Theor Appl Genet (2010) 121:1489–1499
DOI 10.1007/s00122-010-1405-7

ORIGINAL PAPER

Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L.) investigated with joint linkage association mapping

Jochen C. Reif · Wenxin Liu · Manje Gowda · Hans Peter Maurer · Jens Möhring · Sandra Fischer · Axel Schechert · Tobias Würschum

Received: 28 April 2010 / Accepted: 5 July 2010 / Published online: 18 July 2010

Abstract Epistatic interactions may contribute substantially to the hybrid performance of sugar beet. The main goal of our study was to dissect the genetic basis of eight important physiological and agronomic traits using two different biometrical models for joint linkage association mapping. A total of 197 genotypes of an elite breeding population were evaluated in multi-location trials and fingerprinted with 194 SNP markers. Two different statistical models were used for the genome-wide scan for marker–trait associations: Model A, which corrects for the genetic background with markers as cofactors and Model B, which additionally models a population effect. Based on the extent of linkage disequilibrium in the parental population, we estimated that for a genome-wide scan at least 100 equally spaced markers are necessary. We mapped across the eight traits 39 QTL for Model A and 22 for Model B. Only 11% of the total number of QTL were identified based on Models A and B, which indicates that both models are complementary. Epistasis was detected only for two out of the eight traits, and contributed only to a minor extent to the genotypic variance. This low relevance of epistasis implies that in sugar beet breeding the prediction of performance of three-way hybrids is feasible with high accuracy based on the means of their single crosses.

Communicated by I. Mackay.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1405-7) contains supplementary material, which is available to authorized users.

J. C. Reif (✉) · W. Liu · M. Gowda · H. P. Maurer · T. Würschum
State Plant Breeding Institute, University of Hohenheim, 70599 Stuttgart, Germany
e-mail: jochreif@uni-hohenheim.de

J. Möhring
Bioinformatics Unit, Institute for Crop Production and Grassland Research, University of Hohenheim, 70593 Stuttgart, Germany

S. Fischer · A. Schechert
Strube GmbH & Co. KG, 38387 Söllingen, Germany

Introduction

Epistasis refers to interactions among alleles of different genetic loci (Carlborg and Haley 2004). The consequence of epistasis is that the phenotype of an individual cannot be predicted simply by the sum of the single-locus effects, but rather depends on the specific combinations of loci (Lynch and Walsh 1998). Although increasing evidence for the existence of epistasis has been provided at the molecular level (Carlborg and Haley 2004), its importance for the performance of elite breeding germplasm has received little attention.

Joint linkage association mapping was suggested as a strategy to combine the high power of QTL detection from linkage analyses with the fine resolution of association mapping (Yu et al. 2008). The resolution of joint linkage association mapping mainly depends on the extent and distribution of linkage disequilibrium (LD) of parental genotypes of the segregating populations. In elite breeding populations a high extent of LD was observed for various crops such as wheat (Chao et al. 2007), maize (Reif et al. 2005), and barley (Kraakman et al. 2004). Therefore, genome-wide joint linkage association mapping in breeding populations is feasible with the available marker density in many crops of economic importance.

For sugar beet, linkage mapping studies were performed for investigating the genetic basis of sugar yield and its components (Weber et al. 1999, 2000; Schneider et al.

2002). A candidate gene association mapping approach based on 26 SSR markers was employed to map QTL underlying beet yield and sugar content (Stich et al. 2008a). Moreover, a joint linkage association mapping study was conducted using multiple related segregating sugar beet breeding populations with a limited set of 49 SSR markers (Stich et al. 2008b). Nevertheless, this marker density does not facilitate genome-wide QTL scans. In addition, published studies focused rather on main effect QTL than on the detection of epistasis.

Biometric models for joint linkage association mapping should correct for population stratification if the underlying population structure is associated with the trait under consideration. For multiple segregating breeding populations, Stich et al. (2008b) suggested the use of mixed linear models to correct for population structure facilitating a stringent control of the type I error rate. The use of mixed linear models, however, can be computationally challenging for large data sets or for genome-wide scans for digenic or higher order epistatic effects (Zhang et al. 2010). Yu et al. (2008) suggested for the nested-association mapping design the use of cofactors to correct for the genetic background. Inclusion of a general population effect led to a slightly reduced power of QTL detection as well as a smaller genetic variation explained by the final selected significant markers. Nevertheless, all approaches should ultimately be tested with empirical data given that the detection of epistasis may require proper modeling of the genetic background (Yu et al. 2008).

The objectives of our study were to (1) estimate the marker density required for joint linkage association mapping in elite breeding populations of sugar beet, (2) compare empirically different statistical models for joint linkage association mapping in elite breeding populations, and (3) dissect the genetic basis of eight important physiological and agronomic traits.

Materials and methods

Plant materials

Our study was based on 127 S_1, 20 S_2, and 50 doubled haploid (DH) sugar beet (*Beta vulgaris* L.) progenies, which were randomly derived from nine crosses among diploid sugar beet clones (Table 1; Supplementary Table S1). The parental genotypes (G) were all monogerm and belong to the female heterotic pool (further denoted as O pool). The number of progenies from each cross ranged from 12 to 30. The 197 genotypes were crossed to a cytoplasmic male sterile line (CMS-L). Testcross progenies were produced by crossing these single crosses with a diploid pollinator (T) from the male heterotic pool as tester. Thus, entries evaluated in field trials are of the form (G × CMS-L) × T. All material used in this study was provided by the breeding company Strube GmbH & Co. KG.

