**24**	-2.53	-1.78	-3.58	-0.59			0.57
Grain Weight	4H	HORVU4Hr1G007040.1 Start: 17599034 End: 17600737	Transcription factor TEOSINTE BRANCHED 1	4E-17	a		0.8	0.0	0.8	1.0	0.2	1.0	**193**	0.00		0.00	0.00	0.00	0.00	2.20
					b		0.2	0.9	0.1	0.0	0.8	0.0	**168**	5.49		7.86	12.98	10.06	4.74	2.20
					c		0.0	0.1	0.0	0.0	0.0	0.0	**10**	6.06		8.48	10.38	7.73	3.12	2.20
	2H	HORVU2Hr1G103580.1 Start: 700387353 End: 700398095	SIT4 phosphatase -associated family protein isoform 5		a		0.0	0.8	0.0	0.1	0.2	0.0	**115**	0.00		0.00	0.00	0.00	0.00	2.01
					b		0.5	0.1	0.0	0.3	0.2	0.3	**95**	-2.11		-0.76	0.94	0.01	-0.78	2.01
					c		0.4	0.0	0.0	0.0	0.2	0.2	**57**	2.99		4.76	9.30	7.07	1.18	2.01
					d		0.1	0.0	0.0	0.3	0.2	0.3	**51**	-1.43		-0.78	1.24	1.96	0.17	2.01
					e		0.0	0.0	1.0	0.0	0.1	0.1	**32**	-1.98		-3.44	-4.94	-4.56	-2.80	2.01
					f		0.0	0.0	0.0	0.3	0.2	0.0	**21**	-6.88		-3.48	-2.87	-3.52	-4.03	2.01
Heading	2H	HORVU2Hr1G013400 Start: 29123785 End: 29127889	Pseudo-response regulator 7 (HvPPD-H1)	9E-18	a		0.1	0.8	0.4	0.0	0.0	0.0	**111**	0.00	0.00	0.00	0.00	0.00	0.00	1.53
					b		0.2	0.0	0.0	0.2	0.2	0.9	**73**	-2.63	-3.61	-2.09	0.25	1.22	-0.27	1.53
					c		0.3	0.2	0.1	0.2	0.1	0.0	**61**	-3.89	-4.73	-3.39	-1.19	-1.87	-1.70	1.53
					d		0.3	0.0	0.0	0.0	0.0	0.0	**39**	-5.41	-5.35	-1.35	-2.00	0.63	-1.17	1.53
					e		0.0	0.0	0.5	0.4	0.0	0.0	**30**	-5.94	-6.65	-7.03	-3.66	-1.80	-3.01	1.53
					f		0.1	0.0	0.0	0.0	0.3	0.1	**30**	-7.06	-4.29	-0.95	-1.18	2.61	-0.30	1.53
					g		0.0	0.0	0.0	0.2	0.4	0.0	**27**	-6.93	-8.92	-8.56	-7.04	-2.23	-3.25	1.53
	2H	HORVU2Hr1G072750 Start: 523377523 End: 523379139	Protein TERMINAL FLOWER 1 (HvCEN)	3E-13	a		0.5	0.9	0.0	0.1	0.6	0.1	**197**	0.00	0.00	0.00	0.00	0.00	0.00	1.34
					b		0.4	0.1	0.4	0.4	0.1	0.9	**117**	-0.86	-2.69	-4.12	-7.08	-4.01	-2.60	1.34
					c		0.1	0.0	0.0	0.4	0.0	0.0	**34**	-0.35	-1.94	-4.70	-6.47	-3.44	-2.28	1.34
					d		0.0	0.0	0.5	0.0	0.2	0.0	**23**	0.55	-0.02	-1.67	-3.94	-2.15	-1.39	1.34
Height	3H	HORVU3Hr1G090980 Start: 634078038 End: 634081600	gibberellin 20-oxidase 3	9E-16	a		0.6	0.8	0.0	0.0	0.7	1.0	**233**	0.00	0.00	0.00	0.00	0.00	0.00	2.75
					b		0.4	0.1	0.1	0.8	0.2	0.0	**104**	7.74	10.55	11.41	12.56	10.68	11.19	2.75
					c		0.0	0.0	0.9	0.2	0.1	0.0	**34**	3.77	14.37	13.62	15.70	14.31	10.32	2.75
	3H	HORVU3Hr1G084360 Start: 607244129 End: 607247713	ABSCISIC ACID-INSENSITIVE 5-like protein 2	7E-15	a		0.0	0.0	0.3	0.1	0.1	0.1	**153**	0.00	0.00	0.00	0.00	0.00	0.00	3.68
					b		0.3	0.9	0.0	0.1	0.1	0.0	**64**	-5.27	-6.27	-5.99	-9.64	-1.42	-3.70	3.68
					c		0.3	0.0	0.8	0.0	0.4	0.0	**59**	-4.86	1.02	-4.34	-5.85	0.98	-1.22	3.68
					d		0.1	0.0	0.2	0.7	0.2	0.0	**53**	-6.29	0.68	1.01	-0.49	6.75	6.48	3.68
					e		0.1	0.0	0.0	0.0	0.1	0.9	**22**	2.87	4.41	12.70	0.20	13.99	13.96	3.68
					f		0.2	0.0	0.0	0.0	0.0	0.0	**20**	-0.29	10.92	11.62	5.53	9.74	11.93	3.68
						Subpopulation size	**92**	**124**	**42**	**48**	**43**	**22**	**371**							

Table 3. Awn length and **grain weight** Wald statistic and p-value for the fixed model terms, variance components and standard error for the random model terms in seven models differing in the partition of the genotype main effect and genotype by environment interaction. Model terms are indicated as follows; E=Environment, PC=Kinship principal components, R=Row number, S=Subpopulations, AC=adaptation classes, SNP=SNP at the QTL position.

Awn Length																			
		Model 3.1		**Model 3.2**		**Model 3.3**		**Model 3.4**		**Model 3.5**		**Model 3.6**		**Model 3.7**		**Model 3.8**		**Model 3.9**	
		E		S+S.E		R+R.E		PC		PC+PC.E+S+S.E		PC+PC.E+R+R.E		PC+PC.E+AC+EC.E		PC+PC.E+HTL+HTL.E		PC+PC.E+SNP+SNP.E	
Type	Term	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e
Fixed	Group			151.51	<0.001	6.40	0.10			18.70	0.00	13.56	0.00	76.02	<0.001				
(Wald)	Group.E			150.65	<0.001	15.99	0.07			47.08	<0.001	11.37	0.25	78.89	<0.001	130.80	<0.001	178.89	<0.001
Random	Geno	2.14	0.20	1.46	0.14	2.12	0.20	1.36	0.13	1.32	0.13	1.31	0.13	1.09	0.11	1.11	0.12	1.03	0.11
(Comp)	Geno.E	1.74	0.08	1.54	0.07	1.73	0.08	1.50	0.07	1.45	0.07	1.49	0.07	1.40	0.06	1.40	0.06	1.38	0.06
aic		5400		5146		5368		5025		4941		4987		4880		4872		4824	
N groups / loci				6		2				6		2		4		4		4	

Grain Weight																			
		Model 3.1		**Model 3.2**		**Model 3.3**		**Model 3.4**		**Model 3.5**		**Model 3.6**		**Model 3.7**		**Model 3.8**		**Model 3.9**	
		E		S+S.E		R+R.E		PC		PC+PC.E+S+S.E		PC+PC.E+R+R.E		PC+PC.E+AC+EC.E		PC+PC.E+HTL+HTL.E		PC+PC.E+SNP+SNP.E	
Type	Term	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e
Fixed	Group			99.82	<0.001	103.85	<0.001			18.79	0.00	147.82	<0.001						
(Wald)	Group.E			69.65	<0.001	51.53	<0.001			45.74	0.00	31.47	0.00			595.02	<0.001	291.89	<0.001
Random	Geno	34.98	2.94	26.80	2.35	26.65	2.33	24.18	2.15	23.07	2.08	16.06	1.55	22.71	2.02	13.85	1.4	15.91	1.57
(Comp)	Geno.E	22.64	0.86	21.90	0.84	21.99	0.84	21.77	0.84	21.40	0.83	21.46	0.83	19.77	0.76	18.78	0.76	20.20	0.80
aic		11253		11050		11060		10920		10765		10708		10687		10013		10231	
N groups / loci				6		2				6		2		6		10		8	

Table 4. Heading date and **plant height** Wald statistic and p-value for the fixed model terms, variance components and standard error for the random model terms in seven models differing in the partition of the genotype main effect and genotype by environment interaction. Model terms are indicated as follows; E=Environment, PC=Kinship principal components, R=Row number, S=Subpopulations, AC=adaptation classes, SNP=SNP at the QTL position.

Heading date																			
		Model 3.1		**Model 3.2**		**Model 3.3**		**Model 3.4**		**Model 3.5**		**Model 3.6**		**Model 3.7**		**Model 3.8**		**Model 3.9**	
		E		S+S.E		R+R.E		PC		PC+PC.E+S+S.E		PC+PC.E+R+R.E		PC+PC.E+AC+EC.E		PC+PC.E+HTL+HTL.E		PC+PC.E+SNP+SNP.E	
Type	Term	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e
Fixed	Group			166.98	<0.001	24.21	<0.001			6.71	0.25	2.98	0.40	102.20	<0.001				
(Wald)	Group.E			311.05	<0.001	114.04	<0.001			43.71	0.01	48.93	<0.001	355.45	<0.001	756.45	<0.001	497.95	<0.001
Random	Geno	25.93	2.10	17.59	1.47	24.52	1.99	15.95	1.34	15.82	1.34	15.92	1.34	12.49	1.07	10.96	1.00	10.61	0.94
(Comp)	Geno.E	15.35	0.51	13.26	0.44	14.56	0.49	12.68	0.42	12.56	0.42	12.46	0.42	10.70	0.36	10.43	0.37	10.92	0.37
aic		13094		12627		12908		12406		12264		12298		11939		11348		11847	
N groups / loci				6		2				6		2		4		9		9	

Plant height																			
		Model 3.1		**Model 3.2**		**Model 3.3**		**Model 3.4**		**Model 3.5**		**Model 3.6**		**Model 3.7**		**Model 3.8**		**Model 3.9**	
		E		S+S.E		R+R.E		PC		PC+PC.E+S+S.E		PC+PC.E+R+R.E		PC+PC.E+AC+EC.E		PC+PC.E+HTL+HTL.E		PC+PC.E+SNP+SNP.E	
Type	Term	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e	Wald/ σ^2	pval/ s.e
Fixed	Group			67.32	<0.001	22.28	<0.001			19.00	0.00	9.29	0.03	110.80	<0.001				
(Wald)	Group.E			452.94	<0.001	166.21	<0.001			33.92	0.11	18.57	0.24	247.81	<0.001	920.10	<0.001	555.94	<0.001
Random	Geno	111.47	8.91	95.9	7.68	106.35	8.51	92.84	7.43	89.26	7.21	91.16	7.33	70.99	5.79	43.52	3.89	45.82	4.01
(Comp)	Geno.E	56.19	1.87	45.42	1.52	51.84	1.73	43.61	1.46	43.39	1.46	43.53	1.46	38.78	1.31	36.73	1.30	39.46	1.36
aic		15927		15379		15671		15187		15004		15079		14759		13769		14315	
N groups / loci				6		2				6		2		6		11		11	

6.3.6. Adaptation classes based on the multi-HTL model

$\hat{y}_{ij}^{HTL}$ was used to classify genotypes based on the additive effects of the haplotypes across environments. Such a classification is based on the alleles of genes that were shown to have an important phenotypic effect and for that reason, it is expected to give more insight about the GxE patterns in the field trials, than the subpopulations constructed with genome-wide similarity.

The additive effects of the four genes selected for awn length, suggest the presence of four adaptation classes. AC1 was most common in European genotypes, AC2 dominated in Indian and Nepalese genotypes, AC3 was typical for genotypes from Japan, Korea and China, with a few from India. AC4 dominated genotypes from the Fertile Crescent and Central Asia, with also an important representation in genotypes from the UK (Figure 10). The four adaptation classes were more associated to GxE than the six subpopulations, reflected in a better fit and a smaller residual genotype main effect and GxE of the models with adaptation classes and with explicit haplotypes (Tables 3, models 3.4, 3.8 and 3.9). AC2 and AC3, most commonly present in Asian genotypes, showed a positive interaction with BW4TUR, whereas AC1 tended to show some interaction with BS2JHI and BW1CRA.

In the case of grain weight, the additive effects of the 10 HTLs in the multi-HTL model (6) suggest the presence of six adaptation classes. In general, all adaptation classes had both 2- and 6-rowed genotypes, except for AC2, AC5 and AC6 that predominantly had 6-rowed genotypes. These groups were constructed considering a number of loci underlying GxE. For that reason, the large-main effect locus (Transcription factor HvINT-C) does not have a one to one relationship with the adaptation classes and the residual main effect was smaller when including the row type, instead of the ACs. In contrast, the residual GxE was smaller when considering the adaptation classes instead of the row type (Table 4, models 3.7 and 3.6). The reduction of the residual GxE when including the adaptation classes was mainly driven by the positive interaction between AC1 (most commonly found in genotypes from Eastern Europe and Ethiopia) with BW3TUR and by the positive interaction between AC3 (common in the UK) and BW3ATK. A compromise between both models was achieved when explicitly including HTLs (Tables 3 and 4, model 3.8). In this model, the first HTL represents row type (genotype main effect). The additional haplotypes account for genes with a smaller effect contributing to GxE.

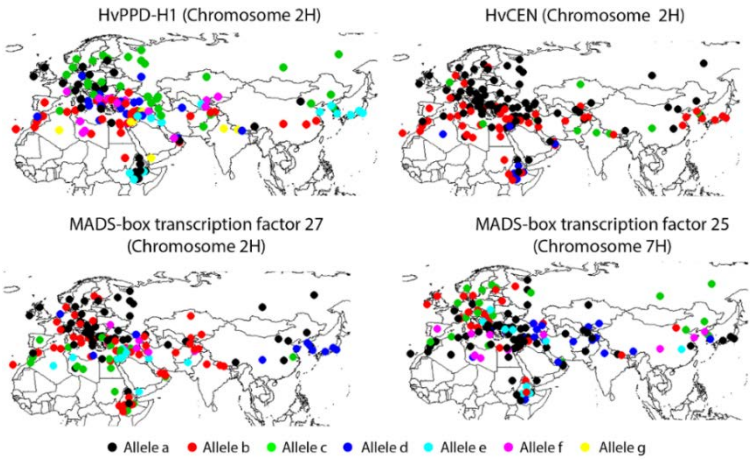


Figure 10. Haplotypes for four of the HTLs used to construct the adaptation classes for heading date.

The additive effects of the nine HTLs included in the multi-HTL model for heading date suggested the presence of four adaptation classes. AC2 was the most common class everywhere, whereas AC1 represented an important proportion of the genotypes in Europe, AC2 occurred often in Chinese genotypes, AC4 was common in central-Asian genotypes and AC5/AC6 occurred often in the fertile crescent and Middle-East (Figure 11). The adaptation classes were effective in capturing 31% of the genotype main effect and 17% of the GxE left after the kinship correction (Table 3, models 3.4 and 3.7). The reduction in GxE was driven by the contrast between AC1 and AC4/AC5 (Figure 2). AC1 was composed by an important proportion of winter formally-bred genotypes and showed a positive interaction with the spring trials. In contrast, AC4/AC5 had an important proportion of spring landraces and showed a positive interaction with the winter trials. The GxE pattern observed in the AMMI biplots in Figure 2 can also be observed when explicitly classifying the genotypes in spring-landraces, formally-bred winter and formally-bred spring genotypes (Figure 3). In that case, the interaction is observed because formally-bred winter genotypes flower earlier than the ideotype and spring landraces flower later than the ideotype.

The 11 loci included in the multi-HTL model for plant height suggested the presence of six adaptation classes. These adaptation classes were very effective in capturing part of the genotype main effect and GxE (the genotype main effect was reduced by 53% and the interaction by 16%, Table 4, models 3.4 and 3.7). AC1 was most common in European genotypes, whereas AC4 was most common in genotypes from the Fertile Crescent and Africa (Figure 10).

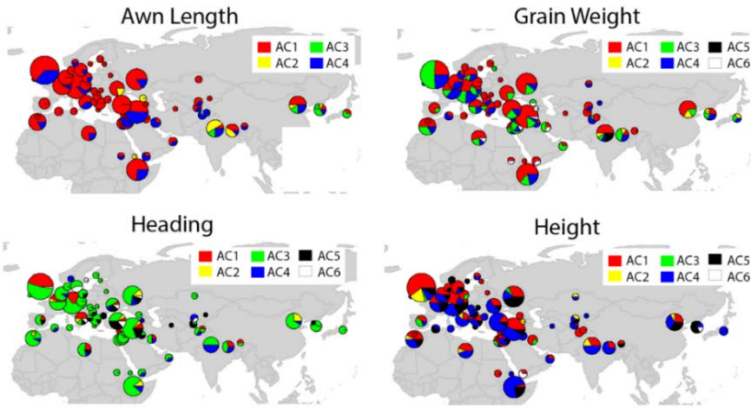


Figure 11. Geographical distribution of the adaptation classes (AC) constructed from the multi-HTL model.

6.4. Discussion

In this paper, we characterise the genetic diversity of 371 of the 511 genotypes belonging to the EU-Whealbi-barley collection, from the perspective of genotype adaptation to contrasting environments in Europe and Eurasia. Genetic diversity studies commonly consider allele frequency as a measure of fitness, assuming that the most frequent allele in a given location contributes most to the adaptation to local conditions (Orr, 2010; Günther & Coop, 2013). However, assessing the adaptation to the TPE is a different problem than the one of searching for local adaptation to the sites of origin because the allele frequency at the sites of origin is not very informative about the value of alleles for adaptation to the TPE. In this paper, we explicitly dissect GxE into its underlying genetic components, assuming that the environments we explored with the field trials are relevant for the European TPE. In this dissection, we used the additive effects of phenotype-related loci as an indication of adaptation. A similar approach has been previously used in genomic prediction to characterise the structure of genotypes and environments based on genome-wide additive effects (Heslot *et al.*, 2013).

As with most diversity collections, EU-Whealbi-barley is largely structured, posing a number of statistical methodological challenges to identify genomic regions regulating the target trait. Genotypes that are genetically similar because of population structure share both causal and non-causal alleles. Furthermore, the highly repetitive nature of the barley genome (Mascher *et al.*, 2017) leads to complex collinearities between different regions in the genome, like the one observed for the four traits we analysed (Figures 5-8). These collinearities and the population structure potentially lead to spurious marker-trait associations (Korte & Farlow, 2013; Vilhjálmsson & Nordborg, 2013). To reduce the

number of false positives, we used a mixed model approach including kinship principal components to account for population structure (Price *et al.*, 2006; Hoffman, 2013; Millet *et al.*, 2016). The kinship represents the genetic similarities between genotypes and reflects population sub-structuring, when it is present, averaged genome-wide (Astle & Balding, 2009). Unfortunately, any relationship matrix used to correct for population structure is only a proxy to the real underlying genetic background (Vilhjálmsson & Nordborg, 2013). It is known that loci with an important effect on genotype performance are under stronger genetic differentiation than the average SNP. The gene HvINT-C is a clear example of this, (Russell *et al.*, 2016), showing that genotypic differentiation is not homogeneous along the genome and that this heterogeneity is not reflected in the kinship, where all SNPs are averaged. An alternative way of accounting for the background structure is by selecting a small number of SNPs that are associated with the phenotype (causal SNPs or SNPs that are nearby causal SNPs, (Listgarten et al., 2011) . Our approach was similar to the method proposed by (Listgarten et al., 2011) because we built the final multi-HTL model in a forward procedure in which we scan in the QTL region, conditioning on HTL that resulted significant in the previous round of testing. Another strategy to deal with the problems of collinearities is by simultaneous estimation of genetic variance explained by all SNPs (Bayes-R, Moser et al., 2015). Unfortunately, the currently Bayes-R method is only available for single-environment data. Thus, such an approach would consider an initial single-environment GWAS with Bayes-R to select candidate loci, followed by the estimation of the additive effects of those candidates across environments in a mixed model like model (3.8).

A second challenge to identify causal loci is how to deal with the complex genetic interactions commonly found in quantitative traits. Examples of these interactions are the pleiotropic effects of row type genes (e.g. HvINT-C) on grain weight, grain number and grain yield (Liller *et al.*, 2015). Another common interaction occurs between flowering time and yield components, as reported by (Maurer *et al.*, 2016; Mikolajczak *et al.*, 2016; Wang *et al.*, 2016). One alternative to model these interactions is to express explicitly the sensitivity of genes to environmental conditions via ecophysiological models, as shown for flowering time in barley by (Yin *et al.*, 2005), or via graphical models and structural relations models (Wang & van Eeuwijk, 2014; Wang *et al.*, 2015; Alimi, 2016). An alternative is the multi-locus HTL model we presented here, which is commonly used in human genetics, the so called SKAT method (sequence kernel association test (Wu *et al.*, 2011)). SKAT is a supervised test for the joint effects of multiple variants in a region on a phenotype. Regions can be defined by genes (as we did) or moving windows across the genome. The main difference between our method and SKAT is that we used linear mixed models, whereas SKAT models epistatic effects within the kernel machine regression framework, showing promise as a tool to deal with the complex gene interactions regulating barley traits.

In spite of all the challenges posed by a highly structured population, we were able to detect promising candidate genes some of which are known to have an important effect on the traits included in this analysis. Examples were the flowering time genes HvPPD-H1, HvPPD-H2 and HvCEN, which were associated with the contrasting genotypic response to spring and winter trials (Turner *et al.*, 2005; Faure *et al.*, 2007; Cuesta-Marcos *et al.*, 2008; Comadran *et al.*, 2012; Russell *et al.*, 2016). We also identified the MADS box genes as having an important effect on flowering time, showing promise for further investigation of the genetic basis of the vernalisation response (Trevaskis *et al.*, 2007). However, as they are in the regions neighbouring the Vrn-H3 gene, it is also possible that MADS box genes are significant because they are in LD with the causal loci. The large number of alleles for HvCEN identified in our study was in line with the 13 haplotypes reported by (Comadran *et al.*, 2012). We used simple clustering methods to reduce the number of alleles with low frequency. However, as some of the alleles we identified showed a similar effect on the phenotype, further research needs to be conducted to assess their functional equivalence (synonymous/ non-synonymous mutations). Other examples of genes were the AP2-like ethylene-responsive transcription factor for heading date (Xue & Loveridge, 2004) and sdw1/denso, reported to regulate plant height (Jia *et al.*, 2009).

The value of a combined genotypic and phenotypic analysis resides in providing the means to prioritise genomic regions that might be related to the phenotype and assess which groups of genotypes are likely to carry contrasting alleles. The identification of contrasting alleles and the adaptation classes also gives insight into which geographical regions are a promising source of alleles useful for European breeding programmes, as shown for the contrast between awn length alleles between Asian genotypes and the rest. In the awn length example, our strategy was able to point out the Asian genotypes as a source of novel alleles, in line with what was reported by (Yuo *et al.*, 2012) for the *Lks2* gene. Such contrasting genotypes can be used for crosses, to search for causal genes (Ellis *et al.*, 2000) and provide breeders with pre-breeding germplasm. In such crosses, the availability of high-density genotyping like exome sequence data becomes a clear advantage for allele mining research, with respect to SNP chips. One strategy to reduce genotyping costs and make allele mining more affordable is by characterising the progeny at a lower density and then impute the missing SNPs from the high-density exome sequence data (van Binsbergen *et al.*, 2014, 2016).

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°FP7- 613556, Whealbi.

Supplementary Material

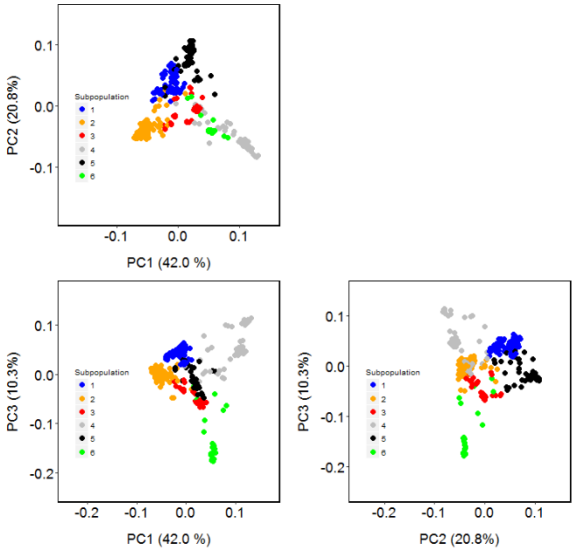


Figure S1. Scatter plots for principal components representing Kinship matrix of the 371 *H. vulgare* genotypes. Symbol colour represents each of the six subpopulations constructed with the cluster analysis.

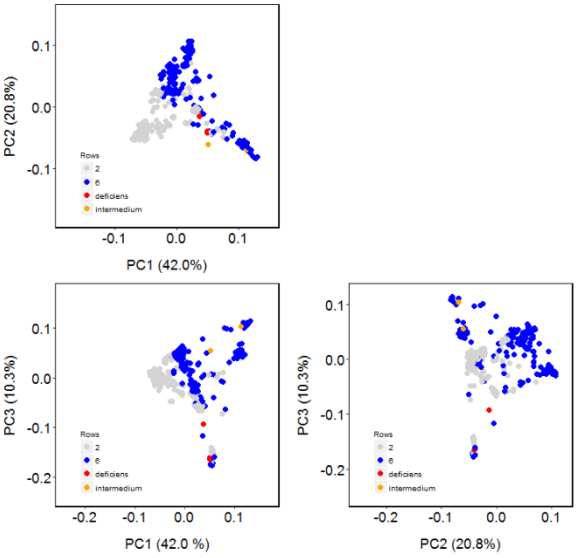


Figure S2. Scatter plots for principal components representing Kinship matrix of the 371 *H. vulgare* genotypes. Grey symbols represent 2-rowed barley, blue symbols represent 6-rowed barley, red symbols represent the deficient type and orange symbols represent the intermedium type.

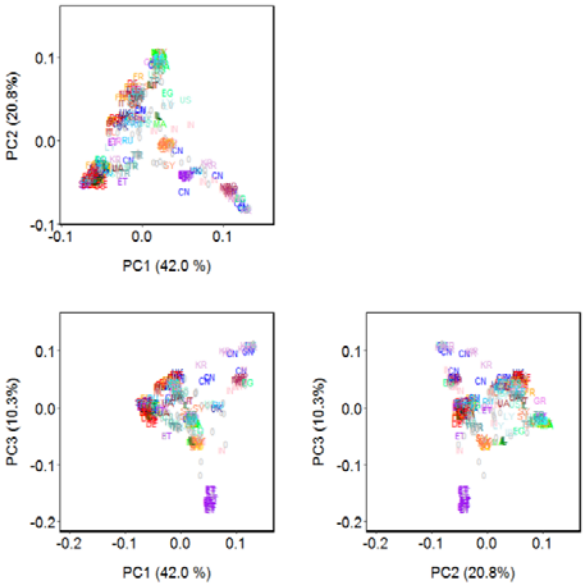


Figure S3. Scatter plots for principal components representing Kinship matrix of the 371 *H. vulgare* genotypes. Different symbols indicate the counties of origin.

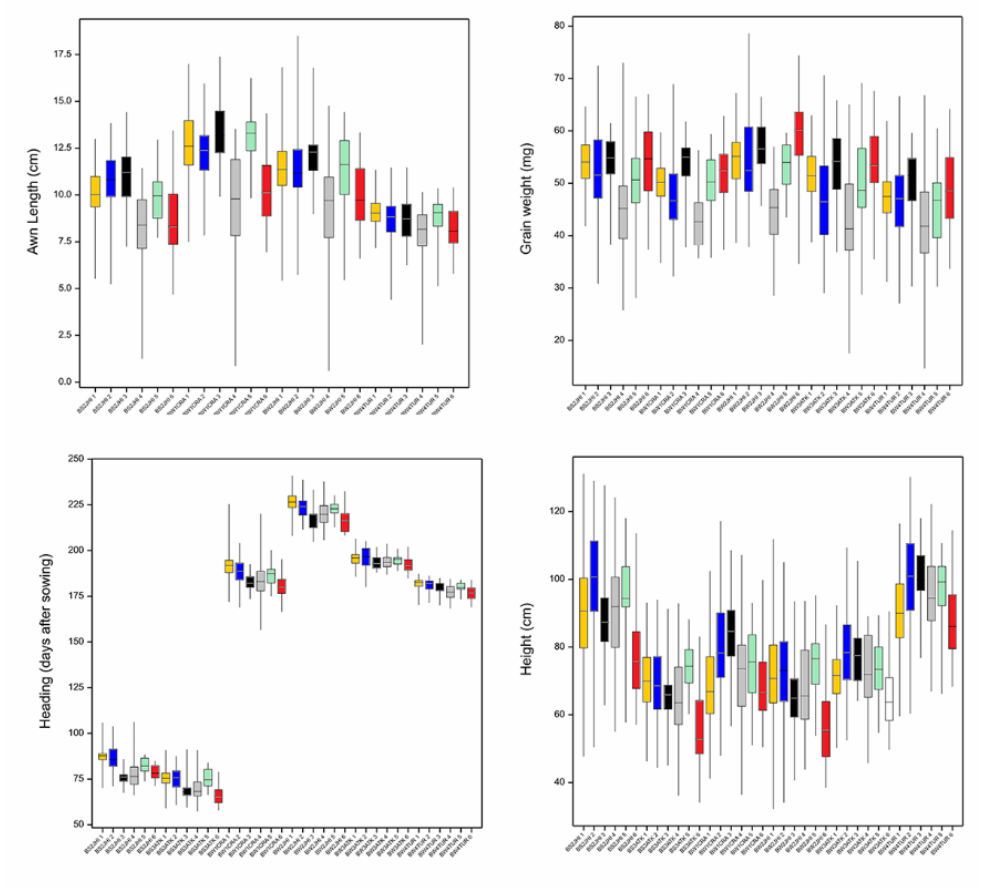


Figure S4. Boxplots for awn length, grain weight, heading date and plant height across environments. Symbol colours represent the six subpopulations (orange=subpopulation 1, blue=subpopulation 2, black= subpopulation 3, grey=subpopulation 4, green=subpopulation 5 and red=subpopulation 6). Environments are presented in the following order: BS2JHI, BS3ATK, BW1CRA, BW2JHI, BW3ATK and BW4TUR.

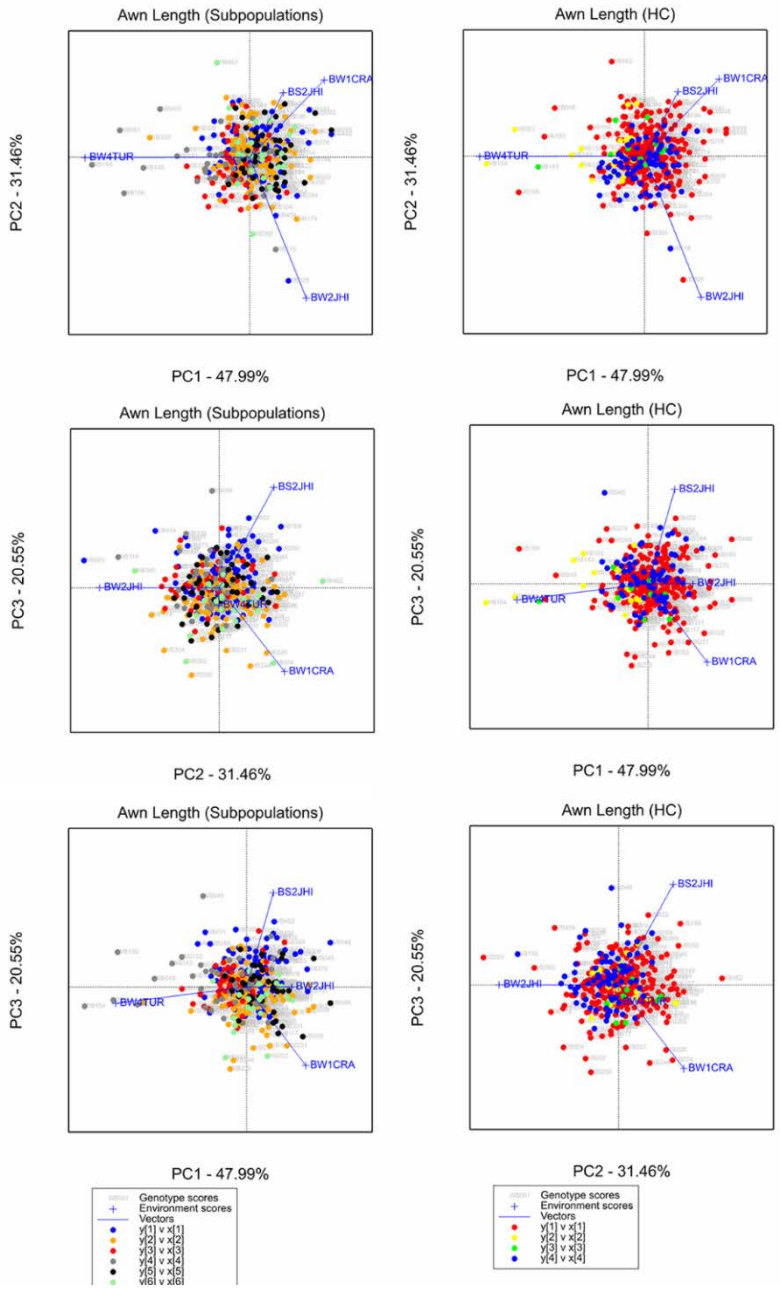


Figure S5. AMMI biplots for awn length. In the left panels, colours show the six subpopulations identified with the cluster analysis. In the right panels, colours show the adaptation classes constructed with the multi-HTL model (same colours as in Figure 9).

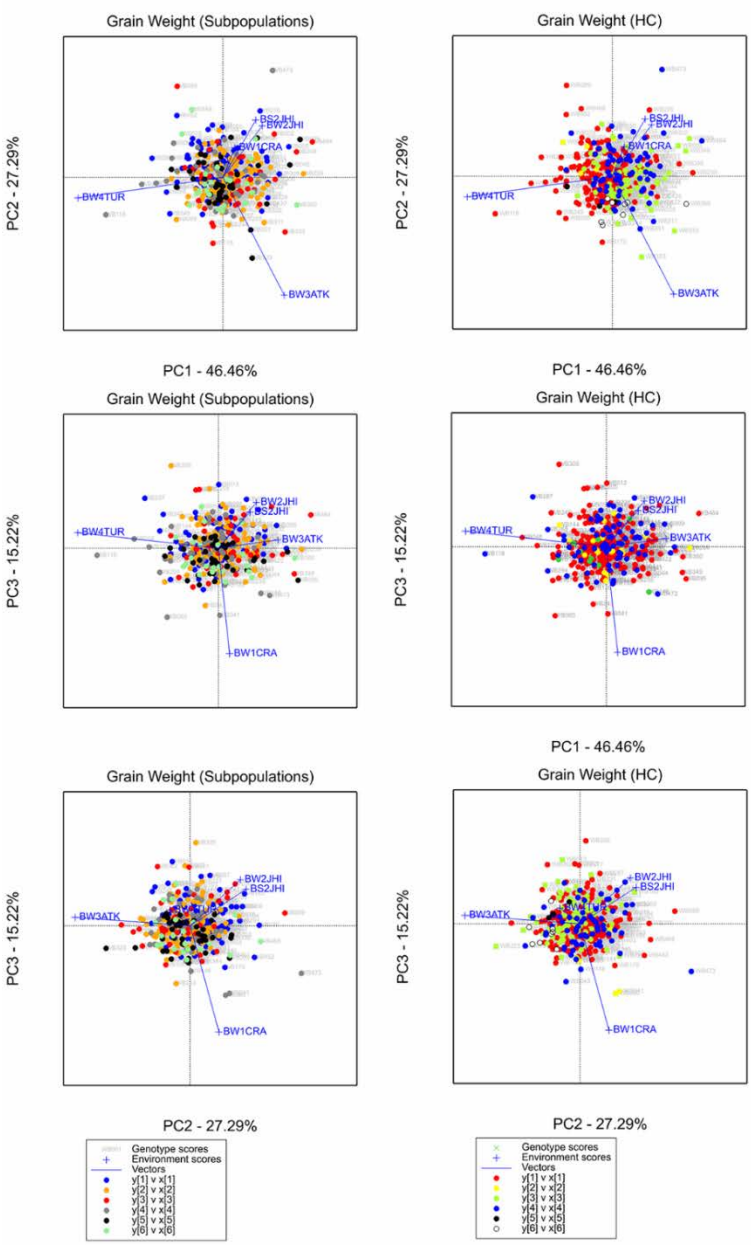


Figure S6. AMMI biplots for grain weight. In the left panels, colours show the six subpopulations identified with the cluster analysis. In the right panels, colours show the adaptation classes constructed with the multi-HTL model (same colours as in Figure 9).

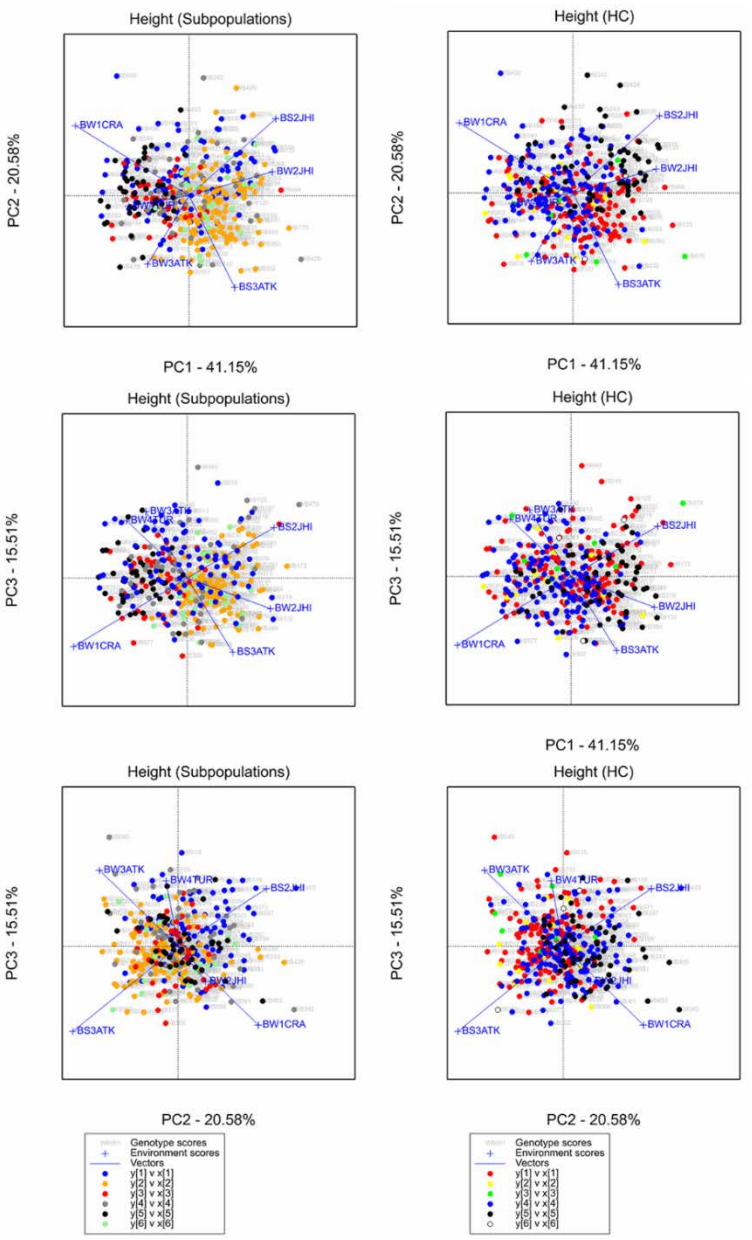


Figure S7. AMMI biplots for plant height. In the left panels, colours show the six subpopulations identified with the cluster analysis. In the right panels, colours show the adaptation classes constructed with the multi-HTL model (same colours as in Figure 9).

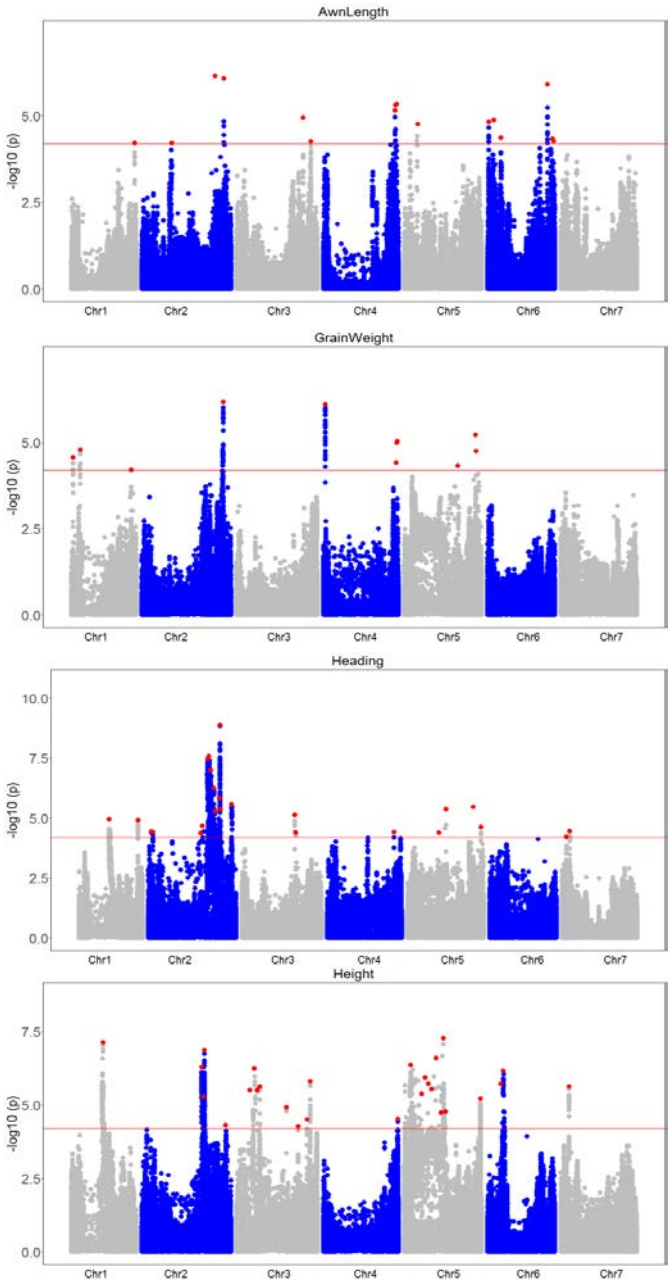


Figure S8. Manhattan plots for the multi-environment GWAS of Awn Length, Grain Weight, Heading date and Plant height. Red dots show the QTL positions.

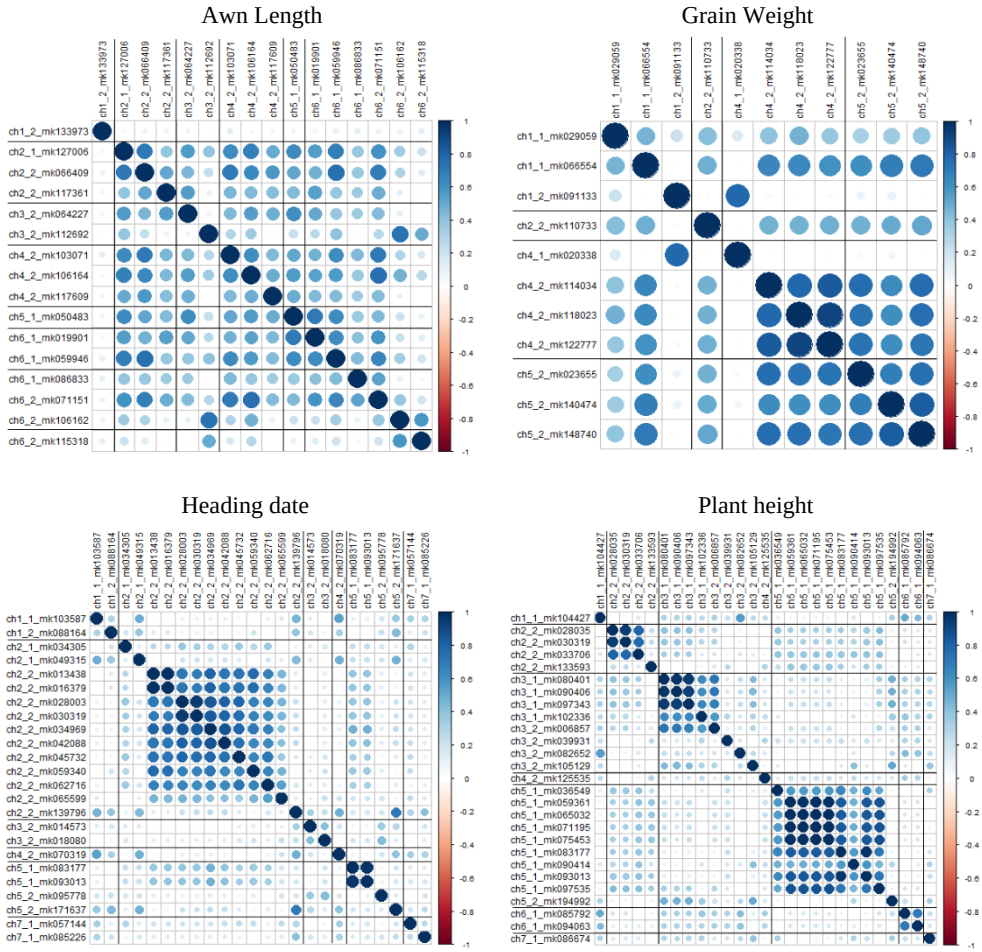


Figure S9. Absolute correlation between SNPs at the QTL positions.

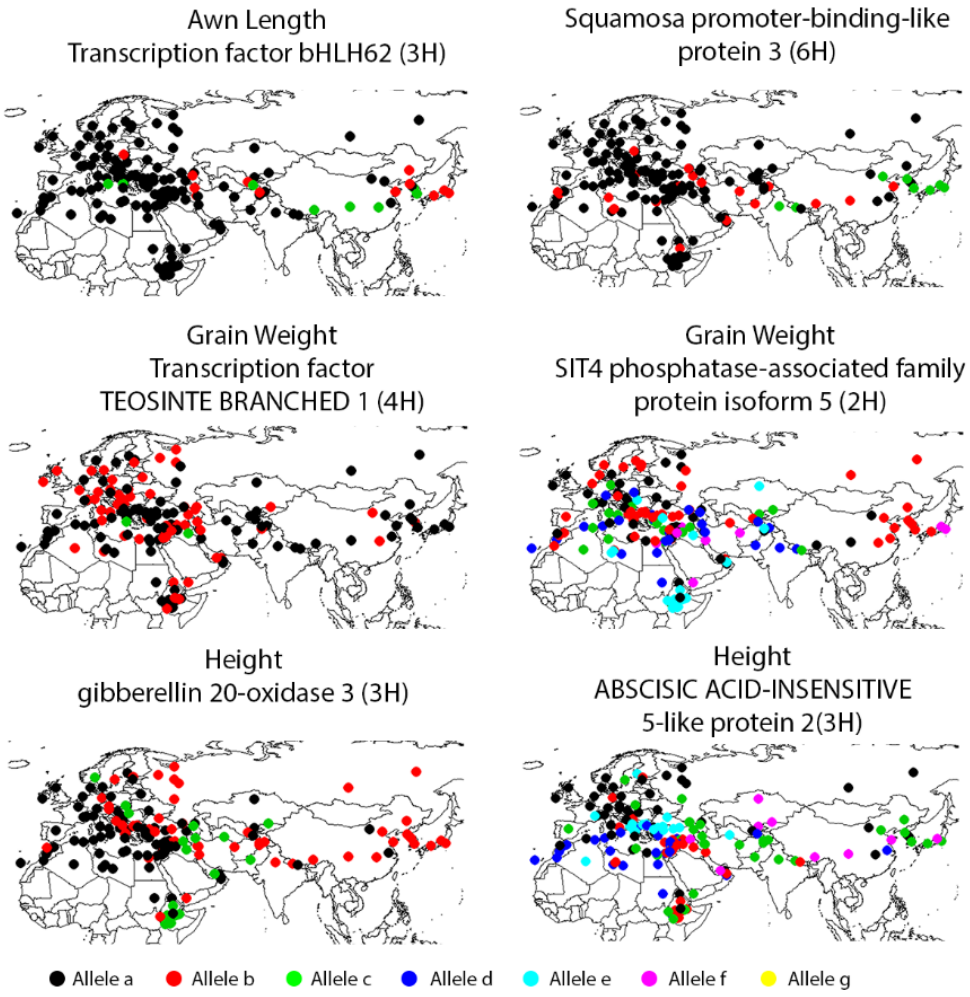


Figure S10. Geographical distribution for haplotypes of two genes per trait that were included in the multi-locus multi-environment model.

Table S1. Heading date of the ideotype and of the genotypes used to calculate the target heading date.

Country	Trial	Variety	Heading
Scotland spring	BS2JHI	**Ideotype**	**88**
		Aurelia (WB017)	86
		Irina (WB016)	89
		Orphelia (WB018)	89
Hungary spring	BS3ATK	**Ideotype**	**76**
		Vada (WB228)	78
		Varunda (WB230)	73
Italy	BW1CRA	**Ideotype**	**186**
		AMILLIS (WB037)	185
		ALDEBARAN (WB020)	186
		KETOS (WB025)	186
		PONENTE (W026)	186
		SIXTINE (WB191)	187
Scotland winter	BW2JHI	**Ideotype**	**222**
		Escadre (WB011)	216
		KW Capella (WB006)	217
		KW Glacier (WB004)	219
		Meridian (WB010)	220
		KW Tower (WB005)	223
		KW Cassia (WB003)	224
		Joy (WB008)	226
		Saffron (WB007)	226
		Wintmalt (WB009)	232
Hungary winter	BW3ATK	**Ideotype**	**197**
		Aldebaran (WB020)	192
		Tif[f]any (WB030)	196
		Saffron (WB007)	198
		Keeper (WB015)	199
		Joy (WB008)	201
Turkey winter	BW4TUR	**Ideotype**	**183**
		Tarm92 (WB031)	183

Table S2. Chromosome, gene position, haplotype block and p-value from the single-locus scan for the haplotypes tested for awn length. Haplotypes retained in the multi-locus multi-environment model and used to construct the adaptation classes are shown in bold.

Chr	gene	GeneStart	GeneEnd	Annotation	block	pvalue
3H	**HORVU3Hr1G077500**	**573268831**	**573269508**	**Protein BIG GRAIN 1**	**1**	**1.97E-11**
3H	**HORVU3Hr1G093310**	**641851932**	**641855001**	**Transcription factor bHLH62**	**2**	**1.37E-09**
3H	HORVU3Hr1G093310	641851932	641855001	Transcription factor bHLH62	1	2.21E-08
6H	HORVU6Hr1G019700	53909817	53916886	Squamosa promoter-binding-like protein 3	2	1.67E-07
2H	HORVU2Hr1G105960.1	708228357	708231094	Transcription factor bHLH96	1	2.38E-06
4H	**HORVU4Hr1G085590**	**632170302**	**632174004**	**Subtilisin-like protease**	**1**	**1.11E-02**
2H	HORVU2Hr1G088460	632848314	632849588	Auxin-responsive GH3 family protein	1	1.71E-02
2H	HORVU2Hr1G087490	628420689	628422821	TRICHOME BIREFRINGENCE-LIKE 11	1	1.89E-02
6H	**HORVU6Hr1G019700**	**53909817**	**53916886**	**Squamosa promoter-binding-like protein 3**	**1**	**2.28E-02**
6H	HORVU6Hr1G074970	517270501	517273811	Ethylene-responsive transcription factor 1	1	3.65E-02
2H	HORVU2Hr1G105980.2	708254388	708259197	RING-H2 finger protein ATL51	1	4.98E-02
2H	HORVU2Hr1G106030.2	708355239	708356004	Transcription factor bHLH96	1	1.18E-01
6H	HORVU6Hr1G074810	515985411	515985781	EPIDERMAL PATTERNING FACTOR-like protein 9	1	1.42E-01
1H	HORVU1Hr1G091110	547422197	547427207	Isoprenylcysteine alpha-carbonyl methylesterase ICME	1	1.45E-01
3H	HORVU3Hr1G077850	575015942	575019199	Calpain-D	1	2.06E-01
2H	HORVU2Hr1G088030	631619646	631620742	proliferating cell nuclear antigen 2	1	2.70E-01
2H	HORVU2Hr1G088030	631619646	631620742	proliferating cell nuclear antigen 2	2	5.12E-01
6H	HORVU6Hr1G075640	520976419	520981459	AP2-like ethylene-responsive transcription factor	1	5.64E-01
4H	HORVU4Hr1G082560	622608023	622610981	Two-component response regulator ORR22	1	6.99E-01
3H	HORVU3Hr1G093110	640891989	640894135	EPIDERMAL PATTERNING FACTOR-like protein 9	1	7.36E-01
1H	HORVU1Hr1G091110	547422197	547427207	Isoprenylcysteine alpha-carbonyl methylesterase ICME	2	8.91E-01
4H	HORVU4Hr1G082560	622608023	622610981	Two-component response regulator ORR22	2	9.16E-01
3H	HORVU3Hr1G078330	577713926	577718729	MORC family CW-type zinc finger protein 3	1	9.48E-01

Table S3. Chromosome, gene position, haplotype block and p-value from the single-locus scan for the 50 most significant haplotypes tested for grain weight. Haplotypes retained in the multi-locus multi-environment model and used to construct the adaptation classes are shown in bold.

