Institut für Nutzpflanzenwissenschaften und Ressourcenschutz

Genomic Prediction and Association Mapping Using Publicly Available Data of German Variety Trials in Spring Barley

Inaugural-Dissertation
zur
Erlangung des Grades

Doktor der Agrarwissenschaften
(Dr. agr.)

der
Landwirtschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Bongsong Kim

aus
Pohang, Südkorea

Referent:	**Professor Dr. Jens Léon**
Korreferent:	**Professor Dr. Heiko Schoof**
Tag der mündlichen Prüfung:	**03.07.2014**
Erscheinungsjahr:	**2014**

Abstract

In recent decades, the implementation of best linear unbiased prediction (BLUP) has been extended beyond its initial purpose for the breeding value (BV) estimation to conduct the association mapping and genomic selection. In this study, the prospect of using BLUP was investigated for the BV estimation, AM and GS in self-pollinating crop with a German barley cultivar collection that is publicly available. Chapter 1 introduces issues of this study and provides a review of the relevant literatures. Chapters 2 and 3 address the application of BLUP with an assembled data set of German spring barley cultivars in unbalanced trials. One issue regarding this work was the absence of a method for computing a numerator relationship matrix (NRM) for selfing crop species. Therefore, the method of constructing the NRM was developed in this study, which is introduced in Chapter 2. Chapter 3 reports the application of the underlying NRM to BLUP for grain yield, scald severity and net blotch severity. Heritabilites resulted in 0.719 for grain yield, 0.491 for scald severity and 0.581 for net blotch severity, which suggests that the given phenotypic data were measured in sufficient level. Spearman's rank correlation between BLUP estimates and mean phenotypes (MPs) were shown to be 0.854 for grain yield, 0.893 for scald severity and 0.940 for net blotch severity, which indicates that the selection depending on the BLUP may respond better than that depending on the phenotypic observation using MPs. Chapter 4 describes the measurement of the marker-trait association for the aforementioned traits in German spring-sown barley cultivars and 1181 diversity array technology (DArT) markers. Two models were fitted: (1) the BLUP that embeds a marker-based kinship matrix and a discriminant analysis of principle component matrix (KD model) and (2) the BLUP that embeds a marker-based kinship matrix and a subpopulation matrix resolved using STRUCTURE software (KS model). For the stringent evaluation of marker-trait association, the significance level of $p < 0.001$ in the Wald test and cross-validation were applied. In total, six marker-trait associations were detected (one for grain yield, four for scald severity and one for net blotch severity). Chapter 5 presents the genomic selection performed using ridge regression BLUP (RR-BLUP) with the same plant materials as used in Chapter 4. The increasing sizes of the training set and marker set were positively correlated with prediction accuracy. As a novel approach, marker sets that were selected based on the strength of marker-trait linkages were examined. To form the sets of markers, p-values obtained from the mapping study were referenced, and ten sets of markers were prepared by applying p-value thresholds of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0. The resulting prediction accuracies ranged from 0.3226 to 0.7323 for grain yield, from 0.3534 to 0.5396 for scald severity and from 0.4340 to 0.8326 for net blotch severity. A marker set formed with a decreasing p-value appeared to provide the higher prediction accuracy for all traits by overcoming the weakness of the small size of marker set, showing that the use of p-values is promising in RR-BLUP.

Zusammenfassung

In den letzten Jahrzehnten wurde das best linear unbiased prediction (BLUP) Verfahren, von seinem ursprünglichen Gebiet zur Zuchtwertschätzung, zusätzlich zur Assoziationskartierung und zur Genomische Selektion angewendet. In dieser Studiewurde BLUP benutzt, um bei einer selbstbefruchtenden deutschen Gerstenpopulation, deren Daten öffentlich zugänglichsind, die Zuchtwertschätzung, Assoziationskartierung und Genomischen Selektion zu untersuchen. Nach einer Einführung in Kapitel 1 wird in den Kapiteln 2 und 3 die Anwendbarkeit von BLUP auf einen unbalancierten Datensatz einer deutschen Sommergerstenpopulation behandelt. Eine Herausforderung in dieser Arbeit war es, dass es bislang keine Methode gab, um eine numerator relationship matrix (NRM) aus den Stammbauminformationen von Selbstbefruchterpopulationen zu berechnen. In Kapitel 2 wird eine Methode zur Erstellung der NRM vorgestellt. In Kapitel 3 wird die Anwendung der zugrundeliegenden NRM mit dem BLUP Verfahren erörtert und zur Berechnung des Kornertrages, sowie der Anfälligkeit für *Rynchosporium* und *Drechslera teres* verwendet. Die berechneten Heritabilitäten waren 0,719 für den Ertrag, 0,491 für die Rynchosporiumanfälligkeit und 0,581 für die Netzfleckenanfälligkeit, was zeigt, dass der verwendete Datensatz für dieses Verfahren geeignet ist. Die Spearman's rank Korrelation zwischen der BLUP-Vorhersage und den phänotypischen Durchschnittswerten zeigte für den Kornertrag 0,854, für die Rynchosporiumanfälligkeit 0,893 und für die Netzfleckenanfälligkeit 0,940. Dieses Ergebnis deutet auf einen besseren Selektionserfolg basierend auf der BLUP-Vorhersage gegenüber der Berechnung mit phänotypischen Durchschnittswerten hin. In Kapitel 4 wird die Messung der Marker- Merkmalsausprägungen beschrieben, die zwischen den zuvor genannten deutschen Sommergerstensorten mit 118 diversity array technology Markern berechnet wurden. Es wurden zwei Modelle verglichen: (1) ein BLUP Verfahren in dem die molekulare Verwandtschaftsmatrix mit einer Diskriminanzanalyse der Hauptkomponentenmatrix verknüpft wird (KD Modell) und (2) ein BLUP Verfahren in dem die molekulare Verwandtschaftsmatrix mit einer Subpopulationsmatrix aus der Software STRUCTURE verknüpft wird (KS Modell). Für die Bewertung der Marker- Merkmalsassoziierung wurde ein Signifikanzniveau von $p < 0{,}001$ und eine Kreuzvalidierung angewandt. Insgesamt wurden sechs QTLs identifiziert (eins für den Ertrag, vier für die Rynchosporiumanfälligkeit und eine für die Netzfleckenanfälligkeit). Kapitel 5 beschreibt die Leistung von ridge regression BLUP (RR-BLUP) für das Verfahren der Genomischen Selektion. Dabei wurde mit demselben Datenmaterial gearbeitet, das bereits in Kapitel 4 verwendet wurde. Eine Vergrößerung der Probenanzahl und der Markermenge war positiv mit der Vorhersagegenauigkeit korreliert. In einem neuen Ansatz wurden die Marker, basierend auf ihrer Korrelation der Marker- Merkmalsausprägung zusammengestellt und untersucht. Um die Markersets zu unterscheiden, wurden die aus der Kartierungsuntersuchung erhaltenen Werte verwendet. Zehn Markersets mit p-Werten von 0,1 bis 1 wurden mit jeweils gleichen Abständen erstellt. Die resultierende Vorhersagegenauigkeit reichte von 0,3226 bis 0,7323 für den Ertrag, 0,3534 bis 0,5396 für die Rynchosporiumanfälligkeit und von 0,4340 bis 0,8326 für die Netzfleckenanfälligkeit. Ein Markerset, das aus den Markern mit den niedrigsten p-Werten erstellt wurde, zeigt eine höhere Vorhersagengenauigkeit für alle Merkmale, obwohl eine Schwächung der Aussagekraft durch eine geringere Markeranzahl besteht. Es wird gezeigt, dass das Einbeziehen der p-Werte in RR-BLUP zu vielversprechenden Ergebnissen führt.

List of contents

1. General Introduction

Best linear unbiased prediction (BLUP) was originally a statistical approach solely for the purpose of breeding value (BV) estimation using phenotypic and pedigree data sets. In recent decades, the application of BLUP has been further expanded to the association mapping (AM) and genomic selection (GS) that require genotypic and phenotypic data sets (Meuwissen et al., 2001; Endelman., 2011; Stich et al., 2008; Stich and Melchinger., 2009; Zhong et al., 2009; Wang et al., 2012a; Zhao et al., 2012; Crossa et al., 2014). The primary focus of this study was to investigate the prospects of BLUP across multiple purposes in self-pollinating crop species by using a data set of the German barley (*Hordeum vulgare* L.) cultivar collection in unbalanced trials. All plant accessions were publicly available from Landessortenversuche (LSV).

In breeding programs of self-pollinating crop, the selection is conventionally made based on mean phenotypes (MPs), general combining ability or mid-parent value (Bernardo., 1994; Panter and Allen., 1995a; Pattee et al., 2001; Oakey et al., 2006; Piepho et al., 2008; Zhong et al., 2009). Such selections based on the phenotypic observation ignore the genetic effect that is not expressed in the present generation but will be expressed in the next generation (Piepho et al., 2008). The BLUP can provide breeders with the predicted BVs that reflect the latent genetic performance of individuals by using a numerator relationship matrix (NRM). In the BLUP, an NRM exhibits a genetic variance-covariance and plays a vital role to capture the genetic potential of an individual from its relatives (Henderson., 1975). The NRM contains the identical by descent probabilities among any two individuals within a population, which is called a relationship coefficient. A method of constructing the NRM was devised by Emik and Terrill., (1949), which requires the pedigrees describing the parent-offspring relationship and assumes a hybrid mating. This method is customized for animal pedigree, which leads to limiting the use of BLUP in plant breeding programs (Bauer et al., 2006). For the BLUP being implemented in self-pollinating crops, the development of a method for constructing an NRM was required. Regarding this, Chapter 2 presents new formulas that define a relationship coefficient using plant pedigree and the number of selfing generation as well as PopKin software tool that constructs an NRM based on the underlying formulas. Chapter 3 shows fitting the BLUP models with the given German barley cultivar collection for the purposes of estimating the BVs and selection response for grain yield, scald severity and net blotch severity, for which an NRM constructed by the PopKin was used. In self-pollinating crops, since the MP of a variety implies the response of the nearly fixed genotypes to diverse environments, the MP reflects the genetic performance of a variety. (Piepho et al., 2008). Therefore, the prospect of the BLUP was investigated by comparing the BLUP estimates with the MPs. Chapters 4 and 5 address the association mapping (AM) and genomic selection (GS), which deals with a kinship matrix (KM) derived from a set of genotypes based on the BLUP. In the AM and GS, the KM corrects a bias that arises from the genetic relationship among individuals (Yu et al., 2005; Stich et al., 2008; Haseneyer et al., 2010; Hayes et al., 2009; Huang et al., 2010; Endelman., 2011; Shi et al., 2011; Wang et al., 2012a; Wang et al., 2012b). Chapter 4 presents the conduction of the AM using a small size of the German barley cultivar collection for grain yield, scald severity and net blotch severity. The AM is an approach to detect trait-associated markers (TAMs) through the comparison of phenotypic variances and genotypic segregations in a diverse panel of existing individuals (Haseneyer et al., 2010; Wang et al., 2012a). In detecting the TAMs, significant amount of linkage disequilibrium (LD) is beneficial because the mapping is based on the detection of markers of LD in association with a trait (Kraakman et al., 2004; Hanseneyer et al., 2010; Shi et al., 2011). Therefore, the AM is also termed LD mapping (Kraakman et al., 2004; Rafalski., 2010; Pasam et al., 2012). A population that contains a number of individuals typically has subpopulations due to selection activities or different originations (Kraakman et al., 2004; Haseneyer et al., 2010; Wang et al., 2012b). Because the subpopulations often have an impact on phenotypic performance through forming genetic stratifications, the precision of the AM can be improved by incorporating a subpopulation structure matrix to BLUP (Kraakman et al., 2004; Pswarayi et al., 2008; Stich et al., 2008; Shi et al., 2011; Wang et al., 2012b). In this study, two different BLUP that contain different subpopulation matrices were modeled. The two subpopulation matrices were obtained by discriminant analysis of principle components (DAPC) and STRUCTURE analysis. The BLUP coupled with DAPC was termed the KD model, and the BLUP coupled with the STRUCTURE analysis was termed the KS model. The objective of Chapter 4 is to investigate the effects of two different subpopulation structures on the resolutions of the AM using the BLUP. Chapter 5 presents the genomic selection (GS) using ridge regression BLUP (RR-BLUP) with the same barley cultivar collection as used in Chapter 4. The GS is a method to predict the unknown phenotypic performance of genotyped individual using the marker estimates of a number of markers.

Typically, the GS requires a training set and a validation set. The training set is used for estimating the marker effects upon a trait, whereas the validation set comprises the genotyped individuals without the phenotypic records. The phenotypic performance can be predicted by applying the estimated marker effects to the genotype of individual in the validation set (Meuwissen et al., 2001; Hayes et al., 2009). Previous studies (Muir., 2007; Sorberg et al., 2008; Zhong et al., 2009; Asoro et al., 2011; Crossa et al., 2014) reported that the precision of genomic estimated breeding values (GEBVs) is positively proportional to the sizes of training set and marker set. The main idea of the GS is the estimation of BVs by additively summing up the effects of markers of LD with QTL (Hayes et al., 2009). Asoro et al., (2011) has explored the estimation of GEBVs under the training sets that contains varying number of QTL markers, from which it was found that a simple increment of the number of QTL markers was not effective in improving the precision of GEBVs. This study presents a new approach that pools the markers by referencing the LD effect of each marker. To reference the LD effect of each marker, p-values obtained from the AM (Chapter 4) were used because the p-valueindicates the association between gene and marker. The objective of Chapter 5 is to determine the optimal conditions to carry out the RR-BLUP (1) under varying sizes of training set, (2) varying sizes of marker set and (3) varying threshold levels of p-value for marker selection.

1.1 Literature reviews

1.1.1 Breeding value estimation using best linear unbiased prediction in self-pollinating crop

Advantages of BLUP

Henderson., (1975) first proposed the best linear unbiased prediction (BLUP), and it was used to increase the response to selection and reduce costs (Bernardo., 1994; Panter and Allen., 1995a; Panter and Allen., 1995b; Durel et al., 1998; Pattee et al., 2001; Purba et al., 2001; Oakey et al., 2006; Piepho et al., 2008). In breeding programs, selection depends on phenotypic observations that ignore the unobservable genetic potential in an individual that could be expressed in its offspring. As a strategy to overcome this problem, the BLUP improves the response to selection in a manner to capture a latent genetic potential of an individual by considering the relatives' phenotype performances. The BLUP can reduce the cost of plant breeding programs by eliminating expanses that are typically associated with evaluations of phenotypic performance, such as investing in large tracts of land and long periods of time to validate the genetic values, which are necessarily expensive in terms of cost and time. The utilization of BLUP is often the smarter approach because it provides predictions without requiring field tests (Bernardo., 1994; Purba et al., 2001).

BLUP based on the mixed linear model

Linear models can be distinguished between fixed model and mixed model. The fixed model only comprises fixed effect variable vectors, whereas the mixed model includes both random effect and fixed effect variable vectors. In this study, the random effect assumes a correlation among variables within a vector, whereas the fixed effect assumes that variables within a vector are independent and unrelated (Crossa et al., 2006). As a statistical approach for estimating breeding values (BVs), the BLUP is based on the mixed model because BLUP assumes that the unknown variables in the BV's vector are genetically related to one another.

Numerator relationship matrix

The numerator relationship matrix (NRM) contains pair-wise additive relationship coefficients among two individuals in a population (Emik and Terrill., 1949). The additive relationship coefficient represents the degree of genetic relationship among two individuals, which is equivalent to twice the inbreeding coefficient from the two individuals' offspring. Therefore, the relationship coefficient ranges between 0 and 2. The method for computing an NRM was devised by Emik and Terrill., (1949). The underlying method requires a tabular pedigree that states the parent-offspring relationship and is sorted from parent to offspring. For example, with a pedigree in Figure 1-1, the pedigree can be tabulated as shown in Table 1-1.

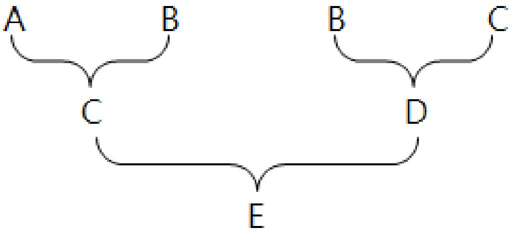


Figure 1-1. A pedigree skeleton of individuals, A, B, C, D and E.

Table 1-1. Pedigree table recorded with pedigree skeleton in Figure 1-1. Pedigree are sorted so that parents precede offspring in an individual column.

Individual	**Father**	**Mother**
A	Unknown	Unknown
B	Unknown	unknown
C	A	B
D	B	C
E	C	D

The equations for constructing an NRM (Emik and Terrill., 1949) can be denoted:

(1) If two different individuals, i and j, are unknown, then

$$f_{i,j} = f_{j,i} = 0$$

(2) If two identical individuals, i and j, are unknown, then

$$f_{i,i} = f_{i,i} = 1$$

(3) If two different individuals, i and j, are known, then

$$f_{i,j} = \frac{1}{2}(f_{p,j} + f_{q,j})$$

(4) If two identical individuals, i and j, are known, then

$$f_{i,i} = 1 + \frac{1}{2}f_{p,q}$$

where $f_{i,j} = f_{j,i}$ = the relationship coefficient between i and j; $f_{i,i}$ = the relationship coefficient between two is; $f_{p,j}$ = the relationship coefficient between i's parent, p, and j; $f_{q,j}$ = the relationship coefficient between i's parent, q, and j; $f_{p,q}$ = the relationship coefficient between i's two parents, p and q. Note that j is not a descendent of i, and vice versa.

When using Table 1-1, the resulting NRM is shown in Table 1-2.

Table 1-2. Numerator relationship matrix constructed via a pedigree in Table 1-1.

	A	B	C	D	E
A	1	0	0.5	0.25	0.375
B	0	1	0.5	0.75	0.625
C	0.5	0.5	1	0.75	0.875
D	0.25	0.75	0.75	1.25	1
E	0.375	0.625	0.875	1	1.375

Initially, an NRM was used for mating designs in breeding programs. Henderson., (1975) then proposed BLUP in which the NRM was embedded for the purpose of minimizing the bias of the estimated breeding values (EBVs). Previous studies (Durel et al., 1998; Bromley et al., 2000; Nunes et al., 2008; Atkin et al., 2009) demonstrated that the utilization of an NRM is effective in increasing the response to selection of breeding. Atkin et al., (2009) further reported that a precision degree of an NRM affects the accuracy of EBVs and variance components.

Heritability

Heritability accounts for the proportion of genetic traits in total phenotype and is defined in two distinct ways: (1) broad-sense heritability and (2) narrow-sense heritabilty (Bernardo., 2002; Oakey et al., 2006; Piepho and Moehring., 2007). Broad-sense heritability represents total genetic variance ($\mathrm{V_g}$) comprising both additive and non-additive genetic variances over the total phenotypic variance ($\mathrm{V_p}$), whereas narrow-sense heritability represents the additive genetic variance ($\mathrm{V_a}$) over the total phenotypic variance ($\mathrm{V_p}$) (Bernardo., 2002; Piepho and Moehring., 2007). The notations for broad-sense heritability (H) and narrow-sense heritability (h^2) are: (1) H = $\frac{\mathrm{V_g}}{\mathrm{V_p}}$, and (2) h^2 = $\frac{\mathrm{V_a}}{\mathrm{V_p}}$, respectively (Bernardo., 2002; piepho and Moehring., 2007). In plant breeding programs, narrow-sense heritability is considered a parameter indicating the response to selection (Durel et al., 1998; Bernardo., 2002; Piepho and Moehring., 2007). The routine equation (Hallauer and Miranda., 1981; Melchinger et al., 1998; Piepho and Moehring., 2007) for computing the narrow-sense heritability can be denoted as follows:

$$\mathrm{h}^2 = \frac{\sigma_\mathrm{g}^2}{\frac{\sigma^2}{\mathrm{rm}} + \frac{\sigma_\mathrm{gv}^2}{\mathrm{m}} + \sigma_\mathrm{g}^2} \qquad \text{(Equation 1-1)}$$

where r = the number of replications; m = the number of environments; σ_g^2 = the genotype variance; σ_gv^2 = the variance of genotype-by-environment interaction; σ^2 = the error variance.

Equation 1-1 can be applied to a balanced data set that has regular numbers of r and m. However, the field trials of plant are often unbalanced because the block designs and the mating schemes are variably adjusted (Piepho and Moehring., 2007). This situation makes it difficult to use Equation 1-1. In BLUP, h^2 is used to calculate the genetic variance in a population, which will be addressed in the next paragraph.

Characteristic of BLUP

A basic mixed model of BLUP can be denoted as follows:

$$y = Xb + Zu + e \qquad \text{(Equation 1-2)}$$

where y is a vector that contains phenotypic observations; X and Z are the incidence matrices; b and u are the vectors that contain the unknown fixed effects and the unknown random effects, respectively; e is a residual vector.

In Equation 1-1, the vector, b, is assumed to be fixed effect, whereas the vectors, u and e, are assumed to be random effect. Elements in the random effect vector are assumed to correlate one another, whereas elements in the fixed effect vector are assumed to be independent and unrelated (Henderson., 1975; Robinson., 1991; Pattee et al., 2001). The random effect vectors, u and e, can be denoted as follows:

$$\mathrm{Var}\begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} A\sigma_g^2 & 0 \\ 0 & I\sigma_e^2 \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \qquad \text{(Equation 1-3)}$$

where u = the vector of genotype effects; e = the vector of random residual effects; A = the numerator relationship matrix; σ_g^2 = the genotype variance; the I = the identity matrix; σ_e^2 = the residual variance; G = Var(u) = $A\sigma_g^2$; R = Var(e) = $I\sigma_e^2$.

In Equation 1-2, BLUP aims to attain the resolution for a random genetic vector, u, and the basic linear model of BLUP can be denoted as follows:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix}\begin{pmatrix} b \\ u \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix} \qquad \text{(Equation 1-4)}$$

where y = the vector that contains phenotypic observations in Equation 1-2; b and u = the vectors that contain the unknown fixed effects and random effects, respectively, in Equation 1-2; X and Z = the incidence matrices in Equation 1-2; G = Var(u) = $A\sigma_g^2$ in Equation 1-3; R = Var(e) = $I\sigma_e^2$ in Equation 1-3.

In Equation 1-4, G^{-1} represents an inverse of the genetic covariance matrix that is defined in Equation 1-3 and relates individuals to one another with minimizing the prediction bias (Tavernier., 1988; Panter and Allen., 1995a; Panter and Allen., 1995b). If the G^{-1} is a null matrix, the Equation 1-4 becomes perfectly compatible with ordinary least squares equation (Robinson., 1991; Piepho et al., 2008). To fit the BLUP model, the G matrix has to be known in advance. However, the G matrix cannot be truly estimated because the BVs within the u vector will be unknown before fitting the BLUP model. Therefore, the G matrix must be presumed based on the reasonable theory. Henderson (1975) defined G as $A\sigma_g^2$, where A is an NRM (Emik and Terrill., 1949); σ_g^2, is a genotype variance that is defined as $\sigma_g^2 = \frac{1-h^2}{h^2}$, and h^2 is a narrow-sense heritability (Panter and Allen., 1995a; Panter and Allen., 1995b; Nielsen et al., 2009). It is noted that BLUP enhances the response to selection by minimizing bias (Robinson., 1991; Verrier et al., 1993; Panter and Allen., 1995a; Panter and Allen., 1995b; Piepho et al., 2008). Regressing the mixed linear model that includes G^{-1} makes a robust shrinkage and allows the imbalance of the data set structure to be unbiased by offsetting the errors (Panter and Allen., 1995a; Piepho and Moehring., 2005; Oakey et al., 2006; Bauer et al., 2008).

Review of BLUP application with crop

With segregating lines of soybean cultivars, Panter and Allen (1995a) conducted a comparison of BLUP estimates with mid-parent values (MPVs) to select the best parents. As a result, it was shown that the rank correlations between the predicted and the realized values were consistently higher in BLUP than in MPV and the standard errors (SEs) were lower in BLUP than in MPV. Purba et al., (2001) applied BLUP to an oil palm population and compared the resulting ranking of BLUP estimates with the realized phenotypic observations across several traits. The resulting correlation coefficients ranged from 0.55 to 0.64 for oil yield and from 0.48 to 0.64 for industrial oil-extraction rate (shown here only for familiar traits. For details, refer to Purba et al., 2001). For the recurrent full-sib selection of maize (*Zea mayus* L.), Fachenecker et al., (2006) carried out selections for grain yield and grain moisture depending on the results from BLUP and MPV. The prediction accuracy of the BLUP estimates was shown to be lower than that of MPV because the small number of related families lessened the advantageous impact of the numerator relationship matrix.

Multiple-trait BLUP outperforms single-trait BLUP

The BLUP model that considers multiple traits is termed a multivariate BLUP. Previous studies demonstrated that multiple-trait BLUP models are more accurate over single-trait BLUP models. This is particularly true when multiple-traits are negatively correlated (Henderson and Quaas., 1976; Bauer and Léon., 2008; Viana et al., 2010) because the residual variance decreases as the negatively related traits are considered (Bauer and Leon., 2008).

Probable long-term side effect resulting from the success of BLUP

In breeding for improving agriculturally useful traits, the BLUP is more effective than traditional selection based on phenotype observation. However, the increasing response to selection when using BLUP does not mean that the BLUP always brings positive effects. The BLUP can depreciate genetic diversity in a population over generations by expanding the symmetric allelic region that contains the selectively favor genes (Bulmer., 1976; Verrier et al., 1993).

1.1.2 Subpopulation analysis

Diversity Array Technology markers

Diversity Array Technology (DArT ^{TM}) marker is a licensed genotyping system (Tinker et al., 2009) developed for barley and wheat. The DArT probes genotypes by hybridizing DNA samples with cloned DNA fragments arrayed on a solid phase slide. As a dominant marker system, the DArT provides the following scores: 1 for a presence and 0 for an absence. The DArT provides a low-cost, high-throughput and comprehensive genome coverage with an even distribution (Jaccoud et al., 2001; Tinker et al., 2009).

Population structure analyses in annual crops

Population analysis is classically conducted in biological studies such as evolution and genetic diversity (Jombart et al., 2010). Recently, the result of population analysis is used to correct the confounded marker-trait association in fitting the model of detecting QTL (Yu et al., 2005; Zhao et al., 2007; Stich et al., 2008; Massman et al., 2011; Pasam et al., 2012). For population analysis, a variety of methods were suggested to date, the most popular of which is the Bayesian clustering algorithm, which can be conducted using STRUCTURE software (Prichard et al., 2000). The STRUCTURE produces the inferred ancestry representing the likelihood of an individual belonging to the defined subpopulations and visualizes the numeric results into graphical bar chart. However, it has several shortcomings: an assumption for the number of populations is required, which is very hard to know in advance (Jombart et al., 2010; Haseneyer et al., 2010); a computation is highly demanding and time-consuming (Jombart et al., 2010; Wang et al., 2012a); individuals within a population are assumed to follow the Hardy-Weinberg equilibrium (Prichard et al., 2000), which is not met in selfing crops (Gao et al., 2007; Haseneyer et al., 2010; Wang et al., 2012b). Therefore, the STRUCTURE analysis must be carefully implemented with crop populations (Gao et al., 2007; Stich et al., 2008). As an alternative to the Bayesian clustering, multivariate methods such as principal component analysis (PCA) and discriminant analysis of principal components (DAPC) are available. The PCA was classically used for decades. In principle, the PCA decomposes the correlation among variables, consequently producing the multi-dimensional variable vectors. Of the multiple vectors, the two largest vectors are taken for the purpose of displaying the two dimensional scatter plot. As a variant of the PCA, the DAPC method maximizes the distance between groups (Jombart et al., 2010), which is advantageous in assigning the population members into a particular genetic group. According to Jombart et al., (2010), the DAPC provides a similar result to that obtained by the STRUCTURE analysis. Contrary to the Bayesian clustering, the multivariate analyses provide the following advantages: the Hardy-Weinberg equilibrium is not assumed; the computation demand is negligible; the dispersion of collection members can be visualized (Jombart et al., 2010; Wang et al., 2012a; Wang et al., 2012b). However, using the two largest variance vectors to explain the total genetic variance of population can be risky because it could not suffice to be the representative of population structure (Wang et al., 2012b). In general, the different approaches of population analysis tend to resolve similar results (Massman et al., 2011).

Genetic diversity in Barley population

Barley's selfing nature combined with its morphologic characters contributed to its diverse subpopulation (Malysheva-Otto et al., 2006; Cockram et al., 2008). In addition, its geographical origin, agronomical traits and kinship among germplasms allow its genetic variation to flourish (Malysheva-Otto et al., 2006; Cockram et al., 2008). Many studies characterized diverse barley collections. Cockram et al., (2008) captured four major subgroups in a population analysis using 423 barley lines collected from EU countries: spring two-row, winter two-row, spring six-row and winter six-row. Comadran et al., (2011) analyzed 192 spring and winter barley lines collected around the Mediterranean basin and revealed five groups (Turkey, Syria & Jordan landraces, North-Europe two-row springs, North-Europe six-row winters, South-West Mediterranean) and revealed that the resulting clusters were closely tied to their geographical origins. Wang et al., (2012) characterized the subpopulation of 615 UK barley cultivars and identified that the cultivars were clustered into the winter- and spring-sown groups.

1.1.3 Mapping of trait-associated markers using BLUP in self-pollinating crop

Linkage disequilibrium and linkage equilibrium

Linkage disequilibrium (LD) occurs when two or more alleles on a single chromosomal block inherit together. In general, alleles that are physically closer have a higher chance of LD. Linkage equilibrium (LE) occurs when the alleles segregate because of breakage of genome through recombination. The LD and LE among alleles are mainly determined because of their distance. However, closely located alleles can appear to be of LE. For example, in a study with 146 European barley collections, Kraakman et al., (2004) identified 19 pairs out of 53 with a distance < 1 cM that showed the LE pattern and one pair of markers in different chromosomes that appeared to show the LD pattern.

Linkage disequilibrium and gene mapping

LD currently provides a useful resource for mapping quantitative trait loci (QTL) because QTL can be positioned approximately depending on the location of markers that are highly correlated with the phenotypic variation under the assumption that a marker is of LD with QTL. Mating schemes determine the extent of LD. In general, random mating decreases the LD, whereas non-random mating increases the persistency of LD (Bernardo., 2002). Likewise, the reproduction system also determines the extent of LD, and the extent of LD is highly variable according to species. Previous studies measuring LD extents found < 10 cM (Kraakman et al., 2004) in barley, < 3 cM in sugar beet (Kraft et al., 2000), roughly 2000 bp in maize (Remington et al., 2001) and 20-30 cM in rice (Agrama et al., 2007). Selfing reproduction causes a significant extent of LD because selfing increases the region of symmetric alleles in a genome (Kraakman et al., 2004; Rafalski., 2010).