Table 1 Description of the nine segregating populations underlying our study

Population	Parent 1	Parent 2	Type of progenies	No. of progenies
A × B	A	B	S_1	29
C × D	C	D	S_2	12
C × E	C	E	S_1	27
A × F	A	F	S_1	19
G × H	G	H	S_1	20
I × J	I	J	DH	20
I × K	I	K	DH	30
L × M	L	M	S_1	20
N × O	N	O	S_2	20
Total				197

Field experiments and phenotypic data analysis

The 197 testcross progenies were divided into four experiments. Each experiment was carried out in 10 locations with a total of 17 locations (Supplementary Table S2). A set of four common parents was included in each experiment. The experimental design was a 12 × 6 α-lattice and a 6 × 6 lattice design with two replicates per location. Plot size was 8.5 m^{-2}. Plant density was 85,000 plants ha^{-1}.

Data were recorded for potassium content (K, decamol Mg^{-1}), sodium content (Na, hectomol Mg^{-1}), α-amino nitrogen content (N, hectomol Mg^{-1}), sugar content (SC, %), white sugar content (WSC, %), sugar yield (SY, Mg ha^{-1}), white sugar yield (WSY, Mg ha^{-1}), and beet yield (BY, Mg ha^{-1}). Methods for measuring the above traits are described in detail by Stich et al. (2008b). Two locations were excluded in the final analyses for sodium content and α-amino nitrogen content because of low quality phenotypic data.

Phenotypic data analyses

The phenotypic data were analyzed based on following statistical model:

$$y_{ijkno} = \mu + c + g_i + l_j + (\mathrm{gl})_{ij} + (\mathrm{cl})_{ij} + t_{jk} + r_{njk} + b_{onjk} + e_{ijkno},$$

where y_{ijkno} was the phenotypic observation for the ith genotype at the jth location of the kth trial of the nth replicate in the oth incomplete block, μ was an intercept term, c was a factor with a single level for each check and a single level for all entries, g_i was the genetic effect of the ith genotype, l_j was the effect of the jth location, $(\mathrm{gl})_{ij}$ was

the interaction effect of the *i*th line and the *j*th location, t_{jk} was the effect of the *k*th trial at the *j*th location, r_{njk} was the effect of the *n*th replicate at the *j*th location of the *k*th trial, b_{onjk} was the effect of the *o*th incomplete block of the *n*th *r*eplicate at the *j*th location of the *k*th trial, and e_{ijkno} was the residual. Dummy variables were used to separate checks and genotypes and estimate variances for each group following Piepho et al. (2006), but for the sake of simplicity we suppressed dummies in the model stated above. For estimating variance components, checks were treated as fixed and the other effects as random. Error and block variances were assumed to be heterogeneous among locations. Variance components were determined by the restricted maximum likelihood (REML) method by using the software ASReml (Gilmour et al. 2006). Significance of the variance component estimates were tested by model comparison with likelihood ratio tests (Stram and Lee 1994). Heritability on an entry-mean basis was calculated as the ratio of genotypic to phenotypic variance according to Melchinger et al. (1998). Furthermore, genotypes were treated as fixed effects, and best linear unbiased estimators (BLUEs) were determined for all testcrosses and traits. Simple correlation coefficients (*r*) were calculated among all traits based on BLUEs of the 197 testcrosses. Significance of *r* was tested by using tabulated values based on Fisher's (1921) *z* transformation. Genotypic correlations were estimated extending the above described model to a bivariate one.

Molecular data analyses

The 15 parents and their 197 progenies were fingerprinted following standard protocols with 194 single nucleotide polymorphisms (SNP) markers. These markers were randomly distributed across the sugar beet genome with an average marker distance of 3 cM (Supplementary Figure S1). Map positions of all markers were based on the linkage map of Strube GmbH & Co. KG. High-quality molecular data were produced for 14 out of the 15 parents and for 196 out of the 197 progenies. Therefore, the analysis of linkage disequilibrium is based on the 14 parents, and the subsequent joint linkage association mapping analyses are based on 196 progenies.

Polymorphic information content of every marker was estimated as $\mathrm{PIC} = 1 - \sum p^2$, where *p* refers to the allele frequencies of the locus under consideration. Associations among the 14 parents and their 196 genotypes were analyzed by applying principal coordinate analysis (PCoA) (Gower 1966) based on the modified Rogers' distances of the individuals (Wright 1978). Moreover, modified Rogers' distances based on the allele frequencies of the nine populations were calculated. Correlation between the resulting distance matrix and the differences in trait means of the nine populations were tested applying a Mantel (1967) test. Extent of LD between all pairs of loci was determined by estimating r^2 as described by Hill and Robertson (1968). Decay of LD with genetic map distance was evaluated by non-linear regression following Remington et al. (2001). The 95th percentile of r^2 estimates between unlinked markers was taken as a population-specific critical value of r^2 due to genetic linkage. LD analyses and PCoA were performed using software Plabsoft (Maurer et al. 2008).