Chr	gene	GeneStart	GeneEnd	Annotation	block	pvalue
4H	HORVU4Hr1G007020.2	17438157	17446158	Transmembrane Fragile-X-F-associated protein	1	2.08E-23
4H	HORVU4Hr1G007050.2	17916120	17919867	Phytanoyl-CoA dioxygenase domain-containing protein 1	1	1.07E-20
4H	**HORVU4Hr1G007040.1**	**17599034**	**17600737**	**HvINT-C**	**1**	**4.32E-17**
2H	HORVU2Hr1G103490.27	700173292	700181605	Disease resistance protein	3	1.61E-16
2H	**HORVU2Hr1G103580.13**	**700387353**	**700398095**	**SIT4 phosphatase-associated family protein isoform 5**	**1**	**1.64E-16**
1H	HORVU1Hr1G019500.3	73807416	73810946	sucrose-phosphatase 1	1	3.04E-15
4H	HORVU4Hr1G086350.1	633907967	633909731	HXXXD-type acyl-transferase family protein	1	7.92E-15
1H	HORVU1Hr1G019380.2	73145936	73147775	HXXXD-type acyl-transferase family protein	1	2.43E-14
2H	HORVU2Hr1G103490.27	700173292	700181605	Disease resistance protein	1	2.60E-14
5H	HORVU5Hr1G105410.8	621489628	621493731	undescribed protein	1	8.18E-14
5H	HORVU5Hr1G105250.2	621065239	621075464	methionine S-methyltransferase	3	8.18E-14
1H	HORVU1Hr1G077600.2	517975530	517979750	F-box family protein	1	8.87E-14
4H	HORVU4Hr1G088060.4	639213994	639219642	Oxysterol-binding protein-related protein 2A	1	1.31E-13
1H	HORVU1Hr1G077820.10	518614904	518621086	Poly [ADP-ribose] polymerase 3	2	6.25E-13
5H	HORVU5Hr1G103330.3	615398310	615401415	40S ribosomal protein S12	2	7.62E-13
2H	HORVU2Hr1G103330.2	699803417	699806529	Transcription factor GTE7	3	1.09E-12
2H	HORVU2Hr1G103340.2	699805285	699808226	unknown function	1	1.55E-12
1H	HORVU1Hr1G019320.3	72455521	72463987	ankyrin repeat-containing protein 2	1	1.99E-12
2H	HORVU2Hr1G103460.3	700117295	700124639	Disease resistance protein	1	2.48E-12
1H	HORVU1Hr1G007480.1	15304943	15305722	Bowman-Birk type trypsin inhibitor	1	5.18E-12
1H	**HORVU1Hr1G017830.1**	**61356166**	**61357141**	**Histone superfamily protein**	**1**	**1.16E-11**
1H	HORVU1Hr1G017770.2	61271314	61272801	Auxin-responsive protein IAA4	1	1.48E-11
1H	HORVU1Hr1G007600.14	15611399	15615821	E3 UFM1-protein ligase 1 homolog	1	1.65E-11
1H	HORVU1Hr1G019480.1	73571033	73572205	Queuine tRNA-ribosyltransferase	1	2.34E-11
1H	HORVU1Hr1G077900.1	518820575	518824592	Transcription initiation factor IIF, beta subunit	1	2.43E-11
5H	HORVU5Hr1G059030.1	460998110	461000300	Protein kinase superfamily protein	1	4.84E-11
1H	HORVU1Hr1G077710.2	518228534	518231920	FAR1-related sequence 5	1	4.85E-11
4H	HORVU4Hr1G087110.9	636203381	636206530	Fatty acyl-CoA reductase 1	3	5.16E-11
2H	HORVU2Hr1G104170.1	702218400	702219149	Nodulation-signaling pathway 1 protein	1	5.33E-11
5H	HORVU5Hr1G060140.8	470768621	470817144	Endoglucanase 11	3	9.22E-11
5H	HORVU5Hr1G059990.5	468849932	468859086	unknown protein	1	1.08E-10
2H	HORVU2Hr1G103330.2	699803417	699806529	Transcription factor GTE7	2	1.23E-10
2H	HORVU2Hr1G103210.2	699436039	699438796	Beta-lactamase domain-containing protein 2	1	1.25E-10
2H	HORVU2Hr1G103220.3	699442141	699447747	polyamine oxidase 2	1	1.62E-10
5H	HORVU5Hr1G103180.1	615129596	615133117	Ferredoxin--NADP reductase	2	1.92E-10
5H	HORVU5Hr1G060110.16	470603383	470607369	malonyl-CoA decarboxylase family protein	1	2.55E-10
1H	HORVU1Hr1G017900.3	61919204	61923605	Transcription factor PIF3	2	3.90E-10
4H	HORVU4Hr1G007090.1	18336047	18342837	Protein of unknown function (DUF1295)	1	4.28E-10
5H	HORVU5Hr1G103080.2	614933658	614938213	Arv1-like protein	1	4.45E-10
1H	HORVU1Hr1G077560.1	517789252	517792319	L-tyrosine decarboxylase	1	4.52E-10
5H	HORVU5Hr1G059930.6	468282593	468287773	PCI domain-containing protein 2	1	6.45E-10
5H	HORVU5Hr1G060090.2	470353664	470367603	Lysine-specific demethylase 3B	1	7.46E-10
2H	HORVU2Hr1G103310.2	699752648	699757948	Nucleoredoxin	1	9.92E-10
4H	HORVU4Hr1G084670.7	629524229	629528665	PLAC8 family protein	1	1.03E-09
4H	HORVU4Hr1G084680.5	629524254	629534446	Homeobox protein knotted-1-like 3	1	1.03E-09
1H	HORVU1Hr1G077680.2	518159277	518160365	Ethylene-responsive transcription factor 1B	1	1.92E-09
5H	HORVU5Hr1G059310.5	463951526	463954263	Elongated mesocotyl1	1	2.18E-09
5H	HORVU5Hr1G060140.8	470768621	470817144	Endoglucanase 11	1	3.98E-09
4H	HORVU4Hr1G007120.2	18550241	18555863	Protein of unknown function, DUF593	1	4.13E-09
1H	HORVU1Hr1G077770.9	518523566	518529444	Outer envelope protein 80, chloroplastic	1	5.07E-09

Table S4. Chromosome, gene position, haplotype block and p-value from the single-locus scan for the haplotypes tested for heading date. Haplotypes retained in the multi-locus multi-environment model and used to construct the adaptation classes are shown in bold.

chr	gene	GeneStart	GeneEnd	Annotation	block	pvalue
2H	**HORVU2Hr1G113880**	**730027508**	**730030208**	**AP2-like ethylene-responsive transcription factor**	**1**	**1.2E-25**
2H	**HORVU2Hr1G072750**	**523377523**	**523379139**	**Protein TERMINAL FLOWER 1**	**1**	**4.5E-25**
2H	**HORVU2Hr1G017020**	**39326257**	**39327426**	**ABSCISIC ACID-INSENSITIVE 5-like protein 3**	**1**	**3.4E-24**
5H	**HORVU5Hr1G114420**	**643795341**	**643796978**	**Pentatricopeptide repeat-containing protein**	**1**	**7.9E-21**
2H	**HORVU2Hr1G080490**	**583370076**	**583390546**	**MADS-box transcription factor 27**	**1**	**9.7E-17**
1H	**HORVU1Hr1G039150**	**273602090**	**273604169**	**CCT motif family protein**	**1**	**7.6E-16**
1H	HORVU1Hr1G076380	514008170	514016820	Homeobox protein HOX1A	1	3.8E-11
2H	**HORVU2Hr1G013400**	**29123785**	**29127889**	**Pseudo-response regulator 7 (Phd-H1)**	**1**	**1.5E-10**
5H	HORVU5Hr1G086780	577242591	577250665	NAC transcription factor	1	1.8E-08
4H	HORVU4Hr1G071420	580566011	580574138	FRIGIDA-like protein	1	4.3E-08
7H	HORVU7Hr1G033820	69932169	69935928	auxin response factor 19	2	2.6E-05
7H	**HORVU7Hr1G023940**	**37608007**	**37620437**	**MADS-box transcription factor 25**	**1**	**3.7E-05**
2H	HORVU2Hr1G113820	729806528	729807892	WUSCHEL related homeobox 4	2	5.7E-05
2H	HORVU2Hr1G065450	449824251	449825110	flowering promoting factor 1	1	9.0E-05
7H	HORVU7Hr1G033820	69932169	69935928	auxin response factor 19	1	3.6E-04
2H	HORVU2Hr1G113820	729806528	729807892	WUSCHEL related homeobox 4	1	7.6E-04
2H	HORVU2Hr1G017270	40107707	40108374	WUSCHEL related homeobox 12	1	1.1E-03
2H	HORVU2Hr1G079220	573587961	573593078	cryptochrome 1	1	1.2E-03
7H	HORVU7Hr1G024000	37888165	37906750	MADS-box transcription factor 25	1	6.1E-03
1H	HORVU1Hr1G076480	514260347	514272318	trehalose phosphate synthase	3	9.1E-03
1H	HORVU1Hr1G076490	514369019	514370593	ovate family protein 4	1	9.2E-03
1H	HORVU1Hr1G076480	514260347	514272318	trehalose phosphate synthase	2	1.4E-02
2H	HORVU2Hr1G087310	627136997	627141298	AP2-like ethylene-responsive transcription factor	2	2.0E-02
1H	HORVU1Hr1G076480	514260347	514272318	trehalose phosphate synthase	1	8.1E-02
1H	**HORVU1Hr1G076430**	**514098183**	**514100647**	**Protein FLOWERING LOCUS T (Phd-H2)**	**1**	**1.4E-01**
2H	HORVU2Hr1G087310	627136997	627141298	AP2-like ethylene-responsive transcription factor	1	3.1E-01
2H	HORVU2Hr1G078880	571078484	571079654	Gibberellin receptor GID1L2	1	3.3E-01
2H	HORVU2Hr1G087260	626863212	626868543	NAC domain containing protein 6	1	3.8E-01
2H	HORVU2Hr1G065450	449824251	449825110	flowering promoting factor 1	2	6.3E-01

Table S5. Chromosome, gene position, haplotype block and p-value from the single-locus scan for the haplotypes tested for plant height. Haplotypes retained in the multi-locus multi-environment model and used to construct the adaptation classes are shown in bold.

chr	gene	GeneStart	GeneEnd	Annotation	block	pvalue
3H	**HORVU3Hr1G069960**	**529679099**	**529680512**	**RING/U-box superfamily protein**	**1**	**1.1E-17**
3H	HORVU3Hr1G030260	143069684	143079659	Cytochrome P450 superfamily protein	1	5.8E-17
3H	**HORVU3Hr1G033740**	**178816679**	**178819800**	**WRKY family transcription factor**	**1**	**1.6E-16**
3H	**HORVU3Hr1G090980**	**634078038**	**634081600**	**sdw1/denso**	**2**	**9.3E-16**
6H	**HORVU6Hr1G028790**	**115444384**	**115446574**	**WRKY family transcription factor**	**1**	**5.9E-15**
3H	**HORVU3Hr1G084360**	**607244129**	**607247713**	**ABSCISIC ACID-INSENSITIVE 5-like protein 2**	**1**	**7.2E-15**
2H	**HORVU2Hr1G072750**	**523377523**	**523379139**	**Protein TERMINAL FLOWER 1**	**1**	**9.7E-15**
3H	HORVU3Hr1G032230	163938676	163943047	Auxin response factor 1	1	2.3E-14
5H	HORVU5Hr1G014300	50382307	50386200	Auxin-responsive protein IAA30	1	4.1E-14
7H	**HORVU7Hr1G034180**	**70963343**	**70964797**	**UDP-Glycosyltransferase superfamily protein**	**1**	**3.8E-10**
6H	HORVU6Hr1G030080	124451142	124455037	HvCO7	1	1.4E-08
3H	HORVU3Hr1G031800	159598302	159604013	elongation defective 1 protein / ELD1 protein	1	2.6E-07
7H	HORVU7Hr1G034170	70870250	70870660	UDP-Glycosyltransferase superfamily protein	1	1.2E-06
5H	HORVU5Hr1G000510	2177242	2178665	Protein FANTASTIC FOUR 3	1	6.3E-06
5H	HORVU5Hr1G001090	3762210	3766599	BEL1-like homeodomain 6	2	4.5E-05
5H	**HORVU5Hr1G000010**	**248988**	**254172**	**sucrose transporter 4**	**1**	**6.5E-05**
3H	HORVU3Hr1G031020	152428027	152430096	Cytochrome P450 superfamily protein	1	9.2E-05
5H	HORVU5Hr1G014290	49955790	49957183	Auxin-responsive protein IAA31	1	1.5E-04
5H	HORVU5Hr1G013990	47175174	47177224	Cytochrome P450 superfamily protein	1	3.2E-04
5H	HORVU5Hr1G013880	47057115	47059536	Cytochrome P450 superfamily protein	1	3.9E-04
2H	**HORVU2Hr1G112280**	**725520002**	**725524192**	**AP2-like ethylene-responsive transcription factor**	**1**	**4.0E-04**
3H	HORVU3Hr1G084450	607796718	607799782	Cytochrome P450 superfamily protein	1	5.5E-04
3H	HORVU3Hr1G027430	116881300	116883595	Cytokinin dehydrogenase 2	1	8.3E-04
3H	**HORVU3Hr1G026300**	**106052467**	**106053272**	**NAM-like protein**	**1**	**9.0E-04**
3H	HORVU3Hr1G027590	119252012	119255903	Protein FLOWERING LOCUS T	1	1.3E-03
7H	HORVU7Hr1G033230	67727548	67732222	sucrose synthase 4	3	1.9E-03
6H	**HORVU6Hr1G031510**	**133941022**	**133941817**	**Auxin transporter-like protein 3**	**1**	**2.7E-03**
7H	HORVU7Hr1G033230	67727548	67732222	sucrose synthase 4	4	4.1E-03
3H	HORVU3Hr1G031460	156625895	156629839	Auxin-responsive protein IAA17	1	4.6E-03
7H	HORVU7Hr1G033230	67727548	67732222	sucrose synthase 4	1	5.1E-03
2H	**HORVU2Hr1G112280**	**725520002**	**725524192**	**(AP2-like ethylene-responsive transcription factor**	**2**	**5.4E-03**
3H	HORVU3Hr1G026990	113025336	113033485	Homeobox-leucine zipper protein family	1	6.0E-03
3H	**HORVU3Hr1G090980**	**634078038**	**634081600**	**gibberellin 20-oxidase 3**	**1**	**7.8E-03**
5H	HORVU5Hr1G001090	3762210	3766599	BEL1-like homeodomain 6	1	2.3E-02
3H	HORVU3Hr1G083820	605272081	605273358	NAC domain protein	1	3.1E-02

Chapter 7

A protocol combining statistical and crop growth modelling to evaluate phenotyping strategies useful for selection under different drought patterns

Daniela Bustos-Korts[1,2], Marcos Malosetti[1], Martin Boer[1], Scott Chapman[3],

Karine Chenu, Fred van Eeuwijk[1*]

1. Biometris, Wageningen University & Research Centre

2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)

3. CSIRO Plant Industry and Climate Adaptation Flagship

4The University of Queensland, CPS, Queensland Alliance for Agriculture and Food Innovation

7.1. Introduction

With the availability of cheaper molecular markers, genomic prediction has become a promising tool to increase the number of genotypes considered for selection (Poland *et al.*, 2012; Crossa *et al.*, 2013; Hickey *et al.*, 2014). In genomic prediction, additive and non-additive effects for the target trait (e.g. yield) are estimated in a training set of genotypes, which has genotypic and phenotypic observations. Those estimates are used to predict the phenotypes of the collection of genotypes for which no phenotypic information is available, the so called test set (Meuwissen, 2007). For predictions to be accurate, training and test sets must belong to the same target population of genotypes (TPG) (Albrecht *et al.*, 2014; Bustos-Korts *et al.*, 2016a). The TPG contains all possible genotypes that could be considered as selection candidates (Jannink *et al.*, 2010; Schulz-Streeck *et al.*, 2012; Albrecht *et al.*, 2014). In the same way as for the TPG, the environments where genotypes were phenotyped should represent well the target population of environments (TPE). The TPE delineates the future growing conditions of the genotypes in the TPG (Comstock & Moll, 1963; Cooper & Hammer, 1996; Cooper *et al.*, 2014a). Breeders aim to identify those genotypes in the TPG that are best adapted to the TPE, where adaptation can be understood as a better performance than a reference genotype in a defined environmental range (van Eeuwijk *et al.*, 2016). The adaptation patterns of the TPG across the TPE during the growing season are also called 'the landscape of trait dynamics' (Chapman *et al.*, 2003; Hammer *et al.*, 2005; Messina *et al.*, 2011; Cooper *et al.*, 2014b,a; Technow *et al.*, 2015).

Complex target traits like yield show low genomic prediction accuracy because they frequently suffer from low heritability and are regulated by a large number of loci with small effects (Crossa *et al.*, 2013; Sorrells, 2015). Complex traits can be decomposed into a number of underlying genetically-correlated traits, called 'intermediate traits' (Yin *et al.*, 2004), 'indicator traits' (Calus & Veerkamp, 2011), 'secondary traits' (Rutkoski *et al.*, 2016) or 'components' (Porter & Gawith, 1999) that might have a simpler genetic basis and larger heritability than the target trait (Yin *et al.*, 2004; Tardieu & Tuberosa, 2010; Cabrera-Bosquet *et al.*, 2016). For simplicity, we will use 'intermediate traits' to refer to all the traits that are genetically correlated to yield. As genetically-correlated traits are informative with respect to each other, modelling yield and its intermediate traits simultaneously allows to achieve larger yield genomic prediction accuracy, compared to single-trait genomic prediction (Dekkers, 2007; Jia & Jannink, 2012; Alimi, 2016; Biscarini *et al.*, 2017; Sun *et al.*, 2017). Another condition for multi-trait genomic prediction to show a larger accuracy than single trait prediction is that the heritability of the intermediate trait is large (Jia & Jannink, 2012).

Phenotyping additional intermediate traits implies an investment that does not always pay off by a larger prediction accuracy. Therefore, it is crucial for breeders to estimate in advance whether their phenotyping strategy for intermediate traits is likely to increase prediction accuracy of the target trait. This is especially relevant for high throughput phenotyping (HTP). HTP makes the phenotyping of additional traits affordable but may suffer from large measurement error. A large measurement error reduces trait heritability and prediction accuracy of the target trait (Cabrera-Bosquet *et al.*, 2012; Araus & Cairns, 2014; Yang *et al.*, 2014; Haghighattalab *et al.*, 2016; Rutkoski *et al.*, 2016).

A strategy to evaluate the potential of phenotyping strategies is by combining crop growth models and statistical-genetic models to characterize trait correlations and heritability over time (Cooper *et al.*, 2002). APSIM is an example of widely-used crop growth models, which characterizes system performance over time with an equal emphasis on crop, weather, soil and agronomic management (Wang *et al.*, 2002; Keating *et al.*, 2003; Holzworth *et al.*, 2014). The algorithms in APSIM predict yield as a nonlinear combination of intermediate phenotypes, which are calculated indirectly from environmental conditions and from a number of physiological parameters (Wang *et al.*, 2002; Keating *et al.*, 2003; Holzworth *et al.*, 2014). APSIM parameters correspond to basic mechanisms, at the bottom of trait hierarchy, that modulate crop response to the environmental conditions and can be regarded as constant across environments (Cooper *et al.*, 2002). APSIM parameters involve development, capture and use efficiency of environmental resources and biomass partitioning to the different plant organs. Genotypes can differ in their APSIM parameter values, leading to phenotypic differences for yield and intermediate traits across environments. Examples of phenotype prediction across environments using APSIM with genotype-dependent parameters can be seen in (Chapman *et al.*, 2003; Chenu *et al.*, 2009, 2011, 2013, Zheng *et al.*, 2012, 2013). More discussion about the combination of crop growth models and statistical models can be found in Bustos-Korts *et al.*, (2016b).

Simulated data of intermediate and target traits during the growing season is a useful resource to evaluate; i) multi-trait prediction using traits measured during the whole growing season and ii) yield predictions from traits measured early in the growing season. In both scenarios, intermediate traits can be measured at a single time point, or they can be monitored at multiple time points during the season to describe their dynamics. Characterizing trait dynamics allows to better capture the genotypic response to the environmental conditions integrated over the growing season and therefore might be more informative about genotypic performance than single-time point measurements (Malosetti *et al.*, 2006; van Eeuwijk *et al.*, 2010; Hurtado *et al.*, 2012; Hurtado-Lopez *et al.*, 2015). Simultaneous modelling of data

points over time is also a strategy to reduce the measurement error and to increase the heritability of traits measured with HTP (Rutkoski *et al.*, 2016).

In this paper, we propose a decision support tool based on the combination of statistical-genetic and crop growth models to design an effective phenotyping schedule across the Australian TPE for wheat. We will compare different strategies to integrate traits over time (i.e. penalized splines and nonlinear regression model), using an Australian wheat panel simulated with APSIM to grow over a sample of 40 environments representing water deficit patterns present in the Australian TPE.

7.2. Methods

7.2.1. Genotypic data

Data consisted of 199 genotypes characterized with 4,002 polymorphic SNPs with less than 1% missing data. Missing markers were replaced by imputed genotypic data using the *missForest* package in R, following the methodology explained in (Bogard *et al.*, 2014). SNPs with minor allele frequency below 0.05 were removed. These 199 genotypes are a sample of the Australian TPG constructed to represent the range in flowering time variation for Australian genotypes (Australian Wheat Flowering time Association Mapping panel, AWFAM). Most of the AWFAM genotypes have been used in previous research about phenotype prediction in Australian environments (Zheng *et al.*, 2013; Bustos-Korts *et al.*, 2016a). To characterise AWFAM population structure, a relationship matrix A was calculated from the SNPs following (Patterson *et al.*, 2006).

$$A = \frac{XX'}{n_m} \qquad (1)$$

The A matrix is of dimensions number of genotypes by number of genotypes and is proportional to the genetic covariance among individuals. X is a matrix of dimensions number of genotypes by number of SNPs, whose entries are marker scores, coded 0, 1 or 2, representing the number of copies of the minor allele. The marker scores were standardised as in (Patterson *et al.*, 2006). In (1), n_m is the number of markers. To infer the number of subpopulations present in AWFAM, we calculated the number of significant principal components after applying a spectral decomposition to the matrix A, where the significance was assessed following (Patterson *et al.*, 2006). The number of subpopulations is then the number of significant components plus one. The calculations were done in Genstat 18 (VSN-International, 2016).

Genotypes were grouped and assigned to subpopulations using a hierarchical clustering procedure applied to the significant principal components, following (Odong *et al.*, 2013). The cut-off for the dendrogram was chosen such that the number of subpopulations was equal to the number of significant ($p<0.05$) principal components plus one.

7.2.2. Phenotypic data

Phenotypic data consisted of the adjusted means for yield and heading date of the 199 genotypes belonging to the AWFAM panel observed in eight environments across the Australian wheat belt (Figure S1). The experimental design was a row-column design with two replicates. Adjusted means were calculated with the following mixed model:

$$y_{ijl(k)} = \mu + Rep_k + R_j + C_{l(k)} + G_i + \epsilon_{ijl(k)} \qquad (2)$$

In model (2), $y_{ijl(k)}$ is the phenotype of genotype i in replicate k, row j and column l within replicate k. μ is the intercept, Rep_k is the fixed effect of replicate k, R_j is the random effect of row j, $C_{l(k)}$ is the random effect of column l within replicate k, G_i is the fixed effect of genotype i. ϵ_{ijkl} is the vector of spatially correlated residuals modelling the local trend, with distribution $\epsilon_{ijkl} \sim N(0, R)$. R represents the Kronecker product of first-order autoregressive processes across rows and columns and σ_e^2 is the residual variance ($R \sim \sigma_e^2 (AR1 \otimes AR1)$).

7.2.3. Environments

APSIM-Wheat simulations were carried out for a TPE represented by four sites (Emerald, Narrabri, Yanco and Merredin) and 21 years (1993-2013), corresponding to a subset of the environments used in (Chenu *et al.*, 2013; Casadebaig *et al.*, 2016). This subset was chosen to represent the most common conditions at important wheat growing areas in the Australian wheat belt. Sowing settings corresponded to control conditions in (Casadebaig *et al.*, 2016), chosen to mimic local farming practices. Sowing date was set at 15th of May for all environments. More details about the sowing settings can be seen in Table 1 and in (Casadebaig *et al.*, 2016).

Table 1. Characteristics of the locations, soils and management regimes representing the target population of environments considering the period 1983-2013. Plant available water capacity (PAWC) is indicated for each soil, as well as the level of initial soil water used in the simulations, following (Casadebaig *et al.*, 2016). Applied nitrogen doses are indicated by 'a/b/c', with the fertilization applied at sowing (a), at the stage of 'end of tillering' (b) and at the stage 'mid-stem elongation, (c). Seasonal data considered the growing period between 15 of May and the maturity date averaged across genotypes.

Variable	Emerald	Merredin	Narrabri	Yanco
Latitude (degree)	-23.53	-31.50	-30.32	-34.61
Longitude (degree)	148.16	118.22	149.78	146.42
Rainfall pattern	summer dominant	winter dominant	summer dominant	evenly distributed
Annual rainfall (mm)	585	313	644	406
Seasonal rainfall (mm)	89	181	202	193
Daily mean temperature (Celsius)	17.5	12.9	13.5	11.9
Daily mean radiation ($MJ.m^{-2}$)	16.7	14.4	15.0	13.6
Soil type	black vertosol	shallow loamy duplex	grey vertosol	brown sodosol
PAWC (mm)	133.5	101.1	217.5	190.8
Sowing date	15/May	15/May	15/May	15/May
Sowing PAWC (mm)	132	39	175	99
Initial nitrogen ($kg\ ha^{-1}$)	30	30	30	50
Applied nitrogen ($kg\ ha^{-1}$)	50/0/0	20/20/30	130/0/0	40/40/40

7.2.4. Simulated phenotypic data

Phenotypic data was simulated in the following steps; 1) generate genotype-specific values for 12 APSIM parameters, regulating phenology, capture of environmental resources, resource use efficiency and biomass partitioning. These APSIM parameters were regulated by 300 SNPs with additive effects sampled from a Gamma distribution (See Figure 1 and section 2.4.1), 2) calculate yield and biomass at harvest from APSIM simulations for the whole TPE (section 2.4.2), 3) select a sample of the TPE that represents environments that contrast in their GxE (section 2.4.3), 4) run APSIM simulations for the sample of the TPE, saving phenology, yield at harvest and the daily output for biomass (section 2.4.3) and 5) add plot and measurement error to the APSIM output (section 2.5). The data generated after step 5) was used as input for multi-trait genomic prediction (section 2.6). For a schematic representation of these simulation steps, see Figure 1.

7.2.4.1. Genotype-specific parameters

The AWFAM panel was assumed to segregate for 12 of the APSIM parameters (Table 2). These parameters were chosen because they have an important impact on grain yield, as shown by global sensitivity analysis (Casadebaig *et al.*, 2016). The range of the parameter values was set to match the ranges that are commonly shown by wheat populations grown in Australia and that have been reported in the literature (Table 2). We estimated SNP effects on real phenotypic data for heading date and yield across the Australian wheat belt, using a single-environment GWAS model (Equation (3)).

$$y_i = \mu + x_{ik}\alpha_k + G_i + \epsilon_i \qquad (3)$$

In model (3), y_i stands for the phenotype of genotype *i*, μ is the intercept, x_{ik} is a vector that represents information of genotype i at marker k (0, 1, or 2 for the number of minor alleles) and α_k is the additive QTL effect (fixed) for marker k. G_i represents a polygenic effect for genotype *i*, with distribution $G_i \sim N(0, A\sigma_g^2)$. A is the additive genetic relationship matrix calculated from the molecular marker information as in (Rincent *et al.*, 2014). In this method, a specific A is calculated for each linkage group by excluding the markers on that particular linkage group. ϵ_i is the residual ($e_i \sim N(0, \sigma_e^2)$). The estimated marker additive effects (α_k) of heading date and yield across the Australian wheat belt were used to were used to define Gamma distributions for the QTL effects underlying individual traits. The estimation procedure was done by maximum-likelihood using the DISTRIBUTION directive in Genstat 18 (VSN-International, 2016). As the distribution parameters (shape and rate) slightly differed between environments, we used the median of the Gamma shape and rate across environments.

The univariate distributions for physiological parameters were turned into a multi-variate distribution by considering the physiological evidence for correlations between some of the APSIM parameters. Most of the parameters were assumed to be uncorrelated, except for the transpiration efficiency coefficient and radiation use efficiency (r= -0.40), the number of grains per gram of stem at flowering and maximum grain size (r=-0.50) and maximum grain size and potential grain filling rate (r=+0.45). These correlations were set to match physiological constraints that have been observed in real experiments (Slafer & Savin, 1994; Monneveux *et al.*, 2006; Sadras & Lawson, 2011; Bustos *et al.*, 2013).

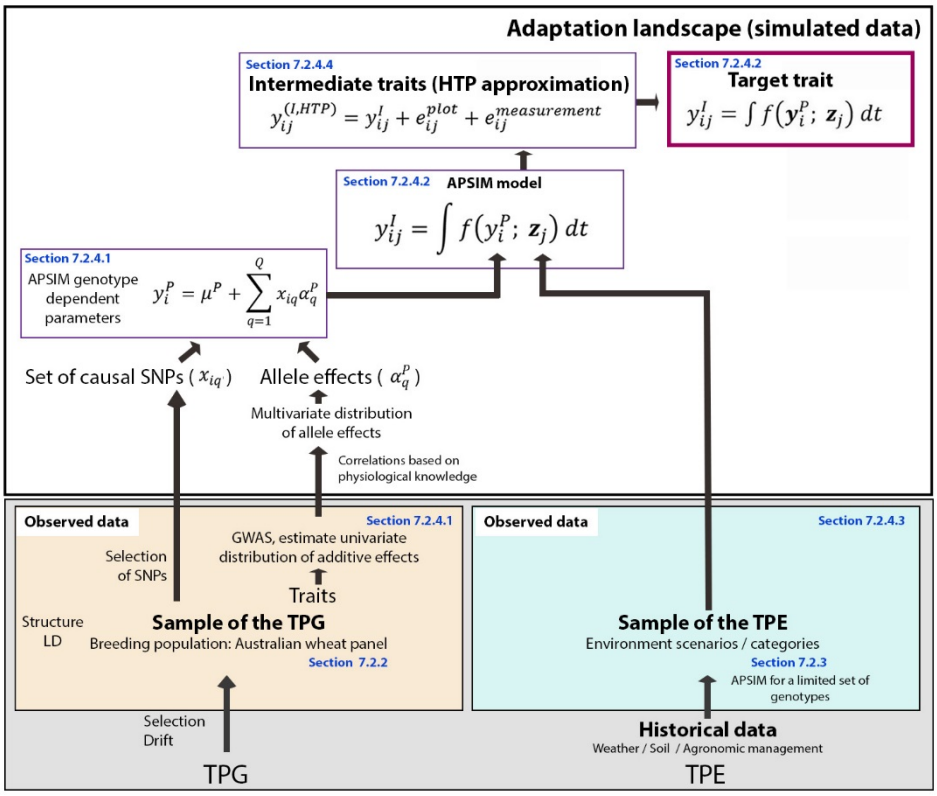


Figure 1. Simulation steps to generate phenotypes for a set of genotypes across environments. Bottom left; an Australian wheat panel is defined as a sample of the target population of genotypes (TPG). For this sample of genotypes, phenotypic data have been collected as well as SNP data. The phenotypic data are associated with SNP data in univariate GWAS analyses. From these analyses, empirical distributions for the additive effects of QTLs underlying these phenotypes are obtained. Physiological knowledge on trait correlations is used to define genetic correlations between APSIM parameters (y_i^P). These correlations are included in a multi-variate description of the QTLs underlying APSIM parameters. From this distribution, genotype specific APSIM parameters (y_i^P) are generated and assigned to a subset of SNPs. Bottom right; we have historical environmental data defining the target population of environments (TPE). We use APSIM to identify environment scenarios (water deficit patterns). The environmental data of the selected scenarios and the genotype-dependent APSIM parameters are used to generate intermediate traits over time (y_{ij}^I). In a breeding programme, these intermediate traits are unknown, but we can approximate intermediate traits by high throughput phenotyping techniques, where the intermediate traits will come with plot (e_{ij}^{plot}) and measurement error ($e_{ij}^{measurement}$). The target trait (y_{ij}^T) is modelled as a function of intermediate traits. For a detailed explanation of each step, see the section numbers written in blue font.

To impose parameter correlations, 300 additive SNP effects were sampled with copulas. We set the number of QTLs to 300 to achieve an average distance between QTLs of about 10 cM. Briefly, the dependence between parameter additive effects is determined by a uniform distribution and then, the marginal Gamma distribution for the additive effects is

imposed on the dependence structure (Nelsen, 2013). Effects were sampled with the R package 'copula' (R Core Team, 2016). Since the Gamma distribution always takes positive values, 50% of the additive effects for each parameter was randomly assigned a negative sign (Pérez-Enciso *et al.*, 2017). Sign allocation was done independently for each parameter, but an exception was made for the following pairs of correlated parameters; a) transpiration efficiency coefficient and radiation use efficiency, b) grains per gram of stem and maximum grain size and c) maximum grain size and potential grain filling rate. In these three exceptional cases, the sign was assigned jointly for a proportion of the additive effects, instead of independently, as for the other parameters. If the parameter correlation was positive a negative sign was assigned jointly to 60% of the loci chosen at random (i.e. negative sign was assigned jointly to the following pairs of parameters; a) transpiration efficiency coefficient and radiation use efficiency and c) maximum grain size and potential grain filling rate). If the target correlation was not achieved by this random allocation, we repeated the sampling for sign allocation until the target correlation was achieved. For all interactions, the percentage of negative signs was equal to 60%. If the correlation was negative (in the case of pair grains per gram of stem and maximum grain size), opposite signs were assigned randomly to a 60% of the additive effects. This means that 60% of the loci would get a negative sign for one parameter and a positive sign for the other parameter, iterating until the negative correlation was achieved.

Additive effects per trait were allocated to a random sample of 300 SNPs out of the 4.002 SNPs available for the AWFAM panel. Because of the random nature of the sampling process, some SNPs were more related to population structure than others. For that reason, SNPs with additive effects differed in their Fst values (corresponding to a neutral genetic architecture, following Pérez-Enciso et al., 2017). Additive effects were assigned at random to the 300 SNPs, except for the largest additive effects regulating photoperiod sensitivity and vernalization requirements. In this case, the largest additive effects were assigned to SNPs at the position of nine known flowering time genes/QTLs, which showed moderate to large Fst (between 0.11 and 0.72 with a median of 0.5). We used the 300 SNPs and their additive effects to generate genotype-specific APSIM parameter values for each genotype (y_i^P), Equation (4). These genotype-specific parameters are constant across environments.

$$y_i^P = \mu^P + \sum_{q=1}^{Q} x_{iq}\alpha_q^P \qquad (4)$$

Where μ^P is the mean of the range for APSIM parameter *P*, x_{iq} is an indicator variable with values -1, 0 and 1 that represents information of genotype *i* at marker *q*. α_q^P is the additive QTL effect (fixed) for marker/QTL *q* and parameter *P*.

Table 2. APSIM genotype-specific parameters, with the shape (k) and scale (b) of the marginal Gamma distribution of the additive effects, lower and upper limit for the simulated population and trait increasing allele for the selective SNPs. For all traits, the marginal Gamma distribution followed k=1.299 and b=13.6, except for photop_sens, tt_floral_initiation and vern_sens, for which k=0.700 and b=13.6. When units are indicated with [ ], it means that the parameter is dimensionless.

Parameter	APSIM name	Units	Default	low_lim	high_lim	References
			APSIM			
Fraction of leaves senescing per main stem node	fr_lf_sen_rate	[ ]	0.035	0.028	0.042	(Christopher *et al.*, 2016)
Number of grains per gram of stem at flowering	grains_per_gram_stem	grains $g_{stem}{}^{-1}$	25	20	25	(Dodig *et al.*, 2012)
Lower limit for water uptake	ll_modifier	[ ]	1	0.9	1.1	Manschadi et al (2006)
Maximum grain size	max_grain_size	mg	0.041	0.03	0.06	(Groos *et al.*, 2003)
Sensitivity to photoperiod	photop_sens	[ ]	2	1.5	3	(Zheng *et al*., 2012)
Potential grain filling rate	potential_grain_filling_rate	g grain^{-1} d^{-1}	0.0025	0.0016	0.0026	(Wang *et al.*, 2009)
Transpiration efficiency coefficient	transp_eff_cf	[ ]	0.006	0.0045	0.0065	(Schoppach & Sadok, 2013)
Thermal time required to reach floral initiation	tt_floral_initiation	^{o}C d	555	455	555	Within the range of (Zheng *et al*., 2013)
Sensitivity to vernalization	vern_sens	[ ]	1.5	1.5	2.5	(Zheng *et al*., 2012)
Canopy extinction coefficient	y_extinct_coef	[ ]	0.45	0.4	0.6	(Isidro *et al.*, 2012)
Biomass partitioning to leaf	y_frac_leaf	[ ]	0.6	0.55	0.65	(Álvaro *et al.*, 2008)
Radiation use efficiency	y_rue	g MJ^{-1}	1.24	1.01	1.4	(Acreche *et al.*, 2009)

7.2.4.2. APSIM simulations

The 12 parameters showing variation for the 199 genotypes were used to simulate intermediate traits using model (5).

$$y_{ij}^{I} = \int f\left(\boldsymbol{y}_i^{P};\, \boldsymbol{z}_j\right) dt \qquad (5)$$

In model (5), y_{ij}^{I} is the intermediate trait of genotype *i* and environment *j*, which is modelled as a function of multiple APSIM parameters $\boldsymbol{y}_i^{P}$ (12 of them are genotype-specific and multiple environmental inputs, $\boldsymbol{z}_j$ integrated over time (Figure 1). A description of the ranges of the genotype-dependent APSIM parameters can be seen in Table 2. For a more detailed description of APSIM see the user manual (http://www.apsim.info/). The function *f* (;) embodies the algorithms that transforms APSIM parameters and environmental inputs into intermediate phenotypes (e.g. biomass). In APSIM, the target trait (yield) for genotype *i* in environment *j* (y_{ij}^{T}) is modelled as a function of intermediate traits (e.g. biomass, grain number, grain weight) and the environment over time, following Equation (6);

$$y_{ij}^{T} = \int f\left(y_{ij}^{I}\right) dt \qquad (6)$$

APSIM simulations were run first for a large sample of the TPE (four locations between 1993-2013), saving only yield and biomass at harvest. These phenotypes were used to characterise the GxE patterns and identify the most important environments driving GxE (Figure 1, bottom right). Once a limited set of environments driving GxE was identified, we ran a more detailed APSIM simulation. The output of this detailed APSIM simulation consisted of phenology, biomass and yield of each genotype, environment and day during the whole growing season and was used to add plot and measurement error and for the genomic prediction steps.

7.2.5. Characterization of the TPE and selection of environments for phenotyping over time

APSIM yield and biomass at harvest simulated for the 199 genotypes, 21 years and four locations were used to characterize the GxE patterns in a sample of the TPG and TPE. The main goal of the characterization of GxE is to select a limited number of environments exhibiting the most important GxE patterns. One way of characterising GxE is by fitting and interpreting an Additive Main Effects and Multiplicative Interaction (AMMI) model (Gauch & Zobel, 1997; Gauch, 2013; van Eeuwijk *et al.*, 2016).

$$y_{ij} = \mu + E_j + G_i + \sum_{m=1}^{M} b_{im} z_{jm} + \epsilon_{ij} \qquad (7)$$

In model (7), y_{ij} represents the phenotype of the i^{th} genotype in the j^{th} environment, μ stands for the intercept, G_i is the fixed effect of the i^{th} genotype and E_j is the fixed effect of the j^{th} environment. The interaction in an AMMI model is described by M multiplicative terms that consist of products of the genotypic sensitivity b_{im} (genotypic score) and an environmental score z_{jm}. Finally, ϵ_{ij} is a residual term, that contains the part of the two-way analysis of variance interaction that is not explained by the AMMI interaction terms and a contribution of the plot error. Genotypic and environmental scores allow visualising the GxE interaction patterns in the form of biplots. We used an AMMI-2 biplot (scatter plot of the first two multiplicative terms) to identify groups of environments that induce similar stress reactions on the genotypes. These groups were identified by assessing the angle between the environmental vectors; if the angle is small, those environments can be interpreted as belonging to the same environmental group (Kempton, 1984; Malosetti *et al.*, 2013; van Eeuwijk *et al.*, 2016).

We also characterized GxE by quantifying the contribution of locations and years to the GxE variance.

$$y_{ijk} = \mu + E_j + G_i + GL_{ij} + GY_{ik} + GLY_{ijk} \qquad (8)$$

In model (8), y_{ijk} is the phenotype (yield or biomass) for genotype *i* in location *j* and year *k*, μ stands for the intercept, E_j is the fixed effect of environment *j*, G_i is the random effect of genotype *i*, GL_{ij} is the random interaction between genotype *i* and location *j*, GY_{ik} is the random interaction between genotype *i* and year *k* and GLY_{ijk} is a random term that contains the residual GxE. In this case, no extra error term was added to the model because as APSIM is fully deterministic, all the residual variance corresponds to GxE.

Besides the implicit environment characterization based on phenotypes across environments, we also used explicit environmental information about the dynamics of water deficit patterns during the growing season. To calculate the water deficit patterns, we ran APSIM for a genotype that had the average population value for each parameter (μ^P). We saved the water supply/demand ratio generated by APSIM to provide an explicit representation of the water availability in the soil, as perceived by the crop (Chenu *et al.*, 2011, 2013). The water supply/demand ratio indicates the degree to which the soil water extractable by the roots ('water supply') is able to match the potential transpiration ('water demand'). The water demand (mm) corresponds to the amount of water the crop would have transpired in the absence of soil water constraint and is estimated daily based on the amount of crop growth on that day (g mm^{-2}), and the atmospheric saturation vapour pressure deficit (kPa).

We used the AMMI biplot, the variance components calculated from model (8) and the supply/demand ratio across the TPE to select three scenarios that represent well the most important GxE patterns. These three scenarios will be used for further analysis, assessing the convenience of phenotyping additional traits (biomass) and integrating them into the genomic prediction for yield. In those selected environments, we quantified for each day the contribution of GxE to the total phenotypic variance for biomass.

$$y_{ij} = \mu + E_j + G_i + GE_{ij} \qquad (9)$$

Model (9), was fitted for each day independently. Here, y_{ij} is the biomass for genotype *i* in environment *j*, μ is the intercept, E_j is the fixed effect of environment *j*, G_i is the random effect of genotype *i* and GE_{ij} is the random GxE. For each environment, we also quantified the autocorrelation for biomass over time. The autocorrelation was calculated as the Pearson correlation coefficient between biomass lagged by 15 days (biomass at day t and day (t-15)).

7.2.6. Plot and measurement error

The three environment scenarios representing the range of GxE patterns present in the TPE were used to simulate a field trials in which biomass dynamics are characterized with HTP. To simulate field biomass, we added an experimental and a measurement error to the APSIM output. This is a necessary step to achieve realistic phenotypes because APSIM is a fully deterministic model.

Experimental error

To calculate the experimental (plot) error, we considered a heritability of 0.50 for yield and 0.70 for biomass. Heritability was calculated as in Equation (10).

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{plot}^2} \qquad (10)$$

The genotypic variance (σ_g^2) was assumed to be equivalent to the variance of APSIM biomass for a given day. The APSIM yield and biomass genotypic values do not contain error because APSIM is a fully deterministic model. Therefore, phenotypic differences in the same environment can be interpreted as been genetic. The experimental error (σ_{plot}^2) was sampled jointly for yield and biomass from a multivariate normal distribution with a covariance of 0.70 and a variance of 1.00. The relatively high covariance for the error of yield and biomass was defined to preserve the large phenotypic correlation for these traits commonly observed

in experimental data (Reynolds *et al.*, 2007; Bustos *et al.*, 2013). The phenotypic value for genotype *i* and day *j* (y_{ij}^{direct}) was calculated with Equation (11):

$$y_{ij}^{(I,direct)} = y_{ij}^{I} + e_{ij}^{plot} \qquad (11)$$

Where y_{ij}^{I} is the APSIM phenotype for an intermediate trait *I*, genotype *i* and day *j* and e_{ij}^{plot} is the experimental (plot) error for genotype *i* and day *j*. As biomass genotypic variance changes over time, we rescaled the size of the plot error (e_{ij}^{plot}) to keep heritability constant and equal to 0.70 during the growing season. To achieve this constant heritability, e_{ij}^{plot} was rescaled according to the phenotypic variance at a particular day during the growing season (systematic error), maintaining the heritability constant.

Homogeneous measurement error over time

Besides the experimental (plot) error, we added a measurement error that simulates the HTP approximation of y_{ij}^{direct};

$$y_{ij}^{(I,HTP)} = y_{ij}^{I} + e_{ij}^{plot} + e_{ij}^{m} \qquad (12)$$

In model (12) $y_{ij}^{(I,HTP)}$ is the phenotype measured by HTP, y_{ij}^{I} and e_{ij}^{plot} are the same as in model (11) and e_{ij}^{m} is the measurement error for genotype *i* and day *j*. Measurement error (e_{ij}^{m}) was sampled independently for each environment, trait and day (random error). We examined eight levels of measurement error size. The size of e_{ij}^{m} was defined to achieve an R^2 between the $y_{ij}^{(I,HTP)}$ and $y_{ij}^{(I,direct)}$ of 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80 and 0.90. For each of these measurement error levels, the relative size of the measurement error with respect to the variance of $y_{ij}^{(I,direct)}$ was constant over time. We combined these eight measurement error levels (R^2) with five levels of measurement intervals (every 5, 10, 15, 20 and 25 days) in a factorial way. Thus, for each environment, we obtained 40 phenotyping schedules differing in their measurement error size and measurement interval.

Measurement error as a function of canopy growth

In the previous section, we assumed that HTP has a homogeneous measurement error over time and across genotypes. However, HTP measurement error size usually changes over time, depending on crop dynamics like canopy closure, i.e. the error gets larger as canopy closes and it gets reduced again with the onset of senescence. These dynamics can be taken into account when simulating measurement error. In this paper, we assumed that the highest

R^2 between HTP phenotyping and direct/destructive phenotyping would be 0.60, with a canopy cover of 10%, as reported by (Grieder *et al.*, 2015). We assumed a quadratic R^2 reduction as a function of canopy cover (Figures 2a and 2b). When combining this R^2 function with the green canopy cover output from APSIM, the R^2 between the HTP measurement and the direct biomass measurement followed a genotype-specific curve, as shown for five genotypes in Figure 2c.

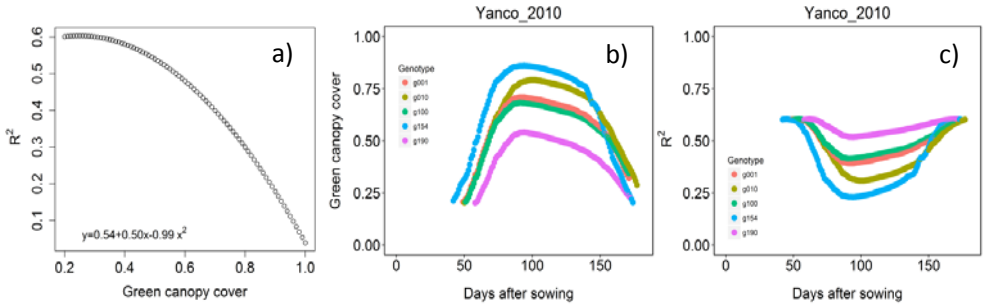


Figure 2. R^2 between the HTP and direct measurement of biomass as a function of green canopy cover (a), Green canopy cover dynamics for a random sample of five genotypes (b) and genotype-specific R^2 between HTP and direct measurement during the growing season (c).

7.2.7. Statistical Modelling of phenotypes over time

The HTP data simulated above was used to compare prediction accuracy calculated from a single-trait (yield) or from multiple traits modelled simultaneously (biomass dynamics and yield). In this section, we describe the statistical models for phenotypes over the growing season.

Logistic regression

A logistic function was fitted independently to the biomass HTP data of each genotype.

$$y_{ij} = \frac{L}{1+e^{-k(t-t_0)}} \qquad (13)$$

In model (13), y_{ij} is the biomass for genotype *i* at day *t*, *L* is the curve's maximum value (asymptote), k is the slope and t_0 is the day at which biomass stops growing exponentially (inflexion point). The curve fit was done with the *nls* function of the *stats* package in R (R Core Team, 2016).

P-Splines

Mixed model P-splines were fitted to the time series data for biomass during the growing season (Eilers & Marx, 1996; Eilers *et al.*, 2015). The biomass inflexion point was calculated as the moment at which the daily biomass change is maximum. The timing of the inflexion point and the phenotypic value at that point were calculated and further used as correlated traits for genomic prediction.

Spline fitted values were calculated with the following model:

$$y_{it} = b_0 + b_i t_t + \sum_k v_k B_k(t_t) + G_i + Glin_{it} + \sum_k \alpha_{ki}\beta_k(t_t) + e_{it} \qquad (14)$$

In Equation (14), y_{it} is biomass of genotype *i* at time point *t*, b_0 is the intercept, $b_i t_t$ is a term representing the linear trend over time for the population mean, $\sum_k v_k B_k(t_t)$ is the non-linear trend for the population mean over time, with v_k indicating the effects for the B-splines with *25* segments (*k*=1,2,…25). G_i stands for the effect of genotype *i*, $Glin_{it}$ is the linear deviation of genotype *i* from the population mean at time point *t*, and $\sum_k \alpha_{ki}\beta_k(t_t)$ is the sum of non-linear (smooth) deviations for genotype *i* over all B-spline segments. e_{it} is the residual.

Heritability for the logistic parameters and spline inflexion point was calculated as the R^2 between the curve characteristics (logistic parameters /spline inflexion point) with and without error.

7.2.8. Genomic prediction

7.2.8.1. Single trait predictions (yield)

Single trait genomic prediction for yield was carried out with the GBLUP model (Equation 14).

$$y_i = \mu + G_i + e_i \qquad (15)$$

In Equation (15), y_i is yield of genotype i, μ is the intercept, G_i stands for the random genotype effects that follow a distribution $G_i \sim N(0, A^{AB}\sigma_g^2)$. A^{AB} is the additive relationship matrix, following (Astle & Balding, 2009). The predictions were calculated using GenStat 18th edition (VSN-International, 2016)

Multi-trait predictions (yield and biomass dynamics)

Multi-trait models considered either two traits or multiple traits. For two-trait models, we used yield and one of the biomass characterizations from the fit of logistic curves or P-splines. For the logistic curves, these characterizations were; asymptote, inflexion day and maximum growth rate. For the spline approximation of biomass we used; spline fitted values at flowering, spline fitted values at maturity, spline inflexion point, spline maximum growth rate. Multi-trait models used yield and a combination of the logistic curves or spline fits. All traits were standardized before entering the genomic prediction models. Predictions were made with the following model:

$$y_{ik} = \mu + G_i + GT_{ik} + e_{ik} \qquad (16)$$

In model (16), y_{ik} is the phenotype for genotype *i* and trait k, μ is the intercept, G_i is the main effect of genotype *i*, following a variance-covariance $G_i \sim MVN(0, A^{AB}\sigma_g^2)$, were A^{AB} is the additive relationship matrix (Astle & Balding, 2009). GT_{ik} is the genotype by trait interaction, with a variance-covariance structure $GT_{ik} \sim MVN(0, I\sigma_{gk}^2)$. e_{it} is the residual.

Early predictions (use early biomass to predict final yield)

For early predictions, we used the P-spline fitted values for biomass accumulated between 20 days after sowing and flowering. The measurement intervals were the same as for the biomass measurements during the whole season (i.e. every 5, 10, 15, 20 and 25 days). Yield predictions from spline fitted values for biomass at flowering were done with a single trait model (15).

7.2.9. Prediction accuracy

Prediction accuracy was calculated as the Pearson correlation coefficient between the APSIM phenotypes (genotypic value) and the predicted phenotypes (Meuwissen *et al.*, 2001), considering a training set of 130 genotypes and a validation set of 69 genotypes. 50 training sets were constructed with the uniform sampling method described by (Jansen & van Hintum, 2007; Bustos-Korts *et al.*, 2016a). We calculated mean predictive ability and standard error across 50 training set realizations. To comply with the normality assumption, correlations were analysed on a transformed scale using Fischer's z transformation, $z = \frac{1}{2}\left(ln\left(\frac{1+r}{1-r}\right)\right)$ and means were back transformed using $r = \frac{\exp(2z)+1}{\exp(2z)-1}$ before reporting them.