Current mapping techniques: bi-parental and association mappings

In QTL detection, two types of mapping methods can be currently selected: (1) bi-parental mapping and (2) association mapping (linkage disequilibrium mapping or LD mapping). As a traditional mapping method, bi-parental mapping dissects the marker-trait association within a population developed from bi-parental crosses (Kraakman et al., 2004; Cockram et al., 2008; Zhang et al., 2009; Yu and Buckler., 2006). Meanwhile, association mapping analyzes the marker-trait association by dissecting the historic patterns of recombination that have occurred within a provided germplasm collection (Kraakman et al., 2004; Cockram et al., 2008; Zhang et al., 2009; Yu and Buckler., 2006).

Association mapping provides several benefits relative to bi-parental mapping: (1) the data from existing germplasm can be used instead of creating new populations through bi-parent crosses (Kraakman et al., 2004; Roy et al., 2010; Massman et al., 2011; Wang et al., 2012b); (2) an unbalanced set of data can be used for mapping because the mixed linear model (MLM) minimizes errors (Wang et al., 2012a); (3) the subpopulation effect, which causes the detection of false positive QTL, can be corrected through fitting the MLM (Yu et al., 2005; Stich et al., 2008; Haseneyer et al., 2010; Massman et al., 2011); (4) the segregation of multiple alleles in a particular locus can be simultaneously observed, which is contrary to the bi-parental mapping's observation of the segregation of alleles from two parents (Flint-Garcia et al., 2003; Kraakman et al., 2004; Yu and Buckler., 2006; Cockram et al., 2008; Stich et al., 2008; Zhang et al., 2009; Rafalski., 2010). However, the bi-parental mapping can provide several benefits over the association mapping. (1) The bi-parental mapping population can show an additional phenotypic variation than an association mapping population by intercrossing or using very large progeny set (Massman et al., 2011). According to previous studies (Massman et al., 2011), the phenotypic variation among association mapping samples accounted for approximately 60 % of phenotypic variation among bi-parental samples, which illustrates that phenotype variation is considerably lower in association mapping and represents the lower mapping efficiency in association mapping than in bi-parental mapping (Massman et al., 2011). (2) The bi-parental mapping also requires less number of markers compared to the association mapping because the additional number of LD can be preserved through bi-parental crossing (Massman et al., 2011). As shown above, the bi-parental and association mapping methods have different benefits. The association mapping has recently become increasingly popular (Cockram et al., 2008; Zhang et al., 2009; Massman et al., 2011). Previous studies have reported that the resulting quality of association mapping is similar or better (Stich et al., 2008; Rafalski., 2010).

Conditions to improve the resolution of association mapping

For a successful association mapping, several conditions are required. (1) The population size has to be large enough (Melchinger et al., 1998; Rafalski., 2010; Massman et al., 2011; Wang et al., 2012a; Wang et al., 2012b). Based on simulation studies, Wang et al., (2012a) suggested that populations comprising greater than 384 individuals are ideal. Rafalski., (2010) stated that a proper population size is approximately 100-500 for mapping complex polygenic traits. (2) The heritability has to be sufficiently high (Melchinger et al., 1998; Yu and Buckler., 2006; Rafalski., 2010). Previous studies have consistently revealed that higher heritability leads to more precise QTL detections (Melchinger et al., 1998; Yu and Buckler., 2006; Massman et al., 2011). Low heritability and small sample size can cause severe upward bias (Melchinger et al., 1998). (3) The estimate of population structure must be taken into account (Yu et al., 2005; Yu and Buckler., 2006; Stich et al., 2008; Rafalski., 2010; Massman et al., 2011; Pasam et al., 2012). The genetic stratification arising from subpopulation structure could create a false positive marker-trait association because the admixture of subpopulations confounds the LD characteristics (Kraakman et al., 2004; Massman et al., 2011). To prevent the false positive QTL from being detected, the genetic stratification effect should be corrected (Yu el al., 2005; Stich et al., 2008; Zhang et al., 2009; Massman et al., 2011). The MLM can be fitted with a subpopulation matrix and a kinship matrix so that pure QTL can be harvested in a condition without external stratification impact. A number of studies showed that the MLM in the association mapping successfully corrected the subpopulation stratification and improved the power of QTL detection (Yu el al., 2005; Stich et al., 2008; Massman et al., 2011). (4) The marker density also has to be even and sufficient (Zhang et al., 2009; Rafalski., 2010). The marker density determines the LD interval as well as the precision of both subpopulation and kinship estimate so that higher marker density improves the mapping resolution (Yu and Buckler., 2006; Stich et al., 2008; Zhang et al., 2009).

Routine model of association mapping: unified mixed model approach

A set of samples for association mapping consists of a number of individuals that have historically different backgrounds (Massman et al., 2011). Diverse genetic backgrounds among individuals within a population for the association mapping necessarily accompany the static subpopulation stratification, which leads to the detection of the false-positive QTL (Kraakman et al., 2004; Massman et al., 2011) or returns the wrong degree of QTL effect (Huang et al., 2010) because the subpopulation structure skews the LD pattern (Yu and Buckler., 2006). To overcome this problem, Yu et al., (2005) presented the unified mixed model that uses kinship (K) and population structure (Q) matrices, which is so called QK model. The authors proved that the QK model improves the control of both type Ⅰ and type Ⅱ error rates in maize and human populations. In crop mapping studies, the QK model is becoming widely used. Previous studies (Kraakman et al., 2004; Pswarayi et al., 2008; Stich et al., 2008; Shi et al., 2011; Wang et al., 2012b) reported that the QK model showed a high performance in association mapping for various traits. Yu et al., (2005) analyzed the marker-trait association of 277 maize inbred lines for quantitative trait dissection and showed that the QK model was effective in reducing the detection of the false positive QTL than other association mapping models. Huang et al., (2010) mapped the marker-trait association for 14 agronomic traits in rice landrace collections and detected 80 marker-trait associations, which were often close to the previously known genes. Roy et al., (2010) mapped the QTL for spot blotch severity in wild barley and detected five previously known and seven new QTL. Massman et al., (2011) captured 12 QTL related to Fusarium head blight severity and deoxynivalenol concentration in barley germplasm and identified a frequent agreement between the resulting and the previous QTL as well as the newly detected QTL. Successful implementations of the QK model were shown in numerous studies (Stich et al., 2008; Shi et al., 2011). The use of K and Q matrices is optional. Previous studies (Stich et al., 2008; Wang et al., 2012b; Comadran et al., 2011) compared the QTL detection powers through the optional use of K and Q matrices. Stich et al., (2008) performed the association analyses using K or Q matrices in wheat population and found that the model including only the marker-based K matrix outperformed over the QK model. Similarly, in an association analysis of a barley collection, Wang et al., (2012b) reported the K model performed better than the QK model. Comadran et al., (2011) compared the K model and QK model and concluded that both models yielded similar results.

1.1.4 Genomic selection using ridge regression BLUP in self-pollinating crop

Concepts of genomic selection

Genomic selection (GS) aims to predict the genomic estimated breeding values (GEBVs) of individuals in descendant generations using the marker effects measured in ancestral generations. The GS requires two subpopulation sets: (1) a training set and (2) a validation set. The training set comprises ancestral individuals within which the measuring of every single marker effect is performed. The validation set contains descendent individuals, for whom the GEBVs are measured. Therefore, the implementation of the GS consists of two steps. In the first step, the marker effects are measured by capturing the marker's allelic variance in response to the phenotypic performance in the training set. In the next step, the prediction of the GEBVs is performed by applying the estimates of the marker effects to the genotyped individuals in the validation set.

Prospect of genomic selection

The GS helps breeders to select superior individuals at early growth stages. According to Schaeffer., (2006), because the correlation between true breeding values (TBVs) and GEBVs accounted for approximately 0.80, the expense could be reduced by 92 % compared with the cost of a traditional progeny test scheme in animal breeding programs. Furthermore, the increasing number of genomic probes and the decreasing cost of genomic profiling make the implementation of the GS more feasible (Meuwissen et al., 2001; Schaeffer et al., 2006; Goddard and Hayes., 2007; Zhong et al., 2009).

Approaches for genomic selection

Several linear models can be implemented for the GS, such as the least square model (LS), ridge regression best linear unbiased prediction (RR-BLUP) and Bayesian methods (Meuwissen et al., 2001; Xu., 2003; Hayes et al., 2009; Lorenzana and Bernardo., 2009; Asoro et al., 2011). These methods have different characteristics depending on the manner of measuring the marker effects. The LS assesses the marker effects based on single-marker regression and selects the markers associated with a trait through significance test (Meuwissen et al., 2001; Lorenzana and Bernardo., 2009). RR-BLUP measures the whole set of marker effects at one time under the assumption that genomic variances for all loci are equal. This assumption is not realistic because it cannot be true that all markers are equally of LD with QTL. However, this method provides sufficient results as well as an ease of implementation (Meuwissen et al., 2001; Lorenzana and Bernardo., 2009). The Bayesian methods are modeled based on BLUP. In fitting the model, the Bayesian methods usually use the Markov Chain Monte Carlo (MCMC) approach. The Bayesian methods routinely combine a prior distribution and data set point through which the unknown QTL-effects are captured with a posterior distribution. This process differentiates the degrees of allelic variance across loci depending on the degree of QTL effect (Meuwissen et al., 2001; Lorenzana and Bernardo., 2009).

Once the estimates of marker effect are obtained using one of the above methods, the GEBVs of individuals can be calculated by using the following equation (Solberg et al., 2008): $\mathrm{EBV_j} = \sum_{i=1}^{n} \mathrm{X_{ji}} \mathrm{g_i}$, where $\mathrm{EBV_j}$ = the GEBV of individual j; $\mathrm{X_{ji}}$ = the marker genotype of individual j; n = the number of markers; g = the estimates of the marker effects.

Ridge regression BLUP

As an approach for GS, the RR-BLUP become increasingly popular (Zhong et al., 2009; Habier et al., 2007; Asoro et al., 2011; Rutkoski et al., 2012; Zhao et al., 2012). The RR-BLUP assumes that the degrees of all marker effects are equal with a mean of zero. In principle, the RR-BLUP is equivalent to the BLUP. However, the RR-BLUP has two advanced features relative to the BLUP (Endelman., 2011): (1) the RR-BLUP is not confined in a condition that the number of markers cannot exceed the number of observations; (2) the RR-BLUP has a stable performance with highly correlated markers.

Cross-validation for measuring of prediction accuracy in RR-BLUP

The predictive ability of the GS model can be estimated using the Spearman correlation between TBVs and GEBVs. To measure the accuracy of the resulting correlation coefficient, previous studies often relied on cross-validation (Lorenzana and Bernardo., 2009; Zhao et al., 2012; Crossa et al., 2014). The cross-validation consists of routine procedures. In the first step, a population is randomly sub-divided into n subsets $\{S_k | S_1, \ldots, S_n\}$. In the second step, one subset (S_k) is allocated to the validation set and the remaining subsets to the training set. In the third step, the marker effects are estimated through fitting the model in the training set. In the fourth step, GEBVs for the genotyped individuals in the validation set are predicted using the marker estimates. In the fifth step, the correlation coefficients between GEBVs and TBVs are calculated and saved. In the sixth step, the procedures from the second to the fifth steps are repeated *n* times with a sequential increase of *k* at S_k. In the seventh steps, all above steps are repeated times of a moderate number, so the correlation coefficient can be calculated through averaging all the saved correlation coefficients at the fifth step.

Address of epistasis in RR-BLUP

The BLUP model primarily assumes that BV is an additive genetic effect, which leads to non-additive genetic effects such as epistasis and dominance being ignored (Schaeffer., 2006; Habier et al., 2007; Lorenzana and Bernardo., 2009). In selfing crop studies, no dominance effect is assumed because of the nearly symmetric structure of the diploid genome. Therefore, the non-additive genetic effect in selfing crops is considered to be epistasis. Previous studies showed that the GS method that considered only the additive genetic effects provided fair results (Meuwissen et al., 2001; Lorenzana and Bernardo., 2009; Asoro et al., 2011). However, this limited the full prediction of GEBVs because the epistasis effect was not considered. To overcome this problem, RR-BLUP uses the realized relationship model and created an appropriate kernel function, which helps capture epistatic effects (Piepho., 2009; Endelman., 2011).

Conditions to improve the prediction accuracy of RR-BLUP

The accuracy of GS is affected by several conditions: (1) Heritability: higher heritability enhances the prediction accuracy (Nielsen et al., 2009; Lorenzana and Bernardo., 2009; Zhong et al., 2009; Albrecht et al., 2011; Riedelsheimer et al., 2012). (2) The quantity of LD: an increasing number of LD steadily improves the prediction accuracy (Meuwissen et al., 2001; Zhong et al., 2009; Asoro et al., 2011). (3) The marker density: dense genome coverage with markers increases the prediction accuracy (Meuwissen et al., 2001; Zhong et al., 2009; Lorenzana and Bernardo., 2009; Nielsen et al., 2009; Asoro et al., 2011; Nagaya and Isobe., 2012; Crossa et al., 2014). (4) The size of a training set: the size of a training set and the prediction accuracy are positively proportional (Lorenzana and Bernardo., 2009; Nielsen et al., 2009; Zhong et al., 2009; Asoro et al., 2011; Heffner et al., 2011; Nagaya and Isobe., 2012; Crossa et al., 2014), particularly in a trait that shows low heritability because the increasing size of a training set efficiently improves the prediction accuracy (Lorenzana and Bernardo., 2009; Hayes et al., 2013). (5) The use of a genetic relationship matrix: the utilization of a genetic relationship matrix in RR-BLUP enhances the response of selection, especially in condition where either a high heritability or close relationship among members is embedded in a population (Habier et al., 2007; Zhong et al., 2009; Crossa et al., 2014). (6) The degree of relatedness between training and validation sets: the increasing degree of relatedness between training and validation sets improves the prediction accuracy (Habier et al., 2007; Muir., 2007; Hayes et al., 2009; Nielsen et al., 2009; Asoro et al., 2011; Crossa et al., 2014). (7) Large size of phenotypic records: the increasing number of phenotypic records improves the prediction accuracy (Meuwissen et al., 2001).

1.2 Objectives of this study

The objective of Chapter 2 is to develop a method to compute an NRM that can be incorporated into the BLUP procedure in self-pollinating crop. The objectives of Chapter 3 are to (1) examine BLUP using data sets from a self-pollinating crop in unbalanced trials, (2) examine BLUP with an NRM that considers self-pollination and (3) correlate the ranking of BLUP estimates with the ranking of phenotypic observations (mean phenotype). The objectives of Chapter 4 are to (1) detect QTL for grain yield, scald severity and net blotch severity and (2) investigate the effects of two different subpopulation structures on the association mapping. The objective of Chapter 5 is to determine the optimum conditions to improve the accuracy of estimating the GEBVs using RR-BLUP in self-pollinating crop.

2. Simple algorithm for enumerating the numerator relationship matrix with a selfing plant pedigree

2.1 Introduction

Relationship coefficients between a pair of individuals represent the degree of common alleles transmitted from common ancestors. By definition, the relationship coefficient between two individuals is equal to twice the inbreeding coefficient of their offspring (Emik and Terrill., 1949). In this respect, the relationship coefficient and the inbreeding coefficient are compatible. As a method to compute the relationship coefficient, Emik and Terrill devised a simple algorithm that fills a square matrix with the relationship coefficients among all pair of individuals within a population. This matrix is termed the numerator relationship matrix (NRM).

In breeding programs, either the relationship coefficient or the inbreeding coefficient conventionally provides the useful index for mating design (Caballero and Toro., 2002; Lynch and Ritland., 1999). In recent years, the usefulness of an NRM has risen particularly in genetic modeling fields for the purposes of breeding value prediction (Henderson., 1975; Oakey et al., 2006; Panter and Allen., 1995a; Panter and Allen., 1995b) and association mapping (Stich et al., 2008). An NRM in the genetic modeling is routinely used in animal studies, whereas its application is unusual in plant studies (Piepho et al., 2008). This fact is due to the following reasons. (1) The current method of constructing the NRM requires compact pedigree records, whereas the plant pedigree often omits the intermediate entries between the crossed progenitors and the progeny; (2) the current NRM assumes that individuals in a population were crossed, whereas selfing often occurs in plant species. To the best of my knowledge, the architecture of the relationship coefficient among individuals in a selfing population has not yet been revealed. It may not be an exaggeration that studies regarding the inbreeding coefficient or relationship coefficient based on the pedigree have not advanced since Emik and Terrill's publication. Modern studies tend to focus on statistical inferences coupled with genotypic observations to obtain an NRM (Gutierrez et al., 2005; Hardy and Vekemans., 2002). Such a circumstance motivated to develop a traditional manner for computing an NRM using plant pedigrees in a selfing population. This study presents new equations to define the relationship coefficient among individuals in a selfing population and a new plant pedigree format that carries all arguments that the equations require. On this basis, a software tool, designated as PopKin, was developed, which is presented in this Chapter. The PopKin could be widely useful for plant studies that demand the NRM for the purpose of genetic modeling or mating design. In this paper, both theories and technical methods concerning the NRM in a selfing plant population are addressed.

Objective

The objective of this study is to develop a method to compute an NRM that can be incorporated into the BLUP procedure in self-pollinating crop.

2.2 Theories and Methods

Rules of plant pedigree notations

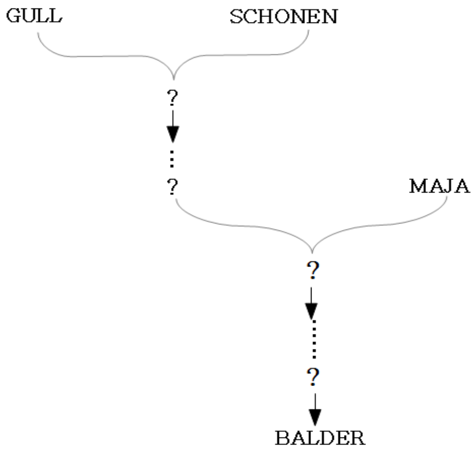
Plant pedigree notation has its own unorthodox rules, which do not represent the complete picture. Let us exemplify with the following real barley pedigree (http://genbank.vurv.cz/barley/pedigree/)

BALDER	GULL / SCHONEN // MAJA
(Offspring)	(Ancestors)

Using the current rules of plant pedigree notation, a symbol “/” represents the mating event, and its count implies the order of mating. The above pedigree notation indicates that GULL and SCHONEN were first crossed. In turn, their progeny was mated with MAJA. However, BALDER might have undergone selfing several times for the purpose of fixation because barley is a selfing species. Here, it can be known that the plant pedigree omits an offspring from the GULL and SCHONEN cross and the selfing entries during a period of a fixation. A pedigree skeleton is given in Figure 2-1.

Figure 2-1. A plant pedigree example. This was referred from http://genbank.vurv.cz/barley/pedigree/. A symbol “/” implies a mating event, and its count represents the mating order. The greater the count of multiple slashes is, the more recently the hybridization was made, and vice versa. A symbol “?” represents an unknown parentage. Plant pedigree does not inform intermediate entries between the crossed progenitors and a progeny as well as the number of selfings that has undergone during a fixation.

Balder: Gull / Schonen // Maja



Emik and Terrill's equations that define the additive relationship coefficient

Emik and Terrill's equations for computing an NRM can be notated as:

(1) If two different individuals, i and j, are unknown

$$f_{i,j} = f_{j,i} = 0 \qquad \text{(Equation 2-1)}$$

(2) If two identical individuals, i and i, are unknown

$$f_{i,i} = 1 \qquad \text{(Equation 2-2)}$$

(3) If two different individuals, i and j, are known

$$f_{i,j} = \frac{1}{2}\,(f_{p,j} + f_{q,j}) \qquad \text{(Equation 2-3)}$$

where $f_{i,j}$ represents the relationship coefficient between two different individuals, i and j; $f_{p,j}$ represents the relationship coefficient between i's parent, p, and j; $f_{q,j}$ represents the relationship coefficient between i's parent, q, and j. Note that j is any individual that satisfies no descendent of i (Chang et al., 1991).

(4) If two identical individuals, i and i, are known

$$f_{i,i} = 1 + \frac{1}{2}\,f_{p,q} \qquad \text{(Equation 2-4)}$$

where $f_{i,i}$ represents the relationship coefficient between two identical individuals, i; $f_{p,q}$ represents the relationship coefficient between i's two parents, p and q.

Creation of empirical pair-wise relationship coefficients

To the best of my knowledge, the pattern of relationship coefficients in a selfing population was unveiled. In this study, however, the pattern was revealed using computer simulations. For this work, the meiosis mechanism was materialized, whose procedures are:

1. Generating founders' diploid chromosomes by both creating and coupling two vectors containing 10,000 of the same codes in each vector. Each pair of vectors is considered a diploid genome of the founder. Note that the codes for each founder's vector should be unique across whole sets of vectors.

2. To imitate the meiosis before crossing two individuals, create a haploid vector by randomly replacing half the number of codes (5,000) in one vector with codes in the same loci of an opposite vector within an individual. Subsequently, choose one vector as a gamete.

3. When two individuals are mated, couple two gamete vectors from both individual. Subsequently, randomly exchange half the number of codes over the coupled gamete vectors. The resultant pair of vectors represents a diploid chromosome of a new descendent.

4. Given any two individuals obtained through the above procedures, the relationship coefficient between two can be computed using Equation 2-5 (Bernardo., 2002).

$$f_{A,B} = \frac{1}{2}\ [P(a1 \equiv b1) + P(a1 \equiv b2) + P(a2 \equiv b1) + P(a2 \equiv b2)] \qquad \text{(Equation 2-5)}$$

where a1 and a2 are gametes of individual, A; b1 and b2 are gametes of individual B; P(a1≡b1) is a proportion that a1 and b1 share the identical alleles from common ancestors; P(a1≡b2) is a proportion that a1 and b2 share the identical alleles from common ancestors; P(a2≡b1) is a proportion that a2 and b1 share the identical alleles from common ancestors; P(a2≡b2) is a proportion that a2 and b2 share the identical alleles from common ancestors.

5. The above steps (1 to 4) have to be conducted throughout whole pair-wise combinations of individuals within a population. Subsequently, an NRM should be built by filling out all elemental positions of the matrix with corresponding values obtained through the above procedures.

6. To reduce simulation error, replicate the above procedures for a moderate number of times.

In the above procedures, it was assumed that irregular chromosomal activities such as insertions, deletions, transposon events, and inversions were absent. For practicing the simulation with a plant pedigree, three sub-pedigrees in Figure 2-2 will be used. The simulation was carried out using R software version 2.15.0 (R Development Core Team., 2012).

Equations defining the relationship coefficients

By analyzing the pattern of the relationship coefficients in a simulated selfing population, it was identified that the relationship coefficients in a selfing population are estimable using Emk and Terrill's method by completing the parents-offspring pedigree table using dummy entries. The manner to record the selfing in a pedigree is to duplicate the same parent's name for paternal and maternal cells in the pedigree table. This revelation opens up the way to compute an NRM in a selfing population. However, the use of Emik and Terrill's equations in a selfing population is accompanied by the following problems. (1) The conversion of a plant pedigree notation to a pedigree table is labor intensive and brings risks introducing mistakes arising from creating and utilizing dummy entries; (2) the use of dummy entries makes the size of an NRM larger, leading to computational burden. To avoid such obstacles, new equations, which define the relationship coefficient between the progenitors and the progeny in the pedigree of a selfing plant, were derived by generalizing Emik and Terrill's equations (Equations 2-3, 2-4), whose notations are given in Equations 2-6 to 2-8.

(1) Relationship coefficient between two different individuals

$$f_{i,j} = \sum_{k=1}^{n} \lambda_j^{d[k]} f_{i,d[k]} \qquad \text{(Equation 2-6)}$$

$f_{i,j}$	Relationship coefficient between i and j
$d[k]$	j's kth progenitor on pedigree
n	Total number of j's progenitors on pedigree
k	The incremental order of the crossed progenitors from left to right
$f_{i,d[k]}$	Relationship coefficient between d[k] and i
$\lambda_j^{d[k]}$	Allele transmission rate from d[k] to j

(2) Relationship coefficient between two identical individuals. Two types of notations are available

$$f_{X[n],X[n]} = \sum_{k=1}^{n} \frac{1}{2^{k-1}} + \frac{1}{2^{n-2}} \sum_{x=1}^{g} \sum_{y=1}^{m} \lambda_{X[1]}^{p[x]} \lambda_{X[1]}^{m[y]} f_{p[x]m[y]} \qquad \text{(Equation 2-7)}$$

n	Value for n in the F_n generation
$X[n]$	Progeny in the F_n generation

$f_{X[n],X[n]}$	Relationship coefficient between two identical progenies, $X[n]$
x	The incremental order of progenitors on the paternal side
y	The incremental order of progenitors on the maternal side
g	The total number of progenitors on the paternal side of the pedigree
m	The total number of progenitors on the maternal side of the pedigree
p[x]	x^th progenitor on the paternal side of the pedigree
m[y]	y^th progenitor on the maternal side of the pedigree
$\lambda_{X[1]}^{p[x]}$	Allele transmission rate from p[x] to X[1]
$\lambda_{X[1]}^{m[y]}$	Allele transmission rate from m[y] to X[1]

$$f_{X[n],X[n]} = 1 + \log_n n \sum_{k=2}^{n} \frac{1}{2^{k-1}} + 2\left(1 - \log_n n \sum_{k=2}^{n} \frac{1}{2^{k-1}}\right) \sum_{x=1}^{g} \sum_{y=1}^{m} \lambda_{X[1]}^{p[x]} \lambda_{X[1]}^{m[y]} f_{p[x]m[y]} \qquad \text{(Equation 2-8)}$$

$$(\log_n n < n)$$

n	Value for n in the F_n generation
X[n]	Progeny in the F_n generation
$f_{X[n],X[n]}$	Relationship coefficient between two identical progenies, $X[n]$
x	The incremental order of progenitors on the paternal side
y	The incremental order of progenitors on the maternal side
g	The total number of progenitors on the paternal side of the pedigree
m	The total number of progenitors on the maternal side of the pedigree
p[x]	x^th progenitor on the paternal side of the pedigree
m[y]	y^th progenitor on the maternal side of the pedigree
$\lambda_{X[1]}^{p[x]}$	Allele transmission rate from p[x] to X[1]
$\lambda_{X[1]}^{m[y]}$	Allele transmission rate from m[y] to X[1]

Equations 2-6 to 2-8 are applicable for non-founder entries. In the case of two parental individuals that are founders, Equations 2-1 and 2-2 should be applied. Above, Equations 2-7 and 2-8 are compatible because both equations always give the same results.

Software, PopKin, for enumerating an NRM with a plant pedigree

This Chapter presents a software tool for computing an NRM with a plant pedigree depending on Equations 2-1, 2-2, 2-6 and 2-7, designated as PopKin (Population Kinship calculator). PopKin was written in C++ and provides an NRM as a result of the computation.

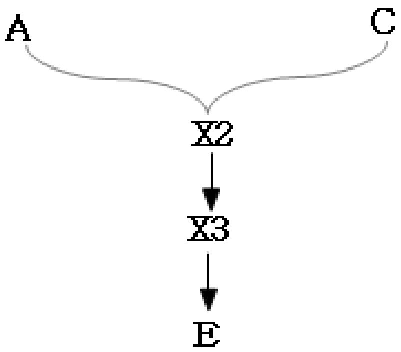
2.3 Results and Discussion

2.3.1 Emik and Terrill's method is applicable to a selfing species

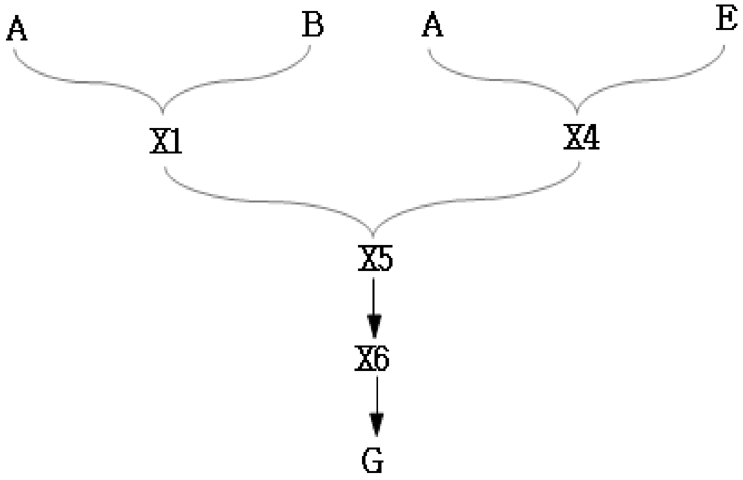
For the sake of practicing the construction of an NRM in a selfing species, plant pedigrees in Figure 2-2 will be used. The notation of each pedigree, following the standard format, is stated above in each sub-figure.

Figure 2-2. Pedigree example for three plant individuals (E, G and J). All the members are closely related because they frequently share common ancestors. The pedigree notation for each pedigree is stated above each pedigree figure following the standard format for plant pedigree notation. In these figures, the entries having with a prefix, X, are unknown parentage, whereas the entries, A, B, C, E, G and J are present in the pedigree notations.

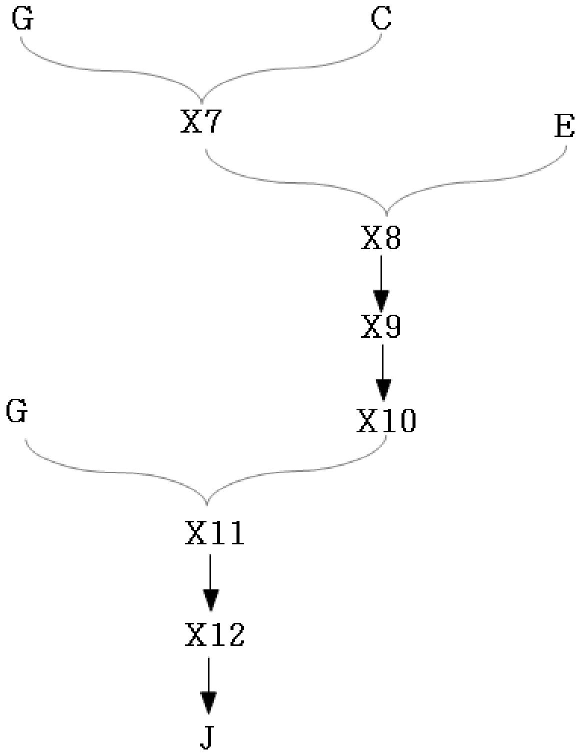
(a) E: A / C



(b) G: A / B // A / E



(c) J: G /// E // C / G



As observed above, plant pedigree notation does not provide any entries between the crossed progenitors and the progeny. In Figure 2-2, every progeny (E, G and J) was in the F_3 generation through two times of selfing. All pedigrees are related to one another through sharing common entries. In Figure 2-2, entries having a prefix, X, represent that they are shown in a pedigree skeleton but absent in the pedigree notation. Using Figure 2-2, the NRMs were constructed using the simulation as well as Emik and Terrill's method, whose results are displayed in Tables 2-1 (simulation) and 2-2 (Emik and Terrill's method). The comparison shows that the both matrices are almost identical, which shows that the relationship coefficients in a selfing species can be obtained using Emik and Terrill's method.