Joint linkage association mapping

For the joint linkage association mapping analyses, an additive genetic model was chosen for the testcross progenies as described by Utz et al. (2000). We applied a two-step procedure for QTL detection. In a first step, stepwise multiple linear regression was used to select a set of cofactors based on the Schwarz (1978) Bayesian Criterion (SBC). In the second step, we calculated a *P* value for the association of each marker with the phenotypic value for the *F* test with a full model (with marker effects) against a reduced model (without marker effects):

$$y = \mu + b_q x_q + \sum_{c \neq q} b_c x_c + e$$

where *y* is the vector of the best linear unbiased estimators of all testcross progenies, μ the intercept, b_q (b_c) is the regression coefficient of the *q*th marker locus (or *c*th cofactor), x_q (x_c) an incidence vector of the genotypes of the testcross progenies at the *q*th marker (*c*th cofactor), and *e* the vector of residual errors. This model was denoted as Model A. In addition, we applied an alternative model (Model B) including an effect for the segregating population (Pop):

$$y = \mu + \mathrm{Pop} + b_q x_q + \sum_{c \neq q} b_c x_c + e.$$

The Bonferroni–Holm procedure (Holm 1979) was used to detect markers with significant ($P < 0.05$) main effects. The proportion of the phenotypic variance explained by QTL was determined by the estimator R^2_{adj} as described by Utz et al. (2000). The proportion of the genotypic variance explained by all detected QTL was estimated from the ratio $p_G = R^2_{\mathrm{adj}}/h^2$.

In addition, we extended Model B and performed a two-dimensional scan for pairwise interaction effects among the 194 SNP markers. The model included the selected co-factors as well as the main and interaction effects of the marker pair under consideration. The Bonferroni–Holm procedure (Holm 1979) was applied to correct for multiple testing. The genotypic variance explained by the epistatic QTL was obtained as the difference in p_G between the full model and a model without

the respective epistatic QTL. Joint linkage association mapping analysis was performed using Proc GLMSELECT, and GLM implemented in the statistical software SAS (SAS Institute 2004) (SAS code is available upon request).

Results

Genotypic variances were significantly larger than zero ($P < 0.01$) for all traits (Table 2). Variances due to genotype × location interactions were also significantly larger than zero ($P < 0.05$) for all the traits except N but of smaller magnitude. Heritabilities ranged between 0.55 for N and 0.93 for K. Moreover, we observed for all traits significant differences in the mean of the populations (data not shown). Absolute values of phenotypic correlations among the eight traits were minimum (0.01) between SY and WSC and maximum (0.99) between SY and WSY (Table 3). Genotypic correlations deviated only slightly from phenotypic correlations.

The first two principal coordinates explained 28.5% of the total variation (Fig. 1). The PCoA revealed a population structure of the 14 parents with the following clusters: (1) E and O, (2) D, M, and I, (3) C, J, and N, (4) G and H, (5) K, and (6) A, B, and F. Moreover, progenies were placed with respect to at least one PC in between their parents. The Mantel test of the distance matrix based on the molecular data and the differences in trait means of the nine populations showed no significant associations for all traits (Table 2). Estimates of r^2 decreased with increasing genetic map distance between marker loci (Fig. 2). The 95th percentile of the distribution of r^2 estimates was used as a population-specific threshold to identify LD due to genetic linkage. By this approach, it was estimated that values of $r^2 > 0.25$ were probably due to genetic linkage. The fitted non-linear regression of r^2 values and genetic map distance passed this threshold at 7 cM.

For all eight traits, the genome-wide scans for main effects based on Models A and B identified QTL distributed throughout the whole genome (Table 4; Fig. 3). The number of detected QTL was higher for Model A (39) compared to Model B (22). The proportion of genotypic variance explained by the detected QTL ranged for Model A from 46.6% for WSC to 100% for N. The proportion of genotypic variance explained by the detected QTL was generally lower for Model B compared to Model A with a range from 0% for BY to 61.2% for N. The allele frequencies may have a substantial impact on the power of QTL detection. Therefore, we investigated the association between P values of marker–trait associations and PIC values for all eight traits (Fig. 4). For Model A, only 10% of the detected QTL had a PIC value lower than 0.2. In contrast, for Model B, 27% of the QTL had a PIC value lower than 0.2.

The two-dimensional genome scan based on Model B revealed seven significant epistatic interactions for K and one for SY. Contribution of epistatic effects to the genotypic variance was small with 10% for K and 0.1% for SY. The highest value for the proportion of the genotypic variance explained by the individual digenic epistatic interactions was 0.1% for SY and 3% for K.

Discussion

For sugar beet breeding, several physiological and agronomic traits with varying complexity are of utmost economic relevance (Draycott 2006). This makes sugar beet as an attractive model crop for genetic studies. We used an

Table 2 Means, ranges, genotypic variances (σ^2_G), genotype × location variances ($\sigma^2_{G\times L}$), error variances (σ^2_E), and broad sense heritabilities (h^2) of 197 sugar beet testcross progenies evaluated in 17 locations for potassium content (K, decamol Mg^{-1}), sodium content (Na, hectomol Mg^{-1}), α-amino nitrogen content (N, hectomol Mg^{-1}), sugar content (SC, %), white sugar content (WSC, %), sugar yield (SY, Mg ha^{-1}), white sugar yield (WSY, Mg ha^{-1}), and beet yield (BY, Mg ha^{-1}). Moreover, the correlation between the genetic distance matrix based on (1) the allele frequencies of the nine populations and (2) the differences in trait means [r(GD, TD)] is given

