

Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.)

Benjamin Stich · Albrecht E. Melchinger ·
Martin Heckenberger · Jens Möhring ·
Axel Schechert · Hans-Peter Piepho

Received: 26 November 2007 / Accepted: 23 July 2008 / Published online: 22 August 2008
© Springer-Verlag 2008

Abstract Association mapping in multiple segregating populations (AMMSP) combines high power to detect QTL in genome-wide approaches of linkage mapping with high mapping resolution of association mapping. The main objectives of this study were to (1) examine the applicability of AMMSP in a plant breeding context based on segregating populations of various size of sugar beet (*Beta vulgaris* L.), (2) compare different biometric approaches for AMMSP, and (3) detect markers with significant main effect across locations for nine traits in sugar beet. We used 768 F_n ($n = 2, 3, 4$) sugar beet genotypes which were randomly derived from 19 crosses among diploid elite sugar beet clones. For all nine traits, the genotypic and genotype \times location interaction variances were highly significant ($P < 0.01$). Using a one-step AMMSP approach, the total number of significant ($P < 0.05$) marker-phenotype associations was 44. The identification of genome regions associated with the traits under consideration indicated that not only segregating populations derived from crosses of

parental genotypes in a systematic manner could be used for AMMSP but also populations routinely derived in plant breeding programs from multiple, related crosses. Furthermore, our results suggest that data sets, whose size does not permit analysis by the one-step AMMSP approach, might be analyzed using the two-step approach based on adjusted entry means for each location without losing too much power for detection of marker-phenotype associations.

Introduction

Since the 1990s, a large number of DNA markers has been available for sugar beet (e.g., Barzen et al. 1992, 1995; Pillen et al. 1992, 1993). These markers have been used to map major genes for bolting (El-Mezawy et al. 2002), fertility restoration (Pillen et al. 1993), hypocotyl color (Barzen et al. 1992), and resistance against nematodes (Cai et al. 1997). However, the majority of traits of economic interest in sugar beet breeding programs such as beet yield, sugar yield, as well as quality parameters, are quantitatively inherited (Weber et al. 2000).

In sugar beet, linkage mapping was employed to dissect quantitative traits into underlying genetic factors, called quantitative trait loci (QTL). Weber et al. (1999, 2000) detected QTL for sugar yield based on two segregating populations grown in different locations. Schneider et al. (2002) identified QTL for sugar yield and quality parameters based on one segregating population using expressed sequence tag related markers.

The major limitations of linkage mapping approaches are the poor resolution in detecting QTL and that only two alleles at any given locus can be studied in biparental crosses of inbred lines (Flint-Garcia et al. 2003).

Communicated by F. van Eeuwijk.

B. Stich · A. E. Melchinger (✉) · M. Heckenberger
Institute for Plant Breeding, Seed Science, and Population
Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

B. Stich
e-mail: stich@uni-hohenheim.de

J. Möhring · H.-P. Piepho
Institute for Crop Production and Grassland Research,
University of Hohenheim, 70593 Stuttgart, Germany

A. Schechert
Fr. Strube Saatzucht GmbH & Co. KG,
38387 Söllingen, Germany

Association mapping methods, which were successfully applied in human genetics to detect genes coding for human diseases (e.g., Ozaki et al. 2002), promise to overcome these limitations (Kraakman et al. 2004). Therefore, in plant genetics several attempts have been made for detecting QTLs by such methods (e.g., Breseghello and Sorrells 2006; Wilson et al. 2004).

In comparison with linkage mapping approaches, however, association mapping approaches have a low power to detect QTL in genome-wide scans (Yu and Buckler 2006). Yu et al. (2008) proposed the nested association mapping (NAM) strategy for plants, which combines the high power to detect QTL in genome-wide approaches of linkage mapping with the high mapping resolution of association mapping approaches. This strategy requires the establishment of segregating populations derived from several crosses of parental inbreds in a systematic manner. For plant breeders, however, it is an appealing idea to exploit information on the individuals routinely derived from multiple, related crosses in plant breeding programs for QTL detection using a similar approach called association mapping in multiple segregating populations (AMMSP). To our knowledge, no study has investigated the applicability of this mapping strategy in a plant breeding context based on empirical data.

The objectives of our research were to (1) examine the applicability of AMMSP in a plant breeding context based on segregating populations of various size of sugar beet, (2) compare different biometric approaches for AMMSP, and (3) detect markers with significant main effect across locations for nine traits in sugar beet.

Materials and methods

Plant materials

Our study was based on 765 F_2 , two F_3 , and one F_4 sugar beet (*Beta vulgaris* L.) genotypes which were randomly derived from 19 crosses among diploid elite sugar beet clones. The number of progenies from each cross ranged from 3 to 72 (Table 1). Testcross progenies were produced by crossing the 768 F_n genotypes ($n = 2, 3, 4$), for which pedigree information was available up to eight generations back (Fig. 1), to one diploid tester. All plant material used in this study was provided by the breeding company Strube–Dieckmann.

Field experiments

In 2004, all 768 testcross progenies were evaluated in a series of 24 field trials that included a set of four common checks. Two of the 24 trials were grown in 12 locations and

the others in 11 locations. Not all series of trials were performed in the same locations and, thus, the total number of locations was 21 with trials located in France (10 locations), Germany (7 locations), Belgium (2 locations), and the Netherlands (2 locations). The number of trials per location ranged from 6 to 24. The experimental design of each trial was a 6×6 lattice design with two replicates per location. The 32 testcross progenies and four checks of each trial were grown in three-row plots of 7 m length and a plant density of 92 063 plants ha^{-1} .

Data was recorded for beet yield (BY; $Mg\ ha^{-1}$). Content of potassium (K; $mmol\ kg^{-1}$), sodium (Na; $mmol\ kg^{-1}$), α -amino nitrogen (N; $mmol\ kg^{-1}$), and sugar (SC; %) were measured as described by Burba and Puszcz (1976). Impurity content (IC; %) was calculated as $[0.343 \times (K + Na) + (0.094 \times N) + 0.29]/SC \times 100$. White sugar content (WSC; %), sugar yield (SY; $Mg\ ha^{-1}$), and white sugar yield (WSY; $Mg\ ha^{-1}$) were calculated as $SC - IC$, $BY \times SC$, and $BY \times WSC$, respectively.

Molecular marker analyses

A subset of the 768 F_n entries comprising 369 F_2 progenies from nine crosses (Table 1) was fingerprinted by

Table 1 Description of the 19 segregating populations underlying our study

Population	Type of progenies	No. progenies with phenotypic data	No. progenies with geno- and phenotypic data
1	S1	72	0
2	S1	64	0
3	S1	64	48
4	S1	8	0
5	S1	8	0
6	S1	8	0
7	S1	40	37
8	S1	64	0
9	S1	20	0
10	S1	36	33
11	S1	63	51
12	S1	63	57
13	S2	2	0
	S3	1	0
14	S1	45	0
15	S1	51	0
16	S1	61	55
17	S1	42	36
18	S1	22	22
19	S1	34	30
Total		768	369

