Theor Appl Genet (2011) 123:1121–1131
DOI 10.1007/s00122-011-1653-1

ORIGINAL PAPER

Genome-wide association mapping of agronomic traits in sugar beet

Tobias Würschum · Hans Peter Maurer · Thomas Kraft · Geert Janssen · Carolina Nilsson · Jochen Christoph Reif

Received: 3 March 2011 / Accepted: 28 June 2011 / Published online: 15 July 2011

Abstract Recent results indicate that association mapping in populations from applied plant breeding is a powerful tool to detect QTL which are of direct relevance for breeding. The focus of this study was to unravel the genetic architecture of six agronomic traits in sugar beet. To this end, we employed an association mapping approach, based on a very large population of 924 elite sugar beet lines from applied plant breeding, fingerprinted with 677 single nucleotide polymorphism (SNP) markers covering the entire genome. We show that in this population linkage disequilibrium decays within a short genetic distance and is sufficient for the detection of QTL with a large effect size. To increase the QTL detection power and the mapping resolution a much higher number of SNPs is required. We found that for QTL detection, the mixed model including only the kinship matrix performed best, even in the presence of a considerable population structure. In genome-wide scans, main effect QTL and epistatic QTL were detected for all six traits. Our full two-dimensional epistasis scan revealed that for complex traits there appear to be epistatic master regulators, loci which are involved in a large number of epistatic interactions throughout the genome.

Communicated by I. Mackay.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1653-1) contains supplementary material, which is available to authorized users.

T. Würschum (✉) · H. P. Maurer · J. C. Reif
State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: tobias.wuerschum@uni-hohenheim.de

T. Kraft · G. Janssen · C. Nilsson
Syngenta Seeds AB, Box 302, 261-23 Landskrona, Sweden

Introduction

Association mapping is emerging as a novel tool in plant genomics (Myles et al. 2009) and has recently been shown to be suited for the analysis of populations from applied breeding programs (e.g. Reif et al. 2011a; Würschum et al. 2011). The detection of QTL in these populations offers the advantage that QTL are identified which are of direct relevance for breeding. A potential problem of association mapping in plant populations is the inherent population structure. Any nonfunctional associations between the trait and the underlying population structure will also be detected as QTL (Zhao et al. 2007). Two methods have recently been suggested to correct for population stratification (*P* matrix) and familial relatedness (*K* matrix) (Price et al. 2006; Yu et al. 2006; Astle and Balding 2009; Cockram et al. 2010). As both, the *K* and the *P* matrix, are estimated based on marker data, their simultaneous use may result in overcorrection for population structure leading to a reduced power to detect QTL. Using experimental data in sugar beet, it has recently been shown that in breeding populations without a major population structure, the *K* matrix is sufficient to control the genetic background (Würschum et al. 2011).

Association mapping is based on linkage disequilibrium (LD) between the examined molecular markers and QTL associated with the trait. The mapping resolution of association mapping is expected to be higher than in classical linkage mapping as it exploits all the recombination events that have occurred during the history of the population. LD is population specific and affected by many genetic factors (Flint-Garcia et al. 2003). Moreover, LD is highly variable across the genome. The power to detect QTL depends on the strength of the LD between the marker and the QTL. High r^2 values are required to detect medium and small

size QTL whereas lower r^2 values will only allow the detection of QTL with large effects. In plant breeding populations LD is expected to be higher than in natural populations due to the selection of favorable genotypes and the shorter history of that germplasm. Recent results, however, have indicated that the extent and structure of LD enables a good mapping resolution also in populations derived from breeding programs (Van Inghelandt et al. 2010).

Epistasis is defined as interactions between alleles from different loci (Carlborg and Haley 2004). Consequently, the presence of epistasis will impede the prediction of the phenotype of an individual simply by the sum of its single-locus effects (Lynch and Walsh 1998). There is accumulating evidence for the presence of epistasis in model organisms at the molecular and biochemical level (e.g. Tong et al. 2004; St Onge et al. 2007; He et al. 2010; Costanzo et al. 2010), but little attention has been paid to the importance of epistasis for the performance of elite breeding germplasm (Li et al. 2010; Miedaner et al. 2010; Reif et al. 2011a, b; Würschum et al. 2011).

Sugar beet (*Beta vulgaris* L.) is of major importance for sugar production worldwide (Draycott 2006). It is also of interest from a scientific point of view, as it serves as an excellent model crop to study the genetic architecture of yield-related and physiological traits such as sodium, potassium and α-amino nitrogen content. Different mapping approaches have identified QTL for the abovementioned traits (Weber et al. 1999; Weber et al. 2000; Schneider et al. 2002; Stich et al. 2008; Reif et al. 2010; Würschum et al. 2011). Nevertheless, advances in sequencing technologies have increased the availability of molecular markers, and an analysis of the genetic architecture of these traits using high-density genetic maps is still lacking.

The objectives of our study were to (1) estimate the marker density required for genome-wide association mapping in a vast elite breeding population of sugar beet, (2) compare biometric models to correct for different levels of population structure with regard to the QTL detection power, and (3) dissect the genetic architecture of agronomically important yield and physiological traits.

Materials and methods

Plant materials, field experiments, and molecular markers

This study was based on 924 diploid elite sugar beet (*B. vulgaris* L.) inbred lines. Testcross progenies were produced by crossing the genotypes to a single-cross hybrid as tester. All material used in this study was provided by the breeding company Syngenta Seeds AB (Sweden).

The 924 genotypes were evaluated in routine plant breeding trials with two replicates at 1–7 locations with on average 3 locations per genotype in 2008. The evaluated traits were white sugar yield (WSY, t ha^{-1}), sugar content (SC, %), root yield (RY, t ha^{-1}), potassium (K, mM), sodium (Na, mM) and α-amino nitrogen (N, mM) (Supplementary Figure 1).