Parameter	K	Na	N	SC	WSC	SY	WSY	BY
Mean	42.2	42.0	82.8	17.1	15.4	14.9	13.63	86.9
Min	37.3	36.8	67.5	16.6	14.9	13.5	12.15	80.6
Max	45.2	53.3	92.9	17.7	15.9	15.9	14.29	92.3
σ^2_G	1.76**	3.21**	10.84**	0.03**	0.03**	0.13**	0.10**	4.56**
$\sigma^2_{G\times L}$	0.25**	0.77**	1.07	0.00**	0.01**	0.02**	0.02**	0.73**
σ^2_E	2.32	51.19	105.55	0.08	0.09	0.34	0.29	11.48
h^2	0.93	0.55	0.67	0.86	0.85	0.87	0.86	0.88
r(GD, TD)a	−0.02	0.38*	0.34	−0.02	0.05	0.23	0.21	0.08

*, ** Significant at the 0.05 and 0.01 probability level, respectively

Table 3 Phenotypic (above diagonal) and genotypic (below diagonal) correlations for testcross performance of 197 sugar beet genotypes evaluated for potassium content (K), sodium content (Na), α-amino nitrogen content (N), sugar content (SC), white sugar content (WSC), sugar yield (SY), white sugar yield (WSY), and beet yield (BY)

	K	Na	N	SC	WSC	SY	WSY	BY
K		0.35**	−0.18*	0.17*	−0.14*	0.48**	0.37**	0.40**
Na	0.18**		−0.60**	−0.35**	−0.50**	0.20**	0.11	0.34**
N	0.12*	−0.37**		0.08	0.16*	−0.14	−0.09	−0.18*
SC	0.16*	−0.48**	0.02		0.95**	0.14*	0.19**	−0.27**
WSC	−0.15*	−0.57**	0.01	1.00**		−0.01	0.07	−0.39**
SY	0.45**	0.13*	−0.13*	0.07	−0.04		0.99**	0.91**
WSY	0.35**	0.06	−0.14*	0.12*	0.03	1.00**		0.88**
BY	0.36**	0.31**	−0.13*	−0.11	−0.40**	0.93**	0.91**	

Values of genotypic correlations exceeding 1 were set to 1

*, ** Significantly different from zero at the 0.05 and 0.01 level of probability, respectively

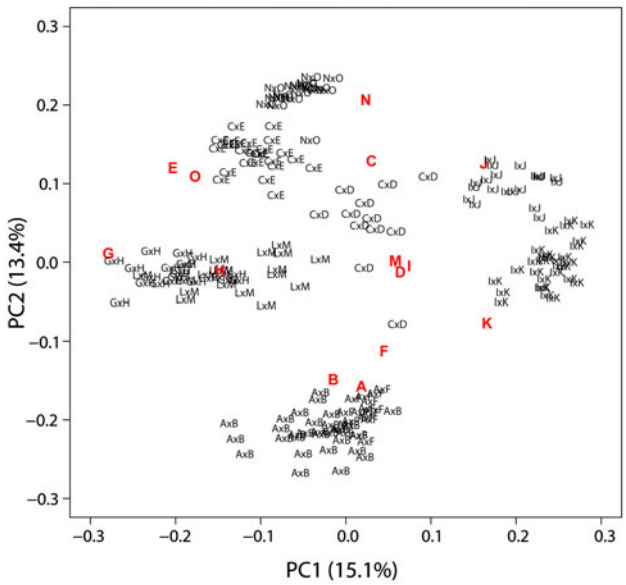

Fig. 1 Principal coordinate analysis of the 14 parents (for detailed information see Table 1) and their 196 progenies based on modified Rogers' distance estimates. Percentages in *parentheses* refer to the proportion of variance explained by the principal coordinate

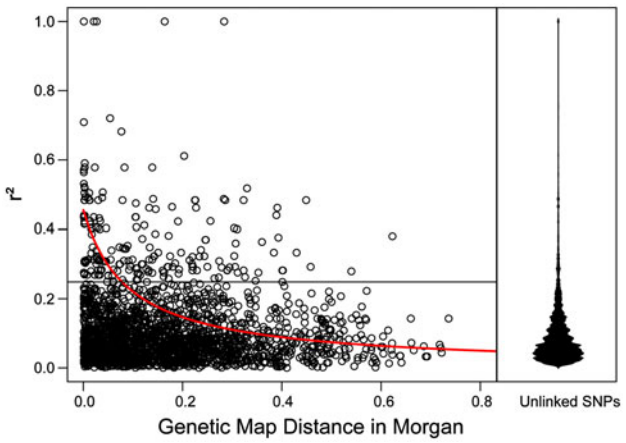

Fig. 2 Linkage disequilibrium (r^2) between linked SNP markers as a function of genetic map distance and the distribution of r^2 values between unlinked SNP markers. *Horizontal line* refers to the 95th percentile of r^2 estimates between unlinked markers. *Curve* was fitted by non-linear regression

elite sugar beet breeding population, and investigated the suitability of joint linkage association mapping to unravel the genetic basis of several complex traits of economic relevance.

Applicability of genome-wide joint linkage association mapping in sugar beet

The power to detect QTL for quantitative traits depends on the sample size and the heritability of the trait under consideration. Our study was based on a set of 196 elite sugar beet testcross progenies (Table 1). Genotypic effects were estimated with high accuracy reflected by heritabilities ranging from 0.55 for Na to 0.93 for K (Table 2). Therefore, the experimental setup should provide insights into the genetic architecture of the eight traits underlying our study.