7.3. Results

7.3.1. Population structure and genotype-specific parameters

A spectral decomposition of the kinship matrix showed four eigenvalues to be significant, suggesting the presence of five subpopulations. The first eigenvector, explained 36.8% of the variation (Figure 3). In general, population structure explained a low to moderate proportion of the variation of the APSIM parameters. In general, APSIM parameters showed low to intermediate correlation with population structure. Correlations with PC1 ranged from -0.46 (ll_modifier) to +0.31 (vern_sens). Correlations with PC2 ranged from -0.27 (y_rue) to +0.69 (transp_eff_cf) and correlations with PC3 ranged from -0.51 (max_grain_size) to +0.40 (grains_per_gram_stem), Figure 3. The intermediate size of the proportion of explained variation of APSIM parameter values by the population structure is in line with the fact that SNPs with additive effects were sampled at random and some of them had moderate to large Fst.

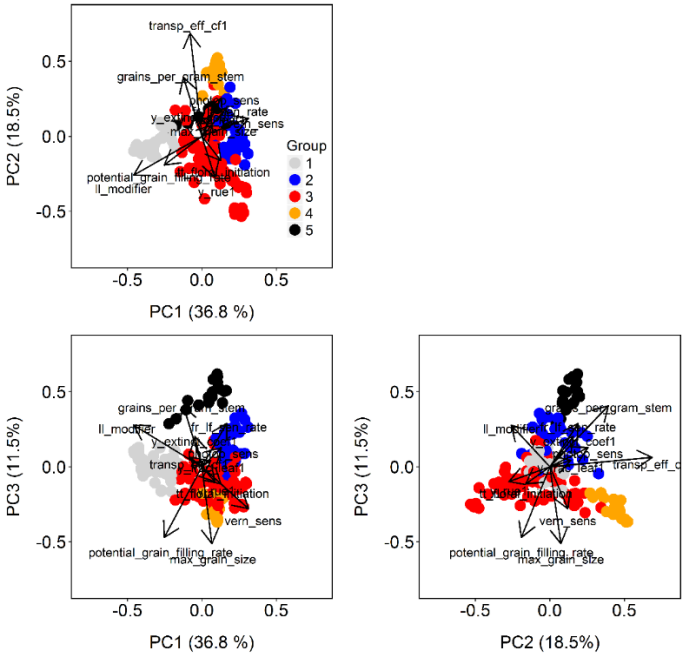


Figure 3. Population structure as revealed by principal components extracted from the matrix of marker scores. Directions of greatest change for a set of physiological parameters have been projected on the biplots to help in interpretation. The length of the physiological parameter representations is proportional to the amount of variation explained by the kinship principal components.

7.3.2. Represent GxE patterns with a sample of the TPE

We used the APSIM-simulated yield of the 199 genotypes grown across a large sample of the TPE (4 locations and 21 years, 1993-2003) to identify a subset of environments that represent the most important growing conditions driving GxE. In the simulated sample of the TPE, yield GxE was almost three-quarters of the total phenotypic variance (Table 3). The largest proportion of the GxE (81%) was driven by GLY_{ijk}, followed by GL_{ij} (33% of the GxE variance) and GY_{ik} with a 17% of the GxE variance. The GxE patterns observed in the AMMI biplot (Figure 4) are closely related to the water deficit dynamics (Figure 5). GxE is driven to a large extent by the contrast between environments that have no/mild water deficit (e.g. most of the environments in Yanco) and those that suffer from severe drought starting before flowering (e.g. most of the environments in Emerald and Merredin). The AMMI biplot also allows examining the contribution of each location to GLY_{ijk}. This contribution can be observed as the changes in the AMMI environment scores for locations across years (Figure 4). Narrabri showed to be an important contributor to GLY_{ijk}. For example, yield in Narrabri_2005 is more correlated to the less dry environments like Yanco_2010, whereas Narrabri_1997 is more correlated to the dry environments like Emerald_1993. The environmental conditions also modified mean yield and biomass across environments. Locations that commonly suffered from drought and that have soils with a lower plant available water capacity, like Emerald and Merredin, showed a lower yield and biomass than Narrabri and Yanco (Figures 5 and S2, Table 1).

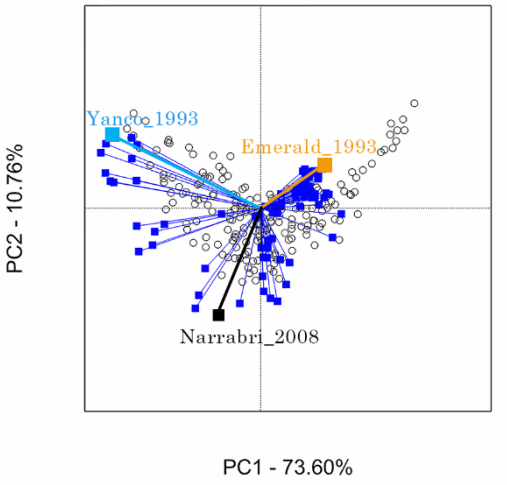


Figure 4. AMMI biplot for grain yield in Emerald, Merredin, Narrabri and Yanco during 1993-2013. White circles represent genotype scores and blue squares represent environment scores. Selected environments for further analysis are: Emerald_1993 (orange square), Narrabri_2008 (black square), Yanco_2010 (cyan square).

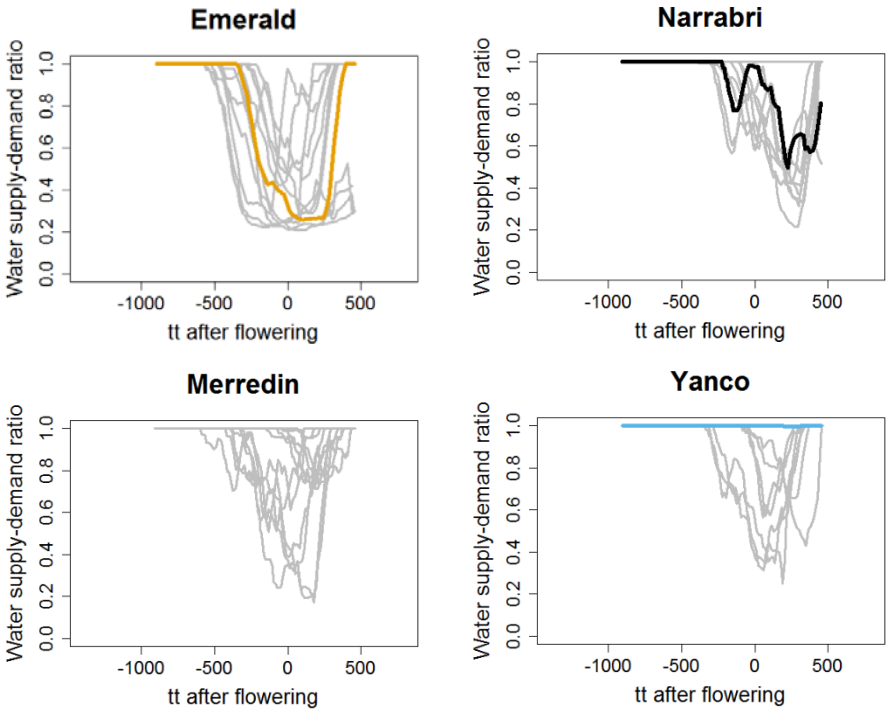


Figure 5. APSIM simulated water supply-demand ratio (stress index) for a genotype with average parameter values growing in Emerald, Narrabri, Merredin and Yanco during 2003-2013. Supply-demand dynamics shown in colours represent the environments used to evaluate the phenotyping strategies (orange=Emerald_1993, black=Narrabri_2008, blue=Yanco_2010).

Table 3. Variance components and standard error as estimated by mixed model analysis of the APSIM output of yield and biomass of the 199 genotypes in the 84 environments used to characterize the TPE (Emerald, Merredin, Narrabri and Yanco between 1993 and 2013).

	Yield		Biomass	
Component	**Variance**	**s.e.**	**Variance**	**s.e.**
Genotype (G_i)	43.89	5.66	349.76	41.94
Genotype.Location (GL_{ij})	41.44	2.58	246.40	14.68
Genotype.Year (GY_{ik})	21.01	0.85	55.51	2.07
Genotype.Location.Year (GLY_{ijk})	62.20	0.81	138.43	1.80

After assessing the GxE patterns and the water deficit dynamics, we selected the following environments that represent common water deficit types driving GxE; a) Emerald_1993 (very dry environment, with drought starting early in the growing season), b) Narrabri_2008 (intermediate post-anthesis drought) and c) Yanco_2010 (no drought). These environments were used for a more detailed phenotyping and genomic prediction (Table S1).

7.3.3. APSIM trait dynamics and correlations

The GxE observed for yield across the three selected environments can be further characterised by assessing the correlations between yield (the target trait) and intermediate traits at the end of the growing season (biomass, grain weight and grain number at harvest) and phenology (flowering time). In the three selected environments, yield showed intermediate to high correlation with biomass at harvest. However, this correlation was larger in the environments that did not suffer from severe drought (Figure 6). The correlation between yield and biomass at harvest was largest in Yanco_2010 (r=0.90), followed by Narrabri_2008 (r=0.72) and Emerald_1993 (r=0.37). Water deficit patterns also modified the correlation between yield, grain weight and grain number at harvest. In Yanco_2010 and Narrabri_2008, yield was mostly correlated to grain number and less to grain weight (Yanco_2010; r=0.91 for grain number and r=0.30 for grain weight. Narrabri_2008; r=0.62 for grain number and r=0.37 for grain weight). In contrast, in the dry environment Emerald_1993, yield was more associated to grain weight than to grain number (r=0.71 for grain weight and r=0.18 for grain number). Flowering time also contributed to the different performance of genotypes across environments. In Emerald_1993, yield was negatively correlated to flowering, in line with the well-known mechanisms of stress avoidance by escape to final drought. In contrast, when drought was mild/absent, a longer cycle allowed to achieve more biomass and therefore, a larger yield.

The way in which GxE is build up during the growing season can also be assessed by observing the trait correlations over time. Figure 7a shows the correlations between biomass during the growing season and final yield. In Yanco_2010, correlation of biomass and yield is large during the whole season (~0.75), with a small increase after flowering. In Narrabri_2008, the correlation of biomass shows a larger change throughout the season, starting in ~0.30, to reach ~0.65 after flowering (Figure 7a). Emerald_2009 shows the largest change in the correlation between biomass and yield. Here, yield is negatively correlated to biomass during the beginning of the growing season (~-0.40), has a correlation close to zero around flowering and reaches ~0.35 during grain filling. The inspection of the correlations between biomass and yield over time suggests that biomass can be used for early selection using correlated traits in humid/mild-stress environments, but not in dry environments.

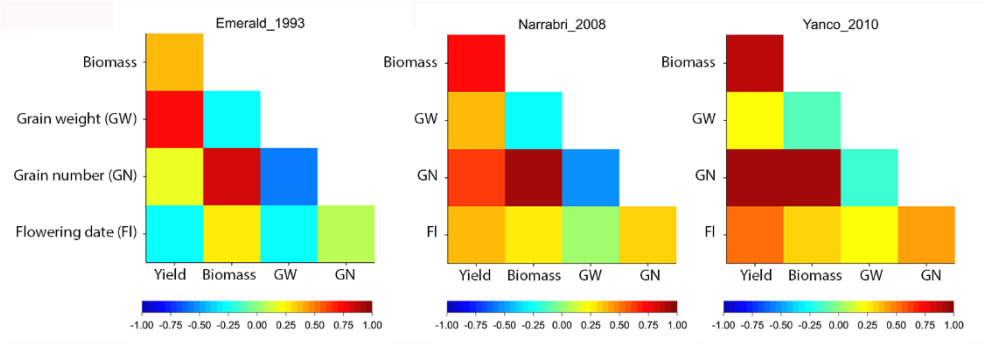


Figure 6. Trait correlations between APSIM yield, biomass, grain weight, grain number and flowering date simulated in Emerald_1993 (very dry environment, with drought starting early in the growing season), Narrabri_2008 (intermediate post-flowering drought) and Yanco_2010 (no drought).

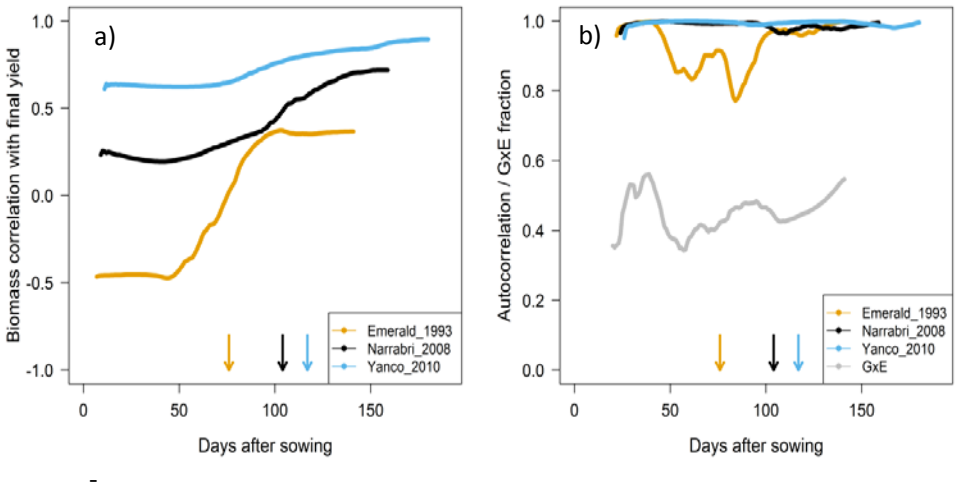


Figure 7. a) Correlation between biomass as measured at individual days and final yield. In both panels, arrows indicate mean flowering date for individual environments. b) Within environment autocorrelation (lag=15 days) of biomass in Emerald_1993 (very dry environment, with drought starting early in the growing season, orange lines), Narrabri_2008 (intermediate post-flowering drought, black lines) and Yanco_2010 (no drought, blue lines). Grey line gives GxE as fraction of the total phenotypic variance for biomass at individual days.

Examining the autocorrelation of intermediate traits like biomass helps to identify critical moments during which genotypic changes occur. The importance and frequency of these changes in genotypic ranking influence the degree of success of using early measurements of intermediate traits to improve the predictions for the target trait. In general, biomass autocorrelation was moderate to large for all the environments that we analysed (Figure 7b). However, the early drought-stress environment Emerald_1993 showed a lower biomass autocorrelation during the first part of the growing season, coinciding with the onset of water stress. For the environment Narrabri_2008 with mild after flowering stress, biomass autocorrelation was reduced after flowering, also coinciding with the start of drought stress in this environment. In the non-stressing environment Yanco_2010, biomass autocorrelations were large and almost constant (around 0.95) during the whole growing season (Figure 7). This means that, in Yanco_2010, biomass accumulation curves for all genotypes run almost parallel. Therefore, biomass in the non-stress environment Yanco_2010, measured early during the growing season, shows promise as an intermediate trait for yield prediction. Changes in biomass genotype ranking over time is related to the partitioning of the variance components over time. Figure 7b shows that GxE for biomass is largest at the beginning of the growing season (more than 50% of the phenotypic variance) and it is slightly reduced around flowering, oscillating at around 40% of the phenotypic variance.

7.3.4. Dynamics and QTLs for biomass accumulation

The shapes of biomass accumulation curves were different across environments; in Emerald_1993, biomass followed a symmetric curve, with a long asymptotic phase that was reached at about 10 days after flowering for most genotypes (Figure 8, upper panels). Most of the small irregularities in the biomass accumulation curve occurred before flowering, coinciding with the lower autocorrelation observed in this period (Figure 7b). The early start of the asymptotic phase in agreement with a very dry grain filling period that limited biomass accumulation. In Narrabri_2008, biomass accumulation over time also followed a symmetric curve. However, the linear biomass accumulation period was longer than in Emerald_1993, with a short asymptotic phase starting about 20 days after flowering. Biomass accumulation over time was more asymmetric in Yanco_2010, compared to Emerald_1993 and Narrabri_2008. Here, biomass had a longer initial lag-phase, followed by a delayed exponential phase and a short asymptotic phase starting at about 20 days after flowering.

To characterize the genetic basis of biomass dynamics during the growing season, we did a GWAS for the daily biomass observations in each environment. In Figure 8 (bottom panels), QTL profiles coincide with the within-environment changes in biomass ranking during the growing season shown in Figure 8 (upper panels) and Figure 7. Although most of

the QTLs shown in Figure 8 coincide across environments, their effects change during the growing season. In environments with lower autocorrelation, like Emerald_1993, QTL effect sizes at the beginning of the growing season are different from the QTL effect sizes at the end of the growing season. Changes in the size of QTL additive effects for biomass can also be interpreted as different intermediate traits and parameters being relevant for different combinations of phenological stages and environmental conditions. In Yanco_2010 (non-stress environment with a large biomass autocorrelation), the QTL effects were more stable over time. An intermediate situation was observed for Narrabri_2008, where the QTL effects were large at the beginning of the season, and got slightly less important towards maturity.

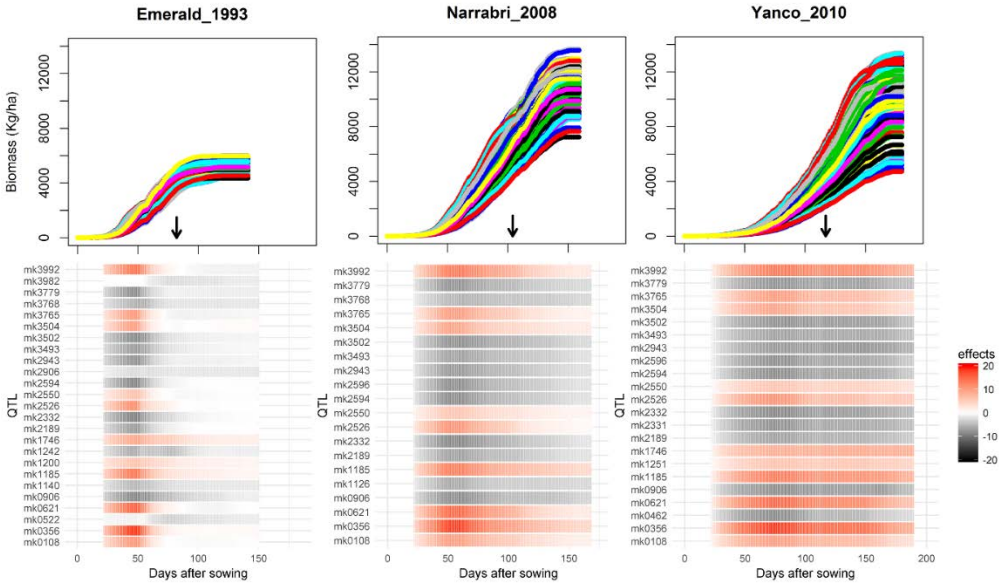


Figure 8. In the upper panels; APSIM output for the biomass of individual genotypes during the growing season in Emerald_1993 (very dry environment, with drought starting early in the growing season), Narrabri_2008 (intermediate post-flowering drought) and Yanco_2010 (no drought, upper panel). In the lower panel, QTL additive effects for biomass estimated from the GWAS for the APSIM output for the biomass during the growing season. QTL additive effects are expressed as a percentage of the population mean at a given environment and day. Arrows show mean flowering date.

We compared the effects of biomass QTLs with the effects of the same QTLs on APSIM parameters. The co-localization of QTLs for biomass and APSIM parameters allows to identify which mechanisms (basic traits/APSIM parameters) underlie biomass genotypic differences over time. The biomass variance explained by the QTL was up to 11% in each of the three environments that we analysed and the largest effects were up to 15% of the

population mean (Figure 8, Yanco_2010). The same QTLs had a similar importance on some of the individual parameters, explaining up to 9% of the variance, Table S2). Most of the QTLs that were relevant for biomass in all environments also regulated radiation use efficiency (y_rue), canopy extinction coefficient (y_extinct_coef) and biomass partitioning to leaf (y_frac_leaf). These parameters represent mechanisms/traits that are known to increase biomass accumulation across a wide range of environmental conditions.

Besides the QTLs that were in common across the three environments, we also identified some environment-specific QTLs. Those QTLs that were specific for water limited environments (Emerald_1993 and Narrabri_2008), were always related to either the lower limit for water uptake (ll_modifier, e.g. mk3765, Table S2) or to the transpiration efficiency coefficient (transp_eff_cf, e.g. mk3982, Table S2). QTLs for the lower limit for water uptake (ll_modifier) co-localized with QTLs that were most relevant during the whole growing season in Emerald_1993 and Narrabri_2008 (mk1140 and mk3768). These QTLs were absent in Yanco_2010 (no water stress). QTLs co-localizing with the transpiration efficiency coefficient were most relevant for biomass accumulation after flowering and were specific for Emerald_1993 (strong drought starting early in the growing season).

7.3.5. Summarizing phenotype dynamics as measured by HTP

We simulated HTP measurements for biomass by adding a plot and a measurement error to the biomass APSIM output. We evaluated different measurement error size in a factorial combination with phenotyping interval (expressed as the number of days between two consecutive measurements). The simulated HTP data were fitted with a logistic curve or with P-splines to evaluate strategies for a better characterization of biomass accumulation over time from HTP data. In this section, we describe the H^2 for parameters of the logistic curve and for parameters defined on the basis of the fitted P-spline function. H^2 for biomass related parameters is an indicator for the potential of that parameter to predict yield

The H^2 for the parameters of the logistic curve fitted to HTP biomass data over time was larger in Emerald_1993 and Narrabri_2008, where biomass curves were more symmetric. In Yanco_2010, biomass accumulation over time was most asymmetric (Figure 9) and H^2 for the parameters of the logistic curve was therefore low in this environment. In the three environments, H^2 increased with more frequent (smaller interval between two consecutive measurements) and with more precise measurements at individual time points (R^2 between the direct phenotypic measurements, Equation 11, and HTP, Equation 12). However, H^2 increased more in Emerald_1993 and Narrabri_2008 than in Yanco_2010 (Figure 9). When comparing the H^2 of the three parameters of the logistic curve fitted to biomass accumulation over time, we observed that the asymptote and the inflexion point showed a more flat H^2

surface, indicating that good estimates of these parameters can still be obtained after reducing measurement frequency and precision. For example, in Emerald_1993, the same H^2 estimate for the asymptote can be obtained (H^2~0.6) from a HTP technology that delivers a precision of 0.5 or with one that has a precision of 0.8. The same applies for measurement intervals; if multiple time points are measured simultaneously, the same H^2 can be obtained measuring every 5 or every 15 days. This highlights the convenience of integrating measurements over time, compared to using single time points independently. In contrast, maximum growth rate for biomass was more difficult to estimate, requiring very frequent and precise measurements to obtain a large H^2. This is especially the case in Yanco_2010, where biomass accumulation over time was asymmetric, the maximum growth rate had a very low H^2.

We also used splines to integrate HTP measurements for biomass over time. In this case, similar H^2 was obtained for curve characteristics across environments (Figure 10), showing that P-splines are more flexible tools than the logistic curve, allowing to better accommodate asymmetries of the biomass accumulation curve. The H^2 achieved for the spline fitted values was also larger than the H^2 of the logistic curve and the H^2 surface was more homogeneous (Figures 9 and 10). The more homogeneous H^2 surface indicates that splines are better than the logistic curve when it comes to remove part of the measurement error by integrating information throughout the season. In practice, this means that, when using a spline to integrate the HTP measurements for biomass, measurements can be done at a lower frequency (larger intervals) and lower precision (lower R^2 between HTP and APSIM biomass) to still obtain large H^2. We characterized the splines as fitted to the HTP measurements for biomass by the following parameters; biomass fit and at flowering, maximum biomass accumulation rate and the inflexion point of biomass accumulation. The largest H^2 was obtained for biomass fit at maturity and flowering (H^2~0.9). The H^2 of maximum rate was slightly lower (H^2~0.6-0.7) and the lowest H^2 was observed for the inflection point (H^2~0.1-0.2).

7.3.6. Multi-trait predictions considering the whole growing season

We used the parameters of logistic curves or spline fitted values for HTP measurements of biomass accumulation during the growing season as correlated traits for yield genomic prediction. In general, multi-trait genomic prediction models showed a larger accuracy than single-trait models (Figure 11). However, prediction accuracy of multi-trait models was highly dependent on the quality (heritability) of the correlated trait; more frequent and more precise HTP measurements showed larger accuracy, compared to less frequent and less precise measurements.

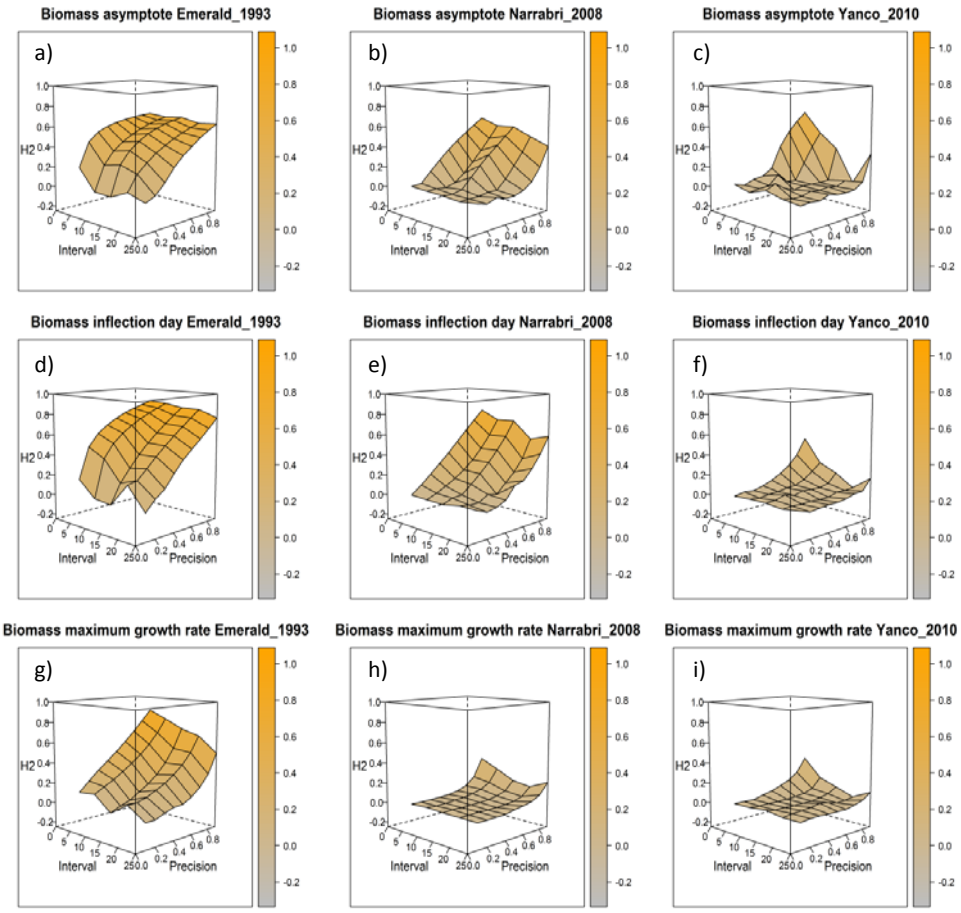


Figure 9. Heritability of the parameters from a logistic curve fitted to the dynamics of biomass accumulation for the collection of genotypes, measured with HTP. The following parameters were calculated; asymptote (a, b, c), biomass inflection day (d, e, f) and maximum biomass accumulation rate (g, h, i). The x-axis indicates the interval, expressed as the number of days between two consecutive HTP measurements. The z-axis (precision) indicates the quality of the HTP measurement, quantified as the R^2 between the direct phenotypic measurements (APSIM biomass plus plot error, Equation 11) and HTP (APSIM biomass plus plot and measurement error, Equation 12). The heritability that is shown is the average across 50 training sets.

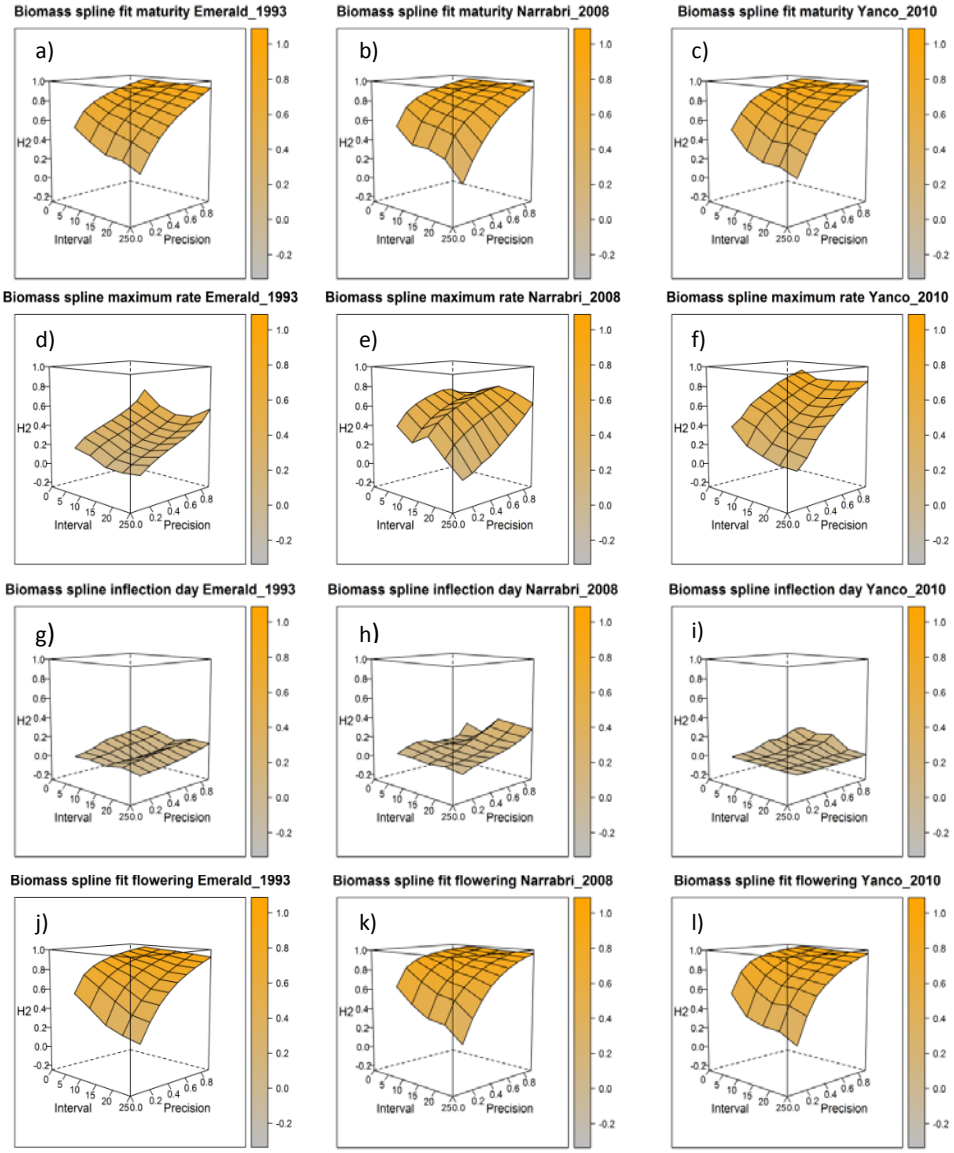


Figure 10. Heritability of splines, fitted to the dynamics of biomass accumulation for the collection of genotypes, measured with HTP. The following values were calculated; fitted values at flowering and maturity (a, b, c), maximum biomass accumulation rate (d, e, f), biomass inflection day (g, h, i) and day at which the maximum biomass accumulation rate is produced (j, k, l).The x-axis indicates the interval, expressed as the number of days between two consecutive HTP measurements. The z-axis (precision) indicates the quality of the HTP measurement, quantified as the R^2 between the direct phenotypic measurements (APSIM biomass plus plot error, Equation 11) and HTP (APSIM biomass plus plot and measurement error, Equation 12). The heritability that is shown is the average across 50 training sets.

Homogeneous measurement error over time

We assessed genomic prediction models that included the parameters of logistic curves fitted to HTP biomass measurements. In the best genomic prediction model, accuracy increased from 0.35 to 0.51 in Emerald_1993, from 0.69 to 0.73 in Narrabri_2008 and from 0.49 to 0.65 in Yanco_2010 (Table 4). However, not all parameters from the logistic curve fitted to the biomass HTP data over time were equally useful to increase genomic prediction. The best predictions in Emerald_1993 and in Narrabri_2008 were obtained with a model that included the asymptote and the inflexion point for biomass as correlated traits (Table 4). Similar prediction accuracy was obtained for the model including only the asymptote as correlated trait, indicating that the inflexion point did not add much additional information to improve yield predictions. A different situation was observed for Yanco_2010, where adding the inflexion point as a correlated trait (in addition to the asymptote) had a lower prediction accuracy than the two-trait prediction using the asymptote only. The maximum biomass accumulation rate did not show to be a useful logistic parameter to increase the prediction accuracy of grain yield in any of the environments, in agreement with the low H^2 observed for the maximum growth rate (Figure 9).

Table 4. Yield prediction accuracy in Emerald_1993, Narrabri_2008 and Yanco_2010 calculated with a prediction model considering a single trait (yield), two traits (yield and one of the **parameters of the logistic curve** fit to HTP biomass over time) and multiple traits (yield and more than one of the parameters fit to HTP biomass over time). Trait combinations showing the best prediction accuracy for each environment are indicated in bold and are also shown in Figure 11, upper panel.

Environment	Single trait		2 traits						Multi-trait			
	Yield		Yield + asymptote		Yield + inflexion day		Yield + max rate		Yield + asymptote + inflexion day + max rate		Yield + asymptote + inflexion day	
	Mean	se	Mean	se	Mean	se	Mean	se	Mean	se	Mean	se
Emerald_1993	0.35	0.02	0.50	0.01	0.44	0.01	0.26	0.01	0.42	0.01	**0.51**	**0.01**
Narrabri_2008	0.69	0.02	0.72	0.01	0.67	0.02	0.49	0.03	0.66	0.01	**0.73**	**0.01**
Yanco_2010	0.49	0.02	**0.65**	**0.01**	0.22	0.02	0.05	0.02	0.26	0.02	0.39	0.01

We also used parameters defining the spline fitted functions of HTP biomass over time as correlated traits in a multi-trait genomic prediction model; fitted values at maturity, fitted values at flowering, maximum biomass accumulation rate and inflexion point. The largest yield prediction accuracy in Emerald_1993 was observed for a multi-trait genomic prediction model using biomass fit at flowering, biomass fit at maturity, inflexion point and the maximum biomass accumulation rate (0.54 vs. 0.35, Table 5). In Narrabri_2008, the best prediction accuracy was observed for a two-trait model with the maximum biomass

accumulation rate (0.76 vs. 0.69). In Yanco_2010, the largest prediction accuracy was observed with a model using biomass fit at maturity and flowering and the maximum biomass accumulation rate (0.73 vs. 0.49). For all environments, the best prediction accuracy using spline information as correlated trait was larger than the best prediction accuracy obtained using the parameters of a logistic curve fitted to biomass HTP measurements over time, indicating the advantage of a more flexible model to integrate biomass HTP over time.

Table 5. Yield prediction accuracy in Emerald_1993, Narrabri_2008 and Yanco_2010 calculated with a prediction model considering a single trait (yield), two traits (yield and biomass over time fitted with a spline) and multiple traits (yield and biomass accumulation fitted with a **spline**). Trait combinations showing the best prediction accuracy for each environment are indicated in bold and are also shown in Figure 11, lower panel.

Environment	Single trait		2 traits								Multi-trait					
	Yield		Yield + fit flowering		Yield + fit maturity		Yield + inflexion day		Yield + Max rate		Yield + fit flowering + fit maturity		yield + fit flowering + fit maturity + inflexion day + max rate		yield + fit flowering + fit maturity + max rate	
	Mean	se	Mean	se	Mean	se	Mean	se	Mean	se	Mean	se	Mean	se	Mean	se
Emerald_1993	0.35	0.02	0.49	0.01	0.48	0.01	0.51	0.01	0.49	0.01	0.49	0.01	**0.54**	**0.01**	0.46	0.01
Narrabri_2008	0.69	0.02	0.64	0.01	0.70	0.01	0.63	0.01	**0.76**	**0.02**	0.71	0.01	0.70	0.01	0.69	0.01
Yanco_2010	0.49	0.02	0.62	0.01	0.66	0.01	0.10	0.01	0.64	0.02	0.67	0.01	0.70	0.01	**0.73**	**0.01**

When assessing the effect of measurement error size on prediction accuracy, we observed that larger prediction accuracy was observed for biomass HTP measurements with a smaller measurement error (Figure 11). However, prediction accuracy showed to be more stable across measurement error sizes when correlated traits for genomic prediction are calculated with a spline, instead of with a logistic curve. When using parameters of a logistic curve fitted to HTP biomass data, the multi-trait models in Emerald_1993 were better than the single trait, when the measurement precision was larger than 0.2. In the other environments (Narrabri_2008 and Yanco_2010), multi-trait genomic prediction models had larger accuracy than single-trait when measurements had a precision of at least 0.6. In contrast, when using the spline fitted values as a correlated trait multi-trait genomic prediction models had larger accuracy than single trait, even if precision was as low as 0.2 (Figure 11). This indicates that using splines to integrate biomass HTP measurements over time is better than using a logistic curve, especially when biomass accumulation curves are not symmetric, like in Yanco_2010. Measurement interval also had a large effect on prediction accuracy, with larger accuracy observed for more frequent measurements (Figure 11). The minimum measurement frequency required for a multi-trait model to show a larger accuracy than a single trait model depended on the environment. In Emerald_1993, where biomass showed symmetric biomass accumulation curve, all measurement frequencies had a similar prediction accuracy.

However, in Narrabri_2009 and Yanco_2010, measuring every 25 days was clearly worse than schedules with a higher measurement frequency (Figure 11). This indicates that when the trait dynamics are more irregular, more frequent measurements are required.

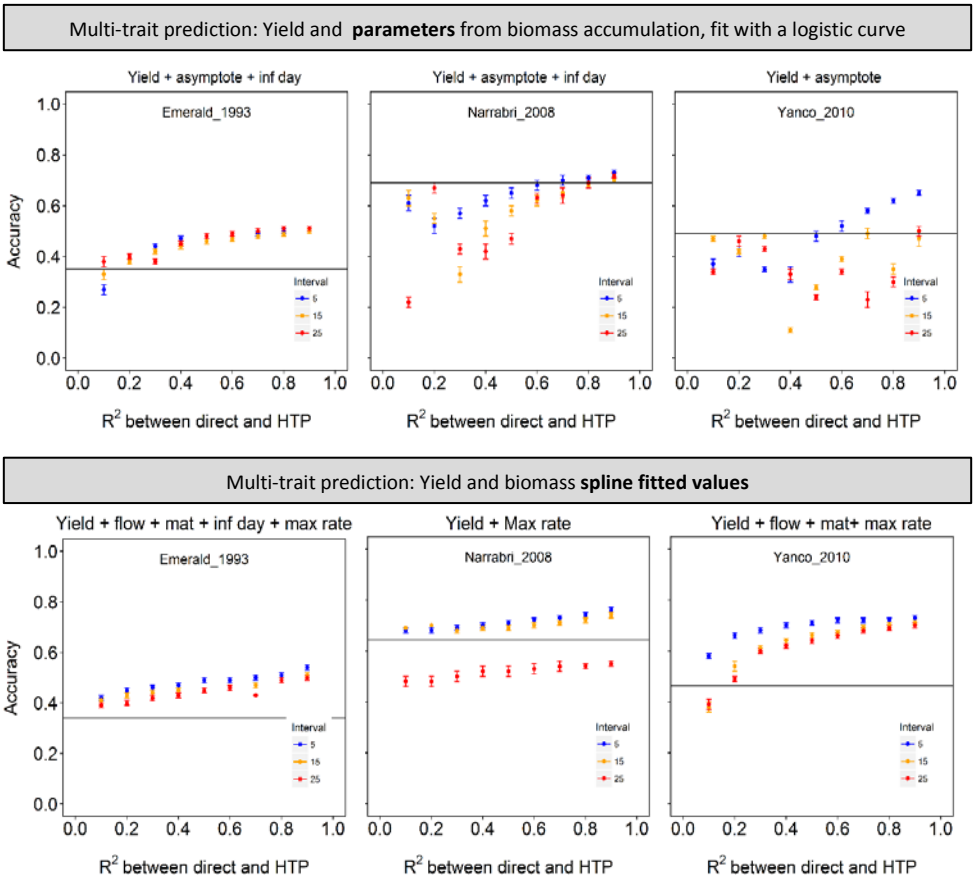


Figure 11. Yield prediction accuracy. Upper panels show multi-trait genomic prediction models that use phenotypic information for yield and for the **parameters from a logistic curve fitted to simulated biomass HTP data**. Lower panels show multi-trait genomic prediction models that combine phenotypic information for yield and for **spline fitted values of simulated biomass HTP data**. The x-axis indicates the quality of the HTP measurement, quantified as the R^2 between the direct phenotypic measurements (APSIM biomass plus plot error, Equation 11) and HTP (APSIM biomass plus plot and measurement error, Equation 12). Symbol colour indicates the interval, expressed as the number of days between two consecutive HTP measurements. Horizontal line indicates prediction accuracy for single-trait genomic prediction model (yield).

Measurement error as a function of canopy growth

In previous section, we presented accuracy for multi-trait genomic prediction models that used traits that were calculated from data that had a homogeneous measurement error over time. Here, we present the results of models that use biomass measured with an error depending on crop dynamics like canopy closure, i.e. the error gets larger as canopy closes and it gets reduced again with the onset of senescence. Prediction accuracy using biomass with heterogeneous error size was not very different from prediction accuracy using biomass with constant error size (Tables 5 and 7). The similar accuracy obtained when using traits with homogeneous vs. heterogeneous error shows that splines are able to remove the transient large measurement error after canopy closure, reinforcing the advantages of the integration of information over time.

7.3.7. Predictions from biomass measured early in the growing season

The last genomic prediction scenario that we assessed considered yield prediction from biomass measured earlier in the growing season. The quality of these early predictions was heavily dependent on the correlation of early biomass and yield. This correlation was very low (even negative) in Emerald_1993 (very dry environment, with drought starting early in the growing season), intermediate in Narrabri_2009 (intermediate post-flowering drought) and large in Yanco_2010 (no water stress, Figure 7a). In Emerald_1993, prediction accuracy was very low, in agreement with the negative correlation between yield and biomass during pre-flowering (Table 6). The largest (negative) correlation was observed for biomass fit at half of the flowering period (day 40), coinciding with the day with the lowest correlation between biomass and yield (Figures 7b and 12). A slightly better prediction accuracy was observed for Narrabri_2008, where yield was predicted from spline fitted values for HTP biomass at flowering. In this environment, prediction accuracy was between 0.2-0.4 (Figure 12). Yanco_2010 was the only environment in which early yield predictions from the spline fit to biomass HTP measurements had a similar prediction accuracy as single trait yield predictions from yield phenotypic data (horizontal line in Figure 12). In the three cases, no clear trend was observed across frequencies and measurement errors, suggesting that, for early yield predictions from HTP biomass measurements, the most limiting factor is the trait correlation, instead of the H^2.

Table 6. Yield prediction accuracy in Emerald_1993, Narrabri_2008 and Yanco_2010 calculated with a prediction model considering the **spline fitted values of pre-flowering biomass over time** (no yield phenotypic information was used). These biomass fitted values were calculated at specific moments, either one time (single trait) or multiple times (multi-trait). Biomass fit showing the best prediction accuracy for each environment is indicated in bold and is also shown in Figure 12

	Single trait (biomass)								Multi-trait (biomass)	
Environment	**fit 1/4 flowering**		**fit 2/4 flowering**		**fit 3/4 flowering**		**fit at flowering**		**4 time points**	
	Mean	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**
Emerald_1993	-0.19	0.02	**-0.24**	**0.02**	-0.19	0.02	0.18	0.02	-0.18	0.02
Narrabri_2008	0.06	0.01	-0.02	0.01	0.06	0.01	**0.27**	**0.01**	-0.01	0.02
Yanco_2010	0.44	0.01	0.39	0.01	0.44	0.01	**0.49**	**0.01**	0.47	0.02

Table 7. Yield prediction accuracy in Emerald_1993, Narrabri_2008 and Yanco_2010 calculated with a prediction model considering **spline fitted values** fitted to HTP biomass data over time. HTP was assumed to have a **measurement error that is a function of canopy growth**.

	Single trait		2 traits								Multi-trait					
Environment	**Yield**		**Yield + fit flowering**		**Yield + fit maturity**		**Yield + inflexion day**		**Yield + max rate**		**Yield + fit flowering + fit maturity**		**yield + fit flowering + fit maturity + inflexion day + max rate**		**yield + fit flowering + fit maturity + max rate**	
	Mean	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**
Emerald_1993	0.35	0.02	0.37	0.01	0.39	0.01	**0.51**	**0.01**	0.49	0.01	0.39	0.01	0.47	0.01	0.30	0.02
Narrabri_2008	0.69	0.02	0.60	0.01	0.64	0.02	0.63	0.02	**0.76**	**0.01**	0.64	0.02	0.65	0.01	0.64	0.01
Yanco_2010	0.49	0.02	0.63	0.01	0.66	0.01	0.10	0.02	0.64	0.01	0.66	0.01	0.69	0.01	**0.72**	**0.01**

Table 8. Yield prediction accuracy in Emerald_1993, Narrabri_2008 and Yanco_2010 calculated with a prediction model considering the parameters of a **logistic curve** fitted to HTP biomass data over time. HTP was assumed to have a measurement error that is a function of canopy growth.

	Single trait		2 traits						Multi-trait			
Environment	**Yield**		**Yield + asymptote**		**Yield + inflexion day**		**Yield + Max rate**		**Yield + asymptote + inflexion day + Max rate**		**Yield + asymptote + inflexion day**	
	Mean	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**	**Mean**	**se**
Emerald_1993	0.35	0.02	0.44	0.01	**0.50**	**0.01**	0.28	0.01	0.41	0.01	0.50	0.01
Narrabri_2008	0.69	0.02	0.72	0.01	0.66	0.01	0.44	0.01	0.67	0.01	**0.73**	**0.01**
Yanco_2010	0.49	0.02	**0.64**	**0.01**	0.24	0.02	-0.02	0.02	0.24	0.02	0.51	0.01

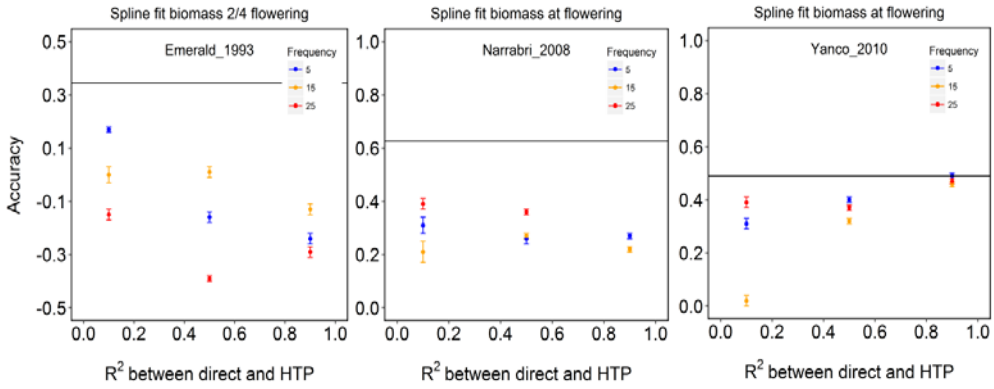


Figure 12. Prediction accuracy for yield predicted from pre-flowering biomass. Multi-trait genomic prediction models that uses phenotypic information for yield and for **spline fitted values of simulated biomass HTP data**. In the left panel, yield was predicted from biomass measured during the whole pre-flowering period, but fit at the first half of the pre-flowering period (when biomass was most correlated to yield, Figure 7a). In the central and right panels, yield was predicted from biomass measured during the whole pre-flowering period, but fit at flowering. The x-axis indicates the quality of the HTP measurement, quantified as the R^2 between the direct phenotypic measurements (APSIM biomass plus plot error, Equation 11) and HTP (APSIM biomass plus plot and measurement error, Equation 12). Symbol colour indicates the interval, expressed as the number of days between two consecutive HTP measurements. Horizontal line indicates prediction accuracy for single-trait genomic prediction model.

7.4. Discussion

In this paper, we illustrate that the APSIM crop growth model with genotype-dependent parameters allows to evaluate the potential of intermediate traits measured with HTP to improve yield genomic prediction accuracy. We used an HTP approximation of biomass as a correlated trait in a multi-trait genomic prediction model. We compared phenotyping schedules across environments varying in their water deficit patterns and statistical models to fit development over time (i.e. splines and logistic model). In our illustration, we used a diversity panel that represents well the spectrum of genotypes that is adapted to Australian environments. We based ourselves on previous research done by (Chenu *et al.*, 2011, 2013; Zheng *et al.*, 2013; Casadebaig *et al.*, 2016) to select a number of environments that represent well the Australian TPE. Similar approaches to characterize the adaptation landscape and identify traits that are useful for selection in different environment types have been previously discussed by (Podlich & Cooper, 1998; Chapman *et al.*, 2002; Hammer *et al.*, 2002, 2005).

7.4.1. Multi-trait predictions

Multi-trait predictions for yield, using biomass as a correlated trait showed in general a larger prediction accuracy than single trait predictions. The exceptions were those cases in which the measurement error was very large or when the curve did not fit well to the biomass data (especially the logistic curves). To avoid the reduced prediction accuracy for yield when biomass contained a large error, an alternative approach would be to use selection indices, giving a larger weight to the target trait than to the intermediate traits (Hazel, 1943; Lande & Thompson, 1990; Dekkers, 2007).

We also evaluated yield predictions from biomass measured early during the growing season. The degree of success of these predictions was largely affected by the correlation between early biomass and yield. As these correlations were associated to the environmental conditions (low for dry environments and medium-sized for humid environments), the phenotyping and prediction strategy needs to consider the type of environments in which genotypes are evaluated. Our modelling approach allows to estimate a priori whether the traits are likely to be correlated with the target trait, allowing to assess the potential of intermediate traits for making predictions.

The third scenario that we evaluated was using biomass that had a measurement error dependent on canopy growth, as most commonly encountered in real experiments (Grieder *et al.*, 2015; Magney *et al.*, 2016). In that way, the error is dynamic over time (heteroscedastic). We used the spline fitted values or the parameters calculated from the logistic curve as correlated traits. These biomass approximations showed similar results, compared to the approximations obtained using a homoscedastic error.

7.4.2. Simultaneous modelling of traits measured with HTP during the growing season

We modelled biomass as measured with HTP during the growing season. We compared the use of logistic functions and splines to characterize biomass dynamics over time. Both, splines and the logistic curves, increased the heritability of biomass, showing that modelling multiple time points simultaneously is a good strategy to reduce the measurement error. Similar results have been observed when using splines to model canopy temperature and NDVI measurements in real wheat experiments (Sun *et al.*, 2017). We observed that splines show a larger heritability than the logistic curve because they can accommodate better the irregularities in the biomass accumulation.

The advantage of using simulated data, is that we can evaluate error sizes and measurement frequencies, allowing to provide recommendations for a phenotyping schedule. In our simulations, we covered a range of phenotyping scenarios, varying in their measurement precision and interval. Prediction accuracy for multi-trait models using biomass information during the whole growing season, or using early biomass measurements to predict final yield, was in line with results obtained for real phenotypic data (Sun *et al.*, 2017). Some of the levels we chose for measurement precision are probably too optimistic, given that biomass approximations with technologies like NDVI usually have R^2 of maximum 0.6 (Marti *et al.*, 2007; Grieder *et al.*, 2015; Magney *et al.*, 2016). Our results indicated that when integrating the information during the growing season, similar prediction accuracy is obtained when using HTP technologies that deliver an R^2 of 0.6 than of 0.8. This suggests that, in principle, if we use the currently available technologies, more can be gained from the integration of multiple observation during the growing season, than from reducing the error of single observations. The next step in terms of integration of HTP data into phenotype prediction might be combining the information from proximal sensing of field trials (e.g. NDVI measured from a drone or helicopter (Chapman *et al.*, 2014)) with remote sensing from the actual wheat production environments (e.g. satellite measurements of wheat paddocks (Perry *et al.*, 2014)).

In section 3.4, we characterize the change of additive effects for biomass QTLs during the growing season. We also quantify how those QTLs relate to each of the APSIM parameters, in an attempt to understand the underlying cause of biomass accumulation over time and across environments. The analysis of QTLs that co-localize certainly provides useful insight about the importance of the APSIM parameters during the growing season. The QTL co-localization analysis allowed us to identify traits that have shown to be related to biomass accumulation in general (Calderini *et al.*, 1997; Sadras & Lawson, 2011) and in water-limited environments (Rebetzke *et al.*, 2012). However, a more elegant and complete representation of the relationship between QTLs, APSIM parameters, intermediate and target traits can be achieved with a network representation (Neto *et al.*, 2010; Alimi, 2016). Best, if multiple layers of information (QTLs, APSIM parameters and final traits) are modelled in a multi-level network (Wang & van Eeuwijk, 2014; Wang *et al.*, 2015).