The 924 genotypes were fingerprinted following standard protocols with 677 single nucleotide polymorphism (SNP) markers. SNP discovery was based on sequencing of RDLP markers and public ESTs. These markers were randomly distributed across the sugar beet genome with an average marker distance of 1 cM and the two largest gaps between adjacent markers had 11 and 23 cM (Supplementary Figure 2). 97% of the adjacent markers had a genetic map distance of <5 cM and 85% of <2 cM. Map positions of all markers were based on the linkage map of Syngenta Seeds AB with a total map length of 698 cM (unpublished data).

Phenotypic and molecular data analyses

The analyses were based on adjusted entry means [best linear unbiased estimates (BLUEs)] calculated for each location. The following linear mixed model was used to estimate variance components of the testcrosses: $y_{ij} = \mu + l_j + g_i + e_{ij}$, where y_{ij} is the adjusted entry mean of the ith sugar beet line at the jth location, μ the intercept term, l_j the effect of the jth location, g_i the genetic effect of the ith sugar beet line, and e_{ij} the error term including the genotype times location interaction effect. Locations and genotype were modeled as a random effect. Variance components were determined by the restricted maximum likelihood (REML) method. Significance for variance component estimates was tested by model comparison with likelihood ratio tests where the halved P values were used as an approximation (Stram and Lee 1994). Heritability (h^2) on an entry-mean basis was estimated as the ratio of genotypic to phenotypic variance according to Melchinger et al. (1998), based on the average replicate number. Furthermore, genotypes were regarded as fixed effects and BLUEs were determined for all genotypes and traits.

Associations among the 924 genotypes were analyzed by applying principal coordinate analysis (PCoA) (Gower 1966) based on the modified Rogers' distances of the individuals (Wright 1978). LD was assessed by the LD measure r^2 (Weir 1996) and significance of LD was tested with Fisher's exact tests (Hill and Robertson 1968). LD and PCoA computations were performed with the software package Plabsoft (Maurer et al. 2008).

Association mapping

The mixed model for the association mapping approach was: $y_{ijp} = \mu + a_p + g_i + l_j + e_{ijp}$, where y_{ijp} is the adjusted entry mean of the ith sugar beet line at the jth location carrying allele p, μ the intercept term, a_p the allele substitution effect of allele p, g_i the genetic effect of the ith sugar beet line, l_j the effect of the jth location, and e_{ijp} the residual. The allele substitution effect a_p was modeled as fixed effect whereas g_i and l_j were regarded as random effects.

Association mapping populations often show a non-functional correlation between the phenotype and the population structure, which will lead to the detection of false positive QTL. It has previously been shown that correction with the kinship matrix, K, is sufficient in populations from applied plant breeding programs (Würschum et al. 2011). Nevertheless, as the population used in this study shows a more pronounced population structure, two additional models correcting for population structure were tested. The K model included only the kinship matrix. The KP_{Reml} model and the KP_{Cor} model included the K matrix and principal coordinates (PCs) (Price et al. 2006). For the KP_{Reml} model PCs were sequentially added as fixed effects to the model until the first PC appeared which was not significant any more at $P < 0.01$ based on the Wald F statistic. The KP_{Cor} model included all PCs which were significantly associated ($P < 0.01$) with the BLUEs of the six traits. In all three models the variance of the random genetic effect was assumed to be $\mathrm{Var}(g) = K\sigma_g^2$, where σ_g^2 refers to the genetic variance estimated by REML and K was a 924×924 matrix of kinship coefficients that define the degree of genetic covariance between all pairs of entries. We followed the suggestion of Bernardo (1993) and calculated the kinship coefficient K_{ij} between inbreds i and j on the basis of marker data as $K_{ij} = 1 + (S_{ij} - 1)/(1 - T_{ij})$, where S_{ij} is the proportion of marker loci with shared variants between inbreds i and j, and T_{ij} is the average probability that a variant from one parent of inbred i and a variant from one parent of inbred j are alike in state, given that they are not identical by descent. The coefficient T_{ij} was estimated separately for each model and trait using a REML method setting negative kinship values between inbreds to zero.

For the detection of main effect QTL, a genome-wide scan for marker-trait associations was conducted. Significance of the allele substitution effect a_p was assessed based on the Wald F statistic. To control for multiple testing, we followed the suggestion of Kraakman et al. (2004) and tested at a false discovery rate (FDR) of 0.20 (Benjamini and Hochberg 1995). A two-dimensional genome scan was performed and all possible marker–marker interactions were tested. The statistical model for the epistasis scan was: $y_{ijvw} = \mu + m_v + m_w + m_v{:}m_w + g_i + l_j + e_{ijvw}$, where m_v and m_w denote the effect of the vth and wth marker genotype and $m_v{:}m_w$ refers to the interaction effect of the vth and wth marker genotype. As for the main effects, epistatic QTL were detected with FDR <0.20. All mixed model calculations were performed using the software ASReml 2.0 (Gilmour et al. 2006).

The allele substitution effect was estimated in the mixed model by a simultaneous fit of all significant QTL. The total proportion of genotypic variance (p_{G}) explained by the detected QTL was calculated by fitting all QTL simultaneously in a linear model to obtain R^2_{adj}. The ratio $p_{\mathrm{G}} = R^2_{\mathrm{adj}}/h^2$ yielded the proportion of genotypic variance (Utz et al. 2000).