The resolution of genome-wide joint linkage association mapping is determined by the extent of LD in the parental population caused by genetic linkage. Besides genetic linkage, several forces can substantially contribute to LD such as population stratification, a high degree of relatedness, and presence of selection as well as genetic drift (Stich et al. 2005). We used the approach suggested by Breseghello and Sorrells (2006), and obtained a critical value of $r^2 = 0.25$ to detect significant LD due to genetic linkage. The non-linear regression of LD and genetic map distance indicated that LD due to genetic linkage can be observed even for marker pairs with a genetic map distance of 7 cM (Fig. 2). This may be interpreted as an indicator that the resolution of genome-wide joint linkage

Table 4 Trait-associated markers, the explained proportion of the genotypic variance p_g, and the allele substitution (α) effects of the joint linkage association mapping analyses applying two different statistical models

Chromosome	Position	Model A		Model B		References
		p_g (%)	α effect	p_g (%)	α effect	
Potassium content (K, decamol Mg^{-1})						
1	58.9	11.5	1.19	–	–	Weber et al. (1999, 2000)
2	56.6	7.7	0.69	28.1	1.11	Schneider et al. (2002)
2	57.2	5.2	0.55	14.5	0.89	
5	18.9	–	–	1.0	−0.33	
5	32.1	–	–	0.0	−0.61	
5	54.1	52.2	0.74	–	–	Weber et al. (1999, 2000)
6	79.9	–	–	3.7	−0.57	
7	34.5	–	–	0.9	−0.15	Weber et al. (1999, 2000)
7	41.3	–	–	6.9	−0.17	Weber et al. (1999, 2000)
Total[a]		70.4		49.0		
Sodium content (Na, hectomol Mg^{-1})						
1	62.3	11.1	1.04	22.5	1.50	
4	52.3	25.3	−0.98	–	–	
4	72.0	50.0	0.65	–	–	
7	38.0	13.9	−0.8	–	–	Weber et al. (2000)
9	13.3	–	–	0.3	−0.11	
9	24.0	5.1	−0.81	–	–	
Total[a]		95.6		19.6		
α-Amino nitrogen content (N, hectomol Mg^{-1})						
2	52.3	54.3	2.78	54.3	4.79	Weber et al. (1999, 2000)
2	54.2	1.6	−1.37	8.4	−3.11	Weber et al. (1999, 2000)
3	22.0	4.7	−1.15	–	–	Schneider et al. (2002)
4	42.3	6.1	1.57	–	–	Weber et al. (1999, 2000)
5	18.8	–	–	4.6	0.88	Schneider et al. (2002)
5	18.9	32.3	−2.64	–	–	Schneider et al. (2002)
7	21.2	19.8	2.12	–	–	Weber et al. (1999, 2000)
Total[a]		100		61.2		
Sugar content (SC, %)						
1	62.3	8.1	−0.08	–	–	
2	56.9	4.9	0.12	0.6	0.03	Weber et al. (1999, 2000), Schneider et al. (2002)
3	24.3	–	–	9.1	0.06	
4	13.5	1.9	0.02	–	–	
4	52.6	–	–	2.8	0.04	
7	41.3	22.4	−0.10	–	–	
7	65.0	15.8	0.12	–	–	
9	12.9	–	–	0.4	0.02	
Total[a]		47.7		9.9		
White sugar content (WSC, %)						
1	62.3	6.5	−0.07	–	–	
2	56.6	–	–	6.3	−0.06	
2	56.9	–	–	0.1	0.02	
3	37.6	9.4	0.06	–	–	
4	52.6	–	–	0.1	−0.01	
4	71.6	16.8	0.06	–	–	
5	10.4	3.8	0.05	–	–	

Table 4 continued

Chromosome	Position	Model A		Model B		References
		p_g (%)	α effect	p_g (%)	α effect	
7	41.3	15.6	−0.06	16.5	−0.10	
Total[a]		46.6		19.4		
Sugar yield (SY, Mg ha^{-1})						
2	52.6	–	–	12.6	0.15	
2	60.5	9.2	0.18	–	–	
3	22.5	25.8	0.18	–	–	Weber et al. (1999)
4	72.0	26.6	0.18	–	–	
6	53.8	3.9	−0.11	–	–	Weber et al. (1999)
Total[a]		59.8		11.1		
White sugar yield (WSY, Mg ha^{-1})						
2	52.6	–	–	15.4	0.14	
3	22.5	24.3	0.17	–	–	
4	72.0	22.3	0.20	–	–	
5	26.8	12.9	−0.13	–	–	
6	34.1	13.8	0.17	–	–	
Total		67.2		13.7		
Beet yield (BY, Mg ha^{-1})						
3	16.4	9.1	1.66	–	–	
3	17.8	12.6	−0.89	–	–	
3	25.6	15.7	0.73	–	–	
4	72.0	24.0	1.22	–	–	
5	26.8	3.9	−0.46	–	–	
6	57.6	2.9	0.56	–	–	
Total[a]		61.7		0.00		

[a] The proportion of the genotypic variance of the final fit

association mapping is limited in the underlying elite population. Nevertheless, many marker pairs with a genetic map distance smaller than 7 cM were below the fitted line suggesting that mapping resolution can be substantially higher in some chromosomal regions. This is in accordance with results of Kraft et al. (2000), who observed LD only for very tightly linked markers (<3 cM). An increase in the mapping resolution for chromosomal regions with a high degree of LD due to linkage may be accomplished by the use of genetic resources such as sea beet with a low extent of linkage disequilibrium between markers (Hansen et al. 2001).