7.4.3. Characterization of the TPG and the TPE

The GxE patterns and partitioning of the phenotypic variance observed for this simulated combination of TPG and TPE samples, are comparable to the ones reported for real field trials (Cullis *et al.*, 2000; Chenu *et al.*, 2011). The large GLY_{ijk} and the clear impact of water deficit patterns on GxE supports the convenience of focusing on the analysis of environment

types (water-deficit patterns) in place of environments defined by years and locations (Chapman *et al.*, 2002; Chenu *et al.*, 2011, 2013; Hammer *et al.*, 2014). The similar GxE patterns between simulated and real wheat data indicate that the APSIM simulations with genotype-dependent parameters are a useful tool to characterize phenotyping and breeding strategies across the Australian TPE.

We selected a small sample of environments that represent the different environment types commonly encountered across the most important wheat growing areas in the Australian wheat belt. This sample was consisted of three locations and different years. However, in practice, breeders are restricted to decide how to best cover the whole range of environmental conditions that pertain to the TPE, sampling locations within the present year. The adequate selection of locations for phenotyping is especially relevant when investing in facilities that cannot be easily transported from one location to another (e.g. LemnaTec Field Scanalyzer). As the Australian year to year variation (GLY_{ijk} and GY_{ij}) is large, there is some risk that, for some years, the locations selected for the field trials might not represent well the whole range of environmental conditions. In our simulated example, this would occur for example for Narrabri. This location does not show a consistent contribution to GxE across years; in some years, it shows no/small drought like Yanco_2010 and in others, it shows a strong post-anthesis drought. The use of APSIM with genotype-dependent parameters allows to assess the potential of locations to represent well the range of the TPE across years, helping to define the phenotyping network.

7.4.4. Combination of statistical and APSIM models

We based our selection of parameter ranges on information from the literature. The selection of parameter ranges is one of the most critical steps, having a large impact on the GxE patterns observed in the APSIM output, where also the TPG is important. The sample of the TPG defines a set of traits and trait ranges that are adaptive across the TPE. Flowering time is one of the most important examples of trait ranges that are specific for the TPE; in some environments, only some flowering time values allow a successful completion of the growing cycle. Examples and discussions on how the relationship between TPG and TPE relates to trait ranges can be found in (Slafer, 2003; Slafer *et al.*, 2005; Zheng *et al.*, 2012, 2013). In our simulations, we also assessed whether the flowering time ranges from the simulations were comparable to those observed in the real phenotypic data and adjusted the parameters accordingly. For example, a slightly larger range of vern_sens, photop_sens and tt_floral_initiation was explored in Zheng *et al.* (2013). We restricted ourselves to a smaller range for photop_sens (we used 1.5 to 3.0 instead of 1.5 to 4.0 as in Zheng *et al.* (2012) and for tt_floral_initiation (we used 455 to 555, instead of 455 to 1025 as in Zheng *et al.* (2013)

because otherwise we would obtain too late flowering dates, leading to crop failure in dry environments. Thus, the sample of the TPG that we constructed had a smaller flowering time range than the range explored by Zheng *et al.*, (2012), matching the phenology that is usually observed in well-adapted genotypes to dry Australian environments.

Combining crop growth models with statistical-genetic models also showed how biological epistasis can arise when scaling up from basic traits (APSIM parameters) to traits that show a larger level of integration of biological processes, like the intermediate or the target traits. Biological epistasis is said to take place when the phenotypic differences among individuals are influenced by other traits via physiological mechanisms (Cheverud & Routman, 1995; Cooper, 2004). If the relationship between the underlying component traits is non-additive, epistatic effects can occur at the phenotypic level of complex traits even if the gene action is purely additive (Holland, 2001; Cooper *et al.*, 2002; Hammer *et al.*, 2006; Technow *et al.*, 2015). For example, all APSIM parameters regulating canopy growth were additive. However, the intermediate trait 'canopy cover' showed a nonlinear response, leading to changes in genotypic ranking during the season (Figure 2). Similar scenarios, in which nonlinear responses arise from crop growth model parameters produced with purely linear statistical models can be seen in (Technow *et al.*, 2015).

Acknowledgements

We thank Bangyou Zheng for his valuable advice regarding the APSIM simulations. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°FP7-613556, Whealbi.

Supplementary Material

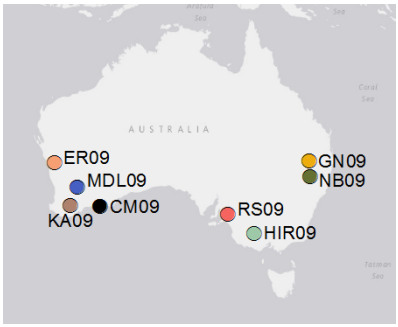


Figure S1. Geographical position of the eight Australian environments with phenotypic data for yield and heading date used to estimate the distribution of the additive effects.

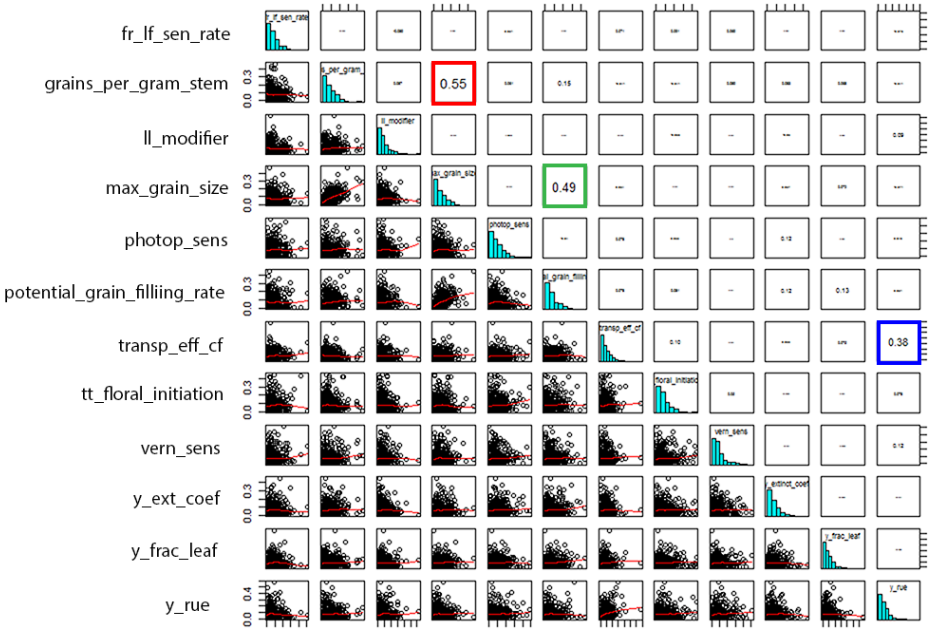


Figure S2. Histograms and observed correlations for the 300 additive effects underlying the 12 APSIM parameters. These effect were sampled with copulas to impose the following correlations on the additive effects of parameters indicated with a coloured frame: a) transpiration efficiency coefficient and radiation use efficiency (r= 0.40, blue), b) number of grains per gram of stem at flowering and maximum grain size (r=0.50, red) and c) maximum grain size and potential grain filling rate (r=0.45, green). All other correlations are lower than 0.20. The shape and rate for each parameter are indicated in Table 2.

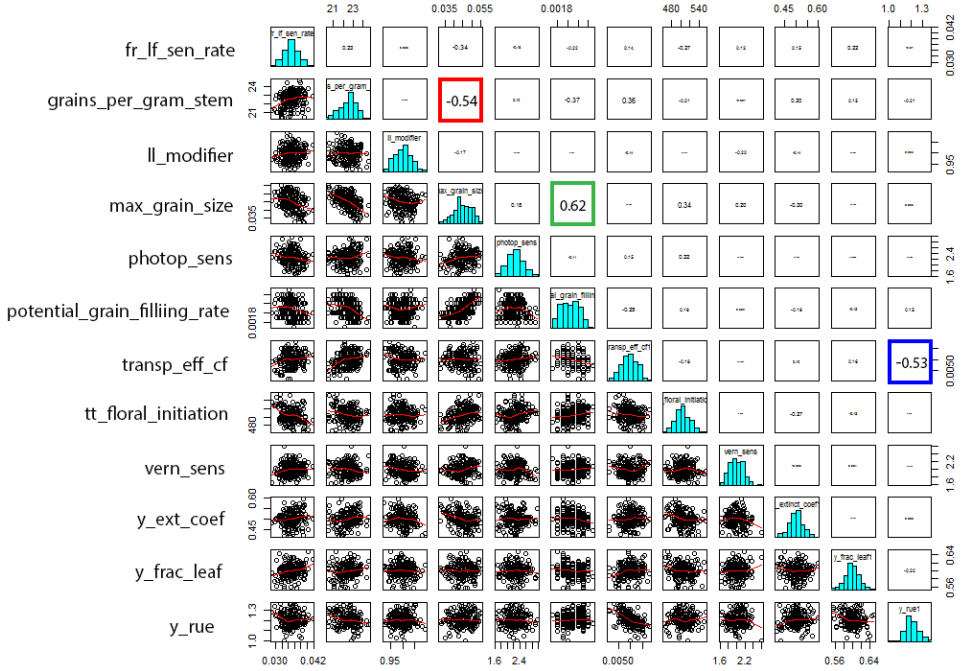


Figure S3. Histograms and correlations for the genotype-dependent APSIM parameters. These parameters were generated from 300 loci with additive effects. We imposed the following correlations on some pairs of parameters: a) transpiration efficiency coefficient and radiation use efficiency (r= -0.40, blue), b) number of grains per gram of stem at flowering and maximum grain size (r=-0.50, red) and c) maximum grain size and potential grain filling rate (r=+0.45, green). Differences between the imposed and realized correlations are product of the sampling process.

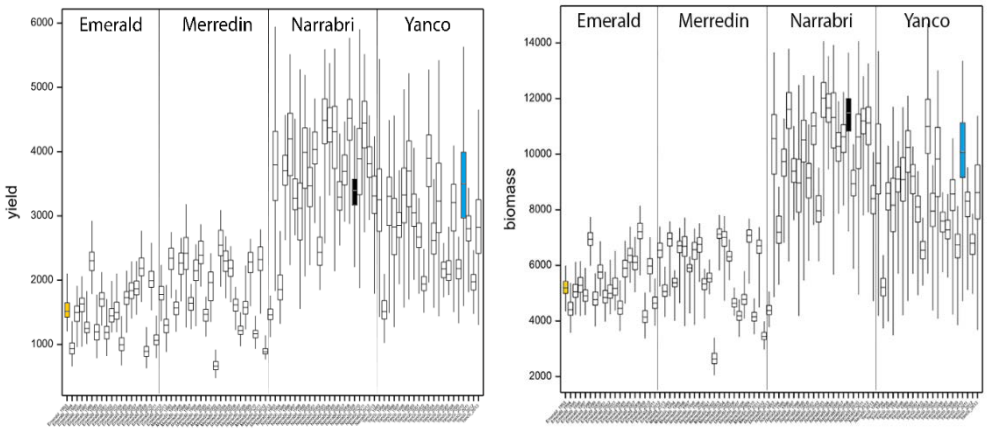


Figure S4. Yield and biomass for 84 environments (Emerald, Merredin, Narrabri and Yanco during 1993-2013). Environments selected to evaluate the integration of biomass as an intermediate trait to improve the quality of yield are shown in coloured boxes (orange=Emerald_1993, black=Narrabri_2008 and blue=Yanco_2010).

Table S1. Summary of the environmental conditions in the selected environments over three phenological periods, following the phenology of a genotype with average parameter values.

Period	Environment	MinT (ºC)	MaxT (ºC)	Radn (MJ m^{-2})	Rain (mm)	vp (hPa)	Day length (h)
Sowing-stem elongation	Emerald_1993	11.93	24.53	14.11	25.3	12.7	11.6
	Narrabri_2008	6.8	20.1	12.0	48.4	10.3	11.2
	Yanco_2010	5.2	15.1	9.3	95.7	9.8	10.9
Stem elongation-flowering	Emerald_1993	11.38	24.73	15.64	12.9	11.6	11.8
	Narrabri_2008	3.1	18.1	14.9	62.4	8.4	11.8
	Yanco_2010	5.7	14.9	12.1	103.3	10.5	12.0
Flowering-maturity	Emerald_1993	12.9	26.8	19.3	132.7	13.3	12.6
	Narrabri_2008	10.8	24.6	20.0	101.7	12.1	13.1
	Yanco_2010	8.4	21.3	21.2	130.7	12.5	13.8

Table S2. Percentage of the APSIM parameter variance explained by the SNPs that were significant for biomass during the growing season.

SNP	fr_lf_sen_rate	grains_per_gram_stem	ll_modifier	max_grain_size	photop_sens	potential_grain_filling_rate	transp_eff_cf	tt_floral_initiation	vern_sens	y_extinct_coef	y_frac_leaf	y_rue
mk3992						1.12						1.36
mk3982							1.13					
mk3779		1.10										3.28
mk3768			3.76									
mk3765			4.43									
mk3504										2.47		5.43
mk3502												
mk3493												2.77
mk2943												
mk2906												
mk2596										3.33		1.52
mk2594					1.58							
mk2550		4.99										2.70
mk2526			1.02	2.84								1.04
mk2332	1.38									1.73		1.74
mk2331		2.34										
mk2189							2.34					
mk1746				1.01						5.16		
mk1251					1.23							5.27
mk1242												
mk1200										4.85	1.06	
mk1185			1.10			2.00				6.43		
mk1140			3.82		2.43					3.69		
mk1126			2.03		1.46	1.27			1.43	6.73		
mk0906		3.63		4.47			2.99	1.58	2.64	8.82		
mk0621										4.07		
mk0522												
mk0462			1.25							2.11	1.23	2.22
mk0356										7.54		2.54
mk0108										2.96		1.35

Chapter 8

What should students in plant breeding know about the statistical aspects of GxE?

Fred van Eeuwijk[1], **Daniela Bustos-Korts**[1,2], Marcos Malosetti[1]

1. Biometris, Wageningen University & Research Centre

2. C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC)

This Chapter is published as:

van Eeuwijk, F.A., D. V Bustos-Korts, and M. Malosetti. 2016. What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? Crop Sci. 56: 2119–2140.

Abstract

A good statistical analysis of genotype by environment interactions (GxE) is a key requirement for progress in any breeding program. Data for GxE analyses traditionally come from multi-environment trials. In recent years, increasingly data are generated from managed stress trials, phenotyping platforms, and high throughput phenotyping techniques in the field. Simultaneously, and complementary to the phenotyping, more elaborate genotyping and envirotyping occur. All of these developments further increase the importance of a sound statistical framework for analyzing GxE. This paper presents considerations on such a framework from the point of view of the choices that need to be made with respect to the content of short academic courses on statistical methods for GxE. Based on our experiences in teaching statistical methods to plant breeders, for specialized GxE courses between three and five days are reserved. The audience in such courses includes MSc students, PhD students, postdocs and researchers at breeding companies. For such specialized courses, we propose a collection of topics to be covered. Our outlook on GxE analyses is two-fold. On the one hand, we see the GxE problem as the building of predictive models for genotype-specific reaction norms. On the other hand, the GxE problem consists in the identification of suitable variance-covariance models to describe heterogeneity of genetic variance and correlations across environments. Our preferred class of statistical models is the class of mixed linear-bilinear models. These statistical models allow us to answer breeding questions on adaptation, adaptability, stability, and the identification and subdivision of the target population of environments. By a citation analysis of the literature on GxE, we show that our preference for mixed linear-bilinear models for analyzing GxE is supported by recent trends in the types of methods for GxE analysis that are most frequently cited.

Keywords: adaptation, adaptability, bilinear model, environmental covariable, factorial regression, genotype by environment interaction, genotypic covariable, genotypic sensitivity, heterogeneity of genetic correlation, heterogeneity of genetic variance, mega-environments, mixed models, phenotypic prediction, plasticity, reaction norm, stability, target population of environments, variance-covariance structure

8.1.1. Setting up a course on GxE

Motivation

About 15 years ago, the prediction of phenotypes from high-density marker information was recognised as a potential game changer in animal breeding (Meuwissen et al., 2001). In plant breeding, the current development of high throughput genotyping techniques alongside with similar techniques for phenotyping and envirotyping (Cobb et al., 2013; Araus and Cairns, 2014; Cooper et al., 2014b) provides opportunities for large scale phenotypic predictions of new, and therefore, untested genotypes in new environments under a wide spectrum of genotype by environment interaction (GxE) scenarios (Burgueño et al., 2008; Dawson et al., 2013; Heslot et al., 2013; Jarquín et al., 2013; Bustos-Korts et al., 2016; Malosetti et al., 2016, this issue). The increased volumes of phenotypic, genotypic and environmental data give a stimulus to the development of new statistical approaches for more precise description and prediction of GxE phenomena. A better modelling of GxE will undeniably contribute to a higher efficiency of breeding programs. In the light of the current developments, it is obvious that the study of GxE will become even more important in the near future than it was already in the past. Future generations of students and researchers in plant breeding will require substantial training in the statistical aspects of GxE. In this paper, we present and discuss our views on the topics that need to be covered in a course on GxE as well as some ideas about how to teach such a course.

From our experience, for specialized courses about the statistical aspects of GxE, typically a period between three to five days is reserved. The population of students for such courses covers the spectrum from MSc students to researchers in plant breeding companies, with in-between PhD students and postdocs. For this paper, we base ourselves on the following experiences. We gave various courses as part of the Integrated Breeding Multi-Year Course (IB-MYC, 2012-2014, https://www.integratedbreeding.net). The students in these courses came from Africa and Asia and ranged from PhD students in plant biology to experienced plant breeders in charge of breeding programs. A course with MSc students from the Mediterranean area was part of the MSc Plant Breeding program at the Mediterranean Agronomic Institute of Zaragoza (http://masters.iamz.ciheam.org/en/plantbreeding/). Further courses were given to MSc and post-doctoral students in plant breeding at Wageningen University (http://www.wageningenur.nl/biometris) and to MSc students in applied statistics at Leiden University (http://en.mastersinleiden.nl/programmes/statistical-science-for-the-life-and-behavioural-sciences/en/introduction). Finally, various in-house courses were organized for employees of international plant breeding companies.

Prerequisites

The target audience for our courses consists of plant breeding students who should be familiar with the following areas and concepts:

- quantitative genetics (estimation of genetic and environmental variances, heritability, genetic correlation, response to selection, correlated response)

- statistical design of plant breeding trials (obligatory: completely randomized designs, randomized complete block designs, incomplete block designs; desired: resolvable designs, cyclic designs, row-and-column designs, p-rep designs, augmented designs, split plot designs)

- statistical analysis of plant breeding trials (one-way and two-way analysis of variance (ANOVA), development and testing of contrasts and interactions, regression, analysis of covariance, one- and two-way linear mixed models, inference for fixed and random terms)

Preferentially, students have followed a course on the design and analysis of individual plant breeding trials before going to a GxE course. On the design side, students should be able to choose an appropriate design for plant breeding experiments given particular objectives (e.g., estimation of heritability versus comparison of a small number of genotypes) and limitations (number of replicates, rows, and columns, spatial variation, trends). On the analysis side, students should be able select an appropriate model for an experiment, making choices on which terms to take fixed and which to take random, and assessing the necessity to include individual model terms by F-tests / Wald tests for fixed terms and log-likelihood ratio / deviance tests for random terms.

Familiarity with quantitative trait locus (QTL) mapping is useful, because it allows GxE analyses to be connected directly to QTL by environment (QTLxE) analyses. For QTL mapping, it is best, when both linkage and linkage disequilibrium mapping have been discussed, for single and multiple traits and environments, preferentially within a mixed model framework. Simple examples of genome-enabled prediction methods can be included in a QTL mapping course as well. From our perspective, an important objective of QTL analysis is the identification of the genetic basis of GxE in the form of QTL by environment interaction (QTLxE).

To finish off the quantitative education of plant breeding students, a course in decision support is desired in which students learn how to integrate information from statistical

analyses and formulate breeding strategies. Specific topics in such a course are marker assisted back crossing, gene pyramiding, marker assisted selection, genomic selection, and various types of index selection. A GxE course will benefit from knowledge on decision support, although a decision support course is not an absolute prerequisite for a GxE course.

In our three-to-five day course on GxE, the emphasis is on the formulation and building of genotype-to-phenotype prediction models. From the identified genotype-to-phenotype models, we estimate genotype-specific statistics for mean performance under general and specific environmental conditions besides genotype-specific stability and risk measures. Selection strategies for genotypic means, sensitivities, stability variances and risks can be part of the GxE course as long as they concern phenotypic selection. We would prefer to discuss marker assisted selection strategies for such GxE related statistics in a separate decision support course.

8.2. GxE concepts and perspectives

Before we describe the statistical-technical details of our proposal for a statistical course on GxE for future plant breeders, we want to introduce a number of breeding concepts that are useful for a better communication between plant breeders and statisticians. These concepts help to define pertinent breeding questions in terms that allow unequivocal translation to statistical models and parameters. We acknowledge that the definitions of the breeding concepts below may look biased toward the direction of statistical clarity at the expense of biological width. We have on purpose narrowed down breeding definitions to guarantee one-to-one relations to statistical parameters.

Target population of genotypes and target population of environments

As a preliminary to models for GxE and breeding concepts related to GxE (adaptation, adaptability, stability, etc.), it is useful to introduce the concepts of target population of genotypes (TPG) and target population of environments (TPE). The TPG and TPE define the set of genotypes and environments for which we want our inference and predictions to be valid and precise. The TPG contains all possible genotypes we hope to develop and grow the coming years. Statistically speaking, we aim at coincidence of the TPG and the genetic design space of our prediction models. In the genomic selection literature, the target population of genotypes coincides with the notion of the set or population of selection candidates (Jannink et al., 2010; Schulz-Streeck et al. 2012; Albrecht et al. 2014). The TPE delineates the future growing conditions of the genotypes in the TPG (Comstock, 1977; Cooper and Hammer, 1996; Cooper et al., 2014b). The TPE can be defined by geography, soil and meteorological

conditions, management choices, and the incidence of biotic and abiotic stresses. We want the TPE to be reflected in the environmental design space of our prediction models.

As the phenotype is an integrated outcome of interactions between genetic and environmental factors during development, TPG and TPE cannot be chosen as to be independent. To give an example for abiotic stress, if we define a TPE on a geographical basis that includes drought and well-watered conditions, we wish to develop genotypes that perform well under both drought stress and well-watered conditions, or, in other words, the TPG consists of genotypes with wide adaptation. However, if we want to interpret drought stressed conditions as a TPE by itself, the composition of the TPG will have to change in reaction to the redefinition of the TPE. Therefore, the width of the TPE has consequences for the definition of the TPG and vice versa. For biotic stresses, the same arguments will hold when we replace the drought stress in the above example with infection pressure for a particular disease.

Reaction norm

For individual genotypes, we want to describe their phenotypic behavior across the full TPE and therefore we introduce the concept of the reaction norm: the genotype-specific functional relationship between phenotype and environmental gradient(s) (Woltereck, 1909; DeWitt and Scheiner, 2004). In practice, environmental gradients are sampled in a limited number of experiments. The observations made in those discrete environments are called character states (Schlichting and Pigliucci, 1998; Pigliucci, 2001).

For phenotypic prediction across a range of environmental conditions, we need to fit statistical models that represent the reaction norms of individual genotypes, that is, the main environmental drivers for phenotypic differences need to be identified together with suitable functional forms for the reaction norms. Phenotypic data can come from a series of field trials that represent a draw from the TPE. Such a draw from the TPE is often equivalent to or part of a multi-environment trial (MET) (Smith et al., 2001, 2005; van Eeuwijk et al., 2010). More informative data for modelling reaction norms can come from managed stress trials (Cooper et al., 2014a) and phenotyping platforms (Tardieu and Tuberosa, 2010; Cobb et al., 2013; Araus and Cairns, 2014; Kuijken et al., 2015).

Adaptedness and adaptation

Within the framework of reaction norms, a genotype shows adaptedness when its reaction norm is superior to that of a standard genotype or when it is close to that of an ideotype (van Oijen and Höglind, 2015). In the plant breeding literature, adaptedness, a state, is not always

distinguished from adaptation, a process. For example, it is common to talk about wide adaptation for genotypes showing adaptedness across the full TPE, versus narrow adaptation for genotypes showing adaptedness for part of the TPE (Ceccarelli, 1989, 1996; Braun et al., 1996; Cooper and Hammer, 1996; Cooper, 1999; Araus et al., 2008; Sadras and Rebetzke, 2013). Below, we will use both terms to some extent interchangeably, but our intention is to refer to a state, so adaptedness would be the more correct term to use.

Adaptation and adaptedness usually pertain to yield or biomass. Traits different from yield itself are instrumental for realizing adaptation in yield, a very important one being phenology or earliness. The development of a genotype should match with the timing of resource availability and the absence or low incidence of stresses in its environment. The reaction norm for yield depends on the reaction norms for the yield components. The joint reaction norm of yield and yield components is a multivariate function of phenotypes that mutually affect each other and genetic and environmental inputs. For interesting elaborations of this multi-trait idea of reaction norms for plant breeding purposes, see Podlich and Cooper (1999) Messina et al. (2011), Cooper et al. (2014a), Cooper et al. (2014b), Harrison et al. (2014). A good understanding of the processes leading to adaptedness and GxE requires observations on yield together with its main component traits as a function of (developmental) time. As multi-trait developmental data are still rare, GxE analysis methods for single traits observed at single time points dominate the literature.

Adaptability and sensitivity

A reaction norm defines a genotype-specific function that translates environmental inputs into a phenotype. GxE occurs when the reaction norms are not parallel, i.e., they intersect, diverge or converge, compare Figure 1a and 1b on the one hand with Figure 1c, 1d, 1e and 1f on the other hand. GxE forces phenotypic prediction models to become more elaborate and to contain genotype-specific parameters; intercepts, slopes and curvatures. These genotype-specific parameters are called sensitivity and adaptability parameters in the plant breeding literature and they facilitate the modelling of non-parallelism of reaction norms to account for GxE (Finlay and Wilkinson, 1963; Bänziger et al., 1997; Bradshaw, 2006; Sadras and Lawson, 2011; Slafer et al., 2014). Sensitivity applies to situations with single and well-identified explicit environmental gradients (drought stress index, temperature), adaptability to less concrete and non-explicit environmental gradients (environmental index based on average performance of all genotypes in a trial).

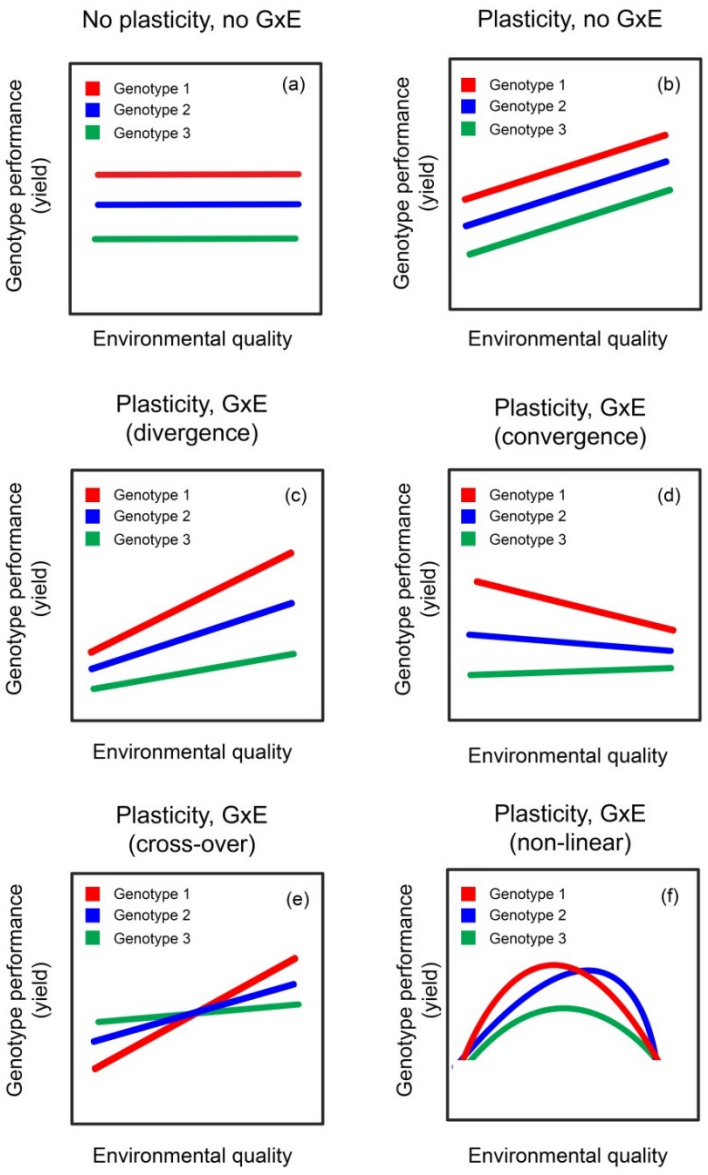


Figure 1. Reaction norms for three genotypes that illustrate various forms of plasticity and Genotype x Environment interaction (GxE). No plasticity in (a) versus plasticity in (b) to (f), no G ´ E in (a) and (b) versus various forms of GxE in (c) till (f).

Stability and risk

Observations on realized phenotypes in field trials will vary around the expected reaction norm for individual genotypes. This variation can be genotype-specific again and then it becomes another expression of GxE, which is captured by stability parameters (Eberhart and Russell, 1966; Wricke, 1966; Shukla, 1972; Fischer and Maurer, 1978; Lin et al., 1986; Lin and Binns, 1988; Piepho, 1998). Stability is called static (Lin et al., 1986), when the reaction norm model does not contain genotype-specific intercept or sensitivity terms and because of that does not correct for the general performance of the genotype or the production level of the environment. When stability is called dynamic, it is defined in terms of the variance of the residuals from more elaborate reaction norm models with genotype-specific parameters. In estimating procedures for dynamic stability variances, we need to account for the environmental factors shaping the reaction norms. For assessments of managed stress on a phenotyping platform, reaction norms are reduced to performance under stress versus control conditions and stability variances express variation within and between runs on the phenotyping platform. To increase the usefulness of stability measures, predictable or repeatable forms of GxE need to be distinguished from non-predictable and non-repeatable forms. For example, repeatable genotype by management interactions need to be distinguished from non-repeatable genotype by time and genotype by management by time interactions. For classical multi-environment trials across locations and years, a similar distinction between repeatable and non-repeatable types of GxE can be made (Chapman et al., 2000a; b). Lin and Binns (1988) proposed to consider genotype by location interactions as predictable and to fit a reaction norm model to such interaction, whereas genotype by location by year interactions were considered to be unpredictable and genotype-specific stability variances were defined on the basis of the latter three-way GxE interactions.

Stability parameters are estimated mainly for yield and yield related traits, and to a lesser extent for quality traits. The concept of stability overlaps with the concepts of homeostasis, (no) plasticity, and resilience, i.e. reverting to equilibrium when perturbed (Lerner, 1954; Hanson, 1970; Govindaraju and Dancik, 1987; Debat and David, 2001; Sadras et al., 2009; Nicotra et al., 2010).

When genotypes differ in both reaction norm and stability, it may be worthwhile to combine these two concepts in a risk concept: the probability to exceed a threshold yield level for part or the whole of the TPE (Eskridge and Mumm, 1992).

Environments

With respect to GxE, the major breeding questions for genotypes concern those related to adaptedness, adaptability and stability. For the environments, the major question is whether a set of trials, experiments, and conditions can be considered to form a sample from one particular target population of environments, or, alternatively, whether to attribute the set of trials, experiments and conditions to two or more target populations of environments. When trials (experiments) are grouped based on phenotypic or environmental information from the trials themselves, such groups are called mega-environments, or, sometimes, adaptation zones and ecological zones (Gauch Jr. and Zobel, 1988; Cornelius et al., 1992; Cooper et al., 1993; Gauch and Zobel, 1997; Atlin et al., 2000, 2011; Yan et al., 2000; Löffler et al., 2005; Chenu et al., 2011, 2013; Zhang et al., 2015). Mega-environments consist of trials and conditions that elicit comparable phenotypic responses in certain groups of genotypes. The question to the identification of mega-environments is the environmental counterpart of the question to genotypic adaptation. Several statistical approaches to GxE problems, like bilinear models (Gauch and Zobel, 1997) and bi-clustering (Corsten and Denis, 1990) approaches, simultaneously identify groups of genotypes with similar adaptedness patterns and groups of environments with similar conditions.

The identification of mega-environments is relevant for collections of trials that are assumed to originate from two or more discrete TPEs. However, even when assuming that trials stem from different TPEs, it may still be the case that genetic correlation exists between the genotypic performances in these trials. In the latter case, it seems better to speak of a subdivided TPE than of different TPEs. When the subdivision of the TPE is geographical, we often speak of regions. When trials are strongly heterogeneous, the central breeding question is whether to focus on specific adaptation or wide adaptation (Cooper and Woodruff, 1993; Braun et al., 1996; Ceccarelli et al., 1998; Atlin et al., 2000; Yan et al., 2000; Trethowan et al., 2001; Piepho and Möhring, 2005; Navabi et al., 2006; Chenu et al., 2011; Windhausen et al., 2012). If the decision is to go for wide adaptation, the most frequent question is how to efficiently combine the information from different regions of the TPE such as to maximize genetic gain across the undivided TPE (Atlin et al., 2000; Piepho and Möhring, 2005; Windhausen et al., 2012).

Another common question on the environmental side of GxE problems is that on the identification of trial locations that best represent the TPE (Heslot et al., 2013; Cooper et al., 2014). Representative locations should have a high correlation with the expectation across the TPE. Of course, this expectation is often unknown, because we are uncertain about which trials to assign to the TPE and which not. Another complicating factor is that genotypes may

be strongly unbalanced across trials. When there is certainty about whether trials are from the TPE, a possible estimator for the expectation across the TPE is the mean of a set of standard or probe genotypes across trials (Mathews et al., 2011). Trials that are chosen to represent the TPE should also have good discriminative properties (high heritability).

8.3. The statistical aspects of a GxE course for plant breeders

8.3.1. Linear, bilinear, and mixed models

In our GxE course of a few days to a week, the emphasis should be on data analysis, which means model formulation and model building, interpretation of statistical parameters in breeding terms and prediction of phenotypic responses to answer breeding questions, all as much as possible on real data. The learning objectives involve the expression of breeding questions in terms of suitable statistical models and parameters, the successful fitting of such models to breeding data, the interpretation of the results of the statistical analysis in breeding terms, and the reporting of the main findings in clear non-statistical language with insightful tables and figures. Statistical and theoretical issues related to parametrization, and estimation and testing procedures are of secondary importance. We believe that the main breeding questions above (adaptation, adaptability, stability, identification of TPE, choice of selection environments) can effectively be addressed by a combination of linear, bilinear, and mixed models. Essentially, the GxE course discusses models that contain linear and bilinear terms that can be fixed or random: linear-bilinear mixed models. The estimation procedures are typically least squares and REstricted Maximum Likelihood in software packages like SAS (SAS Institute, 2015), Genstat (VSN-International, 2015a) and Breeding View (The Breeding Management System Version 3.0.8, 2015), ASReml (VSN-International, 2015b), and ASReml-R (VSN-International, 2015c). Testing of fixed effects takes place by Wald tests or F-tests, while variance components and correlations are tested by likelihood ratio tests (Smith et al., 2005; van Eeuwijk et al., 2010; Gumedze and Dunne, 2011). For us, testing is predominantly an activity within a model building strategy to identify a prediction model.

Linear and linear-bilinear models with only fixed terms are of limited use in the analysis of GxE data. Mixed models have facilities for modelling heterogeneity of genetic variances and correlations between environments as well as for modelling design features and spatial trends in individual trials. Furthermore, prediction of phenotypic traits across environments and estimation of quantitative genetic parameters as genetic variance and correlation, heritability, and responses to direct, indirect and index selection are natural within the context of mixed models, whereas they become contrived in the context of models with only fixed terms.

Nevertheless, for didactical purposes, it is appropriate to start a course on GxE with fixed analysis of variance (ANOVA) models. Inferential procedures by standard F- and t–tests may be more familiar than Wald tests and likelihood ratio tests. ANOVA can also be used to repeat the basics of least squares estimation and testing theory. The ANOVA framework is convenient to introduce model building and different forms of interaction. Later, generalizations to bilinear models and mixed models are easily made.

One stage and two-stage GxE analysis with ANOVA

A good starting point for a GxE course is an ANOVA model for a response at plot level in a multi-environment trial (MET) with randomized complete blocks. The model has the form: $y_{ijk} = \mu + g_i + e_j + B_{k(j)} + ge_{ij} + \epsilon_{ijk}$, with y_{ijk} the phenotypic response in block k for genotype i in environment (trial) j, g_i the fixed genotypic main effect, e_j the fixed environmental main effect, $B_{k(j)}$ the fixed effect of block k in environment j, ge_{ij} the fixed GxE term, and ϵ_{ijk} a random normally distributed error. This model is fitted by least squares. The students will be using specialized software to fit statistical models, so they do not need to be bothered with details of the fitting process. Still, they need to be able to choose least squares for fitting fixed regression and ANOVA models and REML for mixed models and they should be familiar with the expressions for estimators of statistics that are used to answer breeding questions. Familiarity then means that students recognize how the data are used to develop the estimator and how the estimator is related to a statistical model for the data.

In the fixed ANOVA above, the genotypic main effect and the GxE term will be tested against the residual mean square, $var(\epsilon_{ijk}) = \sigma^2_\epsilon$. The standard errors for contrasts on the genotypic main effects and the GxE effect are functions of the residual mean square. It is a good exercise for students to formulate and fit the fixed ANOVA two-way model with randomized complete blocks for a concrete MET dataset and perform a statistical analysis, including diagnostic checks and reporting of results. The analysis can initially concentrate on the test for GxE and ways of quantifying the amount of GxE in comparison to the genotypic main effect, for example, by comparing the magnitude of the sums of squares for those terms. The use of contrasts to investigate the structure of GxE can wait until models for GxE have been discussed (see below). We recommend to follow the protocol described by Welham et al. (2014) for the preparation of data, ANOVA analysis, identification of a predictive model, interpretation of results, and reporting.

For an analysis of MET data in a two-stage approach, in the first stage perform randomized complete block analyses per trial. After checking diagnostics for normality, homogeneity of variance and outliers, vectors of genotypic means are formed together with vectors of weights for subsequent GxE analysis. These weights are functions of the standard

errors of means (Möhring and Piepho, 2009; Welham et al., 2010; Piepho et al., 2012). For the students, the two-stage approach offers an opportunity to rehearse the principles of ANOVA and mixed models on single trials. In the second stage, they fit a two-way ANOVA model without replication to the table of GxE means: $y_{ij} = \mu + g_i + e_j + ge_{ij} + \epsilon_{ij}$. Estimation will follow least squares principles again, as in all fixed ANOVA models. The error term, ϵ_{ij}, has a normal distribution, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, and is confounded with the two-way interaction. An independent error term can be obtained from the errors of the single trial analyses and will allow testing of the GxE term.

In an introductory GxE course, simple examples of one-stage analyses of GxE data can be presented. However, it is convenient to dedicate most of the time to two-stage modelling. The reason is that firstly, the statistical differences between one-stage and two-stage analyses are small in most cases and certainly when appropriate weighting schemes are used (Welham et al., 2010; Piepho et al., 2012), and secondly, the logistics of the two-stage approach are more intuitive. In the two-stage approach, the students first concentrate on the phenotypic analysis of single trials, taking into account design and spatial trends. Subsequently, the students can focus exclusively on the modelling of GxE. The simultaneous handling of large numbers of trials at plot level can be confusing, both at the statistical input and output level. Fewer things can go wrong in a two-stage approach, as compared with a one-stage approach.

Mixed models

The fixed ANOVA model for the MET with randomized complete blocks can be turned into a mixed model by treating blocks as random in the one-stage model. For this kind of design, this will hardly change interpretations on the comparisons of genotypic means, but students will have to realize that the mixed model imposes additional assumptions on the parameters in the model that will require checking. In this case, the extra assumption is that the block effects come from a normal distribution with a particular variance. Because block effects contribute to the variance of observations, the standard error for a genotypic mean will be larger in the model with random blocks. In contrast, for the standard error of a genotypic difference the block effects cancel out, so that this standard error is the same in the fixed and mixed model. Of course, for incomplete block designs, we expect the standard error of a difference to be smaller for random blocks than fixed blocks. These points can be brought to the attention of the students by asking them to perform analyses of MET data with fixed and random blocks and to compare the results. Another point that deserves attention is the inspection of ANOVA tables of degrees of freedom, sums of squares, mean squares and F-tests, versus the inspection of tables of Wald tests for fixed effects and likelihood ratio tests for variance components. As a learning objective, students should be able to interpret these

tables and use them to build prediction models for GxE data. We return to model building in a later section.

A more drastic change occurs when the genotypes are taken random, in both the one-stage and the two-stage model. All basic quantitative genetic concepts can be introduced and discussed: Best Linear Unbiased Predictor (BLUP), genetic variance, GxE variance, genetic correlation, heritability, response to selection, direct response to selection, correlated response to selection. We remark that we will use the word 'genetic' in a rather loose sense, mostly for indicating variance and correlation across genotypes, but also sometimes in a more restricted sense for exclusively additive genetic variance or correlation, as in mixed models with structured variance-covariance matrices.

The one-stage model is easier to work with for the presentation of estimators for all of these parameters. The students should be made familiar with expressions for quantitative genetic parameters. They should learn what the influence is of changing the number of environments and replicates on the heritability. A further objective is to learn how to evaluate the efficiency of selection in one particular environment (e.g. stress) to that in another (e.g. non-stress). For the two-stage modelling, the mixed model for a MET can be defined as $y_{ij} = \mu_j + G_{ij} + \epsilon_{ij}$, with μ_j the fixed intercept for environment *j*, and G_{ij} the random environment-specific genetic effect for genotype *i* in environment *j*. The matrix of random effects G_{ij} will have a multivariate normal matrix distribution with mean zero and variance-covariance matrix (VCOV) $\boldsymbol{\Sigma}$: $\{G_{ij}\} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$ (Cullis et al., 2005). The VCOV, $\boldsymbol{\Sigma}$, is factorized in a VCOV for the genotypes that defines the correlations between the genotypes following from kinship and pedigree, $\boldsymbol{\Sigma^G}$, and another VCOV for the environments that expresses heterogeneity of genetic variance and correlations across environments, $\boldsymbol{\Sigma^E}$: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma^G} \otimes \boldsymbol{\Sigma^E}$, with $\otimes$ a Kronecker product. A major learning objective for a modern course on GxE is the formulation of VCOV structures for the genotypes, $\boldsymbol{\Sigma^G}$ **,** and for the environments, $\boldsymbol{\Sigma^E}$. For the genotypes, common choices are $\boldsymbol{\Sigma^G} = \boldsymbol{I}$ with $\boldsymbol{I}$ an identity matrix for segregating biparental offspring populations and $\boldsymbol{\Sigma^G} = \boldsymbol{A}$ with $\boldsymbol{A}$ the matrix of additive genetic relationships for association panels. For biparental offspring populations, compound symmetry, with a common covariance between pairs of genotypes, is another appropriate formulation. For association panels and for genomic prediction models, $\boldsymbol{\Sigma^G}$ can be based on pedigree information, marker information, or a combination of these two sources of information (Burgueño et al., 2012; Crossa et al., 2010). The specification of $\boldsymbol{\Sigma^G}$ should be based on investigation of the genetic relationships, whereas the specification of $\boldsymbol{\Sigma^E}$ should be based on inspection of patterns in genotypic variances and correlations across environments. We discuss below a number of tools for determining such patterns in the context of the use of bilinar models. For $\boldsymbol{\Sigma^E}$ a number of well known formulations exist. The most simple model

for $\boldsymbol{\Sigma}^E$ has a common genotypic variance, $\{\boldsymbol{\Sigma}^E\}_{jj} = \sigma_G^2$ and a common covariance $\{\boldsymbol{\Sigma}^E\}_{jj'} = \sigma_{GG'}$ across all environments, where $\{\boldsymbol{\Theta}\}_{kk'}$ denotes the entry of symmetric matrix $\boldsymbol{\Theta}$ for the *k*-th row and *k'*-th column. The heterogeneous compound symmetry model has environment-specific genotypic variances, $\{\boldsymbol{\Sigma}^E\}_{jj} = \sigma_j^2$ and a common covariance, while the uniform correlation model has also environment-specific variances, but a common correlation between environments, $\{\boldsymbol{\Sigma}^E\}_{jj'}/\sqrt{\{\boldsymbol{\Sigma}^E\}_{jj}\{\boldsymbol{\Sigma}^E\}_{j'j'}} = \rho$. The unstructured model has unique genotypic variances and covariances, $\{\boldsymbol{\Sigma}^E\}_{jj} = \sigma_j^2$ and $\{\boldsymbol{\Sigma}^E\}_{jj'} = \sigma_{jj'}$. As the latter model requires many parameters for estimation, factor analytic models are used as parsimonious approximations to unstructured models. Covariances are written as products of environmental scores, $\{\boldsymbol{\Sigma}^E\}_{jj'} = \lambda_j\lambda_{j'}$, while variances are sums of squares of environmental scores and environment-specific terms, $\{\boldsymbol{\Sigma}^E\}_{jj} = \lambda_j^2 + \delta_j^2$. More than one product term can be used for the covariances and variances, for example, $\{\boldsymbol{\Sigma}^E\}_{jj'} = \lambda_{1j}\lambda_{1j'} + \lambda_{2j}\lambda_{j2'}$ and $\{\boldsymbol{\Sigma}^E\}_{jj} = \lambda_{1j}^2+\lambda_{2j}^2 + \delta_j^2$. In principle, the residual term, ϵ_{ij}, is confounded with G_{ij}, but the genetic and residual random terms can be separated by imposing a structure on $\boldsymbol{\Sigma}$, like $\boldsymbol{\Sigma}^G = \boldsymbol{A}$ or by using a factor analytic structure on $\boldsymbol{\Sigma}^E$ and interpreting the terms δ_j^2 as non-genetic residuals. As an alternative to imposing structure on $\boldsymbol{\Sigma}$, σ_ϵ^2 can be obtained from single trial analyses. For a single stage MET analysis, the problem of having to separate the genetic variance from the residual variance will not occur. The students should learn how to identify a suitable model for the genetic VCOV, $\boldsymbol{\Sigma}$, preferentially by log-likelihood ratio tests (Gumedze and Dunne, 2011). If these tests cannot be applied, because the VCOV models are not nested, then information criteria like AIC or BIC may be used (Verbeke and Molenberghs, 2009; Müller et al., 2013).

The students need to interpret the structure of $\boldsymbol{\Sigma}^E$ for conclusions about the heterogeneity of the sample of environments included in the MET. This is an excellent moment to discuss the concept of TPE and investigate the question of whether there are indications that the MET contains trials from more than one TPE, or, whether different mega-environments can be distinguished. To identify different mega-environments in the mixed model context, various types of cluster analysis can be applied to the estimate for $\boldsymbol{\Sigma}^E$, as explained in Cullis et al. (2010). Mixed model theory as developed by Piepho and Möhring (2005) will help to establish the gain of subdividing the TPE in different groups or regions. We recommend to analyze not just yield or biomass, but, if possible, also traits related to phenology. As yield is the primary trait of interest, mega-environments should foremost address yield. Nevertheless, inspecting VCOV models for yield and yield components will contribute to the physiological interpretation of mega-environment assignments.

Closely related to the delineation of mega-environments, investigating the structure of $\boldsymbol{\Sigma}^E$ also helps in answering questions about the presence of crossover interactions (Crossa et al.,

2004; Yang, 2007; Burgueño et al., 2008). Environments or trials with high genetic correlations will have few crossover interactions.

From the identified model for the MET data, predictions for genotypic performance can be made in each of the environments. Subsequently, adaptation of genotypes can be studied in the very limited sense of which genotypes do best in individual environments. Environments in which the same genotype or subset of genotypes do best may be members of a common mega environment. Questions on adaptability and stability cannot be answered straightforwardly from models without implicit (bilinear) or explicit (factorial regression) genotypic and environmental covariables.

Interaction in linear models

The two-way ANOVA can serve to introduce the *concept of interaction in linear models*. A plot of estimates of GxE means, $\hat{\mu}_{ij}$ versus environmental main effects, $\hat{e}_j$, or, equivalently, environmental means, $\hat{\mu}_j$, can help in diagnosing non-parallelism of genotypic responses across environments and thereby visualize the presence of GxE. Actually, GxE can equally be inspected via a plot of the genotype by environment means, $\hat{\mu}_{ij}$, versus the genotypic means, $\hat{\mu}_i$, although it is more natural to plot GxE means versus environmental means, as this is the basis for the Finlay Wilkinson regression (Yates and and Cochran, 1938; Finlay and Wilkinson, 1963). Plots of GxE means versus an environmental characterization as the environmental mean can be used to come back to the concept of the reaction norm within the context of ANOVA. When reaction norms are non-constant, genotypes show plasticity, see Figure 1a versus 1b to 1f (Allard and Bradshaw, 1964; DeWitt and Scheiner, 2004; Sadras and Lawson, 2011; Slafer et al., 2014). To elaborate this point statistically, plasticity points to the existence of an environmental main effect. When reaction norms show plasticity and are non-parallel, we have GxE, see Figure 1c to 1f. When reaction norms are non-parallel, but do not intersect, we speak of *quantitative interaction* or *non-crossover interaction*, see Figure 1c and 1d. When the reaction norms do intersect, we speak of *qualitative interaction* or *crossover interaction*, see Figure 1e and 1f. This latter type of GxE has more severe consequences for breeders, as it will change the rank order of genotypes as a function of the environmental conditions.

A formal test for GxE interaction in a two-way ANOVA on GxE means requires an estimate for the error, σ_ϵ^2. Such an error estimate can come from the analysis of single trials. Independent of any formal testing, the ANOVA partitioning of the phenotypic variation into genotypic and environmental main effect and GxE gives a rough indication of the importance of GxE, judged by sum of squares, mean squares or F-values. We prefer the use of variance components in random and mixed models over sums of squares partitioning in fixed

ANOVAs to quantify the contributions of genotype, environment and GxE to the phenotypic variation, even more so when genotypes and environments contain crossed and/or nested factorial structure. The variance components can be expressed on the scale of coefficients of variation to facilitate their interpretation (Gelman, 2005).

Estimation and inference for two-way ANOVA is straightforward for complete GxE tables. For ANOVA interaction parameters, under sum to zero constraints the classical estimator for GxE interaction residuals is $\widehat{ge}_{ij} = y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}$, where the bar indicates averaging over subscripts that are omitted. Such simple looking estimators do not exist for non-orthogonal GxE data. A number of adaptation and stability parameters have been proposed that are functions of these interaction residuals (Wricke, 1962). However, the biological reality and usefulness of such estimators depends on the extent to which the sum to zero identification constraints can be biologically justified (Denis, 1991).

Crossover interactions

In a course on GxE, attention should be given to methods for detecting crossover interactions. The importance of crossovers depends on the magnitude of these interactions and the genotypes and environments involved. The identification of crossovers is a first step in the exploration of the genetic and physiological factors underlying genotypic differences that are conditioned by the environment. Crossover tests make sense for GxE tables with limited numbers of cultivars of contrasting adaptation (tolerant or resistant versus susceptible or sensitive) that are tested under contrasting environmental conditions (e.g., stress versus non-stress). However, for these situations more powerful a priori crossover contrasts can be defined and tested by standard t-test procedures. To emphasize the continuity between ANOVA and mixed models, tests for a priori crossover contrasts should also be demonstrated for mixed models. Students will need to learn to first identify the right fixed or mixed linear model and then answer specific breeding questions by imposing the appropriate contrasts on the levels of genotypic and environmental factors. When a suitable mixed model is identified and specified, mixed model software like ASReml, Genstat and SAS offers the possibility to define and test contrasts that then will automatically have the correct standard errors.

The literature on GxE shows a certain obsession with the identification of crossover interactions. The two-way ANOVA model provides a plain setting to address the issues of the identification of crossover interactions and the assessment of their importance and interpretation. A simple post hoc procedure for testing individual crossover interactions first identifies pairs of genotypes, *i* and *i'*, and pairs of environments, *j* and *j'*, for which either the condition $(y_{ij} - y_{i'j}) > 0$ and $(y_{ij'} - y_{i'j'}) < 0$ is fulfilled, or the condition $(y_{ij} - y_{i'j}) < 0$ and $(y_{ij'} - y_{i'j'}) > 0$. Subsequently, the interaction sum of squares for the

contrast $y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'}$ is calculated with a conservative correction for multiple a posteriori testing according to Scheffé's test for simultaneous inference (Kuehl, 2000). More powerful tests for testing crossovers are the Azzalini-Cox test (Azzalini and Cox, 1984) and the Gail – Simon test (Gail and Simon, 1985). See Baker (1988) for their application to plant breeding examples.