Cross validation was applied to obtain an estimate of the bias in the proportion of genotypic variance explained by the detected QTL. We used fivefold cross validation with 80% of the lines (739) as estimation set in which the effects of the QTL detected in the full data set were estimated. These estimated effects were then used for a prediction in the remaining 20% of lines (185). 1,000 runs were performed and the mean R^2_{adj} between predicted and observed phenotypic values in the test set was then used to estimate the cross validated proportion of genotypic variance.

Results

The genotypic variances estimated in the population of 924 sugar beet lines were significantly larger than zero ($P < 0.01$) for all six traits (Table 1). Heritability ranged from 0.38 for α-amino nitrogen to 0.70 for potassium. Absolute values of phenotypic correlations among the six traits were minimum between WSY and sodium (0.03) and maximum between WSY and RY (0.87) (Fig. 1).

To obtain an estimate for the mapping resolution in our population, we analyzed the extent of LD. The association of LD and genetic map distance indicated that intra-chromosomal LD decayed rapidly within a genetic distance of approximately 10 cM and was rather low even for closely linked markers (Fig. 2, Supplementary Figure 3). The average r^2 between adjacent marker pairs was 0.10 but was variable across the genome (Supplementary Figure 4). The average of the highest r^2 of each marker with any other marker was 0.32.

Principal coordinate analysis revealed a considerable population structure (Fig. 3a). The first PC alone explained 24.0% of the variance (Fig. 3b) and the violin plot revealed that several of the first ten PCs were characterized by a density distribution indicative of a population structure (Fig. 3c). Kinship between the 924 sugar beet lines was estimated by assessing the genetic similarity which revealed presence of a family structure (Fig. 3d).

Table 1 Mean and range of the testcross performance of 924 sugar beet genotypes, variances of the testcrosses for genotypes (σ^2_G) and residuals (σ^2_E), and heritabilities (h^2) for white sugar yield (WSY; t ha^{-1}), sugar content (SC; %), root yield (RY; t ha^{-1}), sodium content (Na; mM), potassium content (K; mM), and α-amino nitrogen content (N; mM)

Parameter	WSY	SC	RY	Na	K	N
Mean	14.63	18.78	86.13	0.34	3.27	0.99
Min	11.89	17.77	72.07	0.15	2.52	0.59
Max	16.35	19.93	98.38	0.68	3.99	1.62
σ^2_G	0.18**	0.059**	8.18**	1.88e−3**	0.029**	7.51e−3**
σ^2_E	0.42	0.096	16.60	4.02e−3	0.030	29.59e−3
h^2	0.51	0.60	0.55	0.53	0.70	0.38

** Significance at $P < 0.01$

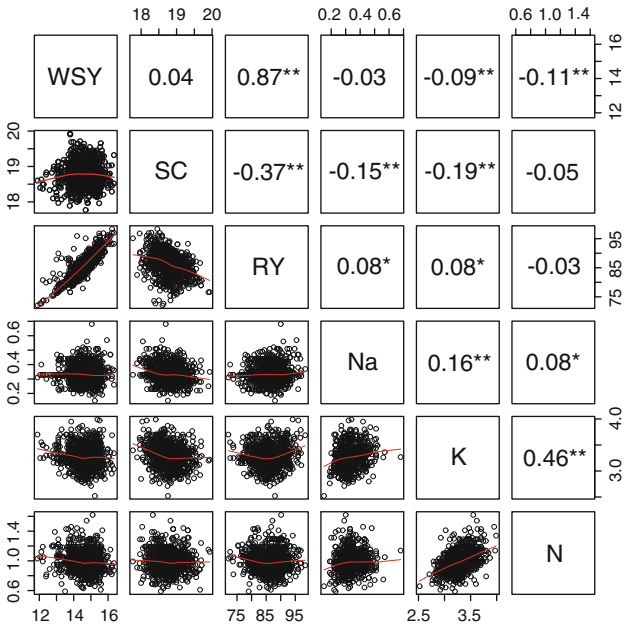

Fig. 1 Correlations between phenotypic values of the testcross performance of 924 sugar beet genotypes evaluated for six traits [*WSY* white sugar yield (t ha^{-1}), *SC* sugar content (%), *RY* root yield (t ha^{-1}), *Na* sodium content (mM), *K* potassium content (mM), *N* α-amino nitrogen content (mM)]. The *lower part* shows the bivariate scatterplots with a fitted line (*, **significantly different from zero with $P < 0.05$ or $P < 0.01$, respectively)

The observed population and family structure prompted us to test three different models to correct for population stratification and familial relatedness. The *K* model included only the kinship (*K*) matrix whereas the KP_{Reml} and the KP_{Cor} models included the *K* matrix and PCs (Supplementary Tables 1, 2). We found that the optimum identity-by-state estimates (*T* value) were almost identical for the three models (Supplementary Figure 5). As a control, we included a simple model without any correction for population structure. The plots of observed versus expected *P* values revealed for all traits strong bulges at the right side of the plot for the simple model (Supplementary Figure 6). The other three models were very similar and distributions closely followed the diagonal indicating a good control for population structure. A comparison between the three models regarding the number of detected QTL and the explained genotypic variance indicated that, in contrast to our expectation, the *K* model was even more conservative and detected fewer QTL for some traits (WSY, SC, N) than the models including the additional *P* matrix (Table 2). Also for those traits for which the *K* model detected more QTL (RY, Na), this difference was negligible given a comparable convergence, indicating that the *K* model is not prone to an enhanced false positive rate due to an insufficient control of the genetic background. The comparison of the detected QTL for traits for which the number of QTL and their explained genetic variance was comparable between the models showed, that most QTL were detected by all three models but single QTL being detected only by one model (Fig. 4). Taking into account the number of markers for which the mixed model converged, the *K* model outperformed the models including a *P* matrix and further analyses were therefore done with the *K* model.