Table 2-1. Numerator relationship matrix obtained by simulation with Figure 2-2. To minimize an error, the simulation was replicated 1,000 times. The entries having a prefix, X, are absent in the plant pedigree notation, whereas the others (A, B, C, E, G and J) are present.

	A	B	C	X1	X2	X3	E	X4	X5	X6	G	X7	X8	X9	X10	X11	X12	J
A	1	0	0	0.5	0.5	0.49984	0.499908	0.749926	0.625049	0.624981	0.624983	0.312504	0.406227	0.406308	0.406328	0.515639	0.515535	0.515503
B	0	1	0	0.5	0	0	0	0	0.249877	0.250028	0.250027	0.12497	0.062524	0.062517	0.06254	0.1563	0.156357	0.156386
C	0	0	1	0	0.5	0.50016	0.500092	0.250074	0.125075	0.124992	0.124991	0.562525	0.531249	0.531174	0.531132	0.328062	0.328108	0.32811
X1	0.5	0.5	0	1	0.250033	0.249984	0.250001	0.375039	0.687457	0.687703	0.687704	0.343816	0.296965	0.296965	0.296951	0.492393	0.49239	0.49239
X2	0.5	0	0.5	0.250033	1	1	1	0.749929	0.500129	0.499952	0.500006	0.500045	0.750088	0.75013	0.750104	0.625011	0.625002	0.625077
X3	0.49984	0	0.50016	0.249984	1	1.500159	1.500159	0.999967	0.625268	0.625044	0.625144	0.562728	1.03154	1.031594	1.031553	0.82828	0.828322	0.828476
E	0.499908	0	0.500092	0.250001	1	1.500159	1.750061	1.124959	0.687668	0.687369	0.687451	0.593855	1.172106	1.172182	1.17205	0.929694	0.929724	0.929821
X4	0.749926	0	0.250074	0.375039	0.749929	0.999967	1.124959	1.249971	0.812585	0.812279	0.812275	0.53122	0.828144	0.828307	0.8283	0.820291	0.820272	0.820302
X5	0.625049	0.249877	0.125075	0.687457	0.500129	0.625268	0.687668	0.812585	1.187503	1.187503	1.187503	0.656277	0.672067	0.67218	0.672192	0.929863	0.92989	0.92985
X6	0.624981	0.250028	0.124992	0.687703	0.499952	0.625044	0.687369	0.812279	1.187503	1.593664	1.593664	0.859324	0.773474	0.773632	0.773692	1.183737	1.183652	1.183621
G	0.624983	0.250027	0.124991	0.687704	0.500006	0.625144	0.687451	0.812275	1.187503	1.593664	1.796835	0.960921	0.824384	0.82444	0.824576	1.31073	1.310687	1.310574
X7	0.312504	0.12497	0.562525	0.343816	0.500045	0.562728	0.593855	0.53122	0.656277	0.859324	0.960921	1.062546	0.828178	0.828152	0.828185	0.894522	0.894467	0.894404
X8	0.406227	0.062524	0.531249	0.296965	0.750088	1.03154	1.172106	0.828144	0.672067	0.773474	0.824384	0.828178	1.296999	1.296999	1.296999	1.060645	1.060561	1.060573
X9	0.406308	0.062517	0.531174	0.296965	0.75013	1.031594	1.172182	0.828307	0.67218	0.773632	0.82444	0.828152	1.296999	1.648444	1.648444	1.236394	1.236313	1.236406
X10	0.406328	0.06254	0.531132	0.296951	0.750104	1.031553	1.17205	0.8283	0.672192	0.773692	0.824576	0.828185	1.296999	1.648444	1.82402	1.324258	1.324206	1.3244
X11	0.515639	0.1563	0.328062	0.492393	0.625011	0.82828	0.929694	0.820291	0.929863	1.183737	1.31073	0.894522	1.060645	1.236394	1.324258	1.412285	1.412285	1.412285
X12	0.515535	0.156357	0.328108	0.49239	0.625002	0.828322	0.929724	0.820272	0.92989	1.183652	1.310687	0.894467	1.060561	1.236313	1.324206	1.412285	1.706269	1.706269
J	0.515503	0.156386	0.32811	0.49239	0.625077	0.828476	0.929821	0.820302	0.92985	1.183621	1.310574	0.894404	1.060573	1.236406	1.3244	1.412285	1.706269	1.853113

Table 2-2. Numerator relationship matrix resolved by Emik and Terrill's algorithm. Entries having a prefix, X, are absent in the plant pedigree notations in Figure 2-2, whereas the others (A, B, C, E, G, J) are present in the pedigree notations. This result is in agreement with the simulation results shown in Table 2-1.

	A	B	C	X1	X2	X3	E	X4	X5	X6	G	X7	X8	X9	X10	X11	X12	J
A	1	0	0	0.5	0.5	0.5	0.5	0.75	0.625	0.625	0.625	0.3125	0.40625	0.40625	0.40625	0.515625	0.515625	0.515625
B	0	1	0	0.5	0	0	0	0	0.25	0.25	0.25	0.125	0.0625	0.0625	0.0625	0.15625	0.15625	0.15625
C	0	0	1	0	0.5	0.5	0.5	0.25	0.125	0.125	0.125	0.5625	0.53125	0.53125	0.53125	0.328125	0.328125	0.328125
X1	0.5	0.5	0	1	0.25	0.25	0.25	0.375	0.6875	0.6875	0.6875	0.34375	0.296875	0.296875	0.296875	0.492188	0.492188	0.492188
X2	0.5	0	0.5	0.25	1	1	1	0.75	0.5	0.5	0.5	0.5	0.75	0.75	0.75	0.625	0.625	0.625
X3	0.5	0	0.5	0.25	1	1.5	1.5	1	0.625	0.625	0.625	0.5625	1.03125	1.03125	1.03125	0.828125	0.828125	0.828125
E	0.5	0	0.5	0.25	1	1.5	1.75	1.125	0.6875	0.6875	0.6875	0.59375	1.17188	1.17188	1.17188	0.929688	0.929688	0.929688
X4	0.75	0	0.25	0.375	0.75	1	1.125	1.25	0.8125	0.8125	0.8125	0.53125	0.828125	0.828125	0.828125	0.820312	0.820312	0.820312
X5	0.625	0.25	0.125	0.6875	0.5	0.625	0.6875	0.8125	1.1875	1.1875	1.1875	0.65625	0.671875	0.671875	0.671875	0.929688	0.929688	0.929688
X6	0.625	0.25	0.125	0.6875	0.5	0.625	0.6875	0.8125	1.1875	1.59375	1.59375	0.859375	0.773438	0.773438	0.773438	1.18359	1.18359	1.18359
G	0.625	0.25	0.125	0.6875	0.5	0.625	0.6875	0.8125	1.1875	1.59375	1.79688	0.960938	0.824219	0.824219	0.824219	1.31055	1.31055	1.31055
X7	0.3125	0.125	0.5625	0.34375	0.5	0.5625	0.59375	0.53125	0.65625	0.859375	0.960938	1.0625	0.828125	0.828125	0.828125	0.894531	0.894531	0.894531
X8	0.40625	0.0625	0.53125	0.296875	0.75	1.03125	1.17188	0.828125	0.671875	0.773438	0.824219	0.828125	1.29688	1.29688	1.29688	1.06055	1.06055	1.06055
X9	0.40625	0.0625	0.53125	0.296875	0.75	1.03125	1.17188	0.828125	0.671875	0.773438	0.824219	0.828125	1.29688	1.64844	1.64844	1.23633	1.23633	1.23633
X10	0.40625	0.0625	0.53125	0.296875	0.75	1.03125	1.17188	0.828125	0.671875	0.773438	0.824219	0.828125	1.29688	1.64844	1.82422	1.32422	1.32422	1.32422
X11	0.515625	0.15625	0.328125	0.492188	0.625	0.828125	0.929688	0.820312	0.929688	1.18359	1.31055	0.894531	1.06055	1.23633	1.32422	1.41211	1.41211	1.41211
X12	0.515625	0.15625	0.328125	0.492188	0.625	0.828125	0.929688	0.820312	0.929688	1.18359	1.31055	0.894531	1.06055	1.23633	1.32422	1.41211	1.70605	1.70605
J	0.515625	0.15625	0.328125	0.492188	0.625	0.828125	0.929688	0.820312	0.929688	1.18359	1.31055	0.894531	1.06055	1.23633	1.32422	1.41211	1.70605	1.85303

2.3.2 Pattern of relationship coefficients in response to selfing

The construction of an NRM constitutes two types. The first type is the calculation of the relationship coefficients among two different individuals. The second type is the calculation of the relationship coefficients among two identical individuals. Here, the pattern of the relationship coefficient in response to the selfing in each part will be addressed.

To view the pattern of the relationship coefficient between two different individuals in response to the selfing, let us choose Figure 2-2 (b). This pedigree contains four crossed progenitors (A, B, A and C) and a progeny (G). G is in the F_3 generation through two times of selfing. Here, to view the relationship coefficients between any one progenitor and several progenies in the same selfing series, the comparison across $f_{A, X5}$, $f_{A, X6}$ and $f_{A, G}$ was made using the NRM in Table 2-2. The comparison resulted in $f_{A, X5} = f_{A, X6,} = f_{A, G} = 0.625$, which indicates no difference. In the case of using other offsprings, B and E, the resultomg relationship coefficients were shown to be $f_{B, X5} = f_{B, X6,} = f_{B, G} = 0.25$ and $f_{E, X5} = f_{E, X6,} = f_{E, G} = 0.6875$. Those results consistently indicate that in computing the relationship coefficient between two different individuals, any individual can be replaced with a different entry in the same selfing series.

To view the pattern of the relationship coefficient between two identical individuals in response to the selfing, let us use Figure 2-2 (c). In this pedigree, two series of selfings exist. Regarding the first series of the selfing, $f_{X11,X11}$ is shown to be 1.41211, which was gained by $1 + 0.5f_{X10,G}$ and dependant on Equation 2-4. An equal value is attainable by replacing X_{10} with either X_8 or X_9 in the same selfing series ($1 + 0.5f_{X8,G} = 1+0.5f_{X9,G} = 1 + 0.5f_{X10,G} = 1.41211$). This result suggests that in computing the relationship coefficient between two identical individuals, if the individual's parents are different, any parent can be replaceable with any entry in a selfing series to which the parent belongs. As another discovery, the values for $f_{X9,X9}$ and $f_{X10,X10}$ in the first selfing series are 1.64844 and 1.82422, respectively, whereas the values are 1.70605 for $f_{X12,X12}$ and 1.85303 for $f_{J, J}$ in the second selfing series. In the four entries (X9, X10, X12 and J), their parents are identical. This result suggests that in each selfing series, if an individual's two parents are identical, the relationship coefficients noticeably increase as the selfing process proceeds.

2.3.3 Method for computing an NRM holds key integrity

With regards to Equation 2-3, Chang et al., (1991) stressed a condition that in $f_{i,j}$, an individual, i, should not be a descendent of an individual, j, and vice versa because Equation 2-3 cannot be applicable to the parent-offspring relationship. The authors described that Emik and Terrill's method systematically avoids the violation of the aforementioned requirement. Equation 2-6 was obtained by extending Equation 2-3. Therefore, the integrity required for Equation 2-3 holds for Equation 2-6.

2.3.4 Advantages and disadvantage of the current plant pedigree

In applying Equations 2-6 to 2-8 to a plant pedigree, the current format of plant pedigree notation has four advantages and one disadvantage. The advantages are that (1) the condition, that the omitted entries should not be duplicated, is never violated; (2) across any two individuals' pedigrees, any common entries are not missed; (3) the distance from the progeny to the crossed progenitors are clearly measurable using multiple slashes (/) under the condition that selfing is ignored; (4) both parental partitions are distinguishable once using the maximum number of successive slashes. Meanwhile, the single disadvantage is that the value for n in the $\mathrm{F_n}$ progeny is not informed. This problem is particularly confined to the pedigree of selfing crops such as rice, wheat and barley.

2.3.5 Identification of the arguments for Equation 2-6 from a plant pedigree

Equation 2-6 defines the relationship coefficient between two different entries. Depending on the aforementioned four advantages, with any given plant pedigree, all arguments for Equation 2-6 can be gained, which are (1) the order of the crossed progenitors (k) on the plant pedigree notation and (2) the allele transmission rate ($\lambda_j^{d[k]}$) from each crossed progenitor (d[k]) to a progeny (j) under the condition that the selfing process is ignored.

For example, in Figure 2-2 (b), a progeny (G) has four crossed progenitors (A, B, A and E) in a grandparental generation. G is in the F_3 generation through twice of selfing. The allele transmission rates from each progenitor to the progeny resulted in λ_G^A = 0.25, λ_G^B = 0.25, λ_G^A = 0.25 and λ_G^E = 0.25 because X5 and X6 in the selfing series should be ignored. Likewise, the arguments for other sub-pedigrees in Figure 2-2 can be obtained in the same manner, summarized in Table 2-3.

Table 2-3. Arguments for Equation 2-6, gathered from pedigrees in Figure 2-2.

j	k	d[k]	$\lambda_j^{d[k]}$
E	1	A	$\lambda_E^A = 0.5$
	2	C	$\lambda_E^C = 0.5$
G	1	A	$\lambda_G^A = 0.25$
	2	B	$\lambda_G^B = 0.25$
	3	A	$\lambda_G^A = 0.25$
	4	E	$\lambda_G^E = 0.25$
J	1	G	$\lambda_J^G = 0.5$
	2	G	$\lambda_J^G = 0.125$
	3	C	$\lambda_J^C = 0.125$
	4	E	$\lambda_J^E = 0.25$

j	A progeny
k	The incremental order of the crossed progenitors from left to right
d[k]	j's kth progenitor on the pedigree
$\lambda_j^{d[k]}$:	Allele transmission rate from d[k] to j

2.3.6 Identification of the arguments for Equations 2-7 and 2-8 from a plant pedigree

Equations 2-7 and 2-8 define the relationship coefficient between the two identical individuals. Although two equations are differently denoted, the resolutions resulting from the two equations are always the same. Equations 2-7 and 2-8 require the arguments: (1) the partition of the crossed progenitors into parental sides; (2) the progenitors' orders within each parental partition (x and y); (3) the number of the crossed progenitors (g and m); (4) the allele transmission rate from each crossed progenitor to a progeny ($\lambda_{X[1]}^{p[x]}$ and $\lambda_{X[1]}^{m[y]}$) under the condition that the selfing process is ignored; (5) the value of n in the F_n progeny. To explain the identification of the above arguments for Equations 2-7 and 2-8, let us use Figure 2-2 (c). In a plant pedigree, the partition of the crossed progenitors into both parental sides can be easily achieved by referring to the maximum count of slashes (/). In Figure 2-2 (c), the maximum number of slashes is three. Therefore, the entries G, C and E on the right hand side belong to a paternal (or a maternal) partition, whereas the entry, G, on the left hand side to a maternal (or a paternal) partition. The order (x and y) of the crossed progenitors can be incrementally given from left to right within each parental partition. In Figure 2-2 (c), the order number within each parental partition can be given as x = 1 for G on the left hand partition, whereas y = 1 for G, y = 2 for C, and y = 3 for E on the right hand partition (x and y are interchangeable). The allele transmission rate can be simply obtained by 2^t, where t is the distance between a crossed progenitor and a progeny under a condition that the selfing is ignored. In Figure 2-2 (c), E is four generations away from X_{11}. Although three entries (X8, X9, X10) between X_{11} and E are located in the first selfing series, this distance should be ignored. Likewise, two entries (X11 and X12) in the second selfing series should not be accounted for. Hence, the allele transmission rates resulted in 0.5 for λ_J^G, 0.125 for λ_J^G, 0.125 for λ_J^C, and 0.25 for λ_J^E. The value for n in the F_n progeny cannot be gained in the plant pedigree notation. However, in the pedigree skeleton of Figure 2-2 (c), J is in the F_3 generation through twice selfing. Hence, n = 3 can be obtained. In the other sub-pedigrees in the Figure 2-2, the arguments for Equations 2-7 and 2-8 can be likewise obtained, which are summarized in Table 2-4.

Table 2-4. Arguments for Equations 2-7 and 2-8, gathered from Figure 2-2.

X[n]	n	x or y	p[x] or m[y]	$\lambda_{X[n]}^{p[x]}$ or $\lambda_{X[n]}^{m[x]}$
E	3	x = 1	p[1] = A	λ_E^A = 0.5
		y = 1	m[1] = C	λ_E^C= 0.5
G	3	x = 1	p[1] = A	λ_G^A = 0.25
		x = 2	p[2] = B	λ_G^B = 0.25
		y = 1	m[1] = A	λ_G^A = 0.25
		y = 2	m[2] = E	λ_G^E = 0.25
J	3	x = 1	p[1] = G	λ_J^G = 0.5
		y = 1	m[1] = G	λ_J^G = 0.125
		y = 2	m[2] = C	λ_J^C = 0.125
		y = 3	m[3] = E	λ_J^E = 0.25

X[n]	Progeny in the F_n generation
n	Value for n in the F_n generation
x	The incremental order of progenitors on the paternal side
y	The incremental order of progenitors on the maternal side
p[x]	xth progenitor on the paternal side of the pedigree
m[y]	yth progenitor on the maternal side of the pedigree
$\lambda_{X[1]}^{p[x]}$	Allele transmission rate from p[x] to X[1]
$\lambda_{X[1]}^{m[y]}$	Allele transmission rate from m[y] to X[1]

2.3.7 New rule of pedigree notation conveying all arguments for Equations 2-6 to 2-8

Equations 2-6 to 2-8 define the relationship coefficient between the crossed progenitors and the progeny in a selfing plant pedigree. The current rule of plant pedigree notation does not convey the value for n in the F_n progeny. For resolving this problem, I propose a new pedigree notation format that carries all arguments for Equations 2-6 to 2-8. In the current format, the plant pedigree depicts the distances from the crossed parents to the progeny by using multiple slashes. Instead, I suggest the use of parentheses and using the slash a single time. This format forms the mating block at every mating event. Here, the value for n in the F_n progeny can be simply added in the right upper corner. Following these rules, the three sub-pedigrees in Figure 2-2 can be denoted as given in Table 2-5.

Table 2-5. Comparison of the current format of plant pedigree and proposed format of plant pedigree in this study. Both formats express the pedigrees in Figure 2-2. The current format of plant pedigree uses multiple slashes. In contrast, the proposed format of plant pedigree uses the slash a single time. Instead, the proposed format of plant pedigree uses parentheses, which make the mating block at each cross. In the proposed format of plant pedigree, the value for n in F_n is marked in the right upper corner.

Progeny	The current format of plant pedigree	The proposed format of plant pedigree
E	A / C	$(A / C)^3$
G	A / B // A / E	$((A / B) / (A / E))^3$
J	G /// G / C // E	$(G / ((G / C) / E))^3$

2.3.8 Syntax of PopKin

The plant pedigree format proposed above can convey all arguments required in Equations 2-6 to 2-8, which makes the computation of an NRM feasible. For this work, the pedigree should be sorted from progenitor to progeny. Hence, pedigrees of the crossed progenitors should precede that of a progeny. This idea is in the same context that Emik and Terrill's method requires parent-offspring sorting.

In this study, a software tool, PopKin, for constructing an NRM is presented. In the PopKin's syntax, the method to block the mating event with parentheses and a single slash is the same as stated above (see Table 2-5). In addition, the PopKin requires symbols, "$" and "^". The "$" separates a progeny and the crossed progenitors in each pedigree record. The "^" indicates that the following number is the n in the F_n progeny. The PopKin is featured to sort the pedigrees from progenitors to progeny. According to these rules, the sub-pedigree notations in Figure 2-2 can be denoted as:

```
E $ (A / C)^3

G $ ((A / B) / (A / E))^3

J $ (((G / C) /E) / G)^3
```

2.3.9 Integrity validation of the present method

Here, the integrity of the present method was demonstrated through comparing an NRM resolved by the PopKin (Table 2-6) with that resolved by Emik and Terrill's method (Table 2-2) with Figure 2-2. The matrix's label in Table 2-6 only comprises entries present in three sub-pedigree notations of Figure 2-2. The size of the matrix is 6 by 6. In contrast, the size of the matrix that includes the omitted entries is 18 by 18 as observed in Table 2-2. Importantly, the relationship coefficients among common entries across Table 2-2 and Table 2-6 are exactly the same. This finding illustrates the usefulness of the proposed method in terms of accuracy as well as the feasibility of constructing a small-sized NRM.

Table 2-6. Numerator relationship matrix constructed using Equation 2-6 to 2-8 with Figure 2-2. The computation was performed using the PopKin software tool. This matrix contains the relationship coefficients among entries present in the pedigree notations, whereas the matrix in Table 2-2 comprises the relationship coefficients among entries present in plant pedigree notations as well as unknown entries between the crossed progenitors and a progeny. Here, the values among common entries across Tables 2-6 and 2-2 are exactly the same. However, the size (6×6) of this matrix is considerably smaller compared with the size (18×18) of Table 2-2.

	A	B	C	E	G	J
A	1	0	0	0.5	0.625	0.515625
B	0	1	0	0	0.25	0.15625
C	0	0	1	0.5	0.125	0.328125
E	0.5	0	0.5	1.75	0.6875	0.929688
G	0.625	0.25	0.125	0.6875	1.79688	1.31055
J	0.515625	0.15625	0.328125	0.929688	1.31055	1.85303

3. Breeding value estimation using BLUP in a German spring barley collection

3.1 Introduction

The best linear unbiased prediction (BLUP) is a statistical approach used to obtain the selection index of germplasms and estimate heritability in breeding programs. BLUP is based on a mixed linear model (MLM) that consists of random effect and fixed effect variables. Random effects assume a correlation among objects to be predicted, whereas fixed effects assume independence and non-relatedness among objects to be predicted. In BLUP, the consideration of a correlation within random variables helps enhance the prediction accuracy (Soh., 1994; Crossa et al., 2006; Piepho et al., 2008; Viana et al., 2009). It is noted that BLUP provides a great selection than phenotypic observation or models that only consider fixed effects (Panter and Allen., 1995a; Panter and Allen., 1995b; Durel et al., 1998; Bromley et al., 2000; Crossa et al., 2006; Bauer et al., 2008; Akin et al., 2009; Piepho et al., 2009). BLUP provides the following advantages. (1) A degree of a genetic correlation among breeding germplasms can be accounted for, which relates relatives to each individual (Soh., 1994; Durel et al., 1998; Crossa et al., 2006; Bauer et al., 2008; Piepho et al., 2008; Viana et al., 2009); (2) breeding value estimates can be corrected by filtering environmental, treatment and replication effects from the phenotypic observations (Soh., 1994; Durel et al., 1998; Crossa et al., 2006); (3) the BV estimation's bias arising from an unbalanced data structure can be refined by shrinking the phenotypic observations towards the overall mean, which reduces the squared errors (Robinson., 1991; Crossa et al., 2006; Piepho et al., 2008).

In an animal breeding program, BLUP is a routine procedure. However, it has not yet been popular for plant breeding (Crossa et al., 2006; Oakey et al., 2006; Piepho et al., 2008). There is even less use of BLUP for selfing species, which include important annual crops such as rice, barley, soybean and wheat. One major reason for this is the lack of a method to compute a numerator relationship matrix (NRM) within a population of self-pollinators (Bauer et al., 2006; Oakey et al., 2006). Previous studies reported that the BV prediction accuracy was decayed when an NRM that ignores the selfing characteristic was used in the BLUP procedure (Atkin et al., 2009). Therefore, the development of a method for computing an NRM for self-pollinators was required.

The method to obtain an NRM in a selfing species is presented in Chapter 2. In this work, the BLUP that embeds an NRM for self-pollinators was modeled for grain yield, scald severity and net blotch severity in an arbitrarily collected data set of German spring barley cultivars, which is publicly available from Landessortenversuche (LSV). In this study, the selection index, variance components and response to selection in the provided population were dissected.

Objectives

The objectives of this study are to (1) examine BLUP using a set of phenotypic data of self-pollinating crop in unbalanced trials, (2) examine BLUP with an NRM that considers self-pollination and (3) correlate the ranking of BLUP estimates with the ranking of phenotypic observations (mean phenotype).

3.2 Materials and Methods

Data set preparation

Ninety-two spring barley (*Hordeum vulgare* L.) varieties that were originated and adapted in Germany were obtained from LSV. In this study, three phenotypes (grain yield, scald severity and net blotch severity) that had been evaluated from 1992 to 2002 in 239 locations in Germany were analyzed. The number of varieties for grain yield, scald severity and net blotch severity were 92, 90 and 88, respectively. The numbers of locations for grain yield, scald severity and net blotch were 184, 181 and 167, respectively. The phenotypic records were observed 2 to 3 times for the replications. The data set for the study was prepared by compiling a number of data sets in unbalanced trials. The grain yield was scored by kg/ha. The infection ratings for scald and net blotch were scored on a scale of 0-9 according to the degree of severity, in which 0 denotes the most positive expression and 9 the most negative expression. A list of the barley accessions is provided in Appendix Ⅰ.

Best Linear Unbiased Prediction

Under an assumption that additive effects solely constitute genetic effect, the basic linear model is as follows:

$$y = Xb + Z_g g + Z_v v + Z_{g.v} g.v + e \qquad \text{(Equation 3-1)}$$

where y = the vector of phenotype observations; b = the vector of constant grand means as fixed effect; g = the vector of breeding values as random effect; v = the vector of environment observations as random effect; g.v = the vector of genotype by environment interaction as random effect; e = the vector of residuals; X, Z_g, Z_v, $Z_{g.v}$ are the design matrices composed of 0s and 1s.

In the above model, the environment variable (v) was assumed to be a product of location-by-year interaction. The genetic effect (g), environment effect (v), environment by genotype interaction (g.v) and residual effect (e) were assumed to be random variables. Of those, the genetic effect has a correlation, whereas the other three effects have no correlation. Therefore, the variances for the aforementioned random effects can be expressed as Var (g) = $A\sigma_g^2$; Var (v) = $I\sigma_v^2$; Var (g.v) = $I\sigma_{g.v}^2$; Var (e) = $I\sigma_e^2$. Based on those expressions, the MME for the BLUP resolution can be denoted as follows:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z'_g & X'R^{-1}Z'_v & X'R^{-1}Z'_{g.v} \\ Z'_gR^{-1}X & Z'_gR^{-1}Z'_g + A^{-1}\sigma^2_g & Z'_gR^{-1}Z'_v & Z'_gR^{-1}Z'_{g.v} \\ Z'_vR^{-1}X & Z'_vR^{-1}Z'_g & Z'_vR^{-1}Z'_v + I\sigma^2_v & Z'_vR^{-1}Z'_{g.v} \\ Z'_{g.v}R^{-1}X & Z'_{g.v}R^{-1}Z'_g & Z'_{g.v}R^{-1}Z'_v & Z'_{g.v}R^{-1}Z'_{g.v} + I\sigma^2_{g.v} \end{pmatrix} \begin{pmatrix} b \\ g \\ v \\ g.v \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'_gR^{-1}y \\ Z'_vR^{-1}y \\ Z'_{g.v}R^{-1}y \end{pmatrix} \qquad \text{(Equation 3-2)}$$

where y = the vector of phenotypic observations; σ^2_g = the overall BV variance; σ^2_v = the overall variance of environment effect; $\sigma^2_{g.v}$ = the overall variance of environment by genotype interaction; A = the numerator relationship matrix; I = the identity matrix; R = the variance-covariance matrix of residual effect; X, Z_g, Z_v, $Z_{g.v}$= the design matrices in Equation 3-1

The above model was fitted to obtain a resolution for a vector, g, using the restricted maximum likelihood (REML).

Numerator relationship matrix

In BLUP, the BVs of varieties within a population were considered to be random and related, and the correlation among BVs can be expressed as an NRM. This matrix can be computed via pedigrees. The pedigree statements were provided by German barley catalogues released from LfL Pflanzenbau (http://www.lfl.bayern.de/ipz/gerste/09740/linkurl_0_9.pdf). The pedigrees for the barley collection are specified in Appendix Ⅰ. To obtain the precise NRM, parental varieties were traced back to their base populations. The NRM was computed using the PopKin software tool, which permits the consideration of the number of selfing generations that are not recorded in real plant pedigree. In this study, the number of selfing generations was assumed to be 10 across entire pedigree records.

Estimations of variance components and heritabilities

In this study, variance components were measured through fitting the BLUP model by using the restricted maximum likelihood (REML) algorithm developed by Patterson and Thomson., (1971). Using the obtained variance components, the narrow-sense heritabilities (h^2) were estimated as proposed by Hallauer and Miranda., (1981):

$$h^2 = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{gv}}{m} + \frac{\sigma^2}{rm}} \qquad \text{(Equation 3-3)}$$

where σ_g^2 = the genotypic variance; $\sigma_{g.v}^2$ = the genotype by environment interaction variance; σ^2 = the residual variance; m = the number of environments; and r = the number of replications.

In practice, the values of r and m in Equation 3-3 were not provided because the structure of data set points was highly unbalanced. As alternatives, the harmonic means obtained by averaging r and m were used.

Software utilization

The general statistical analyses were conducted using R 2.15.1 (R Core Team., 2012) and the numerator relationship matrix was constructed using PopKin software. The BLUP model was fitted using ASReml-R 3.0 library (Butler et al., 2009) on R 2.15.1 platform.

3.3 Results

3.3.1 Measuring of variance component

The variance components were measured for grain yield, scald severity and net blotch severity by fitting the BLUP model using REML (Table 3-1). For all traits, the variance components for the residual (σ_e^2) were the highest, followed by the genotype by environment interaction ($\sigma_{g.v}^2$). The variance component for the environmental effect (σ_v^2) for grain yield and scald severity was larger than that for the genotype effect (σ_g^2) with the reverse result observed for net blotch severity.

3.3.2 Heritability for the three traits

The h^2 for the three traits was measured using Equation 3-3 that requires information about the numbers of replication and environment. However, the provided data set was not produced based on the uniformly structured experimental design, which caused irregular numbers of replication and environment. To overcome this problem, the harmonic means of environment and replication were obtained, which were 55.45 and 2.13 for grain yield, 43.18 and 2.13 for scald severity and 38.72 and 2.09 for net blotch severity, respectively. The estimates of h^2 were variable across traits and resulted in 0.719 for grain yield, 0.491 for scald severity and 0.581 for net blotch severity (Table 3-1).