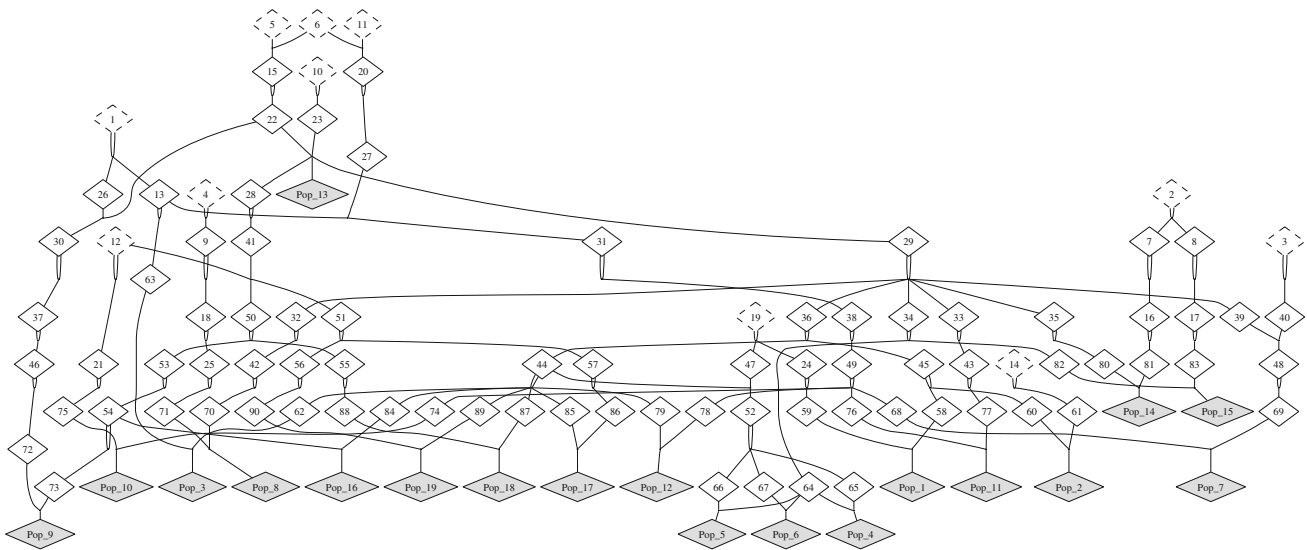


Fig. 1 Pedigree relatedness of the 19 segregating populations of sugar beet underlying our study

Strube–Dieckmann according to standard protocols with 49 simple sequence repeat markers and nine restriction fragment length polymorphism markers. These markers were randomly distributed across the sugar beet genome with an average marker distance of 8.4 cM. Map positions of all markers were based on the linkage map of Strube–Dieckmann (unpublished data).

Statistical analyses

The entries of our study were randomly derived from 19 crosses and, thus, main effects for entries as well as their interaction with the locations were regarded as random in the one-step joint linkage association analyses. Furthermore, we also performed two-step AMMSP in which entries were regarded as random in the second-step but were regarded as fixed in the first step (cf., Piepho and Möhring 2007).

Phenotypic data analysis

The phenotypic data of each of the 21 locations were first analyzed separately based on the statistical model:

$$y_{ikno} = \mu + g_i + t_k + r_{nk} + b_{onk} + e_{ikno}, \tag{1}$$

where y_{ikno} was the phenotypic observation for the i th sugar beet genotype in the o th incomplete block of the n th replication of the k th trial, μ was an intercept term, g_i was the genetic effect of the i th genotype, t_k was the effect of the k th trial, r_{nk} was the effect of the n th replication of the k th trial, b_{onk} was the effect of the o th incomplete block of the n th replication of the k th trial, and e_{ikno} was the residual. Except g_i , all effects were regarded as random.

For the genotypes (entries and checks), adjusted entry means were calculated for the j th location according to:

$$M_{ij} = \hat{\mu}_j + \hat{g}_{ij}, \tag{2}$$

where $\hat{\mu}_j$ was the estimate for the intercept at the j th location and \hat{g}_{ij} was the estimate of the genetic effect of the i th sugar beet genotype at the j th location both estimated based on the statistical model in Eq. (1). A hierarchical cluster analysis was performed on the correlation coefficient of adjusted entry means M_{ij} among all pairs of locations.

A combined analysis of plot-level phenotypic data across locations was performed based on the statistical model:

$$y_{ijkno} = \mu + c_i + d_i g_i + l_j + (cl)_{ij} + d_i (gl)_{ij} + t_{kj} + r_{nj} + b_{onjk} + e_{ijkno}, \tag{3}$$

where y_{ijkno} was the phenotypic observation for the i th sugar beet genotype at the j th location in the o th incomplete block of the n th replication of the k th trial, μ was an intercept term, c_i was a factor with a single level for each check and a single level for all entries, d_i was an indicator variable with $d_i = 0$ for checks and $d_i = 1$ for entries, g_i was the genetic effect of the i th genotype, l_j was the effect of the j th location, $(cl)_{ij}$ was the interaction effect of the i th check and the j th location, $(gl)_{ij}$ was the interaction effect of the i th entry and the j th location, t_{kj} was the effect of the k th trial at the j th location, r_{nj} was the effect of the n th replication of the k th trial at the j th location, b_{onjk} was the effect of the o th incomplete block of the n th replication of the k th trial at the j th location, and e_{ijkno} was the residual. For details regarding the use of dummy coding for separating checks and entries see

Piepho et al. (2006). Because locations were purposefully selected, l_j was regarded as fixed. For estimation of adjusted entry means over all trials and locations, we regarded $d_{1i}c_i$ and $d_{2i}g_i$ as fixed and all other effects as random. Error variances were assumed to be heterogeneous among locations. For each of the 768 entries, an adjusted entry mean M_i over all trials and locations was calculated as:

$$M_i = \hat{\mu} + \hat{g}_i, \tag{4}$$

where $\hat{\mu}$ was the estimate for the intercept and \hat{g}_i was the estimate of the genetic effect of the i th entry, both estimated based on the statistical model in Eq. (3).

For estimation of variance components, the statistical model in Eq. (3) was used, where l_j and $d_{1i}c_i$ were regarded as fixed and all other effects as random. In order to consider the relatedness among entries with (\mathbf{g}_1) and without (\mathbf{g}_2) available marker data, we assumed:

$$\mathbf{g} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \sigma_g^2 \right] \tag{5}$$

where \mathbf{A}_{11} and \mathbf{A}_{22} are the matrices of coancestry coefficients that define the degree of genetic covariance among entries with (\mathbf{A}_{11}) and without available marker data (\mathbf{A}_{22}). \mathbf{A}_{12} is the matrix of coancestry coefficients between entries with available marker data and entries without available marker data and $\mathbf{A}_{21} = \mathbf{A}_{12}$. The coancestry coefficients were estimated based on the available pedigree records, according to the rules described by Falconer and Mackay (1996) and using PROC INBREED in SAS (SAS Institute 2004).

To avoid excessive computing time, we neglected the relatedness of the entries with respect to genotype \times location interaction and assumed:

$$(\mathbf{gl}) = \begin{pmatrix} (\mathbf{g}_1\mathbf{l}) \\ (\mathbf{g}_2\mathbf{l}) \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \sigma_{gl}^2 \right], \tag{6}$$

where \mathbf{I} was an identity matrix. Heritability on an adjusted entry mean basis h_g^2 was calculated according to Emrich et al. (2008). ρ_p was calculated as correlation between the plot-level phenotypic data across locations of two traits and ρ_g as correlation between the adjusted entry means M_i of two traits.

One-step approach for AMMSP

In the current study, not only the entries with available marker data were included in the AMMSP approach but also the entries without available marker data. This was due to the fact that based on a statistical model with random entries and on the modeling of the relatedness among the entries of both sets, the phenotypic observations of the entries without available marker data improve the

estimation of the genotypic value of the entries with available marker data. Thereby, the power for detection of marker-phenotype associations is increased.