The full genome scan for main effect QTL identified QTL for all six analyzed traits (Table 3, Supplementary Figure 7). The proportion of genotypic variance explained by the detected QTL was lowest for α-amino nitrogen (10.4%) and highest for WSY (73.4%). The highest amount of genotypic variance explained by a single marker was found for RY (31.8%; Table 3). The fivefold cross validation resulted in a similar proportion of genotypic variance explained by the detected QTL in the test set, compared to the full data set (Fig. 5). The two-dimensional epistasis scan revealed a high number of significant epistatic interactions (Fig. 6). The proportion of significant epistatic QTL from the total number of tests was 0.6% (WSY), 2.1% (SC), 3.6% (RY), 1.5% (Na), 1.8% (K) and 0.4% (N).

Discussion

The increasing availability of molecular markers and the steadily decreasing costs for genotyping have paved the

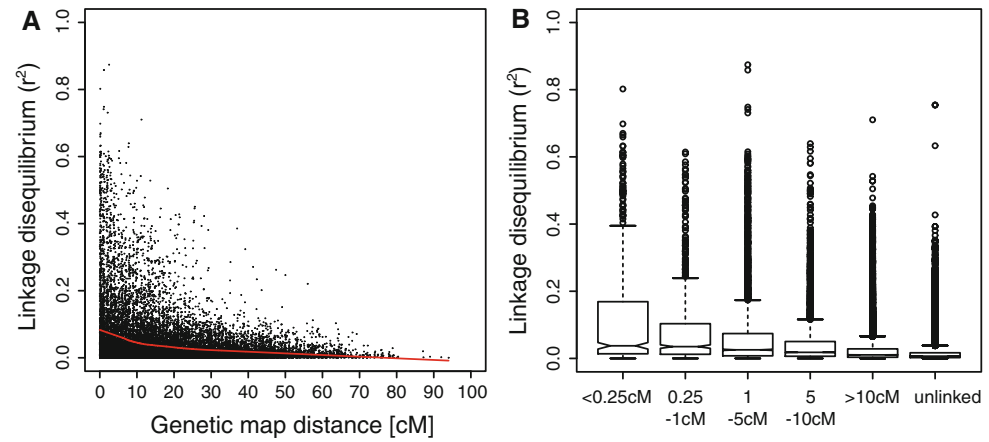

Fig. 2 **a** Linkage disequilibrium (LD) as a function of genetic map distance. Curve was fitted by robust locally fitted regression and shows that LD decays with genetic map distance. **b** Distribution of LD between linked marker pairs for different genetic map distances between the two markers as well as the LD observed between unlinked marker pairs (unlinked)

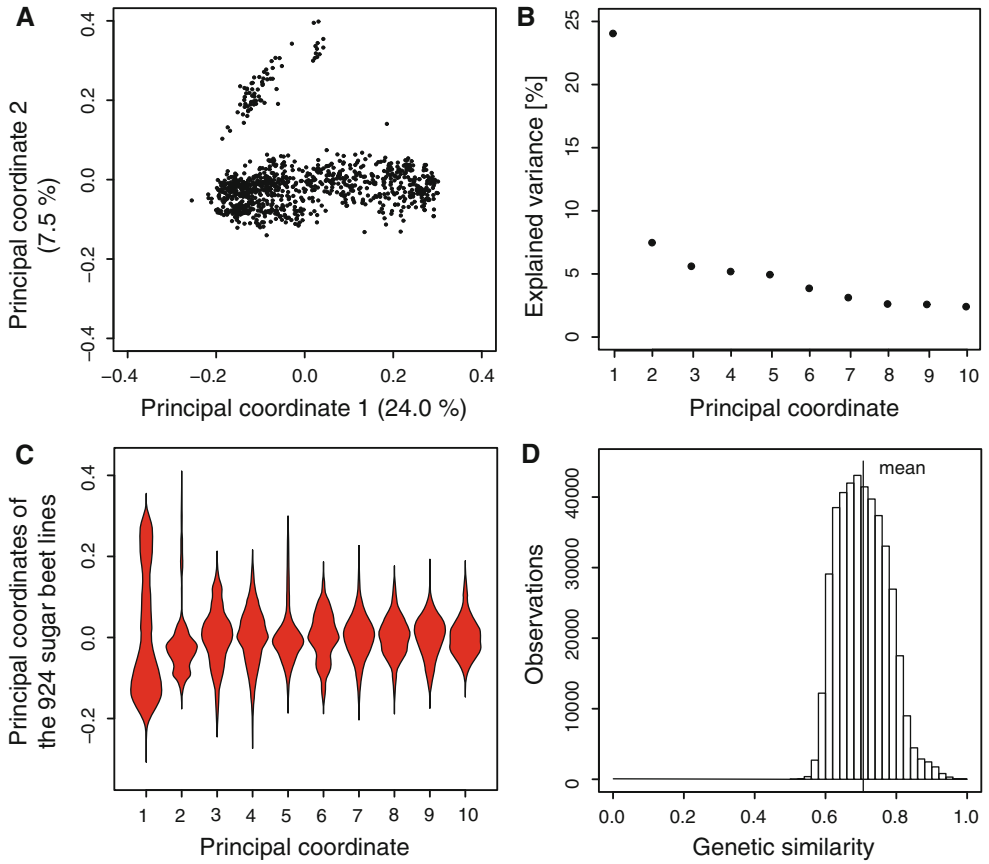

Fig. 3 Analysis of population structure and familial relatedness. **a** Principal coordinate analysis of the 924 genotypes of the population, based on modified Rogers' distance estimates. **b** Explained variance of the first ten principal coordinates. **c** Violin plot showing the density distribution of the first ten principal coordinates for the genotypes from the population. **d** Histogram of the genetic similarities among the 924 genotypes

way for advanced genomic tools also in crop species. Association mapping is a promising avenue to use routinely generated phenotypic data in combination with high-density marker information to investigate the genetic architecture of agronomic traits. This stimulated us to analyze six physiological and agronomic traits in a very large sugar beet association mapping population and to investigate the advantages and limitations of this approach.