In our study, the average genetic map distance between adjacent markers was 3 cM (Supplementary Figure S1). Our result suggests that this marker density can be considered as the lower limit for a genome-wide association mapping survey in this elite sugar beet population. Even though emphasis was given to select equally spaced markers, a few chromosomal regions were not covered with markers. These gaps must be kept in mind when interpreting the QTL results from our study.

Model comparison

For all traits we detected a higher number of QTL and observed a higher proportion of explained genotypic variance for Model A compared to Model B (Table 4). This can be explained by a better exploitation of the genotypic variance among populations for joint linkage association mapping when using only markers as cofactors and disregarding a population effect (Model A) (Yu et al. 2008). Furthermore, we observed lower R^2 values for the fit of the selected co-factors for Model A than for Model B (data not shown). Consequently, the control of the genetic background is less stringent while applying Model A compared to Model B. This harbors the risk to detect a higher number of false positive marker–trait associations. We observed correlations between the genetic distance matrix based on (1) the allele frequencies of the nine populations and (2) the differences in trait means ranging from −0.02 for K and SC to 0.38 for Na (Table 2). Würschum et al. (2010) investigated the consequences of the correlation between genetic distances and the differences in trait means on the number

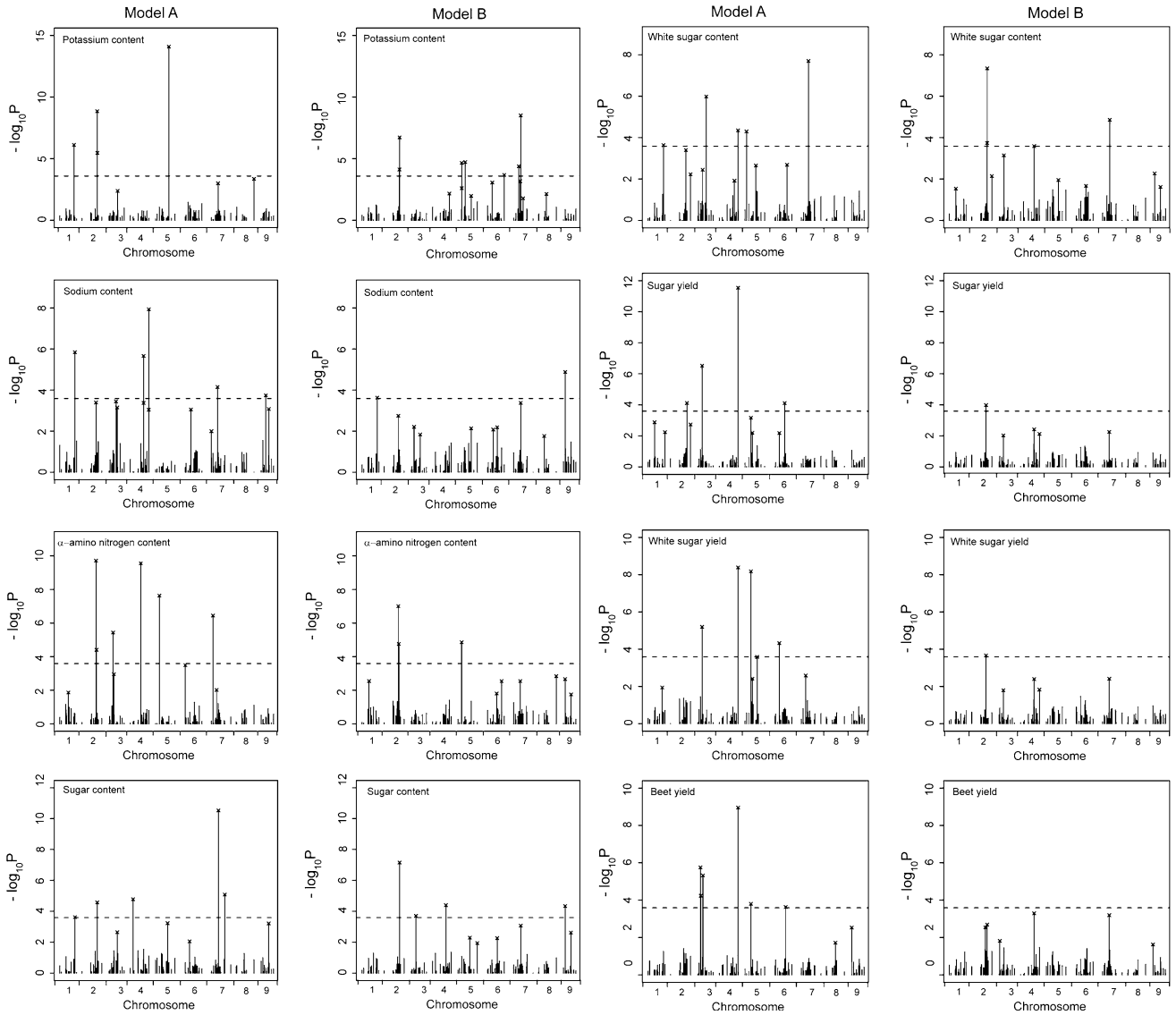

Fig. 3 SNP markers associated with eight traits in sugar beet based on two statistical models. The *horizontal line* refers to a threshold of $P < 0.05$ applying a Bonferroni–Holm correction for multiple tests. *Multiplication symbol* indicates the selected cofactors

of false positives in association mapping designs. Their results indicated that even for traits with a moderate correlation between genetic distance and the trait means correction for population stratification is required to minimize the risk to detect an increased number of false positive QTL. Thus, for Na, N, SY, and WSY, the number of QTL detected based on Model A may be inflated.