3.3.3 Frequency distributions of the mean phenotypes and BLUP estimates for the three traits

The frequency distributions of mean phenotypes (MPs) and BLUP estimates for the three traits are displayed in Figure 3-1. The curve over each frequency distribution shows the normal distribution based on population mean and standard deviation. To examine the normality of the frequency distributions in MPs and BLUP estimates, the Shapiro-Wilk test was performed at a significance level of 0.05 and showed that the MPs for all the traits were found to be normally distributed (p = 3.18e-07 for grain yield, p = 0.004072 for scald severity and p = 0.01093 for net blotch severity). However, the same test with the BLUP estimates revealed a normality for only grain yield (p = 6.163e-07), whereas the BLUP estimates for scald severity (p = 0.8219) and net blotch severity (p = 0.8477) were found to follow a nonnormal distribution. For grain yield, the means of both MPs and SEs (SE_{MP-Y}) were shown to be 656.258 and 157.677, respectively, and the means of the BLUP estimates and SEs (SE_{BL-Y}) were shown to be 654.5 and 35.518, respectively. The ratio of SE_{MP-Y} and SE_{BL-Y} was 4.439:1. For scald severity, the means of both MPs and SEs (SE_{MP-S}) were shown to be 2.991 and 1.457, respectively, whereas the means of both BLUP estimates and SEs (SE_{BL-S}) were 3.029 and 0.223, respectively. The ratio of SE_{MP-S} and SE_{BL-S} was 6.534:1. For net blotch severity, the means of both MPs and SEs (SE_{MP-N}) resulted in 3.022 and 0.491, respectively, whereas the means of both BLUP estimates and SEs (SE_{BL-N}) were 3.036 and 0.254, respectively. The ratio of SE_{MP-N} and SE_{BL-N} was 1.933:1. The above descriptions are summarized in Table 3-2. For all the traits, the SEs in the track of BLUP estimates was considerably lower than the SEs in the MP track.

Table 3-1. Variance components for grain yield, scald severity and net blotch severity in a panel of German spring barley varieties. The variance components were resolved by fitting the BLUP model using REML.

Variance of factor	Traits		
	Grain yield	Scald severity	Net blotch severity
σ_g^2	817.5688	0.0394	0.0504
σ_v^2	1901.3463	0.1224	0.0395
$\sigma_{g.v}^2$	8605.2946	0.7078	0.8351
σ_e^2	19469.4994	1.8300	1.5697
h^2	0.719	0.491	0.581

σ_g^2 = the variance component of genotype effect; σ_v^2 = the variance component of environment effect; $\sigma_{g.v}^2$ = the variance component of genotype by environment interaction; σ_e^2 = the variance component for residual

Table 3-2. Comparison of the mean phenotypes and BLUP estimates for grain yield, scald severity and net blotch severity in terms of mean and SE. Overall, the SEs in BLUP estimate track are considerably lower than those in mean phenotype track for all traits.

Trait	Mean phenotype		BLUP estimate		SE_{MP} vs. SE_{BL}
	$MEAN_{MP}$	SE_{MP}	$MEAN_{BL}$	SE_{BL}	
Grain yield	656.258	157.672	654.5	35.518	4.439 : 1
Scald severity	2.991	1.457	3.029	0.223	6.534 : 1
Net blotch severity	3.022	0.491	3.036	0.254	1.933: 1

$MEAN_{MP}$: mean of mean phenotypes; SE_{MP}: standard error of mean phenotypes; $MEAN_{BL}$: mean of BLUP estimates; SE_{BL}: standard error of BLUP estimates

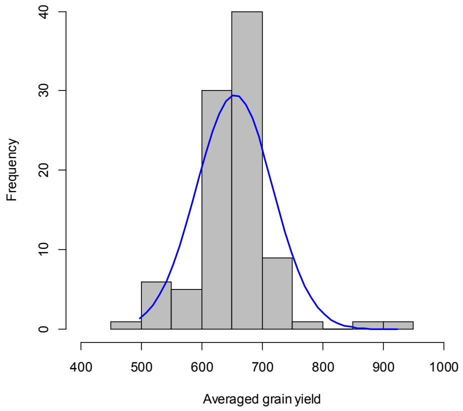
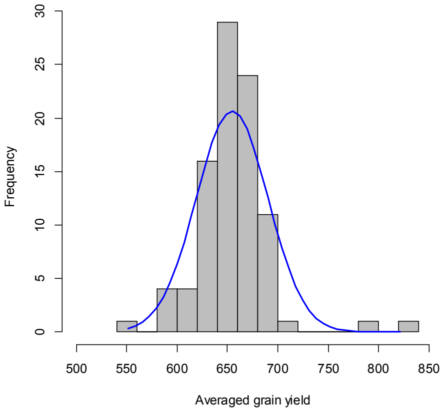
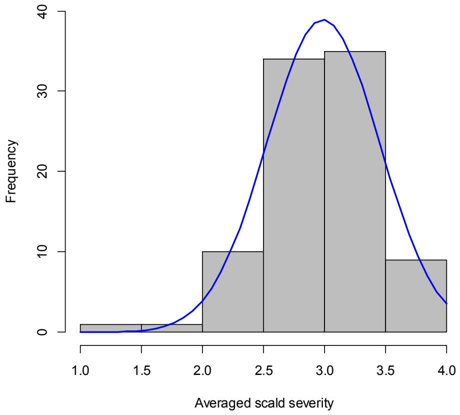
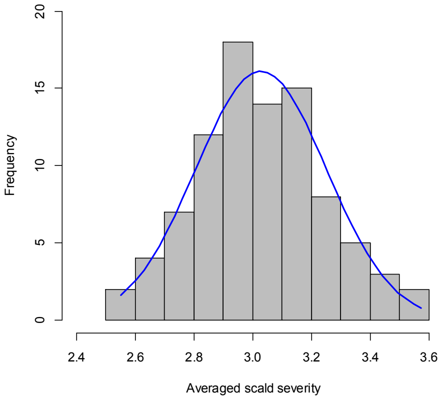
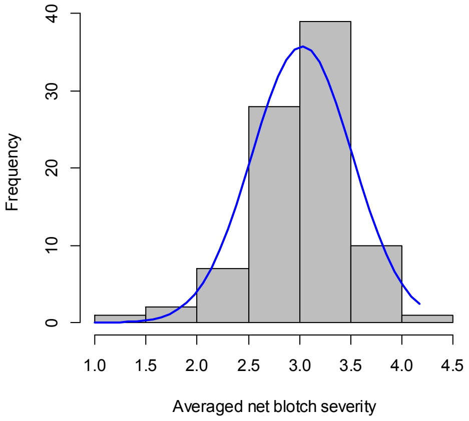
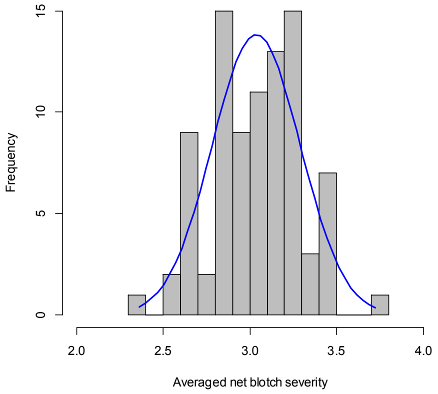
	Frequency distributions of mean phenotypes	Frequency distributions of BLUP estimates
Grain yield		
Scald severity		
Net blotch severity		

Figure 3-1. Frequency distributions of mean phenotypes (left) and BLUP estimates (right) for grain yield (top), scald severity (middle) and net blotch severity (bottom)

3.3.4 The correlation between the BLUP estimates and mean phenotypes

Figure 3-2 displays two tracks of bi-plot for the three traits: (1) bi-plot of BLUP estimates against MPs (left) and (2) bi-plot of the rankings of BLUP estimates against the rankings of MPs (right). Table 3-3 shows the statistical summary of MPs and BLUP estimates in addition to Spearman's rank correlation (rank correlation) and Pearson's correlation (simple correlation) coefficients between BLUP estimates and MPs. The simple correlation coefficients using BLUP estimates and MP estimates were shown to be 0.876 for grain yield, 0.786 for scald severity and 0.832 for net blotch severity. The rank correlation coefficients resulted in 0.854 for grain yield, 0.893 for scald severity and 0.940 for net blotch severity. The rank correlations of MPs against BLUP estimates are similar to higher compared with the simple correlation of MPs against BLUP estimates. The minimum values in BLUP estimates are greater than those in MPs across all traits, whereas the maximum values in BLUP estimates are lower than those in MPs. Therefore, the distributions of BLUP estimates are shown to be narrower than those of MPs. Figure 3-3 shows both the quantile-quantile plot of the MPs' distribution and BLUP estimates' distribution against the barley cultivars, whose results visualize that the BLUP estimates are dispersed in a narrower range compared with the MPs as shown in Table 3-3. Whole BLUP estimates and MPs for all three traits are shown in Appendix Ⅱ.

Table 3-3. Statistical summaries of mean phenotypes, BLUP estimates, Spearman's rank and Pearson's correlation coefficients between the mean phenotypes and BLUP estimates for grain yield, scald severity and net blotch severity.

Traits	Min	Median	Max	Mean	SE	Spearman's rank correlation coefficient	Pearson's correlation coefficient
MP-Y	497.0	659.1	931.9	656.3	157.7	0.854	0.876
BL-Y	551.6	654.1	821.6	654.5	35.518		
MP-S	1.000	3.011	4.000	2.991	1.457	0.893	0.786
BL-S	2.551	3.024	3.574	3.029	0.223		
MP-N	1.000	3.071	4.167	3.022	0.491	0.940	0.832
BL-N	2.363	3.047	3.723	3.036	0.254		

MP-Y: mean phenotypes for grain yield; MP-S: mean phenotypes for scald severity; MP-N: mean phenotype for net blotch severity; BL-Y: BLUP estimates for grain yield; BL-S: BLUP estimates for scald severity; BL-N: BLUP estimates for net blotch severity.

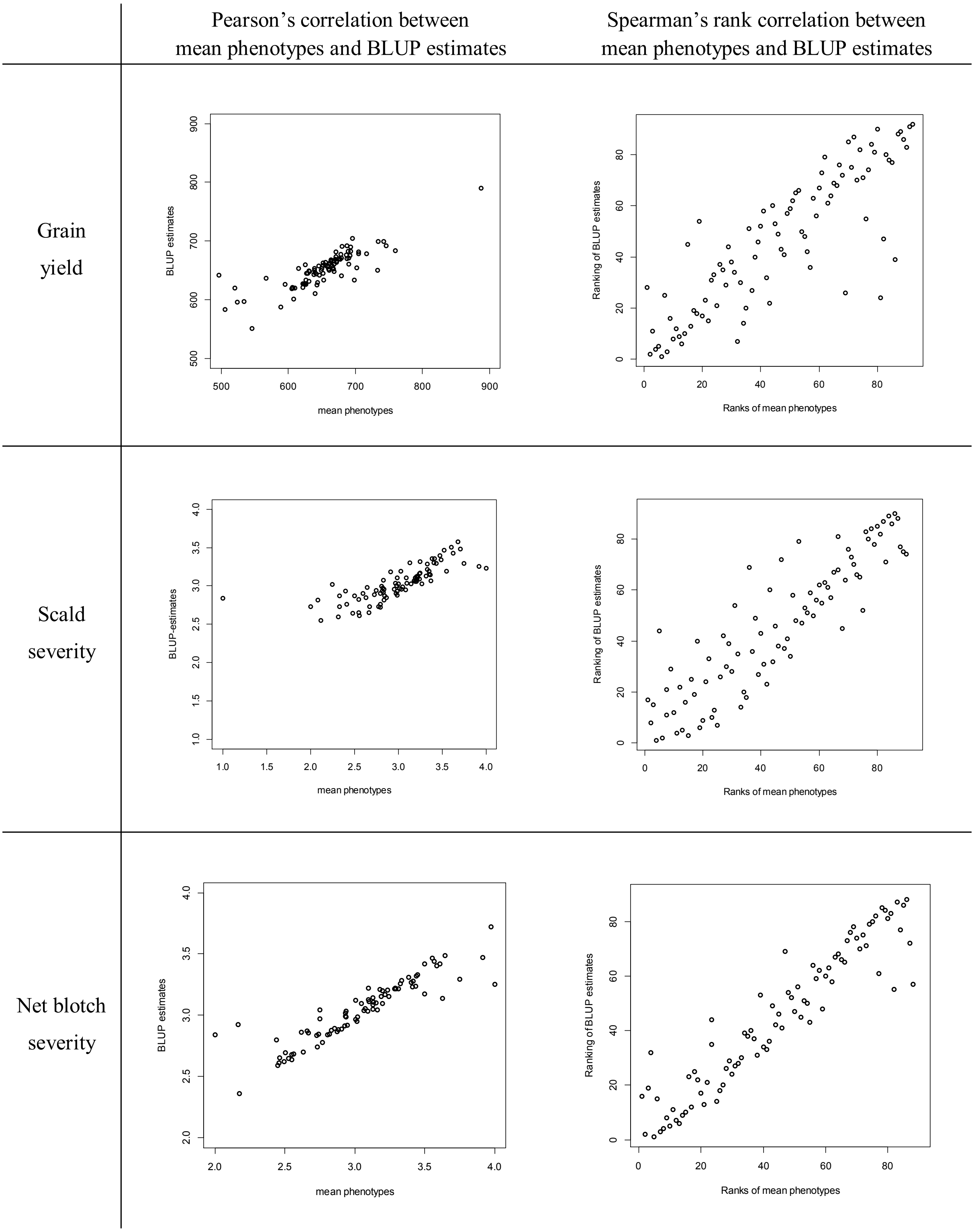


Figure 3-2. Bi-plots displaying Spearman's rank correlation (right) and Pearson's correlation (left) between the BLUP estimates and mean phenotypes for grain yield (top), scald severity (middle) and net blotch severity (bottom). Spearman's correlation coefficients resulted in 0.854, 0.893 and 0.940 for grain yield, scald severity and net blotch severity, respectively. Pearson's correlation coefficient resulted in 0.876, 0.786 and 0.832 for grain yield, scald severity and net bloch severity, respectively.

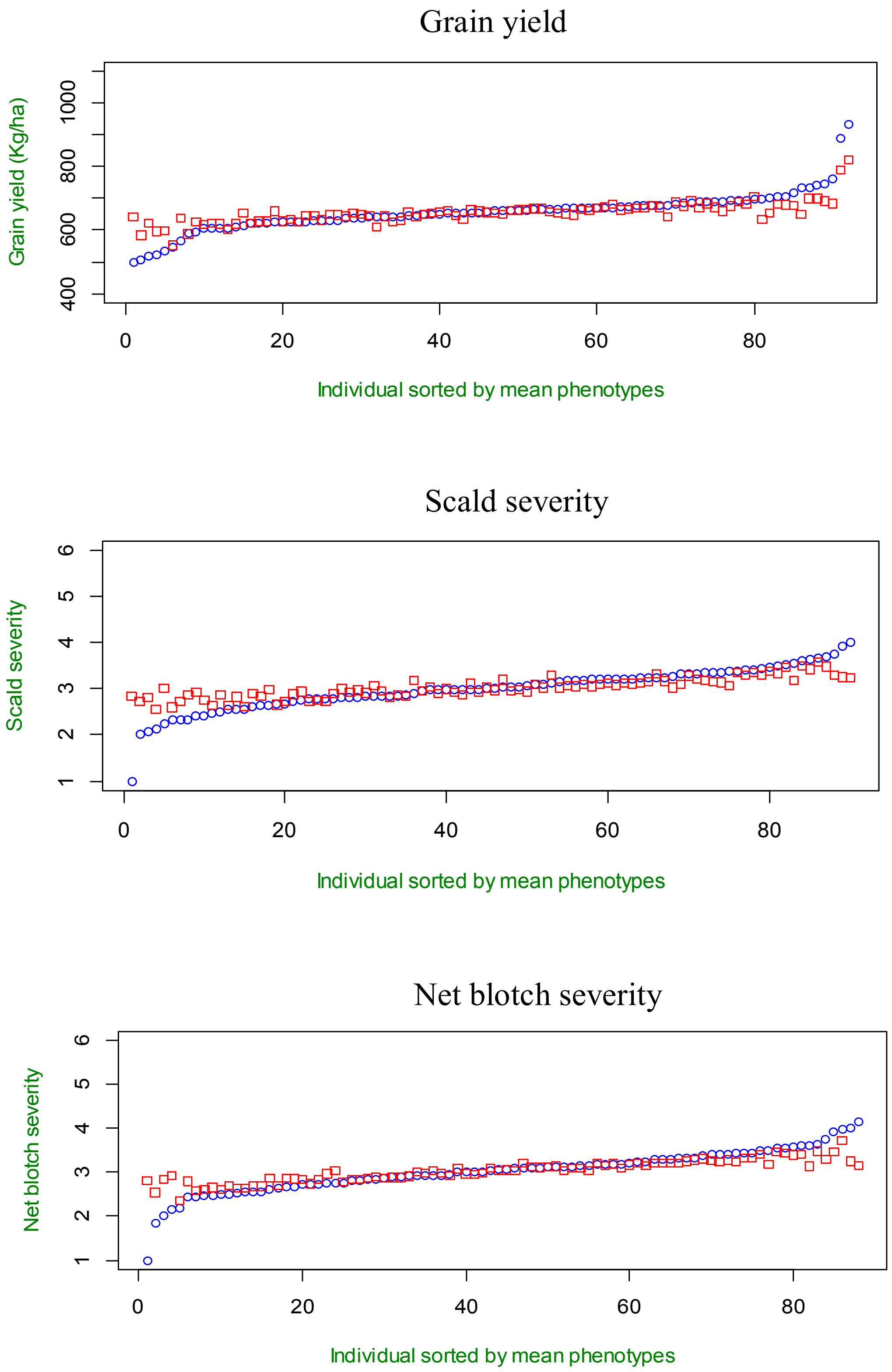


Figure 3-3. Quantile-quantile plot of the distribution of the mean phenotypes (blue circle) and BLUP estimates (red square) for grain yield (top), scald severity (middle) and net blotch severity (bottom) against the barley cultivars.

3.4 Discussion

3.4.1 Estimated variance components

Estimates of variance components are required to estimate the BVs among lines and response to selection (Durel et al., 1998; Bromley et al., 2000; Piepho et al., 2008). Precise estimation of the variance components is a crucial prerequisite for the improved performance of BLUP (Piepho et al., 2008). In this study, the estimates of variance components were obtained using the BLUP that embeds an NRM accounting for selfing reproduction (see Table 3-1). For all the traits, the estimates of residual variance component (σ_e^2) were the largest. This indicates that the non-modeled effects such as the epistatic effect, micro-environment effect and subjectivity of observers might be large (Durel et al., 1998).

3.4.2 Measurement of narrow-sense heritability

The h^2 indicates the response to selection in a breeding practice (Piepho and Moehring., 2007) and is routinely estimated using Equation 3-3 presented by Hallauer and Miranda (1981). To use Equation 3-3, a set of data that provide regular numbers of replication (r) and environment (m) are needed (Piepho and Moehring., 2007). However, because a set of data used in this study was prepared through compiling a number of data sets recorded from numerous environments, the values for m and r were non-uniform. Such a condition often occurs in plant studies (Piepho and Moehring., 2007), which limits the usefulness of Equation 3-3. To overcome this obstacle, the required numbers were approximated through averaging. Therefore, the resulted values for m and r were 55.45 and 2.13 for grain yield, 38.72 and 2.09 for scald severity and 43.18 and 2.13 for net blotch severity, respectively. Using the variance components and approximated values of m and r, the h^2 for the three traits were calculated, and the resulting h^2 ranged from 0.491 to 0.719, which shows that the phenotypic observations were performed in a sufficient level. The estimates of h^2 were similar to higher compared with the heritabilities previously assessed from five traits in an oil palm population (Soh., 1994), 11 traits in an apple breeding population (Durel et al., 1998) and 14 traits in a wheat cultivar collection (Oakey et al., 2006). The estimates of h^2 vary across different traits, which is consistent with observations reported in previous studies (Durel et al., 1998; Oakey et al., 2006).

3.4.3 Utilization of G matrix accounting for selfing in BLUP

In a breeding population, individuals are often genetically related. BLUP provides a manner of incorporating a genetic correlation among individuals using the G matrix (Crossa et al., 2006; Piepho et al., 2008; Viana et al., 2009). By definition, the G matrix represents the variance-covariance among BVs of individuals (Crossa et al., 2006; Bauer et al., 2008; Piepho et al., 2008), which is defined as G = $A\sigma_g^2$ (Henderson., 1975), where A is an NRM and σ_g^2 is the variance of BV in a population. The G matrix relates genetic performances of relatives to an individual's genetic potential, which helps improve the estimations of BVs and response to selection (Panter and Allen., 1995b; Durel et al., 1998; Bromley et al., 2000; Crossa et al., 2006; Piepho et al., 2008; Atkin et al., 2009). As a component of G matrix, the NRM that is obtainable via a pedigree, for which the subsequent recording of parent-offspring relationships is required. However, in plant pedigree, family members are often unknown, and selfing is predominant in some species, which have discouraged the application of BLUP to self-pollinators. The method and tool (PopKin) for constructing the NRM that accounts for the number of selfing generations are presented in Chapter 2. In this study, PopKin software tool was used for computing the NRM under the assumption that the number of selfing for all the cultivars was 10. The precise NRM improves the performance of the BLUP estimates by elevating the precision of G matrix that captures the genetic potential that is not observed in an individual but observed in its relatives (Soh., 1994; Panter and Allen., 1995a; Panter and Allen., 1995b; Piepho et al., 2008; Atkin et al., 2009). Therefore, the BLUP that includes the NRM derived from using PopKin may provide the accurate estimations of BV, particularly in breeding programs of self-pollinating crops.

3.4.4 Shrinkage feature of BLUP

In all analyses for grain yield, scald severity and blotch severity, the comparison of BLUP estimates and MPs shows a considerable reduction of SE in the BLUP estimates compared with the MPs. However, a nearly ignorable shift of mean was observed for all traits (see Table 3-3). The reduction of SEs through BLUP is a typical outcome that arises from BLUP's shrinkage feature regressing the phenotypic observations to a grand mean (Panter and Allen., 1995a; Panter and Allen., 1995b; Crossa et al., 2006; Bauer et al., 2008; Piepho et al., 2008). The rates of SE_{MP} vs. SE_{BL} range between 1.933:1 and 6.534:1 (Table 3-2). The shrinkage of BLUP maximizes the rank correlation between true breeding values (TBV) and BLUP estimates (Searle et al., 1992; Bauer et al., 2008; Piepho et al., 2008).

3.4.5 BLUP provides a ranking index

Distributions of MPs for all three traits are normally distributed according to the Shapiro-Wilk test (Figure 3-1) and is consistent with the typical distribution of quantitative traits, which indicates that all the traits are controlled by polygenes (Durel et al., 1998). The normality of a trait distribution determines the robustness of departure from the REML (Piepho et al., 2008). Therefore, the normality of the traits in the provided set of data might be positively effective in fitting a BLUP model. The resolutions of the BLUP can provide not the future phenotypic performance but ranking index of predicted BVs for individuals (Robinson., 1991).

3.4.6 Selection using BLUP may outperform over selection using phenotypic observation

Typically, annual crops that have a self-pollinating reproductive system maintain the high chromosomal homogeneity within a variety. Therefore, the same varieties are regarded as clones, which enables variety tests to be replicated across locations over years and leads to the relatively precise estimation of BVs by calculating MPs (Piepho et al., 2008; Zhong et al., 2009). Regarding this property, Piepho et al., (2008) described that the BV derived from BLUP does not provide grossly different results from the MPs. In this study, however, the rank correlation coefficients between BLUP estimates and MPs for all the traits ranged from 0.854 to 0.940, which indicates that the BLUP estimates and MPs are variably correlated across the different traits. Non-perfect coincidence between BLUP estimates and MPs illustrates that the selection based on the BLUP may have a potential to elevate the response to selection over the selection based on MPs.

4. Association mapping for three traits of a German spring barley collection

4.1 Introduction

The success of a breeding program relies on the harmonious utilization of polygenes because useful agronomical traits are generally controlled by multiple genes. In crop breeding, the detection of useful genes is challenging (Pasam et al., 2012). In principle, genes in association with a trait can be detected based on the correlation between genotypic pattern and phenotypic variation because high marker-trait correlation indicates that a marker is located within a short range of LD with the gene controlling a trait. Therefore, genes can be mapped by scooping out the mapped markers that show a high correlation with a phenotypic variation (Pasam et al., 2012; Wang et al., 2012a). In this study, this kind of markers is termed quantitative trait loci (QTL). For successful QTL mapping, dense and even genome coverage with a large set of markers and a panel of diverse germplasms are beneficial because these conditions enrich the number of short range LD and increase the allele frequencies at a locus (Haseneyer et al., 2009; Rafalski., 2010; Wang et al., 2012a). Such conditions benefit QTL mapping by facilitating the detection of markers flanking a gene. In breeding, the QTL map is highly informative because it allows the introgression of multiple QTL into a gene pool of elite lineage through backcrossing or gene cloning (Grewal et al., 2008; Rafalski., 2010; Wang et al., 2012a).

There are two available approaches for QTL mapping in crops: (1) bi-parental mapping and (2) association mapping (AM). The bi-parental mapping segregates LD blocks by bi-parental crosses over multiple generations. Accordingly, the bi-parental mapping provides the segregation observations of LD blocks from only bi-parental crosses, which limits the diversity of the alleles (Zhang et al., 2009; Pasam et al., 2012; Wang et al., 2012a). This approach is time-consuming and cost-intensive because of sequent generation advancements (Massman et al., 2011; Pasam et al., 2012; Rode et al., 2012; Wang et al., 2012a). However, the AM excavates the LD blocks including abundant alleles from an existing panel of diverse germplasms, which avoids the aforementioned weaknesses of the bi-parental mapping (Kraakman et al., 2004; Yu and Buckler., 2006; Massman et al., 2011; Pasam et al., 2012; Wang et al., 2012a). However, the diverse genetic backgrounds among individuals can cause genetic stratification in a population because individual germplasms were adapted to independent environments and might have undergone non-random mating and subsequent selection (Kraakman et al., 2004; Haseneyer et al., 2009; Pasam et al., 2012). Such genetic stratification could inflate the detection of spurious marker-trait associations by confounding the subpopulation effect with the marker-trait association. To overcome this effect, Yu et al., (2005) proposed a unified mixed linear model for a robust AM that fits a mixed linear model that embeds a kinship matrix and a subpopulation structure matrix for the purpose of filtering the noise effects arising from a subpopulation structure. Previous studies demonstrated that the underlying model showed a similar to greater performance relative to the bi-parental mapping in crop studies (Stich et al., 2008; Rafalski., 2010).

In this study, the QTL for grain yield, scald severity and net blotch severity were mapped using the unified mixed linear model with a panel of small-sized German spring barley germplasms.

Objective

The objective of this study is to (1) detect QTL for grain yield, scald severity and net blotch severity and (2) investigate the effects of two different subpopulation structures on the AM.

4.2 Materials and Methods

Germplasm and field evaluation

All phenotypic performances for grain yield, scald severity and net blotch severity of German spring barley lines were estimated by German Landessortenvesuche (LSV) from 1992 to 2002. The original phenotypic data contained frequent duplicates of varieties over multiple years across multiple locations. However, all the duplicates were averaged so that a single mean phenotype measurement (MP) was obtained per each variety. The numbers of barley cultivars used in this study were 45 for grain yield, 41 for scald severity and 40 for net blotch severity. All barley cultivars were morphologically two-rowed and spring sown and were geographically released and tested in Germany. The grain yield was scored with kg/ha and the infection ratings for the scald and net blotch were scored on a scale of 0-9 based on the degree of severity, in which 0 denotes the most positive expression and 9 the most negative expression. A list of the barley samples is provided in Appendix Ⅲ.

DArT genotyping

DNA was extracted from the leaf tissue of a single barley sample following the protocol recommended by Triticarte Pty (http://www.triticarte.com.au). The plant samples were genotyped by 1181 Diversity Arrays Technology (DArT) markers covering the whole barley genome: 710 markers were mapped, and 471 were unmapped. The barley germplasms were scored in a binary format: 0 for absent, 1 for present and NA for unknown.

Subpopulation analyses

The characteristics of a subpopulation were analyzed by the discriminant analysis of principal components (DAPC) and Bayesian clustering analysis using the STRUCTURE software (Pritchard et al., 2000). DAPC was carried out with 45 varieties and 44 eigenvectors were obtained. In general, the size of an entire eigenvector is too large to fit a model for mapping QTL, so the use of selective eigenvectors to sufficiently represent a population structure is necessary. To select the eigenvectors, the estimated eigenvalues were referenced. The first two largest eigenvectors were selected and used as X- and Y-coordinates for drawing a bi-plot. The DAPC was performed and visualized using a free package, dapc (Jombart et al., 2010) on the R software environment for statistical computing and graphics

(R Core Team., 2012). The second subpopulation analysis, the Bayesian clustering analysis, was carried out with the DArT marker data genotyped with 242 accessions (Appendix Ⅲ) using STRUCTURE software tool. The set of data included a much greater number of accessions than the number of entries in the phenotype data set. This was performed for the purpose of precisely assigning the used entries into appropriate clusters. The STRUCTURE analysis was run with a burning of 5,000 cycles followed by 100,000 repetitions of the Markov Chain Monte Carlo (MCMC). The STRUCTURE analysis requires the number of inferred clusters (K). To determine the best K value, the STRUCTURE analyses were performed with K = 2 to 10.

Association mapping model

For mapping the grain yield, scald severity and net blotch severity, a single marker regression model was fitted using the BLUP. Because a subpopulation effect often causes a spurious marker-trait association when mapping QTL, the correction of the subpopulation effects is required. Therefore, subpopulation structure matrix was embedded with a kinship matrix in the model. In this study, the kinship matrix was constructed using the TASSEL software tool (Bradbury et al., 2007). The mixed linear model (MLM) equation was denoted as follows:

$$y = Xb + Zu + Qv + Pm + e$$

where y = the vector of phenotypic observations; b = the vector of grand means; u = the vector of random genotype effects; v = the vector of fixed subpopulation effects; m = the vector of fixed marker effects; e = the vector of random residual effects; X, Z, Q, P = the design matrices.

The subpopulation effect (v) and marker effect (m) were assumed to be a fixed effect, and the genetic effect (u) and residual effect (e) were assumed to be a random effect. In this study, the genetic effect has a variance-covariance structure that assumes a distribution of Var(u) ~ N(0, $2K\sigma_g^2$), where K is a kinship matrix and σ_g^2 is the genetic variance. In the above model, two types of subpopulation structure matrices were examined: DAPC and STRUCTURE analysis. Accordingly, two types of models were fitted per each trait: (1) kinship plus DAPC (KD model) and (2) kinship plus STRUCTURE analysis (KS model).