Table 2 Comparison of the number of detected main effect QTL and the explained genotypic variance (p_G) of these QTL for the three statistical models *K*, KP_{Reml} and KP_{Cor} for all six traits

Model	WSY	SC	RY	Na	K	N
K						
conv. (%)	96.0	100	99.9	100	100	85.5
QTL	8	4	10	15	4	9
p_G (%)	73.4	38.1	63.1	58.7	12.8	10.4
KP_{Reml}						
conv. (%)	100	100	99.9	100	100	–
QTL	13	9	8	13	4	–
p_G (%)	80.5	43.7	49.4	47.7	12.8	–
KP_{Cor}						
conv. (%)	100	100	99.9	2.7	100	64.5
QTL	8	4	7	2	2	14
p_G (%)	54.5	19.6	46.4	14.2	5.6	21.4

For better comparison, the percentage of markers for which the model converged is given (conv.). For α-amino nitrogen content no PC was selected for the KP_{Reml} approach (see Supplementary Table 1) and it is therefore similar to the *K* model

WSY white sugar yield (t ha^{-1}), *SC* sugar content (%), *RY* root yield (t ha^{-1}), *Na* sodium content (mM), *K* potassium content (mM), *N* α-amino nitrogen content (mM)

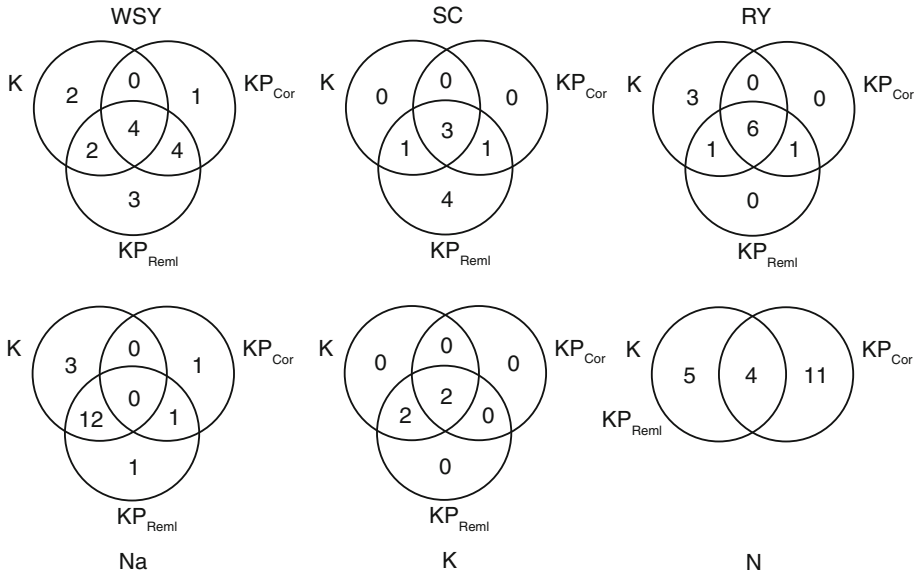

Fig. 4 Venn diagrams each showing common and model-specific main effect QTL detected for the six traits [*WSY* white sugar yield (t ha^{-1}), *SC* sugar content (%), *RY* root yield (t ha^{-1}), *Na* sodium content (mM), *K* potassium content (mM), *N* α-amino nitrogen content (mM)] with the three different models (*K*, KP_{Reml} and KP_{Cor})

Population structure in a large breeding population

The population analyzed in this study represents the largest sugar beet population analyzed in an association mapping approach thus far. The PCoA revealed a pattern characteristic for breeding populations. Whereas most of the lines cluster together in a cloud of more or less closely related genotypes, some genotypes are more distant from the others (Fig. 3a, d). The group of lines in the upper part of Fig. 3a results from crosses between established elite lines and more exotic lines which are being introgressed into the breeding pool. In breeding populations, some elite lines typically are used in many crosses and thus contribute to a higher extent to the population than other lines. This and the fact that breeders favor certain parental combinations can explain the observed population structure. Consequently, a correction for population structure might be required for this data set.

Linkage disequilibrium and mapping resolution

The average r^2 value between neighboring markers was low with 0.1 (Fig. 2) and despite some adjacent marker pairs showing high r^2 values above 0.6, we detected only

Table 3 Trait-associated markers, estimated allele substitution (α) effect, the explained genotypic variance (p_G) and QTL reported in the literature in the same region for the six traits