In studies based on testcross progenies with a common tester, no dominance effects can be estimated, because the effects of allele substitution (α) comprise also the dominance effects between parental alleles and those of the tester (Melchinger 1988). Therefore, we used the following statistical model for one-step AMMSP:

$$y_{ijkno} = \mu + c_i + l_j + (\alpha + \alpha l_j)' \mathbf{x}_i + d_i \check{g}_i + (cl)_{ij} + d_i(\check{g}l)_{ij} + t_{kj} + r_{nj} + b_{onjk} + e_{ijkno}, \tag{7}$$

where αl_j were the interaction effects of the allele substitution effects with the j th location, \mathbf{x}_i was a column vector with the number of copies of the corresponding alleles, and \check{g}_i was the residual genetic effect of the i th entry except for the effect of the marker locus under consideration. We regarded $d_i \check{g}_i$, $(cl)_{ij}$, $d_i(\check{g}l)_{ij}$, t_{kj} , r_{nj} , and b_{onjk} as random and all other effects as fixed. However, the expected value of the genotypic effects of entries without available marker data is not zero any longer, when conditioning on the markers. To overcome this problem, we used the following definition of the variance-covariance matrix (for derivation see Appendix):

$$\begin{pmatrix} \check{\mathbf{g}}_1 \\ \check{\mathbf{g}}_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mathbf{X}_1 \alpha \\ \mathbf{X}_2 \alpha \end{pmatrix}, \begin{pmatrix} (1-p_g)\mathbf{A}_{11} & (1-p_g)\mathbf{A}_{12} \\ (1-p_g)\mathbf{A}_{21} & \mathbf{A}_{22} - p_g \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \end{pmatrix} \sigma_{\check{g}}^2 \right], \tag{8}$$

where \mathbf{X}_1 was a $369 \times (p-1)$ matrix with p being the number of alleles of the marker locus under consideration (i.e., the i th row of \mathbf{X}_1 equalled \mathbf{x}'_i) and p_g the proportion of the genotypic variance by the marker locus under consideration. The m th column of this matrix was calculated as the number of copies of the $(m+1)$ th allele minus the number of copies of the first allele observed for the corresponding genotypes. The matrix \mathbf{X}_2 was calculated as $\mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{X}_1$, and p_g was the proportion of the genotypic variance explained by the locus under consideration.

From Eq. (8) we have $\text{Var}(\check{\mathbf{g}}) = \mathbf{K}_1 \sigma_g^2 + \mathbf{K}_2 \sigma_{g_{p_g}}^2$, where $\check{\mathbf{g}} = (\check{\mathbf{g}}_1, \check{\mathbf{g}}_2)$,

$$\mathbf{K}_1 = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad \mathbf{K}_2 = \begin{pmatrix} -\mathbf{A}_{11} & -\mathbf{A}_{12} \\ -\mathbf{A}_{21} & -\mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \end{pmatrix},$$

and $\sigma_{g_{p_g}}^2 = p_g \sigma_g^2$, such that $p_g = \sigma_{g_{p_g}}^2 / \sigma_g^2$. The proportion of the genotypic variance explained by the marker locus under consideration (\hat{p}_g) can thus be calculated from variance component estimates as

$$\widehat{P}_g = \frac{\widehat{\sigma_{gpg}^2}}{\widehat{\sigma_g^2}} \quad (9)$$

Two-step approaches for AMMSP

Approach based on adjusted entry means for each location: This approach, which was designated as two-step approach A, was based on the adjusted entry means M_{ij} calculated for each of the 21 locations. These adjusted entry means were then used in a second step for joint linkage and association analyses based on the statistical model:

$$M_{ij} = \mu + c_i + l_j + (\boldsymbol{\alpha} + \boldsymbol{\alpha}_j)' \mathbf{x}_i + d_i \check{g}_i + e_{ij}, \quad (10)$$

where all effects except $d_i \check{g}_i$ were regarded as fixed. The assumptions made in the one-step approach for AMMSP concerning $\text{Var}(\check{\mathbf{g}})$ were also made in this approach.

Approach based on adjusted entry means over all locations: This approach, which was designated as two-step approach B, was based on the adjusted entry means M_i calculated over all locations. The M_i of the 369 entries with available marker data were then used in a second step for joint linkage and association analyses based on the statistical model:

$$M_i = \mu + \boldsymbol{\alpha}' \mathbf{x}_i + \check{g}_{1i} + e_i, \quad (11)$$

where \check{g}_{1i} , the residual genetic effect of the i th entry with marker data except for the effect of the marker locus under consideration, was regarded as random and all other effects were regarded as fixed. We assumed $\text{Var}(\check{\mathbf{g}}_1) = 2\mathbf{A}_{11}\sigma_{\check{g}_1}^2$. In contrast to the two-step approach A, the two-step approach B does not allow to make inferences about the interaction effects of the allele substitution effects with the locations.

Weighting method for two-step AMMSP: The covariance matrix for the vector of residuals \mathbf{e} is denoted here as \mathbf{R} (Lynch and Walsh 1998). In two-step mixed-model procedures, the matrix \mathbf{R} may be chosen in different ways to approximate the actual variance–covariance matrix of adjusted entry means denoted here as \mathbf{V} .

We used for both two-step AMMSP approaches the weighting method proposed by Smith et al. (2001). For the two-step approach A, this method is based on the \mathbf{V} matrix calculated for each location (\mathbf{V}_j) with $\mathbf{R}_j^{-1} = D(\mathbf{V}_j^{-1})$, where $D(\mathbf{V}_j^{-1})$ was a diagonal matrix with diagonal elements equal to those of \mathbf{V}_j^{-1} . The matrix \mathbf{R}^{-1} was then calculated as: $\bigoplus_{j=1}^{21} \mathbf{R}_j^{-1}$. For the two-step approach B, $\mathbf{R}^{-1} = D(\mathbf{V}^{-1})$, where \mathbf{V} was the actual variance–covariance matrix of adjusted entry means calculated over all locations.

In a linkage mapping context, the use of the false discovery rate to overcome the multiple test problem is dubious (Chen and Storey 2006). Therefore, for all abovementioned AMMSP approaches we applied the

Bonferroni–Holm procedure (Holm 1979) to detect markers with significant ($P < 0.05$) (1) main effects across locations and (2) marker \times location interactions. The total proportion of the genotypic variance explained by all markers with significant main effect was obtained by fitting a model including all these markers simultaneously. All mixed-model calculations were performed with ASReml release 2.0 (Gilmour et al. 2006).

Linkage disequilibrium was assessed by the composite linkage disequilibrium measure Δ (Weir 1996). Significance of Δ was tested with χ^2 tests. LD computations were performed with the GDA 1.0 software (Lewis and Zaykin 1999).