Model building

Two-way GxE models are generalized to multi-way models when either or both of the genotypic and environmental factor contains itself factorial structure. For example, the genotypic factor may be the factorial product of an earliness classification and a stress tolerance classification. The environmental factor can be a product of location and year, or of management and year. In addition, multi-way factorial structure on the environments is easily imaginable when multiple management factors (drought/irrigated x high/low nitrogen) are crossed with years and/or locations. For analyzing multi-way GxE tables in ANOVA (all factors fixed) a recommended strategy first tests the highest order interaction (Welham et al., 2014, section 8.3.1). When this interaction is significant, the model for prediction will include all model terms up to the highest interaction. We then need to inspect the corresponding multi-way table of means and calculate contrasts on that table for interpreting the interactions. When the highest order interaction is not significant, the corresponding interaction term will not be included in the prediction model and we continue testing the next highest order interactions (Welham et al., 2014, section 11.2.3).

For mixed models, the strategy described by Welham et al., (2014) carries over to the table of fixed terms given that the random structure is kept constant, i.e., the random model is the same for all fixed terms. Testing is done by Wald tests or their F-test equivalents. For mixed models, the situation is complicated by the fact that for individual terms a decision needs to be motivated whether they are fixed or random. This is a subject for which the literature is not very clear and students (and teachers) get easily confused. Choices for whether genotypic, environmental and GxE terms should be fixed or random depend on the question that requires an answer, but also on the assumptions that one is willing to make. Assuming that a term is random imposes additional constraints on the VCOV for the observations and requires verification of the distribution of the corresponding effects. A number of characteristic situations can be presented with the question to formulate reasons for choosing individual model terms as fixed or random and discuss the consequences of choosing particular terms as fixed or random. For METs with segregating populations that will be used for QTL mapping or association mapping, we prefer to take the main effects for genotypes and GxE terms random. If the genotypes consist of a limited set of genotypes late

in the breeding program or of a set of candidate cultivars in an official variety testing cycle, then the genotypic main effects best are chosen fixed. For the environments, managed stress treatments can be treated as fixed when they refer to (hopefully) repeatable environmental conditions. Similarly, repeatable GxE for a small number of genotypes under various levels of a well-defined managed stress factor will be fixed. For locations and years, we prefer to take the main effects, which are a kind of intercept terms, fixed, where the year main effects may be taken random when it concerns many years. For the GxE interactions, the genotype by location interaction is fixed for repeatable locations and selected genotypes (Annicchiarico, 1997; Annicchiarico et al., 2005). The GxE interactions with years are all random. Testing of fixed genotype by management and genotype by location interactions will happen against a background of random genotype by management by year and genotype by location by year interactions. Therefore, tests for these repeatable GxE interactions in mixed models will differ from similar tests in an ANOVA as will standard errors of differences for contrasts. Students will need a series of practical examples and exercises to learn how to define appropriate mixed models and arrive at sound breeding conclusions.

Partitioning GxE in pattern and noise by grouping genotypes and environments

Imposing a categorization or grouping on genotypes and environments allows a partitioning of the initial GxE into an interaction between genotype groups and environment groups and a GxE residual, a deviation from the grouping term: $y_{i(k)j(l)} = \mu + g_i + e_j + rc_{kl} + \delta_{i(k)j(l)} + \epsilon_{i(k)j(l)}$ with rc_{kl} the interaction between the genotype group k and the environment group l (rows and columns, respectively of the two-way GxE table). The term $\delta_{i(k)j(l)}$ is a deviation from the row by column group interaction model that can be considered to represent a random normal variable with zero mean and proper variance. The random term $\epsilon_{i(k)j(l)}$ is confounded with $\delta_{i(k)j(l)}$ in a two-stage analysis of GxE means, but an estimate for its variance can be obtained from intra-trial analyses. The grouping of genotypes and environments is then tested against the mean square for the deviations of the interaction model by an F-test (ANOVA) or Wald test (mixed model) (van Eeuwijk et al., 1996; Kuehl, 2000; Welham et al., 2014). The mean square for the deviations is tested against the estimate for the error, σ_ϵ^2. The choice of testing the groupings against the deviations from the groupings, changes an ANOVA model implicitly in a mixed model. Groupings of the levels of environmental factors can also be included in more complex mixed models with multiple environmental factors, in which some factors and GxE interactions are fixed and others random. For example, in a MET with fixed genotypes and multiple trials across fixed locations and random years, we may want to cluster the locations for the fixed genotype by location interaction, but not the years in the random genotype by year interaction. As a learning objective, students should identify and specify relevant groupings of genotypes and

environments and test these groupings in both ANOVAs and mixed models. Correct specification in ANOVA and mixed model software will lead to the desired F- and Wald tests, respectively.

The success of groupings of genotypes and environments depends on the amount of variation that is described by the interaction between the groupings, the significance of the test on the groupings and the test on the deviations from the groupings. Strong significance of the test for groupings and no significance for the deviations indicate that the pattern in the GxE is captured by the genotype and environment groupings, leaving noise for the deviations from the groupings. The part of the GxE that is covered by genotype and environment groupings is the part that qualifies for being repeatable. Predictions from models with genotype and environment groups are delivered at the level of the groups and not at the level of individual GxE combinations. Partitioning the GxE into a part described by groupings and another part by deviations from grouping reflects a very general principle to reduce the complexity of GxE interactions. Many types of genotypic and environmental groupings can be proposed. Students should become capable of defining and constructing promising groupings from physiological and breeding knowledge. Groupings obtained from exploratory analyses like bi-clustering (Corsten and Denis, 1990) and the application of additive main effects and multiplicative interaction models (Gauch, 1992, 2013) can be inserted in ANOVA models and mixed models to test their contribution to the GxE. Some correction for data snooping and multiple testing is necessary. Again, Scheffé's simultaneous test for a posteriori contrasts is a conservative possibility.

Partitioning GxE in pattern and noise by genotypic and environmental covariables

Grouping of genotypes and environments can be interpreted as the introduction of qualitative covariables on the levels of genotypes and environment. Similarly, quantitative covariables on genotypes and environments can be included to investigate the nature of the GxE. We consider the skillful introduction of covariables for investigating GxE patterns to constitute a major learning objective in a course on GxE. Inclusion of quantitative covariables in models for GxE is equivalent to the definition of contrasts to study interactions in classical ANOVA and mixed models. Various statistical textbooks contain insightful chapters on the use of contrasts to study factorial interactions. We recommend Kuehl (2000) and Welham et al. (2014). In the plant breeding literature, models for GxE using covariables are called factorial regression models. Some papers giving theory and applications of factorial regression are Denis (1988), van Eeuwijk et al. (1996), Vargas et al. (1999), Voltas et al. (1999a), Malosetti et al. (2004), van Eeuwijk et al. (2005). Some nice recent examples of factorial regression can be found in Crossa et al. (2015) and Vargas et al. (2015).

Factorial regression, a single environmental covariable

For a factorial regression model that contains a single environmental covariable, to describe GxE, the fixed GxE term in ANOVA and mixed models, ge_{ij}, is partitioned into a regression part, $\beta_i z_j$, with a genotype-specific slope or sensitivity, β_i, to the environmental covariable, z_j, and a residual or deviation from the regression model, δ_{ij}: $ge_{ij} = \beta_i z_j + \delta_{ij}$. Thus, the double indexed GxE term ge_{ij}, which does not permit predictions to new environments, is replaced by a separable formulation, $\beta_i z_j$, with single indexed values for the genotypic parameter and the environmental covariable, that does offer the possibility for predictions to new environments for genotypes whose sensitivities have been estimated.

It should be emphasized that it is the heterogeneity of the slopes that is of importance for GxE. A sensitivity to an environmental covariable that is constant across genotypes partitions a fixed environmental main effect, $e_j = \beta z_j + \delta_j$, but not a fixed ge_{ij} term, so the environmental covariable would be responsible for plasticity without GxE. A formal test for whether the covariable z_j 'explains' a significant proportion of ge_{ij} can be constructed from testing the mean square for the heterogeneity of the genotypic sensivities over the mean square for the deviations from the factorial regression. In ANOVA, the degrees of freedom for the regressions are *I-1* for the heterogeneity of slopes and *(I-1)(J-2)* for the deviations from the regression.

Factorial regression with multiple environmental covariables

Above, we introduced the concept of factorial regression for a single environmental covariable. However, typically more than one environmental covariable or characterization is required to arrive at an acceptable description of the pattern in the GxE. The factorial regression model can contain multiple environmental covariables that are elements of the environmental covariable set *EC*, $ge_{ij} = \sum_{k \in EC} \beta_{ki} z_{kj} + \delta_{ij}$, or polynomial forms of environmental covariables, $ge_{ij} = \beta_{1i} z_j + \beta_{2i} z_j^2 + \delta_{ij}$. Further elaborations of the factorial regression approach include standard non-linear curves as the logistic, Richards and Gompertz function (Butler and Brain, 1993). For fixed GxE terms, the maximum number of environmental covariables that can be included is equal to the number of environments minus *1*, *J-1*. When more than *J-1* environmental covariables are included, some form of penalization or dimension reduction is required, as in reduced rank regression (Denis (1988), van Eeuwijk (1992)) and partial least squares (Aastveit and Martens, 1986; Vargas et al., 1999, 2015). Recently, interesting suggestions were made to include large numbers of environmental covariables in mixed model factorial regressions where these covariables define an environmental relationship matrix that imposes structure on random G_{ij} terms. As before, we assume that the matrix of random G_{ij} effects has a multivariate normal matrix

distribution: $\{G_{ij}\} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^G \otimes \boldsymbol{\Sigma}^E$, but now environmental covariables define the environmental VCOV $\boldsymbol{\Sigma}^E$ (Jarquín et al., 2013; Bustos-Korts et al., 2016; Malosetti et al., 2016; Pérez-Rodríguez et al., 2015).

Selecting environmental covariables

Classical examples of environmental covariables are soil and meteorological variables. To choose environmental covariables in factorial regression models, standard variable subset selection (forward and stepwise regression) can be demonstrated to students. Nowadays, automatic environmental monitoring protocols produce measurements at short time intervals. It is not immediately obvious how to select from a multitude of short interval measurements the most relevant ones. Routine variable selection procedures will not work well with large numbers of variables to select from. Alternative statistical variable selection methods like penalized and sparse regression methods for high dimensional regression like the Lasso (Tibshirani, 1996; Meinshausen, 2007; Taylor et al., 2012) look interesting, but have not been extended yet to screen large sets of covariables for interaction terms. However, an attractive alternative to statistical selection methods seem integrations over time of multiple environmental variables by crop growth models (Chapman, 2008; Chenu et al., 2011, 2013; Heslot et al., 2013) to produce a limited set of environmental characterizations known to be biologically relevant. Students in a GxE course will benefit from an introduction to the concept of integration of environmental information over time with crop growth models. Instead of running a simple crop growth model to obtain environmental characterizations, students may also be provided with environmental characterizations from earlier crop growth model runs.

Factorial regression with genotypic covariables

Factorial regression under inclusion of genotypic covariables is another useful approach to identify the patterns driving GxE and to search for separable models. The formulation of the model for multiple genotypic covariables, of the type x_{li}, that are all elements of the genotypic covariable set *GC*, in the case of fixed GxE is $ge_{ij} = \sum_{l \in GC} x_{li}\alpha_{lj} + \delta_{ij}$, with the number of genotypic covariables limited to maximally *I-1*. Again, more genotypic covariables can be included, but analogous to the situation for environmental covariables above some form of dimension reduction or penalization should be imposed. Genotypic covariables for yield can be measurements made under managed conditions, including phenotyping platforms, on traits like disease resistances and biotic stress tolerances. The parameters α_{li} then express the severity of the disease or abiotic stress in the experiment *j*. When the genotypic covariables are functions of marker genotypes, the factorial regression model above immediately becomes a model for multiple quantitative trait loci (QTLs) with

the QTLs interacting with the environment (Malosetti et al., 2004, 2013; Boer et al., 2007; van Eeuwijk et al., 2010). For genomic prediction under GxE, the full set of markers can be used to define a genomic relationship matrix on the genotypes. This genomic relationship matrix can be combined with an environmental relationship matrix to produce genomic predictions for new genotypes in new environments (Bustos-Korts et al., 2016; Malosetti et al., 2016).

Linear-bilinear models; an overview

Where factorial regression models contain multiplicative formulations for GxE that use explicit genotypic and environmental covariables, linear-bilinear models use implicit covariables. For our purposes, the essence of analyses by linear-bilinear models is that they generate ideas for groupings and covariables that can be further tested in ANOVA and mixed models. For recent papers on linear-bilinear models, see Gauch et al. (2008), Crossa et al. (2010b), Crossa, (2012) and Gauch (2013). Linear-bilinear models have their name from the fact that in addition to linear terms for genotype and/or environmental main effects they contain bilinear terms for GxE or for combinations of a main effect and GxE. Bilinear terms are separable products of parameters, like $r_i c_j$, with r_i for the sensitivity of genotype i and c_j for the characterization of environment j, that both need to be estimated from the phenotypic data. The name comes from the fact that fixing the genotypic parameter r_i in $r_i c_j$ makes the model linear in c_j and vice versa. A well known member of the linear-bilinear class of models is firstly, the Finlay Wilkinson model (Yates and Cochran, 1938, Finlay and Wilkinson, 1963). A second linear-bilinear model is the additive main effects and multiplicative interaction effects model (AMMI; Gauch Jr., 1988; Gauch, 1992). Thirdly, we mention the Genotype main effects and Genotype by Environment interaction effects model (GGE model; Yan and Kang, 2002). This model is also known in the GxE literature as the site regression model (Crossa and Cornelius, 1997). Actually, the GGE model is equivalent to principal components analysis of the GxE two-way table of means and all theory on principal components carries over to GGE analysis (Gabriel, 1971; Jolliffe, 2013). More linear-bilinear models have been developed and are used for GxE analysis, but would fall outside a GxE course of 3 to 5 days, see for example Cornelius and Seyedsadr (1997) and Crossa (2012).

Finlay Wilkinson model

For a table of two-way GxE means the fixed Finlay Wilkinson model is $y_{ij} = \mu_i + r_i c_j + \delta_{ij} + \epsilon_{ij}$. In comparison with the two-way ANOVA model we rewrite the linear terms $e_j + ge_{ij}$ as $r_i c_j + \delta_{ij}$. In the Finlay Wilkinson model, a single environmental characterization is used, c_j, that either is equal to the average performance of all genotypes in an environment, as in the paper by Finlay and Wilkinson (1963), or is very close to it, as in Mandel (1969). For the latter case, a genotypic intercept term and the genotypic and environmental scores, r_i and c_j, respectively, are found from minimizing by least squares the expression: $\sum_{i,j}^{I,J}\left(y_{ij} - (\mu_i + r_i c_j)\right)^2$. Estimates for the scores are obtained from a singular value decomposition (SVD) of the matrix of GxE means corrected for the genotypic main effect: SVD $(y_{ij} - \mu_i)$ (Gabriel, 1978). Thus, genotypes are characterized by an intercept parameter, μ_i, for general performance and a slope or sensitivity or adaptability parameter, r_i, with unit mean. A large value for the intercept and a sensitivity close to 1 point to a widely adapted genotype. The recommended way for interpreting the results of a Finlay Wilkinson analysis is first to check whether a significant and relevant part of the GxE is explained by the heterogeneity of the slopes and whether the residuals from the model do not show shortcomings of the model. Next the predicted values can be calculated, $\hat{y}_{ij} = \hat{\mu}_i + \hat{r}_i \hat{c}_j$, and the fitted regression lines plotted to identify the superior genotypes for specific environments. Predictions for new environments are possible in so far new environments can be recognized as being similar to trials already included in the set of trials to build the model. The Finlay Wilkinson model is a candidate model for relatively simple environmental configurations in which the environments are homogeneous and differ in a single dimension. For example, the environments may represent optimal conditions except for the level of a limiting factor. When the environments have a more complicated factorial structure, Finlay Wilkinson terms can be embedded in mixed models. For example with a fixed location and a random year factor, a fixed bilinear Finlay Wilkinson term can be proposed that regresses genotype by location means on location means, to study adaptability, while the genotype by year and the genotype by location by year interactions are chosen random. The deviations from the Finlay Wilkinson regression will then be tested over the three-way genotype by location by year interaction. Another generalization of fixed Finlay Wilkinson models are mixed bilinear models in which random bilinear terms are included to describe GxE (Gogel et al., 1995; Nabugoomu et al., 1999): $y_{ij} = \mu_i + G_{ij} + \epsilon_{ij}$ and $\{G_{ij}\} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma^G} \otimes \boldsymbol{\Sigma^E}$ and $\boldsymbol{\Sigma^E}$ having a factor analytic structure; the variance for environment j is $\lambda_j^2 + \delta_j^2$ and the covariance between environments j and j' is $\lambda_j \lambda_{j'}$, with λ_j a score for environment j that defines covariances with other environments and the basis for a shared part of the genetic variance, while δ_j^2 stands for an environment-specific genetic variance part.

The Finlay Wilkinson model is the most frequently used model for the analysis of GxE according to a citation analysis using the Web of Science (see later paragraph). For that reason, it deserves some time dedicated to it in any course on GxE. However, its applicability is limited. Within the class of fixed bilinear models more flexible models are available that allow more than just one bilinear term to be included for a description of GxE. For larger numbers of genotypes, mixed bilinear models seem a more viable modelling option than fixed bilinear models as the Finlay Wilkinson model.

The original Finlay Wilkinson model that aims at an analysis of adaptation and adaptability was extended by Eberhart and Russell (1966) with a stability analysis. To that end, they defined genotype-specific stability variances based on the deviations from the Finlay Wilkinson regression lines. With Gauch (2013) we share some doubts about the utility of stability analyses. First, when calculating stability statistics across a set of trials, it should be verified that these trials belong to a single TPE or mega environment. Second, the analysis of GxE should emphasize the identification of adequate models for the reaction norms and minimization of the deviations from the reaction norms, i.e., minimization of stability variances. When reaction norm models show a good fit to the data with clear genotypic differences for the reaction norm parameters, it is unlikely that simultaneously biologically relevant variation for stability variances will be present (see Kraakman et al., 2004). For historical and didactical reasons, in a GxE course, Finlay Wilkinson regression and Eberhart Russell stability should be presented as early models that connected statistical parameters with breeding concepts such as adaptation, adaptability and stability. A learning objective for students is how to evaluate the merits of classical Finlay Wilkinson and Eberhart Russell approaches in comparison to recent more elaborate fixed and mixed bilinear models.

AMMI models

Fixed bilinear models are and have been very popular for the analysis of GxE. With the Finlay Wilkinson model, the AMMI model, (Gauch Jr., 1988; Gauch, 1992) deserves considerable attention in any GxE course. In the AMMI model, we write the GxE as $ge_{ij} = \sum_{a=1}^{A} r_{ai}c_{aj} + \delta_{ij}$, with the r_{ai}'s genotypic scores or sensitivities and the c_{aj}'s environmental scores or characterizations. The full AMMI model is $y_{ij} = \mu + g_i + e_j + \sum_{a=1}^{A} r_{ai}c_{aj} + \delta_{ij} + \epsilon_{ij}$. For fixed two-way GxE tables, estimates for genotypic and environmental scores are obtained from an SVD of the ANOVA interaction residuals, ge_{ij}. The SVD of the interaction is equivalent to finding the environmental characterizations that best discriminate between genotypes following a least squares criterion.

In an introductory course on GxE it is enough to mention that the multiplicative scores are obtained by minimizing $\sum_{i=1,j=1}^{I,J}\left(ge_{ij}-\sum_{a=1}^{A}r_{ai}c_{aj}\right)^2$. The number of terms, *A*, for 'adequate' description of the GxE can be assessed in a number of ways (Cornelius, 1993; Bro et al., 2008; Josse and Husson, 2012), but for an introductory course, the F-test approximation by Gollob (1968) will do. This approach allocates *(I+J)-(1+2a)* degrees of freedom to the sums of squares explained by *a*-th term to convert it into a mean square. Analogous to earlier partitionings of the GxE, the mean square corresponding to the AMMI model can be tested against deviations from that model, while the deviations are tested against an independent estimate for the error from within trial analyses.

After assessing the dimension of the AMMI model, i.e., establishing the number of multiplicative terms to retain, *A*, adaptation can be investigated by plotting and comparing the environment centred predictions for individual genotypes, $\hat{y}_{ij}=\hat{\mu}+\hat{g}_i+\sum_{a=1}^{A}\hat{r}_{ai}\hat{c}_{aj}$. The predicted reaction norms may be used to find out which environments are similar or part of a mega-environment by comparing the best genotypes (Gauch et al., 2008). The genotypes can further be characterized by the genotypic main effect, g_i, a measure of wide adaptation, and the genotypic sensitivities, r_{1i}, r_{2i},..., r_{Ai}. These genotypic sensitivities can be used to identify groups of genotypes with similar GxE.

The sum of squares of the genotypic scores, $S_i=\sum_{a=1}^{A}r_{ai}^2$, is an approximation to the sum of squares for interaction in a fixed two-way ANOVA, which is equivalent to Wricke's stability statistic (Wricke, 1962): $W_i=\sum_{i=1,j=1}^{I,J}ge_{ij}^2$. The AMMI genotypic scores need to be scaled appropriately for this approximation to work (Gauch et al., 2008). Still, although stability parameters can be defined in fixed ANOVA, we prefer to estimate stability parameters as genotype-specific variances in mixed models.

In the AMMI model, environments have as characteristics the general quality, e_j, and the specific qualities c_{1j}, c_{2j},..., c_{Aj}. Comparable to the genotypes, environments can be grouped on the basis of their environmental scores. To test the contributions of these groupings, contrasts in ANOVA or mixed models can be defined. Another possibility is to apply tests for multiplicative interactions as described by Milliken and Johnson, (1989).

AMMI approaches are remarkably popular. Good predictive properties have been attributed to AMMI models (Gauch, 2006; Gauch et al., 2008). Surely, AMMI predictions will be useful for many breeding purposes, but if bilinear models for GxE can be embedded in mixed models, we would create predictions from the latter. Another asset of AMMI models, and bilinear models in general, is their possibilities to display GxE patterns graphically in biplots (Gauch and Zobel, 1997; Yan et al., 2000; Yan and Rajcan, 2002). For *A*=2, an AMMI biplot contains genotypes with coordinates (r_{1i}, r_{2i}) and environments with

(c_{1j}, c_{2j}). AMMI biplots are highly useful tools to explore GxE patterns. Outlying genotypes and environments can readily be detected, as can groupings of genotypes and groupings of environments. Even, adaptation can be investigated in AMMI biplots. A GxE course will need to reserve time for students to learn the interpretation rules for various types of biplots, but AMMI models should be used primarily for exploration of GxE patterns, with formal testing of GxE structure in mixed models. Recent developments in Bayesian bilinear models (Crossa et al., 2011; Perez-Elizalde et al., 2012; Josse et al., 2014) alleviate the inferential restrictions on bilinear models, but Bayesian bilinear models would fall outside the scope of an introductory GxE course.

GGE models

In the GGE model, the ANOVA terms $g_i + ge_{ij}$ are written as $\sum_{p=1}^{P} r_{pi}c_{pi} + \delta_{ij}$, with P bilinear terms, leading to the model $y_{ij} = \mu_j + \sum_{p=1}^{P} r_{pi}c_{pi} + \delta_{ij} + \epsilon_{ij}$ (Yan and Kang, 2002). With Finlay Wilkinson models and AMMI models, GGE models enjoy a large popularity in the applied literature on GxE. A first reason is that the GGE model produces biplots that cover both the genotypic main effect and the GxE, while AMMI biplots focus on GxE solely. As a second reason, we mention that a particular feature of GGE biplots appears to allow the identification of mega-environments (Yan et al., 2000). The GGE biplot is mostly shown in two dimensions, as more dimensions are difficult to work with. The quality of the two dimensional GGE biplot for inference on genotypic adaptation (predicted values of GGE model, equivalent to lengths of projections of genotypic vectors on environmental vectors), genotypic sensitivities (genotypic scores), genetic variances (squared lengths of environmental vectors), and genetic correlations (angles between environmental vectors) depends on the amount of variation that is represented by the first two dimensions of the GGE model. As the first axis of GGE models tends to mimic the genotypic main effect, which is not present in an additive form in the GGE model, a two dimensional GGE biplot will cover about as much GxE variation as an AMMI model with one bilinear term. Therefore, in two dimensional GGE biplots, less GxE pattern will be shown than in two dimensional AMMI biplots. In a GGE biplot, an average environment axis can be constructed as the average of the vector representations of the environments. Genotypic projections on this average environment axis approximate the genotypic main effect. The distance between a genotype representation in the GGE biplot, $(r_{1i}r_{2i})$, and its projection on the average environment axis is assumed to give an estimate for stability (Yan and Kang, 2002). This type of stability in the GGE biplot will be close to the stability based on the first genotypic score in an AMMI biplot. Various discussions have been published about the comparison of AMMI and GGE model analyses (Gauch et al., 2008; Yang et al., 2009; Gauch, 2013), without the authors getting to an agreement. We feel that both AMMI and GGE are very

useful techniques to explore GxE interactions. There is little reason to prefer one to the other. GGE biplots present a view on the totality of genotype related variation, $g_i + ge_{ij}$, whereas AMMI biplots show more detail for the GxE part of the phenotypic variation, ge_{ij}.

GGE or principal component biplots are excellent ways to explore genetic variances and correlations. The biplot with optimal scaling for the environments can be used alongside with scatterplots matrices for the phenotypic responses across environments to develop ideas about the variance-covariance structure of the data. Hypotheses with respect to patterns in the genetic variances and correlations can be tested in mixed models with likelihood ratio tests.

Delineation of mega-environments based on winning genotypes in individual environments (Yan et al., 2000) in GGE models seems risky (Yang et al., 2009). Suggestions for defining mega-environments, or subdivisions of the TPE, can be obtained in many ways, but we would like to insist on testing the efficiency of such subdivisions in a suitable mixed model, following protocols as described by Piepho and Möhring (2005).

8.3.2. Crop growth models, multi-trait reaction norms and networks

The phenotypes that are analyzed to answer questions with respect to GxE are mostly complex traits, as yield itself. From a statistical perspective, complex traits are traits for which multiple QTLs can be identified with additive, dominance and epistatic effects that can interact with the environment. For complex traits, in a GxE course students learn how to fit suitable models with genotypic and environmental covariables to describe GxE. In our statistical approaches to GxE little attention is given to the fact that phenotypes are products of genetic, physiological and environmental interactions over time. A good way to introduce the developmental aspects of phenotypes is via crop growth models. In this section, we present a statistical-physiological framework for better understanding GxE. Let us label the phenotype we want to predict as the target trait, or focus trait, y_{ij}^{f}. The target trait is now a response trait in a crop growth model, with as inputs a vector of genotype dependent component traits, $\boldsymbol{y}_i^c$, and a vector of environmental variables, $\boldsymbol{z}_j$ (Yin et al., 2000a; Chapman et al., 2002; Snape et al., 2007; Chenu et al., 2009; Malosetti et al., 2016; Technow et al., 2015; van Eeuwijk, 2015). Component traits are related to resource capture (e.g., leaf area index, root architecture), conversion efficiency (e.g., light use efficiency, water use efficiency) and biomass allocation to yield (e.g., harvest index), while environmental variables represent the amount of resource (e.g., light, water, nutrients) and conditions as temperature and CO_2 (Ceccarelli et al., 1991; Slafer and Andrade, 1993; Cooper and Hammer, 1996; Slafer et al., 2014, 1996; Yin et al., 2000b; Reynolds et al., 2009, 2011;

Nicotra and Davidson, 2010; Foulkes et al., 2011; Parry et al., 2011; Sadras and Calderini, 2014). The component traits are integrated over time with the environmental inputs to form the target trait: $y_{ij}^{f} = \int f\left(\boldsymbol{y}_i^c;\ \boldsymbol{z}_j\right)\, dt + \varepsilon_{ij}$ (Chenu et al., 2009; Yin and Struik, 2010; Bustos-Korts et al., 2016). $\int f\left(\boldsymbol{y}_i^c;\ \boldsymbol{z}_j\right)\, dt$ represents the integral over time of the function $f\left(\boldsymbol{y}_i^c;\ \boldsymbol{z}_j\right)$ that converts the inputs $\boldsymbol{y}_i^c$ and $\boldsymbol{z}_j$ into the target trait, y_{ij}^{f}. The term ϵ_{ij} is an error term.

A crop growth model is a complex reaction norm model that describes how to convert component traits and environmental inputs into the target trait yield. Equally, crop growth models can be seen as devices for the dynamic modelling of multiple traits. An interesting extension of classical physiological crop growth models includes DNA marker variation as underlying phenotypic variation in component traits. Effectively this means that the phenotypic values for the component traits are replaced by predictions from QTL models or genomic prediction models (Yin et al., 2000a, 2003, 2005; Reymond et al., 2003; Tardieu et al., 2005; van Eeuwijk et al., 2005; Chenu et al., 2009; Bogard et al., 2014; Malosetti et al., 2016; Technow et al., 2015).

Alternative approaches for modelling target traits as functions of component traits in their joint development over time are given by Sun and Wu (2015). These authors propose a differential equation framework for modelling the dynamics of multiple traits in systems genetics, where the constants in the differential equations are themselves modelled as linear functions of underlying QTL genotypes.

The prediction of phenotypes can be improved by modelling intermediate levels of biological variation in between the DNA level (SNPs and sequence information) and the final phenotype (target trait): gene expression, proteins, metabolites, methylation, etc.. Network models are a popular type of model for combining the variation of different types of traits at multiple levels of biological organization, including target phenotypic traits at the highest level (Welch et al., 2003, 2005; Neto et al., 2008, 2010; Scutari et al., 2014; Wang and van Eeuwijk, 2014; Wang et al., 2015). Network models show the behavior of multiple traits in their dependence on each other. Variation in genetic correlations between traits across environmental conditions is an important form of GxE (Malosetti et al., 2008; Alimi et al., 2013). Network models can make such changes visible in a biologically meaningful way.

We see the multi-trait and dynamical modelling perspective offered by crop growth and differential equation modelling with genetic, genomic and environmental inputs as a benchmark for biologically meaningful modelling of GxE. In this framework GxE arises as an emerging property of the model system as all inputs and/or parameters are exclusively indexed by either genotypes or environments, but not by both (Cooper et al., 2002; Hammer et al., 2005, 2006). This exclusive dependence of the phenotype on either genotype or

environment is referred as separability (Gregorius and Namkoong, 1986; Cornelius et al., 1992). In our statistical approaches to GxE, we aim at identifying predictive models that approach as close as possible this ideal of separability.

Reaction norms as crop growth models and differential equations systems are biologically realistic by the emphasis they place on the multi-trait and dynamic aspects of the phenotype. In a 3 to 5 day course on statistical approaches to GxE, little attention can be given to explicit dynamic (Malosetti et al., 2006; Wu and Lin, 2006) and multi-trait modelling approaches. However, simultaneous univariate analyses of traits, especially yield and phenology, can shed light on trait dependencies, while the dynamic behavior of traits can be represented by slope and curvature parameters of reaction norms (Van Eeuwijk et al., 2007; van Eeuwijk et al., 2010; Hurtado-Lopez et al., 2015).

8.4. GxE models in the literature

8.4.1. Designing a GxE citation index

Previous sections gave an overview of our choices for statistical approaches to model GxE as to be presented in a course on GxE. We recommended an approach departing from concrete breeding questions and think mixed models with linear, bilinear and factorial regression terms are most suitable for a three to five day course. As a closing section of our paper, we compare our choices for particular methods to popularity of methods for GxE analysis in the literature. We analysed the Web of Science citation reports between 1965 and 2015 (Thomson Reuters, 2015). Our search for keywords related to GxE showed a total of 2275 references. We focused on two groups of references; those that had 20 or more citations (highly cited), and those that were published between 2013 and 2015 (recent literature). Based on our subjective judgement, we found that out of the 447 highly cited papers, 302 corresponded to applications and 175 to methodological papers, discussion papers and reviews. Papers could belong to more than one category. Methodological papers were, again subjectively, classified in 11 categories, depending on the model used (Figure 2). Most of the 11 categories corresponded to models discussed in previous sections; e.g. mixed models, AMMI, GGE, ANOVA, stability measures. However, other categories were added, like, for example, papers that contained informal GxE analyses without fitting mixed models with different GxE terms, but that simply compared results of single environment QTL analyses. Of course, such an approach allows to obtain an impression about environment-dependent QTL effects, but ideally one would like to fit a mixed model for simultaneous multi-environment QTL detection (Boer et al., 2007; Alimi et al., 2013).

For Bayesian models and approaches, these papers look at the estimation of genotypic and environment-specific variance components by Bayesian methods (Yang et al., 2007), or propose Bayesian methods to quantify uncertainty for genotypic or environmental scores in AMMI analysis (Josse et al., 2014). Thanks to their heavy use in genomic prediction methods, Bayesian methods are rapidly gaining in popularity in single-environment genetic analysis (Crossa et al., 2010a; Jia and Jannink, 2012; Spindel et al., 2015). For GxE analysis, Bayesian methods are still less frequent (Figure 1).

8.4.2. Results of GxE citation analysis

When evaluating the use of model categories over time, the number of citations obtained by GxE papers was relatively stable between 1965 and 2003, and it rapidly increased afterwards (Figure 2). Papers using stability measures, fixed Finlay Wilkinson models, ANOVA or simple linear regression models showed the largest number of citations during the last part of the 20th Century. However, the ranking of model use changed with the increase in GxE research starting in 2003. Between approximately 2000 and 2005, citations were still dominated by stability measures, Finlay Wilkinson and ANOVA, but alternative linear-bilinear models like AMMI, GGE, SHMM, and SREG models (Crossa, 2012) increased in popularity. After 2005, the impact of mixed models and crop growth models rapidly grew, to become the dominant category in recent years.

Recently published papers are less likely to have a large number of citations, making it difficult to predict their impact. However, the number of methodological papers that proposes a certain class of models gives an indication of the direction current GxE research is taking and about the possible future impact of these model types. The rapid increase in the number of citations obtained by mixed models and crop growth models (Figure 2), together with the large number of papers proposing these methods that were published in the last three years (Figure 3), suggest that in the near future GxE methods will rely less on linear-bilinear models and more on mixed models, crop growth models and Bayesian models. Therefore, our choice in GxE courses for mixed models with bilinear and factorial regression terms for GxE seems to be a good reflection of the current trends in the literature.

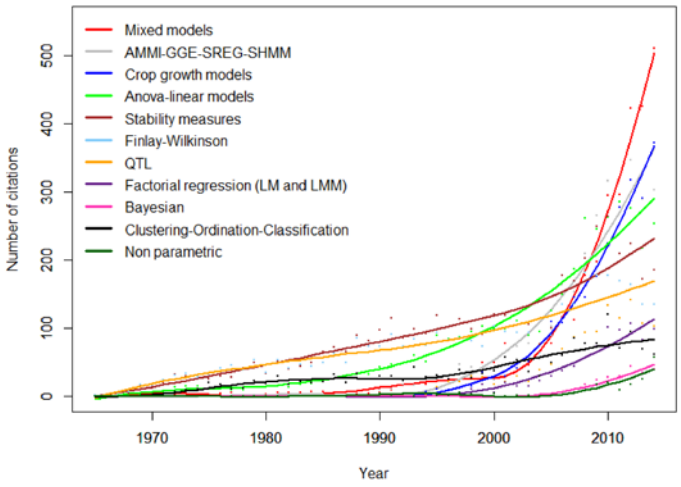


Figure 2. Number of citations per year between 1965 and 2014 of 175 highly cited methodological GxE papers, classified by the model that was used. Lines show the fit of a quadratic smoothing spline with three knots.

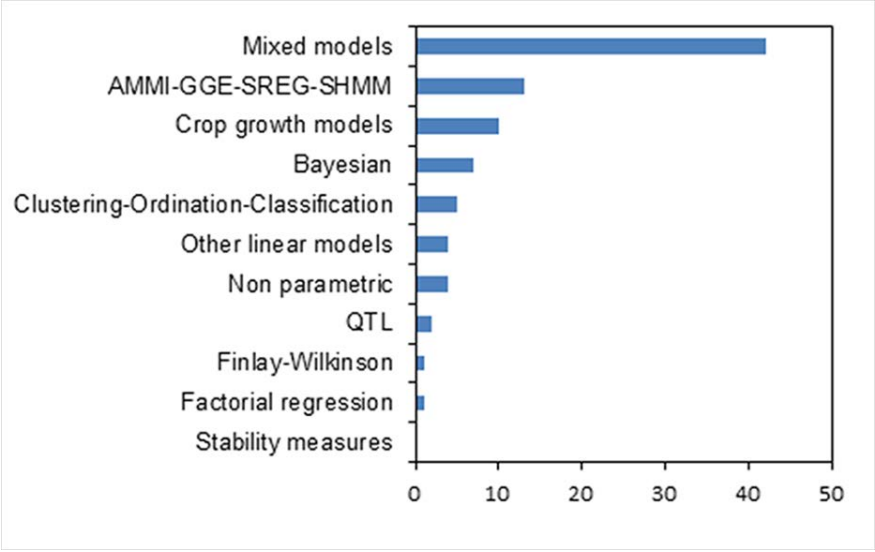


Figure 3. Number of papers published between 2013 and 2015 for several types of analysis on Genotype x Environment interaction.

Acknowledgements

This paper was written as a contribution to the Integrated Breeding Platform, https://www.integratedbreeding.net/ , and as part of the EU DROPS project, FP7-KBBE-2009-3, contract number 244374.

Chapter 9

General Discussion

9.1. Introduction

A major objective of plant breeders is to create and identify genotypes that are well-adapted to the target population of environments (TPE, Cooper & Hammer, 1996). The TPE corresponds to the future growing conditions in which the varieties produced by a breeding program will be grown (Chapter 8, Comstock & Moll, 1963; Cooper & Hammer, 1996; Cooper *et al.*, 2014a). All possible genotypes that could be considered as selection candidates for a specific TPE are said to belong to the target population of genotypes, TPG (Jannink et al. 2010; Schulz-Streeck et al. 2012; Albrecht et al. 2014). A well-adapted genotype can be described as having a better performance than a reference genotype in an environmental range that spans part or whole of the TPE (Chapter 8, van Eeuwijk *et al.*, 2016). The changes in trait configurations in relation to adaptation patterns for the TPG across the TPE over a growing season are referred to as 'the adaptation landscape' (Chapman et al. 2003; Cooper 2004; Hammer et al. 2005; Messina et al. 2011; Cooper et al. 2014b, a; Technow et al. 2015). The adaptation landscape is a further elaboration of the 'fitness landscape', originally proposed in ecology, which relates phenotypes to fitness in a more static way (Wright 1932, 1984; Kauffman 1993). In the landscape metaphor, elevations represent regions with high fitness, and depressions represent regions of low fitness (Gavrilets 2004). In the context of plant breeding, high fitness can be interpreted as a genotype having a high probability of being selected, often equivalent to having a high yield (Cooper 2004).

The identification of genotypes that are well-adapted to the TPE depends critically on the quality of phenotype predictions. These predictions can be made with models that consider a mixture of statistical, genetic and physiological elements. The general objective of this thesis was to propose and evaluate models to predict complex traits across multiple environments. To achieve this goal, we used statistical models like single locus genome-wide association (GWA) models (Chapters 2, 3, 6 and 7), and multi-locus GWA models (Chapter 6) and genomic prediction models (Chapters 2, 3, 5 and 7). We also discussed and used crop growth models (APSIM-wheat, www.apsim.info) and combined crop growth models with genomic prediction and GWA models (Chapters 2 and 7).

Crop growth models like APSIM are useful to characterize the dynamics of traits over time. These traits can be classified into basic traits and intermediate traits, as we did in Chapter 7. We defined basic traits as primary physiological mechanisms that are relatively insensitive to the environment and that allow the integration and modulation of environmental information over time. These basic traits correspond often to crop growth model parameters. Examples of basic traits are sensitivity to photoperiod or water uptake capacity. For each genotype, we assumed the basic traits to be constant over time and across

environments and to be regulated by a limited number of QTLs. In contrast to basic traits that are environment insensitive, we defined intermediate traits as moderately environment sensitive and reflecting a partial integration of a series of basic genotype dependent traits with environmental inputs. An example of an intermediate trait is biomass, which can be measured at multiple times and in multiple environments. A similar distinction between basic and intermediate traits was proposed by Hammer et al. (2016), who define basic traits (crop growth model parameters) as having a short phenotypic distance from their genetic basis because they scale directly from a molecular level to organism. On the other hand, intermediate traits result from the interplay of a larger number of processes, showing a larger level of integration. Intermediate traits can be described as having a larger phenotypic distance from their genetic basis (Hammer et al. 2016). In a modelling context, the number of traits and the degree of detail used to model the trait hierarchy is also described as 'granularity' (Hammer et al. 2006).

In Chapter 8, we provide an overview of concepts and models for genotype adaptation, presented in form of a schedule for a course about GxE for plant breeding students. In the following sections of this Chapter, we will discuss the advantages and challenges of the most important model classes used in this thesis, how these modelling approaches can be implemented given the different sources of genotypic and phenotypic data, and how to design statistical genetic studies to achieve an improved accuracy in phenotype prediction.

9.2. Models for phenotype prediction

9.2.1. QTL models

QTL models aim at identifying genomic regions that are responsible for phenotypic variation in the trait of interest. QTL models can be used for bi-parental populations (Rebai et al. 1995; Lynch and Walsh 1998), multi-parental populations (Cavanagh et al. 2008; McMullen et al. 2009; Garin et al. 2017) or for diversity panels (McCarthy et al. 2008; Huang et al. 2010). QTL mapping is a technique for identifying genomic regions that contain causal genes and is suitable for traits regulated by a small number of loci with relatively large effect size (Kraakman et al. 2004; Malosetti et al. 2007; Donnelly 2008; Pasam et al. 2012; Varshney et al. 2012; Hoffman 2013). The main difference between QTL detection in bi-parental populations and diversity panels is that, in bi-parental populations, QTL detection relies on genetic linkage and segregation of alleles within the family. In contrast, QTL detection in diversity panels, or genome-wide association scans/studies (GWAS), correlates markers with phenotypes across a population (based on linkage disequilibrium, Korte & Farlow, 2013).

In this thesis, we use diversity panels for barley, maize, rice and wheat. Therefore, the QTL results that we present in Chapters 3 and 6 are exclusively examples of GWAS. In both of these Chapters, GWAS showed to be a more promising technique for traits that have a simple genetic basis, like flowering time, compared to grain yield or grain weight, that have a more complex genetic basis. Although GWAS allows overcoming the limitations given by the reduced allelic diversity and the low recombination rate typical for QTL mapping in bi-parental populations, it introduces a number of other challenges (Korte and Farlow 2013; Vilhjálmsson and Nordborg 2013; Xu et al. 2017).

The first challenge is the confounding between the causal loci and the genetic background (Vilhjálmsson and Nordborg 2013). In diversity panels, genotypes are, in general, not equally related, unlike in bi-parental families. Genotypic relationships in diversity panels are typically shaped by domestication, natural and artificial selection and drift (Ross-Ibarra et al. 2007; Flint-Garcia 2013). The outcome of these processes is typically modelled by terms for population structure, complemented by terms for more recent ancestral relationships. Population structure can be represented by a small number of principal components of the genotype by markers matrix. More recent ancestral relationships result in a small number of individuals that are very closely related to each other and unrelated to the rest. Recent ancestral relationships cannot be captured by a small number of principal components. They are modelled via a relationship (kinship) matrix (Hoffman 2013).

In this thesis, we used two different types of correction for genetic relatedness; the first one is by modelling the residual genetic variance via the relationship matrix that accounts for the similarity between individuals (Listgarten et al. 2011; Yang et al. 2014, Chapters 3 and 7). The second type of correction that we used was by including the principal components of the relationship matrix as fixed effects (Chapter 6, Price *et al.*, 2006; Listgarten *et al.*, 2011; Tucker *et al.*, 2014). The main difference between these two methods is that when using the full relationship matrix to model the covariance between pairs of individuals, we account for spurious associations due to the structure and the genetic relatedness, whereas when using the first principal components (like in Chapter 6), we mostly correct for the population structure. A hybrid model, including fixed principal components and a relationship matrix (kinship), has been proposed (Zhang and Pan 2015). Although such a model seems useful to reduce the number of false positives, it contains a certain redundancy because the same information is used to model the fixed and the random terms. Besides, fitting both model terms resulted to be computationally prohibitive. For that reason, we used the first five principal components, plus genomic control a posteriori to correct for inflation that we did not capture by the principal components (Devlin and Roeder 1999). Using the principal components to correct for inflation has been shown to be highly effective, reducing the false

positives and false negatives (Price et al. 2006; Tucker et al. 2014). However, it is not immediately clear how this type of correction works in multi-environment models. Further research is needed to better understand how to account for genetic relatedness when information from multiple environments is combined.

A second issue that deserved further attention are the methods to characterize population structure. Population structure is heterogeneous along the genome, as shown for the Fst variation in barley (Russell et al. 2016), and rare variants can be typically more stratified than common variants (Mathieson and McVean 2012). The relationship matrix is only an approximation of the relationship between genotypes, averaged across the whole genome. For that reason, existing methods like the incorporation of the relationship matrix might not provide the adequate genome-wide correction for genetic relatedness. One alternative approach to overcome the loss of power due to the correction for genetic relatedness and the excess of false positives without using an explicit model term for correction is to calculate a relationship with only those SNPs that show the strongest linear correlation with the trait (Listgarten et al. 2011; Lippert et al. 2013). By conditioning on SNPs that are related to the phenotype, we reduce the noise in the assessment of the association. In Chapter 6, we followed a forward selection procedure to define a multi-locus haplotype model. To add additional SNPs to the multi-locus model, we conditioned on previously selected loci with a large effect, in line with what was proposed in Listgarten *et al.*, (2011) and Yang *et al.*, (2014).

The third challenge for GWAS is that rare variants might be in strong association with many other rare variants along the genome, creating spurious associations between genotypes and phenotypes (Dickson et al. 2010). This is exactly what we observed in the EU-Whealbi-barley collection, where QTLs that are far apart in the genome and even on different chromosomes were correlated. For that reason, there was no unique solution to the identification of a multi-locus model and many multi-locus combinations would provide similar results in terms of explained phenotypic variance. One strategy to reduce this collinearity problem is by simultaneous estimation of genetic variance explained by all SNPs (Bayes-R, Moser et al. 2015). In this technique, only those loci with the largest effect, given the other loci, are selected as QTLs.

Although there are several techniques to reduce the problem of spurious associations in GWAS panels, the complexity of their population structure, especially in such diverse collections like the EU-Whealbi-barley, reduces the chances of finding causal SNPs. In spite of this technical challenge, GWAS and diversity collections are still very useful as a first approach to narrow down the search for the causative SNPs. The second step, after

identifying the largest QTLs in the GWAS, would be to design crosses between genotypes carrying contrasting alleles at those most promising genomic regions. In diversity panels, it is likely that more than one allele is present and therefore multi-parent populations like MAGIC would be preferable for such crosses (Cavanagh et al. 2008; Sannemann et al. 2015).

9.2.2. Genomic prediction

Genomic prediction models aim at predicting genotypic performance, without focusing on the identification of potentially causal loci underlying the trait of interest. For that reason, the most important difference between QTL and genomic prediction models is that in the latter, no explicit SNP selection step is applied (Meuwissen et al. 2001). Prediction quality can be assessed by the prediction accuracy, which corresponds to the correlation between the predictions and the true genetic value of the individuals (Meuwissen et al. 2001), or by the predictive ability, which corresponds to the correlation between the predicted and the observed phenotypic values (Burgueño et al. 2012). As a large number of loci are included simultaneously, genomic prediction delivers higher prediction quality than GWAS. For that reason, complex traits like grain yield or biomass, benefit most from switching from a QTL model to a genomic prediction model.

In this thesis, we used linear models that account for additive marker effects, like QTL and GBLUP models, or a combination of both (QGBLUP, Chapter 3). For all traits and crops that we analysed, models considering a larger number of SNPs, like GBLUP and QGBLUP, always had a larger predictive ability, compared to the QTL model. The phenotype of complex traits often results from epistatic interactions, in addition to the additive effects (McKinney and Pajewski 2012; Langer et al. 2014). Two-locus epistatic effects can be modelled via the adaptive LASSO (Wang et al. 2011). When multiple epistatic interactions occur, the use of Reproducible Kernel Hilbert Space models is a convenient approach (RKHS, Gianola and van Kaam 2008; de los Campos et al. 2009). The RKHS model is a non-parametric model that allows for additive by additive epistatic interactions of various degree (Jiang and Reif 2015). In Chapter 3, we used a RKHS model to predict several traits in wheat, maize and rice and compared it with QTL and GBLUP models. The RKHS model led to larger predictive ability than the other models, coinciding with previous evidence showing that models with non-linear components, like in RKHS and neural networks, consistently had larger accuracy than linear models (Crossa et al. 2010; González-Camacho et al. 2012).

Modelling traits simultaneously allows to achieve larger genomic prediction accuracy for the target trait, compared to single-trait genomic prediction (Dekkers 2007; Jia and Jannink 2012; Alimi 2016; Biscarini et al. 2017; Sun et al. 2017). The benefit from multi-trait

genomic prediction using the target and intermediate traits simultaneously, relative to single trait prediction using the target trait exclusively depends on the trait correlations and trait heritability (Walsh and Lynch; Jia and Jannink 2012), in the same way as discussed in Chapter 4 for the correlated selection response when measuring the same trait in two different regions (Atlin et al. 2000; Piepho and Möhring 2005).

$$\frac{CR}{DR} = r_g \sqrt{\frac{H^2_{intermediate/basic}}{H^2_{target}}} \qquad (1)$$

In Equation (1), response to selection when selecting on the correlated (intermediate or basic) trait, *CR*, will be larger than direct selection on the target trait, *DR*, when the correlation between both traits is large and when the heritability of the intermediate or basic trait is larger than the heritability of the target trait. Basic traits, in general, tend to have a larger heritability than intermediate or target traits. However, the correlation between basic and target traits is lower than the correlation between intermediate and target traits, because the phenotypic distance between basic and target trait is necessarily bigger than the phenotypic distance between intermediate and target trait. Assessing a priori whether it will be convenient to do direct selection on a target trait or to incorporate additional basic and intermediate traits in the phenotype prediction process is challenging. In Chapter 7, we proposed a flexible modelling approach that allows evaluating the potential of basic and intermediate traits to improve prediction accuracy for the target trait.

Different models for multi-trait prediction can be used. In Chapter 7, we used a compound symmetry variance-covariance structure, giving the same weight to yield and biomass measurements. An alternative approach would be to use selection indices, giving a larger weight to the target trait than to the intermediate traits (Hazel 1943; Lande and Thompson 1990; Dekkers 2007). The use of indices may avoid the contamination of the target trait with error from the intermediate trait, avoiding the reduction in prediction accuracy that we observed for some cases in Chapter 7.

9.2.3. Combining crop growth models and QTL or genomic prediction models

In this thesis, we also discussed and evaluated a third modelling strategy for phenotypic prediction; the combination of statistical and crop growth models (Chapters 2 and 7). Crop growth models can range from a simple factorial regression model with QTLs modulating the genotypic sensitivity to the environment, along with more elaborate models like genotype specific growth curves following a logistic model, up to a system of equations that modulates

the genotypic responses over time, like in APSIM. Examples of factorial regression-type of models can be seen in Malosetti et al. (2004) and Boer et al. (2007). The factorial regression-type of models presented by Malosetti et al. (2004) and Boer et al. (2007) present the phenotype as a static variable measured at the end of the growing season. A more elaborate example of factorial regression was presented by Millet (2016), who modelled the number of grains in a maize population as a function of explicit environmental covariables, like the temperature during the night, radiation interception and soil water potential, during different growth stages of each genotype. Considering the developmental stages and other physiological characteristics allowed to calculate environmental characterizations that are genotype and environment specific and explain more GxE than classical environmental characterizations that are based on environment specific averages of soil and meteorological information across the growing season (Millet 2016).

If traits are measured at multiple time-points during the growing season, phenotypic values over time are likely to be correlated. For that reason, prediction of phenotypic values over time benefits from explicit modelling of the response at various time points in a repeated measures framework (Macgregor et al. 2005; Lund et al. 2008). A simpler and more effective alternative is to use mathematical functions that describe trait dynamics, also called 'function-valued trait modelling' (Stinchcombe and Kirkpatrick 2012). The combination of QTLs and mathematical functions that models the QTL effects over time is often called 'functional mapping' (Wu and Lin 2006; Li and Sillanpää 2015). Functional mapping allows to characterize the differences in growth dynamics between genotypes, distinguishing between 'permanent QTLs' (expressed equally during the whole growing season), 'early QTLs' (expressed only during the beginning of the growing season), 'late QTLs' (expressed only at the end of the growing season) and 'inverse QTLs' (a change in the trait-increasing allele during the growing season). An example of time-dependent QTL effects was presented for biomass in Chapter 7, Figure 8. In Chapter 7, a number of QTLs regulate APSIM parameters that have a constant value for each genotype over time and across environments. These constant APSIM genotype-dependent parameters are translated into dynamic biomass curves, allowing to distinguish between time-independent QTLs for physiological parameters and time-dependent parameters for biomass and yield.