Marker	Chr.	Position (cM)	α-Effect	p_G (%)	Reported in literature
WSY					
m183	3	45	−0.04	1.2	Weber et al. (1999)
m192	3	49	0.11	6.1	Weber et al. (1999)
m207	3	71	0.34	11.5	
m259	4	37	0.17	26.4	Schneider et al. (2002)
m279	4	62	−0.11	2.9	Reif et al. (2010)
m377	5	71	−0.16	2.0	
m418	6	57	−0.19	14.8	Weber et al. (1999, 2000)
m553	8	24	−0.12	9.8	Weber et al. (1999, 2000)
SC					
m192	3	49	−0.10	12.9	Weber et al. (1999, 2000)
m224	4	10	0.27	5.6	Weber et al. (2000), Reif et al. (2010)
m233	4	16	−0.12	0.0	Weber et al. (2000), Reif et al. (2010)
m349	5	45	−0.08	20.2	Weber et al. (1999, 2000)
RY					
m192	3	49	0.90	9.1	
m202	3	61	−1.76	7.3	
m203	3	67	0.47	0.0	
m207	3	71	1.90	4.4	
m224	4	10	−2.44	1.8	Schneider et al. (2002)
m259	4	37	0.76	3.1	Schneider et al. (2002)
m411	6	53	0.79	31.8	Reif et al. (2010)
m453	6	74	−0.61	0.8	
m456	6	79	−1.54	0.9	
m553	8	24	−0.70	5.7	
Na					
m141	2	78	−0.01	6.5	Weber et al. (1999)
m213	3	79	0.01	0.0	
m230	4	16	−0.06	0.8	
m260	4	38	−0.01	0.5	Weber et al. (1999, 2000)
m276	4	56	0.01	16.0	Reif et al. (2010)
m325	5	41	−0.05	6.6	
m416	6	56	−0.01	10.8	
m423	6	58	0.01	0.4	
m446	6	65	0.01	0.2	
m519	7	48	−0.00	0.0	Weber et al. (1999, 2000), Reif et al. (2010)
m520	7	49	0.01	1.2	Weber et al. (1999, 2000), Reif et al. (2010)
m547	8	17	−0.01	4.2	Weber et al. (1999, 2000)
m607	9	16	0.02	9.3	Reif et al. (2010)
m618	9	26	0.01	5.0	Reif et al. (2010)
m645	9	49	0.01	0.0	Weber et al. (1999, 2000)
K					
m277	4	58	0.05	0.9	
m325	5	41	−0.14	3.7	Weber et al. (1999, 2000), Reif et al. (2010)
m327	5	42	−0.05	6.0	Weber et al. (1999, 2000), Reif et al. (2010)
m597	8	65	0.08	2.8	Weber et al. (2000)
N					
m197	3	54	−0.02	0.5	Weber et al. (1999, 2000)

Table 3 continued

Marker	Chr.	Position (cM)	α-Effect	p_G (%)	Reported in literature
m232	4	16	0.01	0.0	Weber et al. (1999)
m286	5	0	0.09	3.8	
m397	6	42	0.01	0.0	
m486	7	23	0.01	0.6	Weber et al. (1999), Reif et al. (2010)
m552	8	24	0.03	1.4	Weber et al. (1999)
m554	8	25	−0.06	0.1	Weber et al. (1999)
m577	8	44	−0.03	2.0	
m592	8	57	0.03	4.9	

WSY white sugar yield (t ha^{-1}), *SC* sugar content (%), *RY* root yield (t ha^{-1}), *Na* sodium content (mM), *K* potassium content (mM), *N* α-amino nitrogen content (mM)

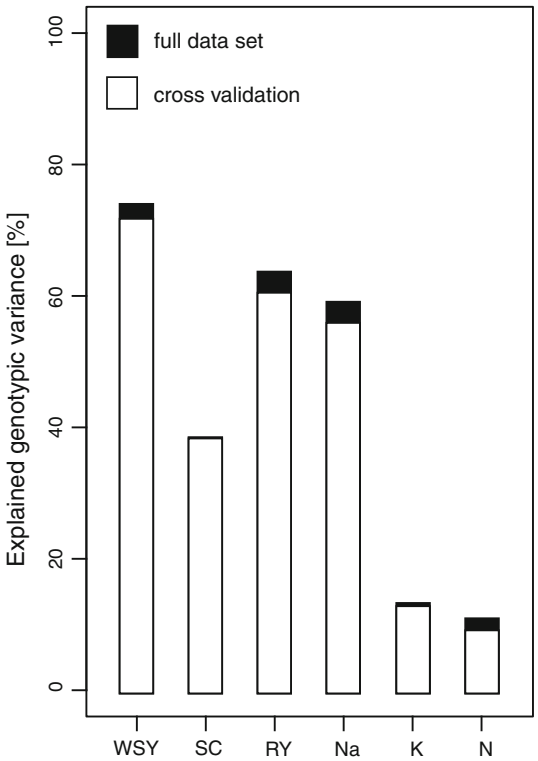

Fig. 5 Explained genotypic variance of the detected QTL in the full data set (924 lines) and in fivefold cross validation. For the cross validation 80% (739 lines) were used to estimate the effects of the QTL, and the prediction was done in the remaining 20% (185 lines). The genotypic variance that was explained with the estimated effects in the prediction is given (*WSY* white sugar yield, *SC* sugar content, *RY* root yield, *Na* sodium content, *K* potassium content, *N* α-amino nitrogen content)

low average LD even for closely linked markers. This is in accordance with results from maize populations from breeding programs where the LD pattern was similar to the LD observed in this study (Van Inghelandt et al. 2010). The proportion of explained genotypic variance that can be explained by a marker compared to that explained by the linked QTL is directly proportional to the r^2 value between the marker and the QTL. The r^2 value of 0.1 between adjacent markers present in this data set has therefore recently been suggested as the minimum threshold to detect associations for comparably large QTL in populations of reasonable size (Ersoz et al. 2007). It must, however, be noted, that LD will always show a more irregular pattern and will not decrease smoothly with genetic map distance (Supplementary Figure 4). The focus of a LD analysis and the deduced threshold determining the QTL detection power should therefore not be limited to adjacent markers only. We found that the average of the highest LD of each marker with any other marker was higher with a r^2 value of 0.32. This represents an estimate of the highest r^2 that can be expected on average for any QTL position with one of the markers in the data set. This is considerably higher than the r^2 between adjacent markers and appears to represent a better estimate of the LD present in association mapping data sets determining the QTL detection power.