Results

The coancestry coefficient calculated from pedigree records (Fig. 1) ranged for the 768 entries from 0.02 to 0.98 with an average of 0.39. The total number of alleles detected for the 58 molecular markers was 155, with the number of alleles per locus ranging from 2 to 6. The allele frequency of the 155 alleles varied between 0.01 and 0.98. Across the nine segregating populations, the percentage of SSR loci pairs with significant ($P < 0.05$) LD was 76.5% (Fig. 2). Within the segregating populations, the

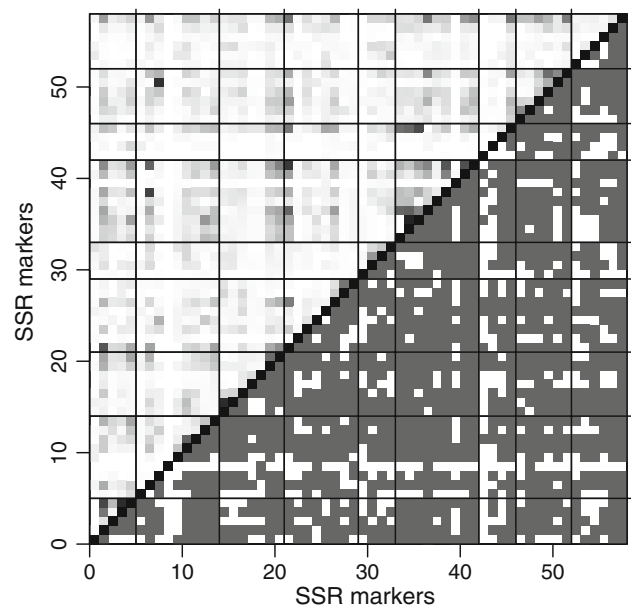


Fig. 2 Linkage disequilibrium (LD; Δ) between pairs of SSR markers (above the diagonal) and its significance ($P < 0.05$) based on χ^2 tests (below the diagonal). Dark-gray coloring indicates high Δ values and significant LD. White coloring indicates $\Delta = 0$ and no significant LD. The thin horizontal and vertical lines mark off the chromosomes

percentage of SSR loci pairs with significant ($P < 0.05$) LD was considerably lower and ranged from 2.2 to 48.6%.

For all traits, the cluster analysis based on the correlation coefficient of adjusted entry means among all pairs of locations revealed the absence of distinct subgroups of locations (data not shown). Significant ($P < 0.01$) genotypic variance and significant ($P < 0.01$) variance of genotype \times location interaction were observed for all nine traits (Table 2). Heritability on an adjusted entry mean basis was high for the nine traits and ranged from 0.82 (WSY) to 0.96 (K, IC). Correlations r_p ranged from -0.56 (Na/WSC) to 0.99 (SC/WSC, SY/WSY) (Table 3). Likewise, correlations r_g varied between -0.46 (Na/WSC) and 0.98 (SC/WSC, SY/WSY).

Using the one-step AMMSP approach, the total number of significant ($P < 0.05$) marker-phenotype associations was 44 (Table 4). The number of markers with significant main effect varied from three for BY, SY, and WSY to nine for Na. The proportion of the genotypic variance explained simultaneously by all markers with significant main effect was lowest for SY (4.4%) and highest for Na (36.8%). The proportion of the genotypic variance explained by the individual markers ranged from 1.5 (Na) to 11.8% (Na). For all markers identified based on their significant main effect, significant ($P < 0.05$) marker \times location interactions were observed, which explained between 7.9 (BY) and 20.8% (SC) of the variance of genotype \times location interactions (σ_{gl}^2).

For the nine traits examined, Spearman rank correlations between the P values calculated for the 58 marker loci by using the one-step AMMSP approach and the P values estimated based on the two-step approach A varied between 0.972 and 0.996 (Table 5). The correlations observed between the P values of the one-step procedure and those of the two-step procedure B were slightly lower and ranged from 0.942 to 0.996 for the nine traits. The same trend was observed for the Spearman rank correlation

between the allele substitution effects calculated by using the one-step approach and those using the two-step procedures for AMMSP. Two-step approach A failed to detect two markers with significant ($P < 0.05$) main effects, which were identified using the one-step approach (Fig. 3). In contrast, the two-step approach B failed to detect five markers previously identified using the one-step approach, but detected three marker-phenotype associations, which were not identified using the one-step approach.

Discussion

The current study was based on 19 segregating populations which were derived from connected crosses of parental genotypes (Fig. 1). In previous studies, such populations were analyzed using linkage mapping methods (e.g., Rebai and Goffinet 1993). These methods, however, use only the LD within each segregating population. Yu et al. (2008) proposed the NAM strategy which in addition to the LD within each segregating population exploits also the LD present in the set of parental genotypes.

When genotypes representing global genetic diversity are used as parents of the segregating populations, NAM approaches promise to result in a mapping resolution considerably higher than that obtained with linkage mapping approaches (Stich et al. 2007). This is due to the fact that LD decays in such germplasm over a short physical distance (Wilson et al. 2004). For AMMSP approaches in a plant breeding context, however, elite genotypes must be used as parental genotypes in order to obtain QTL information of direct use in elite breeding programs (Crepieux et al. 2004). This leads to a reduction in the mapping resolution in comparison with that in the studies of Stich et al. (2007) and Yu et al. (2008), because LD is expected to decay in a set of elite genotypes over a longer physical distance than in a set of diverse genotypes (cf., Palaisa

Table 2 First- and second-degree statistics for sugar beet testcross progenies for potassium content (K), sodium content (Na), α -amino nitrogen content (N), proportion of impurities (IC), sugar content (SC), white sugar content (WSC), beet yield (BY), sugar yield (SY), and white sugar yield (WSY)

Parameter	K (mmol kg ⁻¹)	Na (mmol kg ⁻¹)	N (mmol kg ⁻¹)	IC (%)	SC (%)	WSC (%)	BY (Mg ha ⁻¹)	SY (Mg ha ⁻¹)	WSY (Mg ha ⁻¹)
M_i	35.0–46.5	3.6–8.9	9.3–15.1	1.7–2.2	16.0–17.9	14.0–16.1	66.9–83.8	11.4–14.1	10.1–12.5
σ_g^2	2.7**	0.40**	0.77**	0.004**	0.088**	0.11**	6.6**	0.11**	0.084**
σ_{gl}^2	0.51**	0.093**	0.20**	0.0007**	0.017**	0.021**	1.0**	0.036**	0.031**
σ_e^2	0.83–3.32	0.11–2.36	0.50–2.76	0.001–0.008	0.030–0.239	0.035–0.227	4.6–49.9	0.14–1.39	0.11–0.99
h_g^2	0.96	0.94	0.91	0.95	0.94	0.94	0.90	0.83	0.82

M_i are the adjusted entry means calculated over all locations and σ_g^2 , σ_{gl}^2 , and σ_e^2 are the genotypic, genotype \times location, and error variances, respectively. h_g^2 is the heritability on an adjusted entry mean basis

** Significant at the 0.01 probability level

Table 3 Correlation ρ_p between the plot-level phenotypic data (above diagonal) and correlation ρ_g between the adjusted entry means cross all locations (below diagonal) for potassium content (K), sodium content (Na), α -amino nitrogen content (N), proportion of

impurities (IC), sugar content (SC), white sugar content (WSC), beet yield (BY), sugar yield (SY), and white sugar yield (WSY) based on 768 sugar beet entries

	K	Na	N	IC	SC	WSC	BY	SY	WSY
K		0.08	0.49	0.93	-0.09	-0.25	0.19	0.12	0.01
Na	-0.09		0.18	0.42	-0.52	-0.56	0.30	0.01	-0.07
N	0.32	-0.15		0.60	-0.10	-0.20	0.02	-0.04	-0.11
IC	0.91	0.31	0.35		-0.26	-0.42	0.26	0.10	-0.03
SC	-0.09	-0.42	0.11	-0.25		0.99	-0.08	0.42	0.51
WSC	-0.29	-0.46	0.02	-0.45	0.98		-0.12	0.38	0.48
BY	0.04	0.04	-0.20	0.03	-0.45	-0.42		0.87	0.81
SY	-0.01	-0.20	-0.15	-0.11	0.07	0.09	0.86		0.99
WSY	-0.15	-0.26	-0.19	-0.28	0.17	0.22	0.79	0.98	

et al. 2004). Nevertheless, even in this case exploitation of the LD present in the set of elite genotypes confers AM-MSP approaches a mapping resolution which is higher than that of linkage mapping approaches.