The advantages of identifying the genetic basis for curve parameters, instead of for single time points, relies on the fact that curve parameters are closer to the biological principles underlying trait dynamics, compared to single time points. A second, and a more statistical advantage is related to the estimation of fewer parameters when considering the trait dynamics as a whole, instead of as single time-points. For that reason, functional mapping has increased statistical power to detect QTLs (Ma et al. 2002; Wu and Lin 2006). For

example, Bac-Molenaar et al. (2015) showed that combining multiple time-points for plant size through an exponential curve for drought responses in Arabidopsis allowed to increase the number of detected QTLs, compared to single time points.

One of the aspects that influences QTL detection power is the goodness of fit of the curve selected to characterize trait dynamics. Common mathematical functions used to describe trait dynamics are the logistic growth function (Chapter 7, Ma *et al.*, 2002; Wu *et al.*, 2002; Malosetti et al., 2006) and exponential curves (Bac-Molenaar et al. 2015). Although logistic or exponential functions fit well most of the biomass dynamics, some traits require a more flexible curve shape. In that respect, the use of Legendre polynomials represents an advantage compared to exponential or logistic curve functions because they allow fitting curves of any shape (Macgregor et al. 2005; Yang et al. 2006; Yang and Xu 2007). A more straightforward approach that also allows a large degree of curve flexibility, are splines (Eilers and Marx 1996; Eilers et al. 2015). One example for the use of splines in the literature can be found in Hurtado et al. (2012), who modelled the dynamics of QTLs for haulm senescence in potato using generalized linear models with genotype-dependent slopes. In the context of genomic prediction, Sun et al. (2017) used splines to model the dynamics of canopy temperature and NDVI, integrating them into a multi-trait genomic prediction model, in a similar way as we did in Chapter 7. We compared the use of logistic functions and splines to characterize biomass dynamics over time (Chapter 7). We observed that functions derived from fitted splines and parameters of a logistic curve showed higher heritability than observations at single time points, thereby proving that the simultaneous modelling of multiple time points helps to reduce the measurement error of high-throughput phenotyping (HTP , Figures 9 and 10, Chapter 7). Spline derived functions had higher heritability than logistic curve parameters because they accommodate better the irregularities in the biomass accumulation. The larger heritability for spline derived functions led to a larger prediction accuracy, indicating the potential of HTP to improve prediction accuracy of the target trait by including information of intermediate traits (Chapter 7). In all the examples presented above, the characterization of trait dynamics and the QTL detection (or genomic prediction) were done in a two-step approach. Estimation of the curve parameters and QTL effects can be done in a one-step approach using hierarchical Bayesian methods, as proposed by Sillanpää et al. (2012) and applied to flowering time in rice by Onogi et al. (2016). The main advantage of a single-step is that it leads to a more flexible estimation of time-independent QTLs and to a larger prediction accuracy than the two-step approach.

Biological organizational levels in genotype to phenotype models can be integrated further by the simultaneous modelling of multiple traits via networks (Cooper 2004). Networks can consider processes belonging to the same level of organization, as genes and

gene expression (Mochida et al. 2011; Torres-Sosa et al. 2012; Liseron-Monfils and Ware 2015), or they can model processes belonging to more than one level of biological organization. Examples of the relationship between QTLs and a single level of phenotypic organization can be seen in Neto et al., (2010) and Alimi (2016). If the aim is to characterize the genetic basis for multiple levels of phenotypic organization, multi-level directed networks are necessary. For example, Wang and van Eeuwijk (2014) and in Wang et al. (2015) used multi-level directed networks to characterize the relationships between QTLs, metabolites and sensory traits in tomato.

Networks are a valuable tool to characterize the dependencies between traits and QTLs. In the directed networks, dependencies are inferred from the conditional QTL genotype probabilities, without considering the time dimension (Neto et al. 2010). Longitudinal networks are a strategy to consider the time dimension for multiple traits. Although longitudinal networks have shown promise in other fields of science (De Vos et al. 2017), they haven't been yet assessed in the context of plant breeding, making them worthwhile for exploration in further research.

An increasingly popular approach to explicitly model the functional relationships between plant physiology and the environment is by using dynamic crop growth models (Yin et al. 2000; Tardieu et al. 2005). There is a large range of crop growth models with different levels of complexity and parameterization. Widely-used crop growth models are APSIM (as we used in Chapter 7, Keating et al. 2003; Holzworth et al. 2014), CERES (Ritchie 1985) and DSSAT (Jones et al. 2003). An overview of the models developed at Wageningen can be seen in Bouman et al. (1996) and van Ittersum et al. (2003). Crop growth models can have different uses in the breeding process; to characterize the TPE, as a decision support tool to assess breeding strategies (Cooper et al. 2002; Messina et al. 2011) and for phenotype prediction (Zheng et al. 2013). In Chapter 7, we used the APSIM crop growth model to characterize the TPE. Here, we expanded on the work previously carried out by Chenu et al. (2011, 2013), through the incorporation of an explicit genetic basis for the APSIM parameters. This genetic basis consisted of exclusively additive effects. However, we did observe that trait dynamics show nonlinearities during the growing season (e.g. Figure 2, Chapter 7). These nonlinearities arise as a result of interactions between physiological mechanisms, called 'physiological epistasis' (Cheverud and Routman 1995; Cooper 2004). We also used APSIM to evaluate the potential of biomass measured with different HTP schedules as a correlated trait to improve yield prediction accuracy. In these approaches, we scale up through the levels of organization, from APSIM parameters that represent basic mechanisms with a constant genetic basis over time and environments, to final yield. These parameters modulate the adaptation of genotypes to the wide range of environmental

conditions, where the main environmental driver is the contrast in water deficit patterns. In this approach, GxE arises as an emerging property resulting from the integration of different levels of organization and over time (Cooper et al. 2002; Cooper 2004; Hammer et al. 2006).

Recently, Technow et al. (2015) proposed an alternative strategy, in which phenotypic observations are available for a target trait that follows from a number of component (intermediate and basic) traits without observations, but with the assumption that the component traits obey an additive model at the QTL level. In this case, component traits are treated as latent variables for prediction that arise from the propagation of the marker effects on yield through the non-linear crop growth model structure. Here, crop growth models were used in an Approximate Bayesian computation (ABC) framework, which replaces the calculation of a likelihood function with a simulation step. In this approach, the combination of crop growth models and genome-wide predictions allows to capture non-additive effects, at the target trait level that cannot be captured by models like GBLUP. A promising application of the same method was recently shown for real maize data, sown in environments contrasting for water deficit patterns (Cooper et al. 2016).

A central aspect for our conclusions about phenotyping strategies to be valid is the assumption that APSIM captures the important properties of the natural system composed by the TPG and the TPE. All (crop growth) models are a simplified abstraction of complex natural systems. Although the biological mechanisms are not necessarily exactly described by the crop growth models, the most relevant aspects of the empirical adaptation landscape (i.e. GxE patterns, trait correlations) should be correctly reproduced by crop growth models to be useful tool to assess plant breeding strategies (Cooper et al. 2002; Hammer et al. 2006). There is a large body of evidence showing that crop growth models are a useful tool to characterize the performance of cropping systems over time. However, none of them has been parameterized in a single set of experiments (Asseng et al. 1998; Van Ittersum et al. 2003; Holzworth et al. 2014). In contrast, for most of them, their structure arose as the combination of individual equations derived from individual experiments, and additional sets of equations (‘modules’) were added a posteriori to the original model structure. Some examples of module additions or modifications to APSIM can be seen in Chenu et al. (2009) Zheng et al. (2013) and Inman-Bamber et al. (2016). Although there is nothing wrong per se in this type of model building process, it doesn't take into account the correlation structure between parameters, often leading to overly complex models. The high model complexity adds an additional challenge when parameterizing them for specific genotypes. One example of alternative approach was recently presented by Lamsal et al. (2017), who used image analysis from historical field trials for a simultaneous estimation of crop growth model parameters and environmental quality in a soybean breeding programme. The modelling

shown by Lamsal et al. (2017) is a good example on how the simultaneous modelling of phenotypes, their proxies obtained through image analysis, crop dynamics and genetics can be used as a valuable tool to predict genotype adaptation and assist breeding programmes.

9.3. High-throughput genotyping

The availability of high-throughput genotyping, particularly sequencing techniques, allows to increase the chances of integrating the loci causing the phenotypic differences, into the genetic analysis (van Binsbergen et al. 2016; Uauy 2017). In Chapter 6, we used exome sequences for a large and diverse barley panel belonging to the EU-Whealbi collection (http://www.whealbi.eu/). Exome sequencing is a strategy to reduce genomic complexity, dramatically reducing the sequencing load and cost (Winfield et al. 2012; Uauy 2017). With the barley exome sequences, we constructed haplotypes within the genes in the QTL regions, using single SNPs within annotated genes. The construction of haplotypes is a strategy to get closer to the causal variants related to the phenotype by defining alleles that provide informative contrasts. These informative contrasts contribute to shorten the phenotypic distance between genes and phenotypes by a redefinition of the genetic basis to make it more relevant to the phenotype and allow a more direct scaling from the genetic level to the phenotypic level (Hammer et al. 2016).

The use of haplotypes instead of single SNPs led to larger effects and the multi-haplotype model showed a larger percentage of explained variance, compared to the multi-SNP model (Table 3, Chapter 6). We did not do genomic prediction. However, previous evidence suggests that the use of haplotypes in a genomic prediction model might lead to increased prediction accuracy (Hayes et al. 2007). The gain in terms of explained variance or prediction accuracy when using haplotypes instead of single SNPs depends on the LD in the haplotype region. If neighbouring markers are in high LD, single SNP analysis is likely to produce similar results for SNPs and haplotypes (Calus et al. 2008). However, in diversity panels like the EU-Whealbi collection, the LD decays rather quickly, especially at the end of the chromosomes, where most of the QTLs were detected (Figure 4, Chapter 6), justifying the evaluation of haplotype effects on the phenotype. Genotypic information could have been integrated into an even higher level of organization, assessing the amino acidic composition derived from SNP differences between haplotypes. The comparison of haplotypes in terms of how they are translated into amino acids represents yet another opportunity to shorten the phenotypic distance between the final trait and its genetic basis.

The availability of (exome) sequence information for such a diverse collection as the EU-Whealbi-barley collection is a valuable resource that will facilitate the integration of novel

alleles into breeding programmes via crosses between parents with complementary alleles. The progeny can then be genotypically characterized at a lower intensity and imputation can be used to improve marker coverage (van Binsbergen et al. 2014, 2016). The EU-Whealbi collection possessed phenotypic information for traits that have been important during the barley domestication and adaptation process. The availability of phenotypic information adds value to the exome sequence information, allowing for a targeted characterization of the genetic diversity of this panel (Ross-Ibarra et al. 2007).

9.4. High-throughput phenotyping

After the rapid advancement of genotyping technologies, phenotyping was considered until recently to be a limiting factor in plant breeding. During the last years, a large variety of phenotyping techniques has been developed (Furbank and Tester 2011; White et al. 2012; Chapman et al. 2014). Phenotyping techniques can be applied in facilities under controlled conditions, generally not transportable (platforms, Junker et al. 2015; Neumann et al. 2015) or can be based on mobile devices for field conditions (e.g. drones, helicopters, mobile devices, Chapman et al. 2014; Deery et al. 2014, 2016). In general, there is no clear strategy about which types of facilities can best be used in which breeding situations. Here, we propose that the hierarchy of the levels of organization from genotype to phenotype might provide useful insight on how to make the best use of phenotyping facilities when modelling from genotype to phenotype.

In Chapter 7, we defined basic traits as referring to those mechanisms that modulate the genotypic response to the environment at an elementary level and represented them in the form of crop growth model parameters (little or no GxE). We defined intermediate traits as showing a partial integration of the genotypic response to the environment and as reflecting intermediate steps between the basic traits and a target trait (yield). Intermediate traits, in general, show a larger GxE, compared to basic traits. The notion of differences in scale (basic traits with short phenotypic distance to the genetic basis vs. intermediate traits with larger phenotypic distance to the genetic basis) is useful to organize the phenotyping strategy. We propose the convenience of using platforms, greenhouses or facilities with more controlled conditions for a detailed phenotyping of basic traits, and field phenotyping for the more integrative traits.

A disadvantage of using facilities with controlled conditions to phenotype intermediate traits is that genotype by experimental condition interaction can be large (e.g. biomass ranking in the greenhouse might not correlate well with biomass ranking in the field). Therefore, intermediate traits measured under controlled conditions might fail to predict the

performance in the field. Examples for the use of platforms/controlled conditions to characterize basic traits are wheat early vigour measured in the greenhouse (Duan et al., 2016), the root angle in maize and sorghum measured in greenhouse pots as a trait related to water uptake (Singh et al. 2010), or the sensitivity to photoperiod, vernalization and earliness per se in wheat measured in controlled conditions for photoperiod and temperature (Zheng et al. 2013; Sukumaran et al. 2016). Examples for intermediate traits in field conditions are airborne measurements for wheat NDVI and canopy temperature (Deery et al. 2016; Rutkoski et al. 2016).

Data from field imaging for integrative traits and from platforms for basic traits /crop growth models parameters can be used for predicting target traits. Different approaches are possible. A first type of phenotyping network would rely on a central location to intensively phenotype basic traits in platforms, with some additional phenotyping of integrative traits in the field. As basic traits are commonly difficult to measure, phenotyping could be made on few genotypes, predicting the rest of the TPG (Pauli et al. 2016). This strategy is also highly attractive for genomic prediction, where the expensive basic trait is measured on a wide subset of genotypes and then the rest of the population can be predicted from a HTG data set. This is a way to connect HTP and HTG. Under this scheme, prediction of the target trait would require a good articulation of statistical and crop growth models.

The role of statistical models being the prediction of crop growth model parameters from the experiment in controlled conditions and the role of crop growth models breaching the long non-linear path that connects basic traits, intermediate traits, with eventually the target trait. A second, more statistical approach, would employ the phenotyping network to measure integrative traits with equal detail in all environments, and directly predict the target trait by integrating the information in multi-trait prediction models (e.g. via indices). The recommendation of using one or another modelling approach for subsets of traits and phenotyping facilities is in direct connection with the notion of phenotypic distances proposed by Hammer et al. (2016) and with correlated response to selection theory (Falconer and Mackay 1996). If the distances between genotype and phenotype are short, statistical models are in general sufficient for good prediction quality. One example of prediction for the target trait from an intermediate trait that is at a short phenotypic distance was shown by Montesinos-López et al. (2017), who predicted yield from NDVI canopy measurements using Bayesian models for GxE. If traits are at a long phenotypic distance, the combination of statistical and crop growth models might be preferable.

9.5. Structure of the TPG and TPE in relation to multi-environment predictions

9.5.1. Structure of the TPG

The identification of well-adapted genotypes typically involves the benchmarking of the yield of new genotypes against yield of existing genotypes using multi-environment trials (Hammer *et al.*, 2014; van Eeuwijk *et al.*, 2016, Chapter 8). Breeders have access to a sample from the TPG, which corresponds to the selection candidates for the breeding programme. This sample of genotypes (or part of it) can be regarded as a calibration set for prediction models when both phenotypic and marker data are available. The selection candidates usually show heterogeneous genetic similarities that arose through the crossing schemes and the selection process. To obtain large prediction accuracy, it is convenient to explicitly account for this heterogeneity in genetic similarities. This can be done via strategies for training set construction methods and via the selection of the prediction model. Different methods for training set construction have been proposed; random sampling, stratified sampling and uniform sampling. Conventionally, genomic prediction literature uses random sampling as a strategy to split the calibration set into a training and a validation set (Burgueño et al. 2012; Schulz-Streeck et al. 2012). Although stratified sampling shows larger prediction accuracy than random sampling (Albrecht et al. 2014; Guo et al. 2014), better results have been observed when uniformly covering the genetic space of the TPG (Chapter 3, Bustos-Korts *et al.*, 2016). Similar approaches to represent well the TPG were previously shown by (Rincent et al. 2012; Isidro et al. 2015), who proposed that predictive ability can be improved if genotypes in the training set are chosen in such a way that the precision of the contrasts between each genotype in the validation set and the mean of the calibration set is maximized. Prediction accuracy in structured populations can also be improved by allowing for sub-population specific effects (Schulz-Streeck et al. 2012; Lehermeier et al. 2015; de los Campos et al. 2015). In Chapter 3, we evaluated models with subpopulation specific-effects (not shown). However, they did not provide a clear advantage, probably because the populations we evaluated did not have a very strong structure. The EU-Whealbi-barley panel does show a clear structure driven by the row types, which suggests that it would be interesting to evaluate models allowing for subpopulation specific effects.

9.5.2. Structure of the TPE

The environmental conditions used during the selection process can profoundly affect allele frequency in breeding populations and the stress tolerance of elite commercial products (Campos et al. 2004). Therefore, a central question for testing of genotypes across multiple environments is how to design the network of multi-environment trials in such a way that the

selection environments correspond well to the range of growing conditions in the TPE. The identification of selection environments involves two problems; first, how to define which are the expected environmental conditions for the TPE and, second, how to identify which locations in the present year have a genotypic ranking that predicts well the genotypic ranking in the TPE. The structure of the TPE can be described focusing on the genotype ranking across year-location combinations or focusing on the explicit environment quality (that might lead to changes in genotypic ranking). If we focus on the genotypic ranking, we refer to 'mega-environments', which are defined as groups of trial locations that show a reduced number of crossovers, compared to the locations without grouping (Cornelius et al. 1993; Crossa and Cornelius 1997; Crossa et al. 2004; Atlin et al. 2011). In contrast, 'environment types' is used for a more explicit representation of the environmental quality (e.g. water deficit patterns, Cooper and Fox 1996; Chapman et al. 2000, 2002; Chenu et al. 2011; Cooper et al. 2014b). Of course, mega-environments and environment types commonly overlap to a large extent (Annicchiarico et al. 2005).

There is a large body of literature using multiplicative models like AMMI and GGE to identify mega-environments based on historical yield data across multiple locations and years. To increase response to selection, breeders try to identify repeatable patterns across environments. Those repeatable patterns are commonly associated with the growing conditions at particular locations (e.g. soil, irrigation, Atlin et al. 2011). The grouping patterns of locations and years can be inspected in a biplot (Gauch and Zobel 1997; Gauch 2006), but biplot inspection becomes a difficult and ineffective task when the number of years is large or when there is a large genotype-by-location-by-year interaction. Besides, historical multi-environment data are unbalanced across years and commonly used GxE models like AMMI require balanced data. Mixed models with a factor analytic structure can handle unbalanced data, but they require considerable computational and statistical skills to work with them. In Chapter 4, we proposed a strategy based on statistical tools that are commonly accessible to breeders (AMMI model and clustering), but that still delivers comparable results to the more complex mixed model strategy. A similar approach based on clustering of phenotypes of multiple traits was shown by Bassi and Sanchez-Garcia (2017) for the ICARDA durum wheat breeding programme. The downside of a strategy based exclusively on phenotypic data is that it does not allow to identify the explicit causes responsible for the partitioning of environments into groups.

Environments can also be grouped using explicit environmental information. This grouping can be based on purely statistical models or, on the combination of statistical and crop growth models (Chenu et al. 2013; Cooper et al. 2014b, Chapter 7). In Chapter 4, we relate the mega-environments to explicit information for latitude and longitude. However,

the same type of analysis could be done for the thermal gradient from North to South, for example. When the environmental quality is structured, we can benefit from concepts discussed in Chapter 3 for the genetic space of the TPG. As for the TPG, prediction accuracy is likely to benefit from the homogeneous coverage of the environmental conditions of interest for the TPE. For the selection of locations that should represent well the TPE, we propose to use a set of environmental covariables to construct a similarity matrix for locations within each year ('Kinship of environments', Chapter 5). Then, uniform sampling would lead to a list of locations for each year that represent well the TPE. Those locations that are sampled most often across years can be used to do field trials useful to train multi-environment prediction models.

The identification of the environmental drivers for GxE has a large influence on whether it is convenient to optimize the network of testing sites, relying on the natural year-to-year variation to represent the whole range of environmental conditions in the TPE, or whether using managed stress environments would be more convenient. One example where managed stress trials was proposed to be convenient is presented by Rebetzke et al. (2012) for the water deficit patterns in Australia. Field environments in Australia are variable and unpredictable, making it convenient to evaluate the most common water deficit patterns across the largest wheat breeding areas. The final choice for one or the other strategy will depend on how well locations in an average year do represent the whole range of environmental conditions (= how often the same locations would be chosen in the sampling method proposed above).

The second factor influencing the choice of multi-environment or managed-stress trials is how well do we know which are the environmental variables driving GxE and how well can we reproduce them? One strategy to identify which locations represent the whole range of environmental conditions is to use a combination of crop growth models and meteorological information. Crop growth models allow for a more explicit understanding of the underlying causes of the GxE patterns, under the assumption that the crop growth model is sensitive enough to all the meteorological information that is relevant for GxE. For example, Rincent et al. (2017) proposed an optimization criterion to identify those locations that allow for a better prediction of wheat flowering time in other locations. This optimization criterion was applied to wheat flowering time data simulated with the Sirius model (Jamieson et al. 1998) and validated on real data in field trials across France.

A similar approach aiming at improving the selection of environments that will act as predictors, but from a purely statistical perspective, was presented by Heslot et al. (2013b). This paper presents a method to optimize the composition of the training population for

predicting performance in the target population of environments (TPE). This method does not search to identify mega-environments, but it tries to remove the less predictive environments from the set of environments used to train the model (Heslot et al. 2013b).

If the year-to-year variation is large and there is a high risk of not representing well the TPE range, it might be more convenient to rely on managed environment trials. The degree of success of these managed environment trials will largely depend on how well do we know the environmental drivers of the phenotype and on how easily we can reproduce them in a managed environment trial. The easiest example for phenotype prediction across the TPE from traits measured in managed environment trials is probably flowering time. This trait has a known and relatively simple genetic basis, with clear environment drivers (photoperiod, vernalization and thermal time). Flowering time for wheat across the whole wheat belt was successfully predicted from one field trial with four treatments that covered the environmental range of photoperiod and vernalization in a factorial combination (Zheng et al. 2013). Another example was shown for flowering time in barley. In this case, the managed environment trial consisted of climate chambers differing in temperature and photoperiod, to predict the flowering time in field trials across four German locations (Uptmoor et al. 2017). Here, the growing conditions in the chambers and in the field were very different. However, as the chambers allowed for a good-enough estimation of the crop growth model parameters, those parameters were useful to predict yield in the field. Other examples are for maize grain yield under drought in the field by Cooper et al. (2014a), and for drought and N stress (Weber et al. 2012).

9.5.3. Separability and calculation of the covariance structure

Most of the statistical models treat genotypes and environments as separate sources of information that can be modelled independently. The exclusive dependence of the phenotype on either genotype or environment is referred to as separability (Gregorius and Namkoong 1986; Cornelius et al. 1992). Examples or separable model structures are the genotypic and environmental scores in the AMMI model or modelling the variance-covariance between pairs of observations as the Kronecker product of a genotypic variance-covariance matrix and an environmental variance-covariance matrix. The environmental variance-covariance can be estimated from the phenotypic data, or from explicit environmental covariables (Jarquín et al. 2013). The genotypic variance-covariance is commonly calculated from the molecular markers across the whole genome (Astle and Balding 2009), whereas the environmental variance-covariance matrix is commonly modelled with an unstructured model. To facilitate the estimation process, both genotypic and environmental variance-covariance structures can be simplified. Common approximations for the environmental

variance-covariance structure are the factor analytic models (Burgueño et al. 2008; Beeck et al. 2010). Environment categories (e.g. water deficit patterns) could also be used to group environments into a smaller number of classes. These more simple representations of the genotypic or environmental-variance covariance are also called 'compressed' variance-covariance matrices (Zhang et al. 2010; Huang et al. 2010).

Although separability has proven useful to characterize the adaptation patterns of the TPG across the TPE, it neglects the fact that in the same trial, not all genotypes explore the same environmental conditions. Therefore, the effects of genotypes and environments are not always fully separable. Crop growth models have been proposed as an alternative to deal with the non-separability between genotypes and environments (genotype-specific environments). For example, Heslot et al. (2013a) calculated environmental covariables defined for specific environmental stages, following the development of three reference genotypes with different phenology (early, intermediate and late flowering). Another example can be read in Millet et al. (2016) who expressed yield as a function of its underlying components, their genotype-specific sensitivity to the environment and their genotype-specific environmental characterization. These calculations were possible thanks to the use of a set of equations considering genotype development and the relationships between basic traits and the environment, added in a sequential way. An alternative has been recently proposed by Technow et al. (2015) and Cooper et al. (2016) who used crop growth models in an Approximate Bayesian Computation framework. In approaches combining crop growth models and statistical models, GxE arises as an emerging property from the interplay between crop growth model parameters and environmental information. Therefore, the covariance between pairs of observations is not expressed in terms of an explicit variance-covariance structure, but as a function of latent underlying biological processes.

9.6. Using simulations to evaluate breeding strategies

In Chapter 7, we combined statistical-genetic models and the APSIM crop growth model to simulate yield and its underlying traits across a range of Australian environments. We used simulated data to evaluate phenotyping strategies in contrasting environments, in the context of multi-trait genomic prediction. In this section, we will analyse the limitations of our modelling approach and discuss the challenges and opportunities that deserve exploration in further research.

9.6.1. Genetic architecture

In our simulations, we sampled additive effects using copulas with a Gamma marginal distribution. The marginal distribution followed the same shape and rate as the empirical

additive effects for wheat yield and heading date estimated in Australian environments. We attached the additive effects to a random sample of SNPs, generating genotype-dependent APSIM parameters, in a similar way as proposed by Pérez-Enciso et al. (2017). Main difference between our simulations and the one discussed by (Pérez-Enciso et al. 2017) is that we simulated multiple traits simultaneously. The simultaneous modelling of traits and their combination with a crop growth model like APSIM brings in a number of technical challenges that we will discuss below.

The first challenge is to choose, for each APSIM parameter, the allele that will increase the parameter value (trait). From our simulations, we learned that the APSIM parameter correlations are highly sensitive to the way in which additive effects are assigned to the alleles. In single-trait simulations reported in the literature, the trait-increasing alleles are commonly assigned to the most frequent allele, via the incidence matrix (Kizilkaya et al. 2010; Pérez-Enciso and Legarra 2016; Howard et al. 2017). In other simulations, the trait-increasing allele is allocated at random (Pérez-Enciso et al. 2017). In a random allocation, as we did, both alleles have an equal probability of increasing the trait. For a multi-trait simulation, as our APSIM parameters, it is not immediately obvious which allele should increase which parameter. If for example, the most frequent allele would increase all parameters, the parameters would become highly correlated (Figure 1b), even if the correlation between the additive effects for different APSIM parameters is low (Figure 1a). Random allocation of effects keeps the APSIM parameter correlations low. The APSIM parameter correlations can be manipulated at will, by using the same alleles to increase multiple APSIM parameters simultaneously (Figure 1c). This simultaneous allocation of effects imposes some genetic correlation between traits. Another strategy to impose these correlations would be modifying the APSIM algorithms, integrating explicit relationships between traits. Although we did not explicitly study it here, our methodology allows to assess the response of the target trait to genetic correlations in its underlying traits (APSIM parameters). Such a study would allow to distinguish between trait correlations that arise as a product of physiological processes, from those that are purely genetic. If trait correlations are largely driven by genetics, they could more easily be broken by directed crosses, than physiologically-induced trait correlations. In multi-trait simulations, it is also not clear how an allele that increases parameter values is translated into an additive effect for yield. The final effect on yield of an allele modifying an APSIM parameter depends on the effect of that allele on the parameter, the allele frequency in the population, the sensitivity of yield to the changes in the parameter value, and on the combinations of parameter values from each genotype. Combinations of parameter values indirectly create non-additive effects. A more explicit simulation of additive effects was recently proposed by (Howard et al. 2017), who used solely statistical models to characterize the yield surface of a population, as a function

of additive and epistatic effects. In the example by (Howard et al. 2017), epistasis arises as a purely genetic phenomenon, whereas in the combination of statistical and crop growth models, epistasis can arise as a product of the genetics and/or of the physiological processes.

We also learned that the way of selecting the SNPs that are going to carry the additive effects influences the correlation (confounding) between traits (APSIM parameters) and population structure. If causal SNPs are sampled at random, (equivalent to the neutral genetic architecture proposed by (Pérez-Enciso et al. 2017)), the correlation between APSIM parameters and structure will vary between one realization of the TPG and the next, depending on the Fst of the SNPs that were used to assign the largest additive effects to. Figure 2, upper panels shows how random sampling of SNPs for assigning additive effects to changes the correlation between APSIM parameters and population structure; in some realizations, parameters are more correlated to the first principal component, whereas in others, parameters are more correlated to the second principal component. If selective SNPs are used to assign the largest additive effects to, the APSIM parameters become more correlated to the population structure, but the correlation is more homogeneous across sampling events because we restrict the possible SNPs that can carry additive effects (Figure 2, middle panels). The larger correlations with population structure reduce the parameter space explored by the population, as some parameter combinations might not be present. The opposite will occur if we attach additive effects to SNPs with low Fst; the APSIM parameters will show a lower correlation with population structure and the APSIM parameter space will be explored more homogeneously (Figure 2, lower panels).

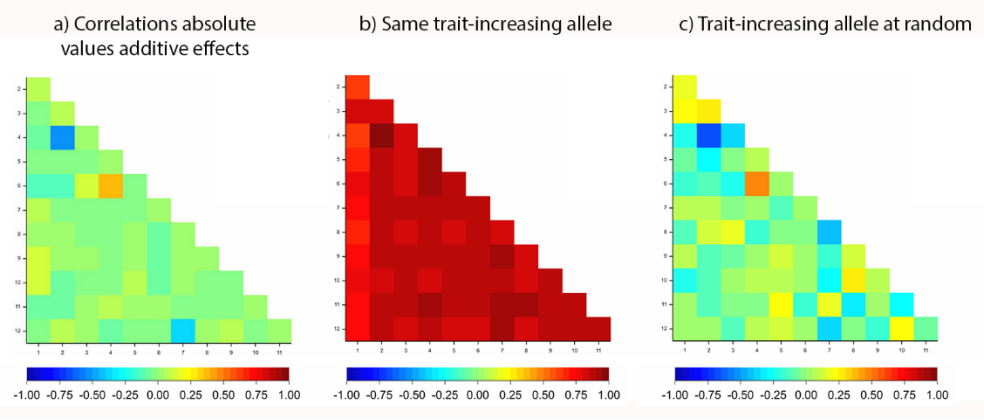


Figure 1. (a) Correlations between the additive effects sampled with copulas with a Gamma marginal, (b) correlations of the APSIM parameter values when the trait-increasing allele is the same for all parameters and (c) correlations of the APSIM parameter values when the trait-increasing allele differs between parameters (random allocation of the trait-increasing allele, except for the correlated parameters).

The way in which the simulation covers the parameter space can modify the GxE patterns for the final trait. Usually, it is the combination of parameter values, more than the individual parameter values which cause GxE to arise. The same holds for real biological traits, where an interaction between environmental conditions, plant phenology and other traits is observed (Fleury et al. 2010). When such an interaction occurs, the effects of QTLs underling traits like water-soluble carbohydrate or osmotic potential on yield depend on flowering time. In our simulations, we chose an intermediate situation, with random SNP selection and low to intermediate correlations between APSIM parameters and population structure. However, these settings can be modified to match the selection history of the TPG. For example, if a TPG, product of the selection history, is composed of two subpopulations differing in flowering time habit, the APSIM parameters for flowering time would become highly correlated with population structure. For our modelling framework to be a realistic characterization of the adaptation landscape, the forces driving the selection process should be taken into account. In Figure 3 and Table 1, we show the AMMI biplots and variance components for a number of realizations of the simulations, differing in the type of SNPs that were used to assign the additive effects to (additive effects allocated to random, selective or neutral SNPs). In this exploratory analysis, it seemed that GxE patterns for yield are quite stable to the changes in genetic architecture and sampling effects. However, here we used the same parameter ranges for all sampling realizations. We based ourselves on information from the literature to set these ranges, but a more detailed experimental confirmation would be needed to make sure that we are fully representing the range of variation present in the Australian TPG.

Genetic architecture for the APSIM parameters did not seem to impact much the shape of the GxE patterns (Figure 3 and Table 1). However, it did have a clear impact on the effective genetic architecture for yield, the SNPs that appear to affect yield. In Figure 4, we show a small example of GWAS in four environments, using a population with a selective genetic architecture and another population with a neutral genetic architecture. Both populations differ in their level of confounding between the causal SNPs and the population structure. The selective genetic architecture led to a stronger confounding than the neutral architecture (Figure 2). Therefore, the QTLs in the neutral architecture showed a larger significance than those in the selective architecture. Although we did not evaluate it here, we expect that a genomic prediction model would show lower prediction accuracy when traits follow a neutral genetic architecture because, in that case, the causal SNPs are not as highly correlated to the population structure (Kinship matrix) as in the selective genetic architecture. The convenience of using either QTL-based models or genome-wide prediction can also be assessed via a combined statistical-genetics and crop growth modelling approach, illustrating the diverse use of the techniques evaluated in this thesis.

The similarity in GxE patterns across realizations allows to suggest that the modelling approach combining statistical-genetic models and APSIM is a promising way to characterize the adaptation landscape across the TPE. However, here we presented only a limited number of sampling realizations. A more intensive sampling scheme and a more formal comparison of the environmental variance-covariance structure would be needed to arrive at solid conclusions. We would also need to evaluate different sets of genotypes, or crossing schemes, as proposed by (Podlich and Cooper 1998) to allow for a broader characterization of the TPG.

9.6.2. Sensitivity analysis of crop growth models and biological sensitivity

In our simulation, APSIM parameters differed in the amount of variation (% of change across the populations with respect to the default parameters value) and followed a normal distribution. These are fundamental differences between our study and a sensitivity analysis. We restricted the parameter space to those combinations that are likely to occur given the constraints imposed by population genetics. This means that, as extreme values are rare in the population, genotypes having extreme values for many parameters simultaneously are less frequent. This restricts the possible outcomes of the GxE patterns. In contrast, sensitivity analysis homogeneously samples the whole parameter space, allowing to estimate marginal parameter effects. In a sensitivity analysis, all parameters also vary in the same amount, usually expressed as a percentage of the default value. Sensitivity analysis is useful to identify the most impactful parameters, as shown by (Casadebaig et al. 2016), but for biological interpretation one should be cautious because some parameter ranges and parameter combinations may be biologically implausible.

Crop growth models like APSIM have shown to be an excellent tool to assess management practices across environments (Asseng et al. 1998; Chenu et al. 2011; Chauhan and Rachaputi 2014). As we illustrated in this thesis, they also allow to evaluate phenotyping strategies, to evaluate the trait correlations and characterize the GxE patters shown by samples of the TPG across the TPE. However, the parameterization of crop growth models should be eventually reconsidered, customizing it to breeding and physiological research (Chenu et al. 2009; Hammer et al. 2010; Zheng et al. 2013). Another reason to re-evaluate the structure of crop growth models like APSIM is that most of the traits considered in its parameterization are extremely hard to phenotype. Thus, the usage of APSIM in breeding programmes would be greatly facilitated if its genotype-dependent parameters were oriented towards proxies of traits with easy phenotyping. An alternative would be to use a simplified structure of the crop growth models; the emulators. Emulators are statistical models that capture the essence of a highly elaborate crop growth model like APSIM (Stanfill et al. 2015). Another approach is to focus only on those parts of the crop growth model that are relevant to a small set of environments, as shown by (Millet 2016).

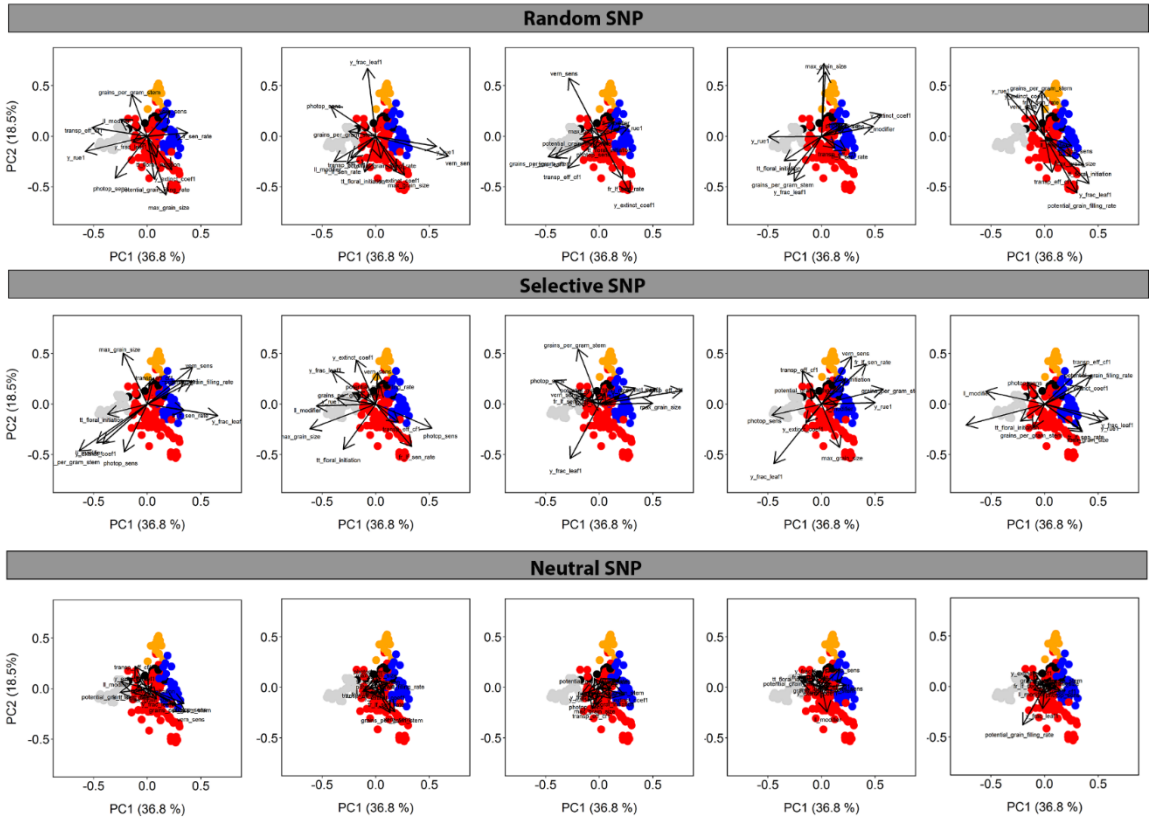


Figure 2 Population structure as revealed by the first two principal components extracted from the matrix of marker scores. Directions of greatest change for a set of physiological parameters simulated using random SNPs (upper panels), selective SNPs (Fst>0.5, middle panels) and neutral SNPs (Fst<0.1, lower panels) have been projected on the biplots to help in interpretation. The length of the arrows indicating each physiological parameter is proportional to the amount of variation explained by the kinship principal components for that parameter.

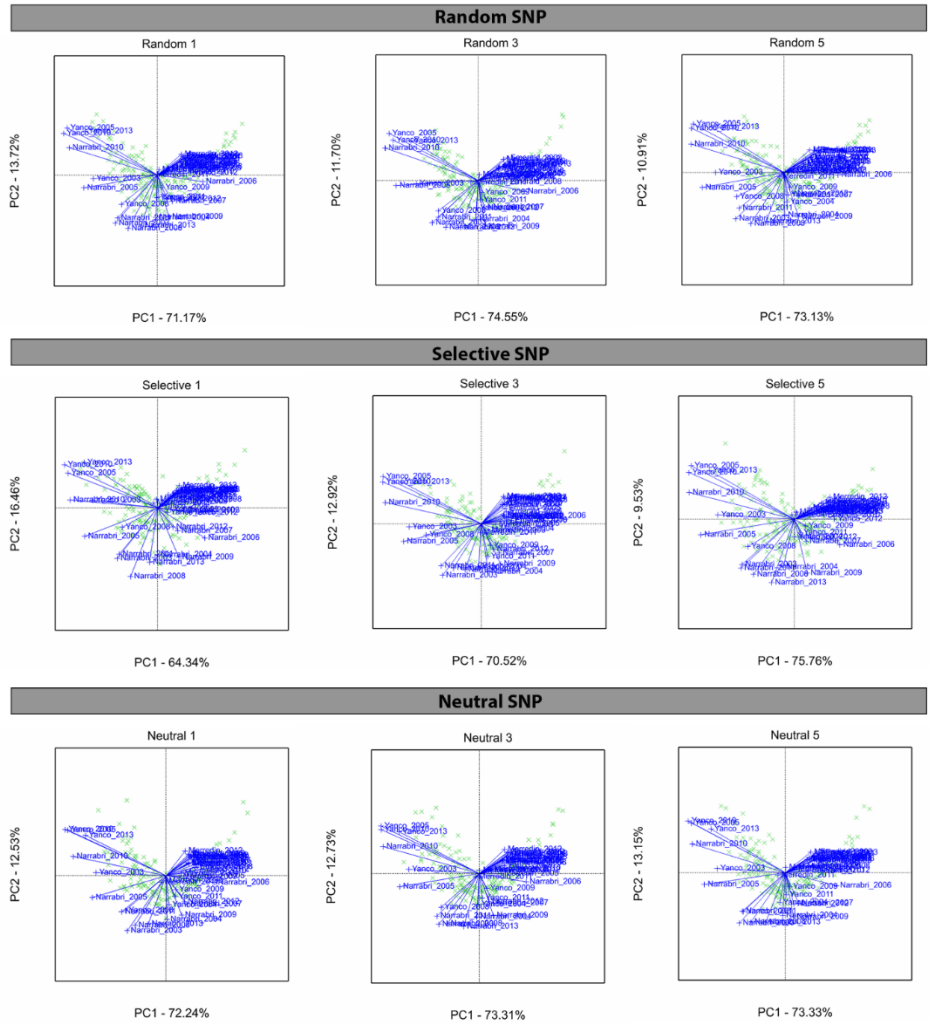


Figure 3. AMMI biplots for yield of three different realizations of the simulated parameters using random SNPs (upper panels), selective SNPs (Fst>0.5, middle panels) and neutral SNPs (Fst<0.1, lower panels).

Table 1. Variance components for the 199 genotypes and 40 environments (Merredin, Narrabri and Yanco during 2003-2013).

	Random 1		**Random 2**		**Random 3**		**Random 4**		**Random 5**	
	Comp	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**
Geno	4.40	0.57	5.64	0.71	4.61	0.62	4.06	0.52	3.45	0.46
Geno.Loc	3.89	0.26	4.94	0.30	4.80	0.31	3.37	0.22	3.35	0.22
Geno.Year	1.63	0.10	2.40	0.09	1.72	0.11	1.00	0.07	1.32	0.09
Geno.Loc.Year	5.99	0.11	6.41	0.08	6.45	0.12	4.69	0.09	5.13	0.09
Total	15.91		19.40		17.57		13.12		13.24	
	Selective 1		**Selective 2**		**Selective 3**		**Selective 4**		**Selective 5**	
	Comp	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**
Geno	4.27	0.51	5.81	0.78	3.42	0.45	4.06	0.52	7.50	0.88
Geno.Loc	2.57	0.17	6.02	0.39	3.03	0.20	3.53	0.23	3.94	0.26
Geno.Year	0.98	0.07	1.94	0.12	1.22	0.08	1.32	0.09	1.46	0.09
Geno.Loc.Year	4.35	0.08	6.94	0.13	5.22	0.10	5.37	0.10	5.59	0.10
Total	12.17		20.71		12.89		14.28		18.49	
	Neutral 1		**Neutral 2**		**Neutral 3**		**Neutral 4**		**Neutral 5**	
	Comp	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**	**Comp**	**s.e.**
Geno	3.07	0.42	4.14	0.54	4.45	0.59	4.98	0.63	4.56	0.62
Geno.Loc	3.28	0.22	3.62	0.24	4.42	0.29	3.84	0.25	4.97	0.32
Geno.Year	1.52	0.10	1.48	0.10	1.65	0.11	1.64	0.10	1.94	0.12
Geno.Loc.Year	5.85	0.11	5.73	0.11	6.48	0.12	6.10	0.11	6.81	0.12
Total	13.72		14.97		17.01		16.56		18.27	

Most of the simulations in the literature are designed for integrative traits and are done in a single-trait fashion. We think that the simulation of multiple traits that arise from a common genetic basic needs further exploration, especially for issues like which distribution to use for the additive effects, how the allocation of the trait-increasing alleles is related to the APSIM parameters and to the final traits, and how to deal with LD and population structure when allocating the additive effects to specific SNPs.

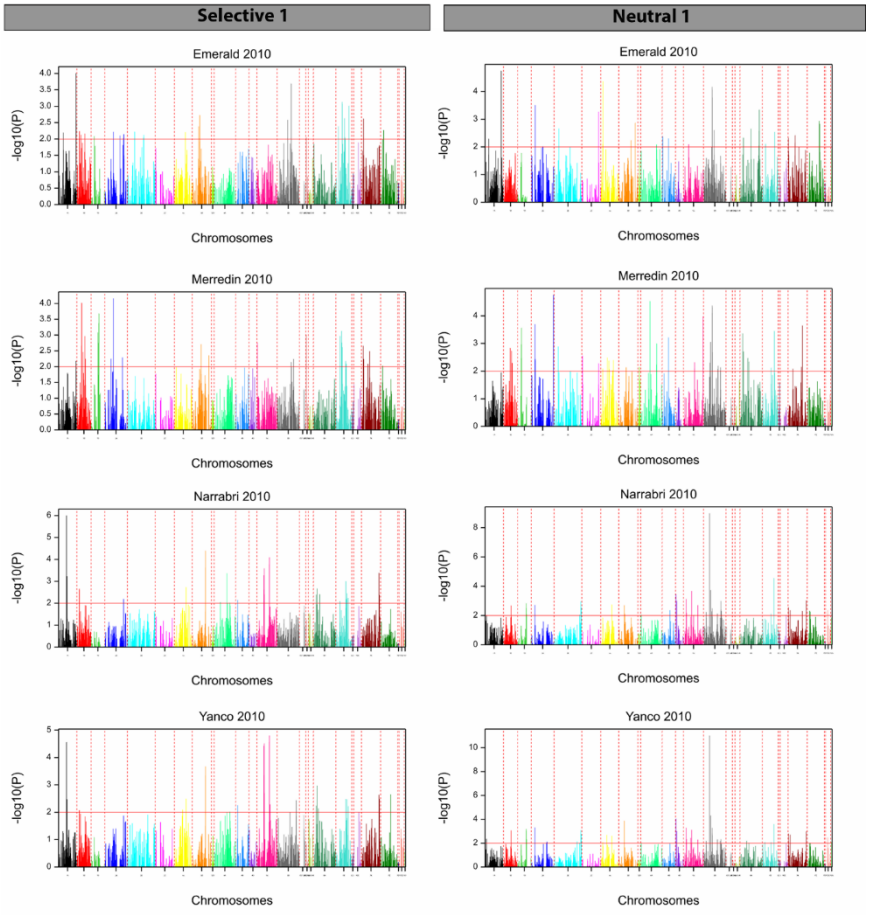


Figure 4. GWAS for yield of 199 genotypes in four locations in 2010. APSIM parameters for the genotypes were simulated using a selective genetic architecture (left panels) and a neutral genetic architecture (right panels). Both simulations used exactly the same additive effects, but they differed in the SNPs used to carry those effects.

List of references

Aastveit AH, Martens H (1986) ANOVA interactions interpreted by partial least squares regression. Biometrics 829–844.

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csáki F, Tsahkadsor (eds) 2nd International Symposium on Information Theory. Budapest, pp 267–281

Alam MM, Mace ES, van Oosterom EJ, et al (2014) QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. TAG 127:2253–2266. doi: 10.1007/s00122-014-2377-9

Albrecht T, Auinger H-J, Wimmer V, et al (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor Appl Genet 1–12. doi: 10.1007/s00122-014-2305-z

Albrecht T, Wimmer V, Auinger H-J, et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350. doi: 10.1007/s00122-011-1587-7

Alimi NA (2016) Statistical methods for QTL mapping and genomic prediction of multiple traits and environments: case studies in pepper.

Alimi NA, Bink MCAM, Dieleman JA, et al (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. Theor Appl Genet 126:2597–2625. doi: 10.1007/s00122-013-2160-3

Allard RW, Bradshaw AD (1964) Implications of genotype-environmental interactions in applied plant breeding. Crop Sci 4:503–508.

Alqudah AM, Koppolu R, Wolde GM, et al (2016) The genetic architecture of barley plant stature. Front Genet 7:1–15. doi: 10.3389/fgene.2016.00117

Annicchiarico P (1997) Additive main effects and multiplicative interaction (AMMI) analysis of genotype-location interaction in variety trials repeated over years. Theor Appl Genet 94:1072–1077. doi: 10.1007/s001220050517

Annicchiarico P, Bellah F, Chiari T (2005) Defining subregions and estimating benefits for a specific-adaptation strategy by breeding programs: A case study. Crop Sci 45:1741–1749. doi: 10.2135/cropsci2004.0524

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci 19:52–61. doi: http://dx.doi.org/10.1016/j.tplants.2013.09.008

Araus JL, Slafer GA, Royo C, Serret MD (2008) Breeding for Yield Potential and Stress Adaptation in Cereals. Crit Rev Plant Sci 27:377–412. doi: 10.1080/07352680802467736

Asseng S, Keating BA, Fillery IRP, et al (1998) Performance of the APSIM-wheat model in Western Australia. F Crop Res 57:163–179. doi: 10.1016/S0378-4290(97)00117-2

Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. Stat Sci 451–471.

Atlin GN, Baker RJ, McRae KB, Lu X (2000) Selection Response in Subdivided Target Regions. Crop Sci 40:7–13. doi: 10.2135/cropsci2000.4017

Atlin GN, Kleinknecht K, Singh GP, Piepho HP (2011) Managing Genotype x Environment Interaction in Plant Breeding Programs: A Selection Theory Approach. J Indian Soc Agric Stat 65:237–247.

Auinger H-J, Schönleben M, Lehermeier C, et al (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (Secale cereale L.). Theor Appl Genet 1–11. doi: 10.1007/s00122-016-2756-5

Austin RB (1999) Yield of Wheat in the United Kingdom: Recent Advances and Prospects. Crop Sci 39:1604–1610. doi: 10.2135/cropsci1999.3961604x

Azzalini A, Cox DR (1984) Two New Tests Associated with Analysis of Variance. J R Stat Soc Ser B 46:335–343. doi: 10.2307/2345519

Bac-Molenaar J a., Granier C, Keurentjes JJB, Vreugdenhil D (2015) Genome wide association mapping of time-dependent growth responses to moderate drought stress in Arabidopsis. Plant Cell Environ 88–102. doi: 10.1111/pce.12595

Baker RJ (1988) Tests for crossover genotype-environmental interactions. Can J Plant Sci 68:405–410. doi: 10.4141/cjps88-051

Bänziger M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. Cimmyt, Mexico, DF (Mexico)

Bänziger M, Betrán FJ, Lafitte HR (1997) Efficiency of High-Nitrogen Selection Environments for Improving Maize for Low-Nitrogen Target Environments. Crop Sci 37:1103–1109. doi: 10.2135/cropsci1997.0011183X003700040012x

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Basford KE, Cooper M (1998) Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia . Aust J Agric Res 49:153–174. doi: http://dx.doi.org/10.1071/A97035

Bassi FM, Sanchez-Garcia M (2017) Adaptation and stability analysis of ICARDA durum wheat elites across 18 countries. Crop Sci 57:1–12. doi: 10.2135/cropsci2016.11.0916

Beeck CP, Cowling WA, Smith AB, Cullis BR (2010) Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. Genome 53:992–1001. doi: 10.1139/g10-051

Beier S, Himmelbach A, Colmsee C, et al (2017) Construction of a map-based reference genome sequence for barley, Hordeum vulgare L. Sci data 4:170044. doi: 10.1038/sdata.2017.44

Bernardo R (2014) Genomewide Selection when Major Genes Are Known. Crop Sci 54:68–75. doi: 10.2135/cropsci2013.05.0315

Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize . Crop Sci 47:1082–1090. doi: 10.2135/cropsci2006.11.0690

Bernardo RN (2010) Breeding for Quantitative Traits in Plants. Stemma Press

Bertin N, Martre P, Génard M, et al (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. J Exp Bot 61:955–967. doi: 10.1093/jxb/erp377

Bessho-Uehara K, Wang DR, Furuta T, et al (2016) Loss of function at RAE2 , a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. Proc Natl Acad Sci 113:8969–8974. doi: 10.1073/pnas.1604849113

Biscarini F, Nazzicari N, Bink M, et al (2017) Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. BMC Genomics 18:432. doi: 10.1186/s12864-017-3781-8

Blanc G, Charcosset A, Mangin B, et al (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224. doi: 10.1007/s00122-006-0287-1

Boer MP, Wright D, Feng L, et al (2007) A Mixed-Model Quantitative Trait Loci (QTL) Analysis for Multiple-Environment Trial Data Using Environmental Covariables for QTL-by-Environment Interactions, With an Example in Maize. Genetics 177:1801–1813. doi: 10.1534/genetics.107.071068

Bogard M, Ravel C, Paux E, et al (2014) Predictions of heading date in bread wheat (Triticum aestivum L.) using QTL-based parameters of an ecophysiological model. J Exp Bot 65:5849–5865. doi: 10.1093/jxb/eru328

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma . doi: 10.1093/bioinformatics/btu170

Bradshaw AD (2006) Unravelling phenotypic plasticity – why should we bother? New Phytol 170:644–648. doi: 10.1111/j.1469-8137.2006.01761.x

Bradshaw AD, Caspari EW, Thoday JM (1965) Evolutionary Significance of Phenotypic Plasticity in Plants. In: Adv. Genet. pp 115–155

Braun H-J, Rajaram S, van Ginkel M (1996) CIMMYT's approach to breeding for wide adaptation. Euphytica 92:175–183. doi: 10.1007/BF00022843

Bro R, Kjeldahl K, Smilde AK, Kiers HAL (2008) Cross-validation of component models: A critical look at current methods. Anal Bioanal Chem 390:1241–1251. doi: 10.1007/s00216-007-1790-1

Buckler ES, Holland JB, Bradbury PJ, et al (2009) The Genetic Architecture of Maize Flowering Time. Science (80-) 325:714–718. doi: 10.1126/science.1174276

Bull JK, Cooper M, DeLacy IH, et al (1992) Utility of repeated checks for hierarchical classification of data from plant breeding trials. F Crop Res 30:79–95. doi: http://dx.doi.org/10.1016/0378-4290(92)90058-H

Burgueño J, Crossa J, Cornelius PL, Yang R-C (2008) Using Factor Analytic Models for Joining Environments and Genotypes without Crossover Genotype × Environment Interaction. Crop Sci 48:1291–1305. doi: 10.2135/cropsci2007.11.0632

Burgueño J, Crossa J, Miguel Cotes J, et al (2011) Prediction Assessment of Linear Mixed Models for Multienvironment Trials. Crop Sci 51:944–954. doi: 10.2135/cropsci2010.07.0403

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype x Environment Interaction using Pedigree and Dense Molecular Markers. Crop Sci 52:707–719. doi: 10.2135/cropsci2011.06.0299

Bustos-Korts D, Malosetti M, Chapman S, et al (2016a) Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. G3 Genes|Genomes|Genetics 6:3733–3747. doi: doi:10.1534/g3.116.035410/-/DC1

Bustos-Korts D, Malosetti M, Chapman S, van Eeuwijk FA van (2016b) Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics. In: Yin X, Struik PC (eds) Crop Systems Biology - Narrowing the Gaps between Crop Modelling and Genetics. Springer, pp 55–82

Bustos D V., Hasan AK, Reynolds MP, Calderini DF (2013) Combining high grain number and weight through a DH-population to improve grain yield potential of wheat in high-yielding environments. F Crop Res 145:106–115. doi: http://dx.doi.org/10.1016/j.fcr.2013.01.015

Butler RC, Brain P (1993) Nonlinear Contrasts in ANOVA. Genstat Newsl 20–27.