Only a small fraction (11%) of the total number of QTL was detected in the joint linkage association mapping based on Models A and B (Table 4). This low overlap can at least partially be explained by the significant differences in the means of the nine populations. Our finding also indicates that both models are complementary: while Model A exploits the genotypic variance among populations more efficiently, Model B facilitates the detection of minor QTL due to a more stringent control of the genetic background (Table 4). Moreover, more QTL were detected at loci with small PIC values with Model B compared to Model A (Fig. 4). Alleles with low frequency, as indicated by low PIC values, can be very likely lost during the phenotypic selection even if the minor allele possesses a positive effect because of genetic drift in elite breeding populations. Therefore, these QTL are of particular interest for marker-assisted selection. Summarizing, the findings of our study clearly suggests that it is advantageous to apply both complementary statistical approaches, Models A and B, for joint linkage association mapping.

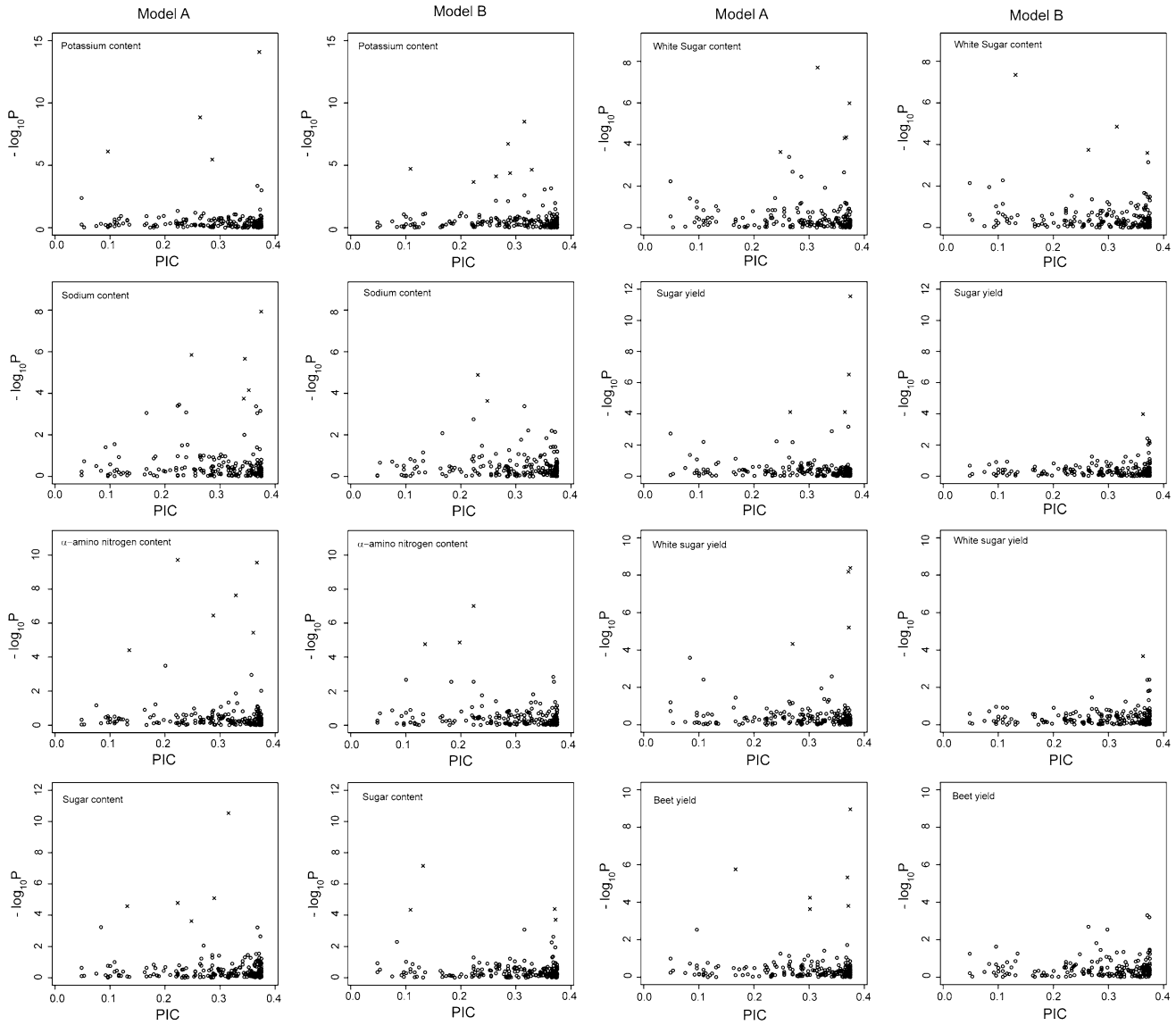

Fig. 4 Relationship between polymorphic information content of SNPs and *P* value for SNP–trait associations based on two statistical models. *Multiplication symbol* refers to identified QTLs

Genetic basis of important physiological and agronomic traits

We performed a literature review for QTL reported for the eight traits K, Na, N, SC, WSC, SY, WSY, and BY (Weber et al. 1999, 2000; Schneider et al. 2002) and compared the published QTL positions with those of our study. Although the lack of an integrated map hinders a precise validation of QTL, it may serve as a rough guideline to confirm our findings. Several of the QTL detected in our study co-located with previously reported QTL (Table 4). Furthermore, we observed several major QTL explaining >10% of the genotypic variance, which were not reported in other studies. This shows that genome-wide joint linkage association mapping holds the potential to detect previously unknown QTL and is well suited to unravel the genetic basis of complex traits.