Statistical approaches for AMMSP

In experiments, in which the genome sequence of all entries is available, the statistical models for AMMSP differ from those for multi-cross linkage mapping in that they neither comprise fixed population effects (e.g., Blanc et al. 2006) nor regard markers as nested within populations. The results of Yu et al. (2008) suggested that approaches neglecting population structure provide a higher power for QTL detection and a lower proportion of false positives than approaches taking population structure into account. This observation might be explained by the fact that differences among the mean performance of segregating populations derived from different crosses of parental genotypes are caused by QTL. In statistical approaches which take population structure into account by fixed population effects, the effects of these QTL are absorbed in the population effect and, thus, these QTL can not be detected.

In the current study, however, no genome sequence was available for the entries and, thus, detection of marker-phenotype association is based on LD between markers and trait coding polymorphisms. Therefore, population structure of the entries must be considered in order to adhere to the nominal α level. In a AMMSP context, this can be achieved by regarding entries as random and defining covariances among the entries based on a numerator relationship matrix.

Stich et al. (2008) proposed the use of marker-based estimates of the numerator relationship matrix optimally adapted to the phenotypic data for association mapping

approaches. This way of calculating the numerator relationship matrix also seems to provide the best estimates of pair-wise coancestry coefficients in the context of AMMSP with respect to adherence to the nominal α level as well as the power for QTL detection (Stich et al. 2008). In the current study, however, the estimation of a marker-based numerator relationship matrix was not possible as marker data were not available for all entries. Therefore, we calculated the numerator relationship matrix from pedigree records.

Comparison of one- and two-step approaches for AMMSP

In all types of genetic mapping experiments, the one-step approach, in which the phenotypic and genotypic data analysis is performed in one step, is the only fully efficient analysis (Stich et al. 2008). Consequently, P values and allele substitution effects calculated for the marker loci under consideration based on such a statistical model are the reference values (Piepho and Pillen 2004). However, AMMSP data sets which are currently under development are of such a size that a one-step analysis of phenotypic and genotypic data might be either impossible or impractical due to excessive computing time. As this problem can be overcome by applying two-step procedures, we examined the consistency of the results of one- and two-step approaches for AMMSP.

Spearman rank correlations observed in the current study between the P values of the two-step approach B and those of the one-step approach (Table 5) were similar to those found by Stich et al. (2008) in an association mapping approach for wheat. Despite the rather high correlation, the two-step approach B failed to detect five marker-phenotype associations which were identified using the one-step procedure (Fig. 3) but identified three

Table 4 Marker loci significantly ($P < 0.05$) associated with potassium content (K), sodium content (Na), α -amino nitrogen content (N), proportion of impurities (IC), sugar content (SC), white sugar content (WSC), beet yield (BY), sugar yield (SY), and white sugar yield

(WSY), the explained proportion of the genotypic variance (\widehat{p}_g), the explained proportion of the genotype \times location interaction ($\widehat{p}_{g \times l}$) by the marker \times location interaction, and the allele substitution effects ($\widehat{\alpha}$)

Trait	Locus	Linkage group	Position (M)	\widehat{p}_g (%)	$\widehat{p}_{g \times l}$ (%)	$\widehat{\alpha}_{1/2} \pm SE$ (mmol kg ⁻¹)	$\widehat{\alpha}_{1/3} \pm SE$ (mmol kg ⁻¹)	$\widehat{\alpha}_{1/4} \pm SE$ (mmol kg ⁻¹)
K	M22	D	0.10	4.0	2.5	-0.23 \pm 0.10	-0.20 \pm 0.08	
	M24	D	0.32	2.8	1.0	0.15 \pm 0.07		
	M25	D	0.37	3.0	1.5	-0.35 \pm 0.07		
	M28	D	0.53	4.4	2.5	-0.04 \pm 0.10		
	M51	H	0.56	7.3	5.5	-0.26 \pm 0.18	-0.51 \pm 0.11	0.26 \pm 0.11
	Total			16.2	13.2			
Na	M05	A	0.48	6.0	1.8	0.04 \pm 0.07	-0.11 \pm 0.05	
	M10	B	0.32	4.5	6.6	-0.17 \pm 0.03		
	M12	B	0.36	4.9	6.9	-0.19 \pm 0.04		
	M19	C	0.58	11.8	14.1	-0.28 \pm 0.03		
	M21	C	0.60	10.8	4.6	-0.12 \pm 0.06	-0.10 \pm 0.12	0.03 \pm 0.13
	M28	D	0.53	2.8	3.5	-0.15 \pm 0.05		
	M38	F	0.28	1.5	11.8	-0.17 \pm 0.07	0.43 \pm 0.13	
	M51	H	0.56	7.2	9.0	-0.01 \pm 0.08	0.20 \pm 0.05	-0.02 \pm 0.05
	M52	H	0.69	5.4	12.1	0.51 \pm 0.08	0.01 \pm 0.06	
	Total			36.8	37.8			
N	M03	A	0.33	3.3	6.5	0.14 \pm 0.13	-0.29 \pm 0.08	
	M05	A	0.48	2.3	0.9	0.13 \pm 0.10	0.20 \pm 0.06	
	M17	C	0.53	2.8	5.3	-0.26 \pm 0.05		
	M19	C	0.58	3.4	1.4	0.14 \pm 0.05		
	M21	C	0.60	9.4	5.9	-0.05 \pm 0.09	0.44 \pm 0.17	0.10 \pm 0.19
	M28	D	0.53	2.5	1.1	-0.15 \pm 0.07		
	Total			16.8	17.0			
Trait	Locus	Linkage group	Position (M)	\widehat{p}_g (%)	$\widehat{p}_{g \times l}$ (%)	$\widehat{\alpha}_{1/2} \pm SE$ (%)	$\widehat{\alpha}_{1/3} \pm SE$ (%)	$\widehat{\alpha}_{1/4} \pm SE$ (%)
IC	M22	D	0.10	4.9	3.3	-0.007 \pm 0.004	-0.009 \pm 0.003	
	M24	D	0.32	3.6	1.0	0.008 \pm 0.003		
	M25	D	0.37	3.3	2.6	-0.017 \pm 0.003		
	M27	D	0.50	3.3	1.7	0.011 \pm 0.003		
	M28	D	0.53	7.9	1.5	-0.009 \pm 0.004		
	M29	D	0.63	2.4	5.8	-0.005 \pm 0.005	0.000 \pm 0.005	
	M46	G	0.42	4.2	0.8	-0.001 \pm 0.007	0.004 \pm 0.007	-0.028 \pm 0.009
	M50	H	0.51	3.1	4.8	-0.002 \pm 0.004	-0.003 \pm 0.005	
Total			15.2	12.3				
SC	M05	A	0.47	6.6	1.4	-0.04 \pm 0.02	0.03 \pm 0.02	
	M38	F	0.28	8.3	7.0	-0.04 \pm 0.03	-0.04 \pm 0.05	
	M39	F	0.34	3.7	1.2	0.04 \pm 0.01		
	M40	F	0.43	8.8	11.1	-0.09 \pm 0.07	0.05 \pm 0.14	
Total			20.8	12.5				
WSC	M05	A	0.47	6.2	1.3	-0.04 \pm 0.03	0.04 \pm 0.02	
	M22	D	0.10	2.6	6.1	0.02 \pm 0.02	0.04 \pm 0.01	
	M38	F	0.28	6.4	7.7	-0.04 \pm 0.03	-0.07 \pm 0.05	
	M40	F	0.43	5.5	11.0	-0.12 \pm 0.08	-0.10 \pm 0.15	
	Total			15.3	12.7			