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, et al (2012) High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. J Integr Plant Biol 54:312–320. doi: 10.1111/j.1744-7909.2012.01116.x

Cabrera-Bosquet L, Fournier C, Brichet N, et al (2016) High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. New Phytol 212:269–281. doi: 10.1111/nph.14027

Calderini DF, Dreccer MF, Slafer GA (1997) Consequences of breeding on biomass, radiation interception and radiation-use efficiency in wheat. F Crop Res 52:271–281. doi: 10.1016/s0378-4290(96)03465-x

Calus M, Veerkamp R (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:26.

Calus MPL, Meuwissen THE, De Roos APW, Veerkamp RF (2008) Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561. doi: 10.1534/genetics.107.080838

Campbell BT, Baenziger PS, Eskridge KM, et al (2004) Using Environmental Covariates to Explain Genotype × Environment and QTL × Environment Interactions for Agronomic Traits on Chromosome 3A of Wheat. Crop Sci 44:620–627. doi: 10.2135/cropsci2004.6200

Campos H, Cooper M, Habben JE, et al (2004) Improving drought tolerance in maize: a view from industry. F Crop Res 90:19–34. doi: http://dx.doi.org/10.1016/j.fcr.2004.07.003

Casadebaig P, Zheng B, Chapman S, et al (2016) Assessment of the potential impacts of plant traits across environments by combining global sensitivity analysis and dynamic modeling in wheat. PLoS One. doi: 10.1371/journal.pone.0146385

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221. doi: 10.1016/j.pbi.2008.01.002

Ceccarelli S (1989) Wide adaptation: How wide? Euphytica 40:197–205. doi: 10.1007/BF00024512

Ceccarelli S (1996) Adaptation to low/high input cultivation. Euphytica 92:203–214. doi: 10.1007/BF00022846

Ceccarelli S, Acevedo E, Grando S (1991) Breeding for yield stability in unpredictable environments: single traits, interaction between traits, and architecture of genotypes. Euphytica 56:169–185. doi: 10.1007/BF00042061

Ceccarelli S, Grando S (1996) Drought as a challenge for the plant breeder. Plant Growth Regul 20:149–155. doi: 10.1007/BF00024011

Ceccarelli S, Grando S, Impiglia A (1998) Choice of selection strategy in breeding barley for stress environments. Euphytica 103:307–318. doi: 10.1023/A:1018647001429

Chapman S (2008a) Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. Euphytica 161:195–208. doi: 10.1007/s10681-007-9623-z

Chapman S, Cooper M, Podlich D, Hammer G (2003) Evaluating Plant Breeding Strategies by Simulating Gene Action and Dryland Environment Effects. Agron J 95:99–113. doi: 10.2134/agronj2003.9900

Chapman SC (2008b) Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. Euphytica 161:195–208. doi: 10.1007/s10681-007-9623-z

Chapman SC, Chakraborty S, Dreccer MF, Howden SM (2012) Plant adaptation to climate change—opportunities and priorities in breeding. Crop Pasture Sci 63:251–268. doi: http://dx.doi.org/10.1071/CP11303

Chapman SC, Cooper M, Butler DG, Henzell RG (2000a) Genotype by environment interactions affecting grain sorghum. I. Characteristics that confound interpretation of hybrid yield. Aust J Agric Res 51:197–208. doi: http://dx.doi.org/10.1071/AR99020

Chapman SC, Cooper M, Hammer GL (2002a) Using crop simulation to generate genotype by environment interaction effects for sorghum in water-limited environments. Aust J Agric Res 53:379–389. doi: http://dx.doi.org/10.1071/AR01070

Chapman SC, Cooper M, Hammer GL, Butler DG (2000b) Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. Aust J Agric Res 51:209–222. doi: http://dx.doi.org/10.1071/AR99021

Chapman SC, Hammer GL, Butler DG, Cooper M (2000c) Genotype by environment interactions affecting grain sorghum. III. Temporal sequences and spatial patterns in the target population of environments. Aust J Agric Res 51:223–234. doi: http://dx.doi.org/10.1071/AR99022

Chapman SC, Hammer GL, Podlich DW, Cooper M (2002b) Linking bio-physical and genetic models to integrate physiology, molecular biology and plant breeding. In: Kang M (ed) Quantitative Genetics, Genomics, and Plant Breeding. CAB International , Wallingford UK, pp 167–187

Chapman SC, Merz T, Chan A, et al (2014) Pheno-copter: a low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. Agronomy 4:279–301.

Chauhan YS, Rachaputi RCN (2014) Defining agro-ecological regions for field crops in variable target production environments: A case study on mungbean in the northern grains region of Australia. Agric For Meteorol 194:207–217. doi: http://dx.doi.org/10.1016/j.agrformet.2014.04.007

Chen A, Baumann U, Fincher GB, Collins NC (2009) Flt-2L, a locus in barley controlling flowering time, spike density, and plant height. Funct Integr Genomics 9:243–254. doi: 10.1007/s10142-009-0114-2

Chen X, Zhang Z, Liu D, et al (2010) SQUAMOSA promoter-binding protein-like transcription factors: Star players for plant growth and development. J Integr Plant Biol 52:946–951. doi: 10.1111/j.1744-7909.2010.00987.x

Chenu K, Chapman SC, Hammer GL, et al (2008) Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. Plant Cell Environ 31:378–391. doi: 10.1111/j.1365-3040.2007.01772.x

Chenu K, Chapman SC, Tardieu F, et al (2009) Simulating the Yield Impacts of Organ-Level Quantitative Trait Loci Associated With Drought Response in Maize: A "Gene-to-Phenotype" Modeling Approach. Genetics 183:1507–1523. doi: 10.1534/genetics.109.105429

Chenu K, Cooper M, Hammer GL, et al (2011) Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. J Exp Bot 62:1743–1755. doi: 10.1093/jxb/erq459

Chenu K, Deihimfard R, Chapman SC (2013) Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt – spatial and temporal trends. New Phytol 198:801–820. doi: 10.1111/nph.12192

Cheverud JM, Routman EJ (1995) Epistatis and its contribution to genetic variance components. Genetics 139:1455–1461.

Close TJ, Bhat PR, Lonardi S, et al (2009) Development and implementation of high-throughput SNP genotyping in barley.

Cobb JN, Declerck G, Greenberg A, et al (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126:867–887. doi: 10.1007/s00122-013-2066-0

Cockram J, Chiapparino E, Taylor SA, et al (2007) Haplotype analysis of vernalization loci in European barley germplasm reveals novel VRN-H1 alleles and a predominant

winter VRN-H1/VRN-H2 multi-locus haplotype. Theor Appl Genet 115:993–1001. doi: 10.1007/s00122-007-0626-x

Cockram J, Horsnell R, Soh E, et al (2015) Molecular and phenotypic characterization of the alternative seasonal growth habit and flowering time in barley (Hordeum vulgare ssp. vulgare L.). Mol Breed 35:165. doi: 10.1007/s11032-015-0359-5

Comadran J, Kilian B, Russell J, et al (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 44:1388–1392.

Comstock R (1977) Quantitative genetics and the design of breeding programme. In: Proceedings of the International Conference on Quantitative Genetics. University, Iowa State Press, Ames, IA, pp 705–718

Comstock RE, Moll RH (1963) Genotype-Environment Interactions. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding: A Symposium and Workshop. National Academy of Sciences-National Research Council, pp 164–196

Cooper M (1999) Concepts and strategies for plant adaptation research in rainfed lowland rice. F Crop Res 64:13–34. doi: http://dx.doi.org/10.1016/S0378-4290(99)00048-9

Cooper M (2004) Complex Trait Genetics and Gene-to-Phenotype Models. 895–918.

Cooper M, Brennan PS, Sheppard JA (1996) A Strategy for Yield Improvement of Wheat which Accomodates Larce Genotype by Environment Interactions. In: Cooper M, Hammer GL (eds) Plant adaptation and Crop Improvement. University Press, Cambridge, pp 487–512

Cooper M, Byth DE, DeLacy IH (1993) A procedure to assess the relative merit of classification strategies for grouping environments to assist selection in plant breeding regional evaluation trials. F Crop Res 35:63–74. doi: http://dx.doi.org/10.1016/0378-4290(93)90137-C

Cooper M, Chapman SC, Podlich DW, Hammer GL (2002) The GP Problem: Quantifying Gene-to-Phenotype Relationships. In Silico Biol 2:151–164.

Cooper M, Fox P. (1996) Environmental characterisation based on probe and reference genotypes. In: Cooper M, Hammer G. (eds) Plant Adaptation and Crop Improvement. CAB International, Wallingford, UK, pp 529–547

Cooper M, Gho C, Leafgren R, et al (2014a) Breeding drought-tolerant maize hybrids for the US corn-belt: discovery to product.

Cooper M, Hammer GL (1996) Plant Adaptation and Crop Improvement. CAB International, Wallingford, UK

Cooper M, Messina CD, Podlich D, et al (2014b) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 65:311–336. doi: http://dx.doi.org/10.1071/CP14007

Cooper M, Technow F, Messina C, et al (2016) Use of crop growth models with whole-genome prediction: Application to a maize multienvironment trial. Crop Sci 56:2141–2156. doi: 10.2135/cropsci2015.08.0512

Cooper M, van Eeuwijk FA, Hammer GL, et al (2009) Modeling QTL for complex traits: detection and context for plant breeding. Curr Opin Plant Biol 12:231–240. doi: 10.1016/j.pbi.2009.01.006

Cooper M, Woodruff DR (1993) Predicting grain yield in Australian environments using data from CIMMYT international wheat performance trials 3. Testing predicted correlated response to selection. F Crop Res 35:191–204. doi: http://dx.doi.org/10.1016/0378-4290(93)90153-E

Cornelius PL (1993) Statistical Tests and Retention of Terms in the Additive Main Effects and Multiplicative Interaction Model for Cultivar Trials. Crop Sci 33:1186–1193. doi: 10.2135/cropsci1993.0011183X003300060016x

Cornelius PL, Seyedsadr M, Crossa J (1992) Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theor Appl Genet 84:161–172. doi: 10.1007/bf00223996

Cornelius PL, Seyedsadr MS (1997) Estimation of general linear-bilinear models for two-way tables. J Stat Comput Simul 58:287–322. doi: 10.1080/00949659708811837

Cornelius PL, Van Sanford DA, Seyedsadr MS (1993) Clustering Cultivars into Groups without Rank-Change Interactions. Crop Sci 33:1193–1200. doi: 10.2135/cropsci1993.0011183X003300060017x

Corsten LCA, Denis JB (1990) Structuring Interaction in Two-Way Tables by Clustering. Biometrics 46:207–215. doi: 10.2307/2531644

Crossa J (2012) From Genotype x Environment Interaction to Gene x Environment Interaction. Curr Genomics 13:225–244.

Crossa J, Beyene Y, Kassa S, et al (2013a) Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. G3 Genes Genom Genet 3:1903–1926. doi: 10.1534/g3.113.008227

Crossa J, Burgueño J, Cornelius PL, et al (2006) Modeling Genotype × Environment Interaction Using Additive Genetic Covariances of Relatives for Predicting Breeding Values of Wheat Genotypes. Crop Sci 46:1722–1733. doi: 10.2135/cropsci2005.11-0427

Crossa J, Cornelius PL (1997) Sites Regression and Shifted Multiplicative Model Clustering of Cultivar Trial Sites under Heterogeneity of Error Variances. Crop Sci 37:406–415. doi: 10.2135/cropsci1997.0011183X003700020017x

Crossa J, Cornelius PL, Yan W (2002) Biplots of Linear-Bilinear Models for Studying Crossover Genotype × Environment Interaction The investigation reported in this paper (No. 00-06-100) relates to a project of the Kentucky Agric. Exp. Stn. and is published with approval of the Director. . Crop Sci 42:619–633. doi: 10.2135/cropsci2002.6190

Crossa J, de los Campos G, Perez P, et al (2010a) Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. Genetics 186:713-U406. doi: 10.1534/genetics.110.118521

Crossa J, Fox PN, Pfeiffer WH, et al (1991) AMMI adjustment for statistical analysis of an international wheat yield trial. Theor Appl Genet 81:27–37. doi: 10.1007/bf00226108

Crossa J, Perez-Elizalde S, Jarquin D, et al (2011) Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. Crop Sci 51:1458–1469. doi: 10.2135/cropsci2010.06.0343

Crossa J, Perez P, Hickey J, et al (2013b) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb) 112:48–60. doi: 10.1038/hdy.2013.16

Crossa J, Vargas M, Cossani CM, et al (2015) Evaluation and Interpretation of Interactions. Agron J 107:736–747. doi: 10.2134/agronj2012.0491

Crossa J, Vargas M, Joshi AK (2010b) Linear, bilinear, and linear-bilinear fixed and mixed models for analyzing genotype × environment interaction in plant breeding and agronomy. Can J Plant Sci 90:561–574. doi: 10.4141/CJPS10003

Crossa J, Vargas M, van Eeuwijk FA, et al (1999) Interpreting genotype × environment interaction in tropical maize using linked molecular markers and environmental covariables. Theor Appl Genet 99:611–625. doi: 10.1007/s001220051276

Crossa J, Yang R-C, Cornelius P (2004) Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. JABES 9:362–380. doi: 10.1198/108571104x4423

Cuesta-Marcos A, Igartua E, Ciudad FJ, et al (2008) Heading date QTL in a spring × winter barley cross evaluated in Mediterranean environments. Mol Breed 21:455–471. doi: 10.1007/s11032-007-9145-3

Cullis BR, Smith A, Hunt C, Gilmour A (2000) An examination of the efficiency of Australian crop variety evaluation programmes. J Agric Sci 135:213–222.

Cullis BR, Smith AB, Beeck CP, Cowling WA (2010) Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. Genome 53:1002–1016.

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Env Stat 11:381–393. doi: 10.1198/108571106x154443

Cullis BR, Thomson FM, Fisher JA, et al (1996) The analysis of the NSW wheat variety database. II. Variance component estimation. Theor Appl Genet 92:28–39. doi: 10.1007/bf00222948

Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ (2012) Components of the accuracy of genomic prediction in a multi-breed sheep population. J Anim Sci 90:3375–3384. doi: 10.2527/jas.2011-4557

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. PLoS One 3:e3395.

Dawson IK, Russell J, Powell W, et al (2015) Barley: a translational model for adaptation to climate change. New Phytol 206:913–931. doi: 10.1111/nph.13266

Dawson JC, Endelman JB, Heslot N, et al (2013) The use of unbalanced historical data for genomic selection in an international wheat breeding program. F Crop Res 154:12–22.

de la Vega AJ, Chapman SC (2010) Mega-Environment Differences Affecting Genetic Progress for Yield and Relative Value of Component Traits All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including phot. Crop Sci 50:574–583. doi: 10.2135/cropsci2009.04.0209

De los Campos G, Gianola D, Rosa G, et al (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res (Camb) 92:295–308. doi: doi:10.1017/S0016672310000285

de los Campos G, Gianola D, Rosa GJM (2009) Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. J Anim Sci 87:1883–1887. doi: 10.2527/jas.2008-1259

de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345. doi: 10.1534/genetics.112.143313

de los Campos G, Veturi Y, Vazquez AI, et al (2015) Incorporating Genetic Heterogeneity in Whole-Genome Regressions Using Interactions. J Agric Biol Environ Stat 20:467–490. doi: 10.1007/s13253-015-0222-5

De Vos S, Wardenaar KJ, Bos EH, et al (2017) An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. PLoS One 12:1–18. doi: 10.1371/journal.pone.0178586

Debat V, David P (2001) Mapping phenotypes: canalization, plasticity and developmental stability. Trends Ecol Evol 16:555–561. doi: http://dx.doi.org/10.1016/S0169-5347(01)02266-2

Deery D, Jimenez-Berni J, Jones H, et al (2014) Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. Agronomy 4:

Deery DM, Rebetzke GJ, Jimenez-Berni JA, et al (2016) Methodology for High-Throughput Field Phenotyping of Canopy Temperature Using Airborne Thermography. Front Plant Sci 7:1808. doi: 10.3389/fpls.2016.01808

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331–341. doi: 10.1111/j.1439-0388.2007.00701.x

Denis J baptiste (1988) Two way analysis using covariates. Statistics (Ber) 19:123–132.

Denis JB (1991) Ajustements de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. Rev Stat Appliquée 39:5–24.

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004.

DeWitt TJ, Scheiner SM (2004) Phenotypic plasticity: functional and conceptual approaches. Oxford University Press Oxford

Dickson SP, Wang K, Krantz I, et al (2010) Rare Variants Create Synthetic Genome-Wide Associations. PLoS Biol. doi: 10.1371/journal.pbio.1000294

Dobzhansky T, Spassky B (1963) Genetics of Natural Populations. Xxxiv. Adaptive Norm, Genetic Load and Genetic Elite in Drosophila Pseudoobscura. Genetics 48:1467–1485.

Dockter C, Gruszka D, Braumann I, et al (2014) Induced Variations in Brassinosteroid Genes Define Barley Height and Sturdiness, and Expand the Green Revolution Genetic Toolkit 1[C][W][OPEN]. Plant Physiol 166:1912–1927. doi: 10.1104/pp.114.250738

Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. Nature 456:728–31. doi: 10.1038/nature07631

Drosse B, Campoli C, Mulki AM, von Korff M (2014) Genetic control of reproductive development. In: Kumlehn J, Stein N (eds) Biotechnological Approaches to Barley Improvement. Springer, Berlin, Heidelberg, Germany, pp 81–98

Eberhart SA, Russell WA (1966) Stability Parameters for Comparing Varieties. Crop Sci 6:36–40. doi: 10.2135/cropsci1966.0011183X000600010011x

Eberius M, Lima-Guerra J (2009) High-Throughput Plant Phenotyping – Data Acquisition, Transformation, and Analysis. In: Edwards D, Stajich J, Hansen D (eds) Bioinformatics. pp 259–278

Edmeades GO, McMaster GS, White JW, Campos H (2004) Genomics and the physiologist: bridging the gap between genes and crop response. F Plant Biol 90:5–18. doi: http://dx.doi.org/10.1016/j.fcr.2004.07.002

Eilers PHC, Marx BD (1996) Flexible Smoothing with B-splines and Penalties. Stat Sci 11:89–102. doi: 10.2307/2246049

Eilers PHC, Marx BD, Durbán M (2015) Twenty years of P-splines. SORT-Statistics Oper Res Trans 39:149–186.

Ellis RP, Forster BP, Robinson D, et al (2000) Wild barley: a source of genes for crop improvement in the 21st century? J Exp Bot 51:9–17. doi: 10.1093/JEXBOT/51.342.9

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6:e19379. doi: 10.1371/journal.pone.0019379

Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. 250–255.

Eskridge KM, Mumm RF (1992) Choosing plant cultivars based on the probability of outperforming a check. Theor Appl Genet 84:494–500. doi: 10.1007/BF00229512

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, Fourth. Longman

FAO (2013) http://faostat.fao.org/site/368/DesktopDefault.aspx?PageID=368#ancor.

FAO/IIASA (2011) Global Agro-Ecological Zones (GAEZ v. 3.0).

Faure S, Higgins J, Turner A, Laurie DA (2007) The FLOWERING LOCUS T -Like Gene Family in Barley (Hordeum vulgare). Genetics 176:599–609. doi: 10.1534/genetics.106.069500

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14:742–754. doi: http://dx.doi.org/10.1071/AR9630742

Fischer RA, Maurer R (1978) Drought resistance in spring wheat cultivars. I. Grain yield responses. Aust J Agric Res 29:897–912. doi: http://dx.doi.org/10.1071/AR9780897

Fleury D, Jefferies S, Kuchel H, Langridge P (2010) Genetic and genomic tools to improve drought tolerance in wheat. J Exp Bot 61:3211–3222. doi: 10.1093/jxb/erq152

Flint-Garcia SA (2013) Genetics and Consequences of Crop Domestication. J Agric Food Chem 61:8267–8276. doi: 10.1021/jf305511d

Forkman J, Amiri S, Rosen D (2012) Effect of region on the uncertainty in crop variety trial programs with a reduced number of trials. Euphytica 186:489–500. doi: 10.1007/s10681-012-0646-8

Foulkes MJ, Slafer GA, Davies WJ, et al (2011) Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. J Exp Bot 62:469–486. doi: 10.1093/jxb/erq300

Franco J, Crossa J, Taba S, Shands H (2005) A Sampling Strategy for Conserving Genetic Diversity when Forming Core Subsets. Crop Sci 45:1035–1044. doi: 10.2135/cropsci2004.0292

Frascaroli E, Schrag TA, Melchinger AE (2012) Genetic diversity analysis of elite European maize (Zea mays L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. Theor Appl Genet 126:133–141. doi: 10.1007/s00122-012-1968-6

Furbank RT, Tester M (2011) Phenomics - technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16:635–644. doi: 10.1016/j.tplants.2011.09.005

Gabriel KR (1971) The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. Biometrika 58:453–467. doi: 10.2307/2334381

Gabriel KR (1978) Least Squares Approximation of Matrices by Additive and Multiplicative Models. J R Stat Soc Ser B 40:186–196. doi: 10.2307/2984755

Gabriel SB, Schaffner SF, Nguyen H, et al (2002) The structure of haplotype blocks in the human genome. Science (80-) 296:2225–2229.

Gail M, Simon R (1985) Testing for Qualitative Interactions between Treatment Effects and Patient Subsets. Biometrics 41:361–372. doi: 10.2307/2530862

Ganal MW, Durstewitz G, Polley A, et al (2011) A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. PLoS One 6:e28334.

García GA, Hasan AK, Puhl LE, et al (2013) Grain Yield Potential Strategies in an Elite Wheat Double-Haploid Population Grown in Contrasting Environments. Crop Sci 53:2577–2587. doi: 10.2135/cropsci2012.11.0669

Garin V, Wimmer V, Mezmouk S, et al (2017) How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. Theor Appl Genet 130:1–12. doi: 10.1007/s00122-017-2923-3

Gauch HG (1992) Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs. Elsevier

Gauch HG (2013) A Simple Protocol for AMMI Analysis of Yield Trials. Crop Sci 53:1860–1869. doi: 10.2135/cropsci2013.04.0241

Gauch HG (2006) Statistical Analysis of Yield Trials by AMMI and GGE. Crop Sci 46:1488–1500. doi: 10.2135/cropsci2005.07-0193

Gauch HG, Piepho H-P, Annicchiarico P (2008) Statistical Analysis of Yield Trials by AMMI and GGE: Further Considerations. Crop Sci 48:866–889. doi: 10.2135/cropsci2007.09.0513

Gauch HG, Zobel RW (1997) Identifying Mega-Environments and Targeting Genotypes. Crop Sci 37:311–326. doi: 10.2135/cropsci1997.0011183X003700020002x

Gauch Jr. HG, Gauch Jr. HG (1988) Model Selection and Validation for Yield Trials with Interaction. Biometrics 44:705–715. doi: 10.2307/2531585

Gauch Jr. HG, Zobel RW, Gauch Jr. HG, Zobel RW (1988) Predictive and postdictive success of statistical analyses of yield trials. Theor Appl Genet 76:1–10. doi: 10.1007/BF00288824

Gavrilets S (2004) Fitness Landscapes and the Origin of Species. Princeton University Press

Gelman A (2005) Analysis of variance--why it is more important than ever. Ann Stat 33:1–53. doi: 10.1214/009053604000001048

Gianola D, van Kaam JBCHM (2008) Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics 178:2289–2303. doi: 10.1534/genetics.107.084285

Giudici P (2003) Applied Data Mining. Wiley

Gogel BJ, Cullis BR, Verbyla AP (1995) REML Estimation of Multiplicative Effects in Multienvironment Variety Trails. Biometrics 51:744–749. doi: 10.2307/2532960

Gollob H (1968) A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika 33:73–115. doi: 10.1007/bf02289676

González-Camacho JM, de los Campos G, Pérez P, et al (2012) Genome-enabled prediction of genetic values using radial basis function neural networks. Theor Appl Genet 125:759–771. doi: 10.1007/s00122-012-1868-9

Gore MA, Chia J-M, Elshire RJ, et al (2009) A First-Generation Haplotype Map of Maize. Science (80-) 326:1115–1117.

Govindaraju DR, Dancik BP (1987) Allozyme heterozygosity and homeostasis in germinating seeds of jack pine. Heredity (Edinb) 59:279–283.

Gower JC (1971) A General Coefficient of Similarity and Some of Its Properties. Biometrics 27:857–871. doi: 10.2307/2528823

Gregorius HR, Namkoong G (1986) Joint analysis of genotypic and environmental effects. Theor Appl Genet 72:413–422. doi: 10.1007/bf00288581

Grieder C, Hund A, Walter A (2015) Image based phenotyping during winter: a powerful tool to assess wheat genetic variation in growth response to temperature. Funct. Plant Biol. 42:387–396.

Gruszka D, Gorniak M, Glodowska E, et al (2016) A reverse-genetics mutational analysis of the barley HvDWARF gene results in identification of a series of alleles and mutants with short stature of various degree and disturbance in BR biosynthesis allowing a new insight into the process. Int J Mol Sci. doi: 10.3390/ijms17040600

Gu B, Zhou T, Luo J, et al (2015) An-2 Encodes a Cytokinin Synthesis Enzyme that Regulates Awn Length and Grain Production in Rice. Mol Plant 8:1635–1650. doi: 10.1016/j.molp.2015.08.001

Gu J, Yin X, Zhang C, et al (2014) Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (Oryza sativa) under drought stress. Ann Bot 114:499–511. doi: 10.1093/aob/mcu127

Gumedze FN, Dunne TT (2011) Parameter estimation and inference in the linear mixed model. Linear Algebra Appl 435:1920–1944.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. Genetics 195:205–220. doi: 10.1534/genetics.113.152462

Guo G, Zhao F, Wang Y, et al (2014a) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:1–7. doi: 10.1186/1471-2156-15-30

Guo Z, Tucker D, Basten C, et al (2014b) The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127:749–762. doi: 10.1007/s00122-013-2255-x

Guo Z, Tucker DM, Wang D, et al (2013) Accuracy of Across-Environment Genome-Wide Prediction in Maize Nested Association Mapping Populations. G3 Genes Genom Genet 3:263–272. doi: 10.1534/g3.112.005066

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12:186. doi: 10.1186/1471-2105-12-186

Habier D, Tetens J, Seefried F-R, et al (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 42:5. doi: 10.1186/1297-9686-42-5

Haghighattalab A, González Pérez L, Mondal S, et al (2016) Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods 12:35. doi: 10.1186/s13007-016-0134-6

Hammer G, Cooper M, Tardieu F, et al (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593. doi: http://dx.doi.org/10.1016/j.tplants.2006.10.006

Hammer G, Messina C, van Oosterom E, et al (2016) Molecular Breeding for Complex Adaptive Traits: How Integrating Crop Ecophysiology and Modelling Can Enhance Efficiency BT - Crop Systems Biology: Narrowing the gaps between crop modelling and genetics. In: Yin X, Struik PC (eds). Springer International Publishing, Cham, pp 147–162

Hammer GL, Chapman S, van Oosterom E, Podlich DW (2005) Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Aust J Agric Res 56:947–960. doi: http://dx.doi.org/10.1071/AR05157

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling—from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18:15–31. doi: http://dx.doi.org/10.1016/S1161-0301(02)00093-X

Hammer GL, McLean G, Chapman S, et al (2014) Crop design for specific adaptation in variable dryland production environments. Crop Pasture Sci. 65:614–626.

Hammer GL, van Oosterom E, McLean G, et al (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61:2185–2202. doi: 10.1093/jxb/erq095

Hanson WD (1970) Genotypic stability. Theor Appl Genet 40:226–231.

Harrison MT, Tardieu F, Dong Z, et al (2014) Characterizing drought stress and trait influence on maize yield under current and future conditions. Glob Chang Biol 20:867–878. doi: 10.1111/gcb.12381

Hartmann A, Czauderna T, Hoffmann R, et al (2011) HTPheno: An image analysis pipeline for high-throughput plant phenotyping. BMC Bioinformatics 12:148.

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning, Second. Springer

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92:433–443. doi: http://dx.doi.org/10.3168/jds.2008-1646

Hayes BJ, Chamberlain a J, McPartlan H, et al (2007) Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. Genet Res 89:215–220. doi: 10.1017/S0016672307008865

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

Hayes PM, Liu BH, Knapp SJ, et al (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm. Theor Appl Genet 87:392–401. doi: 10.1007/BF01184929

Hazel LN (1943) The genetic basis for constructing selection indexes. Genetics 28:476 LP-490.

Heffner EL, Jannink J-L, Sorrells ME (2011a) Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. Plant Gen 4:65–75. doi: 10.3835/plantgenome2010.12.0029

Heffner EL, Jannink JL, Iwata H, et al (2011b) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. Crop Sci 51:2597–2606. doi: 10.2135/cropsci2011.05.0253

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding With Genomic Selection: Gain Per Unit Time And Cost. Crop Sci 50:1681–1690. doi: 10.2135/cropsci2009.11.0662

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

Heslot N, Akdemir D, Sorrells M, Jannink J-L (2013a) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 1–18. doi: 10.1007/s00122-013-2231-5

Heslot N, Jannink J-L, Sorrells ME (2013b) Using Genomic Prediction to Characterize Environments and Optimize Prediction Accuracy in Applied Breeding Data. Crop Sci 53:921–933. doi: 10.2135/cropsci2012.07.0420

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci 52:146–160. doi: 10.2135/cropsci2011.06.0297

Hickey JM, Dreisigacker S, Crossa J, et al (2014) Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. Crop Sci 54:1476–1488. doi: 10.2135/cropsci2013.03.0195

Hill CB, Li C (2016) Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. Front Plant Sci 7:1906. doi: 10.3389/fpls.2016.01906

Hoffman GE (2013) Correcting for Population Structure and Kinship Using the Linear Mixed Model: Theory and Extensions. PLoS One. doi: 10.1371/journal.pone.0075707

Holland JB (2001) Epistasis and Plant Breeding. In: Plant Breeding Reviews. John Wiley & Sons, Inc., pp 27–92

Holzworth DP, Huth NI, deVoil PG, et al (2014) APSIM – Evolution towards a new generation of agricultural systems simulation. Env Model Soft. doi: http://dx.doi.org/10.1016/j.envsoft.2014.07.009

Howard R, Carriquiry AL, Beavis WD (2017) Application of Response Surface Methods To Determine Conditions for Optimal Genomic Prediction. G3: Genes|Genomes|Genetics g3.117.044453. doi: 10.1534/g3.117.044453

Hua L, Wang DR, Tan L, et al (2015) LABA1 , a Domestication Gene Associated with Long, Barbed Awns in Wild Rice. Plant Cell 27:1875–1888. doi: 10.1105/tpc.15.00260

Huang X, Wei X, Sang T, et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967.

Hurtado-Lopez PX, Tessema BB, Schnabel SK, et al (2015) Understanding the genetic basis of potato development using a multi-trait QTL analysis. Euphytica 1–13. doi: 10.1007/s10681-015-1431-2

Hurtado P, Schnabel S, Zaban A, et al (2012) Dynamics of senescence-related QTLs in potato. Euphytica 183:289–302. doi: 10.1007/s10681-011-0464-4

Igartua E, Gracia MP, Lasa JM, et al (1998) The Spanish barley core collection. Genet Resour Crop Evol 45:475–481. doi: 10.1023/A:1008662515059

Inman-Bamber NG, Jackson PA, Stokes CJ, et al (2016) Sugarcane for water-limited environments: Enhanced capability of the APSIM sugarcane model for assessing traits for transpiration efficiency and root water supply. F Crop Res 196:112–123. doi: 10.1016/j.fcr.2016.06.013

Isidro J, Jannink J-L, Akdemir D, et al (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet 128:145–158. doi: 10.1007/s00122-014-2418-4

Jamieson PD, Semenov MA, Brooking IR, Francis GS (1998) Sirius: a mechanistic model of wheat response to environmental variation. Eur J Agron 8:161–179. doi: http://dx.doi.org/10.1016/S1161-0301(98)00020-3

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Br Funct Genomics 9:166–177. doi: 10.1093/bfgp/elq001

Jansen J, van Hintum T (2007) Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. Theor Appl Genet 114:421–428. doi: 10.1007/s00122-006-0433-9

Janss L, de los Campos G, Sheehan N, Sorensen D (2012) Inferences from Genomic Models in Stratified Populations. Genetics 192:693–704. doi: 10.1534/genetics.112.141143

Jarquín D, Crossa J, Lacaze X, et al (2013) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 3:1–13. doi: 10.1007/s00122-013-2243-1

Jia QJ, Zhang JJ, Westcott S, et al (2009) GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Funct Integr Genomics 9:255–262. doi: 10.1007/s10142-009-0120-4

Jia Y, Jannink J-L (2012) Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. Genetics 192:1513–1522. doi: 10.1534/genetics.112.144246

Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. Genetics 201:759–768.

Jin J, Hua L, Zhu Z, et al (2016) GAD1 Encodes a Secreted Peptide That Regulates Grain Number, Grain Length, and Awn Development in Rice Domestication. Plant Cell 28:2453–2463. doi: 10.1105/tpc.16.00379

Johnson R, Wichern D (2007) Applied Multivariate Statistical Analysis, Sixth. Pearson International Edition

Jolliffe IT (2013) Principal Component Analysis. Springer New York

Jones JW, Hoogenboom G, Porter CH, et al (2003) The DSSAT cropping system model.

Josse J, Husson F (2012) Selecting the number of components in principal component analysis using cross-validation approximations. Comput Stat Data Anal 56:1869–1879.

Josse J, van Eeuwijk F, Piepho H-P, Denis J-B (2014) Another Look at Bayesian Analysis of AMMI Models for Genotype-Environment Data. J Agric Biol Environ Stat 19:240–257. doi: 10.1007/s13253-014-0168-z

Junker A, Muraya MM, Weigelt-Fischer K, et al (2015) Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. Front Plant Sci 5:770. doi: 10.3389/fpls.2014.00770

Kang HM, Zaitlen NA, Wade CM, et al (2008) Efficient Control of Population Structure in Model Organism Association Mapping. Genetics 178:1709 LP-1723.

Kang MS, Gauch HG (1996) Genotype-by-environment interaction. CRC press

Kauffman SA (1993) The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press

Keating BA, Carberry PS, Hammer GL, et al (2003) An overview of APSIM, a model designed for farming systems simulation. Eur J Agron 18:267–288. doi: http://dx.doi.org/10.1016/S1161-0301(02)00108-9

Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. J Agric Sci 103:123–135.

Khoury CK, Achicanoy HA (2016) Origins of food crops connect countries worldwide. Proc R Soc B 283:468–74. doi: 10.1126/science.174.4008.468

Kippes N, Zhu J, Chen A, et al (2014) Fine mapping and epistatic interactions of the vernalization gene VRN-D4 in hexaploid wheat. Mol Genet Genomics 289:47–62. doi: 10.1007/s00438-013-0788-y

Kizilkaya K, Fernando RL, Garrick DJ (2010) Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J Anim Sci 88:544–551. doi: 10.2527/jas.2009-2064

Kleinknecht K, Möhring J, Singh KP, et al (2013) Comparison of the Performance of Best Linear Unbiased Estimation and Best Linear Unbiased Prediction of Genotype Effects from Zoned Indian Maize Data. Crop Sci 53:1384–1391. doi: 10.2135/cropsci2013.02.0073

Knüpffer H (2009) Triticeae Genetic Resources in ex situ Genebank Collections BT - Genetics and Genomics of the Triticeae. In: Muehlbauer GJ, Feuillet C (eds). Springer US, New York, NY, pp 31–79

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29. doi: 10.1186/1746-4811-9-29

Kraakman ATW, Niks RE, Van den Berg PMMM, et al (2004) Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. Genetics 168:435–446. doi: 10.1534/genetics.104.026831

Kuehl RO (2000) Design of Experiments: Statistical Principles of Research Design and Analysis. Duxbury/Thomson Learning

Kuijken RP, van Eeuwijk FA van, Marcelis LFM, Bouwmeester H (2015) Root phenotyping: from component trait in the lab to breeding.

Laidig F, Drobek T, Meyer U (2008) Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breed 127:541–547. doi: 10.1111/j.1439-0523.2008.01564.x

Lamsal A, Welch SM, Jones JW, et al (2017) Efficient crop model parameter estimation and site characterization using large breeding trial data sets. Agric Syst 157:170–184. doi: 10.1016/j.agsy.2017.07.016

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits.

Langer SM, Longin CFH, Würschum T (2014) Flowering time control in European winter wheat. Front Plant Sci 5:537. doi: 10.3389/fpls.2014.00537

Laperche A, Brancourt-Hulmel M, Heumez E, et al (2007) Using genotype × nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115:399–415. doi: 10.1007/s00122-007-0575-4

Lee D-J, Durbán M (2011) P-spline ANOVA-type interaction models for spatio-temporal smoothing. Stat Model 11:49–69.

Lehermeier C, Schön C-C, de los Campos G (2015) Assessment of genetic heterogeneity in structured plant breeding populations using multivariate whole-genome regression models.

Lerner IM (1954) Genetic Homeostasis. Oliver and Boyd

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinforma 25:1754–1760. doi: 10.1093/bioinformatics/btp324

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 95:221–227.

Li X, Duan X, Jiang H, et al (2006) Genome-Wide Analysis of Basic / Helix-Loop-Helix Transcription Factor Family in Rice and Arabidopsis. Plant Physiol 141:1167–1184. doi: 10.1104/pp.106.080580.2001

Li Z, Sillanpää MJ (2015) Dynamic Quantitative Trait Locus Analysis of Plant Phenomic Data. Trends Plant Sci 20:822–833. doi: 10.1016/j.tplants.2015.08.012

Liller CB, Neuhaus R, Von Korff M, et al (2015) Mutations in barley row type genes have pleiotropic effects on shoot branching. PLoS One 10:1–20. doi: 10.1371/journal.pone.0140246

Liller CB, Walla A, Boer MP, et al (2017) Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. Theor Appl Genet 130:269–281. doi: 10.1007/s00122-016-2807-y

Lin CS, Binns MR (1988) A method of analyzing cultivar x location x year experiments: a new stability parameter. Theor Appl Genet 76:425–430. doi: 10.1007/bf00265344

Lin CS, Binns MR, Lefkovitch LP (1986) Stability Analysis: Where Do We Stand? Crop Sci 26:894–900. doi: 10.2135/cropsci1986.0011183X002600050012x

Lippert C, Quon G, Kang EY, et al (2013) The benefits of selecting phenotype-specific variants for applications of mixed models in genomics. Sci Rep 3:1815. doi: 10.1038/srep01815

Liseron-Monfils C, Ware D (2015) Revealing gene regulation and associations through biological networks. Curr Plant Biol 3–4:30–39. doi: 10.1016/j.cpb.2015.11.001

Listgarten J, Lippert C, Kadie CM, et al (2011) Improved linear mixed models for genome-wide association studies. Nat Methods 9:525–526. doi: 10.1038/nmeth.2037

Löffler CM, Wei J, Fast T, et al (2005) Classification of Maize Environments Using Crop Simulation and Geographic Information Systems. Crop Sci 45:1708–1716. doi: 10.2135/cropsci2004.0370

Longin CF, 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. doi: 10.1007/s00122-015-2505-1

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3 Genes|Genomes|Genetics. doi: 10.1534/g3.114.016097

Lund MS, Sorensen P, Madsen P, Jaffrezic F (2008) Detection and modelling of time-dependent QTL in animal populations. Genet Sel Evol 40:177–194. doi: 10.1051/gse:2007043

Luo J, Liu H, Zhou T, et al (2013) An-1 Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. Plant Cell 25:3360–3376. doi: 10.1105/tpc.113.113589

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Sunderland, MA

Ma CX, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161:1751–1762.

Macgregor S, Knott SA, White I, Visscher PM (2005) Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. Genetics 171:1365–1376. doi: 10.1534/genetics.105.043828

Mackay I, Horwell A, Garner J, et al (2011) Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. Theor Appl Genet 122:225–238. doi: 10.1007/s00122-010-1438-y

Magney TS, Eitel JUH, Huggins DR, Vierling LA (2016) Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. Agric For Meteorol 217:46–60. doi: http://dx.doi.org/10.1016/j.agrformet.2015.11.009

Magwa RA, Zhao H, Yao W, et al (2016) Genomewide association analysis for awn length linked to the seed shattering gene qSH1 in rice. J Genet 95:639–646. doi: 10.1007/s12041-016-0679-1

Makumburage GB, Richbourg HL, LaTorre KD, et al (2013) Genotype to Phenotype Maps: Multiple Input Abiotic Signals Combine to Produce Growth Effects via Attenuating

Signaling Interactions in Maize. G3 Genes Genom Genet 3:2195–2204. doi: 10.1534/g3.113.008573

Malosetti M, Abadie T (2001) Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. Genet Resour Crop Evol 48:381–390. doi: 10.1023/A:1012003611371

Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting Responses in Multiple Environments: Issues in Relation to Genotype × Environment Interactions. Crop Sci 56:2210–2222. doi: 10.2135/cropsci2015.05.0311

Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4:1–17. doi: 10.3389/fphys.2013.00044

Malosetti M, Ribaut J, 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. doi: 10.1007/s10681-007-9594-0

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A Mixed-Model Approach to Association Mapping Using Pedigree Information With an Illustration of Resistance to Phytophthora infestans in Potato. Genetics 175:879–889. doi: 10.1534/genetics.105.054932

Malosetti M, van Eeuwijk F, Boer M, et al (2011) Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. Theor Appl Genet 122:1605–1616. doi: 10.1007/s00122-011-1558-z

Malosetti M, Visser RGF, Celis-Gamboa C, Eeuwijk FA (2006) QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. Theor Appl Genet 113:288–300. doi: 10.1007/s00122-006-0294-2

Malosetti M, Voltas J, Romagosa I, et al (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139–145. doi: 10.1023/B:EUPH.0000040511.46388.ef

Mandel J (1969) The partitioning of interaction in analysis of variance. J Res Natl Bur Stand Ser B 73:309–328.

Mangin B, Siberchicot A, Nicolas S, et al (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity (Edinb) 108:285–291. doi: 10.1038/hdy.2011.73

Marti J, Bort J, Slafer GA, Araus JL (2007) Can wheat yield be assessed by early measurements of Normalized Difference Vegetation Index? Ann Appl Biol 150:253–257. doi: 10.1111/j.1744-7348.2007.00126.x

Mascher M, Gundlach H, Himmelbach A, et al (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544:427–433.

Mascher M, Richmond TA, Gerhardt DJ, et al (2013) Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. Plant J 76:494–505. doi: 10.1111/tpj.12294

Mathews K, Malosetti M, Chapman S, et al (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117:1077–1091. doi: 10.1007/s00122-008-0846-8

Mathews KL, Trethowan R, Milgate AW, et al (2011) Indirect selection using reference and probe genotype performance in multi-environment trials. Crop Pasture Sci 62:313–327.

Mathieson I, McVean G (2012) Differential confounding of rare and common variants in spatially structured populations. Nat Genet 44:243–6. doi: 10.1038/ng.1074

Maurer A, Draba V, Pillen K (2016) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot 67:2507–2518. doi: 10.1093/jxb/erw070

Mayer KFX, Waugh R, Langridge P, et al (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi: 10.1038/nature11543

McCarthy MIMI, Abecasis GRGR, Cardon LRLR, et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356–69. doi: 10.1038/nrg2344

McKinney BA, Pajewski NM (2012) Six degrees of epistasis: Statistical network models for GWAS. Front Genet 2:1–6. doi: 10.3389/fgene.2011.00109

McMullen MD, Kresovich S, Villeda HS, et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–40. doi: 10.1126/science.1174320

Meinshausen N, Bühlmann P (2010) Stability selection. J R Stat Soc Ser B (Statistical Methodol 72:417–473. doi: 10.1111/j.1467-9868.2010.00740.x

Messina CD, Podlich D, Dong Z, et al (2011) Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. J Exp Bot 62:855–868. doi: 10.1093/jxb/erq329

Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11:31–46.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829.

Mikolajczak K, Ogrodowicz P, Gudys K, et al (2016) Quantitative trait loci for yield and yield-related traits in spring barley populations derived from crosses between European and Syrian cultivars. PLoS One. doi: 10.1371/journal.pone.0155938

Millet E (2016) Genetic variability of maize grain yield as affected by drought and heat: analysis of a multi-site network of experiments.

Millet E, Welcker C, Kruijer W, et al (2016) Genome-wide analysis of yield in Europe: allelic effects as functions of drought and heat scenarios. Plant Physiol 172:pp.00621.2016. doi: 10.1104/pp.16.00621

Milliken GA, Johnson DE (1989) Analysis of Messy Data: Nonreplicated Experiments. Taylor & Francis

Mochida K, Uehara-Yamaguchi Y, Yoshida T, et al (2011) Global landscape of a Co-expressed gene network in barley and its application to gene discovery in triticeae crops. Plant Cell Physiol 52:785–803. doi: 10.1093/pcp/pcr035

Möhring J, Piepho H-P (2009) Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. Crop Sci 49:1977–1988. doi: 10.2135/cropsci2009.02.0083

Monneveux P, Rekika D, Acevedo E, Merah O (2006) Effect of drought on leaf gas exchange, carbon isotope discrimination, transpiration efficiency and productivity in field grown durum wheat genotypes. Plant Sci 170:867–872. doi: 10.1016/j.plantsci.2005.12.008

Montesinos-López A, Montesinos-López OA, Cuevas J, et al (2017) Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. Plant Methods 13:62. doi: 10.1186/s13007-017-0212-4

Moser G, Lee SH, Hayes BJ, et al (2015) Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. PLoS Genet 11:1–22. doi: 10.1371/journal.pgen.1004969

Muir W, Nyquist WE, Xu S (1992) Alternative partitioning of the genotype-by-environment interaction. Theor Appl Genet 84:193–200. doi: 10.1007/bf00224000

Muller S, Scealy JL, Welsh AH (2013) Model Selection in Linear Mixed Models. Stat Sci 28:135–167. doi: 10.1214/12-STS410

Muñoz-Amatriaín M, Cuesta-Marcos A, Endelman JB, et al (2014) The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. PLoS One 9:1–13. doi: 10.1371/journal.pone.0094688

Nabugoomu F, Kempton RA, Talbot M (1999) Analysis of Series of Trials Where Varieties Differ in Sensitivity to Locations. J Agric Biol Environ Stat 4:310–325. doi: 10.2307/1400388

Navabi A, Yang R-C, Helm J, Spaner DM (2006) Can Spring Wheat-Growing Megaenvironments in the Northern Great Plains Be Dissected for Representative Locations or Niche-Adapted Genotypes? Crop Sci 46:1107–1116. doi: 10.2135/cropsci2005.06-0159

Nelsen RB (2013) An Introduction to Copulas. Springer New York

Neto EC, Ferrara CT, Attie AD, Yandell BS Inferring Causal Phenotype Networks From Segregating Populations. Genetics 179:1089–1100.

Neto EC, Keller MP, Attie AD, Yandell BS (2010) Causal Graphical Models in Systems Genetics. A Unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. Ann Appl Stat 4:320–339.

Neumann K, Klukas C, Friedel S, et al (2015) Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis. Plant, Cell Environ 38:1980–1996. doi: 10.1111/pce.12516

Nicotra AB, Davidson A (2010) Adaptive phenotypic plasticity and plant water use. Funct. Plant Biol. 37:117–127.

Oakey H, Verbyla A, Pitchford W, et al (2006) Joint modeling of additive and non-additive genetic line effects in single field trials. Theor Appl Genet 113:809–819. doi: 10.1007/s00122-006-0333-z

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL (2013a) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet 126:289–305. doi: 10.1007/s00122-012-1971-y

Odong TL, van Heerwaarden J, Jansen J, et al (2011) Statistical Techniques for Defining Reference Sets of Accessions and Microsatellite Markers. Crop Sci 51:2401–2411. doi: 10.2135/cropsci2011.02.0095

Odong TL, van Heerwaarden J, van Hintum TJL, et al (2013b) Improving Hierarchical Clustering of Genotypic Data via Principal Component Analysis. Crop Sci 53:1546–1554. doi: 10.2135/cropsci2012.04.0215

Olson KM, VanRaden PM, Tooker ME (2012) Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J Dairy Sci 95:5378–5383. doi: http://dx.doi.org/10.3168/jds.2011-5006

Onogi A, Watanabe M, Mochizuki T, et al (2016) Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theor Appl Genet 129:805–817. doi: 10.1007/s00122-016-2667-5

Orr HA (2010) Fitness and its role in evolutionary genetics. Nat Rev Genet 10:531–539. doi: 10.1038/nrg2603.Fitness

Parry MAJ, Reynolds M, Salvucci ME, et al (2011) Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. J Exp Bot 62:453–467. doi: 10.1093/jxb/erq304

Pasam RK, Sharma R, Malosetti M, et al (2012) Genome-wide association studies for agronomical traits in a world-wide spring barley collection. BMC Plant Biol. doi: 1610.1186/1471-2229-12-16

Passioura JB (2012) Phenotyping for drought tolerance in grain crops: when is it useful to breeders? Funct Plant Biol 39:851–859. doi: http://dx.doi.org/10.1071/FP12079

Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. PLoS Genet 2:e190. doi: 10.1371/journal.pgen.0020190

Pauli D, Chapman SC, Bart R, et al (2016) The quest for understanding phenotypic variation via integrated approaches in the field environment. Plant Physiol 172:pp.00592.2016. doi: 10.1104/pp.16.00592

Perez-Elizalde S, Jarquin D, Crossa J (2012) A General Bayesian Estimation Method of Linear–Bilinear Models Applied to Plant Breeding Trials With Genotype × Environment Interaction. J Agric Biol Environ Stat 17:15–37. doi: 10.1007/s13253-011-0063-9

Pérez-Enciso M, Forneris N, de los Campos G, Legarra A (2017) Evaluating Sequence-Based Genomic Prediction with an Efficient New Simulator. Genetics 205:939 LP-953.

Pérez-Enciso M, Legarra A (2016) A combined coalescence gene-dropping tool for evaluating genomic selection in complex scenarios (ms2gs). J Anim Breed Genet 133:85–91. doi: 10.1111/jbg.12200

Pérez-Rodríguez P, Crossa J, Bondalapati K, et al (2015) A Pedigree-Based Reaction Norm Model for Prediction of Cotton Yield in Multienvironment Trials. Crop Sci 55:1143–1151. doi: 10.2135/cropsci2014.08.0577

Perez P, De los Campos G, Crossa J, Gianola D (2010) Genomic-Enabled Prediction Based on Molecular Markers and Pedigree Using the Bayesian Linear Regression Package in R. Plant Genome 3:106–116. doi: 10.3835/plantgenome2010.04.0005

Perry EM, Morse-Mcnabb EM, Nuttall JG, et al (2014) Managing wheat from space: Linking MODIS NDVI and crop models for predicting Australian dryland wheat biomass. IEEE J Sel Top Appl Earth Obs Remote Sens 7:3724–3731. doi: 10.1109/JSTARS.2014.2323705

Piepho H-P (2000) A Mixed-Model Approach to Mapping Quantitative Trait Loci in Barley on the Basis of Multiple Environment Data. Genetics 156:2043–2050.