The observed LD has implications for the interpretation of our results as the applied marker density will only facilitate the detection of QTL with strong or medium effects whereas QTL with smaller effect size will remain undetected unless by chance they are in close proximity to a marker. The positive aspect of this finding is that on average LD appears to decay rapidly in this population which enables a high-mapping resolution. To achieve this and to detect also QTL with smaller effect size, the marker density must be tremendously increased. To obtain a desired r^2 threshold between adjacent markers of 0.8, as currently practiced in human genetics (Barett and Cardon 2006), a much higher marker density must be applied. With the availability of the full genome sequence of sugar beet, however, this goal can realistically be achieved in the near future.

The analysis of the extent of LD along the chromosomes indicated that there are chromosomal regions with a LD

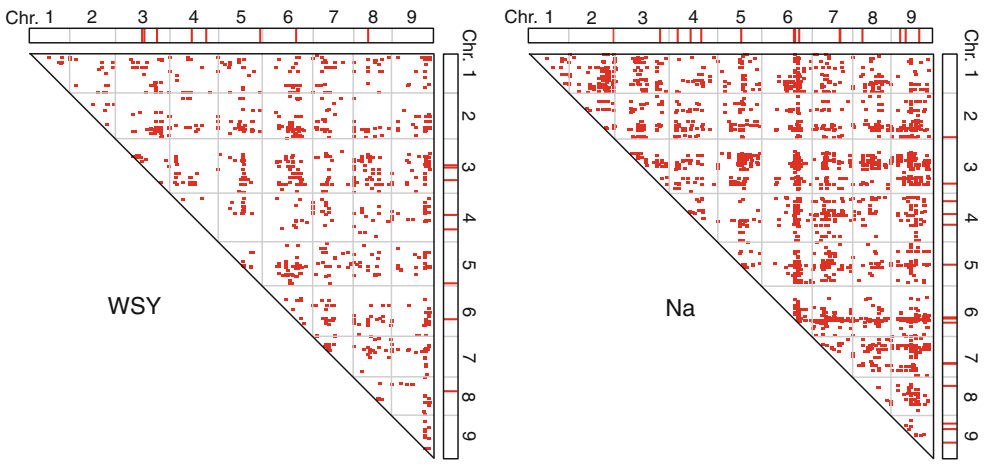

Fig. 6 Results from the genome scan for main effect QTL with the model including the *K* matrix (*above* and *right side*) and epistatic QTL for white sugar yield (WSY, t ha^{-1}) and sodium (Na, mM). *Thick lines* and *points* indicate significant QTL (FDR <0.20)

extent higher than average (Supplementary Figure 4). This is in part due to a higher marker density in these regions, but also reflects the history of the population, for example regions where QTL for introgressed traits are located and the consequences of co-selection of favorable gene combinations. The selection pressure in breeding populations will make it unlikely that such combinations are broken but rather are maintained for many generations.

In summary, excluding the few chromosomal regions which were not sufficiently covered by markers, the overall genome coverage and the observed LD enable association mapping in this population with the restriction that only QTL with large or medium effects can be identified.

Comparison of models for association mapping

Correction for the confounding effects of population structure present in plant populations is essential for association mapping if the population structure is associated with the trait (Zhao et al. 2007). To reduce the probability to detect false positive marker-trait associations, a correction for population stratification (*P* matrix, Price et al. 2006) and familial relatedness (*K* matrix; Yu et al. 2006) have been suggested. The plot of observed versus expected *P* values revealed that a simple model without any correction for population structure was not suited for the analysis of this data set (Supplementary Figure 6). Recent studies have shown that correction for familial relatedness is sufficient in populations without a major population structure (Reif et al. 2011a; Würschum et al. 2011). In this data set it was, however, not obvious whether the kinship matrix alone would be sufficient to correct for the present population structure, or if an additional correction by the inclusion of the *P* matrix was necessary. As the by default inclusion of the first ten PCs appears somewhat arbitrary, we tested two approaches including PCs based on either their significance in the model or on their association with the trait (KP_{Reml} and KP_{Cor}). The rationale for these approaches was that either as few PCs as possible should be included to avoid overcorrection (KP_{Reml}) or that all PCs associated with the trait must be incorporated to minimize the detection of false positives (KP_{Cor}). The scan for the optimum identity-by-state probability (*T* value) revealed an identical *T* value estimate for all three models. We then performed a full genome scan for main effect QTL with these three models and found that in terms of model stability (convergence of models as shown in Table 2), number of detected QTL and the explained genotypic variance explained by these QTL, the model including only the kinship matrix performed best. We speculate that, even though the PCoA revealed the presence of a population structure in this data set, this is sufficiently controlled by the kinship matrix making the incorporation of a *P* matrix nonessential. Both, the *K* matrix and the *P* matrix are based on marker data and their simultaneous use may result in an overcorrection reducing the QTL detection power. As the *P* matrix appeared to be not required for the control of population structure and as these models had more problems with model convergence, we did the further analyses with the *K* model.

Detection of main effect QTL

In the genome-wide scan we detected main effect QTL for all three yield-related traits (WSY, SC, RY) and for the three physiological traits (Na, K, N). We performed a literature review for QTL reported for these traits in linkage mapping studies (Weber et al. 1999, Weber et al. 2000, Schneider et al. 2002) and one joint linkage association mapping study (Reif et al. 2010) and compared the published QTL with those detected in this study. As no common integrated map exists for sugar beet, such a comparison must be interpreted cautiously as no precise comparison of the QTL positions can be done due to the different maps applied in the different studies. Nevertheless, the chromosomal regions harboring the QTL should in

most cases be comparable across studies and many of the QTL detected in this study collocated with previously described QTL. Interestingly, we also detected a number of QTL which have not been described so far, some of which being major QTL with an explained genotypic variance >5% (Table 3). With the exception of α-amino nitrogen the total genotypic variance explained by the detected QTL was high, explaining up to 73% (WSY, Table 2).