Table 4 continued

Trait	Locus	Linkage group	Position (M)	\widehat{p}_g (%)	$\widehat{p}_{g \times i}$ (%)	$\widehat{\alpha}_{1/2} \pm \text{SE}$ (Mg ha ⁻¹)	$\widehat{\alpha}_{1/3} \pm \text{SE}$ (Mg ha ⁻¹)	$\widehat{\alpha}_{1/4} \pm \text{SE}$ (Mg ha ⁻¹)
BY	M05	A	0.47	5.2	1.9	0.36 ± 0.30	-0.57 ± 0.20	
	M38	F	0.28	4.8	6.5	0.36 ± 0.31	0.18 ± 0.56	
	M54	I	0.30	3.3	0.5	-0.63 ± 0.20		
	Total			12.6	7.9			
SY	M50	H	0.51	3.2	7.5	0.08 ± 0.03	-0.15 ± 0.04	
	M51	H	0.56	3.5	15.0	0.01 ± 0.06	0.06 ± 0.04	0.01 ± 0.04
	Total			4.4	17.1			
WSY	M49	H	0.42	2.3	3.9	-0.09 ± 0.03		
	M50	H	0.51	4.1	6.1	0.07 ± 0.02	-0.13 ± 0.03	
	M51	H	0.56	4.1	14.0	0.02 ± 0.05	0.07 ± 0.03	0.00 ± 0.03
	Total			6.4	12.5			

markers which were not detected using the one-step procedure. In addition, a rather low correlation was observed between the allele substitution effects of the two-step procedure B and the one-step procedure (Table 5). These observations suggested that for AMMSP, use of the two-step approach B, frequently applied in linkage mapping (e.g., Schön et al. 2004) and association mapping (e.g., Breseghello and Sorrells 2006) experiments, is questionable.

Spearman rank correlations between P values calculated using the two-step approach A and the one-step approach were only slightly higher than the correlation observed between two-step approach B and the one-step approach (Table 5). Nevertheless, the two-step approach A failed only to detect two marker-phenotype associations which were identified using the one-step approach and detected only marker-phenotype associations which were also identified based on the one-step approach. These observations indicated that the two-step approach A has a considerably higher power for QTL detection and results in a lower proportion of false positives than the two-step approach B. Therefore, for data sets in which the one-step analysis of phenotypic and genotypic data might be impossible or impractical, we suggest using the two-step approach A. Our further discussion, however, is restricted to results from the one-step approach due to its more desirable properties.

Detected marker-phenotype associations

In our study, the lowest number of markers with significant ($P < 0.05$) main effect was detected for BY, SY, and WSY, whereas a considerably higher number of markers were associated with K, Na, and N (Table 4). This observation is in accordance with findings of Schneider et al. (2002) and might be explained by the fact that BY, SY, and

Table 5 Spearman rank correlation between P values (ρ_P) and allele substitution effects (ρ_z) calculated using the one-step AMMSP approach and the two-step AMMSP approaches A and B

Trait	Two-step AMMSP approach			
	A		B	
	ρ_P	ρ_z	ρ_P	ρ_z
K	0.996	0.853	0.991	0.346
Na	0.993	0.874	0.975	0.870
N	0.978	0.940	0.991	0.583
IC	0.987	0.853	0.942	0.565
SC	0.994	0.934	0.995	0.788
WSC	0.996	0.948	0.996	0.822
BY	0.984	0.876	0.970	0.748
SY	0.983	0.746	0.955	0.473
WSY	0.972	0.763	0.959	0.401

For a detailed description of the different methods see “[Materials and methods](#)”

WSY are complex traits for which no QTL with major effects were expected but rather a large number of QTL with small individual effects (Moreau et al. 2004). However, the power to detect such QTL is low. In contrast, for less complex traits influenced by a low number of QTL, such as K, Na, and N, the individual QTL explain a high proportion of the variance and, thus, a high power for QTL detection is anticipated.

Proportion of the genotypic variance explained by markers: In the current study, the proportion of the genotypic variance explained by the markers with significant main effect ranged from 1.5 to 11.8% and the proportion explained by all significant markers together varied for the nine traits from 4.4 to 36.8% (Table 4). These estimates of the proportion of the explained variance were considerably lower than those observed by Schneider et al. (2002) for

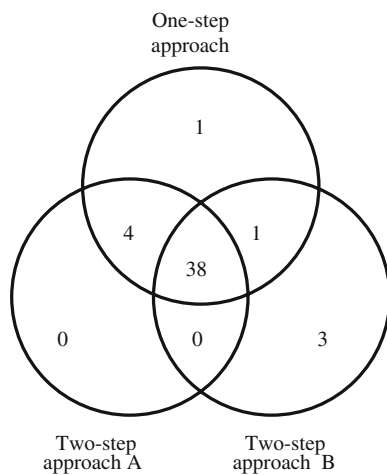


Fig. 3 Number of markers with significant ($P < 0.05$) main effect detected for the nine traits under consideration using the (1) one-step joint linkage and association mapping (AMMSP) approach and (2) two-step AMMSP approaches A and B. For a detailed description of the different methods see “Materials and methods”

the same traits in a linkage mapping experiment of sugar beet. This finding might be explained by the different concepts in choosing the parental genotypes of the segregating populations underlying these studies. In the present study, current sugar beet elite genotypes were used as parents of the segregating populations while in the study of Schneider et al. (2002) the parents of the mapping population were chosen in such a way that they maximally differed for the traits under consideration. The latter approach increases the probability of detecting QTL explaining a large proportion of the genotypic variance (Lander and Botstein 1989). However, in contrast to the current study, this procedure leads to QTL information which might be worthless for marker-assisted selection in elite plant breeding programs because the favorable QTL allele might already be fixed in the elite germplasm pool.

A further reason for the considerable discrepancy in the proportion of the explained variance observed in our study and the biparental cross of Schneider et al. (2002) might be the difference in allele frequencies expected for the germplasm in both studies. This is due to the fact that the proportion of genotypic variance explained by a marker is a function of the allele frequency and the allele substitution effect. Under the assumption of a fixed allele substitution effect, the maximum of the proportion of genotypic variance explained by a marker is observed for an allele frequency of 0.5, as expected for the entries derived from a biparental cross. In contrast, for progenies derived from several crosses of parental genotypes, like those examined in our study, the allele frequencies might be considerably different from 0.5 and, thus, the proportion of the genotypic variance explained by a marker might be notably lower

despite the same underlying allele substitution effect. Because allele frequencies of plant breeding populations are expected to be more similar to that of populations derived from several crosses of parental inbreds than to those of a population derived from a biparental cross (Crepieux et al. 2004), we suggest that in comparison to the latter approach the former approach leads to a more representative estimate of the variance accounted for by a marker in the breeding population.