Piepho H-P, Laidig F, Drobek T, Meyer U (2014) Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials. Theor Appl Genet 127:1009–1018. doi: 10.1007/s00122-014-2275-1

Piepho H-P, Möhring J, Schulz-Streeck T, Ogutu JO (2012) A stage-wise approach for the analysis of multi-environment trials. Biometrical J 54:844–860. doi: 10.1002/bimj.201100219

Piepho H-PP (1998) Methods for Comparing the Yield Stability of Cropping Systems. J Agron Crop Sci 180:193–213. doi: 10.1111/j.1439-037X.1998.tb00526.x

Piepho HP (2009) Ridge Regression and Extensions for Genomewide Selection in Maize. 1165–1176.

Piepho HP (1997) Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. Biometrics 53:761–766. doi: 10.2307/2533976

Piepho HP, Möhring J (2005) Best Linear Unbiased Prediction of Cultivar Effects for Subdivided Target Regions. Crop Sci 45:1151–1159. doi: 10.2135/cropsci2004.0398

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. doi: 10.1007/s10681-007-9449-8

Pigliucci M (2001) Phenotypic plasticity: beyond nature and nurture. JHU Press

Podlich DW, Cooper M (1998) QU-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653. doi: 10.1093/bioinformatics/14.7.632

Podlich DW, Cooper M (1999) Modelling Plant Breeding Programs as Search Strategies on a Complex Response Surface. In: McKay B, Yao X, Newton C, et al. (eds) Simulated Evolution and Learning SE - 23. Springer Berlin Heidelberg, pp 171–178

Podlich DW, Cooper M, Basford KE, Geiger HH (1999) Computer simulation of a selection strategy to accommodate genotype environment interactions in a wheat recurrent

selection programme. Plant Breed 118:17–28. doi: 10.1046/j.1439-0523.1999.118001017.x

Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Gen 5:92–102. doi: 10.3835/plantgenome2012.05.0005

Poland J, Endelman J, Dawson J, et al (2012a) Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome J. doi: 10.3835/plantgenome2012.06.0006

Poland J, Endelman J, Dawson J, et al (2012b) Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Gen 5:103–113. doi: 10.3835/plantgenome2012.06.0006

Porter JR, Gawith M (1999) Temperatures and the growth and development of wheat: a review. Eur J Agron 10:23–36. doi: http://dx.doi.org/10.1016/S1161-0301(98)00047-1

Prasanna B, Araus J, Crossa J, et al (2013) High-Throughput and Precision Phenotyping for Cereal Breeding Programs. In: Gupta PK, Varshney RK (eds) Cereal Genomics II. Springer Netherlands, pp 341–374

Price AL, Price AL, Patterson NJ, et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–9. doi: 10.1038/ng1847

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959.

Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci 95:389–400. doi: http://dx.doi.org/10.3168/jds.2011-4338

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing,.

Rajaram S, Van Ginkel M, Fischer RA (1993) CIMMYT's wheat breeding mega-environments (ME). In: 8th International Wheat Genetics Symposium. pp 1101–1106

Ramsay L, Comadran J, Druka A, et al (2011) INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. Nat Genet 43:169–172.

Rebai A, Goffinet B, Mangin B (1995) Comparing power of different methods for QTL detection. Biometrics 51:87–99. doi: 10.2307/2533317

Rebetzke GJ, Chenu K, Biddulph B, et al (2012a) A multisite managed environment facility for targeted trait and germplasm phenotyping. Funct Plant Biol 40:1–13. doi: http://dx.doi.org/10.1071/FP12180

Rebetzke GJ, Rattey AR, Farquhar GD, et al (2012b) Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. Funct Plant Biol 40:14–33.

Reif J, Maurer H, Korzun V, et al (2011) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theor Appl Genet 123:283–292. doi: 10.1007/s00122-011-1583-y

Reymond M, Muller B, Leonardi A, et al (2003) Combining Quantitative Trait Loci Analysis and an Ecophysiological Model to Analyze the Genetic Variability of the Responses of Maize Leaf Growth to Temperature and Water Deficit. Plant Physiol 131:664–675. doi: 10.1104/pp.013839

Reymond M, Muller B, Tardieu F (2004) Dealing with the genotype×environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. J Exp Bot 55:2461–2472. doi: 10.1093/jxb/erh200

Reynolds M, Bonnett D, Chapman SC, et al (2011) Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. J Exp Bot 62:439–452. doi: 10.1093/jxb/erq311

Reynolds M, Calderini DF, Condon AG, Vargas M (2007) Association of source/sink traits with yield, biomass and radiation use efficiency among random sister lines from three wheat crosses in a high-yield environment. J Agric Sci 145:3–16. doi: 10.1017/S0021859607006831

Reynolds M, Foulkes MJ, Slafer GA, et al (2009a) Raising yield potential in wheat. J Exp Bot 60:1899–1918. doi: 10.1093/jxb/erp016

Reynolds M, Manes Y, Izanloo A, Langridge P (2009b) Phenotyping approaches for physiological breeding and gene discovery in wheat. Ann App Biol 155:309–320. doi: 10.1111/j.1744-7348.2009.00351.x

Reynolds MP, Trethowan R, Crossa J, et al (2002) Physiological factors associated with genotype by environment interaction in wheat. F Crop Res 75:139–160.

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220.

Rincent R, Kuhn E, Monod H, et al (2017) Optimization of multi-environment trials for genomic selection based on crop models. Theor Appl Genet 1–18. doi: 10.1007/s00122-017-2922-4

Rincent R, Laloë D, Nicolas S, et al (2012) Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (Zea mays L.). Genetics 192:715–728. doi: 10.1534/genetics.112.141473

Rincent R, Moreau L, Monod H, et al (2014a) Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics 197:375–387. doi: 10.1534/genetics.113.159731

Rincent R, Nicolas S, Bouchet S, et al (2014b) Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. Theor Appl Genet 127:2313–2331. doi: 10.1007/s00122-014-2379-7

Ritchie JT (1985) Description and performance of CERES Wheat: A user-oriented wheat yield model. ARS wheat yield Proj 159–175.

Rodríguez-Álvarez MX, Boer MP, Eeuwijk FA van, Eilers PHC (2016) Spatial Models for Field Trials.

Romagosa I, Borràs-Gelonch G, Slafer G, Eeuwijk F (2013) Genotype by Environment Interaction and Adaptation. In: Christou P, Savin R, Costa-Pierce B, et al. (eds) Sustainable Food Production. Springer New York, pp 846–870

Romagosa I, Fox PN (1993) Genotype× environment interaction and adaptation. In: Hayward MD, Bosemark NO, Romagosa T (eds) Plant Breeding. Springer, pp 373–390

Romagosa I, Ullrich S, Han F, Hayes P (1996) Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. 93:30–37. doi: 10.1007/bf00225723

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci U S A 104 Suppl:8641–8. doi: 10.1073/pnas.0700643104

Russell J, Mascher M, Dawson IK, et al (2016) Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. Nat Genet 48:1024–1030.

Rutkoski J, Poland J, Mondal S, et al (2016) Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat.

Saatchi M, McClure MC, McKay SD, et al (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet Sel Evol 43:40. doi: 10.1186/1297-9686-43-40

Sadras VO, Calderini D (2014) Crop Physiology: Applications for Genetic Improvement and Agronomy. Elsevier Science

Sadras VO, Lawson C (2011) Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. Crop Pasture Sci 62:533–549. doi: http://dx.doi.org/10.1071/CP11060

Sadras VO, Rebetzke GJ (2013) Plasticity of wheat grain yield is associated with plasticity of ear number. Crop Pasture Sci 64:234–243. doi: http://dx.doi.org/10.1071/CP13117

Sadras VO, Rebetzke GJ, Edmeades GO (2013) The phenotype and the components of phenotypic variance of crop traits. F Crop Res 154:255–259. doi: http://dx.doi.org/10.1016/j.fcr.2013.10.001

Sadras VO, Reynolds MP, de la Vega AJ, et al (2009) Phenotypic plasticity of yield and phenology in wheat, sunflower and grapevine. F Crop Res 110:242–250.

Sannemann W, Huang BE, Mathew B, Léon J (2015) Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Mol Breed. doi: 10.1007/s11032-015-0284-7

Sarkar S (1999) From the Reaktionsnorm to the Adaptive Norm: The Norm of Reaction, 1909–1960. Biol Philos 14:235–252. doi: 10.1023/a:1006690502648

SAS Institute (2015) SAS/STAT.

Schlichting CD, Pigliucci M (1998) Phenotypic evolution: a reaction norm perspective. Sinauer Associates Incorporated

Schulz-Streeck T, Ogutu JO, Karaman Z, et al (2012) Genomic Selection using Multiple Populations. Crop Sci 52:2453–2461. doi: 10.2135/cropsci2012.03.0160

Scutari M, Howell P, Balding DJ, Mackay I (2014) Multiple Quantitative Trait Analysis Using Bayesian Networks. Genetics 198:129–137. doi: 10.1534/genetics.114.165704

Searle SR, Casella G, McCulloch CE (2009) Variance Components. Wiley

Sellers W.D. (1965) Physical Climatology, University. Chicago, Illinois

Shukla GK (1972) Some statistical aspects of partitioning genotype-environmental components of variability. Heredity (Edinb) 29:237–245.

Sillanpää MJ, Pikkuhookana P, Abrahamsson S, et al (2012) Simultaneous estimation of multiple quantitative trait loci and growth curve parameters through hierarchical Bayesian modeling. Heredity (Edinb) 108:134–146. doi: 10.1038/hdy.2011.56

Singh V, van Oosterom EJ, Jordan DR, et al (2010) Morphological and architectural development of root systems in sorghum and maize. Plant Soil 333:287–299. doi: 10.1007/s11104-010-0343-0

Slafer GA (2003) Genetic basis of yield as viewed from a crop physiologist's perspective. Ann Appl Biol 142:117–128. doi: 10.1111/j.1744-7348.2003.tb00237.x

Slafer GA, Andrade FH (1993) Physiological attributes related to the generation of grain yield in bread wheat cultivars released at different eras. F Crop Res 31:351–367.

Slafer GA, Araus JL, Royo C, Del Moral LFG (2005) Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. Ann Appl Biol 146:61–70. doi: 10.1111/j.1744-7348.2005.04048.x

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: breking the barriers. CIMMYT,

Slafer GA, Rawson HM (1994) Sensitivity of Wheat Phasic Development to Major Environmental Factors: a Re-Examination of Some Assumptions Made by Physiologists and Modellers. Funct Plant Biol 21:393–426. doi: http://dx.doi.org/10.1071/PP9940393

Slafer GA, Savin R (1994) Source—sink relationships and grain mass at different positions within the spike in wheat. F Crop Res 37:39–49. doi: 10.1016/0378-4290(94)90080-9

Slafer GA, Savin R, Sadras VO (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. F Crop Res 157:71–83. doi: http://dx.doi.org/10.1016/j.fcr.2013.12.004

Smith A, Cullis B, Gilmour A (2001a) Applications: The Analysis of Crop Variety Evaluation Data in Australia. Aust N Z J Stat 43:129–145. doi: 10.1111/1467-842x.00163

Smith A, Cullis B, Thompson R (2001b) Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. Biometrics 57:1138–1147. doi: 10.1111/j.0006-341X.2001.01138.x

Smith AB, Cullis BR, Appels R, et al (2001c) The statistical analysis of quality traits in plant improvement programs with application to the mapping of milling yield in wheat. Aust J Agric Res 52:1207–1219. doi: http://dx.doi.org/10.1071/AR01058

Smith AB, Cullis BR, Thomson R, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agr Sci 143:449–462. doi: doi:10.1017/S0021859605005587

Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. J Agr Sci 144:393–409. doi: doi:10.1017/S0021859606006319

Snape JW, Butterworth K, Whitechurch E, Worland AJ (2001) Waiting for fine times: genetics of flowering time in wheat. Euphytica 119:185–190. doi: 10.1023/a:1017594422176

Snape JW, Foulkes MJ, Simmonds J, et al (2007) Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. Euphytica 154:401–408. doi: 10.1007/s10681-006-9208-2

Sokal RR (1958) A statistical method for evaluating systematic relationships. Univ Kans Sci Bull 38:1409–1438.

Sorrells ME (2015) Genomic Selection in Plants: Empirical Results and Implications for Wheat Breeding BT - Advances in Wheat Genetics: From Genome to Field: Proceedings of the 12th International Wheat Genetics Symposium. In: Ogihara Y, Takumi S, Handa H (eds). Springer Japan, Tokyo, pp 401–409

Speed D, Balding DJ (2014) MultiBLUP: improved SNP-based prediction for complex traits. Genome Res 24:1550–1557.

Spindel J, Begum H, Akdemir D, et al (2015) Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Bre. PLoS Genet 11:e1004982.

Spindel J, Wright M, Chen C, et al (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 126:2699–2716. doi: 10.1007/s00122-013-2166-x

Stanfill B, Mielenz H, Clifford D, Thorburn P (2015) Simple approach to emulating complex computer models for global sensitivity analysis. Environ Model Softw 74:140–155. doi: 10.1016/j.envsoft.2015.09.011

Stange M, Schrag TA, Utz HF, et al (2013) High-density linkage mapping of yield components and epistatic interactions in maize with doubled haploid lines from four crosses. Mol Breed 32:533–546. doi: 10.1007/s11032-013-9887-z

Stephens M (2013) A Unified Framework for Association Analysis with Multiple Related Phenotypes. PLoS One 8:e65245. doi: 10.1371/journal.pone.0065245

Stinchcombe JR, Kirkpatrick M (2012) Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. Trends Ecol Evol 27:637–647. doi: 10.1016/j.tree.2012.07.002

Sukumaran S, Lopes MS, Dreisigacker S, et al (2016) Identification of Earliness Per Se Flowering Time Locus in Spring Wheat through a Genome-Wide Association Study. Crop Sci 56:2672–2962. doi: 10.2135/cropsci2016.01.0066

Sun J, Rutkoski JE, Poland JA, et al (2017) Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. Plant Genome. doi: 10.3835/plantgenome2016.11.0111

Sun L, Wu R (2015) Mapping complex traits as a dynamic system. Phys Life Rev. doi: 10.1016/j.plrev.2015.02.007

Tanaka W, Toriba T, Ohmori Y, et al (2012) The YABBY Gene TONGARI-BOUSHI1 Is Involved in Lateral Organ Development and Maintenance of Meristem Organization in the Rice Spikelet. Plant Cell 24:80–95. doi: 10.1105/tpc.111.094797

Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9–14. doi: http://dx.doi.org/10.1016/S1360-1385(02)00008-0

Tardieu F, Reymond M, Muller B, et al (2005) Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. Aust J Agric Res 56:937–946. doi: http://dx.doi.org/10.1071/AR05156

Tardieu F, Tuberosa R (2010) Dissection and modelling of abiotic stress tolerance in plants. Curr Opin Plant Biol 13:206–212. doi: http://dx.doi.org/10.1016/j.pbi.2009.12.012

Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring Coalescence Times From DNA Sequence Data. Genetics 145:505 LP-518.

Tayeh N, Klein A, Le Paslier M-C, et al (2015) Genomic Prediction in Pea: Effect of Marker Density and Training Population Size and Composition on Prediction Accuracy. Front. Plant Sci. 6:

Taylor JD, Verbyla AP, Cavanagh C, Newberry M (2012) Variable Selection in Linear Mixed Models Using an Extended Class of Penalties. Aust N Z J Stat 54:427–449. doi: 10.1111/j.1467-842X.2012.00687.x

Technow F, Messina CD, Totir LR, Cooper M (2015) Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS One 10:1–20. doi: 10.1371/journal.pone.0130855

Technow F, Totir LR (2015) Using Bayesian Multilevel Whole Genome Regression Models for Partial Pooling of Training Sets in Genomic Prediction. G3 Genes|Genomes|Genetics 5:1603 LP-1612.

Tester M, Langridge P (2010) Breeding Technologies to Increase Crop Production in a Changing World. Science (80-) 327:818–822. doi: 10.1126/science.1183700

The Breeding Management System Version 3.0.8 (2015) The Integrated Breeding Platform.

Thoen MPM, Davila Olivas NH, Kloth KJ, et al (2017) Genetic architecture of plant stress resistance: Multi-trait genome-wide association mapping. New Phytol 213:1346–1362. doi: 10.1111/nph.14220

Thomson Reuters (2015) Web of Science. http://apps.webofknowledge.com/.

Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B 58:267–288. doi: 10.2307/2346178

Toriba T, Hirano HY (2014) The DROOPING LEAF and OsETTIN2 genes promote awn development in rice. Plant J 77:616–626. doi: 10.1111/tpj.12411

Torres-Sosa C, Huang S, Aldana M (2012) Criticality Is an Emergent Property of Genetic Networks that Exhibit Evolvability. PLoS Comput Biol. doi: 10.1371/journal.pcbi.1002669

Trethowan RM, Crossa J, van Ginkel M, Rajaram S (2001) Relationships among Bread Wheat International Yield Testing Locations in Dry Areas. Crop Sci 41:1461–1469. doi: 10.2135/cropsci2001.4151461x

Trevaskis B, Hemming MN, Dennis ES, Peacock WJ (2007) The molecular basis of vernalization-induced flowering in cereals. Trends Plant Sci 12:352–357. doi: 10.1016/j.tplants.2007.06.010

Tucker G, Price AL, Berger B (2014) Improving the power of GWAS and avoiding confounding from population stratification with PC-select. Genetics 197:1045–1049. doi: 10.1534/genetics.114.164285

Turner A, Beales J, Faure S, et al (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science (80-) 310:1031–1034. doi: 10.1126/science.1117619

Uauy C (2017) Wheat genomics comes of age. Curr Opin Plant Biol 36:142–148. doi: 10.1016/j.pbi.2017.01.007

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, et al (2013) A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. PLoS One 8:e62355. doi: 10.1371/journal.pone.0062355

Uptmoor R, Pillen K, Matschegewski C (2017) Combining genome-wide prediction and a phenology model to simulate heading date in spring barley. F Crop Res 202:84–93. doi: 10.1016/j.fcr.2016.08.006

Utz HF, Melchinger AE, Schön CC (2000) Bias and Sampling Error of the Estimated Proportion of Genotypic Variance Explained by Quantitative Trait Loci Determined From Experimental Data in Maize Using Cross Validation and Validation With Independent Samples. Genetics 154:1839–1849.

van Binsbergen R, Bink M, Finkers HJ, et al (2016) Utilizing low-coverage sequence data in tomato recombinant inbred lines (S. lycopersicum x S. pimpinellifolium). In: Plant Breeding. p 19

van Binsbergen R, Bink MC, Calus MP, et al (2014) Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Genet Sel Evol 46:41. doi: 10.1186/1297-9686-46-41

van der Heijden G, Song Y, Horgan G, et al (2012) SPICY: towards automated phenotyping of large pepper plants in the greenhouse. F Plant Biol 39:870–877. doi: http://dx.doi.org/10.1071/FP12019

van Eeuwijk F (2015) How to dissect complex traits and how to choose suitable mapping resources for system genetics?: Comment on "Mapping complex traits as a dynamic system" by L. Sun and R. Wu. Phys Life Rev. doi: 10.1016/j.plrev.2015.04.035

van Eeuwijk FA (1992) Interpreting genotype-by-environment interaction using redundancy analysis. Theor Appl Genet 85:89–100. doi: 10.1007/BF00223849

van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13:193–205. doi: 10.1016/j.pbi.2010.01.001

van Eeuwijk FA, Bustos-Korts D V, Malosetti M (2016) What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? Crop Sci 56:2119–2140. doi: 10.2135/cropsci2015.06.0375

van Eeuwijk FA, Denis JB, Kang MS (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: Kang MS, Gauch Jr. HG (eds) Genotype-by-environment interaction. Taylor & Francis Group, pp 15–50

van Eeuwijk FA, Malosetti M, Yin X, et al (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust J Agric Res 56:883–894. doi: http://dx.doi.org/10.1071/AR05153

Van Ittersum MK, Leffelaar PA, Van Keulen H, et al (2003) On approaches and applications of the Wageningen crop models. Eur J Agron 18:201–234. doi: 10.1016/S1161-0301(02)00106-5

Van Oijen M, Höglind M (2016) Toward a Bayesian procedure for using process-based models in plant breeding, with application to ideotype design. Euphytica 207:627–643. doi: 10.1007/s10681-015-1562-5

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–23. doi: 10.3168/jds.2007-0980

Vargas M, Crossa J, van Eeuwijk FA, et al (1999) Using Partial Least Squares Regression, Factorial Regression, and AMMI Models for Interpreting Genotype × Environment Interaction. Crop Sci 39:955–967. doi: 10.2135/cropsci1999.0011183X003900040002x

Vargas M, Glaz B, Alvarado G, et al (2015) Analysis and Interpretation of Interactions in Agricultural Research. Agron J 107:748–762. doi: 10.2134/agronj13.0405

Varshney RK, Paulo MJ, Grando S, et al (2012) Genome wide association analyses for drought tolerance related traits in barley (Hordeum vulgare L.). F Crop Res 126:171–180. doi: 10.1016/j.fcr.2011.10.008

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. PLoS Biol 12:e1001883. doi: 10.1371/journal.pbio.1001883

Velazco JG, Rodríguez-Álvarez MX, Boer MP, et al (2017) Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. Theor Appl Genet 130:1375–1392. doi: 10.1007/s00122-017-2894-4

Verbeke G (1997) Linear Mixed Models for Longitudinal Data BT - Linear Mixed Models in Practice: A SAS-Oriented Approach. In: Verbeke G, Molenberghs G (eds). Springer New York, New York, NY, pp 63–153

Verbyla A, Cullis B (2012) Multivariate whole genome average interval mapping: QTL analysis for multiple traits and/or environments. Theor Appl Genet 125:933–953. doi: 10.1007/s00122-012-1884-9

Vilhjálmsson BJ, Nordborg M (2013) The nature of confounding in genome-wide association studies. Nat Rev Genet 14:1–2. doi: 10.1038/nrg3382

Voltas J, Romagosa I, Lafarga A, et al (1999a) Genotype by environment interaction for grain yield and carbon isotope discrimination of barley in Mediterranean Spain. Aust J Agric Res 50:1263–1271. doi: http://dx.doi.org/10.1071/AR98137

Voltas J, van Eeuwijk FA, Araus JL, Romagosa I (1999b) Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain filling of barley II.: Grain growth. F Crop Res 62:75–84. doi: http://dx.doi.org/10.1016/S0378-4290(99)00007-6

VSN-International (2016) GenStat 18th Edition.

VSN-International (2015) ASReml-R. Version 3.0.

Walsh B, Lynch M Evolution and Selection of Quantitative Traits: II. Advanced Topics in Breeding and Evolution . http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html#2B,

Wang E, Robertson MJ, Hammer GL, et al (2002) Development of a generic crop model template in the cropping system model APSIM. Eur J Agron 18:121–140. doi: http://dx.doi.org/10.1016/S1161-0301(02)00100-4

Wang H, Paulo J, Kruijer W, et al (2015) Genotype-phenotype modeling considering intermediate level of biological variation: a case study involving sensory traits, metabolites and QTLs in ripe tomatoes. Mol Biosyst 11:3101–3110. doi: 10.1039/C5MB00477B

Wang H, van Eeuwijk FA (2014) A New Method to Infer Causal Phenotype Networks Using QTL and Phenotypic Information. PLoS One 9:e103997.

Wang J, Sun G, Ren X, et al (2016) QTL underlying some agronomic traits in barley detected by SNP markers. BMC Genet 17:103. doi: 10.1186/s12863-016-0409-y

Weber VS, Melchinger AE, Magorokosho C, et al (2012) Efficiency of managed-stress screening of elite maize hybrids under drought and low nitrogen for yield under rainfed conditions in Southern Africa. Crop Sci 52:1011–1020. doi: 10.2135/cropsci2011.09.0486

Weber WE, Westermann T (1994) Prediction of Yield for Specific Locations in German Winter-wheat Trials. Plant Breed 113:99–105. doi: 10.1111/j.1439-0523.1994.tb00711.x

Welch SM, Dong Z, Roe JL, Das S (2005) Flowering time control: gene network modelling and the link to quantitative genetics. Aust. J. Agric. Res. 56:919–936.

Welch SM, Roe JL, Dong Z (2003) A Genetic Neural Network Model of Flowering Time Control in. Agron J 95:71–81. doi: 10.2134/agronj2003.7100

Welham SJ, Gezan SA, Clark SJ, Mead A (2014) Statistical Methods in Biology: Design and Analysis of Experiments and Regression. Taylor & Francis

Welham SJ, Gogel BJ, Smith AB, et al (2010) A Comparison of Analysis Methods for Late-Stage Variety Evaluation Trials. Aust N Z J Stat 52:125–149. doi: 10.1111/j.1467-842X.2010.00570.x

Welham SJ, Thompson R (1997) Likelihood Ratio Tests for Fixed Model Terms Using Residual Maximum Likelihood. J R Stat Soc Ser B Stat Methodol 59:701–714. doi: 10.2307/2346019

West BT, Welch KB, Galecki AT (2014) Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition. Taylor & Francis

White JW, Andrade-Sanchez P, Gore MA, et al (2012) Field-based phenomics for plant genetics research. F Crop Res 133:101–112. doi: 10.1016/j.fcr.2012.04.003

Wientjes YCJ, Bijma P, Veerkamp RF, Calus MPL (2016) An Equation to Predict the Accuracy of Genomic Values by Combining Data from Multiple Traits, Populations, or Environments. Genetics 202:799–823.

Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinforma 28:2086–2087. doi: 10.1093/bioinformatics/bts335

Windhausen VS, Atlin GN, Hickey JM, et al (2012a) Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. G3 Genes Genom Genet 2:1427–1436. doi: 10.1534/g3.112.003699

Windhausen VS, Wagener S, Magorokosho C, et al (2012b) Strategies to Subdivide a Target Population of Environments: Results from the CIMMYT-Led Maize Hybrid Testing Programs in Africa. Crop Sci 52:2143–2152. doi: 10.2135/cropsci2012.02.0125

Winfield MO, Wilkinson PA, Allen AM, et al (2012) Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol J 10:733–742. doi: 10.1111/j.1467-7652.2012.00713.x

Woltereck R (1909) Weitere experimentelle Untersuchungen über Artveränderung, speziel über das Wesen quantitativer Artunterschiede bei Daphnien. Verhandlungen der Dtsch Zool Gesellschaft 110–173.

Wricke G (1966) Eine biometrische Methode zur Erfassung der ökologischen Anpassung. Qual Plant Mater Veg 13:318–319. doi: 10.1007/BF01103423

Wricke G (1962) Uber eine methode zur erfassung der okologischen streubreite in feldversuchen. Zeitschrift fur Pflanzenzuchtung 47:92–96.

Wright S (1984) Evolution and the Genetics of Populations, Volume 3: Experimental Results and Evolutionary Deductions. University of Chicago Press

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. Sixth Int. Congr. Genet. 1:356–366.

Wright S (1931) Evolution in Mendelian Populations. Genetics 16:97–159.

Wu MC, Lee S, Cai T, et al (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89:82–93. doi: 10.1016/j.ajhg.2011.05.029

Wu R, Lin M (2006) Functional mapping-how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7:229–237.

Wu R, Ma C-X, Littell RC, et al (2002) A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories. Genet Res 79:235–45. doi: 10.1017/s0016672302005633

Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. Crop J 5:175–184. doi: 10.1016/j.cj.2016.06.003

Xue GP, Loveridge CW (2004) HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. Plant J 37:326–339. doi: 10.1046/j.1365-313X.2003.01963.x

Yan W, Hunt LA (2001) Interpretation of Genotype × Environment Interaction for Winter Wheat Yield in Ontario. Crop Sci 41:19–25. doi: 10.2135/cropsci2001.41119x

Yan W, Hunt LA, Sheng Q, Szlavnics Z (2000) Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. Crop Sci 40:597–605. doi: 10.2135/cropsci2000.403597x

Yan W, Kang MS (2002) GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists. CRC Press

Yan W, Rajcan I (2002) Biplot Analysis of Test Sites and Trait Relations of Soybean in Ontario. Crop Sci 42:11–20. doi: 10.2135/cropsci2002.1100

Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23:1527–1536. doi: 10.1093/bioinformatics/btm143

Yang R-C (2007) Mixed-Model Analysis of Crossover Genotype–Environment Interactions All rights reserved. Crop Sci 47:1051–1062. doi: 10.2135/cropsci2006.09.0611

Yang R-C, Crossa J, Cornelius PL, Burgueño J (2009) Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution . Crop Sci 49:1564–1576. doi: 10.2135/cropsci2008.11.0665

Yang R, Tian Q, Xu S (2006) Mapping quantitative trait loci for longitudinal traits in line crosses. Genetics 173:2339–2356. doi: 10.1534/genetics.105.054775

Yang R, Xu S (2007) Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. Genetics 176:1169–1185. doi: 10.1534/genetics.106.064279

Yang W, Guo Z, Huang C, et al (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat Commun 5:5087.

Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci 28:556–580.

Yin X, Chasalow SD, Dourleijn CJ, et al (2000a) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity (Edinb) 85:539–549. doi: 10.1046/j.1365-2540.2000.00790.x

Yin X, Schapendonk AHCM, Kropff MJ, et al (2000b) A Generic Equation for Nitrogen-limited Leaf Area Index and its Application in Crop Growth Models for Predicting Leaf Senescence. Ann Bot 85:579–585. doi: 10.1006/anbo.1999.1104

Yin X, Stam P, Kropff MJ, Schapendonk AHCM (2003) Crop Modeling, QTL Mapping, and Their Complementary Role in Plant Breeding . Agron J 95:90–98. doi: 10.2134/agronj2003.9000

Yin X, Struik PC (2010) Modelling the crop: from system dynamics to systems biology. J Exp Bot 61:2171–2183.

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9:426–432. doi: http://dx.doi.org/10.1016/j.tplants.2004.07.007

Yin X, Struik PC, van Eeuwijk FA, et al (2005) QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. J Exp Bot 56:967–976. doi: 10.1093/jxb/eri090

Yuo T, Yamashita Y, Kanamori H, et al (2012) A SHORT INTERNODES (SHI) family transcription factor gene regulates awn elongation and pistil morphology in barley. J Exp Bot 63:5223–5232.

Zhang X, Perez-Rodriguez P, Semagn K, et al (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity (Edinb) 114:291–299.

Zhang Y, Pan W (2015) Principal Component Regression and Linear Mixed Model in Association Analysis of Structured Samples: Competitors or Complements? Genet Epidemiol 39:149–155. doi: 10.1002/gepi.21879

Zhang Z, Ersoz E, Lai C, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Publ Gr 42:355–360. doi: 10.1038/ng.546

Zheng B, Biddulph B, Li D, et al (2013) Quantification of the effects of VRN1 and Ppd-D1 to predict spring wheat (Triticum aestivum) heading time across diverse environments. J Exp Bot 64:3747–3761. doi: 10.1093/jxb/ert209

Zheng B, Chenu K, Fernanda Dreccer M, Chapman SC (2012) Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivum) varieties? Glob Chang Biol 18:2899–2914. doi: 10.1111/j.1365-2486.2012.02724.x

Summary

Summary

The main goal of plant breeders is to create and select genotypes that are well-adapted to the future growing conditions as defined by the meteorological, soil and management factors in the growing area of interest. Genotypes commonly show different sensitivities to environmental changes and genotype by environment interaction (GxE) is observed. GxE can lead to changes in genotypic ranking, complicating the breeding process. Usually, GxE patterns are analysed for a set of environments that correspond to the future growing conditions of the genotypes created by the breeding programme (the target population of environments, TPE). The selection strategy chosen by breeders, together with the crossing scheme, shape the target population of genotypes (TPG), which corresponds to all possible genotypes that the breeding programme for the TPE hopes to develop during the coming years. If the phenotypic responses at particular locations have a certain degree of repeatability across years, these locations may be classified into 'mega-environments', which correspond to sets of environments of similar quality that show a reduced number of cross-over interactions. Within a mega-environment, similar genotypes can be recommended, simplifying the selection and the recommendation processes for better adapted varieties.

Selection processes in plant breeding depend critically on the quality of phenotype predictions across a sample of target environments. The phenotype is classically predicted as a function of genotypic and environmental information. Models for phenotype prediction contain a mixture of statistical, genetic and physiological elements. In this thesis, we discuss prediction from linear mixed models (LMMs), with an emphasis on statistics, prediction from crop growth models (CGMs) and the combination of LMMs and CGMs. For LMMs, the genotypic input information includes molecular marker variation, while the environmental input can consist of meteorological, soil and management variables. Common LMMs for phenotype prediction consider QTL and genotypic prediction models. For a second type of phenotype prediction, we consider crop growth models. CGMs predict a target phenotype as a non-linear function of underlying intermediate phenotypes. The intermediate phenotypes are outcomes of functions defined on genotype dependent CGM parameters and classical environmental descriptors. Besides the objective of phenotype prediction, the combination of LMMs and CGMs can be useful to characterize the TPE and to evaluate breeding strategies. Both LMMs and CGMs require extensive characterization of genotypes and environments. High-throughput

technologies for genotyping and phenotyping provide new opportunities for upscaling phenotype prediction and increasing the response to selection in the breeding process.

The first specific objective of this thesis was to set the scene for prediction scenarios that are interesting for breeders and to discuss how either statistical models, CGMs or the combination of both types of models can be used to predict phenotypes across environments. In **Chapter 2**, we discuss the following scenarios; prediction of unobserved genotypes in observed environments, the prediction of observed genotypes in unobserved environments and finally, the prediction of unobserved genotypes in unobserved environments. Each of these prediction scenarios are discussed from the perspective of the LMMs, CGMs and combination of LMMs and CGMs. For each model category and scenario, we show a number of model adaptations, discussing and illustrating models with examples from the literature.

In **Chapter 3**, we propose a strategy to construct the training set of genotypes, improving the prediction accuracy of unobserved genotypes in observed environments. The main underlying hypothesis for our proposed training set construction method was that a homogeneous representation of the genetic diversity (genetic space) in the TPG leads to larger prediction accuracy. We illustrate our ideas with data for wheat, maize and rice. Training sets were constructed using uniform sampling, stratified-uniform sampling, stratified sampling and random sampling. We compared these methods with a method that maximizes the generalized coefficient of determination (CD). Several training set sizes were considered. We investigated four genomic prediction models: multi-locus QTL models, GBLUP models, combinations of QTL and GBLUPs, and Reproducing Kernel Hilbert Space (RKHS) models. The results of our proposed method, uniform sampling, were similar to those of the CD method, and showed a larger prediction accuracy than that for random or stratified sampling.

The first step for multi-environment predictions is to characterize the structure of GxE, quantifying the importance of repeatable versus non-repeatable patterns. In **Chapter 4**, we assessed strategies based on the additive main effects and multiplicative interactions model (AMMI) as applied to repeatable genotype by location interaction and on mixed models to identify regions that are internally more homogeneous. We compared two strategies for grouping trial locations into regions, one based on a full mixed model analysis, and one based on a relatively simple, yet robust two-step approach based on fitting AMMI models to within year genotype by location tables of means. The AMMI predictions were then used to cluster locations

within years. Consistent clustering of locations over years was used to assign locations to regions. The mixed model approach uses the parameters of a factor analytic model to classify locations in regions. We presented examples for the identification of regions in historical multi-environment trials for wheat in Denmark, Germany, The Netherlands and the United Kingdom. We identified regions in Denmark, Germany and the United Kingdom that coincided with latitudinal and longitudinal gradients.

Chapter 5 illustrates the concepts discussed in Chapters 2, 3 and 4, applying them to a multi-environment genomic prediction context using the 'Steptoe' x 'Morex' barley population. We compared a range of multi-environment genomic models differing in the way environment specific genetic effects were modeled. Information among environments was modelled by structuring the variance-covariance matrix between environments, either implicitly by factor analytic models or explicitly by similarities calculated from environmental covariables. We discuss the models not only in the light of their accuracy, but also in their ability to predict the different parts of the incomplete G×E table. We show the advantage of multi-environment genomic models that account for genotype by environment interactions. In addition, for our example data set, we show that for prediction in the most challenging scenario of untested environments, the use of explicit environmental information is preferable over the simpler approach of predicting from a main effects model.

In **Chapter 6**, we apply multi-environment mixed models to characterize the EU-Whealbi-barley germplasm collection as a source of valuable alleles for adaptation to European environments. EU-Whealbi was genotypically characterized with exome sequence data and phenotypically characterized for awn length, grain weight, heading date and plant height in six very diverse environments in Hungary, Italy, Scotland and Turkey. We compared the effects of haplotypes with those of single SNPs, adding an extra layer of biological complexity to the statistical analysis. Our results show that the EU-Whealbi barley collection has a large diversity of promising alleles regulating the four traits we analysed. We also show that haplotypes of genes underlying QTLs explain a larger proportion of the phenotypic variance, compared to the single SNPs. Other methodological issues related to the identification of the genetic basis for adaptation across multiple environments are also discussed. For example, the characterization of the heterogeneity of the LD across the genome, how to deal with collinearities between SNPs in the context of multi-locus models and the estimation of additive effects with multi-environment models. The allelic diversity is also discussed in relation to the geographical origin of barley genotypes.

Chapters 3 to 6 discuss methods for single-trait phenotype prediction. In **Chapter 7**, we move to a multi-trait prediction context. In this chapter, we propose the combination of statistical-genetic models and the APSIM crop growth model as a strategy to assess the traits and phenotyping strategies to improve the prediction accuracy of the target trait. We assess the potential of the combined modelling approach to characterize a sample of the TPE and TPG. We also illustrate how trait correlations are modified by environmental conditions and by the genetic architecture of the sample of the TPG. We contextualize the results for multi-trait genomic prediction in relation to the environmental quality. As an illustration, we show how biomass can be useful to improve yield predictions in some environments, but not in others. We compare different strategies to model biomass measured at multiple time-points (logistic curves and splines), from the perspective of their heritability and impact on prediction accuracy. We also evaluate models for yield prediction from biomass measured early in the growing season.

Many of the concepts and statistical models discussed in Chapters 2 to 7 are presented in **Chapter 8** from an educational/didactical perspective. In **Chapter 8**, we propose a schedule of topics that should be covered in a GxE course for plant breeders. In addition, we provide an overview of the trends in the usage of different GxE models in the literature over time.

Finally, in **Chapter 9**, we discuss the convenience and limitations of modelling approaches based on statistical models, crop growth models or a combination of statistical and crop growth modelling for different breeding situations. We further discuss how high throughput genotyping and high throughput phenotyping can be used to increase the chances of selecting better-adapted varieties. We present a number of technical considerations regarding the combination of statistical-genetic and crop growth models to assess breeding strategies. Other topics covered in the general discussion are how does the structure of TPE and TPG affect the choice of phenotyping strategies and prediction models. We conclude that the physiological hierarchy of traits should be the guiding principle when choosing the phenotyping strategy and modelling approach. Subsequently, we illustrate the concept of separability of genetic and environmental effects from a statistical and physiological perspective. Finally, we discuss challenges and opportunities presented by the characterization of the TPE and TPG when using simulations based on statistical and crop growth models.

Curriculum Vitae

About the Author

List of Publications

About the author

Daniela Bustos-Korts was born on the 4th of July 1986 in Temuco, Chile. In 2004 she graduated from the German School in Temuco. In 2008 she obtained the degree of Bachelor in Agricultural Sciences from the Universidad Austral de Chile, Valdivia. After doing a specialization in Plant Production in the same university, in 2009 she obtained the title of Agronomy Engineer, with the highest distinction. Her thesis was about developing a calibration curve for Near Infrared Spectroscopy (NIRS) to quantify the anthocyanin concentration in grains of purple wheat. She was distinguished with the price to the best graduate from her promotion of Agricultural Sciences. She also obtained the Bayer Crop Science award to the best graduate of the specialization in Plant Production. She continued her postgraduate studies with an MSc in Plant Sciences, specialization Crop Physiology at the Universidad Austral de Chile. Her MSc thesis involved the investigation of physiological and management drivers of the anthocyanin concentration in grains of purple wheat. She also investigated the physiological drivers of yield components in a doubled haploid population derived from parents with contrasting grain weight and grain number. She obtained her MSc in 2011, with the highest distinction. In the same year, she obtained a PhD grant from the Becas Chile programme (CONICYT, Chile). She started her PhD at Biometris, Wageningen University in October 2012. The results of her PhD are described in this thesis. In 2015, she obtained the Huub and Julliene Spiertz fund, allowing her to visit CSIRO in Brisbane, Australia, to learn about the APSIM crop growth model. From April 2016 she started working on the EU-Whealbi project at Biometris, Wageningen University. Her work at the EU-Whealbi project focuses on the statistical analysis of barley and wheat phenotypic and exome sequence data. She will continue with a post-doctoral position in the same project, until December 2018.

List of publications

Bustos, D. V., R. Riegel, and D.F. Calderini. 2012. Anthocyanin content of grains in purple wheat is affected by grain position, assimilate availability and agronomic management. J. Cereal Sci. 55(3): 257–264.

Bustos, D. V., A.K. Hasan, M.P. Reynolds, and D.F. Calderini. 2013. Combining high grain number and weight through a DH-population to improve grain yield potential of wheat in high-yielding environments. F. Crop. Res. 145(0): 106–115.

Bustos-Korts, D., M. Malosetti, S. Chapman, B. Biddulph, and F. van Eeuwijk. 2016. Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. G3 Genes|Genomes|Genetics 6(11): 3733–3747.

Bustos-Korts, D., M. Malosetti, S. Chapman, and F.A. van Eeuwijk. 2016. Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics. p. 55–82. *In* Yin, X., Struik, P.C. (eds.), Crop Systems Biology - Narrowing the Gaps between Crop Modelling and Genetics. Springer.

van Eeuwijk, F.A., **D. V Bustos-Korts**, and M. Malosetti. 2016. What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? Crop Sci. 56: 2119–2140.

Malosetti, M., **D. Bustos-Korts**, M.P. Boer, and F.A. van Eeuwijk. 2016. Predicting Responses in Multiple Environments: Issues in Relation to Genotype × Environment Interactions. Crop Sci. 56: 2210–2222.

(https://scholar.google.nl/citations?user=hGKCIegAAAAJ&hl=en)

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

- Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics (2016)

Writing of project proposal (4.5 ECTS)

- Statistical modelling of genetic and physiological processes underlying GxE in wheat (2012)

Post-graduate courses (8.7 ECTS)

- Introduction to R for statistical analysis; Biometris/PE&RC (2012)
- Advanced methods and algorithms in animal breeding with focus on genomic selection; WIAS 92012)
- Selection response in quantitative traits; Summer school Synbreed project TUM, Germany (2013)
- Analysis of linear mixed models by ASREML-R with applications in plant breeding; Biometris
- Introduction to Bayesian methods for quantitative geneticists; Synbreed project TUM, Germany (2015)
- Statistical uncertainty analysis of dynamic models; PE&RC/SENSE 92015)

Laboratory training and working visits (3 ECTS)

- Modelling with the APSIM crop growth model; CSIRO, Queensland, Australia (2015)

Invited review of (unpublished) journal manuscript (4 ECTS)

- Euphytica: GxE interactions organic production (2013)
- Annals of Applied Biology: combining abilities and heterosis in medicago (2016)
- Field Crops Research: APSIM simulations (2016)
- Theoretical and Applied Genetics: genomic prediction across multiple breeding cycles (2016)

Deficiency, refresh, brush-up courses (6 ECTS)

- Modern statistics for the life sciences; Biometris/ABG (2012)

Competence strengthening / skills courses (2.7 ECTS)

- Scientific writing; Wageningen in'to Languages (2014)
- Orientation on teaching for PhD candidates; Educational Staff Development, WUR (2016)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.2 ECTS)

- PE&RC Day (2013)
- PE&RC Weekend (2013)

Discussion groups / local seminars / other scientific meetings (4.5 ECTS)

- Modelling and statistic network (2014)
- Biometris statistical genetics meetings (2012-2015)
- Modelling of multi-parent populations (2015-2016)

International symposia, workshops and conferences (17.1 ECTS)

- Eucarpia Wheat section; poster presentation; Wernigerode, Germany (2014)
- Symposium all-inclusive breeding; poster presentation; Biometris, Plant Breeding, Genetics and Du-Pont Pioneer Wageningen, the Netherlands (2014)
- EUCARPIA Conference; oral presentation; Biometrics section, Wageningen, the Netherlands (2015)
- International wheat conference; poster presentation; Sydney, Australia (2015)
- 20th Eucarpia General congress; oral presentation; Zürich, Switzerland (2016)
- 4th International Plant Phenotyping symposium; oral presentation (2016)
- InterDrought V; poster presentation; Hyderabad, India (2017)
- Young investigator meeting, within the project flowering time control: from natural variation to crop improvement; oral presentation; Gatersleben, Germany (2017)

Lecturing / supervision of practical's / tutorials (6 ECTS)

- Basic statistics (2013)
- Modern statistics for the life sciences (2013-2015)
- Generation challenge programme (2014)

Supervision of a MSc student (3 ECTS)

- Monique Gosseling: Using meteorological information to characterize GxE in the Australian wheat belt

Acknowledgements

The completion of my PhD thesis was possible thanks to the contribution and support from a large number of people. I especially want to thank my promotor, Fred van Eeuwijk, and my co-promotor, Marcos Malosetti. Thanks to both of you, I dared to explore outside of my comfort zone, integrating statistics into my rather biological/agronomical background. One of the things I appreciate the most of working with you is that you made me feel part of the team, you allowed for an atmosphere that promoted creativity, you were enthusiastic about many of my suggestions and were also patient and kind when I felt the load was getting large. I will also treasure the good moments we spend together, with entertaining conversations during dinners, conferences or drinks. I'm very happy and grateful I could have both of you as my supervisors. Fred, thank you for encouraging me to believe in myself and to dare dreaming about the possibility of a scientific career. Thank you for allowing me to complement the statistical/modelling challenges with some organizational challenges, in a balanced way. That surely made my PhD more exciting and interesting. I also appreciate the multiple opportunities you gave me, for example, by trusting me to work on the Whealbi project of by inviting me to participate in the organization of the Eucarpia conference. Marcos, thank you for your important contribution to the more technical aspects of my thesis, especially when I was just starting. You are very good at finding the essence of the problem, keeping things simple and interesting at the same time. That, together with your positive attitude and your kindness, helped me to overcome many of the challenges that arose when implementing our ideas.

I would also like to thank other colleagues that were very important during my PhD. Special thanks to Martin Boer. We only collaborated in a specific part of my thesis (Chapter 7). However, you were always very generous with your knowledge and open to interesting discussions. You were particularly important during the last weeks, before handing in my thesis. You encouraged me to stay strong, to keep writing, even if the nice summer and the empty Radix were tempting me to better go on a holiday. Many thanks for being there to cheer me up.

Thanks to one of the key Biometris components; our dear secretaries Dinie and Hanneke. I appreciate your kindness and dedication helping us to sort out many little details, while making us feel at home at the same time. I also would like to thank you for encouraging me to practice my Dutch and being patient when I didn't get the grammar right. Dank jullie wel!

Thanks to all Biometris PhDs, office mates and other colleagues, especially to those with whom we shared for the longest time; Vincent, Julio, Renaud, Emilie,

Dominique, Nadia, George K., George A., Rianne J. Rianne vB, Jeroen, Stefan, Frederik. You were an important part of my social life, I enjoyed the conversations, Friday drinks, dinners and PhD parties. From each of you, I could learn that we all go through academic and personal challenges, that it will get difficult at times, but that those harder moments will pass more smoothly and that life will look sweeter when we share it. Thank you for sharing your time and company with me.

I would like to thank other Biometris-colleagues with whom I had nice conversations during these PhD years, making the time in Radix more enjoyable. Thanks to Joaõ, Sabine, Willem, Cajo, Paul K., Paul G., Bas, Evert-Jan, Maikel, Elly, Chaozhi, Marco, Lia and all other colleagues that were kind, positive and supportive in one or another way.

There were also some other PhDs and friends from outside Biometris that were very important to me, especially during the first year in the Netherlands. Thank you, Andrés T., Sanne M., Paula H., Vicky A., Pieter, Marcela V., Rafa Ch., Christos K. and Myluska C., for giving me a warm welcome, helping me to get around and for sharing your experiences.

I would like to thank Gilles Charmet and the other members of the EU-Whealbi project for trusting me to take care of many of the data analyses. Thereby, you gave me the opportunity to broaden my network, allowing me to have stimulating discussions. It was also a very enriching experience to interact with those of you that work in disciplines other than statistical genetics. The annual meetings were also a great opportunity to visit interesting countries like Israel and Italy and learn about other production systems and breeding situations.

Thanks Scott Chapman for giving me the opportunity of visiting his lab in Queensland and for his valuable feedback. I also thank Bertrand Schuiling for his contribution to the very first part of my PhD, sharing the data that allowed me to start thinking about questions related to GxE.

There is a number of friends and colleagues in Chile, who, in spite of the geographical distance, were close to me. Thank you my friends, Javier, Susana, Carolina, Beatriz, Uschi and Gerardo. I would also like to thank Daniel, Ricardo, Suzanne, Ignacio and Hugo Campos; you encouraged me to start this adventure in Wageningen and motivated me to keep on going. Thanks to all of you for staying in touch and for showing interest in my activities.

Further, I would like to thank the Chilean team at (or around) WUR. Especially Yélica, Daniel, Grace, Mauricio, Marcia, Pame, Loreto, Henk, Roberto, Fran, Julian,

Marcela, Andrés, Francisco and Lena. All of you, in different moments and ways, contributed finding in Wageningen a second home. I'm very grateful for that.

I would also like to thank the Ambassador of Chile to the Netherlands, Ms María Teresa Infante and her team, Juan-Enrique, Edgard and María Olivia, for your deep, genuine and warm interest in the contribution of science and culture to help Chilean society moving forward. Thank you for the interesting discussions and for constantly supporting the network of Chilean students and researchers in the Netherlands.

I want to finalize thanking those that have been always the closest to me; my family. I want to thank my dear parents, Marianne and Juan de Dios; since a little girl, you showed to me how fascinating nature is, motivating me to ask questions, to try out, to see what happens, to be a curious. I feel extremely privileged for the education I received. The fascination and respect for nature I learned from you are without any doubts the foundation of this PhD. Thank you for encouraging our extended family to get together every weekend at the farm. That surely contributed to give us a happy childhood. I also want to thank you, dear mamá and papá, for sending me your love during these years in the Netherlands. Thanks to my 'little' sisters Mariana, María Paz and Paulina. I appreciate all your messages, pictures, letters and calls, thanks for sharing your own adventures with me, making me a participant of your life, even if I was far away. Thank you, Mariana, for joining me in the Netherlands. I'm happy I can have your good company at our Wageningen home and I wish you all the best for your experiences in Europe, during your MSc. También quiero agradecer a mi familia por parte de Marcelo, Elisia, Carlos y Pamela, por su cariño y preocupación durante todos estos años. Gracias también a los demás miembros de mi familia, especialmente a mis queridos Oma, Opa, tías y primos. Muchas gracias, querida familia, por su cariño incondicional. Los quiero mucho y los extraño.

Last, to the one that is closest to my heart; thank you, Marcelo, for believing with me in this crazy adventure of coming to the Netherlands. We had to sort out all kind of challenges to be here, but that was also rewarded by many beautiful experiences during the process of 'discovering' Europe together. Thank you for being patient, positive and loving, for listening to me when I had to get rid of my worries by verbalising them. I feel extremely lucky and happy to have your company. Thank you for being here with me.

Colophon

Funding

Chapters 1, 2, 3 and 4 in this thesis were funded by the Becas Chile Programme (CONICYT, Chile).

The research leading to the results of chapters 6, 7 and 9 has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°FP7- 613556, Whealbi.

The work of MM and FvE for chapters 5 and 8 was conducted as part of the Integrated Breeding Platform phase II project (Bill & Melinda Gates Foundation).

Thesis cover

The thesis cover was designed by Daniel Romero and Daniela Bustos-Korts. The four figures correspond to different interpretations of the Moray Inca ruins in Peru. The site contains several terraced circular depressions, with an irrigation system. The terrace depth, design, and orientation with respect to wind and sun, create a temperature difference of as much as 15 °C between the top and the bottom. It is possible that this large temperature difference was used by the Inca to study the effects of different climatic conditions on crops, probably being one of the first managed environment trials made by humankind. The four-images collage symbolizes the main topics addressed in this thesis; genotype to phenotype modelling across multiple environments. The first image in the collage (starting from the left) shows the observed terraces with some noise, symbolising the imperfect phenotype predictions. The second image represents the temperature gradient along the terraces (the environment), the third image symbolizes the different levels of biological complexity between genotype and phenotype and the fourth image represents the phenotype, as directly observed in an experiment.

Printing

This thesis was printed by Digiforce, Vianen, the Netherlands