Every fitting of a model was followed by the Wald test to verify if the marker-trait association is statistically significant. The threshold ($-log_{10}p$) for identifying QTL was set at 3, which is equivalent to p =0.001. The AM models were fitted using the ASReml-R software tool (Butler et al., 2009).

Cross-validation

For detecting the stringent marker-trait association, a cross-validation was performed for a set of the trait-associated markers with $p < 0.001$. Eighty percent of barley accessions were randomly selected, upon which the marker-trait association analysis was carried out 100 times. For a set of 100 p-values via the cross-validation, if a median value was significant ($p < 0.001$), the marker was determined to be in association with a trait. The cross-validation analyses were conducted using the ASReml-R software tool (Butler et al., 2009).

Estimation of allelic effects

The marker effects on a trait were estimated per a different bi-allelic score (0/1) via resolving a regression coefficient on the marker variable. If the estimated effect is positive non-zero for an observed DArT score (0 or 1), the provided score denotes a positive effect on the trait. Likewise, the negative non-zero represents a negative effect on a trait. The allelic effects were measured using the ASReml-R software tool (Butler et al., 2009).

Analysis of linkage disequilibrium

To observe the characteristics of LD within a provided population, the squared correlation coefficient r^2 (Pritchard and Przeworski., 2001) was calculated for all pairs of markers. The visualizations of the LD plot and local weighted scatterplot smoothing (LOESS) curve were fitted using a software tool, GenStat (Payne et al., 2006).

Mapping and Manhattan plotting of the trait-associated markers

The QTL were mapped using the program, MapChart (Voorrips., 2002). A Manhattan plot based on the resulting p-values was constructed using a free software package called gap (Zhao et al., 2013) for the R software environment for statistical computing and graphics (R Core Team., 2012).

4.3 Results

4.3.1 Characterization of subpopulation structure using the DAPC

For the purpose of characterizing the subpopulation structure, DAPC was performed using 45 barley varieties and resolved 44 dimensions of eigenvectors (data not shown). To form a subpopulation matrix to be used for mapping, selection of eigenvectors were required, so the eigenvalues were referenced. Figure 4-1 exhibits the bar graph of resulting eigenvalues in descending order. The largest three Eigenvalues comprised over 60 % of the levels on a vertical scale and resulted in 95.85 %, 67.57 % and 64.00 %. Therefore, the corresponding eigenvectors were selected to form a subpopulation matrix for the AM. The largest two eigenvectors were selected for bi-plotting, where each vector represented X- and Y-coordinates. The resulting plot is shown in Figure 4-2. Except two outliers (Baronesse and Nevada), the barley cultivars formed two respective clusters on the left hand and right hand sides of y axis. The position of each variety on the plot is specified in Table 4-1.

4.3.2 Characterization of subpopulation structure using the Bayesian clustering

The Bayesian clustering analysis was used as the second trial for characterizing a subpopulation and was conducted using the software tool, STRUCTURE (Pritchard et al., 2000). In this study, the determination of the appropriate sub-population number (K) is vital but difficult to attain (Jombart et al., 2010). To verify the appropriate K value, two approaches were attempted. In the first approach, the clustered result derived from the DAPC analysis was referenced as prior knowledge. Since the DAPC resolved two major clusters, K= 2 was examined. The resulting graphical barchart is shown in Figure 4-3. The clusters resulting from the STRUCTURE analysis were compared with those derived from the bi-plots resolved using the DAPC (see Table 4-1). Except for two outliers (Baronesse and Nevada) found in the DAPC analysis, 34 entries out of 43 were found to be allocated in the same cluster, which accounted for 79.07 % agreement. This result supports the value K = 2 being a reasonable parameter for the STRUCTURE analysis. In the second approach, the selection of K that provides the highest likelihood value was attempted. In this study, the STRUCTURE software tool was run with K = 2 to 10 and the resulting likelihood values were distributed as shown in Figure 4-4. This result shows that the likelihood values increase monotonously as the value of K increases and failed to provide any outstanding number for K because the K value that provided the greatest likelihood was thought to be too large (K = 10) under the factors of the geological origins being confined to Germany, samples being morphologically monotonous and the sample size being small. The likelihood values could become larger when K > 10. Therefore, according to the result from the first approach, K = 2 was confirmed and the second approach was ignored.

In this Chapter, 45, 41 and 40 accessions were used for the AM for grain yield, scald severity and net blotch severity, respectively. However, for the purpose of providing precise clustering of the accessions, the Bayesian clustering analysis was performed with a panel of 242 barley varieties that comprise 45 accessions.

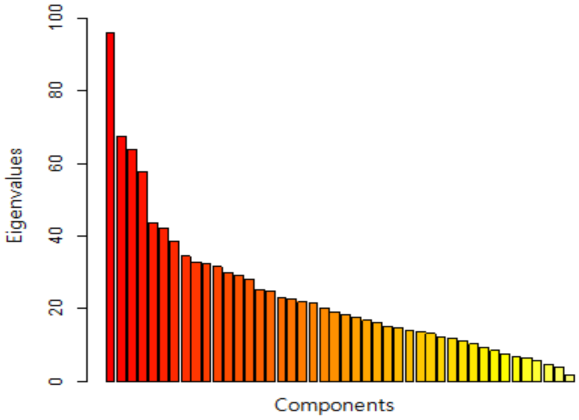


Figure 4-1. Bar graph representing in descending order of the eigenvalues obtained using the DAPC. The first three bars show over 60 % levels on a vertical scale, and the resulting values are 95.85 %, 67.57 %, and 64.5 %. Accordingly, the largest three vectors were taken to form a subpopulation matrix to use for an association mapping.

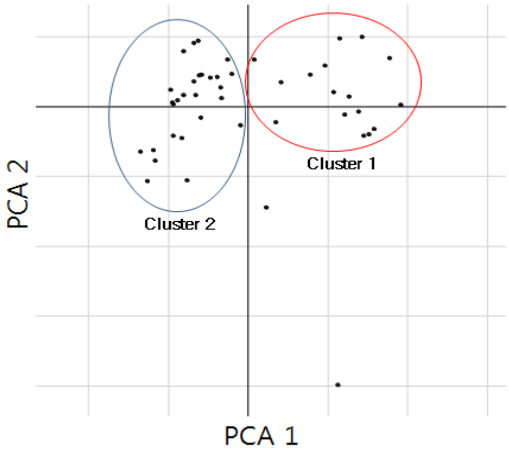


Figure 4-2. Bi-plot scattered using the largest two eigenvectors from the DAPC. Except two outliers (Baronesse and Nevada), the remaining individuals formed two distinct groups, whose borders are drawn by blue and red colored circles.

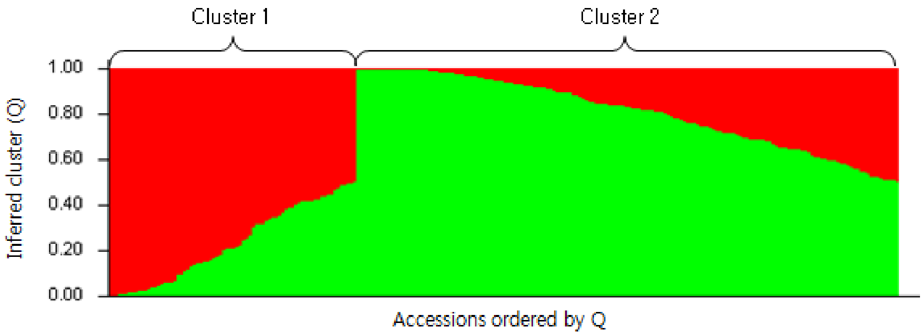
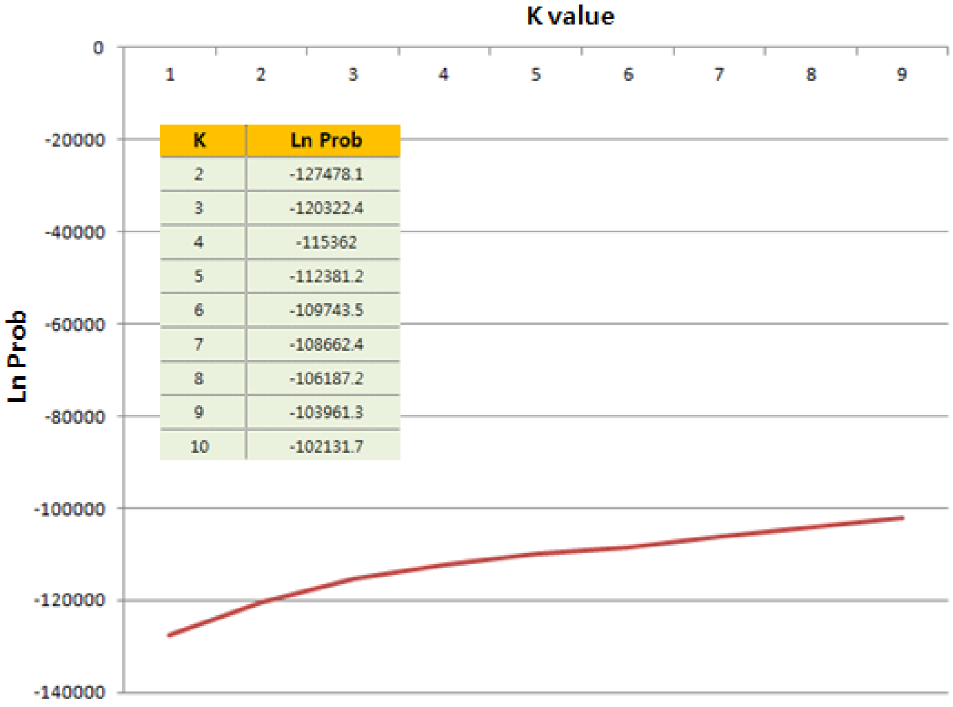


Figure 4-3. Bar chart resulting from Bayesian clustering analysis using the STRUCTURE software tool. The inferred number of clusters (K) was selected through a comparison of clusters obtained using DAPC and STRUCTURE with K = 2. The comparison showed a 79.07 % agreement.



K	Ln Prob
2	-127478.1
3	-120322.4
4	-115362
5	-112381.2
6	-109743.5
7	-108662.4
8	-106187.2
9	-103961.3
10	-102131.7

Figure 4-4. The distribution of log likelihood probability (Ln Prob) resulting from the STRUCTURE analysis with K =2 to 10. As the value of K increases, Ln Prob steadily rises. When K > 10, Ln Prob could further increase, therefore, the outstanding K value was not determined.

Table 4-1. Comparison of two clusters derived from the DAPC and STRUCTURE analysis. As a result of DAPC, two clusters were formed, excluding two outliers (Baronesse and Nevada). As a result of comparison of both methods, 34 out of a total 43 accessions were found to be allocated in the same cluster, which accounted for an agreement of 79.02 %.

Accession	DAPC (1st and 2nd Coordinates)		STRUCTURE (Inferred ancesty)		Resulting group via DAPC	Resulting group via STRUCTURE
	Coordinate 1	Coordinate 2	Cluster 1	Cluster 2		
ADONIS	-1.027	-2.581	0.563	0.437	2	1
ALEXIS	-6.325	9.476	0.217	0.783	2	2
ALONDRA	-3.419	1.301	0.253	0.747	2	2
ANNABELL	-7.758	-10.481	0.102	0.898	2	2
APEX	15.098	-3.898	0.556	0.444	1	2
AURIGA	-4.770	4.187	0.232	0.768	2	2
BARKE	-6.854	9.218	0.041	0.595	2	2
BARONESSE	-6.854	9.218	0.492	0.508	-	2
BELLA	2.252	-14.411	0.584	0.416	1	1
BESSI	14.446	-4.086	0.432	0.568	1	2
BITRANA	9.580	5.886	0.152	0.848	2	2
BRENDA	-11.980	-6.171	0.002	0.998	2	2
CAMINANT	-9.749	2.443	0.227	0.773	2	2
CELLAR	0.676	6.844	0.462	0.538	1	2
CITY	11.435	9.802	0.734	0.266	1	1
CORA	7.776	4.628	0.574	0.426	1	1
DERKADO	-9.465	0.427	0.033	0.967	2	2
DIAMALTA	-5.911	4.596	0.284	0.714	2	2
DITTA	13.799	-0.625	0.508	0.492	1	1
ESCADA	-8.101	8.037	0.173	0.827	2	2
EUNOVA	12.624	1.533	0.780	0.220	1	1
EXTRACT	3.407	-2.135	0.497	0.503	1	1
GOLF	14.240	10.080	0.939	0.061	1	1
HANKA	-13.589	-6.426	0.005	0.995	2	2
KATHARINA	-6.645	1.767	0.074	0.926	2	2
KORINNA	-9.558	0.596	0.018	0.982	1	2
KRONA	-11.668	-7.734	0.003	0.997	2	2
LARISSA	-6.256	4.547	0.251	0.749	2	2
LENKA	-3.981	4.357	0.163	0.837	2	2
MARESI	-6.897	3.630	0.081	0.919	2	2
MARINA	-8.946	0.942	0.027	0.973	2	2
MARNIE	-5.940	-1.554	0.067	0.933	2	2
MELTAN	-2.585	6.819	0.275	0.725	2	2
NANCY	17.695	7.019	0.657	0.343	1	1
NEVADA	11.245	-39.825	0.752	0.248	-	1
OLGA	12.060	-1.103	0.634	0.366	1	1
PASADENA	-8.408	-4.415	0.007	0.993	2	2
POMPADUR	15.720	-3.189	0.794	0.206	1	1
RIA	-9.402	-4.109	0.063	0.937	2	2
SCARLETT	-3.550	2.746	0.056	0.944	2	2
SISSY	-2.114	4.790	0.081	0.919	2	2
STEFFI	10.613	2.154	0.239	0.761	1	2
TEO	19.121	0.274	0.977	0.023	1	2
THURINGIA	4.014	3.504	0.309	0.691	1	2
URSA	-12.731	-10.567	0.006	0.994	2	2

4.3.3 Linkage disequilibrium

There were 710 markers used for investigating LD characteristics that were spanned approximately 1116.7 cM across seven chromosomes. The average distance between marker loci was 1.57 cM. The extents of LD were quantified by measuring the squared correlation (r^2) between paired marker intensities, which were plotted against the genetic distance (Figure 4-5). A reference value for r^2 of 0.2 on the LOESS curve indicates that LD is decayed at approximately 2.5 cM.

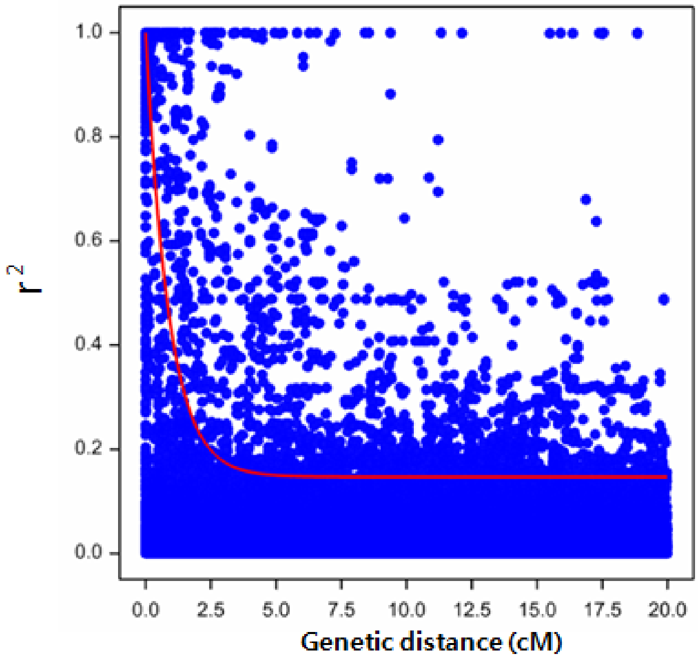


Figure 4-5. Linkage disequilibrium plot constructed with 45 German spring barley cultivars using 710 DArT markers and showing the genetic distance (cM) between markers. The markers were spanned across approximately 1116.7 cM on 7 chromosomes of barley. Each plot shows all the pair-wise comparisons for the 710 DArT markers. The LOESS curve (red line) indicates that LD is decayed at approximately 2.5 cM.

4.3.4 QTL detection for the three traits using the two models

For mapping the provided traits, an MLM-based single marker regression was performed and followed by two stepwise tests to determine a solid marker-trait association. In the first test, markers with $p < 0.001$ were selected through the Wald test after fitting the regression model. In the second test, a cross-validation was conducted with the markers selected in the first step. Table 4-2 shows an overview about trait-associated markers. In this study, the trait-associated markers were sub-divided depending on whether the marker position is known. Hereafter, a trait-associated marker that has a known map location will be termed quantitative trait loci (QTL). The unmapped trait-associated markers will be termed UTAM. And both UTAM and QTL will be comprehensively termed trait-associated markers (TAM).

The detected TAMs were shown to vary depending on the trait and the subpopulation matrix (DAPC and STRUCTURE) that was embedded in the BLUP model. The model with the DAPC subpopulation matrix was termed the KD model. The model with the STRUCTURE subpopulation matrix was termed the KS model. For grain yield, a single QTL (bPb-8962) was detected across the KD and KS models. For scald severity, one QTL (bPb-8445) was found from the KD model, whereas three QTL (bPb-8445, bPb-6264 and bPb-5458) and one UTAM (bPb-2018) were detected from the KS model. For net blotch severity, only a single QTL (bPb-1946) was detected across the KD and KS models. In addition, TAM effect per each bi-allele was estimated. The list of both the detected TAMs and their bi-allelic effect on TAMs is provided in Table 4-2, and the positive or negative effects on each TAM in relation with a trait at each variety are provided in Table 4-3.

4.3.5 Simplified allelic effects

Allelic effects were tabulated using "+" for positive effect and "–" for negative effect depending on the regression coefficient for marker and genotyped digits for each barley cultivar (Table 4-3). The allelic effects, expressed as "+" and "–", mostly were in accordance with the rising and falling pattern of mean phenotypes. This result is natural because QTL were mapped based on the pattern between phenotype variation and marker genotypes. However, the allelic effects for some cultivars did not represent the variation of mean phenotypes, such as the BARONESSE cultivar in a table of grain yield and the ADONIS cultivar in a table of net blotch severity.

Table 4-2. Trait-associated markers for grain yield, scald severity and net blotch severity resulting from both a kinship plus DAPC (KD) model and a kinship plus STRUCTURE (KS) model.

DArT	Chr	Pos	KD model									KS model								
			Grain yield			Scald severity			Net blotch severity			Grain yield			Scald severity			Net blotch severity		
			p-value	Marker (0)	Marker (1)	p-value	Marker (0)	Marker (1)	p-value	Marker (0)	Marker (1)	p-value	Marker (0)	Marker (1)	p-value	Marker (0)	Marker (1)	p-value	Marker (0)	Marker (1)
bPb-8962	3H	178.59789	1.7E-05	47.18	-47.18	-	-	-	-	-	-	9.7E-03	45.53	-45.53	-	-	-	-	-	-
bPb-8445	2H	5.02763	-	-	-	0.00013	0.22	-0.22	-	-	-	-	-	-	8.2E-05	0.21	-0.21	-	-	-
bPb-6264	6H	98.70832	-	-	-	-	-	-	-	-	-	-	-	-	0.00073	0.23	-0.23	-	-	-
bPb-5458	7H	82.60586	-	-	-	-	-	-	-	-	-	-	-	-	1.1E-05	-0.14	0.14	-	-	-
bPb-2018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8.9E-05	-0.11	0.11	-	-	-
bPb-1946	7H	82.60586	-	-	-	-	-	-	3.7E-06	-0.23	0.23	-	-	-	-	-	-	1.2E-05	-0.24	0.24

For detecting the stringent marker-trait association, two stepwise tests were implemented. The first step was the identification of the trait-associated markers with $p < 0.001$ through the Wald test after fitting the single-marker model (KD or KS model). The second step was to filter the markers that did not meet $p < 0.001$ at a median value out of 100 p-values obtained through 100 repetitions of cross-validation with a random selection of 80 % of the members of a population. Within each trait track, the second and the third columns are the regression coefficients that represent the estimates of QTL effects on a trait. A positively increasing value of an observed genotype (0 or 1) denotes the positively increasing magnitude of the effect on a trait. Negatively increasing values denote the increasingly negative effects on a trait.

Table 4-3. Distribution of positive and negative effects of trait-associated markers in mapping populations for grain yield (up), scald severity (middle) and net blotch severity (bottom). “+” and “-” represent the positive and negative effects of an observed bi-allele upon a trait. X represents a missing genotypic value. The estimates of allelic effect for each marker are provided in Table 4-2. For a comparison of the marker’s effect and phenotypic performance, the mean phenotypes are specified at the bottom of each table.

Grain yield	DArT	Chromosome	ADONIS	ALEXIS	ALONDRA	ANNABELL	APEX	AURIGA	BARKE	BARONESSE	BELLA	BESSI	BITRANA	BRENDA	CAMINANT	CELLAR	CITY	CORA	DERKADO	DIAMALTA	DITTA	ESCADA	EUNOVA	EXTRACT	GOLF	HANKA	KATHARINA	KORINNA	KRONA	LARISSA	LENKA	MARESI	MARINA	MARNIE	MELTAN	NANCY	NEVADA	OLGA	PASADENA	POMPADUR	RIA	SCARLETT	SISSY	STEFFI	TEO	THRUINGIA	URSA
	bPb-8962	3H	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
	Mean phenotypes	-	666.1	629.3	605.2	687.5	609.7	692.5	674.8	661.2	698.9	647.7	666.1	685.3	671.8	664.2	657.7	510.9	726.5	625.1	672.2	701.5	627.1	638.8	679.5	665.4	640.7	583	659.0	506.8	623.5	647.0	661.5	717.1	668.4	530.3	640	617.6	671.4	520.4	655.3	645.3	545.7	607.7	626.0	672.4	689.0

Scald severity	DArT	Chromosome	ADONIS	ALEXIS	ALONDRA	ANNABELL	APEX	AURIGA	BARKE	BARONESSE	BELLA	BESSI	BITRANA	BRENDA	CAMINANT	CELLAR	CITY	CORA	DERKADO	DIAMALTA	DITTA	ESCADA	EUNOVA	EXTRACT	GOLF	HANKA	KRONA	LENKA	MARESI	MARINA	MELTAN	NANCY	NEVADA	OLGA	PASADENA	POMPADUR	RIA	SCARLETT	SISSY	STEFFI	TEO	THRUINGIA	URSA
	bPb-8445	2H	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	-	+	-	+	+	+	+	+	+	+	+	-	-	+	+	-	+	+	+	-	+	+	-
	bPb-6264	6H	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	-	+	-	+	+	+	+	+	+	X
	bPb-5458	7H	-	+	-	+	-	+	+	-	-	-	+	+	+	+	-	X	+	-	-	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	+	-	-	+	+	+
	bPb-2018	-	+	-	-	+	+	+	+	+	-	+	+	+	+	+	-	-	+	-	+	+	-	+	-	+	+	-	-	+	-	-	-	-	+	-	+	+	-	+	-	+	+
	Mean phenotypes	-	2.833	3.446	3.194	3.606	3.207	3.25	2.852	3.049	2.5	3.273	3.627	3.208	2.778	3.206	3.363	2.6	3.404	2.548	2.837	2.887	2.796	3.209	2	3.358	3.254	3.25	3.330	3.692	2.836	2.083	2.333	2.727	3.411	2.4	2.774	3.147	2.56	2.957	2.828	2.966	2.912

Net blotch severity	DArT	Chromosome	ADONIS	ALEXIS	ALONDRA	ANNABELL	APEX	AURIGA	BARKE	BARONESSE	BELLA	BESSI	BITRANA	BRENDA	CAMINANT	CELLAR	CITY	CORA	DERKADO	DIAMALTA	DITTA	ESCADA	EUNOVA	EXTRACT	GOLF	HANKA	KRONA	LENKA	MARESI	MARINA	MELTAN	NANCY	OLGA	PASADENA	POMPADUR	RIA	SCARLETT	SISSY	STEFFI	TEO	THRUINGIA	URSA
	bPb-1946	7H	-	+	-	+	-	+	+	-	+	-	+	+	+	+	-	X	+	-	-	+	-	+	X	+	+	+	+	+	-	-	-	+	-	+	+	-	-	+	+	-
	Mean phenotypes	-	3.5	2.965	2.494	3.236	2.564	3.216	3.210	2.729	3.241	2.517	3.566	3.184	3.609	3.313	3.080	2.833	2.936	2.460	2.175	3.362	3.131	3.554	2	2.952	3.107	2.938	3.060	3.181	2.740	1.9	2.463	3.104	2.389	3.151	2.892	2.832	2.548	2.853	2.858	2.657

4.3.6 Chromosomal positions of QTL for the three traits

As a result of mapping the three traits, five QTL were found across four chromosomes: bPb-8962 for grain yield on 3H, bPb-8445 for scald severity on 2H, bPb-6264 for scald severity on 6H, bPb-5458 for scald severity on 7H and bPb-1946 for net blotch severity on 7H. The physical positions of QTL are graphically displayed in Figure 4-6. At 82.6 cM on chromosome 7H, two QTL (bPb-5458, bPb-1946) were shown to co-locate. However, it was found that each marker was linked to different traits: bPb-5458 is linked to scald severity and bPb-1946 to net blotch severity.

The estimates of $-log_{10}p$ obtained from the KD and KS models for the three traits were scattered against seven chromosomes of barley using the Manhattan plot method (Figure 4-7). To determine the QTL, a threshold value of 3 (= $-log_{10}0.001$) was set. In the first step, plots above a threshold line were filtered as putative QTL. In the second step, the markers that passed the criterion of the first step were filtered using a cross-validation. Finally, the remaining markers that passed above two tests were determined to be QTL, which are highlighted with a red-dotted circle in Figure 4-7.

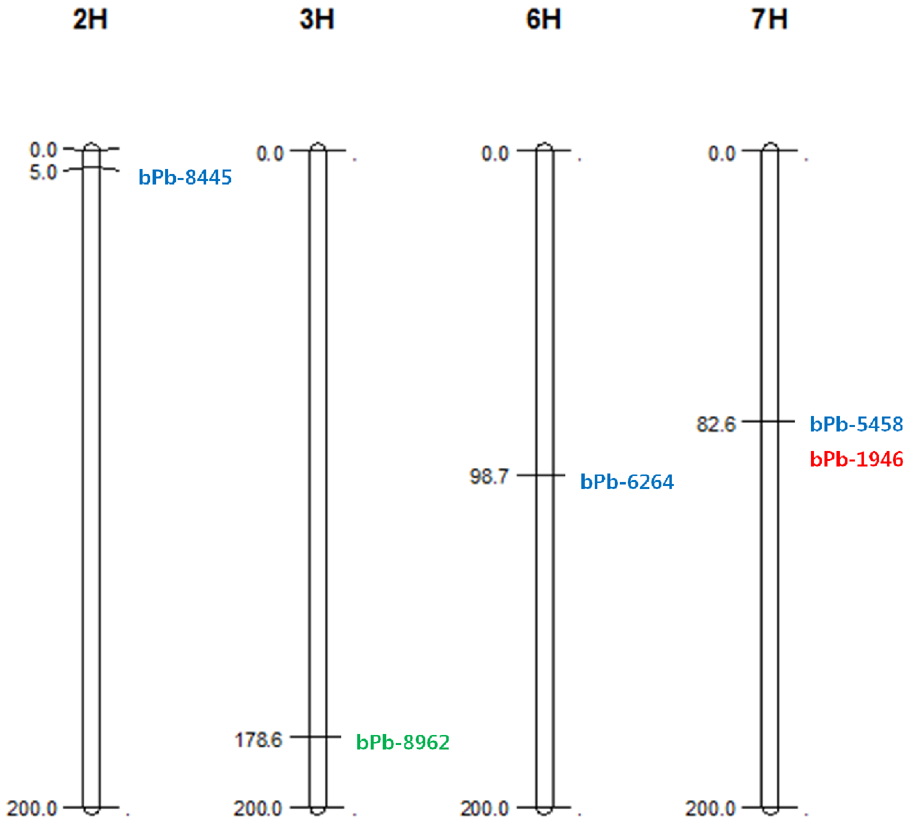


Figure 4-6. Location of quantitative trait loci (QTL) for grain yield (green), scald severity (blue) and net blotch severity (red) detected using 710 DArT markers. Note regarding the detected QTL, (1) 2H: bPb-8445 is the putative QTL flanking the gene for scald severity; (2) 3H: bPb-8962 is the putative QTL flanking the gene for grain yield; (3) 6H: bPb-6264 is the putative QTL flanking the gene for scald severity; (4) 7H: bPb-5458 and bPb-1946 co-locate on an identical locus. However, bPb-5458 was detected as the putative QTL for scald severity and bPb-1946 as the putative QTL for net blotch severity. This detection of different QTL in a co-locus occurred because of the non-identical genotypes between bPb-5458 and bPb-1946.