Furthermore, the differing results between our study and that of Schneider et al. (2002) with respect to the proportion of the variance explained by markers might be attributable to the fact that the number of genotypes underlying the study of Schneider et al. (2002) was considerably lower than that of the current study. With a limited sample size, however, model selection generally causes an overestimation of the genotypic variance explained by markers (Beavis 1994; Utz and Melchinger 1994; Schön et al. 2004).

Markers associated with multiple traits: Our results suggested that several markers were not only associated with one trait but with two to five traits (Table 4). This observation is in accordance with the results of Schneider et al. (2002) and can be explained by a single gene with pleiotropic effects on several traits. The same phenomenon can be the result of tightly linked genes, which due to the limited mapping resolution are detected as one locus, but affect different traits. For some markers (e.g., M28), this discrimination is of less interest to plant breeders if the same allele affects the phenotype for all traits in the desired direction. However, for most markers associated with several traits in our study (e.g., M5), the positive allele for one trait had a negative effect for the other trait. In this case, the explanation for colocalization of marker-phenotype associations for different traits by pleiotropy or linkage determines whether marker-assisted selection can be used to improve both traits concurrently. Based on our results, however, this discrimination can not be made.

The high extent of LD observed for the markers of our study across the nine segregating populations suggested that the marker density should be sufficient for genome-wide QTL detection. Nevertheless, colocalization of QTL for different traits might be further analyzed based on the germplasm of our study genotyped in the identified genome regions with a higher marker density and applying multivariate association mapping approaches. Subsequently, experiments for a direct proof of the allele function might be performed, because AMMSP studies provide only statistical, i.e., indirect evidence for the function of the identified genome regions (Andersen and Lübberstedt 2003). This can be obtained by comparing isogenic genotypes which can be produced by recurrent backcrossing or

based on targeting induced local lesions in genomes (Jung and Hohmann 2006).

Marker × location interactions: For all markers with significant main effects identified in the current study, also significant marker × location interactions were observed (Table 4). Our observation is in contrast to results from most linkage mapping studies reported in the literature (e.g., Cockerham and Zeng 1996; Melchinger et al. 1998), which rarely found significant marker × location interactions despite the presence of significant genotype × location interactions. This difference might be explained by the fact that in contrast to previous studies the entries of our study were grown in a large number of locations allowing a more reliable estimation of marker × location interactions. Furthermore, the high number of entries examined in our study increases the power for detection of marker × location interactions (cf. Boer et al. 2007).

Marker × location interactions can not be directly used in marker-assisted selection programs. They rather contribute to the instability of QTL effects across multiple environments (Piepho 2000) and thereby reduce the efficiency of marker assisted selection with respect to a broad adaptation of the germplasm (e.g., Bouchez et al. 2002). One way to use marker × location or genotype × location interactions for breeding progress is the development of specific varieties for specific groups of locations (Curnow 1988). In contrast to results of Moreau et al. (2004), however, a clustering approach did not reveal the presence of distinct subgroups of locations for any trait. Therefore, we did not further examine the possibility of unraveling marker × location interactions in marker × mega-environment interactions.

Conclusions

We proposed an approach for AMMSP which is not only based on entries with available marker data but also on

manner, as suggested by Yu et al. (2008), can be used for QTL detection, but also populations routinely derived in plant breeding programs from multiple, related crosses. Furthermore, our results indicate that data sets, which are of a size that does not permit analysis by a one-step AMMSP approach, might be analyzed based on the two-step approach A without losing too much power for detection of marker-phenotype associations. By contrast, the commonly employed two-step approach B suffers from a more severe loss of power.

Acknowledgments This research was conducted within the Breeding and Informatics (BRAIN) project of the Genome Analysis of the Plant Biological System (GABI) initiative (<http://www.gabi.de>). The authors appreciate the editorial work of Dr. J. Muminović, whose suggestions considerably improved the style of the manuscript. The authors thank two anonymous reviewers for their valuable suggestions.

Appendix

Statistical models with random effects for genotypes can be used in AMMSP approaches to model the relatedness among genotypes. In this context it is assumed that

$$\text{Var}(\mathbf{g}) \sim \mathbf{A}\sigma_g^2,$$

where \mathbf{g} is the vector containing the genotypic effects of all entries and \mathbf{A} is the numerator relationship matrix. However, in AMMSP studies in which marker data are not available for all entries, separate genotypic effects have to be assumed for entries with (\mathbf{g}_1) and without (\mathbf{g}_2) available marker data:

$$\mathbf{g} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \sigma_g^2 \right].$$

Assume $\mathbf{g}_1 = \mathbf{f}_1 + \mathbf{h}_1$, where \mathbf{f}_1 is the vector of genotypic effects of the marker locus under consideration, then:

$$\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{h}_1 \\ \mathbf{g}_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} p_g \mathbf{A}_{11} & \mathbf{0} & p_g \mathbf{A}_{12} \\ \mathbf{0} & (1-p_g) \mathbf{A}_{11} & (1-p_g) \mathbf{A}_{12} \\ p_g \mathbf{A}_{21} & (1-p_g) \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \sigma_g^2 = \begin{pmatrix} \mathbf{G} & \mathbf{H} \\ \mathbf{H}' & \mathbf{K} \end{pmatrix} \sigma_g^2 \right],$$

related entries without available marker data. Thereby, the estimation of the genotypic value of the entries with available marker data is improved and the power for detection of marker-phenotype associations is increased. Based on this approach, we identified several genome regions associated with the traits under consideration, which are promising for marker-assisted selection. This observation indicated that not only segregating populations derived from crosses of parental genotypes in a systematic

where p_g is the proportion explained genotypic variance of the marker locus under consideration, $\mathbf{G} = p_g \mathbf{A}_{11}$, $\mathbf{H} = (\mathbf{0} \ p_g \mathbf{A}_{12})$, and

$$\mathbf{K} = \begin{pmatrix} (1-p_g) \mathbf{A}_{11} & (1-p_g) \mathbf{A}_{12} \\ (1-p_g) \mathbf{A}_{12} & \mathbf{A}_{22} \end{pmatrix}.$$

Based on the formulas for the conditional multivariate normal distribution (Searle et al. 1992), it can be concluded that conditioning on \mathbf{f}_1 leads to:

$$(\mathbf{h}_1, \mathbf{g}_1 | \mathbf{f}_1) \sim \left[(\mathbf{H}'\mathbf{G}^{-1}\mathbf{f}_1), (\mathbf{K} - \mathbf{H}'\mathbf{G}^{-1}\mathbf{H})\sigma_g^2 \right],$$

where

$$\mathbf{H}'\mathbf{G}^{-1}\mathbf{f}_1 = (\mathbf{0}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{f}_1)$$

and

$$\begin{aligned} \mathbf{K} - \mathbf{H}'\mathbf{G}^{-1}\mathbf{H} &= \begin{pmatrix} (1-p_g)\mathbf{A}_{11} & (1-p_g)\mathbf{A}_{12} \\ (1-p_g)\mathbf{A}_{12} & \mathbf{A}_{22} \end{pmatrix} \\ &\quad - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_g\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix} \\ &= \begin{pmatrix} (1-p_g)\mathbf{A}_{11} & (1-p_g)\mathbf{A}_{12} \\ (1-p_g)\mathbf{A}_{12} & \mathbf{A}_{22} - p_g\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix}. \end{aligned}$$