Due to the inherent population structure of association mapping populations, cross validation approaches as done in linkage mapping (Utz et al. 2000), are more problematic in association mapping approaches (Müller et al. 2010). Nevertheless, to obtain a first estimate of the prediction power of the detected QTL, we performed a validation approach in which the QTL detected in the full data set were used to estimate their effects in an estimation set (80% of the lines). These effects were then used for the prediction in the test set (remaining 20% of the lines). We observed that a large proportion of the explained genotypic variance from the full data set was also observed in the validation approach (Fig. 5). Cross validation in association mapping certainly warrants further research, but our results suggest, that the estimated effects of the QTL may not be as strongly overestimated as in linkage mapping. This shows that the association mapping approach used in this study is a powerful tool to detect previously unknown QTL. The advantage of this approach is that many lines are included representing the full genotypic variance, increasing the probability that QTL alleles are present in the population under consideration.

Genetic architecture of physiological and agronomic traits

Association mapping approaches have recently shown that epistasis contributes to the expression of some traits in wheat and sugar beet (Miedaner et al. 2010; Reif et al. 2011a; Würschum et al. 2011). We therefore also performed a full two-dimensional genome scan and detected epistatic QTL for all six traits. The large number of significant QTL detected in this study compared to a previous study in sugar beet (Würschum et al. 2011) may be attributed to the applied significance threshold (FDR <0.20) which was suited for the detection of main effects, but appears too liberal for the detection of epistatic QTL. Nevertheless, the plot of the detected significant epistatic QTL revealed a genome-wide landscape of epistatic interactions (Fig. 5). Interestingly, there appear to be interaction hotspots, loci which are involved in a large number of interactions with other loci across the whole genome. These epistatic key regulators were sometimes found in regions where a main effect was detected (i.e. Na chromosome 6), but also on chromosomes without a main effect (i.e. WSY chromosome 9). A recent study on flowering time in wheat has revealed *Vrn-A1* as such a key regulator involved in many epistatic interactions though not as a main effect (Reif et al. 2011a). Our study highlights the importance of such loci which appear to be central to the genetic interaction landscape and which were found for both, agronomic and physiological traits. It is tempting to speculate that such epistatic master regulators exist in all species and for many quantitative traits and it will be interesting to identify the molecular nature of such loci.

Our results confirm that the genetic architecture of important agronomic and physiological traits in sugar beet is defined by main effect QTL and epistatic QTL resulting in complex networks that have to be considered in breeding programs. The promise of QTL mapping studies for plant breeding lies in the implementation of the detected QTL in marker-assisted breeding programs to increase selection gain and shorten the time required for the establishment of new varieties. For all traits there are promising candidate loci which explain a considerable proportion of the genotypic variance (Table 3) and which after validation could be used in knowledge-based breeding.

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. We greatly appreciate the helpful comments and suggestions of two anonymous reviewers.

References

Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. Stat Sci 24:451–471

Barett JC, Cardon LR (2006) Evaluating coverage of genome-wide association studies. Nat Genet 38:659–662

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57:289–300

Bernardo R (1993) Estimation of coefficient of coancestry using molecular markers in maize. Theor Appl Genet 85:1055–1062

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5:618–625

Cockram J, White J, Zuluaga DL, Smith D, Comadran J et al (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. PNAS 107: 21611–21616

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED et al (2010) The genetic landscape of a cell. Science 327:425–431

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

Ersoz ES, Yu J, Buckler ES (2007) Application of linkage disequilibrium and association mapping in crop plants. Genomics-assisted crop improvement: genomics approaches and platforms, vol 1, pp 97–119

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

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

He X, Qian W, Wang Z, Li Y, Zhang J (2010) Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. Nat Genet 42:272–276

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

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

Li L, Paulo MJ, van Eeuwijk F, Gebhardt C (2010) Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. Theor Appl Genet 121:1303–1310

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

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

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

Miedaner T, Würschum T, Maurer HP, Korzun V, Ebmeyer E, Reif JC (2010) Association mapping for Fusarium head blight resistance in soft European winter wheat. Mol Breed. doi: 10.1007/011032-010-9516-Z

Müller BU, Stich B, Piepho HP (2010) A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. Heredity 106:825–831

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Reif JC, Liu W, Gowda M, Maurer HP, Möhring J et al (2010) Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L.) investigated with joint linkage association mapping. Theor Appl Genet 8:1489–1495

Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T, Würschum T (2011a) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theor Appl Genet. doi:10.1007/s00122-011-1583-y

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V et al (2011b) Association mapping for quality traits in soft winter wheat. Theor Appl Genet 122:961–970

Schneider K, Schäfer-Pregl R, Borchardt C, 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

St Onge RP, Mani R, Oh J, Proctor M, Fung E et al (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet 39:199–206

Stich B, Melchinger AE, Heckenberger M, Möhring J, Schechert A et al (2008) 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

Tong AH, Lesage G, Bader GD, Ding H, Xu H et al (2004) Global mapping of the yeast genetic interaction network. Science 303:808–813

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

Van Inghelandt D et al (2010) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123:11–20

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

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

Würschum T, Maurer HP, Schulz B, Möhring J, Reif JC (2011) Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. Theor Appl Genet 123:109–118

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C et al (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4