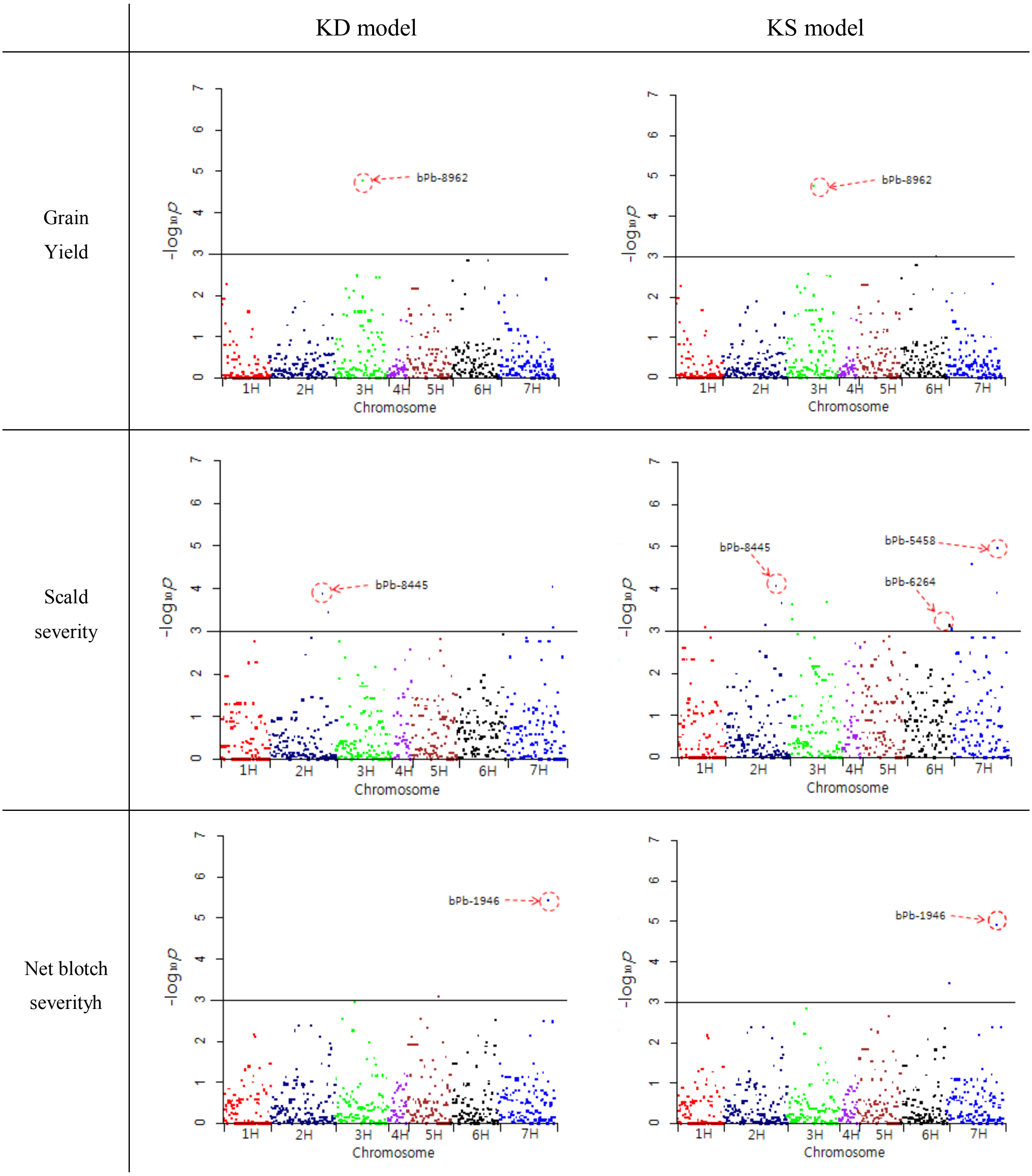


Figure 4-7. Manhattan plots from the kinship plus DAPC model (left) and the kinship plus STRUCTURE model (right) for grain yield (top), scald severity (middle) and net blotch severity (bottom). On each plot, the X-axis and Y-axis show the sequentially arranged chromosome numbers and the rating scale of $-log_{10}p$. Each dot represents a single marker. The vertical position of each dot indicates the QTL-marker linkage degree expressed as a form of $-log_{10}p$. Points above the solid line are significant markers of LD with QTL at the p < 0.001 level. Points in the dotted circle are the putative QTL filtered using the cross-validation.

4.4 Discussion

4.4.1 Correction of population stratification in association mapping

Based on the bi-plot derived from the DAPC, two groups were characterized in a panel of 45 spring barley varieties (Figure 4-2). The bi-plot explains 15.53 % of variation in the total phenotype. In the Bayesian clustering analysis using STRUCTURE, validating an appropriate K is vital but very difficult (Jombart., 2010). Generally, the K value is determined by referring to the estimated likelihood values resulting under various K values (Yu et al., 2005; Massman et al., 2011; Shi et al., 2011). In this study, the Bayesian clustering model was fitted with K = 2 to 10, and the results show that the likelihood values rise steadily with increasing K. This pattern was unlikely to give the proper resolution because K could become too large. According to Malysheva-Otto et al., (2006), European barley collections exhibited a narrow genetic diversity, and distinct clusters were identified according to morphological characteristics such as the number of spikes (two-rowed and six-rowed) and the seasonal types (spring sown and winter sown). In this work, the entirety of the barley cultivars was morphologically two-rowed and spring sown, and geographically released from Germany. In addition, the population size was small. Under these conditions, a small K might be convincing. Therefore, the manner of determining the K values based on the likelihood values was ignored in this work. As an alternative strategy, a different manner was considered. Previous studies (Massman et al., 2011; Pasam et al., 2012) described that the population structures characterized by different population analyses tend to show consistent results. Under this premise, it was attempted to determine the K value using prior knowledge of the subpopulation structure. In general, geographical information and morphological characteristics are often used as prior knowledge (Malysheva-Otto et al., 2006; Massman et al., 2011; Pasam et al., 2012). However, in this study, the provided barley collection could not be sub-divided in the general manner because both the origin and morphology of the germplasms were monotonous. Instead, the subpopulation that was derived from the DAPC was referenced as prior knowledge. As the DAPC resolved two main clusters, K=2 was examined using the STRUCTURE software tool. The comparison of the results from the STRUCTURE and DAPC analyses showed a 79.07% agreement (Table 3-1). Such a degree of agreement is obviously high. According to Jombart et al., (2010), the result from the DAPC produces a cluster that is similar to the result produced by the STRUCTURE analysis, which supports the value of K = 2 being the appropriate number for the inferred ancestry in the provided barley population.

4.4.2 Definition of QTL

In this study, 1181 DArT markers were used to genotype the barley cultivar collections for mapping. Of these, 710 markers were mapped, and 471 were unmapped. In the mapping study, the mapped markers are addressed and termed as QTL. The knowledge of the marker positions increases the efficiency of gene isolation through map-based cloning and marker assisted selection (MAS) because the mapped markers improve the chance of identifying polymorphic markers in genetic backgrounds (Hearnden et al., 2007).

4.4.3 Strategy to overcome a weakness from a small population

A large sample size is beneficial in QTL mapping by providing an abundance of observations of LD block segregation (Wang et al., 2012a). In this work, the size of the barley collection was small, which violated the ideal conditions for AM. According to Melchinger et al., (1998), small sample size causes an upward bias in mapping resolution. In addition, the barley germplasm generally shows a long range extent of LD because of its selfing nature, which can reduce the QTL detection power (Kraakman et al., 2004; Zhang et al., 2009; Massman et al., 2011). The above conditions indicate that the mapping trial in this study might be vulnerable to the detection of spurious marker-trait associations, therefore a stringent level of p-value can be effective (Massman et al., 2011). To do so, the confidence level of $p < 0.001$ was set, which is a similar to higher level than the values used in other studies (Kraakman et al., 2004; Xue et al., 2009; Roy et al., 2010; Shi et al., 2011; Looseley et al., 2012; Rode et al., 2012). Subsequently, the cross-validation was performed.

4.4.4 Variation of the detected QTL depending on the measurements of subpopulation

The evolution and breeding activities in crops generate a subpopulation structure, which often prevents observations of the pure phenotype distribution for a particular trait (Yu et al., 2005) because subpopulation structure causes the mapping experiment to catch the spurious marker-trait associations (Yu et al., 2005). For improving the power of TAM detection, the estimate of the subpopulation profile needs to be accounted for in the mapping to correct confounded LD status (Yu et al., 2005; Cockram et al., 2008), which implies that the list of TAMs could vary depending on the subpopulation matrix. Indeed, this study showed that the TAMs detected from the KD and KS models were found to be asymmetric in the analysis of scald severity (Table 4-2). This suggests that the TAMs detected from using the different subpopulation analyses do not provide precisely identical resolutions but similar resolution (Massman et al., 2011; Pasam et al., 2012). In mapping TAM for scald severity, although the KS model yielded a larger number of TAMs over the KD model, the robustness between the two models could not be determined because some TAMs may have a latent risk of a false positive. Therefore, the comparison between the two models based on efficiency was not discussed in this study.

4.4.5 Comparison of the presently and previously detected QTL for grain yield

As a popular trait for mapping, a number of QTL for grain yield were previously found across the seven barley chromosomes (Hayes et al., 1993; Bezant et al., 1997; Kraakman et al., 2004; Xue et al., 2009; Comadran et al. 2011). In this work, a single QTL (bPb-8962) was detected that was simultaneously found by both the KD and KS models. In addition, p-values for bPb-8962 were shown to be highly significant for both models (see Table 3-2), likely indicating that bPb-8962 co-locates with an extremely strong gene controlling the grain yield on a short range LD block. bPb-8962 was mapped on the long arm of 3H (178.6 cM), and Hayes et al., (1993) showed that 3H contains an abundance of QTL for grain yield, which may be related to the fact that the single QTL detected in this study for grain yield was detected on 3H. Specifically, QTL for grain yield on 3H have a high chance of duplication with QTL for plant height because grain yield and plant height are strongly correlated and because QTL for plant height have mainly been identified on 3H (Laurie et al., 1993; Hayes et al., 1993; Thomas et al., 1995; Li et al., 2009; Xue et al., 2009). The mutual correlation between grain yield and plant height makes sense because short plants are additionally resistant to lodging (Laurie et al., 1993; Hayes et al., 1993). Furthermore, short barleys had a high yield even in environments where lodging no longer occurred (Hayes et al., 1993). This suggests that the QTL for grain yield could be related to that for plant height at the level of genetic mechanism. To verify whether bPb-8962 confers plant height, marker-trait associations for plant height were estimated using a provided data set. Across the KD and KS models, two QTL (bPb-5899 and bPb-0990) were detected in common. However, their locations remained unmapped, so the comparison of the QTL for plant height with bPb-

8962 was not conducted in this study. A previous study using bPb-8962 for the purpose of mapping barley plant height reported that bPb-8962 was found to be out of the QTL's region (Li et al., 2009).

4.4.6 Comparison of the presently and previously detected QTL for scald severity

In previous studies, QTL for scald (*Rhynchosporium secalis*) resistance were reported on chromosomes 2H, 3H, 6H, and 7H (Backes et al., 1995), on chromosomes 1H, 2H, 3H, 6H and 7H (Thomas et al., 1995) and on chromosomes 3H, 4H and 6H (Jensen et al., 2002). Zhan et al., (2008) reviewed the characteristics of scald resistance on barley and described QTL for scald resistance being rich on chromosomes 3H, 6H, and 7H, few on chromosomes 1H, 2H and 4H, and absent on chromosome 5H. In this work, three QTL for scald severity were mapped across three chromosomes: bPb-8445 on chromosome 2H, bPb-6264 on chromosome 6H and bPb-5458 on chromosome 7H. Consistent with the previous report, this study revealed no QTL on chromosome 5H. Looseley et al., (2012) reported that the most effective QTL for this trait was detected in a region 107-111 cM on chromosome 7H. However, the resulting QTL (bPb-5458) in this study was mapped at 82.6 cM on the same chromosome.

4.4.7 Comparison of the presently and previously detected QTL for net blotch severity

A single QTL (bPb-1946) was identified for net blotch (*Pyrenophorateres Drechs.*) severity. The detected QTL was located on 7H at 82.6 cM (bPb-1946). Historically, the resistance to net blotch severity was intensively developed in six-rowed barley (Wilcoxson et al., 1990; Fetch and Steffenson., 1994; Steffenson et al., 1996), which was likely related to the low number of QTL detected in this work because the barley cultivar collection in the present study was solely made up of two-rowed barley. In previous studies, QTL for net blotch severity were analyzed at two growth stages of plants: (1) the seedling stage and (2) adult plant stage (Steffenson et al., 1996). Interestingly, the QTL detected in the two stages have been demonstrated to differ (Steffenson et al., 1996; Grewal et al., 2008), which indicates that the change of gene expression at the different growth stages may affect the detection of QTL for net blotch resistance. In previous studies, the QTL for net blotch severity were extensively identified across the seven barley chromosomes. Steffenson et al., (1996) mapped three QTL on chromosome 4H and 6H at the seedling stage and seven QTL on chromosome 1H, 2H, 3H, 4H, 5H, and 7H at the adult plant stage (Steptoe/Morex). William et al., (1999) mapped a single gene (Rpt4) conferring resistance to the net blotch on chromosome 7H in the 'Galleon/Haruna Nijo' cross. Grewal et al., (2008) detected 12 QTL for net blotch resistance on chromosome 2H, 3H, 4H, 5H, 6H and 7H in 150 DH lines from the cross, CDC Dolly/TR251.

In this discussion, the comparison of the detected QTL (bPb-1946) and the prior QTL was not addressed because the QTL analysis for net blotch severity has a sensitive response under the different germplasm sets and the different monoconidial isolates (Graner et al., 1996), so such circumstances require a great caution in comparing the presently detected QTL with the previously detected QTL (Graner et al., 1996; Mannien et al., 2000).

5. Genomic estimated breeding value prediction using RR-BLUP in a German spring barley collection

5.1 Introduction

Genomic selection (GS) is an approach for predicting the unobserved phenotypic performance for genotyped individuals by applying the estimates of marker effects in relation to a trait (Meuwissen et al., 2001). The GS routinely requires two subpopulations: a training set and a validation set. In the training set, the markers' effects for a trait are measured, and in the validation set, the phenotypic performances of individuals are calculated by applying the estimated effects of the markers to the individuals' genotypes. The GS predicts the unobserved phenotype by summing up a number of the estimated effects of markers at one time. This facilitates the GS to measure the quantitative traits more precisely than marker assisted selection (MAS) that only uses low numbers of markers of LD with QTL (Solberg et al., 2008).

Several methods for the GS are available and include least-square analysis (LS), ridge regression best linear unbiased prediction (RR-BLUP), Bayes-A analysis and Bayes-B analysis. The above methods have their own approaches for estimating the marker effects. The LS estimates the marker effects based on a single-marker regression, selecting the large-effect markers through a significance test (Meuwissen et al., 2001; Habier et al., 2007). The RR-BLUP estimates the marker effects by fitting a model with the whole marker genotype matrix at once under the assumption that every genetic variance per locus is equal (Meuwissen et al., 2001; Habier et al., 2007; Zhong et al., 2009; Asoro et al., 2011; Endelman., 2011; Rutkoski et al., 2012; Zhao et al., 2012). The Bayesian analyses such as the Bayes-A and Bayes-B are based on the BLUP model and usually fit the model using the Markov Chain Monte Carlo (MCMC) method. The popular algorithms for implementing the MCMC are the Gibbs sampling and Metropolis-Hastings algorithm. The Bayesian method combined with MCMC summarizes the unknown QTL-effects with posterior distribution gained by processing a prior distribution and data point at the same time. Such a procedure differentiates the degree of allelic variances across loci depending on the degree of QTL-effects. Of the above methodologies, the RR-BLUP and Bayesian analyses represent the major approaches to the GS because both show reasonably fair performances (Meuwissen et al., 2001). However, in terms of ease of use, the RR-BLUP is superior to the Bayesian methods because it is simple to implement and demands less computations (Lorenzana and Bernardo., 2009). For this reason, the RR-BLUP has become increasingly popular.

In this study, the conditions that increase the RR-BLUP accuracy were examined with German spring barley germplasms for grain yield, scald severity and net blotch severity. The experimental conditions to determine the desirable conditions were: (1) varying marker density, (2) varying size of training set and (3) varying QTL-marker association threshold to select markers.

Objective

The objective of Chapter 5 is to determine the optimum conditions to improve the accuracy of estimating the GEBVs using RR-BLUP in self-pollinating crop.

5.2 Materials and Methods

Phenotypic data

Phenotypic performances for grain yield, scald severity and net blotch severity for 45 spring barley lines were estimated by German Landessortenvesuche (LSV) from 1992 to 2002. The grain yield was scored with kg/ha. The infection ratings for scald and net blotch severities were scored on a scale of 0-9 according to the degree of severity in which 0 denotes the most positive expression and 9 the most negative expression. A list of the barley samples used in this study is provided in Appendix Ⅲ.

Genotype data

DNA samples of the 45 barley entries were extracted following the protocol recommended by Triticarte Pty. The DNA samples were sent to Triticarte Pty for genotyping of the DArT markers. A total of 1181 loci were scored for the panel of the barley collection. Because DArT markers have a dominant system, the genotypes were scored 1 for present, 0 for absent and NA for unknown.

Basic model and kinship matrix

For estimating the marker effects on the three traits, the ridge regression best linear unbiased prediction (RR-BLUP) method was used. The basic model can be denoted as follows:

$$y = WGu + e \qquad \text{(Equation 5-1)}$$

where y = the vector of the phenotype observations; W = the design matrix that relates the lines to observations; G = the DArT genotype data; u = the vector of the unknown marker effects; e = the residual vector.

The RR-BLUP assumes that all markers have an equal genetic variance ($\frac{V_g}{n}$, where V_g = the total genetic variance and n = the number of markers) and that g ~ $N(0, K\sigma_g^2)$, where K is the marker-based relationship matrix and σ_g^2 is the total variance of marker effects (Endelman., 2011). The marker-based relationship matrix was obtained as follows:

$$K = GG' \qquad \text{(Equation 5-2)}$$

where G is the DArT genotype data. The marker effects were measured by means of the rrBLUP package (Endelman., 2011) for the R software platform (R Development Core Team., 2012).

Estimation of mean phenotypes in barley

As a selfing species, barley has nearly homogeneous genomes within a same variety group so that they are regarded as clones, which makes it feasible for phenotypes of barley varieties to be measured over years across locations and facilitates the estimation of the mean phenotype (MP) for a particular barley variety. It is noted that the MP highly approximates the phenotypic performance in general environments so that the use of the MP can increase the response to selection (Piepho et al., 2008; Zhong et al., 2009), which indicates that the MP can represent the approximate true breeding values (TBVs). In this study, the MPs were used to overcome the lack of provision of TBVs.

Cross-validation of the effect of the marker density on GEBV accuracy

The effect of marker density on GEBVs accuracy was estimated through cross-validation. This work constitutes a series of two stepwise random samplings. In the first step, a set of data was randomly divided into training and validation subpopulations, and the size of the training and validation sets were 42:3 for grain yield, 38:3 for scald severity and 37:3 for net blotch severity, respectively. In the second step, partial markers were randomly selected in the training subpopulation to vary the marker density. The number of markers to be selected were 1,181 for whole set, 1,063 for 90 %, 945 for 80 %, 827 for 70 %, 709 for 60 %, 590 for 50 %, 472 for 40 %, 354 for 30 %, 236 for 20 % and 118 for 10 %. The measurement of GEBVs accuracy was conducted using Spearman's rank correlations between the MPs and the GEBVs for the validation entries. The first and the second steps are repeated 50 and 1,000 times, respectively. Therefore, the GEBVs accuracy was estimated 50,000 times and averaged.

Cross-validation of the effect of the training set size on GEBVs accuracy

The effect of the training set size on GEBVs accuracy was observed through the cross-validations. For this work, the reducing size of samples was randomly divided into training and validation data sets, and the size of the samples was reduced from 45 to 36 for grain yield, from 41 to 32 for scald severity and from 40 to 31 for net blotch severity. Within sets of samples, three barley lines were randomly sampled to be the validation entries, and the remaining lines were used as the training entries. To implement the cross-validation, the random divisions into the training and validation sets were repeated 10,000 times per each size of training set. To obtain the GEBVs accuracy, the Spearman's rank correlation coefficient between MPs and GEBVs for the validation entries was measured.

Cross-validation of the effect of various degree of LD on GEBV accuracy

The effect of marker selection based on marker-trait association on the GEBVs accuracy was investigated. In this trial, two cross-validation tests were implemented. In the first test, a set of barley lines was randomly divided into the training and validation sets and the proportions were 42:3 for grain yield, 38:3 for scald severity and 37:3 for net blotch severity, respectively. To see the degree of marker-trait association, p-values obtained via the AM resulting from the KS and KD models were referenced (see Chapter 4). The p-value thresholds for marker selection were given with an interval of 0.1 between 0.1 and 1.0, and a list of p-values gained by taking a greater value between the p-values obtained from the KD and KS models was used. For the experiments, 10 sets of markers were prepared per every trait. With each marker set, RR-BLUP was performed in the training set. The estimates of marker-effect were subsequently applied to the entries in the validation set to rank the entries. The GEBVs accuracy was measured using the Spearman's rank correlation coefficient between MPs and GEBVs in the validation set, and the measurement was replicated 10,000 times of allocating the samples into the training and validation sets. The second cross-validation test was conducted to determine if the results from marker selection based on p-values are more accurate than the results from random marker selection. For this experiment, the cross-validation was performed in two stepwise random samplings. In the first step, a set of data was randomly divided into training and validation sets, and the sizes of the training and validation sets were 42:3 for grain yield, 38:3 for scald severity and 37:3 for net blotch severity. In the second step, partial markers were randomly selected in the training set. The sizes of the selected marker set corresponded to the sizes of the marker sets gathered based on p-value. The GEBVs accuracy was conducted by measuring the Spearman's rank correlation coefficient between MPs and GEBVs for the validation entries. The first and the second steps were repeated 100 and 2,000 times, respectively. Therefore, the GEBVs accuracy was estimated through 200,000 times of repetition. In the end, the GEBVs accuracy based on p-value and the GEBVs accuracy based on random selection of marker were compared.

5.3 Results

5.3.1 Precision of GEBVs under varying marker densities

The prediction of GEBVs under varying marker densities using the RR-BLUP was derived from a cross-validation test. The sampled sizes of marker sets varied from 100 % to 10 % with a 10 % interval. In this study, the random divisions of samples and the random marker selections were performed 50 and 1,000 times, respectively. The final prediction accuracies were obtained by averaging the 50,000 estimations.

The resulting prediction accuracies are summarized in Table 5-1 and graphically displayed in Figure 5-1. The highest prediction accuracies were observed at a level of 70 % (0.3219) for grain yield, at a level of 100 % for scald severity and at a level of 90 % (0.4344) for net blotch severity. In contrast, the values for the lowest prediction accuracy were found at the smallest level (10 %) for all three traits: 0.2774 for grain yield, 0.3168 for scald severity, and 0.3599 for net blotch severity. Across all the traits, the distributions of prediction accuracy were shown to have both a plateau and decreasing section. The sections of plateau were found in approximately 100-30 % for all traits, whereas the decreasing sections were found in approximately 30-10 % for all traits. Overall, the patterns of prediction accuracy were observed to gradually decrease for all traits in response to the decreasing marker density.

5.3.2 Precision of GEBVs under varying sizes of training set

The prediction accuracy of GEBVs under varying sizes of training sets using the RR-BLUP was measured depending on the cross-validation test. For this experiment, the sample sizes of training sets were varied from 42 to 33 for grain yield, 38 to 29 for scald severity, and 37 to 28 for net blotch severity, and the size of the validation sets was consistently three. The training and validation sets were randomly allocated 10,000 times within each size of training set. Thus, the estimates of prediction accuracy were conducted 10,000 times, so the final prediction accuracy was gained by averaging the estimates. The size of the DArT marker panel was 1,181, and the estimates of prediction accuracy and the graphical distributions are shown in Table 5-2 and in Figure 5-2, respectively. The highest prediction accuracies for the three traits were observed in the largest training set for grain yield (0.3221) and scald severity (0.3666) and in the fourth largest training set for net blotch severity

(0.4281). The lowest prediction accuracy was 0.3117 at the lowest sizes of training sets for grain yield, whereas it was 0.3373 and 0.3959 at the ninth largest set for scald and net blotch severities. The distributions of prediction accuracy for the three traits generally tended to decrease with decreasing size of training set.

5.3.3 Precision of GEBVs with markers selected based on QTL-marker association

The RR-BLUP was conducted for the three traits using markers selectively chosen based on the level of marker-trait association. For the trials, two cross-validations were performed to (1) measure the prediction accuracy when using the marker collection selected based on marker-trait association and (2) measure the prediction accuracy when using the marker collection selected at random. The superiority of the former trial was investigated by comparing with the result from the latter trial. In the first cross-validation, 10 sets of markers were prepared, in which markers were selected depending on the degree of marker-trait association. The degree of marker-trait association was referenced by the p-values obtained by the AM (see Chapter 4). Sets of makers were varied with p-value thresholds from p = 0.1 to 1.0 with an interval of 0.1. A subset of markers collected at lower p-value threshold contains the lower number of markers, whereas the higher p-value increases the size of a marker set. The cross-validation was first performed to divide the whole set into the training and validation sets at random. Subsequently, the prediction accuracy in the validation set was measured. For each set size, sub-division was performed 10,000 times. The estimates of prediction accuracy ranged between 0.3226 and 0.7323 for grain yield, between 0.3534 and 0.5394 for scald severity and between 0.4431 and 0.8326 for net blotch severity. A subset of markers that pooled at lower p-value threshold provided the considerably high prediction accuracy. However, the fall of the estimated prediction accuracy was observed in a section of 0.4-0.3 for grain yield and 0.2-0.1 for scald severity and net blotch severity. In the second cross-validation, the sizes of the sample sets corresponded with the sizes from the first cross-validation. Across all traits, the estimates of prediction accuracy formed a plateau or gradually decreased as the size of the marker subsets became smaller. The first cross-validation test produced ratios of maximum vs. minimum prediction accuracies, which were 2.270 (= 0.7323/0.3226) for grain yield, 1.527 (= 0.5396/0.3534) for scald severity and 1.879 (= 0.8326/0.4431) for net blotch severity, whereas the second cross-validation tests produced ratios of 1.158 (= 0.3160/0.2730) for grain yield, 1.047 (= 0.3641/0.3477) for scald severity and 1.156 (= 0.4340/0.3753) for net blotch severity. A difference between the former and latter trials was the greatest for grain yield, followed by net blotch severity and scald severity. The above results are summarized in Table 5-3 and graphically displayed in Figure 5-3.

Table 5-1. Rank correlation coefficients between mean phenotypes and genomic estimated breeding values for grain yield, scald severity and net blotch severity under different sizes of the marker set. The values can be regarded as the prediction accuracy of GEBVs, whose measurements were performed using cross-validation.

Proportion of the sampled markers	The number of markers	The prediction accuracies of GEBV under different sizes of marker data		
		Grain yield	Scald Severity	Net blotch severity
100 %	1181	0.3130	0.3661	0.4333
90 %	1063	0.3178	0.3598	0.4344
80 %	945	0.3189	0.3583	0.4321
70 %	827	0.3219	0.3565	0.4329
60 %	709	0.3175	0.3575	0.4229
50 %	590	0.3142	0.3506	0.4231
40 %	472	0.3106	0.3518	0.4157
30 %	354	0.3015	0.3472	0.4073
20 %	236	0.2946	0.3347	0.3930
10 %	118	0.2774	0.3168	0.3599

Table 5-2. Rank correlation coefficients between mean phenotypes and genomic estimated breeding values for grain yield, scald severity and net blotch severity under different sizes of training set. These values can be regarded as the prediction accuracy of GEBVs, whose measurements were performed using cross-validation.

The prediction accuracies of GEBV under different sizes of training set					
Size of training set	Grain yield	Size of training set	Scald severity	Size of training set	Net blotch severity
42	0.3221	38	0.3666	37	0.4261
41	0.3195	37	0.3571	36	0.4276
40	0.3188	36	0.3567	35	0.4234
39	0.3167	35	0.3520	34	0.4281
38	0.3163	34	0.3540	33	0.4258
37	0.3170	33	0.3520	32	0.4210
36	0.3160	32	0.3478	31	0.4193
35	0.3151	31	0.3394	30	0.4176
34	0.3131	30	0.3373	29	0.3959
33	0.3117	29	0.3408	28	0.3965

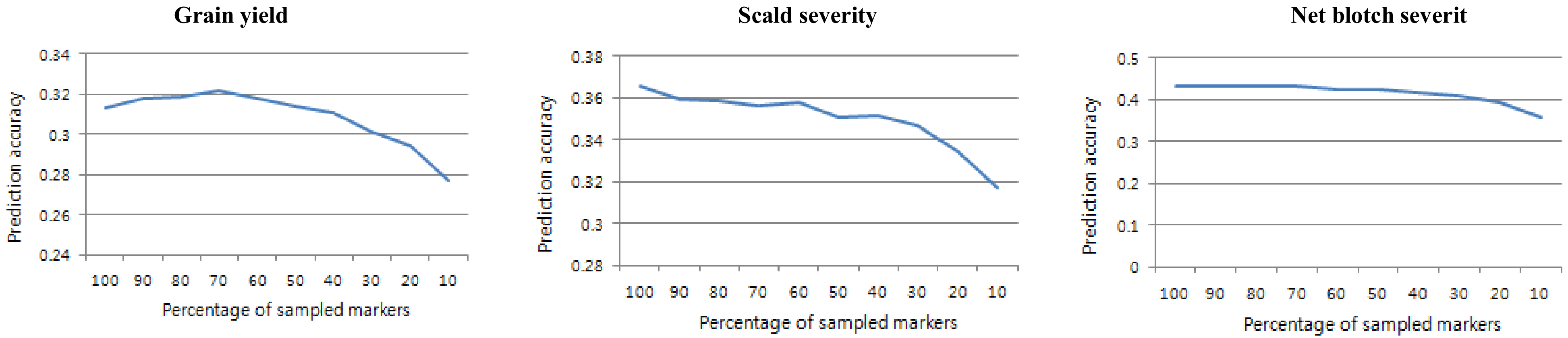


Figure 5-1. Distritution of GEBVs under varing sizes of marker sets with an interval of 10 % from 100 % to 10 % for grain yield (left), scald severity (middle) and net blotch severity (right).

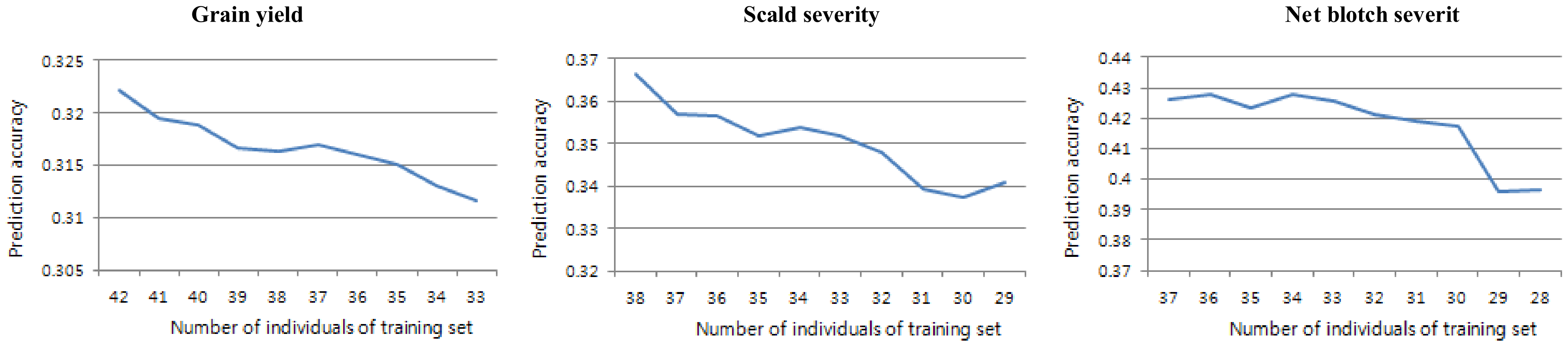


Figure 5-2. Distritution of GEBVs under decreasing sizes of the training sets for grain yield (left), scald severity (middle) and net blotch severity (right).