Under the assumption that conditioning on a marker locus is equivalent to conditioning on \mathbf{f}_1 , this result suggests that in the conditional density of all entries, the expected value of the entries without available marker data is shifted by conditioning on the marker data of the entries with available marker data. This problem can be overcome by replacing \mathbf{f}_1 by a regression approach on the marker ($\mathbf{f}_1 = \mathbf{X}_1\boldsymbol{\alpha}$), which leads to:

$$(\mathbf{h}_1, \mathbf{g}_2 | \mathbf{f}_1 \approx \text{Marker}) \sim \text{MVN} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1\boldsymbol{\alpha} \end{pmatrix}, (\mathbf{K} - \mathbf{H}'\mathbf{G}^{-1}\mathbf{H})\sigma_g^2 \right]$$

and

$$(\mathbf{g}_1, \mathbf{g}_2 | \mathbf{f}_1 \approx \text{Marker}) \sim \text{MVN} \left[\begin{pmatrix} \mathbf{X}_1\boldsymbol{\alpha} \\ \mathbf{X}_2\boldsymbol{\alpha} \end{pmatrix}, (\mathbf{K} - \mathbf{H}'\mathbf{G}^{-1}\mathbf{H})\sigma_g^2 \right],$$

where $\mathbf{X}_2 = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1$ and $\sigma_g^2 = (1-p_g)\sigma_g^2$.

References

- Andersen JR, Lübberstedt T (2003) Functional markers in plants. *Trends Plant Sci* 8:554–560
- Barzen E, Mechelke W, Ritter E, Seitzer JF, Salamini F (1992) RFLP markers for sugar beet breeding: chromosomal linkage maps and location of major genes for rhizomania resistance, monogerm and hypocotyl colour. *Plant J* 2:601–611
- Barzen E, Mechelke W, Ritter E, Schulte-Kappert E, Salamini F (1995) An extended map of sugar beet genome containing RFLP and RAPD loci. *Theor Appl Genet* 90:189–193
- Beavis WD (1994) The power and deceit of QTL-experiments: lessons from comparative QTL studies. In: 49th annual corn and sorghum industry research conference. American Seed Trade Association, Washington, DC, pp 250–266
- Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theor Appl Genet* 113:206–224
- Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics* 177:1801–1813
- Bouchez A, Hospital F, Causse M, Gallais A, Charcosset A (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite inbred lines. *Genetics* 162:1945–1959
- Breseghele F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177
- Burba M, Puszcz W (1976) Über die Verwendung von Aluminiumsalzen an Stelle von basischen Bleiacetaten zur Klärung von kalten wäßrigen Breiextrakten der Rübe. *Zuckerindustrie* 26:249–251
- Cai D, Kleine M, Kifle S, Harloff H-J, Sandal NN, Marcker KA, Klein-Lankhorst RM, Salentijn EMJ, Lange W, Stiekema WJ, Wyss U, Grundler FMW, Jung C (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science* 275:832–834
- Chen L, Storey JD (2006) Relaxed significance criteria for linkage analysis. *Genetics* 173:2371–2381
- Cockerham CC, Zeng Z-B (1996) Design III with marker loci. *Genetics* 143:1437–1456
- Crepieux S, Lebreton C, Servin B, Charmet G (2004) Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information from marker data. *Genetics* 168:1737–1748
- Curnow RN (1988) The use of correlated information on treatment effects when selecting the best treatment. *Biometrika* 75:287–293
- El-Mezway A, Dreyer F, Jacobs G, Jung C (2002) High-resolution mapping of the bolting gene *B* of sugar beet. *Theor Appl Genet* 105:100–105
- Emrich K, Wilde F, Miedaner T, Piepho H-P (2008) REML approach for adjusting the Fusarium head blight rating to a phenological date in inoculated selection experiments of wheat. *Theor Appl Genet* 117:65–73
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd, London
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, UK
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Jung C, Hohmann U (2006) Establishment of a TILLING platform for sugar beet. In: Proceedings of the 14th plant and animal genome conference, San Diego
- 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
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Lewis PO, Zaykin D (1999) Genetic data analysis. Computer program for the analysis of allelic data. Version 1.0.
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Assoc., Sunderland
- Melchinger AE (1988) Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent population. In: Weir BS, Eisen EJ, Goodman MM, Nomkoong G (eds) Proceedings of the 2nd international conference on quantitative genetics. Sinauer Associates, Raleigh, pp 400–415
- 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 larger bias in estimates of QTL effects. *Genetics* 149:383–403
- Moreau L, Charcosset A, Gallais A (2004) Use of trail clustering to study QTL \times environment effects for grain yield and related traits in maize. *Theor Appl Genet* 110:92–105
- Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunado T, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 32:650–654
- Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *PNAS (USA)* 101:9885–9890
- Piepho H-P (2000) A mixed model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics* 156:253–260
- Piepho H-P, Möhring J (2007) On weighting in two-stage analyses of series of experiments. *Biul Oceny Odmian* 32:109–212
- Piepho H-P, Pillen K (2004) Mixed-modelling for QTL \times environment interaction analysis. *Euphytica* 137:147–153
- Piepho H-P, Williams ER, Fleck M (2006) A note on the analysis of designed experiments with complex treatment structure. *Hort-Science* 41:446–452
- Pillen K, Steinrücken G, Wricke G, Herrmann RG, Jung C (1992) A linkage map of sugar beet (*Beta vulgaris* L.). *Theor Appl Genet* 84:129–135
- Pillen K, Steinrücken G, Wricke G, Herrmann RG, Jung C (1993) A extended linkage map of sugar beet (*Beta vulgaris* L.) including nine putative lethal genes and restorer gene X. *Plant Breed* 111:265–272
- Rebai A, Goffinet B (1993) Power of tests for QTL detection using replicated progenies derived from a diallel cross. *Theor Appl Genet* 86:1014–1022
- SAS Institute (2004) SAS Version 9.1. SAS Institute, Cary
- Schneider K, Schäfer-Pregl R, Borchardt DC, Salamini F (2002) Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. *Theor Appl Genet* 104:1107–1113
- Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167:485–498
- Searle SR, Casella G, and McCulloch CE (1992) Variance components. Wiley, New York
- Smith A, Cullis B, Gilmour A (2001) The analysis of crop variety evaluation data in Australia. *Aust N Z J Stat* 43:129–145
- Stich B, Yu J, Melchinger AE, Piepho H-P, Utz HF, Maurer HP, Buckler ES (2007) Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* 176:563–570
- Stich B, Möhring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. *Genetics* 178:1745–1754
- Utz HF, Melchinger AE (1994) Comparison of different approaches to interval mapping of quantitative trait loci. In: Van Ooijen JW, Jansen J (eds) *Biometrics in plant breeding: applications of molecular markers*. Proceedings of the ninth meeting of the EUCARPIA section biometrics in plant breeding. CPRO-DLO, Wageningen, Netherlands
- Weber WE, Borchardt DC, Koch G (1999) Combined linkage maps and QTLs in sugar beet (*Beta vulgaris* L.) from different populations. *Plant Breed* 118:193–204
- Weber WE, Borchardt DC, Koch G (2000) Marker analysis for quantitative traits in sugar beet. *Plant Breed* 119:97–106
- Weir BS (1996) Genetic data analysis II, 2nd edn. Sinauer Associates, Sunderland
- Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16:2719–2733
- Yu J, Buckler E (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551