In total, ten markers were associated with at least two traits (Table 4). Multi-trait QTL were also reported in previous studies in sugar beet (e.g., Schneider et al. 2002; Stich et al. 2008b), and can be explained either by pleiotropic effects of a single gene or by tightly linked genes. Differentiation between both causes is of relevance if the positive allele for one trait possesses a detrimental effect for the other trait. For the detected multi-trait QTL, the majority showed allele substitution effects in the same desired direction. Consequently, the discrimination between linkage and pleiotropy is of less interest.

In accordance with previous studies on agronomic important traits in sugar beet (e.g., Schneider et al. 2002),

the effects of single QTL were of considerable size explaining up to 54.3% of the genotypic variation (Table 4). QTL with large effects are promising candidates for marker-assisted selection. Nevertheless, large estimates of QTL size may also reflect small QTL effects estimated with a large bias (Utz et al. 2000). Cross-validation (Melchinger et al. 1998) or validation in an independent sample has been suggested (Lande and Thompson 1990) to obtain unbiased estimates of QTL effects in the context of linkage mapping. Cross-validation, especially for large sample sizes, is an efficient validation approach (Melchinger et al. 1998). When using joint linkage association mapping with an inherent population structure, however, it is not obvious from which pool to draw samples for estimating the QTL parameters in a resampling strategy. Therefore, we suggest validating the effect sizes of QTL in an independent experiment.

Epistasis among QTL is expected to be less relevant in studies based on germplasm selected from one heterotic group because of the narrow genetic base and the low probability to disrupt co-adapted gene complexes in the parents (Melchinger et al. 1998). The power to detect epistasis with testcross performance is also low because of masking effects of the tester (Gallais and Rives 1993). Furthermore, genetic contribution of the 196 genotypes to the testcross progenies was 25%. Consequently, relative contribution of additive variance to the genetic variance among testcross progenies is eight times larger than for additive × additive variance resulting in a lower power to detect epistatic compared to main effects. In accordance with this expectation, digenic epistatic interactions were detected only for two out of the eight traits and contributed only little to the explained genotypic variation. This finding is in accordance with the results reported for testcross performance of complex traits in other cross-pollinating species such as maize (Mihaljevic et al. 2005). Consequently, our results may be interpreted as an indicator that epistasis can be ignored in sugar beet breeding. One important implication of this is that predicting the performance of three-way hybrids—the predominating variety types in sugar beet—from the means of their single crosses is very efficient (Melchinger et al. 1987). The genetic contribution of male lines to testcross progenies (50%) is owing to the common mating design (female × CMS-L) × male, higher than that of the female lines (25%). Therefore, mapping epistasis within the male compared to the female heterotic pool possesses a higher power to detect epistasis and represents a promising strategy to validate the findings of our study.

Acknowledgments This research was conducted within the Biometric and Bioinformatic Tools for Genomics based Plant Breeding project supported by the German Federal Ministry of Education and Research (BMBF) within the framework of GABI-FUTURE initiative.

References

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5:618–624

Chao S, Zhang W, Dubcovsky J, Sorrells ME (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci 47:1018–1030

Draycott AP (2006) Sugar beet, 1st edn. Blackwell, Oxford

Fisher RA (1921) On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1:1–32

Gallais A, Rives M (1993) Detection, number and effects of QTLs for a complex character. Agronomie 13:723–738

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml user guide release 2.0. VSN International Ltd, Hermel Hempstead, UK

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Hansen M, Kraft T, Ganestam S, Sall T, Nilsson NO (2001) Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. Genet Res 77:61–66

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

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

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Maurer HP, Melchinger AE, Frisch M (2008) Population genetic simulation and data analysis with Plabsoft. Euphytica 161:133–139

Melchinger AE, Geiger HH, Seitz G, Schmidt GA (1987) Optimum prediction of three-way crosses from single crosses in forage maize (*Zea mays* L.). Theor Appl Genet 74:339–345

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

Mihaljevic R, Utz HF, Melchinger AE (2005) No evidence for epistasis in hybrid and per se performance of elite European flint maize inbreds from generation means and QTL analyses. Crop Sci 45:2605–2613

Piepho H-P, Williams ER, Fleck M (2006) A note on the analysis of designed experiments with complex treatment structure. HortScience 41:446–452

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE (2005) Temporal trend of genetic diversity in European maize germplasm. Theor Appl Genet 111:838–845

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

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

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111: 723–730

Stich B, Piepho H-P, Schulz B, Melchinger AE (2008a) Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). Theor Appl Genet 117:947–954

Stich B, Melchinger AE, Heckenberger M, Möhring J, Schechert A, Piepho H-P (2008b) Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). Theor Appl Genet 117:1167–1179

Stram DO, Lee JW (1994) Variance components testing in longitudinal mixed effects model. Biometrics 50:1171–1177

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849

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

Wright S (1978) Evolution and genetics of populations, variability within and among natural populations, 4th edn. The University of Chicago Press, Chicago, p 91

Würschum T, Maurer HP, Schulz B, Möhring J, Reif JC (2010) Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. Theor Appl Genet (in review)

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–363