Table 5-3. Comparison of prediction accuracy between GEBVs derived from using markers selected at the given p-value thresholds and GEBVs derived from using markers selected at random. To level the LD effects for markers, p-values obtained from the association mapping (Chapter 4) were referenced. The p-values as a criterion for the marker collection were varied, with an interval of 0.1 between 0.1 and 1.0.

p-value	Grain yield			Scald severity			Net blotch severity		
	The number of markers	Prediction accuracy		The number of markers	Prediction accuracy		The number of markers	Prediction accuracy	
		P-value based subset	Random subset		P-value based subset	Random subset		P-value based subset	Random subset
≤ 1.0	1181	0.3226	0.3160	1181	0.3534	0.3641	1181	0.4431	0.4340
≤ 0.9	957	0.3654	0.3164	1035	0.3809	0.3582	960	0.5147	0.4306
≤ 0.8	813	0.4537	0.3189	978	0.4023	0.3588	846	0.5978	0.4323
≤ 0.7	697	0.5511	0.3150	919	0.4383	0.3582	748	0.6467	0.4267
≤ 0.6	596	0.6021	0.3178	857	0.4613	0.3578	663	0.7051	0.4276
≤ 0.5	497	0.6394	0.3125	775	0.4912	0.3585	569	0.7459	0.4235
≤ 0.4	399	0.6934	0.3099	682	0.5141	0.3555	484	0.7744	0.4192
≤ 0.3	317	0.6835	0.3049	595	0.5323	0.3501	394	0.7949	0.4115
≤ 0.2	235	0.6968	0.2963	517	0.5396	0.3544	283	0.8326	0.3992
≤ 0.1	137	0.7323	0.2730	369	0.4652	0.3477	160	0.7601	0.3753

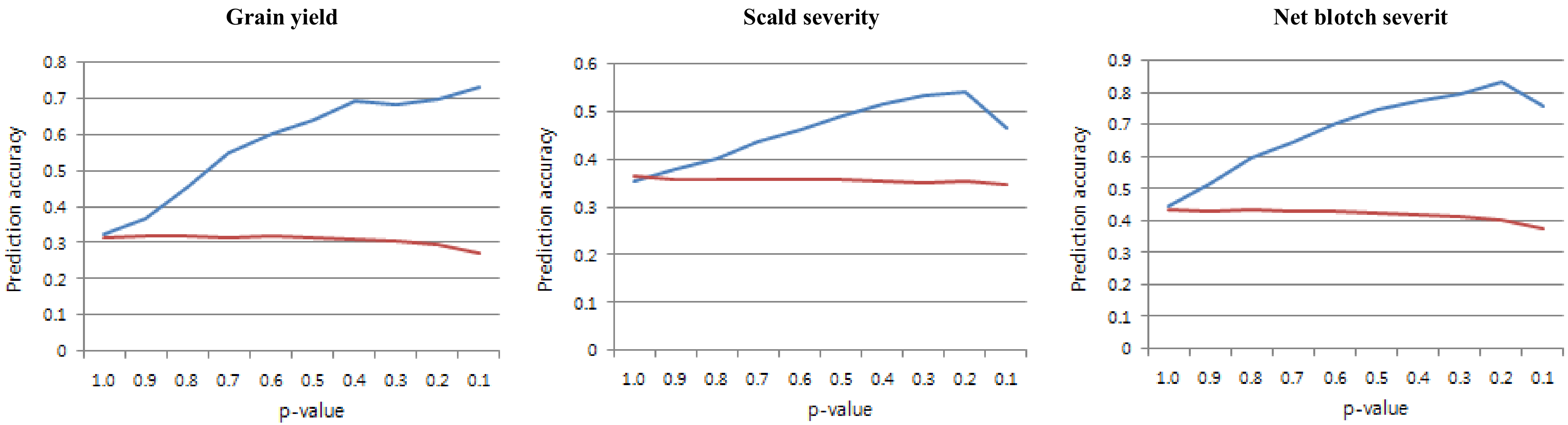


Figure 5-3. Comparison between the distribution of GEBVs obtained using markers pooled based on the p-values (blue) and distribution of GEBVs obtained using markers pooled at random (red) for grain yield (left), scald severity (middle) and net blotch severity (right). Within each figure, the blue line shows the distribution of GEBVs accuracies obtained by performing the RR-BLUP with the markers that remained after filtering below a provided p-value. The red line provides a control of the blue line, which shows the distribution of GEBVs measured using the same size of marker sets selected at random. In the above figure, the X-axis represents the p-value as a criterion for filtering markers to prepare the subsets of markers. The p-values were referenced from the association mapping study (Chapter 4). The Y-axis indicates the prediction accuracy on a scale of 0-1.

5.4 Discussion

5.4.1 Difference between the genomic selection and traditional BLUP

The purpose of breeding value prediction can be divided into two categories to (1) predict the unobserved phenotypic performance of individuals or (2) obtain the selection index for choosing the superior parents (Meuwissen et al., 2001; Oakey et al., 2006; Schaeffer., 2006; Goddard and Hayes., 2007; Muir., 2007; Piepho et al., 2008; Nielsen et al., 2009). The GS determines purpose (1) and is a useful approach for reducing selection cost because it can select superior individuals by predicting phenotypic performances at an early growth stage (Schaeffer., 2006). In contrast, the traditional BLUP pursues purpose (2) because it focuses on the selection of the best parents by disclosing the latent BVs (Robinson., 1991; Panter and Allen., 1995a; Panter and Allen., 1995; Pattee et al., 2001; Purba et al., 2001).

5.4.2 Impact of marker density on the prediction accuracy

In RR-BLUP, it is known that marker density has an impact on the prediction accuracy (Bernardo and Yu., 2007; Solberg et al., 2008; Zhong et al., 2009; Nielsen et al., 2009; Nakaya and Isobe., 2012; Zhao et al., 2012; Crossa et al., 2014). To examine this, the prediction accuracy of RR-BLUP under varying densities of markers was explored using the cross-validation. For this work, the size of the marker panels varied from 100 % to 10 % with an interval of 10 %, and within each marker subset, the rank correlation coefficient between GEBVs and MPs for the validation entries was observed. The resulting rank correlation coefficients (= prediction accuracy) ranged between 0.2774 (10 %) and 0.3219 (70 %) for grain yield, between 0.3168 (10 %) and 0.3661 (100 %) for scald severity and between 0.3599 (10 %) and 0.4334 (90 %) for net blotch severity. For all three traits, it was shown that the distributions of prediction accuracy remained at a plateau in the range of 100-30 %, whereas the prediction accuracy fell in the range of 30-10 %. This result is consistent with previous findings (Habier et al., 2007; Lorenzana and Bernardo., 2009; Nielsen et al., 2009; Daetwyler et al., 2010; Asoro et al., 2011; Zhao et al., 2012). The stable prediction accuracy of the RR-BLUP in the modest-sized marker set is beneficial because it provides workers with a comparable prediction accuracy without increasing the cost for genotyping.

5.4.3 Impact of the size of training set on the prediction accuracy

To examine the prediction accuracy of the RR-BLUP under varying sizes of training set, cross-validations for grain yield, scald severity and net blotch severity were performed as the training set size was decreased one by one so that 10 levels of training sets were prepared for the three traits. In contrast, the size of the validation set was consistently three. The resulting values for prediction accuracy ranged between 0.3221 and 0.3117 for grain yield, between 0.3666 and 0.3373 for scald severity and between 0.4281 and 0.3959 for net blotch severity. The estimates of prediction accuracy were shown to diminish gradually as the size of a training set decreased (Table 5-2). This is consistent with previous findings and implies that the prediction accuracy would increase if the size of the training set increases (Zhong et al., 2009; Asoro et al., 2011; Heffner et al., 2011; Nakaya and Isobe., 2012; Zhao et al., 2012; Crossa et al., 2014). Considering that the sizes of the training sets were small compared to previous studies (Habier et al., 2007; Lorenzana and Bernardo., 2009; Daetwyler et al., 2010; Zhao et al., 2012; Schulz-Streeck and Piepho et al., 2009), it is likely that there is a large potential to increase the prediction accuracy by increasing the size of the training set.

5.4.4 Impact of markers of LD with QTL on the prediction accuracy

Conceptually, the GS predicts GEBVs by aggregating the infinitesimal QTL effects captured by the numerous markers throughout the entire genome (Meuwissen et al., 2001; Goddard and Hayes., 2007). Therefore, an abundance of markers of LD with large-effect QTL supposes to be advantageous in increasing the efficiency of the GS (Zhong et al., 2009; Habier et al., 2007; Muir., 2007). In a similar context, Habier et al., (2007) described that the RR-BLUP elevated the prediction accuracy of GEBVs with an increment of the number of markers of LD with QTL. Regarding this, in this study, it was hypothesized that a set of markers with a high degree of marker-trait association may increase the prediction accuracy. To verify this hypothesis, the GEBVs predictions were examined using the RR-BLUP method with sets that have varying degrees of marker-trait associations. To vary the degree of marker-trait association among sets of markers, p-values from the AM (Chapter 4) were referenced because p-values represent the degree of non-random association between markers and traits by linkage. In this study, the p-value provides a criterion for filtering the markers below a particular degree of marker-trait association. In this study, 10 levels of marker sets were prepared in which the p-value thresholds were set at between 0.1 and 1.0 with an interval of 0.1. As a result of experiments for grain yield, scald severity and net blotch severity, marker sets selected at decreasing p-value thresholds predominantly provided a considerable improvement in the prediction accuracy. This indicates that a marker set that comprises markers with high marker-trait associations can increase prediction accuracy, which suggests that the previously stated hypothesis is true and additionally supports Habier et al., (2007)'s statement. Based on previous literatures, two reasons regarding this phenomenon were supposed. The first reason is that the stronger LDs within a marker set lead to the improved estimates of marker effects. The second reason is that the stronger LDs lead to the improved precision for genetic relationships among individuals (Habier et al., 2007; Zhong et al., 2009; Asoro et al., 2011). However, in a decreasing pattern of p-values, sudden falls in the prediction accuracy were found in a section of 0.4-0.3 for grain yield, 0.2-0.1 for scald severity and net blotch severity. For grain yield, the decreased amount of prediction accuracy was approximately 1 % with a reduction of the size of marker set from 399 to 317. Meanwhile, the decreased amounts accounted for 7.44 % and 7.25 % with reducing size of marker set from 517 to 369 for scald severity and from 283 to 160 for net blotch severity, respectively. The significant reductions of prediction accuracy might happen because the small-sized marker set offsets an advantage of the lower p-value threshold. Previous studies correlated between a quantity of QTL markers and prediction accuracy of GEBVs using a simulation (Habier et al., 2007). This is the first study to use the p-values resolved by the AM for the purpose of predicting GEBVs with real sets of data.

5.4.5 Further investigation

Previous studies (Asoro et al., 2011; Crossa et al., 2014) revealed that closed multi-parental population provides the improved prediction accuracy of GS over a population comprising lines that had diverse genetic backgrounds. In fact, a majority of studies (Meuwissen et al., 2001; Habier et al., 2007; Zhong et al., 2009; Asoro et al., 2011; Endelman., 2011; ; Zhao et al., 2012; Crossa et al., 2014) estimated the GEBVs of progenies by using the lines generated from several founders as a training set, which led to a close relatedness between the training and validation sets. In this study, however, the training and validation sets were not in a genealogical relationship, which apparently lessens the relatedness between the two sets. If the training and validation sets were genealogically related, the improved resolution of the RR-BLUP could be obtained. Therefore, it may be worth testing the RR-BLUP in a closed multi-parent population.

6. General discussion

In this study, a set of equations (Equations 2-6, 2-7 and 2-8) were derived for the purpose of defining the relationship coefficient between any two individuals by using the parentage information in a plant pedigree. Depending on this, a software tool called PopKin was developed, whose strengths are the capacity to build an NRM with the consideration of the number of self-pollination and the use of plant pedigree notation.

For implementing BLUP, German barley cultivar collection that was publicly available from LSV was used. A data set was prepared by compiling the multiple data sets in unbalanced trials. The NRM with pedigrees of the provided accessions was constructed using PopKin, and the resulting NRM was included in the BLUP model for grain yield, scald severity and net blotch severity using the ASReml-R package. The hetitabilities for the three traits resulted in 0.719 for grain yield, 0.491 for scald severity and 0.581 for net blotch severity, respectively, which appeared to be similar to higher compared with the previously reported heritabilities (Durel et al., 1998; Oakey et al., 2006). The resulting heritabilities reflect that the phenotypic observations recorded in unbalanced trials were sufficient to use. Traditionally, the BVs for breeding lines were estimated by calculating the MP, general combining ability or mid-parent value (Bernardo., 1994; Panter and Allen., 1995a; Pattee et al., 2001; Oakey et al., 2006; Piepho et al., 2008; Zhong et al., 2009). For viewing the accuracy of the BLUP estimates, the MPs were obtained as traditional BVs and compared with the BLUP estimates. The rank correlation coefficients between the BLUP estimates and MPs were shown to be 0.854 for grain yield, 0.893 for scald severity and 0.940 for net blotch severity. Such discrepancy between the MPs and BLUP estimates could be interpreted as a margin to improve the response to selection using the BLUP (Pattee et al., 2001). In fact, previous studies reported that the BLUP provides an improved response to selection than the methods based on phenotypic observations or linear models without an NRM (Panter and Allen., 1995a; Panter and Allen., 1995b; Pattee et al., 2001a; Oakey et al., 2006). For the better validation, however, it will be necessary to estimate the prediction accuracy between the TBVs and BLUP estimates.

For conducting the AM for grain yield, scald severity and net blotch severity with German spring barley cultivars, a single marker regression has been conducted using BLUP. It is noted that a risk to detect the spurious TAMs becomes escalated when using a low number of accessions (Melchinger et al., 1998; Massdam et al., 2011), which implies that the given numbers of accessions (45 for grain yield, 41 for scald severity and 40 for net blotch severity) could increase a chance to detect the spurious TAMs. To filter the spurious TAMs, the Wald test and cross-validation were employed, from which three TAMs (one for grain yield, one for scald severity and one for net blotch severity) and six TAMs (one for grain yield, four for scald severity and one for net blotch severity) were detected in the KD and KS models, respectively. The number of TAMs in this study is lower than that in previous studies (Pswarayi et al., 2008; Xue et al., 2009), which is because of stringent significance tests such as the Wald test at a significant level of $p < 0.001$ and cross-validation test. All the TAMs detected in the KD model are found in the KS model, which indicates that the use of different subpopulation matrices provides not the same but similar resolution in detecting TAMs (Massdam et al., 2011; Pasam et al., 2012). The TAMs detected in this study were not previously reported, which limited validating a robustness of the present mapping result. The six TAMs detected in this study may be useful for breeding practices using marker assisted selection for improving grain yield, scald severity and net blotch severity.

Using the RR-BLUP, GEBVs were measured with the same barley collections as used for the AM. It was observed that the sizes of marker set and training set are positively proportional to the accuracy of GEBVs, which is in agreement with the previous studies (Solberg et al., 2008; Lorenzana and Bernardo., 2009; Asoro et al., 2011; Zhao et al., 2012; Crossa et al., 2014). In this study, an experiment to form the training sets was attempted by referencing the p-values obtained from the AM. The p-value thresholds to form the 10 levels of marker set were given with $p < 0.1$, $p < 0.2$, $p < 0.3$, $p< 0.4$, $p < 0.5$, $p < 0.6$, $p < 0.7$, $p < 0.8$, $p < 0.9$ and $p < 1.0$. It was found that a training set formed at lower p-value predominantly provided higher accuracy of GEBVs despite the size of the training set is decreasing. However, it was found that the accuracy of GEBVs dramatically fell down in scald severity and net blotch severity when $p < 0.1$, which is presumably because the small size of the training set suppressed the increment of the accuracy of GEBVs. This illustrates that GEBVs accuracy can be improved in lower level of the p-value and the moderate size of the training set. In contrast, Asoro et al., (2011) reported that a simple addition of QTL markers to a list of markers does not improve the prediction accuracy of the RR-BLUP. Hence, it is important to note that the level of LD with QTL for entire markers in a training set needs to be controlled by referencing the p-values. Above, a lack of the same QTL between the previous and present AM studies discouraged an appeal of the robustness of the resulting TAMs. Alternatively, a tight correlation between the level of p-value threshold and accuracy of GEBVs validates that the AM was properly performed.

Overall, this study showed the usefulness of the BLUP across the traditional BLUP, AM and GS in self-pollinating plant. All the studies fundamentally pursue improving the selection efficiency of breeding in different manners: the traditional BLUP provides a manner to estimate BVs using the phenotypic data and pedigree, the AM facilitates the MAS by detecting the markers that are highly associated with trait, and the GS reduces a breeding cost and improves a selection response by applying the estimates of marker effect to genotyped individuals at an early growth stage. As a future study, the presently tested approaches might be worthy of comparison in terms of the selection efficiency of breeding.

7. Summary

During a past decade, an application of the BLUP became varied for the diverse purposes. In this study, the prospect of the BLUP was explored in terms of its multiple purposes: BV estimation, AM and GS for grain yield, scald severity and net blotch severity in a German barley cultivar collection.

Chapter 1 provides the review of previous literatures and objectives regarding this study. Chapter 2 introduces a new method for computing an NRM with pedigree of self-pollinating plant. The manner of computing an NRM in a self-pollinating plant was not revealed. To develop the manner, the architecture of an NRM in a population of self-pollinators was simulated. Based on a pattern of the simulation, three equations (2-6, 2-7 and 2-8) that define the relationship coefficient among self-pollinators were formulated, which can define the relationship coefficient among any two individuals within a plant pedigree with a consideration of the number of selfing generations. To the best of my knowledge, this is the first method for constructing an NRM using a plant pedigree. Based on the above equations, PopKin software tool was developed. The strengths of the PopKin are the abilities to (1) construct an NRM by using syntax that has the similar format to plant pedigree and (2) provide small-sized and accurate NRM. The PopKin was used for implementing the BLUP in Chapter 3.

Chapter 3 reports the estimation of BVs for grain yield, scald severity and net blotch severity using the BLUP with a panel of German spring barley cultivars. The population sizes were 92, 90 and 88 for grain yield, scald severity and net blotch severity, respectively. The phenotypic observations for all the traits were normal distributed. The BLUP is a modeling approach that integrates an NRM into MLM, which provides precise BVs by capturing the genetic performances of relatives of an individual. A precise measurement of an NRM is essential in performing the BLUP, for which PopKin software tool was employed. The BLUP was examined for the three traits using the provided accessions. The basic model can be denoted as follows:

$$y = Xb + Z_g g + Z_v v + Z_{g.v} g.v + e$$

where y = the vector of phenotype observations; b = the vector of constant grand means as fixed effect; g = the vector of breeding values as random effect; v = the vector of environment observations as random effect; g.v = the vector of genotype by environment interactions as random effect; e = the vector of residuals as random effect; X, Z_g, Z_v, $Z_{g.v}$ are the design matrices.

For all three traits, the BLUP models were smoothly fitted. The narrow-sense heritabilies resulted in 0.719 for grain yield, 0.491 for scald severity and 0.581for net blotch severity, which indicats that a quality of the data set in unbalanced trials was sufficient to use. As prediction accuracy, a correlation coefficient between TBVs and the BLUP estimates can be ideally used. However, because of no information about the TBVs, the MPs were alternatively used under an assumption that the MPs represent the TBVs. The estimates of prediction accuracy resulted in 0.854 for grain yield, 0.893 for scald severity and 0.940 for net blotch severity. The discrepancy between the resulting prediction accuracy and 1.0 represents that the BLUP procedure may improve a response to selection.

Chapter 4 presents the mapping of marker-trait associations for grain yield, scald severity and net blotch severity with the small numbers of German spring barley cultivar collection using 1181 DArT genotypes. As a statistical model, single marker regression based on the BLUP that embedded a marker-derived kinship matrix and a subpopulation matrix was adapted, which is called QK model (Q and K represent a subpopulation matrix and a kinship matrix, respectively). The mapping was modeled in two tracks: (1) marker-based empirical kinship and subpopulation matrix from discriminant analysis of principle component (KD model) and (2) marker-based empirical kinship and subpopulation matrix from using STRUCTURE software (KS model). The basic model can be denoted as follows:

$$y = Xb + Zu + Qv +Pm +e$$

where y = the vector of phenotypic observations; b = the vector of grand means; u = the vector of random genotype effects; v = the vector of fixed subpopulation effects; m = the vector of fixed marker effects; and e = the vector of random residual effects; X, Z, Q, P = the design matrices.

The population sizes for mapping were 45 for grain yield, 41 for scald severity and 40 for net blotch severity. These small population sizes increase a detection of spurious QTL by causing an upward bias. To avoid this risk, the Wald test at the significance level of p = 0.001 and cross-validations were conducted. For grain yield, bPb-8962 was mapped on chromosome 3H across the KD and KS models. According to previous studies (Laurie et al., 1993; Hayes et al., 1993; Thomas et al., 1995; Li et al., 2009; Xue et al., 2009), TAMs for grain yield are often linked to the genes controlling a plant height on chromosome 3H. To validate this, a plant height trait was additionally mapped. As a result, two UTAMs (bPb-5899 and bPb-0990) were detected, and the association of bPb-8962 with the plant height was not found, so the relatedness between the grain yield and plant height was not identified in this study. For scald severity, bPb-8445, bPb-6264, bPb-5458 and bPb-2018 were detected. Of these, only bPb-8445 was found across the KD and KS models, and bPb-6264, bPb-5458 and bPb-2018 were found only in the KS model. The bPb-8445, bPb-6264 and bPb-5458 were distributed on chromosomes 2H, 6H and 7H, respectively, whereas the bPb-2018 was unmapped. Zhan et al., (2008) described that the TAMs for the scald resistance were not found in chromosome 5H, which is supported by this study. For net blotch severity, bPb-1946 was identified across the KD and KS models. Previous studies extensively detected TAMs for net blotch resistance across seven chromosomes. However, no TAM was previously found in the proximity of bPb-1946. In this study, all the TAMs from the KS model were found from those from the KD model, which indicates that the use of different subpopulation in the QK model give not the same but similar resolution.

Chapter 5 shows an exploration of the BLUP for the GS for grain yield, scald severity and net blotch severity using the same barley collections and markers as used in Chapter 4. As a statistical model, the RR-BLUP was implemented. The basic model was denoted as follows:

$$y = WGu + e$$

where y = the vector of the phenotype observations; W = the design matrix that relates the lines to observations (y); G = the DArT marker genotype data; u = the vector of the unknown marker effects; e = the residual vector.

The main focus of Chapter 5 is to determine the conditions to improve the prediction accuracy of RR-BLUP. To achieve this, three experiments were conducted. In the first experiment, the impact of the size of marker set on the prediction accuracy was investigated. An interval of 10 % was given between 100 and 10 % in marker set size. The results shows a plateau of prediction accuracy in a range of 100-30 %. However, the rapid reductions of prediction accuracy are observed for all the traits below the level of 30 %, which suggests that RR-BLUP can be economically implemented by using the moderate size of marker set. In the second experiment, the impact of the size of training set on prediction accuracy was investigated. As a result, it was found that the prediction accuracy declines as the size of the training set becomes smaller. This observation shows that the improved prediction accuracy can be obtained with increasing sizes of training set. The results that have been found in the first and second experiments were in agreement with previous studies (Solberg et al., 2008; Lorenzana and Bernardo., 2009; Asoro et al., 2011; Zhao et al., 2012; Crossa et al., 2014). In the third experiment, an impact of controlling of the marker-trait association in a marker set on the prediction accuracy was investigated. To examine this, sets of markers were prepared by applying p-value thresholds with a reference of p-values of markers obtained from the AM (see Chapter 4). Lower p-value indicates that the marker and gene are stably linked, whereas greater p-value represents that the marker-gene linkages are fragile by a recombination. In this study, 10 levels of marker sets were formed with a p-value interval of 0.1 from 0.1 to 1.0. The estimates of prediction accuracy were shown to increase from 0.3226 to 0.7323 for grain yield, from 0.3534 to 0.5396 for scald severity and from 0.4431 to 0.8326 for net blotch severity, respectively, as the p-value threshold decreases. This pattern violates the fact that the prediction accuracy of RR-BLUP decreases as the size of marker set becomes smaller. In scald severity and net blotch severity, however, a rapid drop in the section of 0.2-0.1 was observed, which might be caused because the small number of markers offsets the advantage of the low p-value threshold. In conclusion, the ideal conditions for the RR-BLUP are thought to be a large size of marker set consisted of the markers selected at a low p-value threshold and a large training set.

8. References

Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. Mol Breeding 19:341-356

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

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

Atkin FC, Dieters MJ, Stringer JK (2009) Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. Theor Appl Genet 119:555-565

Bauer AM, Reetz TC, Léon J (2006) Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. Crop Sci 46:2685-2691

Bauer AM, Reetz TC, Léon J (2008) Prediction breeding values of spring barley accessions by using the singular value decomposition of genetic similarities. Plant Breeding 127:274-278

Bauer AM, Léon J (2008) Multiple-trait breeding values for parental selection in self-pollinating crops. Theor Appl Genet 116:235-242

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20-25

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

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

Bezant J, Laurie D, Pratchett N, Chojecki J, Kearsey M (1997) Mapping QTL controlling yield and yield components in a spring barley (*Hordeum vulgare* L.) cross using marker regression. Mol Breed 3:29-38

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Bromley CM, Van Vleck LD, Johnson BE, Smith OS (2000) Estimation of genetic variance in corn from F_1 performance with and without pedigree relationships among inbred lines. Crop Sci 40:651-655

Bulmer MG (1976) The effect of selection on genetic variability. Am Nat 105:201-211

Butler D, Cullis BR, Gilmour AR, Gogel BJ (2009) Analysis of mixed models for S-language enviromnents: ASReml-R Reference Manual. Queensland DPI, Brisbane, Australia. URL http://www.vsni.co.uk/resources/doc/asreml-R.pdf.

Caballero A, Toro MA (2002) Analysis of genetic diversity for the management of conserved subdivided populations. Conserv Genet 3:289-299

Chang HL, Fernando RL, Grossman M (1991) On the principle underlying the tabular method to compute coancestry. Theor Appl Genet. 81:223-238

Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O'Sullivan DM (2008) Association mapping of partitioning loci in barley. BMC Genetics, 9:16

Comadran J, Russel JR, Booth A, Pswarayi A, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, van Eeuwijk FA, Thomas WTB, Romagosa I (2011) Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. Theor Appl Genet 122:1363-1373

Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2006) Modeling genotype x environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46:1722-1733

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

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021-1031

Durel CE, Laurens F, Fouillet A, Lespinasse Y (1998) Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. Theor Appl Genet 96:1077-1085

Emik LO, Terrill CE (1949) Systematic procedures for calculating inbreeding coefficients. J Hered 40:51-55

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4:250-255

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, Essex, UK.

Fetch TG Jr, Steffenson BJ (1994) Identification of *Cochliobolus sativus* isolates expressing differential virulence on two-row barley genotypes from North Dakota. Can J Plant Pathol 16:202-206

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374

Gao H, Williamson S, Bustamante CD (2007) A markov chain monte carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635-1651

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323-330

Graner A, Foroughi-Wehr B, Tekauz A (1996) RFLP mapping of a gene in barley conferring resistance to net blotch (*Pyrenophora teres*). Euphytica 91:229-234

Grewal TS, Rossnagel BG, Scoles GJ (2008) The utility of molecular markers for barley net blotch resistance across geographic regions. Crop Sci 48:2321-2333

Gutiérrez JP, Royo LJ, Alvarez I, Goyache F (2005) Molkin v2.0: A computer program for genetic analysis of populations using molecular coancestry information. J Hered 96:718-721

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa state university press, Ames, IA.

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

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618-620

Haseneyer G, Stracke S, Paul C, Einfeldt C, Broda A, Piepho HP, Graner A, Geiger HH (2010) Population structure and phenotypic variation of a spring barley world collection set up for association studies. Plant Breeding 129:271-279

Hayes PM, Liu BH, Knapp SJ, Chen F, Jones B, Blake T, Franckowiak J, Rasmusson D, Sorrells M, Ullrich SE, Wesenberg D, Kleinhofs A (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm. Theor Appl Genet 87:392-401

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, and Goddard ME (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol, 41:51

Hayes BJ, Cogan NOI, Pembleton LW, Goddard ME, Wang J, Spangenberg GC, Forster JW (2013) Prospects for genomic selection in forage plant species. Plant Breeding 132:133-143

Hearnden PR, Eckermann PJ, McMichael GL, Hayden MJ, Eglinton JK, Chalmers KJ (2007) A genetic map of 1,000 SSR and DArT markers in a wide barley cross. Theor Appl Genet 115:383-391

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

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69-83

Henderson CR, Quaas RL (1976) Multiple trait evaluation using relatives' records. J Anim Sci 43:1188-1197

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961-967

Jensen J, Backes G, Skinnes H, Giese H (2002) Quantitative trait loci for scald resistance in barley localized by a non-interval mapping procedure. Plant Breeding 121:124-128

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11:94

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

Kraft T, Hansen M, Nilsson NO (2000) Linkage disequilibrium and fingerprinting in sugar beet. Theor Appl Genet 101:323-326

Laurie DA, Pratchett N, Romero C, Simpson E, Snape JW (1993) Assignment of the *denso* dwarfing gene to the long arm of chromosome 3 (3H) of barley by use of RFLP markers. Plant Breeding 111:198-203

LfL Pflanzenbau (2013) http://www.lfl.bayern.de/ipz/gerste/09740/linkurl_0_9.pdf. Visited on 21st March 2013.

Li HB, Zhou MX, Liu CJ (2009) A major QTL conferring crown rot resistance in barley and its association with plant height. Theor Appl Genet 118:903-910

Looseley ME, Newton AC, Atkins SD, Fitt BDL, Fraaije BA, Thomas WTB, Keith R, Macaulay M, Lynott J, Harrap D (2012) Genetic basis of control of *Rhynchosporium secalis* infection and symptom expression in barley. Euphytica 184:47-56

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

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genetics 152:1753-1766

Malysheva-Otto LV, Ganal MW, Röder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genetics 7:6

Massman J, Cooper B, Horsley R, Neate S, Dill-Macky R, Chao S, Dong Y, Schwarz P, Muehlbauer GJ, Smith KP (2011) Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. Mol Breeding 27:439-454

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

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

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Amin Breed Genet 124:342-355

Nakaya A, Isobe SN (2012) Will genomic selection be a practical method for plant breeding? Ann Bot 110:1303-1316

Nielsen HM, Sonesson AK, Yazdi H, Meuwissen THE (2009) Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. Aquaculture 289:259-264

Nunes JAR, Ramalho MAP, Ferreira DF (2008) Inclusion of genetic relationship information in the pedigree selection method using mixed models. Genet Mol Biol 31:73-78

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modeling of additive and non-additive genetic line effects in single field trials. Theor Appl Genet 113:809-819

Panter DM, Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean: Ⅰ. Choosing parents. Crop Sci 35:397-405

Panter DM, Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean: Ⅱ. Selection of superior crosses from a limited number of yield trials. Crop Sci 35:405-410

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biology, 12:16

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545-554

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM (2006) GenStat for Wondows (9th Edition) introduction. VSN International, Hemel Hempstead.

Piepho HP, Moehring J (2005) Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Sci 45:1151-1159

Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881-1888

Piepho HP, Möhring J, Melchinger AE (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209-228

Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49:1165-1176

Plum M (1954) Computation of inbreeding and relationship coefficients. J Hered 45:92-94

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Purba AR, Flori A, Baudouin L, Hamon S (2001) Prediction of oil palm (*Elaeis guineensis*, Jacq.) agronomic performances using the best linear unbiased predictor (BLUP). Theor Appl Genet 102: 787-792

R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBS 3-900051-07-0, URL http://www.R-project.org/.

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174-180

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler Ⅳ ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479-11484

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

Robinson GK (1991) That BLUP is a good thing: The estimation of random effects. Statistical science 6:15-32

Rode J, Ahlemeyer J, Friedt W, Ordon F (2012) Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulare* L.). Mol Breeding 30:831-843

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656-18661

Roy JK, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ (2010) Association mapping of spot blotch resistance in wild barley. Mol Breeding 26:243-256

Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells M (2012) Evaluation of genomic prediction methods for fusarium head bright resistance in wheat. The Plant Genome 5:51-61

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123: 218-223

Schulz-Streeck T, Piepho HP (2010) Genome-wide selection by mixed model ridge regression and extensions based on geostatistical models. BMC Proceedings, 4 (Suppl. 1):S8

Searle SR, Casella G, McCulloch CE (1992) Variance components. John Wiley and Sons, New York.

Shi C, Navabi A, Yu K (2011) Association mapping of common bacterial blight resistance QTL in Ontario bean breeding population. BMC Plant Biology, 11:52

Soh AC (1994) Ranking parents by best linear unbiased prediction (BLUP) breeding values in oil palm. Euphytica 76:13-21

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. J ANIM SCI 86:2447-2454

Steffenson BJ, Hayes PM, Kleinhofs A (1996) Genetics of seedling and adult plant resistance to net blotch (*Pyrenophora teeres* f. *teres*) and spot blotch (*Cochliobolus sativus*) in barley. Theor Appl Genet 92:552-558

Stich B, Möhring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. Genetics 178:1745-1754

Stich B, Melchinger AE (2009) Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics, 10:94

Tavernier A (1988) Advantages of BLUP animal model for breeding value estimation in horses. Livest Prod Sci 20:149-160

Thinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW, Bjørnstad A, Howarth CJ, Jannink JL, Anderson JM, Rossnagel BG, Stuthman DD, Sorrells ME, Jackson EW, Tuvesson S, Kolb FL, Olsson O, Federizzi LC, Carson ML, Ohm HW, Molnar SJ, Scoles GJ, Eckstein PE, Bonman JM, Ceplitis A, Langdon T (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics, 10:39

Thomas WTB, Powell W, Waugh R, Chalmers KJ, Barua UM, Jack P, Lea V, Forster BP, Swanston JS, Ellis RP, Hanson PR, Lance RCM (1995) Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.). Theor Appl Genet 91:1037-1047

Verrier E, Colleau JJ, Foulley JL (1993) Long-term effects of selection based on the animal model BLUP in a finite population. Theor Appl Genet 87:446-454

Viana JMS, Almeida ÍF de, Resende MDV de, Faria VR, Silva FF (2009) BLUP for genetic evaluation of plants in non-inbred families of annual crops. Euphytica 174:31-39

Viana JMS, Sobreira FM, Resende MDV de, Faria VR (2010) Multi-trait BLUP in half-sib selection of annual crops. Plant Breeding 129:599-604

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Wang H, Smith KP, Combs E, Blake T, Horsley RD, Muehlbauer GJ (2012) Effect of population size and unbalanced data sets of QTL detection using genome-wide association mapping in barley breeding germplasm. Theor Appl Genet 124:111-124

Wang M, Jiang N, Jia T, Leach L, Cockram J, Waugh R, Ramsay L, Thomas B, Luo Z (2012) Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. Theor Appl Genet 124:233-246

Wilcoxon RD, Rasmusson, DC, Miles MR (1990) Development of barley resistant to spot blotch and genetics of resistance. Plant Dis 74:207-210

Xu S (2003) Estimating polygenic effects using markers of the entire genome. Genetics 163: 789-801

Xue D, Huang Y, Zhang X, Wei K, Westcott S, Li C, Chen M, Zhang G, Lance R (2009) Identification of QTLs associated with salinity tolerance at late growth stage in barley. Euphytica 169:187-196

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155-160

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208

Zhan J, Fitt BDL, Pinnschmidt HO, Oxley SJP, Newton AC (2008) Resistance, epidemiology and sustainable management of *Rhynchosporium secalis* populations on barley. Plant Pathology 57:1-14

Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. Theor Appl Genet 119:43-52

Zhang Z, Liu J, Ding X, Bijma P, Koning DJ de, Zhang Q (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PLoS ONE 5:e12648

Zhao JH (2013) gap: Genetic analysis package. R package version 1.1-9

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An *Arabidopsis* example of association mapping in structured samples. PLoS Genet 3:e4

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769-776

Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. Genetics 182: 355-364

9. List of Tables

10. List of Figures

11. Abbreviations

A matrix	Numerator relationship matrix
AM	Association mapping
BLUE	Best linear unbiased estimation
BLUP	Best linear unbiased prediction
BV	Best value
cM	Centimorgan
DAPC	Discriminant analysis of principal components
DArT	Diversity array technology
GEBV	Genetic estimated breeding value
GS	Genomic selection
H	Broad-sense heritability
h^2	Narrow-sense heritability
K	Kinship matrix
KD	Kinship plus DAPC model
KM	Kinship matrix
KS	Kinship plus STRUCTURE model
LD	Linkage Disequilibrium
LE	Linkage Equilibrium
LOESS	Local weighted scatterplot smoothing
LS	Least square
LSV	Landessortenversuche
MCMC	Markov Chain Monte Carlo
MLM	Mixed linear model
MP	Mean phenotype
MPV	Mid-parent value
NRM	Numerator relationship matrix
PCA	Principle component analysis
PopKin	Population kinship calculator
Q	Population structure matrix
QK	Population plus kinship model
QTL	Quantitative trait loci
r^2	Squared correlation
REML	Restricted maximum likelihood
RR-BLUP	Ridge regression best linear unbiased prediction
SE	Standard error
TAM	Trait-associated marker
TBV	True breeding value
UTAM	Unmapped trait-associated marker

12. Appendices

Appendix Ⅰ. The list of accessions used for BLUP procedure. Totally 94 spring barley cultivars released from Germany are included.

No	Accession	Seasonal type (S: spring / W: winter)	Number of ear	Pedigree
1	ADONIS	S	2	Wren / Trinity
2	ALEXIS	S	2	BREUN-1622 / TRUMPF
3	ALONDRA	S	2	TRUMPF / NORDAL // KORAL
4	ANNABELL	S	2	KRONA / 90014-DH
5	APEX	S	2	VOLLA / L-100 // JULIA /// CEBECO-6721 ////ARAMIR
6	ASPEN	S	2	VINTAGE / CHARIOT
7	AURIGA	S	2	VISKOSA / KRONA // ANNABELL
8	BACCARA	S	2	RS-112 / ARAMIR // KORALLE /// GOLF //// CANDICE ///// GOLF
9	BARKE	S	2	LIBELLE / ALEXIS
10	BARONESSE	S	2	343-6 / V-34-6 // J-427 /// ORIOL / LBV-6153-P40
11	BELLA	S	2	HOCKEY / APEX
12	BESSI	S	2	GOLF / AC-77-1798-1
13	BIRTE	S	2	GOLDIE / CORK
14	BITRANA	S	2	SALOME / HVS-18709-78
15	BRAEMAR	S	2	NFC-5563 / NFC-94-20
16	BRENDA	S	2	NEBI / 11827-80 // GIMPEL
17	BRITTA	S	2	KRONA / HADM.59789-85
18	CAMBRINUS	S	2	BALDER / STRENG-FRANKEN-Ⅲ
19	CAMINANT	S	2	ANT-28-484 / BLENHEIM
20	CELLAR	S	2	NFC-94-20 / CORK // NFC-94-11
21	CHALICE	S	2	COOPER / NFC-514-5 // CHARIOT
22	CHANTAL	S	2	CHARIOT / KRONA
23	CHARIOT	S	2	DERA // CARNIVAL / ATEM
24	CHALOTT	S	2	-

(Continued)

No	Accession	Seasonal type (S: spring / W: winter)	Number of ear	Pedigree
25	CHERI	S	2	TRUMPF // MEDUSA / DIAMANT
26	CITY	S	2	VISTA / THEMIS
27	CLAUDINE	S	2	ROMI / ROLAND
28	CORA	S	2	ROMI-ABED / ROLAND
29	DANOR	-	-	-
30	DANUTA	S	2	SALOME / MARESI // 90014-DH
31	DERKADO	S	2	LADA / SALOME
32	DIAMALTA	S	2	10100-80 / 45465-78 // 21275-82
33	DITTA	S	2	APEX / 76-1754-6
34	ESCADA	S	2	NRPB-87-3277-B / ALEXIS
35	EUNOVA	S	2	H-53-D / CF-79
36	EXTRACT	S	2	CASK / CHARIOT // AMBER
37	FERMENT	S	2	NFC-327-10 / COOPER // CHARIOT
38	FORUM	S	2	H-387-75 / HORPATSI-KETSCOROS // 044-78
39	GOLF	S	2	ARMELLE / LUD // LUKE
40	HALLA	S	2	STEFFI / GERLINDE // 243-4 / SALOME
41	HANKA	S	2	HADMERSLEBEN-59473-85 / HADMERSLEBEN-96677-87
42	HAVANNA	S	2	BREUN-3556-A / BREUN-3192-F
43	HENDRIX	S	2	MADRAS / S90772
44	HENNI	S	2	BARONESSE / 84160.1.3.3
45	JACINTA	S	2	ALEXIS / MELTAN // CANUT
46	JERSEY	-	-	APEX / ALEXIS
47	JULIA	S	2	BULGARISCHE-468 / ERFA // MASTO
48	KATHARINA	S	2	HVS-1129-79 / 1057-81 // DERA
49	KRONA	S	2	NEBI / TRUMPF // UNION /// GIMPEL
50	LENKA	S	2	HVS-5013-74 / Q-496-72

(Continued)

No	Accession	Seasonal type (S: spring / W: winter)	Number of ear	Pedigree
51	MADEIRA	S	2	HADMERSLEBEN-12939-82 / HADMERSLEBEN-63787-83
52	MADONNA	S	2	MARINA / KRONA
53	MADRAS	S	2	R-62761 / 4.2606 // ALEXIS
54	MARESI	S	2	CEBECO-6801 / GB-1605 // HA-46459-68
55	MARINA	S	2	ASTRA / LICHTIS DN
56	MARNIE	S	2	HAVANNA // PRISMA / BR4714A
57	MAUD	S	2	VEB-813 / FLARE
58	MELTAN	S	2	D-80-20 / TELLUS-MMMDDN
59	MENTOR	S	2	KARA / ARIEL
60	MINNA	S	2	TRUMPF / CARINA // BREUN-2357-B-33 / BREUN-853-B-12
61	NANCY	S	2	INGRID-M / ANSGAR // ARAMIR /// YRJAR
62	NERUDA	S	2	NOMAD / GOLF // ALEXIS /// CHARIOT
63	NEVADA	S	2	DELTA / TRUMPF
64	NOMAD	S	2	KYM / TRUMPF
65	OLGA	S	2	BENEDICTE / KORU
66	OPTIC	S	2	CHAD // CORNICHE / FORCE
67	ORTHEGA	S	2	CEBECO-7931 / POMPADOUR // S.77323 / GOLF
68	OTIRA	S	2	BARTOK / SJ-930331
69	OTIS	S	2	ST.08020 / EUROPA // ATEM
70	PASADENA	S	2	MARINA / KRONA
71	PEGGY	S	2	RS-112 / ARAMIR // KORALLE /// GOLF / CANDICE //// GOLF
72	PEWTER	S	2	NFC-94-20 / NFC-94-11
73	POMPADUR	S	2	FD-0192 / PATTY
74	PONGO	S	2	PL-1587-87 / 88008
75	PRESTIGE	S	2	CORK / CHARIOT
76	PROLOG	S	2	ETNA / MELTAN

(Continued)

No	Accession	Seasonal type (S: spring / W: winter)	Number of ear	Pedigree
77	RIA	S	2	HADMERSLEBEN-55648-85 / HADMERSLEBEN-96677-87
78	RICARDA	S	2	NRPB-83-1083 / CHARIOT
79	RIVIERA	S	2	STANZA / CEBECO-8331
80	ROXANA	S	2	BR.3556-A / KORINNA // ALEXIS
81	SALLY	S	2	RS-112 / ARAMIR // KORALLE /// GOLF //// CANDICE ///// GOLF
82	SALOON	S	2	CORK / HIND
83	SCARLETT	S	2	AMAZONE / BREUN-2730-E // KYM
84	SIGRID	S	2	FORESTER / NAIRN // CARNIVAL
85	SISSY	S	2	FRANKENGOLD / MONA // TRUMPF
86	STEFFI	S	2	STAMM-101 / ARAMIR // STAMM-210
87	TEO	S	2	CLARET / KYM
88	THERESA	S	6	FRANKA / 943-77 // CORONA
89	THURINGIA	S	2	STEFFI / GERLINDE // 243-4 /SALOME
90	TOLAR	S	2	HE-4710 / HVS-78267-83
91	URSA	S	2	THURINGIA / HANKA // ANNABELL
92	VISKOSA	S	2	90014-DH // MARESI / SALOME

Appendix Ⅱ. Comparison of mean phenotypes and BLUP estimates for grain yield, scald severity and net blotch severity

No.	Parent	Grain yield				Scald severity				Net blotch severity			
		Mean phenotype		Breeding value		Mean phenotype		Breeding value		Mean phenotype		Breeding value	
		Mean	SE	Prediction	SE	Mean	SE	Prediction	SE	Mean	SE	Prediction	SE
1	ADONIS	666.143	116.284	654.805	28.227	2.833	1.581	3.075	0.222	3.500	1.633	3.176	0.244
2	ALEXIS	626.472	177.046	626.449	7.691	3.431	1.902	3.292	0.077	2.931	1.659	3.020	0.081
3	ALONDRA	605.192	163.876	618.850	11.329	3.194	1.920	3.060	0.111	2.494	1.435	2.621	0.131
4	ANNABELL	687.463	169.881	692.410	10.684	3.606	1.705	3.503	0.104	3.236	1.509	3.207	0.111
5	APEX	609.717	151.769	620.192	13.889	3.207	1.668	3.091	0.146	2.564	1.284	2.683	0.147
6	ASPEN	670.986	177.299	673.068	11.759	2.991	1.465	2.933	0.114	3.326	1.808	3.256	0.121
7	AURIGA	692.533	146.013	689.909	20.505	3.250	1.842	3.320	0.170	3.216	1.813	3.168	0.181
8	BACCARA	631.066	141.643	631.489	11.968	2.794	1.302	2.721	0.118	3.014	1.535	2.951	0.127
9	BARKE	672.027	180.195	664.397	8.295	2.845	1.325	2.868	0.082	3.196	1.652	3.198	0.085
10	BARONESSE	661.884	149.225	662.288	8.594	3.049	1.751	2.988	0.090	2.733	1.408	2.740	0.092
11	BELLA	698.864	154.251	633.941	19.037	2.500	1.403	2.873	0.183	3.241	1.770	3.153	0.186
12	BESSI	649.591	156.102	653.631	12.893	3.273	1.772	3.024	0.133	2.524	1.380	2.648	0.136
13	BIRTE	742.069	208.877	699.712	21.680	2.413	0.909	2.760	0.192	3.405	1.768	3.266	0.206
14	BITRANA	667.910	172.348	648.332	16.443	3.627	2.211	3.427	0.154	3.566	1.882	3.440	0.167
15	BRAEMAR	677.706	161.464	668.996	26.631	3.353	1.631	3.154	0.215	3.438	1.917	3.236	0.230
16	BRENDA	679.019	194.899	671.504	11.171	3.214	1.690	3.082	0.110	3.127	1.752	3.145	0.121
17	BRITTA	651.807	167.724	662.554	15.482	3.350	1.621	3.188	0.143	3.291	1.538	3.220	0.153
18	CAMBRINUS	567.333	149.638	637.186	35.588	3.000	0.632	3.024	0.262	4.167	2.229	3.150	0.289
19	CAMINANT	653.189	175.161	633.283	19.612	2.633	1.235	2.848	0.183	3.587	1.562	3.402	0.191
20	CELLAR	664.176	147.143	667.679	25.735	3.206	1.473	3.059	0.209	3.313	1.615	3.213	0.224
21	CHALICE	631.241	170.181	649.013	15.666	2.986	1.362	2.933	0.144	3.020	1.341	2.986	0.156
22	CHANTAL	639.087	167.915	650.104	13.624	3.179	1.520	3.064	0.124	3.646	1.900	3.485	0.135

(Continued)

No.	Parent	Grain yield				Scald severity				Net blotch severity			
		Mean phenotype		Breeding value		Mean phenotype		Breeding value		Mean phenotype		Breeding value	
		Mean	SE	Prediction	SE	Mean	SE	Prediction	SE	Mean	SE	Prediction	SE
23	CHARIOT	638.207	189.705	643.322	9.341	3.320	1.846	3.128	0.089	3.004	1.712	3.121	0.099
24	CHARLOTT	645.616	167.969	657.366	11.865	3.239	1.358	3.093	0.112	3.127	1.550	3.084	0.124
25	CHERI	589.272	97.433	587.681	18.125	2.314	1.198	2.600	0.166	3.196	1.510	3.094	0.187
26	CITY	660.905	174.671	650.875	15.524	3.363	2.076	3.141	0.163	3.094	1.485	3.035	0.165
27	CLAUDINE	606.713	153.665	619.828	14.555	2.974	1.385	2.904	0.163	3.442	1.851	3.319	0.162
28	CORA	506.067	137.817	583.441	26.559	2.600	0.843	2.900	0.245	2.750	1.342	2.972	0.250
29	DANOR	759.941	179.741	683.477	25.024	2.647	1.115	2.979	0.187	2.167	1.030	2.924	0.212
30	DANUTA	680.505	160.198	690.669	13.873	2.819	1.443	2.933	0.133	3.074	1.376	3.053	0.136
31	DERKADO	733.764	146.373	699.203	18.851	3.404	2.195	3.300	0.183	2.936	1.420	3.033	0.200
32	DIAMALTA	625.131	183.632	627.386	12.873	2.548	1.482	2.649	0.128	2.460	1.414	2.613	0.143
33	DITTA	672.176	174.820	666.024	12.848	2.837	1.799	2.958	0.124	2.175	1.182	2.363	0.135
34	ESCADA	690.058	186.518	671.196	15.518	2.809	1.274	2.998	0.150	3.286	1.878	3.218	0.152
35	EUNOVA	627.099	135.171	644.644	15.930	2.796	1.353	2.904	0.152	3.131	1.429	3.049	0.170
36	EXTRACT	638.777	182.472	653.048	10.594	3.209	1.379	3.122	0.106	3.554	1.565	3.468	0.110
37	FERMENT	702.250	190.283	654.534	32.927	2.250	0.500	3.021	0.237	2.750	0.957	3.044	0.265
38	FORUM	615.125	40.923	653.341	35.266	3.375	1.685	3.064	0.262	4.000	1.773	3.250	0.288
39	GOLF	679.500	33.234	640.648	20.452	2.000	0.000	2.729	0.154	2.000	0.000	2.841	0.171
40	HALLA	667.836	190.030	652.506	12.451	3.081	1.580	2.945	0.122	2.769	1.418	2.777	0.133
41	HANKA	660.192	175.136	652.561	10.109	3.334	1.684	3.217	0.101	2.947	1.436	2.921	0.102
42	HAVANNA	704.613	191.100	679.025	21.127	2.333	1.243	2.734	0.195	2.727	1.301	2.834	0.208
43	HENDRIX	687.850	151.164	670.455	24.769	3.382	2.015	3.352	0.202	3.914	1.821	3.470	0.216

(Continued)

No.	Parent	Grain yield				Scald severity				Net blotch severity			
		Mean		Breeding value		Mean		Breeding value		Mean		Breeding value	
		Mean	SE	Prediction	SE	Mean	SE	Prediction	SE	Mean	SE	Prediction	SE
44	HENNI	664.176	171.559	667.490	9.275	3.708	1.971	3.478	0.094	3.067	1.532	3.036	0.100
45	JACINTA	643.689	116.985	629.024	22.642	2.824	1.167	2.974	0.196	2.667	1.028	2.854	0.215
46	JERSEY	732.813	217.358	650.200	24.704	3.923	1.038	3.257	0.193	2.875	1.553	2.865	0.218
47	JULIA	887.478	207.371	789.913	20.157	3.475	1.851	3.391	0.185	3.972	1.610	3.723	0.185
48	KATHARINA	642.815	173.220	625.295	15.857	2.480	1.586	2.641	0.153	2.505	1.259	2.694	0.172
49	KRONA	658.087	177.742	656.522	7.621	3.241	1.722	3.163	0.075	3.096	1.617	3.129	0.080
50	LENKA	623.455	81.024	627.807	29.009	3.250	1.888	3.164	0.232	2.938	1.569	2.988	0.257
51	MADEIRA	649.080	170.935	650.324	15.233	3.223	1.423	3.127	0.153	3.333	1.597	3.285	0.160
52	MADONNA	675.968	187.178	670.181	14.581	3.134	1.501	3.299	0.137	3.095	1.736	3.220	0.141
53	MADRAS	667.928	176.492	665.827	13.568	3.522	1.591	3.469	0.136	3.189	1.649	3.154	0.139
54	MARESI	645.994	172.866	641.694	7.366	3.334	1.805	3.284	0.073	3.049	1.597	3.095	0.077
55	MARINA	662.195	170.179	663.039	11.174	3.681	2.121	3.574	0.107	3.181	1.751	3.213	0.119
56	MARNIE	717.111	204.220	678.075	31.975	-	-	-		1.000	0.000	2.805	0.283
57	MAUD	497.000	7.071	642.247	37.887	1.000	0.000	2.839	0.269	-	-	-	-
58	MELTAN	668.571	172.000	660.757	10.501	2.839	1.529	2.817	0.112	2.740	1.418	2.847	0.116
59	MENTOR	745.721	177.117	691.850	16.233	2.991	1.360	3.108	0.153	3.294	1.739	3.217	0.159
60	MINNA	651.980	175.788	644.974	16.710	2.971	1.604	3.036	0.161	2.549	1.361	2.679	0.184
61	NANCY	534.250	175.862	596.982	26.150	2.083	0.900	2.818	0.242	1.833	0.857	2.541	0.246
62	NERUDA	678.788	165.850	677.383	13.796	3.094	1.382	3.103	0.131	3.162	1.579	3.044	0.138
63	NEVADA	640.000	13.229	611.062	34.430	2.333	0.577	2.869	0.247	-	-	-	-
64	NOMAD	606.500	89.803	621.254	33.214	-	-	-	-	-	-	-	-

(Continued)

No.	Parent	Grain yield				Scald severity				Net blotch severity			
		Mean		Breeding value		Mean		Breeding value		Mean		Breeding value	
		Mean	SE	Prediction	SE	Mean	SE	Prediction	SE	Mean	SE	Prediction	SE
65	OLGA	621.521	176.342	621.664	18.422	2.727	1.442	2.883	0.204	2.462	1.527	2.651	0.190
66	OPTIC	669.930	172.763	668.650	21.775	3.750	1.320	3.291	0.209	3.415	1.936	3.232	0.209
67	ORTHEGA	692.111	161.927	675.614	9.795	2.982	1.523	3.000	0.104	3.008	1.443	2.965	0.108
68	OTIRA	629.101	162.658	645.609	15.537	3.485	1.836	3.338	0.151	3.610	1.569	3.420	0.166
69	OTIS	663.621	166.666	664.820	12.609	3.013	1.696	2.966	0.124	2.449	1.470	2.589	0.138
70	PASADENA	671.429	180.053	681.411	9.438	3.411	1.607	3.355	0.093	3.104	1.346	3.111	0.097
71	PEGGY	622.098	137.666	627.828	11.751	2.794	1.258	2.764	0.115	2.806	1.439	2.839	0.125
72	PEWTER	692.906	189.876	682.405	18.353	2.115	1.093	2.551	0.170	3.387	1.881	3.312	0.178
73	POMPADUR	523.767	167.841	596.583	26.431	2.400	1.430	2.932	0.242	2.438	1.153	2.801	0.248
74	PONGO	686.921	191.583	676.077	24.711	3.548	1.338	3.192	0.213	2.619	1.284	2.862	0.245
75	PRESTIGE	704.490	185.525	681.490	17.200	2.550	1.359	2.824	0.153	3.416	1.765	3.280	0.161
76	PROLOG	594.809	138.976	626.824	18.486	3.038	1.298	2.960	0.176	3.750	1.557	3.294	0.194
77	RIA	655.261	176.768	659.231	12.173	2.774	1.171	2.732	0.120	3.151	1.258	3.103	0.130
78	RICARDA	651.590	171.550	658.918	11.915	3.100	1.440	3.031	0.116	3.497	1.623	3.419	0.126
79	RIVIERA	655.143	170.408	663.959	10.262	3.196	1.481	3.104	0.103	2.929	1.440	2.914	0.108
80	ROXANA	520.000	77.527	620.614	32.335	4.000	0.894	3.233	0.235	2.907	1.347	2.888	0.148
81	SALLY	639.978	153.502	645.752	13.983	2.667	1.270	2.652	0.131	-	-	-	-
82	SALOON	624.960	167.247	659.701	17.800	2.863	1.412	2.847	0.172	2.933	1.212	2.990	0.179
83	SCARLETT	641.498	168.101	643.849	7.672	3.145	1.659	3.029	0.077	2.884	1.548	2.882	0.081
84	SIGRID	690.187	172.606	660.395	11.863	2.672	1.309	2.729	0.123	2.633	1.438	2.699	0.131
85	SISSY	545.682	105.988	551.615	13.343	2.560	1.461	2.614	0.140	2.832	1.342	2.876	0.154

(Continued)

No.	Parent	Grain yield				Scald severity				Net blotch severity			
		Mean		Breeding value		Mean		Breeding value		Mean		Breeding value	
		Mean	SE	Prediction	SE	Mean	SE	Prediction	SE	Mean	SE	Prediction	SE
86	STEFFI	607.726	155.760	601.506	13.746	2.957	2.002	2.959	0.143	2.548	1.456	2.640	0.155
87	TEO	625.990	148.933	633.907	17.332	2.828	1.454	2.917	0.171	2.853	1.259	2.891	0.168
88	THERESA	931.940	215.400	821.564	19.060	3.028	1.621	3.193	0.180	3.449	1.823	3.332	0.180
89	THURINGIA	665.760	182.926	657.239	10.668	2.987	1.555	2.877	0.106	2.817	1.448	2.845	0.113
90	TOLAR	631.500	47.713	648.175	34.603	2.750	1.488	2.938	0.258	3.625	1.847	3.136	0.283
91	URSA	689.047	139.307	682.992	22.120	2.912	1.865	3.186	0.187	2.657	1.235	2.874	0.196
92	VISKOSA	696.060	175.860	704.521	12.317	3.223	1.500	3.114	0.121	3.132	1.620	3.105	0.126

Appendix Ⅲ. The list of accessions used for association mapping (Chapter 4) and genome-wide breeding value estimation (Chapter 5). Totally 45 accessions, released from Germany, are included, all of which have the genotype and phenotype data.

No.	Accession	Seasonal type	Number of ear	Pedigree
1	ADONIS	S	2	WREN / TRINITY
2	ALEXIS	S	2	BREUN-1622 / TRUMPF
3	ALONDRA	S	2	TRUMPF / NORDAL // KORAL
4	ANNABELL	S	2	90014-DH / KRONA
5	APEX	S	2	L-100 / VOLLA // JULIA /// CEBECO-6721 //// ARAMIR
6	AURIGA	S	2	VISKOSA / KRONA // ANNABELL
7	BARKE	S	2	LIBELLE / ALEXIS
8	BARONESSE	S	2	343-6 / V-34-6 // J-427 /// ORIOL / LBW-6153-P-40
9	BELLA	S	2	HOCKEY / APEX
10	BESSI	S	2	GOLF / AC-77-1798-1
11	BITRANA	S	2	SALOME / HVS-18709-78
12	BRENDA	S	2	NEBI / 11827-80 // GIMPEL
13	CAMINANT	S	2	ANT-28-484 / BLENHEIM
14	CELLAR	S	2	NFC-94-20 / CORK // NFC-94-11
15	CITY	S	2	VISTA / THEMIS
16	CORA	S	2	ROMI / ROLAND
17	DERKADO	S	2	LADA / SALOME
18	DIAMALTA	S	2	10100-80 / 45465-78 // 21275-82
19	DITTA	S	2	APEX / 76-1754-6
20	ESCADA	S	2	NRPB-87-3277-B / ALEXIS
21	EUNOVA	S	2	H-53-D / CF-79
22	EXTRACT	S	2	CASK / CHARIOT
23	GOLF	S	2	ARMELLE / LUD // LUKE
24	HANKA	S	2	HADMERSLEBEN-59473-85 / HADMERSLEBEN-97777-87
25	KATHARINA	S	2	HVS-1129-79 / 1057-81 // DERA
26	KORINNA	S	2	62397-73 / 64045-74 // DORINA
27	KRONA	S	2	NEBI / TRUMPF // UNION /// GIMPEL
28	LARISSA	S	2	1097-77 / 23807-78
29	LENKA	S	2	HVS-5013-74 / Q-496-72
30	MARESI	S	2	CEBECO-6801 / GB-1605 // HA-46459-68
31	MARINA	S	2	ASTRA / LICHTIS DN
32	MARNIE	S	2	HAVANNA // PRISMA / BR4714A

(Continued)

No.	Accession	Seasonal type	Number of ear	Pedigree
33	MELTAN	S	2	D-80-20 / TELLUS-MMMDDN
34	NANCY	S	2	INGRID-M / ANSGAR // ARAMIR /// YRJAR
35	NEVADA	S	2	DELTA / TRUMPF
36	OLGA	S	2	BENEDICTE / KORU
37	PASADENA	S	2	MARINA / KRONA
38	POMPADUR	S	2	FD-0192 / PATTY
39	RIA	S	2	HADMERSLEBEN-55648-85 / HADMERSLEBEN-96677-87
40	SCARLETT	S	2	AMAZONE / BREUN-2730-E
41	SISSY	S	2	FRANKENGOLD / MONA // TRUMPF
42	STEFFI	S	2	STAMM-101 / ARAMIR // STAMM-210
43	TEO	S	2	CLARET / KYM
44	THURINGIA	S	2	STEFFI / GERLINDE // SALOME
45	URSA	S	2	THURINGIA / HANKA // ANNABELL